

Uso de Deep Learning en sistemas de recomendación para reducir la deserción en Educación
Superior Colombiana

Jorge Alvarez

Maestría en Inteligencia de Negocios

Universidad EAN

Bogotá, noviembre 2020

Resumen

El anteproyecto de monografía planteado aquí, busca encontrar cómo la implementación de una herramienta de Deep Learning puede ser empleada para mitigar de forma efectiva, la deserción en Instituciones de educación superior, al reducir el tiempo de detección y recomendación de apoyos necesarios para que los alumnos puedan continuar su proceso educativo.

El gobierno colombiano es consciente de la problemática que envuelve la deserción en educación superior, y por esto, ha generado iniciativas que buscan bajar las cifras del problema, pero en la actualidad los esfuerzos no han tenido el resultado esperado.

La investigación se centrará en determinar la viabilidad de implementación de la herramienta en las plataformas en la nube con reconocimiento como las de mayor capacidad y completitud de su servicio, y en determinar si las Instituciones de Educación Superior tienen la capacidad de generar la información que permita alimentarla y la posibilidad de mantener el costo de una herramienta implementada en una plataforma en la nube.

Palabras clave: Anteproyecto, Deep Learning, Deserción, Instituciones de Educación Superior

Tabla de contenido

Resumen	2
1. Introducción	7
2. Antecedentes de la investigación	7
3. Planteamiento del problema de investigación	8
3.1 Descripción	8
3.2 Árbol de problemas	9
3.2.1 Efectos	9
3.2.2 Causas	10
4. Formulación de la pregunta de investigación	13
5. Objetivos de investigación	14
5.1 Objetivo general	14
5.2 Objetivos específicos	14
6. Justificación de la investigación	14
7. Viabilidad de la investigación	16
8. Marco teórico de la propuesta	17
8.1 Inteligencia artificial	17
8.1.1 Machine Learning	18
8.2 Deep Learning	20
8.2.1 Estructura	20
8.2.2 Usos	21
8.2.3 Metodología general para construir modelos de Deep Learning	22
8.2.4 Deep Learning en el contexto educativo	22
8.3 Deserción en la Educación Superior	22
8.3.1 Concepto de deserción	23
8.3.2 Causas de deserción en la educación superior	23
8.3.3 Modelos para el análisis de la deserción	26
8.3.4 El sector de la educación superior en Colombia	27

9	Hipótesis de trabajo	29
9.1	Hipótesis 1	29
9.1.1	Variables hipótesis 1	29
9.2	Hipótesis 2	29
9.2.1	Variables hipótesis 2	29
10	Diseño metodológico	30
10.1	Tipo de investigación	30
10.2	Instrumentos a utilizar	30
10.3	Población y muestra	31
11	Contribuciones originales esperadas	32
12	Cronograma	32
13	Referencias	34

Tabla de Figuras

Figura 1	Árbol del problema	13
Figura 2	Tasa de deserción por nivel de formación	15
Figura 3	Divisiones de ML	19
Figura 4	Factores de deserción según el MEN	26
Figura 5	Cuadrante mágico de Gartner, IA en la nube	31

Tabla de tablas

Tabla 1 Análisis de viabilidad	16
Tabla 2 Tamaño de muestras representativas	32
Tabla 3 Cronograma de la investigación	33

1. Introducción

La presente investigación tiene como objetivo proponer una forma de mitigar un problema de alto impacto para Colombia, como lo es la deserción en las Instituciones de Educación Superior IES, por medio de la uso de una herramienta que implemente Deep Learning, con el objetivo de identificar posibles estudiantes con tendencia a desertar y recomendar planes de acción de forma oportuna.

2. Antecedentes de la investigación

Según el informe de desarrollo del Banco Mundial (BM, 2018, p. 3), la educación permite a los países mejorar en innovación, fortalecimiento de instituciones y cohesión social, mientras que para los individuos promueve el empleo, aumenta la remuneración, mejora el acceso a salud y reduce de la pobreza, todas estas ventajas se ven afectadas por la deserción, ya que al no culminar sus estudios, los alumnos y sus familias ven que la inversión hecha no es compensada por un mayor ingreso, por lo cual el gobierno colombiano ha manifestado interés en la reducción de este fenómeno mediante programas como el publicado por el Ministerio de Educación Nacional (MEN, 2015) tendientes a recolectar información y aplicar medidas que mitiguen las variables que generan la deserción en estudiantes de educación terciaria.

Los porcentajes de deserción de estudiantes en Instituciones de Educación Superior, IES, eran 26.1% para técnica profesional, 16.7% para tecnológica y 9% en universitaria, según el Sistema para la Prevención y Análisis de la Deserción en las Instituciones de Educación Superior (SPADIES, 2017) antes de la pandemia, pero aún no se tiene claro cómo estas cifras se pueden empeorar con las dificultades sociales y económicas generadas en 2020.

Los estudios que se han hecho sobre deserción, ha permitido entender las variables que la generan como en los estudios de Fontalvo (2015), Chinome et al. (2016), Barragán y Urrego (2019), Barrera et al. (2015), Yepes D et al. (2007), Sánchez et al. (2009) y Londoño (2013), pero otros estudios también permitieron entender que hay condiciones clave, aplicables por las IES, que pueden incrementar la retención, como lo explica el sociólogo y experto en el tema Tinto (MEN, 2010, p. 18).

Las IES siempre tienen la posibilidad de hacer cotejos y seguimientos manuales, pero por lo general, las instituciones no cuentan con una planta de personal tan extensa que les permita hacer este proceso de forma efectiva, es por esto que se plantea la necesidad de usar un motor de procesamiento analítico de datos que genere resultados dentro de un marco de oportunidad pertinente, para que las ayudas alcancen a tener un efecto en un marco de tiempo corto, inferior a un semestre.

Tomando en cuenta la restricción de tiempo de obtención de resultados, se plantea la necesidad de utilizar una de las mejores tecnologías para análisis y correlación de datos que existe actualmente, es por esto que en este estudio se explora el uso de motores de *Deep Learning*, DL en adelante, como una tecnología de propósito general para encontrar soluciones óptimas, que hoy en día está campos de investigación de alta diversidad (Klinger, Mateos-García y Stathoulopoulos, 2018), y que en este caso puede tener un impacto muy positivo.

3. Planteamiento del problema de investigación

3.1 Descripción

La deserción implica el abandono voluntario o no, por parte de los alumnos, de los programas educativos antes de su culminación exitosa, tomando en cuenta que es un problema de múltiples raíces y con consecuencias de alto impacto para los estudiantes, sus familias y en general la sociedad, el gobierno colombiano se ha propuesto metas para disminuirla (MEN, 2015).

Aunque el gobierno ha establecido el programa SPADIES, para las IES sigue habiendo una carencia de análisis de información, para lograr detectar a tiempo a los alumnos que cumplen patrones de riesgo, y ofrecerles un apoyo pertinente, que resulte en una disuasión efectiva para evitar el abandono del programa de educación superior, por lo cual, se requiere encontrar una alternativa que correlacione grandes cantidades de información sin intervención humana y a un costo viable para las IES.

También es importante determinar si las IES tienen claro el problema de deserción, y si existe la capacidad de invertir en una herramienta para solucionarlo, ya que sin esta capacidad, los costos asociados impedirían que una solución automática sea realizable.

Al revisar la tecnología existente, se encuentra que es posible aplicar un reconocimiento de patrones a gran escala, con *Machine Learning* pero esta solución sólo permitiría detectar casos que cumplan con los antecedentes previamente establecidos, mientras que una solución de *DL* puede encontrar patrones no detectados por los expertos, y llegar a soluciones que tarden menor tiempo para ser efectivas.

3.2 Árbol de problemas

3.2.1 Efectos

Efecto 1: Baja competitividad

Descripción: Para Colombia hay un impacto negativo debido a que la educación es un factor determinante de la competitividad de los países (Heredia Gonzalez et al., 2017), ya que implica frenar la oportunidad de tener mayor inversión extranjera, y disminución de la capacidad de producir bienes de alto valor agregado.

Efecto 2: Baja creación de empresas

Descripción: Al existir baja capacidad de innovación hay menor confianza para crear empresas competitivas que aporten al mercado laboral.

Efecto 3: Pérdidas económicas

Descripción: Las pérdidas económicas para las universidades se dan por los costos fijos que tienen que asumir para mantener la planta física y de personal, adicionados a los costos de promoción requeridos para obtener nuevos estudiantes. También para las familias de los estudiantes y los mismos estudiantes, se dan pérdidas económicas debido a que no se recupera la inversión en matrículas, materiales y sostenimiento durante el periodo de estudio.

Efecto 4: Problemas financieros para las universidades

Descripción: Los costos de las universidades para atraer estudiantes, más los costos de mantener una planta física y el personal sólo se ven compensados con las matrículas que pagan

los estudiantes, al abandonar sus estudios, los estudiantes generan pérdidas económicas debido a que no se pueden cumplir los presupuestos establecidos por las instituciones.

Efecto 5: Deudas para los estudiantes

Descripción: Los estudiantes y sus familias tienen que recurrir por lo general a créditos, ya sea con el Instituto Colombiano de Crédito Educativo y Estudios Técnicos, ICETEX o con entidades bancarias para poder pagar los estudios, pero al no poder terminarlos, no se produce la compensación esperada por el esfuerzo.

Efecto 5: Baja movilidad social

Descripción: Al no terminar los estudios y no tener una compensación por su esfuerzo, las personas tienen menor acceso a empleos mejor remunerados o a la producción de sus propias empresas, limitando su capacidad de mejorar su estatus social.

Efecto 6: Aumento de la desigualdad

Descripción: El esfuerzo de una familia de bajos recursos para lograr la graduación de uno de sus miembros, es cada vez mayor, comparado con el de una familia de los estratos más altos, debido a los aumentos continuos de los costos de las matrículas, sumado a la necesidad de tener a la mayoría de sus miembros produciendo ingresos durante el mayor tiempo posible, hacen cada vez más probable que las personas de bajos recursos tengan que suspender sus estudios.

3.2.2 Causas

Causa 1: Problemas psicosociales

Descripción: Los alumnos se ven expuestos a diferentes situaciones desde familiares, problemas con compañeros y profesores, hasta depresión y otros problemas psicosociales que los fuerzan a abandonar sus estudios.

Causa 2: Aplicación ineficiente de apoyos

Descripción: En el SPADIES constan los apoyos brindados por las diferentes entidades a los estudiantes, sin embargo, revisando las cifras, no han tenido el impacto deseado sobre la decisión de culminar los estudios.

Causa 3: Aplicación de apoyos tardíos

Descripción: Cuando no se hace una detección a tiempo de los problemas y la intervención se hace después de ciertos momentos, la decisión de abandonar los estudios no puede ser modificada.

Causa 4: Dificultad para detectar alumnos con probabilidad de deserción

Descripción: La detección de alumnos con probabilidad de deserción, requiere una cantidad de procesamiento de información que en la mayoría de los casos, no poseen las entidades de estudios superiores, que les permitan cotejar datos y detectar a tiempo cuando un estudiante cumple uno de los patrones de abandono de sus estudios.

Causa 5: No tener claro el apoyo necesario

Descripción: Aunque un alumno sea detectado como posible desertor, es posible que el apoyo requerido para revertir su estado, no sea del todo claro para la institución.

Causa 6: Problemas económicos

Descripción: Los problemas para continuar pagando los estudios son una de las causas detectadas en todos los estudios analizados, ya que al no ser gratuito el tener a un miembro de la familia estudiando durante un largo tiempo, así la matrícula sea de bajo costo como en el caso de las entidades públicas, mantener una persona improductiva puede ser un costo no asumible para muchas familias de bajos ingresos.

Causa 7: Necesidad de trabajar

Descripción: En varios casos los estudiantes tienen que dedicarse a trabajar, ya sea para mantener a sus padres y hermanos, ellos mismos o personas dependientes, forzándolos a abandonar sus estudios por incompatibilidad de horarios o simplemente por no tener el tiempo necesario para cumplir con sus obligaciones académicas.

Causa 8: Personas a cargo

Descripción: El tener personas a cargo obliga a los alumnos a dedicar tiempo y recursos económicos a sus dependientes, imposibilitando continuar con los estudios.

Causa 9: Problemas económicos familiares

Descripción: Los problemas económicos familiares pueden dejar sin la oportunidad de continuar pagando los costos de matrículas y el mantenimiento requerido por los estudiantes para asistir a las clases o adquirir los útiles y materiales necesarios para continuar estudiando.

Causa 10: Falta de acceso a becas

Descripción: El acceso a becas es limitado por las capacidades académicas de los estudiantes, pero también por el acceso a la información y la poca cultura de uso de esta herramienta para culminar los estudios.

Causa 11: Problemas académicos

Descripción: Los problemas académicos conducen a la deserción porque imposibilitan culminar los programas académicos, ya sea por repetición de materias, o por incumplir normas de las instituciones como el mantener un promedio académico por encima de un nivel.

Causa 12: Alta repetición de materias

Descripción: El repetir múltiples veces una o más materias imposibilita a los estudiantes continuar avanzando en los programas académicos que cursan, llegando a puntos donde se hace inviable continuar estudiando.

Causa 13: Bajo rendimiento académico

Descripción: Ya sea por la falta de preparación, acceso a información, capacidad o desmotivación, es posible que los estudiantes abandonen sus estudios forzados por su bajo desempeño.

Causa 14: Baja preparación en el colegio

Descripción: El bajo nivel de preparación académica en los colegios produce estudiantes sin los conocimientos requeridos para tener buenos resultados en las instituciones de educación superior, lo cual se vuelve un problema de falta de conocimiento acumulado desde tempranas edades en muchos casos, lo cual genera una incapacidad de cumplir con los requisitos para poder culminar los estudios.

Causa 15: Orientación profesional equivocada

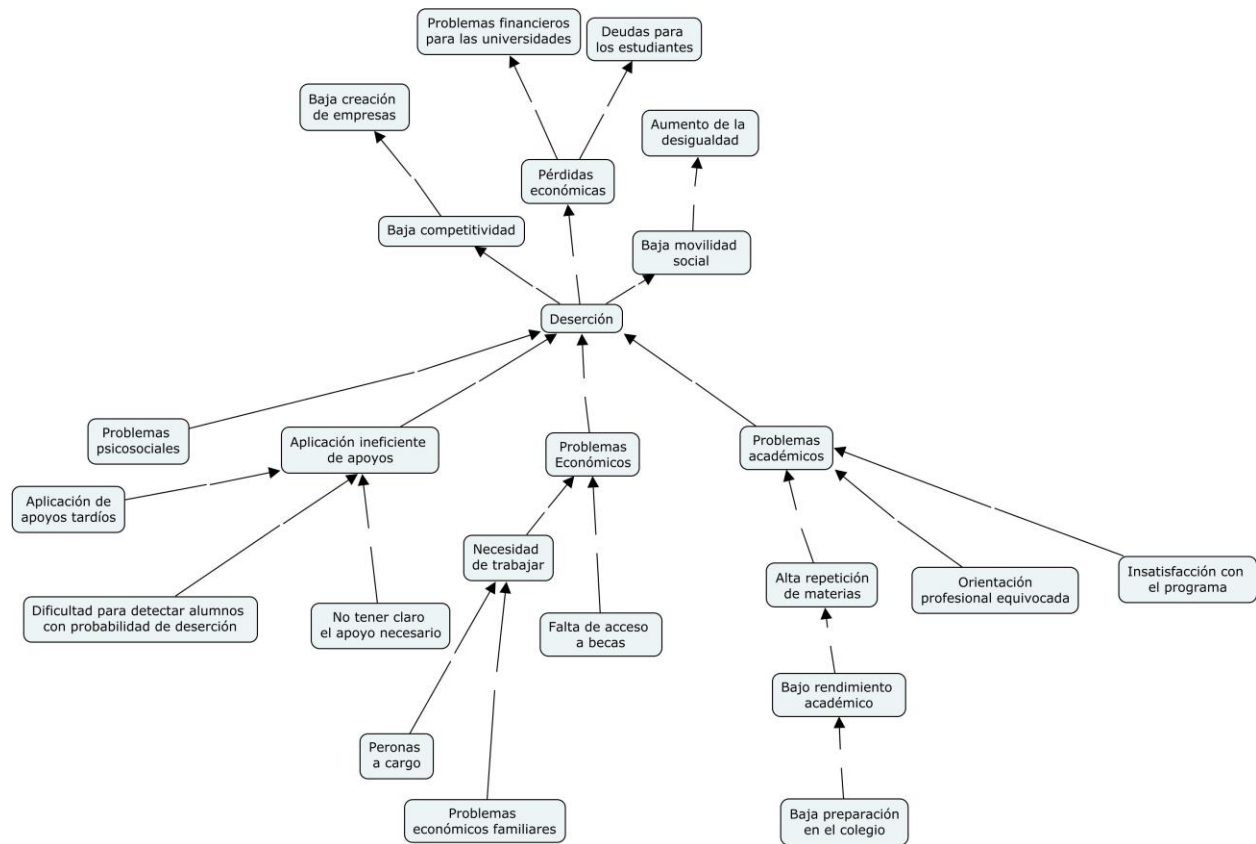
Descripción: El estudiar carreras para las cuales el estudiante tiene baja motivación crea insatisfacción y termina produciendo abandono en algunos casos, o la transferencia hacia otro programa.

Causa 16: Insatisfacción con el programa

Descripción: El nivel académico o el contenido que ofrecen las instituciones pueden no estar de acuerdo con las expectativas de los alumnos, lo cual puede llevar a una desmotivación que produzca la deserción.

El árbol de problema en la Figura 1, a continuación, se basa en las investigaciones que se han hecho por diferentes autores y las ventajas de la educación que son eliminadas por la deserción.

Figura 1 Árbol del problema



Fuente propia.

4 Formulación de la pregunta de investigación

¿Cómo se puede aplicar *DL* en un sistema de recomendación que permita reducir la deserción en la educación superior colombiana?.

5 Objetivos de investigación

5.1 Objetivo general

Proponer una alternativa para aplicar DL en un sistema de recomendación que permita disminuir la deserción en la educación superior en Colombia

5.2 Objetivos específicos

1. Identificar la percepción que tienen las Instituciones de Educación Superior respecto a las posibles soluciones para disminuir la deserción estudiantil
2. Determinar las posibles variables a evaluar dentro de la data proveniente del SPADIES, que permitan sugerir posibles acciones que disminuyan la deserción
3. Plantear recomendaciones que permitan mejorar la probabilidad de permanencia en los programas de educación superior desde el uso del *DL*
4. Determinar la plataforma óptima en la cual se pueda hacer una implementación viable
5. Definir la ruta de aplicación del *DL* que permita disminuir la deserción en la educación superior en Colombia

6 Justificación de la investigación

Esta investigación tiene como propósito, determinar la forma de ayudar a la IES a procesar los datos de los estudiantes, de forma que pueda no sólo determinar qué alumnos se encuentran en riesgo de deserción sino las alternativas de acción, dentro de un plazo oportuno para su aplicación.

Si bien es posible que hoy en día se esté recolectando la información necesaria, la cantidad de personal para procesarla sin un motor de análisis, no permite obtener resultados y aplicar cambios dentro del mismo semestre, de forma que sea oportuna la intervención para obtener los resultados que espera la IES.

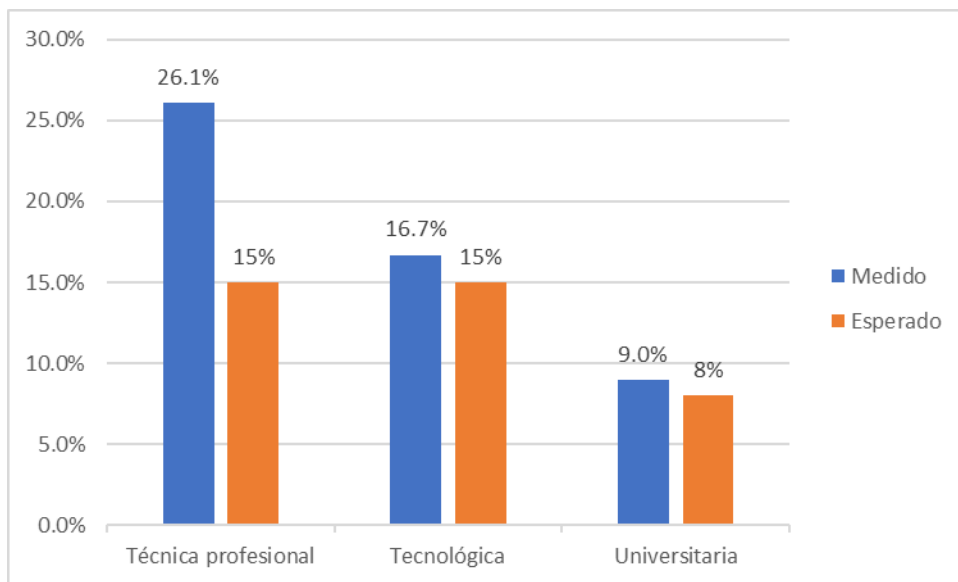
Las causas de la deserción son variadas y cubrir todas en una sola solución no es viable, por lo cual es importante encontrar la solución o una mitigación a un grupo de estas, como lo son

la detección temprana de estudiantes con patrones de deserción y la recomendación de una acción preventiva a aplicar.

Los efectos en la sociedad colombiana son bastante importantes, sobre todo para un país con una de las mayores desigualdades sociales del mundo, por lo cual esta investigación es relevante para lograr reducir el impacto de este problema.

Es importante hacer énfasis, en que el gobierno nacional ha hecho esfuerzos para implantar un sistema que permita disminuir la deserción, pero la detección tardía de los estudiantes en riesgo y la aplicación de apoyos fuera de tiempo sigue sin hacer que el sistema sea lo suficientemente eficiente para las metas propuestas originalmente de: 8% a nivel universitario y 15% en el nivel técnico profesional y tecnológico (MEN, 2015), pero como lo muestra la Figura 2, a continuación, las cifras en 2016 no habían llegado a los niveles esperados.

Figura 2 Tasa de deserción por nivel de formación



Fuente MEN-SPADIES (2019, a corte 2016).

Esta investigación es conveniente porque permitirá vincular una de las tecnologías de mayor evolución en los últimos años, como lo manifiestan Klinger, Mateos-Garcia y Stathoulopoulos (2018), como una tecnología de propósito general, comparable en su impacto al uso del motor a combustión, los computadores o la electricidad, con un problema de alto impacto para el país, para lograr la mitigación de por lo menos uno de los factores que lo originan al

proporcionar un poder de cómputo superior encontrando patrones, a otras herramientas aplicables.

Este trabajo tiene relevancia social, porque el impacto que tiene en la sociedad colombiana la deserción es un obstáculo para los planes de eliminación de la pobreza que ha planteado el gobierno, debido a que como lo explica el MEN(2015, p. 17), “..., la equidad pasa por la posibilidad real de brindar una educación con los más altos estándares de calidad para toda la población...”, y sin un sistema que permita a las instituciones ayudar de forma oportuna a los alumnos, las ventajas que se espera brinde la educación superior a la sociedad llegan sólo de forma parcial, frenando el desarrollo del país.

7 Viabilidad de la investigación

Tabla 1 Análisis de viabilidad

Criterio	Viabilidad o Facilidad (siendo 1 menor y 5 mayor)
Acceso a la información	3
Procesamiento estadístico o disponibilidad de software	5
Desarrollo de entregables en el tiempo	4
Calidad de la información	3
Costos de la investigación	4
Promedio	3.8

Fuente propia

La información que se posee es la que tiene recolectada el sistema SPADIES y la que se puede encontrar en Datos abiertos Colombia DAC, sobre las IES, pero de acuerdo a lo que evalúe el procesamiento de DL se puede determinar si la información es suficiente o no.

El software para aprendizaje y procesamiento de DL se encuentra para evaluación en diferentes plataformas en la nube, como: IBM, Google Cloud, AWS o Azure, adicionalmente, se pueden usar herramientas gratuitas para procesamiento o depuración de data si es necesario.

El periodo planteado para la investigación es de nueve meses, el cual se espera que sea suficiente para recolectar la información y hacer la experimentación necesaria, sin embargo el riesgo se encuentra en la accesibilidad para hacer entrevistas en las IES, la calidad y actualidad de la información en SPADIES, y el conocimiento en DL y las plataformas, que debe ser adquirido para realizar la experimentación.

La calidad y actualidad de la información en SPADIES es un tema que se sale del control de esta investigación, por lo cual es un riesgo, es necesario revisar la data que se puede descargar del portal y la que se encuentra en DAC para determinar si es suficiente, o si es necesario gestionar con varias IES el acceso a sus registros anonimizados, lo cual puede ser más complejo de conseguir.

Se espera cubrir la investigación con las capas gratuitas de entrenamiento que ofrecen las plataformas de DL en la nube, sin embargo, es posible que se deba incurrir en costos para almacenamiento y procesamiento, dichos costos servirán para establecer el valor aproximado que debería poder pagar una IES, si quiere hacer uso de esta tecnología. Adicionalmente, se deben hacer encuestas a instituciones de diferentes regiones del país, por lo cual debe adicionarse el costo de las llamadas telefónicas necesarias.

El resultado total de 3.8, indica que aunque existe una viabilidad para su ejecución, la investigación tiene unos riesgos específicos en el acceso a la información, que pueden afectar su cronograma esperado.

8 Marco teórico de la propuesta

8.1 Inteligencia artificial

La Inteligencia Artificial o IA, tiene orígenes antiguos y se puede referenciar en periodos tan lejanos como la Grecia antigua (Mijwel, 2015) o las máquinas de Babbage del siglo 18 (Aspray, Bromley, Campbell-Kelly, Ceruzzi y Williams, 1990), aunque en términos generales, la IA moderna se reconoce como inicialmente propuesta por Turing (1936), al determinar que una máquina tiene la capacidad de hacer cálculos de forma repetitiva, comparando los resultados, hasta encontrar una solución acorde a una serie de restricciones planteadas inicialmente, en algo que él denominó una máquina universal.

La IA se puede definir de forma sencilla como lo propuso McCarthy en 1956: “La ciencia e ingeniería de hacer máquinas inteligentes” (Herrera y Analide, 2012), o como lo complementó Andrew Moore (High, 2017) : “Inteligencia Artificial es la ciencia y la ingeniería de hacer que los computadores se comporten de formas que, hasta recientemente, pensábamos que requerían de inteligencia humana para hacerlo”, pero este principio básico ha avanzado hasta convertirse en uno de los motores de la tecnología humana más ampliamente utilizado, y en uno de los campos de la ciencia de computación más atrayente para los investigadores y la industria.

En los años 80 del siglo pasado, ya investigadores ubicaban a la IA como una ciencia a la altura de la matemática, la física o la sociología, que puede servir fundamento para nuevos conocimientos humanos (Campbell, 1986). Y aunque siempre han existido temores sobre su uso, pronto fue notoria la necesidad de crear lenguajes de programación y tecnología propicia para su desarrollo (Partridge, 1986), como por ejemplo la programación orientada a restricciones (Meseguer, 2012) o los algoritmos evolucionarios de múltiple objetivo, EMOA por su sigla en Inglés, utilizados para optimización de procesos (Zhao et al., 2019).

Actualmente la IA se estudia en campos que a su vez son suficientemente profundos para generar grandes cantidades de conocimiento, como lo son *Machine Learning* y *DL*, los cuales serán analizados a continuación.

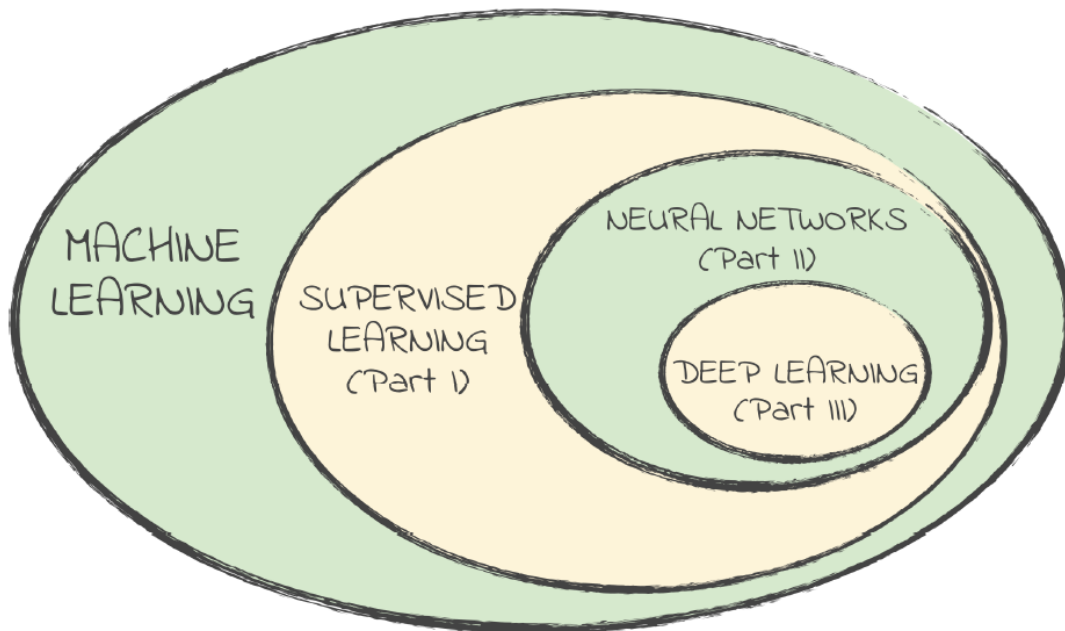
8.1.1 Machine Learning

Machine Learning ML se puede definir como los procesos mediante los cuales un sistema inteligente aprende de forma autónoma, mejorando su desempeño a través del tiempo (Langley, 1986), o como lo define Mitchell (1997), es el estudio de los algoritmos de computación que permiten a los programas de computación mejorarse automáticamente a través de la experiencia.

Al ser una rama de la IA, debe diferenciarse en que su propósito es el aprendizaje y mejora autónomas mediante experiencia, y no solamente generar comportamientos inteligentes mediante decisiones de reconocimiento de patrones, como lo define Mitchell(1997): un programa de computación se dice que aprende de la experiencia E, con respecto a una clase de tareas T y una medida de desempeño P, si su desempeño P para realizar las tareas T, mejora al tener la experiencia E.

Dentro de ML se encuentran otras subramas de la IA, como lo explica la figura a continuación, pero en general se divide en tres ramas: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje de refuerzo (Skansi, 2018, p. 51)

Figura 3 Divisiones de ML



Fuente: (Perrotta, 2020, p. viii)

A partir del estudio de ML se han generado algoritmos relevantes para la clasificación de fenómenos (Smola y Vishwanathan, 2008, p.20), como Naive Bayes, Nearest Neighbors, Mean Classifier, Perceptron o K-Means, los cuáles son usados para resolver múltiples problemas de clasificación y predicción, desde el envío de correos spam, hasta la sugerencia de artículos en tiendas en línea.

8.1.1.1 Deep Learning

DL es una subdivisión de ML, basada en Redes Neuronales (Skansi, 2018, p. 8), las cuales empezaron a ser proyectadas desde los años 50 del siglo pasado, por Marvin Minsky, como una aproximación para solucionar problemas donde la simulación de neuronas interactuando pudiera ser más adecuada que el uso de procesadores normales, como la

interpretación de textos, pero que debido a la incapacidad para solucionar momentáneamente problemas de ambigüedad, casi es abandonada en los años 70, hasta que fue viable con el aumento de la capacidad de las máquinas en los años 90, y finalmente con el artículo de Hinton, Osindero y Teh (2006), donde proponen un modelo de red de creencia profunda, con mejores resultados que los obtenidos previamente para el reconocimiento de patrones, el paso para la aparición de DL fue posible.

8.2 Deep Learning

Para entender mejor qué es DL, se puede explicar que es la implementación de procesos cognitivos mediante redes neuronales, donde por medio de unas reglas establecidas se logran razonamientos, y aunque las reglas se establecen a nivel neuronal, el razonamiento sigue siendo simbólico (Skansi, 2018, p. 14), la principal diferencia con ML es que no requiere un paso intermedio que en ML se llama la extracción de cualidades para comparar, sino que este paso viene implícito en el proceso de aprendizaje, y la máquina detecta las características que debe buscar para hacer la identificación correcta.

8.2.1 Estructura

Debido a que DL se basa en el uso de redes neuronales para su lograr sus resultados, normalmente la estructura que tiene es la de la definición de las neuronas con las reglas de interpretación que debe seguir y luego la especificación de la red, donde se establecen las interacciones que deben tener las neuronas (Michelucci, 2018).

La arquitectura de una red neuronal para hacer DL se compone de varias capas de interpretación, de acuerdo al propósito buscado (Singaravel et al., 2018), normalmente se reconocen tres tipos de capas: capas de entrada, capas ocultas y capas de salida. Donde el número de capas ocultas define la profundidad de la arquitectura.

En las capas ocultas se definen las funciones de activación, las cuales definen las relaciones no lineales, y con los resultados de estas capas ocultas, las capas de salida generan las predicciones. Dentro de las funciones de activación comunes se encuentran binarias, lineales

sigmoides, de tangente hiperbólica, lineal rectificadora, lineal rectificadora agujereada, lineal rectificadora configurada, exponencial lineal unitaria, Swish y Softmax.

Otra parte de la estructura importante son los parámetros internos de las capas que determinan la complejidad de interpretación en las capas internas y los hiper parámetros, que controlan el número de interacciones y la tasa de aprendizaje (Michelucci, 2018, p. 90), dichos hiper parámetros se deben ajustar para lograr la optimización de las predicciones.

8.2.2 Usos

DL es una tecnología relativamente nueva para la humanidad, en el artículo de Klinger, Mateos-García y Stathoulopoulos (2018) se hace un análisis del impacto que va a tener como una tecnología de uso general, tal como fueron el motor de vapor y las computadoras, cambiando a múltiples industrias.

En la actualidad DL se utiliza en sectores muy diversos como la salud, en identificación de enfermedades mediante clasificación de imágenes (Ker et al., 2018), predicciones de resultados en el uso de medicinas (Aliper et al., 2016), en oftalmología se usa para detección temprana de tumores mediante clasificación de imágenes (Rahimy, 2018), en biomedicina se usa para segmentación de imágenes diagnósticas, secuenciación genética y predicción de estructura de proteínas (Cao et al., 2018), en arquitectura se ha aplicado para predecir consumo de energía en edificios (Singaravel et al., 2018), en astronomía se está usando para detección de vida extraterrestre y exoplanetas (Gheller et al., 2018), y muchos otros usos donde la gran capacidad de DL para detectar patrones y deducir relaciones es funcional.

La relación fundamental de esta investigación con el uso de *DL*, se da con la capacidad que tienen estos para generar sistemas de recomendación efectivos, y de esta forma se encuentran varios estudios como el de Da'u, A. y Salim, D. (Da'u y Salim, 2020), donde se analiza el avance de esta tecnología y se hace un análisis sistemático de la literatura existente, mientras que en el artículo de Batmaz, Z. et al. (Batmaz et al., 2019) se analizan los retos y opciones que se presentan al afrontar un sistema de recomendación con esta tecnología, por otra parte en el estudio de Singhal, A., Sinha, P. y Pant R. (Singhal et al., 2017) hace una compilación de trabajos recientes para la época, también el artículo de Naumov, M. et al. (Naumov et al., 2019) hace un análisis de la incidencia de esta tecnología en los sistemas de recomendación, y

finalmente, en el artículo de Cheng, H. et al. (Cheng et al., 2016) generado por Google Inc. se hace un análisis del uso de Wide y DL para la generación de sistemas de recomendación.

8.2.3 Metodología general para construir modelos de Deep Learning

La metodología general para construir modelos de DL se basa en los siguientes pasos (Oppermann, 2019): a) Definición del set de datos de entrenamiento, b) Definición de las capas que va a tener la red neuronal, c) Definición de la función de activación adecuada por capa, d) Entrenamiento del modelo, e) Comparación de resultados, f) Definición de la función de pérdida, g) Aplicación de cambios a parámetros y pesos para minimizar la función de pérdida.

8.2.4 Deep Learning en el contexto educativo

DL se ha utilizado para hacer estudios del entorno educativo, como en el artículo de Doleck et al. (2020) donde se analiza la aplicación de diferentes plataformas de DL usando datos de estudiantes que recibieron clases por video.

Doleck et al. refieren que esta clase de estudios se han hecho antes, como el realizado por Piech et al.(2015) usando Deep knowledge para medir la efectividad del uso de plataformas como Khan Academy, Coursera y EdX.

Por otra parte, Wilson et al.(2016) hicieron un estudio del posible desempeño de los alumnos ante el uso de sistemas de educación en línea, haciendo un análisis bayesiano, y de nuevo aplicando redes neuronales convolutivas con Deep Knowledge Tracing.

Otros estudios relacionados son los que hicieron los equipos de Botelho et al. (2017) y Jiang et al. (2018), para estudiar el rendimiento de los estudiantes de escolaridad media en plataformas de enseñanza abierta, usando DL.

8.3 Deserción en la Educación Superior

La deserción en educación superior en Colombia ha sido estudiada en varias ocasiones desde el punto de vista de cada institución, como en el caso de la Universidad de Ibagué con la investigación de Betacur y González (2008), la Universidad de Rosario con el artículo de

Restrepo (2016), la Universidad Distrital Francisco José de Caldas con el artículo de Flórez (2016), la Fundación universitaria católica del norte con el artículo de Rodríguez (2011), y la Universidad industrial de Santander con el artículo de Zárata y Mantilla(2015), pero en otros artículos se busca hacer un estudio comprensivo del fenómeno, como en el de Páramo y Correa (2012) que estudia variables y fenómenos de incidencia para estructurar indicadores de deserción, el de Barrero (2015) analiza estudios anteriores y generar propuestas para mitigar el fenómeno, mientras que Areth (Areth et al., 2015) genera una análisis de modelos de deserción para comprender el fenómeno aunque indica que en Colombia los estudios están sesgados por modelos teóricos de educación a distancia, el artículo de Castillo y Mejía (2010) habla sobre las causas sociales del fenómeno y el artículo de Barragán y Patiño (2016) que interpreta la situación de la deserción en Colombia desde un contexto latinoamericano, pero tal vez la guía de implementación del modelo de gestión de permanencia y graduación estudiantil en IES (MEN, 2015), es la que tiene un estudio más completo alrededor del fenómeno y la forma de abordarlo desde el punto de vista del gobierno nacional de Colombia.

8.3.1 Concepto de deserción

En el artículo de Rodríguez y Londoño(2011) se define que la deserción es “Fenómeno de abandono de los estudios debido a la confluencia de causas internas y externas a las IES”, y como lo manifiestan Sánchez, Navarro y García (2009) la diferencia de la mortalidad académica se puede manifestar así: “Esto lleva a definir Deserción como el abandono académico pero voluntario que el estudiante hace de un programa o de la universidad y mortalidad académica como el retiro forzoso”. Otro aspecto que se analiza en el trabajo de Gutiérrez et al. (2012), es que la deserción se acota como el abandono por dos periodos consecutivos del programa académico, ya que cuando este se produce por causas temporales y solucionables sólo se da por un periodo y luego se retoma el programa por parte del estudiante.

8.3.2 Causas de deserción en la educación superior

Varias investigaciones abordan las causas de la deserción, algunas enfocadas en una IES en particular, como el de Fontalvo, Castillo y Polo (2015), donde se analiza la correlación de

diferentes factores: ambientales universitarios, acceso a la ayuda, razones económicas, ambiente familiar, nivel de satisfacción, situación financiera externa, capacidad académica y elección de programa académico.

En el artículo de Barrera, Casallas y Sastre (2015) enfocado en la Universidad Piloto, se llega a la conclusión que algunos alumnos desertan de la universidad pero culminan sus estudios en otra IES, cuando el factor más influyente es académico o personal, pero que cuando es una deserción total del sistema educativo por lo general está asociado a un factor económico, y finalmente reconoce que la mejor forma de disminuir el problema es haciendo una detección temprana de los alumnos que presenten las características que puedan llevar a la deserción, aún en el proceso de admisión, y tener un sistema que monitoree y alerte oportunamente sobre las ayudas que deben darse a los alumnos.

En su investigación, Osorio, Bolancé y Castillo (2012), trabajan con datos de la Universidad Javeriana de Cali, y concluyen que hay razones que se suman para hacer más factible la salida del programa académico, como bajo nivel educativo de los padres y bajo rendimiento académico que los lleva a múltiple repetición de materias y finalmente a desistir, también señalan que según estudios de la Universidad Nacional y el Instituto Colombiano para el Financiamiento de la Educación Superior, ICFES (2002), se unen variables como el género, la edad, y las condiciones académicas y económicas para generar la deserción. En el análisis de su población objetivo, llegan a la conclusión que son más relevantes los factores académicos que las individuales, y destacan que la unión de los factores: ser hombre y estudiar Ciencias económicas o Ingenierías, lleva a tener el mayor factor de probabilidad para no graduarse, concluyendo la necesidad de usar un sistema que ayude a determinar la trayectoria universitaria como un modelo con múltiples causas de finalización.

La investigación de Lopera (2012) enfocada en la Universidad del Rosario, determina que los hombres, que deben trabajar y estudiar, y que provienen de otras regiones, tienen la mayor probabilidad de abandonar sus estudios, especialmente si son jóvenes, pero disminuye bastante cuando aumenta la edad.

En el artículo de Timarán(2009), se utiliza minería de datos para determinar que los factores de deserción son: la carrera cursada, con alta incidencia en carreras de la facultad de ciencias naturales y matemáticas si los estudiantes eran originarios de la ciudad donde estudiaban, y el factor económico, en un estudio posterior utilizando la misma herramienta, en un

artículo, Timarán, Calderón y Jiménez (2013) concluyen que existen factores de tipo académico adicionales a los anteriormente encontrados, como que si se supera en 4 el número de materias perdidas en los primeros semestres aumenta la probabilidad de retiro del programa, y en el artículo de Timarán y Jiménez (2014) con la misma herramienta, se encuentra que hay patrones de deserción muy marcados en menores de edad, que provienen de colegios públicos, que estudian Ingenierías, de estrato bajo social, promedio entre 3.7 y 4.0 y que han perdido entre 1 y 2 materias de ciencias básicas.

En algunos artículos se ha hecho un estudio general de las variables que causan la deserción, como en los artículos de: Fontalvo (2015), Chinome et al. (2016), Barragán y Urrego (2019), Barrera et al. (2015), Yepes D et al. (2007), Sánchez et al. (2009) y Londoño (2013), pero en el artículo de Acosta (Acosta, 2010) se estudia la influencia de los procesos de aprendizaje en la deserción, y en el estudio de Suárez-Montes y Díaz-Subieta (2015) y el de Reyes et al. (2019) se hace un análisis de los factores psicosociales.

Finalmente, en la Guía de implementación del modelo de gestión para disminuir la deserción, el MEN (2015) resalta que hay factores de tipo psicológicos, sociológicos, económicos, organizacionales e interaccionistas, como se puede observar en la figura 3 a continuación.

Figura 4 Factores de deserción



Fuente: MEN(2015, p. 13)

8.3.3 Modelos para el análisis de la deserción

De acuerdo con el artículo de Areth, Castro y Rodríguez (2015) se han generado dos modelos diferentes para analizar la deserción, un con respecto al tiempo y otro con respecto al espacio. El modelo del tiempo indica que existen tres tipos de deserción: precoz, mediana y tardía, y en lo referente al modelo espacial se habla de institucional e interna o del programa académico, es decir que el estudiante se retira del programa pero continua con otro programa de la misma IES o se retira definitivamente.

Por otra parte hay dos enfoques que se han establecido desde los años 70 y 80 del siglo anterior, inicialmente Tinto (1975) sugiere que el vínculo entre el estudiante y la entidad educativa es lo que permite al estudiante mantenerse en el programa y finalizarlo. Pero en contraste, primero Sweet (1986) y luego Kember (1989), complementando el trabajo de Tinto en

un enfoque de educación a distancia, agregan variables como el entorno familiar y la capacidad para organizar el tiempo que tiene el estudiante.

Posteriormente otros autores como Rovai (2003) proponen modelos con múltiples variables de análisis, separándolas en aquellas que se deben tomar en cuenta antes de la entrada del alumno a la institución, tales como: nivel educativo, género, edad, situación laboral y habilidades para el uso de la tecnología (Se hace énfasis en este aspecto porque su estudio se centra en la educación a distancia), y otras que se deben abordar cuando el alumno ya hace parte de entidad, diferenciándolas en externas como horas de trabajo y aspectos financieros, y un segundo grupo interno, relacionadas con autoestima, relaciones interpersonales, asesoramiento y otras variables psicosociales.

8.3.4 El sector de la educación superior en Colombia

Según los datos entregados por el MEN a diciembre de 2019 a la página Datos abiertos (Gobierno de Colombia, 2019a), son 589 IES, de las cuales 165 son Universidades, 253 son Instituciones universitarias o Escuelas tecnológicas, 100 Instituciones tecnológicas y 71 son Instituciones técnicas profesionales, de las cuales 427 son privadas y 162 son públicas, un gran avance en número ya que de acuerdo a Melo, Ramos y Hernández (2017, p. 73), en 2015 el país contaba con 290 instituciones, es decir que el número se duplicó en cuatro años, pero en 2020 aún no se tienen las cifras post pandemia para ver cómo se vio afectado el sector.

Adicionalmente, 371 de las IES se encuentran en las cuatro principales ciudades del país lo cual significa que el 63% de estas entidades se encuentran centralizadas en los centros de población más grandes del país, implicando que para poder estudiar una carrera de Educación superior, para muchas personas todavía es necesario hacer una migración a estos centros urbanos, sin embargo el panorama no es tan complejo, tomando en cuenta que 37% de las entidades se encuentran en ciudades secundarias ofreciendo alternativas a las personas de ingresos bajos.

Tomando otro grupo de datos del MEN a septiembre de 2019, con información consolidada hasta 2018 (Gobierno de Colombia, 2019), sobre las personas matriculadas en las IES, se puede observar que hay 8067 programas académicos ofertados, de los cuales 7429 son universitarios, 523 tecnológicos y 115 técnicos, y de los universitarios 3090 son de pregrado

(624 licenciaturas), 2491 de especialización, 1558 de maestría y 290 de doctorado, esto nos da a concluir que los programas universitarios de pregrado son los de mayor demanda por parte de los estudiantes con un 92%, mientras que los programas técnicos y tecnológicos no tienen tanta solicitud, ya que sólo son el 7.9%, y en el otro lado del espectro, los doctorados sólo son el 3.5% de los programas, pero tiene la explicación del prerrequisito de estudios previos y su costo.

En estos programas académicos se matricularon en 2018, 3.324.809 estudiantes, de los cuales 1.696.463 lo hicieron en el primer semestre y 1.628.346 en el segundo semestre, lo cual indica que están balanceados los dos semestres en número de matrículas, aunque se esperaría que al tener la mayoría de colegios calendario A, el primer semestre tenga mayor cantidad de nuevos estudiantes. Otro dato interesante encontrado es que 1.509.523 hombres se inscribieron, y 1.815.286 mujeres lo hicieron, lo cual es una buena noticia para el país, ya que se ha encontrado que el número de mujeres que desertan es menor que el de los hombres ICFES (2002), pero lo mejor de todo, es que de acuerdo a las metas establecidas por la Organización de las Naciones Unidas, ONU(2020), la quinta meta que es la paridad de géneros, ayuda a tener un planeta más pacífico, próspero y sostenible.

En la investigación de Melo, Ramos y Hernández (2017, p. 102) se determina que el sector tiene una cobertura baja comparada con otros países de la región, aunque resalta que la cobertura ha ido aumentando a partir de la constitución de 1991, y el establecimiento de las pruebas Saber cómo obligatorias a partir de la ley 1324 de 2009 permite tener medidas comparativas más exactas sobre la calidad. Adicionalmente el factor económico es muy decisivo, tanto para la tasa de deserción como para la calidad educativa, agudizando la desigualdad que existe en el país.

9 Hipótesis de trabajo

9.1 Hipótesis 1

Es posible generar recomendaciones de apoyo en tiempos adecuados, para detener el proceso de deserción usando DL

9.1.1 Variables hipótesis 1

Capacidad de generar resultados: Variable dependiente. Capacidad de un sistema basado en DL para soportar un sistema de recomendación para un problema social con las características de la deserción en la educación superior colombiana. Depende del tiempo requerido para generar una recomendación y de los costos que implique para una universidad utilizarlo. Su valor es booleano, positivo o negativo. Se podrá definir mediante la experimentación y revisión de la generación de la información en un sistema.

Tiempo de entrega apropiado de la información: Variable independiente. Debe ser determinado con las universidades. Se mide en días. Se medirá mediante la recolección en entrevistas con personal de las entidades de estudios superiores.

Cantidad de datos necesarios para que un sistema genere una recomendación acertada: Variable independiente. Es necesario saber cuántas mediciones de cambio de estado por persona se requiere para que el sistema haga un pronóstico. Debe ser medido en fracciones de Gigabytes. Se recolectará mediante experimentación con las plataformas.

9.2 Hipótesis 2

Las universidades están interesadas en usar un sistema que les permita tener recomendaciones de forma automática.

9.2.1 Variables hipótesis 2

Valor que una institución está dispuesta a pagar por estudiante: Variable independiente. Este costo es una condicional de viabilidad de la implementación del sistema en forma real. Debe ser determinado en las entrevistas a universidades y se mide en dólares debido a que los costos de las plataformas vienen en esa moneda. Se recolectará mediante entrevistas a las universidades.

10 Diseño metodológico

10.1 Tipo de investigación

El de tipo de investigación es descriptivo, porque permitirá medir varios aspectos como:

1. Capacidad de una plataforma *DL* para generar resultados en un problema social como es el propósito de esta investigación.
2. Determinar la capacidad de pago de las IES alrededor de una herramienta como esta para determinar la viabilidad de implementación.
3. Encontrar la plataforma de *DL* en la nube más adecuada para este sistema

Esta investigación tiene un enfoque mixto (Cualitativo y cuantitativo), ya que busca encontrar la plataforma más adecuada para implementar un sistema *DL* que ayude a solucionar el problema objetivo, pero a su vez servirá para medir el interés que tienen las IES en implementar o utilizar un sistema de este tipo.

El campo de investigación de esta monografía, es Ciencia tecnología e innovación, el grupo es ONTARE y la línea de investigación es Tecnología de la información.

10.2 Instrumentos a utilizar

Se utilizará una encuesta para las IES, en la cuál se busca explorar la recolección de la información acerca de deserción de sus alumnos, y la capacidad de inversión para mitigar este problema, medido en una cantidad de dólares por semestre que está dispuesta a invertir en la herramienta.

Para esta encuesta se busca entrevistar a las personas relacionadas con bienestar universitario o admisiones, que tengan como prioridad la atención del problema.

Para la parte de la herramienta, se hará una experimentación en las herramientas de DL en el mercado, esta investigación se enfocará en las principales plataformas en la nube que ofrecen este servicio de IA de acuerdo a Gartner, que son: AWS, Azure, Google Cloud e IBM, como se aprecia en la figura a continuación.

Figura 5 Cuadrante mágico de Gartner, IA en la nube



Fuente: Gartner 2020

10.3 Población y muestra

La investigación que se llevará a cabo se enfoca en las IES, que según las cifras del MEN a diciembre de 2019 (Gobierno de Colombia, 2019a), la población objetivo es de 589 IES, por lo cual aplicando la fórmula para encontrar la muestra representativa, tenemos que con un margen de error del 5% y un nivel de confianza del 90%, pero como la población no es homogénea, se hará una división por tipo de Institución con una muestra probabilística estratificada, de acuerdo a Fernández, Hernández y Baptista (1991, p. 217), se puede tomar un factor entre la muestra

calculada y la población, para encontrar los valores, como se observa en la tabla 2 a continuación.

Tabla 2 Tamaño de muestras representativas

Tipo de Institución	Total	Universidades	Instituciones universitarias o Escuelas tecnológicas	Instituciones tecnológicas	Instituciones técnicas profesionales
Población	589	165	253	100	71
Muestra	72	20	31	12	9

Fuente: propia

11 Contribuciones originales esperadas

Las implicaciones prácticas de esta monografía son la recolección de las necesidades de las IES, confrontándolas con las oportunidades que puede ofrecer una herramienta de *DL* y el planteamiento de uso de una plataforma comercial en la nube, también permitirá entender si las instituciones pueden invertir en esta herramienta, para verificar la viabilidad de la solución en un ambiente real, y si es viable, generar una oportunidad de negocio o una iniciativa liderada por el estado colombiano.

El valor teórico que tiene esta investigación es explorar las capacidades de la tecnología *DL*, enfocándose en la posibilidad que ofrece para hacer recomendaciones, para mitigar problemas con factores psicológicos, socio económicos y educativos, tomando en cuenta el estado del arte de las principales plataformas en la nube que ofrecen el servicio de IA.

La investigación permitirá examinar la población de estudiantes usando una correlación de patrones, que posiblemente no han sido tomados en cuenta en investigaciones anteriores, y es posible que genere una nueva herramienta para analizar el fenómeno de la deserción.

12 Cronograma

A continuación se encuentran las actividades, con lo cual se puede revisar en el cuadro el tiempo estimado: a) Selección de data para la herramienta, b) Presentación de primer informe c) Generación de herramienta para las cuatro plataformas que ofrecen servicios de IA en la nube, d) Revisión de resultados, e) Presentación de segundo informe, f) Preparación de la encuesta, g) Aplicación de la encuesta, h) Tabulación de resultados, i) Presentación de informe, j) Corrección de observaciones, k) Publicación de artículo y l) Presentación de informe final.

Tabla 3 Cronograma de la investigación

Actividades a desarrollar		Semanas	Tiempo del proyecto en meses (9 meses)																	
			Mes 1	Mes 2	Mes 3	Mes 4	Mes 5	Mes 6	Mes 7	Mes 8	Mes 9									
1	a	1	■																	
2	b	1		■																
3	c	10		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
4	d	1					■													
5	e	1					■													
6	f	1					■	■												
7	g	16					■	■	■	■	■	■	■	■	■	■	■	■	■	■
8	h	1																■		
9	i	1																	■	
10	j	1																		■
11	k	3																		■
12	l	1																		■

Fuente: propia

Dentro de las tareas principales, que son la generación de la herramienta en las plataformas en la nube, y la encuesta a IES se espera ir adelantando el trabajo escrito correspondiente a la monografía, es necesario acordar con el asesor los hitos de revisión.

13 Referencias

- Acosta, M. (2010). Los procesos de aprendizaje y su incidencia en la deserción estudiantil en el Programa de Química Farmacéutica de la Universidad de Cartagena. In *Psicología desde el Caribe* (Vol. 0, Issue 0).
- Aliper, A., Plis, S., Artemov, A., Ulloa, A., Mamoshina, P., y Zhavoronkov, A. (2016). Deep Learning Applications for Predicting Pharmacological Properties of Drugs and Drug Repurposing Using Transcriptomic Data. *Molecular Pharmaceutics*, 13(7), 2524–2530. Recuperado de: <https://doi.org/10.1021/acs.molpharmaceut.6b00248>
- Areth, J., Castro, J., y Rodríguez, H. (2015). La educación virtual en Colombia: exposición de modelos de deserción. *Apertura: Revista de Innovación Educativa*, 7(1), 10. Recuperado de: <http://search.ebscohost.com/login.aspx?direct=true&db=ehh&AN=102888761&lang=es&site=ehost-live>
- Aspray, W., Bromley, G., Campbell-Kelly, M., Ceruzzi, P., y Williams, M. (1990). *Computing before computers*. Iowa State University Pres.
- Banco mundial. (2018). Overview: Learning to realize education's promise. In *World Development Report 2018: Learning to Realize Education's Promise* (pp. 1–35). Recuperado de: https://doi.org/10.1596/978-1-4648-1096-1_ov
- Barragán, D., y Patiño, L. (2016). Elementos para la comprensión del fenómeno de la deserción universitaria en Colombia. Más allá de las mediciones. *Cuadernos Latinoamericanos de Administración*, 9(16), 55. Recuperado de: <https://doi.org/10.18270/cuaderlam.v9i16.1248>
- Barrera, O., Casallas, P., y Sastre, A. (2015). Factores asociados a la deserción estudiantil en la universidad piloto de Colombia. *V Clabes*.
- Barrero, F. (2015). Investigación en deserción estudiantil universitaria: educación, cultura y significados. *Revista Educación y Desarrollo Social*, 9(2), 86–101. Recuperado de: <https://doi.org/10.18359/reds.948>
- Batmaz, Z., Yurekli, A., Bilge, A., y Kaleli, C. (2019). A review on deep learning for

- recommender systems: challenges and remedies. *Artificial Intelligence Review*, 52(1), 1–37. Recuperado de: <https://doi.org/10.1007/s10462-018-9654-y>
- Betancur, M., y González, D. (2008). Deserción estudiantil en la universidad de Ibagué, colombia: una lectura histórica en perspectiva cuantitativa. *Revista Del Instituto de Estudios En Educación*, 9.
- Botelho, A., Baker, R., y Heffernan, N. (2017). *Improving Sensor-Free Affect Detection Using Deep Learning BT - Artificial Intelligence in Education*. pp. 40–51. Springer International Publishing.
- Bravo, M., y Mejía, A. (2010). Los retos de la educación superior en Colombia: Una reflexión sobre el fenómeno de la deserción universitaria. *Revista Educación En Ingeniería*, 5(10), 85–98. Recuperado de: <http://www.educacioneningenieria.org/index.php/edi/article/view/101>
- Campbell, J. (1986). On artificial intelligence. *Artificial Intelligence Review*, 1(1), 3–9. Recuperado de: <https://doi.org/10.1007/BF01988524>
- Cao, C., Liu, F., Tan, H., Song, D., Shu, W., Li, W., Zhou, Y., Bo, X., y Xie, Z. (2018). Deep Learning and Its Applications in Biomedicine. *Genomics, Proteomics and Bioinformatics*, 16(1), 17–32. Recuperado de: <https://doi.org/10.1016/j.gpb.2017.07.003>
- Cheng, H., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., Anderson, G., Corrado, G., Chai, W., Ispir, M., Anil, R., Haque, Z., Hong, L., Jain, V., Liu, X., y Shah, H. (2016). Wide & deep learning for recommender systems. *ACM International Conference Proceeding Series, 15-Septemb*, 7–10. Recuperado de: <https://doi.org/10.1145/2988450.2988454>
- Chinome, P., Ruiz, C., y Fernández, L. (2016). Priorización de variables en el diseño de un sistema de gestión integral de la deserción estudiantil. *Revista Educación En Ingeniería*, 11(22), 69–77. Recuperado de: <https://doi.org/10.26507/rei.v11n22.683>
- Da’u, A., y Salim, N. (2020). Recommendation system based on deep learning methods: a systematic review and new directions. In *Artificial Intelligence Review* (Vol. 53, Issue 4). Springer Netherlands. Recuperado de: <https://doi.org/10.1007/s10462-019-09744-1>
- Doleck, T., Lemay, D., Basnet, R., y Bazelais, P. (2020). Predictive analytics in education: a comparison of deep learning frameworks. *Education and Information Technologies*, 25(3), 1951–1963. Recuperado de: <https://doi.org/10.1007/s10639-019-10068-4>

- Flórez, E., y Carrascal, J. (2016). Estudio de la deserción estudiantil de la Licenciatura en Ciencias Naturales y Educación Ambiental de la Universidad de Córdoba-Colombia- 2011 – 2015. - Study of student desertion of the Degree in Natural Sciences and Environmental Education of the Univer. *Revista Científica*, 4(27), 340. Recuperado de: <https://doi.org/10.14483/udistrital.jour.rc.2016.27.a4>
- Fontalvo, W. (2015). Análisis multivariado para determinar los factores más relevantes de deserción estudiantil presentes en el programa de Ingeniería Industrial de una Universidad del Caribe colombiano. *Prospectiva*, 13(1), 86. Recuperado de: <https://doi.org/10.15665/rp.v13i1.363>
- Gheller, C., Vazza, F., y Bonafede, A. (2018). Deep learning based detection of cosmological diffuse radio sources. *Monthly Notices of the Royal Astronomical Society*, 480(3), 3749–3761. Recuperado de: <https://doi.org/10.1093/mnras/sty2102>
- Gobierno de Colombia. (2019a). *INSTITUCIONES DE EDUCACION SUPERIOR*. Datos Abiertos. Recuperado de: <https://www.datos.gov.co/Educacion/INSTITUCIONES-DE-EDUCACION-SUPERIOR/iwb2-skc4>
- Gobierno de Colombia. (2019b). *MEN_MATRICULA-ESTADISTICA_ES*. Datos Abiertos. Recuperado de: https://www.datos.gov.co/Educacion/MEN_MATRICULA-ESTADISTICA_ES/5wck-szir
- Gutiérrez, M. y Zea, E. (2012). Análisis de la deserción estudiantil en los programas de pregrado de la Universidad EAFIT. *Cuadernos de Investigación*, 0(81).
- Heredia, A., Silvestre, J., y Cruz, B. (2017). Educación y competitividad; un análisis de los problemas y necesidades que imperan en los distintos entornos, para la competitividad de los programas de negocio con base a las demandas de los sectores productivos. *Red Internacional de Investigadores En Competitividad*, 4(1), 2100–2119. Recuperado de: <https://riico.net/index.php/riico/article/view/837/505>
- Hernández, R., Fernández, C., y Baptista, P. (1991). *Metodología de la investigación*. McGraw Hill.
- Herrera, F. y Analide, C. (2012). Editors' Preface. *Progress in Artificial Intelligence*, 1(1), 1. Recuperado de: <https://doi.org/10.1007/s13748-012-0009-7>
- High, P. (2017). Carnegie Mellon Dean Of Computer Science On The Future Of AI. *Forbes*. Recuperado de: <https://www.forbes.com/sites/peterhigh/2017/10/30/carnegie-mellon-dean->

of-computer-science-on-the-future-of-ai/#62f617e42197

- Hinton, G. y Osindero, S. (2006). A Fast Learning Algorithm for Deep Belief Nets Geoffrey. *MIT Press, 1554*, 341–346. Recuperado de: <https://doi.org/10.7763/ijesd.2010.v1.67>
- ICFES, y Colombia, U. N. de. (2002). *Estudio de la deserción estudiantil en la educación superior en Colombia*.
- Isaza Restrepo A., Enríquez Guerrero C., y Pérez I. (2016). Deserción y rezago académico en el programa de medicina de la Universidad del Rosario, Bogotá, Colombia. *Ciencias de La Salud, 14*(2), 231–245. Recuperado de: <https://doi.org/10.12804/revsalud14.02.2016.08>
- Jiang, Y., Bosch, N., Baker, R., Paquette, L., Ocumpaugh, J., Andres, J., Moore, A. y Biswas, G. (2018). *Expert Feature-Engineering vs. Deep Neural Networks: Which Is Better for Sensor-Free Affect Detection? BT - Artificial Intelligence in Education* pp. 198–211). Springer International Publishing.
- Kember, D. (1989). A longitudinal-process model of drop-out from distance education. *Journal of Higher Education, 60*(3), 278–301.
- Ker, J., Wang, L., Rao, J., y Lim, T. (2018). Deep Learning Applications in Medical Image Analysis. *IEEE Access, 6*, 9375–9389. Recuperado de: <https://doi.org/10.1109/ACCESS.2017.2788044>
- Klinger, J., Mateos-Garcia, J., y Stathoulopoulos, K. (2018). Deep Learning, Deep Change? Mapping the Development of the Artificial Intelligence General Purpose Technology. *SSRN Electronic Journal*. Recuperado de: <https://doi.org/10.2139/ssrn.3233463>
- Langley, P. (1986). Editorial: On Machine Learning. *Machine Learning, 1*(1), 5–10. h Recuperado de: <https://doi.org/10.1023/A:1022687019898>
- Londoño Ardila, L. (2013). Factores de riesgo presentes en la deserción estudiantil en la Corporación Universitaria Lasallista. *Revista Virtual Universidad Católica Del Norte, 1*(38), 183-194–194.
- Lopera, C. (2012). Determinantes de la deserción universitaria en la Facultad de Economía de la Universidad del Rosario. *Reponame:Repositorio Institucional EdocUR*, 254. Recuperado de: <http://repository.urosario.edu.co/handle/10336/3740>
- Melo, L., Ramos, J., y Hernández, P. (2017). La educación superior en Colombia: situación actual y análisis de eficiencia. *Desarrollo y Sociedad, 2017*(78), 59–111. Recuperado de: <https://doi.org/10.13043/DYS.78.2>

- MEN. (2015). Guía Para La Implementación del modelo de gestión de permanencia y graduación estudiantil en instituciones de Educación Superior. In *Mineeducacion, todos por un nuevo pais. paz, equidad, educacion*. Recuperado de:
http://www.colombiaaprende.edu.co/html/micrositios/1752/articles-355193_guia_.pdf
- Meseguer, P. (2012). Towards 40 years of constraint reasoning. *Progress in Artificial Intelligence*, 1(1), 25–43. Recuperado de: <https://doi.org/10.1007/s13748-011-0006-2>
- Michelucci, U. (2018). Applied Deep Learning. In *Applied Deep Learning*. Recuperado de: <https://doi.org/10.1007/978-1-4842-3790-8>
- Mijwel, M. (2015). History of Artificial Intelligence. *Computer Science, April 2015*, 3–4.
- Ministerio de Educación Nacional. (2010). Diagnóstico de la deserción en Colombia. *Educación Superior*, 4–6.
- Mitchell, T. (1997). *Machine Learning*. McGraw-Hill Science/Engineering/Math.
<https://doi.org/10.1109/ICDAR.2019.00014>
- Naumov, M., Mudigere, D., Shi, H., Huang, J., Sundaraman, N., Park, J., Wang, X., Gupta, U., Wu, C., Azzolini, A., Dzhulgakov, D., Malleovich, A., Cherniavskii, I., Lu, Y., Krishnamoorthi, R., Yu, A., Kondratenko, V., Pereira, S., Chen, X. y Smelyanskiy, M. (2019). *Deep Learning Recommendation Model for Personalization and Recommendation Systems*. Recuperado de: <http://arxiv.org/abs/1906.00091>
- ONU. (2020). *Goal 5: Achieve gender equality and empower all women and girls*. Recuperado de: <https://www.un.org/sustainabledevelopment/gender-equality/>
- Oppermann, A. (2019). *What is Deep Learning and How does it work?* Towardsdatascience.Com. Recuperado de: <https://towardsdatascience.com/what-is-deep-learning-and-how-does-it-work-2ce44bb692ac>
- Osorio, A., Bolancé, C., y Castillo, M. (2012). Deserción y graduación estudiantil universitaria: una aplicación de los modelos de supervivencia. *Revista Iberoamericana de Educación Superior*, iii, 31–57. Recuperado de: <https://doi.org/10.22201/iissue.20072872e.2012.6.55>
- Paramo, G., y Correa, C. (2012). Deserción estudiantil universitaria. Conceptualización. *Revista Universidad EAFIT* (Vol. 35, Issue 114, pp. 65–78).
- Partridge, D. (1986). Engineering artificial intelligence software. *Artificial Intelligence Review*, 1(1), 27–41. Recuperado de: <https://doi.org/10.1007/BF01988526>
- Perrotta, P. (2020). *Programming Machine Learning: From Coding to Deep Learning*.

Pragmatic Bookshelf.

- Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L., y Sohl-Dickstein, J. (2015). Deep knowledge tracing. *Advances in Neural Information Processing Systems, 2015-Janua*, 505–513.
- Rahimy, E. (2018). Deep learning applications in ophthalmology. *Current Opinion on Ophthalmology, 29*(3), 254–260.
- Reyes, L., Castañeda, E., y Pabón, D. (2019). Causas psicosociales de la deserción universitaria. *Revista Logos, Ciencia & Tecnología, 12*(1). Recuperado de: <https://doi.org/10.22335/rlct.v4i1.179>
- Rodríguez, L., y Londoño, F. (2011). Estudio sobre deserción estudiantil en los programas de Educación de la Católica del Norte Fundación Universitaria A Study on Student ' s Dropout in Education Programs of the Catholic of the North University Foundation Étude sur l ' abandon scolaire aux p. *Revista Virtual Universidad Católica Del Norte, 33*, 328–355. Recuperado de: <http://www.redalyc.org/articulo.oa?id=194218961018%0ACómo>
- Rodríguez, M. (2019). La investigación sobre deserción universitaria en Colombia 2006-2016. Tendencias y resultados. *Pedagogía y Saberes, 51*, 49–66. Recuperado de: <https://doi.org/10.17227/pys.num51-8664>
- Rovai, A. (2003). In search of higher persistence rates in distance education online programs. *Internet and Higher Education, 6*, 1–16.
- Sánchez, G., Navarro, W., y García, A. (2009). Factores de deserción estudiantil en la Universidad Surcolombiana. *Paideia Surcolombiana, 1*(14), 97. Recuperado de: <https://doi.org/10.25054/01240307.1083>
- Singaravel, S., Suykens, J., y Geyer, P. (2018). Deep-learning neural-network architectures and methods: Using component-based models in building-design energy prediction. *Advanced Engineering Informatics, 38*(May), 81–90. Recuperado de: <https://doi.org/10.1016/j.aei.2018.06.004>
- Singhal, A., Sinha, P., y Pant, R. (2017). Use of Deep Learning in Modern Recommendation System: A Summary of Recent Works. *International Journal of Computer Applications, 180*(7), 17–22. Recuperado de: <https://doi.org/10.5120/ijca2017916055>
- Skansi, S. (2018). *Introduction to Deep Learning*. Springer.
- Smola, A., y Vishwanathan, S. (2008). Introduction to machine learning. In *Methods in*

- Molecular Biology* (Vol. 1107). Cambridge University Press. Recuperado de:
https://doi.org/10.1007/978-1-62703-748-8_7
- SPADIES. (2017). *Estadísticas de desercion*. Recuperado de:
<https://www.mineducacion.gov.co/sistemasinfo/spadies/Informacion-Institucional/357549:Estadisticas-de-Desercion>
- Suárez, N., y Díaz, L. (2015). Estrés académico, deserción y estrategias de retención de estudiantes en la educación superior. *Revista de Salud Publica*, 17(2), 300–313. Recuperado de: <https://doi.org/10.15446/rsap.v17n2.52891>
- Sweet, R. (1986). Student dropout in distance education: An application of Tinto's model. *Distance Education*, 7(2), 201–213.
- Timarán, R. (2009). Detección de Patrones de Bajo Rendimiento Académico y Deserción Estudiantil con Técnicas de Minería de Datos. *Iiisorg*. Recuperado de:
<http://www.iiis.org/CDs2009/CD2009CSC/CISCI2009/PapersPdf/C692YV.pdf>
- Timarán, R., Calderón, A., y Jiménez, J. (2013). Descubrimiento De Perfiles De Deserción Estudiantil Con Técnicas De Minería De Datos. *Revista Vínculos*, 10(1), 373–383. Recuperado de: <https://doi.org/10.14483/2322939X.4687>
- Timarán, R., y Jiménez, J. (2014). Detección de Patrones de Deserción Estudiantil en Programas de Pregrado de Instituciones de Educación Superior con CRISP-DM. *Formacion Universitaria*, 1–19. Recuperado de:
<https://www.oei.es/historico/congreso2014/memoriactei/758.pdf>
- Tinto, V. (1975). Dropout from higher education, a theoretical synthesis of recent research. *Review of Educational Research*, 45(1), 88–125.
- Turing, A. (1936). On computable numbers. *Proceedings of the London Mathematical Society*, 42, 230–265. Recuperado de: <https://doi.org/10.2307/2268810>
- Wilson, K., Karklin, Y., Han, B., y Ekanadham, C. (2016). Back to the basics: Bayesian extensions of IRT outperform neural networks for proficiency estimation. *Proceedings of the 9th International Conference on Educational Data Mining, EDM 2016*, 539–544.
- Yepes D, , Beltrán, M., Arrubla, J., Marín, L., Martínez, M., y Hoyos, A. (2007). Factores causales de la deserción estudiantil en el pregrado de la Facultad de odontología de la Universidad de Antioquia de 1997 a 2004. *Rev. Fac. Odontol. Univ. Antioq*, 19, 35–48.
- Zárate, R., y Mantilla, E. (2015). The student dropout UIS, a view from the university social

responsibility. *Zona Próxima*, 21, 121–134. Recuperado de:
<https://doi.org/10.14482/zp.21.6061>

Zhao, H., Zhang, C., Ning, J., Zhang, B., Sun, P., y Feng, Y. (2019). A comparative study of the evolutionary many-objective algorithms. *Progress in Artificial Intelligence*, 8(1), 15–43. Recuperado de: <https://doi.org/10.1007/s13748-019-00174-2>

Sugerencia (1) de director de trabajo de grado:	
Información	Nombre: Luis Cobo
del docente	Correo institucional: lacobo@universidadean.edu.co
Sugerencia (2) de director de trabajo de grado:	
Información	Nombre: Eduard Galvis
del docente	Correo institucional: egalvis@universidadean.edu.co