# Regresión Lineal Múltiple en Python

## 1 Introducción

La regresión lineal múltiple es un método estadístico que modela la relación entre una variable dependiente y múltiples variables independientes. Se expresa con la ecuación:

$$Y = b + m_1X_1 + m_2X_2 + ... + m_nX_n \tag{1}$$

Esta técnica permite mejorar la precisión de las predicciones al considerar múltiples factores en el análisis de datos.

## 2 Metodología

Los pasos seguidos en la implementación fueron:

1. Cargar los datos desde un archivo CSV.
2. Crear nuevas características combinando enlaces, comentarios e imágenes.
3. Separar las variables predictivas de la variable objetivo.
4. Entrenar un modelo de regresión lineal múltiple utilizando `scikit-learn`.
5. Evaluar el modelo mediante métricas como el error cuadrático medio y el puntaje de varianza.
6. Visualizar los resultados en un gráfico 3D.

### 2.1 Código Implementado

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn import linear_model
from sklearn.metrics import mean_squared_error, r2_score

df = pd.read_csv("archivo.csv")

suma = (df["# of Links"] + df['# of comments'].fillna(0) + df['# Images video'])
```

```
dataX2 = pd.DataFrame()
dataX2["Word count"] = df["Word count"]
dataX2["suma"] = suma

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

regr2 = linear_model.LinearRegression()
regr2.fit(XY_train, z_train)

z_pred = regr2.predict(XY_train)

print('Coeficientes:', regr2.coef_)
print("Error cuadrático medio:", mean_squared_error(z_train, z_pred))
print('Puntaje de varianza:', r2_score(z_train, z_pred))

fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')
ax.scatter(XY_train[:, 0], XY_train[:, 1], z_train, c='blue', marker='o', label="Datos reale
ax.scatter(XY_train[:, 0], XY_train[:, 1], z_pred, c='red', marker='^', label="Predicciones'
ax.set_xlabel('Cantidad de Palabras')
ax.set_ylabel('Suma de Links, Comentarios, Imágenes')
ax.set_zlabel('Compartido en Redes')
ax.set_title('Regresión Lineal Múltiple')
plt.legend()
plt.show()
```

# 3 Resultados

Los coeficientes obtenidos indican la influencia de cada variable en la cantidad de veces que un artículo es compartido. Se observaron los siguientes valores:

- Coeficientes del modelo: Indican el impacto de cada variable independiente.
- Error cuadrático medio: Mide la precisión del modelo.
- Puntaje de varianza: Evalúa qué tan bien se ajusta el modelo a los datos.

# 4 Conclusión

La regresión lineal múltiple permitió mejorar la predicción de compartidos en redes sociales en comparación con una regresión simple. Sin embargo, los resultados pueden mejorarse con técnicas avanzadas como selección de características y reducción de dimensionalidad.