

Prediccion de la facturacion de la energia electrica en electro puno usando regresion lineal y python

Jose Angel Condori Ccapa, *angel.cc@upeu.edu.pe*

26 de junio de 2023

1. Resumen

En este artículo, se presenta un estudio sobre la predicción de la facturación de la energía eléctrica en el departamento de Puno, utilizando técnicas de regresión lineal y programación en Python. La energía eléctrica es un recurso vital en la región de Puno, y es importante poder predecir la facturación para fines de planificación y gestión eficiente.

En primer lugar, se aplicó la regresión lineal simple para modelar la relación entre el consumo de energía y la facturación. Se encontró una relación lineal positiva entre estas variables, lo que indica que a medida que aumenta el consumo, también lo hace la facturación. Se calcularon los coeficientes de la ecuación de regresión, incluyendo el intercepto y la pendiente.

Posteriormente, se llevó a cabo una regresión lineal múltiple para incorporar una segunda variable independiente: el ubigeo de la persona. Se encontró que el consumo y el ubigeo son factores significativos en la predicción de la facturación. Se calcularon los coeficientes correspondientes a cada variable independiente en la ecuación de regresión.

Se implementó el análisis utilizando Python y diversas bibliotecas, como numpy, pandas y matplotlib. Los datos fueron cargados, limpiados y procesados para su análisis. Se generaron gráficos de dispersión y se trazaron las líneas de regresión para visualizar la relación entre las variables.

palabras clave: Ciencia de datos, Consumo de energia electrica, Regresion lineal simple y Reresion lineal multiple

2. Introduccion

La predicción precisa de la facturación de la energía eléctrica es esencial para las empresas de distribución eléctrica, ya que les permite planificar y gestionar de manera eficiente la producción y distribución de energía, optimizar los recursos y mejorar la rentabilidad. En este contexto, el uso de técnicas de análisis de datos y modelado se ha vuelto fundamental para realizar pronósticos precisos y confiables.

En este estudio, se propone la aplicación de la regresión lineal como una herramienta para predecir los niveles de facturación de la energía eléctrica en la empresa Electro Puno. La regresión lineal es un método estadístico que busca establecer una relación lineal entre una variable dependiente (la facturación eléctrica en este caso) y una o más variables independientes (como el consumo de energía, la temperatura, la estacionalidad, entre otros)

El objetivo principal de este estudio es evaluar la eficacia de la regresión lineal en la predicción de la facturación de la energía eléctrica en Electro Puno, así como identificar las variables más influyentes en dicho proceso. Los resultados obtenidos en este análisis podrían proporcionar a Electro Puno información valiosa para la toma de decisiones relacionadas con políticas tarifarias, eficiencia energética y estrategias de demanda.

3. Marco Teorico

3.1. Electro puno

Electro Puno es una empresa estatal peruana que se dedica a la generación, transmisión y distribución de energía eléctrica en la región de Puno, ubicada en el sur del Perú. Es una de las empresas eléctricas del país y forma parte del Grupo Electro Perú, junto con otras empresas estatales del sector eléctrico.

Electro Puno se encarga de operar y mantener las instalaciones eléctricas en su área de concesión, que abarca la región de Puno y parte de la región de Madre de Dios. Su objetivo principal es garantizar el suministro de energía eléctrica a los usuarios de la zona, tanto residenciales como comerciales e industriales.

3.2. Regresion lineal simple

La regresión lineal simple es un método estadístico utilizado para modelar la relación entre una variable dependiente (variable a predecir) y una única variable independiente (variable predictora). Se basa en la suposición de que existe una relación lineal entre las dos variables.

En la regresión lineal simple, se busca encontrar la línea recta que mejor se ajusta a los puntos de datos en un gráfico de dispersión. Esta línea recta

se conoce como "línea de regresión" se representa mediante una ecuación de la forma:

$$y = \beta_0 + \beta_1 x.$$

- y = variable dependiente o variable a predecir
- β_0 = El intercepto
- β_1 = La pendiente
- x = variable independiente

3.3. Regresión lineal múltiple

La regresión lineal múltiple es una extensión de la regresión lineal simple que permite modelar la relación entre una variable dependiente y múltiples variables independientes. En lugar de utilizar una única variable predictora como en la regresión lineal simple, la regresión lineal múltiple utiliza varias variables independientes para predecir la variable dependiente.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

