Maestría en Inteligencia Artificial Aplicada

UTPL
UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA

Análisis de datos y Visualización

Trabajo final. Mayo 2024

Integrantes:

- Henry Guarnizo
- Eduardo Araujo
- Astrid Medina

Tema

Sistema Inteligente

- Predecir el abandono escolar y el éxito académico de los estudiantes

UTPL
UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA
DECIDE SER más

DATA SET USADO

Tema	Descripción
Tamaño	Contiene 4424 filas y 37 columnas.
Variables	Las variables incluyen detalles demográficos, rendimiento académico y factores socioeconómicos como "Estado civil", "Modo de aplicación", "Curso", "Calificación previa", "Calificación de la madre", "Calificación del padre", "Nota de admisión", "Desplazado", "Necesidades educativas especiales", "Deudor", "Cuotas al día", "Género", "Becario", "Edad al inscribirse", "Unidades curriculares 1er semestre (aprobadas)", "Unidades curriculares 2do semestre (aprobadas)", "Tasa de desempleo", "Tasa de inflación" y "PIB".
Distribución	La mayoría de las variables son numéricas. No hubo valores nulos.
Patrones	Correlaciones altas entre "Notas de unidades curriculares" y la variable objetivo "Target".

1. Carga de datos

Importar librerías:

```
[ ] import pandas as pd
    import plotly.express as px
    import seaborn as sns
    import matplotlib.pyplot as plt
    from sklearn.model_selection import train_test_split
    from sklearn.metrics import accuracy_score,f1_score,precision_score,recall_score

    from sklearn.tree import DecisionTreeClassifier
    from sklearn.ensemble import RandomForestClassifier
    from sklearn.linear_model import LogisticRegression
    from sklearn.neighbors import KNeighborsClassifier
    from sklearn.ensemble import AdaBoostClassifier
    from xgboost import XGBClassifier
    from sklearn import svm
    from sklearn.metrics import classification_report
    import optuna
    from sklearn.model_selection import cross_val_score
    from sklearn.preprocessing import LabelEncoder


    from sklearn.model_selection import GridSearchCV, RandomizedSearchCV
    from sklearn.ensemble import VotingClassifier
```

1. Carga de datos:

```
[ ] df = pd.read_csv('/content/data.csv',sep=',')
```

1. Carga de datos

	Marital status	Application mode	Application order	Course	Daytime/evening attendance\t	Previous qualification	Previous qualification (grade)	Nacionality	Mother's qualification	Father's qualification	...	Curricular units 2nd sem (credited)	Curricular units 2nd sem (enrolled)
0	1	17	5	171	1	1	122.0	1	19	12	...	0	0
1	1	15	1	9254	1	1	160.0	1	1	3	...	0	6
2	1	1	5	9070	1	1	122.0	1	37	37	...	0	6
3	1	17	2	9773	1	1	122.0	1	38	37	...	0	6
4	2	39	1	8014	0	1	100.0	1	37	38	...	0	6
...	...	...	...	...	...	...	...	...	...	...	...	...	...
4419	1	1	6	9773	1	1	125.0	1	1	1	...	0	6
4420	1	1	2	9773	1	1	120.0	105	1	1	...	0	6
4421	1	1	1	9500	1	1	154.0	1	37	37	...	0	8
4422	1	1	1	9147	1	1	180.0	1	37	37	...	0	5
4423	1	10	1	9773	1	1	152.0	22	38	37	...	0	6

4424 rows × 37 columns

1. Carga de datos

Curricular units 2nd sem (evaluations)	Curricular units 2nd sem (approved)	Curricular units 2nd sem (grade)	Curricular units 2nd sem (without evaluations)	Unemployment rate	Inflation rate	GDP	Target
0	0	0.000000	0	10.8	1.4	1.74	Dropout
6	6	13.666667	0	13.9	-0.3	0.79	Graduate
0	0	0.000000	0	10.8	1.4	1.74	Dropout
10	5	12.400000	0	9.4	-0.8	-3.12	Graduate
6	6	13.000000	0	13.9	-0.3	0.79	Graduate
...	...	...	...	...	...	...	...
8	5	12.666667	0	15.5	2.8	-4.06	Graduate
6	2	11.000000	0	11.1	0.6	2.02	Dropout
9	1	13.500000	0	13.9	-0.3	0.79	Dropout
6	5	12.000000	0	9.4	-0.8	-3.12	Graduate
6	6	13.000000	0	12.7	3.7	-1.70	Graduate

