# PROTOTIPO PARA GESTIONAR BAJAS ACADÉMICAS EN ENTORNOS UNIVERSITARIOS

## *PROTOTYPE TO MANAGE ACADEMIC WITHDRAWALS IN UNIVERSITY ENVIRONMENTS*

***Aldonso Becerra Sánchez***
Universidad Autónoma de Zacatecas, México
*a7donso@uaz.edu.mx*

***Viridiana Murillo Leija***
Universidad Autónoma de Zacatecas, México
*viridiml50@gmail.com*

***Nancy Delgado Salazar***
Universidad Autónoma de Zacatecas, México
*nancydesal@uaz.edu.mx*

***Gustavo Zepeda Valles***
Universidad Autónoma de Zacatecas, México
*gzepeda@uaz.edu.mx*

***Santiago Esparza Guerrero***
Universidad Autónoma de Zacatecas, México
*chago@uaz.edu.mx*

**Recepción:** 17/marzo/2023 **Aceptación:** 17/abril/2023

## Resumen

La deserción escolar es un problema social que afecta a las universidades de México, generando una eficiencia terminal alrededor del 50% [Urbina-Nájera, 2020]. En muchas escuelas se realizan seguimientos de las bajas académicas para conocer los motivos que llevan a los alumnos a tomar la decisión de desertar, promoviendo así su retención. El objetivo de este trabajo es desarrollar un prototipo para automatizar el proceso de control de bajas académicas dentro de un entorno universitario, a la vez que se permite analizar datos de perfiles de ingreso para generar la predicción de posibles deserciones. Empleando las metodologías Prototyping y KDD, en registros de bajas académicas en la Unidad Académica de Ingeniería Eléctrica, se obtuvieron valores de precisión del 92.64% y exactitud del 96.34%. Los resultados mostraron que este prototipo es capaz de generar

información para sustentar la toma de decisiones dentro de entornos universitarios, encaminando a mejorar su eficiencia terminal.

**Palabras Clave:** Análisis de datos educativos, control de bajas académicas, deserción escolar, eficiencia terminal, machine learning.

## *Abstract*

*School desertion is a social problem that affects universities in Mexico, generating a terminal efficiency of around 50% [Urbina-Nájera, 2020]. In many schools, academic withdrawals are monitored to find out the reasons that lead students to make the decision to drop out, thus promoting their retention. The aim of this work is to develop a prototype to automate the process of controlling academic withdrawals within a university environment, while allowing the analysis of data from admission profiles to generate possible desertions predictions. Using the Prototyping and KDD methodologies, in records of academic withdrawals in the Unidad Académica de Ingeniería Eléctrica, precision values of 92.64% and accuracy of 96.34% were obtained. The results showed that this prototype can generate information to support decision-making within university environments, aiming to improve its terminal efficiency.*

***Keywords*****:** *Analysis of educational data, control of academic withdrawals, machine learning, school desertion, terminal efficiency.*

## 1. Introducción

La deserción escolar universitaria en México es un problema latente, el cual se ha estudiado desde diversas perspectivas, llegando incluso a encontrar ciertas variables ya sea en las personas o el entorno que las propician [Rochin, 2021]. De acuerdo con datos reportados por el Banco Mundial, la mitad de los estudiantes entre 25 y 29 años terminan sus estudios universitarios; mientras que el 50% de los abandonos ocurre en el primer año de estudios [Urbina-Nájera, 2020]. Se debe entender que la deserción escolar es una opción del estudiante, influenciada por factores externos o internos, la cual comprende los casos en los que el estudiante abandona la carrera universitaria en el que realiza sus estudios para continuarla en

otra institución, o cambiar de carrera ya sea en la misma u otra institución, además del caso de abandono definitivo de las clases [Rochin, 2021]. Se destaca que al hablar de deserción no se toman en cuenta factores como la mortalidad estudiantil y el retiro forzoso (cuando la institución lo da de baja) [Paramo, 1999].

Es claro que las universidades se esfuerzan por lidiar con el abandono estudiantil, pues más allá de generar conflictos administrativos, la deserción afecta la reputación de la institución [Rochin, 2021]. Para evaluar la eficiencia de las instituciones de educación superior (IES) se usa la Eficiencia Terminal (ET), medida que la UNESCO ha definido como el grado en el cual un sistema educativo consigue optimizar la relación inversión-resultado en la educación [López, 2008]. Se mide entonces cuántos estudiantes egresados obtiene una institución de educación superior en relación con los alumnos que ingresaron.

Estudios realizados con información de la carrera se han centrado en encontrar el perfil del desertor, así como los factores externos que están presentes en su decisión de abandonar los estudios. Al ser la deserción también un problema social, se le considera como “el problema de las mil causas” dados los factores que la originan, como de tipo vocacional, académico, físicos, económicos, familiares, sociales, institucionales, pedagógicos, de transporte, laborales, de salud, de creencias, entre otros [Romero, 2012], [Urbina-Nájera, 2020].

Diversas IES han optado por realizar análisis de datos de manera particular para evaluar sus resultados en cuanto a eficiencia terminal. Esto permite ampliar el conocimiento del contexto en el que se encuentra la institución. De manera especial se han orientado hacia la predicción con el fin de mejorar las técnicas de retención estudiantil, así como generar estrategias de prevención contra el fracaso escolar [Romero, 2012], [Ruiz, 2016], [Ponce, 2016]. En este sentido, algunas universidades han realizado estudios para determinar causas de abandono escolar [Rodríguez, 2011], otras en cambio han buscado predecir la posibilidad de que un alumno se de baja [Valero, 2010], mientras que también se buscan patrones para prevenir la deserción universitaria [Urbina-Nájera, 2020]. Por otro lado, existen estudios para analizar la relación entre el desempeño académico de los estudiantes en el primer año de universidad y los instrumentos de selección de aspirantes [Vergara-Díaz,

2017]. Tal que también se han realizado análisis encaminados a la identificación de los perfiles asociados a la deserción universitaria de aspirantes [Chaves, 2016].

Es decir, el fenómeno no puede ser explicado solamente por las limitaciones socioeconómicas de las familias, ni por la falta de integración del estudiante al ambiente universitario, como tampoco exclusivamente por el desempeño escolar, sino que se define por varios de ellos de modo compuesto. Las primeras investigaciones para identificar causas de la deserción se realizaban por medio del análisis de datos estadísticos, basados por ejemplo en análisis correlacionales y en la comparación de modelos de regresión. No obstante, en los últimos 10 años incrementaron los estudios que aplican técnicas novedosas en el campo de la educación, como son diferentes métodos basados en la minería de datos, los cuales tienen como principal ventaja su alta capacidad predictiva [Bernardo, 2015].

Estudios de este tipo suelen centrarse en analizar factores que influyen en los estudiantes a lo largo de su desempeño dentro de la carrera, sin embargo, también es posible trabajar con factores que se presentan desde una etapa más temprana como el perfil de ingreso de los estudiantes [Dieser, 2017]. Normalmente las IES realizan un proceso de selección de aspirantes, en el cual se incluyen resultados de exámenes académicos, diversos tipos de pruebas psicométricas y vocacionales, así como datos académicos provenientes del bachillerato. Comenzar a analizar datos de los estudiantes desde el proceso de admisión es una opción viable, tomando en cuenta que entre más pronto se encuentran las posibles causas de un problema, se pueden hallar un mayor número de soluciones a éste, incursionando en acciones preventivas [Sánchez, 2014].

Con el constante incremento de deserción en las carreras universitarias, todo el proceso de registro y seguimiento de bajas se vuelve más pesado cada año, pues es más información la que debe ser procesada. Desde esta perspectiva, el objetivo de este trabajo es plantear una solución que automatice la mayor parte de los procesos de baja académica, que efectúe un análisis de los factores en las deserciones y además funja como medio para la implementación de acciones de tomas de decisión preventivas a la deserción. Como caso de estudio se tomó de referencia la información del área de ingeniería de la Universidad Autónoma de

Zacatecas (UAZ), la cual, durante el proceso de selección de aspirantes, obtiene datos derivados de preparatoria, de la prueba psicométrica y el examen CENEVAL, creando un perfil académico y vocacional de cada posible aspirante.

## 2. Métodos

Para el desarrollo de este prototipo se optó por una metodología ágil conocida como Prototyping, ver figura 1, la cual se define como un método de desarrollo de sistemas en el que un prototipo es construido, probado y luego reinventado según sea necesario.

El proceso se repite iterativamente hasta que finalmente se logre un prototipo aceptable a partir del cual el sistema o producto completo puede ser ahora desarrollado [Weitzenfeld, 2008], [Sommerville, 2011], [Becerra, 2018]. Este modelo de prototipo permite obtener retroalimentación de algunos de los interesados en el sistema en etapas tempranas, haciendo que funcionalidades del sistema fueran descartadas o suprimidas mientras que nuevas funcionalidades y necesidades fueran agregadas conforme van siendo requeridas.

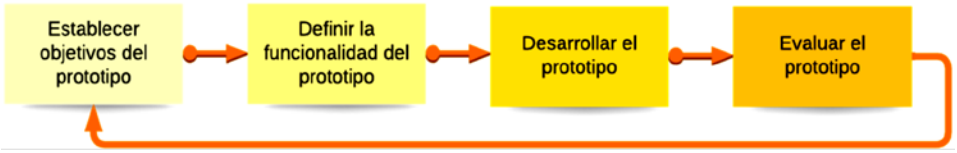



Figura 1 Ciclo de vida de Prototyping.

### Establecimiento de los objetivos del prototipo

Para llevar a cabo el prototipo, se establecen objetivos extraídos del caso de estudio definido dentro del área de las ingenierías de la UAZ (pudiendo aplicar a cualquier área), siendo un apoyo en la evaluación y mejora continua del prototipo:

- Registrar solicitudes y bajas académicas.
- Análisis estadístico de los datos de deserción.
- Predecir la posibilidad de que un alumno deserte o no de la carrera.
- Predecir la posibilidad de que un alumno deserte por motivos vocacionales.
- Generar y exportar reportes.
- Autenticación de usuarios.

## Definición de la funcionalidad del prototipo

El prototipo ofrece la funcionalidad necesaria para coadyuvar en un proceso eficiente de solicitud y control de bajas académicas. También procesará la información recabada para presentarla en reportes que cuenten con la posibilidad de ser exportados en distintos formatos, los cuales son usados en la entrega constante de informes. Se usará análisis de datos para predecir la probabilidad de que un alumno se dé de baja de la carrera con base en su información de ingreso. Para poder permitir todo este proceso, el prototipo será restringido mediante el acceso controlado de usuarios. Este prototipo propone ser un apoyo a los encargados del seguimiento de las bajas de la UAZ con la organización y eficiente uso de la información de la que se tiene registro. Se espera que sea utilizado como herramienta que oriente en la identificación de situaciones referentes a la deserción escolar y sea un antecedente para la planeación de estrategias de retención y prevención del fracaso académico.

La figura 2 muestra el diagrama de casos de uso que comprende la funcionalidad del prototipo según los usuarios que lo pueden emplear.

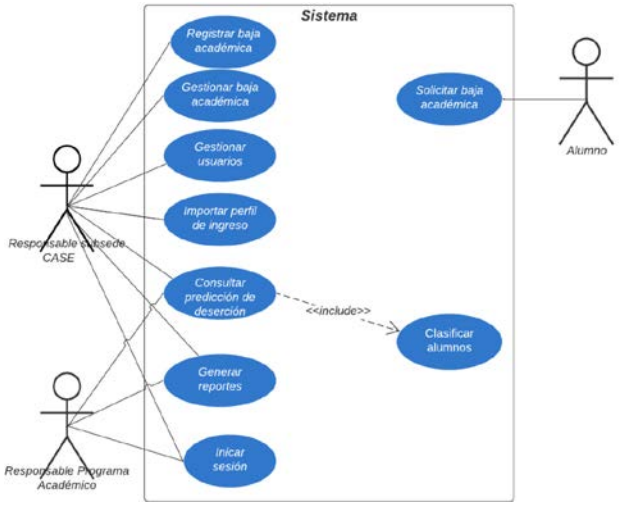



Figura 2 Diagrama de casos de uso.

En donde el usuario CASE (Centro de Aprendizaje y Servicios Estudiantiles) pertenece al departamento que brinda programas académicos y asistenciales a la comunidad universitaria para mejorar su rendimiento académico, buscando abatir el

rezago, reprobación y deserción. Mientras que el Responsable de Programa (RP) corresponde a un encargado de carrera en el área. En donde el caso de uso *importar perfil de ingreso* añade información del perfil de los aspirantes en un archivo de Excel; y en donde *realizar solicitud de baja, registrar baja académica y gestionar baja académica* permiten dar seguimiento al proceso de baja académica. Mientras que se tienen casos de uso para *generar reportes* y *gestionar usuarios* que emplean el sistema, en el cual se ocupa el proceso de *iniciar sesión.* Asimismo, se tiene el proceso para *consultar predicción de baja académica.*

## Desarrollo del prototipo

El prototipo está compuesto por dos módulos principales (módulo A para el control de bajas y módulo B para el análisis de datos), que a la vez cuentan con sub-módulos que alojan funcionalidades específicas, ver figura 3.

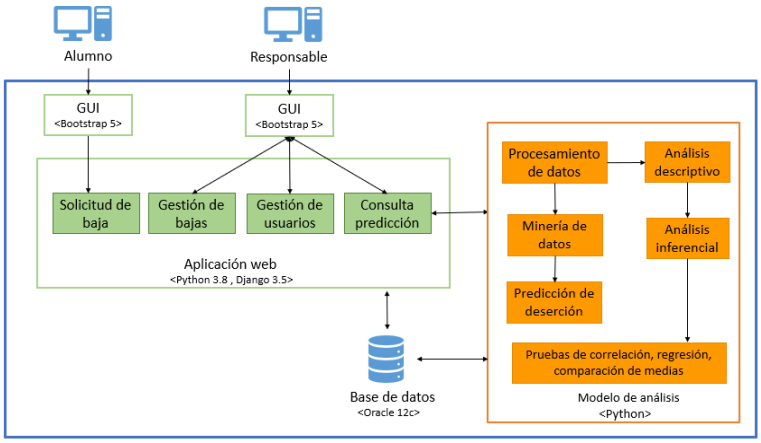



Figura 3 Arquitectura general del prototipo propuesto.

En el módulo A, el alumno se conecta a la aplicación para realizar una solicitud de baja. Las labores de control asignadas a un administrador le son disponibles mediante un inicio de sesión con su cuenta de usuario. Una vez ingresado a la aplicación, el responsable puede hacerse cargo de la gestión de las bajas escolares revisando las solicitudes, y si es el caso, procesarlas como una baja. El usuario administrador es capaz de crear y eliminar usuarios, así como otorgarles permisos dentro de la aplicación. Para la consulta de predicciones, el módulo de consultas se

conecta al módulo B, en este módulo se realiza el entrenamiento de modelo, los análisis de datos y las predicciones.

**Módulo A: control de bajas**

El proceso de desarrollo es iterativo incremental siguiendo los elementos definidos en el diseño, creando una versión del prototipo que evoluciona según el avance del desarrollo tomando en cuenta el cumplimiento de los objetivos y la retroalimentación del usuario final. El desarrollo del prototipo basado en una aplicación web fue hecho con el framework Django 3.2.7 haciendo uso de Python 3.8.10. Este framework trabaja sobre una arquitectura MVT (Model-View-Template) para proponer la separación clara de los tipos de lógica al programar. Así, se crea una estructura específica para el modelado de datos, así como para las operaciones que se deban aplicar a estos y su posterior presentación en la aplicación, lo que facilita el mantenimiento y la escalabilidad del sistema. La base de datos que se emplea es relacional y está desarrollada en Oracle 12c, donde se almacena la información escolar de los estudiantes, la pertenencia de un alumno a determinada carrera, el proceso de solicitud y registro de baja, así como sus tipos.

**Módulo B: análisis de datos**

Para el desarrollo de un modelo de predicción se sigue la metodología KDD (Knowledge Discovery in Databases), donde una de sus premisas mayores conlleva el descubrimiento del conocimiento usando técnicas de aprendizaje inteligente al examinar datos a través de procesos automatizados. KDD es un proceso iterativo que involucra decisiones que deben ser tomadas por el usuario, y se estructura en las siguientes etapas [Molina, 2006]:

- Comprensión del dominio de la aplicación.
- Creación, limpieza y pre-procesamiento del conjunto.
- Elegir la tarea de minería de datos, así como la aplicación del algoritmo apropiado.
- Interpretación de los patrones encontrados.
- Consolidación del conocimiento descubierto.

Los pasos resumidos comprenden:

- La información pre-seleccionada incluye bases de datos de la Unidad Académica de Ingeniería Eléctrica (2016-2019), correspondientes a los instrumentos de perfil de ingreso que se aplican a los aspirantes a entrar a los Programas Académicos de esta Unidad.
- Se deciden qué factores de riesgo tomar en cuenta, aplicando un proceso de selección, recolección y preparación de datos en el análisis estadístico, el cual incluye el uso de la librería Pandas 1.3.4 para Python.
- Se analizaron 170 registros de baja dados entre el 2016 al 2019 en la Unidad Académica de Ingeniería Eléctrica, además del conjunto de datos del perfil de ingreso. Un 86.74% de los alumnos dados de baja indicó que continuará con sus estudios en diferente carrera o universidad, el 10.75% dijo no continuará con sus estudios y 2.51% desconoce si continuará su formación académica, ver figura 4.

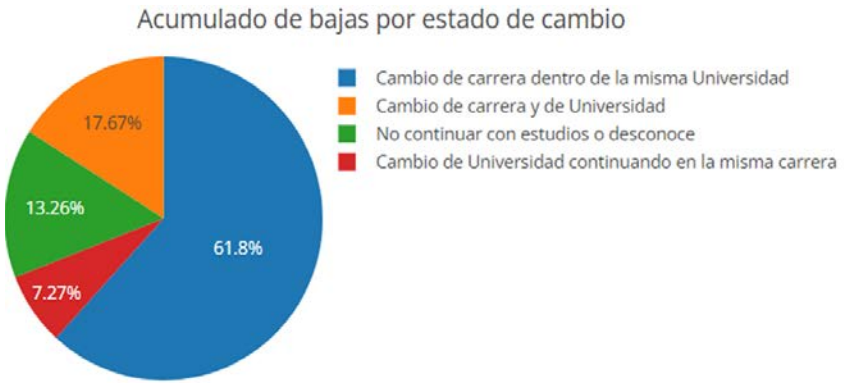



Figura 4 Estado del cambio de alumnos en la UAIE en el seguimiento de bajas.

Con esto resaltamos que la mayor parte de las deserciones registradas no van orientadas hacia el abandono escolar, sino que el estudiante decide continuar su educación en un ambiente o área del conocimiento distinta. En ese sentido, el 58.23% de las bajas se dan durante el primer año de la carrera, otro 28.82% del total de bajas se registró durante el segundo año de la carrera. Dado que la gran mayoría de bajas se da durante los primeros semestres de la carrera, es importante identificar el riesgo de deserción de

manera oportuna a fin de que las estrategias de retención y prevención sean implementadas.

- Por otro lado, el 48.83% de los desertores obtuvieron en los instrumentos de ingreso un área vocacional distinta a "Ingenierías y Tecnología", 58.6% de los desertores tienen puntaje CENEVAL por debajo de la media; también 90% de los desertores tienen su promedio de bachillerato por debajo de la media. Por ende, incluso desde antes de que el estudiante entre a la carrera, ya tiene un perfil propenso a la deserción. Además, se encontró que el 51.18% de las bajas indican la “vocación” como motivo de deserción, seguido de "reprobación" y "dificultad en las materias" con el 24% cada una. A esto le siguen los motivos "desempeño académico inadecuado" y "perfil de la carrera" con el 18.5 y 17.4% respectivamente, ver figura 5.

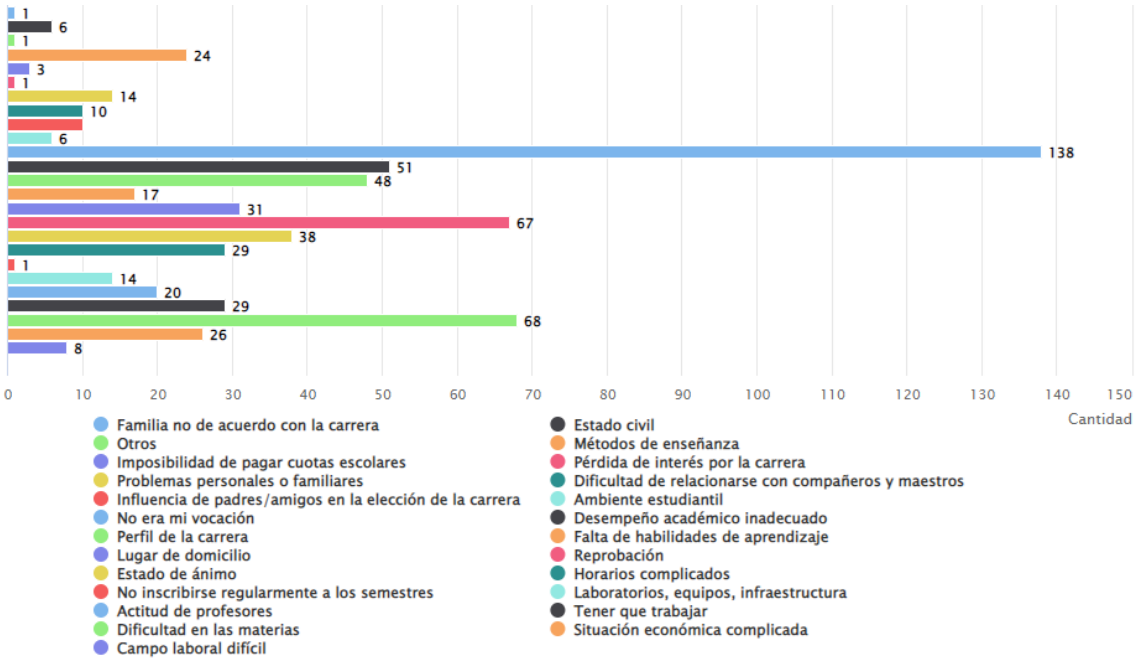



Figura 5 Motivos de baja de alumnos de la UAIE.

Es importante recalcar que normalmente el estudiante indica más de un motivo por el cual se da de baja, siendo el más recurrente "no era mi vocación", acompañado de uno o varios factores disponibles en la lista de motivos. Estos resultados muestran que la deserción escolar es un fenómeno

multifactorial, donde predomina en este objeto de estudio el ámbito académico, relacionado directamente con el área de dominio de la carrera.

- En cuanto al análisis inferencial, se pueden definir dos objetivos que guían el aspecto de análisis a futuro de alumnos en cuanto a las bajas:
  - ✓ Probabilidad de un alumno de darse de baja según su perfil de ingreso.
  - ✓ Probabilidad de un alumno de darse de baja por el motivo "vocación".

  Para cada objetivo de predicción se utilizó un conjunto de datos distinto. Para la predicción de un alumno de darse de baja se usó como target el valor binario de "Baja_Si", que indica si el registro corresponde a un alumno que se dio de baja. Este conjunto consta de 1316 registros, de los cuales 153 pertenecen a alumnos que se dieron de baja de la carrera. Con una representación del 11.6%, es necesario realizar un balanceo de datos con el fin de equilibrar las entradas que recibirá el modelo [Riquelme, 2012]. En este caso, como la cantidad de registros es pequeña, es más propicio realizar un sobremuestreo, que consiste en aumentar el número de datos que contiene la clase minoritaria [Mera, 2015]. Para esto se utilizó la librería imblearn, con el método "Random Over Sampled", la cual replica registros de la clase minoritaria hasta conseguir el conjunto de datos equilibrado. Después de este procedimiento el dataset cuenta con 2326 registros.
- Para la predicción de bajas por motivo de vocación se contempla un target que representa los valores binarios indicando si la baja fue por este motivo o no (otro factor). Estos registros son 162, 51% de los cuales corresponden a las bajas por vocación.
- En el dataset, los atributos categóricos fueron sometidos a tratamiento para pasarlos a una escala numérica. Se transformaron a columnas binarias los datos categóricos que no tienen una jerarquía, mientras que a escala de valores en la misma columna los que sí cuentan con una [Valdovinos, 2006]. También se realizó el escalamiento de los datos haciendo uso de las librerías Pandas y Numpy para Python, ya que disponen de características distintas en magnitudes, unidades y rangos [Hernández, 2008].

- Las técnicas de minería de datos (rama del análisis de datos) intentan obtener patrones o modelos a partir de los datos con los que se trabaja [García, 2018]. Los modelos de predicción utilizados son árboles de decisión, bosque aleatorio, y regresión logística; para su implementación se usaron los modelos proporcionados por la librería Scikit-learn: LogisticRegression, DecisionTreeRegressor y RandomForestRegressor respectivamente. Cada conjunto de datos fue separado para entrenamiento y prueba usando la librería sklearn. Para cada modelo fue entrenado variando sus hiperparámetros con el fin de mejorar el rendimiento. En el caso de los árboles de decisión y bosque aleatorio se trabajó con los parámetros de profundidad del árbol y número de árboles dentro del bosque.

### Evaluación del prototipo

En esta sección se presentan las iteraciones que sufre el prototipo con el fin de comprobar que el desarrollo fue acorde a los objetivos establecidos, ver tabla 1. Una vez finalizados los módulos para el control de bajas académicas, así como la consulta de predicciones de deserción escolar por medio del perfil de ingreso, fueron probados y validados por los usuarios del CASE. Cumpliéndose con los 6 objetivos establecidos para el prototipo según sus propios requerimientos.

Tabla 1 Evaluación de objetivos cumplidos en el prototipo.

| Iteración | Objetivo |
|---|---|
| 1 | Registrar solicitudes y bajas académicas. |
| 2 | Autenticación de usuarios. |
| 3 | Generar y exportar reportes. |
| 4 | Análisis estadístico de los datos de deserción. |
| 5 | Predecir la posibilidad de que un alumno deserte o no de la carrera. |
| 6 | Predecir la posibilidad de que un alumno deserte por motivos vocacionales. |

## 3. Resultados

### Módulo A: control de bajas

La figura 6 corresponde a la interfaz vista por un usuario "Alumno", cuya única interacción con la aplicación es para solicitar su baja académica, y en el cual se

ingresan los motivos de baja que él considera, que deben ser revisados por el responsable del CASE antes de continuar con el proceso de baja.

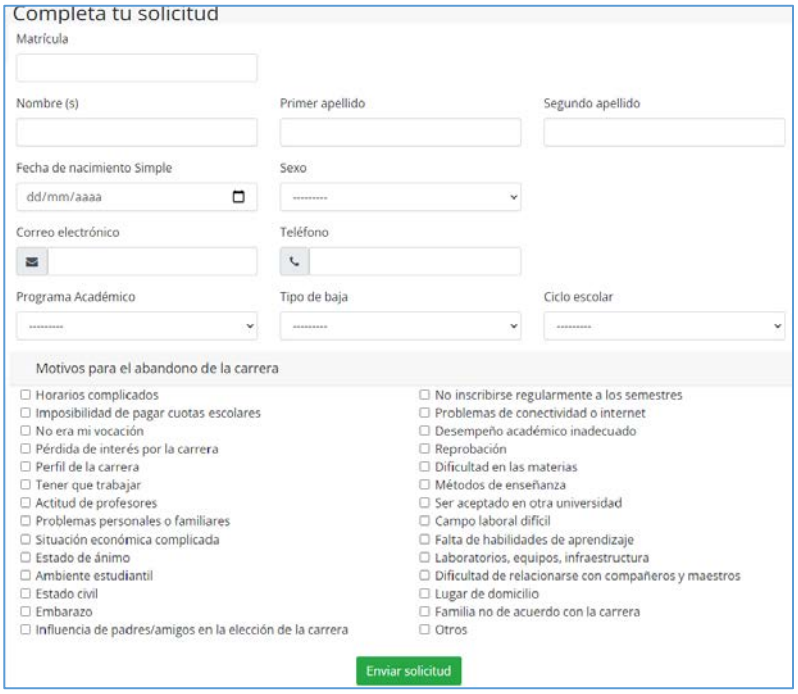
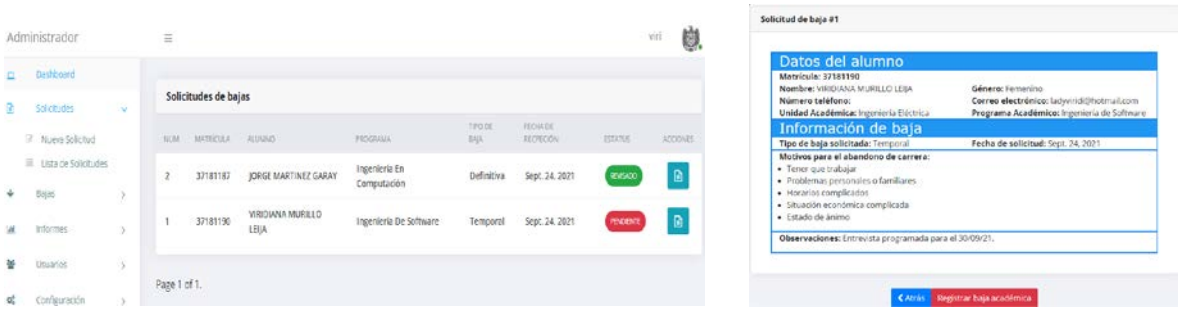



Figura 6 Solicitud de baja académica.

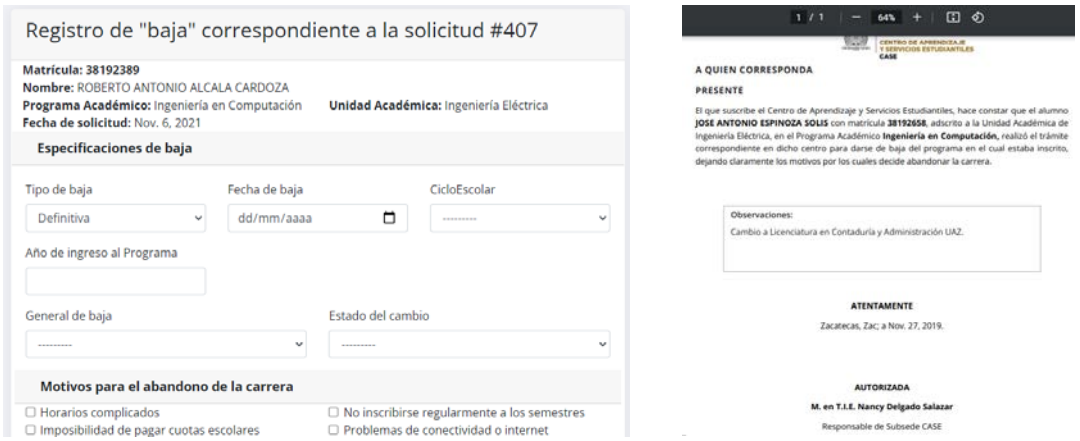
El usuario administrador "responsable de subsede de CASE" tiene la facultad de ver todas las solicitudes que se han registrado hasta la fecha, ver figura 7a. En el detalle se encontrará la ficha de solicitud, que contiene datos académicos y de contacto del alumno, además de los motivos especificados de baja; a partir de ahí es posible registrar una baja académica, ver figura 7b.




*a) Vista de solicitudes* *b) Ficha de solicitud*

Figura 7 Registro de bajas académicas.

Después de realizada la entrevista con el estudiante, en caso de que no se haya logrado la retención, el responsable de subsede debe registrar la baja en el formulario correspondiente, ver figura 8a, el cual importa los datos iniciales de la solicitud. Finalizado este proceso, la aplicación genera en PDF el oficio a nombre del CASE para hacer válida la baja ante la Universidad, ver figura 8b.




*a) Formulario de baja* *b) PDF del oficio de baja*

Figura 8 Capitalización del oficio de baja.

Para la visualización rápida de la información se tiene la sección de gráficos porcentuales de bajas por Programa Académico, bajas por ciclo escolar, tipo de baja, general de baja, baja por sexo, motivos de baja, entre otros, ver figura 9.

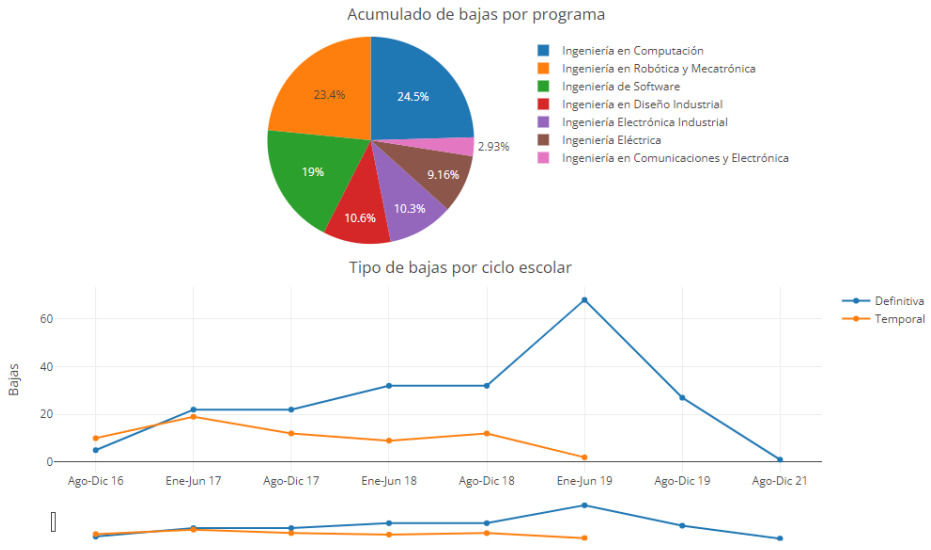



Figura 9 Ejemplos de gráficas generadas por el prototipo.

**Módulo B: análisis de datos**

Dentro del módulo de análisis de datos se tiene la funcionalidad que permite importar la información de los perfiles de ingreso de los alumnos, ver figura 10a. Estos datos se cargan desde archivos de Excel con formato especificado por una plantilla compatible. Posteriormente estos datos son pre-procesados de manera interna (de acuerdo a KDD), tal que con ellos se definen los pasos a seguir en el módulo correspondiente. En la sección de consulta de predicciones se muestra la información para identificar un alumno, ver figura 10b, así como el resultado de la predicción que la aplicación hizo para él según la información cargada.

*a) Importación de perfil de ingreso* *b) Resultados de predicción*

Figura 10 Predicciones de deserción.

De acuerdo con la tabla 2, el modelo de bosque aleatorio fue el que mejor rendimiento dio en los resultados de las predicciones.

Tabla 2 Métricas obtenidas en los modelos de predicción de bajas en fase de pruebas.

| Modelo | Precisión | Exactitud | Sensibilidad | F1 | Matriz de Confusión |
|---|---|---|---|---|---|
| Regresión logística | 0.6597 | 0.7060 | 0.7429 | 0.6989 | [170, 82] [55, 159] |
| Árbol de decisión (profundidad de 20) | 0.8359 | 0.9098 | 1 | 0.9106 | [210, 42] [0 , 214] |
| Bosque Aleatorio (400 árboles, profundidad de 20) | 0.9264 | 0.9635 | 1 | 0.9617 | [235, 17] [0 , 214] |

En la matriz de confusión se aprecia que 449 casos de prueba fueron clasificados correctamente de un total de 466, es decir una exactitud de 96.3%. Con base en estos datos se observa que el modelo tiende a cometer mayor error de predicción

hacia los falsos positivos, mientras que para predecir a los estudiantes que no desertan, así como los que se dan de baja, el algoritmo tiene un mejor desempeño. Cuando el algoritmo predice estudiantes que desertan, así como los que no, clasifica el 96.3% de las muestras totales correctamente. Eso significa que el clasificador obtiene una buena cantidad de predicciones positivas (bajas y permanencias reales).

Por otro lado, el modelo obtiene una precisión en la que cuando predice la baja de un alumno, el modelo acierta el 92.64% de las veces. En cuanto a la sensibilidad, el modelo aprende a identificar correctamente a los alumnos que se dan de baja en un 100%. Al usar F1-score, importan tanto los alumnos que se dan de baja como los que no, por lo que esta métrica logra el equilibrio. Es así como el modelo logra diferenciar las clases con una confiabilidad del 96.17%.

De igual manera, para las predicciones basadas en el factor que indica si el alumno se dio de baja por factores vocacionales, la tabla 3 muestran los mejores resultados. En esta ocasión el modelo de árbol de decisión es el algoritmo que mejores resultados devuelve. En la matriz de confusión se aprecia que 25 casos fueron clasificados correctamente, en un total de 33 casos de prueba, es decir, una exactitud de 75.75%. Eso significa de manera general que este algoritmo está haciendo un buen trabajo en la identificación de la baja por motivos vocacionales, ya que obtiene una cantidad mayor de predicciones positivas acertadas.

Tabla 3 Métricas de los modelos de predicción de bajas por vocación en fase de pruebas.

| **Modelo** | **Precisión** | **Exactitud** | **Sensibilidad** | **F1** | **Matriz de Confusión** |
|---|---|---|---|---|---|
| Regresión logística | 0.6366 | 0.66 | 0.5 | 0.56 | [15, 4]<br>[7, 7] |
| Árbol de decisión (profundidad de 5) | 0.7826 | 0.7575 | 0.8117 | 0.7994 | [18, 5]<br>[3 , 7] |
| Bosque Aleatorio (100 árboles, profundidad de 15) | 0.666 | 0.666 | 0.42 | 0.52 | [16, 3]<br>[8 , 5] |

Por otro lado, el modelo obtiene una precisión que implica que, al predecir una baja por vocación, el algoritmo acierta el 78.26% de las veces. Además, la sensibilidad

obtenida muestra que aprende a identificar correctamente a los alumnos que se dan de baja en un 81.17%. Mientras que para la puntuación de F1, el modelo identifica correctamente la clase el 79.94% de los casos.

## 4. Discusión

Los resultados demuestran que el modelo desarrollado puede predecir con precisión el riesgo de deserción de estudiantes basándose en el perfil de ingreso de los aspirantes. Además, es capaz de generalizar para resolver situaciones no vistas, ya que no memorizaron los datos de entrenamiento, es decir, el algoritmo es capaz de aplicar lo aprendido con nuevos datos. La precisión obtenida para los modelos de árbol de decisión y bosque aleatorio indica que para este contexto los árboles son el mejor algoritmo para datasets pequeños y equilibrados.

Tomando en cuenta que el 50% de las bajas académicas se registran en los primeros semestres de la Universidad, la pronta identificación de los alumnos que presenten un riesgo de abandonar la carrera, incluso cuando aún son aspirantes a ingresar, supone beneficios para las instituciones de educación superior. Entre los cuales se encuentran la posibilidad de ampliar los filtros del proceso de admisión, considerando el riesgo de deserción que presenta un aspirante por su perfil de ingreso. Además, se puede realizar la implementación de estrategias de prevención al abandono escolar una vez que los alumnos ya están inscritos.

## 5. Conclusiones

La deserción escolar, un problema constante en las instituciones de educación superior del país, afecta a la Unidad Académica de Ingeniería Eléctrica de la UAZ, quien, con el propósito de mejorar su eficiencia terminal, ha implementado medidas para gestionar las bajas académicas. Para automatizar estos procesos, se trabajó con la metodología Prototyping y KDD en el desarrollo de una aplicación web por iteraciones, en donde se manejan datos libres de inconsistencias, a su vez permite que la información sea transparente y fácil de compartir.

Para el análisis de datos se utilizó la información de las evaluaciones aplicadas a los aspirantes como instrumentos de ingreso, así como los registros de bajas

académicas de la Unidad en el periodo 2016-2021. Las predicciones para baja realizadas por la aplicación tienen una precisión del 92.64%, mientras que las de baja por motivos vocacionales son del 78.26%, las cuales han sido desarrolladas con un modelo de bosque aleatorio y árbol de decisión respectivamente. Estos resultados se contrastan, en contextos diferentes, con algunos del estado del arte [Valero, 2010], donde el cálculo de la métrica de precisión para predecir la deserción usando técnicas de minería de datos obtuvo un 98.98% como máximo.

Estos resultados indican que se cuenta con una herramienta confiable para el pronóstico de deserción de estudiantes basados en su perfil de ingreso. Además, se puede tener en cuenta el contexto del estudiante que ingresa para entender la manera en la que se desenvolverá a lo largo de la carrera, aplicando en ello estrategias que impulsen el desarrollo educativo de la Universidad.

Como trabajo futuro se incluye el crecimiento de los datos a través de los años, donde sería posible probar otras técnicas de análisis de datos. Otro de las adiciones posibles radica en la identificación más profunda de los factores de deserción, en la que, considerando aspectos éticos, se pueden buscar los mecanismos para que los alumnos puedan concluir sus estudios.

## 6. Bibliografía y Referencias

[1] Becerra, A., Zepeda, G., Pérez, A. I., & Ramírez-García, U. Learning Content Management Software Personalized for a University Environment. Pist. Educ., Vol. 40, No. 130, 347-362, 2018.

[2] Bernardo, A., Cerezo, R., Rodríguez-Muñiz, L. J., Núñez, J. C., Tuero, E., & Esteban, M. Predicción del abandono universitario: variables explicativas y medidas de prevención. Rev. Fuentes, Vol. 16, 63-84, 2015.

[3] García, J., Molina J., Berlanga, A., Bustamante, A., Padilla, W., & Patricio, M. Ciencia de datos: técnicas analíticas y aprendizaje estadístico. Bogotá: Alfaomega, 2018.

[4] Chaves, L. J., & José, P. A. C. Aplicativo web para el análisis, selección y admisión de aspirantes al programa de ingeniería de sistemas de la universidad Cundinamarca en la extensión Facatativá utilizando modelos

predictivos de minería de datos. J. Chem. Inf. Model., Vol. 53, No. 9, 1689-1699, 2016.

[5] Dieser, M. P., Cavero, L. V., Martín, C., Schlaps, E., Titionik, D., & Wagner, L. Evaluación de Técnicas de Clasificación para Predecir el Rendimiento Académico de Ingresantes a la Universidad en Temas de Matemática-Memorias XIX Work. Investig. en Ciencias la Comput. (WICC 2017). Buenos Aires, Argentina. Abril, 2017.

[6] Hernández, C., & Rodríguez, J. Preprocesamiento de datos estructurados. Vínculos, Vol. 4, No. 2, 27-48, 2008.

[7] López, A., Albiter, A., & Ramírez, L. Eficiencia terminal en la educación superior, la necesidad de un nuevo paradigma. Rev. la Educ. Super., Vol. 37, No. 146, 135-151, 2008.

[8] Mera, C., & Arrieta, J. Estudio Comparativo de Técnicas de Balanceo de Datos en el Aprendizaje de Múltiples Instancias. LACNEM2015 - Lat. Am. Conf. Networked Electron. Media, Vol. 6, 2015.

[9] Molina, J. M., & García, J. Técnicas de análisis de datos aplicaciones prácticas utilizando microsoft excel y weka. 2006.

[10] Paramo, G., & Correa, C. Deserción estudiantil universitaria. Conceptualización. Rev. Univ. EAFIT, Vol. 35, 65-78, 1999.

[11] Ponce, J. C., & Toscano de la Torre, B. A. Análisis de la Eficiencia Terminal Basado en Técnicas de Minería de Datos para Identificar los Factores que Determinan su Logro. En: Tendencia de las Tecnologías de la Información y Comunicaciones. Alfa Omega, 309-322, 2016.

[12] Riquelme, J., & Ruíz, R. Minería de Datos: Conceptos y Tendencias. Rev. Iberoam. Intel. Artif., Vol. 10, No. 29,11-18, 2012.

[13] Rodríguez, J., & Hernández, J. M. La deserción escolar universitaria en México. La experiencia de la Universidad Autónoma Metropolitana, Campus Iztapalapa. Actual. Investig. en Educ., Vol. 8, No. 1, 2011.

[14] Romero, C., Márquez, C., & Ventura, S. Predicción del Fracaso Escolar Mediante Técnicas de Minería de Datos. IEEE-Rita, Vol. 7, No. 3, 109-117, 2012.

[15] Rochin, F. Deserción escolar en la educación superior en México: revisión de literatura. RIDE. Rev. Iberoam. Investig. Desarro., Vol. 11, No. 22, 1-11, 2021.

[16] Ruiz, J., Jiménez, A., & Rodríguez-Maya, N. Hacia un modelo predictivo de la deserción escolar: El caso del Instituto Tecnológico de Zitácuaro. Congr. Ing. en Sist. Comput. Mecatrónica y Telemática [Online]. Disponible en: https://promep.sep.gob.mx/archivospdf/MEMORIAS/Producto2814830.PDF, 2016.

[17] Sánchez Santamaría, J. Modelos predictivos para el estudio del abandono en centros universitarios, Tesis de Licenciatura, Universidad Politécnica de Madrid. Repositorio institucional, 2014.

[18] Sommerville, I. Software engineering. Boston: Addison-Wesley, 2011.

[19] Urbina-Nájera, A. B., Camino-Hampshire, J. C., & Cruz, R. Deserción escolar universitaria: Patrones para prevenirla aplicando minería de datos educativa. Reli. - Rev. Electrónica Investig. y Evaluación Educ., Vol. 26, No. 1, 2020.

[20] Valdovinos, R. M. Técnicas de submuestreo, toma de decisiones y análisis de diversidad en aprendizaje supervisado con sistemas múltiples de clasificación, Tesis de Doctorado, Universitat Jaume I. Repositorio institucional, 2006.

[21] Valero Orea, S., Salvador, A., & García Alonso, M. Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos. Recur. Digit. para la Educ. y la Cult., 33-39, 2010.

[22] Vergara-Díaz, G., & Peredo López, H. Relación del desempeño académico de estudiantes de primer año de universidad en Chile y los instrumentos de selección para su ingreso. Rev. Educ., Vol. 41, No. 2, 95-104, 2017.

[23] Weitzenfeld, A., & Guardati, S. Ingeniería de Software: El proceso para el desarrollo de software. Inst. Tecnológico Autónomo México, 355–396, 2008.