

**Instituto Tecnológico de Estudios Superiores de Occidente**

# **Construcción de modelos predictivos del precio del maíz utilizando dos métodos convexos: $SVR$ y regresión lineal múltiple**

**Departamento de Matemáticas y Física**

**Optimización convexa**

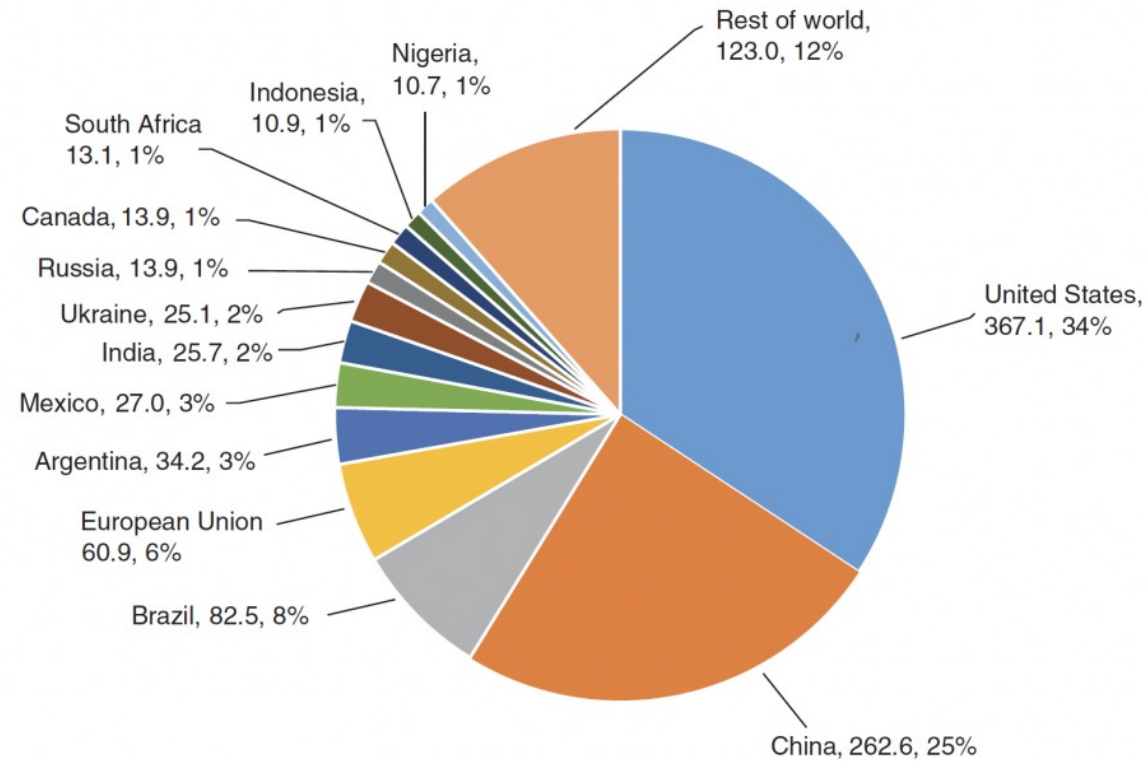
**Dr. Juan Diego Sánchez Torres**

## **INTEGRANTES**

Emilio Carranza Ávila  
Gustavo Ibáñez Sosa  
Gabriela Lozano Orozco  
Karen Manguart Páez  
11 de mayo del 2022

# 1. INTRODUCCIÓN

The United States and Mexico are among the world's largest corn producers



Note: Figure depicts average annual corn production by country during U.S. marketing years 2015/16 to 2017/18. Data labels indicate country, average annual production (millions of metric tons), and share of world production (percent).

# 1. INTRODUCCIÓN

- Importancia del precio del maíz en México.
  - México segundo importador de maíz en el mundo.
  - Séptimo productor mundial.
  - Gran volatilidad en los precios del maíz.
- 
- Datos a usar.
  - Serie de tiempo de precios del maíz.
  - 1959 hasta la actualidad.