Información del dataset

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4424 entries, 0 to 4423
Data columns (total 37 columns):
 #   Column                                          Non-Null Count  Dtype
---  ------                                          --------------  -----
 0   Marital status                                  4424 non-null   int64
 1   Application mode                                4424 non-null   int64
 2   Application order                               4424 non-null   int64
 3   Course                                          4424 non-null   int64
 4   Daytime/evening attendance                      4424 non-null   int64
 5   Previous qualification                          4424 non-null   int64
 6   Previous qualification (grade)                  4424 non-null   float64
 7   Nacionality                                     4424 non-null   int64
 8   Mother's qualification                          4424 non-null   int64
 9   Father's qualification                          4424 non-null   int64
 10  Mother's occupation                             4424 non-null   int64
 11  Father's occupation                             4424 non-null   int64
 12  Admission grade                                 4424 non-null   float64
 13  Displaced                                       4424 non-null   int64
 14  Educational special needs                       4424 non-null   int64
 15  Debtor                                          4424 non-null   int64
 16  Tuition fees up to date                         4424 non-null   int64
 17  Gender                                          4424 non-null   int64
 18  Scholarship holder                              4424 non-null   int64
 19  Age at enrollment                               4424 non-null   int64
 20  International                                   4424 non-null   int64
 21  Curricular units 1st sem (credited)             4424 non-null   int64
 22  Curricular units 1st sem (enrolled)             4424 non-null   int64
 23  Curricular units 1st sem (evaluations)          4424 non-null   int64
 24  Curricular units 1st sem (approved)             4424 non-null   int64
 25  Curricular units 1st sem (grade)                4424 non-null   float64
 26  Curricular units 1st sem (without evaluations)  4424 non-null   int64
 27  Curricular units 2nd sem (credited)             4424 non-null   int64
 28  Curricular units 2nd sem (enrolled)             4424 non-null   int64
 29  Curricular units 2nd sem (evaluations)          4424 non-null   int64
 30  Curricular units 2nd sem (approved)             4424 non-null   int64
 31  Curricular units 2nd sem (grade)                4424 non-null   float64
 32  Curricular units 2nd sem (without evaluations)  4424 non-null   int64
 33  Unemployment rate                               4424 non-null   float64
 34  Inflation rate                                  4424 non-null   float64
 35  GDP                                             4424 non-null   float64
 36  Target                                          4424 non-null   object
dtypes: float64(7), int64(29), object(1)
memory usage: 1.2+ MB
```

UTPL
DECIDE SER más

Análisis Exploratorio de datos

- Se inició con la carga y verificación del dataset, incluyendo la inspección de valores nulos y la distribución de las variables.
- La visualización inicial mostró que no había valores nulos en el dataset.
- Se realizó un análisis de correlación para identificar cómo las características se relacionan con la variable objetivo (Target). Esto ayudó a comprender qué variables tenían mayor impacto en la predicción de deserción escolar.

EDA

Comprobar valores nulos

```
df.isnull().sum()

Marital status                                    0
Application mode                                  0
Application order                                 0
Course                                            0
Daytime/evening attendance\t                      0
Previous qualification                            0
Previous qualification (grade)                    0
Nacionality                                       0
Mother's qualification                            0
Father's qualification                            0
Mother's occupation                               0
Father's occupation                               0
Admission grade                                   0
Displaced                                         0
Educational special needs                         0
Debtor                                            0
Tuition fees up to date                           0
Gender                                            0
Scholarship holder                                0
Age at enrollment                                 0
International                                     0
Curricular units 1st sem (credited)               0
Curricular units 1st sem (enrolled)               0
Curricular units 1st sem (evaluations)            0
Curricular units 1st sem (approved)               0
Curricular units 1st sem (grade)                  0
Curricular units 1st sem (without evaluations)    0
Curricular units 2nd sem (credited)               0
Curricular units 2nd sem (enrolled)               0
Curricular units 2nd sem (evaluations)            0
Curricular units 2nd sem (approved)               0
Curricular units 2nd sem (grade)                  0
Curricular units 2nd sem (without evaluations)    0
Unemployment rate                                 0
Inflation rate                                    0
GDP                                               0
Target                                            0
dtype: int64
```

Preparación de datos

- La columna Target, indicaba si un estudiante se graduó o abandonó, fue codificada de valores categóricos a numéricos.
- Se eliminaron características con baja correlación con la variable objetivo para simplificar el modelo y mejorar su rendimiento.
- El dataset fue dividido en dos subconjuntos: uno para el modelo de árbol de decisiones (df_decision_tree) y otro para el modelo de bosque aleatorio (df_random_forest).

Codificación