- y = variable dependiente o variable a predecir
- β_0 = El intercepto
- β_1, \dots, β_n = La pendiente
- x_1, \dots, x_n = variable independiente

4. Metodología

los datos mostrados en este estudio están en el siguiente enlace: Electro Puno
Diccionario de datos

Diccionario de Datos				
Dataset	Consumo de energía eléctrica de los clientes de Electro Puno S.A.A. - [Electro Puno S.A.A.]			
Variable	Tipo de Dato	Descripción	Tamaño	Información Adicional
CÓDIGO_CLIENTE	Númérico	Código de suministro del usuario anonimizado	6	Número Entero
UBIGEO	Alfanumérico	Código de ubicación geográfica según INEI	6	Códigos del Catálogo de UBIGEO del INEI
DEPARTAMENTO	Alfanumérico	Departamento donde se ubica el usuario	200	En Mayúsculas
PROVINCIA	Alfanumérico	Provincia donde se ubica el usuario	200	En Mayúsculas
DISTRITO	Alfanumérico	Distrito donde se ubica el usuario	200	En Mayúsculas
FECHA_ALTA	Datetime	Fecha de alta o fecha de instalación del suministro	10	Formato: ddmmaaaa
TARIFA	Alfanumérico	Tarifa elegida por el cliente, entre: MT2,MT3,MT4, opciones de media tensión. BT2,BT3,BT4,BT5 y BT6, opciones de baja tensión	4	Entre Letras y números, en Mayúsculas.
PERIODO	Númérico	Año y mes facturado, AAAA-MM.	6	Formato: aaaamm
CONSUMO	Númérico	Consumo de energía eléctrica en Kwh	9	Número real con dos decimales, en Kwh
FACTURACIÓN	Moneda	Consumo de energía eléctrica monetizado. Expresado en soles (S./.)	9	Número real con dos decimales, en soles.
ESTADO_CLIENTE	Alfanumérico	Estado del servicio suministro eléctrico entre Normal y Anulado.	7	En Mayúsculas
FECHA_CORTE	Datetime	Fecha en que se generó la data.	10	Formato: ddmmaaaa

Figura 1: Diccionario de datos

Tabla de datos

CÓDIGO_CLIEN	UBIGEO	DEPARTAMENT	PROVINCIA	DISTRITO	FECHA_ALTA	TARIFA	PERIODO	CONSUMO	FACTURACIÓN	ESTADO_CLIEN	FECHA_CORTE
1	210701	PUNO	LAMPA	LAMPA	02/05/2011	BT5B	202301	0	7.9	NORMAL	07/02/2023
2	210701	PUNO	LAMPA	LAMPA	11/01/2008	BT5B	202301	0	7.7	NORMAL	07/02/2023
3	210701	PUNO	LAMPA	LAMPA	24/08/2021	BT5B	202301	0	7.6	NORMAL	07/02/2023
4	211101	PUNO	SAN ROMAN	JULIACA	23/02/2018	BT5B	202301	6	10.1	NORMAL	07/02/2023
5	210701	PUNO	LAMPA	LAMPA	11/01/2008	BT5B	202301	0	7.8	NORMAL	07/02/2023
6	210701	PUNO	LAMPA	LAMPA	11/01/2008	BT5B	202301	27	18.6	NORMAL	07/02/2023
7	210701	PUNO	LAMPA	LAMPA	11/01/2008	BT5B	202301	0	7.6	NORMAL	07/02/2023
8	210701	PUNO	LAMPA	LAMPA	16/07/2020	BT5B	202301	12	12.6	NORMAL	07/02/2023
9	210701	PUNO	LAMPA	LAMPA	11/01/2008	BT5B	202301	22	16.6	NORMAL	07/02/2023
10	210701	PUNO	LAMPA	LAMPA	11/01/2008	BT5B	202301	19	15.4	NORMAL	07/02/2023
11	210701	PUNO	LAMPA	LAMPA	11/01/2008	BT5B	202301	1	8.1	NORMAL	07/02/2023
12	210701	PUNO	LAMPA	LAMPA	11/01/2008	BT5B	202301	7	10.7	NORMAL	07/02/2023
13	210701	PUNO	LAMPA	LAMPA	11/01/2008	BT5B	202301	17	14.6	NORMAL	07/02/2023
14	210701	PUNO	LAMPA	LAMPA	11/01/2008	BT5B	202301	42	37.5	NORMAL	07/02/2023
15	210701	PUNO	LAMPA	LAMPA	11/01/2008	BT5B	202301	0	7.7	NORMAL	07/02/2023
16	210701	PUNO	LAMPA	LAMPA	11/01/2008	BT5B	202301	16	14.1	NORMAL	07/02/2023
17	210701	PUNO	LAMPA	LAMPA	11/01/2008	BT5B	202301	3	9.1	NORMAL	07/02/2023
18	210701	PUNO	LAMPA	LAMPA	11/01/2008	BT5B	202301	0	8.9	NORMAL	07/02/2023
19	210701	PUNO	LAMPA	LAMPA	11/01/2008	BT5B	202301	0	7.7	NORMAL	07/02/2023
20	210701	PUNO	LAMPA	LAMPA	04/09/2021	BT5B	202301	580	681.1	NORMAL	07/02/2023
21	210701	PUNO	LAMPA	LAMPA	11/01/2008	BT5B	202301	3	8.9	NORMAL	07/02/2023
22	210701	PUNO	LAMPA	LAMPA	11/01/2008	BT5B	202301	0	7.6	NORMAL	07/02/2023
23	210701	PUNO	LAMPA	LAMPA	11/01/2008	BT5B	202301	1	8.2	NORMAL	07/02/2023
24	210701	PUNO	LAMPA	LAMPA	11/01/2008	BT5B	202301	0	8.3	NORMAL	07/02/2023
25	210701	PUNO	LAMPA	LAMPA	11/01/2008	BT5B	202301	0	7.7	NORMAL	07/02/2023

Figura 2: Muestra de datos

4.1. Regresión lineal simple

La regresión lineal simple es un modelo estadístico usado para relacionar una variable independiente(X) con una variable dependiente(y). Es decir, en una regresión lineal simple solo hay dos variables (la variable explicativa X y la variable respuesta Y) y se intenta aproximar la relación que hay entre ambas variables.

Nuestra variable independiente(X) es consumo

Nuestra variable dependiente(y) es facturacion

La ecuacion de la regresión lineal simple es:

$$y = \beta_0 + \beta_1 x.$$

- y = variable dependiente o variable a predecir
- β_0 = El intercepto
- β_1 = La pendiente

El intercepto (β_0) se puede calcular utilizando la ecuacion:

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

La pendiente (β_1) se puede calcular utilizando la fórmula:

$$\beta_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

4.1.1. Usando Python

Como estamos usando un rango de datos de 640212 para ayudarnos en el trabajo usaremos python

- importamos las librerias que usaremos

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from scipy.stats import linregress
```

- cargamos los datos y lo concatenamos

```
pd.options.display.float_format = '{:.2f}'.format
enero = pd.read_csv('Enero2023.csv', encoding="ISO-8859-1", sep=";")
febrero = pd.read_csv('febrero2023.csv', encoding="ISO-8859-1", sep=";")
marzo = pd.read_csv('Marzo2023.csv', encoding="ISO-8859-1", sep=";")
abril = pd.read_csv('Abril2023.csv', encoding="ISO-8859-1", sep=";")

datos = pd.concat([enero, febrero, marzo, abril])
```

- hacemos limpieza de datos y elegimos la variable dependiente e independiente

```
datos = datos.fillna(0)
x = datos['CONSUMO']
y = datos['FACTURACION']
```

- la funcion `linregress()` nos permite calcular la pendiente
- la funcion `linregress.intercept` nos permite calcular el intercepto
- despues `x` genera un arreglo de valores en la columna consumo

```
#parametros de la recta
b1 = linregress(x, y).slope #pendiente
b0 = linregress(x, y).intercept #intercepto

x = np.linspace(0, datos['CONSUMO'].max())
```

- `valor_x` es el consumo
- usamos la ecuacion y creamos el grafico
- hacemos una predicción usando la ecuacion