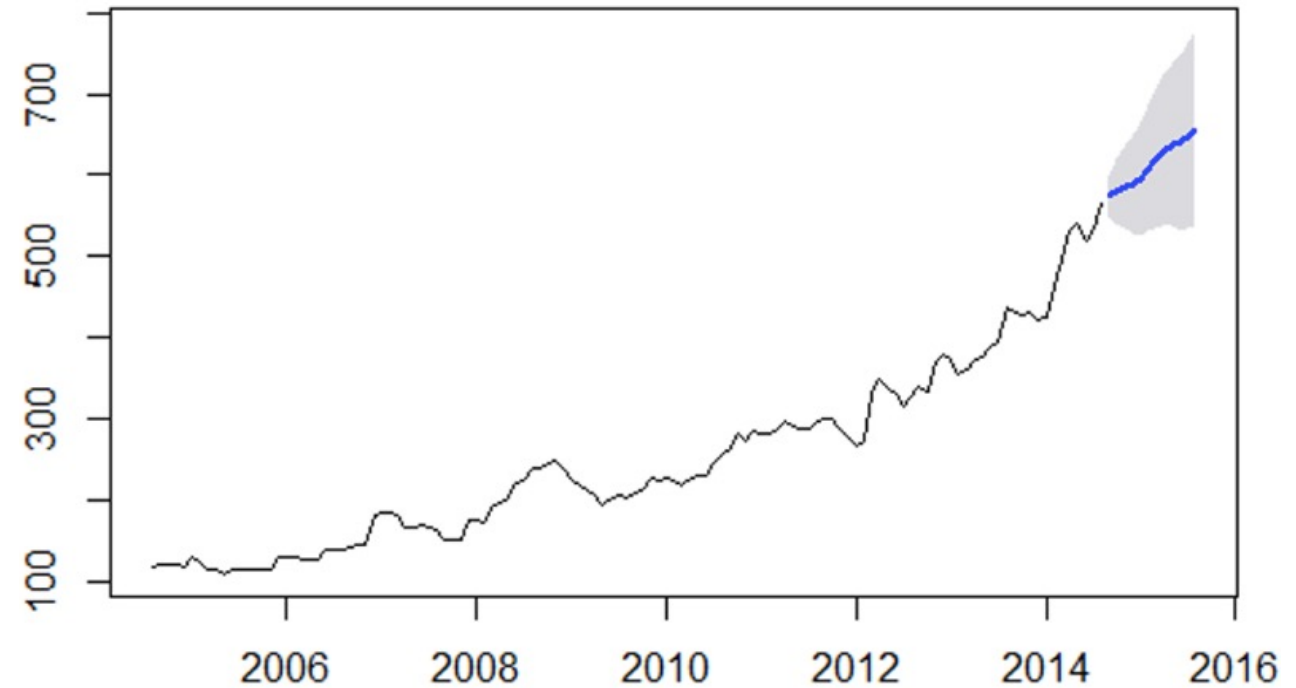
## 2. OBJETIVOS

- Construir dos modelos predictivos de una serie de tiempo del precio del maíz utilizando dos métodos de optimización convexa:
  - *SVR*
  - Regresión lineal múltiple.
- Calcular métricas de error para poder determinar cuál de los dos métodos predice con mejor exactitud con respecto al tiempo.

# 3. MARCO TEÓRICO

## 3.1 Series de Tiempo

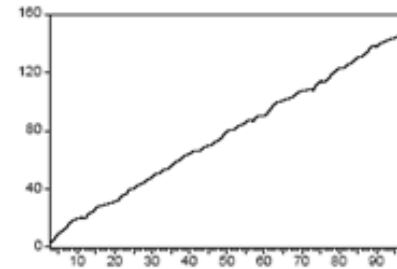
- Secuencia de observaciones ordenadas cronológicamente
- Datos espaciados entre sí de forma uniforme
- Objetivo: Realizar pronósticos



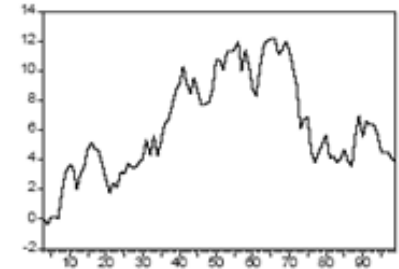
# 3. MARCO TEÓRICO

## 3.1 Series de Tiempo

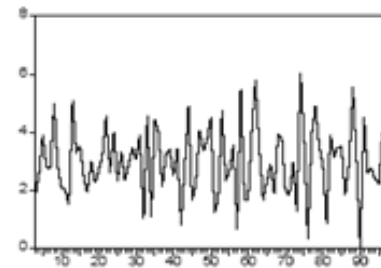
- Aplicaciones en economía, ingeniería y finanzas.
- Modelos estadísticos tradicionales:
  - **ARIMA**
  - Funciones de transferencia
  - Modelos para predicciones de series no lineales (**Redes neuronales** y **Máquinas de Soporte**)