```
encoder = LabelEncoder()
df['Target'] = encoder.fit_transform(df['Target'])
df['Target'].value_counts()
```

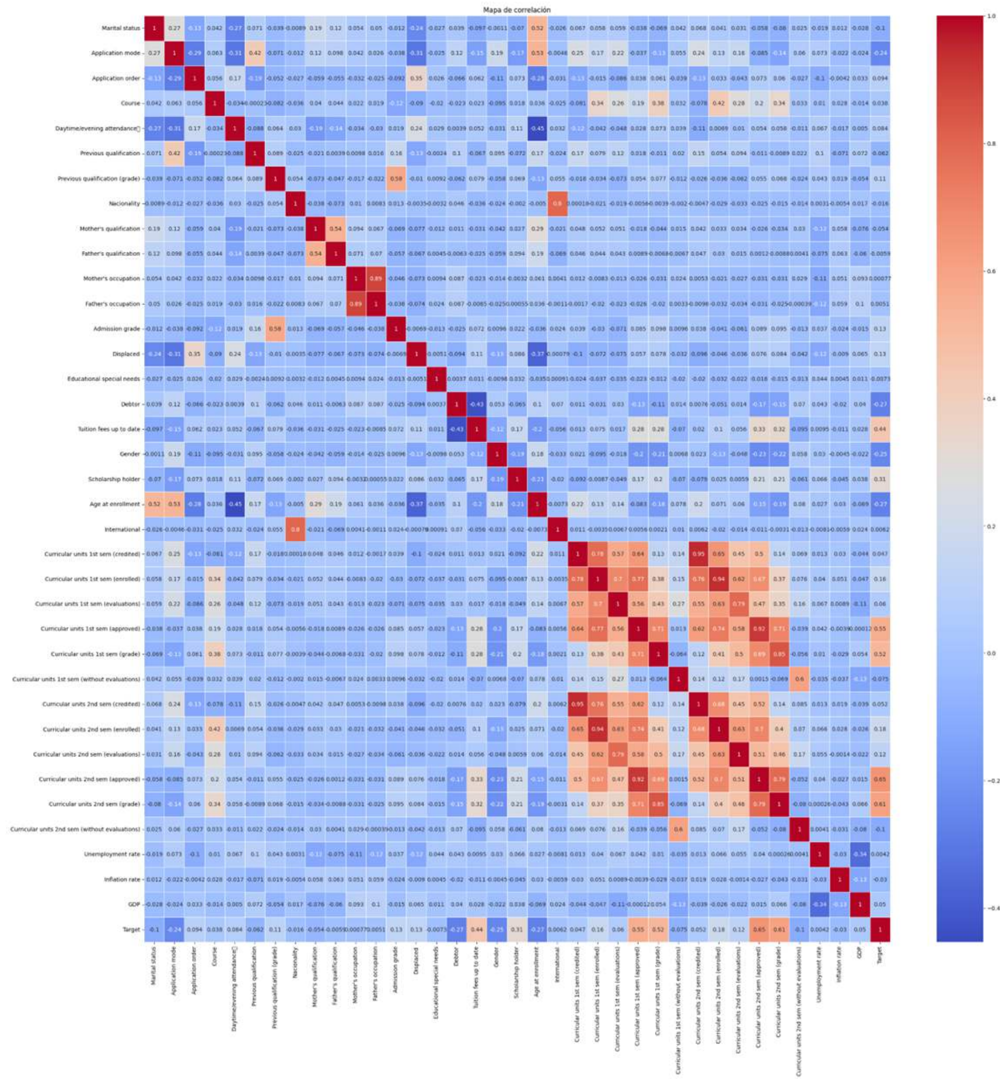
```
<ipython-input-162-ad1808a1de4f>:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df['Target'] = encoder.fit_transform(df['Target'])
Target
1    2209
0    1421
Name: count, dtype: int64
```

Luego de la codificación tenemos los siguientes valores para la variable target:

- Dropout (Abandono) - 0
- Graduate (Graduado) - 1

EDA Correlación de variables con target



EDA Selección de características

```python
#Realizamos una copia del archivo antes de decidir que variables eliminar, para cada modelo de predicción
df_decision_tree = df.copy()
df_random_forest=df.copy()
```

