

Identificación de patrones de consumo mediante aprendizaje autónomo localizando posibles focos de pérdidas no técnicas

PROBLEMA

La capacidad de identificar adecuadamente clientes que presentan consumos anómalos en la ciudad de Guayaquil es inexistente, adicionalmente el único método actual es el levantamiento de información por parte de cuadrillas, lo cual se resume en gastos económicos, así como levantamiento de información incompleta e imprecisa.

OBJETIVO GENERAL

Entrenar diferentes modelos de aprendizaje autónomo para detección de posibles pérdidas no técnicas (comerciales) de energía, en una región específica de la ciudad de Guayaquil.

PROPUESTA

Para mejorar el resultado de este proceso poco preciso, se ha codificado un programa dentro del lenguaje Python, en específico, se implementaron tres modelos de aprendizaje autónomo: árboles de decisión, KNN y regresión logística, capaces de aprender y ser entrenados gracias a datos proporcionados directamente por las respectivas empresas distribuidora, y no por medio de información levantada por cuadrillas.

Para lo cual, una vez entrenado y testeado, el mejor modelo será capaz de identificar a los diferentes clientes asociados a pérdidas comerciales con un alto grado de efectividad.

```
Projecto Ximena Last Checkpoint: 07/08/2021 (unsaved changes)
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3
In [273]: import pandas as pd
import numpy as np
from numpy import mean
import matplotlib.pyplot as plt
%matplotlib inline

from sklearn.linear_model import SGDClassifier
from sklearn.preprocessing import StandardScaler
from sklearn.pipeline import make_pipeline

from sklearn.model_selection import train_test_split
from sklearn.model_selection import learning_curve

from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix, accuracy_score
from sklearn.metrics import plot_confusion_matrix

from sklearn.model_selection import cross_val_score
from sklearn.model_selection import RepeatedStratifiedFold
from imblearn.ensemble import BalancedRandomForestClassifier
from sklearn.linear_model import LogisticRegression
```

RESULTADOS

Se obtiene una matriz por cada muestra de prueba llena de unos y ceros, donde “1” representa a usuarios detectados con pérdidas comerciales y “0” los que no representan.