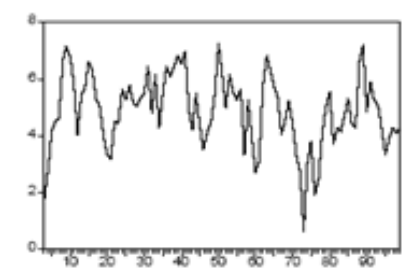
Serie no estacionaria en media



Serie no estacionaria en media y varianza



Serie estacionaria en media y varianza

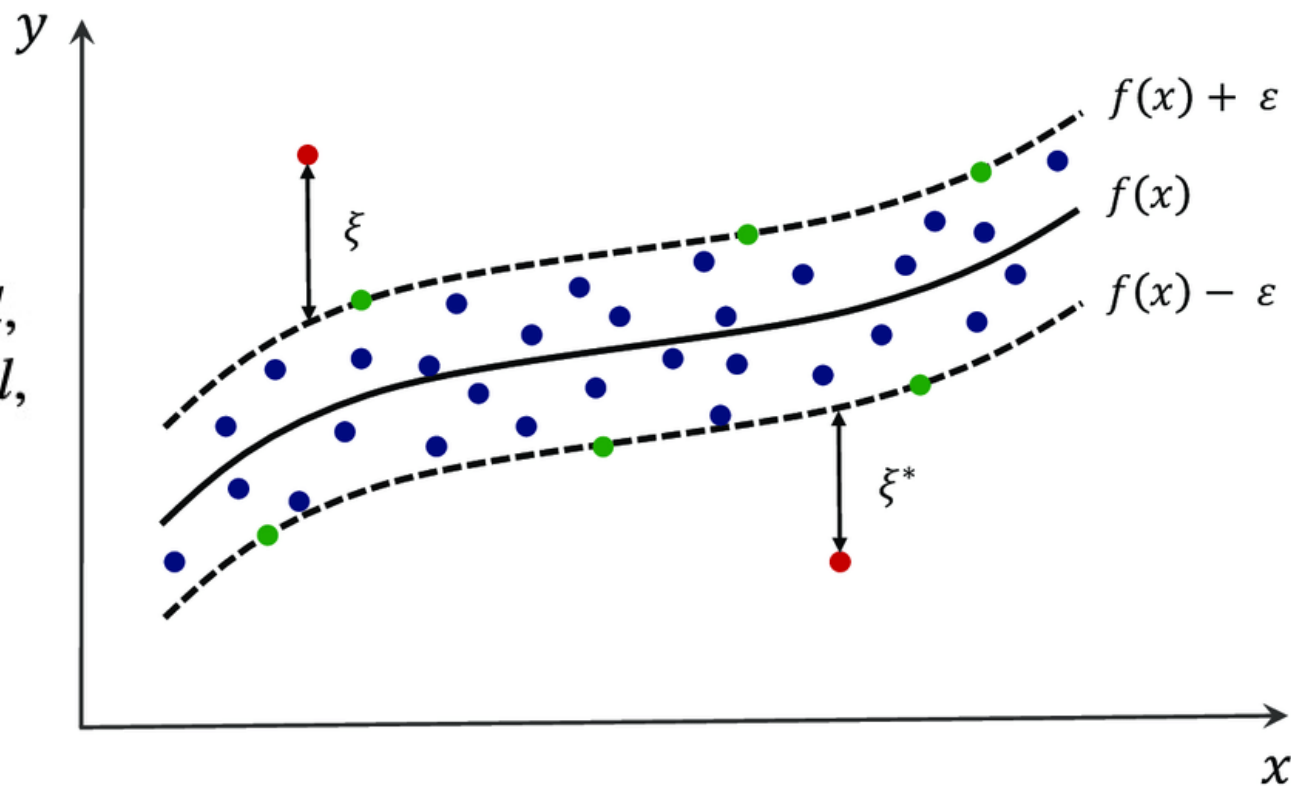


Serie estacionaria en media pero no en  
varianza

# 3. MARCO TEÓRICO

## 3.2 SVR (Support Vector Regression)

$$\begin{aligned} \min_{w, b, \xi^{(*)}} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*), \\ \text{s.t.} \quad & (w \cdot x_i) + b - y_i \leq \varepsilon + \xi_i, i = 1, \dots, l, \\ & y_i - (w \cdot x_i) - b \leq \varepsilon + \xi_i^*, i = 1, \dots, l, \\ & \xi_i^{(*)} \geq 0, i = 1, \dots, l, \end{aligned}$$

