

# **Prototipo de sistema de aprendizaje automático de apoyo en la identificación de riesgos que llevan al fracaso escolar a los alumnos del IPN**

***Trabajo Terminal No.2020 -A125***

***Alumnos: \*Domínguez Morán Joaquín, Mendez Alvarez Jorge Luis, Ochoa Ginera Carlos Armando.***

***Directores: Sánchez Garfias Flavio Arturo.***

***Mendoza Macías Elba.***

***\*e-mail: jdominguezm1300@alumno.ipn.mx***

**Resumen** - Se desarrollará un prototipo de sistema con algoritmos de aprendizaje automático que identifique situaciones de riesgo de los alumnos de la ESCOM que provocan fracaso escolar, con el fin de proveer a las autoridades del plantel de la información necesaria para que puedan tomar decisiones en beneficio del estudiantado.

**Palabras clave** - Aprendizaje automático, fracaso escolar, deserción escolar.

## **1.Introducción**

La educación superior se ha concebido a lo largo del siglo XXI como un componente principal para el desarrollo de la sociedad tal y como se planteó a principios del siglo en la Conferencia Mundial sobre la Educación Superior para el Siglo XXI (París, noviembre de 1998) donde se enfatizó que la educación superior “debe reforzar sus funciones de servicio a la sociedad, y más concretamente sus actividades encaminadas a erradicar la pobreza, la intolerancia, la violencia, el analfabetismo, el hambre, el deterioro del medio ambiente y las enfermedades”[1].

Sin embargo, la educación superior ha presentado problemas para lograr cumplir su función en la sociedad como es por un lado, la deserción escolar, la cual incluye situaciones como el abandono involuntario, dejar una carrera para iniciar otra, no concluir la formación universitaria para incorporarse al mundo laboral o interrumpirla por un periodo de tiempo y por otro lado el fracaso escolar que implica el retraso y las calificaciones reprobatorias obtenidas durante la trayectoria escolar.[2]

Dichos problemas antes mencionados, se han hecho presentes en las diferentes universidades alrededor del mundo, como es el caso de América Latina, tal y como señala la UNESCO solo uno de cada 10 jóvenes de 25 a 29 años de edad había completado cinco años de educación superior para el 2010. Y aunque en México la educación superior se ha ido ampliando como muestran las cifras entre los años 2016 y 2017 en los cuales había 4.4 millones de estudiantes matriculados en instituciones de nivel superior, se estima que solo 26% de los jóvenes mexicanos obtienen un título de educación superior en algún momento de su vida [3].

En el Anuario Estadístico del Instituto Politécnico Nacional (IPN) del año 2018 se presentaron estadísticas en las que se muestra que el 39.24% de los estudiantes están en situación reprobatoria, el 7.24% se encuentran en abandono definitivo en la modalidad escolarizada y la eficiencia terminal en el periodo entre 2017 y 2018 en el IPN fue del 51.56%.

Así mismo, en nuestro foco de interés, en el área de ingeniería y ciencias físico matemático se tiene una eficiencia terminal del 46.81%, lo que se traduce en que solo uno de cada dos estudiantes que ingresan terminan la carrera. Por otro lado, en el caso particular de la Escuela Superior de Cómputo (ESCOM) solo el 35.66% logra graduarse y el 50.07% de la población estudiantil se encuentra en situación académica irregular, es decir, que tienen al menos una unidad de aprendizaje reprobada. También es necesario puntualizar que 7.35% de los estudiantes abandonan de forma definitiva la carrera. [4]

Se considera que la ciencia de datos es una herramienta de utilidad para abordar problemáticas de esta índole, pues aprovecha la información que ofrecen los volúmenes de datos, mediante el uso de técnicas variadas de aprendizaje automático, análisis de la información y reconocimiento de patrones. Lo anterior se fundamenta en anteriores investigaciones realizadas por diversas instituciones que a continuación presentaremos con el objetivo de que sirvan como referencia para nuestro trabajo y la comparación de estos con nuestra solución.

<b>Tipo</b>	<b>Título</b>	<b>Descripción</b>	<b>Comparación con nuestra solución</b>
Investigación	<i>Using data mining to predict secondary school student performance.</i>	Se trató de predecir el desempeño de los estudiantes de educación superior de Portugal utilizando técnicas de Business Intelligence (BI) y Data Mining (DM), en dos materias principales que fueron matemáticas y portugués.[5]	En el artículo solo se hace enfoque a dos materias específicas tratando de optimizar los resultados en estas mientras que en nuestra solución se busca encontrar maneras de mejorar el desempeño escolar en general.
Investigación	<i>Predicting the Efficiency of Colombian Higher Education Institutions with Data Envelopment Analysis and Data Mining</i>	El propósito de esta investigación fue evaluar la eficiencia técnica de las instituciones de educación superior en Colombia mediante la aplicación del análisis envolvente de datos y técnicas de minería de datos. [6]	El enfoque de este proyecto está totalmente dirigido a las autoridades académicas para tomar decisiones sobre políticas educativas mientras que nuestra solución también tiene un enfoque estudiantil.
Investigación	<i>Accuracy Comparison of Machine Learning Algorithms for Predictive Analytics in Higher Education</i>	El objetivo fue hacer una combinación de algoritmos clásicos de aprendizaje automático como Naive Bayes y Random Forest con varios métodos de conjunto, para detectar patrones y predecir posibles factores de abandono escolar. [7]	El propósito de este proyecto es identificar herramientas de aprendizaje automático que reduzcan exclusivamente la cantidad de alumnos que abandonan la escuela mientras que nuestra solución busca mejorar el nivel de aprendizaje.
Aplicación	<i>Preventing Student Dropout in Distance Learning Using Machine Learning Techniques</i>	El propósito de esta investigación fue estudiar si el uso de técnicas de aprendizaje automático es útil para reducir la cantidad de alumnos que abandonan sus estudios a distancia, el resultado fue un prototipo de herramienta de soporte basada en la WEB, que puede reconocer automáticamente a los estudiantes con alta probabilidad de deserción basado en estos algoritmos.[8]	Las variables a estudiar en este proyecto son distintas debido a que la modalidad es no escolarizada.

**Tabla 1. Cuadro comparativo**

**Fuente: Creación propia**

## **2. Objetivo.**

Implementar un prototipo de sistema de aprendizaje automático que identifique posibles riesgos de caer en situaciones que lleven a la deserción y fracaso escolar los estudiantes de la carrera de Ingeniería en Sistemas Computacionales (ISC) de la ESCOM con el fin de que las autoridades correspondientes puedan tomar las decisiones pertinentes y propongan acciones para ayudar a estudiantes

## **3. Justificación.**

En la ESCOM, se observa que en la carrera de Ingeniería en Sistemas Computacionales uno de cada dos alumnos tiene reprobada por lo menos una materia y aproximadamente uno de cada catorce se dan de baja definitiva en algún punto de la carrera[3], por otro lado, se encuentra también que aproximadamente uno de cada tres alumnos que ingresan a ESCOM se gradúa satisfactoriamente [4], lo que nos permite visualizar que estas cifras son solamente un indicador de que hay un problema de fracaso escolar en la unidad académica.

Lo anterior es un problema debido a que analizando el comportamiento de estas cifras según el Anuario Estadístico del IPN se puede observar que es una tendencia que se ha mantenido por lo menos los últimos 6 años, lo que significa que la institución requiere de nuevas estrategias para intentar reducir estas cifras y con ello impactar positivamente en la vida estudiantil de la unidad académica mejorando el desempeño general de la escuela frente a los retos que implica la educación.

Por ello, para prevenir y corregir las situaciones que conllevan al fracaso y por ende a la deserción escolar se propone hacer uso de las tecnologías emergentes con respecto al aprendizaje automático con lo cual se puede obtener una mejor comprensión de las necesidades de los estudiantes identificar los posibles factores que conllevan al fracaso y deserción escolar.

Por consiguiente, se propone un sistema de aprendizaje automático que con base al historial de las trayectorias de los estudiantes de la ESCOM de la carrera de ISC de los pasados seis años de las cuales se toman las estadísticas antes mencionadas , se identifiquen relaciones significativas que quizás no hayan sido contempladas con anterioridad para así recopilar, analizar y correlacionar las situaciones que conllevan al fracaso y la deserción escolar y con ello proponer acciones preventivas y correctivas concretas a los estudiantes y al personal académico de la ESCOM para impulsar políticas educativas internas que logren combatir eficazmente el fracaso y la deserción escolar.

Existen otros trabajos similares en el área, pero se limitan a determinar el desempeño de una institución según su eficiencia terminal, sin generar ningún otro conocimiento de valor para las autoridades educativas de las universidades que permitan la mejora continua. Los principales beneficiarios del sistema propuesto son los estudiantes de la ESCOM, pues se pretende que la información resultante sea de vital importancia para las autoridades del plantel, quienes son nuestros usuarios finales, para impulsar técnicas y políticas a nivel interno que permitan mejorar la situación de los estudiantes en irregularidad.

Finalmente es necesario recalcar la complejidad del Trabajo Terminal, pues esta radica en el tratamiento de la información con la que se entrenará al sistema, manejar estas cantidades de información, determinar qué variables son las que nos interesa evaluar, determinar qué algoritmos de aprendizaje automático son los mejores para nuestro propósito, el diseño del sistema para que sea accesible, asegurar que nuestro modelo de aprendizaje sea generalizable y cómo serán protegidos los datos de los estudiantes.

## **4. Productos a entregar.**

Como producto final de nuestro Trabajo Terminal se planea entregar un sistema que permita que los estudiantes puedan consultar posibles riesgos de caer en fracaso escolar y recibir sugerencias de cómo mantener o mejorar en su caso su situación académica a nivel individual, mientras que a los directivos del plantel les permita consultar estos datos para individuos o bien para grupos completos de alumnos, con el fin de recibir sugerencias a este nivel de consulta para la toma de decisiones.

Por supuesto se plantea que el sistema cuente con la correspondiente visualización de los datos que se están consultado, de forma clara, útil y agradable, en distintas presentaciones de tipo estadístico.

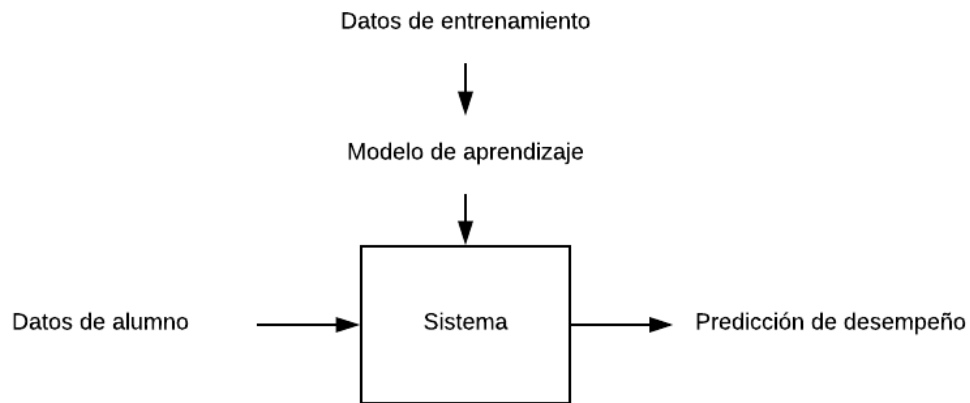


Figura 1. Arquitectura del sistema.  
Fuente: Creación propia

Así mismo se hará entrega de:

1. Reporte técnico.
2. Manual de usuario.
3. Artículo técnico.
4. Código fuente.

## 5. Metodología.

Decidimos trabajar sobre un modelo de procesos espiral, el cual consta de fases evolutivas, dado que los flujos de los procesos evolutivos realizan actividades de forma circular, cada círculo lleva a una versión más completa del software. El modelo busca minimizar los riesgos en el desarrollo del software. Este comienza analizando las posibles alternativas de desarrollo y se opta por la de menor riesgo asumible y se hace un ciclo espiral. Los espirales se repiten hasta que se pueda entregar un producto terminado lo que permite que el producto se trabaje continuamente en forma de prototipos. [11]

Entendemos en este contexto que el proceso principal a desarrollar para el proyecto es el del núcleo de aprendizaje automático, en el cual su desarrollo, consiste en un proceso de análisis de datos que consta de comprensión y procesamiento de los datos, selección y evaluación del modelo de aprendizaje y finalmente su visualización e implementación. [12]

## 6. Cronograma.

Nombre del alumno: Domínguez Morán Joaquín.

TT No.: 2020 - A125

Título del TT: Prototipo de sistema de aprendizaje automático de apoyo en la identificación de riesgos que llevan al fracaso escolar a los alumnos del IPN.

[illegible]

TT No.: 2020 - A125

Título del TT: Prototipo de sistema de aprendizaje automático de apoyo en la identificación de riesgos que llevan al fracaso escolar a los alumnos del IPN.

[illegible]

TT No.: 2020 - A125

Título del TT: Prototipo de sistema de aprendizaje automático de apoyo en la identificación de riesgos que llevan al fracaso escolar a los alumnos del IPN.

[illegible]

## 7. Referencias.

- [1] Tünnermann Bernheim, Carlos (2010) Las conferencias regionales y mundiales de la UNESCO y su impacto en la Educación Superior en América Latina. Disponible en : <https://www.redalyc.org/pdf/373/37318570005.pdf> Consultado en 9 de Septiembre de 2019.
- [2] Cabrera, Lidia, Bethencourt, José Tomás, Alvarez Pérez, Pedro y González Afonso, Míriam (2006). El problema del abandono de los estudios universitarios. [http://www.uv.es/RELIEVE/v12n2/RELIEVEv12n2\\_1.htm](http://www.uv.es/RELIEVE/v12n2/RELIEVEv12n2_1.htm). Consultado en 9 de Septiembre de 2019.
- [3] OECD (2019), Higher Education in Mexico: Labour Market Relevance and Outcomes, Higher Education, OECD Publishing. Disponible en: [https://www.oecd.org/centrodemexico/medios/educacion\\_superior\\_en\\_mexico.pdf](https://www.oecd.org/centrodemexico/medios/educacion_superior_en_mexico.pdf) Consultado en 11 de septiembre de 2019.
- [4] IPN (2018) Anuario General Estadístico. Disponible en: <https://ipn.mx/assets/files/coplaneval/docs/Evaluacion/AnuarioGeneralEstadistico2018ok.pdf> Consultado en 10 de septiembre de 2019.
- [5] Paulo Cortez y Alice Silva, Using Data Mining to Predict Secondary School Student Performance. Disponible en : <http://www3.dsi.uminho.pt/pcortez/student.pdf> . Consultado en 5 de septiembre de 2019.
- [6] Delimiro Visbal Cadavid, Adel Mendoza Mendoza y Sonia Jacqueline Orjuela Pedraza, Predicción de la eficiencia de las instituciones de educación superior colombianas con análisis envolvente de datos y minería de datos. Disponible en : <http://rcientificas.uninorte.edu.co/index.php/pensamiento/article/viewFile/9511/11030> Consultado en 12 de septiembre de 2019.
- [7] Brohi S.N., Pillai T.R., Kaur S., Kaur H., Sukumaran S., Asirvatham D. (2019) Accuracy Comparison of Machine Learning Algorithms for Predictive Analytics in Higher Education. In: Miraz M., Excell P., Ware A., Soomro S., Ali M. (eds) Emerging Technologies in Computing. iCETiC 2019. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 285. Springer, Cham Consultado en 12 de septiembre de 2019.
- [8] Kotsiantis S.B., Pierrakeas C.J., Pintelas P.E. (2003) Preventing Student Dropout in Distance Learning Using Machine Learning Techniques. In: Palade V., Howlett R.J., Jain L. (eds) Knowledge-Based Intelligent Information and Engineering Systems. KES 2003. Lecture Notes in Computer Science, vol 2774. Springer, Berlin, Heidelberg Consultado en 12 de septiembre de 2019.
- [9] IPN (2018). Agenda estadística 2018. Disponible en : <https://www.ipn.mx/assets/files/gestionestrategica/docs/Evaluacion/1AgendaEstadistica18.pdf> Consultado en 13 de septiembre de 2019.
- [10] Rodríguez Lagunas Javier., Leyva Piña Marco Antonio. (2007) La deserción escolar universitaria. La experiencia de la UAM. Entre el déficit de la oferta educativa superior y las dificultades de la retención escolar. Available: <https://www.redalyc.org/pdf/325/32514212.pdf> . Consultado en 13 de septiembre de 2019.
- [11] Boehm, B. «A Spiral Model of Software Development and Enhancement.» ACM SIGSOFT Software Engineering Notes, ACM, 11(4):14-24, agosto de 1986.
- [12] Shearer, C. (2000) The CRISP-DM Model: The New Blueprint for Data Mining. Journal of Data Warehousing, 5, 13-22.



## 8. Alumnos y Directores.

CARÁCTER: Confidencial

FUNDAMENTO LEGAL: Artículo 11  
Fracc. V y Artículos 108, 113 y 117 de la  
Ley Federal de Transparencia y Acceso a  
la Información Pública.

Mendez Alvarez Jorge Luis.

- Alumno de la carrera de Sistemas Computacionales en ESCOM,  
Especialidad Sistemas Boleta: 2012050943, Tel.:57739907,  
email: [jorgeluispolaris@gmail.com](mailto:jorgeluispolaris@gmail.com)

Firma:



Ochoa Giner Carlos Armando.

- Alumno de la carrera de Sistemas Computacionales en ESCOM,  
Especialidad Sistemas Boleta: 2014081103, Tel.:5613036384,  
email: [caog.freeman@gmail.com](mailto:caog.freeman@gmail.com)

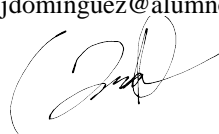
Firma:



Domínguez Morán Joaquín.

- Alumno de la carrera de Sistemas Computacionales en ESCOM,  
Especialidad Sistemas Boleta: 2014130330, Tel.:5560892652,  
email: [jdominguez@alumno.ipn.mx](mailto:jdominguez@alumno.ipn.mx)

Firma:



*Flavio Arturo Sánchez Garfias.*- Doctorado en ciencias de la  
computación por el CIC del IPN. Áreas de interés: machine  
learning, image analysis, data science. Tel. 57296000 ext 52032  
E-mail: [fsanchezga@ipn.mx](mailto:fsanchezga@ipn.mx)

Firma:



*Elba Mendoza Macias.*-Maestría en Dirección Empresarial.  
Áreas de interés: administración, finanzas y sistemas de información  
Tel. 57296000 ext 52049  
E-mail: [emendozam64@gmail.com](mailto:emendozam64@gmail.com) y [emendozam@ipn.mx](mailto:emendozam@ipn.mx)

Firma:

