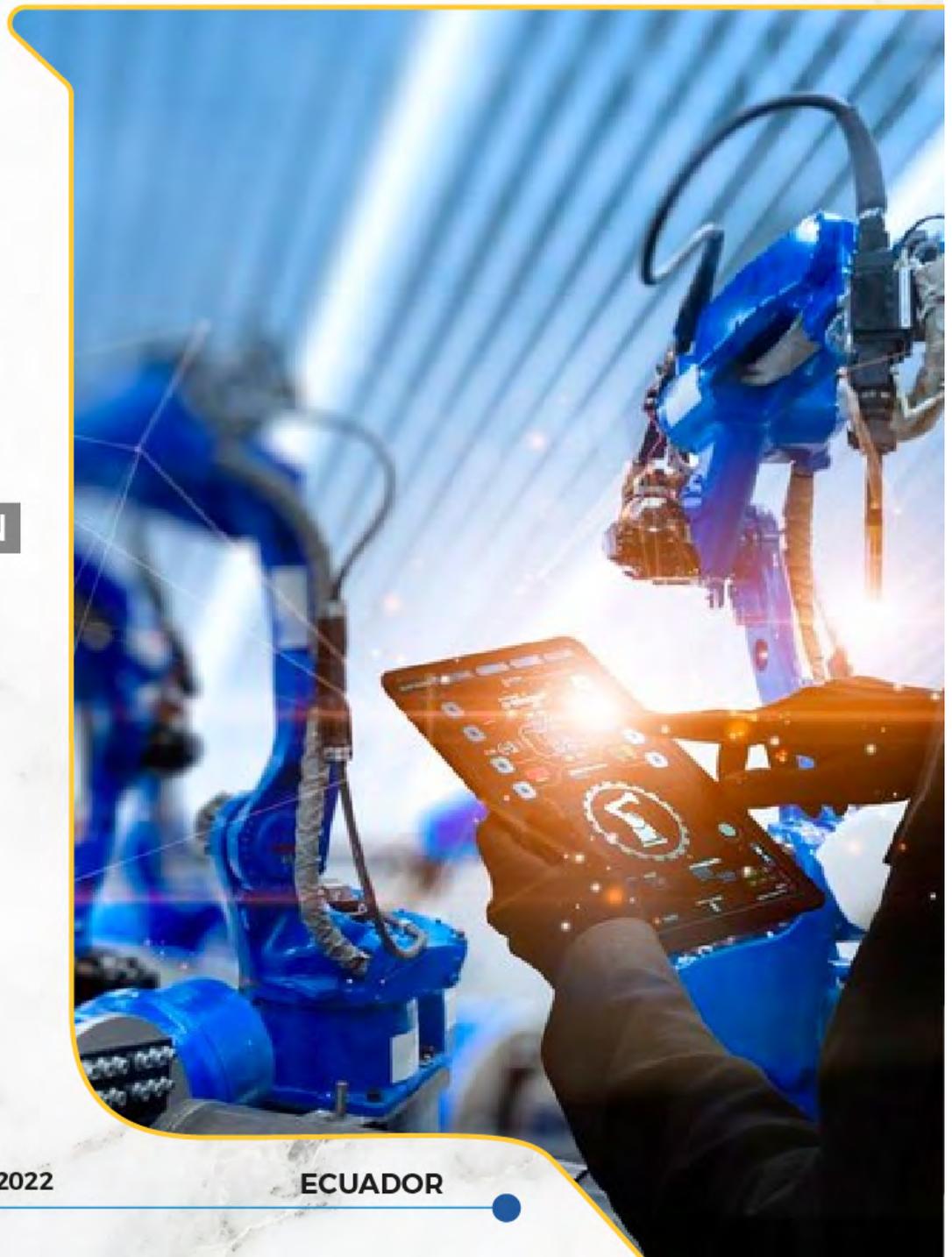


e-ISSN 2550-6730



UNIVERSIDAD
TÉCNICA DE
MANABÍ
Fundada en 1952

6
VOLUMEN
Núm. 1



ENERO - JUNIO 2022

ECUADOR

INDICE

1-11

Jhonny Xavier Chavez Calderón, Walter Daniel Zambrano Romero, Emilio Antonio Cedeño Palma, Dannyll Michellc Zambrano Zambrano, Gabriel Agustín Coterá Ramírez

El Frontend. Diseño web adaptativo y diseño web responsivo para el desarrollo de aplicaciones web

The Frontend. Adaptive Web Design and Responsive Web Design for Web Application Development

12-23

Juan-Carlos Zambrano, Patricia Quiroz-Palma, Alex Santamaría-Philco, Willian Zamora

Covid-19 en Ecuador: Aplicación de minería de datos

Covid-19 in Ecuador: Application of data mining

24-30

Teófilo Abdul Vera Peralta, Andrea Katherine Alcívar Cedeño, Tatiana Elizabeth Cobeña Macías

Revisión de los mecanismos de QoS en las instituciones públicas del Ecuador

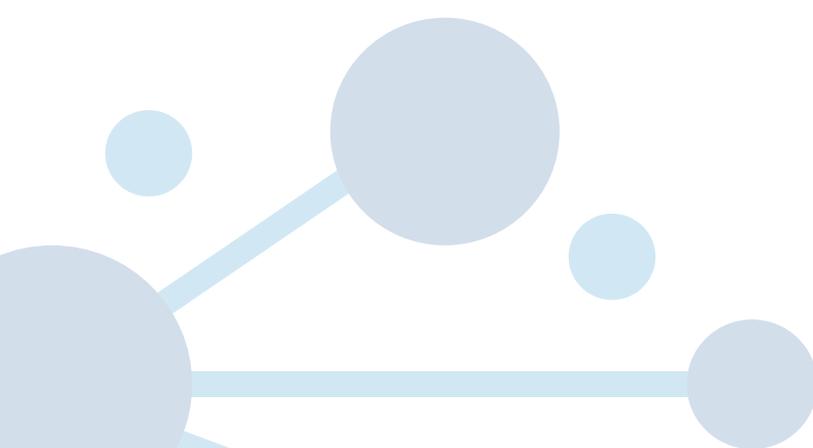
Review of QoS mechanisms in Public Institutions in Ecuador

31-37

Lizbeth Pacheco-Guevara, Ruth Reátegui, Priscila Valdiviezo-Díaz

Identificación de temas del blog de noticias en idioma español

Topic identification from news blog in Spanish language



38-44

Maria Jose Guerrero Garcia, Jorge Rodas-Silva

Análisis comparativo de metodologías y herramientas tecnológicas para procesos de Business Intelligence orientado a la toma de decisiones

Comparative analysis of methodologies and technological tools for Business Intelligence processes aimed at decision making

45-51

Ingrid Yadira Álava Chávez, Dolores Esperanza Muñoz Verduga

La brecha digital en las instituciones educativas fiscales de Manta: Situación actual, necesidades y desafíos

The Digital Divide in Fiscal Educational Institutions of Manta: Current Situation, Needs and Challenges

52-59

Manuel Eduardo Vinces Mendieta, Darwin Patricio Loor Zamora, Marelly del Rosario Cruz Felipe

Análisis de rendimiento en red de sensores inalámbrica con distintas topologías y cantidades de nodos

Wireless sensor network performance analysis with different topologies and number of nodes

60-70

María Angélica Velepucha Sánchez, Jéssica Morales Carrillo, Marco Fernando Pazmiño Campuzano

Análisis y evaluación de riesgos aplicados a la seguridad de la información bajo la norma ISO

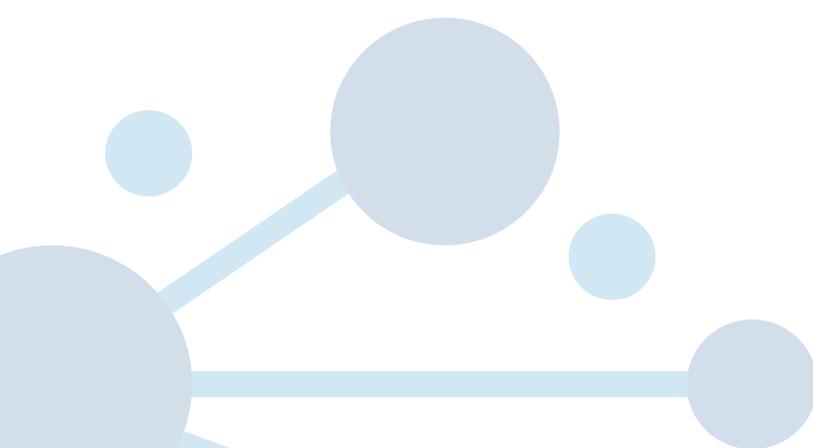
Analysis and risk assessment applied to information security under the ISO/IEC 27002 standard: Case study Bravel distributor

71-77

Vladimir Javier Rojano Guamaní, Ginger Lissbeth Jaramillo Tenezaca, Karla Susana Cantuña Flores, Gustavo Adolfo Sandoval Rulova, José María Bengochea Guevara

Desarrollo de una aplicación móvil para el monitoreo de la fenometría vegetativa del maíz amarillo, en la sierra central ecuatoriana

Development of a mobile application for the monitoring of the vegetative phenomenon of yellow corn, in the central Ecuadorian highlands





El Frontend. Diseño Web Adaptativo y Diseño Web Responsivo para el desarrollo de aplicaciones web

The Frontend. Adaptive Web Design and Responsive Web Design for Web Application Development

Autores

- ✉ * Jhonny Xavier Chávez Calderón
- ✉ Walter Daniel Zambrano Romero
- ✉ Emilio Antonio Cedeño Palma
- ✉ Dannyll Michelle Zambrano Zambrano
- ✉ Gabriel Agustín Cotera Ramírez



Resumen

El desarrollo web ha evolucionado a gran escala, el presente artículo realiza una comparativa entre dos importantes marcos de diseño que son indispensables para el desarrollo de interfaces web, como son el marco responsivo y adaptativo, destacando sus principales ventajas y diferencias a través del método experimental, propiamente enfocado en el desarrollo de un portal web, que se desarrollará para mostrar información sobre la cantidad de investigaciones científicas que publican las Facultades de la Universidad Técnica de Manabí, para el estudio comparativo se usó la herramienta Google Page Speed Insights, donde se aplicaron métricas de evaluación impuestas por W3C, con respecto al rendimiento y optimización. Para garantizar la calidad de la investigación se aplicó el método Delphi, en un cuestionario dirigido a 50 Docentes de diferentes Universidades Ecuatorianas, con la finalidad de conocer el marco de diseño más adecuado al momento de desarrollar y producir sistemas web.

Palabras clave: Diseño web responsivo, Diseño web adaptativo, Desarrollo web, Frontend, Métricas.

¹Universidad Técnica de Manabí,
Portoviejo – Ecuador.

* Autor para correspondencia

Comó citar el artículo: Chávez Calderón, J.X., Zambrano Romero, W.D., Cedeño Palma, E.A., Zambrano Zambrano, D.M. & Cotera Ramírez, G.A. (2022). El Frontend: Diseño web adaptativo y diseño web responsivo para el desarrollo de aplicaciones web. *Informática y Sistemas: Revista de Tecnologías de la Informática y las Comunicaciones*, 6(1), 1-11. DOI: <https://doi.org/10.33936/isrtic.v6i1.4625>

Enviado: 26/04/2022
Aceptado: 04/06/2022
Publicado: 09/07/2022

Abstract

Web development has evolved on a large scale, this article makes a comparison between two important design frameworks that are essential for the development of web interfaces, such as the responsive and adaptive framework, highlighting its main advantages and differences through the experimental method, properly focused on the development of a web portal, The Google Page Speed Insights tool was used for the comparative study, where evaluation metrics imposed by W3C were applied, with respect to performance and optimization. To guarantee the quality of the research, the Delphi method was applied in a questionnaire directed to 50 teachers from different Ecuadorian Universities, with the purpose of knowing the most adequate design framework at the moment of developing and producing web systems.

Keywords: Responsive web design, Adaptive web design, Web development, Frontend, Metrics.



1. Introducción

En los últimos años, la producción de sistemas informáticos para móviles y tabletas se ha visto agigantado en diversas áreas. En este artículo se indaga sobre los mecanismos que existen para desarrollar interfaces gráficas para sistemas en línea que constantemente buscan métodos innovadores con soluciones pertinentes para presentar sus contenidos en dispositivos de diversas escalas, siempre poniendo la mira en el usuario final, para de esa manera captar mayor volumen de usabilidad en Internet.

Los sistemas web han evolucionado de manera acelerada, por lo que han reducido la producción de los sistemas de escritorio, las grandes empresas han migrado sus sistemas locales a la web, ya que los servicios en línea brindan a los usuarios una mayor disponibilidad y rapidez para obtener información, es ahí donde entra en juego el diseño web como parte del desarrollo, es muy importante porque es parte de la experiencia que tendrá el cliente al navegar por un sitio (Manso Guerra, 2016).

El presente tema es innovador desde el punto de vista de producción de sistemas informáticos orientados a la web, ya que es parte del proceso de mantenimiento de software. Para esto se analizan la usabilidad del Diseño Web Adaptativo (AWD) y del Diseño Web Responsivo (RWD), que son considerados como los dos grandes mecanismos para el desarrollo de interfaces gráficas, basándose en resultados obtenidos a través de experimentación de autores, se manifiesta que las aplicaciones son altamente disponibles y accesibles para los usuarios, cuando se usa estos tipos de diseño, requiriendo sólo un dispositivo que tenga un navegador web para acceder (Adenan, 2020).

Adicional a lo anterior, este tema es relevante porque busca conocer el marco de diseño más ventajoso a la hora de desarrollar interfaces gráficas, aplicando para ello un método científico de evaluación, una parte importante de este tema se debe a que el Frontend es actualmente un tema interesante y que ha venido evolucionando a través de los Frameworks, aunque es un tema que ya ha sido explorado, en ocasiones los investigadores desconocen el origen para el apareamiento de nuevos lenguajes gráficos, como por ejemplo: Bootstrap, Bulma CSS, Pure CSS, Tailwind CSS, entre otros.

La finalidad de la presente investigación es demostrar que marco de diseño se adapta a mejor a las buenas prácticas, usabilidad y experiencia del usuario

2. Materiales y Métodos

La metodología de esta investigación es de carácter cuantitativo, bibliográfico y experimental. Se ha realizado una revisión bibliográfica en bases de datos de alto nivel, como IEEE, Scopus, ACM y diversos congresos científicos donde se ha extraído

información sobre la usabilidad que ofrecen AWD y RWD.

Como parte del método científico aplicado en el presente proyecto, se utilizará la experimentación, la cual consistirá en el desarrollo de un portal web aplicando RWD y AWD, el sitio será un prototipo de portal que en el futuro tomará un valor significativo para la Universidad Técnica de Manabí (UTM), debido a que será un espacio dedicado a la información de investigaciones científicas publicadas por las Facultades de la UTM. En la experimentación se harán comparaciones según métricas de evaluación impuestas por W3C a través de la herramienta de Google “Page Speed Insights”, los aspectos considerados serán: carga, velocidad de navegación, presentación, volumen de código, entre otros.

Otro método utilizado para garantizar la calidad de la investigación es el Delphi o consulta a expertos, que es utilizado para conocer cuál es el marco de diseño recomendado para desarrollar interfaces web. Los expertos son 100 docentes especializados en el área de Ingeniería de Software, los cuales son pertenecientes a diferentes Universidades Ecuatorianas. De los 100 docentes, se procedió a trabajar con una muestra de tipo finita sobre la población total, aplicando la siguiente fórmula:

$$n = \frac{Z^2 * N * P * Q}{E^2 (N-1) + Z^2 * P * Q} \quad n = \frac{95^2 * 100 * 0.5 * 0.5}{10^2 (100-1) + 95^2 * 0.5 * 0.5}$$
$$n = \frac{225.625}{4.731,25}$$
$$n = 49,7 \rightarrow 50 \text{ docentes}$$

Figura 1. Cálculo de población para el estudio de la presente investigación.

Fuente: Los autores

Datos

Z = Coeficiente de confiabilidad.

P = Proporción de éxito.

Q = Proporción de fracaso (Q = 1-P).

E = Margen de error.

Las herramientas utilizadas en esta investigación son: Google Form para desarrollar los formularios (método Delphi), Visual Studio Code como IDE para la codificación de las interfaces, en la parte AWD se utilizará CSS, HTML y JavaScript, y se modelarán diferentes aspectos de presentación para los tamaños generales 360x640, 600 x 1024, 800x1280, y finalmente 1366x768, en la

parte RWD se utilizará el framework Bootstrap en su versión 5.0.

Para un resultado eficiente de la experimentación se formularán cuatro hipótesis que se deberán evaluar al final de la presente investigación, estas son:

- H0: El diseño web adaptativo (AWD) posee mejores utilidades para el desarrollo de interfaces web que infieren en la calidad de navegación del usuario según expertos.
- H1: El diseño web responsivo (RWD) posee mejores utilidades para el desarrollo de interfaces web que infieren en la calidad de navegación del usuario según expertos.
- H2: El diseño web adaptativo (AWD) posee una mejor valoración según Page Speed Insights.
- H3: El diseño web responsivo (RWD) posee una mejor valoración según Page Speed Insights.

2.1 El Frontend como carta de presentación de la web

Dentro del mundo informático, el desarrollo web, se ha convertido en una de las áreas con mayor demanda laboral para un Ingeniero en Sistemas (Pérez Ibarra, 2021). Dentro de esta área informática, se encuentran dos vertientes muy importantes para los desarrolladores, el elegante FrontEnd y el robusto Backend.

El Frontend, es un componente visual dentro del desarrollo web, está compuesto por todas aquellas tecnologías que se ejecutan del lado del cliente, específicamente visualizadas en el navegador web, de manera general, los lenguajes de diseño pioneros en el desarrollo web son: HTML, CSS y JavaScript, cuya implementación se da, en el desarrollo de interfaces gráficas adecuadas para los usuarios (Valdivia-Caballero, 2016).

Un profesional encargado del FrontEnd, debe trabajar regularmente con estas tres tecnologías, aunque es elemental que conozca la manera de trabajar del BackEnd (lado servidor), debido a que es importante para el diseño del maquetado en HTML y su debida estilización en CSS.

2.1.1 Principales objetivos del Frontend

Busca implementar la responsividad de las interfaces, para que una aplicación web sea multiplataforma (Alonso Vega, 2013). Permite la creación de API's para el consumo de datos de manera fácil y a su vez mejora la experiencia del usuario (Berbel Marín, 2018). Presentar la información, de manera dinámica y estilizada para los diversos tipos de usuarios (Deshmukh, 2019, March).

2.1.2 Evolución acelerada del Frontend

Las tendencias en desarrollo web, a través de la aparición de frameworks han marcado una evolución muy acelerada en el componente Frontend, debido a que todo empezó por los lenguajes de diseño clásicos: HTML, CSS y JavaScript, que

han sido la base para la construcción de nuevas mejoras en el diseño web (Zelaya, 2020). En la actualidad se puede encontrar diversos marcos visuales, que mejoran la calidad de las interfaces tomando en consideración diversos aspectos de las aplicaciones web, tales como: formularios, botones, tablas, banners, fondos, colores, alertas, cargas, tipografías, adaptabilidad de interfaces a móviles, efectos e interactividad con el usuario, entre otros.

2.1.3 Frameworks destinados al Frontend (Responsividad de interfaces web)

Los frameworks que en la actualidad son más usados por los desarrolladores web (Zelaya, 2020), son:

- Bootstrap.
- Bulma.
- Skeleton.
- W3CSS.
- Angular JS.
- React JS.
- Vue JS.
- Vuetify, entre otros.

2.2 Aproximación a los marcos de diseño web

Para la creación de interfaces web, es necesario en primera instancia conocer el alcance que tendrá la página o portal que se vaya a realizar, en caso de que el usuario requiera que sea ejecutado en diversos tamaños de pantalla, se podrá abordar en dos temáticas, el AWD y RWD, que son marcos para el diseño de interfaces, ambas con características y maneras diferentes de implementación pero que tienen el mismo fin, brindar una experiencia de navegación óptima para el usuario final. Bajo esta premisa se plantea realizar una comparativa entre ambos marcos para el diseño de interfaces.

2.2.1 Enfoque del diseño web adaptativo (AWD)

AWD es un marco de diseño relativamente nuevo para la mejora progresiva de sitios web, enfocado a la accesibilidad, usando hojas de estilo con tecnología de scripting para diseñar diversas páginas web que se adaptan según el dispositivo que esté usando el usuario en tiempo real (Yousaf, 2018).

Se necesitan los siguientes recursos para realizar un diseño adaptativo:

- Diseño fluido con rejillas flexibles y Media Queries.
- Imágenes, objetos, vídeos o medios similares flexibles.
- Fuentes tipográficas con valores relativos.

Enfoque del diseño web adaptativo (AWD) La técnica



utilizada en el diseño web adaptativo, se basa en ajustar el contenido de los sitios web al ancho del área de visualización de cada dispositivo personalizado. Presenta algunas características, como: Diseño fluido basado en porcentajes y píxeles.

```
estilos.css > ...  
1 .contenedor {  
2   max-width: 1024px;  
3   background-color: aquamarine;  
4 }
```

Figura 2. Anchura mediante píxeles en CSS
Fuente: Los autores

El uso de Media Queries Para configurar la altura, la anchura y la resolución del dispositivo móvil en el que se quiere ejecutar. Por ejemplo, si desea consultar el sitio en una resolución móvil de 450px. Cabe destacar que en el archivo estilo_movil.css se ha colocado algunos estilos sencillos a modo de demostración.

Cabe destacar que en el archivo estilo_movil.css se ha colocado

```
<link rel="stylesheet" href="estilo_movil.css" media="only screen and (max-width:450px)">
```

Figura 3. Condición de las consultas de medios para configurar el ancho de la pantalla
Fuente: Los autores

algunos estilos sencillos a modo de demostración

```
estilo_movil.css > ...  
1 .contenedor {  
2   max-width: 40%;  
3   background-color: rgb(255, 72, 0);  
4   color: white;  
5 }
```

Figura 4. Estilos aplicados al div contenedor desde el archivo CSS
Fuente: Los autores

En esta sección aún no se hace ningún cambio porque no llega al límite máximo de 450px. Se presenta a continuación



Figura 5. Presentación de la interfaz en el navegador a escala 468x522. El recuadro rojo identifica la anchura y el azul la altura.
Fuente: Los autores

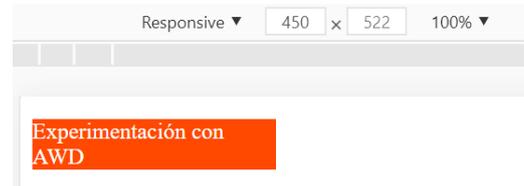


Figura 6. Presentación de la interfaz en el navegador cuando es inferior a 450px.
Fuente: Los autores

Si apreciamos en el caso anterior, el cambio se realiza cuando es menor a 450px, se espera a través de este ejemplo muy sencillo resaltar la funcionalidad de AWD, obviamente se complica un poco cuando se comienzan a colocar dentro del HTML más etiquetas de recursos dependiendo de la página a desarrollar.

En resumidos términos el marco AWD se maneja a través de media queries para la presentación de interfaces según el programador lo haya establecido, en el caso de la tienda de Amazon, el portal web de ellos funciona bajo el marco AWD, lo que quiere decir que existen 3 tipos de interfaces creadas específicamente por el desarrollador para dispositivos como ordenadores, tabletas y móviles. La funcionalidad AWD se da al momento que un usuario accede a la tienda, y el servidor detecta la pantalla del dispositivo que está usando el usuario, en tal caso lo redirecciona a la presentación que se ajusta de manera eficiente a su equipo.

Ventajas del diseño web adaptativo El desarrollo de interfaces web mediante el diseño adaptativo es ventajoso, y contribuye a los siguientes beneficios (Labrada, 2013):

- Personalización de interfaces: Con AWD, los desarrolladores podrán hacer una interfaz diferente para cada resolución, en la que se podrán colocar los recursos óptimos para cada pantalla del dispositivo.
- Adaptación de la velocidad de la interfaz: Se notará la fluidez, la carga de recursos será más rápida, debido a que ya está especificado para qué tipo de dispositivo, eliminando recursos que ralentizan la ejecución, por ejemplo: en dispositivos móviles no se mostrarán imágenes masivas como en un dispositivo de escritorio.

Desventajas del diseño web adaptativo La desventaja de aplicar el AWD es que hay que ajustar correctamente las escalas de resolución para cada dispositivo, hay que hacer algunas presentaciones de la misma plantilla, nunca se sabe en qué dispositivo navegará el cliente, por lo que las páginas deben hacerse según las resoluciones generales emitidas por el W3C. Además, se genera una masiva codificación para cada interfaz (Labrada, 2013).

2.2.2 Enfoque del diseño web responsivo (RWD)

El Internet ha entrado en su versión Web 3.0, a su vez ha incorporado el desarrollo de la tecnología móvil, debido a que cada vez se utilizan más terminales móviles inteligentes (Li, 2019).

Es allí donde aparece el RWD que es un marco de diseño que resuelve el conflicto de algunos sitios al no poder adaptarse a los teléfonos móviles u otras interfaces de dispositivos. Por ello, el RWD se ha convertido en un tema de investigación de gran interés para los desarrolladores y diseñadores de experiencia de usuario (UX), el objetivo del marco de diseño RWD, es que las páginas web tradicionales puedan mejorar la experiencia del usuario.

El RWD incorpora varias tecnologías para poder cumplir con el desarrollo de interfaces (Li, 2019):

- HTML5
- CSS3
- Rejillas de fluidos
- Imágenes flexibles

Características del diseño web responsivo A su vez posee las siguientes características:

- Ajuste automático del tamaño de una imagen u objeto a la pantalla del dispositivo que se esté usando.
- Uso de paquetes de media queries minimizados.
- Automatización para el proceso de la grid automática, a través de Frameworks, tales como: Bootstrap, Materialize, Foundation, entre otros.

La esencia del RWD es permitir el acceso al contenido web para todos los dispositivos existentes a través de la capacidad de respuesta de las interfaces, hoy en día existen varios tipos de dispositivos que presentan diferentes dimensiones de pantalla, de ahí la necesidad de adaptar el diseño del contenido web (Pérez Ibarra, 2021).

Los dispositivos móviles de pantalla pequeña están ganando popularidad a medida que avanza el tiempo. También se puede hablar de las consolas de juegos y las SmartTV, que son cada vez más comunes entre los consumidores, ya que permiten navegar por Internet en dispositivos de alta resolución con comodidad. La capacidad de respuesta responde a los puntos de ruptura establecidos por el W3C en 2014, cuando comenzó el movimiento de la Web 3.0 (Otto, 2014).

Dentro de la web responsiva hoy en día se pueden encontrar algunos frameworks que facilitan la tarea de desarrollo frontend, algunos de ellos son: Bootstrap, Materialize, Skeleton, Angular, React, entre otros, pero en esta investigación se utilizará Bootstrap para hacer la comparación contra AWD.

Bootstrap como marco de trabajo para el desarrollo de interfaces responsivas Bootstrap (BS) es un framework de diseño que permite desarrollar sitios web adaptativos y modernos, la idea básica de BS es utilizar las características empaquetadas

Tabla 1. Resolución de la pantalla de los dispositivos que acceden a la web.

Fuente: Tomado de (Peltomäki, 2014)

Tipo de resoluciones de pantalla (px)	Popularidad (%)
1366x768	20.74
1024x768	9.61
768x1024	5.26
320x480	2.90

de CSS para trabajar la adaptabilidad a través de puntos de ruptura y mediante el uso de una retícula de doce columnas que ofrece BS (Tapani, 2014). Este artículo no pretende profundizar en Bootstrap, sólo se considera dentro del RWD como uno de los frameworks más utilizados por los desarrolladores web.

Ventajas del diseño web responsivo Bootstrap tiene magníficos beneficios para el desarrollo Frontend, ya que se acopla visualmente a las necesidades y experiencias del usuario (Rodríguez, 2018). Estos son:

- Permite la inserción de elementos inline: iconos, menús desplegables, formularios, entre otros, utilizando sintaxis cortas que se pueden encontrar en su documentación.
- El diseño será adaptable, no se verá afectado por la escala o resolución del dispositivo móvil.
- El diseño de columnas es más rápido y fácil de configurar.
- Permite el uso del lenguaje de preprocesamiento Less.

Puntos de interrupción ofrecidos por Bootstrap Bootstrap permite una adaptabilidad dinámica a través de la inserción de filas y columnas, lo cual es una ventaja importante que presenta ya que el desarrollador, conociendo la sintaxis básica de sus componentes, podrá crear rápidamente formularios, botones, galerías, entre otros elementos de las páginas web (W3schools., 2018).

3. Resultados y Discusión

Antes de utilizar las métricas, se desarrollará como ejemplo un prototipo de portal web donde se evidencie información de cantidades de artículos publicados por Facultad dentro de la Universidad Técnica de Manabí, el portal se desarrolla en base a los marcos de diseño AWD y RWD. Una vez desarrollados, ambos sitios serán subidos a un servidor de prueba donde a través de la herramienta PageSpeed Insights (se medirá el tiempo de carga y la velocidad de navegación) y a su vez también con el módulo de Google Test Mobile (se mide el porcentaje de adaptabilidad).

Implementación del sitio con AWD y medición de los problemas de rendimiento



Tabla 2. Aceptación de puntos para ruptura en píxeles.
Fuente: Los autores.

Categoría	Anchura	Dirigido a
.col - lg- *	> 1200	Escritorios grandes.
.col - md- *	> 992	Escritorios pequeños.
.col - sm- *	> 768	Tabletas y pantallas pequeñas.
.col - xs- *	< 768	Teléfonos inteligentes.

Resultado e interpretación: El sitio web en modo escritorio tiene un tiempo de renderizado de imágenes y texto de 0,8 s. La velocidad de visualización del contenido de cada página es de 1,0 s. El renderizado de la imagen gif principal es de 0,9 s. El tiempo que tarda la página en ser totalmente interactiva es de 0,8 s. La página no experimenta caídas en la carga. Y finalmente los movimientos de los elementos en la carga es de 0,077 s

Resultado e interpretación: El sitio web en resolución Tablet presenta una lentitud en el renderizado de la imagen y el texto debido a que supera los 2,4 segundos. La velocidad con la que se observa el contenido de cada página es de 2,5 segundos. El renderizado de la imagen gif principal es de 2,4 segundos. El tiempo que tarda la página en ser totalmente interactiva es de 2,4 segundos. La página no experimenta caídas en la carga. Y finalmente los movimientos de los elementos al cargar es de 0,139 s

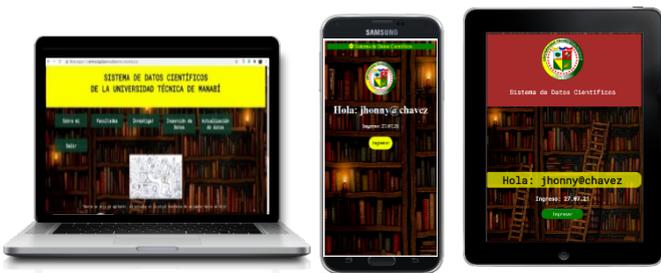


Figura 7. Ordenador portátil (1366x768px), Tableta (768x1024px), Teléfono inteligente (360x640px).
Fuente: Los autores.

Resultado e interpretación: El sitio web para dispositivos móviles con AWD tiene un renderizado de imágenes y textos de 1,6 s, considerado como promedio. La velocidad con la que se observa el contenido de cada página es de 2,4 segundos. El renderizado de la imagen gif principal es de 1,6 segundos. El tiempo que tarda la página en ser totalmente interactiva es de 1,6 s. La página experimenta mucho bloqueo en la carga, con un tiempo total de 0 ms. Y finalmente los movimientos de los

Tabla 3. Métricas en modo escritorio AWD con PageSpeed Insights y Google Test Mobile. (Pantalla general de 1366x768 de ordenadores portátiles y de sobremesa)
Fuente: Los autores.

Categorías	Valores	Ponderación	Adaptabilidad
Primer cuadro de contenido.	0,8 s		
Índice de velocidad.	1,0 s		
Mayor contenido de pintura.	0,9 s	98%	91%
Tiempo de interacción.	0,8 s		
Tiempo total de bloqueo.	0 ms		
Desplazamiento acumulado del trazado.	0,077 s		

Tabla 4. Métricas en modo tableta AWD con PageSpeed Insights y Google Test Mobile. (768x1024 Tablet iPad o Samsung).
Fuente: Los autores.

Categorías	Valores	Ponderación	Adaptabilidad
Primera pintura de contenido	2,4 s		
Índice de velocidad	2,5 s		
Mayor contenido de pintura	2,4 s	92%	95%
Tiempo de interacción	2,4 s		
Tiempo total de bloqueo	0 ms		
Desplazamiento acumulado del trazado	0,139 s		

elementos en la carga es de 0,046 s.

Comparativa propuesta - Sitio en relación al marco AWD

Con respecto a la renderización de imágenes y texto, mientras menor sea el tiempo de renderizado, mejora la calidad de respuesta de una interfaz, y es lo que los usuarios buscan, una experiencia agradable y rápida en la web (Gustafson, 2016). El sitio muestra una renderización de imagen y texto más eficiente

Tabla 5. Métricas en modo móvil AWD con PageSpeed Insights y Google Test Mobile. (360x640 para Samsung Galaxy S5)
 Fuente: Los autores.

Categorías	Valores	Ponderación	Adaptabilidad
Primera pintura de contenido	1,6 s		
Índice de velocidad	2,4 s		
Mayor contenido de pintura	1,6 s	99%	97%
Tiempo de interacción	1,6 s		
Tiempo total de bloqueo	0 ms		
Desplazamiento acumulado del trazado	0,046 s		

cuando es presentado en la resolución de escritorio, debido a que solo necesita 0,8 s en comparación a la resolución de tableta con 2,4 s y la de móvil con 1,6s.

Uno de los aspectos esenciales en un portal o aplicación web, es la velocidad de respuesta del contenido (texto), por lo que se recomienda siempre que no exceda de los tres segundos (Perez, 2020). El contenido aparece más rápido en la resolución de escritorio con un tiempo de 1,0 s, en comparación a la respuesta de la tableta con 2,5 s y el móvil con 2,4 s.

En una aplicación o página web, se recomienda usar siempre resoluciones optimizadas de poco tamaño, que tengan un renderizado de entre 0,3 s a 1,0 s (Veloz, 2016). Por lo tanto, la resolución de escritorio, cumple con lo establecido con un renderizado de 0,9 s, mientras que en tableta tiene 2,4 s y en el móvil 1,6 s.

La interactividad es una de las propiedades que toda aplicación web debe tener, por ende, se recomienda que su tiempo de respuesta sea menor a 1,0 s (Barba, 2015). En la implementación AWD del sitio, la resolución de escritorio tarda 0,8 s en ser interactiva, lo que refiere a un tiempo eficiente para la calidad y experiencia que el usuario tenga del sitio. No siendo así en la resolución de tableta con 2,4 s y la de móvil con 1,6 s.

Implementación del sitio con RWD y medición de los problemas de rendimiento

Resultado e interpretación: El sitio web para dispositivos móviles y tabletas en modo RWD tiene un renderizado de imágenes y textos de 2,0 s. La velocidad con la que se visualiza el contenido de cada página es de 2,6 s. El renderizado de la imagen gif principal es de 2,0s. El tiempo que tarda la página en ser totalmente interactiva es de 2,1 s. La página experimenta un bloqueo en la carga de 30 ms. Y finalmente los movimientos de los elementos en la carga es de 0 s.

Resultado e interpretación: El sitio web para la resolución de los ordenadores (portátiles y de sobremesa), tiene una muy buena representación de la imagen y el texto con 0,7 seg. La velocidad con la que se observa el contenido de cada página es de 0,9 segundos. El renderizado de la imagen gif principal es de 0,7s. El tiempo que tarda la página en ser totalmente interactiva es de 0,7s. La página experimenta un bloqueo en la carga de 3 ms. Y finalmente los movimientos de los elementos en la carga es de 0 s.

Comparativa propuesta - Sitio en relación al marco RWD

Con el marco RWD, se logra una renderización de imágenes y texto, de 2,0 s, por lo que excede al 1,0 s, recomendados para la mejora de la calidad de respuesta, los usuarios tendrían que esperar un pequeño lapso de tiempo para que aparezca la imagen, por lo que la experiencia se vería afectada (Gustafson, 2016). El sitio muestra una renderización de imagen y texto más eficiente cuando es presentado en la resolución de escritorio, debido a que solo necesita 0,7 s en comparación a la resolución de tableta y móvil con 2,0 s.



Figura 8. Ordenador Portátil (1366x768px), Tableta (768x1024px), Teléfono inteligente (360x640px).
 Fuente: Los autores.

Tabla 6. Métricas en modo móvil y Tableta RWD con PageSpeed Insights y Google Test Mobile
 Fuente: Los autores.

Categorías	Valores	Ponderación	Responsividad
Primera pintura de contenido	2,0 s		
Índice de velocidad	2,6 s		
Mayor contenido de pintura	2,0 s	96%	99%
Tiempo de interacción	2,1 s		
Tiempo total de bloqueo	30 ms		
Desplazamiento acumulado del trazado	0 s		

La velocidad de respuesta del contenido (texto), es algo muy necesario para la experiencia del usuario en un sitio web, por lo que se recomienda siempre que no exceda de los tres segundos (Perez, 2020). El texto aparece más rápido en la resolución de escritorio con un tiempo de 0,9 s, en comparación a la respuesta de la tableta y el móvil con 2,6 s. Con RWD, también se deben colocar imágenes de poco tamaño, que no afecten la navegación y el renderizado que debe estar en un rango de 0,3 s a 1,0 s (Veloz,

Tabla 7. Métricas en modo escritorio RWD con PageSpeed Insights y Google Test Mobile
 Fuente: Los autores.

Categorías	V a - P o n - Responsivi- lores deración dad
Primera pintura de contenido	0,7 s
Índice de velocidad	0,9 s
Mayor contenido de pintura	0,7 s 99% 99%
Tiempo de interacción	0,7 s
Tiempo total de bloqueo	3 ms
Desplazamiento acumulado del trazado	0 s

2016). En las pruebas con RWD, la resolución de escritorio, permite un renderizado de 0,7 s, mientras que en tableta y móvil tiene una duración de 2,0 s.

Una de las propiedades que toda aplicación web debe presentar, es la carga interactiva de todo su contenido, incluyendo tipografías, efectos, y cualquier archivo multimedia, por lo que se recomienda que su tiempo de respuesta deba ser menor a 1,0 s (Barba, 2015). Con la resolución de escritorio en RWD, la respuesta a la carga interactiva de los elementos en pantalla es de 0,7 s, lo que es considerado como un tiempo eficiente y a su vez mejora la experiencia del usuario. Por otra parte, para la resolución de tableta y móvil, la página tarda 2,1 s en ser interactiva.

Método Delphi

Resultado e interpretación: La figura 9 muestra el número de profesores que participaron en la encuesta. Cabe señalar que las universidades se eligieron acorde a la especialidad establecida (Ingeniería de Software) y que se solicitó el permiso debido a las autoridades académicas.

Resultado e interpretación: 43 docentes consideran que es preferible el marco RWD para desarrollar interfaces web.

Mientras que 7 están a favor de AWD.

1. Seleccionar las características, por las cuales usted utilizaría el marco de diseño escogido en la pregunta # 1

Resultado e interpretación: Para 41 docentes, RWD tiene mejor rendimiento en el momento de la ejecución, debido a que la mayor parte de los frameworks tienen sus componentes incrustados en la nube o también de manera minimizada, esto influye en la velocidad de ejecución de los elementos en pantalla, ya que se sincronizan de manera rápida, mientras que 9 docentes consideran que AWD rinde más, porque se diseña una interfaz para cada dispositivo, en este caso se puede simplificar algunos aspectos visuales mientras la pantalla del dispositivo sea más pequeña.

Con respecto a la originalidad de la interfaz, 34 docentes mencionan que RWD es mejor debido a la gran cantidad de elementos que ofrecen los Frameworks, y 16 afirman que AWD tiene mayor originalidad, debido a que pueden desarrollar elementos desde cero.

Tabla 8. Datos básicos del Test – Método Delphi.
 Fuente: Los autores

Objeto de estudio	Docentes informáticos con experiencia en Ingeniería de Software, pertenecientes a Universidades Ecuatorianas.
Instrumento	Cuestionario virtual de preguntas abiertas y cerradas.
# Preguntas	4



Figura 9. Número de profesores especializados en Desarrollo Web, desde distintas IES
 Fuente: Los autores

Otra característica es la reducción del código CSS, 43 docentes afirman que RWD es más simplificado, ya que se apoya en el uso de librerías y extensiones, que simplifican la redundancia de código, mientras que 7 docentes mencionan que AWD es más reducido, ya que por cada interfaz según el dispositivo en el cual

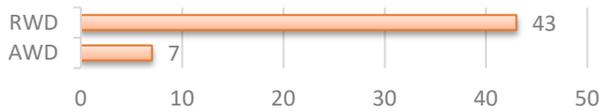


Figura 10. Marco de diseño preferido para el desarrollo de aplicaciones o sitios web

Fuente: Los autores

se vaya a ejecutar, se puede quitar código no necesario, en el caso de dispositivos móviles, se podría quitar imágenes grandes.