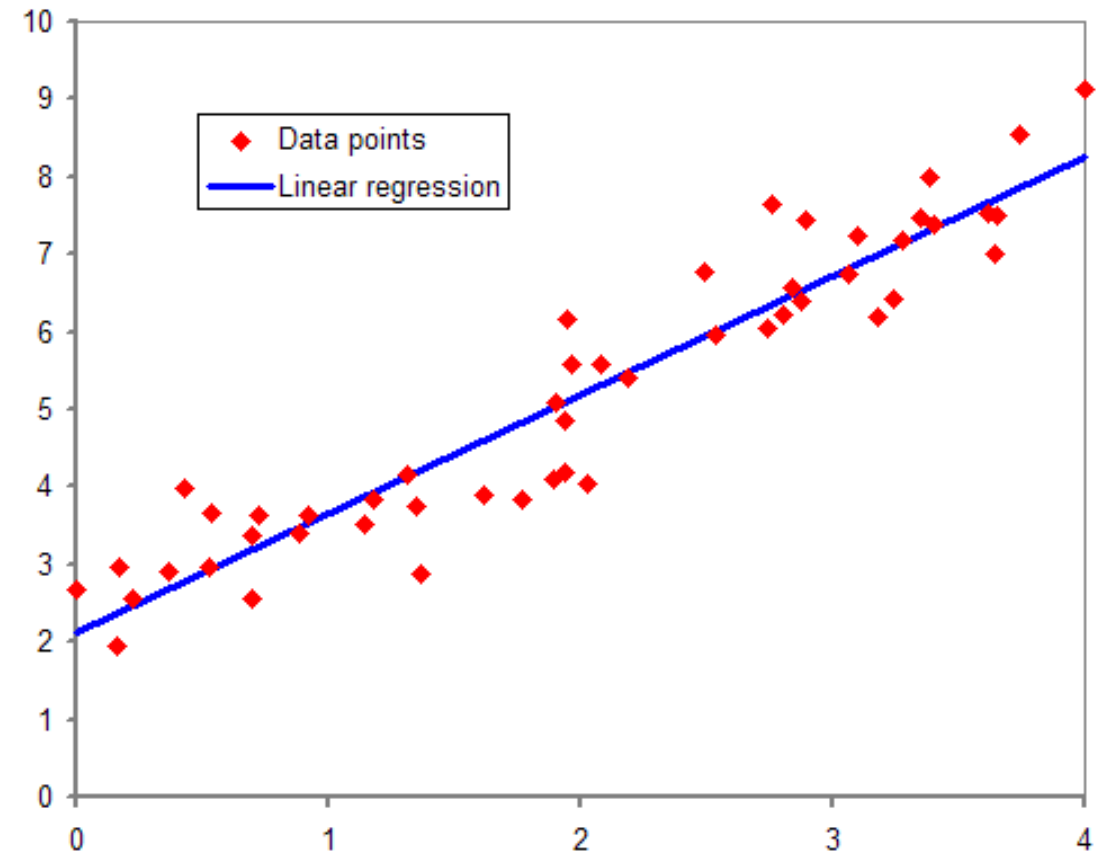


# 3. MARCO TEÓRICO

## 3.3 Regresión lineal

$$\hat{Y} = bX + a$$

predicted values of Y      slope = rate of increase/decrease of Y hat for each unit increase in X      Y-intercept = level of Y when X is 0.