```python
df_decision_tree
```

	Marital status	Application mode	Application order	Course	Daytime/evening attendance\t	Previous qualification	Previous qualification (grade)	Nacionality	Mother's qualification	Father's qualification	...	Curricular units 2nd sem (credited)	Curricular units 2nd sem (enrolled)
0	1	17	5	171	1	1	122.0	1	19	12	...	0	0
1	1	15	1	9254	1	1	160.0	1	1	3	...	0	6
2	1						122.0	1	37	37	...	0	6
3	1						122.0	1	38	37	...	0	6
4	2												6

Mejorar el rendimiento del modelo, reducir la complejidad y mejorar la interpretabilidad.

Eliminar características irrelevantes, redundantes o ruidosas

Elegir el subconjunto de características más relevante del conjunto de datos original.

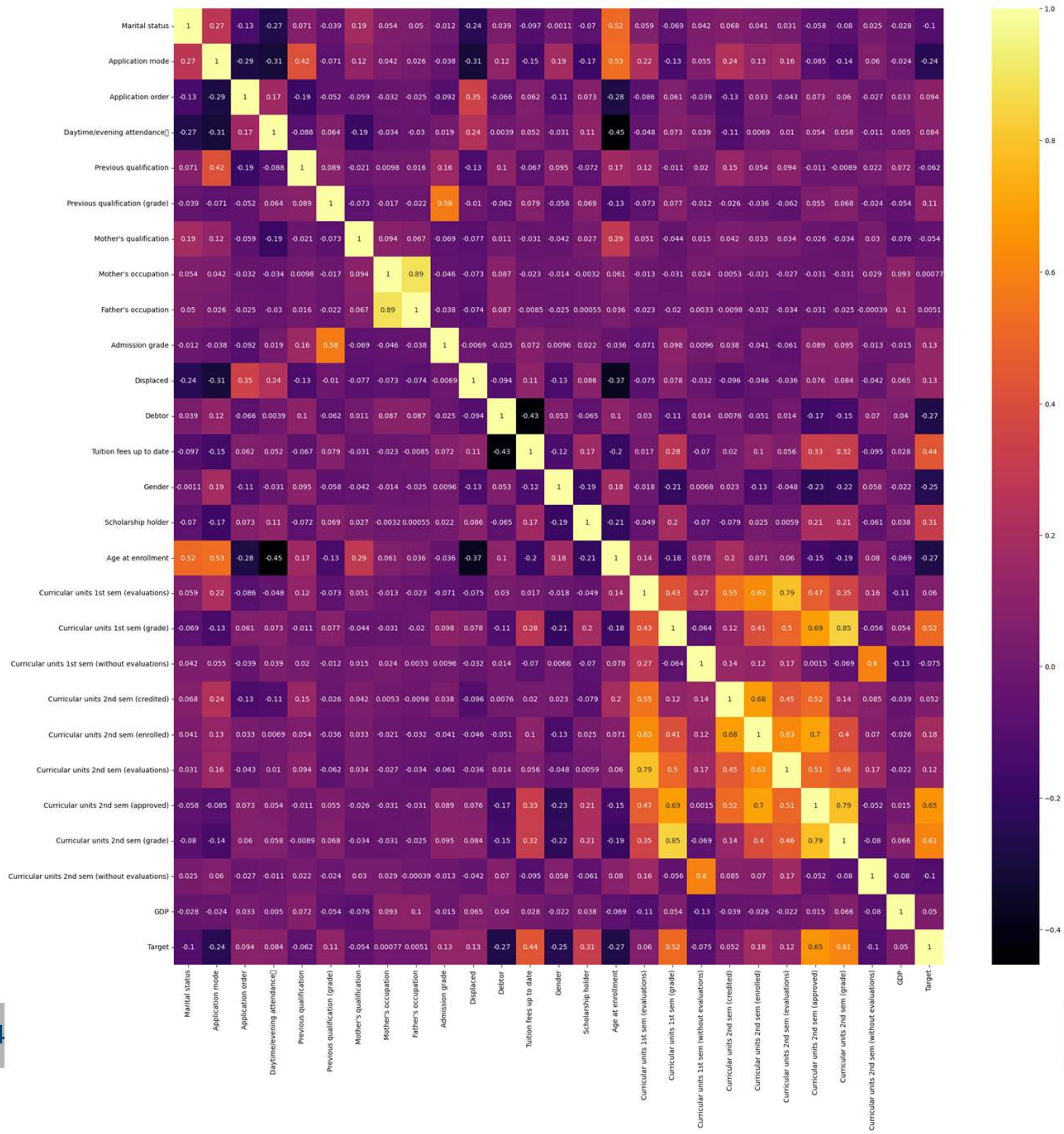
Selección de características

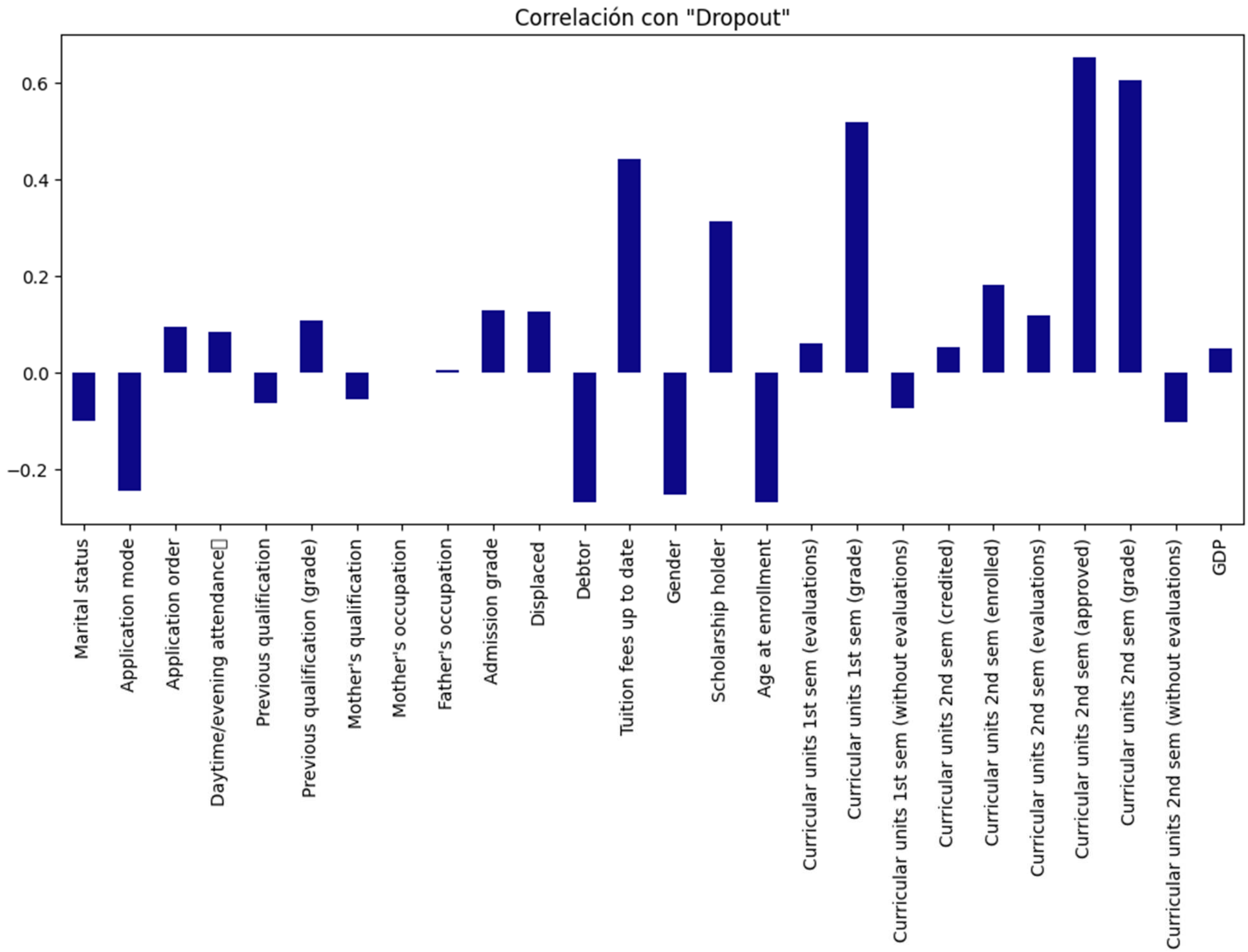
Esto tien req er eval