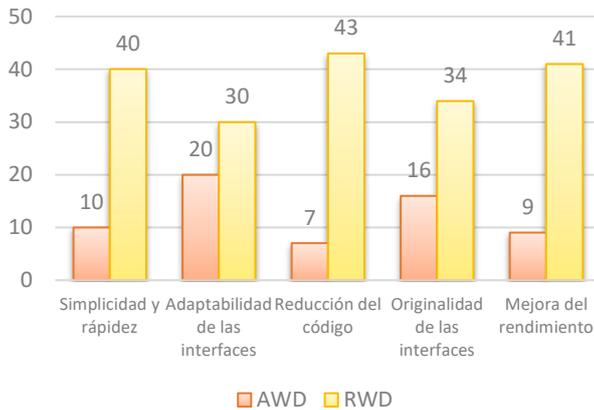


Figura 11. Razones para implementar el marco RWD o AWD.

Fuente: Los autores

Por otro lado, acorde a la adaptabilidad de las interfaces, 30 docentes consideran que RWD permite la presentación multiplataforma con tan solo diseñar una interfaz, mientras que 20 docentes mencionan que AWD permite un mejor diseño para cada dispositivo, ya que en este marco se deberá crear diferentes interfaces según la dimensión de la pantalla.

Por último, con respecto a la simplicidad y rapidez, RWD es aceptado por 40 docentes, debido a la simplicidad que los frameworks otorgan, en el caso de Bootstrap con su grid y sus múltiples elementos HTML y CSS empaquetados, por otra parte 10 docentes afirman que AWD es muy rápido a la hora de ser ejecutado en la web, debido a que al crear una interfaz para cada dimensión de pantalla de los dispositivos se puede añadir o quitar elementos dinámicos que perjudiquen la carga de la página.

Resultado e interpretación: 37 docentes afirman que el marco RWD tiene una mayor autonomía y por lo tanto se convierte en el diseño adecuado para desarrollar interfaces haciendo uso de

buenas prácticas, finalmente 13 docentes afirman que el marco AWD es más rígido y que requiere de un mayor esfuerzo a la hora de las maquetaciones, ya que se deberán desarrollar diversas plantillas.



Figura 12. Marco de diseño con más expectativa en la actualidad

Fuente: Los autores

Discusión

Evidentemente el marco RWD, es la modalidad actual para el diseño de interfaces web, debido al acelerado avance de los dispositivos móviles, obviando de cierta manera la presentación de las interfaces tradicionales que solo podían ser visualizadas desde los ordenadores.

Para concretar los resultados originados de la encuesta de expertos y de la experimentación, se puede resumir las diferencias principales que existen entre AWD y RWD (Tabla 9):

Usar uno de los dos marcos de diseño dependerá netamente del desarrollador, ambos marcos poseen ventajas y desventajas. Pero acorde a los resultados de la investigación, el marco que más se encuentra estable en la actualidad para el desarrollo ágil de interfaces es el RWD, debido a que la mayor parte de los usuarios utilizan múltiples dispositivos, y por lo tanto el RWD brinda un mayor despliegue para las aplicaciones en el aspecto de adaptabilidad. Por otra parte, las empresas para evitar gastos innecesarios en desarrollo de interfaces, recomiendan la utilización de frameworks que pertenecen al marco RWD. Algunos ejemplos: Angular, React, Bootstrap, Bulma CSS, entre otros. Las actuales aplicaciones web utilizan las tecnologías antes mencionadas con el objetivo de simplificar el desarrollo de interfaces y mejorar la calidad de experiencia que tenga el usuario usando algún servicio

4. Conclusiones

Luego de haber estudiado la eficacia del marco de diseño web adaptativo (AWD) y del diseño web responsivo (RWD) para el desarrollo de sitios web dinámicos, se concluye que el diseño RWD es más óptimo para el desarrollo de interfaces web considerando los valores originados desde la herramienta Google Speed Insights con relación a las métricas de la W3C.

Por otro parte, acorde al método Delphi los docentes afirman que RWD es y será el marco de diseño más utilizado por los desarrolladores en la actualidad y a futuro, debido a la constante actualización de recursos, su simplicidad al momento de codificar y también por la reducción de código CSS, además

Tabla 9. Diferencias entre AWD y RWD

Fuente: Los autores

Diseño Web Adaptativo	Diseño Web Responsivo
Principales diferencias	
<ul style="list-style-type: none">• La URL deberá ser distinta, dependiendo el dispositivo en el cual se esté navegando.• Involucra más tiempo de desarrollo.• Adaptación ideal a la pantalla del usuario. Esto dependerá de la experticia en el diseño que provea el desarrollador.• Las páginas web cargan más rápido debido a que están optimizadas únicamente para un solo dispositivo. Sea este un móvil o una tableta.	<ul style="list-style-type: none">• La URL será la misma para todos los dispositivos.• No involucra excesivo tiempo de desarrollo, debido a la simplicidad de código que ofrecen los Frameworks.• Las páginas cargan un poco más lento. Esto dependerá de los elementos gráficos que se estén usando en la interfaz.• Al ser un marco de diseño reciente, pueden darse incompatibilidades con ciertos navegadores antiguos.

por su rapidez para implementar interfaces a través del uso de componentes ya diseñados.

Sin menospreciar el marco AWD, algunos docentes se manifestaron, afirmando que ofrece una gran diversificación a la hora de desarrollar diferentes interfaces para cada dispositivo, aunque es un poco tradicional, tiene algunas opciones que le permiten posicionarse aún como un buen marco de diseño.

Referente a las hipótesis de investigación, se cumplen la H1, afirmando que el marco RWD posee mejores utilidades para el desarrollo de interfaces web que inferen en la calidad de navegación del usuario según expertos. A su vez la H3, que considera que el diseño web responsivo (RWD) posee una mejor valoración según Page Speed Insights.

Para trabajos futuros se pretende realizar una comparativa entre algunos frameworks pertenecientes al marco RWD, con el objetivo de conocer las principales potencialidades y popularidad en entornos empresariales.

Agradecimientos

Agradezco a mi Dios Todopoderoso por brindarme la salud e inteligencia, con las cuales pude desarrollar este artículo científico. Además, estoy muy agradecido por la ayuda de mi tutor el Ing. Walter Zambrano, quien siempre estuvo allí asesorando de manera correcta. Un agradecimiento especial a mis padres quienes siempre me han brindado su apoyo y estuvieron allí ayudándome a alcanzar la meta de poder graduarme. Gracias a la FCI por abrirme las puertas al conocimiento

Contribución de los autores

Jhonny Xavier Chávez Calderón, Emilio Antonio Cedeño Palma: Conceptualización, Metodología, Software, Análisis formal. **Jhonny Xavier Chávez Calderón, Gabriel Agustín Cotera Ramírez:** Redacción – borrador original del artículo. **Walter Daniel Zambrano Romero:** Conceptualización, Visualización, Investigación. **Dannyl Michelle Zambrano Zambrano:** Supervisión, Redacción – revisión y edición del artículo.

Conflictos de interés

Los autores declaran no tener ningún conflicto de interés.

Referencias bibliográficas

- Adenan, A. H., Abdullah, N. A. S., Hussain, M., & Sahudin, S. (2020, November). A Comparative Study of Web Design for Reengineering Pharmacokinetics Application. In *2020 IEEE Conference on Open Systems (ICOS)* (pp. 53-58). IEEE. doi:10.1109/ICOS50156.2020.9293641
- Alonso Vega, A. (2013). Responsive Web Design interfaces web adaptables al dispositivo empleando HTML5 y CSS3.
- Barba Soler, J. P. (2015). Diseño y desarrollo web. *Análisis de casos (Doctoral dissertation, Universitat Politècnica de València)* <http://hdl.handle.net/10251/49757>
- Berbel Marín, P. (2018). Desarrollo de un frontend en ReactJS. https://repositori.tuji.es/xmlui/bitstream/handle/10234/179297/Memoria_Pablo_Berbel.pdf?sequence=1&isAllowed=y
- Deshmukh, S. M. (2019, March). Building a single page

- application web front-end for E-learning site. In *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC.) IEEE*, 985-987. doi:10.1109/ICCMC.2019.8819703
- Gustafson, A. (2016). *Adaptive Web Design*, Second Edition.
- Labrada, E., & Salgado, (2013). Diseño web adaptativo o responsivo. *Revista Digital Universitaria* Vol(14).
- Li, N., & Zhang, B. (2019, November). The design and implementation of responsive web page based on HTML5 and CSS3. In *2019 International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)* (pp. 373-376). IEEE.
- Manso Guerra, Y. C., & Pedro Febles, J. (2016). Diseño web adaptativo para la plataforma educativa ZERA. *Revista Cubana de Ciencias Informáticas*, 10(2), 100-115.
- Otto, M., & Thornton, J. (2014). *Bootstrap*.
- Peltomäki, V. (2014). *Bootstrap Framework as a web design tool*.
- Pérez Ibarra, S. G. (2021). Herramientas y tecnologías para el desarrollo web desde el FrontEnd al BackEnd. In *XXIII Workshop de Investigadores en Ciencias de la Computación (WICC 2021, Chilecito, La Rioja)*.
- Perez, A. (2020). Características y funcionalidades de interacción e interfaz en el prototipo de visualización de los recursos culturales emergentes en el proyecto de investigación creando paz. <http://hdl.handle.net/20.500.11912/6155>
- Rodríguez, R. V. (2018). Análisis de frameworks web adaptativos basados en HTML5. .
- Tapani, C. (2014). *Study and Integrate Bootstrap 3 for OpixManager*.
- Valdivia-Caballero, J. J. (2016). Modelo de procesos para el desarrollo del front-end de aplicaciones web. *Interfases*, (009), 187-208. DOI: <https://doi.org/10.26439/interfases2016.n009.1245>
- Veloz, A. (2016). De la interfaz del usuario al responsive web design. . *Editorial UCSG*, 59-66.
- W3schools. (2018). *Bootstrap 3 Tutorial*.
- Yousaf, N. A. (2018). Towards adaptive and responsive web design: A systematic literature review. *Language, International Conference on Engineering, Computing & In-formation Technology (ICECIT 2017)*, 55-60.
- Zelaya, C. (2020). *Nuevas tendencias en desarrollo Web*.





Covid-19 en Ecuador: Aplicación de minería de datos

Covid-19 en Ecuador: Aplicación de minería de datos

Autores

- ✉ * Juan Carlos Zambrano
- ✉ Patricia Quiroz-Palma
- ✉ Alex Santamaría-Philco
- ✉ Willian Zamora

¹Facultad de Ciencias Informáticas,
Universidad Laica Eloy Alfaro de
Manabí, Manta, Ecuador.

* Autor para correspondencia

Comó citar el artículo: Zambrano, J. C., Quiroz-Palma, P., Santamaría-Philco, A., & Zamora, W. (2022). Covid-19 en Ecuador: Aplicación de minería de datos. *Informática y Sistemas: Revista de Tecnologías de la Informática y las Comunicaciones*, 6(1), 12-23. DOI: <https://doi.org/10.33936/isrtic.v6i1.4366>

Enviado: 31/01/2022
Aceptado: 09/05/2022
Publicado: 29/05/2022

Resumen



El COVID-19 se introdujo rápidamente como una pandemia global, la cual necesita ser tratada con respuesta inmediatas e integradas a todos los sistemas nacionales que requieran de ellas. Con la llegada del COVID-19 el mundo vio la necesidad de respuestas oportunas y el intercambio de datos sobre ésta y futuras pandemias globales de rápida propagación. Este estudio se enfoca en predecir la incidencia del COVID-19 en Ecuador. Se realizó Minería de Datos de los registros proporcionados por instituciones públicas del estado ecuatoriano con información oficial y actualizada del COVID-19 en el Ecuador. Se experimento con modelos de regresión y memoria a largo plazo obteniendo como resultado el modelo óptimo para estimar el número de casos positivos de COVID-19. Para el modelo matemático se hizo uso del error cuadrático medio como métrica del rendimiento. Del análisis de los datos sobre el COVID-19 en Ecuador el modelo de regresión lineal predijo la incidencia con un error cuadrático medio de 0.54. siendo los factores más efectivos la incidencia de días anteriores y el número de población de cada una de las provincias afectadas.

Palabras clave: Covid-19; Minería de datos; Metodología KDD; Algoritmos series-temporales.

Abstract

COVID-19 was quickly introduced as a global pandemic, which needs to be addressed with immediate and integrated responses to all national systems that require them. With the advent of COVID-19 the world saw the need for timely responses and data sharing on this and future rapidly spreading global pandemics. This study focuses on predicting the incidence of COVID-19 in Ecuador. Data mining was performed on records provided by public institutions of the Ecuadorian state with official and updated information on COVID-19 in Ecuador. We experimented with regression and long-term memory models, obtaining as a result the optimal model to estimate the number of positive cases of COVID-19. For the mathematical model, the mean square error was used as a performance metric. From the analysis of the data on COVID-19 in Ecuador, the linear regression model predicted the incidence with a mean square error of 0.54, the most effective factors being the incidence of previous days and the population in each of the affected provinces.

Keywords: Covid-19; Data mining, KDD methodology, Time-series algorithms.



1. Introducción

El mundo entero ha sido afectado de los efectos producidos por la pandemia provocada por coronavirus (COVID-19). La mayoría de los países se registran miles de infectados, bajas humanas, pacientes con secuelas y gran afectación económica. El coronavirus surgió en Wuhan, China y se extendió por todo el mundo. El análisis genómico reveló que el SARS-CoV-2 está relacionado filogenéticamente con los virus de murciélagos, similares al síndrome respiratorio agudo severo (similar al SARS), determinándose que estos mamíferos podrían ser el posible reservorio primario. Se desconoce la fuente intermedia de origen y transmisión a humanos, sin embargo, la rápida transferencia de humano a humano se ha confirmado ampliamente. (Muhammad, Suliman, Abeer, Nadia, & Rabeea, 2020).

En Ecuador se han reportado alrededor de 859 mil casos para el mes de marzo del 2022 desde el inicio de la emergencia sanitaria, según el informe emitido por el Ministerio de Salud Pública. Siendo Pichincha la provincia más afectada con 64469 casos en total representando el 36.99% de los casos registrados a escala nacional. En tanto, hay 35.421 fallecidos en el contexto de la pandemia en Ecuador. En Ecuador según los datos de (MSP, 2022) se han aplicado más de 30 millones de dosis de la vacuna contra el virus. Los resultados de esta se evidenciaron al registrar un número menor de casos confirmados por Covid-19 en comparación a fechas anteriores a la vacuna. Sin embargo, entre la segunda y tercera semana del mes de enero del año 2022 se registraron más de 100 mil casos, mientras que durante la última semana del 2021 se registraron 9092 casos, evidenciándose un gran incremento en de nuevos casos. Estos datos representan una valiosa fuente de información para estudiar el compartimiento del virus dentro del contexto ecuatoriano. Sin embargo, son pocos los estudios que apliquen modelos predictivos que generen pronósticos y que estos a su vez respalden a la toma de decisiones de los ministerios y organismos para salvaguardar la seguridad ciudadana.

En la India, (ubicar el espacio temporal) que al alcanzar las cifras de 536 casos y 11 muertos tomo la decisión de ejecutar un bloqueo nacional y rápidamente llamo al campo a las ciencias de datos para reunir información e implementar estrategias de predicción. Una de ellas fue una extensión bayesiana del modelo Susceptible-Infected-Removed (eSIR) diseñado para el pronóstico de intervenciones para estudiar el impacto a corto y largo plazo de un bloqueo inicial de 21 días en el número total de infecciones por COVID-19 (Debashree, Maxwell, Rupam, & Lili, 2020). En Ecuador se realizó un modelo predictivo de los casos de contagio por Covid-19 para la provincia de Loja presentado por los autores (Salcedo & Salcedo, 2021), donde se aplica método numérico de diferencias divididas, método logístico y método de mínimos cuadrados para predecir el nivel de contagio. Para ello se hizo uso de los datos proporcionados

por el Ministerio de Salud pública del Ecuador en un periodo de 399 días.

En este trabajo se presenta el proceso de minería de datos para predecir la incidencia del Covid-19 en el Ecuador, para se realizaron cada una de las etapas de la metodología Knowledge Discover Database (KDD). Los datos analizados durante un año nos permitieron predecir la proyección de incidencia de contagios en el Ecuador y sus provincias. En la sección 2, se describen los materiales y métodos usados en la solución propuesta mediante procesos de minería de datos. En la sección 3, se describen los resultados obtenidos de las predicciones; en la sección 4, se describen la discusión y en la sección 5 se describen las conclusiones y trabajos futuros de este trabajo de investigación.

2. Materiales y Métodos

Los investigadores de todo el mundo están utilizando el aprendizaje automático (Machine Learning) para desarrollar modelos que simulan y predicen la propagación del virus en un intento por identificar patrones que puedan revelar las debilidades y los peligros de esta pandemia. Robert H (Jason & Robert, 2020) expresa lo siguiente: “Taiwán aprovechó su base de datos de seguros de salud y la integró con su inmigración y base de datos aduanera para iniciar la creación de big data para análisis; generó alertas en tiempo real durante una visita clínica basada en el historial de viajes y síntomas para ayudar a la identificación del caso”

Tomando en cuenta lo expuesto y considerando el estado actual de la crisis sanitaria en Ecuador es necesario preparar un modelo de minería de datos que prestará los siguientes beneficios:

- Se recopilaron los registros diarios del Ministerio de Salud Pública del Ecuador y se lo modelara en un conjunto de datos el mismo que será accesible para que cualquier investigador o entidad a quien interese haga uso de los datos.
- Los modelos de regresión lineal indicarán las provincias presentarán mayor incidencia, lo que apoyaría a la toma de decisiones de las prefecturas y municipios competentes.

La investigación, tiene un enfoque descriptivo e interpretativo, con información de tipo cuantitativo, que busca predecir el número de contagios por COVID-19 en Ecuador. Para ello se consideran los datos proporcionados por el ministerio de Salud Pública del Ecuador, que posteriormente son recopilados entre el 13 de marzo del 2020 hasta 28 de marzo del 2022 el en un conjunto de datos.

2.1. Métodos de Investigación

Los métodos de investigación que se emplearon para predecir

la incidencia del COVID-19 en el Ecuador se describen a continuación:

- **Método Analítico:** Consiste en la desmembración de los elementos que componen la totalidad de las técnicas de minería de datos para la predicción, también instaura las relaciones causa, efecto y la naturaleza del problema identificado.
- **Método Sintético:** Con este método se pretende reconstruir en un todo a partir los elementos distinguidos por el análisis de minería de datos para estudiarlos y conseguir un extracto minucioso para la predicción de la incidencia del virus con técnicas de minería de datos.
- **Método Deductivo:** Se emplea este método para analizar la minería de datos, yendo de lo general a lo particular para definir conclusiones que ayudan a la predicción de la incidencia del COVID-19 utilizando registros históricos de los casos que se presentaron y se siguen presentando en el país, además de los factores que tengan relación con el incremento de contagios.
- **Método Inductivo:** Este método se operó realizando observaciones de la teoría y trabajo de otros investigadores entre estos el modelo predictivo para la incidencia del COVID-19 en Iran mediante el análisis de datos de Google Trends (Ayyoubzadeh, Zahedi, & R, 2020), el modelo predictivo de la salud de pacientes por COVID-19 aplicando algoritmos de bosques aleatorios (Celestine, Ali, Atharva, & R, 2020) en ellos se utilizaron técnicas de minería de datos para predecir la incidencia del COVID-19, mediante el razonamiento a partir de premisas particulares para sustentar conclusiones.

2.2 Herramientas de Recolección de datos

Las herramientas de recolección de datos empleadas para esta investigación son las siguientes.

El análisis documental: Se recopila información de diversas fuentes como lo son, reportes del MSP, COE Nacional, libros, sitios web, artículos científicos, etc., que tenga un enfoque en el problema identificado. Esta exploración de contextos, además de usarse para el desarrollo del marco teórico conceptual referente a minería de datos, incluida las fases, técnicas, algoritmos y demás; también tuvo su lugar en la extracción de los datos con relevancia para formar la conjunto de datos fundamental en el entrenamiento de los algoritmos de predicción empleados.

La observación: El investigador hace uso de su razonamiento luego de observar y percibir los acontecimientos locales y nacionales que se difunden por diversos medios y que tienen relación con el comportamiento de la pandemia a causa del COVID-19, permitiéndole identificar los objetos de estudios como patrones de propagación, incrementos en las tasas de contagio, aplicación de medidas de bioseguridad, etc., estos factores se correlacionan con la incidencia del virus.

La encuesta: Como herramienta adicional de recolección de datos para la investigación se empleó la encuesta, la misma que es apoyada por cuestionarios, como instrumento de recolección de la información. Estos cuestionarios fueron aplicados a través

de la herramienta Google Form y repartidos en grupos de redes sociales, Whatsapp y Facebook cuyos grupos están integrados por miembros de la comunidad estudiantil de las Facultades de Ciencias Informáticas, Ciencias Médicas y Enfermería de la Universidad Laica Eloy Alfaro de Manabí, además un pequeño número de formularios fueron extendidos hasta personas que laboran en primera línea en lucha contra la pandemia de COVID-19. Se optó por tomar una muestra de este universo puesto que varias de las preguntas incluidas en los formularios son de campo y requerían de un poco de conocimiento técnico para ser respondidas de manera confiable.

2.3 Fuentes de información de datos

Fuentes primarias: Para el desarrollo y propósito de esta investigación se optó por tomar el análisis documental como herramienta principal de información, siendo las fuentes más importantes los boletines epidemiológicos emitidos diariamente por el MSP, y COE Nacional esta información se encuentra documentada en los repositorios oficiales de las instituciones antes mencionadas, de donde se extrajeron datos de gran valor para el estudio de la predicción de la incidencia del virus en el país, como lo son el número de nuevos casos de cada provincia, dadas de altas, fallecimientos, vacante hospitalaria, género y rango de edad de los pacientes intervenidos.

Fuentes Secundarias: Como fuentes secundarias de información se designan los resultados obtenidos en las encuestas realizadas al muestreo definido de la población, que en conjunto con una observación estructurada realizada por parte del investigador hacia los acontecimientos nacionales relacionados a la pandemia que enfrenta el Ecuador, se logra identificar coincidencias en algunas de las variables consideradas como factores que intervienen en la incidencia del COVID-19 en el país.

2.4 Proceso de Minería de Datos

El proceso Knowledge Discover Database (KDD), tal como se presenta en (Fayyad, Piatetsky, & Smyth, 1996), es el proceso de usar métodos de Data Mining (DM) para extraer conocimiento, utilizando una base de datos junto con cualquier preprocesamiento, submuestreo y transformación requeridos de la base de datos. Se consideran cinco etapas como se muestra en la Figura 1.

1. **Selección.** - Esta etapa consiste en crear un conjunto de datos objetivo, o enfocarse en un subconjunto de variables o muestras de datos, en las que se debe realizar el descubrimiento.
2. **Procesamiento Previo.** - Esta etapa consiste en la limpieza y preprocesamiento de datos de destino para obtener datos consistentes
3. **Transformación.** - Esta etapa consiste en la transformación de los datos usando métodos de reducción de dimensionalidad o transformación.
4. **Minería de Datos.** - Esta etapa consiste en la búsqueda de patrones de interés en una forma representacional particular,



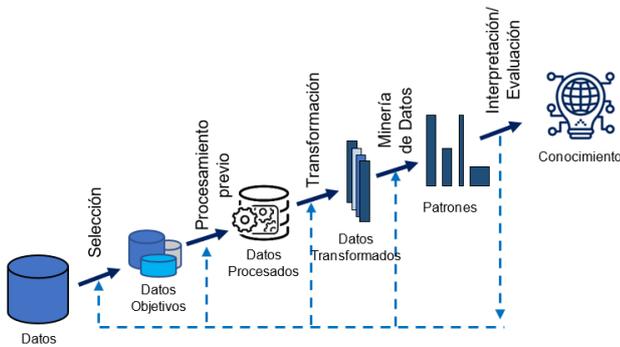


Figura 1. Etapas en el procedimiento KDD.

dependiendo del objetivo de minería de datos (por lo general, la predicción).

5. Interpretación / Evaluación. - Esta etapa consiste en la interpretación y evaluación de los patrones.

2.4.1. Selección de Datos

Para la obtención de la información se procedió a realizar el análisis de las infografías emitidas por el Servicio Nacional de Gestión de Riesgos y Emergencias desde su repositorio digital (COE, 2022). A través del cual se obtuvieron trescientos ochenta y dos registros en un periodo de 365 días. Para los mismos se definió una estructura adaptada a un conjunto de datos tabulados para su procesamiento.

La siguiente fase de esta etapa correspondió a la creación del conjunto de datos siendo la fuente de datos los informes por casos de COVID-19 del COE Nacional, dichos informes se emitieron en documentos digitales con extensión de archivo '.pdf' que representa una fuente de datos no procesados. Además de esto, todos los informes presentan un formato general que comprende su contenido, esto incluye: textos, tablas, imágenes, gráficos estadísticos, marcas de agua etc., Las condiciones de los documentos dificultó la extracción automática de los datos. Por este motivo, se planteó un método alternativo para la extracción de los datos, el cual implica: la extracción por captura de imagen del contenido relevante de los informes, aplicación de filtro de color para minimizar las marcas de agua, la aplicación de una herramienta de reconocimiento óptico de caracteres y libros de Excel. Para la comprensión de esta fase se detalla el proceso en la Figura 2 con la descripción de cada uno de los pasos.

2.4.2. Procesamiento previo

Después del análisis y selección de Python con sus librerías para el análisis de datos como herramienta de minería de datos, se procede a continuar con esta etapa de la metodología. De este

modo se realiza la carga de información de los conjuntos de datos "Covid-19EC_dataset.csv" y "Datos provincia.csv" los mismo que fueron almacenados en variables de nombre "df1" y "df2" respectivamente, esto se logró mediante mediante el paquete de Python Pandas por lo que se apoya en el uso del módulo 'pandas.read_csv', que proporciona estructuras de datos similares a los marcos de datos (ver Figura 3).

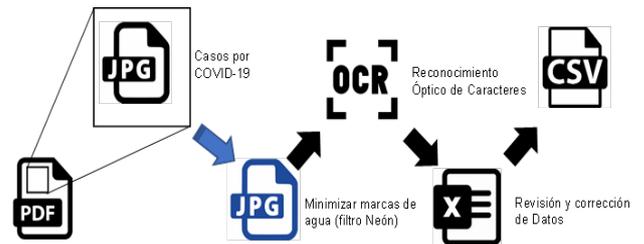


Figura 2. Proceso de extracción de datos.

Así los datos pasan de ser visualizados como se muestra en la Figura 4, en su estructura original del conjunto de datos en formato CSV delimitado por comas, a su adaptación dentro de un marcos de datos proporcionado por la librería Pandas como se observa en las Figuras 5 y 6.

Después de cargar el conjunto de datos y, siguiendo las fases de la metodología se procedió con el preprocesamiento de los datos el cual consiste en dos etapas: 'limpieza' y 'procesamiento'.

Se procedió a concatenar los conjuntos de datos cargados previamente, mediante el uso del módulo 'pd.DataFrame.merge'

```

jupyter Modelo predictivo COVID-19 EC Last Checkpoint: 28/07/2021
File Edit View Insert Cell Kernel Widgets Help
+ -> <-> Run Code
In [1]: import pandas as pd
In [2]: # Carga del dataset de los casos por COVID-19
df1 = pd.read_csv("Covid-19EC_Dataset.csv")
In [3]: # Carga del dataset con los datos de las provincias
df2 = pd.read_csv("Datos provincia.csv")
    
```

Figura 3. Fase de Selección de datos.

definiendo los datos del campo "provincia" como indicador común y se lo almacenó en la variable "df1". Tras la ejecución de este módulo se obtuvo un nuevo marcos de datos el mismo que contiene los campos de ambos conjuntos de datos (ver Figura 7).

Para la siguiente etapa se dio paso a las tareas de limpieza y depuración de datos. Por ello se empleó el módulo '.info()' el cual proporciona la información de los campos del conjunto de datos

```
[fecha,día,mes,año,casosConfirmados,fallecimientos,provincia,región
30/04/2020,30,4,2020,187,19,Esmeraldas,Costa
29/04/2020,29,4,2020,173,19,Esmeraldas,Costa
28/04/2020,28,4,2020,164,17,Esmeraldas,Costa
27/04/2020,27,4,2020,143,15,Esmeraldas,Costa
23/04/2020,23,4,2020,115,12,Esmeraldas,Costa
22/04/2020,22,4,2020,104,7,Esmeraldas,Costa
21/04/2020,21,4,2020,102,7,Esmeraldas,Costa
20/04/2020,20,4,2020,102,7,Esmeraldas,Costa
19/04/2020,19,4,2020,89,7,Esmeraldas,Costa
18/04/2020,18,4,2020,86,7,Esmeraldas,Costa
17/04/2020,17,4,2020,85,7,Esmeraldas,Costa
16/04/2020,16,4,2020,53,7,Esmeraldas,Costa
15/04/2020,15,4,2020,44,7,Esmeraldas,Costa
14/04/2020,14,4,2020,41,7,Esmeraldas,Costa
13/04/2020,13,4,2020,38,7,Esmeraldas,Costa
12/04/2020,12,4,2020,38,7,Esmeraldas,Costa
11/04/2020,11,4,2020,31,7,Esmeraldas,Costa
10/04/2020,10,4,2020,31,7,Esmeraldas,Costa
```

Figura 4. Estructura del conjunto de datos original.

```
In [7]: # Visualizamos el dataset con las datos de las provincias
df2
```

provincia	región	poblacion	superficieKm2	densidadPoblacion	
0	Azuay	Sierra	881394	8309.58	106.069521
1	Bolívar	Sierra	209803	3645.38	53.209830
2	Cañar	Sierra	281306	3146.08	89.443371
3	Carchi	Sierra	198669	3700.45	49.430359
4	Chimborazo	Sierra	624004	6469.72	80.619473
5	Cotacachi	Sierra	488719	6108.23	80.009430
6	El Oro	Costa	715751	5766.88	124.119383
7	Esmeraldas	Costa	643664	16132.73	39.898638
8	Galápagos	Insular	33042	8010.00	4.125064
9	Guayas	Costa	4387434	15430.40	284.337023
10	Imbabura	Sierra	478257	4687.51	103.816013
11	Loja	Sierra	521154	11062.73	47.108967
12	Los Ríos	Costa	921763	7205.27	127.929002
13	Morona Santiago	Sierra	1562079	18639.80	82.478974
14	Morona Santiago	Amazonia	198535	24059.40	8.168741
15	Napo	Amazonia	133705	12542.50	10.660156
16	Orellana	Amazonia	161338	21662.10	7.437938
17	Pastaza	Amazonia	114202	29641.37	3.852781
18	Pichincha	Sierra	3228233	9535.91	338.534130
19	Santa Elena	Costa	401178	3690.17	108.715317
20	Santo Domingo	Costa	458580	3446.65	133.050934
21	Sucumbios	Amazonia	230503	18064.42	12.745844

Figura 5. Previsualización de “Covid-19EC_dataset.csv”.

```
In [18]: # Concatenamiento de los datasets
df1 = pd.DataFrame.merge(df1, df2, how="left", on="provincia")
df1
```

fecha	día	mes	año	casosConfirmados	casosDiarios	fallecimientos	nuevosFallecimientos	provincia	región_x	región_y	poblacion	superficieKm2
0	2020-03-13	13	3	2020	0	0	0	0	Azuay	Sierra	881394	8309.58
1	2020-03-14	14	3	2020	1	1	0	0	Azuay	Sierra	881394	8309.58
2	2020-03-15	15	3	2020	1	0	0	0	Azuay	Sierra	881394	8309.58
3	2020-03-16	16	3	2020	1	0	0	0	Azuay	Sierra	881394	8309.58
4	2020-03-17	17	3	2020	5	4	0	0	Azuay	Sierra	881394	8309.58
2480	2020-06-26	26	6	2020	486	0	16	0	Zamora Chiriquí	Amazonia	120416	10548.28
2481	2020-06-27	27	6	2020	486	30	16	0	Zamora Chiriquí	Amazonia	120416	10548.28
2482	2020-06-28	28	6	2020	523	27	16	0	Zamora Chiriquí	Amazonia	120416	10548.28
2483	2020-06-29	29	6	2020	523	0	16	0	Zamora Chiriquí	Amazonia	120416	10548.28
2484	2020-06-30	30	6	2020	536	13	16	0	Zamora Chiriquí	Amazonia	120416	10548.28

Figura 6. Previsualización de “Datos provincia.csv” cargados a Pandas mediante el módulo ‘pd.read_csv’.

su estructura. Con ello se pudo observar que el nuevo marco de datos está estructurado con 14 columnas cada una de ellas con # de entradas lo cual descarta la existencia de campos vacíos. Además de esto se observó que la columna del campo “fecha” está definida como ‘object’, lo cual representaría inconvenientes al momento de generar series temporales ya que es necesaria la presencia de la variable de tipo ‘date’.

```
In [4]: # Visualizamos el dataset de los casos por COVID-19
df1
```

fecha	día	mes	año	casosConfirmados	casosDiarios	fallecimientos	nuevosFallecimientos	provincia	región	
0	2020-03-13	13	3	2020	0	0	0	0	Azuay	Sierra
1	2020-03-14	14	3	2020	1	1	0	0	Azuay	Sierra
2	2020-03-15	15	3	2020	1	0	0	0	Azuay	Sierra
3	2020-03-16	16	3	2020	1	0	0	0	Azuay	Sierra
4	2020-03-17	17	3	2020	5	4	0	0	Azuay	Sierra
2480	2020-06-26	26	6	2020	486	0	16	0	Zamora Chiriquí	Amazonia
2481	2020-06-27	27	6	2020	486	30	16	0	Zamora Chiriquí	Amazonia
2482	2020-06-28	28	6	2020	523	27	16	0	Zamora Chiriquí	Amazonia
2483	2020-06-29	29	6	2020	523	0	16	0	Zamora Chiriquí	Amazonia
2484	2020-06-30	30	6	2020	536	13	16	0	Zamora Chiriquí	Amazonia

Figura 7. Concatenación de los conjuntos de datos “Covid-19EC_dataset.csv” y “Datos provincia.csv”.

2.4.3. Transformación de datos

Esta etapa de la metodología se realizaron las modificaciones necesarias al conjunto de datos con la finalidad proveer una estructura de los datos adecuadas para el modelo propuesto. Se identificó un defecto en el tipo de variable del campo “fecha”, fue necesario transformar dicha variable a tipo ‘date’ con la ayuda del módulo ‘pd.to_datetime’ como se observa en la Figura 8.

Se realizó la conversión de la columna ‘fecha’ de ser una columna que contiene una variable a ser un indicador de las demás columnas, para ello se empleó el módulo ‘set_index()’ sobre el campo mencionado. Este proceso tuvo lugar con el fin de que la estructura del conjunto de datos permita generar un modelo con series temporales más adelante.

2.4.4. Minería de Datos

Concluidas las fases anteriores se obtuvo una base de datos lista para continuar el proceso de minería de los datos, el mismo que se comprendió de tres etapas, partiendo desde la búsqueda de patrones entre el conjunto de datos, los mismos que permitan identificar una correlación entre ellos. Para cumplir esta tarea se empleó la librería Seaborn y Matplotlib de Python.

Se identificaron tres correlaciones importantes siendo estas: casos confirmados-fallecimientos, casos confirmados-población y casos confirmados-densidad de la población, las mismas que poseen un grado de correlación equivalente a 0.91, 0.76 y 0.61 correspondientemente. Ver Figura 9.

Después de analizar los primeros resultados se identificó que el número de los casos confirmados (“casosConfirmados”)

```
In [12]: df1["fecha"] = pd.to_datetime(df1["fecha"], errors = 'coerce')

In [13]: df1.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 2465 entries, 0 to 2464
Data columns (total 14 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   fecha                 2465 non-null  datetime64[ns]
1   dia                  2465 non-null  int64
2   mes                  2465 non-null  int64
3   año                  2465 non-null  int64
4   casosConfirmados    2465 non-null  int64
5   casosDiarios        2465 non-null  int64
6   fallecimientos      2465 non-null  int64
7   nuevosFallecimientos 2465 non-null  int64
8   provincia            2465 non-null  object
9   región_x             2465 non-null  object
10  región_y             2465 non-null  object
11  población            2465 non-null  int64
12  superficieKm2        2465 non-null  float64
13  densidadPoblacion    2465 non-null  float64
dtypes: datetime64[ns](1), float64(2), int64(8), object(3)
```

Figura 8. Transformación Variable 'fecha' de tipo 'object' a tipo 'datetime'.

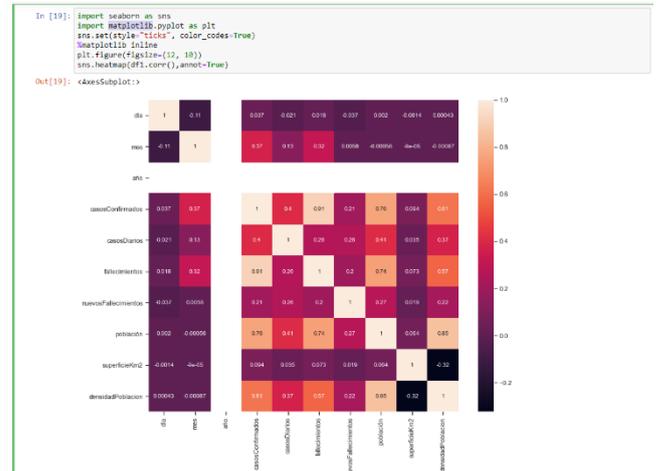


Figura 9. Gráfica de calor con los índices de correlación.

es la variable que se encuentra presente en cada una de las correlaciones (Ver Figura 10), además de ser esta el objetivo a predecir de este modelo. Por esta razón, se realizó un segundo análisis comparando la dependencia de una variable a otra, para realizar esta tarea se empleó el módulo 'sns.pairplot' de la librería Matplotlib.

Del análisis de los resultados, se identificó que a mayor número de habitantes y un índice más elevado de la densidad poblacional en una provincia mayor resulta el incremento de los casos confirmados por COVID-19 y esto representa también una elevación en la tasa de fallecimiento por el virus. En la siguiente sección, se relatan los resultados obtenidos a partir de la aplicación del modelo seleccionado.

3. Resultados y Discusión

Posterior al análisis e interpretación de toda la información, se prosigió a buscar un modelo que genere las predicciones en razón a las variables previamente identificadas. Para ello, se consideró trabajar primero sobre un número reducido de datos a través del módulo '.sample()' que permitió tomar una muestra, siendo esta el diez por ciento de los registros del conjunto de datos. A continuación, se describen los experimentos con algoritmos de predicción para identificar cuál de ellos se adapta mejor al modelo iniciando por los algoritmos de regresión.

3.1. Primer experimento (regresión lineal)

Considerando los resultados obtenidos de la matriz de correlación, decidió realizar pruebas con modelos de regresión lineal y polinomial, para ello se seleccionaron las variables con

mayor correlación siendo estas la 'densidad poblacional' y los 'casos confirmados'.

Para este modelo se utilizó la librería Scikit-learn y el módulo 'LinearRegression', con el fin de encontrar la tendencia de la variable 'casos confirmados' con dependencia en la variable 'densidad poblacional', expresada en habitantes por metro cuadrado. Para calcular el grado de precisión del modelo se utilizó el error medio cuadrático como métrica de rendimiento, siendo esta 0.38 como podemos observar en la Figura 11.

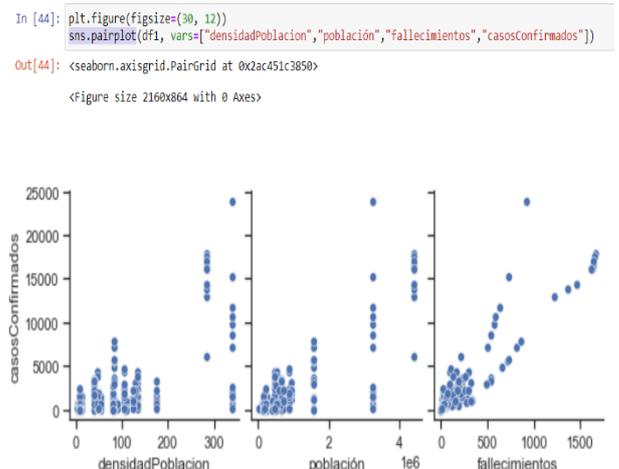


Figura 10. Dependencia de la variable "casosConfirmados" con respecto a las variables "densidadPoblacion", "población" y "fallecimientos".