# 3. MARCO TEÓRICO

## 3.3.1 Regresión lineal múltiple

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon$$

donde,

$i = n$  observaciones

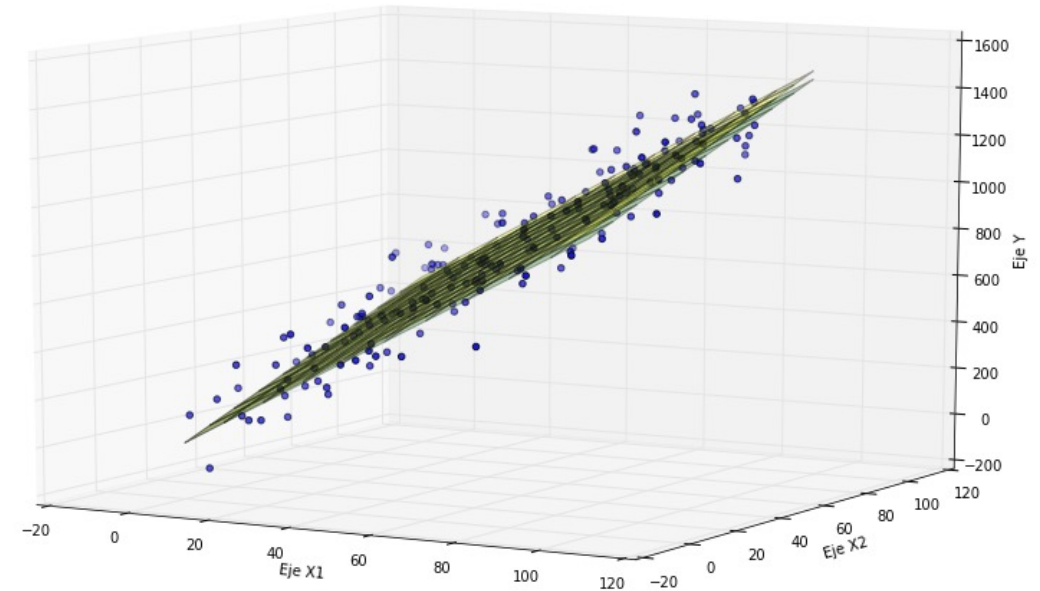
$y_i$  = variable dependiente

$x_i$  = variables explicativas

$\beta_0$  = intercepto

$\beta_p$  = coeficientes de pendiente para  $p$  variables explicativas

$\epsilon$  = termino de error del modelo.



# 3. MARCO TEÓRICO

## 3.3.1 Regresión lineal múltiple

- **Problema de mínimos cuadrados:**

**Función objetivo:**

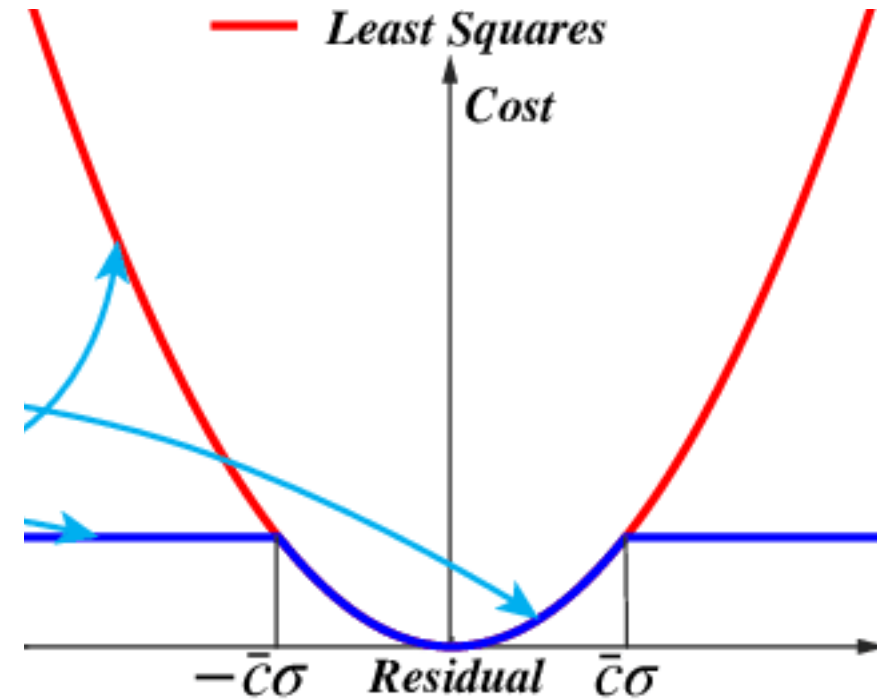
$$\min \|Ax - b\|^2$$

donde,

A es una matriz de variables explicativas, b de variables de salida y x un vector de pesos.



$$\nabla \min \|Ax - b\|^2 = 0$$



# 3. MARCO TEÓRICO

## 3.4 Transformador simbólico

- Genera distintos ajustes por medio de un algoritmo genético
- Se crea una población original aleatoria con la cual se iniciará un proceso de “evolución” para encontrar la mejor población que estará conformada por los individuos con el mejor ajuste y los coeficientes de correlación más cercanos a cero.
- Parámetros
  - Número de generaciones, porcentaje de mutaciones, número de individuos que sobreviven al cambio generacional, métrica, entre otros.
- Uno de los parámetros que se deben de tener en cuenta es el de métrica (*metric*), este parámetro se vuelve relevante en el resultado final ya que con él se define qué coeficiente de correlación se utilizará para corroborar los resultados. Específicamente en gplearn y en la función de *Symbolic Transformer* se cuenta con dos tipos de coeficiente de correlación: Pearson y Spearman.

# 3. MARCO TEÓRICO

## 3.5 MAPE

El MAPE es un indicador del desempeño de un pronóstico. Mide el tamaño del error absoluto en términos porcentuales y eso lo hace fácil de interpretar. La fórmula para calcularlo es la siguiente:

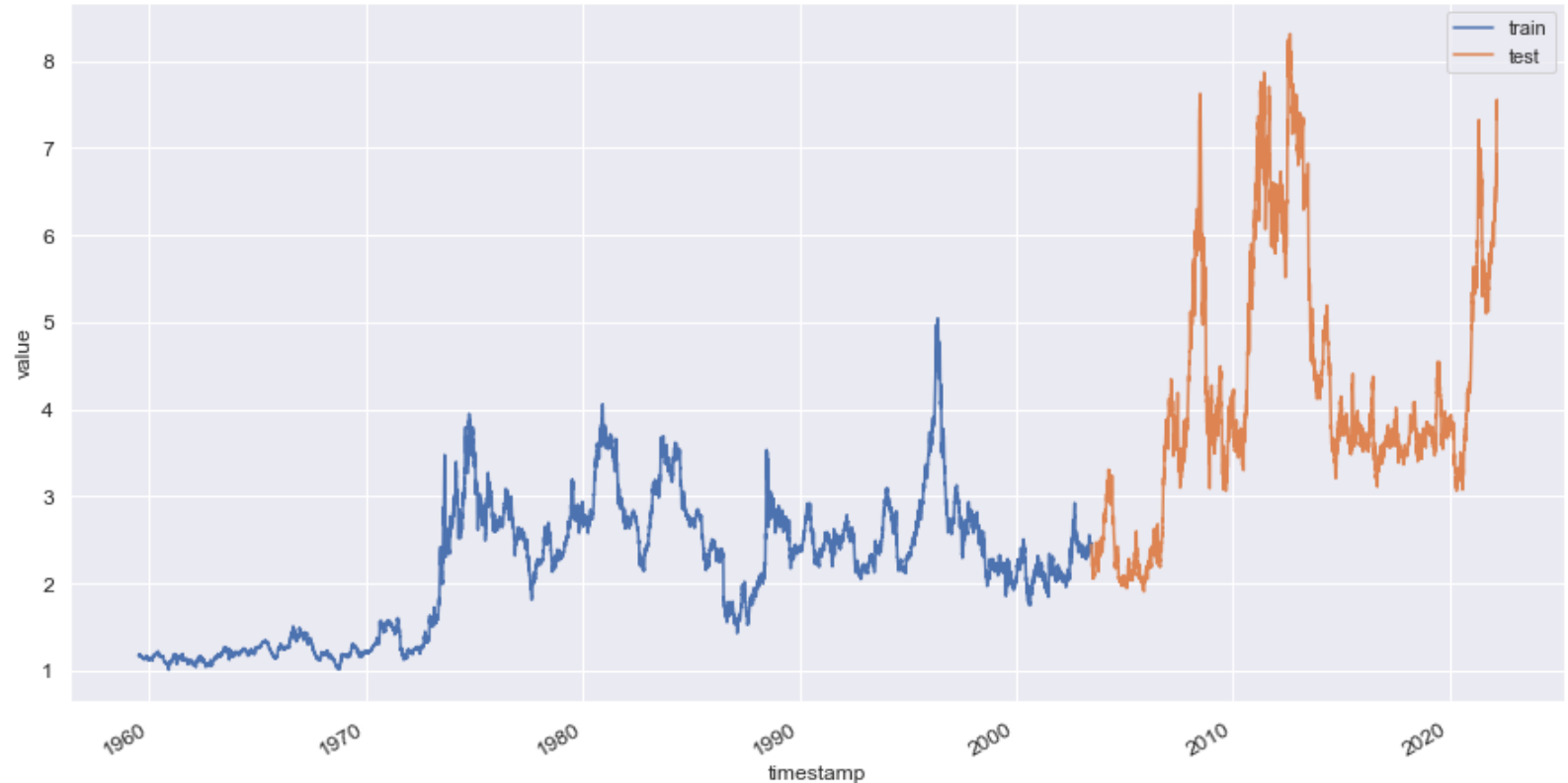
$$\text{MAPE} = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|$$

Donde  $A_t$  es el valor real y  $F_t$  es el valor de pronóstico.

# 4. DESARROLLO

## 4.1 División de la base de datos:

- Separación en dos conjuntos: entrenamiento y prueba.
- Variable explicativa: tiempo
- Variable de respuesta: precio
- Separación lineal de datos 70-30 (no aleatoria).
- Datos dependientes entre sí.
- Normalización *MinMax*.



**Fig. 1. División lineal (no aleatoria) de la base de datos en conjuntos de entrenamiento y prueba.**

# 4. DESARROLLO

## 4.2 SVR

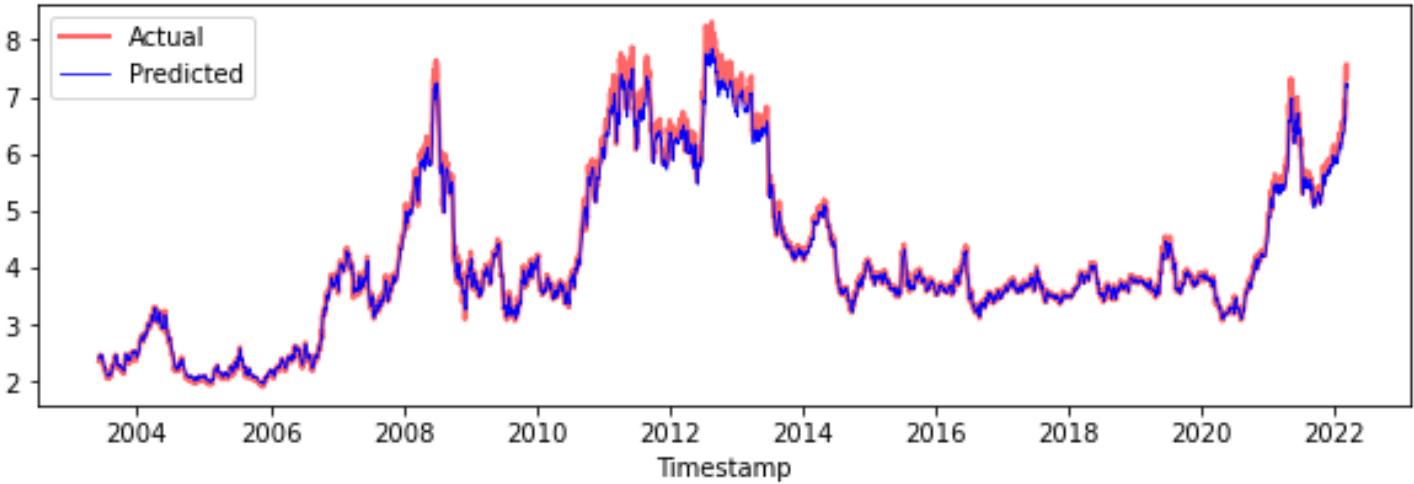
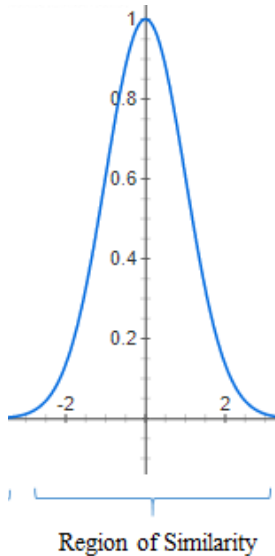
- Kernel: RBF (*Radial Basis Function*) ,  $\gamma=0.01$
- $C=10$
- $\epsilon = 0.05$

### Resultados:

- **MAPE= 1.973%.**

Tabla 1. Resultados de MAPE de 5 modelos SVR de kernel RBF variando gamma.

kernel	gamma	C	$\epsilon$	MAPE (%)
RBF	0.5	10	0.05	5.841
RBF	0.25	10	0.05	4.016
RBF	0.10	10	0.05	3.348
RBF	0.01	10	0.05	<b>1.973</b>
RBF	0.0005	10	0	3.146



# 4. DESARROLLO

## 4.2 Regresión lineal múltiple

- Serie de tiempo - una sola variable explicativa (tiempo)
- Se corrió transformador simbólico
- Se dividieron los datos en conjuntos de entrenamiento y prueba
- Se corrió el algoritmo de ajuste de regresión lineal.
- El mejor MAPE obtenido fue 1.59% y se logró con los siguientes parámetros:
  - generations=20
  - population\_size=2000
  - hall\_of\_fame=100
  - n\_components=10

# 5. CONCLUSIONES

- Importancia de predecir los precios del maíz.
- Resultados de 1.97% SVR
- Resultados de 1.59% Regresión lineal múltiple.
- Por qué es preferible el SVR.



# 6. REFERENCIAS

- Brownlee, J. (09 de 05 de 2022). *Machine Learning Mastery*. Obtenido de Linear Regression for Machine Learning: [https://machinelearningmastery.com/linear-regression-for-machine-learning/#:~:text=Linear%20regression%20is%20a%20linear,the%20input%20variables%20\(x\).](https://machinelearningmastery.com/linear-regression-for-machine-learning/#:~:text=Linear%20regression%20is%20a%20linear,the%20input%20variables%20(x).)
- gplearn. (2022). *Symbolic Transformer*. Obtenido de gplearn API reference: <https://gplearn.readthedocs.io/en/stable/reference.html#gplearn.genetic.SymbolicTransformer>
- Hayes, A. (09 de 05 de 2022). *Investiopedia*. Obtenido de Regresión Lineal Múltiple: [https://www.investopedia.com/terms/m/mlr.asp#:~:text=Key%20Takeaways-,Multiple%20linear%20regression%20\(MLR\)%2C%20also%20known%20simply%20as%20multiple,uses%20just%20one%20explanatory%20variable](https://www.investopedia.com/terms/m/mlr.asp#:~:text=Key%20Takeaways-,Multiple%20linear%20regression%20(MLR)%2C%20also%20known%20simply%20as%20multiple,uses%20just%20one%20explanatory%20variable).
- Minitab. (2022). *Relaciones lineales, no lineales y monótonas*. Obtenido de Support Minitab: <https://support.minitab.com/es-mx/minitab/19/help-and-how-to/statistics/basic-statistics/supporting-topics/basics/linear-nonlinear-and-monotonic-relationships/#:~:text=En%20una%20relaci%C3%B3n%20mon%C3%B3tona%2C%20las,direcci%C3%B3n%20a%20un%20ritmo%20con>
- Minitab. (2022). *Una comparación de los métodos de correlación de Pearson y Spearman-Minitab*. Obtenido de Support Minitab: <https://support.minitab.com/es-mx/minitab/18/help-and-how-to/statistics/basic-statistics/supporting-topics/correlation-and-covariance/a-comparison-of-the-pearson-and-spearman-correlation-methods/#:~:text=Los%20coeficientes%20de%20correlaci%C3%B3n%20de%20P>
- Ping-Feng, P., Kuo-Ping, L., & Chi-Shen, L. (2009). Pronóstico de series de tiempo mediante un modelo de regresión de vector de soporte estacional. *Elsevier Ltd.* , 4261–4265.
- Velázquez, J., Olaya, Y., & Franco, C. (2010). Predicción de series temporales usando máquinas de vectores de soporte. *Ingeniare*, 64-75.
- Villavicencio, J. (09 de 05 de 2022). *Introducción a Series de Tiempo*. Obtenido de Instituto de Estadísticas de Puerto Rico: [http://www.estadisticas.gobierno.pr/iepr/LinkClick.aspx?fileticket=4\\_BxecUaZmg%3D](http://www.estadisticas.gobierno.pr/iepr/LinkClick.aspx?fileticket=4_BxecUaZmg%3D)
- Xuchan, J., Manjin, C., Yuhong, X., Fuqiang, Q., & Yingjie, T. (2014). Regresión vectorial y análisis de series temporales para el Pronóstico del requerimiento total de agua de Bayannur. *Elsevier*, 523 – 531 .