

## **El aprendizaje automático en entornos educativos universitarios: Caso deserción académica**

### **Application of Machine Learning in university educational environments: Academic dropout case**

**Carlos Henriquez<sup>1</sup>, Dixon Salcedo<sup>2</sup>, German Sanchez-Torres<sup>3</sup>**

<sup>1</sup> Ph.D. Grupo de Investigación en Sistemas Inteligentes y Nuevas Tecnologías - SINT, Facultad de Ingeniería, Universidad Autónoma del Caribe, Barranquilla, Colombia.

<sup>2</sup>Ph.D Grupo de Investigación en Ingeniería del Software y Redes, Departamento de Ciencias de la Computación y Electrónica, Universidad de la Costa, Barranquilla, Colombia

<sup>3</sup>:Ph.D. Grupo de Investigación y Desarrollo en Sistemas y Computación, Universidad del Magdalena, Santa Marta, Colombia

Email: carlos.henriquez@uac.edu.co

*Cite this article as: C. Henriquez, D. Salcedo y G. Sánchez-Torres, "El aprendizaje automático en entornos educativos universitarios: Caso deserción académica.", Prospectiva, Vol 20, N° 1, 2022.*

**Recibido: 10/08/2021**

**Aceptado: 18/11/2021**

<http://doi.org/10.15665/rp.v20i1.2736>

#### **RESUMEN**

Uno de los principales problemas que enfrenta el sistema de educación superior en distintos países se relaciona con los altos niveles de deserción académica al cursar una carrera de pregrado. El número de alumnos que logra culminar sus estudios superiores no es alto, evidenciando que una gran parte de éstos abandona sus estudios principalmente en los primeros semestres. Las nuevas tecnologías como el aprendizaje automático han permitido la creación de aplicaciones de software que ayudan a la comprensión y solución de muchos problemas actuales. En el campo de la educación, estas tecnologías se han aplicado en los procesos administrativos y académicos durante mucho tiempo. Este artículo presenta el diseño de un modelo que permite predecir si un estudiante desertará o no. Para la validación del modelo se construyó un prototipo y se realizaron varios experimentos con datos reales de una institución educativa. Los resultados obtenidos son prometedores y sientan las bases para futuras investigaciones en este campo.

**Palabras clave:** Aprendizaje automático; Predicción; Entornos universitarios; Prototipo; Deserción universitaria.

## ABSTRACT

One of the main problems faced by the higher education system in different countries concerns the high levels of academic dropout when studying for an undergraduate degree. The number of students who manage to complete their higher studies is not high, suggesting that a large part of them abandon their studies, mainly in the first semesters. New technologies such as machine learning have allowed the creation of software applications that help understand and solve many current problems. In the field of education, these technologies have been applied in administrative and academic processes for a long time. This article presents the design of a model that allows predicting whether a student will drop out or not. For the validation of the model, a prototype was built and several experiments were carried out with real data from an educational institution. The results obtained are promising and lay the foundations for future research in this field.

**Key words:** machine learning; Prediction; University environments; Prototype; college dropout.

## 1. INTRODUCCIÓN

El aprendizaje automático (AA) se centra en desarrollar sistemas que aprenden, o mejoran el rendimiento, en función de los datos que consumen. Es uno de los campos técnicos de más rápido crecimiento en la actualidad [1]. Este se encuentra ubicado en la intersección de la informática y las estadísticas, y en el núcleo de la inteligencia artificial y la ciencia de la información. El progreso reciente en el aprendizaje automático ha sido impulsado tanto por el desarrollo de nuevos algoritmos y teoría de aprendizaje como por la continua explosión en la disponibilidad de datos en línea y el cálculo de bajo costo [2]. La adopción de métodos de aprendizaje de máquina con uso intensivo de datos se puede encontrar a lo largo de la ciencia, la tecnología y el comercio. Lo que ha permitido aumentar el número de decisiones basadas en la evidencia en diferentes ámbitos de la sociedad; incluida la atención médica, manufactura, educación, modelos financieros, políticas y mercadeo [3].

Uno de los principales problemas que enfrenta el sistema de educación superior en distintos países concierne a los altos niveles de deserción académica al cursar una carrera de pregrado. El número de alumnos que logra culminar sus estudios superiores no es alto, evidenciando que una gran parte de éstos abandona sus estudios, principalmente en los primeros semestres [4].

El uso de técnicas de aprendizaje automático para propuestas educativas es un campo emergente destinado para desarrollar métodos para explorar datos desde entornos educativos computacionales y descubrir patrones significativos [5]. Los datos almacenados pueden ser útiles para los algoritmos de aprendizaje automático, por ejemplo, las características demográficas claves de los estudiantes pueden constituir un conjunto de datos que alimenten una máquina y poder así, realizar distintas predicciones sobre el comportamiento de un estudiante.

Tanto en Colombia, como en varios países de América Latina, se evidencia que la deserción universitaria ha ido en aumento durante los últimos 10 años, y a diferencia del comportamiento que presenta en distintos países, en Colombia la cifra de estudiantes que desertan por algún motivo específico luego de matricularse a una carrera universitaria asciende a 37%, y a su vez, son alrededor del 36% de los estudiantes desertores los que presentan este comportamiento a final del primer año [6]. Por lo que, es posible identificar que una cantidad cercana a la mitad de los estudiantes que ingresan a los programas de educación superior se retiran sin terminar su proceso de formación. Además de lo anterior, la deserción estudiantil universitaria en Colombia ha presentado distintas tasas o intensidad que varían respecto a cada departamento. Durante los últimos años, el departamento del Atlántico ha presentado una tasa de deserción estudiantil en el nivel de educación superior del 46%, la cual es una cantidad alta a tener en cuenta en este aspecto [7].

Las universidades del departamento del Atlántico presentan índices de deserción estudiantil alto, lo que hace importante tomar este ámbito como área de estudio. En lo que respecta a la Universidad Autónoma del Caribe como institución educativa, perteneciente al departamento del Atlántico, se tomó el área de Ingeniería de Sistemas como referencia para estudiar el comportamiento de esta problemática en la institución.

Por todo lo anterior, este artículo busca construir un prototipo basado en aprendizaje automático para caracterizar la permanencia estudiantil en un programa académico. Este estudio es el resultado de un proyecto de investigación a nivel de semillero con la participación del semillero de investigación del programa de Ingeniería de Sistemas de la Universidad Autónoma del Caribe, Barranquilla, Colombia.

El resto del artículo se estructura de la siguiente manera. En la sección 2, se presentan los antecedentes y trabajos similares. En la sección 3, se describe el modelo propuesto para el desarrollo de este estudio, en la sección 4 los resultados y discusión. Finalmente, se presentan las conclusiones en la sección 5.

## **2. ANTECEDENTES Y OBRAS AFINES**

Diversos trabajos se encuentran en la literatura relacionados con la educación, la deserción y el aprendizaje automático. Por ejemplo, en [8] se exploró el comportamiento de abandono escolar de los estudiantes en los cursos en línea masivos y abiertos (MOOC), aquí, se analizaron los factores relacionados con el comportamiento de los estudiantes y el posicionamiento social en los foros de discusión utilizando técnicas analíticas de redes sociales estándar. El análisis reveló varios predictores significativos de abandono escolar.

En el estudio de [9] se realizó un análisis de deserción-permanencia de estudiantes universitarios utilizando técnica de clasificación en minería de datos. Aquí, a través de la aplicación de esta técnica se identificaron los factores que influyeron sobre la deserción de los estudiantes de la carrera de Ingeniería en Informática de la Universidad Gastón Dachary en Argentina. Asimismo, en [10] se analizaron tres métodos de aprendizaje automático para

predecir la deserción en estudiantes de una amplia muestra de alumnos inscritos en programas de educación a distancia en China.

Adicionalmente, en [11] se presentó un trabajo para predecir la deserción de los estudiantes en entornos de aprendizaje en línea basados en suscripciones. Este enfoque, muestra la comparación de un algoritmo de modelo de hoja Logit utilizando un conjunto de datos de la vida real, en donde se especifica los impactos de la demografía de los estudiantes; características del aula; y variables de participación académica, cognitiva y conductual sobre la deserción de los estudiantes.