```python
df_random_forest.drop(['International',
                       'Nacionality',
                       "Father's qualification",
                       'Curricular units 1st sem (credited)',
                       'Curricular units 1st sem (enrolled)',
                       'Curricular units 1st sem (approved)','Course',
                       'Educational special needs','Unemployment rate',
                       'Inflation rate'],axis=1,inplace=True)
```

```python
df_decision_tree.drop(['International',
                       'Nacionality',
                       "Father's qualification",
                       'Curricular units 1st sem (credited)',
                       'Curricular units 1st sem (enrolled)',
                       'Curricular units 1st sem (approved)','Course',
                       'Educational special needs','Unemployment rate',
                       'Inflation rate'],axis=1,inplace=True)
```

EDA Correlación





Correlación con "Dropout"
0.6
0.4
0.2
0.0
−0.2
Marital status
Application mode
Application order
Daytime/evening attendance
Previous qualification
Previous qualification (grade)
Mother's qualification
Mother's occupation
Father's occupation
Admission grade
Displaced
Debtor
Tuition fees up to date
Gender
Scholarship holder
Age at enrollment
Curricular units 1st sem (evaluations)
Curricular units 1st sem (grade)
Curricular units 1st sem (without evaluations)
Curricular units 2nd sem (credited)
Curricular units 2nd sem (enrolled)
Curricular units 2nd sem (evaluations)
Curricular units 2nd sem (approved)
Curricular units 2nd sem (grade)
Curricular units 2nd sem (without evaluations)
GDP

Modelos

- Se definieron varios modelos de machine learning, incluyendo Árbol de Decisión, Bosque Aleatorio, Regresión Logística, KNN, AdaBoost, XGBoost y SVM.
- Durante el entrenamiento, se ajustaron los hiperparámetros de los modelos para optimizar su rendimiento.
- La precisión y otros métricos de rendimiento fueron evaluados para cada modelo.

DIVIDIR CONJUNTO