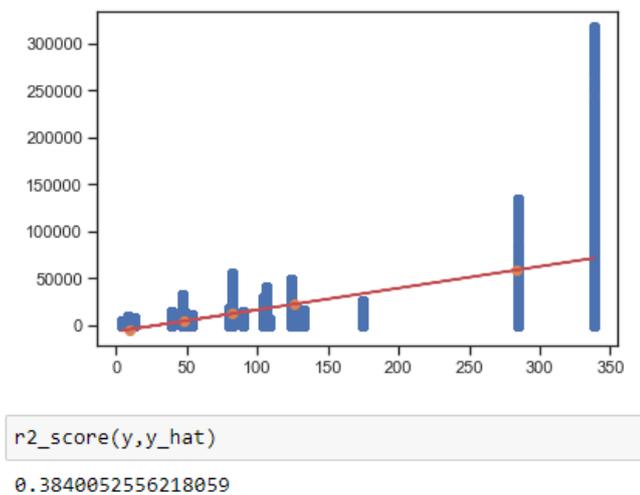


Figura 11. Primer experimento, definición de campos para regresión lineal.

Debido a los resultados de la métrica de rendimiento, se procedió a ejecutar el modelo de regresión polinomial, escalando con polinomios del grado 1 al 10 con el fin de encontrar el mejor ajuste para el modelo comparando cada uno de ellos. Siendo el polinomio de tercer grado el que presentó una métrica más alta con 0.54 de rendimiento en comparación a los demás polinomios como se puede observar en la Figura 12.

Posteriormente se procedió a generar las predicciones de los casos confirmados con relación a la cantidad de habitantes por metro cuadrado, a partir del modelo de regresión polinomial con polinomio de tercer grado. Se obtuvieron pronósticos de los casos confirmados en relación con la cantidad de habitantes por metros cuadrados, como se observa en la Figura 13.

Para evaluar los resultados del modelo se tomaron como muestra las provincias con una densidad poblacional aproximada y se compararon los casos confirmados del pronostico con los datos reales con fecha 28 de marzo del 2022. Los mismo que se detallan en la Tabla 1.

Llegado a este punto se implementó el modelo secuencial de memoria de largo plazo utilizando la librería Keras con una capa de entrada, dos capas oculta y una capa de salida, donde las capas ocultas son capas LSTM, además la primera de ella pose 4 unidades ocultas y la segunda solo pose 1 unidad. Posterior a ello se evaluó el modelo mediante un historial de perdida, el cual se configuró para el calculo de dos mil iteraciones parada uno de los datos del conjunto de entrenamiento equivalente al 70% del total de los datos, tal como se observa en la ilustración 37. Una vez procesado los datos se pudo observar que el nivel perdida del modelo era satisfactorio, como se observa en la ilustración 38, ya que una menor perdida proporciona una mejor predicción.

Una vez evaluado el modelo y obtenido un índice de perdida bajo, se procede a realizar las predicciones, basadas en los casos confirmados. Para ello se tomó de muestra las provincias de: Bolívar, Cañar, Loja y Pichincha como se observa en la Figura 14.

Las predicciones se realizaron a partir de los registros de casos confirmados por COVID 19 durante el primer trimestre del año 2022, con extensión hasta el 28 marzo. Se obtuvo el pronostico de 7 días en el futuro (04/04/2022) para las provincias del conjunto de muestra. Se observa que el modelo secuencial LSTM proporciona una precisión del %, para predecir la tendencia de casos confirmados por COVID 19, el desglose de los resultados se puede observar en la Tabla 2.

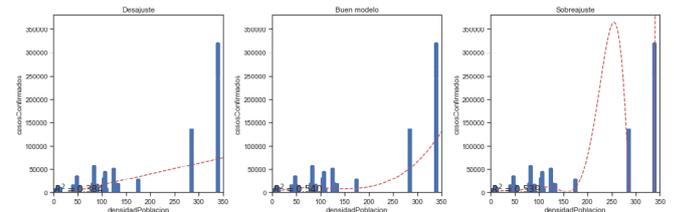


Figura 12. Regresión Polinomial. Ajuste polinomio grado (1-10).

Este trabajo de investigación permitió analizar la información respecto a la crisis sanitaria por Covid-19 y predecir la incidencia de esta pandemia en el Ecuador mediante el uso de procesos y metodologías de minería de datos que se describen en las secciones anteriores. Esta aplicación se constituye en una

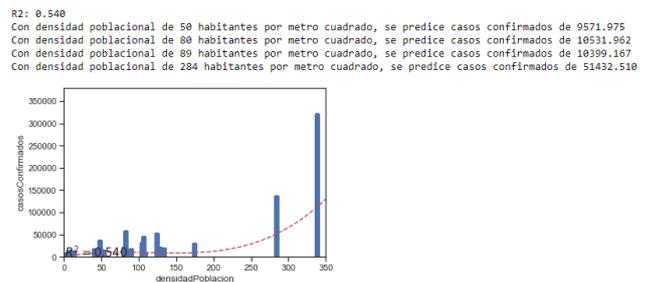


Figura 13. Regresión Polinomial, pronóstico de los casos confinados en relación con la densidad poblacional.

herramienta que aporta a mejor toma de decisiones para prevenir y precautelar la salud de los ecuatorianos.

Tras el análisis correlacional entre los datos de las provincias del Ecuador y los casos confirmados por COVID-19 en cada una de ellas, se identificó que, para los casos particulares entre las provincias de Pichincha y Guayas, el factor más influyente no fue el total de la población, si no la cantidad de habitantes por metros cuadrado dentro de la provincia. Para el caso de las Galápagos, ninguno de los factores mencionados anteriormente desempeño un papel crucial en la propagación del virus, ya que se observó que los casos nuevos por contagios de COVID 19 se prolongaron en la región y dentro algunos periodos fueron nulos. Se presume que los factores más influyentes fueron las medidas de bioseguridad rigurosas que se venían tomando incluso antes de iniciar la pandemia por medidas de conservación de las especies endémicas de la región. Al no contar con un indicador que pueda medir este factor no fue posible realizar el estudio de este dentro del impacto de los contagios. Los resultados obtenidos en

Tabla 1. Resultados del modelo predictivo: Regresión polinomial.

Fuente: Los autores.

Provincia	Datos reales 23/03/2022		Predicción: Regresión polinomial		Precisión
	Hab. m ²	Casos conf.	Hab. m ²	Casos conf.	
Carchi	49.43	13.111	50	9.572	73 %
Chimbo-razo	80.62	15.866	80	10.532	66 %
Cotopaxi	80.01	17.899	80	10.532	59%
Cañar	89.44	14.335	89	10.399	73%
Guayas	284.34	134.054	284	51.433	38%

ambos experimentos nos permiten concluir que las herramientas tecnológicas y, en este caso con la aplicación de ciencias de datos cumplen un papel importante en el aporte científico para el desarrollo de soluciones importantes para el país y el mundo.

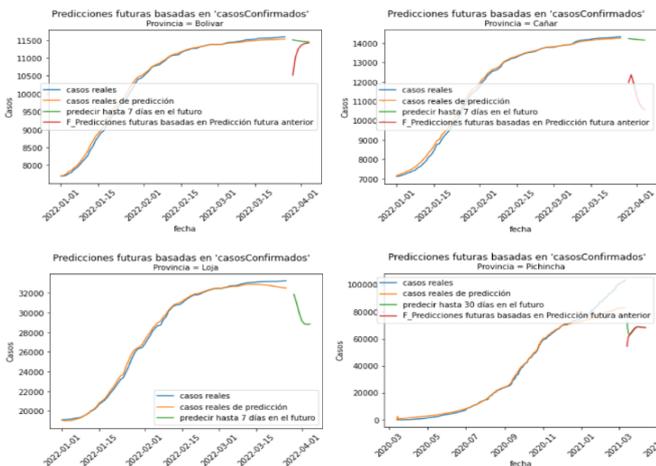


Figura 14. Predicciones conjunto de muestra.

Tabla 2. Resultados del modelo secuencial LSTM.

Fuente: Los autores.

Provincia	Casos confirmados (04/04/2022)	Predicción	Diferencia	Porcentaje precisión
Bolívar	11640	11447	193	98.34%
Cañar	14388	14160	228	98.42%
Loja	33357	31896	1461	95.62%
Pichincha	319676	298791	20885	93.47%
PROMEDIO PRECISIÓN DEL MODELO				96.46%

4. Conclusiones

El estado del arte en técnicas de minería de datos y modelos predictivos fue estudiado y revisado, con el fin de explorar y comprender la manera en que estas técnicas han sido aplicadas en respuesta a la emergencia sanitaria que se presentó por el Covid-19. Para ello, en la investigación fue documentada la metodología KDD junto a las técnicas predictivas que se aplicaron en este proyecto.

Se preparó un conjunto de datos a partir de los datos de Covid-19, emitidos por el Ministerio de Salud Pública del Ecuador mediante sus boletines epidemiológicos, con un periodo desde el 13 de marzo del 2020 al 28 de marzo del 2022. Con el fin de contar con un conjunto de datos que se puedan procesar e implementar en modelos predictivos. Las limitaciones se dieron debido a la irregularidad de los datos en los registros oficiales, obligó a implementar técnicas que ayudaron a detectar las anomalías y que estas sean depuradas en el producto final.

Se aplicaron algoritmos de minería de datos de regresión y redes neurales recurrentes para implementarlos en los modelos predictivos de regresión polinomial y memorias de corto y largo plazo respectivamente. Entre estos modelos se pudo observar que las memorias de corto y largo plazo generan mejores pronósticos de casos confirmados por Covid-19, alcanzando una precisión general de hasta el 96.46%. Como trabajos futuros se propone actualizar los datos para realizar predicciones futuras incrementando las variables que se presenten en la realidad futura. Además de ampliar los experimentos a más provincias y la mejora continua de los modelos aplicados.

Agradecimientos

El trabajo de los autores es parcialmente soportado por la Universidad Laica Eloy Alfaro de Manabí.

Contribución de los autores

Juan Carlos Zambrano: Conceptualización, Metodología, Análisis formal. **Patricia Quiroz-Palma:** Redacción – borrador original del artículo, Metodología, Análisis formal. **Alex Santamaría-Philco:** Metodología, Revisión y edición del artículo. **William Zamora:** Revisión y edición del artículo.

Conflictos de interés

Los autores declaran no tener ningún conflicto de interés.

Anexos

Análisis e interpretación de las encuestas aplicadas

Pregunta 1.- ¿Pertenece o perteneció usted al personal de primera línea (médicos, enfermeras, paramédicos, etc....) frente al COVID-19?

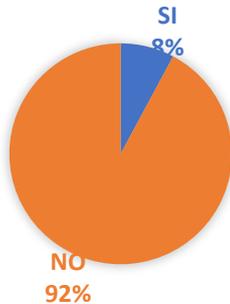


Figura A1. Gráfico pregunta 1.

Análisis:

De un total de 128 encuestados que representan al 100%, el 92% que corresponde a 118 encuestados han respondido que no pertenecen al personal de primera línea, y el 8% que pertenece a 10 encuestados han respondido que si pertenecen o pertenecieron al personal que labora o laboró en primera línea en la pandemia de COVID-19.

Interpretación:

La mayor parte de los encuestados pertenecen al grupo de la comunidad estudiantil con conocimientos de campo respecto a la investigación realizada, mientras que la parte restante son profesionales de primera línea en la pandemia de COVID-19.

Pregunta 2.- ¿Conoce usted que es la Minería de Datos?

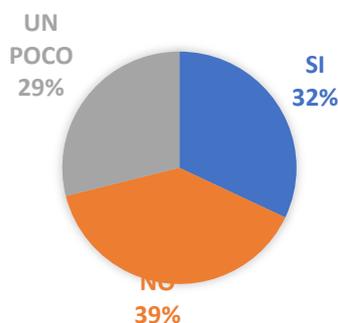


Figura A2. Gráfico pregunta 2.

Análisis:

De un total de 128 encuestados que representan al 100%, el 39% que corresponde a 50 encuestados han respondido que no conocen que es la minería de datos, y el 32% que pertenece a 41 encuestados han respondido que si conocen y el 29% que corresponde a 37 encuestados respondieron que tienen un conocimiento limitado del tema.

Interpretación:

Más de la mitad de los encuestados tienen la noción de lo que es la minería de datos sus tendencias y alcances a la hora de resolver problemas en diversos campos de acción.

Pregunta 3.- ¿Sabía usted que en el área de la salud la Minería de Datos puede tener un valor científico o investigador que ayude a determinar causas de determinadas patologías o a identificar poblaciones de riesgo?

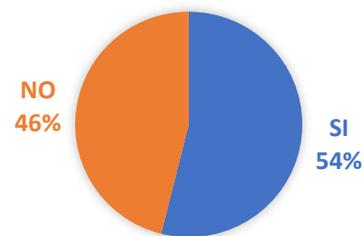


Figura A3. Gráfico pregunta 3.

Análisis:

De un total de 128 encuestados que representan al 100%, el 54% que corresponde a 69 encuestados han respondido que sí estaban familiarizados con aplicación de la minería de datos en estas temáticas, y el 46% que pertenece a 59 encuestados han respondido que no estaban informados al respecto.

Interpretación:

Un poco más de la mitad de los encuestados están al tanto de los avances de la minería de datos en las investigaciones de causas patológicas, esto es un indicador de que la mayor parte de los encuestados tienen la noción de que se presente con lograr con la esta investigación.

Pregunta 4.- ¿Cree usted que se esté aplicando Minería de Datos para estudiar el COVID-19 en Ecuador?

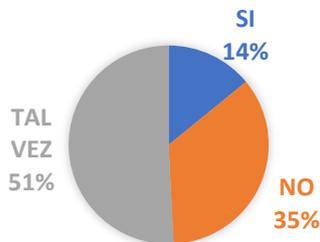


Figura A4. Gráfico pregunta 4.

Análisis:

De un total de 128 encuestados que representan al 100%, el 51% que corresponde a 65 encuestados han respondidos que no están seguros de si se aplica esta tecnología en el país, el 36% que pertenece a 45 encuestados han respondido que no se está aplicando esta mientras que el 14% restante y que corresponde a 18 encuestados consideran que si se está aplicando minería de datos para estudiar el COVID-19 en el país.

Interpretación:

La mitad de los encuestados desconocen si se está aplicando minería de datos para estudiar la pandemia en el país, la mitad restante se reparte entre grupos que afirman si se aplica o no minería de datos como herramienta de estudio.

Pregunta 5.- Si se empleara Minería de Datos y algoritmos para predecir la incidencia del COVID-19 en Ecuador. ¿Cuál considera usted que sería su impacto?

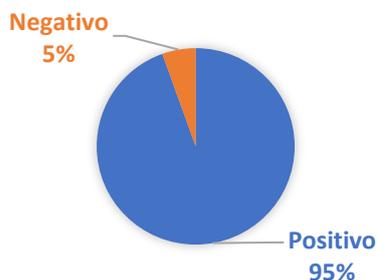


Figura A5. Gráfico pregunta 5.
Fuente: Encuestas-Elaboración propia

Análisis:

De un total de 128 encuestados que representan al 100%, el 95% que corresponde a 121 encuestados han respondido que una vez que se logre predecir la incidencia del COVID-19 esta investigación tendría un impacto positivo, y el 5% que pertenece a 7 encuestados han respondido que no tendrá un buen impacto.

Interpretación:

Con las respuestas obtenidas en esta sección de la encuesta se logra determinar el índice de aceptación de esta investigación, como un apoyo a los aconteceres que si ven en el país a casusa de la pandemia.

Pregunta 6.- Observando su propio entorno. ¿Qué factores considera usted que están relacionados con la incidencia del COVID-19 en su localidad?

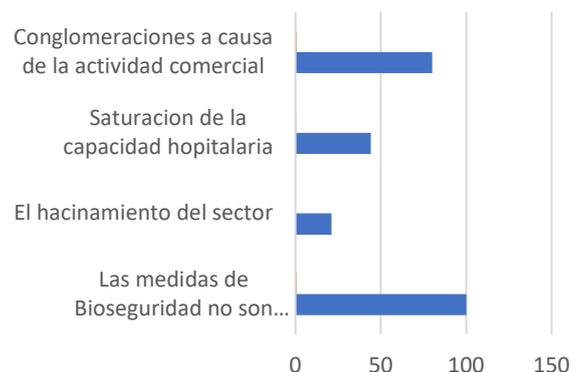


Figura A6. Gráfico pregunta 6.

Análisis:

De un total de 128 encuestados que representan al 100%, en esta serie de pregunta de opciones múltiples el 78% que corresponde a 100 encuestados han respondido que el mal cumplimiento de las medidas de bioseguridad es el factor que tiene una relación más estrecha con la incidencia del COVID-19 en el País, el 62% que corresponde a 80 encuestados señalaron también que las conglomeraciones a causa de actividades de carácter económico son el segundo factor responsable de la incidencia investigada, el 34% que corresponden a 44 encuestas señalan la saturación en los hospitales como otro factor de relevancia mientras que el 16% que equivale a 21 de las encuestas repartidas indican que el hacinamiento de la población es otro de los factores que participa con un menor grado de efectividad en la incidencia del virus en el país.

Interpretación:

La escasa ejecución de las medidas de bioseguridad sanitaria sería el factor con mayor relación en las incidencias del COVID-19 presentadas en el país según las encuestas, a esto se le suman los demás factores que tienen participación en este acontecimiento.

Pregunta 7.- Una vez que se haya implementado la minería de Datos y se logre anticipar la incidencia del COVID-19. ¿De qué manera considera usted que beneficiaría esto al Ecuador?

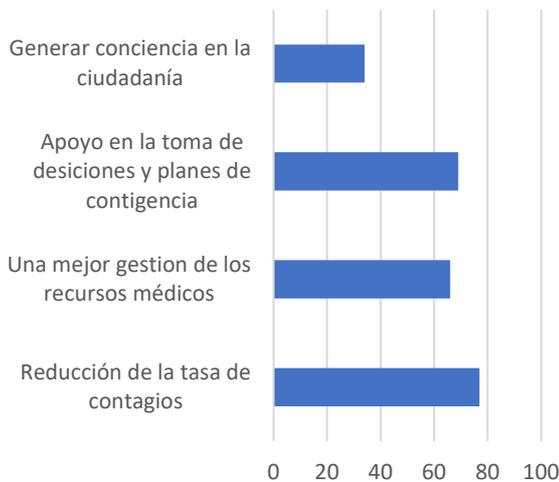


Figura A7. Gráfico pregunta 7.

Análisis:

De un total de 128 encuestados que representan al 100%, en esta serie de pregunta de opciones múltiples el 60.2% que corresponde a 77 encuestas responden a que mayor beneficio una vez implementada la propuesta es la reducción en la tasa de contagios por COVID-19, el 53.9% que corresponde a 69 de las encuestas señalan también que implementación de esta propuesta apoyara a las autoridades en la toma de decisiones, el 51.6% que corresponden a 66 encuestas señalan que mejorará la gestión de los recursos médicos el 26.6% que equivale a 34 indican que predecir la incidencia ayudaría a generar conciencia en la ciudadanía

Interpretación:

La mayoría de los encuestados coinciden en que la aplicación de este estudio jugaría un papel en la reducción de la tasa de contagios con COVID-19, gran parte de los encuestados señalan también predecir la incidencia del virus apoyaría a las autoridades a tomar mejores decisiones en esta lucha contra la pandemia.

Pregunta 8.- ¿Qué tan necesario cree usted que se debe aplicar Minería de Datos para predecir la incidencia del COVID-19 en el País?

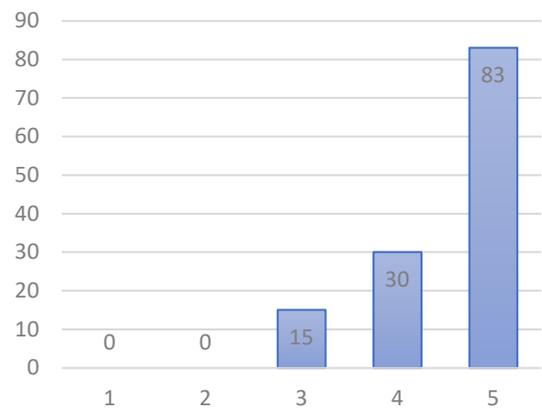


Figura A8. Gráfico pregunta 8.

Análisis:

De un total de 128 encuestados que representan al 100%, el 64.8% que corresponde a 83 de las encuestas indican con el mayor nivel de necesidad la implementación minería de datos para predecir la incidencia del COVID-19 en el País, el 23% que equivale a 30 de las encuestas reflejan un grado de necesidad de nivel 4, mientras que el 11% restante que señala que no están necesario aplicar Minería de datos para la predecir la incidencia del virus.

Interpretación:

En esta sección de la encuesta mediante el empleo de preguntas con respuestas de escala lineal donde 1 significa el más bajo nivel de necesidad y 5 el nivel de necesidad más alto destaca que la mayoría de los encuestados coinciden en que se debe aplicar minería de datos para predecir la incidencia del COVID-19 en el país.

Referencias bibliográficas

Ayyoubzadeh, S. M., Ayyoubzadeh, S. M., Zahedi, H., Ahmadi, M., & Kalhori, S. R. N. (2020). Predicting COVID-19 incidence through analysis of google trends data in Iran: data mining and deep learning pilot study. *JMIR public health and surveillance*, 6(2), e18828.

Celestine, I., Ali, K. B., Atharva, P., & R, S. (2020). COVID-19 Patient Health Prediction Using Boosted Random Forest Algorithm. *Frontiers in Public Health*, 357.

Ray, D., Salvatore, M., Bhattacharyya, R., Wang, L., Du, J., Mohammed, S., ... & Mukherjee, B. (2020). Predictions, role of interventions and effects of a historic national lockdown in India's response to the COVID-19 pandemic: data science call to arms. *Harvard data science review*, 176(3), 139-148.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI*





magazine, 17(3), 37-37.

Wang, C. J., Ng, C. Y., & Brook, R. H. (2020). Response to COVID-19 in Taiwan: big data analytics, new technology, and proactive testing. *Jama*, 323(14), 1341-1342.

MSP. (2022). Vacunómetro. Obtenido de

[https://app.powerbi.com/view?](https://app.powerbi.com/view?r=eyJrIjoiYTtkzNTFkMmUtZmUzNi00NDcwLTg0MDEtNjFkNzhhZTg5ZWYyIiwidCI6IjcwNjlyMGRiLTliMjktNGU5MS1hODIhLTlhNmIwNmQyNjlmMyJ9&pageName=ReportSection)

[r=eyJrIjoiYTtkzNTFkMmUtZmUzNi00NDcwLTg0MDEtNjFk](https://app.powerbi.com/view?r=eyJrIjoiYTtkzNTFkMmUtZmUzNi00NDcwLTg0MDEtNjFkNzhhZTg5ZWYyIiwidCI6IjcwNjlyMGRiLTliMjktNGU5MS1hODIhLTlhNmIwNmQyNjlmMyJ9&pageName=ReportSection)

[NzhhZTg5ZWYyIiwidCI6IjcwNjlyMGRiLTliMjktNGU5MS](https://app.powerbi.com/view?r=eyJrIjoiYTtkzNTFkMmUtZmUzNi00NDcwLTg0MDEtNjFkNzhhZTg5ZWYyIiwidCI6IjcwNjlyMGRiLTliMjktNGU5MS1hODIhLTlhNmIwNmQyNjlmMyJ9&pageName=ReportSection)

[1hODIhLTlhNmIwNmQyNjlmMyJ9&pageName](https://app.powerbi.com/view?r=eyJrIjoiYTtkzNTFkMmUtZmUzNi00NDcwLTg0MDEtNjFkNzhhZTg5ZWYyIiwidCI6IjcwNjlyMGRiLTliMjktNGU5MS1hODIhLTlhNmIwNmQyNjlmMyJ9&pageName=ReportSection)

[=ReportSection](https://app.powerbi.com/view?r=eyJrIjoiYTtkzNTFkMmUtZmUzNi00NDcwLTg0MDEtNjFkNzhhZTg5ZWYyIiwidCI6IjcwNjlyMGRiLTliMjktNGU5MS1hODIhLTlhNmIwNmQyNjlmMyJ9&pageName=ReportSection)

Shereen, M. A., Khan, S., Kazmi, A., Bashir, N., & Siddique, R. (2020). COVID-19 infection: Emergence, transmission, and characteristics of human coronaviruses. *Journal of advanced research*, 24, 91-98.

Salcedo, F., & Salcedo, G. (2021). Modelos predictivos de los contagios de la COVID-19 para la provincia de Loja-Ecuador. *Revista Digital Novasinergia*, 4(2), 62-77





Revisión de los mecanismos de QoS en las Instituciones Públicas del Ecuador

Review of QoS mechanisms in Public Institutions in Ecuador

Autores

- ✉ * **Teófilo Abdul Vera Peralta**
- ✉ **Andrea Katherine Alcívar Cedeño**
- ✉ **Tatiana Elizabeth Cobeña Macías**

¹ Facultad de Ciencias Informáticas, Universidad Técnica de Manabí, Portoviejo, Ecuador.

* Autor para correspondencia

Comó citar el artículo: Vera Peralta, T.A., Alcívar Cedeño, A.K. & Cobeña Macías, T.E. (2022). Revisión de los mecanismos de QoS en las instituciones públicas del Ecuador. *Informática y Sistemas: Revista de Tecnologías de la Informática y las Comunicaciones*, 6(1), 24-30. DOI: <https://doi.org/10.33936/isrtic.v6i1.4565>

Enviado: 21/02/2022
Aceptado: 09/03/2022
Publicado: 09/07/2022

Resumen



Hoy en día, las instituciones u organizaciones se han enfocado en obtener calidad de servicio(QoS) para el desarrollo de sus redes, satisfaciendo la demanda de tráfico de aplicaciones confidenciales, como voz y video en tiempo real, y así evitar la degradación de la calidad de servicio debido a la pérdida de paquetes, ancho de banda, retardo, jitter y otros. La QoS enfocada al conjunto de tecnologías en las instituciones ha permitido manejar los efectos de la congestión del tráfico usando óptimamente los diferentes recursos de la red en lugar de ir aumentando continuamente. Recuerde que la QoS no es aumentar ancho de banda sino distribuirlo de una mejor manera de acuerdo a las necesidades de la institución. En la actualidad, es imperioso tener o implementar Calidad de Servicio (QoS) en las instituciones para las diferentes actividades que desarrollan dentro de una red, y así de esta manera poder ofrecer servicios prioritarios tales como telefonía IP, transmisión de datos, envío de información en tiempo real, de acuerdo al manejo de los parámetros críticos que causan problemas de rendimiento en ellos. El objetivo de este estudio se centró en contestar las preguntas de investigación, realizando un filtro de acuerdo a lo más relevante encontradas en ellas previo su análisis metódico. Con la finalidad de cumplir con dicho objetivo se ha realizado una revisión sistemática de la literatura (SLR). Los resultados muestran información importante sobre la comparación de los diferentes mecanismos de QoS en las instituciones públicas.

Palabras clave: QoS, tráfico, delay, jitter, mecanismos

Abstract

Nowadays, institutions or organizations have focused on obtaining quality of service (QoS) for the development of their networks, satisfying the traffic demand of confidential applications, such as voice and video in real time, and thus avoid the degradation of the network. Quality of service due to packet loss, bandwidth, delay, jitter, and others. The QoS focused on the set of technologies in the institutions has made it possible to manage the effects of traffic congestion by optimally using the different network resources instead of continually increasing them. Remember that QoS is not about increasing bandwidth but distributing it in a better way according to the needs of the institution. Currently, it is imperative to have or implement Quality of Service (QoS) in institutions for the different activities that they develop within a network, and thus be able to offer priority services such as IP telephony, data transmission, sending information in real time, according to the management of critical parameters that cause performance problems in them. The objective of this study focused on answering the research questions, making a filter according to the most relevant found in them prior to their methodical analysis. In order to meet this objective, a systematic review of the literature (SLR) has been carried out. The results show important information about the comparison of the different QoS mechanisms in public institutions.

Keywords: QOS, traffic, delay, jitter, mechanisms.



1. Introducción

En la actualidad las redes informáticas se han convertido en una necesidad muy importante, tanto para las personas como para las empresas, ya que por medio de estas nos permite comunicarnos ya sea por motivos de trabajo o algo personal. El manejo de los sistemas informáticos hoy en día es uno de los elementos indispensables para la interacción y progreso en las instituciones.

La Ley Orgánica de Telecomunicaciones en su artículo 20 afirma: Las empresas públicas que presten servicios de telecomunicaciones y las personas naturales o jurídicas delegatarias para prestar tales servicios, deberán cumplir las obligaciones establecidas en esta Ley, su reglamento general y las normas emitidas por la Agencia de Regulación y Control de las Telecomunicaciones para garantizar la calidad, continuidad, eficacia, precios y tarifas equitativas y eficiencia de los servicios públicos (Barragán-Martínez & Guevara-Viejó, 2016). Con los servicios de telecomunicaciones se puede indicar que una principal parte de la satisfacción de los usuarios tendrá relación con el funcionamiento de la red, es por ellos que determinados trabajos con mecanismos de QoS orientan aspectos técnicos de la red.

(Juca Jara, 2016) Afirma: “Calidad de Servicio QoS es la habilidad de un elemento de red (una aplicación, un proveedor) para mantener un nivel de tratamiento del tráfico y del servicio satisfactorio”, los mecanismos de calidad de servicio (QoS - Quality of Service) implementados en cualquier tipo de red proporcionan niveles de funcionamientos adecuados, maximizando eficazmente el reparto del ancho de banda del enlace.

En la actualidad es muy necesaria la calidad de servicios (QoS) en las redes convergentes, por lo que en las instituciones actuales se maneja varios tipos de tráfico. (Rodríguez, 2017) menciona: “El impacto de QoS en redes ha sido analizado en otros trabajos relacionados en los que se han utilizado equipos reales en entornos de laboratorio o emulaciones con herramientas de software”. Según (Machini, 2020) la QoS nos proyecta a Garantizar la disponibilidad de la red de la organización y las aplicaciones que se ejecutan en esa red. Proporciona una transmisión de datos segura y eficiente a través de esta red.

El tráfico en la red está formado por diferentes flujos de datos. A partir de un flujo de información dado, los mecanismos de QoS proporcionan a la red de datos la capacidad de asegurar, con un grado de fiabilidad preestablecido, que se cumplan los requisitos de tráfico necesarios en términos de perfil y ancho de banda, con el fin de conseguir servicios útiles (Solís, 2018). En cada flujo, la transmisión de paquetes en una dirección tiene unas características significativas. Estas características pueden especificarse en términos de: caudal (throughput), retardo (delay), variación del retardo (jitter) y/o pérdidas. Los mecanismos de

QoS miden, mejoran y garantizan dichas características a un nivel determinado. Por tanto, los datos en las grandes redes necesitan control (Pacheco García, 2017).

El objetivo de esta revisión es indagar el tema central en investigaciones técnicas especializadas, realizando un filtro de acuerdo a lo más relevante encontradas en ellas previo su análisis metódico.

Este trabajo se encuentra organizado de la siguiente manera: En la sección 2 se explica la metodología y materiales aplicada para el desarrollo del trabajo; en la sección 3 se presenta los resultados, en la sección 4 se discute sobre los trabajos relacionados; y en la sección 5 se desarrollan las conclusiones.

2. Materiales y Métodos

2.1. Aplicación de la metodología mobile D

Para la presente investigación se emplea una revisión sistemática de la literatura (SLR) cualitativa, cuyo propósito es recopilar información más valiosa referente a los mecanismos de QoS, para lo cual se realizó una búsqueda de trabajos relacionados, los cuales aporten conocimientos relevantes para la investigación, a través de la discusión de referencias de alto nivel.

Las SLR se emplean para identificar, evaluar e interpretar los datos disponibles dentro de un periodo de tiempo de un determinado campo de investigación. El proceso de esta revisión está sustentado, en términos generales, en las directrices establecidas por (Brereton, 2007), enfocadas en la realización de SLR en ingeniería de software. Las tres fases de la revisión se muestran en la. Figura 1.

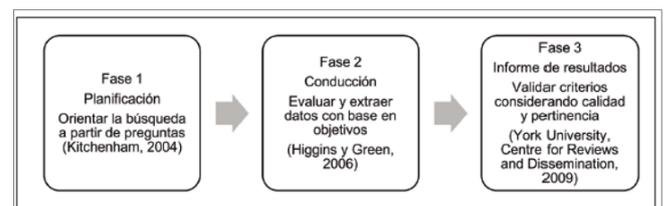


Figura 1. Proceso de revisión sistemática de literatura

Fuente: Los autores.

3. Resultados y Discusión

A continuación, se describen los pasos que se llevaron a cabo para cada fase y los resultados que fueron procediendo:

3.1. Fase 1: Planificación de la revisión.

La fase de la planificación consistió en realizar la estrategia para orientar las búsquedas en bases de datos, lo cual implicó partir

de preguntas que emanaron de las ausencias de conocimiento detectadas, palabras clave, elección de bases de datos especializadas y criterios de delimitación de las búsquedas. El objetivo de este estudio se centró en contestar las preguntas de investigación que se encuentran detalladas en la Tabla 1.

Tabla 1. Preguntas de investigación
 Fuente: Los autores.

Nº	Preguntas
RQ1	¿Cuántos estudios hay en las bases de datos Google Scholar, ACM Digital Library, Springer, LA Referencia, REDIB y Dialnet acerca de mecanismos, parámetros y modelos de QoS, desde enero de 2018 a enero de 2022?
RQ2	¿Cuáles son los principales parámetros de QoS en una red?
RQ3	¿Qué modelo (Best Effort, IntServ, Diff-Serv) han sido objeto de estudio para aplicar políticas de calidad de Servicio (QoS)?

Como protocolo para la revisión y las pautas de como seleccionar los estudios pertinentes se desarrolló de la siguiente manera:

Se seleccionaron los repositorios, los cuales fueron de la biblioteca digital como: Google Scholar, ACM Digital Library, Springer, LA Referencia, REDIB y Dialnet. Para la elaboración de la cadena de búsqueda, se realizó en los repositorios mencionados con trabajos relacionados con la investigación, posteriormente se efectuó la combinación de palabras para la conformación de la cadena de búsqueda más acertada, la Tabla 2 muestra las palabras clave que se utilizaron en los repositorios.

Tabla 2. Palabras claves de la búsqueda en los repositorios
 Fuente: Los autores

Keyword
Mecanismos de QoS
Parámetros de QoS
Modelos de QoS

Se empleó una búsqueda para escoger de estudios candidatos en tesis doctorales y de maestría, en portales de tecnología, así como en diferentes revistas con alto nivel de influencia en el mundo debido a sus inciviles artículos sobre todo en el área de ingeniería.

Para la investigación se consideró establecer un rango de búsqueda de los últimos cinco años en los diferentes repositorios o bases de datos antes mencionadas, rango establecido entre enero del 2018 y 2022, por lo que se consideró criterios de inclusión e exclusión. Los criterios que se utilizaron se alinean hacia el hallazgo de información relacionada a las QoS como se muestra en la Tabla 3.

Tabla 3. Criterios de inclusión y exclusión
 Fuente: Los autores

Inclusión	Exclusión
Artículos publicados dentro del rango de enero de 2018 a enero de 2022	Artículos publicados con más de 5 años de antigüedad.
Artículos en el campo de tecnología como estudios relacionados con los mecanismos, parámetros y modelos de QoS.	Artículos que no guarden relación con los mecanismos, parámetros y modelos de QoS
Tesis y artículos de alto nivel sobre los modelos de QoS.	Artículos duplicados

3.2. Fase 2: Conducción.

La etapa de conducción se dio para dar respuesta a la RQ1 ¿Cuántos estudios hay en las bases de datos Google Scholar, ACM Digital Library, Springer, LA Referencia, REDIB y Dialnet acerca de mecanismos, parámetros y modelos de QoS, desde enero de 2018 a enero de 2022? El proceso se dio a través de la revisión y extracción de los datos de los artículos. La primera búsqueda integró las palabras claves en las bases de datos como se muestra en la Tabla 4 ; en la segunda se seleccionaron únicamente por año, donde no pueden ser más de 5 años de antigüedad con eso se obtuvieron en cada repositorio los artículos con información actual como se puede observar en la Tabla 5, se procedió a descartar todos los artículos científicos duplicados y los que cumplan con el segundo criterio de exclusión, cuyo artículos que no guarden relación con los mecanismos, parámetros y modelos de QoS como se muestra en la Tabla 6. Al finalizar se obtuvo como resultado únicamente los artículos de selección como se puede apreciar en la columna Selección de la Tabla 7, artículos que contienen información referente a la QoS.

Tabla 4. Búsqueda de investigaciones con palabras claves
 Fuente: Los autores

Repositorio	Resultado de búsqueda de artículo		
	Mecanismo de QoS	Parámetros de QoS	Modelos de QoS
Google Scholar	21200	15800	15200
ACM Digital Library	14813	14623	15494
Springer	16	6	79
La Referencia	307	341	470
REDIB	15	34	28
Dialnet	80	87	169

En la Tabla 4 podemos observar el resultado de la primera búsqueda realizada con las palabras claves en los repositorios seleccionados. Con la palabra clave “mecanismo de QoS” en Google Scholar con 21200 artículo, ACM Digital Library 14813,



Springer 16, LA Referencia 307, REDIB 15 y Dialnet 80. Con la palabra clave “Parámetros de QoS” en Google Scholar con 15800 artículo, ACM Digital Library 14623, Springer 6, LA Referencia 341, REDIB 34 y Dialnet 87. Con la palabra clave “modelos de QoS” en Google Scholar con 15200 artículo, ACM Digital Library 15494, Springer 79, LA Referencia 470, REDIB 28 y Dialnet 169.

Tabla 5. Filtro de búsqueda dentro del rango de 5 años
Fuente: Los autores

Repositorio	Filtro (5 años)		
	keyword		
	Mecanismo de QoS	Parámetros de QoS	Modelos de QoS
Google Scholar	7800	6350	4240
ACM Digital Library	3618	3577	4002
Springer	5	2	62
La Referencia	74	90	155
REDIB	5	11	9
Dialnet	20	17	37

En la Tabla 5 podemos observar el resultado del filtro realizado en la búsqueda por año dentro del rango de enero de 2018 a enero de 2022. Con la palabra clave “mecanismo de QoS” en Google Scholar con 7800 artículo, ACM Digital Library 3618, Springer 5, LA Referencia 74, REDIB 5 y Dialnet 20. Con la palabra clave “Parámetros de QoS” en Google Scholar con 6350 artículo, ACM Digital Library 3577, Springer 2, LA Referencia 90, REDIB 11 y Dialnet 17. Con la palabra clave “modelos de QoS” en Google Scholar con 4240 artículo, ACM Digital Library 4002, Springer 62, LA Referencia 155, REDIB 9 y Dialnet 37.

En la Tabla 6 podemos observar el resultado de los artículos duplicados y los que no guarden relación con los temas de mecanismos, parámetros y modelos de QoS. Con la palabra clave “mecanismo de QoS” en Google Scholar con 34 artículo, ACM Digital Library 9, Springer 0, LA Referencia 5, REDIB 1 y Dialnet 4. Con la palabra clave “Parámetros de QoS” en Google Scholar con 18 artículo, ACM Digital Library 12, Springer 4, LA Referencia 4, REDIB 3 y Dialnet 5. Con la palabra clave “modelos de QoS” en Google Scholar con 23 artículo, ACM Digital Library 6, Springer 2, LA Referencia 3, REDIB 0 y Dialnet 0.

En la Tabla 7 se muestra el total de investigaciones seleccionadas en los repositorios con sus palabras claves, y el total general de todas las indagaciones estudiadas para dar respuestas a las preguntas de investigación que se presentan en la Tabla 1.

Tabla 6. Información duplicada y que no guarde relación con el tema
Fuente: Los autores

Repositorio	keyword		
	Mecanismo de QoS	Parámetros de QoS	Modelos de QoS
Google Scholar	34	18	23
ACM Digital Library	9	12	6
Springer	0	4	2
La Referencia	5	4	3
REDIB	1	3	0
Dialnet	4	5	0

Tabla 7. Total de artículos seleccionados
Fuente: Los autores

Repositorio	keyword		
	Mecanismo de QoS	Parámetros de QoS	Modelos de QoS
Google Scholar	34	18	23
ACM Digital Library	9	12	6
Springer	0	4	2
La Referencia	5	4	3
REDIB	1	3	0
Dialnet	4	5	0
Selección	4	5	0
Total			133

En la Tabla 8 podemos observar los datos obtenidos por año dentro del rango de enero de 2018 a enero 2022 que cubrieron los criterios delimitados para la investigación.

3.3. Fase 3: Informe de los resultados.

Finalmente, para generar el reporte de RQ2 ¿Cuáles son los principales parámetros de QoS en una red? y RQ3 ¿Qué modelo (Best Effort, IntServ, DiffServ) han sido objeto de estudio para aplicar políticas de calidad de servicio (QoS)?, se realizó un proceso de validación donde se usaron los criterios de inclusión y exclusión de la revisión, que se muestran en la Tabla 3.

En la pregunta RQ2 ¿Cuáles son los principales parámetros de QoS en una red? se realizó un análisis de contenido para identificar los parámetros y se diagramó en Power BI, por la cantidad total de artículos de estudio seleccionados dentro del rango de enero de 2018 a enero de 2022, que a su vez, fueron organizados por parámetros como se muestra en la Figura 2.

Tabla 8. Estudios encontrados por año (enero 2018 a enero 2022)

Fuente: Los autores

Año	Mecanismo de QoS	Parámetros de QoS	Modelos de QoS	Total
Google Scholar	34	18	23	
ACM Digital Library	9	12	6	
Springer	0	4	2	
La Referencia	5	4	3	
REDIB	1	3	0	
Dialnet	4	5	0	
Selección	4	5	0	
Total	53	46	34	133

En la Figura 2 se identifica por color los parámetros de QoS de las 46 investigaciones que fueron seleccionadas para el estudio donde hablan de los parámetros de QoS, muestra:

Nueve (9) documentos de investigación hablaron de pérdida de paquetes (Pino), (Espinoza Obando, 2021), (Solís, 2018), (García Pardo, 2009), (Onofre Caicedo, 2018), (López Casa, 2016), (Jiménez Velásquez, 2019), (Felipe, 2020), (Sabloak, 2018)).

Diez (10) documentos de investigación hablaron de ancho de banda (Niño Sepúlveda), (Criollo Bustamante, 2020), (Vargas Rodríguez), (Muñoz, 2016), (Sabloak, 2018), (Vargas Rodríguez), (Ariza Florez), (Villamarín, 2022), (Ortega, 2019), (Cayetano Mancha, 2018)).

Ocho (8) documentos de investigación hablaron de jitter (Espinoza Obando, 2021), (Chávez Zambrano, 2016) (Ali & Roh, 2020), (Sabloak, 2018), (Ortega, 2019), (Cayetano Mancha, 2018), (Molina Merchán, 2018), (Sobreviela Blasco, 2019)).

Ocho (8) documentos de investigación hablaron de latencia (Vargas Rodríguez), (Pibaque Villacrese, 2019), (Pino), (Cayetano Mancha, 2018), (Villamarín, 2022), (Ortega, 2019), (Molina Merchán, 2018), (Sobreviela Blasco, 2019)).

Cinco (5) documentos de investigación hablaron de retardo (Cacuango Lagla, 2019), (Martínez Fernández, 2022), (Ortega, 2019), (Castro Sandoval & Orjuela Gutierrez), (Molina Merchán, 2018)).

Seis (6) documentos de investigación hablaron de throughput (Vargas Rodríguez), (Azamuddin, 2020.), (Ortega, 2019),

(Cayetano Mancha, 2018), (Molina Merchán, 2018), (Sobreviela Blasco, 2019).))

En la pregunta RQ3 ¿Qué modelo (Best Effort, IntServ, DiffServ) han sido objeto de estudio para aplicar políticas de calidad de servicio (QoS)?, se realizó un análisis de contenido y se delimitaron palabras clave de los modelos de calidad de servicio.

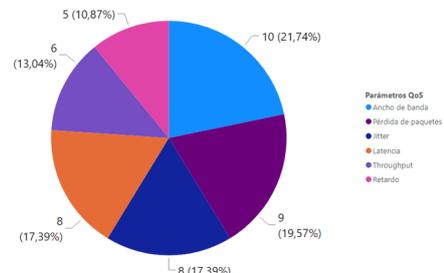


Figura 2. Parámetros de QoS

Fuente: Los autores

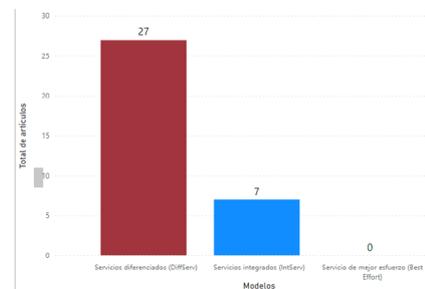


Figura 3. Modelos de políticas de QoS

Fuente: Los autores.

Estos datos fueron cruzados con los contextos identificados y se diagramó en Power BI, como se muestra en la Figura 3.

En la Figura 3 según las investigaciones seleccionadas sobre los tres modelos de políticas de calidad de servicio (QoS), servicio de mejor esfuerzo (Best Effort), servicios integrados (IntServ) y servicios diferenciados (DiffServ), muestran una mayor incidencia en el modelo DiffServ con un total de 27 archivos de investigación, dando a entender que el que más se utiliza, dejando a tras al IntServ con 7 artículos y a Best Effort con ninguno.

Discusión

La calidad de servicio está dando resultados favorables en todas las organizaciones, según lo analizado en los trabajos consultados para la presente investigación.

La ITU-T define diferentes parámetros de calidad de servicio (QoS) que son la base para definir los requerimientos de las distintas aplicaciones, estos parámetros varían de tráfico en tráfico y de cliente en cliente según los requerimientos y los aspectos técnicos de la red. Los parámetros que se mencionan en el estudio se pueden utilizar para los diferentes tipos de

especificaciones y hasta para evaluaciones de calidad de servicio en las instituciones.

La IETF ha propuesto modelos que ofrecen calidad de servicio (QoS) en diferentes modalidades como el modelo IntServ que pretendía ampliar la arquitectura IP existente para soportar actividades en tiempo real, el servicio Best-Effort que existía hasta el momento y el modelo Diffserv que es una de las propuestas más alentadoras para redes grandes como por ejemplo el Internet, si hoy en día se aplica políticas de calidad de servicio en una red, según el estudio realizado lo recomendable es trabajar con el modelo de Diffserv por su flexibilidad ofreciendo distintos niveles de calidad de servicio para diferentes clases.

La QoS abre la puerta como su nombre lo indica a entregar una mejor calidad de servicio dando una experiencia al cliente final inigualable y si está bien aplicado serás destacado sobre los demás por el buen rendimiento de su red.

4. Conclusiones

Los mecanismos de calidad de servicio (QoS) aplicadas a las instituciones públicas son de vital importancia, ya que permite seguridad, confiabilidad, integridad de datos y un escalamiento creciente de la red por su fácil adaptación a cualquier tecnología.

La revisión sistemática de la literatura permitió identificar algunos estudios que comprueban la posibilidad de aplicar las diferentes técnicas y parámetros que deben tomarse en cuenta al momento de construir un SLR.

Será necesario realizar un análisis más profundo a los mecanismos de calidad de servicio (QoS), con el fin de reconocer las técnicas y metodologías que podrían servir de base para continuar esta investigación.

Con los resultados obtenidos, se espera continuar ampliando información en investigaciones futuras, dado que el campo de la calidad de servicio (QoS) avanza cada vez más, aún hay mucho que conocer sobre este tema, sobre su adaptabilidad en la difusión de información, a las nuevas exigencias que se van presentando. Por el momento esperamos que la información contenida en el presente artículo, pueda ser de utilidad a investigaciones que requieran información relevante a los mecanismos de QoS.

Contribución de los autores

Teófilo Abdul Vera Peralta: Conceptualización, Metodología, Software, Análisis formal, Redacción – borrador original del artículo. **Andrea Katherine Alcívar Cedeño:** Supervisión, Redacción – revisión y edición del artículo. **Tatiana Elizabeth Cobeña Macías:** Supervisión, Redacción – revisión y edición del artículo.

Conflictos de interés

Los autores declaran no tener ningún conflicto de interés.

Referencias bibliográficas

- Ali, J., & Roh, B. H. (2020). An effective hierarchical control plane for software-defined networks leveraging TOPSIS for end-to-end QoS class-mapping. *Ieee Access*, 8, 88990-89006.
- Ariza Florez, E. M. (s.f.). Ajuste de políticas de QoS para plataformas de voz, video y aplicaciones compartidas del Banco de Occidente.
- Azamuddin, W. M.-K. (2020.). Quality of service (Qos) management for local area network (LAN) using traffic policy technique to secure congestion. *Computers*, 9(2), 39.
- Barragán-Martínez, X., & Guevara-Viejó, F. (2016). El gobierno electrónico en Ecuador. *Revista Ciencia UNEMI*, 9(19), 110-127.
- Brereton, P. K. (2007). Lessons from applying the systematic literature review process within the software engineering domain. *Journal of systems and software*, 80(4), 571-583.
- Cacuango Lagla, S. E. (2019). Evaluación de una Red LAN para el establecimiento de las Políticas de la Calidad de Servicio. *Master's thesis, Quito*. <http://repositorio.uisrael.edu.ec/handle/47000/2021>
- Castro Sandoval, W., & Orjuela Gutierrez, C. D. (s.f.). Estudio para mejorar el comportamiento de una red de datos WAN en ambiente productivo. <http://hdl.handle.net/11349/7811>
- Cayetano Mancha, X. M. (2018). Modelo de Intranet para Mejorar la Calidad de Servicio (Qos) de la Red de Datos en la Ie Jec Tupac Amaru Lircay-Huancavelica. <http://repositorio.unh.edu.pe/handle/UNH/2620>
- Chávez Zambrano, G. K. (2016). Propuesta de red de datos para la gestión de los servicios de red en el campus politécnico de la Espam MFL. *Bachelor's thesis Calceta: Espam*. <http://repositorio.esпам.edu.ec/handle/42000/319>
- Criollo Bustamante, M. D. (2020). Políticas de QoS en redes empresariales para el análisis de rendimiento, en entornos convencionales y SDN. <http://dspace.espace.edu.ec/handle/123456789/14090>
- Espinoza Obando, J. R. (2021). Implementación de un prototipo de telefonía IP con funcionalidades extendidas para el

- Sector hotelero aplicando políticas de QoS. *Doctoral dissertation, Universidad de Guayaquil. Facultad de Ciencias Matemáticas y Físicas. Carrera de Ingeniería en Networking y Telecomunicaciones Guayaquil*. <http://repositorio.ug.edu.ec/handle/redug/52273>
- Felipe, M. D. (2020). Evaluación de soluciones de QoS para una Red de Área Local. *Revista Ibérica de Sistemas e Tecnologías de Informação, (E29), 27-40*.
- García Pardo, S. M. (2009). Implementación de mecanismos de calidad de servicio con equipos Nortel. <http://hdl.handle.net/10317/1901>
- Jiménez Velásquez, Á. L. (2019). Implementación y evaluación de una estrategia para garantizar mantenimiento de QoS en la transmisión de video en tiempo real en redes WLAN bajo el esquema de gestión SDN. <https://hdl.handle.net/10495/13873>
- Juca Jara, P. S. (2016). Estudio de la implementación de Calidad de Servicio (QoS) para el mejoramiento de la red de datos que optimice el acceso a los servicios en la Planta de Producción de la Compañía Yanbal Ecuador SA. *B.S. thesis. PUCE*. <http://repositorio.puce.edu.ec/handle/22000/12463>
- López Casa, B. A. (2016). Diseño e implementación de un prototipo de calidad de servicio (QoS) sobre IPV6. *Bachelor's thesis*. <http://dspace.ups.edu.ec/handle/123456789/12395>
- Machini, A. E. (2020). Nexo: Una herramienta para la visualización y análisis de indicadores QoS y QoE móviles. *Informes Científicos Técnicos-UNPA, 12(2), 47-62*. <https://doi.org/10.22305/ict-unpa.v12.n2.731>
- Martínez Fernández, A. (2022). Diseño de una red para analizar el efecto de los tipos de colas. *Doctoral dissertation, Universitat Politècnica de València*.
- Molina Merchán, J. (2018). Medidas de calidad para la transmisión de vídeo en entornos con restricciones de baja latencia. <http://hdl.handle.net/10486/685319>
- Muñoz, L. F. (2016). Regiones de eficiencia espectral asociadas a satisfacción de QoS basadas en estrategias de ancho de banda. *Ingeniería y Ciencia, 12(24), 149-168*.
- Niño Sepúlveda, Y. R. (s.f.). Configuración IPTV y QoS. <https://repository.unad.edu.co/handle/10596/44546>
- Onofre Caicedo, J. G. (2018). Análisis de la calidad del servicio (QoS) en la red de internet de la ciudadela el Chorrillo en la ciudad de Babahoyo para una empresa de telecomunicaciones. (*Doctoral dissertation, Universidad de Guayaquil. Facultad de Ciencias Matemáticas y Físicas. Carrera de Ingeniería En Networking y Telecomunicaciones*). <http://repositorio.ug.edu.ec/handle/redug/32544>
- Ortega, J. (2019). Diseño e Implementación de un sistema de control y balanceo de carga, en Routers MIKROTIK para mejorar la calidad de servicio (QoS) de la empresa Zona VIP, ubicada en el distrito de Amarilis, Provincia de Huánuco 2015.
- Pacheco García, D. (2017). Propuesta de prácticas de laboratorio de QoS en VoWiFi. (*Doctoral dissertation, Universidad Central "Marta Abreu" de Las Villas, Facultad de Ingeniería Eléctrica, Departamento de Electrónica y Telecomunicaciones*).
- Pibaque Villacrese, M. J. (2019). Red de Datos con QoS y Balanceo de datos con QoS y Balanceo de carga mediante la Tecnología MIKROTIK que comunique las dependencias fuera DELGAD del Cantón Jipijapa. (*Bachelor's thesis, JIPIJAPA-UNESUM*). <http://repositorio.unesum.edu.ec/handle/53000/1573>
- Pino, F. A. (s.f.). Implementación de calidad de servicio (QoS) en redes tácticas de gran unidad. <https://zaguan.unizar.es/record/96790?ln=es#>
- Rodríguez, M. Z. (2017). Evaluación de parámetros de QoS en una red VPN-MPLS Diffserv bajo un entorno completo de emulación de software libre. *Revista Científica y Tecnológica UPSE, 4(3), 74-82*.
- Sabloak, S. W. (2018). Analisis Pemantauan LAN Menggunakan Metode QoS dan Pengklasifikasian Status Jaringan Internet Menggunakan Algoritma Naive Bayes. *Jurnal Ilmiah Teknologi Infomasi Terapan, 4(2)*.
- Sobreviela Blasco, L. M. (2019).). Calidad de servicio (QoS) con routers Cisco. *Doctoral dissertation, Universitat Politècnica de València*.
- Solís, R. A. (2018). Calidad de servicios en redes WLAN utilizando acceso al medio basado en tiempos de contención. *Telemática, 17(3), 55-64*.
- Vargas Rodríguez, M. A. (s.f.). Lineamiento técnico para la provisión de Calidad de Servicio (QoS) extremo a extremo en interconexión bajo IP Multimedia Subsystem (IMS). (*Doctoral dissertation, Universidad Nacional de Colombia*).
- Villamarín, E. J.-T. (2022). Diseño de redes para Instituciones Académicas con criterios de QoS. *Revista Ibérica de Sistemas e Tecnologías de Informação, (E47), 170-183*.





Identificación de temas del blog de noticias en idioma español

Topic identification from news blog in Spanish language

Autores

✉ *Lizbeth Pacheco-Guevara*



✉ **Ruth Reátegui*



✉ *Priscila Valdiviezo-Díaz*



Resumen

Actualmente existe una gran cantidad de noticias en formato digital que necesitan ser clasificadas o etiquetadas automáticamente según su contenido. La Asignación Latente de Dirichlet (LDA por sus siglas en inglés) es una técnica no supervisada que crea automáticamente tópicos a partir de las palabras de los documentos. El presente trabajo tiene como objetivo aplicar LDA para analizar y extraer temas de noticias digitales en español. Se recogieron un total de 198 noticias digitales de un blog de noticias universitario. Se realizó un preprocesamiento de los datos y su representación en espacios vectoriales y se seleccionaron los valores de k en base a la métrica de coherencia. Con una matriz frecuencia de términos – frecuencia inversa del documento (TF_IDF por sus siglas en inglés) y una combinación de unigramas y bigramas se obtienen temas con una variedad de términos y tópicos relacionados con actividades universitarias como programas de estudio, investigación, proyectos de innovación y responsabilidad social. Además, con el proceso de validación manual, los términos de los temas se corresponden con los hashtags escritos por los profesionales de la comunicación.

Palabras clave: LDA, modelado de tópicos, noticias, blog.

¹ Universidad Técnica Particular de Loja,
Loja, Ecuador.

* Autor para correspondencia

Abstract

Currently, exist a large amount of news in a digital format needs to be classified or labeled automatically according to its content. Latent Dirichlet Allocation (LDA) is an unsupervised technique that automatically creates topics based on words in documents. The present work aims to apply LDA in order to analyze and extract topics from digital news in the Spanish language. A total of 198 digital news were collected from a university news blog. A data pre-processing and representation in vector spaces were carried out and k values were selected based on the coherence metric. A term frequency – inverse document frequency (TF_IDF) matrix and a combination of unigrams and bigrams produce topics with a variety of terms and topics related to university activities like study programs, research, projects for innovation, and social responsibility. Furthermore, with the manual validation process, terms in topics correspond with hashtags written by the communication professionals.

Keywords: LDA; topic modeling; news, blog.

Comó citar el artículo: Pacheco-Guevara, L., Reátegui, R. y Valdiviezo-Díaz, P. (2022). Topic identification from news blog in Spanish language. *Informática y Sistemas: Revista de Tecnologías de la Informática y las Comunicaciones*. 6, (1), 31–37. DOI:<https://doi.org/10.33936/isrtic.v6i1.4514>

Enviado: 18/03/2022

Aceptado: 24/04/2022

Publicado: 27/05/2022



1. Introduction

Currently it exists a large amount of news in a digital format. Text mining and natural language processing (NLP) allow us for transforming and analyzing unstructured information such as digital news. Topic modeling extracts the main topics from a collection of documents; this task can be done with the Latent Dirichlet Allocation (LDA) model.

LDA is a generative probabilistic model where documents are represented as random mixture over latent topics, and each topic is represented by a distribution of words (Blei, Ng, & Jordan, 2003). LDA is based on three levels: words, topics and documents.

In order to identify topics from news, some researches have been done. In (Xu, Meng, Chen, Qiu, Wang, & Yao, 2019) the authors proposed a method to analyze the evolution of news topics over time. They applied LDA method to extract topics from news, with LDA they obtained more reliable results than K-means clustering. Moreover, Misztal-Radecka (2018) worked in create user profiles based on news articles. This author used LDA and Word2Vec to represent information and demonstrated that Word2Vec works better in comparing small text like titles. However, for a classification task, a combination of topic identification and word embedding had a better performance.

Furthermore, in (Larsen & Thorsrud, 2019) the authors applied LDA models to analyze the role of business newspaper in predicting and explaining economic fluctuations. Moreover, Oliveira Capela & Ramirez-Marquez (2019) proposed a method to detect and to compare identities of cities using news articles from thirty-six cities from the United States of America. They used LDA to group words in topics. Also, Xu, Yeo, Hwang & Kim (2020) used LDA and network analysis to explore social media discourse on hyper-connected society base on news articles (Xu, Yeo, Hwang & Kim, 2020).

In addition, in (Guangce & Lei, 2021) the authors used LDA and the Apriori algorithm to improve media monitoring reports based on news articles. LDA allowed to identify topics, meanwhile Apriori allowed to discover relationships of topics of words. A recent work, focused on COVID-19, applied LDA in traditional and social media news to characterize and to compare early coverage of COVID-19 (Chipidza, Akbaripourdibazar, Gwanzura & Gatto, 2021).

Considering the importance of topics identification from unstructured documents, the present work aims to automatically classify and to assign topics to a corpus of news obtained from a university in Ecuador.

This paper is organized as follows: the section 2 presents the methodology for topic modeling; section 3 presents results and discussion, and finally, conclusions in section 4.

2. Methodology

In the reviewed literature, it was not found a predetermined methodology for topic modeling; in general, it is defined according to the type of data to analyze; however, some important steps mentioned in some related works were taken. Figure 1 shows five important steps mentioned in some related works to obtain a LDA model.

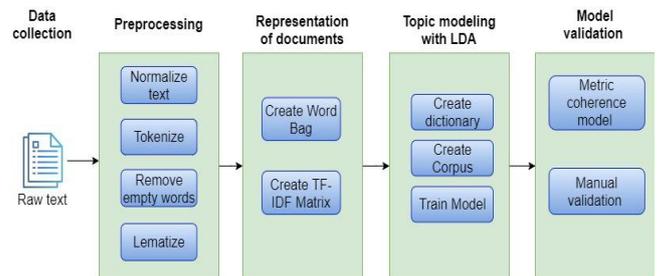


Figure 1. Flow for obtaining a LDA model.
Source: Authors

2.1. Data collection

In this step the data is gathering. In our case the data consists of digital news obtained from a news blog of a university in Ecuador. The university considered as a case of study has a news blog with updated information on the most relevant events of different academic activities that take place inside and outside of the institution. The 200 links were collected from October 2019 to May 2020, which are initially stored in a xml format. Next, the links were analyzed to verify they are working correctly and to eliminate duplicate information. Finally, 198 news were stored in a txt format to be analyzed by LDA. Figure 2 shows the news's HTML structure, it consists of a title and several paragraphs that make up the body of the news. It was verified that all news in the corpus have the HTML structure.



Figure 2. General structure of the news blog.
Source: Authors

2.2. Preprocessing

The preprocessing ensures data quality that influences LDA performance. Figure 3 shows the steps following in the preprocessing task: text normalization, tokenization, elimination of stop words and lemmatization.

Text normalization: It consists of removing numbers, alphanumeric values, special characters and punctuation marks (*, /, %, \$, #, “”, @, °), among others. Also, in our cases letters were transformed from uppercase to lowercase and dates were eliminated. Denny and Spirling (2017) mention that numbers have relevance according to the domain of analysis. In this study numbers are not relevant because the objective is to identify topics that could show the main activities of the university in its different departments. Also, words in topics could help to select hashtags for each news, and as we can see in our dataset number are not used as hashtags, unless they are a compound word like COVID-19.

Tokenization: This step is used to separate words contained in each news.

Elimination of Stop Words: This step is used to select the most important words in the news. We used the list of stop words defined in Python spaCy library for the Spanish language.

Lemmatization: Our corpus is in the Spanish language; therefore, when processing text in Spanish it is necessary to rely on libraries with pre-trained statistical models for this language. In this work, the `es_core_news_sm` model was used, which basically performs multitask work in Spanish trained in UD Spanish AnCora and WikiNER. It assigns token vectors, POS tags, dependency parsing, and context-specific named entities (Kim, Seo, Cho, & Kang, 2019). In addition, we used the `token.lemma` tag to obtain the word in its base form and the `token_pos` tag to return the word lemmatized by nouns, verbs, and adjectives.

2.3. Document representation

Additionally, Bags of words and Inverse document frequency matrix are two of the most widely used methods to represent documents with matrices.

Bag of words (BoW) counts the occurrence of each word and annotates them into a vector. One of the first tasks for the preparation of the model is the vectorization of the corpus.

The Gensim function called `doc2bow`, document to document representation, was used to transform words within news into a tuple for numerical representation (Oliveira Capela & Ramirez-Marquez, 2019). To create the dictionary, we applied the filters and default values `no_below = 5`, which keeps the tokens contained in at least 5 documents, and `no_above = 0.5`, which keeps the tokens above 0.5 of the total size of the corpus. Finally, the dictionary consists of 1861 words.

Inverse document frequency matrix (TF-IDF) that gives each word a weight according to the frequency in a document (TF), and the rarity of its occurrence in the corpus (IDF). In our case, terms with a scarcity greater than 0.99 and terms with a very low frequency were eliminated. Through inverse term-frequency analysis of documents, the most relevant terms within the corpus were identified.

Figure 4 shows an example of the result of the two matrices.

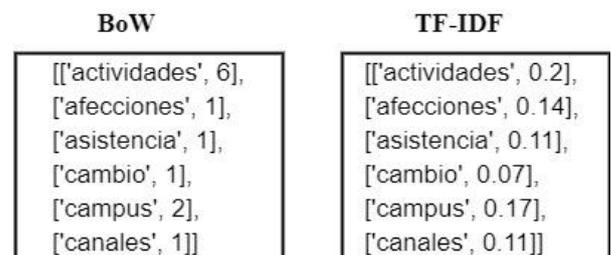


Figure 4. BoW and TF-IDF representation
 Source: Authors

2.4. Topic Modelling

The topic modeling step with LDA requires inputs for specific hyper-parameters such as α (alpha), β (beta), and the number of topics (k). Alpha refers to the distribution of documents per theme and beta refers to the distribution of words per topic (Blei, Ng, & Jordan, 2003, Xu, Meng, Chen, Qiu, Wang, & Yao, 2019).

An exhaustive search for the number of topics k was performed and the hyperparameter α was optimized for the LDA algorithm. The number of topics was explored from $k = 2$ to $k = 40$.

2.5. Model Validation

The main challenge when modeling topics with LDA is to define the appropriate number of topics (k) to represent the whole corpus and the interpretability of the topics (Misztal-Radecka, 2018, Wang, Feng & Dai, 2018). Therefore, the coherence metric developed by Mimno & Wallach (2011) was applied to validate

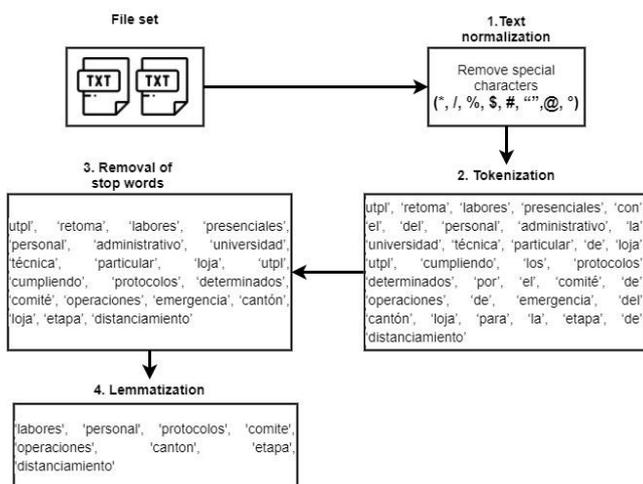


Figure 3. Preprocessing steps for Spanish news documents.
 Source: Authors

the tuples (k, α) .

The coherence metric takes into account the co-occurrence statistics of words collected from the corpus. It calculates the benchmark score in a range from 0 to 1, higher values represent better subject quality (Oliveira Capela & Ramirez-Marquez, 2019).

Alpha values distributed between 0 and 1 were tested. A text is considered coherent if all or most of the main words of the text are related to each other (Buenano-Fernandez, Gonzalez, Gil & Lujan-Mora, 2020). Figure 5 shows the results of the coherence metric for different numbers of topics. According to the behavior of the coherence values, the maximum value occurs for $k = 10$, $\alpha = 0.01$ and $\beta = 0.1$ with a coherence of 0.35, the model was carried out with these parameters.

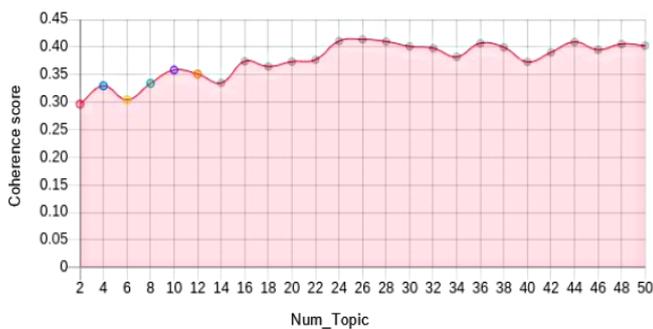


Figure 5. Hyper-parameter optimization for the number of topics k and LDA parameter α .
 Source: Authors

Additionally, most of the news has some hashtags written by communication professionals. Considering these hashtags, we made a manual validation of the result through a comparison between the hashtags and terms of the topic assigned to each news. In the next section, we show the topics identified and the comparison as a manual validation.

3. Results and Discussion

Once the reference values for k were obtained, some experiments were made considering the BoW and TF-IDF matrixes and different k values from 8 to 12. Matrixes with TF-IDF gave us the best coherence results; therefore, we decide to explore a visual representation of the topics using the pyLDAvis with the following TF-IDF matrixes:

- Matrix TF-IDF, $K= 8$, with unigrams and bigrams
- Matrix TF-IDF, $K= 10$, with unigrams and bigrams

The pyLDAvis package represents topics by circles in 4 quadrants.

As long as the circles are more dispersed between the quadrants and between each other, we have a good model. Also, this graph allows the option to visualize the terms contained in a topic, this appears on the right side in red color, and the terms are ordered based on their relevance. Figures 6 and 7 shows the pyLDAvis results, when $k=8$ exists some topics overlapped, while when $k=10$ presents a good distribution of topics.

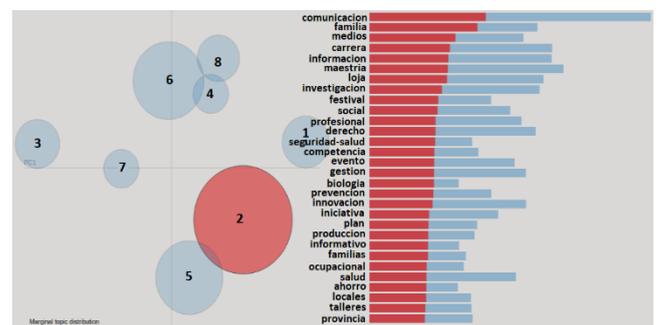


Figure 6. pyLDAvis graph with TF-IDF and $k=8$
 Source: Authors

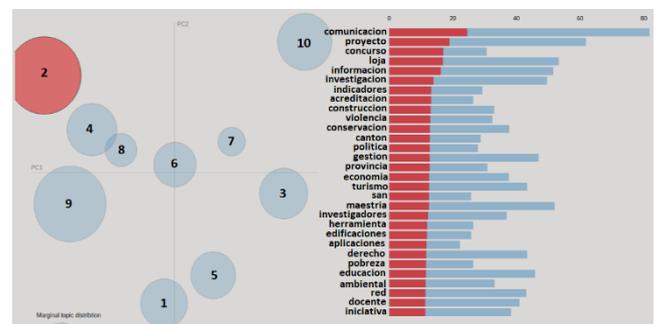


Figure 7. pyLDAvis graph with TF-IDF and $k=10$
 Source: Author

Figure 7 shows the predominant topic 2 with terms: “comunicacion”, “proyecto”, “concurso”, “loja”, “informacion”, “investigacion”, “indicadores”, “acreditacion”, “construccion”, “violencia”; followed by topic 9 with the terms: “proyecto”, “salud-mental”, “cultural”, “conservacion”, “comunicacion”, “gastronomia”, “especies”, “museo”, “mujer”, “virtual”, and topic 3 with the terms: “derecho-penal”, “radio”, “covid”, “prendho”, “crisis”, “pandemia”, “personal”, “familia”, “graduados” and “instituto”.

Considering that the value of $k=10$ has the best topics distribution, next, we will present in Table 1 the topics obtained with TF-IDF matrix when $k= 10$. Topics listed from 0 to 9 correspond with topics 1 to 10 in pyLDAvis Figure 7. The number before each term represents the probability of the term belong to the topic.

All the texts analyzed were in a university environment; therefore, the topics are related to the activities developed in an academic field. Topics 5, 7, and 9 refers to different carriers that the UTPL offers. Topic 4 refers to master programs, topic 6 mentions two modalities of study such as open university and continuing education. Topic 3 refers to a family event as well as studies and researches. Topics 0, 1, 2 y 8 presents words related to studies, projects and researches developed by the UTPL.

In Table 2, we verified that the topics assigned by the model are related to the text within the news. This table shows some news of the corpus, columns Num_News represents the news number,

Dominant_Topic refers to the most relevant topic assigned to news, Term_in_Topics represents the terms in the dominant topic, and Terms_in_New presents some terms within the news.

As an example, news number 85 and 119 refer to “concurso”, news 187 refers to “construccion”. Also, news 13 refers to “conservación”, news 7 refers to “gastronomia”, and news 15 and 131 refers to “cultural”.

In general, after analyzing the results, the TF-IDF topics include a variety of terms facilitating and making a distinction among topics.

Table 1. Topics obtained with the TF-IDF matrix with k=10
Source: Authors

Topics	Terms
Topic 0	0.003**paz” + 0.002**campus” + 0.002**comunicacion” + 0.002**violencia” + 0.002**educacion-superior” + 0.002**ninios” + 0.002**investigacion” + 0.002**red” + 0.002**mejora” + 0.002**educacion”
Topic 1	0.003**comunicacion” + 0.003**proyecto” + 0.002**concurso” + 0.002**loja” + 0.002**informacion” + 0.002**investigacion” + 0.002**indicadores” + 0.002**acreditacion” + 0.002**construccion” + 0.002**violencia”
Topic 2	0.002**estudios” + 0.002**mujer” + 0.002**educacion” + 0.002**universidad” + 0.002**feria” + 0.002**innovacion” + 0.002**prueba” + 0.002**museo” + 0.002**sur” + 0.002**software”
Topic 3	0.005**familia” + 0.003**comunicacion” + 0.002**radio” + 0.002**congreso” + 0.002**ods” + 0.002**educativa” + 0.002**sector” + 0.002**medios” + 0.002**pobreza” + 0.002**cambios”
Topic 4	0.004**maestria” + 0.003**festival” + 0.003**suelo” + 0.003**comunicacion” + 0.003**paz” + 0.002**mencion” + 0.002**informacion” + 0.002**responsabilidad-social” + 0.002**laboratorio” + 0.002**equipo”
Topic 5	0.003**derecho” + 0.003**mujeres” + 0.003**ods” + 0.002**artes” + 0.002**aplicacion” + 0.002**mayo” + 0.002**loja” + 0.002**muerte” + 0.002**donaciones” + 0.002**carrera”
Topic 6	0.003**distancia” + 0.003**ambiental” + 0.002**virtual” + 0.002**secretario” + 0.002**carreras” + 0.002**ayudar” + 0.002**ideas” + 0.002**educacion-continua” + 0.002**seminario” + 0.002**dolor”
Topic 7	0.003**turismo” + 0.003**personas-discapacidad” + 0.003**comunicacion” + 0.002**social” + 0.002**psicologia” + 0.002**convenio” + 0.002**maestria” + 0.002**libros” + 0.002**experiencia” + 0.002**reconocimiento”
Topic 8	0.003**proyecto” + 0.003**salud-mental” + 0.003**cultural” + 0.003**conservacion” + 0.003**comunicacion” + 0.002**gastronomia” + 0.002**especies” + 0.002**museo” + 0.002**mujer” + 0.002**virtual”
Topic 9	0.004**derecho-penal” + 0.003**radio” + 0.003**covid” + 0.003**prendho” + 0.002**crisis” + 0.002**pandemia” + 0.002**personal” + 0.002**familia” + 0.002**graduados” + 0.002**instituto”

2.1. Manual validation

Coherence value is a metric that allow us to validate the model, but considering that the university news blog has hashtags placed at the end of each news, we decided to make a comparison of the hashtags and the topics terms obtained in each news. Table 3 shows this comparison, where the coincidences are highlighted in two colors: pink color for exact terms and green color for similar terms. For example, the relevant topics for news number 1 are 9, 0, and 1; in topic 9 the terms “covid” and “pandemia” are related to “semaforo amarillo”. In Ecuador due to the covid-19 pandemic, a traffic light system advises about the level of

contagion. In topic 0 the term “educacion” appears highlighted in pink because it coincides in both columns. In news number 6 the relevant topics are 5 and 1, in this case, the word “donacion” are highlighted with pink, and the words “loja” and “utpl” with green because the principal campus of the university analyzed is in Loja. In this way, it can be concluded that the topics identified with LDA are quite good since most of them are related to the hashtags and the words within the news. We must to remark that all the news did not have hashtags; therefore, we did not calculate a numeric representation of the coincidence between hashtags and the terms of the topics.



Table 2. Examples of topics assigned to documents
Source: Authors

Num_ News	Dom - inant_ Topic	Terms_in_Topics	Terms_in_News
85	1	comunicacion, indicadores, informacion, acreditacion, loja, concurso, proyecto, violencia, investigacion, construccion	utpl, destaca, concurso, arquitectura, urbanistica, quitoel, noviembre, teatro, sucre, quito, sede, premiacion, finalistas, concurso, arquitectura, diseno, plan, especial, ...
119	1	comunicacion, indicadores, informacion, acreditacion, loja, concurso, proyecto, violencia, investigacion, construccion	estudiantes, utpl, obtienen, competencia, nacional, pasteleriaen, competencia, nacional, pasteleria, expo, sweet, desarrollo, centro-convenciones, metropolitano, quito, rosalia-arteaga, veronica, abad, estudiantes, carrera,...
187	1	comunicacion, indicadores, informacion, acreditacion, loja, concurso, proyecto, violencia, investigacion, construccion	constructores, fortalecen, conocimientos, edificaciones, lojapara, garantizar, calidad, construccion, obras, civiles, loja, universidad-tecnica, particular-loja, utpl,...
13	8	comunicacion, especies, virtual, salud-mental, proyecto, conservacion, museo, cultural, gastronomia, mujer	turismo, aporta, conservacion, medioambienteel, turismo, industria, peso, economia, mundial, convertido, sector, crecimiento, mundo, emplea, millones, personas, nivel, global, datos, organizacion-mundial, ...
7	8	comunicacion, especies, virtual, salud-mental, proyecto, conservacion, museo, cultural, gastronomia, mujer	copa, culinaria, utpl, reto, expone, talento, juvenesla, gastronomia, profesion, innovadora, emplea, atributos, creatividad, destrezas, motrices, elaborar, postre, sofisticado, ...
15	8	comunicacion, especies, virtual, salud-mental, proyecto, conservacion, museo, cultural, gastronomia, mujer	estudiantes, utpl, innovan, museo-matilde, hidalgo, lojael, impacto, figura, ecuatoriana, matilde-hidalgo, mujer, latinoamerica, votar, eleccion, nacional, motivo, ...
131	8	comunicacion, especies, virtual, salud-mental, proyecto, conservacion, museo, cultural, gastronomia, mujer	conferencias, virtuales, arte, cultura, desarrollan, difundir, proyectos, estrategias, fortalezcan, sector, cultural, ecuador, ...

4. Conclusions

When analyzing the corpus of digital news with the LDA model, a better topic coherence was obtained using the TF-IDF matrix constructed with unigrams, bigrams, words lemmatized by nouns, adjectives and verbs, and with a value of $k=10$. Considering the news correspond to an academic environment, the topics obtained refer to a greater visibility of the work and activities carried out by the communication, tourism and law careers. In addition, considering that the university has two modalities of study most of the news focused on the open modality. Furthermore, some topics show the interest of the university in research and social activities that involve the promotion of a culture of peace and non-violence, social responsibility, health and so on.

In the manual validation process many words or terms in topics assigned to news correspond with news hashtags, then we pretend to implement a web application to recommend hashtags according to the news. Also, for future work, others topic modeling algorithms could be applied in order to improve the results.

Authors' contribution

Lizbeth Pacheco-Guevara: Methodology, experimentation, results, article editing. **Ruth Reátegui:** Introduction, conclusions, supervision, article editing. **Priscila Valdiviezo-Díaz:** Conclusions, supervision, article editing.

Table 3. Comparison between the topics emitted by LDA and hashtags given from the news
Source: Authors

Num_ News	Terms_in_Topics	Hashtags_News
1	Topic 9: derecho-penal, radio, covid, prendho, crisis, pandemia, personal, familia, graduados, instituto	
	Topic 0: paz, campus, comunicacion, violencia, educacion-superior, ninos, investigacion, red, mejora, educacion	#Retorno, #semáforo amarillo #educacion
	Topic 1: comunicacion, proyecto, concurso, loja, informacion, investigacion, indicadores, acreditacion, construccion, violencia	
6	Topic 5: derecho, mujeres, ods, artes, aplicacion, mayo, loja, muerte, donaciones, carrera	#donacion #campanaDalesuna patita
	Topic 1: comunicacion, proyecto, concurso, loja, informacion, investigacion, indicadores, acreditacion, construccion, violencia	#relacionespublicas utpl #utpl solidaria”
11	Topic 9: derecho-penal, radio, covid, prendho, crisis, pandemia, personal, familia, graduados, instituto	“teleasistencia médica #COVID-19 #asistencia médica gratuita
	Topic 1: comunicacion, proyecto, concurso, loja, informacion, investigacion, indicadores, acreditacion, construccion, violencia	#Loja #utpl #Hospital UTPL”

Conflicts of interest

The authors declare no conflict of interest

Bibliographic references

- Blei, D.M., Ng, A.Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Buenano-Fernandez, D., Gonzalez, M., Gil, D., & Lujan-Mora, S. (2020). Text Mining of Open-Ended Questions in

Self-Assessment of University Teachers: An LDA Topic Modeling Approach. *IEEE Access*, 8, 35318–35330. doi: 10.1109/ACCESS.2020.2974983.

- Chipidza, W., Akbaripourdibazar, E., Gwanzura, T., & Gatto, N.M. (2021). A topic analysis of traditional and social media news coverage of the early COVID-19 pandemic and implications for public health communication. *Disaster Medicine and Public Health Preparedness*, 3, 1-8. doi:10.1017/dmp.2021.65
- Denny, M. J., & Spirling, A. (2018). Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it. *Political Analysis*, 26(2), 168-189. doi:10.1017/pan.2017.44
- Guangce, R., and Lei, X. (2021). Knowledge discovery of news text based on artificial intelligence. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 20(1), 2021, doi:10.1145/3418062
- Kim, D., Seo, D., Cho, S., & Kang, P. (2019) Multi-co-training for document classification using various document representations: TF-IDF, LDA, and Doc2Vec. *Inf. Sci. (Ny)*, 477, 15–29. doi: 10.1016/j.ins.2018.10.006.
- Larsen, V.H., & Thorsrud, L. A. (2019). The value of news for economic developments. *Journal of Econometrics*, 210(1), 203–218, doi: 10.1016/j.jeconom.2018.11.013
- Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. EMNLP 2011 - Conf. Empir. *Methods Nat. Lang. Process. Proc. Conf.*, 2, 262–272.
- Misztal-Radecka, J. (2018). Building semantic user profile for polish web news portal. *Computer Science*, 19(3), 307–332, 2018.
- Oliveira Capela F. de, & Ramirez-Marquez, J. E. (2019). Detecting urban identity perception via newspaper topic modeling. *Cities*, 93, 72–83. doi: 10.1016/j.cities.2019.04.009.
- Xu, G., Meng, Y., Chen, Z., Qiu, X., Wang, C., and Yao, H. (2019). Research on Topic Detection and Tracking for Online News Texts. *IEEE Access*, 7, 58407–58418, doi: 10.1109/ACCESS.2019.2914097.
- Xu, L., Yeo, H., Hwang, H., & Kim, K.O. (2020). 5G service and discourses on hyper-connected society in south Korea: Text mining of online news. *Advances in Intelligent Systems and Computing*, 1120, 892-897, doi:10.1007/978-3-030-39442-4_68
- Wang, W., Feng, Y., & Dai, W. (2018). Topic analysis of online reviews for two competitive products using latent Dirichlet allocation. *Electron. Commer. Res. Appl.*, 29, 142–156. doi: 10.1016/j.elerap.2018.04.003.





Análisis comparativo de metodologías y herramientas tecnológicas para procesos de Business Intelligence orientado a la toma de decisiones

Comparative analysis of methodologies and technological tools for Business Intelligence processes aimed at decision making

Autores

✉¹* *María José Guerrero García*

✉² *Jorge Rodas-Silva*



¹Maestría en Tecnologías de la información, Universidad Laica Eloy Alfaro de Manabí, Manta, Ecuador.

²Facultad de Ciencias e Ingeniería, Universidad Estatal de Milagro, Guayas, Ecuador.

* Autor para correspondencia

Comó citar el artículo: Guerrero García, M. J., & Rodas Silva, J. (2022). Análisis comparativo de metodologías y herramientas tecnológicas para procesos de Business Intelligence orientado a la toma de decisiones. *Informática y Sistemas: Revista de Tecnologías de la Informática y las Comunicaciones*, 6(1), 38-44. DOI: <https://doi.org/10.33936/isrtic.v6i1.4522>

Enviado: 21/03/2022
Aceptado: 19/05/2022
Publicado: 31/05/2022

Resumen

La inteligencia de negocios a medida que pasa el tiempo adquiere mayor importancia y exige mayor competencia en las organizaciones, hoy en día las empresas tienen la necesidad de aplicar metodologías que se adapten e integren de manera positiva, tomando en consideración que cada negocio tiene sus propios procesos y determinadas actividades, es por ello que este artículo se centra en determinar como arma estratégica una comparativa de las metodologías Bill Inmon y Ralph Kimball, se aplicó el método deductivo el cual permitió especificar los atributos importantes de la inteligencia de negocios a su vez mediante un análisis de las características más importantes tomando parámetros de valoración las capacidades de mejoras, el grado de cumplimiento, objetivos y satisfacción, los avances en tecnologías de información ofrecen herramientas de gran capacidad que se han desarrollado para ello se pretende establecer un análisis estratégico de las herramientas tecnológicas Business Intelligence más utilizadas con el fin de evaluar sus fortalezas y debilidades para así alcanzar la mejora en la toma de decisiones y visión estratégica en cuanto a productividad de las organizaciones. Se obtuvo como resultados la metodología Ralph Kimball con mayor capacidad y características para garantizar la utilización de los datos de manera rápida y en menor tiempo de implementación, así mismo la herramienta BI con mayor porcentaje en fortalezas fue PowerBI, debido a su visión empresarial, facilidad de uso y visualización e integración de datos. Las metodologías y herramienta de inteligencia de negocios son de vital importancia por que permiten una buena planificación de la producción en cada gestión de procesos, y optimización de costos dando así una mejora en la toma de decisiones.

Palabras clave: Metodologías; Herramientas; Comparativa; Toma de decisiones.

Abstract

As time goes by, business intelligence acquires greater importance and demands greater competence in organizations. Nowadays, companies have the need to apply methodologies that adapt and integrate in a positive way, taking into consideration that each business has its own processes and certain activities, that is why this article focuses on determining as a strategic weapon a comparison of the Bill Inmon and Ralph Kimball methodologies, the deductive method was applied which allowed specifying the important attributes of business intelligence in turn through An analysis of the most important characteristics, taking evaluation parameters, the capacity for improvement, the degree of compliance, objectives and satisfaction, advances in information technology offer tools of great capacity that have been developed for this purpose, it is intended to establish a strategic analysis of the Business Intelligence technology tools most used in order to evaluate their strengths and weaknesses in order to achieve improvement in decision-making and strategic vision in terms of productivity of organizations. As a result, the Ralph Kimball methodology was obtained with greater capacity and characteristics to guarantee the use of data quickly and in less implementation time, likewise the BI tool with the highest percentage of strengths was PowerBI, due to its business vision, ease of use and visualization and integration of data. The methodologies and business intelligence tools are of vital importance because they allow good production planning in each process management, and cost optimization thus giving an improvement in decision making.

Keywords: Methodologies; Tools; Comparative; Decision making.



1. Introducción

Las soluciones de inteligencia de Negocios buscan transformar a las organizaciones en entidades analíticas, proporcionando destrezas y dinamismo para la investigación de la información ya que es uno de los activos más preciados en una organización para tener un elevado grado de competitividad y maneras de desarrollo. La utilización adecuada de la herramienta desarrollada beneficia a todo el equipo de trabajo de la empresa.

El manejo, la administración, la gestión y el control de la información como un arma estratégica forman parte de la inteligencia del negocio (IN), con apoyo de herramientas informáticas y analíticas, que ayudan a las organizaciones a maximizar su rendimiento, generando la eficiencia operativa. Así mismo, la gestión de conocimiento ayuda a obtener mayor comprensión y entendimiento del entorno y de los procesos desde la propia experiencia de las personas y organizaciones. (Dávila Ladrón de Guevara, 2005).

El conjunto de herramientas y aplicaciones para la ayuda a la toma de decisiones que posibilitan acceso interactivo, análisis y multiplicación de la información corporativa de misión crítica. Estas aplicaciones aportan un conocimiento valioso sobre la información operativa identificando problemas y oportunidades de negocio. Con ellas los usuarios son capaces de acceder a grandes cantidades de información para establecer y analizar relaciones y comprender tendencias que, a la postre, soportaran decisiones de negocio. (Méndez del Río, 2006).

Entre los trabajos de investigación en el área que destacamos los siguientes: (Basantes Espinoza & López Galarza, 2012). Ellos desarrollaron un estudio para la Universidad Politécnica Salesiana en Guayaquil, “Estudio De La Aplicación De Inteligencia De Negocios En Los Procesos Académicos”, Dentro de los objetivos lo que desea es implementar una aplicación de Business Intelligence que pueda brindar a los directivos de la universidad información detallada y consolidada del rendimiento por curso de cada estudiante para brindar una precisa información que permita conocer a que alumnos se les debe reforzar y orientar en las asignaciones más bajas (p.1).

(Muñoz Hernández, Osorio Mass, & Zúñiga Pérez, 2016) Este artículo tiene por objetivo describir y clasificar de una forma más concreta los sistemas de información, los cuales se encuentran enmarcados en lo que se denomina propiamente como “la inteligencia de los negocios”. Finalmente se concluye describiendo la relevancia que ha tenido la implementación de por lo menos alguno de estos sistemas de información en la planeación estratégica de las organizaciones, especialmente en Colombia y Latinoamérica.

2. Materiales y Métodos

Se aplicó el método deductivo el cual permitió especificar los atributos importantes de la inteligencia de negocios, el análisis de los conceptos y la información permitió detallar de forma general las características y criterios de las metodologías, la información permitió identificar las herramientas como alternativas para la implementación de un proyecto Business Intelligence, a su vez de integro el método comparativo mediante una búsqueda de patrones y semejanza con el objetivo de establecer las herramientas que coincidan con características ideales para ser aplicadas de forma sistemática.

2.1. Inteligencia de los negocios

Son herramientas de soporte de decisiones que permiten en tiempo real, acceso interactivo, análisis y manipulación de información crítica para la empresa. Estas aplicaciones proporcionan a los usuarios un mayor entendimiento que les permite identificar las oportunidades y los problemas de los negocios. Los sistemas de soporte de decisiones constituyen una tendencia creciente en empresas medianas y grandes que desean una gerencia eficiente. (Alameda, 2011).

Business Intelligence (BI) es la combinación de prácticas, capacidades y tecnologías usadas por las compañías para recopilar e integrar la información, aplicar reglas del negocio y asegurar la visibilidad de la información en función de una mejor comprensión de este y, en última instancia, para mejorar el desempeño. (Muñoz Hernández et al., 2016).

La inteligencia de los negocios y su aplicabilidad en los mismos sirven de soporte especial, a la organización para los procesos y operaciones, e indudablemente para la toma de decisiones; este apoyo que le otorga la BI a las organizaciones se da de manera organizada y estructurada, y se logra teniendo en cuenta tres frentes fundamentales en los cuales la BI se desarrolla: a nivel estratégico, táctico y operativo como se muestra en la Figura 1 (Zarate Gallardo, 2013)

2.2. Data Warehouse

El Data Warehouse es una pieza básica, fundamental e indispensable de todo sistema Business Intelligence porque todos los listados y análisis que se hagan se harán a partir de esta única base de datos.

Según (Inmon, 2002) un Data Warehouse comienza a diseñar el modelo de datos del almacén de datos corporativos, que identifica las principales áreas temáticas y entidades con las que trabaja la empresa, como clientes, producto, proveedor, y así. Por lo que se considera que es una “colección de datos orientada al

sujeto, no volátil, integrada y variable en el tiempo en apoyo de las decisiones de la gerencia”.

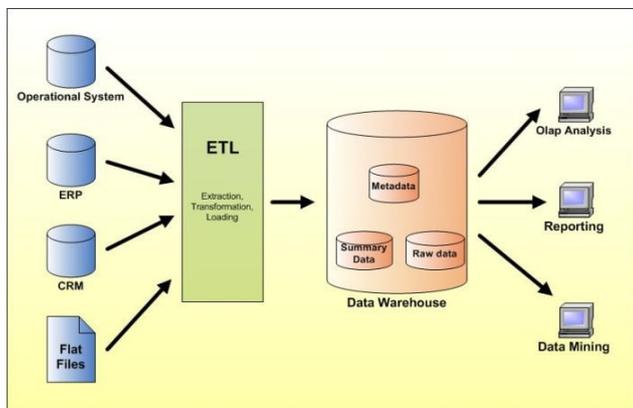


Figura 1. DataWarehouse Estructura.

2.3. Metodologías de Inteligencia de negocios

Sabemos que cada día la tecnología crece con mucha rapidez, por lo cual ya encontramos variedades de metodologías de BI que se ajustan a muchas necesidades que facilitan a organizaciones en áreas específicas, en este caso se realizara una comparativa de dos de ellas Inmon y Kimball, por lo que debemos considerar cuál es la más adecuada para elección en aspectos tales como las necesidades y su estructura de datos almacenados que requiere la organización.

Las metodologías nos pueden orientar a conceptualizar de manera correcta los objetivos y sobre todo a conceptualizarlos, para los profesionales de Business Intelligence, ya sean con experiencia o interesados en la materia, el tema de la metodología de Business Intelligence es considerada parte medular de la formación profesional. Muchos profesionales se guían por metodologías basadas en las corrientes de Kimball y/o Inmon, así como muchos otros diseñan sus propias metodologías basadas en alguna de las corrientes mencionadas o que sean afines a la tecnología seleccionada. (Cavazos, 2008)

2.4. Bill Inmon

La metodología Bill Inmon propone un modelo normalizado basado en la empresa, con una arquitectura de varios niveles y áreas de interés, Data Marts dependientes, poblando el Data Warehouse con datos a nivel atómico. (Inmon, 2002).

Inmon defiende una metodología descendente (top-Down) a la hora de diseñar un almacén de datos, ya que de esta forma se considerarán mejor todos los datos corporativos. En esta metodología los Data Marts se crearán después de haber terminado el Data Warehouse completo de la organización (Kimball, 2012).

Bill Inmon fue uno de los primeros autores en escribir sobre el tema de los almacenes de datos, (Vela Pizango, 2019), esta metodología es adecuada para sistemas complejos, sobre todo

cuando exista algún cambio o variación en los procesos de la organización se requiera de estabilidad y seguridad. Ver Tabla 1 presentada a continuación.

Tabla 1. Características de la Metodología Inmon.

Fuente: Los autores.

Características	Detalles
Enfocado a temas	La información y recursos concernientes a un mismo asunto tienen que estar organizados de tal forma que ellos queden juntos entre si dentro del almacén de datos
Variación en el tiempo	Los cambios elaborados en los datos a lo largo del tiempo quedan registrados para que los informes que se logren reflejen aquellas variaciones.
No volátil	La información no se modifica ni se elimina, una vez guardado un dato, éste se convierte en información de únicamente lectura, y se preserva para futuras consultas.
Integrado	La base de datos tiene los datos de los sistemas operacionales de la organización, y estos datos tienen que ser consistentes.

El enfoque de Inmon es una metodología de trabajo Top-Down por que se encarga de centrar una visión global de la empresa y luego se va distribuyendo en pequeñas secciones de datos departamentales, como se muestra en la Figura 2.

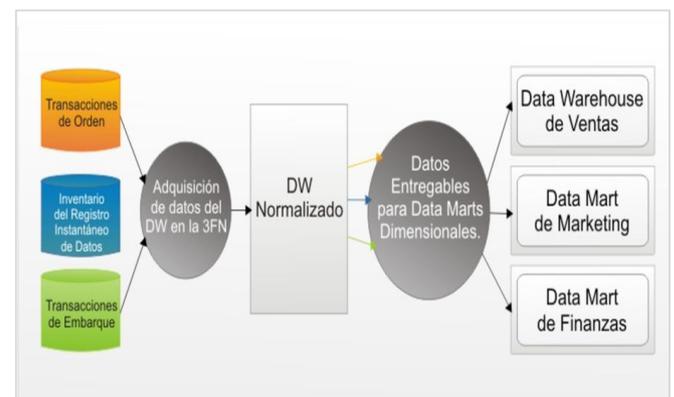


Figura 2. Enfoque Inmon.

2.5. Metodología Kimball

La metodología de Ralph Kimball propone un modelo dimensional de Data Marts, utilizando un esquema de estrella, con una arquitectura basada en las áreas de interés y Data Marts (Esparza, 2012), poblando el Data Warehouse con datos atómicos y sumariados. (Kimball, 1996)

Kimball proporciona una base empírica y metodológica perfectamente adecuada para las implementaciones de almacenes de datos, dada su gran versatilidad y su enfoque ascendente, que permite construir los almacenes en forma escalonada integra fases y artefactos para obtener mejores resultados: análisis de



requerimientos, análisis de los sistemas transaccionales, modelo lógico del ETL, procesos ETL, performance y mantenimiento del DW.

El Modelo Dimensional es una técnica de diseño lógico que tiene como objetivo presentar los datos dentro de un marco de trabajo estándar e intuitivo, para permitir su acceso con un alto rendimiento. Ver Tabla 2 a continuación.

Tabla 2. Características Metodología Kimball.
Fuente: Los autores.

Características	Detalles
Enfoque en el negocio	Se inicia con una identificación de los requerimientos del negocio y su valor asociado, y usar estos esfuerzos para desarrollar relaciones sólidas con el negocio.
Infraestructura de información	Diseñar una base de información única, integrada, fácil de usar, de alto rendimiento donde se reflejará la amplia gama de requerimientos de negocio identificados en la empresa.
Realizar entregas en incrementos significativos	Crear DataWarehouse en incrementos entregables en plazos de 6 a 12 meses. Hay que usar el valor de negocio de cada elemento identificado para determinar el orden de aplicación de los incrementos.
Solución completa	Proporcionar todos los elementos necesarios para entregar valor a los usuarios de negocios. Para comenzar, esto significa tener un almacén de datos sólido, bien diseñado, con calidad probada, y accesible.

El enfoque de Kimball es una metodología de trabajo Bottom-up se caracteriza por unir diferentes DataMarts que se encuentren estructurados, a su vez crear DataMart y luego ir agregando otros que comparten las dimensiones que estén establecidas u otras nuevas. Ver Figura 3.

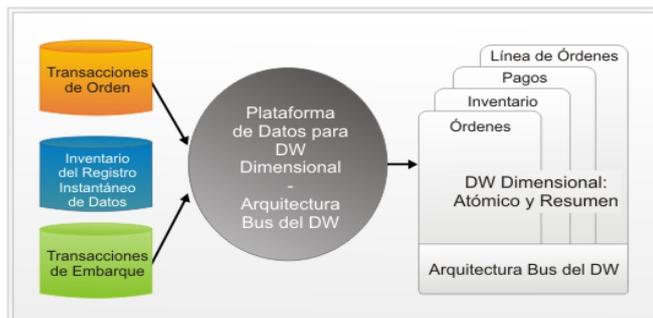


Figura 3. Enfoque Kimball.

2.6. Comparativa de metodologías Business Intelligence

Para la realización de un proyecto es fundamental tener definida la más adecuada metodología para el entorno del negocio, contar con ella nos ayuda a manejar procesos de manera sistemática facilitando a la organización mediante normativas de planificación y ejecución de proyectos. Dentro de esta comparativa se ha considerado las dos metodologías Inmon y Kimball en la Tabla 3.

Tabla 3. Comparativa de las Metodologías de BI.
Fuente: Los autores.

	R. Kimball	W. Inmon
Objetivo	Entregar una solución que facilite la labor de los usuarios finales para consultar directamente los datos	Almacenar, analizar e interpretar los datos que van generando y acumulando de una forma global
Enfoque	Bottom - Up	Top - Down
Estructura	Modelamiento de una DataMart por proceso de negocio el data Warehouse se obtiene con la unión de data Marts de la empresa	Un data Warehouse que abarque toda la empresa y que se alimenta de las bases de los departamentos
Herramientas	Modelo dimensional	Tradicional Diagramas E-R
Tiempo Implementación	Es rápido de implementar	Demora en implementar demanda más tiempo.
Usuarios	Finales	Área de T. I
Costos	Desarrollo de solución por DataMart no es un costo alto	Costos aumentan por cantidad de datos

2.7. Herramientas tecnológicas para Inteligencia de negocios

Las herramientas de inteligencia de negocios están programadas para ayudarte a entender las tendencias y extraer información a partir de los datos para que puedas tomar decisiones empresariales tácticas y estratégicas. (Montaño Alvarez, 2017).

“La disciplina denominada como Business Intelligence nos acerca a los sistemas de información que nos ayudan a la toma de decisiones. La PYME dispone, como todas las empresas sin importar su tamaño, de sistemas de información más o menos sofisticados y que es conveniente analizar y optimizar”. (Cano, 2007)

Análisis detallado de algunas de las herramientas, Business Intelligence las que más aceptación están teniendo últimamente, en la Tabla 4 se determina cuáles son sus fortalezas y debilidades de cada herramienta:

Tabla 4. Fortalezas y debilidades de las Herramientas BI.

Fuente: Los autores.

Herramientas	Fortalezas	Debilidades
Tableau	<ul style="list-style-type: none"> ✓ Funciones de exploración interactive ✓ Visualización 	<ul style="list-style-type: none"> ✓ Deficiente Reporting ✓ Análisis predictivo
Qlikview	<ul style="list-style-type: none"> ✓ Rápida implementación ✓ In-memory ✓ Rendimiento ✓ Buena funcionalidad ✓ Experiencia con el cliente ✓ Éxito del proveedor 	<ul style="list-style-type: none"> ✓ Costo de licencia ✓ Reporting deficiente
IBM	<ul style="list-style-type: none"> ✓ Buena infraestructura ✓ Amplitud de usuarios 	<ul style="list-style-type: none"> ✓ Costo de licencia ✓ Rendimiento
Oracle	<ul style="list-style-type: none"> ✓ Excelente integración ✓ Despliegues grandes en usuarios ✓ Despliegues grandes en volumen de datos 	<ul style="list-style-type: none"> ✓ Difícil administración ✓ Cuadro de mando parametrizados
BITAM	<ul style="list-style-type: none"> ✓ Fácil de usar ✓ Satisfacción del cliente 	<ul style="list-style-type: none"> ✓ Poca estabilidad ✓ Muchos números de errores

2.8. Beneficios de las herramientas de Inteligencia de negocios

Como beneficios tenemos los siguientes detalla (Pesce & Pons, 2012):

Disponer de la capacidad de extraer, depurar, consolidar, sintetizar y presentar datos, información, conocimiento de forma automatizada.

Poner a disposición de los usuarios que la necesiten la información necesaria para el análisis y la toma de decisiones: ágil, flexible y fiable, en el formato adecuado a cada uno.

La información se obtiene sin dependencias de otros departamentos, con posibilidad de navegación OLAP por los propios usuarios, que permite profundizar en el análisis de forma interactiva en base a cualquiera de las dimensiones disponibles.

Homogeneidad en la utilización de la información (interna y externa): utilización de la misma información al medir las mismas cosas.

3. Resultados y Discusión

El análisis de la comparativa de las metodologías más utilizadas las cuales son Inmon y Kimball se propone que esta metodología cumpla y se adapte a la gestión de datos, necesidades y optimización de recursos en los proyectos que requieran las empresas, se analizaron variables que ayudo a dar claridad dando a conocer cual metodología es la más adecuada mediante indicadores de las características metodológicas como el grado de cumplimiento de objetivos, capacidad de mejoras y satisfacción.

Se observa en la Tabla 5 y la Figura 4, que la metodología Kimball cumple con el mayor porcentaje de los indicadores establecidos para dicha evaluación con un total de 22 puntos obteniendo el primer lugar, esta metodología según sus características está diseñada para garantizar la utilización de los datos de manera rápida en un menor tiempo de implementación, dando facilidad a cualquier usuario, a su vez se adapta a proyectos pequeños y es el más usado en el mundo. En el caso de la metodología Inmon sus resultados obtuvieron una puntuación de 17 puntos quedando en el segundo lugar, esta metodología integra todos los datos de la compañía es más compleja de utilizar por que necesariamente requiere de usuarios especializados, porque está diseñada para almacenar grandes volúmenes de datos por lo que su costo es mayor.

Se obtuvo como resultados en la Tabla 6 y la Figura 5, en base a las fortalezas y debilidades de las herramientas Business Intelligence y una respectiva revisión literaria de las más utilizadas en el ámbito empresarial, teniendo como propósito ayudar a conocer más de ellas y saber cuál es la que más se adapta a las necesidades del usuario, la herramientas con más valoración son: PowerBI con un total 19,35% en fortalezas ofrece facilidad de uso, visualización e integración de datos y sobre todo visión empresarial muy completa, sobresale en términos de precio y uso compartido de datos, sus debilidades son de 7,14% por su desventaja de atraso en cuanto sus limitaciones en el modelado de datos y manejo On-premises, la segunda herramienta más utilizada es Qlikview con un 16,13% de fortalezas ofrece BI de autoservicio, visualización, estructura solida de derechos de usuarios, sus debilidades son de 14,29 % por limitaciones en cuanto a el análisis de AD HOC simples, licencia compleja muy costosa.

4. Conclusiones

La inteligencia de negocios tiene como vital apoyo las metodologías y herramientas para la mejora de gestión de datos en las organizaciones, analizando varias revisiones literarias se determina como las entidades solucionan problemas en la entrega de información en menor tiempo de reportes mediante la decisión de utilizar metodologías de Business Intelligence que ha llevado al éxito y mejora en la toma de decisiones.

La metodología Inmon y Kimball garantizan la integración y explotación de información como apoyo a las organizaciones para tomar mejores decisiones en cuanto a la inteligencia de negocios ajustándose a las necesidades empelando normativas y planificaciones que facilitan el sistema de almacenamiento de datos, cada una tiene sus principios y desempeños con propósitos



Tabla 5. Indicador de característica de Inmon y Kimball.
Fuente: Los autores.

Indicadores	Kimball	Inmon
Adaptable	3	3
Fácil entendimiento	2	1
Mas usado	3	2
Tamaño del proyecto	2	3
Tiempo en análisis, diseño y construcción	2	2
Tiempo de implementación	3	1
Costos	2	1
Comunicación con el cliente	2	2
Total	22	17

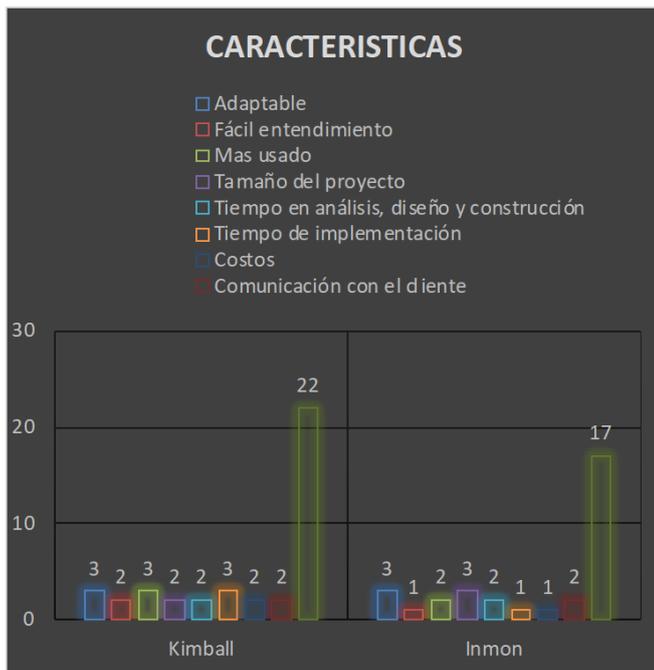


Figura 4. Indicador de características de metodologías Inmon y Kimball.

que se pueden acoplar a la necesidad que se requiera los proyecto. De acuerdo con el estudio comparativo que se llevó a cabo se recomienda la metodología Kimball como las más adecuada para proyectos pequeños, para realizar un mejor uso de la información y permitir la integración de BI en las actividades estratégicas, tácticas y operacionales.

Mediante el análisis de las fortalezas y debilidades de las

herramientas Business Intelligence se logró determinar que hay muchas herramientas en el mercado ya que estas tecnologías son apropiadas y dan acceso a información e integración de datos en las áreas de la organización, al conocer sus ventajas y desventajas las más competentes y con mayor beneficio en cuanto a sus funciones de desempeño y costos, fueron Power BI y Qlikview de gran aporte a las necesidades del usuario.

Tabla 6. Valoración de Fortalezas y Debilidades de herramientas BI.
Fuente: Los autores.

Herramientas	Fortalezas	Debilidades	%F	%D
Tableau	5	2	16,13%	14,29%
Qlikview	5	2	16,13%	14,29%
IBM	2	2	6,45%	14,29%
Oracle	3	2	9,68%	14,29%
BITAM	2	2	6,45%	14,29%
MicroStrategy	5	1	16,13%	7,14%
PowerBI	6	1	19,35%	7,14%
Information Builders	3	1	9,68%	7,14%
Total	31	13	100,00%	100,00%



Figura 5. Resultados Fortalezas y Debilidades de las herramientas BI.

Contribución de los autores

María José Guerrero García: Conceptualización, Metodología, Análisis formal. **Jorge Rodas-Silva:** Revisión y edición del artículo

Conflictos de interés

Los autores declaran no tener ningún conflicto de interés.

Referencias bibliográficas

- Alameda, C. (2011). *Business Intelligence*. Obtenido de Business Intelligence: <https://www.dorjansistemas.com/business-intelligence/>
- Basantes Espinoza, G. P., & López Galarza, D. E. (2012). *Estudio de la aplicación de inteligencia de negocios en los procesos académicos caso de estudio Universidad Politécnica Salesiana*. Guayaquil, Ecuador. Obtenido de <https://dspace.ups.edu.ec/bitstream/123456789/3164/1/UPS-GT000322.pdf>
- Cano, J. L. (2007). Business Intelligence: competir con información
- Cavazos, E. (2008). *Metodología para Business Intelligence*. Obtenido de Business Intelligence, Data Warehouse, Tech & Biz: <https://gravitar.biz/bi/metodologia-business-intelligence/>
- Dávila Ladrón de Guevara, F. (2005). *Hacia La Inteligencia Del Negocio Con Excel 2003 (Vol. 148)*. (E. N. Acevedo, Ed.) Bogotá, D.C. Colombia: Editorial Politécnico Grancolombiano. Obtenido de <http://hdl.handle.net/10823/794>
- Esparza, C.A. (2012). *Análisis, Diseño e Implementación de un Data Mart Utilizando Herramientas Open Source Para las Unidades Administrativa y Financiera De La Espe*. Biblioteca Universidad de las Fuerza Armadas ESPE. <https://repositorio.espe.edu.ec/bitstream/21000/8786/1/AC-SI-ESPE-048075.pdf>
- Inmon, W.H. (2002). *Building the Data Warehouse* (Third Edition). John Wiley & Sons, Inc. New York, NY, USA 2002
- Kimball, G. (1996). *The data warehouse toolkit: practical techniques for building dimensional data warehouses*. John Wiley & Sons, Inc., 1996.
- Kimball, G. (2012). *Kimball group*. Obtenido de <http://www.kimballgroup.com/data-warehousebusiness-intelligence-resources/books/kimball-reader>
- Méndez del Río, L. (2006). *Más allá del Business Intelligence. 16 experiencias de éxito. España: Gestión 2000*. Obtenido de <https://dialnet.unirioja.es/servlet/libro?codigo=267353>
- Muñoz Hernández, H., Osorio Mass, R. C., & Zúñiga Pérez, L. M. (08 de Abril de 2016). Inteligencia de los negocios. Clave del Éxito en la era de la información. *Clío América*, 10(20), 194-211. doi:<http://dx.doi.org/10.21676/23897848.1877>
- Pesce, P., & Pons, C. F. (2012). Fundamento arquitectónico para herramientas de gestión del rendimiento empresarial. En: XIII Argentine Symposium on Software Engineering, (XLII JAIIO, La Plata, 27 al 31 de agosto de 2012)
- Vela Pizango, D. G. (2019). *Solución de inteligencia de negocio para la toma de decisiones en la empresa Milenium Electronics SAC*. <http://hdl.handle.net/11458/3515>
- Zarate Gallardo, E. D. (2013). *Inteligencia de negocios*. Obtenido de Obtenido de Gestipolis: <http://www.gestipolis.com/inteligencia-de-negocios/>





La Brecha Digital en las Instituciones Educativas Fiscales de Manta: Situación actual, necesidades y desafíos

The Digital Divide in Fiscal Educational Institutions of Manta: Current Situation, Needs and Challenges

Autores

✉ * *Ingrid Yadira Álava Chávez*

✉ *Dolores Esperanza Muñoz Verduga*

Universidad Laica “Eloy
Alfaro” de Manabí, Manta,
Ecuador.

* Autor para correspondencia

Resumen

Para ciertos grupos vulnerables de la población, les es difícil e incluso imposible acceder a la educación virtual, las desigualdades socioeconómicas acrecientan la brecha digital, lo que impide el desarrollo. Esta investigación describe como el analfabetismo tecnológico amplía la brecha digital en las instituciones educativas fiscales de Manta Post COVID periodo 2021-2022. Se realizó un estudio de tipo descriptivo documental, utilizando el método de análisis-síntesis y el deductivo-inductivo, que ayudaron en la revisión de la literatura científica. Se demuestra que, la gestión realizada por el MINEDUC ha logrado dar acceso al internet al 97 % de las instituciones educativas, lo que contrasta respecto a la dotación de equipos tecnológicos, que alcanza un 2% de la población educativa de Manta. Se debe destinar una mayor asignación económica para educación, enfocado a la adecuación de espacios y dotación de equipos informáticos, y en especial una capacitación efectiva a sus docentes.

Palabras clave: Brecha digital, analfabetismo tecnológico, Educación.

Abstract

For certain vulnerable groups of the population, has been difficult and at times almost impossible have access to virtual education, the socioeconomic differences have increased the digital gap, that has caused the lack of developing on this matter. This investigation will help to discover how digital illiteracy widens the digital bridge in the fiscal year for the education sector of Manta post-COVID from the 2021-2022 period. Recently there was a study made, a descriptive documentary, utilizing a synthesis analysis method and deductive-inductive method that helped to build the revision of the scientific literature. This study demonstrates that even after all the efforts done by MINEDUC to reduce the digital bridge, even though they were able to give 97 % internet access to all the educational institutions, they were short with respect to the endowment of technical equipment that only cover 2 % of the student population from the city of Manta. They need to improve a major economic assignment for the education sector, focus on spacious rooms and donations of technological equipment, specially an effective capacitation for the teachers.

Keywords: Digital Divide, Technical Illiteracy, Education.

Comó citar el artículo: Álava Chávez, I. Y. & Muñoz Verduga, D.E. (2022). La Brecha Digital en las Instituciones Educativas Fiscales de Manta: Situación actual, necesidades y desafíos. *Informática y Sistemas: Revista de Tecnologías de la Informática y las Comunicaciones*, 6(1), 45-51. DOI: <https://doi.org/10.33936/isrtic.v6i1.4536>

Enviado: 25/03/2022

Aceptado: 14/06/2022

Publicado: 09/07/2022



1. Introducción

A partir de los años 80' y 90', la nueva generación nace inmersa en el mundo de la Tecnología de la Información y Comunicación (TIC) los muy conocidos Millennial, es así como, para ellos, su comprensión y uso se les da de forma natural, pero sería muy apresurado suponer que todos poseen el mismo nivel de acceso a una computadora, al internet y en general a las nuevas tecnologías. El COVID-19 obligó a reformular, de manera general, las necesidades, para tratar de reducir la brecha digital. Debido a esta emergencia que ocasionó el cierre de las instituciones educativas, tanto docentes como estudiantes, quedaron desprovistos de sus espacios y se vieron abocados, obligatoriamente, a utilizar entornos virtuales desconocidos para ellos. Todo esto se convirtió en un verdadero desafío, en especial para aquellos docentes analfabetos tecnológicos, porque en general ninguno de ellos estaba preparado para desarrollar su labor profesional en escenarios como los actuales.

Por lo que (Moreira & López, 2017) define a la brecha digital como un fenómeno complejo debido a la variedad de variables económicas, demográficas, individuales y sociales asociadas a ella, las que clasifica por factores: 1) Brecha digital de acceso - Situación financiera, ingresos, nivel educativo, ocupación, ubicación geográfica. 2) Brecha digital de capacidad - Habilidad digital, nivel educativo, control sobre la tecnología disponible, y 3) Brecha digital de resultado - Actitud y motivación, comportamiento y voluntad, naturaleza del uso de la tecnología y capacidad de darle sentido. Es decir que los grupos vulnerables de la población afrontan todos estos factores en diferentes niveles de incidencia, por lo que se les hace difícil, e incluso imposible acceder a la educación virtual, generando desigualdades socioeconómicas y de oportunidades que acrecientan la brecha digital. En este sentido, se entiende por Brecha digital a la distancia que existe entre las personas y la tecnología; entre más individuos tengan acceso a la tecnología, la comprendan y la usen, este margen ira disminuyendo. Compartiendo el criterio de (Santoyo & Martínez, 2003) se define como la separación que existe entre las personas (comunidades, estados, países) que utilizan las Tecnologías de Información y Comunicación (TIC) como una parte rutinaria de su vida diaria y aquellas que no tienen acceso a las mismas, y que, aunque las tengan no saben cómo utilizarlas. Sin embargo, en todo el mundo hay millones de personas que viven sin acceso al internet y a las tecnologías; esta desigualdad afecta directamente al crecimiento económico y social de una nación.

Según (Marín, 2003) estableció que un "analfabeto tecnológico. es aquella persona que no sabe manejar un equipo de uso común: un celular, o bien, una computadora, e identifica dos tipos de analfabetos tecnológicos: el Pleno o Absoluto que son aquellas personas que ignoran por completo el manejo de una computadora o un equipo, e incluye a las personas que se resisten

a aprender un sistema (tecnofobia); y el Relativo o Funcional que consideran a aquellas personas que utilizan rudimentariamente una computadora o un equipo, es decir, sus funciones básicas, o bien, personas que aprendieron a manejar un viejo sistema y que por carecer de práctica no pueden comprender los nuevos adelantos, o las que por distintos motivos (miedos, inseguridades, etc.) abandonaron sus esfuerzos por continuar aprendiendo. Por lo tanto, cuando una persona desconoce de los avances tecnológicos, no utiliza el computador, software, el internet y demás, se convierte en un Analfabeto Tecnológico. Aunque en el Ecuador un Analfabeto Digital según datos del INEC (2020) está entre 15 a 49 años, y cumple simultáneamente tres características: 1) No tiene celular activado 2) En los últimos 12 meses no ha utilizado computadora 3) En los últimos 12 meses no ha utilizado internet. Pero este concepto, va mucho más allá de la accesibilidad y el uso de internet y de equipos tecnológicos; a diario la tecnología va innovando y actualizándose, cada vez de manera más acelerada, por lo que es probable que quien no es un "analfabeto tecnológico" hoy, puede serlo mañana.

Por otro lado, (Díaz Lazo & Florido Bacallao, 2011) resalta la importancia de superar la brecha digital, y menciona que la capacitación junto con una alfabetización digital se podría transformar, no sólo en mejores oportunidades de desarrollo, sino que también en una mejor calidad de vida, al contar con servicios electrónicos, acceso a información, cursos de capacitación, intercambio de conocimientos entre especialistas en distintas temáticas, consulta entre personas y muchas otras, que nos posibilitan hoy las TIC. En este sentido la problemática principal ha sido el uso obligatorio de las TIC en sus quehaceres laborales, esto evidenció aún más el analfabetismo tecnológico en los docentes. Por lo tanto, el trabajo para mitigar la brecha digital conlleva a reducir las barreras del conocimiento y el acceso a las tecnologías, solucionar este problema traerá consigo un mayor desarrollo social y económico para el país.

Con este enfoque, para el proceso de construcción de este trabajo surgieron interrogantes que a la vista de las autoras son importantes: el plantearse la forma de reflexión sobre el analfabetismo tecnológico, como algo que desde la óptica de la política pública tiene que abordarse con mayor rigurosidad, en este sentido, se parte desde lo siguiente ¿Cómo el analfabetismo tecnológico amplía la brecha digital en las instituciones educativas fiscales de Manta?, ¿La brecha digital causa el analfabetismo tecnológico?, ¿Cómo mitigar la brecha digital en las instituciones educativas fiscales de Manta?, ¿Cuáles son las políticas públicas, enfocadas en democratizar el acceso a las nuevas tecnologías en el Ecuador?, ¿Qué hacer para disminuir la brecha digital y obtener el máximo provecho de la educación virtual?. Por tanto, el objetivo de este trabajo se estableció en describir el analfabetismo tecnológico y sus consecuencias en el incremento de la brecha digital en las instituciones educativas

fiscales de Manta Post COVID periodo 2021-2022.

La brecha Digital

De acuerdo al Informe de la (UNESCO, 2005) en pro de una auténtica solidaridad digital, el reducir la brecha digital es un objetivo prioritario si queremos que las nuevas tecnologías contribuyan al desarrollo social y económico, y propicien el surgimiento de auténticas sociedades del conocimiento. En donde, se establece como elemento central la capacidad para identificar, producir, tratar, transformar, difundir y utilizar la información con vistas a crear y aplicar los conocimientos necesarios para el desarrollo humano.

A pesar del tiempo transcurrido y de la emergencia sanitaria COVID-19, a marzo del 2021 según datos (World Internet Users Statistics and 2022 World Population Stats, 2021) tan solo un 65,6 % de la población mundial tenía acceso a internet: más del 85 % corresponde a las regiones desarrolladas (Europa, Norteamérica), pero menos del 50 % en regiones más pobres (como África), lo que evidencia que aún falta mucho trabajo para reducir la brecha digital y convertirnos en una verdadera sociedad del conocimiento.

El Ecuador retrocedió en la meta de reducir la brecha digital según datos del Banco Interamericano de Desarrollo (2021) (IADB, 2022) el país cayó seis puestos en el índice de Desarrollo de la Banda Ancha. Las falencias en infraestructura y acceso a tecnología para los estudiantes ubican al país en el puesto 54 entre 65 economías, aunque Ecuador ha tenido avances en normativa y regulación, los resultados siguen siendo deficientes en infraestructura, acceso y calidad de conexión. Se deduce que, si hubiera más población con acceso a internet y con servicios de calidad en conexión en general, reducir la brecha digital tendría un gran impacto positivo con relación a lo económico y social.

Para (Guallo Paca & Guadalupe Arias, 2018) en su artículo Alfabetización Digital en el Ecuador en el siglo XXI, con el objetivo de conocer el grado de discernimiento y proporcionar alternativas de enseñanza – aprendizaje que permitan la transmisión de conocimientos, ideas, imaginación, creatividad, entre otros, utilizaron en su investigación el método descriptivo, que permite conocer las condiciones de las personas acerca de la brecha digital, donde plantearon varias estrategias para que las personas que habitan en lugares con escasos recursos tecnológicos y económicos tengan acceso a ellas. Estrategias como promocionar el precio de dispositivos móviles, equipos activos y pasivos de red, la instalación de redes WLAN, lo que da pauta para que se involucre a las instituciones educativas, a fomentar de esta manera a través de la educación como parte de la formación en la vida del ser humano, por lo que se podría establecer a la brecha digital como una de las principales causas del analfabetismo digital.

(Moreno Inte & Ramón Naula, 2019) en su proyecto de investigación Factores Determinantes del Analfabetismo Digital del Ecuador, que tuvo como objetivo analizar los factores determinantes del analfabetismo digital en el país, realizó un análisis descriptivo para caracterizar a la población ecuatoriana

considerada como analfabeta digital, en donde se halló como determinantes a los factores de: sexo (mujeres), edad (adulto-mayor), etnia (indígena), área (rural), nivel de instrucción (ninguno), y condición de actividad (ocupados).

Lo que resalta (Women., 2021) que existe la desigualdad tecnológica por género, ya que la mayoría de las personas que no tienen acceso a la red son mujeres. A través de este análisis se establece que existen desigualdades en el acceso a las Tic, que tienen consecuencias importantes en el desarrollo social y económico del país, generando que la grieta de la brecha digital se extienda.

De acuerdo con (Camacho, 2005) la brecha digital resulta de las brechas sociales, es decir, que la reducción de la brecha digital implicará más desarrollo en los países, regiones y personas con mejores oportunidades de acceso en detrimento de quienes tienen menos. Por lo que, quienes tienen un mayor nivel de formación con respecto al uso de las TICs, tendrán mejores condiciones y oportunidades de progresar y optimizar su calidad de vida. Según (Trilling & Fadel, 2009) las habilidades del siglo XXI están vinculadas a la alfabetización digital, aprender e innovar para la vida y la profesión, las instituciones educativas no son la excepción y obligatoriamente entraron al proceso digital, Yos (2021) (Pandemia, 2021) indicó en el Congreso Educación Post Pandemia, que toca “Desaprender para volver a aprender porque este contexto es diferente y tenemos que adaptarnos a la nueva normalidad, que no va a ser lo que estábamos haciendo en las presenciales, la nueva normalidad es trabajar con los entornos tecnológicos”. Así mismo (Ullauri, 2016) indica que en la actualidad la educación debe ser un espacio de construcción de saberes no tradicionales, debe comprender y apropiarse de las estrategias y herramientas de enseñanza-aprendizaje idóneas que faciliten y potencien el desarrollo de aprendizajes auténticos.

La preparación de los docentes del siglo XXI debe comprender el manejo y empleo correcto de las diferentes herramientas TICs para su uso en las aulas de clase y en su trabajo diario, el nuevo rol del docente es trasladar sus conocimientos a través de este entorno virtual, construir y aprovechar al máximo la tecnología para hacer sus clases amenas, ya no puede venir a dar una clase magistral donde el alumno solo escucha, debe buscar estrategias para que el alumno pueda participar e interactuar y a su vez aprenda a utilizar estas herramientas, la construcción del conocimiento demanda alfabetizarse digitalmente.

El Ministerio de Educación (MINEDUC) implementó el plan nacional de formación permanente y durante el 2020, que comprenden capacitaciones dirigidas a todos los docentes del magisterio fiscal, en este sentido por la emergencia sanitaria que atravesó el país, se implementaron nuevas capacitaciones sobre el uso de herramientas tecnológicas. De acuerdo con los cursos disponibles en la plataforma MeCapacito, se realizó una revisión de los cursos existentes, se constata que existen 30 cursos disponibles, uno sobre innovación tecnología y creación de contenido digital y los restantes orientados al campo pedagógico. A nivel nacional a septiembre del 2020 en el campo de Tic se han capacitado 11866 docentes lo que contrasta con los resultados del curso de actualización curricular en el que se capacitaron



187362 docentes (Educación, 2021).

De acuerdo a lo descrito anteriormente la brecha tecnológica ha sido tratada desde diferentes aristas en las distintas investigaciones que se han planteado en este apartado, por lo tanto una conclusión previa sería que ante la ausencia del conocimiento tecnológico la brecha digital en los profesionales de la educación va incrementado con el pasar del tiempo, ante ello es fundamental que se establezcan líneas de mejora (capacitación, talleres, seminarios) para ir reduciendo esta problemática del no manejo y uso de las herramientas tecnológicas de forma óptima y adecuada.

2. Materiales y Métodos

El presente trabajo realizará un énfasis en el análisis de la situación del analfabetismo tecnológico, en la población docente, determinando así, la existencia de una segregación de estos sectores en el acceso y uso de las tecnologías. Por lo cual se ha considerado estudiar a la población docente del cantón Manta (zona urbana y rural). En este sentido, esta investigación se fundamenta desde el paradigma interpretativo, ya que asume desde diferentes puntos de vista ontológico la realidad interpretando y comprendiendo a través de métodos multimetódicos, naturalista y reflexión del objeto de estudio.

El trabajo de investigación se planteará bajo un enfoque cuantitativo con un diseño metodológico no experimental de carácter transversal, en este aspecto, el alcance de la investigación será descriptivo, para identificar las características más importantes de actividades, procesos, personas, comunidades o cualquier otro hecho, fenómeno u objeto que se estudie. En cuanto a la población de estudio, este está conformada por 81 instituciones de carácter fiscal según consta en el registro del Distrito de Educación 13D02, las mismas que fueron seleccionadas con base a los siguientes criterios: sostenimiento fiscal, población estudiantil de más de 500 estudiantes ya que estas instituciones cuentan con laboratorio de computación y, que tengan un docente TIC dentro de la plantilla óptima, por otro lado se tomó en cuenta centros educativos del sector rural del distrito educativo que cumplieran con los criterios expuestos, para la selección de la muestra se empleó un tipo de muestreo no probabilístico intencional o por conveniencia, ya que este tipo de técnica permitió que los casos o los individuos se seleccionan en función de la intención del investigador bajo criterios preestablecido.

Para la obtención de la información se empleó la encuesta, que en su definición según palabras de (Arias, 2012) es una técnica comúnmente utilizada en distintas ciencias del conocimiento, en este aspecto, la misma se desarrolló tomando en cuenta las dos categorías de análisis del estudio, la brecha digital y el

analfabetismo tecnológico, esta a su vez, en palabras del mismo autor es una técnica desarrollada mediante un instrumento llamado “cuestionario” el mismo es definido por (Bolívar, 2013) como un instrumento de recolección de datos, de lápiz y papel, integrado por preguntas que solicitan información referida a un problema, objeto o tema de investigación, el cual es normalmente aplicado a un grupo de personas.

El instrumento permitió caracterizar las categorías analizadas en este estudio ya entre sus ítems se propuso aspectos como: formación profesional, años de experiencia en relación al empleo de las TIC en el proceso de enseñanza-aprendizaje, características del laboratorio de cómputo, niveles de satisfacción en cuanto a los desarrollado por ente de la política pública en educación en capacitación continua. Para (Sambache & Cevallos, 2020) en su artículo Brecha digital: conectividad y equipamiento en instituciones de educación fiscal en Ecuador, indagó en los antecedentes sobre la dotación de conectividad a Internet y equipamiento tecnológico, para las instituciones educativas del Ecuador, utilizó un diseño metodológico con las técnicas de investigación documental, donde revisan las causas para los ascensos y descensos de la inversión gubernamental, al dotar de conectividad y equipamiento a las instituciones educativas, y se establecen los argumentos por los cuales el uso pedagógico de las TIC es limitado en Ecuador.

Por lo que se deduce que los factores que influyeron en este desarrollo sin duda tienen una amplia relación con la tendencia política-ideológica que experimentó el país en la última década. Mediante Acuerdo Ministerial 011-2017, publicado en el Registro Oficial No 15 del 15 de junio del 2017, se expidió las Políticas Públicas del Sector de las Telecomunicaciones y de la Sociedad de la Información 2017-2021, en cuanto a lo que tiene que ver con la reducción de la brecha digital en la población, fomentar el uso de las TIC y la incorporación tecnológica de las empresas como apoyo al desarrollo de los negocios para impulsar la productividad y competitividad, así como también continuar en la construcción de las bases para promover el desarrollo de la industria de la TIC en el país. Es decir que toda esta instrumentación de gobernanza es altamente desarrollada. Según (Bravo Ávalos, 2019) en su investigación Análisis de las Políticas Públicas y la Brecha Digital en el Ecuador Período 2009-2016 donde analiza las Políticas Públicas para el desarrollo de las Tecnologías de la Información y Comunicación (TIC) en el Ecuador, específicamente para mitigar la Brecha Digital en el Ecuador, en el período 2009 – 2016 con un enfoque metodológico basado en el Modelo NATO, por sus siglas en inglés (Nodality, Authority, Treasure, Organización) que contiene un compendio estructurado y metodológico de las decisiones gubernamentales referentes a las TIC y puntualmente a la Brecha Digital.

3. Resultados y Discusión

Descripción del Contexto de la Brecha Digital en el Ecuador del Distrito de Educación Manta

En el Ecuador sí existió una Política Pública marcada y claramente definida respecto a la reducción de la brecha digital, implementada e instrumentada dentro del período (2009 – 2016) que corresponde a la administración del Gobierno de la Revolución Ciudadana. La planificación y el orden institucional que experimentó el Estado permitió desarrollar un esquema favorable en la construcción de Políticas Públicas enfocadas a la reducción de la brecha digital, sin embargo, la cantidad de políticas y normativas por parte del estado puede generar sensibilidad en los actores sociales del sector de las Telecomunicaciones. En la administración del Lcdo. Lenin Moreno en noviembre del 2019, lanzó el proyecto “Internet para todos” para reducir la brecha digital en zonas rurales donde existe bajo índice de penetración del internet, en busca de garantizar el internet fijo y móvil con precios económicos, se aspiró a cubrir un 98 % de territorio nacional integrando a varios proveedores de internet al proyecto, como Claro, Telefónica y CNT-EP, quienes comparten infraestructura, con lo que beneficiarían a 16 millones de personas según lo indicado por el Ministro de Telecomunicación del momento Andrés Michelena, quien resaltó que la transformación en el sector se da en cumplimiento a la disposición presidencial de conectar vidas, conectar territorios, conectar con tecnología de alta velocidad y conectar a precios bajos. Pero con el nuevo cambio de gobierno y la emergencia sanitaria, surgieron otras prioridades y estos proyectos perdieron fuerza, lo que se traduce que en Ecuador no se haya podido cumplir con la meta de reducir la citada brecha para que las instituciones educativas puedan beneficiarse de la tecnología. La realidad de cada país difiere en cuanto a brecha digital se trata, actualmente el 53.20 % de los hogares ecuatorianos cuentan con acceso a internet; mientras el porcentaje se reduce a 25.30 % si se analiza si poseen una computadora personal según datos del INEC (2020) (Tecnologías de la Información y Comunicación-TIC , s.f.), lo que se traduce que un 46.80 % de la población nacional del Ecuador no tienen acceso al internet y lo que es aún más grave que solo un cuarto de la población haya podido adquirir una computadora personal, en la lo que deja en evidencia la brecha digital que existe con respecto a la adquisición de equipos tecnológicos, en la Figura 1 se muestran porcentajes de la disponibilidad de recursos tecnológicos con los que cuentan docentes de las escuelas fiscales de Manta, donde se observa según los datos obtenidos por medio de encuestas, que aún existe un porcentaje de la población de docentes con la falta de estos recursos tecnológicos.

Para Tejedor et al. (2020) (Santiago & Parola, 2020), los docentes, en el escenario de la tele formación, deben mostrar conocimientos no solo sobre el contenido de la materia, sino también conocimiento tecnológico y pedagógico - digital. Por lo tanto, los docentes deben conocerla y aprender a usarla, identificar cuál es la más amigable y, sobre todo, más accesible para él y sus alumnos, lamentablemente hay escuelas, más aún, de sostenimiento fiscal que no poseen los recursos, o poseen equipos tecnológicos, pero ya obsoletos, no tienen un buen servicio de internet o no lo tienen, en la figura 2 se muestra el

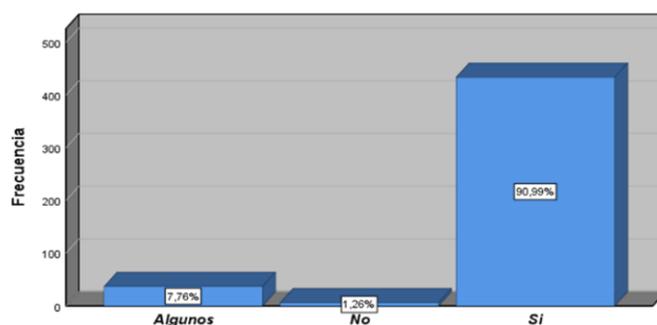


Figura 1. Estadística de docentes de las escuelas fiscales de Manta con falta de recursos tecnológicos.

Fuente: Las autoras.

porcentaje de la disponibilidad de recursos tecnológicos de los profesores, resultados que se obtuvieron de una encuesta realizada a los profesores de los colegios fiscales de Manta.

De acuerdo con datos estadístico del Área de Planificación de la Dirección Distrital de Educación 13D02 a Julio del 2021, el cantón Manta cuenta con 81 instituciones fiscales, 58.164 estudiantes, 2.098 docentes, y de acuerdo con el plan de conectividad de la Unidad Distrital de Tic a mayo del 2021, de las 81 instituciones educativas fiscales 30 no cuentan con equipos tecnológicos, y 2 aun no tienen internet (por ser zona rural de difícil acceso, los costos de instalación son muy elevados), ver en la Tabla 1.

Si bien el Ministerio de Educación y la Dirección Distrital han

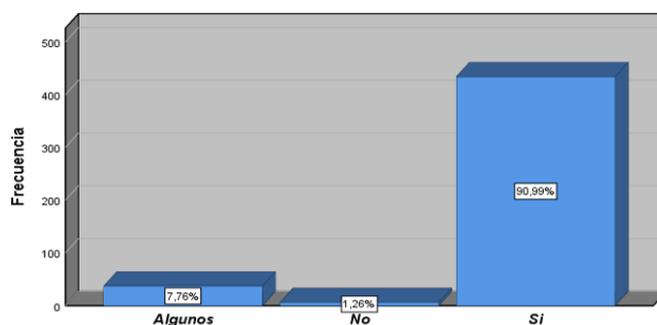


Figura 2. Estadística de la disponibilidad de recursos tecnológicos para el desarrollo de clases virtuales en los docentes de las instituciones educativas fiscales del cantón Manta periodo lectivo 2021-2022.

Fuente: Las autoras.

trabajado para la reducción de la brecha y dar accesibilidad al internet a toda la comunidad educativa con respecto a la dotación de equipos tecnológicos, de acuerdo con la Unidad Distrital de tecnología de la Información y Comunicación del Distrito 13D02 Manta-Montecristi-Jaramijó, no se ha visto dotación de equipos tecnológicos en los últimos 6 años a excepción de las instituciones que fueron reconstruida por el terremoto 16A, donde 12 unidades educativas del cantón Manta recibieron dotación de equipos informáticos en el año 2017, y las instituciones restantes poseen equipos que se pueden considerar obsoletos. Actualmente con

Tabla 1. Detalle de la Población e infraestructura de las Instituciones Educativas del Cantón Manta.

Fuente: Los siguientes datos fueron obtenidos por la autora del artículo, que es la directora del Departamento de Tecnología y Planificación del Distrito 13D02 (2021).

Detalle	Zona	Zona	Total
	Urbana	Rural	
Instituciones Educativas	68	13	81
Número de Docentes	1996	102	2098
Número de Estudiantes	55976	2191	58167
Instituciones Educativas con Laboratorio de computación	42	8	50
Instituciones Educativas con acceso a internet	68	11	79
Número de computadoras en las instituciones educativas	1123	118	1241

la emergencia sanitaria COVID-19 se ha observado el trabajo de varios organismos particulares (Fundación Telefónica, OIM, Rotary Club, etc.) en coordinación con el Distrito de Educación, para dotar de tabletas a estudiantes y docentes, pero no alcanzan a cubrir ni el 1% de la población estudiantil del cantón Manta, por lo que la entrega se prioriza por alguna condición de vulnerabilidad.

Discusión

La brecha digital en las instituciones educativas fiscales de Manta aún tiene mucho trabajo por delante, a pesar de los esfuerzos realizados por mitigarla de parte del gobierno, éste no ha sido suficientemente efectivo, ya que la brecha digital se orienta también a factores socioeconómicos, puesto que no solo comprende accesibilidad al internet o adquisición de equipo tecnológicos. Por lo tanto, el gobierno en turno debe mejorar la distribución de los recursos y destinar una mayor asignación económica para la educación, enfocado a la infraestructura de las instituciones educativas fiscales, en particular las de la zona rural, con dotación de bibliotecas, equipos tecnológicos, acceso a un buen servicio de internet y adecuación de espacios para los estudiantes que se encuentran en retorno progresivo, que brinde todo el apoyo socioemocional, académico y capacitación efectiva a sus docentes.

La pandemia dejó en evidencia una gran falta de autonomía de los estudiantes, ya que requieren de la presencia del docente para que les dicten o les den todas las indicaciones, y entre otros factores que han complicado la educación virtual, por lo que se debe potenciar las habilidades de autoaprendizaje.

Convirtiéndose en un verdadero desafío lograr que los miembros de las comunidades educativas de las instituciones fiscales de Manta adquieran habilidades de autoaprendizaje para la utilización de recursos tecnológicos disponibles.

En el Ecuador los instrumentos de gobernanza para la reducción de la brecha digital son altamente desarrollados y completos, pero por factores en relación con la tendencia política-ideológica que experimento el país en la última década, no se ha visto una constancia en su seguimiento y cumplimiento. Las clases virtuales con todos sus componentes llegaron para quedarse, y a pesar del retorno a clases presencial, que se está desarrollando de a poco a través del retorno progresivo, se podrá siempre recurrir a esta modalidad para complementar lo trabajado en las aulas regulares, tener como perspectiva a futuro que, minimizando el índice de analfabetismo digital o tecnológico en docentes y estudiantes, ya sea mediante capacitaciones, o de algún otro proceso pedagógico, reducirá también la brecha digital.

Agradecimientos

Se agradece a los docentes y compañeros por la dedicación y apoyo que ha brindado a este trabajo, por respetar las sugerencias e ideas, por la dirección en el desarrollo de este artículo.

Contribución de los autores

Ingrid Yadira Álava Chávez: Conceptualización, Metodología, Redacción – borrador original del artículo, Revisión y edición del artículo. **Dolores Esperanza Muñoz Verduga:** Resultados, Redacción – borrador original del artículo, Revisión y edición del artículo.

Conflictos de interés

Los autores declaran no tener ningún conflicto de interés.

Referencias bibliográficas

- Arias, F. G. (2012). El proyecto de investigación. Introducción a la metodología científica. 6ta. Fidas G. Arias Odón.
- Bolívar, C. R. (2013). Instrumentos y técnicas de investigación educativa. USA: Dananga Training and Consulting.
- Bravo Ávalos, S. P. (2019). Análisis de las políticas públicas y la brecha digital en el Ecuador periodo 2009-2016. *Master's thesis, Universidad Andina Simón Bolívar, Sede Ecuador.* <http://hdl.handle.net/10644/7041>

- Camacho, K. (2005). La brecha digital. Palabras en juego: enfoques multiculturales sobre las sociedades de la información, 61-71. <https://analfatecnicos.net/archivos/96.LaBrechaDigital-PalabrasEnJuego-KenlyCamacho.pdf>
- Díaz Lazo, J. P., & Florido Bacallao, R. (2011). Impacto de las tecnologías de la información y las comunicaciones (TIC) para disminuir la brecha digital en la sociedad actual. *Cultivos tropicales*, 32(1), 81-90.
- Educación, M. d. (2021). Retrieved from Mecapacito, Plataforma: <https://educacion.gob.ec/plataformamecapacito/>.
- Guallo Paca, J. F., & Guadalupe Arias, S. E. (2018). La alfabetización digital en Ecuador en el siglo XXI. *Atlante Cuadernos de Educación y Desarrollo*, (noviembre).
- IADB, P. G. (2022). Retrieved from <https://www.iadb.org/es/paises/ecuador/perspectivageneral>.
- Marín, J. (2003). El analfabetismo tecnológico. *Monografías.com. Lucas Morea/Sinexi, SA*. <http://www.monografias.com/trabajos12/elanalfc2.shtml>.
- Moreira, J. P., & López, J. (2017). *Un breve análisis de la brecha digital de acceso en el Ecuador*. Jornadas SARTECO.
- Moreno Inte, G. A., & Ramón Naula, L. C. (2019). Factores determinantes del analfabetismo digital del Ecuador en el año 2017. *Bachelor's thesis, Quito: UCE*.
- Pandemia, C. s. (2021).
- Sambache, D. D., & Cevallos, J. V. (2020). Brecha digital: conectividad y equipamiento en instituciones de educación fiscal en Ecuador. *Gigapp Estudios Working Papers*, 7 (166-182), 758-770.
- Santiago, T. L., & Parola, A. (2020). Educación en tiempos de pandemia: reflexiones de alumnos y profesores sobre la enseñanza virtual universitaria en España, Italia y Ecuador. *Revista latina de comunicación social*, 78, 1-21.
- Santoyo, A. S., & Martínez, E. M. (2003). La brecha digital: mitos y realidades. *Uabc*.
- Tecnologías de la Información y Comunicación-TIC*. (n.d.). Retrieved from <https://www.ecuadorencifras.gob.ec/tecnologias-de-la-informacion-y-comunicacion-tic/>
- Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times*. John Wiley & Sons.
- Ullauri, J. I. (2016). La tecnología: experiencias exitosas para la educación. Proyecto integrador de saberes, construcción desde la base del analfabetismo digital. *Mamakuna*, (2), 16-25.
- UNESCO, I. M. (2005). *Hacia las sociedades del conocimiento. Publicaciones Unesco. París*.
- Women., T. I.-G. (2021). Retrieved from <https://www.globalfundforwomen.org/initiatives/technologyinitiative/>
- World Internet Users Statistics and 2022 World Population Stats*. (2021, 07 08). Retrieved from <https://www.internetworldstats.com>.





Análisis de rendimiento en red de sensores inalámbrica con distintas topologías y cantidades de nodos

Wireless sensor network performance analysis with different topologies and number of nodes

Autores

✉^{1*} Manuel Eduardo Vinces Mendieta



✉² Darwin Patricio Loor Zamora



✉² Marely del Rosario Cruz Felipe



¹Instituto de Posgrado, Universidad Técnica de Manabí, Portoviejo, Ecuador.

²Facultad de Ciencias Informáticas, Universidad Técnica de Manabí, Portoviejo, Ecuador.

* Autor para correspondencia

Comó citar el artículo: Vinces Mendieta, M.E., Loor Zamora, D.P. & Cruz Felipe, M. del R. (2022). Análisis de rendimiento en red de sensores inalámbrica con distintas topologías y cantidades de nodos. *Informática y Sistemas: Revista de Tecnologías de la Informática y las Comunicaciones*. 6 (1),52-59. DOI: <https://doi.org/10.33936/isrtic.v6i1.4469>

Enviado: 04/03/2022

Aceptado: 08/05/2022

Publicado: 17/05/2022

Resumen

Las redes de sensores inalámbricas se basan en dispositivos que están interconectados entre sí, capaces de obtener información de su entorno; estas se construyen con varias tecnologías, entre ellas encontramos a Zigbee como protocolo de comunicación a corta distancia utilizado en aplicaciones de redes inalámbricas, maneja diferentes configuraciones sobre topologías de red en estrella, árbol y malla, soportando un gran número de sensores; el objetivo de esta investigación es analizar el rendimiento de la red de sensores inalámbrica comparando topologías con distintas cantidades de nodos; la metodología empleada se realizó sobre un ambiente de simulación utilizando la herramienta OPNET; creando escenarios para medir la latencia en la capa de aplicación sobre las topologías con diferentes números de nodos. Como resultado se determinó que mientras más nodos se aumente dentro de la red de sensores inalámbrica mejor rendimiento presenta la topología malla.

Palabras clave: Latencia; OPNET; WSN; ZigBee.

Abstract

Wireless sensor networks are based on devices that are interconnected with each other, capable of obtaining information from their environment; these are built with several technologies, among them, we find Zigbee as a short distance communication protocol used in wireless network applications, it handles different configurations on star, tree and mesh network topologies, supporting a large number of sensors; The objective of this research is to analyze the performance of the wireless sensor network comparing topologies with different numbers of nodes; the methodology used was performed on a simulation environment using the OPNET tool; creating scenarios to measure the latency in the application layer on topologies with different numbers of nodes. As a result, it was determined that the more nodes are increased within the wireless sensor network, the better the performance of the mesh topology.

Keywords: Latency; OPNET; WSN; ZigBee.



1. Introducción

Las redes de sensores inalámbrica (WSN) se basan en una serie de pequeños dispositivos electrónicos que tienen acceso al mundo exterior por medio de sensores, permiten monitorizar cualquier entorno por inaccesible y recóndito que parezca tal y como lo haría una red de insectos, los nodos sensoriales se distribuyen en el terreno y trabajan de forma coordinada para monitorizar los parámetros elegidos (Gascón, 2010).

Según la Unión Internacional de Telecomunicaciones sobre redes (UIT-T) recomienda que, una WSN es una red de nodos de sensores interconectados para el intercambio inalámbrico de datos detectados; el cual consta de varios nodos sensores que van desde unas pocas decenas hasta miles de nodos, que trabajan juntos para monitorear un área en busca de datos ambientales, además, los autores identificaron dos tipos de redes de sensores: estructuradas y no estructuradas (Rueda & Portocarrero, 2017).

Sobre la localización en WSN, se efectuó un estudio sobre algoritmos, técnicas de medición, aplicaciones y desafíos en donde dividieron los algoritmos enfocados a dos categorías: las centralizadas cuyo algoritmo permite medir todas las distancias de los sensores a una ubicación central permitiendo conocer todas las posiciones de los sensores y las distribuidas cuyo algoritmo permite a cada sensor individual localizar su propia posición mediante la medición de distancia de otros nodos ancla, estos algoritmos a su vez se sub dividen en algoritmos dentro y fuera de rango (Paul & Sato, 2017).

Con el tiempo a las WSN se las reconoce como estructuras de red distribuidas y autoconfiguradas que constan de un gran número de nodos sensores, creando distintos escenarios en topologías estrella, árbol y malla; mediante simulación donde se lleva a cabo con una cantidad de 3, 5, 10, 15, 20, 25, 30, 35 y 40 nodos, para observar el rendimiento de la red sobre las métricas como el retardo de extremo a extremo (ETD), el rendimiento y el número de saltos (Mounika, 2018).

Zigbee es el nombre de la especificación de un conjunto de protocolos de comunicación de alto nivel basado en el estándar IEEE 802.15.4. El cual se utiliza para crear redes inalámbricas de área personal de bajo consumo energético, en comparación con otras tecnologías similares como Bluetooth y Wifi. Tiene amplia utilización en la automatización de casas sobre domótica, obtención de datos de equipo médico, así como cualquier otro dispositivo de bajo consumo de energía y ancho de banda que requiera conexión inalámbrica (Alonso, 2020).

Para la evaluación del rendimiento del protocolo de enrutamiento Zigbee bajo varias condiciones, en un trabajo investigativo se modelaron y simularon dos escenarios diferentes, en el primero, se probó las características topológicas y el IEEE 802.15.4 en términos de rendimiento, retardo de nodo a nodo y la cifra de

routers sobre tres diseños de red (estrella, árbol y malla); y el segundo escenario investigó la función de auto-reparación en una topología de malla desde el diseño anterior hasta el soporte de un gran número de puntos finales con enrutamiento dinámico (Abdulrahman & Ali Yaseen, 2015).

En el trabajo publicado “Simulación de redes de sensores inalámbricos: Un modelo energético a nivel nodo-sensor bajo las especificaciones IEEE 802.15.4 y Zigbee”, se diseñó e implementó utilizando el simulador para redes de eventos discretos OPNET, basado en tres modelos de nodo, cada uno consta de un modelo de flujo de datos y un modelo energético (Amarillo Rojas & Trujillo Arboleda, 2015).

La comunicación se ha convertido en una parte muy importante sobre un sistema de WSN en cuanto a estar libre de interferencias, atenuación, diafonía y desvanecimiento; cualquiera de estos factores es un problema grave en sistemas de comunicación; para resolver estos problemas, se consideraron redes de nodos móviles y fijos para el funcionamiento eficiente empleando la tecnología Zigbee y su diseño fue realizado con la herramienta de simulación OPNET; se basaron en topologías de árbol, estrella y de malla entre los dos sistemas (Adaramola & Olasina, 2018).

La validación de los nodos sensores es una tarea desafiante por la naturaleza del entorno, el reemplazo o la recarga de la batería de este pequeño dispositivo en un lugar inaccesible, y la distribución de un gran número de nodos sensores; en esta investigación, propusieron el diseño e implementación de un modelo de validación para WSNs y poder descubrir nodos malos en cualquier distribución WSN, dando como resultado obtenido de la simulación que, al aumentar el número de nodos, el número de nodos malos aumentará en un objetivo (Almajidi & Pawar, 2019).

Es por ello que la presente investigación con base en los antecedentes de trabajos mencionados anteriormente se va a comparar redes inalámbricas utilizando varios dispositivos sensores, con el objetivo de analizar el rendimiento de una WSN sobre distintas topologías que estén soportadas con diferentes cantidades de nodos; el diseño para cada escenario será realizado mediante simulación, utilizando la tecnología Zigbee para evaluar el rendimiento de la red basándose en el parámetro de latencia.

2. Materiales y Métodos

Se realizó una investigación de tipo cuasi-experimental para analizar el rendimiento de una WSN utilizando distintas topologías de red con diferentes cantidades de nodos, diseñado sobre un área de 140x160 m² en donde los escenarios serán creados en un ambiente de simulación para eventos discretos soportado por el protocolo de comunicaciones inalámbricas

Zigbee.

2.1 Topologías de redes Zigbee

La tecnología Zigbee utiliza tres tipos de topologías, estas son la de estrella, malla y árbol tal como se muestra en la figura 1 en donde se puede observar cómo estarían desplazados los dispositivos coordinadores, router y dispositivos finales, basado en el estándar de comunicaciones para redes inalámbricas IEEE 802.15.4, es importante resaltar que la administración de los caminos dentro de la red de sensores es tarea del coordinador.

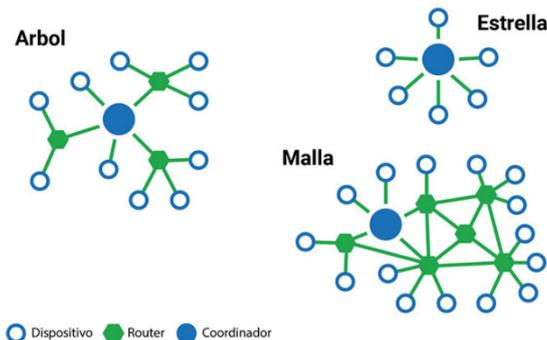


Figura 1. PTopologías estrella, malla y árbol.

La topología inalámbrica en estrella es la más sencilla del resto de topologías por ser su tecnología definida de un solo salto; esta topología contiene un grupo de dispositivos, o más nodo finales llamados dispositivos miembros. El dispositivo nodo coordinador es el responsable de establecer y crear la red, se utiliza para gestionar toda la red, comprobar el estado de cada nodo sensor, y almacenar toda la información de la red, como la detección, la ubicación de los nodos sensores, batería, etc.

La topología inalámbrica árbol, se compone de tres dispositivos: el nodo coordinador, el nodo router y los dispositivos del nodo final (Cuomo et al., 2013). El nodo coordinador actúa como la raíz del árbol, y el nodo router representa la raíz de los sub-árbol, mientras que el nodo final representa las hojas del árbol. La topología de malla, es la topología de red inalámbrica más complicada en su creación, se compone de tres dispositivos diferentes, estos son nodo coordinador, nodo router, y el nodo final. Tanto el nodo coordinador como el nodo router se consideran conexiones múltiples entre uno y otro, los dispositivos de los nodos finales son hijos de los nodos coordinadores y router (Guo et al., 2011).

2.2 Nodos de red IEEE 802.15.4/Zigbee

Cada uno de los nodos de red representa un sensor inalámbrico, el cual basa su configuración y demás en el estándar descrito en la norma Zigbee del 2007, la cual contempla tres tipos de nodos de red: dispositivo final, coordinador PAN y enrutador (Alliance, 2015).

Es preciso mencionar que tanto la capa física como la capa

MAC se han construido de tal manera que cumplan con las especificaciones impuestas por el estándar IEEE 802.15.4-2003 (IEEE SA - IEEE 802, 2003).

- **Dispositivo final.** Es un dispositivo de la norma 802.15.4-2003 con funciones reducidas (RFD) o completas (FFD), y que no es un coordinador Zigbee ni tampoco un enrutador Zigbee. Este tipo de nodo tiene las siguientes características (Cunha et al., 2007):

- Componente de red opcional.
- No permite asociación.
- No participa en el enrutamiento.
- Puede ser un RFD que implementa un subconjunto reducido de la pila de protocolos.

- **Coordinador Zigbee.** Es un coordinador PAN según la norma 802.15.4-2003, el cual trabaja en la capa de red del protocolo Zigbee. Sus características principales son (Cunha et al., 2007):

- Uno para cada red Zigbee.
- Inicia y configura la formación de la red.
- Es un coordinador de red de área personal de la norma IEEE 802.15.4.
- Es el enrutador cuando la red está formada.
- Es un FFD, el cual implementa toda la pila de protocolos.

- **Enrutador.** Es un FFD, según la norma 802.15.4-2003, el cual, aunque no es un coordinador Zigbee, puede actuar como un coordinador PAN de la norma 802.15.4-2003 dentro de su espacio de operación. Además, este dispositivo es capaz de enrutar mensajes entre dispositivos y asociar dispositivos tanto en su capa MAC como en su capa de red. Sus principales características son (Cunha et al., 2007):

- Componente opcional de red.
- Se asocia con el coordinador Zigbee (ZC) o con un router Zigbee previamente asociado (ZR).
- Puede ser un coordinador PAN según la norma IEEE 802.15.4.
- Participa en el enrutamiento multi-salto de los mensajes.
- Es un FFD, el cual implementa toda la pila de protocolos.

2.3 Red de Sensores Inalámbrica (WSN)

La WSN está constituida por sensores de bajo costo con una interfaz inalámbrica que nos permiten, mediante elementos sensores, recolectar todo tipo de información del mundo físico. Son utilizadas para supervisar de manera cooperativa cualquier fenómeno que pueda ser medido, por ejemplo, condiciones ambientales tales como temperatura, sonido, vibración o, en este caso, agentes contaminantes en las plantas de tratamiento de agua automatizada (Sohraby et al., 2007). La WSN puede incluir nodos sensores que son, a su vez, actuadores y que permiten controlar diferentes dispositivos utilizando el mismo medio de

comunicación inalámbrico.

Actualmente, existen muchos tipos de redes que pueden ser enmarcados dentro de las WSNs. Estos son tipos de WSNs definidos para aplicaciones particulares, como, por ejemplo, las redes inalámbricas de sensores multimedia (wireless sensor multimedia networks, WSMNs), las redes de sensores submarinas (underwater sensor networks, USNs), las redes de área corporal (body area networks, BANs), las redes de área vehicular (vehicular área sensor networks, VANs), entre otros (Almajidi & Pawar, 2019).

Las redes de sensores inalámbricos (WSN) son capaces de detectar eficientemente diversos parámetros con alta precisión. Básicamente, una red de sensores inalámbricos se compone de tres elementos principales; unidad de sensor (para tomar mediciones), unidad de cálculo (para procesar los datos) y la unidad de comunicación (para permitir la comunicación entre los nodos inalámbricos) (Risteska Stojkoska et al., 2014).

2.4 Simulador para la WSN

El método empleado para esta investigación fue la simulación, donde se pudo evidenciar que existen varios simuladores que permiten diseñar un modelo de red completo. Entre estas herramientas de simulación encontramos al Network Simulator 2 (NS-2), NS3, OMNET ++, OPNET, MATLAB, que soportan las funciones de desarrollo del protocolo de comunicaciones inalámbricas Zigbee; todos los simuladores nombrados proporcionan la simulación de eventos discretos para evaluar diversos parámetros dentro de la red que es muy empleado por investigadores.

La simulación con OPNET es la plataforma que se convierte en la principal herramienta proveedora de soluciones para la estimación de diseños y aplicaciones de redes inalámbricas; soporta el modelo de red inalámbrica basado en Zigbee dando varios elementos de una red inalámbrica Zigbee (Coordinador Zigbee, Router Zigbee, Dispositivo final Zigbee) estos elementos pueden ser establecidos como fijos o móviles (Zhao & Xie, 2011).

Para el análisis dentro de la simulación de la WSN se crearon nueve escenarios, las topologías fueron combinadas de acuerdo a la distancia entre nodos, a continuación, se detalla en la Tabla 1 los parámetros a utilizar para cada escenario de simulación dentro de la WSN.

3. Resultados y Discusión

Para la realización de la presente investigación se comparó el rendimiento dentro de una WSN utilizando tres topologías de red combinadas a distintas distancias manejada por la tecnología

Tabla 1. Parámetros para cada escenario de simulación.

Fuente: Los autores.

Parámetros de Topología	Escenario 1	Escenario 2	Escenario 3
Distancia entre nodos (metros)	40 metros	20 metros	10 metros
Números de Nodos	25	49	195
Coordinador	1	1	1
Router	8	8	8
Dispositivo final	16	40	186
Poder de transmisión	0.05	0.05	0.05
Banda de Transmisión	2.4 Ghz	2.4 Ghz	2.4 Ghz

Zigbee; utilizando como parámetro de medición la latencia sobre la capa de aplicación en cada uno de los escenarios que fueron diseñados mediante la herramienta de simulación OPNET y serán expuestos a continuación.

3.1 Primer Escenario

Sobre el primer escenario, la Figura 2 nos muestra cómo se elaboró la WSN dentro de un área aproximada de 140 x 160 m², donde se aplicó una topología de red en estrella, la misma que fue duplicada para una topología de red en árbol y en malla creándose de esta forma las tres topologías para este primer escenario combinado, en donde se distribuyeron dispositivos entre ellos 1 coordinador Zigbee, 8 dispositivos router y 186 dispositivos finales, manteniéndose a una distancia entre nodos de 10 metros.

3.2 Segundo Escenario

Para el segundo escenario la Figura 3, muestra la WSN con un área aproximada de 140 x 160 m², donde se diseñó una red en estrella, y se la duplicó para una red en árbol y en malla creando así las tres topologías para este segundo escenario combinado, en donde se distribuyeron dispositivos entre ellos 1 coordinador Zigbee, 8 dispositivos router y 40 dispositivos finales, manteniendo una distancia entre nodos a 20 metros.

3.3 Tercer Escenario

Para el tercer escenario la figura 4, muestra la WSN con un área aproximada de 140 x 160 m², donde se diseñó una red en estrella, y se la duplicó para una red en árbol y en malla creando así las tres topologías para este tercer escenario compartido, en donde se distribuyeron dispositivos entre ellos 1 coordinador Zigbee, 8 dispositivos router y 16 dispositivos finales, manteniendo una

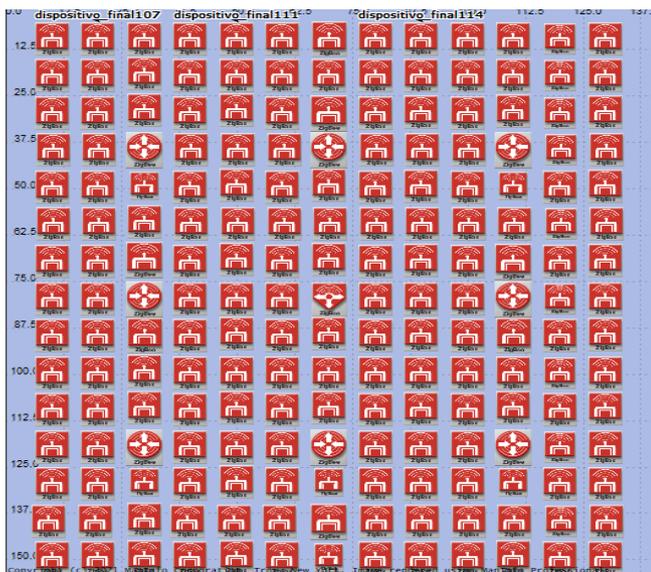


Figura 2. Escenario WSN a una distancia de 10 metros

Fuente: Los autores.



Figura 4. Escenario WSN a una distancia de 40 metros.

Fuente: Los autores.

3.4 Latencia en topologías a 10 metros de distancia

La simulación de retardo extremo a extremo (latencia), se lo identifica como el retardo de los paquetes recibidos por la capa MAC 802.15.4 sobre los nodos de la red que son reenviados a la capa superior; refiriéndonos sobre este escenario se ha combinado a las tres topologías a una misma distancia de 10 metros entre nodos, donde se pudo evidenciar los resultados comparativos como lo muestra la Figura 5 en función del número de nodos dentro de la WSN.

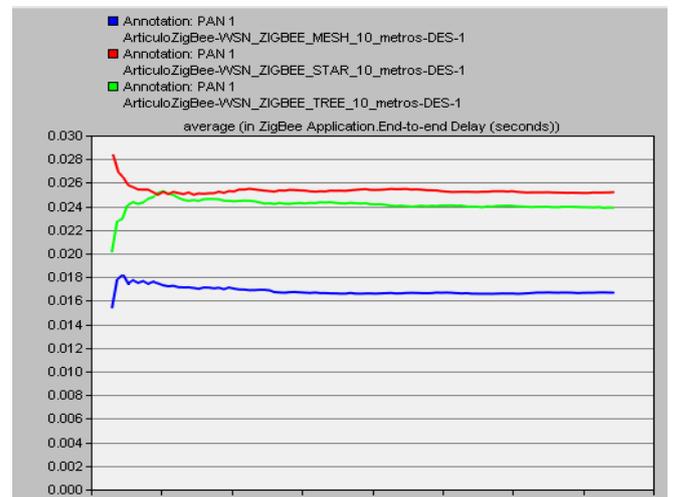


Figura 5. Combinación de 3 escenarios utilizando las topologías estrella, árbol y malla a 10 metros de distancia

Fuente: Los autores.

En la Figura 5 se observa que la topología de malla tiene un menor retardo llegando a los 0.01669 segundos, evidenciando mejor rendimiento de la red basándose en el número de nodos dentro de la WSN al estar entre cortas distancias sobre este escenario combinado a 10 metros, seguido de la topología árbol que tiene

distancia entre nodos a 40 metros.

En una red en estrella, todos los nodos finales se comunican directamente con la estación base. En una topología de árbol, el coordinador y el enrutador pueden tener hijos y actuar como padres. Utilizando una red de malla, los paquetes pasan por una serie de rutas para llegar a su destino utilizando la ruta más viable para evitar las zonas muertas. Una vez configurado el parámetro definido para cada uno de los escenarios utilizando la tecnología Zigbee, se procede a comparar el retardo de la red dentro de la WSN a distintas distancias.



Figura 3. Escenario WSN a una distancia de 20 metros

Fuente: Los autores.



un retardo de los 0,02389 segundos y por último la topología estrella aparece con un retardo de los 0,02520 segundos.

3.5 Latencia en topologías a 20 metros de distancia

Refiriéndonos sobre este segundo escenario se ha combinado a las tres topologías a una misma distancia de 20 metros entre nodos, donde se pudo evidenciar los resultados comparativos como lo muestra la Figura 6 en función del número de nodos dentro de la WSN.

Sobre la Figura 6 se observa que la topología de malla tiene un menor retardo llegando a los 0.02956 segundos, evidenciando mejor rendimiento de la red basándose en el número de nodos dentro de la WSN al estar entre cortas distancias sobre este escenario combinado a 20 metros, esta vez cambió el orden de retardo en donde identificamos a la topología estrella que tiene un retardo del 0,03743 segundos y la topología árbol tiene un retardo del 0.04238 segundos, dando a denotar un cambio significativo en este escenario de la topología malla con la topología árbol, en donde la cantidad de nodos ha influido en comparación del retardo con el primer escenario.

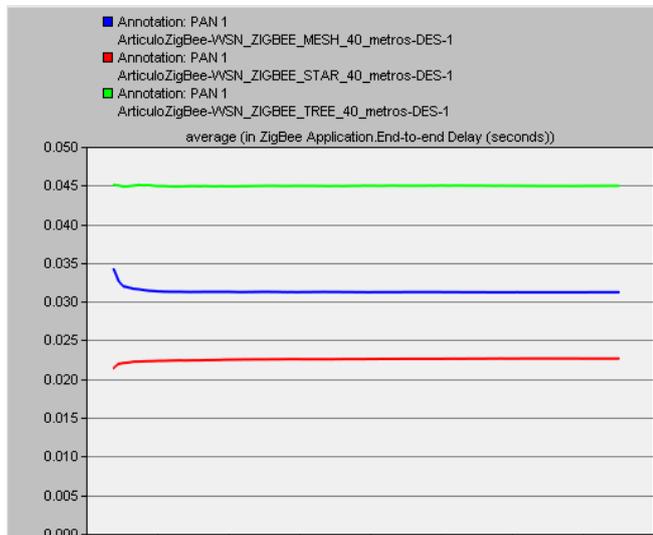


Figura 6. Combinación de 3 escenarios utilizando las topologías estrella, árbol y malla a 20 metros de distancia.
Fuente: Los autores.

3.6 Latencia en topologías a 40 metros de distancia

Sobre este escenario se ha combinado a las tres topologías a una misma distancia de 40 metros entre nodos, donde se pudo evidenciar los resultados comparativos como lo muestra la Figura 7 en función del número de nodos dentro de la WSN.

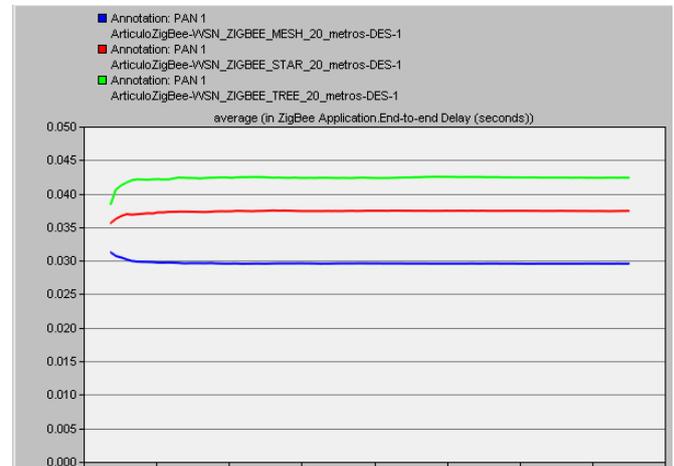


Figura 7. Combinación de 3 escenarios utilizando las topologías estrella, árbol y malla a 40 metros de distancia.
Fuente: Los autores.

En la Figura 7 se observa que ahora aparece la topología estrella con menor retardo sobre sus antecesores llegando a los 0.02265 segundos, esto se refleja basándose en la cantidad menor de nodos dentro de la WSN al estar a una distancia de 40 metros entre nodos, mientras que la topología malla tiene un retardo del 0,03121 segundos y la topología árbol tiene un retardo del 0.04497 segundos, dando a denotar un cambio significativo de la topología estrella sobre las otras topologías donde la cantidad de nodos ha influido en este tercer escenario.

De los resultados obtenidos podemos decir que, la de mejor rendimiento dentro de la WSN entre una distancia de 10 y 20 metros es la topología malla, mientras que en el tercer escenario a 40 metros se pudo denotar un cambio significativo en donde resalto con menor retardo la topología estrella; sin embargo, se realizó un análisis estadístico a base de los datos exportados de cada uno de los escenarios comparados, en donde se utilizó un método de validación el cual permitió definir si en realidad la topología estrella, malla o árbol es la recomendada para la WSN.

3.7 Análisis estadístico comparativo entre las topologías estrella árbol y malla.

Sobre el análisis estadístico realizado se puede observar en la Figura 8 una comparativa entre las topologías estrella, árbol y malla a 10, 20 y 40 metros de distancia entre los nodos aplicado sobre la WSN, mediante la exportación de los datos obtenidos de la simulación en OPNET.

Podemos evidenciar una comparativa de las topologías basándose en un gráfico de cajas, en donde se puede denotar que a mayor número de nodos mejor será la latencia utilizando la topología malla sobre esta WSN cuando se manejan gran número de nodos

distribuidos a 10 metros, mientras que en la topología estrella se puede denotar sobre la Figura 8 que la latencia a 10 y 40 metros no es mucha su diferencia significativa sobre la topología malla.

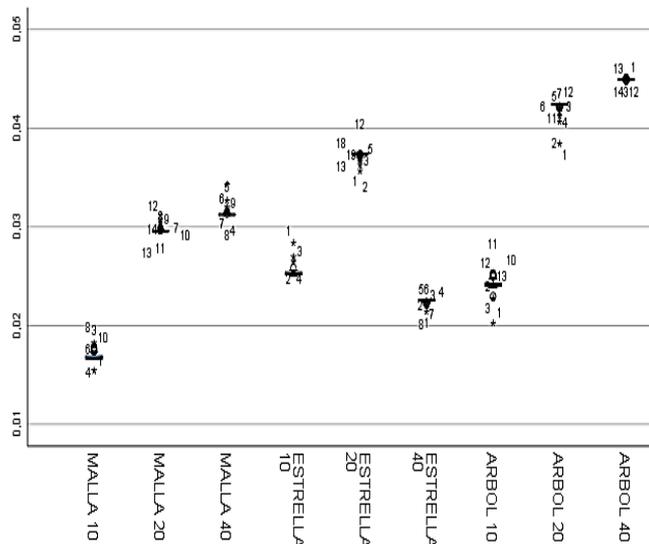


Figura 8. Combinación de 9 escenarios utilizando la topología estrella, árbol y malla a 10, 20 y 40 metros de distancia.

Fuente: Los autores.

4. Conclusiones

A partir de la conclusión de la WSN utilizando el protocolo de comunicaciones inalámbrica ZigBee se compararon las diferentes topologías soportadas por esta tecnología obteniendo que la más idónea es la topología en malla.

A partir de la comparación empleando la topología en malla, la simulación con diferentes cantidades de nodos se obtuvo que el mejor escenario fue el primero con 195 nodos a una distancia de 10 metros dentro de la WSN.

Se comprueba, a partir del análisis estadístico, una diferencia significativa entre los datos exportados de los 9 escenarios simulados entre los nodos corroborando que la topología en malla a mayor cantidad de nodos tiene un menor retardo, comprobándose que es la mejor topología dentro de la WSN.

Contribución de los autores

Manuel Eduardo Vincés Mendieta: Conceptualización, metodología y análisis formal del artículo; Diseño y configuración del software de simulación; Redacción del borrador original del artículo. **Darwin Patricio Loor Zamora:** Redacción del borrador original del artículo, Conceptualización y metodología, Participación en la implementación de la simulación. **Marely del Rosario Cruz Felipe:** Conceptualización y la investigación, Visualización y análisis de los resultados, Supervisión del proyecto y revisión y edición del artículo.

Conflictos de interés

Los autores declaran no tener ningún conflicto de interés.

Referencias bibliográficas

- Abdulrahman, N., & Ali Yaseen, F. (2015). Performance Evaluation of Zigbee Routing Protocol under Various Conditions using OPNET Modeler. *International Journal of Computer Applications*, 117(18), 25-31. <https://doi.org/10.5120/20656-3284>
- Adaramola, O. J., & Olasina, J. R. (2018). Network Investigation and Performance Analysis of ZigBee Technology using OPNET. *Journal of Advances in Computer Engineering and Technology*, 4(4), 209-218.
- Alliance, Z. (2015). ZigBee Specification. *Standard*, Oct.
- Almajidi, A. M., & Pawar, V. P. (2019). A New System Model for Sensor Node Validation by Using OPNET. *Wireless Personal Communications*, 108(4), 2389-2401. <https://doi.org/10.1007/s11277-019-06527-3>
- Alonso, C. C. (2020). *Proyecto de Investigación*.
- Amarillo Rojas, M. O., & Trujillo Arboleda, L. C. (2015). Simulación de Redes de Sensores Inalámbricos: un Modelo Energético a Nivel de Nodo-Sensor Bajo las Especificaciones Ieee 802.15.4tm y Zigbee. *Publicaciones e Investigación*, 9(0), 13. <https://doi.org/10.22490/25394088.1430>
- Cunha, A., Alves, M., & Koubâa, A. (2007). *Implementation details of the time-division beacon scheduling approach for ZigBee cluster-tree networks*. <https://recipp.ipp.pt/handle/10400.22/3940>
- Cuomo, F., Abbagnale, A., & Cipollone, E. (2013). Cross-layer network formation for energy-efficient IEEE 802.15.4/ZigBee Wireless Sensor Networks. *Ad Hoc Networks*, 11(2), 672-686. <https://doi.org/10.1016/J.ADHOC.2011.11.006>
- Gascón, D. (2010). *Redes de Sensores Inalámbricos, la tecnología invisible*. 180-182. Gascón, D. (2010). *Redes de Sensores Inalámbricos, la tecnología invisible*. Tecnología y Sociedad. Artículo de la revista BIT del Colegio Oficial de Ingenieros de Telecomunicación (COIT),[consulta: 21 febrero 2019]. Available: <http://www.libelium.com/libelium-downloads/libelium-bit-coit.pdf>.
- Guo, L., Peng, Y., Wang, X., Jiang, D., & Yu, Y. (2011). Performance evaluation for on-demand routing protocols based on OPNET modules in wireless mesh networks. *Computers & Electrical Engineering*, 37(1), 106-114. <https://doi.org/10.1016/J.COMPELECENG.2010.10.002>
- Hammoodi, I. S., Stewart, B. G., Kocian, A., & McMeekin, S. G. (2009). A Comprehensive Performance Study of



- OPNET Modeler For ZigBee Wireless Sensor Networks. NGMAST 2009 - *3rd International Conference on Next Generation Mobile Applications, Services and Technologies*, 357-362. <https://doi.org/10.1109/NGMAST.2009.12>
- IEEE SA - IEEE 802.* (2003). <https://standards.ieee.org/ieee/802.15.4/3388/>
- Mounika, P. (2018). Performance analysis of wireless sensor network topologies for Zigbee using riverbed modeler. *Proceedings of the 2nd International Conference on Inventive Systems and Control, ICISC 2018, Icisc*, 1456-1459. <https://doi.org/10.1109/ICISC.2018.8399050>
- Paul, A. K., & Sato, T. (2017). Localization in wireless sensor networks: A survey on algorithms, measurement techniques, applications and challenges. *Journal of Sensor and Actuator Networks*, 6(4). <https://doi.org/10.3390/JSAN6040024>
- Risteska Stojkoska, B., Popovska Avramova, A., & Chatzimisios, P. (2014). Application of wireless sensor networks for indoor temperature regulation. *International Journal of Distributed Sensor Networks*, 2014. <https://doi.org/10.1155/2014/502419>
- Rueda, J. S., & Portocarrero, J. M. T. (2017). Similitudes y diferencias entre Redes de Sensores Inalámbricas e Internet de las Cosas: Hacia una postura clarificadora. *Revista Colombiana de Computación*, 18(2), 58-74. <https://doi.org/10.29375/25392115.3218>
- Sohraby, K., Minoli, D., & Znati, T. (2007). Basic Wireless Sensor Technology. En *Wireless Sensor Networks* (pp. 75-92). <https://doi.org/10.1002/9780470112762.ch3>
- Zhao, W., & Xie, J. (2011). OPNET-based modeling and simulation study on handoffs in Internet-based infrastructure wireless mesh networks. *Computer Networks*, 55(12), 2675-2688. <https://doi.org/10.1016/J.COMNET.2011.04.013>



Análisis y evaluación de riesgos aplicado a la seguridad de la información bajo la norma ISO/IEC 27002: Caso de estudio Distribuidora Bravel

Analysis and risk assessment applied to information security under the ISO/IEC 27002 standard: Case study Bravel distributor

Autores

✉^{1*} *María Angélica Velepucha Sánchez*



✉² *Jessica Morales Carrillo*



✉ *Marco Fernando Pazmiño Campuzano*



^{1,3} Instituto de Posgrado, Universidad Técnica de Manabí, Portoviejo, Ecuador.

² Grupo de Investigación SISCOM, Carrera de Computación, Escuela Superior Politécnica Agropecuaria de Manabí Manuel Félix López, Ecuador.

* Autor para correspondencia

Comó citar el artículo: Velepucha Sánchez, M. A., Morales Carrillo, J., & Pazmiño Campuzano, M. F. (2022). Análisis y evaluación de riesgos aplicado a la seguridad de la información bajo la norma ISO/IEC 27002: Caso de estudio Distribuidora Bravel. *Informática y Sistemas: Revista de Tecnologías de la Informática y las Comunicaciones*, 6(1), 60-70. DOI: <https://doi.org/10.33936/isrtic.v6i1.4473>

Enviado: 21/03/2022

Aceptado: 28/04/2022

Publicado: 02/06/2022

Resumen

El propósito del presente estudio fue diagnosticar el grado de los riesgos que sufren la información y activo que se encuentra en la distribuidora Bravel, ya que muchas empresas carecen de controles de seguridad por lo que no pueden garantizar la seguridad de la información. Esta investigación se encaminó al análisis y evaluación para la gestión de activos de información mediante las secciones que tiene la norma ISO/IEC 27002. Se presentan los resultados aplicando la metodología del análisis y evaluación de riesgos con el diseño de diversos instrumentos como cuestionarios de los ítems que ha establecido la norma, en la cual nos permitirá conocer el riesgo que tiene la organización con respecto a la seguridad de la información como: clave de seguridad, datos del personal del área informática y usuarios de los sistemas, y realizar un testeó que permitieron establecer el diagnóstico de seguridad actual. Finalmente, y de acuerdo a los resultados del análisis y evaluación de los riesgos, se proponen los controles de seguridad para que sean integrados hacia el futuro dentro de un SGSI que responda a las necesidades de seguridad informática y de la información acorde a sus necesidades.

Palabras claves: Análisis; Información; Riesgo; Seguridad.

Abstract

The purpose of this study was to diagnose the degree of risks suffered by the information and assets found in the Bravel distributor, since many companies lack security controls, so they cannot guarantee the security of the information. This research was directed to the analysis and evaluation for the management of information assets through the sections that the ISO / IEC 27002 standard has. The results are presented applying the methodology of analysis and risk evaluation with the design of various instruments such as questionnaires of the items that the standard has established, in which it will allow us to know the risk that the organization has with respect to information security, such as: security key, data of the personnel of the computer area and users of the systems, and carry out a test that allowed to establish the current safety diagnosis. Finally, and according to the results of the analysis and evaluation of the risks, the security controls are proposed to be integrated into the future within an ISMS that responds to the needs of computer and information security according to their needs.

Keywords: Analysis; Information; Risk; Security.



1. Introducción

Esta investigación tuvo como objetivo específico diagnosticar el índice de riesgo que sufre la información procesada y activo que se encuentra dentro de la Distribuidora Bravel, ya que actualmente la información son los activos más valiosos que se presenta dentro de una organización. Por ende, este estudio tiene un aporte teórico al campo de conocimiento; debido a los resultados obtenidos de alguna forma aportarán significativamente a otras respectivas investigaciones.

El análisis y evaluación de riesgos, la verificación de la existencia de controles de seguridad existentes, las pruebas con software y el monitoreo de los sistemas de información permiten establecer el estado actual de la organización, identificar las causas de vulnerabilidades y proponer soluciones de control que permitan su mitigación (Solarte Solarte, 2015). Estas acciones deben estar enmarcadas en un proceso lógico, sistemático, documentado, que pueda ser difundido interno para garantizar la gestión correcta de la seguridad informática y de la información, siguiendo el ciclo de mejora continua (planear, hacer, verificar y actuar - PHVA) (Salazar Choez, 2018).

Inicialmente se trata de comprender la norma ISO/IEC27002 en cada uno de los dominios, para determinar el alcance de su aplicabilidad. Una vez definidos los dominios y determinados los activos existentes, se aplica la metodología para realizar análisis y evaluación de riesgos respecto a los tres criterios de información que son la confidencialidad, la integridad y la disponibilidad de la información.

La siguiente tarea consiste en la verificación de la existencia de controles de seguridad existentes en la empresa y su aplicación; ya que pueden estar incluidos dentro de los procesos de calidad organizacionales. Estos deben ser comparados los resultados con el software eMarisma e Isotools, definidos en la norma ISO/IEC 27002 como políticas y procedimientos; el resultado servirá de base para el diseño, la implementación e implantación futura de un SGSI como respuesta a los riesgos encontrados (Enríquez Collaguazo, 2018).

En el artículo se muestra un conjunto de instrumentos que posibilitan realizar el análisis y evaluación de riesgos, las técnicas utilizadas para conocer y comprender el estado actual de las organizaciones empresariales evaluadas y que pueden ser aplicados para realizar procesos de auditoría a la seguridad. Finalmente se explica la metodología para aplicar el proceso de análisis y evaluación de los riesgos desde la fase inicial de conocimiento del sistema, la fase de identificación de las vulnerabilidades, amenazas y riesgos de seguridad determinando el nivel de riesgo a que se ve expuesta la organización, por probabilidad e impacto en los criterios de confidencialidad, integridad y disponibilidad de la información, para luego aplicar el software que mejor resultado arroje.

2. Desarrollo

2.1. La seguridad

La seguridad, aún existe divergencia en los criterios de almacenamiento, acceso y transmisión de información de los pacientes porque los requerimientos físicos y lógicos varían para cada empresa, equipo desarrollador o intereses particulares. En general, las amenazas y ataques sobre una red de datos, obligan a establecer parámetros para prevenir o mitigar estas falencias, por medio de regulaciones y estándares (Enríquez Collaguazo, 2018).

La seguridad informática ha tomado gran auge, debido a las cambiantes condiciones y nuevas plataformas tecnológicas disponibles. La posibilidad de interconectarse a través de redes, ha abierto nuevos horizontes a las empresas para mejorar su productividad y poder explorar más allá de las fronteras nacionales, lo cual lógicamente ha traído consigo, la aparición de nuevas amenazas para los sistemas de información (Armendáriz, 2017).

Actualmente las TICS (Tecnologías de la información y de la comunicación) es la base fundamental para el desarrollo y superación de un país, la información que en ellas se opera son consideradas como datos muy valiosos para todas organizaciones públicas como privadas con el fin de que tengan éxito, es por eso que debemos ofrecer seguridad a la información (Ladino, 2011).

Los modelos de evaluación de sistemas de seguridad permiten conocer, dentro de una organización, la madurez con la cual la organización lleva adelante sus políticas, actividades, usa sus herramientas y métodos, etc., en pos de su seguridad. Por lo tanto, para una organización preocupada en su seguridad informática, resulta necesario contar con estándares, así como herramientas apropiadas para evaluar el grado de adecuación con dichos estándares.

Los procesos de admisión y nivelación han operado normalmente sin que exista una evaluación o auditoría a las aplicaciones que los soportan situación que incrementa los riesgos a los que cada proceso y actividades están expuestas. La evaluación de dichos procesos con el objetivo de establecer recomendaciones que permitan mejorar los controles o deficiencias encontradas y comunicarlas formalmente a los directivos (Velasco, 2008).

En la actualidad todas las organizaciones dentro de nuestro medio se basan en la información para tomar decisiones que permitan la continuidad de los negocios, “innovar así los activos importantes para la empresa, siendo necesario protegerla ante cualquier evento que puede causar corrupción en los datos, dando la importancia de la información, organizaciones internacionales de estandarización han elaborado normas de buenas prácticas para seguridad y buen uso de la información y de los activos en general” (Figueroa Moran, 2017).

La masiva incorporación de las tecnologías de las comunicaciones y la información al mundo de los negocios ha generado profundas transformaciones en las responsabilidades de la alta gerencia en las organizaciones. Es importante entender que hoy la inversión en TI no trata solamente de implementar soluciones tecnológicas, sino que se focalizan en implementar cambios y transformaciones en la organización posibilitados por TI. Esto genera mayor complejidad y mayores riesgos que en el pasado, y la aplicación de las prácticas tradicionales de gestión no son suficientes (Mora Palacios, 2017).

El amplio uso de las tecnologías de información en los negocios hace que cada vez sea más fácil la expansión de éstos. La comunicación con clientes que se encuentran en una ciudad o país diferente al de ubicación de la empresa, la posibilidad de realizar transacciones comerciales vía web y en general, la facilidad del uso de la tecnología y la globalización de la información para todas las personas ha contribuido a que las organizaciones crezcan cada vez más rápido (Pazmiño Flores, 2019).

La seguridad informática.

La seguridad informática se define como la protección de los recursos valiosos, que corresponden a un ente legítimo propietario, de los posibles riesgos y ataques efectuados por agentes no autorizados. De igual manera la seguridad informática tiene como finalidad en proteger los recursos de un sistema informático como la Información, Servicios y Arquitecturas. ISOTOOLS, (Bermudez Molina, 2015) Software de Gestión para la excelencia empresarial define que la seguridad informática es una rama de la ingeniería de sistemas que se encarga de coordinar acciones para proteger la integridad y la privacidad de la información que ha sido almacenada en un sistema informático.

Afirman que: “La seguridad informática se puede definir como un conjunto de procedimientos, dispositivos y herramientas encargadas de asegurar la integridad, disponibilidad y privacidad de la información en un sistema informático e intentar reducir las amenazas que pueden afectar al mismo” (Bermudez Molina, 2015).

Al mismo tiempo la seguridad informática consiste en proteger la integridad de la información no sólo es una cuestión de ordenarlas, clasificarlas y almacenarlas porque no es suficiente. Las amenazas que velan a los datos tienen estrategias sofisticadas. Los virus informáticos, por ejemplo, son programas perjudiciales que se sitúan en la memoria RAM de los ordenadores del usuario, impidiendo el normal acceso a los datos que allí reposa.

Asimismo, existen los mencionados hackers o peritos del saqueo informático, que se encargan de bloquear los sistemas para acceder a bases de datos confidenciales y utilizar dicha información para fines desconocidos. Toda empresa u organización tienen información de carácter confidencial, de mayor o menor medida, con la finalidad de optimizar la seguridad requerida para la protección de la integridad de los recursos informáticos de las actividades de la misma (López Uriarte, 2017).

En la actualidad considera que la seguridad de los datos y la información comprenden 3 principios esenciales que son: Confidencialidad se trata de la cualidad que debe poseer un

documento o archivo para que éste solo se entienda de manera comprensible o sea leído por la persona o sistema que esté autorizado; Integridad es la cualidad que posee un documento o archivo que no ha sido alterado y que además permite comprobar que no se ha producido manipulación alguna en el documento original; Disponibilidad se trata de la capacidad de un servicio, de unos datos o de un sistema a ser accesible y utilizable por los usuarios o procesos autorizados cuando lo requieran. También se refiere a la capacidad de que la información pueda ser recuperada en el momento que se necesite.

Como resultado la Seguridad Informática se preocupa de que la información operada por un ordenador no sea dañada o alterada, que esté disponible y en condiciones de ser procesada en cualquier momento y se mantenga confidencial.

El riesgo informático.

Los riesgos se pueden especificar como aquellas casualidades que impiden el cumplimiento de un objetivo ya que afectaría al total funcionamiento de la organización considerado un riesgo o amenaza para la entidad.

En lo relacionado con la tecnología, señala que totalmente el riesgo se traza directamente como amenaza, estableciendo el nivel de muestra a la ocurrencia de una pérdida, como, por ejemplo, el riesgo de datos debido a rotura de disco, virus informáticos, etc. (Silva Coelho, 2018).

Resalta que la Organización Internacional por la Normalización (ISO) define el riesgo como “La probabilidad de que una amenaza se materialice, utilizando vulnerabilidades existentes de un activo o un grupo de activos, generándole pérdidas o daños”. Por otra parte, los riesgos informáticos se mencionan a la inseguridad existente por la posible realización de un suceso concerniente con la amenaza de daño relacionado a los bienes o servicios informáticos (Periféricos, instalaciones, proyectos, programas de cómputo, archivos, información, datos confidenciales, entre otros) (Silva Coelho, 2018).

Elemento de riesgo.

Existen algunos tipos de riesgos que corren los sistemas informáticos y para los cuales los responsables del manejo de la información deben tomar acciones para corregirlos, los riesgos se pueden clasificar de la siguiente manera: “Respecto a los equipos, respecto a los programas, respecto a las personas, y respecto a los trabajos”

Activos.

Es cualquier elemento que posee valor para la organización, sus operaciones comerciales o su continuidad, incluidos los recursos de información que apoyan la misión de la organización. Se pueden distinguir dos clases de activos: los activos primarios que incluyen a los procesos del negocio, actividades e información; y los activos de apoyo, que incluyen hardware (equipos de procesamiento de datos, periféricos y medios de comunicación), software (sistema operativo, servicio, software de aplicación), redes, personal (directores, usuarios, personal de operación y desarrolladores), lugar y estructura de la organización (proveedores y fabricantes) (Bermudez Molina, 2015).



Los activos forman uno de los 14 dominios que trata el estándar ISO/IEC 27002, el cual contiene 3 objetivos de control y 10 controles, siendo la finalidad de este dominio que la organización tenga un conocimiento preciso sobre los activos que posee, su responsabilidad y su clasificación como parte importante de la gestión de riesgos. Según el estándar ISO/IEC 27002, los activos de información deben ser clasificados de acuerdo a la sensibilidad y criticidad de la información que contienen de acuerdo a la funcionalidad que cumplen y rotulados en función a ello, con el objetivo de indicar cómo ha de ser tratada y protegida dicha información (ISO/IEC 27002, 2013) (Disterer, 2013).

La información es un activo que, como otros activos comerciales importantes, tiene valor para la organización y en consecuencia necesita ser protegido adecuadamente. La seguridad informática protege la información de un amplio rango de amenazas con la finalidad de asegurar la continuidad de los negocios, minimizar el daño comercial y maximizar el reembolso de las inversiones y oportunidades comerciales. La información puede existir en muchas formas; puede ser de forma escrita, impresa, electrónica, transmitida por correo o usando medios electrónicos o hablado en una conversación (Bermudez Molina, 2015).

Amenazas

Las amenazas siempre estarán presentes en los sistemas informáticos debido a la fragilidad que muchos de éstos presentan y se lo puede definir de la siguiente manera:

En sistemas de información se entiende por amenaza la presencia de uno o más factores de diversa índole (personas, máquinas o sucesos) de tener la oportunidad afectarían al sistema produciéndole daños aprovechándose de su nivel de vulnerabilidad. Hay diferentes tipos de amenazas de las que hay que proteger al sistema, desde las físicas como cortes eléctricos, fallos del hardware o riesgos ambientales hasta los errores intencionados o no de los usuarios, la entrada de software malicioso (virus, troyanos, gusanos) o el robo, destrucción o modificación de la información (Chamorro, 2015).

Vulnerabilidades.

Los activos se ven influidos por una serie de amenazas; la probabilidad de que se materialice una de dichas amenazas y la degradación que le supone a un activo es lo que se conoce como vulnerabilidad según la metodología MAGERIT (Silva Coelho, 2018). “Las vulnerabilidades deben ser clasificadas de acuerdo a la clase de activos, es decir: hardware (susceptibilidad a la humedad, polvo, suciedad, almacenamiento sin protección), software (falta de pruebas del software, falta de seguimiento de auditoría), red (líneas inadecuadas, falta de seguridad), sitio (ubicación en un área susceptible a inundaciones, red de energía inestable), y organización (falta de auditorías periódicas, falta de planes de continuidad del negocio)”. (López Uriarte, 2017)

Impacto.

Al respecto de lo que son los impactos y las consecuencias que puede ocasionar, se puede mencionar que: “Son la consecuencia de la materialización de una o más amenazas sobre uno o varios activos aprovechando la vulnerabilidad del sistema o, dicho de otra manera, el daño causado. Los impactos pueden ser cuantitativos, si los perjuicios pueden cuantificarse económicamente, o cualitativos, si suponen daños no cuantificables, como los causados contra los derechos fundamentales de las personas” (Ladino, 2011). Todos los riesgos mencionados pueden suceder dentro de los sistemas informáticos, por lo que se deben tomar las respectivas precauciones para proteger la información.

Análisis de riesgo.

El modelo PDCA

Algunos autores como (Disterer, 2013) y (Bermudez Molina, 2015) coinciden en que, para ejecutar el análisis y posterior gestión del riesgo, se tiene que continuar un ciclo con cuatro etapas conocido por sus siglas en inglés como PDCA (Plan-Do-Check-Act) o Planificar-Ejecutar-Verificar-Actuar, que se ilustra en la Figura 1. “Al igual que con otros estándares de TI, la familia de estándares ISO 27000 se refiere directamente al ciclo “Plan-Do-Check-Act” (ciclo PDCA), conocido por la gestión clásica de calidad de Deming, que enfatiza la necesidad de la orientación al proceso, así como la integración del planeamiento de las operaciones y la verificación constante de la implementación conforme a la planificación.” (Disterer, 2013). “Los sistemas de gestión de la seguridad de la información formalizan cuatro etapas cíclicas donde el análisis de riesgos es parte de las actividades de planificación, se toman decisiones de tratamiento y estas decisiones se materializan en la etapa de implantación, en la cual se despliegan elementos que acceden la monitorización de las medidas tomadas para poder evaluar la efectividad de las mismas y actuar dependiendo de éstas, dentro de un círculo de excelencia o mejora continua” (Figueroa Moran, 2017).



Figura 1. Etapas del ciclo PDCA según ISO.
Fuente: (Disterer, 2013)

Según (Disterer, 2013), en un Sistema de Gestión de Seguridad de la Información (SGSI), en la etapa de planificación es donde se definirán los requisitos para la protección de la información, se identificarán y evaluarán los riesgos y se desarrollarán los procedimientos y medidas adecuados para reducir los riesgos. Estos procedimientos y medidas se implementan durante la etapa de implementación y operación (o etapa de ejecución). Los informes generados a través del monitoreo continuo de las operaciones (etapa de verificación) se utilizarán en la última etapa (Actuar) para obtener las mejoras y el desarrollo posterior y continuo del SGSI. Estas etapas se resumen en la Figura 1. La mayoría de las metodologías de gestión de riesgos utilizan como base el modelo PDCA (Disterer, 2013), y lo realizan con la propósito de establecer un proceso de gestión que se enfoque en la mejora continua, siguiendo teniendo las siguientes actividades en cada una de sus etapas:

- Planificar: “Se establecen los objetivos, procesos y procedimientos para la gestión de riesgos tecnológicos. El objetivo de esta etapa es la entrega de resultados acordes con las políticas y objetivos globales de la organización. Además, se forma el plan de comunicaciones y el análisis del contexto organizacional actual para limitar el alcance de la gestión de riesgos tecnológicos”. (Disterer, 2013)
- Hacer: “Se realiza la implementación y operación de los controles, procesos y procedimientos e incluye además la operación e implementación de las políticas definidas y la valoración y tratamiento de los riesgos” (Disterer, 2013).
- Verificar: “En esta etapa se evalúa y se mide la ocupación de los procesos contra la política y los objetivos de seguridad. Además, se debe informar los resultados obtenidos” (Silva Coelho, 2018).
- Actuar: “En esta etapa se establece la política para la gestión de riesgos tecnológicos y se realizan los cambios solicitados para la mejora de los procesos. En las etapas verificar y actuar, se incluye el monitoreo y la mejora continua, donde se verifican los cambios y los cumplimientos de indicadores establecidos en la etapa de planificación” (Guamán, 2019).

Gestión de riesgo.

La gestión de riesgos radica en el proceso de analizar, evaluar, tratar, monitorizar y comunicar los riesgos encontrados. “La gestión de los riesgos es un desafío estratégico para las organizaciones, las cuales enfrentan amenazas cada vez más complejas y diversas. Caldera y Watkins indican que todas las organizaciones se enfrentan diariamente a riesgos de un tipo u otro” (Silva Coelho, 2018). Se definen a la Gestión de Riesgos como una disciplina no especulativa, que son aquellos riesgos de los cuales sólo puede ocurrir una pérdida; en cambio, los riesgos especulativos son aquellos a partir de los cuales se puede producir una ganancia o una pérdida, que a menudo son estrategias de negocio de la organización (Enríquez Collaguazo, 2018).

La gestión de riesgos, suelen tener cuatro objetivos vinculados (Silva Coelho, 2018), los cuales son:

- Eliminar los riesgos

- Reducir a niveles “aceptables” aquellos riesgos que no pueden eliminarse; y entonces
- Convivir con ellos, ejerciendo cuidadosamente los controles que los mantienen en niveles “aceptables”; o
- Transferirlos, por medio de aseguradoras, por ejemplo, a otra instancia u organización.

Según (Bermudez Molina, 2015), la gestión del riesgo en general consiste en seis procesos: establecimiento del contexto, evaluación del riesgo, tratamiento del riesgo, aceptación de riesgos, comunicación y consulta de riesgos, revisión y seguimiento del riesgo (Bermudez Molina, 2015). Estos procesos y su interrelación se los puede apreciar en la Figura 2.



Figura 2. Proceso de la gestión de riesgo.

Fuente: (Bermudez Molina, 2015).

2.2. Metodología de gestión de riesgo.

Magerit.

Es una de las metodologías más utilizadas en la gestión de riesgos de los Sistemas de Información; fue creada por el Consejo Superior de Administración Electrónica del Ministerio de Hacienda y Administraciones Públicas de España (Ministerio de Hacienda y Administraciones Públicas de España, 2012) para minimizar los riesgos de la implantación y uso de las Tecnologías de la Información siguiendo la terminología de la norma ISO 31000 y, en el año 2012 se actualizó a la versión 3 (Genova Garcia, 2017). La entidad que creo esta metodología, la define como: “Una metodología que ha sido elaborada como respuesta a la percepción de la administración pública (y en general toda la sociedad), depende de forma creciente de los sistemas de información para alcanzar sus objetivos (De la Torre, 2018). Así, menciona que el uso de tecnologías de la información y comunicaciones (TIC) supone unos beneficios evidentes para los ciudadanos; pero también da lugar a ciertos riesgos que deben gestionarse prudentemente con medidas de seguridad que sustenten la confianza de los usuarios de los servicios” (Velasco, 2008). Según (Bermudez Molina, 2015), “los objetivos que busca alcanzar esta metodología son los siguientes:

- Innovar que los responsables de los sistemas de información sean conscientes de la existencia de riesgos y de la

necesidad de conocer a tiempo.

- Dar un método sistemático para el análisis de riesgos.
- Ayudar en la diseño y planificación de las medidas adecuadas para mantener los riesgos bajo control.
- De forma indirecta, preparar la organización de los procesos de evaluación, auditoría, certificación o acreditación, según corresponda en cada caso”.

La metodología se resume en el modelo de la Figura 3.



Figura 3. Modelo Magerit.

El desarrollo de esta metodología contempla las siguientes fases:

- Fase 1: definir el alcance. “El primer paso a la hora de llevar a cabo el análisis de riesgos es establecer el alcance del estudio. Vamos a considerar que este análisis de riesgos forma parte del Plan Director de Seguridad. Por lo tanto, recomendamos que el análisis de riesgos cubra la totalidad del alcance del PDS, dónde se han seleccionado las áreas estratégicas sobre las que mejorar la seguridad.
- Fase 2: Identificar los activos. Una vez definido el alcance, debemos identificar los activos más importantes que guardan relación con el departamento, proceso, o sistema objeto del estudio.
- Fase 3: Identificar amenazas. Habiendo identificado los principales activos, el siguiente paso consiste en identificar las amenazas a las que estos están expuestos. Tal y como imaginamos, el conjunto de amenazas es amplio y diverso por lo que debemos hacer un esfuerzo en mantener un enfoque práctico y aplicado.
- Fase 4: Identificar vulnerabilidades y salvaguardas. La siguiente fase consiste en estudiar las características de nuestros activos para identificar puntos débiles o vulnerabilidades.

- Fase 5: Evaluar el riesgo. Disponemos de los siguientes elementos (inventarios de activos, conjunto de amenazas a las que están expuestas los activos, conjunto de vulnerabilidades asociadas a cada activo, conjunto de medidas de seguridad implantadas)” (Chamorro, 2015).

Elementos de Magerit

Los elementos considerados significativos por Magerit para el estudio de los sistemas de información:

- **ACTIVOS:** “Recursos del sistema de información o relacionados con este, necesarios para que la organización funcione correctamente y alcance los objetivos propuestos por la dirección”. (López Uriarte, 2017)
- **AMENAZAS:** “Eventos que pueden desencadenar un incidente en la organización, produciendo daños materiales o pérdidas inmateriales en sus activos”. (Bermudez Molina, 2015)
- **VULNERABILIDAD DE UN ACTIVO:** “Potencialidad o posibilidad de ocurrencia de la materialización de una amenaza sobre dicho activo”. (Bermudez Molina, 2015)
- **IMPACTO DE UN ACTIVO:** “Consecuencia sobre este de la materialización de un activo.
- **RIESGO:** Posibilidad de que se produzca un impacto determinado en un activo, en un dominio o en toda la organización.
- **SERVICIO DE SALVAGUARDIA:** Acción que reduce el riesgo.
- **MECANISMO DE SALVAGUARDA:** Procedimiento, dispositivo, físico o lógico, que reduce el riesgo”. (Silva Coelho, 2018)

La Figura 4 muestra los elementos y sus interrelaciones:

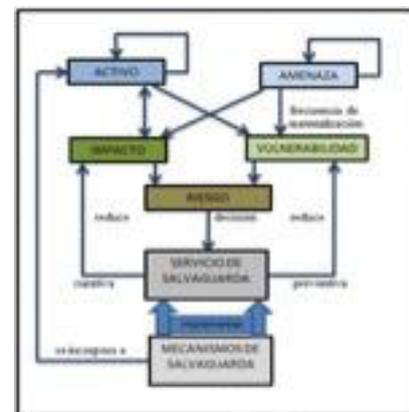


Figura 4. Elementos del Magerit.

Norma ISO/IEC 27002

“La norma internacional ISO/IEC 27002, que se centra en las buenas prácticas para gestión de la seguridad de la información (Armendáriz, 2017). Esta norma se basa en el código de las buenas prácticas para la gestión de la seguridad. Se puede dar recomendaciones sobre qué medidas tomar para asegurar los sistemas de información de una organización. Esta norma también describe los objetivos de control (aspectos para garantizar la seguridad de la información) y especifica los controles recomendables a implantar (medidas tomar)” (Sanchez, 2017).

Principales ítems que componen la ISO 27002:

“La parte principal de la norma se encuentra distribuida en las siguientes secciones, que corresponden a controles de seguridad de la información. Es importante recordar que la organización puede utilizar esas directrices como base para el desarrollo del SGSI” (Silva Coelho, 2018).

Sección 5 – Política de Seguridad de la Información

“Se debe crear un documento sobre la política de seguridad de la información de la empresa, que debe contener los conceptos de seguridad de la información, una estructura para establecer los objetivos y las formas de control, el compromiso de la dirección con la política, entre tantos otros factores” (Genova Garcia, 2017).

Sección 6 – Organización de la Seguridad de la Información

“Para implementar la Seguridad de la Información en una empresa, es necesario establecer una estructura para gestionarla de una manera adecuada. Para ello, las actividades de seguridad de la información deben ser coordinadas por representantes de la organización, que deben tener responsabilidades bien definidas y proteger las informaciones de carácter confidencial” (De la Torre, 2018).

Sección 7 – Gestión de activos

“Activo, según la norma, es cualquier cosa que tenga valor para la organización y que necesita ser protegido. Pero para ello los activos deben ser identificados y clasificados, de modo que un inventario pueda ser estructurado y posteriormente mantenido. Además, deben seguir reglas documentadas, que definen qué tipo de uso se permite hacer con dichos activos” (López Uriarte, 2017).

Sección 8 – Seguridad en recursos humanos

“Antes de la contratación de un empleado – o incluso de proveedores – es importante que sea debidamente analizado, principalmente si se trata de información de carácter confidencial. La intención de esta sección es mitigar el riesgo de robo, fraude o mal uso de los recursos. Y cuando el empleado esté trabajando en la empresa, debe ser consciente de las amenazas relativas a la seguridad de la información, así como de sus responsabilidades y obligaciones” (Romo Villafuerte, 2012).

Sección 9 – Seguridad física y del medio ambiente

“Los equipos e instalaciones de procesamiento de información crítica o sensible deben mantenerse en áreas seguras, con niveles

y controles de acceso apropiados, incluyendo protección contra amenazas físicas y ambientales” (Chamorro, 2015).

Sección 10 – Seguridad de las operaciones y comunicaciones

“Es importante que estén definidos los procedimientos y responsabilidades por la gestión y operación de todos los recursos de procesamiento de la información. Esto incluye la gestión de servicios tercerizados, la planificación de recursos de los sistemas para minimizar el riesgo de fallas, la creación de procedimientos para la generación de copias de seguridad y su recuperación, así como la administración segura de las redes de comunicaciones” (Montoya, 2017).

Sección 11 – Control de acceso

“El acceso a la información, así como a los recursos de procesamiento de la información y los procesos de negocios, debe ser controlado con base en los requisitos de negocio y en la seguridad de la información. Debe garantizarse el acceso de usuario autorizado y prevenido el acceso no autorizado a los sistemas de información, a fin de evitar daños a documentos y recursos de procesamiento de la información que estén al alcance de cualquiera” (López Uriarte, 2017).

Sección 12 – Adquisición, desarrollo y mantenimiento de sistemas

“Los requisitos de seguridad de los sistemas de información deben ser identificados y acordados antes de su desarrollo y/o de su implementación, para que así puedan ser protegidos para el mantenimiento de su confidencialidad, autenticidad o integridad por medios criptográficos” (Pacheco Villamar, 2018).

Sección 13 – Gestión de incidentes de seguridad de la información

“Los procedimientos formales de registro y escalonamiento deben ser establecidos y los empleados, proveedores y terceros deben ser conscientes de los procedimientos para notificar los eventos de seguridad de la información para asegurar que se comuniquen lo más rápido posible y corregidos en tiempo hábil” (Disterer, 2013).

Sección 14 – Gestión de continuidad del negocio

“Los planes de continuidad del negocio deben ser desarrollados e implementados, con el fin de impedir la interrupción de las actividades del negocio y asegurar que las operaciones esenciales sean rápidamente recuperadas” (Genova Garcia, 2017).

Sección 15 – Conformidad

“Es importante evitar la violación de cualquier ley criminal o civil, garantizando estatutos, regulaciones u obligaciones contractuales y de cualesquiera requisitos de seguridad de la información. En caso necesario, la empresa puede contratar una consultoría especializada, para que se verifique su conformidad y adherencia a los requisitos legales y reglamentarios” (Guamán, 2019).

2. Materiales y Métodos

La metodología propuesta consistió en la detección, análisis a las



vulnerabilidades a las que se ven expuesto la distribuidora, ya que nos permite conocer cada factor que pueda estar poniendo en riesgo a la organización que se está elaborando el estudio.

Una de la parte importante es la información de la empresa, donde se evaluará los puntos críticos en la hora de manejar el riesgo que tiene cada activo dentro del departamento de TI, de esta manera nos permitirá conocer los niveles de riesgo de cada uno de los activos.

Las fases que se van a realizar son las siguientes:

- **Definir el alcance:** Se implementa el ciclo PDCA (Planifica, Hacer, Verificar, Actuar), dentro de este ciclo se logra obtener la información que se necesita para la evaluación del riesgo de los activos, Y poder llevar acabo las acciones correctivas y verificar las planificaciones que se tiene la empresa con el respecto del cuidado de la información.
- **Identificar los activos:** Una vez ya teniendo hasta donde se puede llegar, con la evaluación dentro de la empresa, podemos identificar los activos que están en riesgo como se evidencia en la Tabla 1.

Tabla 1. Identificación de activo de la distribuidora Bravel.
 Fuente: Los autores.

ITEM	TIPO DE ACTIVO	ACTIVO
1	Información	Datos de los clientes
2	Equipo informático (hardware)	Servidor de base de dato, Pc's, impresora, Servidor de Intranet, Firewall, Router.
3	Red de comunicación	Red inalámbrica
4	software	Herramienta a utilizar para el desarrollo de las actividades (sistema).
5	Equipamiento Auxiliar	Fibra Óptica, Aire Acondicionado
6	Personal	Responsables del área de ti

Amenazas: Acorde a la metodología Magerit catálogos de elementos, hace referencia a:

D=disponibilidad, C=confidencialidad, A=autenticidad y

T=trazabilidad para dar significado al activo de la empresa. En donde cada nivel de riesgo toma un valor de una escala de 1-5. Las amenazas se presentan en la Tabla 2.

Tabla 2. Identificación de valoración de las amenazas de la distribuidora Bravel.

Fuente: (Armendáriz, 2017).

Ítem	Activos críticos	Dimensiones de valoración				
		C	I	D	A	T
1	Servidor de base de datos	5	5	5	5	5
2	Redes de datos	5	5	5	5	5
3	Servidor proxy	5	5	5	5	5

Identificar vulnerabilidad y salvaguardas: La información suministrada por el personal encargado del área de TI, se pudieron identificar las siguientes salvaguardas: Backup Se tiene como procedimiento establecido las copias de seguridad como factor fundamental para salvaguardar la información como se presenta en la Tabla 3.

Mantenimiento de hardware: se tiene como prioridad los días o fechas que se tiene para el óptimo Funcionamiento.

En el estudio que se realizó en la empresa, se pudieron identificar las amenazas y vulnerabilidades con respecto, a los activos críticos identificados.

Tabla 3. Identificación de Vulnerabilidades de la distribuidora Bravel.

Fuente: Los autores.

Item	Activos	Vulnerabilidades	Amenazas
1	Servidor de base de datos	Fallos del sistema eléctrico	Perdida de la información Incendio
2	Red de datos	Manipulación de la red	Interrupción en las actividades Perdida de la información
4	Servidor proxy	Ataques internos	Manipulación de la información

Evaluar el riesgo: Dentro de la evaluación que se realizó dentro de la empresa, se estable los valores del riesgo, que se presenta en los activos de la organización que se realizó el estudio, como se observa en la Figura 5 y Tabla 4 a continuación.

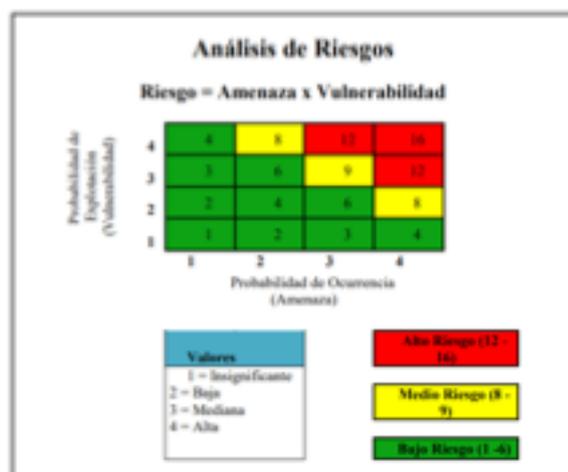


Figura 5. Análisis de riesgo.
Fuente: (Genova Garcia, 2017)

Tabla 4. Identificación de Riesgo de la distribuidora Bravel.
Fuente: Los autores.

Item	Activos	Vulnerabilidades	Amenazas	Probabilidad	Item
1	Servidor de base de datos	Fallos del sistema eléctrico	Deterioro del gabinete cortocircuito	3	3
2	Red de datos	Manipulación de la red	Interrupción en las actividades Perdida de la información	3	4
3	Servidor proxy	Ataques internos	Manipulación de la información	3	4

3. Resultados y Discusión

Después de la evaluación de las 3 formas para poder evaluar la distribuidora, nos permite llegar a una nueva metodología. Que consta en los siguientes:

Fase 1: Revisión manual: En esta fase se realizará el proceso de análisis y evaluación de riesgos de acuerdo al estándar MAGERIT que permite valorar los riesgos en cada uno de los criterios de información evaluados, identificando las posibles causas que los originan y que posteriormente permitan definir un sistema de control de seguridad de acuerdo a los hallazgos confirmados, lo que permitirá disminuir el impacto en la organización y probabilidad de ocurrencia de los mismos. El proceso de análisis y evaluación de los riesgos se lleva a cabo teniendo en cuenta el estándar MAGERIT versión 3.0, que

permite hacer la clasificación de amenazas y riesgos, los activos informáticos, muestra las escalas de valoración y los criterios de información que será evaluados.

Posteriormente se aplican las listas de chequeo que son utilizadas para verificar y determinar la existencia de controles de seguridad informática y de la información, diseñadas de acuerdo a la norma ISO/IEC 27002, la cual observaremos los riesgos que está presentando a organización y también si cumple con cada uno de los objetivos de control que tiene la norma como se observa en la Figura 6.



Figura 6. ISO/IEC 27002, dominio, objetivo de control y controles.

Fuente: www.iso27000.es/assets/files/ControlesISO27002-2013.pdf.

Fase 2: Elección del software a aplicar: En esta fase se va tener en cuenta 2 software para la respectiva comparación, que son eMarisma e ISOTolls, la cual nos permite conocer el riesgo de la organización de cada uno de los activos y el manejo de la información que se encuentra en el departamento de TI, y conocer el riesgo que tiene la distribuidora.

Fase 3: implementación del software eMarisma: en esta fase se implementa el software la cual se debe realizar unos pasos antes de la realización de la evaluación, debemos realizar primeramente el proyecto de estudio que se va a realizar, una vez ya realizado se debe realizar una nueva auditoría para proceder a crear cada uno de los activos que tiene la organización para después realizar un agrupamiento para poder tenerlo relacionado con cada uno del proceso a examinar.

Por último, se debe realizar el SOA (arquitectura orientada a servicios), es el nexo que une las metas de negocio con el sistema de software. Su papel es el de aportar flexibilidad, desde la automatización de las infraestructura y herramientas necesarias consiguiendo, al mismo tiempo, reducir los costes de integración. SOA se ocupa del diseño y desarrollo de sistemas distribuidos y es un potente aliado a la hora de llevar a cabo la gestión de grandes volúmenes de datos, datos en la nube y jerarquías de datos. Ver Figuras 7, 8 y 9 a continuación.



Figura 7. Ingreso de los activos que tiene la distribuidora.



Figura 7. Activos para el uso del caso agregados en la herramienta eMARISMA.

4. Conclusiones

La investigación realizada demuestra que no existe una cultura de seguridad dentro de la distribuidora; por esta razón tampoco existe sistemas de control de seguridad informática y de información, y mucho menos, procesos y procedimientos documentados para protección de la información.

Por tanto, es fundamental que las organizaciones cuenten con un marco normativo de seguridad, que permita aplicar la auditoría basada en la norma ISO/IEC 27002.

La seguridad de la información se concluye que este proceso



Figura 9. Cuadro de mando generado por la herramienta eMARISMA para nuestro caso de estudio.

debe ser continuo y que debe ser realizado por los entes de control interno de la organización, y periódico, implementación de un SGSI adecuado a sus necesidades.

Es necesario realizar una constante actualización y búsqueda de las mejores herramientas disponibles para la detección de amenazas a las que se exponen los activos.

Contribución de los autores

María Angélica Velepucha Sánchez: Conceptualización, Metodología, Análisis formal. **Jéssica Morales Carrillo:** Metodología, Análisis formal, Revisión y edición del artículo. **Marco Fernando Pazmiño Campuzano:** Metodología, Revisión y edición del artículo.

Conflictos de interés

Los autores declaran no tener ningún conflicto de interés.

Referencias bibliográficas

- Armendáriz, D. N. L. (2017). Modelo de gestión de los servicios de tecnología de información basado en COBIT, ITIL e ISO/IEC 27000. *Revista Tecnológica-ESPOL*, 30(1).
- Bermúdez Molina, K. G. (2015). Análisis en seguridad informática y seguridad de la información basado en la norma ISO/IEC 27001-sistemas de gestión de seguridad de la información dirigido a una empresa de servicios financieros. <https://dspace.ups.edu.ec/handle/123456789/10372>
- Aguayo Chamorro, C. R. (2015). Propuesta para un adecuado

- manejo de la seguridad de la información en base a la norma ISO 27002 para la Dirección de Gestión Tecnológica del Ministerio del Deporte, Quito. <http://bibdigital.epn.edu.ec/handle/15000/12650>
- De la Torre, C. (2018). *SCP progreso*. Obtenido de <https://www.scprogress.com/NOTICIAS/CyberNoticia47-20170824.pdf>
- Disterer, G. (2013). ISO/IEC 27000, 27001 and 27002 for Information. *Scientific Research An Academic publisher*; 92-100. <http://dx.doi.org/10.4236/jis.2013.42011>
- Enríquez Collaguazo, A. A. (2018). *Repositorio Digital Universidad Técnica del Norte*. Obtenido de <http://repositorio.utn.edu.ec/handle/123456789/8572>
- Figueroa Moran, G. L. (2017). *Repositorio digital UNESUM*. Obtenido de <http://repositorio.unesum.edu.ec/handle/53000/980>
- Silva Coelho, F. E. (2018). *Gestión de la Seguridad de la información*. Colombia: ebook. Obtenido de <https://www.cedia.edu.ec/assets/docs/publicaciones/libros/GTI8.pdf>
- Genova Garcia, L. Z. (2017). Tecnología de la Información. Técnicas de seguridad. Código de prácticas para los controles de seguridad de la información. *UNE-EN normalizacion español*, 13-22. <https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma?c=N0058429>
- Guamán, V. L. (2019). *Evaluación de seguridad de la información aplicado al sistema de evaluación de docentes de la Universidad Técnica del Norte basado en la ISO 27002:2017 con la metodología Magerit V3*. Ibarra: Universidad Técnica del Norte. <http://repositorio.utn.edu.ec/handle/123456789/9535>
- Mora Palacios, J. P. (2017). Desarrollo y Adaptación de COBIT 5 como metodología de gestión de riesgos a la norma ISO/IEC 27001, utilizando el proceso APO12. *GIS Gestión, Ingenio y Sociedad*, 18-37. <http://gis.unicafam.edu.co/index.php/gis/article/view/22/70>
- López Uriarte, E. D. (2017). *Evaluación de la Red Inalámbrica en el Hospital Escuela Cesar Amador Molina, basado en la norma IEEE 802.11 y controles de seguridad del estándar ISO 27002-2013 Matagalpa, I semestre 2015*. Managua: Universidad Nacional Autónoma de Nicaragua. <http://repositorio.unan.edu.ni/id/eprint/3198>
- Ladino, M. I., Villa, P. A., & López, A. M. (2011). Fundamentos de ISO 27001 y su aplicación en las empresas. *Scientia et Technica*, 17(47), 334-339. <https://www.redalyc.org/pdf/849/84921327061.pdf>
- Montoya, Y. A. (2017). *Repositorio Escuela Superior Politécnica del Litoral*. Obtenido de <http://www.dspace.espol.edu.ec/xmlui/handle/123456789/38692>
- Pacheco Villamar, R. A. (2018). *Repositorio Digital de la Universidad Espiritu Santo*. Obtenido de <http://repositorio.uees.edu.ec/handle/123456789/3059>
- Pazmiño Flores, C. D. (2019). *Repositorio de la Universidad Internacional SEK Ecuador*. Obtenido de <https://repositorio.uisek.edu.ec/handle/123456789/3345>
- Romo Villafuerte, D. &. (2012). *Repositorio Institucional de la Universidad Politécnica Salesiana*. Obtenido de <https://dspace.ups.edu.ec/handle/123456789/3163>
- Sanchez, J. D. (2017). *Secretaria general de industria y de la pequeña y mediana empresa*. Obtenido de https://static.eoi.es/inline/une-en_iso-iec_27002_norma_mincotur.pdf
- Solarte, F. N. S., Rosero, E. R. E., & del Carmen Benavides, M. (2015). Metodología de análisis y evaluación de riesgos aplicados a la seguridad informática y de información bajo la norma ISO/IEC 27001. *Revista Tecnológica-ESPOL*, 28(5).
- Salazar Choez, T. K. (2018). *REPOSITORIO UNESUM*. Obtenido de <http://repositorio.unesum.edu.ec/bitstream/53000/1469/1/UNESUM-ECU-REDES-2017-01.pdf>
- Velasco, M. A. (2008). El derecho informático y la gestión de la seguridad de la información una perspectiva con base en la Norma ISO 27 001. *Revista de derecho*, 320. <https://www.redalyc.org/articulo.oa?id=85102913>





Desarrollo de una aplicación móvil para el monitoreo de la fenometría vegetativa del maíz amarillo, en la sierra central ecuatoriana

Development of a mobile application for the monitoring of the vegetative phenomenon of yellow corn, in the central Ecuadorian highlands

Autores

✉ * *Vladimir Javier Rojano Guamaní*



✉ *Ginger Lissbeth Jaramillo Tenezaca*



✉ *Karla Susana Cantuña Flores*



✉ *Gustavo Adolfo Sandoval Ruilova*



✉ *José María Bengochea Guevara*



Universidad Técnica de Cotopaxi (UTC),
Latacunga, Ecuador.

* Autor para correspondencia

Comó citar el artículo: Rojano Guamaní, V.J., Jaramillo Tenezaca, G.L., Cantuña Flores, K.S., Sandoval Ruilova, G.A. & Bengochea Guevara, J.M. 2022. Desarrollo de una aplicación móvil para el monitoreo de la fenometría vegetativa del maíz amarillo, en la sierra central ecuatoriana. *Informática y Sistemas: Revista de Tecnologías de la Informática y las Comunicaciones*. 6(1) 71-77. DOI: <https://doi.org/10.33936/isrtic.v6i1.4447>

Enviado: 21/02/2022

Aceptado: 09/03/2022

Publicado: 17/05/2022

Resumen

En este trabajo se presenta una aplicación móvil en la que se realizan observaciones de los parámetros de la fenometría vegetativa del maíz, es decir, valorar cuantitativamente el crecimiento de la planta teniendo presente las medidas de longitud y diámetro del tallo. Para el desarrollo de la aplicación móvil se emplea la metodología Mobile-D, la medición de la longitud y el diámetro del tallo en imágenes se basa en la colocación de un objeto de referencia junto a la planta y en el uso de técnicas del área del procesamiento digital de imágenes a través de la librería OpenCV. La precisión de la medición de la longitud y el diámetro se evalúa utilizando el estimador RMSE (error cuadrático medio) en un banco de cincuenta imágenes obteniendo una desviación estándar de la varianza de 5,41 milímetros para la longitud y de 5,27 milímetros para el diámetro.

Palabras claves: Fenometría; maíz; procesamiento digital de imágenes; OpenCV; Mobile-D.

Abstract

In this work, a mobile application is presented in which observations of the parameters of the vegetative phenometry of corn are made, that is, quantitatively assessing the growth of the plant taking into account the measurements of length and diameter of the stem. For the development of the mobile application, the Mobile-D methodology is used, the measurement of the length and diameter of the stem in images is based on the placement of a reference object next to the plant and on the use of techniques from the area of the digital image processing through the OpenCV library. The precision of the length and diameter measurement is evaluated using the RMSE (mean square error) estimator in a bank of fifty images, obtaining a standard deviation of the variance of 5,41 millimeters for the length and 5,27 millimeters for the diameter.

Keywords: Phenometry; corn; digital image recognition; OpenCV; Mobile-D.



1. Introducción

En la actualidad se resalta la importancia que tiene la tecnología en la generación de competitividad en la producción y comercialización agrícola, tanto en el mercado local como en el mercado nacional. Así también, la tecnología puede incrementar la sostenibilidad y la mejora de la seguridad alimentaria de la población (Carpio Santos, 2018). Por otra parte, el uso de la tecnología en el sector agrícola ha servido históricamente como herramienta mediadora entre el hombre y la naturaleza, su función básica es contribuir sustancialmente a transformar la naturaleza para beneficio de la gente que vive en el campo (Tapia, 2006).

El cultivo de maíz es de suma importancia en el Ecuador debido al rol significativo que cumple en seguridad alimentaria de la población (Zambrano Mendoza et al., 2019). La producción de maíz en Ecuador es de 1.2 millones de toneladas, 5.6 toneladas por hectárea de las cuales 900 000 toneladas son adquiridas por la industria nacional para la elaboración de elementos proteicos. El Ministerio de Agricultura y Ganadería (MAG) establece que la provincia de Los Ríos registra la mayor producción de maíz con un 85% del rendimiento a nivel nacional. Además, afirma que la producción del cereal se ha ido incrementando, pero sigue siendo deficitario, por lo que se deben buscar estrategias para prevenir daños en el cultivo y mejorar su producción.

El conocimiento de la fenometría o fenología cuantitativa del maíz es una valiosa herramienta de manejo agronómico, pues permite, predecir la calidad del cultivo a lo largo de sus fases vegetativas y con ello establecer un método de control efectivo que garantice su ciclo vegetativo.

Existen pocos estudios referentes a la fenometría empleando herramientas informáticas, la mayoría de estos se centran en el área de la agronomía. Zari Arévalo (2014) en su estudio registra los datos fenométricos cada siete días en plantas seleccionadas al azar, las estructuras de la planta se midieron mediante un calibrador obteniendo así, la información fenométrica de la altura, número de hojas verdes, ancho del tallo, diámetro y longitud de la mazorca. En el trabajo de López Castañeda (2015) se evaluó la altura de la planta con un flexómetro tomando el dato desde la base del suelo hasta la hoja más joven ubicada en el tercer tercio de la planta, el diámetro del tallo con un vernier tomando como referencia el primer nudo de la planta y por conteo el número de tallos secundarios. Ventura Elías (2016), en su estudio, desarrollo un muestreo al azar con frecuencia de 3 por semana en cultivos de frijol durante las cuales se midió y cuantificó la altura a partir del nudo cotiledonal, hojas verdes existentes, guías secundarias, flores, vainas llenas y maduras.

La revisión anterior sugiere la necesidad de desarrollar una aplicación informática que permita determinar de manera automática la información fenométrica de la planta de maíz con

el objeto de tomar decisiones oportunas sobre el crecimiento y rendimiento del cultivo. Por tal motivo, el objetivo del presente estudio es desarrollar una aplicación móvil que permita observar los aspectos fenométricos a partir de la longitud y el diámetro del tallo empleando técnicas de procesamiento digital de imágenes y el uso de la ingeniería de software para el desarrollo de la aplicación.

El estudio se encuentra organizado de la siguiente forma, en la sesión 2 se presenta el método de investigación, en la sección 3 se discuten los resultados y en la sección 4 se abordan las conclusiones del trabajo de investigación.

2. Materiales y Métodos

2.1. Aplicación de la metodología mobile D

Mobile-D es una metodología ágil orientada al desarrollo de aplicaciones móviles desarrollado por Pekka Abrahamsson y su equipo de trabajo del VTT (Valtion Teknillinen Tutkimuskeskus, en inglés Technical Research Centre of Finland) en Finlandia, orientados a ciclos de desarrollos muy rápidos y grupos pequeños de desarrolladores (Zambrano, 2017). La metodología Mobile-D se basa en las prácticas de eXtreme Programming (XP), metodología Crystal para la escalabilidad de los métodos y Rational Unified Process para el diseño completo de ciclo de vida.

Esta metodología está compuesta por 5 fases:

Fase de Exploración. En esta fase se define el alcance del aplicativo, el cual aborda el desarrollo de una aplicación móvil que mide la longitud y el diámetro del tallo de la planta de maíz en imágenes digitales. También aquí se especifican los requerimientos de usuario los mismos que fueron descritos de manera detallada para evitar ambigüedades Tabla 1.

Tabla 1. Requerimientos de usuario de la aplicación móvil.

Fuente: Los autores.

Descripción de requerimientos de usuario
El sistema deberá permitir la segmentación de la planta de maíz.
El sistema deberá permitir medir la longitud de la planta de maíz
El sistema deberá permitir medir el diámetro del tallo de la planta de maíz

Fase de Inicialización. En esta fase se redactan las historias de usuarios con el fin de determinar la funcionalidad de la aplicación móvil entre ellas: segmentación de la planta de maíz, medición de la longitud del tallo y medición del diámetro del tallo del maíz, en la Tablas 2 se detalla una de las historias de usuario.

Tabla 2. Historia de usuario medición de la longitud del tallo.

Fuente: Los autores.

Historia de usuario			
Medición de la longitud del tallo			
Código HU:	HU0002	Fecha:	12/04/2021
Sprint:	1	Prioridad:	Alta
Actores:	Usuario	Puntos:	3
Descripción: Como usuario quiero que la aplicación móvil mida la longitud del maíz para observar los parámetros de la fenometría vegetal de la planta.			
Detalles de la HU:			
<ul style="list-style-type: none"> El usuario ingresa a la aplicación El usuario toma la fotografía de la planta El usuario ingresa los datos del objeto referencial La aplicación calcula la longitud del tallo de la planta de maíz 			
Restricciones: las medidas del alto y ancho del objeto referencial deben ser ingresadas en milímetros.			
Criterios de aceptación: los campos alto y ancho del objeto referencial deben ingresarse obligatoriamente			
DoD: la prueba funcional aplicada a esta historia de usuario generó un resultado exitoso			

Fase de Producción. En esta fase se presenta el diseño de las interfaces y la codificación de las historias de usuario. La codificación de la aplicación móvil conllevó el uso del entorno de desarrollo integrado oficial para la plataforma Android (Android Studio), el lenguaje de programación Java y la librería OpenCV para la segmentación de imágenes digitales, en la Figura 1 se muestra parte de la codificación y en la Figura 2 el diseño de la interfaz inicial.

```

public void calcular_dimensiones(Integer alto_ref, Integer ancho_ref, Integer alto_tallo, Integer ancho_tallo,
int alto_print=0; int ancho_print=0; int ancho_print_2=0; int ancho_print_3=0;
String ancho_p=" ";
if((alto_tallo>0)&&(alto_ref>0)) {
    alto_print = ((objeto.alto + alto_tallo) / alto_ref); //alto del tallo
}
if((ancho_tallo>0)&&(ancho_ref>0)) {
    ancho_print=(objeto.ancho*ancho_tallo)/ancho_ref; //ancho del tallo
    ancho_p=ancho_print+"mm - ";
}
if(ancho_ref>0) {
    if (ancho_tallo_2 > 0) {
        ancho_print_2 = (objeto.ancho + ancho_tallo_2) / ancho_ref; //ancho del tallo
        ancho_p += ancho_print_2 + "mm - ";
    }
    if (ancho_tallo_3 > 0) {
        ancho_print_3 = (objeto.ancho + ancho_tallo_3) / ancho_ref; //ancho del tallo
        ancho_p += ancho_print_3 + "mm - ";
    }
}
    
```

Figura 1. Codificación en Android Studio de la historia de usuario.

Fuente: Los autores.



Figura 2. Interfaz inicial del aplicativo móvil.

Fuente: Los autores.

Fase de Estabilización. Aquí se llevó a cabo la integración de las funcionalidades establecidas en la fase de inicialización conjuntamente con el desarrollo del código fuente con el fin de obtener una aplicación amigable para el usuario. Así como también, se procede con la documentación del aplicativo, en la Figura 3 se muestra la interfaz de ayuda para el usuario.



Figura 3. Interfaz de ayuda para el usuario

Fuente: Los autores.

Fase de Pruebas. Empleando los casos de prueba se determina si las historias de usuario son completamente satisfactorias, en la Tabla 3 se muestra el caso de pruebas correspondiente a la historia de usuario medición de la longitud del tallo.

2.2. Implementación de procesamiento digital de imágenes

El procesamiento o tratamiento digital de imágenes consiste en procesos algorítmicos que transforman una imagen en otra en donde se resalta cierta información de interés, y/o se atenúa o elimina información irrelevante para la aplicación. Así, las tareas del procesamiento de imágenes comprenden la supresión de ruido, mejoramientos de contraste, eliminación de efectos no deseados en la captura como difuminaciones o distorsiones por efectos ópticos o de movimiento, mapeos geométricos,



Tabla 3. Caso de prueba historia de usuario medición de la longitud del tallo
Fuente: Los autores.

Caso de prueba			
Medición de la longitud del tallo			
Código HU:	HU0002	Autor del Caso de Prueba:	D i a n a Cando
ID/Nombre Caso de Prueba:	CP002	Fecha de Creación:	1 de Junio del 2021
Versión 1.0	Usuario	Fecha de Ejecución:	08 de Julio del 2021
Flujo de casos de prueba:			
Nro.	Descripción del paso	Resultado Esperado	Resultado Obtenido
1	La aplicación realiza el cálculo de la longitud de maíz.	La aplicación realiza el cálculo de la longitud de maíz.	La aplicación si realiza satisfactoriamente el cálculo de la longitud del maíz.
2	No se captura la planta.	La aplicación no podrá medir la longitud.	La aplicación si permite medir la longitud.
Decisión de Aprobación del Caso de Prueba: Aprobó: <u>X</u> Fallo: <u> </u>			
Nombre y firma del Probador		Diana cando	

transformaciones de color, etc., (Alvarado, 2012).

La articulación del procesamiento digital de imágenes en la aplicación móvil implicó seguir el siguiente protocolo:

Adquisición de las imágenes:

En la fase de adquisición de datos se estimó un banco de 60 imágenes de plantas de maíz en formato jpg. Las imágenes se tomaron en condiciones naturales de iluminación (iluminación no controlada), en diversos estados de crecimiento del cultivo, empleando la cámara posterior del teléfono móvil (Figura 4).



Figura 4. Interfaz inicial del aplicativo móvil
Fuente: Los autores.

Pre procesamiento

Una vez realizada la captura de las imágenes se seleccionan las fotografías de la planta de maíz con el objeto referencial, obteniendo así, un banco de aproximadamente 50 imágenes. Durante la limpieza de los datos no se consideraron las imágenes borrosas, imágenes en las que el tallo de la planta de maíz presenta malformaciones, o aquellas en las que la cámara haya sufrido vibraciones o movimientos bruscos al capturar las fotografías, en la Figura 5 se observa una imagen seleccionada.



Figura 5. Imagen de la planta de maíz seleccionada
Fuente: Los autores.

Procesamiento

Utilizando la librería OPENCV se segmentó la imagen digital para extraer la planta de maíz. Además, se empleó el detector de bordes Canny creando una traza de rectángulos en color rojo sobre la imagen con el objeto de establecer el cálculo de la

longitud y el diámetro del tallo de la planta de maíz, en la Figura 6 se aprecia la imagen segmentada de la planta.



Figura 6. Imagen de la planta de maíz segmentada
Fuente: Los autores.

Extracción de características:

Durante esta fase se comparan las medidas de longitud y ancho del objeto de referencia con las de la planta de maíz segmentada. A través de la librería OPENCV se dibuja un rectángulo de color verde sobre la imagen el cual identifica la longitud del tallo y uno de color naranja el cual determina el diámetro del tallo, a continuación, se calculan los valores correspondientes Figura 7.



Figura 7. Cálculo de la longitud y ancho de la planta de maíz
Fuente: Los autores.

2.3. Evaluación de la longitud y el diámetro de la planta de maíz

Para evaluar la validez de las medidas de longitud y diámetro del maíz obtenidas mediante la aplicación móvil se utiliza el error cuadrático medio (RMSE), el cual mide la cantidad de error entre dos conjuntos de datos, en este caso los valores predichos (longitud y tallo obtenidos por medio de la aplicación móvil) y los valores observados (longitud y tallo reales de las plantas de maíz). Para la construcción del conjunto de datos observados se midieron en milímetros la longitud y el diámetro del tallo en cada una de las 50 plantas de maíz utilizando un flexómetro (Figura 8). A continuación, se obtuvo el RMSE, empleando la fórmula de la Figura 9.



Figura 8. Obtención de medidas reales o valores observados de longitud y tallo
Fuente: Los autores.

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}}$$

Figura 9. Fórmula error cuadrático medio (RMSE)
Fuente: Los autores.

3. Resultados

3.1. Resultados de la metodología Mobile-D

La Figura 10a presenta la ventana que permite la captura de la imagen digital a través de la activación de la cámara del dispositivo móvil. Para la captura de la imagen se recomienda colocar un objeto referencial junto a la planta, es necesario conocer las medidas del objeto. En la Figura 10b se observa la interfaz que solicita el ingreso de las medidas de longitud y ancho del objeto de referencia, de la misma forma y de manera opcional se solicita la medida del soporte, todas estas en milímetros. La Figura 10c muestra la interfaz con los resultados en milímetros de la longitud y el diámetro del tallo de la imagen de la planta de maíz capturada por medio de la cámara posterior.

4.2. Resultados de la evaluación de la longitud y diámetro de la planta de maíz

En la Figura 11 se observan los valores de la longitud del tallo del conjunto de datos predichos y del conjunto de datos observados de las 50 imágenes digitales, como se aprecia en la gráfica de línea apilada existe una variabilidad entre los dos conjuntos de datos. La figura 12 muestra los valores del diámetro del tallo del conjunto de datos predichos y del conjunto de datos observados de las imágenes digitales, de igual forma los datos presentan cierta variabilidad. El ponderado que se obtiene de la aplicación de RMSE tanto para la longitud como para el diámetro es de 5,41 y 5,27 milímetros respectivamente, estos valores determinan que la aplicación móvil calcula de manera adecuada y confiable las medidas de longitud y diámetro del tallo del maíz en la imagen.

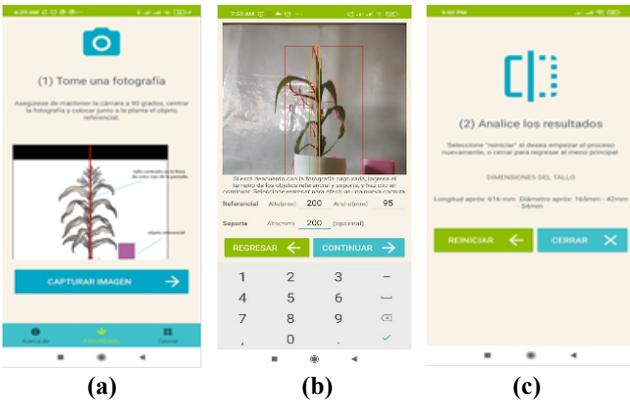


Figura 10. Interfaces aplicación móvil: a) Ventana captura de imagen. b) Ventana ingreso de alto y ancho de objeto referencial. c) Ventana resultados de longitud y diámetro de planta de maíz en imagen digital.

Fuente: Los autores.

4. Conclusiones

Se considera de suma importancia el desarrollo de aplicaciones móviles en el área de la agricultura, específicamente para observar la información fenométrica de la planta del maíz, esta aplicación es una aproximación a dos de los aspectos que conllevan las variables fenométricas: la altura y el diámetro del tallo, el uso del aplicativo permitirá tomar decisiones oportunas para mejorar el crecimiento y rendimiento del cultivo.

Por otra parte, la técnica seleccionada para el cálculo de la longitud y el diámetro del tallo conlleva el uso de un objeto referencial colocado junto a la planta, así como también el uso de técnicas de procesamiento de imágenes como la segmentación y Canny para la detección de bordes. Un RMSE para la longitud de 5,41 milímetros y para el diámetro de 5,27 milímetros garantiza

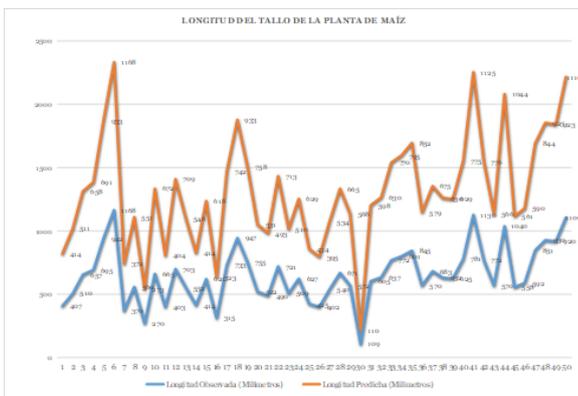


Figura 11. Conjunto de datos predichos y observados de la longitud del tallo
Fuente: Los autores.

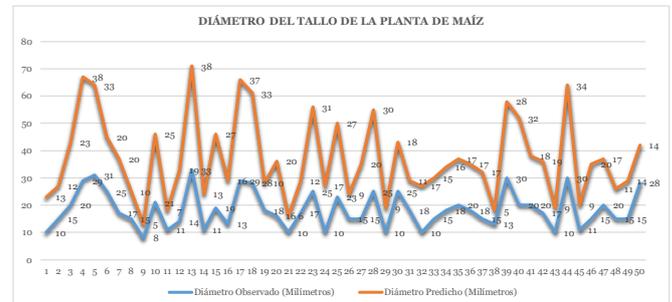


Figura 12. Conjunto de datos predichos y observados del diámetro del tallo
Fuente: Los autores.

el cálculo de la longitud y el diámetro a través de la aplicación móvil.

Contribución de los autores

Vladimir Javier Rojano Guamaní: Conceptualización, Metodología, Software, pruebas. **Ginger Lissbeth Jaramillo Tenezaca:** Conceptualización, Metodología, pruebas. **Karla Susana Cantuña Flores:** Redacción – borrador original del artículo. **Gustavo Adolfo Sandoval Ruilova:** Metodología, Redacción, revisión y edición del artículo. **José María Bengochea Guevara:** Supervisión, Redacción – revisión y edición del artículo.

Conflictos de interés

Los autores declaran no tener ningún conflicto de interés.

Referencias bibliográficas

Alvarado, J. P. (2012). Procesamiento y análisis de imagen digitales. Instituto Tecnológico de Costa Rica.

Carpio Santos, L. K. (2018). El uso de la tecnología en la agricultura. *Pro Sciences: Revista de Producción, Ciencias e Investigación*, 2(14), 25–32. <https://doi.org/10.29018/ISSN.2588-1000VOL2ISS14.2018PP25-32>

López Castañeda, M. (2015). Las Apps y tipos de Apps. Universidad Tecnológica de Pereira. <http://univirtual.utp.edu.co/pandora/recursos/2000/2591/2591.pdf>

Tapia, F. (2006). Innovaciones tecnológicas en la agricultura empresarial mexicana: Una aproximación teórica. http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S1315-85972006000100005

Ventura Elías, R. (2016). Fenología y fenometría de una variedad y una línea de frijol (*Phaseolus vulgaris*) en la zona occidental de El Salvador. *Agronomía Mesoamericana*, 2, 56. <https://doi.org/10.15517/AM.V2I0.25224>

Zambrano Mendoza, J. L., Yáñez, C., Sangoquiza Caiza, C. A., Limongi Andrade, R., Alarcón Cobeña, F., Zambrano Zambrano, E. E., ... & Pinargote García, L. F. (2019). *Situación del cultivo de maíz en Ecuador: investigación y desarrollo de tecnologías en el Iniap*. <http://repositorio.iniap.gob.ec/handle/41000/5457>

Zambrano, W. (2017). Estudio comparativo de metodologías de desarrollo ágil en base al desarrollo de una aplicación móvil. *Ekp*, 13(3), 1576–1580. <http://www.dspace.uce.edu.ec/handle/25000/9445>

Zari Arévalo, B. D. (2014). Determinación de las fases fenológicas, fenométricas e índice de balance hídrico en el cultivo de maíz duro (*Zea mays* L.), bajo condiciones de secano, en el cantón Pindal provincia de Loja. <http://dspace.utpl.edu.ec/handle/123456789/11136>

