



Publicación Cuatrimestral. Vol. 7, No 3, Septiembre/Diciembre, 2022, Ecuador (p. 69-88). Edición continua

<https://revistas.utm.edu.ec/index.php/Basedelaciencia/index>

[revista.bdlaciencia@utm.edu.ec](mailto:revista.bdlaciencia@utm.edu.ec)

Universidad Técnica de Manabí

DOI: <https://doi.org/10.33936/revbasdelaciencia.v7iESPECIAL.3529>

## PREDICCIÓN DEL NIVEL DE DESARROLLO DE UNA CADENA DE SUMINISTRO UTILIZANDO TÉCNICAS DE MACHINE LEARNING

Douglas Andrés Verduga Alcívar<sup>1\*</sup>, Víctor Ernesto Márquez Pérez<sup>2</sup>, Emanuel Guillermo Muñoz Muñoz<sup>3</sup>

<sup>1</sup> Instituto de Posgrado. Instituto de Ciencias Básicas. Universidad Técnica de Manabí. Ecuador. Correo electrónico: [douglas.verduga@utm.edu.ec](mailto:douglas.verduga@utm.edu.ec)

<sup>2</sup> Instituto de Ciencias Básicas. Universidad Técnica de Manabí. Ecuador. Correo electrónico: [victor.marquez@utm.edu.ec](mailto:victor.marquez@utm.edu.ec)

<sup>3</sup> Instituto de Ciencias Básicas. Universidad Técnica de Manabí. Ecuador. Correo electrónico: [emanuel.munoz@utm.edu.ec](mailto:emanuel.munoz@utm.edu.ec)

\*Autor para la correspondencia: [douglas.verduga@utm.edu.ec](mailto:douglas.verduga@utm.edu.ec)

Recibido: 9-6-2021 / Aceptado: 12-12-2022 / Publicación: 26-12-2022

Editor Académico: Carmen Judith Vanegas Espinoza 

### RESUMEN

Las técnicas de predicción del nivel de desarrollo en las cadenas de suministro son importantes para la toma de decisiones ante diferentes escenarios; por esta razón el objetivo de este trabajo es predecir el nivel de desarrollo de una cadena de suministro manabita de clasificación agroalimentaria, cuyo producto fundamental es el agua de coco artesanal. Usando datos históricos, se aplican los métodos de Máquinas de Soporte Vectorial en Regresión y Regresión Lineal Múltiple. Como parte del trabajo se lleva a cabo un estudio de correlación para mejorar el desempeño de los métodos. Los resultados muestran que las medidas de desempeño exponen el éxito predictivo de las SVR, con la configuración correcta del Kernel Lineal a Polinomial mejorando los resultados y el ajuste de los hiperparámetros, logrando una mejor precisión que la solución estándar tradicional de Regresión Lineal Múltiple y el cumplimiento de los supuestos para que sus resultados sean tomados como válidos, a deferencia de las SVM. La metodología utilizada para la evaluación de los métodos implementados a través de un conjunto de entrenamiento y prueba, permite obtener una mejor predicción con relación a estas dos técnicas, obteniendo un modelo que se puede aplicar en otros estudios de casos.

**Palabras clave:** Máquinas de Soporte Vectorial en Regresión, Regresión Lineal Múltiple, Cadenas de suministro, entrenamiento.

### PREDICTION OF THE LEVEL OF DEVELOPMENT OF A SUPPLY CHAIN USING MACHINE LEARNING TECHNIQUES

### ABSTRACT

The techniques for predicting the level of development in supply chains are important for decision-making in different scenarios; For this reason, the aim of this work was to predict the level of development of a Manabi supply chain of agri-food classification, whose main product is artisanal coconut water. Using historical data, the methods of Support Vector Machines in Regression and Multiple Linear Regression were applied. As part of the work, a correlation study is carried



out to improve the performance of the methods. The results showed that the performance measures expose the predictive success of the SVR, with the correct configuration of the Linear to Polynomial Kernel improving the results and the adjustment of the hyperparameters, achieving a better precision than the traditional standard solution of Multiple Linear Regression and the compliance with the assumptions so that their results are taken as valid, unlike the SVM. The methodology used for the evaluation of the methods implemented through a set of training and testing, allowed to obtain a better prediction in relation to these two techniques, obtaining a model that can be applied in other case studies.

**Keys word:** Support Vector Regression Machines, Multiple Linear Regression, Supply chains, training

## PREVISÃO DO NÍVEL DE DESENVOLVIMENTO DE UMA CADEIA DE SUPRIMENTOS USANDO TÉCNICAS DE MÁQUINA DE APRENDIZAGEM

### RESUMO

As técnicas para prever o nível de desenvolvimento em cadeias de abastecimento são importantes para a tomada de decisões em diferentes cenários; por este motivo, o objetivo deste trabalho é prever o nível de desenvolvimento de uma cadeia de abastecimento em Manabi de classificação agroalimentar, cujo produto principal é a água de coco artesanal. Utilizando dados históricos, são aplicados os métodos das Máquinas de Regressão Vetorial de Apoio e Regressão Linear Múltipla. Como parte do trabalho, é realizado um estudo de correlação para melhorar o desempenho dos métodos. Os resultados mostram que as medidas de desempenho expõem o sucesso preditivo do SVR, com a configuração correta do Kernel Linear ao Polinomial melhorando os resultados e o ajuste do hiperparâmetro, conseguindo uma precisão melhor que a solução padrão tradicional de Regressão Linear Múltipla e o cumprimento das suposições para que seus resultados sejam considerados válidos, em deferência às SVMs. A metodologia utilizada para a avaliação dos métodos implementados através de um conjunto de treinamento e testes, permite obter uma melhor previsão em relação a estas duas técnicas, obtendo um modelo que pode ser aplicado em outros estudos de caso.

**Palavras chave:** Máquinas de Regressão de Vetores de Suporte, Regressão Linear Múltipla, Cadeias de suprimentos, treinamento

---

Citación sugerida: Verduga Alcívar, D., Márquez Pérez, V., Muñoz Muñoz, E. (2022). PREDICCIÓN DEL NIVEL DE DESARROLLO DE UNA CADENA DE SUMINISTRO UTILIZANDO TÉCNICAS DE MACHINE LEARNING. Revista Bases de la Ciencia, 7 (No Especial), Diciembre, 69-88. DOI: <https://doi.org/10.33936/revbasdelaciencia.v7iESPECIAL.3529>

---



## 1. INTRODUCCIÓN

La cadena de suministro es una red de proveedores, fábricas, almacenes, centros de distribución y ventas al por menor y por mayor; a través de los cuales se adquieren y se transforman las materias primas, para entregar al cliente productos o servicios (Bautista et.al, 2015; Acevedo et. al. 2020).

El diseño de la cadena de suministro incluye cooperación, localización de planta, sistemas de producción, y la distribución física de la planta, entre otros. Las tareas de planificación a mediano plazo engloban las compras, la planeación de requerimientos de materiales y demanda (pronóstico) e inventarios; mientras que la planificación a corto plazo contempla la planificación del transporte, la distribución y programación de la producción (Albretch, 2010, Sanchez et.al 2020). En aras de mejorar la competitividad de la cadena de suministros es importante predecir el nivel de desarrollo de cada uno de estos elementos que la conforman.