```
[ ] x = df_random_forest.drop
    y = df_random_forest['Tar
```

```
[ ] x1 = df_decision_tree.drc
    y1 = df_decision_tree['Ta
```

```
[ ] x.shape, y.shape
((3630, 26), (3630,))
```

```
[ ] x1.shape, y1.shape
((3630, 26), (3630,))
```

Datos de entrenamiento y datos de prueba

Random Forest

```
[ ] x_train, x_test, y_train, y_test = train_test_split(x,
                                                        y,
                                                        test_size = 0.3,
                                                        stratify = y,
                                                        random_state = 42)
```

```
[ ] x_train.shape, x_test.shape, y_train.shape, y_test.shape
((2541, 26), (1089, 26), (2541,), (1089,))
```

Decision Tree

```
[ ] x1_train, x1_test, y1_train, y1_test = train_test_split(x1,
                                                            y1,
                                                            test_size = 0.3,
                                                            stratify = y1,
                                                            random_state = 42)
```

```
[ ] x1_train.shape, x1_test.shape, y1_train.shape, y1_test.shape
((2541, 26), (1089, 26), (2541,), (1089,))
```

Escalado de datos

```
[ ] x_train = scaler.fit_transform(x_train)
    x_test = scaler.transform(x_test)
```

```
x_train
```

```
array([[-0.3002885 , -0.99691783, -0.56574082, ...,  0.54376269,
        -0.18874023,  0.79649683],
       [-0.3002885 , -0.12941428,  0.18220269, ...,  0.39318043,
        -0.18874023,  0.77418582],
       [-0.3002885 , -0.12941428,  0.18220269, ...,  0.39318043,
        -0.18874023, -1.39444413],
       ...,
       [-0.3002885 , -0.64991641, -0.56574082, ..., -1.80532043,
        -0.18874023,  0.35027667],
       [-0.3002885 , -0.99691783, -0.56574082, ...,  0.64917026,
        -0.18874023,  1.56399549],
       [ 1.37670726,  1.20075784, -0.56574082, ..., -1.80532043,
        -0.18874023, -0.41275979]])
```

```
[ ] x_test
```

```
array([[-0.3002885 , -0.99691783,  0.18220269, ...,  0.31272649,
        -0.18874023, -0.76081151],
       [-0.3002885 , -0.07158071, -0.56574082, ...,  0.18236528,
        -0.18874023,  0.35027667],
       [-0.3002885 ,  1.48992569, -0.56574082, ...,  0.15655118,
        -0.18874023, -1.81389107],
       ...,
       [-0.3002885 , -0.99691783,  0.18220269, ...,  0.430826  ,
        -0.18874023,  0.35027667],
       [-0.3002885 , -0.07158071, -0.56574082, ...,  0.18236528,
        -0.18874023,  0.1405532 ],
       [-0.3002885 , -0.07158071,  0.18220269, ...,  0.42329688,
        -0.18874023,  0.35027667]])
```

Selección del modelo y entrenamiento

```
[ ] x_train.shape, x_test.shape
```

```
((2541, 26), (1089, 26))
```

```
[ ] clf = RandomForestClassifier()
```

```
[ ] clf.fit(x_train, y_train)
```

```
▾ RandomForestClassifier
RandomForestClassifier()
```

```
y_pred = clf.predict(x_test)
y_pred
```

```
array([1, 0, 0, ..., 1, 0, 1])
```

```
[ ] clf.score(x_test, y_test)
```

```
0.9044995408631772
```

Mejorando el modelo con Hyper-parameter Tuning

```
param_grid = {
    'n_estimators': [50, 100],
    'max_depth': [None , 10, 20],
    'min_samples_split': [2,4,5],
    'min_samples_leaf': [1,2,4],
}
```

```
clf_grid=GridSearchCV(estimator=clf,
                      param_grid=param_grid,
                      cv=3,
                      verbose=0,
                      n_jobs=-1,
                      return_train_score=False
```

```
clf_grid.fit(x_train,y_train)
```

```
▸ GridSearchCV
▸ estimator: RandomForestClassifier
  ▸ RandomForestClassifier
```

```
clf_grid.best_params_
```

```
{'max_depth': None,
 'min_samples_leaf': 4,
 'min_samples_split': 4,
 'n_estimators': 50}
```

```
random_forest = RandomForestClassifier(**clf_grid.best_params_)
```

```
random_forest.fit(x_train, y_train)
```

```
▾ RandomForestClassifier
RandomForestClassifier(min_samples_leaf=4, min_samples_split=4, n_estimators=50)
```

Métricas de evaluación del modelo