Entre otros trabajos referentes están, el Modelo predictivo de deserción estudiantil basado en arboles de decisión que muestra la construcción de un modelo para pronosticar la probabilidad, que un estudiante abandone su programa académico, mediante técnicas de clasificación, basadas en árboles de decisión. La metodología utilizada, se basó en *Knowledge Discovery in Database*, con cinco etapas: selección, procesamiento, transformación, minería de datos y evaluación [12]. También, se encuentra [13] que buscó la implementación de un método para predecir la deserción de los estudiantes en los años superiores de un programa de fisioterapia. Aquí se accedió a una base de datos de 336 estudiantes universitarios en sus cursos superiores, en donde se eligieron cinco variables cuantitativas y 11 cualitativas, asociadas a la deserción de estudiantes universitarios. Con esta base de datos, se probaron 23 clasificadores basados en aprendizaje automático supervisado.

Finalmente, otros estudios pertinentes como en [14], que analizaron el rendimiento de cuatro algoritmos de aprendizaje automático con diferentes perspectivas para definir archivos de datos, en la predicción de la deserción de estudiantes universitarios. Estos algoritmos fueron bosque aleatorio, redes neuronales, máquinas de vectores de soporte y regresión logística. Asimismo, en el estudio realizado por [15], que presentó una clasificación basada en árboles de decisión (CBAD) con parámetros optimizados para predecir la deserción de los estudiantes universitarios. El estudio analizó 5288 casos de estudiantes pertenecientes a una universidad pública chilena. Para la técnica CBAD se optimizaron los parámetros para mejorar la predicción usando el software RapidMiner.

Los anteriores referentes, sentaron las bases para diseñar el modelo propuesto de aprendizaje automático, así como para analizar las variables estudiadas para la deserción, conocer los distintos algoritmos de aprendizaje utilizados y las métricas más usadas.

### **3. MODELO PROPUESTO**

El modelo consta de las siguientes fases: preprocesamiento, selección de algoritmos, entrenamiento, evaluación y ejecución y despliegue de la predicción.

#### **3.1 Fase de preprocesamiento**

Para la ejecución de todas las fases, es necesario tener en cuenta la consideración de un conjunto específico de variables que influyen en mayor medida al comportamiento desertor,

al resultado final que se obtendrá en la predicción. Asimismo, en esta fase se realiza la extracción de datos desde la fuente original y se realiza un proceso de limpieza y transformación de los datos. Finalmente, se cargan los datos finales para su análisis.

### **3.2 Fase selección algoritmos**

La selección de los algoritmos a utilizar se llevó a cabo teniendo en cuenta las dos categorías posibles de algoritmos de Aprendizaje automático, las cuales son regresión o clasificación. Se optó por utilizar algoritmos de clasificación por las características de los datos manejados. Para establecer específicamente qué algoritmos se deberían utilizar, se tuvo en cuenta investigaciones anteriores realizadas por otros autores con un objetivo similar a la presente investigación.

### **3.3 Fase de entrenamiento**

Una vez seleccionados los algoritmos, se procedió a dar lugar al entrenamiento de estos, donde se escogió únicamente una fracción de los datos equivalente al 20% de los datos totales ya normalizados con el fin de ser utilizados como datos de entrenamiento y que de esta forma la máquina sea capaz de obtener conclusiones acertadas. Posteriormente, se utilizó el 40, 60 y 80 respectivamente, de los datos para ser utilizados como entrenamiento de la máquina.

### **3.4 Fase de Evaluación**

Una vez entrenado el modelo con los porcentajes determinados de datos de entrenamiento con respecto a los datos totales, se utilizó porcentaje restante como datos de prueba. Así mismo, una vez que se realizó la prueba con estos datos, se evaluó qué tan acertada fue la predicción utilizando distintas métricas de evaluación, como son las siguientes: Precisión, Sensibilidad, Exactitud y Puntaje F1 [16]. Con base a estas métricas aplicadas en cada algoritmo, se escogerá aquel que obtenga un mejor resultado.

### **3.5 Fase de ejecución de la predicción y despliegue**

Una vez escogidos los algoritmos de predicción, estos se evaluaron para ser utilizado en la versión final del proyecto. En esta fase se diseña una interfaz gráfica que tendrá campos disponibles para que el usuario ingrese las variables del estudiante que desea conocer si desertará o no. Estas variables ingresadas por el usuario serán enviadas al algoritmo de predicción final, y los resultados que este proporcione (Desertará o no).