Una forma de determinar este nivel de desarrollo es usar datos históricos y llevar a cabo una regresión que permita predecir y establecer estrategias. Dentro de las técnicas de regresión se encuentran la tradicional Regresión Lineal Múltiple y las Máquinas de Soporte Vectorial en Regresión. En este trabajo serán usados ambos métodos y los resultados serán comparados con la finalidad de observar cuál de ellos posee un mejor rendimiento al momento de predecir el nivel de desarrollo.

De esta manera, el objetivo de este artículo es predecir el nivel de desarrollo de una cadena de suministro manabita con clasificación agroalimentaria, usando Máquina de Soporte vectorial en Regresión y Regresión Lineal Múltiple.

## 2. MATERIALES Y MÉTODOS

El problema de regresión consiste en lo siguiente. Dados un conjunto de datos  $\{(x_1, y_1), \dots, (x_i, y_i)\}$  con  $x_i \in \mathbb{R}^d$  para  $d$  es cualquier espacio de dimensiones y  $y_i$  en  $\mathbb{R}$ , se desea establecer una función  $f$  (llamada función de generalización o predicción) tal que el valor de  $f(x_i)$  se aproxime a  $y_i$ . Con esta función se predice el valor de cualquier dato nuevo.

En este trabajo se utiliza la metodología de entrenamiento de las Máquinas de Soporte Vectorial en Regresión de (Cortes & Vapnik, 1995); y el método tradicional de Regresión Lineal Múltiple (Rodrigo, 2016) para el problema de regresión asociado a la cadena de suministros. A continuación, se presenta una introducción a ambas metodologías.

### Máquina de Vectores de Soporte en Regresión

Las Máquinas de Vectores de Soporte (SVM, del inglés Support Vector Machines) tienen su origen en los trabajos sobre la teoría del aprendizaje estadístico y fueron introducidas en los años 90 por (Cortes & Vapnik, 1995). Aunque originariamente las SVM fueron pensadas para resolver problemas de clasificación binaria y de regresión. En el trabajo de (Wang, 2012) se utilizan el método para la previsión de la demanda de la cadena de suministro basada en el método de regresión de vectores de soporte.

Las Máquinas de Soporte Vectorial también se puede utilizar como método de regresión, manteniendo todas las características principales que describen al método. Las Máquinas de Soporte Vectorial en Regresión (SVR) utiliza los mismos principios del método que la SVM para la clasificación, con solo algunas diferencias menores. En primer lugar, debido a que la salida es un número real, resulta muy difícil predecir la información disponible, que tiene infinitas posibilidades. En el caso de la regresión, se establece un margen de tolerancia  $\varepsilon$  (épsilon) y el algoritmo es más complicado; sin embargo, la idea principal es siempre la misma: minimizar el error, individualizando el hiperplano que maximiza el margen, teniendo en cuenta que se tolera parte del error. (Jos et al., 2016).

Una de las características importantes de las Máquinas de Soporte Vectorial es el hecho de que se puede hacer regresión no lineal de manera implícita tal como se describe a continuación.

Consideramos el conjunto de datos  $\{(x_1, y_1), \dots, (x_i, y_i)\}$ , donde  $x_i \in \mathbb{R}^d$  e  $y_i \in \mathbb{R}$ . Sea  $\Phi: \mathbb{X} \rightarrow \mathcal{F}$  la función que hace corresponder a cada punto de entrada  $x$  un punto en el espacio de características  $\mathcal{F}$ , donde  $\mathcal{F}$  es un espacio de Hilbert. Este espacio de características puede ser de dimensión elevada o incluso infinita

El objetivo va a ser trasladar el conjunto de entrenamiento a un nuevo espacio de características, para encontrar la función que mejor se aproxime a los datos del conjunto. Para este caso la función que estamos buscando será de la ecuación 1:

$$f(x) = \langle w, \phi(x) \rangle + b \tag{Ecuación 1}$$

Planteamos el problema primal, en este caso no depende directamente de las muestras del conjunto, sino de sus imágenes por una cierta función  $\phi$ . como se observa en la ecuación 2

$$\begin{aligned} \min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) & \tag{Ecuación 2} \\ \text{s. a} \quad y_i - \langle w, \phi(x_i) \rangle \leq \varepsilon + \xi_i \quad i = 1, \dots, n & \\ \langle w, \phi(x_i) \rangle - y_i \leq \varepsilon + \xi_i^* \quad i = 1, \dots, n & \\ \xi_i \geq 0, \quad \xi_i^* \geq 0 \quad i = 1, \dots, n & \end{aligned}$$

La constante  $C > 0$  determina el equilibrio entre la regularidad de  $f(x)$  y la cantidad hasta la cual toleramos desviaciones mayores que  $\varepsilon$ . Consideraremos  $\xi_i$  y  $\xi_i^*$  las variables que controlan el error cometido por la función de regresión al aproximar el  $i$ -ésimo ejemplo. Si el valor de la constante  $C$  es muy elevado en el caso límite cuando ( $C \rightarrow \infty$ ) estaríamos considerando que el conjunto está perfectamente representado por nuestro hiperplano predictor ( $\xi_i \rightarrow 0$ ). Por contrario, un número demasiado pequeño para  $C$  permitiría valores de  $\xi_i$  elevados, es decir, estaríamos admitiendo un número muy elevado de datos mal representados.

La complejidad de este problema va a depender de la dimensión en la que se encuentre nuestro estudio, y en este caso, tras ser transformados por la función  $\phi$ , tendríamos una dimensión alta, lo que complicaría en exceso la resolución de este problema primal.

Tras la obtención del problema primal pasamos a plantear el problema dual asociado. Para ello determinamos la función lagrangiana, la idea es construir una función de Lagrange con la función objetivo y las restricciones correspondientes, mediante la introducción de un conjunto de variables duales. El objetivo es obtener el problema dual a partir de las condiciones de dualidad; obteniendo el siguiente problema de optimización dual como está planteado en la ecuación 3:

$$\max -\frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*) (\alpha_i - \alpha_i^*) \langle \phi(x_i), \phi(x_j) \rangle - \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) + \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) \quad \text{Ecuación 3}$$

$$s. a \quad \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) = 0$$

$$\alpha_i, \alpha_i^* \in [0, C]$$

Los problemas primal y dual descritos anteriormente, obedecen una serie de parámetros que pueden hacer variar la solución. Además, cuando introducimos el Kernel, éste depende a su vez de unos parámetros de los cuales se deriva nuestra solución.

### Artificio del Kernel

Tras plantear el problema dual, se observa que la función objetivo solo depende del producto interno de las imágenes de nuestros ejemplos. El truco del Kernel es ampliamente usado en los algoritmos de cálculo de productos internos de la forma  $\langle \phi(x), \phi(x') \rangle$  en el espacio de características  $\mathcal{F}$ .

El truco consiste en que, en lugar de calcular estos productos internos en  $\mathcal{F}$ , se define una función *Kernel*,  $K : X \times X \rightarrow \mathbb{R}$  que asigna a cada par de elementos del espacio de entrada  $X$ , un valor real correspondiente al producto escalar de las imágenes de dichos elementos en el nuevo espacio  $\mathcal{F}$ , es decir,  $K(x, x') = \langle \phi(x), \phi(x') \rangle$ , donde  $\phi: X \rightarrow \mathcal{F}$ , se sustituye la función Kernel en el problema dual asociado. Los primeros en aplicar este tipo de problemas fueron (Cortes & Vapnik, 1995).

El caso lineal es un caso particular donde la función  $\phi$  es igual a la identidad por lo cual se denota Kernel Lineal al definido por el producto escalar de las variables. Para el caso no lineal existen varias funciones Kernel que son usadas. En este trabajo consideraremos el Kernel Polinomial tal como se define a continuación.

### Kernel lineal

Hay varias funciones de Kernel conocidas que se usan comúnmente en conexión con máquina de soporte vectorial.

El primero ejemplo bastante trivial es un Kernel lineal. Se define simplemente como en la ecuación 4:

$$k(x, z) = \phi(x) \cdot \phi(z) = \langle x, z \rangle \quad \text{Ecuación 4}$$

En este caso, el espacio de características y el espacio de entrada son las mismas; es decir que ésta función trabaja en su espacio original.

**Kernel Polinomial** ( $d \geq 2, c > 0$ ) es:

$$k(x, z) = (\langle x, z \rangle + c)^d \quad \text{Ecuación 5}$$

La ecuación 4 mapea el espacio de entrada original en un espacio de características de dimensionalidad  $\binom{n+d}{d}$ , donde  $d$  es el grado del Kernel y  $n$  es la dimensión del espacio de entrada original.

### Regresión Lineal Múltiple

La Regresión Lineal Múltiple permite generar un modelo lineal donde el valor de la variable dependiente o respuesta ( $y$ ) se determina a partir de un conjunto de variables independientes llamadas predictores  $(x_1, x_2, x_3, \dots, x_n)$ . Es una extensión de la Regresión Lineal Simple, por lo que es fundamental comprender esta última. Los modelos de Regresión Múltiple pueden emplearse para predecir el valor de la variable dependiente o para evaluar la influencia que tienen los predictores sobre. (Rodrigo, 2016)

Los modelos Lineales Múltiples siguen la ecuación 6:

$$y_i = (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_n x_{ni}) + e_i \quad \text{Ecuación 6}$$

- $\beta_0$ : es la ordenada en el origen, el valor de la variable dependiente y cuando todos los predictores son cero.



- $\beta_i$ : es el efecto promedio que tiene el incremento en una unidad de la variable predictora  $x_i$  sobre la variable dependiente  $y$ , manteniéndose constantes el resto de variables. Se conocen como coeficientes parciales de regresión.
- $e_i$ : es el residuo o error, la diferencia entre el valor observado y el estimado por el modelo.

### Supuestos del modelo de regresión lineal múltiple

Para poder crear un modelo de regresión lineal es necesario que se cumpla con los siguientes supuestos:

- **Linealidad:** Que la relación entre las variables sea lineal.
- **Independencia:** Que los errores en la medición de las variables explicativas sean independientes entre sí.
- **Homocedasticidad:** Que los errores tengan varianza constante.
- **Normalidad:** Que las variables sigan la Ley Normal.
- **No colinealidad:** Que las variables independientes no estén correlacionadas entre ellas.

Es importante tener en cuenta que la magnitud de cada coeficiente parcial de regresión depende de las unidades en las que se mida la variable predictora a la que corresponde, por lo que su magnitud no está asociada con la importancia de cada predictor. Para poder determinar qué impacto tienen en el modelo cada una de las variables, se emplean los coeficientes parciales estandarizados, que se obtienen al estandarizar (sustraer la media y dividir entre la desviación estándar) las variables predictoras previo al ajuste del modelo. (Rodrigo, 2016).

El entrenamiento de Máquinas de Soporte Vectorial en Regresión de (Cortes & Vapnik, 1995), y la comparación con el método tradicional de Regresión Lineal Múltiple, se realizan en el programa Rstudio<sup>1</sup> de IDE (en inglés Integrated Development Environment) de Neuralnet.

## 3. RESULTADOS Y DISCUSIÓN

La cadena objeto de estudio tiene 160 actores y 7 eslabones (Romero et.al.2020). Estos últimos se enfocan en: proveedores, productores, comerciantes mayoristas, distribuidores, fabricantes, vendedores y clientes. La misma se clasifica como agroalimentaria porque su producto fundamental es el agua de coco de forma artesanal.

A partir de la identificación de los actores, se aplicó la lista de chequeo (Sablón-Cossío et al., 2014) para la evaluación del nivel de desarrollo y se analizaron los valores de las variables de entrada y de salida.

---

<sup>1</sup>Es una aplicación informática que proporciona servicios integrales para facilitarle al desarrollador o programador el desarrollo de software

El nivel de desarrollo de la cadena se verá reflejado en los valores estimados por los modelos de Máquinas de Soporte Vectorial en Regresión y Regresión Lineal Múltiple. El 30% de los datos de los valores reales se compararán con los estimados obtenidos por cada método para determinar cuál de las metodologías empleadas provee mejor predicción.

La cadena de suministro utilizada en esta investigación cuenta con 121 ítems agrupados en 17 variables, las cuales se enuncian a continuación: configuración de las redes de valor, gestión de la integración, planificación colaborativa, gestión de capacidades, gestión de la demanda, gestión de inventarios, gestión de pedidos, tecnología de información y comunicaciones, desarrollo gerencial, gestión de la innovación, coordinación estratégica, servicio al cliente, desarrollo del personal, desempeño de la red de valor, y desarrollo del producto o servicio. Estas variables son las consideradas para medir el nivel de desarrollo de la cadena de suministro. A continuación en la **tabla 1**, se describen los valores de la aplicación de la lista de chequeo y el análisis de las 16 dimensiones. Con 16 variables de entrada y una de salida cada una con su respectiva correspondencia (**Tabla 2**).

**Tabla 1.** Variables de entrada y de respuesta.

ACTOR	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	NC
1 Actor 1	1.36	2.80	1.46	1.17	2.17	2.8	1.78	1.50	1.22	1.00	1	2.22	1.57	4.00	1.17	1.82
2 Actor 2	1.36	2.80	1.46	1.17	2.17	2.8	1.78	1.50	1.22	1.00	1	2.22	1.57	3.91	1.17	1.81
3 Actor 3	1.36	2.80	1.46	1.17	2.17	2.8	1.78	1.50	1.22	1.00	1	2.22	1.57	3.82	1.17	1.80
4 Actor 4	1.36	2.80	1.46	1.17	2.17	2.8	1.78	1.50	1.22	1.00	1	2.22	1.57	3.82	1.17	1.80
5 Actor 5	1.36	2.80	1.46	1.17	2.17	2.8	1.78	1.50	1.22	1.00	1	2.22	1.57	3.82	1.17	1.80
6 Actor 6	1.36	2.80	1.46	1.17	2.17	2.8	1.78	1.50	1.22	1.00	1	2.22	1.57	3.82	1.17	1.80
7 Actor 7	1.36	2.80	1.46	1.17	2.17	2.8	1.78	1.50	1.22	1.00	1	2.22	1.57	3.82	1.17	1.80
8 Actor 8	1.36	2.80	1.46	1.17	2.17	2.8	1.78	1.50	1.22	1.00	1	2.22	1.57	3.82	1.17	1.80
9 Actor 9	1.36	2.80	1.46	1.17	2.17	2.8	1.78	1.50	1.22	1.00	1	2.22	1.57	3.82	1.17	1.80
10 Actor 10	1.36	2.80	1.46	1.17	2.17	2.8	1.78	1.50	1.22	1.00	1	2.22	1.57	3.82	1.17	1.80
11 Actor 11	1.36	2.80	1.46	1.17	2.17	2.8	1.78	1.50	1.22	1.00	1	2.22	1.57	3.82	1.17	1.80
12 Actor 12	1.36	2.80	1.46	1.17	2.17	2.8	1.78	1.50	1.22	1.00	1	2.22	1.57	3.82	1.17	1.80
13 Actor 13	1.36	2.80	1.46	1.17	2.17	2.8	1.78	1.50	1.22	1.00	1	2.22	1.57	3.82	1.17	1.80
14 Actor 14	1.36	2.80	1.46	1.17	2.17	2.8	1.78	1.50	1.22	1.00	1	2.22	1.57	3.82	1.17	1.80
15 Actor 15	1.36	2.80	1.46	1.17	2.17	2.8	1.78	1.50	1.22	1.00	1	2.22	1.57	4.00	1.17	1.82
16 Actor 16	1.36	2.80	1.46	1.17	2.17	2.8	1.78	1.50	1.22	1.00	1	2.22	1.57	3.82	1.17	1.80
17 Actor 17	1.36	2.80	1.46	1.17	2.17	2.8	1.78	1.50	1.22	1.00	1	2.22	1.57	3.82	1.17	1.80

Showing 1 to 17 of 160 entries

**Fuente:** Elaboración propia.

Debido a la complejidad de la lista de chequeo por su amplitud, se aplica un análisis de correlación, para investigar la asociación entre dos o más variables y así intentar reducir su número. Se procedió a hacer la matriz de correlación (**Tabla 3**) para analizar la correlación entre múltiples variables al mismo tiempo. La distribución de cada variable se muestra en la diagonal, en la parte inferior de la diagonal se muestran los gráficos de dispersión bivariados con una línea ajustada, en la parte superior de la diagonal el valor de la correlación más el nivel de significancia como estrellas y cada nivel de significancia está asociado a un símbolo: valores p (0, 0.001, 0.01, 0.05, 0.1, 1)  $\Leftrightarrow$  símbolos (“\*\*\*”, “\*\*”, “\*”, “.”, “ ”).

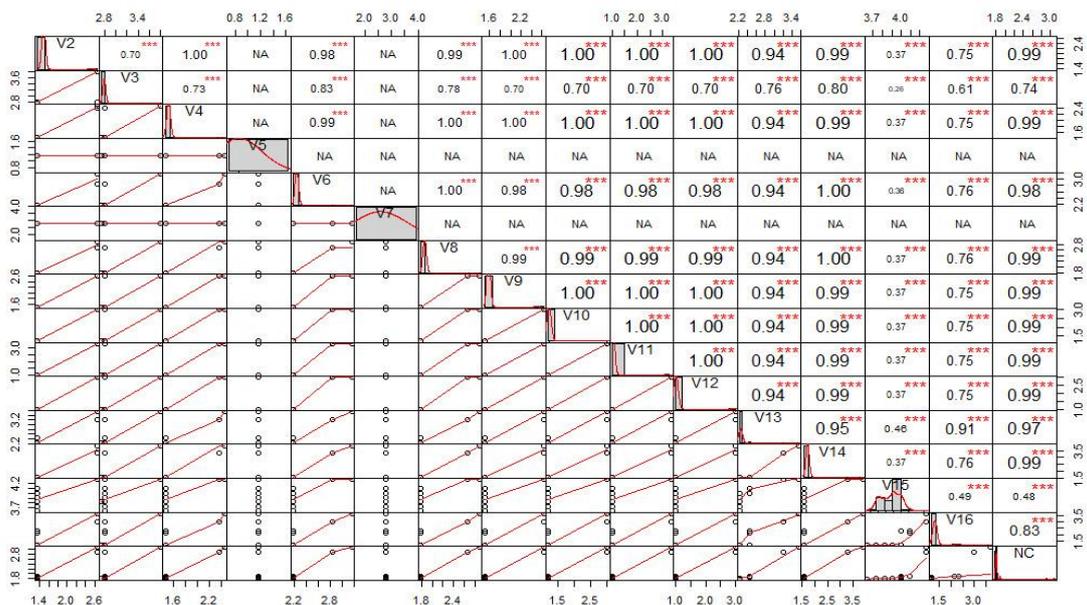
La idea es descartar las variables correlacionadas por encima del 80% positivo o por debajo del 80% negativo con el análisis de correlación.

**Tabla 2.** Correspondencia de las variables.

Variables	Nombres de las Variables
V1	Elemento
V2	Configuración de la cadena
V3	Gestión de capacidades
V4	Gestión de la integración
V5	Planificación colaborativa
V6	Gestión de la demanda
V7	Gestión de inventarios
V8	Gestión de pedidos
V9	Tecnología de información y comunicaciones
V10	Desarrollo gerencial
V11	Gestión de la innovación
V12	Coordinación estratégica
V13	Servicio al cliente
V14	Desarrollo del personal
V15	Desempeño de la cadena de suministro
V16	Desarrollo del producto o servicio
V17	Nivel de la cadena

Fuente: Elaboración propia.

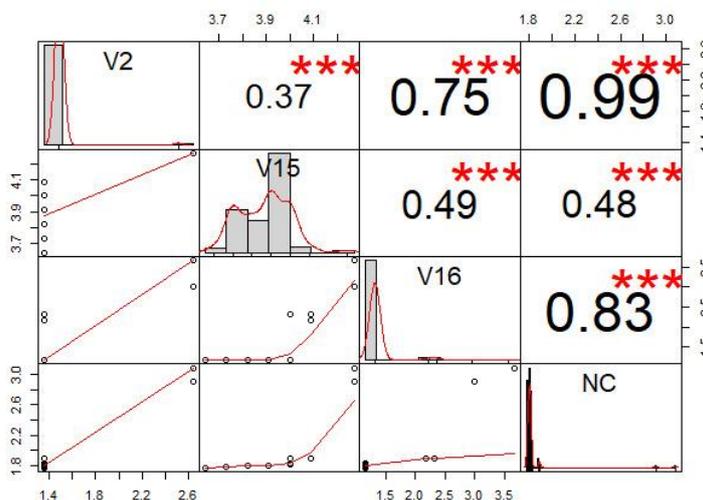
**Tabla 3.** Matriz de correlación.



Fuente: Elaboración propia.

Una vez realizado el análisis de correlación de las 17 variables, se obtuvo como resultado 4 variables (Configuración de la cadena, Desempeño de la cadena de suministro, Desarrollo del producto o servicio y Nivel de la cadena) las cuales aportan significativamente al estudio. (**Tabla 4**).

**Tabla 4.** Matriz de variables resultantes.



Fuente: Elaboración propia.

Se realizó el entrenamiento al **modelo tradicional de Regresión Lineal Múltiple** con el 70% de los datos, el otro 30% se utilizó para probar el modelo y obtener la estimación de los valores. Los resultados estimados de la variable resultante (**Tabla 5**) (**Ver anexos**), demuestran que la utilización del modelo de Regresión Lineal Múltiple realiza mejores estimaciones de las variables de entrada en la etapa inicial. Esos estimadores se comparan con los valores reales y se calcula el coeficiente de correlación de Pearson (Restrepo et al., n.d.), con el fin de establecer que los datos tengan valores parecidos y puedan tomar un rango de valores de +1 a -1. Un valor de 0 indica que no hay asociación entre las dos variables y un valor mayor que 0 indica una asociación positiva. El resultado del coeficiente de correlación de la cadena es de 0.8145, lo cual indica que existe una correlación alta y que la estimación del modelo es óptima (**Figura 1**) (**Ver anexos**), además se aplicaron otras medidas de desempeño para estimar el rendimiento y evaluar el ajuste del modelo tales como: el error cuadrático medio (RMSE, por sus siglas en inglés, root mean squared error), error absoluto medio (MAE, mean absolute error) y R-cuadrado o Coeficiente de determinación (**Tabla 6**).

**Tabla 6.** Medidas de desempeño del modelo de Regresión Lineal Múltiple.

Medidas de Desempeño	Valores
Error absoluto medio (MAE)	0.0249
Error cuadrático medio (RMSE)	0.0887
Coeficiente de determinación (R cuadrado)	0.6584
Coeficiente de correlación de Pearson	0.8145

Fuente: Elaboración propia.

**El error cuadrático medio (RMSE)** indica el ajuste absoluto del modelo a los datos, cuán cerca están los puntos de datos observados de los valores predichos del modelo. El RMSE es una medida absoluta de ajuste. Como la raíz cuadrada de una varianza, RMSE se puede interpretar como la desviación estándar de la varianza inexplicada, y tiene la propiedad útil de estar en las mismas unidades que la variable de respuesta. Los valores más bajos de RMSE indican un mejor ajuste. RMSE es una buena medida de la precisión con que el modelo predice la respuesta, y es el criterio más importante para ajustar si el propósito principal del modelo es la predicción.

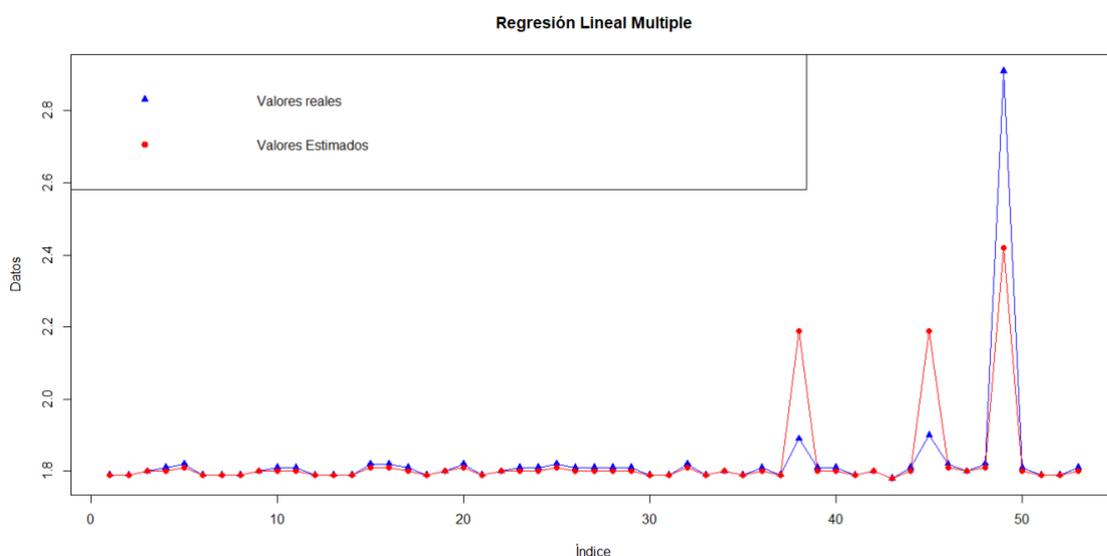
**Error absoluto medio (MAE)** es el promedio de la diferencia absoluta entre el valor observado y los valores predichos. El error absoluto medio o MAE es un puntaje lineal, lo que significa que todas las diferencias individuales se ponderarán por igual en el promedio.

**R-cuadrado** indica la bondad o la aptitud del modelo, tiene la propiedad útil de que su escala es intuitiva, va de 0 a 1, con 0 indicando que el modelo propuesto no mejora la predicción sobre el modelo medio y 1 indica una predicción perfecta. La mejora en el modelo de regresión da como resultado aumentos proporcionales en R-cuadrado. Estas métricas fueron tomadas (Métricas En Regresión | Interactive Chaos, n.d.).

Aplicando estas medidas de desempeño se evalúa el modelo tradicional de Regresión Lineal Múltiple obteniendo los resultados (**Tabla 6**) los mismos que permiten determinar que el rendimiento del modelo es óptimo.

Una vez aplicado el modelo de Regresión Lineal Múltiple a los datos, se puede observar el comportamiento entre los valores reales y los estimados por el modelo (**Figura 2**), además se puede apreciar que existen un comportamiento muy similar entre ellos.

**Figura 2.** Comportamiento de los valores reales y los estimados del modelo de Regresión Lineal Múltiple.



**Fuente:** Elaboración propia.

A partir del análisis del modelo de **Máquinas de Soporte Vectorial en Regresión** y la utilización del **Kernel Lineal** se obtiene el modelo de entrenamiento. El cual se entrenó en el **software R** utilizando la **librería e1071** que contiene el paquete SVM donde se estableció el conjunto de datos de entrenamiento (data = entrenar), y se realizó el ajuste de los parámetros como el tipo de SVM en este caso Regresión (SVM-Tipo: eps-regresión), el kernel (Kernel = "lineal"), el parámetro de regularización (coste: 1), y el  $\epsilon$  ( $\epsilon$ : 0.1) especificando el tubo  $\epsilon$  dentro del cual no se asocia ninguna penalización en la función de pérdida de entrenamiento con puntos predichos dentro de una distancia  $\epsilon$  del valor real, una vez entrenado el modelo se obtuvieron 8 vectores de soporte.

Los valores reales corresponden al 30% de los datos utilizados para probar el modelo, el tiempo de ejecución del algoritmo de aprendizaje de las Máquinas de Soporte Vectorial en Regresión es de 3 segundos. Se puede observar que los resultados estimados de la variable resultante (**Tabla 7**) (**Ver anexos**), indican que el uso del modelo de Máquinas de Soporte Vectorial en Regresión con el Kernel Lineal realiza estimaciones que no superan al modelo de Regresión Lineal Múltiple.

Este modelo se evalúa bajo los mismos parámetros del modelo anterior, calculando el coeficiente de correlación de Pearson, cuyo resultado del coeficiente de correlación de la cadena es de 0.8121, que aun siendo alto no supera al modelo tradicional de Regresión Lineal Múltiple (**Figura 3**) (**Ver anexos**), además se aplicaron las medidas de desempeño como el error cuadrático medio (RMSE) con 0.1268, error absoluto medio (MAE) de 0.0275 y R-cuadrado de 0.3010 (**Tabla 8**).

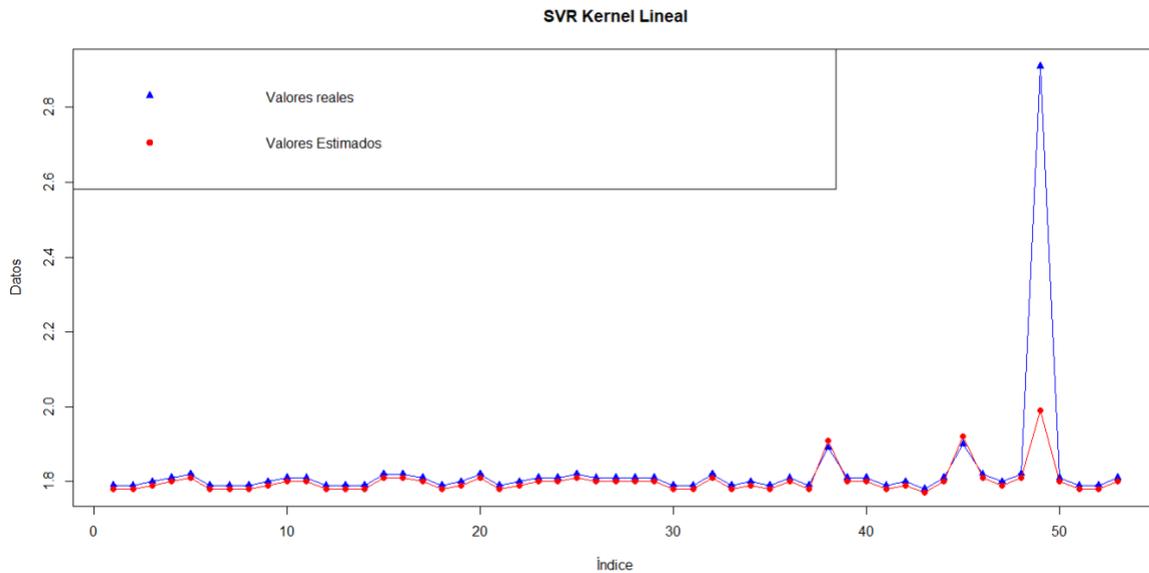
**Tabla 8.** Medidas de desempeño del modelo de Máquinas de Soporte Vectorial en Regresión con el Kernel Lineal.

<b>Medidas de Desempeño</b>	<b>Valores</b>
Error absoluto medio (MAE)	0.0275
Error cuadrático medio (RMSE)	0.1268
Coefficiente de determinación (R cuadrado)	0.3010
Coefficiente de correlación de Pearson	0.8121

**Fuente:** Elaboración propia.

Una vez aplicado el modelo de Máquinas de Soporte Vectorial en Regresión con el Kernel Lineal a los datos, se puede observar el comportamiento entre los valores reales y los estimados por el modelo (**Figura 4**), además se puede apreciar que existen un comportamiento muy similar entre ellos, variando significativamente en uno de los picos de los datos.

**Figura 4.** Comportamiento de los valores reales y los estimados del modelo de Máquinas de Soporte Vectorial en Regresión con el Kernel Lineal.



**Fuente:** Elaboración propia.

Los valores de los parámetros considerados para el **Kernel Polinomial** fueron el tipo de **SVM** en este caso Regresión (SVM-Tipo: eps-regresión), el kernel (Kernel = " polinomial"), el parámetro de regularización (coste: 1), grado de la función del núcleo Polinomial (grado: 3), Coeficiente de Kernel (medida de similitud entre dos puntos) (gama: 1), término independiente en función del Kernel (coeficiente.0: 0) y el épsilon (épsilon: 0.1) especificando el tubo épsilon dentro del cual no se asocia ninguna penalización en la función de pérdida de entrenamiento con puntos predichos dentro de una distancia épsilon del valor real, una vez entrenado el modelo se obtuvieron 10 vectores de soporte, dado los mismos criterios de evaluación de los modelos anteriores, se procedió a realizar el cambio del **Kernel de Lineal a Polinomial** para ver si se mejoraban los resultados obtenidos.

En los resultados estimados de la variable resultante (**Tabla 9**) (**Ver anexos**), se puede observar mejores estimaciones que en los modelos anteriores. Cabe mencionar que el coeficiente de correlación de la cadena muestra un resultado de 0.9612, (**Figura 5**) (**Ver anexos**) revelando un valor muy superior a diferencia de los modelos de SVR con Kernel Lineal y el tradicional de Regresión Lineal Múltiple. También se aplicaron las mismas medidas de desempeño (**Tabla 10**), las cuales permiten observar que el desempeño es mejor en comparación a los otros modelos.

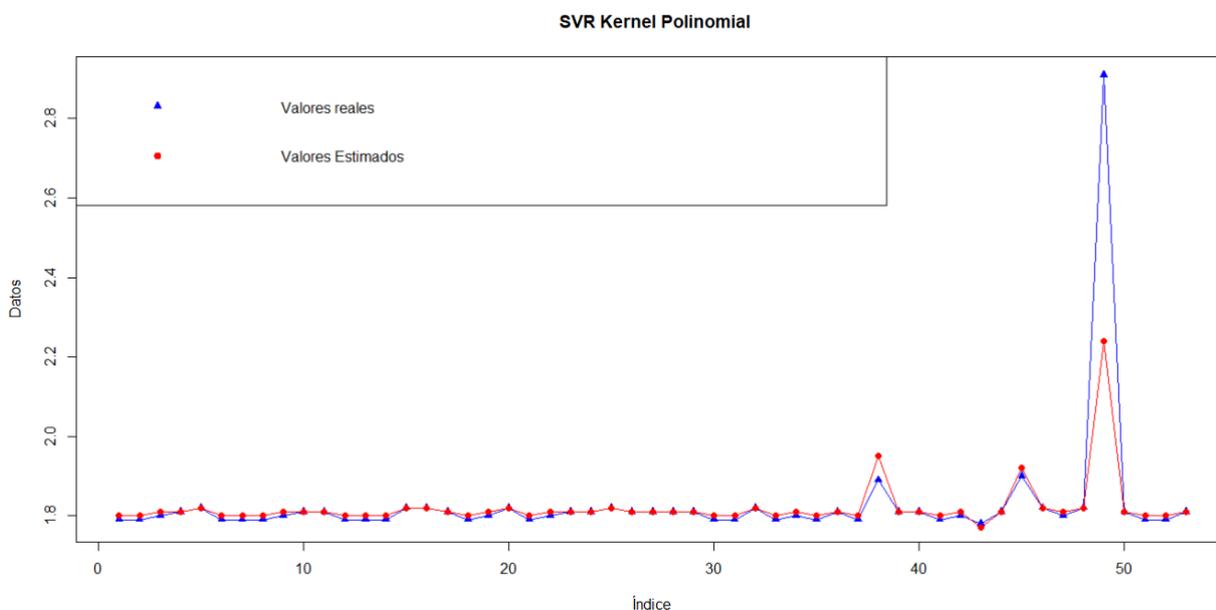
**Tabla 10:** Medidas de desempeño del modelo de Máquinas de Soporte Vectorial en Regresión con el Kernel Polinomial.

Medidas de Desempeño	Valores
Error absoluto medio (MAE)	0.0927
Error cuadrático medio (RMSE)	0.0191
Coefficiente de determinación (R cuadrado)	0.6264
Coefficiente de correlación de Pearson	0.9612

Fuente: Elaboración propia.

Una vez aplicado el modelo de Máquinas de Soporte Vectorial en Regresión con el Kernel Polinomial a los datos, se puede observar el comportamiento entre los valores reales y los estimados por el modelo (**Figura 6**), además se puede apreciar que existen un comportamiento muy similar entre ellos y con más exactitud, variando en uno de los picos de los datos.

**Figura 6:** Comportamiento de los valores reales y los estimados del modelo de Máquinas de Soporte Vectorial en Regresión con el Kernel Polinomial.



Fuente: Elaboración propia.

#### 4. CONCLUSIONES

Las técnicas de predicción del nivel de desarrollo en las cadenas de suministro de tipo agroalimentaria, son una necesidad para la toma de decisiones ante diferentes escenarios, principalmente por las características de los productos que se enfocan en los alimentos y sus derivados.

Esta investigación se desarrolló usando datos reales de una cadena de suministro manabita de producción de agua de coco, y la predicción de sus niveles de desarrollo aplicando dos técnicas de regresión: las Máquinas de Soporte Vectorial y la tradicional Regresión Lineal Múltiple. Los resultados obtenidos a través de las medidas de desempeño como Error absoluto medio (MAE), Error cuadrático medio (RMSE), Coeficiente de determinación (R cuadrado) y Coeficiente de correlación de Pearson demostraron el éxito predictivo de las Máquinas de Soporte Vectorial para Regresión en este contexto, con la configuración correcta del Kernel Lineal a Polinomial para mejorar los resultados obtenidos y el ajuste de los hiperparámetros, logrando una mejor precisión que la solución estándar tradicional de Regresión Lineal Múltiple, que por ser una técnica paramétrica necesita del cumplimiento de los supuestos para que sus resultados sean tomados como válidos, a diferencia de las SVM, que por ser una técnica de inteligencia artificial el cumplimiento de estos supuestos no son necesarios. Es importante mencionar que la configuración de los parámetros y la elección del Kernel es de relevancia para la aplicación de las técnicas de regresión consideradas. Su combinación con técnicas de selección de variables, como la de análisis de correlación, es crucial para la obtención de resultados significativos de interés práctico.

## 5 DECLARACIÓN DE CONFLICTO DE INTERÉS DE LOS AUTORES

Los autores declaran no tener conflicto de intereses

## 6 REFERENCIAS

- Acevedo Urquiaga, J. A., Sablón Cossío, N., Acevedo Suárez, J. A., González Rodríguez, Y. & Martín Marrero, A. (2020). Analysis of the hospital supply chain in Cuban territory. *DYNA*, 87(215), 9-17.
- Albrecht, M., Supply chain coordination mechanisms, *New Approaches for Collaborative Planning*, *LecturNotes in Economics and Mathematical Systems*, Ed. Springer Germany, 2010, pp. 5-62.
- Angel, I. J., Trejo, R., Yu, W., & México, L. (2006). Las maquinas de vectores de soporte para identificación en línea.
- Bautista Santos, H., Martínez Flores, J. L., Fernández Lambert, G., Bernabé Loranca, M. B., Sánchez, F., & Sablón Cossío, N. (2015). Integration model of collaborative supply chain. *DYNA Medellín*, Vol 83. No. 193. DOI: <http://dx.doi.org/10.15446/dyna.v82n193.47370>.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/bf00994018>
- Jos, J., Bravo, R. B., & Priego, E. C. (2016). Trabajo Fin De Grado Support Vector Regression: Propiedades Y Aplicaciones. Universidad de Sevilla. Departamento de Estadística e Investigación Operativa, 60.
- López-Joy, T., Acevedo-Urquiaga, A. J., Gómez-Acosta, M. I., Neira-Mugercia, A. N., Acevedo-Suárez, J. A., & Peña-García, C. (2020). Medicines value chain management level in Cuban health system. *Ingeniería y Universidad*, 24.
- Métricas en regresión / Interactive Chaos*. (n.d.). Retrieved April 19, 2022, from <https://interactivechaos.com/es/manual/tutorial-de-machine-learning/metricas-en-regresion>
- Restrepo, L. F., Julián González L., Est, E., & Bioma, E. (n.d.). SELECCIONES De Pearson a Spearman From Pearson to Spearman.
- Rodrigo, J. A. (2016). Introducción a la Regresión Lineal Múltiple. [https://www.cienciadedatos.net/documentos/25\\_regresion\\_lineal\\_multiple](https://www.cienciadedatos.net/documentos/25_regresion_lineal_multiple).
- Romero Delgado, V. M., Rosado Zambrano, G. V., Sablón Cossío, N., & Burbano Mera, L. (2020). Análisis de la cadena agroalimentaria del coco (cocos nucifera) en la provincia de Manabí, Ecuador. *La Técnica*. (Etapa de publicación)

Sánchez-Galván, F., Bautista-Santos, H., Martínez-Flores, J. L., Ireta-Paredes, A. D. R., & Sánchez-Partida, D. (2020). Cadena de suministro de productos agrícolas de traspatio. *AgroProductividad*, 13(8).

Sablón-Cossío, N. I., Julia Acevedo-Urquiaga, A. I., & Antonio Acevedo-Suárez, J. I. (2014). Propuesta para la evaluación de la planificación colaborativa de la cadena de suministro Proposal for the evaluation of supply chain collaborative planning. [www.vics.org](http://www.vics.org)

Shawe-Taylor, J., & Cristianini, N. (2004). Kernel Methods for Pattern Analysis. In *Kernel Methods for Pattern Analysis*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511809682>

Wang, G. (2012). Demand forecasting of supply chain based on Support Vector Regression method. *Procedia Engineering*, 29, 280–284. <https://doi.org/10.1016/J.PROENG.2011.12.707>

**Contribución de autores**

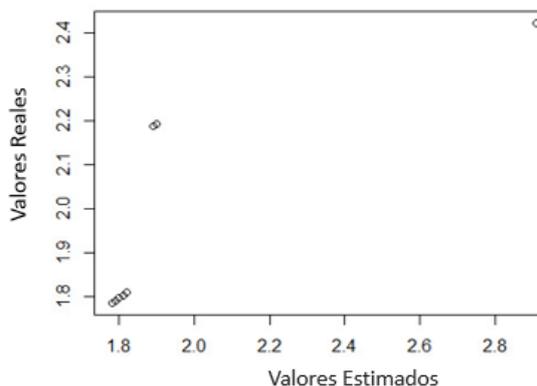
Autor	Contribución
Douglas Verduga	Metodología, revisión, búsqueda bibliográfica y diseño del artículo
Víctor Márquez	Concepción, redacción y análisis microbiológicos
Emanuel Muñoz	Diseño experimental y análisis estadístico

**7 ANEXOS**

**Tabla 5:** Estimadores de la variable resultante de la cadena de suministro manabita con Regresión Lineal Múltiple.

Registro	Valores Reales	Valores Estimados
72	1.79	1.79
63	1.79	1.79
77	1.80	1.80
145	1.81	1.80
132	1.82	1.81
81	1.79	1.79
68	1.79	1.79
53	1.79	1.79
3	1.80	1.80
18	1.81	1.80
104	1.81	1.80
39	1.79	1.79
34	1.79	1.79
51	1.79	1.79
115	1.82	1.81
123	1.82	1.81
112	1.81	1.80
56	1.79	1.79
20	1.80	1.80
127	1.82	1.81
80	1.79	1.79
6	1.80	1.80
154	1.81	1.80
49	1.81	1.80

**Figura 1:** Coeficiente de correlación de la cadena de suministro manabita con Regresión Lineal Múltiple.



**Fuente:** Elaboración propia.

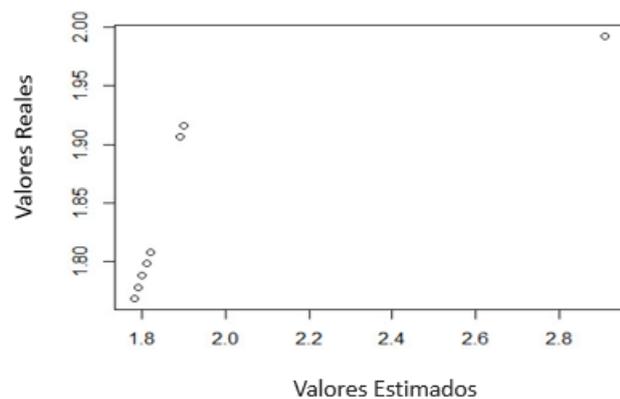
156	1.82	1.81
107	1.81	1.80
140	1.81	1.80
48	1.81	1.80
101	1.81	1.80
35	1.79	1.79
65	1.79	1.79
103	1.82	1.81
57	1.79	1.79
43	1.80	1.80
79	1.79	1.79
151	1.81	1.80
69	1.79	1.79
114	1.89	2.19
148	1.81	1.80
144	1.81	1.80
40	1.79	1.79
17	1.80	1.80
73	1.78	1.78
45	1.81	1.80
89	1.90	2.19
97	1.82	1.81
78	1.80	1.80
119	1.82	1.81
130	2.91	2.42
150	1.81	1.80
66	1.79	1.79
67	1.79	1.79
126	1.81	1.80

Fuente: Elaboración propia.

**Tabla 7:** Resultados de la variable resultante de la cadena de suministro manabita con modelo SVR Kernel Lineal.

Registro	Valores Reales	Valores Estimados
72	1.79	1.78
63	1.79	1.78
77	1.80	1.79
145	1.81	1.80
132	1.82	1.81
81	1.79	1.78
68	1.79	1.78
53	1.79	1.78
3	1.80	1.79

**Figura 3:** Coeficiente de correlación de la cadena de suministro manabita con modelo SVR Kernel Lineal.



Fuente: Elaboración propia.

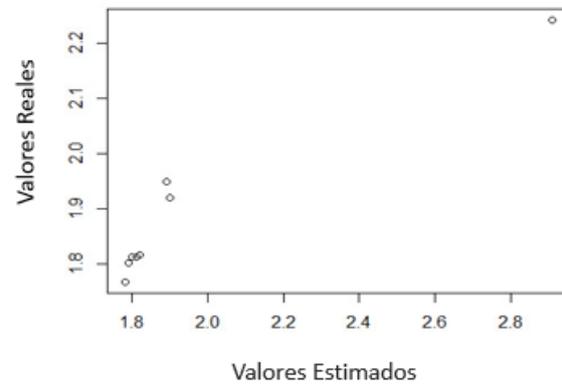
18	1.81	1.80
104	1.81	1.80
39	1.79	1.78
34	1.79	1.78
51	1.79	1.78
115	1.82	1.81
123	1.82	1.81
112	1.81	1.80
56	1.79	1.78
20	1.80	1.79
127	1.82	1.81
80	1.79	1.78
6	1.80	1.79
154	1.81	1.80
49	1.81	1.80
156	1.82	1.81
107	1.81	1.80
140	1.81	1.80
48	1.81	1.80
101	1.81	1.80
35	1.79	1.78
65	1.79	1.78
103	1.82	1.81
57	1.79	1.78
43	1.80	1.79
79	1.79	1.78
151	1.81	1.80
69	1.79	1.78
114	1.89	1.91
148	1.81	1.80
144	1.81	1.80
40	1.79	1.78
17	1.80	1.79
73	1.78	1.77
45	1.81	1.80
89	1.90	1.92
97	1.82	1.81
78	1.80	1.79
119	1.82	1.81
130	2.91	1.99
150	1.81	1.80
66	1.79	1.78
67	1.79	1.78
126	1.81	1.80

**Fuente:** Elaboración propia.

**Tabla 9:** Resultados de la variable resultante de la cadena de suministro manabita con modelo SVR Kernel Polinomial.

Registro	Valores Reales	Valores Estimados
72	1.79	1.80
63	1.79	1.80
77	1.80	1.81
145	1.81	1.81
132	1.82	1.82
81	1.79	1.80
68	1.79	1.80
53	1.79	1.80
3	1.80	1.81
18	1.81	1.81
104	1.81	1.81
39	1.79	1.80
34	1.79	1.80
51	1.79	1.80
115	1.82	1.82
123	1.82	1.82
112	1.81	1.81
56	1.79	1.80
20	1.80	1.81
127	1.82	1.82
80	1.79	1.80
6	1.80	1.81
154	1.81	1.81
49	1.81	1.81
156	1.82	1.82
107	1.81	1.81
140	1.81	1.81
48	1.81	1.81
101	1.81	1.81
35	1.79	1.80
65	1.79	1.80
103	1.82	1.82
57	1.79	1.80
43	1.80	1.81
79	1.79	1.80
151	1.81	1.81
69	1.79	1.80
114	1.89	1.95
148	1.81	1.81
144	1.81	1.81

**Figura 5:** Coeficiente de correlación de la cadena de suministro manabita con modelo SVR Kernel Polinomial.



**Fuente:** Elaboración propia.

40	1.79	1.80
17	1.80	1.81
73	1.78	1.77
45	1.81	1.81
89	1.90	1.92
97	1.82	1.82
78	1.80	1.81
119	1.82	1.82
130	2.91	2.24
150	1.81	1.81
66	1.79	1.80
67	1.79	1.80
126	1.81	1.81

**Fuente:** Elaboración propia.