

# Regresión Lineal Múltiple en Python

[Carlos Oswaldo Gonzalez Garza]

March 28, 2025

## 1 Introducción

La regresión lineal múltiple es una técnica estadística utilizada para modelar la relación entre una variable dependiente y dos o más variables independientes. En este trabajo, implementamos un modelo de regresión lineal múltiple en Python utilizando la biblioteca `sklearn`.

## 2 Metodología

Se siguieron los siguientes pasos para realizar la regresión lineal múltiple:

1. Carga y preprocesamiento de datos.
2. Definición de variables predictivas.
3. Entrenamiento del modelo de regresión lineal.
4. Evaluación del modelo.
5. Visualización en 3D.
6. Realización de predicciones con nuevos datos.

### 2.1 Código en Python

```
# Importamos librerías necesarias
import numpy as np
import pandas as pd
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D

# Definición de variables predictivas
suma = (filtered_data['# of Links'] +
        filtered_data['# of comments'].fillna(0) +
```

```

        filtered_data['# Images video'])

dataX2 = pd.DataFrame()
dataX2['Word count'] = filtered_data['Word count']
dataX2['suma'] = suma

XY_train = np.array(dataX2)
z_train = filtered_data['# Shares'].values

# Entrenamiento del modelo
regr2 = LinearRegression()
regr2.fit(XY_train, z_train)

# Coeficientes del modelo
print('Coefficients:', regr2.coef_)
print("Mean squared error: %.2f" % mean_squared_error(z_train, regr2.predict(XY_train)))
print("Variance score: %.2f" % r2_score(z_train, regr2.predict(XY_train)))

# Predicción para un nuevo artículo
z_Dosmil = regr2.predict([[2000, 10+4+6]])
print(int(z_Dosmil))

```

### 3 Resultados

El modelo de regresión lineal generó los siguientes coeficientes:

```

Coefficients: [ 6.63216324 -483.40753769]
Mean squared error: 352122816.48
Variance score: 0.11

```

Se obtuvo una predicción de aproximadamente 20518 shares para un artículo con 2000 palabras, 10 enlaces, 4 comentarios y 6 imágenes.

### 4 Conclusión

Se implementó con éxito un modelo de regresión lineal múltiple en Python, permitiendo estimar la cantidad de "Shares" basada en la cantidad de palabras y la suma de enlaces, comentarios e imágenes. A pesar de la mejora respecto al modelo de una sola variable, el puntaje de varianza sigue siendo bajo, indicando que el modelo podría beneficiarse de métodos adicionales, como la reducción de dimensiones o la inclusión de nuevas variables.