```
valor_x = 1000
prediccion_y = b1 * valor_x + b0

datos.plot.scatter(x='CONSUMO', y='FACTURACION')
plt.plot(x, y, '-r')
plt.ylim(0, datos['FACTURACION'].max() * 1.1)
plt.show()
```

4.2. Regresion Lineal Multiple

La regresión lineal múltiple es un modelo de regresión en cual se incluyen dos o más variables independientes. Es decir, la regresión lineal múltiple es un modelo estadístico que permite relacionar varias variables explicativas con una variable respuesta de manera lineal.

usaremos la siguiente ecuacion para hallar la facturacion dependiendo al consumo y el ubigeo de la persona

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

Son 2 variables independientes la ecuacion queda asi

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

para hallar

$$\beta_0, \beta_1, \beta_2$$

$$\left(\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \begin{bmatrix} n & \sum x_1 & \sum x_2 \\ \sum x_1 & \sum x_1^2 & \sum x_1 x_2 \\ \sum x_2 & \sum x_1 x_2 & \sum x_2^2 \end{bmatrix} \begin{bmatrix} \sum y \\ \sum x_1 y \\ \sum x_2 y \end{bmatrix} \right)$$

reemplazaremos la matriz

$$\left(\begin{bmatrix} R1 \\ R2 \\ R3 \end{bmatrix} \quad \begin{bmatrix} a & d & c \\ d & b & d \\ c & d & c \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right)$$

rempezamos convirtiendo la primera columna en 1, 0, 0

$$\left(\begin{bmatrix} R1/a \\ -d(R1) + R2 \\ -c(R1) + R3 \end{bmatrix} \quad \begin{bmatrix} 1 & e & f \\ 0 & h & i \\ 0 & k & l \end{bmatrix} \quad \begin{bmatrix} g & 0 & 0 \\ j & 1 & 0 \\ m & 0 & 1 \end{bmatrix} \right)$$

convertimos la segunda columna en 0, 1, 0

$$\left(\begin{bmatrix} -e(R2) + R1 \\ R2/h \\ -k(R2) + R3 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & n \\ 0 & 1 & x \\ 0 & 1 & B \end{bmatrix} \quad \begin{bmatrix} o & p & q \\ y & Z & A \\ C & D & E \end{bmatrix} \right)$$

convertimos la tercera columna en 0, 0, 1

$$\left(\begin{bmatrix} -h(R3) + R1 \\ -x(R3) - R2 \\ R3/B \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} F & H & I \\ J & L & M \\ N & O & P \end{bmatrix} \right)$$

Multiplicamos las matrices

$$\left(\begin{bmatrix} F & H & I \\ J & L & M \\ N & O & P \end{bmatrix} \quad \begin{bmatrix} \sum y \\ \sum x_1 y \\ \sum x_2 y \end{bmatrix} \right)$$

$$\left(\begin{bmatrix} \sum y * F & \sum y * H & \sum y * I \\ \sum x_1 y * J & \sum x_1 y * L & \sum x_1 y * M \\ \sum x_2 y * N & \sum x_2 y * O & \sum x_2 y * P \end{bmatrix} \right)$$

resultado

$$\left(\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \quad \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \right)$$

4.2.1. Error estandar de la estimacion multiple

Es el error tipico cuando se emplea la ecuacion de regresion multiple para predecir la facturacion

$$\sqrt{\frac{SCE}{n - (k + 1)}}$$

- n = es la cantidad de la poblacion o Muestra
- k = es el numero de variables independientes

- SCE = suma de cuadrados del error o residuo
- para hallar SCE usamos la siguiente ecuacion

$$SCE = \sum_i (y_i - \hat{y}_i)^2$$

4.2.2. Coeficiente de determinacion multiple

es el porcentaje de variacion de la variable dependiente(Y) explicada por el conjunto de variables independientes(X1,X2)

$$R^2 = \frac{SCR}{STC}$$

- SCR = suma de cuadrados del error o residuo

$$SCR = \sum (\hat{y} - \bar{y})^2$$

- STC = suma total de cuadrados

$$STC = SCE + SCR$$

4.2.3. Coeficiente ajustado de determinacion multiple

El coeficiente de multiple de determinacion R^2 modificado para justificar el numero de variables y el tamaño de la muestra

$$R^2_{ajustada} = 1 - (1 - R^2) \frac{n - 1}{n - k - 1}$$

Despues de hacer la metodologia y haber usado las ecuaciones usaremos python para hallar rapidamente

cargamos las librerias que usaremos

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.api as sm

importamos los datos y los concatenamos
hacemos limpieza de datos

enero = pd.read_csv('Enero2023.csv', encoding="ISO-8859-1", sep=";")
febrero = pd.read_csv('febrero2023.csv', encoding="ISO-8859-1", sep=";")
marzo = pd.read_csv('Marzo2023.csv', encoding="ISO-8859-1", sep=";")
abril = pd.read_csv('Abril2023.csv', encoding="ISO-8859-1", sep=";")

datos = pd.concat([enero, febrero, marzo, abril])
datos = datos.fillna(0)
```


Con el metodo OLS(Ordinary Least Squares) hallamos los valores de

$$\beta_0, \beta_1, \beta_2$$

obtenemos los valores

```
model = sm.OLS(y, X)
results = model.fit()

# Obtener los coeficientes
intercept = results.params[0]
B1 = results.params[1]
B2 = results.params[2]

prediciremos la facturacion con lo siguiente

x1_new = 210306 # ubigeo
x2_new = 1000 # consumo

# calculamos
y= intercept + B1*x1_new + B2 * x2_new

Error estandar de la estimacion multiple

longitudD = len(datos)
residuals = results.resid
SCE = np.sum(residuals**2)

eeem = np.sqrt(SCE/(longitudD-(2+1)))
print('Error estandar de la determinacion multiple',eeem)

Coeficiente de determinacion multiple

promedio_facturacion = datos['FACTURACION'].mean()
y_pred = results.predict(X)

SCE = np.sum(residuals**2)
y_pred_resta_promedio = y_pred - promedio_facturacion
SCR = (y_pred_resta_promedio ** 2).sum()

STC= SCR + SCE
CDM= SCR/STC
print('Coeficiente de determinacion multiple',CDM)

Coeficiente ajustado de determinacion multiple

coeficiente_adj_r2 = results.rsquared_adj
ccm= np.sqrt(coeficiente_adj_r2)
```

5. Resultados

5.1. Regresion lineal

Despues de hacer la metodologia y haber usado las formulas usando python podemos ver los resultados

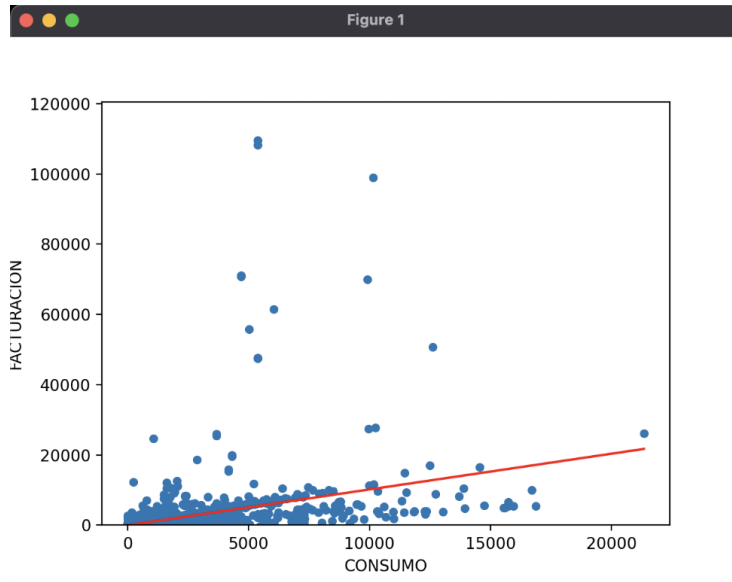


Figura 3: Grafico de dispercion

- en la figura 1 se puede ver el grafico de dispercion y la linea

usando la formula que se ejecuto en python podemos ver el resultado en este caso prediciremos la facturacion cuando el consumo sea en 1000

```
valor_x = 1000 # Valor de x para la prediccion  
prediccion_y = b1 * valor_x + b0  
print('Prediccion:', prediccion_y)
```

ejecutando la aplicacion valor_x es el valor del consumo el resultado es.

Predicción: 1016.309361317152

Una persona que consume 1000 kwh estaria pagando 1016.30 soles

5.2. Regresion Multiple

en el departamento de puno provincia de carabaya distrito crucero mientras el consumo sea de 1000 llegara de facturacion 1016.27 soles

5.2.1. Error estandar de la determinacion multiple

el error estandar es de 314.0039251060518 de facturacion

5.2.2. Coeficiente de determinacion multiple

indica que el 27.60 de la variacion de la facturacion puede explicarse por el consumo y el ubigeo. En otras palabras el 72.4 se debe a otras fuentes como, como el error aleatorio o variables no incluidas en el analisis

5.2.3. Coeficiente ajusta de determinacion multiple

el 52.54 de la variacion de la facturacion puede explicarse por el consumo y el ubigeo. En otras palabras el 47.36 se debe a otras fuentes como, como el error aleatorio o variables no incluidas en el analisis

5.2.4. Coeficiente de correlacion multiple

Esta correlacion R de 0.52 es una correlacion moderada

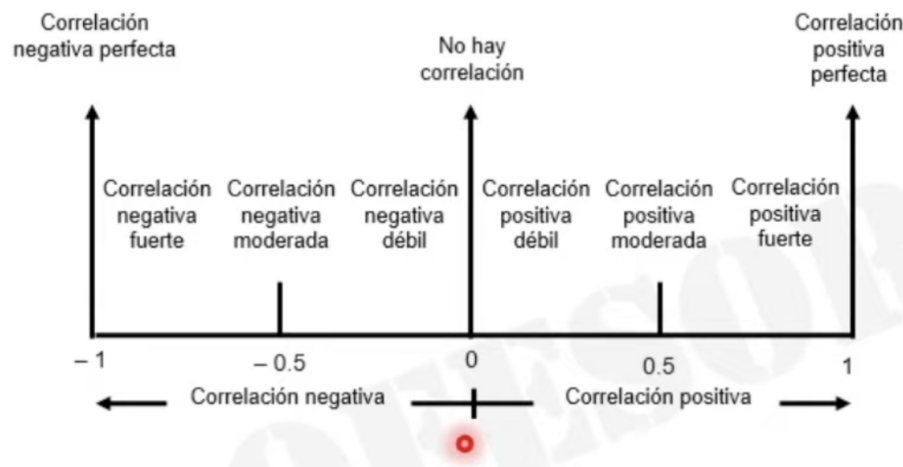


Figura 4: Tabla

6. Conclusiones

Las predicciones que nos da como resultado en regresion lineal y multiple son parecidos para que la prediccion sea mas exacta tenemos que usar mas datos en regresion multiple para tener predicciones mas exactas

La regresión lineal aplicada al análisis de la facturación de la energía eléctrica en Electro Puno demostró ser una herramienta efectiva para predecir los niveles de facturación futuros.

El modelo de regresión lineal desarrollado en Python mostró un buen ajuste a los datos históricos de facturación, lo que indica que puede utilizarse como una herramienta confiable para la predicción de la facturación eléctrica.

Los resultados obtenidos en el estudio pueden ser utilizados por Electro Puno para planificar y gestionar de manera más eficiente la producción y distribución de energía eléctrica, optimizando los recursos y mejorando la rentabilidad.

La implementación del modelo de regresión lineal en Python proporcionó una solución práctica y accesible para la predicción de la facturación eléctrica, lo que demuestra el potencial de la programación y el análisis de datos en el campo de la gestión de la energía.

7. Referencias

Concentraci3n horizontal en un ambiente regulado. El caso de la distribuci3n de electricidad en el gran buenos aires — Revista de la Competencia y la Propiedad Intelectual. (n.d.). Retrieved June 18, 2023, from <https://revistas.indecopi.gob.pe/index.php/rcpi/article/view/100>

Cutipa Huarsaya, M. W. (2016). Los estados financieros y su influencia en la toma de decisiones de la Empresa Regional de Servicio P3blico de Electricidad - Electro Puno S. A. A. periodos 2014 - 2015. Uancv [Http://Repositorio.Uancv.Edu.Pe/](http://Repositorio.Uancv.Edu.Pe/). <https://renati.sunedu.gob.pe/handle/sunedu/2895750>

Cutipa Ticona, A. C. (2016). Incidencia de la morosidad en la cartera de clientes de Electro Puno S.A.A. y su efecto en la liquidez y rentabilidad en el 2014-2015. Repositorio Institucional - UNAP. <https://renati.sunedu.gob.pe/handle/sunedu/3221488>
Granados , R. M. (n.d.). Modelos de regresi3n lineal m3ltiple.

Irene Moral Pel3ez. (n.d.).

Registro Nacional de Trabajos de Investigaci3n: Incidencia de la morosidad en la cartera de clientes de Electro Puno S.A.A. y su efecto en la liquidez y rentabilidad en el 2014-2015. (n.d.). Retrieved June 18, 2023, from <https://renati.sunedu.gob.pe/handle/sunedu/3221488>

Registro Nacional de Trabajos de Investigaci3n: Los estados financieros y su influencia en la toma de decisiones de la Empresa Regional de Servicio P3blico de Electricidad - Electro Puno S. A. A. periodos 2014 - 2015. (n.d.). Retrieved June 18, 2023, from <https://renati.sunedu.gob.pe/handle/sunedu/2895750>