



**Universidad
Andrés Bello®**

UNIVERSIDAD ANDRES BELLO

Facultad de Ingeniería

**DETECCIÓN TEMPRANA DE TITULACIÓN TARDIA EN ESTUDIANTES DE
POSTGRADOS APLICANDO MACHINE LEARNING**

Tesis para la obtención del grado de Magíster en Gestión en Tecnologías de
Información y Telecomunicaciones

Autores:

Francisco Javier Ojeda Yáñez
David Alberto Palominos Fajardo

Profesor Tutor:

PhD. David Ruete Zúñiga

Índice de Contenido

1. Introducción	1
2. Estado del Arte	3
3. Diagrama de Ishikawa	8
4. Hipótesis, Objetivo General y Objetivos Específicos	11
4.1. Hipótesis	11
4.2. Objetivo General	11
4.3. Objetivos Específicos	11
5. Alcances	13
6. Marco Teórico	14
6.1 Inteligencia Artificial	14
6.2 Machine Learning	15
6.4.1 RStudio	21
7. Marco Metodológico	22
7.1 Metodología de investigación cuantitativa	23
8. Resultados Preliminares	26
9. Artículos Propuestos	29
9.1 Introducción	29
9.2. Trabajos relacionados	29
9.3. Métodos experimentales	32
9.3.1. Obtención de Base de Datos	32
9.3.2. Análisis descriptivo	33
9.3.3. Análisis predictivo	36
9.3.4. Entrenamiento y Prueba del Modelo	36
9.4. Resultados y Discusión	38
9.4.1. Regresión Logística (LR)	38
9.4.2. Máquina de vector de soporte (SVM)	40
9.4.3. Árbol de decisión (Decision Tree, DT)	42

9.4.4. Nayve Bayes (NB)	43
9.4.5. Random Forest (RF)	44
9.4.6. K vecinos más cercanos (KNN)	46
9.5. Comparación de resultados	48
9.6. Conclusiones	49
10. Conclusiones Proyecto de Grado	51

Índice de Figuras

Figura 1. Diagrama de Ishikawa (Fuente: elaboración propia)	8
Figura 2. Diagrama de Ishikawa (Fuente: elaboración propia)	8
Figura 3. Libro Online IAAR (Fuente: https://iaarbook.github.io/deeplearning/)	15
Figura 4. Libro Online IAAR (Fuente: https://iaarbook.github.io/machine-learning/)	16
Figura 5. Libro Online IAAR (Fuente: https://iaarbook.github.io/machine-learning/)	16
Figura 6. Libro Online IAAR (Fuente: https://iaarbook.github.io/machine-learning/)	17
Figura 7. Libro Online IAAR (Fuente: https://iaarbook.github.io/machine-learning/)	18
Figura 8. Libro Online IAAR (Fuente: https://iaarbook.github.io/deeplearning/)	19
Figura 9 Libro Online IAAR (Fuente: https://iaarbook.github.io/deeplearning/)	19
Figura 10. Libro Online IAAR (Fuente: https://iaarbook.github.io/deeplearning/)	20
Figura 11. RStudio (Fuente: https://www.rstudio.com/)	21
Figura 13 (Fuente creación propia)	26
Figura 14 (Fuente creación propia)	27
Figura 15 (Fuente creación propia)	27
Figura 16 (Fuente creación propia)	28
Figura 17 (Fuente creación propia)	28
Figura 18 Distribución de Graduación tardía (Fuente creación propia)	34
Figura 19 Distribución de Graduación tardía por género (Fuente creación propia)	34
Figura 20. Distribución de Graduación tardía por modalidad de estudio (Fuente creación propia) .	35
Figura 21 (fuente https://aws.amazon.com/es/blogs/big-data/create-train-and-deploy-machine-learning-models-in-amazon-redshift-using-sql-with-amazon-redshift-ml/)	37
Figura 22 Ecuación para encontrar la precisión del modelo	39
Figura 23 Gráfico Matriz de confusión Regresión Logística, Accuracy: 0,7574	39
Figura 24 Curvas ROC – Regresión Logística	40
Figura 25 Gráfico Matriz de confusión SVM, Accuracy: 0,8441	41
Figura 26 Curvas ROC – SVM	41
Figura 27 Gráfico Matriz de confusión DT, Accuracy: 0,7871	42
Figura 28 Curvas ROC – DT	43
Figura 29 Gráfico Matriz de confusión NB, Accuracy: 0,5074	44
Figura 30 Curvas ROC – NB	44
Figura 31 Gráfico Matriz de confusión RF, Accuracy: 0,8688	45
Figura 32 Curvas ROC – RF	46
Figura 33 Gráfico Matriz de confusión KNN, Accuracy: 0,854	47
Figura 34 Curvas ROC – KNN	47
Figura 35 Verdaderos (+) + Verdaderos (-) vs Falsos (+) + Falsos (-)	48
Figura 36 Nivel de Accuracy por modelo	49

Índice de Tablas

Tabla 1. Cantidad de alumnos de postgrado por especialidad	38
Tabla 2. Variables de entrada para ejecución modelo predictivo	43
Tabla 3. Matriz de Confusión Regresión Logística	43
Tabla 4. Matriz de Confusión SVM.....	45
Tabla 5. Matriz de Confusión DT.....	47
Tabla 6. Matriz de Confusión NB	48
Tabla 7. Matriz de Confusión RF.....	50
Tabla 8. Matriz de Confusión KNN.....	51
Tabla 9. Resultado de precisión de modelos predictivos.....	53

RESUMEN

El aumento a nivel global en la oferta de nuevos programas académicos que sean atractivos y actualizados ha generado como consecuencia, un incremento directo en la cantidad de estudiantes inscritos en el mercado de la educación. Este incremento exponencial en la oferta y demanda de nuevos y mejores programas académicos, ha incorporado nuevos desafíos y problemas que deben ser abordados e investigados para lograr resolver de mejor forma los nuevos paradigmas que presenta este cambio. Dos de los problemas ya detectados, tiene relación con el aumento en el nivel de deserción de estudiantes de forma temprana y un aumento en la graduación tardía de aquellos que sí han logrado completar su ciclo de estudios en los plazos establecidos por cada institución.

Chile, no ha estado aislado de esta tendencia, generando los mismos síntomas identificados a nivel global, generando consecuencias negativas importantes para los establecimientos educacionales, dado que esta tendencia, está impactando de forma directa en los procesos de acreditación de sus carreras educativas. Dado lo anterior tanto docentes como investigadores se han interesado en participar de nuevas investigaciones que permitan ofrecer estrategias que aborden nuevas soluciones para disminuir este fenómeno, incorporando mecanismos de análisis respecto a los factores de riesgo de los estudiantes durante su proceso de formación.

Este estudio se focalizará en presentar la elaboración de un modelo de detección temprana basado en el aprendizaje automático – Machine Learning – para la problemática de graduación tardía en alumnos de Postgrado. El estudio está basado en dos partes, en primer lugar, se presenta un análisis descriptivo que permite examinar la información existente para lograr comprender la situación actual respecto de la situación de los estudiantes y los programas de estudios ofertados por las instituciones educacionales. En segundo lugar, se presenta un análisis predictivo, utilizando técnicas de aprendizaje automático, permitiendo documentar los factores de riesgo que son parte importante en la explicación de la titulación tardía en los alumnos de postgrado. La base de información analizada contiene

datos que representan factores de riesgos de los estudiantes y su programa de estudio, toda esta información está disponible en los sistemas de información de la universidad. Para el análisis, se considera una muestra de los alumnos de postgrado de la universidad, específicamente la Facultad de Economía, de los últimos 5 años.

Los resultados de los hallazgos proponen mejorar el apoyo de las instituciones hacia los alumnos, permitiendo desarrollar herramientas que permitan intervenir en el momento preciso a los alumnos que presenten factores de riesgos respecto a la graduación tardía. El acompañamiento de forma cercana y apoyo de los profesores tutores a los alumnos durante el proceso de desarrollo de su tesis, es un factor importante para el logro del objetivo en la graduación a tiempo.

Proponemos que los resultados de este estudio sean utilizados por las instituciones de educación superior para desarrollar nuevas iniciativas proactivas, que ayuden a disminuir el aumento de titulación tardía de los estudiantes.

1. Introducción

La deserción y la graduación tardía son problemas críticos en los programas de magíster en todas las disciplinas y universidades. Estas variables se pueden explicar con diferentes supuestos o apreciación del comportamiento de los estudiantes. Sin embargo, es relativamente poca la investigación que evalúa las motivaciones o percepciones de los estudiantes de magíster para abandonar los programas a los cuales están adscritos.

Existen varios estudios que buscan identificar las diferentes razones de la deserción estudiantil en pregrado y posgrado, incluyendo Cursos en Línea Masivos (MOOCs), estos últimos han tenido una explosiva y creciente aceptación por estudiantes que no requiere una exigencia académica previa, por tanto, su popularidad entre las universidades por brindar este tipo de estudios y de quienes se inscriben en ellos, ha aumentado de forma explosiva y considerable en los últimos años. Esta explosión repentina de los MOOCs también ha generado un aumento explosivo de la deserción de los estudiantes que se inscriben en ellos (Hanan A., Hosam A., Ibrahim A., Nasser A., 2020).

Un análisis de 1167 estudiantes de pregrado muestra una relación interesante respecto de los estudiantes que viajan diariamente versus los que residen en el campus, de acuerdo con esta relación los primeros tienden a funcionar mejor en los estudios, teniendo una probabilidad mayor de completar el programa de pregrado, a diferencia que el segundo grupo que su probabilidad es mayor a no completar el programa de pregrado (Anna B., Claudio D., Laura S., 2020).

Otro de los aspectos importantes a analizar es la deserción en los postgrados, dado que el alumno de máster presenta un perfil diferente al alumno de pregrado, pues, en buena parte de los casos, compatibiliza sus estudios con su actividad profesional. Este elemento parece no ser un inconveniente determinante durante el período lectivo, en el que el alumno es guiado y se programa su tiempo en función de las clases y exámenes. Sin embargo, cuando se encuentra ante el desarrollo del proceso de Título de fin de Máster, que requiere de una organización autónoma, le

resulta complejo compatibilizar ambas actividades. Esta causa destaca notablemente sobre el resto como motivo principal del retraso en la graduación (Rafaela A., Pedro G., María del Mar G., Carmen M., Macarena S., 2017).

Dentro del comportamiento de los desertores, se puede ver que de una muestra de 9 de cada 10 estudiantes que se desvinculan la primera vez, no retornan a la universidad. Del total de desertores el 30% se retira durante el segundo semestre y el 20% en el tercer semestre, es decir al menos el 50% de los estudiantes se retiran dentro de los primeros tres semestres de estudios (Bibiana A., Rubén D., Difariney G., Angela M., 2020). Esto nos hace pensar en la necesidad de explorar y analizar las causas que llevan a los estudiantes a desertar, buscando modelos que nos permitan predecir con cierto grado de certeza la cantidad de estudiantes que se retirarán en los primeros semestres de estudios, y cuáles son los factores que aportan a esta decisión.

Esta investigación proporciona un estudio, focalizado en la etapa de elaboración de la tesis de grado, a través de un análisis descriptivo y modelado del perfil del estudiante, de cómo las variables cuantitativas de la experiencia académica pueden ser una proyección de las intenciones de abandono de los estudiantes.

2. Estado del Arte

Muchos investigadores y universidades se han interesado en estudiar el fenómeno de la graduación tardía y deserción de los estudiantes en los primeros años de estudios, utilizando para ello, diferentes modelos y técnicas de Machine Learning, regresión logística o algoritmos de redes neuronales, que permiten determinar en algunos casos, la probabilidad de deserción para un grupo de estudiantes y en otro análisis determinar las causas que llevan a los jóvenes estudiantes de tomar la determinación de abandonar o retrasar su titulación. Para el caso de los países que pertenecen a la OCDE, las tasas de graduación en adultos jóvenes indica que el 39% completan su educación universitaria en el transcurso de sus vidas (Francesco A., Mauro M., Gianmarco B., 2020). En 2017 el 10,6% de los jóvenes de la Unión Europea abandonaron sus carreras en los primeros años de formación. Un estudio en la Universidad de Jerusalén encontró que las tasas de deserción de los estudiantes de maestrías son del 12% (Nir R., Gad Y., Elad S., 2020).

La deserción temprana de los estudiantes ha sido materia de estudios de investigadores y universidades, lograr generar programas desde las propias universidades para mejorar la retención y la titulación, ha sido otro punto importante en materia de estudio. En el ámbito de las universidades privadas, estas no están exentas del abandono temprano y la titulación tardía en alumnos de pregrado, postgrado y doctorados (Alejandra V., Gabriel L., (2020).

Las tasas de graduación se utilizan comúnmente para evaluar la calidad de las instituciones de educación superior. Mantener un adecuado equilibrio entre el número de nuevos estudiantes inscritos y el número de graduados cada año se ha convertido en un gran desafío para las universidades (Kelzang T., Nittaya M., Phattrawan T, 2019). La graduación oportuna a menudo se ignora porque una gran mayoría de los estudiantes no se gradúa dentro del tiempo estipulado.

Las instituciones deben estudiar y analizar estos factores que están presente en la realidad actual del mundo estudiantil, estas deben encontrar e identificar cual es mejor momento para intervenir e incorporar recursos que apoyen a los estudiantes

para mantener los rangos bajos de deserción (José O., Antonio R., Paloma B., Martí C., 2020). Utilizando metodologías predictivas pueden implementar modelos que les permitan llegar a predecir con un grado de confianza suficientemente óptimo, cuando el mejor momento de intervenir.

Con la utilización de procesos de minería de datos, es posible identificar los perfiles de los estudiantes de primer año que tienen grandes dificultades para completar su primer año académico (Anne H., Michaël S., 2017).

Un punto de vista diferente pero no menos importante es la deserción de estudiantes de grados académicos en línea. utilizando técnicas similares, pero variables diferentes relacionadas con la demografía, características de las aulas, formas de participación cognitivas, académicas y conductuales, es posible predecir la deserción de los estudiantes (Kristof C., Minh P., Arno De C., Dries F., Annelies R., 2020), que, en términos comparativos, este grupo de estudiantes en línea tiene una tasa de deserción mucho mayor a los estudiantes de las aulas, estando entre el 25% y 90%.

Las direcciones de programa de magíster deben aumentar la sostenibilidad del desarrollo de capacidades en la mejora de los planes de estudio que proporcionará a los estudiantes una mayor oportunidad de aprendizaje (E. G. Publishing, E. Group, P. Limited, A. R. Reserved, O. Url y E. Uri, 2017). En este sentido las instituciones de nivel superior que ofrecen estudios de postgrado, como magísteres, deben trabajar en el desarrollo y la mejora de los modelos educativos mediante la incorporación de tecnologías de la información y las comunicaciones que servirán como herramienta de apoyo a las distintas iniciativas propuestas por la dirección de programa.

Estas herramientas proporcionarán la capacidad de análisis que hoy las herramientas convencionales no son capaces de tratar por la gran cantidad de datos existentes. Para realizar un análisis, con grandes volúmenes de datos y variables, deben considerarse tecnologías como sistemas inteligentes, minería de datos, máquina de aprendizaje, entre otras, transformando los datos de educación en una

nueva forma observable de información que sea más significativa y útil (S. Luj, 2018).

En (A. M. Shahiri, W. Husain y N. A. Rashid, 2015) las variables más utilizadas para alimentar el modelo predictivo aplicado al problema de deserción son las variables relacionadas con el rendimiento de los estudiantes, como el promedio acumulativo, las pruebas, el trabajo de laboratorio, los exámenes y la asistencia. También se utilizaron variables demográficas como la edad, el sexo, antecedentes familiares y la discapacidad. Otros trabajos utilizan variables relacionadas a actividades extracurriculares como su red de interacción social. Árbol de decisión (decision tree), K – vecino más cercano (KNN), redes neuronales (Neuronal Network), Neive Bayes y máquina vectorial de apoyo (Support Vector Machine) fueron los algoritmos utilizados para generar los modelos predictivos aplicados al problema de deserción. De estos algoritmos los que obtuvieron un mayor rendimiento fueron la Red Neuronal con un 98% de precisión, y el Árbol de decisión con un 91% de precisión. El algoritmo de Árbol de Decisión es el más utilizado para desarrollar modelos predictivos cuyo objetivo es identificar el abandono de los estudiantes en las instituciones de educación superior (M. Alban, D. Mauricio, 2019). Una explicación recomendada de cómo funciona una red neuronal o un Árbol de decisión la pueden encontrar en (T. Barros, P. SouzaNeto, I. Silva y L. A. Guedes, 2019).

En (I. Lykourantzou, I. Giannoukos, V. Nikolopoulos, G. Mpardis y V. Loumos, 2009) los autores utilizan una combinación de técnicas de máquina de aprendizaje automático de redes neuronales multicapa (MPL), máquina de vector de apoyo y análisis probabilístico, mediante tres esquemas de decisión. Se utilizaron como variables de entrada el género, residencia, experiencia laboral (variables demográficas), nivel educativo, puntaje de la prueba de elección múltiple, carrera, fecha de presentación del proyecto, actividad de la sección (variables de rendimiento académico previo). Para medir el rendimiento del modelo se utilizaron métricas de exactitud, precisión y sensibilidad.

En (V. R. D. C. Martinho, C. Nunes y C. R. Minussi, 2013) los autores presentan un modelo basado en una red neuronal Fuzzy-ARTMAP utilizando solo los datos de

matrícula recopilados para un período de 7 años. Los resultados del rendimiento del modelo mostraron una precisión superior al 85%.

En (A. A. Mubarak, H. Cao y W. Zhang, 2020), los autores utilizan modelos ocultos de Markov (HMM: Hidden Markov Model) y regresión logística secuencial para generar el modelo predictivo. El conjunto de datos está conformado por variables como número medio de sesiones de cada participante por semana, números de comportamiento de acceso, problema, video wiki, discusión de la semana, número de comportamiento de todos los eventos por semana, número de días activos por semana, número de comportamientos de eventos de video, consumo de tiempo por semana, y cuestionarios semanales, para 32,593 estudiantes, sumando un total de 9,966,292 registros. El modelo tuvo una exactitud del 84%. Comparándose con otros modelos, se obtuvieron resultados similares con máquinas de vectores de apoyo, árboles de decisión, regresión lineal y bosque aleatorio (Random Forrest).

En (S. B. Kotsiantis, C. J. Pierrakeas y P. E. Pintelas, 2003), los autores realizan experimentos con redes neuronales, SVM, NB y LR para predecir el riesgo de abandonar la educación a distancia. Los conjuntos de datos mezclan variables demográficas y de rendimiento académico de los estudiantes. Como resultados obtuvieron y concluyeron que la precisión de NB era significativamente eficiente y oscilaba entre el 60,42% y 83,67%.

En otro estudio, (K. T. Chui, D. C. L. Fung, M. D. Lytras y T. M. Lam, 2018) propuso un algoritmo de SVM basado en vectores de entrenamiento reducido (RTV-SVM) para predecir el riesgo tanto a los estudiantes universitarios como a los marginales. Los resultados demostraron que la precisión alcanzó un 91,2% para ambos casos.

En (C. Burgos, M. L. Campanario, D. de la Peña, J. A. Lara, D. Lizcano y M. A. Martínez, 2018) los autores utilizaron el algoritmo de regresión logística para crear un modelo predictivo, considerando los datos académicos de los estudiantes. El objetivo del modelo era como mecanismo de apoyo o ayuda a la toma de decisiones de una política de retención de estudiantes. Para comprobar el rendimiento del

modelo se utilizaron matrices de confusión y precisión. Como resultado se obtuvo una reducción del 14% en la tasa de deserción.

En (R. Asif, A. Merceron, S. A. Ali y N. G. Haider, 2017), los autores proponen un método de dos etapas: predecir el rendimiento académico y estudiar las progresiones típicas de los estudiantes al final del programa de estudios de cuatro años. Ambos métodos son combinados para realzar la predicción aplicada al problema de deserción. Se utilizó para esto el algoritmo de árbol de decisión con clasificación jerárquica. El conjunto de datos está compuesto por únicamente variables de rendimiento académico como las notas de admisión, notas finales de los cursos del programa de estudio, para primer y segundo año de universidad. Para evaluar el rendimiento del modelo se utilizaron las matrices de precisión y confusión.

Se puede observar que los estudios se focalizan en la deserción de los estudiantes a distintos niveles, pero no con foco en la etapa de graduación. Una de las características de este estudio, es que los alumnos de magíster no abandonan el programa de estudio en cursos regulares (cursos tradicionales dictados por docente) sino que en los cursos dedicados a la etapa de elaboración de la tesis o proyecto final de grado de magíster. Debido a este particular hecho, este estudio pretende entregar una alerta temprana del riesgo de deserción de los alumnos de magíster, en la etapa de elaboración de su proyecto final de grado, utilizando algoritmos de machine Learning para realizar la predicción, y un conjunto de datos cognitivos conformados por variables demográficas y académicas. Por cuestiones de tiempo, se ha decidido dejar de lado variables no cognitivas, pues requieren de la aplicación de encuestas o herramientas para conocer el comportamiento no cognitivo de las personas como, por ejemplo, el liderazgo.

3. Diagrama de Ishikawa

Para analizar el problema central, causas y subcausas que lo afectan y categorías asociadas, se desarrolla el Diagrama de Ishikawa, Figura 1, el cual consiste en la representación gráfica de una línea horizontal, correspondiente al problema central a resolver, sobre la cual conectan flechas rectas oblicuas (espinas), que representan las causas del problema, por lo que en esta iteración permite encontrar el conjunto de problemas que representan la causa raíz.

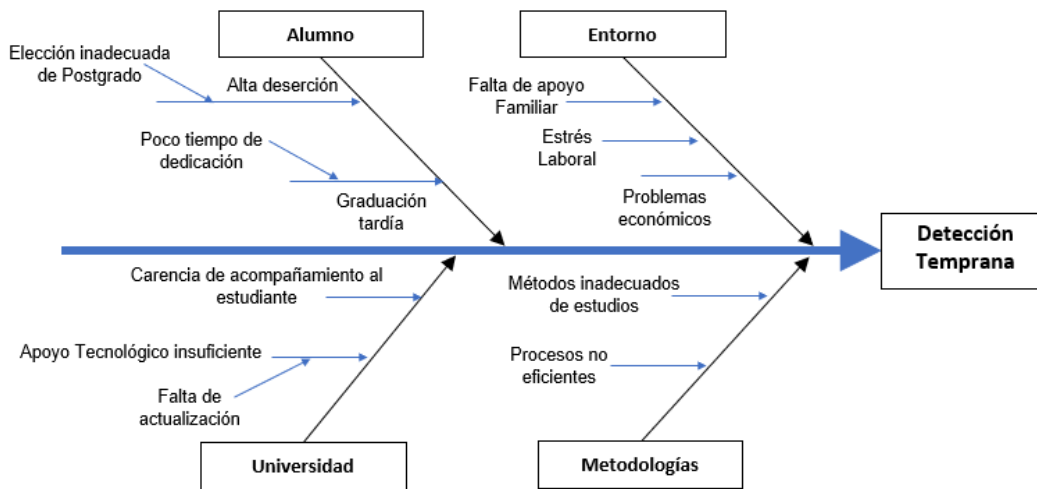


Figura 1. Diagrama de Ishikawa (Fuente: elaboración propia)

En el segundo diagrama, figura 2, se destacan en verde las causas que serán parte de esta investigación y en rojo las que no serán parte del alcance.

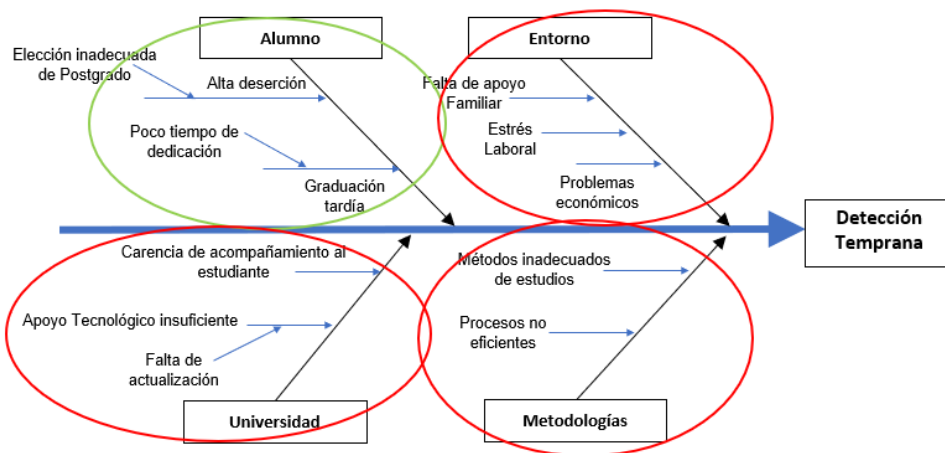


Figura 2. Diagrama de Ishikawa (Fuente: elaboración propia)

Como se puede observar, la cabeza del diagrama de Ishikawa representa el efecto de las causas que se explicitan en las espinas. En este caso son la deserción y graduación tardía en estudiantes de Maestría, que pasará a ser el problema principal por resolver.

Luego tenemos las entidades involucradas en el diagrama de causa y efecto: Alumno, entorno, universidad y métodos de estudio. En las entidades se producen las causas que producen el efecto. De estas se puede decir:

- **Alumno.** Es el principal interesado en finalizar su proceso de titulación en un tiempo acotado y de acuerdo con las directrices de la universidad, es quien tomó en la mayoría de los casos la decisión de estudiar, puede también que por su trabajo le solicitarán el magister. Se identifican como principales causas la alta deserción (C01) y la graduación tardía (C02).
- **Entorno.** El entorno del alumno tiene directa relación con su situación económica, laboral y también del apoyo que reciba de su familia para que se pueda dedicar e invertir tiempo en la ejecución de sus estudios y trabajo de título.
- **Universidad.** Entidad que debe tener un programa adecuado y actualizado para que se puedan lograr los objetivos propuestos al inicio y planificación del magíster, es importante que el plan considere el acompañamiento hasta el final del proceso a los estudiantes tanto para modalidad online como presencial. Se consideran las herramientas tecnológicas actuales para que el estudiante pueda realizar sus estudios y trabajo de título, en la mayoría de los casos los alumnos dominan y tienen acceso a las nuevas tecnologías lo que les permite avanzar de forma adecuada y apoyarse en presentaciones y estudios, sin embargo, también existe un porcentaje de personas que no tienen acceso o dominio de herramientas que pudiesen facilitar la ejecución y culminación en el tiempo adecuado de sus estudios.

- **Métodos de Estudio.** Otra entidad detectada como causa de deserción y demora en el proceso de titulación son los métodos de estudio que tenga el estudiante y/o también los de la universidad, los cuales si no son adecuados pueden provocar que los resultados no sean los esperados.

4. Hipótesis, Objetivo General y Objetivos Específicos

4.1. Hipótesis

Mejorar los programas y procesos de enseñanza puede ser una realidad, apoyados con herramientas tecnológicas y el uso de técnicas de Inteligencia Artificial, en este caso Machine Learning, permitirá generar modelos de predicción que contribuyan en esta labor. Con el uso de Machine Learning y una buena selección de dataset que incluyan variables de rendimiento académico y otros factores relevantes de los alumnos, se pueda comprobar la detección de forma temprana de aquellos alumnos que están en riesgo de transitar por una titulación tardía.

4.2. Objetivo General

El objetivo general del trabajo es diseñar una Metodología aplicada mediante el uso de Machine Learning para la Detección Temprana de deserción y graduación tardía de Magíster.

4.3. Objetivos Específicos

Es importante que los objetivos del proyecto se puedan medir, para poder verificar cuantitativa y/o cualitativamente que se cumple con el objetivo general, como herramienta para la construcción del objetivo específico medible se considera la técnica S.M.A.R.T.;

- Específico (Specific)
- Medible (Mensurable)
- Alcanzable (Achievable)
- Realizable (Realistic)
- Limitado por el Tiempo (Timed)

El objetivo específico se ha obtenido a partir de las causas establecidas en el Diagrama de Ishikawa, el que ha sido confeccionado para representar el análisis del problema.

Para establecer los objetivos específicos, se consideró aquellos problemas que fueron identificados dentro del alcance de las Causas identificadas del Alumno, que en el caso de medibles son dos.

- Disminuir la alta tasa de deserción (OEM1)
- Reducir los tiempos de graduación tardía (OEM2)

Serán también parte de los objetivos específicos que los resultados obtenidos y proyectados puedan ser medidos en cuanto a precisión, eficiencia, exactitud y rendimiento.

5. Alcances

De acuerdo con lo analizado en el diagrama de Ishikawa y en los objetivos de este proyecto podemos definir los siguientes alcances de la investigación:

Será parte de esta investigación:

- Analizar estadísticamente la información disponible en la Base de Datos de la Universidad Nacional Andrés Bello (UNAB), referente a Magister.
- Aplicar algoritmos que permitan predecir qué perfil de alumnos y carreras tendrán problemas para culminar a tiempo el proceso de titulación.
- Investigar modelos de Machine Learning que permitan mejorar los tiempos actuales asociados a los procesos de titulación de postgrados.
- Generar una guía de reacción que apoye en evitar deserciones a las propias universidades y alumnos.

No son parte de este trabajo:

- El entorno del estudiante.
- Situaciones motivacionales del alumno.
- Infraestructura de la Universidad.
- Planes de estudio de la Universidad.

6. Marco Teórico

Para comprender la solución desarrollada en esta investigación, es necesario conocer algunos aspectos teóricos, para esto se definen los principales conceptos base de la Detección temprana de deserciones y graduación tardía en estudiantes de Maestría.

Los procesos de detección temprana buscan anticiparse a conductas mediante la identificación de factores de vulnerabilidad existentes. Con esta formación se espera contribuir a un proceso más eficiente de acompañamiento a las personas e instituciones que requieran apoyo en esta línea, identificando situaciones de riesgo, así como también recursos personales y ambientales para afrontar el tema.

La Tasa de Deserción, es la razón entre la diferencia de los alumnos que iniciaron un proceso y los que llegaron al final de este dentro de los plazos considerados para completar el estudio e investigación en el caso de la tesis.

Graduación tardía concierne a unos de los principales aspectos de esta investigación y corresponde al tiempo excesivo y fuera de plazo en realizar principalmente el trabajo de Tesis, afectando el normal proceso de titulación.

6.1 Inteligencia Artificial

La inteligencia artificial (IA) es una de las ramas de la informática, con fuertes raíces en otras áreas como la lógica y las ciencias cognitivas, existen muchas definiciones de lo que es la inteligencia artificial. Sin embargo, todas ellas coinciden en la necesidad de validar el trabajo mediante programas. Figura 3.

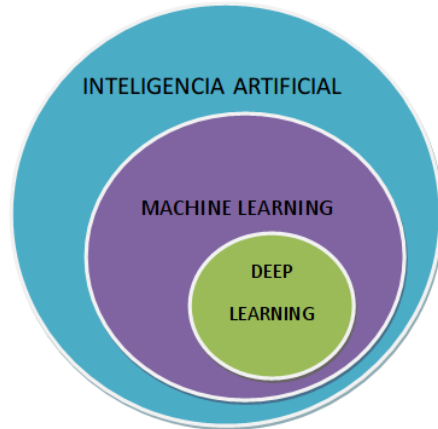


Figura 3. Libro Online IAAR (Fuente: <https://iaarbook.github.io/deeplearning/>)

6.2 Machine Learning

Es una disciplina científica del ámbito de la Inteligencia Artificial que crea sistemas que aprenden automáticamente. Aprender en este contexto quiere decir identificar patrones complejos en millones de datos. Cabe conocer los tres tipos de algoritmos de aprendizaje automático que existen: supervisado, no supervisado y por refuerzo.

6.2.1 Algoritmo de aprendizaje supervisado

En el aprendizaje supervisado, la máquina se enseña con el ejemplo. De este modo, el operador proporciona al algoritmo de aprendizaje automático un conjunto de datos conocidos que incluye las entradas y salidas deseadas, y el algoritmo debe encontrar un método para determinar cómo llegar a esas entradas y salidas.

Algoritmos Supervisados:

- Regresión Lineal
- Regresión Logística
- Árboles de Decisión
- Redes Neuronales
- Random Forest
- SVM o Máquinas de vectores de soporte

- KNN o K vecinos más cercanos

A. Regresión Lineal

De acuerdo con la figura 4, para representar Regresión Lineal, esta se utiliza para estimar los valores reales (costo de las viviendas, el número de llamadas, ventas totales, etc.) basados en variables continuas. La idea es tratar de establecer la relación entre las variables independientes y dependientes por medio de ajustar una mejor línea recta con respecto a los puntos. Esta línea de mejor ajuste se conoce como línea de regresión y está representada por la siguiente ecuación lineal:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

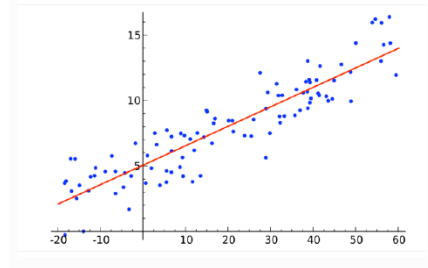


Figura 4. Libro Online IAAR (Fuente: <https://iaarbook.github.io/machine-learning/>)

B. Regresión Logística

Los modelos lineales, representado en la figura 5, también pueden ser utilizados para clasificaciones; es decir, que primero ajustamos el modelo lineal a la probabilidad de que una cierta clase o categoría ocurra y, a luego, utilizamos una función para crear un umbral en el cual especificamos el resultado de una de estas clases o categorías

$$f(x) = \frac{1}{1 + e^{-x}}$$

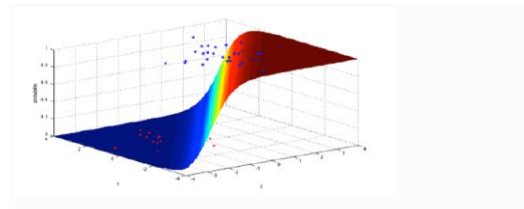


Figura 5. Libro Online IAAR (Fuente: <https://iaarbook.github.io/machine-learning/>)

C. Árboles de decisión

Son diagramas muy similares a los sistemas de predicción basados en reglas, ver figura 6, que sirven para representar y categorizar una serie de condiciones que ocurren de forma sucesiva, para resolver un problema.

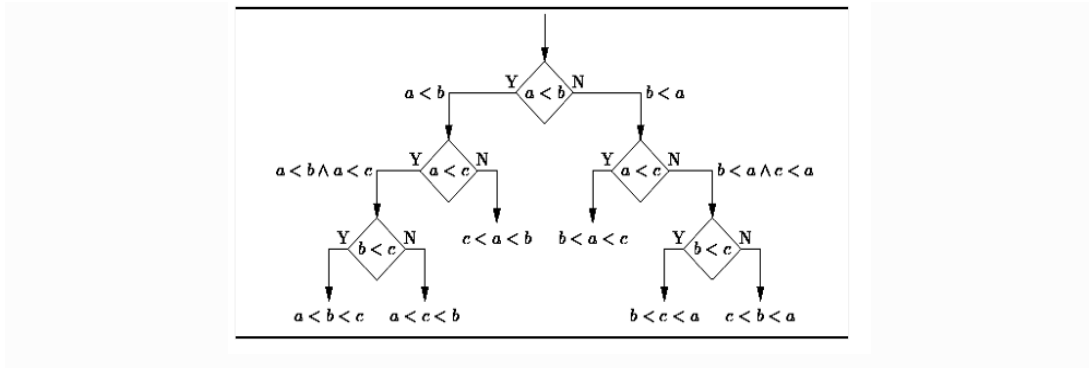


Figura 6. Libro Online IAAR (Fuente: <https://iaarbook.github.io/machine-learning/>)

6.2.2 Algoritmo de aprendizaje sin supervisión

Aquí, el algoritmo de aprendizaje automático estudia los datos para identificar patrones. No hay una clave de respuesta o un operador humano para proporcionar instrucción. En cambio, la máquina determina las correlaciones y las relaciones mediante el análisis de los datos disponibles.

En un proceso de aprendizaje no supervisado, se deja que el algoritmo de aprendizaje automático interprete grandes conjuntos de datos y dirija esos datos en consecuencia. Así, el algoritmo intenta organizar esos datos de alguna manera para describir su estructura. Esto podría significar la necesidad de agrupar los datos en grupos u organizarlos de manera que se vean más organizados.

Algoritmos no Supervisados

- Algoritmos Clustering K-means
- Análisis de Componentes Principales

A. Algoritmos Clustering K-means

Figura 7, representa una de las técnicas de aprendizaje no supervisadas más conocidas. Es un algoritmo muy simple que sirve para reducir al mínimo la suma de las distancias cuadradas de la media dentro del agrupamiento.

$$\min_{\mathcal{S}} E(\mu_i) = \min_{\mathcal{S}} \sum_{i=1}^k \sum_{\mathbf{x}_j \in \mathcal{S}_i} \|\mathbf{x}_j - \mu_i\|^2 \quad (1)$$

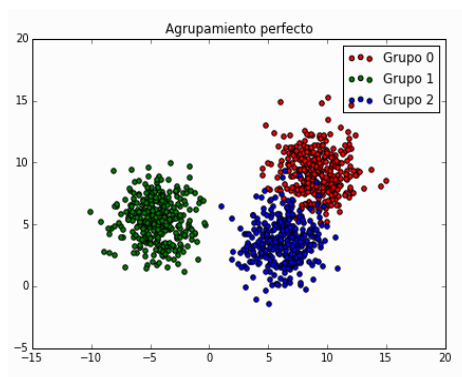


Figura 7. Libro Online IAAR (Fuente: <https://iaarbook.github.io/machine-learning/>)

6.2.3 Algoritmo de aprendizaje por refuerzo

El aprendizaje por refuerzo se centra en los procesos de aprendizajes reglamentados, en los que se proporcionan algoritmos de aprendizaje automáticos con un conjunto de acciones, parámetros y valores finales.

Al definir las reglas, el algoritmo de aprendizaje automático intenta explorar diferentes opciones y posibilidades, monitorizando y evaluando cada resultado para determinar cuál es el óptimo.

En consecuencia, este sistema enseña la máquina a través del proceso de ensayo y error. Aprende de experiencias pasadas y comienza a adaptar su enfoque en respuesta a la situación para lograr el mejor resultado posible.

6.3 Deep Learning

Es una de las áreas de investigación más popular dentro del campo de la Inteligencia Artificial. Gran parte de las nuevas investigaciones, trabajan con modelos basados en técnicas de Deep Learning, logrando resultados sorprendentes en áreas de procesamiento de lenguaje natural y Visión por computadora. En los últimos años Deep Learning ha producido una revolución en el campo del Machine Learning.

6.3.1 Arquitecturas de Deep Learning

- Redes neuronales prealimentadas: Es el modelo más sencillo. Estas redes se mueven en un solo sentido, hacia adelante. La figura 8 representa un ejemplo de arquitectura del Deep Learning. Se suelen utilizar en problemas de selección simples.

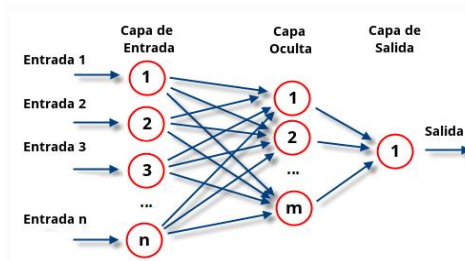


Figura 8. Libro Online IAAR (Fuente: <https://iaarbook.github.io/deeplearning/>)

- Redes neuronales convolucionales: Lo que diferencia a este tipo de redes, es que suponen explícitamente que las entradas son imágenes, lo que permite codificar ciertas propiedades en la arquitectura, permitiendo ganar en eficiencia y reducir la cantidad de parámetros en la red. Ver figura 9.

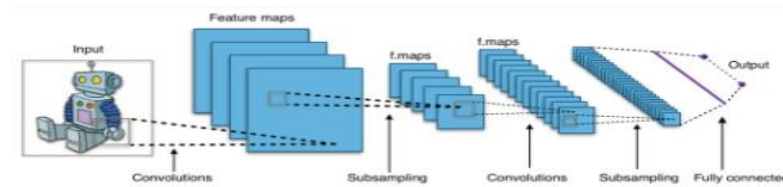


Figura 9 Libro Online IAAR (Fuente: <https://iaarbook.github.io/deeplearning/>)

- Redes neuronales recurrentes: En la figura 10, se pueden apreciar redes en bucle de retroalimentación, que permiten que la información persista. Aplicadas con éxito en problemas de reconocimiento de voz, modelado de lenguaje, subtítulos de imágenes, entre otros.

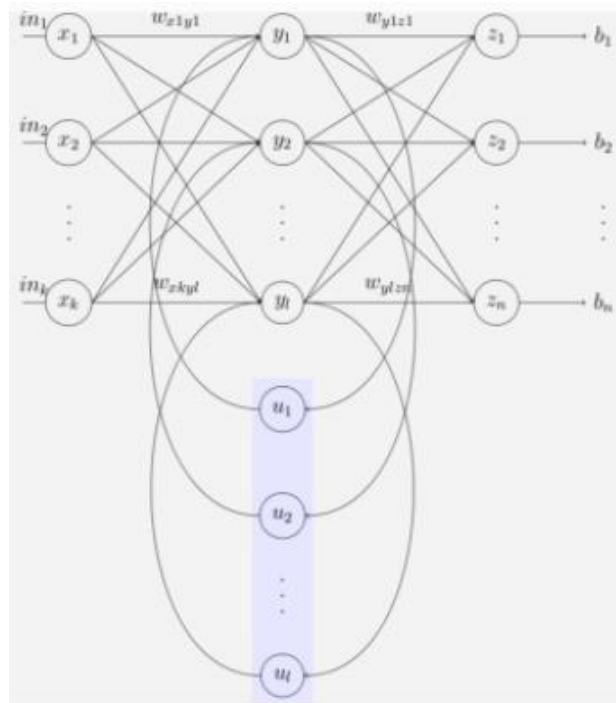


Figura 10. Libro Online IAAR (Fuente: <https://iaarbook.github.io/deeplearning/>)

6.4 Lenguaje de Programación R

R es un sistema para análisis estadísticos y gráficos creado por Ross Ihaka y Robert Gentleman. R tiene una naturaleza doble de programa y lenguaje de programación. Es de distribución libre bajo licencia GNU y funciona para distintos sistemas operativos.

Es un lenguaje orientado a Objetos: bajo este complejo término se esconde la simplicidad y flexibilidad de R. El hecho que R es un lenguaje de programación puede desanimar a muchos usuarios que piensan que no tienen “alma de programadores”. Esto no es necesariamente cierto por dos razones. Primero R es un lenguaje interpretado (como Java) y no compilado (como C, C++, Fortran,

Pascal, entre otros), lo cual significa que los comandos escritos en el teclado son ejecutados directamente sin necesidad de construir ejecutables.

6.4.1 RStudio

Es un entorno de desarrollo integrado para el lenguaje de programación R, dedicado a la computación estadística y gráficos. Incluye una consola, editor de sintaxis que apoya la ejecución de código, así como herramientas para el trazado, la depuración y la gestión del espacio de trabajo. La figura 11 se puede observar el entorno RStudio.

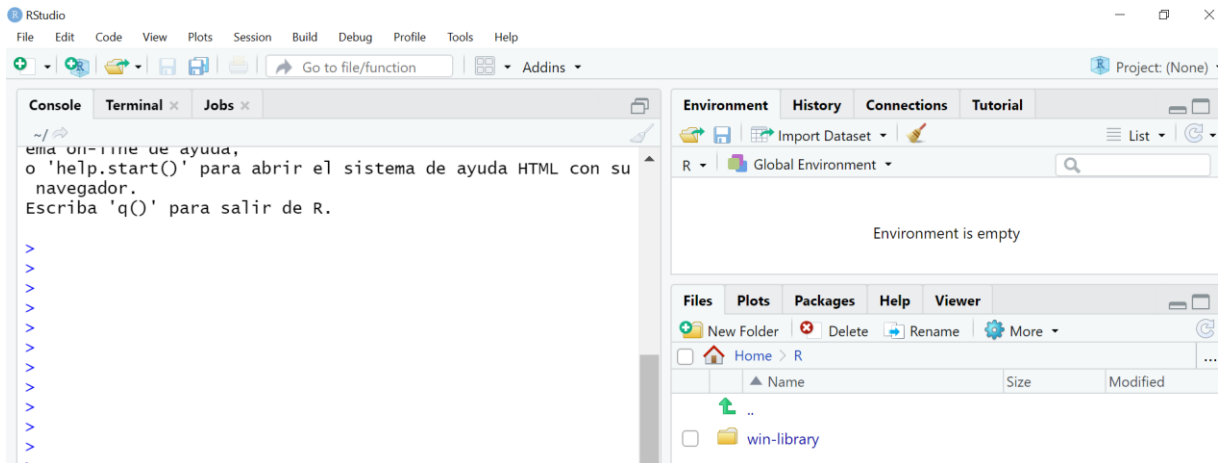


Figura 11. RStudio (Fuente: <https://www.rstudio.com/>)

7. Marco Metodológico

La figura 12, permite identificar el marco metodológico de investigación Cuantitativa (Hernández R., 2010), debido a las facilidades que esta presenta, se tomó como referencia esta metodología, dado que su sistema de recolección de datos es más fiable y representa de forma más exacta los resultados para un desarrollo del tipo de proyecto presentado:

- Requiere mantener un orden en el cumplimiento de cada etapa del proyecto, es decir, antes de dar comienzo a una nueva etapa, se debe finalizar la etapa inmediatamente anterior, esto asegura que no queden etapas con tareas pendientes, permitiendo asegurar que los entregables de cada etapa se cumplan en los plazos comprometidos.
- Orientado a la creación de una estructura documental.
- Los elementos pueden ser tratados de forma separada y luego ser integrados de acuerdo con la estructura especialmente elaborada, para que cada una de las partes puedan quedar perfectamente combinadas.



Figura 12. Metodología de Investigación Cuantitativa (Figura: H., F. y B., 2014: 5)

7.1 Metodología de investigación cuantitativa

El enfoque cuantitativo es secuencial y probatorio. Cada etapa debe terminar para comenzar otra y no se pueden saltar pasos del modelo.

Se comienza por la elaboración de una idea que debe definir su alcance. Una vez realizada esta tarea, se derivan los objetivos y preguntas de investigación, si ese fuese el caso, se revisa la literatura asociada al tema y se construye el marco teórico. Se establece el aporte de valor del trabajo y se determinan las variables relevantes que servirán para validar las metas propuestas. Se desarrolla un plan, para probar las variables (diseño), se miden las variables en un escenario controlado, se analizan los resultados obtenidos de la medición de variables y se establecen las conclusiones respecto a las metas planteadas.

La primera fase es el inicio de la investigación o proyecto, donde se presenta la idea que será estudiada. La idea representa el primer acercamiento a la realidad que se quiere estudiar, sin importar qué tipo de paradigma fundamente el estudio ni el enfoque que se adopte. En esta fase no se puede descartar la idea por buena o mala, pues hace falta más análisis e información para decidir.

En la segunda fase, se plantea el problema de investigación, se establecen los objetivos de investigación, las preguntas de investigación (si fuese necesario), se justifica la investigación y se evalúan las deficiencias en el conocimiento del problema. De nada sirve contar con un buen método, si no se tiene claro qué problema se quiere resolver. El tiempo de pasar de la fase 1 a la fase 2 es relativa y depende de cuán familiarizado se encuentre el investigador con el tema tratado. En esta etapa, se necesita el problema específico en términos concretos y explícitos, de manera que sea susceptible de investigarse con procedimientos científicos (Selltiz, 1980). Ahora bien, como dice (Ackoff, 1967), un problema bien planteado está parcialmente resuelto.

En la siguiente fase, se revisa la literatura asociada al tema y se construye el marco teórico. Según (Hernández R., 2010) el desarrollo de esta fase da como resultado,

por un lado, un proceso de inmersión en el conocimiento existente correlacionado con el problema planteado, y por otro lado un producto, el marco teórico ((Eds.), 2005). Este paso sustenta, válida y encuadra teóricamente el estudio (Hernández Sampieri, 2009) (Rojas Soriano, 2002).

En la fase cinco, se define el fin del estudio. Además, se definen las variables que ayudarán a medir y validar el fin del estudio. La variable es una propiedad de variación que puede medirse u observarse.

En la fase seis, el investigador debe visualizar la manera práctica y concreta de medir las variables definidas en la fase anterior, para validar los objetivos específicos del proyecto. En esta fase se diseña el experimento, se fija el contexto basado en los alcances fijados para el proyecto.

La fase siete se centra en “qué o quiénes”, es decir, en los participantes, objetos, sucesos o comunidades de estudio, de las cuales depende del planteamiento del proyecto. Se define la población, se elige el método de selección de la muestra (probabilístico o no probabilístico), se define el tamaño de la muestra, el proceso de selección y la obtención de la muestra.

En la siguiente fase se recolectan los datos pertinentes sobre los atributos, conceptos o variables del caso de análisis. Recolectar los datos implica la elaboración de un plan que permita conseguir este objetivo. Para esto se debe tener en cuenta y determinar las fuentes de obtención de datos, la localización de estas fuentes, el proceso de recolección de los datos y cómo finalmente los analizaremos para responder al planteamiento del problema. Esta fase es fundamental para la obtención de resultados.

En la fase nueve, se realiza la tarea de análisis de los datos obtenidos en la fase anterior. Actualmente, este proceso se realiza por medio de computadores y softwares especializados. Es por esto que además del análisis de los datos, esta fase se encarga de la interpretación de los resultados del análisis.

La última fase, se encarga de presentar los resultados del análisis de los datos realizados en la fase anterior. Se comunicarán los resultados mediante un reporte, el cual puede presentarse en distintos formatos: un libro, un artículo para una revista o congreso, un diario, una presentación, un documento técnico, una tesis, entre otras posibilidades. En cualquiera de los casos, debemos describir la investigación realizada y los descubrimientos producidos.

8. Resultados Preliminares

De acuerdo con la información proporcionada, sobre la cual se realizará la investigación, podemos destacar algunos resultados preliminares.

8.1 Distribución de estudiantes por Género

Figura 13, se puede observar que existe una cantidad mayor de alumnas mujeres inscritas en programas de postgrado, siendo la distribución de la siguiente forma:

- 44% del alumnado son hombres
- 56% de alumnado es mujer

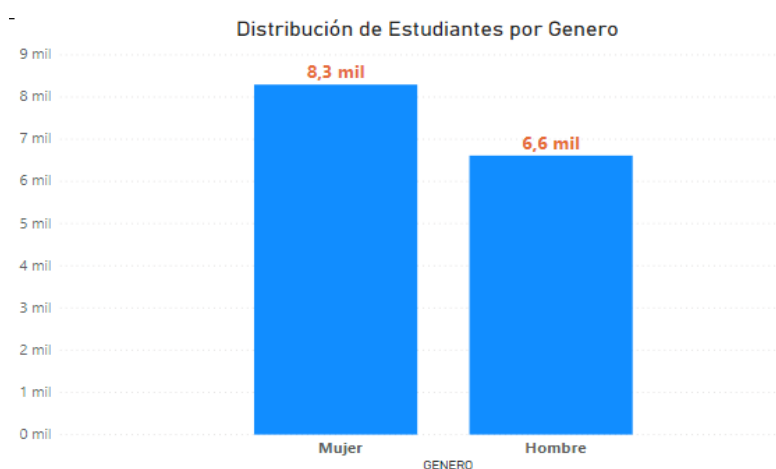


Figura 13 (Fuente creación propia)

8.2 Distribución Alumnos activos versus desertores

Figura 14, se puede determinar la cantidad de alumnos activos y desertores de acuerdo con el año de ingreso al postgrado, pudiendo inferir lo siguiente:

- Aumenta la cantidad de alumnos año sobre año
- Se mantiene estable la cantidad de Deserción durante los primeros 4 años de ingreso de alumnos.

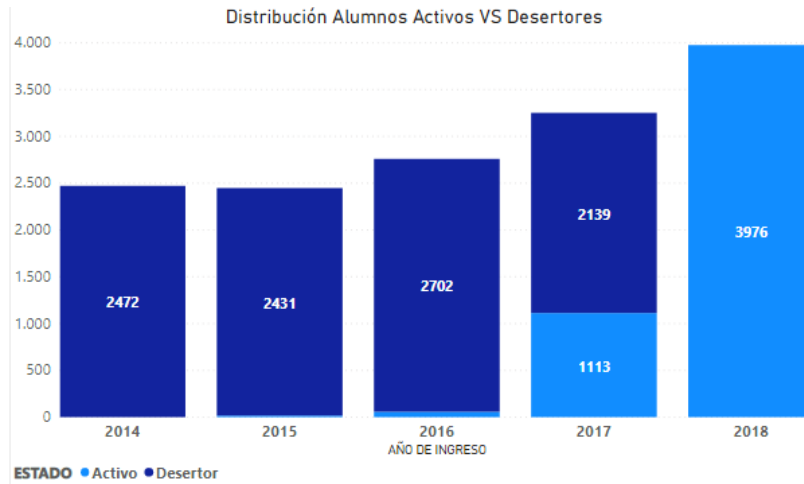


Figura 14 (Fuente creación propia)

8.3 Distribución de alumnos desertores

Figura 15, se puede observar la distribución por género de alumnos que desertaron en sus postgrados entre los años 2014 al 2017, quedando de la siguiente forma:

- Existe un Total de 9700 alumnos que desertaron de los programas de postgrado entre el 2014 al 2017
- 54% de los alumnos son mujeres desertoras
- 46% de los alumnos son hombres desertores

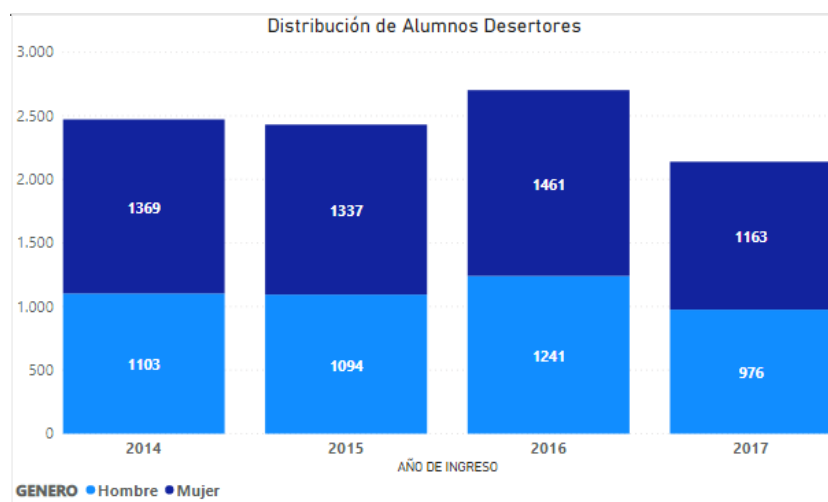


Figura 15 (Fuente creación propia)

8.4 Distribución alumnos desertores por sede

Figura 16 permite observar la distribución de alumnos desertores por sede. En Santiago existe la mayor cantidad de alumnos desertores de postgrado

- Las sedes de Viña del Mar y Valparaíso tienen un número similar de desertores

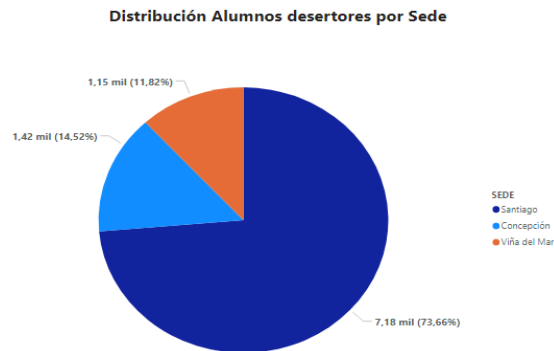


Figura 16 (Fuente creación propia)

8.5 Distribución desertor por género por sede

Figura 17, se observa la distribución por género para cada una de las sedes observadas arroja los siguientes resultados:

- Santiago concentra la mayor cantidad de alumnos desertores.
- En todas las sedes las alumnas mujeres concentran el mayor número de deserción en postgrado.

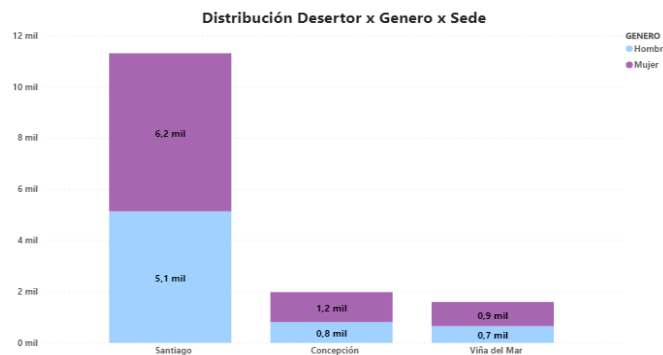


Figura 17 (Fuente creación propia)

9. Artículos Propuestos

9.1 Introducción

La deserción y la graduación tardía son problemas críticos en los programas de magíster en todas las disciplinas y universidades. Estas variables se pueden explicar con diferentes supuestos o apreciación del comportamiento de los estudiantes. Sin embargo, es relativamente poca la investigación que evalúa las motivaciones o percepciones de los estudiantes de magíster para abandonar los programas a los cuales están adscritos.

Existen varios estudios que buscan identificar las diferentes razones de la deserción estudiantil en pregrado y posgrado, incluyendo Cursos en Línea Masivos (MOOCs), estos últimos han tenido una explosiva y creciente aceptación por estudiantes que no requiere una exigencia académica previa, por tanto, su popularidad entre las universidades por brindar este tipo de estudios y de quienes se inscriben en ellos, ha aumentado de forma explosiva y considerable en los últimos años. Esta explosión repentina de los MOOCs también ha generado un aumento explosivo de la deserción de los estudiantes que se inscriben en ellos (Hanan A., Hosam A., Ibrahim A., Nasser A., 2020).

Un análisis de 1167 estudiantes de pregrado muestra una relación interesante respecto de los estudiantes que viajan diariamente versus los que residen en el campus, de acuerdo con esta relación los primeros tienden a funcionar mejor en los estudios, teniendo una probabilidad mayor de completar el programa de pregrado, a diferencia que el segundo grupo que su probabilidad es mayor a no completar el programa de pregrado (Anna B., Claudio D., Laura S., 2020).

9.2. Trabajos relacionados

Muchos investigadores y universidades se han interesado en estudiar el fenómeno de la graduación tardía y deserción de los estudiantes en los primeros años de estudios, utilizando para ello, diferentes modelos y técnicas de Machine Learning, regresión logística o algoritmos de redes neuronales, que permiten determinar en

algunos casos, la probabilidad de deserción para un grupo de estudiantes y en otro análisis determinar las causas que llevan a los jóvenes estudiantes de tomar la determinación de abandonar o retrasar su titulación. Para el caso de los países que pertenecen a la OCDE, las tasas de graduación en adultos jóvenes indica que el 39% completan su educación universitaria en el transcurso de sus vidas (Francesco A., Mauro M., Gianmarco B., 2020). En 2017 el 10,6% de los jóvenes de la Unión Europea abandonaron sus carreras en los primeros años de formación. Un estudio en la Universidad de Jerusalén encontró que las tasas de deserción de los estudiantes de maestrías son del 12% (Nir R., Gad Y., Elad S., 2020).

La deserción temprana de los estudiantes ha sido materia de estudios de investigadores y universidades, lograr generar programas desde las propias universidades para mejorar la retención y la titulación, ha sido otro punto importante en materia de estudio. En el ámbito de las universidades privadas, estas no están exentas del abandono temprano y la titulación tardía en alumnos de pregrado, postgrado y doctorados (Alejandra V., Gabriel L., (2020).

Las tasas de graduación se utilizan comúnmente para evaluar la calidad de las instituciones de educación superior. Mantener un adecuado equilibrio entre el número de nuevos estudiantes inscritos y el número de graduados cada año se ha convertido en un gran desafío para las universidades (Kelzang T., Nittaya M., Phattrawan T, 2019). La graduación oportuna a menudo se ignora porque una gran mayoría de los estudiantes no se gradúa dentro del tiempo estipulado.

Las instituciones deben estudiar y analizar estos factores que están presente en la realidad actual del mundo estudiantil, estas deben encontrar e identificar cual es mejor momento para intervenir e incorporar recursos que apoyen a los estudiantes para mantener los rangos bajos de deserción (José O., Antonio R., Paloma B., Martí C., 2020). Utilizando metodologías predictivas pueden implementar modelos que les permitan llegar a predecir con un grado de confianza suficientemente óptimo, cuando el mejor momento de intervenir. Con la utilización de procesos de minería de datos, es posible identificar los perfiles de los estudiantes de primer año que

tienen grandes dificultades para completar su primer año académico (Anne H., Michaël S., 2017).

Un punto de vista diferente pero no menos importante es la deserción de estudiantes de grados académicos en línea. utilizando técnicas similares, pero variables diferentes relacionadas con la demografía, características de las aulas, formas de participación cognitivas, académicas y conductuales, es posible predecir la deserción de los estudiantes (Kristof C., Minh P., Arno De C., Dries F., Annelies R., 2020), que, en términos comparativos, este grupo de estudiantes en línea tiene una tasa de deserción mucho mayor a los estudiantes de las aulas, estando entre el 25% y 90%.

Las direcciones de programa de magíster deben aumentar la sostenibilidad del desarrollo de capacidades en la mejora de los planes de estudio que proporcionará a los estudiantes una mayor oportunidad de aprendizaje (E. G. Publishing, E. Group, P. Limited, A. R. Reserved, O. Uri y E. Uri, 2017). En este sentido las instituciones de nivel superior que ofrecen estudios de postgrado, como magísteres, deben trabajar en el desarrollo y la mejora de los modelos educativos mediante la incorporación de tecnologías de la información y las comunicaciones que servirán como herramienta de apoyo a las distintas iniciativas propuestas por la dirección de programa.

Estas herramientas proporcionarán la capacidad de análisis que hoy las herramientas convencionales no son capaces de tratar por la gran cantidad de datos existentes. Para realizar un análisis, con grandes volúmenes de datos y variables, deben considerarse tecnologías como sistemas inteligentes, minería de datos, máquina de aprendizaje, entre otras, transformando los datos de educación en una nueva forma observable de información que sea más significativa y útil (S. Luj, 2018).

En (A. M. Shahiri, W. Husain y N. A. Rashid, 2015) las variables más utilizadas para alimentar el modelo predictivo aplicado al problema de deserción son las variables relacionadas con el rendimiento de los estudiantes, como el promedio acumulativo, las pruebas, el trabajo de laboratorio, los exámenes y la asistencia. También se utilizaron variables demográficas como la edad, el sexo, antecedentes familiares y

la discapacidad. Otros trabajos utilizan variables relacionadas a actividades extracurriculares como su red de interacción social. Árbol de decisión (decision tree), K – vecino más cercano (KNN), redes neuronales (Neuronal Network), Neive Bayes y máquina vectorial de apoyo (Support Vector Machine) fueron los algoritmos utilizados para generar los modelos predictivos aplicados al problema de deserción. De estos algoritmos los que obtuvieron un mayor rendimiento fueron la Red Neuronal con un 98% de precisión, y el Árbol de decisión con un 91% de precisión. El algoritmo de Árbol de Decisión es el más utilizado para desarrollar modelos predictivos cuyo objetivo es identificar el abandono de los estudiantes en las instituciones de educación superior (M. Alban, D. Mauricio, 2019). Una explicación recomendada de cómo funciona una red neuronal o un Árbol de decisión la pueden encontrar en (T. Barros, P. SouzaNeto, I. Silva y L. A. Guedes, 2019).

9.3. Métodos experimentales

9.3.1. Obtención de Base de Datos

Para el presente estudio se considera un dataset proporcionado por la universidad, que está compuesto por 14.903 registros con diferentes campos relacionados al alumno de la casa de estudios. De este universo de alumnos, seleccionaremos la Facultad de Economía y Negocios, con 2.020 registros como área de investigación considerando para ello dos fases de estudio.

Fase 1, análisis descriptivo, que examina la relación entre las características de los estudiantes de la Facultad de Economía y Negocios y las características del programa de Postgrado, donde presentan retraso en su proceso de graduación.

Fase 2, análisis predictivo, donde se utilizarán diferentes algoritmos de aprendizaje automático, que permitirán entregar información predictiva de la graduación tardía de los alumnos de la facultad de economía.

La información analizada corresponde a una muestra de estudiantes de los últimos 5 años.

9.3.2. Análisis descriptivo

Se realizará un análisis descriptivo respecto de la base de datos obtenida, relacionada con la Facultad de Economía y Negocios, permitiendo capturar las principales conclusiones relativas a la relación entre las variables de entrada y salida. Los resultados evaluarán la entrada y sus variables de salida para formar la base de datos final para el análisis predictivo.

De la distribución de alumnos por especialidad se aprecia en la tabla 1, que los programas que aportan con el 80% de alumnos registrados corresponden a los programas de Magíster en Administración de Empresas, Magister en Dirección de empresas (MBA Executive), Master en Dirección de Recursos Humanos y Habilidades Directivas y Magíster en Tributación.

Programa - Facultad Economía y Negocios	Alumnos	% Distribución
Magister en Administración de Empresas	762	37,7%
Magister en Direccion de Empresas (MBA Executive)	394	19,5%
Magister en Direccion de Recursos Humanos y Habilidades Directivas	274	13,6%
Magíster en Tributación	206	10,2%
Magister en Direccion Comercial y Marketing Estrategico	183	9,1%
Magíster en Finanzas Aplicadas	88	4,4%
Magister Internacional en Administración de Empresas	58	2,9%
Magíster en Dirección y Gestión de Empresas (MBA Executive)	25	1,2%
Master International Business (MIB)	20	1,0%
Magíster en Dirección y Gestión de Comunicación Marketing Estrategico	10	0,5%
TOTAL	2020	100%

Tabla 1. Cantidad de alumnos de postgrado por especialidad

Se puede apreciar en la figura 18, que el 23,4% de los alumnos de la Facultad de Economía y Negocios no lograron la titulación en los plazos establecidos como oportuno, mientras que el 76,6% si logro el objetivo. La Tabla 1 refleja la magnitud del problema que están enfrentando las universidades respecto a este fenómeno.

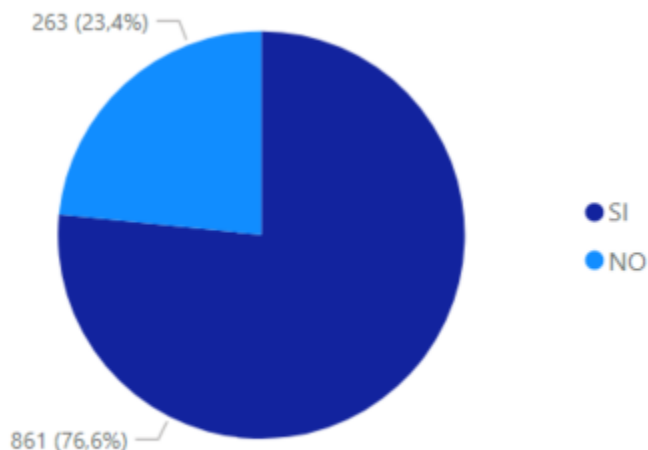


Figura 18 Distribución de Graduación tardía (Fuente creación propia)

Del universo de estudiantes que lograron y que no lograron su titulación en los plazos establecidos (Figura 18), podemos apreciar que la distribución de acuerdo con su género corresponde a un 48,49% de estudiantes hombre y un 28,11% de estudiantes mujeres, que, sí lograron titularse dentro de los plazos establecidos, mientras que el 14,32% de estudiantes hombre y el 9,07% de estudiantes mujeres, no lograron titularse oportunamente. (Figura 19).

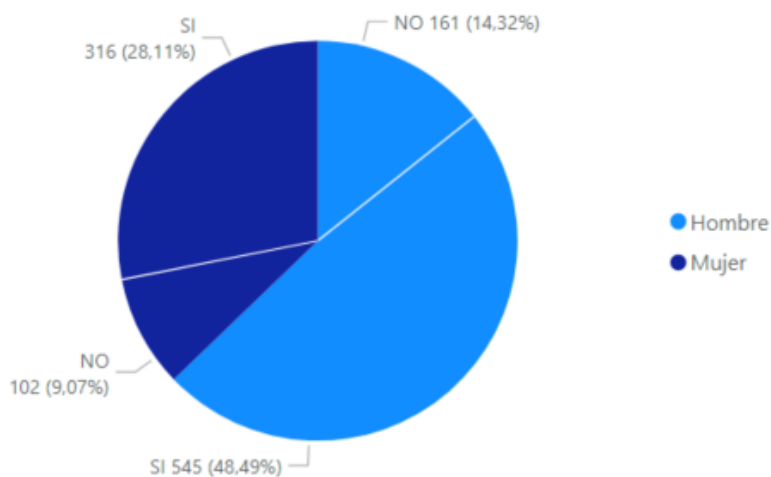


Figura 19 Distribución de Graduación tardía por género (Fuente creación propia)

Otro de los factores interesantes a analizar está relacionado con la modalidad de estudio, lo que podría entregar información relevante respecto al nivel de dedicación de los estudiantes de acuerdo con su programa de estudios y su modalidad (Figura 20). Se aprecia que la modalidad vespertina aporta la mayor cantidad de alumnos con un 67,26% de alumnos con titulación dentro de los plazos, mientras que el 21,89% de los alumnos de esta modalidad, corresponden a titulación tardía, como consecuencia la modalidad de estudio vespertino aporta también con el mayor número de alumnos con titulación tardía.

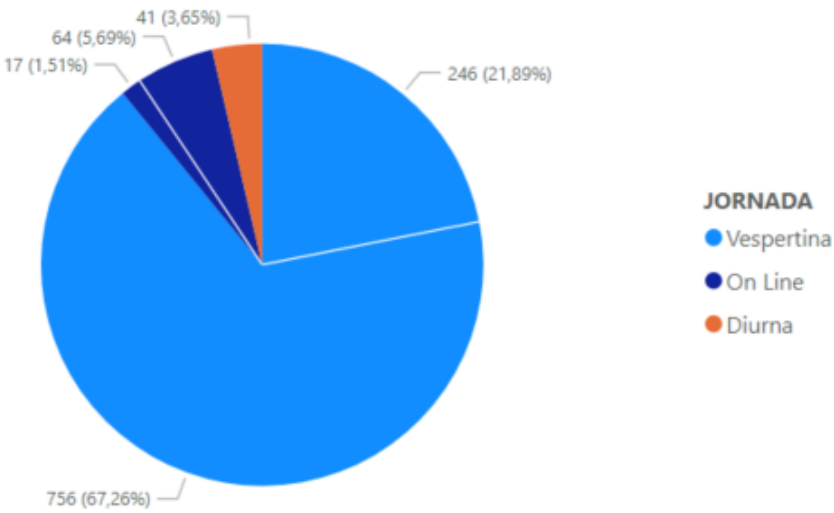


Figura 20. Distribución de Graduación tardía por modalidad de estudio (Fuente creación propia)

El análisis anterior nos presenta una realidad importante a ser considerada por la casa de estudios, debido al alto porcentaje de casos que alumnos que están manifestando un proceso de graduación tardía. En promedio un 30% de alumnos de la facultad de Economía y Negocios, son parte de las estadísticas de graduación tardía, en la casa de estudios Universidad Nacional Andrés Bello.

9.3.3. Análisis predictivo

Para el análisis, se han preparado seis experimentos, los cuales están programados en lenguaje R.

Utilizando la base de datos de la sección anterior, se han diseñado seis experimentos para detectar de forma temprana y oportuna los tiempos de graduación de los estudiantes. Si este problema se pudiese detectar al principio del ciclo académico, los estudiantes pueden recibir apoyo de forma oportuna, que permita corregir su condición a través de procesos de mitigación, con la finalidad de asegurar la graduación oportuna. En la actualidad, este tipo de intervenciones se aborda como una medida de contingencia, de forma reactiva, y a menudo llega demasiado tarde.

Para los experimentos, se utilizarán seis algoritmos de aprendizaje automático supervisados para generar los modelos de predicción: Regresión Logística, Máquina de vector de soporte, Árbol de decisión, Naive Bayes, Random Forest y K Vecinos más cercanos. Los resultados de los diferentes modelos se compararán para seleccionar aquel que presente mejor desempeño. Las descripciones de todos estos algoritmos de aprendizaje automático se pueden encontrar en el Sitio web de R-Project.

9.3.4. Entrenamiento y Prueba del Modelo

La base de datos será dividida en dos secciones, primero está el set de pruebas seguido del set de entrenamiento. La distribución corresponde al 80% para el set de entrenamiento y un 20% para el set de pruebas.

El modelo general considera variables de entrada y objetivo, aplicando técnicas de aprendizaje automático, nos puede entregar como resultado una predicción que finalmente si esta es óptima y con un nivel de confianza aceptable, pasa a ser el objetivo final del modelo. En Figura 21 podemos apreciar un modelo de aprendizaje automático con variables de entradas y salidas.

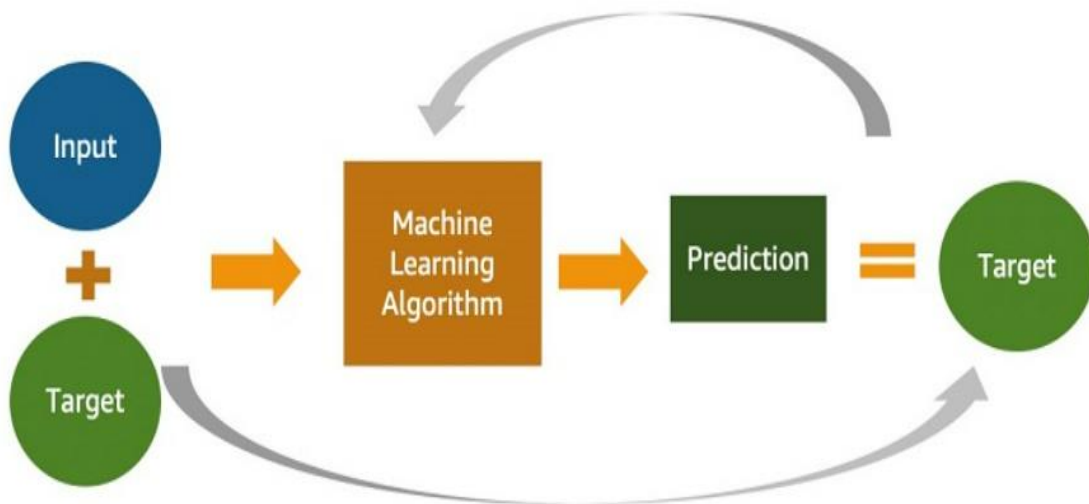


Figura 21 (fuente <https://aws.amazon.com/es/blogs/big-data/create-train-and-deploy-machine-learning-models-in-amazon-redshift-using-sql-with-amazon-redshift-ml/>)

La base de datos se divide en 3 variables de entrada para el set de entrenamiento, y una variable de salida, esta última con 4 estados, los cuales son:

- Graduación titulación-oportuna-verdadero (GTOV)
- Graduación titulación-oportuna-Falso (GTOF)
- No Graduado titulación-oportuna-verdadero (NGTOV)
- No Graduado titulación-oportuna-Falso (NGTOF)

De acuerdo con lo observado en la Tabla 2, es posible describir las variables de entrada, las que corresponden a los siguientes valores:

VARIABLE	DESCRIPCIÓN	VALOR
EST_EGRASADO	Esta variable contiene el estado de egreso de los alumnos, egresados para los alumnos que cursaron sus asignaturas y no egresados para aquellos aún está pendiente.	Egresado: Correspondiente a un alumno que curso todas sus asignaturas. NoEgresado: Correspondiente a un alumno que no curso todas sus asignaturas.
EST_GRADUADO	Esta variable está compuesta por los valores asociados a estado de titulado (graduado) o no titulado de los alumnos.	Graduado: Corresponde a los alumnos que se encuentran titulados. NoGraduado: Corresponde a los alumnos que se no encuentran titulados.
ALCANZA_ANO_APORT-TITULACIÓN OPORTUNA	Esta variable corresponde al estado de titulación oportuna de un alumno, es decir, si cumple con el criterio de finalización de su postgrado en los años asignados sin presentar ningún retraso.	SI: Correspondiente a alumno que cumple con su titulación dentro del plazo estipulado por la institución. NO: Correspondiente a alumno que no cumple con su titulación dentro del plazo estipulado por la institución.

Tabla 2. Variables de entrada para ejecución modelo predictivo

9.4. Resultados y Discusión

9.4.1. Regresión Logística (LR)

Los resultados de la aplicación del algoritmo Regresión Logística, es posible ver en Tabla 3, comparados contra el set de pruebas. Se considera Graduación titulación-oportuna (GTO) y No Graduación titulación-oportuna (NGTO)

Set de Prueba	Predicciones	
	GTO	NGTO
GTO	92	69
NGTO	29	214

Tabla 3 Matriz de Confusión Regresión Logística

Los resultados indican de un total de 404 predicciones, 98 de ellas eran completamente erróneas, mientras que 306 predicciones fueron consideradas correctas. Para encontrar el grado de precisión del modelo, se utilizará la ecuación de la figura 22.

$$Accuracy = \frac{VP + VN}{VP + FP + FN + VN}$$

Figura 22 Ecuación para encontrar la precisión del modelo

Donde:

- VP es el número de resultados Verdaderos Positivos
- VN es el número de resultados Verdaderos Negativos
- FP es el número de resultados Falsos Positivos
- FN es el número de resultados Falsos Negativos

De acuerdo con la ecuación de la Figura 22, el grado de precisión para el modelo Regresión Logística obtenido es del 75,74%. (Figura 23)

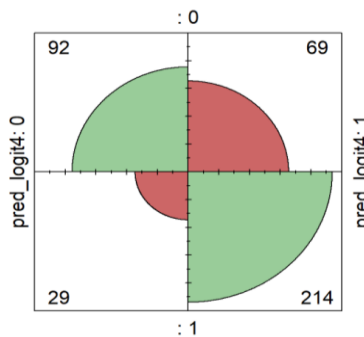


Figura 23 Gráfico Matriz de confusión Regresión Logística, Accuracy: 0,7574

La figura 24 muestra la curva ROC del modelo Regresión Logística, observándose una curva pronunciada, lo que indica que su efectividad no es lo suficientemente buena para el modelo. Mientras más cercano al 1 se grafique la curva ROC, mayor será su nivel de precisión.

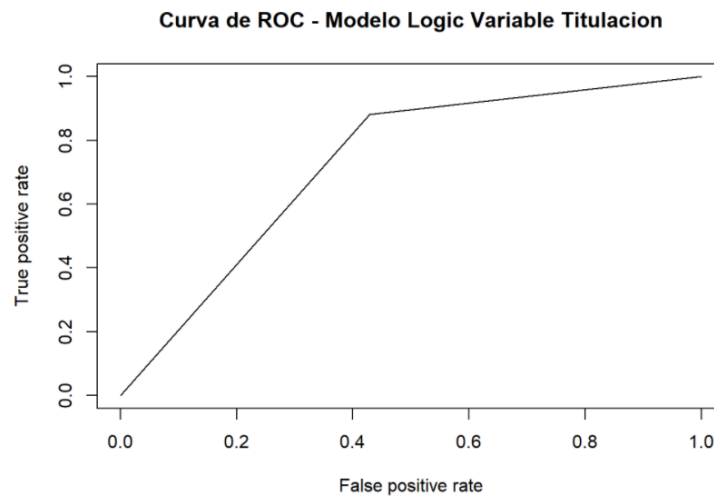


Figura 24 Curvas ROC – Regresión Logística

9.4.2. Máquina de vector de soporte (SVM)

Los resultados de la aplicación del algoritmo Máquina de vector de soporte, es posible ver en Tabla 4, comparados contra el set de pruebas. Se considera Graduación titulación-oportuna (GTO) y No Graduación titulación-oportuna (NGTO)

Set de Prueba	Predicciones	
	GTO	NGTO
GTO	120	41
NGTO	22	221

Tabla 4 Matriz de Confusión SVM

Los resultados indican de un total de 404 predicciones, 63 de ellas eran completamente erróneas, mientras que 341 predicciones fueron correctas.

De acuerdo con la ecuación de la Figura 21, el grado de precisión para el modelo Regresión Logística obtenido es del 84,41%. (Figura 25).

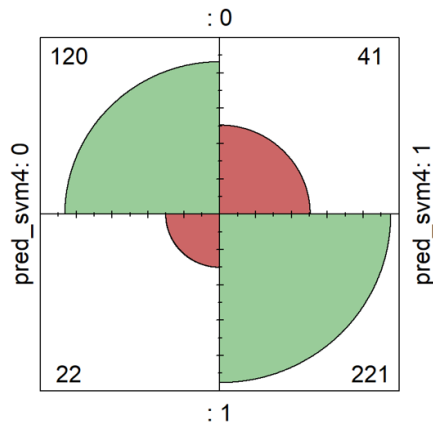


Figura 25 Gráfico Matriz de confusión SVM, Accuracy: 0,8441

La figura 26 muestra la curva ROC del modelo Máquina de Vector de Soporte (SVM), observándose una curva menos pronunciada que el modelo **RL**, lo que indica que su efectividad es ligeramente mayor para este modelo. Mientras más cercano al 1 se grafique la curva ROC, mayor será su nivel de precisión.

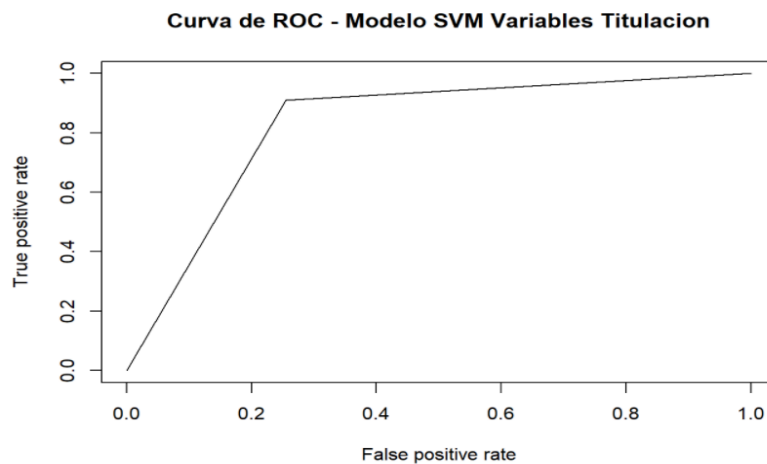


Figura 26 Curvas ROC – SVM

9.4.3. Árbol de decisión (Decision Tree, DT)

Los resultados de la aplicación del algoritmo Árbol de Decisión, es posible ver en Tabla 5, comparados contra el set de pruebas. Se considera Graduación titulación-oportuna (GTO) y No Graduación titulación-oportuna (NGTO)

Set de Prueba	Predicciones	
	GTO	NGTO
GTO	120	41
NGTO	45	198

Tabla 5 Matriz de Confusión DT

Los resultados indican de un total de 404 predicciones, 86 de ellas eran completamente erróneas, mientras que 318 predicciones fueron correctas.

De acuerdo con la ecuación de la Figura 21, el grado de precisión para el modelo Árbol de Decisión obtenido es del 78,71%. (Figura 27).

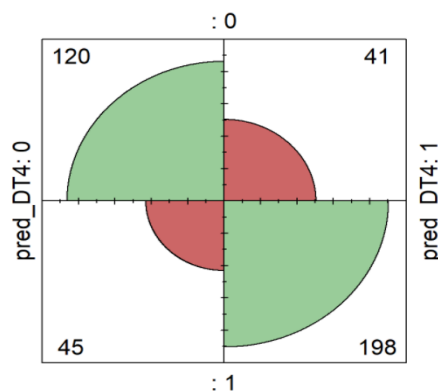


Figura 27 Gráfico Matriz de confusión DT, Accuracy: 0,7871

La figura 28 muestra la curva ROC del modelo Árbol de Decisión, observándose una curva más pronunciada que el modelo **SVM**, lo que indica que su efectividad es ligeramente menor para este modelo. Mientras más cercano al 1 se grafique la curva ROC, mayor será su nivel de precisión.

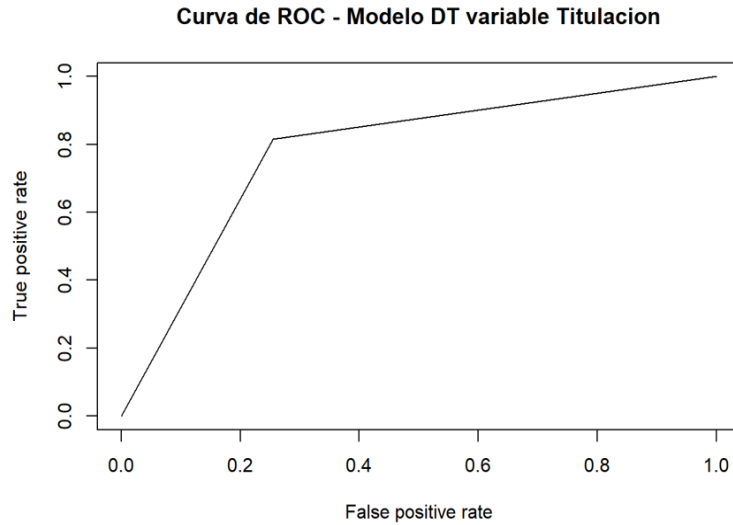


Figura 28 Curvas ROC – DT

9.4.4. Nayve Bayes (NB)

Los resultados de la aplicación del algoritmo Nayve Bayes, es posible ver en Tabla 6, comparados contra el set de pruebas. Se considera Graduación titulación-oportuna (GTO) y No Graduación titulación-oportuna (NGTO)

Set de Prueba	Predicciones	
	GTO	NGTO
GTO	161	0
NGTO	199	44

Tabla 6 Matriz de Confusión NB

Los resultados indican de un total de 404 predicciones, 199 de ellas eran completamente erróneas, mientras que 205 predicciones fueron correctas.

De acuerdo con la ecuación de la Figura 21, el grado de precisión para el modelo Nayve Bayes obtenido es del 50,74%. (Figura 29).

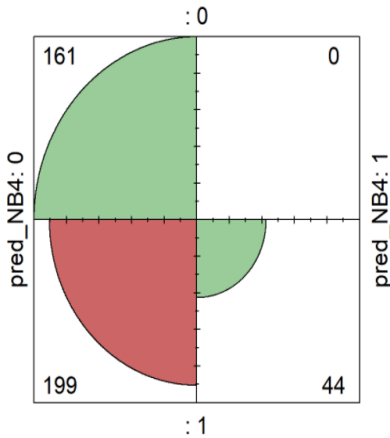


Figura 29 Gráfico Matriz de confusión NB, Accuracy: 0,5074

La figura 30 muestra la curva ROC del modelo Regresión Logística, observándose una curva muy pronunciada respecto a todos los anteriores modelos, lo que indica que su efectividad es muy por debajo respecto a lo esperado. Mientras más cercano al 1 se grafique la curva ROC, mayor será su nivel de precisión.

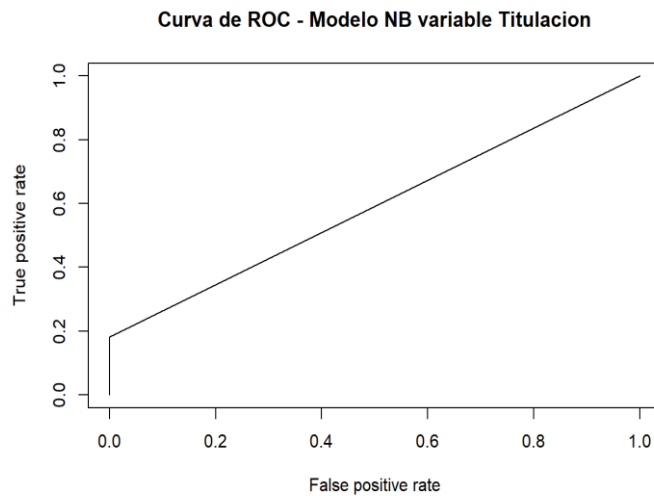


Figura 30 Curvas ROC – NB

9.4.5. Random Forest (RF)

Los resultados de la aplicación del algoritmo Random Forest, es posible ver en Tabla 7, comparados contra el set de pruebas. Se considera Graduación titulación-oportuna (GTO) y No Graduación titulación-oportuna (NGTO)

Set de Prueba	Predicciones	
	GTO	NGTO
GTO	127	34
NGTO	19	224

Tabla 7 Matriz de Confusión RF

Los resultados indican de un total de 404 predicciones, 53 de ellas eran completamente erróneas, mientras que 351 predicciones fueron correctas.

De acuerdo con la ecuación de la Figura 21, el grado de precisión para el modelo Random Forest obtenido es del 86,88%. (Figura 31).

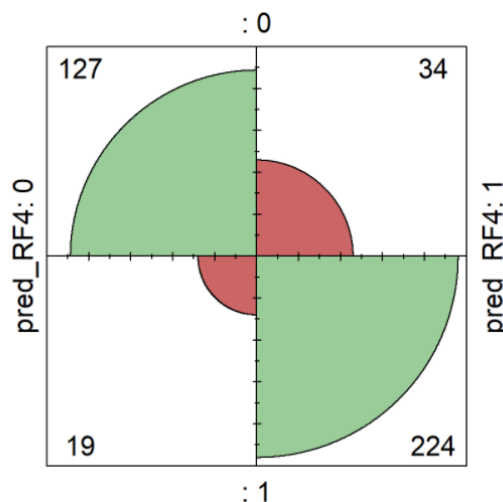


Figura 31 Gráfico Matriz de confusión RF, Accuracy: 0,8688

La figura 32 muestra la curva ROC del modelo Random Forest, observándose una curva menos pronunciada respecto a los anteriores modelos, lo que indica que su efectividad es mayor para este modelo. Mientras más cercano al 1 se grafique la curva ROC, mayor será su nivel de precisión.

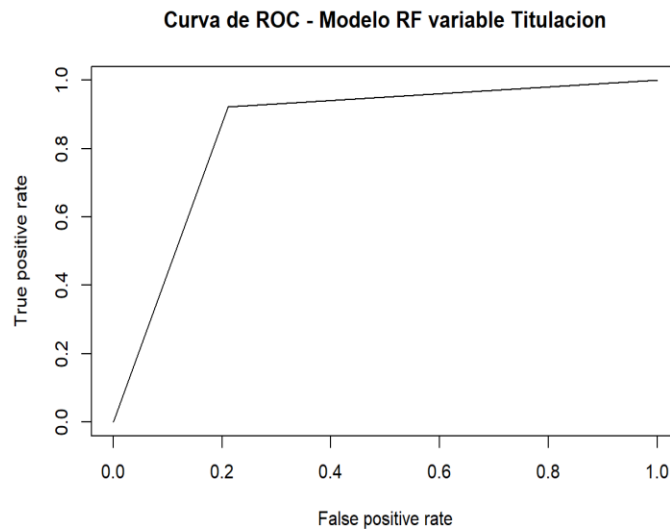


Figura 32 Curvas ROC – RF

9.4.6. K vecinos más cercanos (KNN)

Los resultados de la aplicación del algoritmo K Vecinos más Cercanos, es posible ver en Tabla 8, comparados contra el set de pruebas. Se considera Graduación titulación-oportuna (GTO) y No Graduación titulación-oportuna (NGTO)

Set de Prueba	Predicciones	
	GTO	NGTO
GTO	121	40
NGTO	19	224

Tabla 8 Matriz de Confusión KNN

Los resultados indican de un total de 404 predicciones, 59 de ellas eran completamente erróneas, mientras que 345 predicciones fueron correctas.

De acuerdo con la ecuación de la Figura 21, el grado de precisión para el modelo K Vecinos más Cercanos obtenido es del 85,4%. (Figura 33).

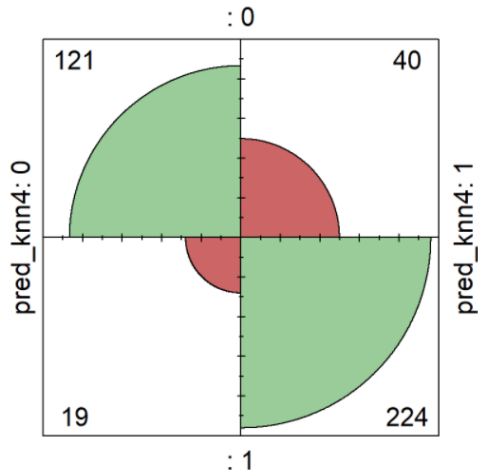


Figura 33 Gráfico Matriz de confusión KNN, Accuracy: 0,854

La figura 34 muestra la curva ROC del modelo K Vecinos más Cercanos, observándose una curva ligeramente más pronunciada que el modelo **RF**, lo que indica que su efectividad es ligeramente menor para este modelo. Mientras más cercano al 1 se grafique la curva ROC, mayor será su nivel de precisión.

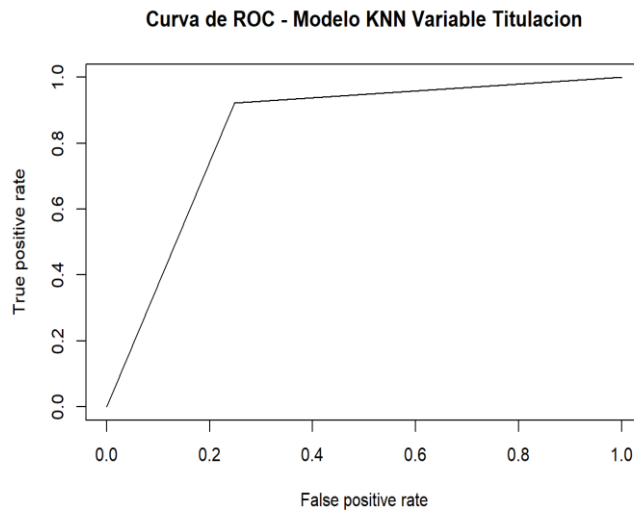


Figura 34 Curvas ROC – KNN

9.5. Comparación de resultados

Como resultado de las observaciones, mediante la ejecución de modelos de aprendizaje automático, para el set de entrenamiento, podemos observar para la columna accuracy de la Tabla 9, la cual nos indica el grado de precisión que obtuvieron como resultado cada uno de los modelos de predicción utilizados en este experimento.

Modelos	Predicciones			Elección Modelo
	VP+VN	FP+FN	Accuracy	
LR	306	98	75,74%	NO
SVM	341	63	84,41%	NO
DT	318	86	78,71%	NO
NB	205	199	50,74%	NO
RF	351	53	86,88%	SI
KNN	345	59	85,40%	SI

Tabla 9 Resultado de precisión de modelos predictivos

Figura 35, muestra la relación entre los resultados de los valores Verdaderos positivos (VP) más los resultados de los valores de verdaderos negativos (VN), versus los resultados de los valores de falsos positivos (FP) más los resultados de los valores de falsos negativos (FN).

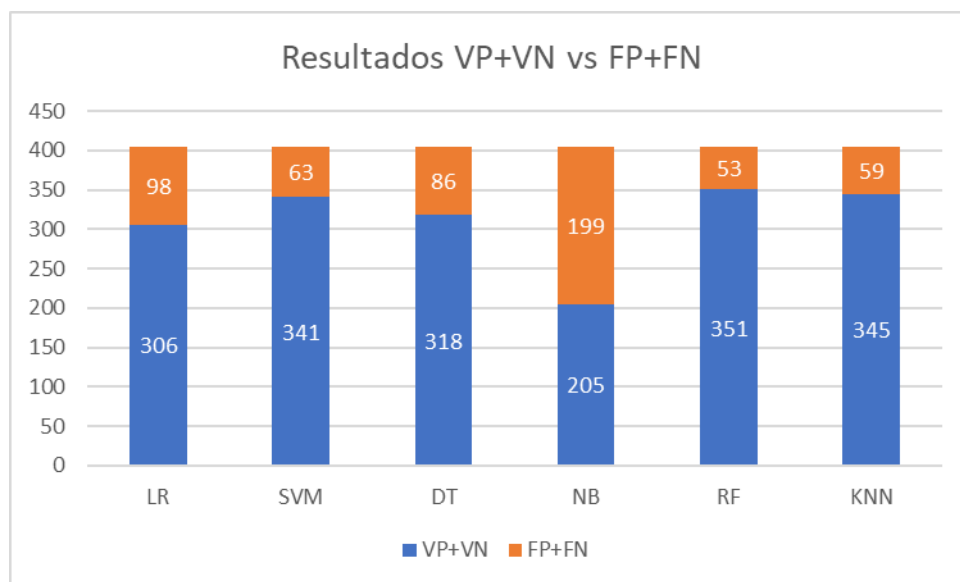


Figura 35 Verdaderos (+) + Verdaderos (-) vs Falsos (+) + Falsos (-)

En la figura 36 podemos apreciar una comparativa de los niveles de Accuracy para cada uno de los algoritmos trabajados en las secciones anteriores.

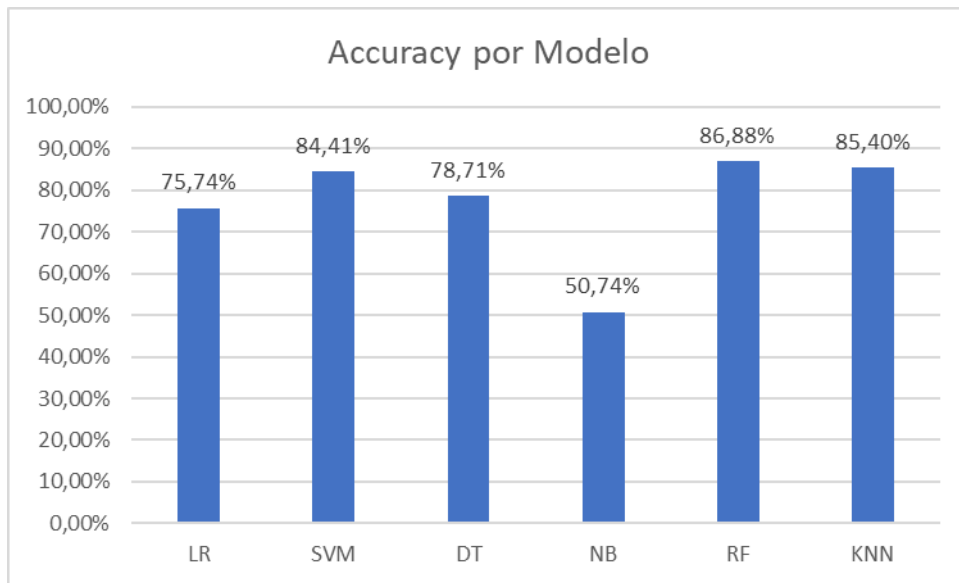


Figura 36 Nivel de Accuracy por modelo

9.6. Conclusiones

Se puede observar del conjunto de resultados analizados para los diferentes modelos de Aprendizaje Automático, que el modelo **NB** cuenta con la precisión más baja del conjunto de modelos, con un 50,74% de precisión, claramente es un valor muy alejado respecto al esperado, considerando como valor esperado lo más cercano al 90% de precisión.

Respecto a los modelos **LR**, **DT** y **SVM**, si bien sus valores de precisión no fueron satisfactorios, estos presentaron resultados mucho mayores que el proceso **NB**, a pesar de lo anterior siguen siendo valores no aceptables, por tanto, quedan fuera de la selección como modelos a utilizar para la detección temprana de graduación tardía.

Finalmente, los **RF** y **KNN** fueron los que obtuvieron una mejor precisión, con una exactitud de 86,88% y 85,40% respectivamente. De acuerdo con los resultados descritos anteriormente, podemos concluir que estos dos modelos nos pueden

entregar buenos resultados a la hora de predecir de forma temprana graduación tardía de los alumnos de postgrado.

Como resultado final, podemos indicar que, para esta investigación, el modelo **RF** es el que entrega un porcentaje de error más bajo con un 13,12% de error en su calidad de predicción. Esto nos indica que el estudio en general cumple con el objetivo propuesto, logrando una predicción del 86% al utilizar el modelo Random Forest.

10. Conclusiones Proyecto de Grado

En la presente tesis de investigación se diseñaron modelos de predicción utilizando aprendizaje automático – Machine Learning – con la finalidad de buscar un mecanismo que permita alertar de forma temprana a las instituciones académicas, respecto a cuáles podrían ser los potenciales alumnos que presentan características de sufrir una graduación de forma tardía. La base del estudio pretende entregar herramientas preliminares de análisis que permita a las instituciones educacionales indagar con mayor profundidad mecanismos que logren disminuir este factor, y como efecto directo, contribuir en el aumento de los niveles de prestigio y eficiencia en sus programas educacionales, incrementando así la cantidad de profesionales certificados en las materias cursadas durante su periodo de postgrado.

En la hipótesis, hemos planteado que el uso de la tecnología como herramienta de apoyo para avanzar en la búsqueda de nuevas formas de ayuda hacia la comunidad estudiantil, el uso del aprendizaje automático forma parte de las nuevas tecnologías que están en pleno desarrollo, y pueden perfectamente ser aplicadas en el mundo de la educación.

El aporte de esta investigación junto al desarrollo práctico de la hipótesis, nos ha demostrado que el uso del aprendizaje automático – Machine Learning – es un campo que puede ser ampliamente explorado y explotado para fines que permitan beneficiar tanto las economías de las entidades de educación y su eficiencia en sus programas impartidos, como las inversiones de los propios alumnos que como meta debiese ser siempre el lograr obtener el grado académico.

Como resultado final, podemos concluir que el campo de la Inteligencia Artificial, específicamente en el área de la educación, aún tiene mucho que aportar, es necesario continuar estudiando e investigando diferentes usos y aplicaciones en este campo, que permitan a las instituciones educacionales generar nuevos y modernos métodos de estudio y apoyo a la comunidad estudiantil.

11. Referencias

Hanan Aldowah, Hosam Al-Samarraie, Ahmed Ibrahim Alzahrani & Nasser Alalwan, (2020). "Factores que afectan la deserción de los estudiantes en los MOOCs: un modelo de toma de decisiones de causa y efecto", *Journal of Computing in Higher Education*, pp, 429-454.

Anna Bussu, Claudio Detotto and Laura Serra, (2020). "Indicators to prevent university drop-out and delayed graduation: an Italian case", *Journal of Applied Research in Higher Education*, Vol 12, n°2, pp.230-249.

Rafaela Alfalla-Luquea, Pedro Garrido-Vegab, María del Mar González-Zamorac, Carmen Medina-Lópezd, Macarena Sacristán-Díaze, (2017). "An analysis of the causes of Master's Thesis dropouts", *Working Papers on Operations Management*, Vol. 8, pp. 32-49.

Bibiana A. Castro-Montoya, Rubén D. Manrique-Hernández, Difariney Gonzalez-Gómez, Angela M. Segura-Cardona, (2020). "Trayectoria académica y factores asociados a graduación, deserción y rezago en estudiantes de programas de pregrado de una universidad privada de Medellín", *Formación Universitaria* Vol. 13(1), pp. 43-54.

Francesco Agrusti, Mauro Mezzini, Gianmarco Bonavolontà, (2020). "Deep learning approach for predicting university dropout: a case study at Roma Tre University", *Journal of e-learning and knowledge society*, Vol. 16, No. 01, pp. 44-54, 2020.

Nir Rotem , Gad Yair & Elad Shustak, (2020). "Dropping out of master's degrees: objective predictors and subjective reasons", <https://doi.org/10.1080/07294360.2020.1799951>

Alejandra Valencia Vargas, Gabriel Jaime López Palacio, (2020). "Student Dropout at a Private University: A Phenomenon Not Foreign to Graduate Students. A

Quantitative Assessment Based on Survival Analysis”, *Revista Educación*, 44(2).

Kelzang Tentsho, Nittaya McNeil, Phattrawan Tongkumchum, (2019). “Examining timely graduation rates of undergraduate students”, *Journal of Applied Research in Higher Education*, Volume 1.

José María Ortiz-Lozano, Antonio Rua-Vieites, Paloma Bilbao-Calabuig & Martí Casadesús-Fa, (2020). “University student retention: Best time and data to identify undergraduate students at risk of dropout”, *INNOVATIONS IN EDUCATION AND TEACHING INTERNATIONAL*, VOL. 57, NO. 1, 74–85.

Anne-Sophie Hoffait, Michaël Schyns, (2017). “Early detection of university students with potential difficulties”, *HEC Management School, University of Liège, Decision Support Systems* 101 pp.1–11.

Kristof Coussementa, Minh Phana, Arno De Caigny, Dries F. Benoit, Annelies Raes, (2020). “Predicting student dropout in subscription-based online learning environments: The beneficial impact of the logit leaf model”, *Decision Support Systems*, 135.

E. G. Publishing, E. Group, P. Limited, A. R. Reserved, O. Url y E. Uri, (2017). “Catalysing Change in Higher Education for Sustainable Development : A Review of Professional Development Initiatives for University Educators”, *International Journal of Sustainability in Higher Education*, vol. 18, nº 5, p. 798.

S. Luj, (2018). “Big Data, the Next Step in the Evolution of Educational Data Analysis” *Icits*, vol. 1.

A. M. Shahiri, W. Husain y N. A. Rashid, (2015). “A Review on Predicting Student’s Performance Using Data Mining Techniques”, *Procedia Comput. Sci.*, vol. 72, p. 414–422.

M. Alban, D. Mauricio, D. Mauricio, National University of San Marcos y P. Artificial Intelligence Group, (2019). “Predicting University Dropout through Data Mining:

- A systematic Literature”, Indian Journal of Science and Technology, vol. 12, n° 4, p. 1–12.
- T. Barros, P. SouzaNeto, I. Silva y L. A. Guedes, (2019). “Predictive Models for Imbalanced Data: A School Dropout Perspective”, Education Sciences, vol. 9, p. 275.
- I. Lykourantzou, I. Giannoukos, V. Nikolopoulos, G. Mpardis y V. Loumos, (2009). “Dropout prediction in e-learning courses through the combination of machine learning techniques”, Comput. Educ, vol. 53, p. 950–965.
- V. R. D. C. Martinho, C. Nunes y C. R. Minussi, (2013). “An intelligent system for prediction of school dropout risk group in higher education classroom based on artificial neural networks”, IEEE 25th International Conference on Tools with Artificial Intelligence, p. 159–166.
- A. A. Mubarak, H. Cao y W. Zhang, (2020). “Prediction of students’ early dropout based on their interaction logs in online learning environment”, Interactive Learning Environments, pp. 1-20.
- S. B. Kotsiantis, C. J. Pierrakeas y P. E. Pintelas, (2003). “Preventing student dropout in distance learning using machine learning techniques”, International conference on knowledge-based and intelligent information and engineering systems, p. 267–274.
- K. T. Chui, D. C. L. Fung, M. D. Lytras y T. M. Lam, (2018). “Predicting at-risk university students in a virtual learning environment via a machine learning algorithm”, Computers in Human Behavior, vol. 13, n° 1, p. 63–75.
- C. Burgos, M. L. Campanario, D. de la Peña, J. A. Lara, D. Lizcano y M. A. Martínez, (2018). “Data mining for modeling students’ performance: A tutoring action plan to prevent academic dropout”, Comput. Electr. Eng., vol. 66, p. 541–556.

- R. Asif, A. Merceron, S. A. Ali y N. G. Haider, (2017). "Analyzing undergraduate students' performance using educational data mining", *Comput. Educ.*, vol. 113, p. 177–194.
- Hernández R., & F. (2010). *Metodología de la Investigación* (5ta. Edición). México: Ed. Mc Graw Hill.
- Selltiz, J. D. (1980). *Métodos de investigación en las relaciones sociales*. Madrid: Rialp.
- Ackoff, R. L. (1967). *Management Misinformation Systems*. *Management Sciences*, Vol. 14, No. 4.
- (Eds.), R. G. (2005). *Social work: Research and evaluation*. New York, NY, EE.UU: Oxford University Press.
- Hernández Sampieri, R. y. (2009). *Marco teórico*. Guanajuato, México: Universidad de Celaya.
- Rojas Soriano, R. (2002). *Guía para realizar investigaciones sociales* 34a Ed. México: Plaza y Valdés.