```
from sklearn.metrics import classification_report, confusion_ma
```

```
accuracy_score = accuracy_score(y_test, y_pred)
print('Accuracy_score: ', accuracy_score)
```

```
Accuracy_score:  0.9044995408631772
```

```
precision_score = precision_score(y_test, y_pred)
print('Precision_score: ', precision_score)
```

```
Precision_score:  0.8909090909090909
```

```
f1_score = f1_score(y_test, y_pred)
print('f1_score: ', f1_score)
```

```
f1_score:  0.9245283018867925
```

```
print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
0	0.93	0.82	0.87	426
1	0.89	0.96	0.92	663
accuracy			0.90	1089
macro avg	0.91	0.89	0.90	1089
weighted avg	0.91	0.90	0.90	1089

Matríz de confusión final

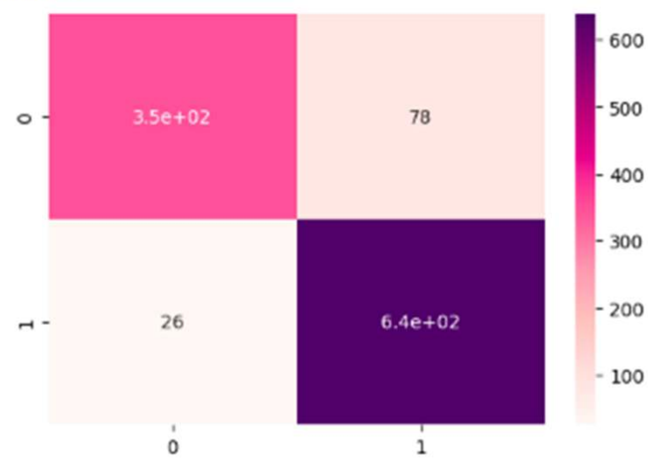
```
[ ] confusion_matrix = confusion_matrix(y_test, y_pred)
```

```
confusion_matrix
```

```
array([[348,  78],
       [ 26, 637]])
```

```
plt.figure(figsize = (6, 4))
sns.heatmap(confusion_matrix,
            annot = True,
            cmap = 'RdPu')
```

```
<Axes: >
```



Modelo predictivo con árboles de decisión

```
Mejores hiperparámetros: {'max_depth': 4, 'min_samples_leaf': 10, 'min_samples_split': 2}
Precisión: 0.8856749311294766
Reporte de clasificación:
              precision    recall  f1-score   support

           0       0.88      0.82      0.84       277
           1       0.89      0.93      0.91       449

    accuracy                           0.89       726
   macro avg       0.88      0.87      0.88       726
weighted avg       0.89      0.89      0.88       726
```

El modelo de clasificación presenta un buen rendimiento con una precisión global del 89%. Las métricas de precisión, recall y F1-score son altas para ambas clases, lo que indica que el modelo es eficaz tanto en la identificación de estudiantes que se gradúan como en aquellos que abandonan. Los hiperparámetros optimizados ayudan a mantener un equilibrio entre la complejidad del modelo y el riesgo de sobreajuste.