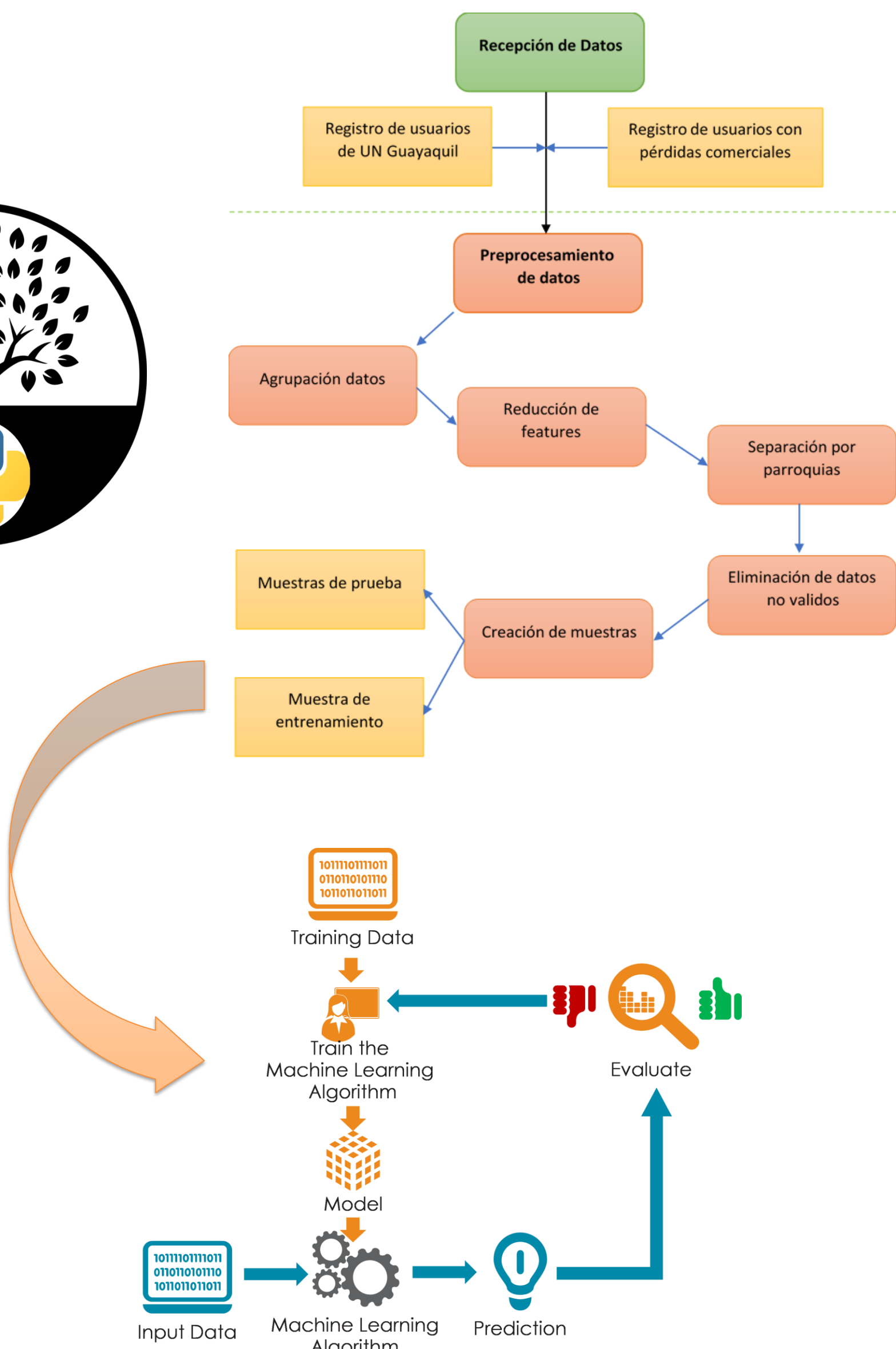
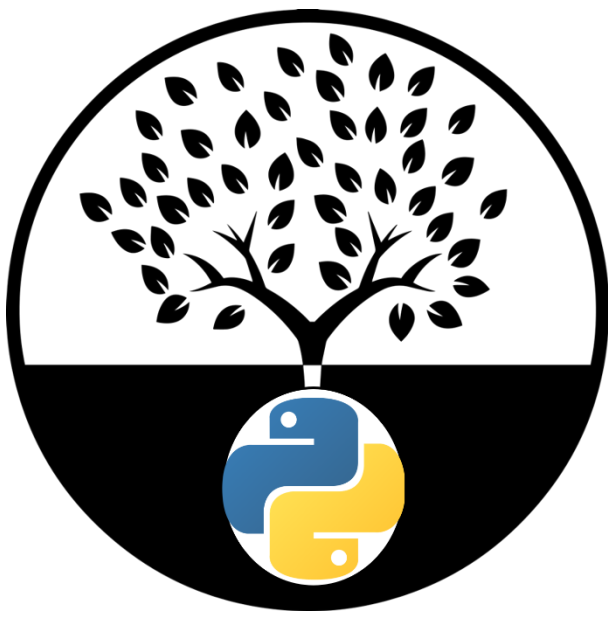
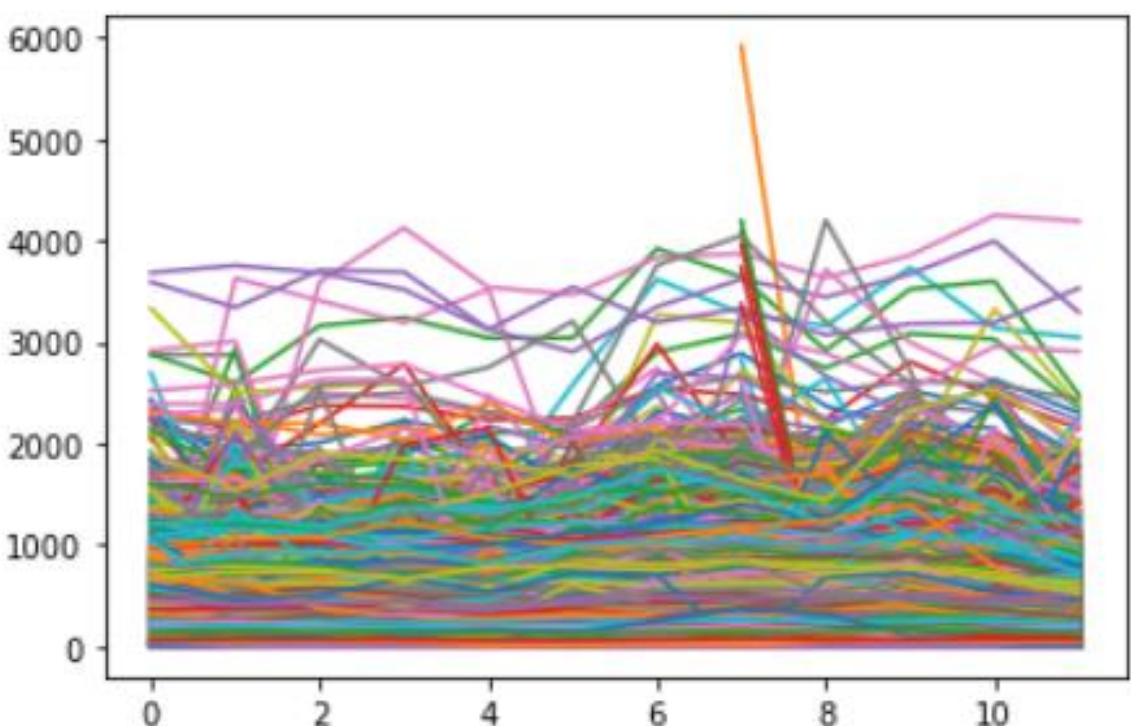
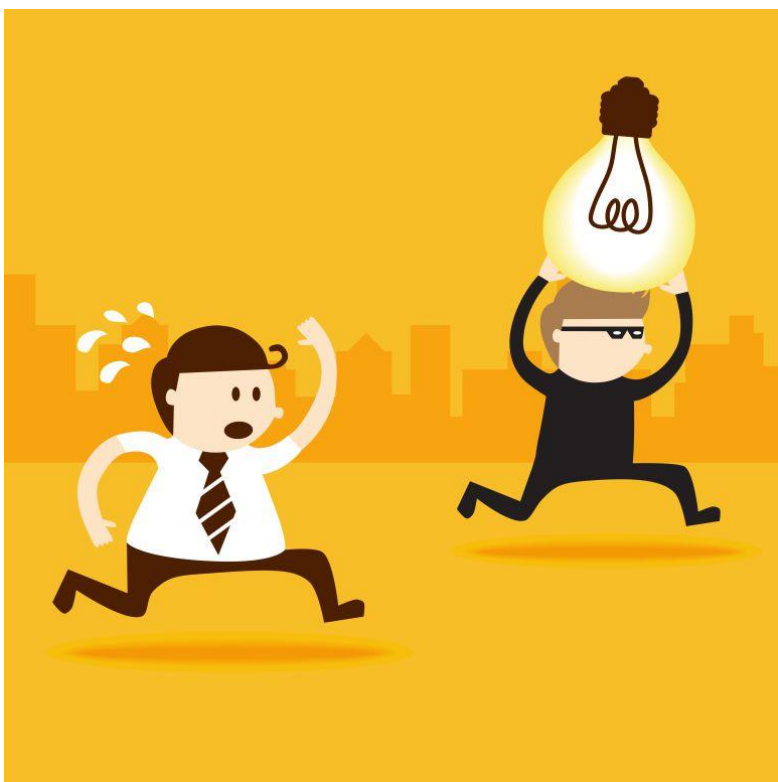
Luego se hace una tabulación de los resultados de las matrices donde se aprecia el desempeño de cada modelo de machine learning implementado con muestras de pruebas.

```
array([1, 0, 0, 0, 1, 1, 1, 0, 0, 0, 1, 0, 1, 0, 0, 1, 0, 1, 1, 1, 1, 1,
1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 0, 1, 1, 0, 0, 1, 1,
1, 0, 0, 1, 1, 1, 0, 1, 0, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 1,
1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 0, 0, 1, 1,
0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
dtype=int64)
```

Figura 1: Matriz de resultados

CONCLUSIONES

El Modelo de Balanced Random Forest, se encuentra en un punto medio entre los modelos evaluados y es la mejor opción para ser implementada, ya que es el que mejor identifica y comprende los patrones dentro de la muestra de entrenamiento y así consigue detectar con mejor precisión a usuarios con pérdidas no técnicas.



	Porcentaje de acierto de "1"	Porcentaje de acierto de "0"
Balanced Random Forest	61.9%	90.4%
Logistic Regresion	61.9%	85.7%
KNN	57.14%	95.2%

Figura 2: Resultados de la muestra 1

	Porcentaje de acierto de "1"	Porcentaje de acierto de "0"
Balanced Random Forest	64.7%	92%
Logistic Regresion	70.5%	82.9%
KNN	63%	88.6%

Figura 3: Resultados de la muestra 2

Los resultados de las predicciones de los modelos están altamente influenciados por la calidad de la muestra de entrenamiento, mientras más representativos sean los datos y mejor proporcionada sea la información con la que se entrenan los modelos de aprendizaje autónomo, mejor será el desempeño de los modelos