## **4. RESULTADOS Y VALIDACIÓN**

Para la validación del modelo se tomaron los registros de los estudiantes matriculados desde los periodos 2017-01 hasta 2020-02 de la facultad de ingeniería de la Universidad Autónoma del Caribe. En la fase de preprocesamiento, para el desarrollo de la investigación se

definieron y seleccionaron un conjunto de variables partir de la disponibilidad de datos. Estas fueron:

- Tipo de Ingreso del Estudiante.
- Sexo del Estudiante.
- Estrato del estudiante
- Estado de la inscripción
- Ciudad de Procedencia del Estudiante

Se aplicaron distintos métodos para el análisis, transformación y limpieza de los datos, utilizando el lenguaje de programación Python, implementando diferentes librerías [17]. Estos métodos fueron:

- Método de extracción de datos: Por medio de las funciones que nos proporciona la librería Pandas, se realizó la lectura de los registros provenientes de archivos en formato de CSV (*comma-separated values*), para ser incorporados al proyecto.
- Método de limpieza y transformación de los datos: Una vez obtenidos los datos categorizados, se normalizaron y se llevó a cabo una limpieza donde se filtraron únicamente las variables influyentes mencionadas con anterioridad. A partir de esto, se generó un nuevo archivo de registros en formato csv, el cual fue utilizado para realizar la conversión de datos a tipo numérico con el fin de que pudiesen ser utilizados en los algoritmos de Aprendizaje de Máquina. Dicha conversión de datos se llevó de la siguiente manera:
  - Los datos de la variable sexo se cambiaron de F (Femenino) y M (Masculino), a 0 y 1, respectivamente.
  - Los datos de la variable Tipo de Ingreso se cambiaron de Antiguo, Por Primera vez, Reingreso, Transferencia Externa, y Transferencia Interna, a 1, 2, 3, 4, y 5, respectivamente.
  - Los datos faltantes en la variable estrato fueron asignados a partir de los distintos valores del rango aleatorio entre la diferencia de la media y la desviación estándar hasta la suma entre estas mismas.
  - Los datos de la variable Ciudad de Procedencia Geográfica se cambiaron de Barranquilla a 1, y todos aquellos lugares diferentes de Barranquilla a 0.
  - Los datos de la variable Estado Inscripción se cambiaron por Admitido = 1, Inscrito = 2.
- Método de Carga de datos preparados: Una vez ya preparados los datos, se realizó la exportación de los registros por medio de funciones que nos proporciona la librería Pandas para este fin.

Para establecer específicamente cuales algoritmos utilizar, se consideró reportes del estado del arte de investigaciones anteriores realizadas por otros autores con un objetivo similar a la presente investigación. A partir de dicha indagación, se decidió optar por los algoritmos: Regresión logística, Árboles de decisión y Clasificador Bayesiano, debido a que permiten mayor interpretabilidad de los resultados, los que nos permite contribuir con la comprensión del fenómeno de la deserción, igualmente, estos algoritmos se presentan mayor frecuencia de uso en las distintas investigaciones y se corresponden con la mayor precisión para este tipo de problemas. Una vez seleccionados los algoritmos, se procedió a dar lugar al entrenamiento de estos.

A continuación, se muestra partes del Código creado en el lenguaje Phyton y sus diferentes librerías como Pandas, Numpy, matplotlib, seaborn, sklearn, entre otras. Se muestra el código para la limpieza y transformación de los datos y partes de la implementación de un algoritmo seleccionado. En la implementación del proceso se cambia la instancia del modelo en Phyton de cada uno de los algoritmos.

#### ALGORITMO LIMPIEZA Y TRANSFORMACIÓN DE LOS DATOS:

```
#!/usr/bin/env python
# coding: utf-8

# In[48]:
import pandas as pd
import sys
import numpy as np
# In[22]:
data=pd.read_csv('CSV-DIRTY-JUNTOS/Registro1402.csv')
data
# In[37]:
def ciudad_fix(CIUDAD):
    if CIUDAD!='BARRANQUILLA':
        return 0
    return 1
# In[49]:
def estrato_fix(ESTRATO_PR):
    if ESTRATO_PR==0.0:
        return None
    return ESTRATO_PR

# In[66]:
def transform(ruta):
    dataframe = pd.read_csv('CSV-DIRTY-JUNTOS/'+ruta+'.csv',encoding="latin-1")
    TotalDF = pd.DataFrame()
    TotalDF
    [['TIPOINGRESO', 'GENERO', 'SEMESTRE', 'ESTRATO_PR', 'ICFES_PUNTAJE', 'ICFES_TIPO', 'CIUDAD_PROCEDENCIA
    GEOGR', 'DESERTOR']] =
    dataframe[['TIPOINGRESO', 'GENERO', 'SEMESTRE', 'ESTRATO_PR', 'ICFES_PUNTAJE', 'ICFES_TIPO', 'CIUDAD_PR
    OCEDENCIAGEOGR', 'DESERTOR']]
    TotalDF ['GENERO'] = TotalDF ['GENERO'].replace({'M':1, 'F':0})
    TotalDF ['ESTRATO_PR']=TotalDF['ESTRATO_PR'].fillna(0)
    TotalDF ['ICFES_TIPO']= TotalDF ['ICFES_TIPO'].fillna(0)
    TotalDF ['ICFES_PUNTAJE']= TotalDF ['ICFES_PUNTAJE'].fillna(0)
    #ASIGNANDO A CADA CIUDAD, ---crea una nueva columna
    TotalDF['CIUDAD']=TotalDF.apply(lambda x: ciudad_fix(x['CIUDAD_PROCEDENCIAGEOGR']), axis=1);
    #ASIGNANDO A CADA ESTRATO, ---crea una nueva columna
    TotalDF['ESTRATO_new']=TotalDF.apply(lambda x: estrato_fix(x['ESTRATO_PR']), axis=1);

#SACAR VALOR RANDOM PARA ESTRATO_NEW
```

```

estrato_avg = TotalDF['ESTRATO_new'].mean()
estrato_std = TotalDF['ESTRATO_new'].std()
estrato_null_count = TotalDF['ESTRATO_new'].isnull().sum()
estrato_null_random_list = np.random.randint(estrato_avg - estrato_std, estrato_avg +
estrato_std, size=estrato_null_count)
conValoresNulos = np.isnan(TotalDF['ESTRATO_new'])
TotalDF.loc[np.isnan(TotalDF['ESTRATO_new']), 'ESTRATO_new'] = estrato_null_random_list
TotalDF['ESTRATO_new'] = TotalDF['ESTRATO_new'].astype(int)
#SACAR VALOR RANDOM PARA ESTRATO_NEW

TotalDF ['ICFES_TIPO'] = TotalDF ['ICFES_TIPO'].replace({'EXAMEN TIPO 1(ANTES DEL
2000)':1,'EXAMEN TIPO 2(ANIO 2000)':2, 'EXAMEN TIPO 3(ANTES DEL 2000)':3,'EXAMEN TIPO 4 ( ANIO
2006)':4,'EXAMEN TIPO 5 (ANIO 2014)':5} )
TotalDF ['TIPOINGRESO'] = TotalDF ['TIPOINGRESO'].replace({'ANTIGUO':1,'POR PRIMERA VEZ':2,
'REINGRESO':3,'TRANSFERENCIA EXTERNA':4,'TRANSFERENCIA INTERNA':5} )
[TotalDF.ESTRATO_PR,TotalDF.GENERO,TotalDF.SEMESTRE, TotalDF.TIPOINGRESO, TotalDF.CIUDAD,
TotalDF.ESTRATO_new] =
[TotalDF.ESTRATO_PR.astype(float),TotalDF.GENERO.astype(float),TotalDF.SEMESTRE.astype(float),Tot
alDF.TIPOINGRESO.astype(float),TotalDF.CIUDAD.astype(float),TotalDF.ESTRATO_new.astype(float)]
#TotalDF
[['TIPOINGRESO', 'GENERO', 'SEMESTRE', 'ESTRATO_new', 'ICFES_PUNTAJE', 'ICFES_TIPO', 'CIUDAD',
'DESERTOR']] =
TotalDF[['TIPOINGRESO', 'GENERO', 'SEMESTRE', 'ESTRATO_new', 'ICFES_PUNTAJE', 'ICFES_TIPO', 'CIUDAD',
'DESERTOR']]
#TotalDF.drop(['ESTRATO_PR', 'CIUDAD_PROCEDENCIAGEOGR'], axis=1)
del TotalDF['ESTRATO_PR']
del TotalDF['CIUDAD_PROCEDENCIAGEOGR']
TotalDF.to_csv("cleaned_data/"+ruta+"CLEANED.csv",encoding="utf-8")
return TotalDF

# In[77]:
data = transform('Registro1501')

```

### ALGORITMO NAIVE BAYES:

```

#!/usr/bin/env python
# coding: utf-8

# In[1]:

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from matplotlib import colors
import seaborn as sb
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score

get_ipython().run_line_magic('matplotlib', 'inline')
plt.rcParams['figure.figsize'] = (18, 9)
plt.style.use('ggplot')

from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
from sklearn.naive_bayes import GaussianNB
from sklearn.feature_selection import SelectKBest

# In[5]:

cd C:\Users\LENOVO\Desktop\Investigacion\csv\registros\cleaned_data

# In[6]:

data = pd.read_csv('registro1402cleaned.csv')
data

```



```

# In[7]:

olormap = plt.cm.viridis
plt.figure(figsize=(12,12))
plt.title('Pearson Correlation of Features', y=1.05, size=15)
sb.heatmap(data.astype(float).corr(),linewidths=0.1,vmax=1.0, square=True, cmap=colormap,
linecolor='white', annot=True)

# In[9]:

X=data.drop(['DESERTOR'], axis=1)
y=data['DESERTOR']

best=SelectKBest(k=6)
X_new = best.fit_transform(X, y)
X_new.shape
selected = best.get_support(indices=True)
print(X.columns[selected])

# In[10]:

used_features =X.columns[selected]

colormap = plt.cm.viridis
plt.figure(figsize=(12,12))
plt.title('Pearson Correlation of Features', y=1.05, size=15)
sb.heatmap(data[used_features].astype(float).corr(),linewidths=0.1,vmax=1.0, square=True,
cmap=colormap, linecolor='white', annot=True)

# In[11]:
# Split dataset en training y test
X_train, X_test = train_test_split(data, test_size=0.45, random_state=6)
y_train =X_train["DESERTOR"]
y_test = X_test["DESERTOR"]

# In[12]:

gnb = GaussianNB()
gnb.fit(X_train[used_features].values,y_train)
y_pred = gnb.predict(X_test[used_features])

print('Precisión en el set de Entrenamiento: {:.2f}'
      .format(gnb.score(X_train[used_features], y_train)))
print('Precisión en el set de Test: {:.2f}'
      .format(gnb.score(X_test[used_features], y_test)))

# In[13]:

print('Exactitud: ', format(accuracy_score(y_test, y_pred)))
print('Precisión: ', format(precision_score(y_test, y_pred)))
print('Sensibilidad: ', format(recall_score(y_test, y_pred)))
print('F1: ', format(f1_score(y_test, y_pred)))

```

Los resultados de la implementación de los algoritmos seleccionados se presentan en la Tabla 1. Por el lado de la Sensibilidad, que compara la cantidad de casos clasificados como verdaderos positivos sobre todo lo que realmente era positivo, los resultados no son los esperados para ninguno de los algoritmos tratados. Esto puede ser causa del desbalanceo de los datos, ya que existen más registros de datos No desertores que desertores. Asimismo, se puede interpretar, que para el modelo las variables seleccionadas no son lo suficiente

explicativas para predecir el comportamiento de matrícula o no. Sin embargo, es un buen inicio para explorar otros algoritmos y experimentar con las diferentes variables que aparecen en los informes iniciales.

Tabla 1. Resultados de la implementación de los algoritmos

Table 1: Results of the algorithm implementation

Algoritmo	Precisión		Exactitud:	Precisión	Sensibilidad	F1
	Entrenamiento	Prueba				
Naive Bayes	0,84	0,84	0,84	0,99	0,05	0,11
Regresión Logística	0,78	0,77	0,78	0,78	0,05	0,10
Árboles de Decisión	0,85	0,82	0,82	0,49	0,11	,018

## 5. CONCLUSIÓN

- La deserción estudiantil en los estudiantes de educación superior que puede ser explicado por diferentes causas como las individuales, la familia, el estado además de otras variables como las académicas, las financieras, las socioeconómicas, las psicológicas y las sociales. El análisis de estas variables y sus categorías depende del contexto en el que se realicen debido a que la complejidad de cada situación está determinada por muchos factores que son únicos para cada entorno y hacen que exista una gran variedad de causas de deserción.
- De los resultados es posible concluir que las nuevas tecnologías, específicamente la inteligencia artificial y el aprendizaje automático nos pueden ayudar mucho para desarrollar este tipo de modelos en los cuales se necesite predecir sobre datos y así poder llegar a soluciones para mejorar estos problemas que se presentan como en este caso el abandono del proceso de matrícula o inscripción por diferentes motivos que se le presenten al estudiante.

## 6. REFERENCIAS

- [1] C. Henríquez, F. Briceño, and D. Salcedo, “Unsupervised Model for Aspect-Based Sentiment Analysis in Spanish,” *IAENG Int. J. Comput. Sci.*, no. 3, pp. 430–438, 2019.
- [2] C. Henríquez and J. Guzmán, “Las ontologías para la detección automática de aspectos en el análisis de sentimientos,” *Rev. Prospect.*, vol. 14, no. 2, pp. 90–98, 2016.
- [3] M. I. Jordan and T. M. Mitchell, “Machine learning: Trends, perspectives, and prospects,” *Science (80-. )*, vol. 349, no. 6245, pp. 255–260, 2015.
- [4] L. Paura and I. Arhipova, “Cause analysis of students’ dropout rate in higher education study program,” *Procedia-Social Behav. Sci.*, vol. 109, pp. 1282–1286,

2014.

- [5] S. B. Kotsiantis, “Use of machine learning techniques for educational proposes: a decision support system for forecasting students’ grades,” *Artif. Intell. Rev.*, vol. 37, no. 4, pp. 331–344, 2012.
- [6] Banco Mundial, “Graduarse: solo la mitad lo logra en América Latina,” <https://www.bancomundial.org/es/news/feature/2017/05/17/graduating-only-half-of-latin-american-students-manage-to-do-so>, 2017. .
- [7] Ministerio Educación, “Estrategias para la Permanencia en Educación Superior: Experiencias Significativas,” [https://www.mineducacion.gov.co/1759/articles-356276\\_recurso.pdf](https://www.mineducacion.gov.co/1759/articles-356276_recurso.pdf), 2015. .
- [8] D. Yang, D. Adamson, T. Sinha, and C. P. Rose, “‘Turn on, Tune in, Drop out’: Anticipating Student Dropouts in Massive Open Online Courses Investigating Individual Learner Behavior in Educational Technologies View project Investigating the Impact of Social and Interpersonal Factors on Learning View project ‘Turn on, Tune in, Drop out’: Anticipating student dropouts in Massive Open Online Courses,” 2013.
- [9] K. B. Eckert and R. Suénaga, “Análisis de deserción-permanencia de estudiantes universitarios utilizando técnica de clasificación en miner\’\ia de datos,” *Form. Univ.*, vol. 8, no. 5, pp. 3–12, 2015.
- [10] M. Tan and P. Shao, “Prediction of student dropout in e-Learning program through the use of machine learning method.,” *Int. J. Emerg. Technol. Learn.*, vol. 10, no. 1, 2015.
- [11] K. Coussement, M. Phan, A. De Caigny, D. F. Benoit, and A. Raes, “Predicting student dropout in subscription-based online learning environments: The beneficial impact of the logit leaf model,” *Decis. Support Syst.*, vol. 135, p. 113325, 2020.
- [12] B. Cuji, W. Gavilanes, and R. SANCHEZ, “Modelo predictivo de deserción estudiantil basado en arboles de decisión,” *Espacios*, vol. 38, no. 55, p. 17, 2017.
- [13] M. Barramuño, C. Meza-Narváez, and G. Gálvez-Garc\’\ia, “Prediction of student attrition risk using machine learning,” *J. Appl. Res. High. Educ.*, 2021.
- [14] M. Sol\’\is, T. Moreira, R. Gonzalez, T. Fernandez, and M. Hernandez, “Perspectives to predict dropout in university students with machine learning,” in *2018 IEEE International Work Conference on Bioinspired Intelligence (IWOBI)*, 2018, pp. 1–6.
- [15] P. E. Ram\’\irez and E. E. Grandón, “Predicción de la Deserción Académica en una Universidad Pública Chilena a través de la Clasificación basada en Árboles de Decisión con Parámetros Optimizados,” *Form. Univ.*, vol. 11, no. 3, pp. 3–10, 2018.
- [16] A. Mujumdar and V. Vaidehi, “Diabetes prediction using machine learning algorithms,” *Procedia Comput. Sci.*, vol. 165, pp. 292–299, 2019.

- [17] NumFOCUS sponsored project., “Pandas,” 2021. [Online]. Available: <https://pandas.pydata.org/>. [Accessed: 02-Sep-2021].