Optimización con Optuna

```python
print(dtree_study.best_value)
print(dtree_study.best_params)
```

```
0.9026635195168449
{'max_depth': 4, 'min_samples_leaf': 13, 'criterion': 'log_loss'}
```

```python
def dtree_objective(trial):

    md = trial.suggest_int('max_depth'
    mi = trial.suggest_int('min_sample
    crit = trial.suggest_categorical("

    clf =  DecisionTreeClassifier(max_
    scores = cross_val_score(clf, x1_t

    return scores.mean()

dtree_study = optuna.create_study(direction='maximize')
dtree_study.optimize(dtree_objective, n_trials=20)
```

```
[I 2024-05-31 22:14:50,550] A new study created in memory with name: no-name-885eda61-3d56-4900-959a-c2c4a99587af
[I 2024-05-31 22:14:51,019] Trial 0 finished with value: 0.9026635195168449 and parameters: {'max_depth': 4, 'min_samples_leaf': 13, 'criterion': 'log_loss'}. Best is trial 0 with value: 0.9026635195168449.
[I 2024-05-31 22:14:51,502] Trial 1 finished with value: 0.8868885390162029 and parameters: {'max_depth': 57, 'min_samples_leaf': 30, 'criterion': 'gini'}. Best is trial 0 with value: 0.9026635195168449.
[I 2024-05-31 22:14:51,755] Trial 2 finished with value: 0.8863844160924181 and parameters: {'max_depth': 14, 'min_samples_leaf': 23, 'criterion': 'gini'}. Best is trial 0 with value: 0.9026635195168449.
[I 2024-05-31 22:14:52,158] Trial 3 finished with value: 0.89273015627015 and parameters: {'max_depth': 25, 'min_samples_leaf': 30, 'criterion': 'log_loss'}. Best is trial 0 with value: 0.9026635195168449.
[I 2024-05-31 22:14:52,556] Trial 4 finished with value: 0.8863149493412745 and parameters: {'max_depth': 59, 'min_samples_leaf': 16, 'criterion': 'gini'}. Best is trial 0 with value: 0.9026635195168449.
[I 2024-05-31 22:14:53,073] Trial 5 finished with value: 0.8815391205014436 and parameters: {'max_depth': 58, 'min_samples_leaf': 12, 'criterion': 'log_loss'}. Best is trial 0 with value: 0.9026635195168449.
[I 2024-05-31 22:14:53,473] Trial 6 finished with value: 0.8854386831625762 and parameters: {'max_depth': 57, 'min_samples_leaf': 27, 'criterion': 'gini'}. Best is trial 0 with value: 0.9026635195168449.
[I 2024-05-31 22:14:53,832] Trial 7 finished with value: 0.8815391205014436 and parameters: {'max_depth': 45, 'min_samples_leaf': 12, 'criterion': 'log_loss'}. Best is trial 0 with value: 0.9026635195168449.
[I 2024-05-31 22:14:54,191] Trial 8 finished with value: 0.8937158209599125 and parameters: {'max_depth': 20, 'min_samples_leaf': 17, 'criterion': 'log_loss'}. Best is trial 0 with value: 0.9026635195168449.
[I 2024-05-31 22:14:54,522] Trial 9 finished with value: 0.8920939307633187 and parameters: {'max_depth': 16, 'min_samples_leaf': 20, 'criterion': 'log_loss'}. Best is trial 0 with value: 0.9026635195168449.
[I 2024-05-31 22:14:54,795] Trial 10 finished with value: 0.8880171953638805 and parameters: {'max_depth': 3, 'min_samples_leaf': 3, 'criterion': 'entropy'}. Best is trial 0 with value: 0.9026635195168449.
[I 2024-05-31 22:14:55,139] Trial 11 finished with value: 0.9001903556371428 and parameters: {'max_depth': 4, 'min_samples_leaf': 7, 'criterion': 'log_loss'}. Best is trial 0 with value: 0.9026635195168449.
[I 2024-05-31 22:14:55,400] Trial 12 finished with value: 0.8880171953638805 and parameters: {'max_depth': 3, 'min_samples_leaf': 4, 'criterion': 'entropy'}. Best is trial 0 with value: 0.9026635195168449.
[I 2024-05-31 22:14:55,786] Trial 13 finished with value: 0.8665274198528532 and parameters: {'max_depth': 38, 'min_samples_leaf': 8, 'criterion': 'log_loss'}. Best is trial 0 with value: 0.9026635195168449.
[I 2024-05-31 22:14:56,168] Trial 14 finished with value: 0.8823479692054648 and parameters: {'max_depth': 9, 'min_samples_leaf': 8, 'criterion': 'log_loss'}. Best is trial 0 with value: 0.9026635195168449.
[I 2024-05-31 22:14:56,603] Trial 15 finished with value: 0.8505511242597903 and parameters: {'max_depth': 29, 'min_samples_leaf': 1, 'criterion': 'log_loss'}. Best is trial 0 with value: 0.9026635195168449.
[I 2024-05-31 22:14:56,933] Trial 16 finished with value: 0.8752229825075872 and parameters: {'max_depth': 11, 'min_samples_leaf': 8, 'criterion': 'entropy'}. Best is trial 0 with value: 0.9026635195168449.
[I 2024-05-31 22:14:57,180] Trial 17 finished with value: 0.8815391205014436 and parameters: {'max_depth': 37, 'min_samples_leaf': 12, 'criterion': 'log_loss'}. Best is trial 0 with value: 0.9026635195168449.
[I 2024-05-31 22:14:57,415] Trial 18 finished with value: 0.8862485021117544 and parameters: {'max_depth': 23, 'min_samples_leaf': 16, 'criterion': 'log_loss'}. Best is trial 0 with value: 0.9026635195168449.
[I 2024-05-31 22:14:57,625] Trial 19 finished with value: 0.8861199677688198 and parameters: {'max_depth': 7, 'min_samples_leaf': 7, 'criterion': 'entropy'}. Best is trial 0 with value: 0.9026635195168449.
```

Conclusiones

- La preparación adecuada de los datos y la selección de características relevantes son cruciales para el rendimiento del modelo.
- Diferentes modelos de machine learning pueden ser evaluados y comparados para seleccionar el más adecuado según las métricas de interés.
- La combinación de técnicas de visualización y análisis de datos facilita la comprensión de los factores que influyen en la deserción escolar.

Gracias…

UTPL

UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA