

**DEFINICIÓN DE UN MODELO PREDICTIVO PARA LA DESERCIÓN
ESTUDIANTIL EN EDUCACIÓN VIRTUAL Y A DISTANCIA**

**IVÁN DARÍO ACOSTA CONTRERAS
GIOVANNY ALBERTO PUENTES**

**Escuela Colombiana de Ingeniería Julio Garavito
Decanatura de Ingeniería de Sistemas
Maestría Gestión de Información
BOGOTÁ D.C., Colombia 2019**

**DEFINICIÓN DE UN MODELO PREDICTIVO PARA LA DESERCIÓN
ESTUDIANTIL EN EDUCACIÓN VIRTUAL Y A DISTANCIA**

IVÁN DARÍO ACOSTA CONTRERAS

GIOVANNY ALBERTO PUENTES

Trabajo de investigación para optar al título de

Magíster en Gestión de Información

Modalidad: Profundización

Director

Oswaldo Castillo

Doctor en Ingeniería de Sistemas

Escuela Colombiana de Ingeniería Julio Garavito

Decanatura de Ingeniería de Sistemas

Maestría Gestión de Información

BOGOTÁ D.C., Colombia 2019

© Únicamente se puede usar el contenido de las publicaciones para propósitos de información. No se debe copiar, enviar, recortar, transmitir o redistribuir este material para propósitos comerciales sin la autorización de la Escuela Colombiana de Ingeniería. Cuando se use el material de la Escuela se debe incluir la siguiente nota “Derechos reservados a Escuela Colombiana de Ingeniería” en cualquier copia en un lugar visible. Y el material no se debe notificar sin el permiso de la Escuela.

Publicado en 2019 por la Escuela Colombiana de Ingeniería Julio Garavito. Avenida 13 No 205-59 Bogotá. Colombia
TEL: +57 – 1 668 36 00

Aceptación del jurado

El Trabajo de grado de maestría titulada “Definición de un Modelo Predictivo para la Deserción Estudiantil en Educación Virtual y a Distancia”, presentada por Iván Acosta Contreras y Giovanni Alberto Puentes, cumple con los requisitos establecidos para optar al título de Magíster en Gestión de Información.

Oswaldo Castillo Navetty
Director trabajo de grado

Nombre
Jurado

Nombre
Jurado

Bogotá, D.C., 16 de diciembre de 2019.

Reconocimiento o Agradecimientos

Como siempre gracias a mis padres por su incondicional apoyo en la educación ya que es un “tesoro que nadie te puede quitar” por eso es importante “Nunca pare de aprender en la Vida”, a mi esposa y a mi hijo por que se que muchos de los peldaños recorridos los hago y los hare en compañía de ustedes. A mi amigo @gururea por un ejemplo a seguir y finalmente mil gracias a mi compañero y al director de tesis Oswaldo Castillo, por confiar en la idea y ser una guía durante el proceso.

Resumen

La deserción universitaria es una de las preocupaciones más relevante e importantes que evalúan las Instituciones de Educación Superior (IES) en Colombia. Es por ello que este documento describe y propone un modelo que permite entender e identificar requerimientos institucionales acerca de la deserción estudiantil y además se mapeen todos los cambios que se les requerirán a la institución para lograr los beneficios y resultados esperados, también se ilustra como a través de tecnologías digitales y metodologías de análisis de datos se logra identificar niveles de precisión más confiable a fin de predecir posibles estudiantes que pueden caer en el ausentismo estudiantil y/o deserción estudiantil.

Con esto en contexto, se ha realizado en un escenario de la vida real un enfoque practico de la metodología propuesta y en compañía de una universidad en Colombia que ha sido pionera en la modalidad de educación virtual, en donde se realizo el entendimiento estratégico y operativo de las áreas de negocio que participan en los procesos de análisis de información para la deserción estudiantil y junto con ello se aplico algoritmos y técnicas de análisis de datos a los alumnos permitiendo obtener resultados experimentales que lograron niveles confiables de precisión para identificar predicciones de ausentismo y deserción estudiantil. El resultado final se presenta a discusión con la institución a fin de que la universidad pueda definir iniciativas que permitan disminuir la tasa de abandono mediante la identificación de posibles causas de deserción estudiantil.

Palabras Clave

Deserción, Retención, Persistencia Estudiantil

Abstract

University desertion is one of the most relevant and important concerns evaluated by Higher Education Institutions in Colombia. For this reason, the purpose of this document is to describe a proposed model that allows for the understanding and identification of institutional requirements regarding student dropout and also pathway all the changes that will be required of the institution in order to achieve the expected benefits and results. It also illustrates how, through digital technologies and data analysis methodologies, it is possible to identify more reliable levels of precision in order to predict possible students who may fall into student absenteeism and/or dropout.

With this context, a practical approach of the proposed methodology has been carried out in a real life scenario and in the company of a university in Colombia that has been a pioneer in the virtual education modality, where the strategic and operative understanding of the business areas that participate in the processes of information analysis for student dropout was carried out and together with this algorithms and techniques of data analysis were applied to the students allowing to obtain experimental results that achieved reliable levels of precision to identify predictions of student absenteeism and dropout. The final result is presented for discussion with the institution so that the university can define initiatives to reduce the dropout rate by identifying possible causes of student dropout.

Keywords

Dropout, Retention, Student Persistence

Contenido

1	Introducción	13
2	Descripción del problema	15
3	Objetivos, alcance y limitaciones	20
3.1	Objetivo general	20
3.2	Objetivos específicos	20
3.3	Alcance y limitaciones	21
4	Metodología	21
5	Marco teórico	23
5.1	Tinto	23
5.2	Spady	25
6	Estado del arte	26
6.1	La deserción estudiantil.	26
6.2	Predicción del rendimiento académico mediante minería de datos	29
7	Marco metodológico y caso de estudio	34
7.1	Fase No.1 - Marco metodológico	34
7.1.1	Alineamiento conceptual	35
7.1.2	Análisis de capacidades institucionales	37
7.1.3	Formato de entendimiento de deserción	38
7.1.4	Marco de trabajo	39

7.2	Fase No. 2 - Caso de estudio	46
7.2.1	Institución Educativa: Universidad Minuto de Dios (Rectoría de Educación Virtual y a Distancia)	47
7.2.2	Alineamiento conceptual (Rectoría de Educación Virtual y a Distancia)	50
7.2.3	Desarrollo del trabajo	54
7.2.3.1	Definición de deserción	54
7.2.3.2	Entendimiento de los datos	55
7.2.3.3	Preparación de los datos	58
7.2.3.4	Modelamiento	62
7.3	Despliegue de componentes tecnológicos	72
7.4	Conclusiones del modelo	73
8	Conclusiones y Recomendaciones	73
8.1	Conclusiones	73
8.2	Trabajos Futuros	74
9	Referencias	76

Lista de tablas

Tabla 1. <i>Técnicas de minería de datos</i>	31
Tabla 2. <i>Descripción de Hojas de Artefactos</i>	45
Tabla 3. <i>Limpieza de atributos de DataSet</i>	58
Tabla 4. <i>Descripción pasos de Preparación de Tabla Resumen</i>	59
Tabla 5. <i>Descripción de atributos de Tabla Resumen</i>	60
Tabla 6. <i>Descripción de Iteraciones de Grupo</i>	66
Tabla 7. <i>Descripción de Segmentos Detectados - kmeans</i>	67

Lista de figuras

<i>Figura 1.</i> Porcentaje de desertores según género entre 1998-2014 (SNIES)	16
<i>Figura 2.</i> Porcentaje de desertores según clasificación de exámenes entre 1998-2014 (SNIES)	17
<i>Figura 3.</i> Porcentaje de desertores según clasificación de exámenes entre 1998-2014 (SNIES)	18
<i>Figura 4.</i> Porcentaje de desertores según clasificación de exámenes entre 1998-2014 (SNIES)	19
<i>Figura 5.</i> Desarrollo de las fases.....	22
<i>Figura 6.</i> Esquema conceptual de abandono escolar	23
<i>Figura 7.</i> Modelo de Spady. Abandono escolar basado en Durkheim	26
<i>Figura 10.</i> Factores comunes en la deserción	33
<i>Figura 11.</i> Proceso y artefactos usados en el marco metodológico	35
<i>Figura 12.</i> Red de dependencia de beneficios - Deserción estudiantil	37
<i>Figura 13.</i> Artefacto de entendimiento de capacidades.....	38
<i>Figura 14.</i> Artefacto resumen de alineamiento conceptual de deserción estudiantil	39
<i>Figura 15.</i> Ciclo de vida de metodología CRISP-DM.....	40
<i>Figura 16.</i> Marco de trabajo propuesto.....	41
<i>Figura 17.</i> Mapa estratégico - Uniminuto.....	48
<i>Figura 18.</i> Participación de los estudiantes por modalidad de estudios	49
<i>Figura 19.</i> Cifras de deserción de Rectoría IEVD - Uniminuto	50
<i>Figura 20.</i> Artefacto - Análisis de capacidades en Gestión de Comunicaciones - Uniminuto	51
<i>Figura 21.</i> Artefacto - Análisis de capacidades en bienestar estudiantil - Uniminuto	52

<i>Figura 22.</i> Artefacto - Análisis de capacidades en asesores académicos - Uniminuto.....	52
<i>Figura 23.</i> Artefacto - Análisis de capacidades en asesores académicos - Uniminuto.....	53
<i>Figura 24.</i> Artefacto - Alineamiento conceptual de deserción estudiantil - Uniminuto.....	54
<i>Figura 25.</i> Validación orígenes de datos del caso de estudio	55
<i>Figura 26.</i> Variables Significativas	63
<i>Figura 27.</i> Variables Definitivas.....	64
<i>Figura 28.</i> Histograma Residaj	64
<i>Figura 29.</i> División por ausentismo.....	70
<i>Figura 30.</i> División por estrato socioeconómico	70
<i>Figura 31.</i> Ramas del árbol.....	71
<i>Figura 32.</i> Diagrama de despliegue de componente de tecnología	72

1 Introducción

La deserción y permanencia estudiantil constituyen un problema que impacta a las IES alrededor del mundo, y la preocupación por este tema es compartida por los diferentes sistemas de calidad que evalúan la educación superior. Para tener una magnitud o dimensión del problema existen varios indicadores (Ministerio de Educación Nacional, 2011), pero quizás el más relevante es el que permite evidenciar las tasas de deserción anual en las IES, el cual se mide en términos de porcentaje de estudiantes activos que se encontraban como matriculados un año antes y que al siguiente año o periodo académico figuran como desertores. Por ejemplo, para Colombia es una tasa superior al 10% al año 2014, y es una no tan alarmante si se compara con países en la región, como Brasil (18%) y Estados Unidos (17%).

En Colombia, desde hace una década aproximadamente, el Ministerio de Educación Nacional (MEN) ha comenzado a trabajar en las IES a fin de que, dentro de los planes estratégicos institucionales, logren promover y llevar a cabo iniciativas que busquen reducir las tasas de deserción e incrementar la cantidad de estudiantes que comienzan a gozar de su título profesional. Pero para llevar a cabo estos propósitos institucionales, es necesario identificar las causas o motivos de por qué los jóvenes no logran obtener su título profesional en los tiempo debidos; combatir y contrarrestar el acelerado ritmo y las altas tasas de deserción universitaria que se presentan en los programas de las diferentes modalidades de educación, como lo son pregrado, posgrado y educación continua; y reconocer cómo afecta este problema a estudiantes presenciales y de educación virtual y a distancia.

Específicamente en este país existían diversos enfoques metodológicos para analizar la deserción, sin embargo, el MEN puso al servicio y al alcance de las IES, desde el año 2006, el Sistema de Prevención y Análisis de la Deserción en las Instituciones de Educación Superior

(SPADIES), que para efectos prácticos solamente se considera o evalúa el modelo de “First drop Out” (Hernández, 2009) es decir, el abandono de un programa académico por dos semestres consecutivos. Esta circunstancia también es conocida como primera deserción, lo cual sirve como señal o alerta para que las instituciones puedan diseñar estrategias a fin de evitar la posible deserción de estudiantes en el futuro.

Pese a ello, en la mayoría de las ocasiones, las instituciones realizan este análisis de manera post mortem, es decir, se analiza la información cuando el estudiante ya abandonó o desertó del programa académico o la institución, muchas veces sin entender o tener el conocimiento de cuáles fueron los motivos o causas por las que el estudiante se fue de la universidad.

Para finalizar, es importante mencionar que la deserción es un fenómeno que se ha venido contrarrestando y en muchas instituciones en Colombia ha venido decreciendo gracias a diferentes medidas nacionales y a las iniciativas institucionales, mas es preciso detectar a tiempo a aquellos estudiantes que pueden clasificar como posibles desertores. De ahí la necesidad de que las universidades desarrollen estrategias que motiven a los estudiantes y que permitan una retroalimentación oportuna y continua a los mismos; además de ofrecer asistencia adecuada a los requerimientos de la comunidad estudiantil, por medio de las oficinas o áreas de éxito estudiantil, para potencializar estas metodologías.

En particular cabe implementar dichas estrategias en la educación en ambientes virtuales o digitales, puesto que estos son uno de los canales o espacios más prometedores como nuevas fuentes de ingresos en las universidades, donde se da la utilización del conocimiento como actor principal del desarrollo social y productivo.

2 Descripción del problema

La deserción estudiantil es un problema que impacta las IES, tanto en su labor social como en sus iniciativas y estrategias, económicas y de sostenibilidad. Afortunadamente, cada vez son más los Gobiernos y los ministerios de educación de cada país que brindan u ofrecen herramientas a favor de fortalecer la permanencia de los estudiantes en las instituciones, y de esta manera lograr un efecto positivo en el incremento de las tasas de graduación.

Al respecto es pertinente aclarar que hasta hace algunos años las universidades solamente hablaban de deserción, lo cual hacía evidenciar cifras negativas en su gestión al problema, por ese motivo, ahora han cambiado de enfoque y se habla de permanencia estudiantil en donde el principal objetivo es prevenir que el estudiante abandone sus estudios universitarios a partir de la motivación del estudiante para lograr alcanzar sus propias metas educativas. No obstante, las IES tienen varios retos para enfrentar la problemática de la deserción, entre ellos, enfocar esfuerzos en promover la continuidad de los estudios para aquellos estudiantes que pueden ser posibles desertores del sistema educativo universitario, y desarrollar metodologías o planes de trabajo que contribuyan a mejorar el rendimiento académico de los estudiantes.

El MEN, junto con el sistema de información SNIES (Sistema Nacional de Información de la Educación Superior), ha comenzado a trabajar en elementos fundamentales alrededor de la problemática de la deserción estudiantil; a partir de diferentes estudios y análisis que se han realizado durante los últimos años, se ha logrado identificar los diferentes factores o elementos que intervienen en la permanencia estudiantil, y a su vez, se han reconocido en detalle los aspectos y las variables asociadas a dicho fenómeno. Lo anterior se ha hecho con el propósito de que las universidades puedan trabajar en la estimación del riesgo de deserción de los estudiantes,

para facilitar el seguimiento a estos y así definir estrategias que contrarresten la deserción estudiantil.

Al profundizar en cada uno de estos grupos de variables se puede ver una diferente caracterización de los estudiantes, en donde se evidencia un comportamiento de la deserción. Por ejemplo, en el grupo de aspectos personales, como lo es el género, se observa en la siguiente figura que la mayor población de estudiantes desertores está conformada por hombres, con una diferencia de casi 10% por encima de las mujeres.

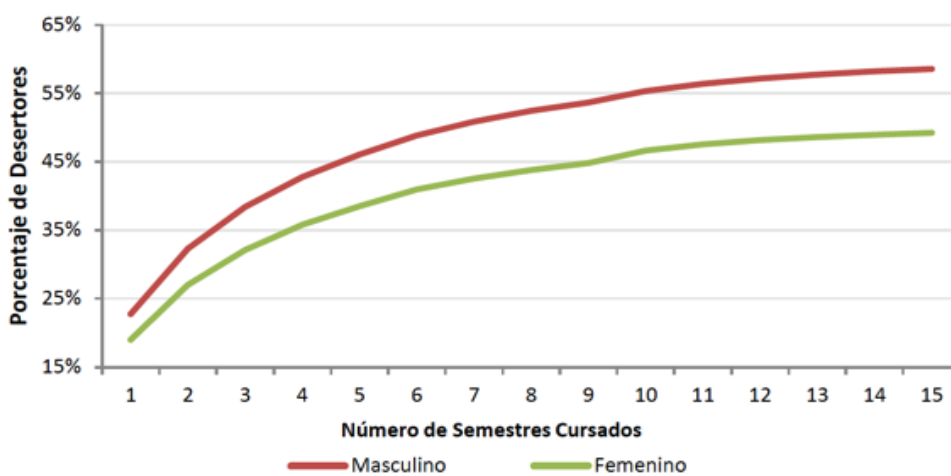


Figura 1. Porcentaje de desertores según género entre 1998-2014 (SNIES)

Fuente: (Ministerio de Educación, 2014)

Aún con base en el análisis realizado por el SNIES, respecto a los aspectos académicos, se halla una relación inversa entre la clasificación de los estudiantes según los resultados en las pruebas realizadas en la universidad y la tasa de deserción, pues se demuestra que los estudiantes con bajo desempeño académico presentan mayor tasa de deserción, y los de mejor desempeño académico las menores tasas.

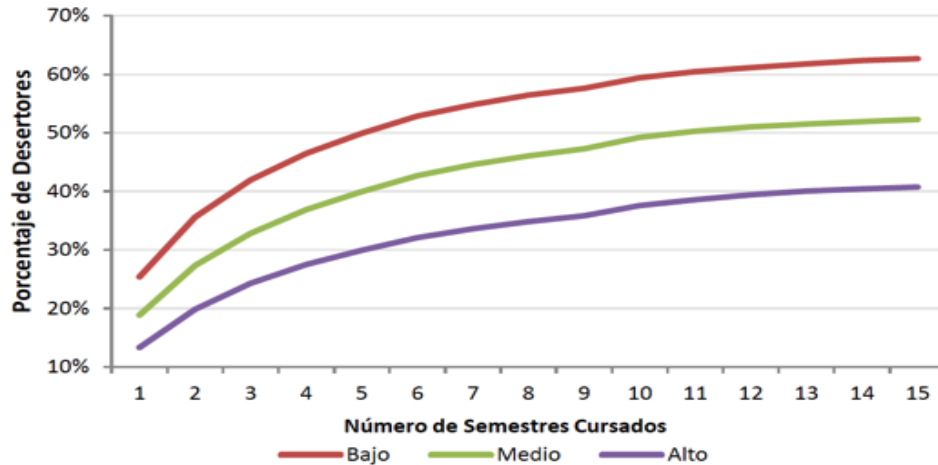


Figura 2. Porcentaje de desertores según clasificación de exámenes entre 1998-2014 (SNIES)

Fuente: (Ministerio de Educación, 2014)

Sobre el tercer aspecto, relacionado con variables de índole socioeconómica, y que según análisis realizados por el SNIES es uno de los criterios que eleva la tasa o porcentaje de estudiantes desertores, se puede observar en la figura siguiente una estrecha relación de manera inversa entre el ingreso y la deserción. El rango salarial de las familias para el cual hay mayor tasa de deserción es para aquellas que ganan uno o menos de un salario mínimo legal vigente, y a medida que se aumenta el rango salarial de la familia, disminuye la tasa de deserción de estudiantes.

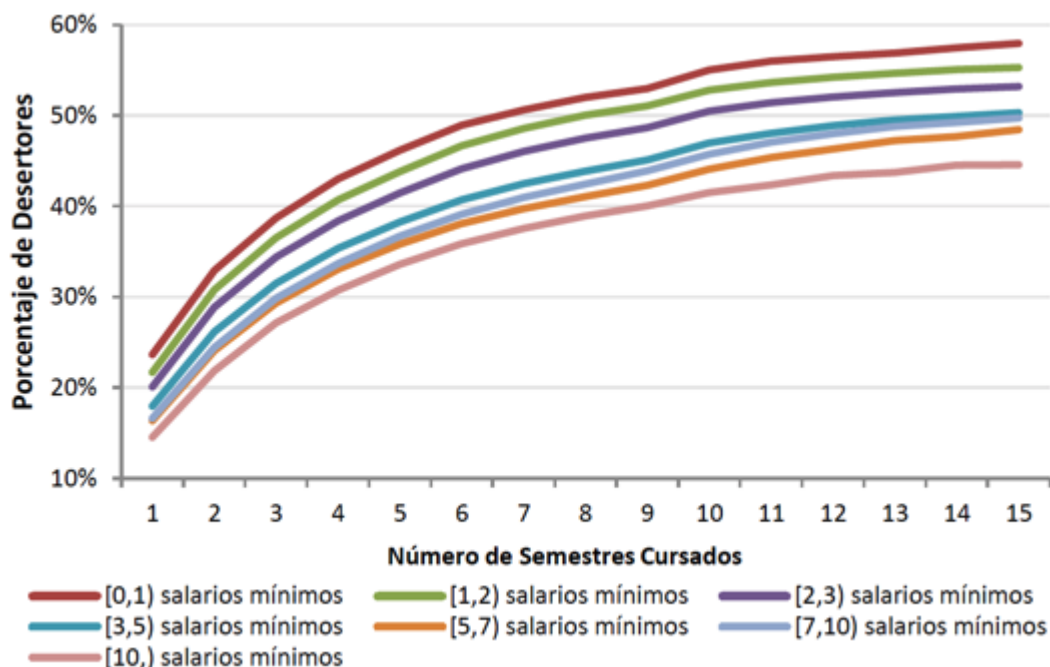


Figura 3. Porcentaje de desertores según clasificación de exámenes entre 1998-2014 (SNIES)

Fuente: (Ministerio de Educación, 2014)

El último dominio o aspecto evaluado por el SNIES hace referencia a las variables institucionales, en ese caso, se toma como base y se evalúa información del estudiante en cuanto a apoyos o ayudas financieras (académicas y otros apoyos), lo que en efecto propende o busca disminuir el riesgo de deserción. Asimismo, el origen (pública o privada) de la IES afecta de manera significativa el riesgo de deserción, debido a que las instituciones técnicas de carácter público aumentan de manera considerable el riesgo de deserción –un comportamiento que también se presenta en las instituciones tecnológicas–, pero en las universidades dicho riesgo se disminuye de manera notable. Este comportamiento se evidencia en la siguiente figura.

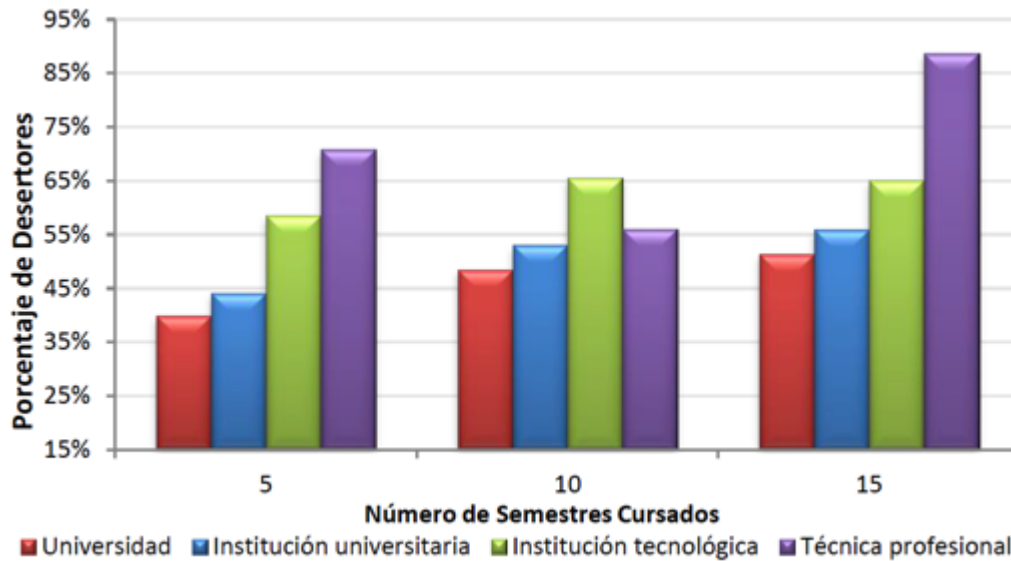


Figura 4. Porcentaje de desertores según clasificación de exámenes entre 1998-2014 (SNIES)

Fuente: (Ministerio de Educación, 2014)

De este modo, se ha logrado determinar que dicho estudio se encuentra enfocado en las IES, sin tener en cuenta la modalidad (presencial, semipresencial, virtual, a distancia) en la cual se imparte el programa de estudios de la universidad. Es por ello que el presente documento tiene como hipótesis analizar la deserción de los estudiantes de educación virtual, para este contexto, seguramente los posibles estudiantes desertores puedan tener un comportamiento totalmente diferente al de quienes se encuentran en programas universitarios de modalidad presencial. Por lo tanto, este proyecto busca responder la siguiente pregunta: ¿Es posible construir un modelo predictivo para deserción estudiantil a partir de una metodología planteada?, y una serie de preguntas específicas:

- ¿Por medio del modelo predictivo, que permite identificar patrones de deserción, se pueden crear estrategias y herramientas de apoyo institucional para que un estudiante no abandone sus estudios?

Las respuestas a estas preguntas ayudarán a demostrar que, con un entendimiento de las capacidades institucionales, el alineamiento en las áreas de negocio de la institución superior conceptual alrededor de la deserción y junto con técnicas de estadísticas y/o minería de datos junto con la analítica de datos, es posible generar modelos predictivos de deserción, capaces de contribuir al mejoramiento de la intervención temprana a nivel institucional con esos estudiantes que le apostaron a mejorar su proyecto de vida.

3 Objetivos, alcance y limitaciones

En este apartado del documento de tesis se define, a través de los objetivos, la finalidad y el alcance de este proyecto de tesis, a fin de identificar de manera clara las tareas que se realizan para llegar al resultado principal, cuyo desarrollo se plasma a lo largo del documento. Además de los objetivos, por medio del alcance se definen los límites de este proyecto, en donde se especifica cuáles son los resultados esperados y qué se entregará para soportarlos.

3.1 Objetivo general

Construir un modelo predictivo de la deserción estudiantil para la modalidad de educación virtual y a distancia.

3.2 Objetivos específicos

- Realizar un diagnóstico por medio de una metodología de análisis de deserción particular en la institución para estudiantes de educación virtual y a distancia.
- Definir las fuentes de información y las variables que inciden en la deserción de estudiantes en la modalidad de educación virtual y a distancia.

- Establecer un modelo predictivo que permita identificar la deserción de estudiantes en la modalidad de educación virtual y a distancia.
- Construir y validar el modelo predictivo con datos reales de una institución de educación superior con modalidad de educación virtual y a distancia.

3.3 Alcance y limitaciones

El presente proyecto de tesis tiene como propósito presentar un enfoque teórico, práctico, y a la vez dinámico, que le permita a las IES (especialmente a las que tienen programas de educación bajo la modalidad virtual y a distancia) contar con una metodología para realizar un autodiagnóstico en el que se pueda predecir con anticipación a aquellos estudiantes que tienen una mayor tendencia de abandono de estudio. Lo anterior con el objeto de que las mismas instituciones puedan analizar los resultados e identificar factores que les ayuden a construir un plan de mejora continua, para aumentar las tasas de retención de los estudiantes.

4 Metodología

Debido a la naturaleza y el alcance del proyecto, el planteamiento metodológico que se desarrolla en el presente estudio se encuentra dividido en dos momentos o fases principalmente, estas son:

Marco de metodológico. El principal objetivo de esta fase es construir un alineamiento conceptual con las IES, en especial con aquellas que están enfocadas en programas de modalidad virtual y a distancia, para entender el enfoque o la estrategia particular del modelo de persistencia estudiantil propuesto por la universidad, y en conjunto definir unos resultados esperados a partir

del diagnóstico realizado. De ahí en adelante la universidad podrá definir estrategias que hagan posible la creación o fortalecimiento de la persistencia o la retención estudiantil.

Caso de estudio. Para este segundo momento, se llevarán a cabo varias actividades, como son: contexto del caso de estudio, en donde se integrará de manera activa la Universidad Minuto de Dios, específicamente el trabajo a realizar con la rectoría de la sede virtual y a distancia. En esta fase se ejecutará el marco de trabajo propuesto en la fase anterior, y se espera tener un entregable con el análisis predictivo que permitirá detectar a aquellos posibles desertores estudiantiles de la universidad.

A continuación, se ilustran las fases que hacen parte del alcance del presente documento y los pasos a seguir en cada una de ellas.



Figura 5. Desarrollo de las fases

Fuente: elaboración propia

5 Marco teórico

5.1 Tinto

En el estudio desarrollado por Tinto (1975) se desarrolló un modelo que explicó la relación entre el estudiante y la universidad, y está basado en el involucramiento de los estudiantes donde las metas y el compromiso organizacional son los valores raíz y son facilitados por los académicos y la integración social modelo Tinto en la participación de los estudiantes donde el objetivo y el compromiso institucional son valores fundamentales y son facilitados por la integración académica y social. En ese sentido, el estudiante ve influenciado, positiva o negativamente, su compromiso académico con su desempeño académico en la institución, y la integración con compañeros y docentes.

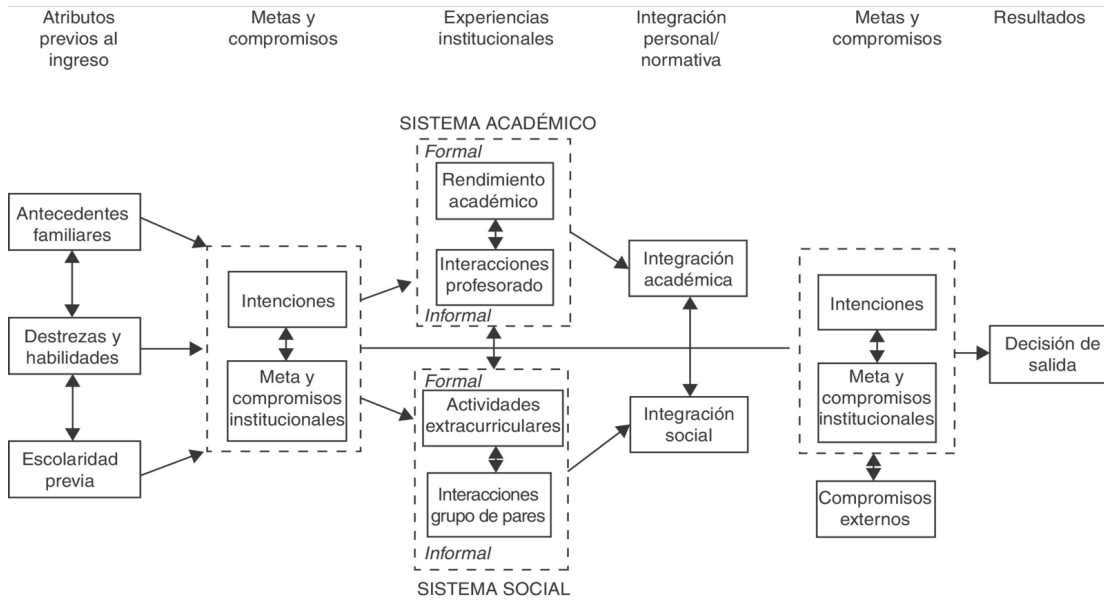


Figura 6. Esquema conceptual de abandono escolar

Fuente: (Tinto, 1975)

Esta figura muestra un modelo longitudinal de relaciones entre los sistemas individuales y académicos; de ahí se desprende que el compromiso de la universidad incide fuertemente en el

estudiante, puede lograr que un estudiante comprometido con su estudio desista, o, al contrario, que uno con bajo nivel de compromiso individual se mantenga estudiando.

Conviene distinguir que los estudiantes ingresan a las universidades con una variedad de características (sexo, raza, capacidad), experiencias previas a la universidad (promedios de calificaciones en la secundaria, logros académicos y sociales) y antecedentes familiares (estatus social, ingresos, cantidad de familiares), y que cada una tiene una incidencia directa e indirecta en su desempeño en la universidad. Más aún, estas características de fondo y los atributos de la persona también influyen en las expectativas y compromisos educativos que esta trae al entorno universitario. En ese escenario, dichos objetivos y compromisos de la universidad son importantes en la predicción y reflexión en torno a la experiencia de la persona, sus decepciones y satisfacciones.

El modelo de deserción planteado toma esta idea como se indica y sugiere que dichas evaluaciones se verán reflejadas en los compromisos del estudiante, con el objetivo de terminar la universidad, y en la institución en la que está matriculado. En suma, estos compromisos reflejan su integración en la parte académica y social en la institución, son el resultado de la percepción de los beneficios (mejores notas, satisfacciones personales, amistades) y los costos (costos de estudiar, tiempo, insatisfacciones, fallas académicas) de su asistencia a la universidad.

De otro lado, el modelo teórico de Tinto toma en cuenta elementos externos que afectan la decisión del estudiante de permanecer en la universidad. Por ejemplo, permite incluir los efectos de la oferta y la demanda en el mercado laboral en las tasas de deserción, y tiene en cuenta que hay restricciones (por ejemplo, la discriminación) que limitan la capacidad de las personas para invertir en alternativas percibidas como potencialmente más gratificantes.

La deserción entonces es el resultado de un proceso que involucra la interacción entre el estudiante y la universidad, y no sorprende que las características de la universidad se relacionen con tasas diferenciales de retiro. Es la institución –sus recursos, instalaciones, arreglos estructurales y la composición de sus miembros– la que limita el desarrollo e integración de los estudiantes en la institución y la que lleva al desarrollo del clima social y académico.

5.2 Spady

El modelo es un resumen y extensión de conceptos relacionados con la teoría del equilibrio, la teoría del suicidio de Durkheim y el estudio sobre las deserciones de la universidad. Este considera la decisión de retirarse de un sistema social particular, como la consecuencia de un proceso complejo a nivel social que incluye antecedentes familiares y educativos, potencial académico, conveniencia normativa, apoyo de amigos, desarrollo intelectual, desempeño, integración social, satisfacción y compromiso con la institución. En este caso, el análisis de regresión múltiple se utiliza para calificar la contribución independiente de cada factor en la explicación de resultados importantes. La integración social, la satisfacción y el compromiso con la institución se pueden explicar sobre la base de recompensas asociadas a las relaciones interpersonales y el desarrollo intelectual. Además, la decisión de deserción escolar a corto plazo está muy influenciada por los criterios de desempeño entre los hombres, pero no es así para las mujeres. Sin embargo, en un periodo de cuatro años, el rendimiento académico es claramente el factor dominante en la deserción tanto de hombres como de mujeres.

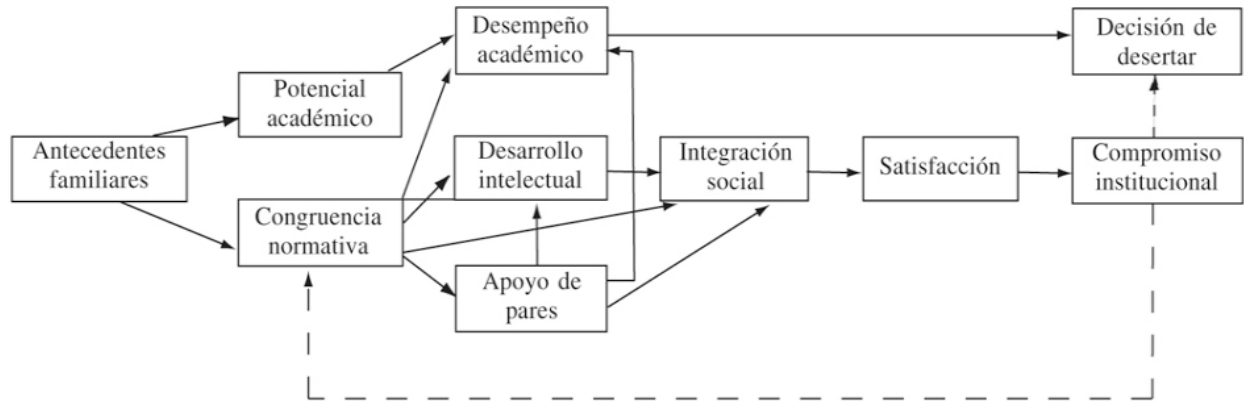


Figura 7. Modelo de Spady. Abandono escolar basado en Durkheim

Fuente: (Spady, 1971)

6 Estado del arte

El enfoque propuesto en este apartado del documento, buscar compartir un marco de referencia donde se busca contextualizar como se ha venido dado la deserción en la educación superior en donde factores individuales, socioeconómico, institucionales y académicos se debe tener en cuenta para el correcto avancen dentro del ciclo de vida del estudiante en la universidad. Pero antes de avanzar es importante socializar de manera concreta algunas definiciones relacionadas a la deserción y junto con ello algunas las técnicas de análisis de datos que han venido evolucionando e implementado en educación superior.

6.1 La deserción estudiantil.

Para comenzar y de acuerdo con Tinto (1989) la definición de la deserción estudiantil puede analizarse desde varias perspectivas y de acuerdo con los diferentes tipos de abandono que se puede dar a lo largo de ciclo académico del estudiante. Estas perspectivas dependen de los actores involucrados en el proceso, como son los estudiantes, el personal administrativo de las

instituciones de educación superior y los responsables de la política institucional en la universidad. En este sentido, el concepto de deserción puede tener diferentes matices como el institucional, el individual, el académico y el económico.

De manera general, la deserción es considerada como el abandono temporal o definitivo del aula de clase por parte de los estudiantes, de los programas académicos para los cuales se matricularon. A partir de esta definición las instituciones han venido trabajando y perfeccionando modelos de deserción que ayudan a estimar la probabilidad de que un alumno se retire en cierto periodo, es allí donde a partir de los datos e información que se recolecta del estudiante desde que es el prospecto o inscrito hasta que se convierte en estudiante toma relevancia ya que es el punto de partida para la definición y construcción de dichos modelos, además se ha logrado evidenciar en los diferentes estudios e investigaciones que los datos más usados para la construcción de dichos modelos, son los siguientes:

- A. Prueba de habilidades a los nuevos estudiantes. En él, el razonamiento verbal verifica las habilidades del estudiante para entender conceptos, además se evalúa la capacidad de abstraer y pensar constructivamente.
- B. Encuesta a los nuevos estudiantes de sus características y opiniones. La encuesta tiene preguntas acerca de las condiciones de vida del alumno, lugar de origen, entorno social y caracterización socioeconómica de su grupo familiar.
- C. Encuestas institucionales. Al terminar cada periodo académico se realiza una encuesta de los servicios académicos y de la satisfacción con los profesores.
- D. Datos de los alumnos. En los sistemas de información que sirven como apoyo a la operación para la recolección de datos del estudiante, normalmente las universidades extraen estos tipos

datos como que se nombran a continuación y lo hacen parte de los modelos donde análisis información de deserción. Estos datos son:

- Individuales
 - Edad al ingreso.
 - Sexo
 - Lugar de origen.
 - Grado de responsabilidad.
 - Autopercepción como alumno.
- Socioeconómicas
 - Estrato social
 - Situación laboral
 - Entorno familiar
- Académicas
 - Resultado de pruebas de ingreso - Razonamiento verbal.
 - Rendimiento académico del periodo académico.
- Institucionales
 - Satisfacción con la universidad.
 - Evaluación de los docentes.
- Contextuales
 - Tasa de crecimiento anual del PIB en el año anterior.

Para Colombia, la entidad encargada de analizar las cifras de deserción en educación superior es el Ministerio de Educación Nacional, para ello la entidad creó una plataforma tecnológica llamada SPADIES la cual se utiliza para hacer seguimiento sobre las cifras de deserción de estudiantes de la educación superior en sus diferentes modalidades. Este seguimiento se realiza analizando los datos suministrados por las instituciones de educación superior, se identifican y se ponderan los comportamientos, las causas y riesgos determinantes para desertar. A partir de este

análisis de información que realiza el Ministerio de Educación, el fenómeno de la deserción se presenta por tres aspectos fundamentales, el rendimiento académico, el cual hace referencia al conocimiento adquirido por los estudiantes durante la educación básica y media; la capacidad económica; y la orientación vocacional de los estudiantes.

6.2 Predicción del rendimiento académico mediante minería de datos

La minería y el análisis de datos es un conjunto de técnicas y procedimientos que puede ser desarrolladas a partir de diferentes fuentes o repositorios de datos, en donde estos datos, sirve para realizar análisis predictivos mediante técnicas de análisis estadístico para predecir medidas de certeza que se basa en datos o hechos reales en la institución.

En diferentes estudios que realizan las universidades donde se utilizaron técnicas de minería de datos (Yamao, 2018), las cuales tienen como meta predecir el rendimiento académico de los alumnos; las fases del desarrollo del proyecto fueron las siguientes:

Recolección de datos: se usó información de las bases de datos de las diferentes áreas de negocio de la universidad como admisión, gestión académica, bienestar universitario entre otras. Algunos datos se describen a continuación.

- Semestre de ingreso.
- Código del alumno.
- Apellido paterno
- Apellido materno
- Nombres
- Fecha de nacimiento
- Sexo

- Modalidad de ingreso.
- Nombre del colegio de procedencia.
- Departamento, provincia, distrito del colegio.
- Tipo de colegio.
- Puntaje obtenido en el examen de admisión.
- Estado de matrícula.
- Escala de pensión.
- Dirección de domicilio.
- Cursos llevados
- Sección
- Nota

Técnicas aplicadas: se aplicaron diferentes métodos de minería de datos definiendo como atributo objetivo una variable definida como Aprobado, entre las técnicas usadas están las referidas a continuación.

Regresión lineal: para la selección de las variables que tienen mayor influencia en la variable dependiente se utilizó el método de regresión paso a paso. Así, las variables que tienen mayor relación con la condición de aprobado son: nota de examen de admisión, edad, sexo y distancia. No se logró demostrar que las variables relacionadas con el colegio de procedencia tengan alguna influencia sobre la posibilidad de ingreso a la universidad del estudiante.

De igual modo, para realizar la predicción con este modelo, se dividieron los datos en 75% para entrenamiento y 25% para probar. En las pruebas de validación de los resultados se ve que hay mucha variación en ellos según la selección aleatoria que se hizo cuando se dividieron los datos.

Árbol de decisiones: normalmente se hace uso del algoritmo C5.0 en la creación de los modelos de árbol de decisiones; estos permitieron establecer reglas para realizar la predicción y analizar la importancia de las variables utilizadas. Las variables de mayor uso fueron el examen de admisión, el sexo y el colegio de origen, especialmente si proviene de colegio privado.

Support Vector Machines: se usó con el fin de crear modelos, basados en vector de soporte, para la creación de hiperplanos que corresponden a un punto de corte en pro de realizar la clasificación. Por otra parte, en el análisis de relación de las categorías social, económica y académica versus el rendimiento académico, se encontró que el carácter social, como razón de retiro, tuvo mayor número de estudiantes.

Cabe añadir que las variables que tienen alta relación concuerdan con otros estudios, donde el rendimiento académico de la secundaria es el que tiene mayor importancia en el rendimiento académico posterior.

Es importante mencionar que las IES han logrado proporcionar elementos con capacidad de evaluar las debilidades y fortalezas que ayudan a disminuir la deserción de los estudiantes, en los diferentes programas académicos. Es por ello que uno de los puntos a evaluar en el presente documento de tesis se basa en el análisis de técnicas de minería de datos que usan diferentes tipos de modelos predictivos, cuya finalidad es identificar la mejor técnica para predecir la probabilidad de que un estudiante deserte del programa académico.

En la siguiente tabla se recolectan y se consolidan varios estudios realizados al respecto, mencionando las técnicas utilizadas.

Tabla 1. *Técnicas de minería de datos*

Estudio o paper relacionado	Técnica de minería de datos aplicada
New Directions in Education Research: Using Data Mining	Regresión logística

Techniques to Explore Predictors of Grade Retention	Árboles de clasificación
Predicting Students' Academic Performances – A Learning Analytics Approach using Multiple Linear Regression	Regresión lineal
Modelo para la automatización del proceso de determinación de riesgos de deserción en estudiantes universitarios	Árboles de decisión Clusters (K-medianas)
Mining Educational Data to Analyze Students Performance	Árboles de decisión
Determining factors affecting student retention in a higher education institute in Taiwan and building a prediction model using logistic regression and support vector machine	Regresión Logit Support Vector Machine
Student Performance Evaluation in Education Sector using Prediction and Clustering Algorithms	Clustering
A Model to Predict Ohio University Student Retention From Admissions and Involvement Data	Regresión lineal Regresión logística Árboles de decisión
Predicción del rendimiento académico de alumnos de primer año de la FACENA (UNNE) en función de su caracterización socioeducativa	Regresión logística
Data Mining in the Classroom: Discovering Groups' Strategies at a Multi-tabletop Environment.	Differential Sequence Mining
Una lectura sobre deserción universitaria en estudiantes de pregrado desde la perspectiva de la minería de datos	Árboles de decisión C 4.5

Fuente: elaboración propia

De esta interrelación de los modelos mencionados por diferentes exponentes y autores, se presenta en la siguiente figura un resumen de los criterios que afectan de manera directa la permanencia de los estudiantes en un programa académico, en particular en la modalidad de educación virtual y a distancia:



Figura 8. Factores comunes en la deserción

Fuente: elaboración propia

Finalmente, la identificación de criterios o variables en cada uno de los estudios y métodos propuestos, a lo largo de la historia, ha permitido consolidar los factores y dominios que son determinantes para la deserción (MEN, 2009). Estos elementos descritos, junto con los dominios que afectan directamente la deserción, se convierten en el marco de referencia para el análisis y el entendimiento de la altísima complejidad del fenómeno de la deserción. Además, le permiten a las IES contar con una visión holística de las causas, y así identificar posibles acciones que se pueden promover, desde la institución, para lograr que los estudiantes culminen con éxito su programa académico en la modalidad virtual y a distancia.

7 Marco metodológico y caso de estudio

Como parte de la contextualización general para el desarrollo del planteamiento metodológico, el cual fue presentado en el apartado de la metodología, esta sección tiene como objeto principal definir lineamientos y artefactos, a fin de fortalecer las capacidades institucionales en la ejecución de un modelo que haga posible evaluar la deserción en los estudiantes de la modalidad de educación virtual y a distancia.

7.1 Fase No.1 - Marco metodológico

Para las IES es fundamental contar con elementos o herramientas que les permitan realizar una autoevaluación o un autodiagnóstico. Este programa o iniciativa debe hacerse periódicamente, pueden revisar factores internos y externos referentes a la deserción estudiantil en la modalidad de educación virtual y a distancia. Para ello, y con el fin de realizar en este documento una guía práctica y dinámica, se propone como punto de partida un ejercicio de autoevaluación con la institución, en el que se haga uso de elementos utilizados en la arquitectura empresarial para Educación Superior (Santiago, 2013). En la figura siguiente se mencionan los pasos a realizar en la Fase No.1, junto con los artefactos que fueron creados en cada uno de los pasos.

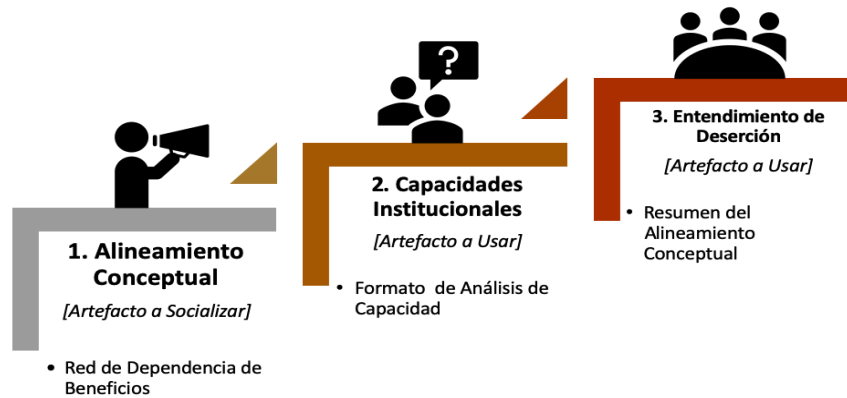


Figura 9. Proceso y artefactos usados en el marco metodológico

Fuente: elaboración propia

7.1.1 Alineamiento conceptual

El alineamiento conceptual establece lineamientos con el fin de que las universidades puedan crear y establecer un proyecto de permanencia, el cual les ayudará a evaluar la capacidad institucional para gestionar la permanencia estudiantil y dar cumplimiento a la reducción o prevención de las políticas institucionales de deserción estudiantil. Estos lineamientos tienen las siguientes características:

- Un propósito: el programa de autoevaluación debe estar centrado en el estudiante y ser diseñado para facilitar la persistencia estudiantil, con el objeto de reducir la tasa de deserción.
- Proactivo: el programa de autoevaluación requiere acciones tempranas y preventivas que sean definidas por la institución para abordar necesidades educativas y los problemas de los estudiantes de manera anticipatoria, antes de que se conviertan en conflictos que requieran intervención reactante.
- Colaborativo: el programa de Autoevaluación debe ser iniciado por la institución y luego llegar a los estudiantes para brindar el apoyo requerido.

Por lo anterior, y con el propósito de apoyar la formulación de esta hipótesis, se usan elementos o artefactos provistos por la arquitectura empresarial, puesto que esta disciplina de planeación o de dirección establece una hoja de ruta más exacta y precisa frente a los elementos más relevantes dentro de los procesos de negocio. En este caso, dichos elementos se enfocan en las IES para el proceso de persistencia estudiantil.

Para este primer alineamiento conceptual de la fase 1 del marco metodológico, se usará como referencia un diagrama de Dependencia de Beneficios (BDNg) (Peppard, 2016), que se utiliza para socializar junto con la institución las relaciones causa-efecto organizadas en capacidades, a fin de obtener cambios que se requieren para los beneficios y resultados esperados.

Adicionalmente, lo que busca este artefacto es ilustrar toda la alineación que se necesita obtener, a través de la combinación de la tecnología, procesos de negocio e iniciativas institucionales que buscan reducir la deserción estudiantil, las cuales se relacionan entre sí a fin de tener un mapa de cambios que puedan ser adoptados por la institución. Ese insumo o modelo le da la posibilidad a la IES de una comprensión cada vez más integral de la problemática de deserción estudiantil que puede estar presente en los estudiantes de la modalidad virtual y a distancia.

Red de Dependencia de Beneficios

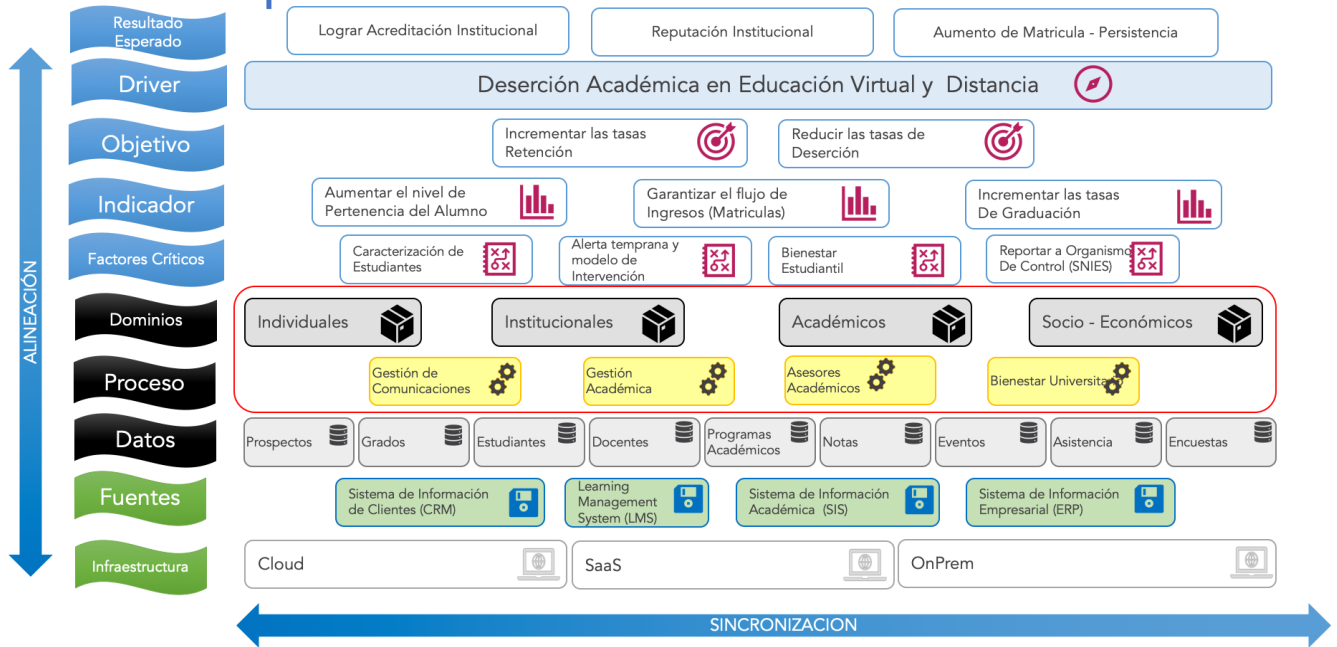


Figura 10. Red de dependencia de beneficios - Deserción estudiantil

Fuente: elaboración propia

Una vez finalizada la socialización y el alcance con las unidades de negocio que intervienen en los procesos académicos que hacen parte de la metodología, se continúa con el siguiente paso a nivel más funcional y operativo.

7.1.2 Análisis de capacidades institucionales

A partir de la socialización de la Red de dependencia de beneficios de deserción académica en Educación Virtual y a Distancia, hecha al equipo de trabajo de las áreas de la institución que se van a intervenir, se definió el artefacto que permite realizar un diagnóstico sobre los procesos académicos más relevantes dentro de la institución, y en los cuales se concentra mayor uso de la información académica. Lo anterior se hizo con el propósito de observar cuáles son los elementos que se pueden articular con el marco de trabajo, y que apoyen la identificación y evaluación de las entidades de negocio y de las variables que pueden ser utilizadas y asociadas a

la deserción y permanencia estudiantil. Paralelamente, tales elementos contribuyen al reconocimiento de las acciones que están llevando a cabo las IES a fin de mitigar las problemáticas alrededor del seguimiento en los procesos.

En la Figura 13 se encuentra una breve descripción de los campos que se deben completar por parte de la IES.

NOMBRE DE LA CAPACIDAD				< Capacidad a ser Analizada >			
DESCRIPCIÓN							
<i>Descripción del proceso de negocio a ser intervenido – Solamente aplica (Gestión de Comunicaciones – Gestión Académicas – Asesores Académicos – Gestión de Bienestar Universitario)</i>							
PROBLEMÁTICA				OPORTUNIDADES			
<i>Listar las problemáticas identificadas en el proceso de negocio</i>				<i>Listar las oportunidades de mejoras identificadas en el proceso de negocio</i>			
SUB-PROCESOS		INFORMACIÓN		PERSONAS (ROLES)		TECNOLOGÍA	
<i>Listar Sub-procesos identificados</i>		<i>Entidades Basicas que se cubren en el proceso de negocio</i>		<i>Listar roles o personas que intervienen en el proceso</i>		<i>Listar sistemas de información que apoyan los procesos</i>	
INICIATIVAS							
<i>Listar proyectos a realizar</i>							
RESULTADOS DEL AREA / INDICADORES							
<i>Listar las problemáticas identificadas en el proceso de negocio</i>							

Figura 11. Artefacto de entendimiento de capacidades

Fuente: elaboración propia

7.1.3 Formato de entendimiento de deserción

Una vez finalizado el proceso de entendimiento correspondiente al análisis de capacidades institucionales en los procesos definidos, se procede a elaborar y validar con la IES el entendimiento que esta tiene sobre deserción estudiantil. Con ese objeto, se ha elaborado este artefacto resumen donde se relacionan las políticas, las estrategias, las acciones e incluso instrumentos que ayudan al fortalecimiento y al alcance que se le debe dar al modelo propuesto.

En ese sentido, se busca con dicho artefacto tener un plan unificado, integrado y coherente al Plan Estratégico Institucional a fin de que se pueda asegurar que los objetivos básicos propuestos se logren.

En la Figura 14 se encuentra una breve descripción de los campos que se deben completar por parte de la IES.

Resumen – Alineamiento Conceptual









Definición  <i>Definición concreta, simple y resumida (¿qué entiende la Universidad por Deserción? ¿Por qué le podría interesar ?)</i>		Stakeholder (Área de Negocio)  <i>Específicamente a quién le interesa / duele dentro de la universidad ? (rol)</i>	
Problemática Identificada  <i>¿Cuáles son los dolores (pains), riesgos, desafíos típicos que afronta la universidad en temas de Deserción? ¿Qué les preocupa en la Graduación y Permanencia? ¿Qué pasa si no lo resuelven? ¿Cuánto les cuesta la deserción en la modalidad de educación virtual y a distancia?</i>		Beneficios Esperados  <i>¿Cuáles son los beneficios típicos que espera la universidad? ¿Cómo se sentirían tranquilos? ¿Cómo se podrían medir la retención?</i>	
Población Objetivo  <i>¿Cuál es el grupo, programa o facultad objetivo para el análisis de la deserción? ¿Cuales son las características de la población en términos Socioeconómicos y Académicos? Es fácil identificar atributos potenciales de análisis</i>		Beneficios Esperado  <i>¿Qué beneficios específicos se obtiene con el uso de la metodología propuesta?</i>	
Factores Críticos de Éxito  <i>Lista de factores que permitan el éxito de la metodología propuesta</i>		Factores Críticos de Riesgo  <i>Lista de factores que ponen en riesgo la metodología propuesta</i>	

Figura 12. Artefacto resumen de alineamiento conceptual de deserción estudiantil

Fuente: elaboración propia

7.1.4 Marco de trabajo

Cuando la institución de educación superior toma la decisión de apropiar la metodología propuesta, es necesario que tengan en cuenta aspectos mínimos que aseguren el éxito de su ejecución. Dicho esto se presentan los beneficios que fortalecen el marco metodológico propuesto:

- A. Garantizar el cumplimiento de las actividades propuestas.
- B. Aumentar los niveles de retención (Persistencia).

- C. Garantizar el cumplimiento de las políticas institucionales.
- D. Aumentar el nivel de pertenencia del alumno.
- E. Minimizar el riesgo de pérdida de acreditación de los programas de la modalidad de educación virtual y a distancia.
- F. Alto nivel de pertinencia en el mercado laboral.

Ahora bien, para lograr realizar el marco de trabajo, es preciso apropiarse de los elementos de la metodología CRISP-DM (Cross Industry Standard Process for Data Mining), puesto que es un modelo de proceso de minería de datos con una guía estructurada en seis pasos; estos pueden ser bidireccionales, es decir que de un paso en concreto se puede volver a un paso anterior para poder revisar. En la Figura 15 se describen los pasos o fases que propone la metodología CRISP-DM.

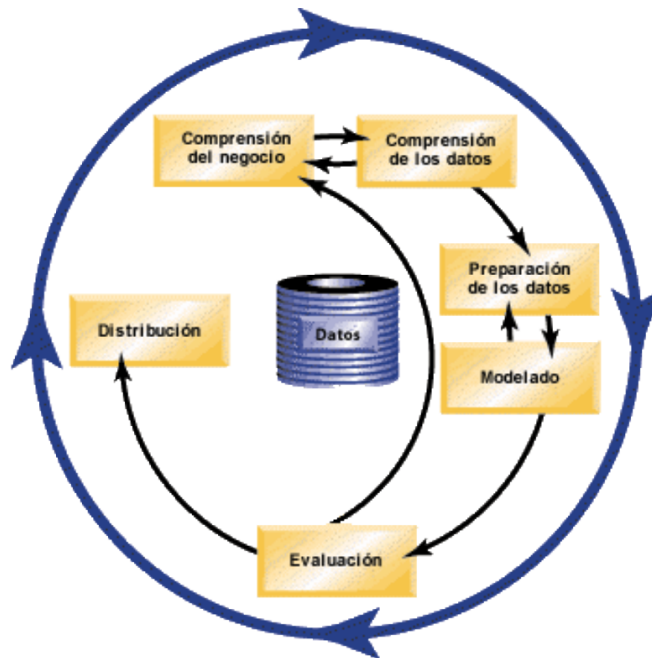


Figura 13. Ciclo de vida de metodología CRISP-DM

Sin embargo, y con el entendimiento o acuerdos que se logran previamente con la IES en el alineamiento conceptual, varios pasos o fases que son propuestas en la metodología CRISP-DM

son limitadas en el marco de trabajo propuesto en este documento. Por esa razón, se han fijado cuatro pasos que orientan a la institución en el entendimiento de los datos, preparación de los datos, aplicación de modelamiento y, por último, una evaluación del modelo predictivo de deserción estudiantil para estudiantes de modalidad virtual y a distancia.

Dado lo anterior, a continuación, se propone el marco de trabajo por medio del cual se lleva a cabo el análisis de cada uno de los cuatro pasos allí descritos.

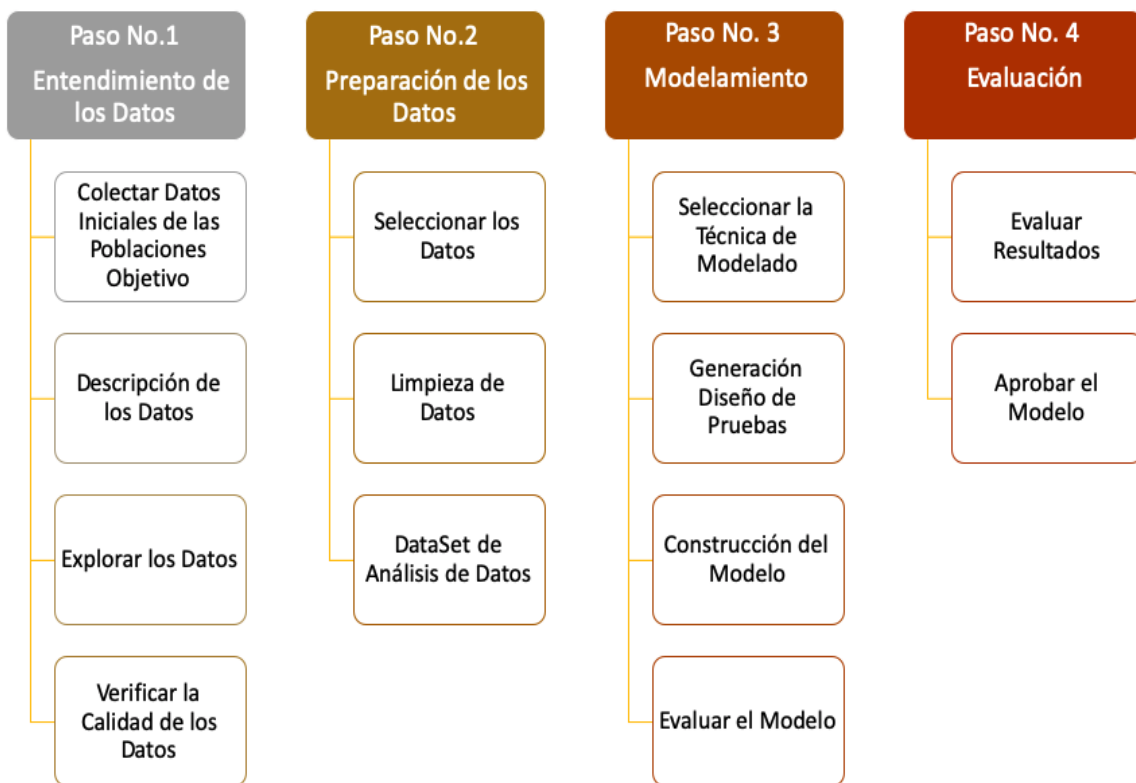


Figura 14. Marco de trabajo propuesto

Fuente: elaboración propia

Entendimiento de los datos: el principal objetivo de establecer un primer contacto con el problema y con los datos mínimos viables, determinados para construir el modelo de predicción

de deserción de estudiantes, es identificar la calidad de los datos, y además apoyar la creación de relaciones que permitan definir las primeras hipótesis a trabajar en el modelo predictivo. Por consiguiente, las actividades a realizar se especifican en los siguientes ítems.

- a) Colección de datos iniciales de las poblaciones objetivo: en esta primera actividad se elaboran informes por medio de listas de los datos que son proporcionados por la IES, en donde se menciona el dato, la fuente o sistema de información en que este reside, y la técnica de recolección.
- b) Descripción de los datos: en esta segunda actividad se detallan los datos, volúmenes de registros y campos por registro, así como su significado y la descripción del formato.
- c) Exploración de datos: en este paso se procede a encontrar una estructura general de los datos, por medio de pruebas estadísticas (tablas de frecuencia y gráficos de distribución), para posteriormente elaborar un informe de exploración de datos.
- d) Verificar la calidad de datos: se efectúan verificaciones sobre los datos para determinar o identificar valores nulos o datos de valores de rango que pueden generar ruido en el desarrollo de modelo.

Preparación de los datos: una vez finalizado el proceso de recolección de los datos, es necesario prepararlos a fin de aplicar técnicas de minería de datos; a su vez, con el apoyo de herramientas de visualización se podría realizar búsqueda de relaciones entre las variables u otras medidas interesantes que pueden ser aplicadas al modelo. Aquí se incluyen actividades generales como las descritas a continuación.

- a) Seleccionar los datos: en esta actividad se determina y se selecciona el conjunto de datos o un subconjunto de datos que fueron definidos en actividades anteriores.

- b) Limpieza de datos: esta actividad lo que pretende es identificar la normalización de los datos, así como el tratamiento de valores que puedan faltar y que pueden ser esenciales al modelo, de hecho, esta es una actividad complementaria de la actividad de selección de datos.
- c) DataSet de análisis de datos: esta actividad implica la creación y definición de los campos que se incluirán en la tabla resumen. Esta tabla se crea a parte de los múltiples registros fuentes y campos que son trabajados en la fase inicial de entendimiento de los datos. Asimismo, es posible que se den ajustes a los campos y al formato, para facilitar el proceso de modelamiento en la herramienta tecnológica que apoya el proceso de modelamiento predictivo.

Modelamiento: en la fase de modelamiento, según la metodología CRISP-DM, se deben seleccionar las técnicas más apropiadas para elaborar el modelo predictivo. Normalmente las técnicas de modelado se eligen en función de diferentes criterios, pero es sumamente importante determinar un método de evaluación de modelos que permita garantizar el modelo que más sea adecuado, con el objetivo de reconocer los estudiantes que puedan tener mayor probabilidad de desertar de la universidad. En este caso se incluyen cuatro actividades.

- a) Seleccionar técnica de modelado: se selecciona la técnica de minería más apropiada, coherente con el problema que se quiere resolver, y que sobre todo radique en detectar a aquellos estudiantes más propensos a ser desertores. Para esta selección es posible evaluar modelos de clasificación o modelos de árboles de decisión, y se deben incluir técnicas de visualización de datos.
- b) Generación del diseño de pruebas: en esta actividad es necesario verificar la validez y contundencia del modelo usado. Por ello, al utilizar algoritmos supervisados, como la clasificación, es común separar los datos en dos subconjuntos, uno de estos es usado para

entrenamiento y el otro para pruebas. Posteriormente se construye un modelo basado en entrenamiento, de esta manera es factible medir el nivel de calidad del modelo.

- c) **Construcción del modelo:** se ejecuta la técnica seleccionada en las actividades anteriores, y se preparan los datos para generar uno o más modelos, donde cada una de las técnicas de modelado tienen características del modelo a generar, el cual debe ser aplicado al problema que se desea resolver.
- d) **Evaluación de modelo:** allí se interpretan los modelos de acuerdo con la problemática de deserción de estudiantes que se quiere solucionar, y en donde los factores críticos de éxito tienen un peso significativo.

Evaluación: en esta fase se evalúa el modelo, teniendo en cuenta el objetivo y el cumplimiento de los factores críticos de éxito. Igualmente, se debe considerar la confiabilidad de que el modelo solamente aplica para el conjunto de datos en el cual se realizó el análisis. Por último, es conveniente revisar el proceso, a partir de los resultados obtenidos, para identificar posibles errores. Las actividades involucradas en esta última fase se detallan en los ítems que siguen.

- a) **Evaluar resultados:** en esta actividad de evaluación del modelo se estudia la relación de los objetivos propuestos en este documento de tesis, en búsqueda de determinar la eficiencia del modelo, o si por el contrario es oportuno replantearlo si el tiempo y las restricciones que tenga la IES lo permiten.
- b) **A probar el modelo:** si se ha determinado que las actividades hasta este momento han generado los resultados esperados, la IES estará en libertad de elaborar un plan de acción que mitigue la deserción estudiantil de la modalidad de educación virtual y distancia, dirigido a la población que fue objeto de estudio.

Finalmente, cuando las IES decidan poner en marcha el modelo propuesto, es necesario que tengan en cuenta aspectos mínimos para asegurar el éxito en su ejecución, y que permitan conseguir un diagnóstico más detallado de los datos que posee la institución, por cada uno de los dominios a revisar. De este modo se obtendría una línea base de las variables a trabajar en el modelo predictivo. Por lo anterior, se ha creado un artefacto que apoya todo el proceso de entendimiento, preparación y modelamiento de los datos que hacen parte del marco de trabajo.

El formato de apoyo se ha llamado Preparación de datos en formato Excel, y tiene varias hojas anexas para el diligenciamiento por parte de la institución. A continuación, se describe cada una de las hojas a completar en el artefacto.

Tabla 2. *Descripción de Hojas de Artefactos*

Hoja del artefacto	Descripción y propósito
00 - Informativo	Es una plantilla control en donde se indica cuál es la versión del documento y se detallan los dominios que hacen parte de la deserción.
01 - Dashboard	Muestra gráficos de poblamiento de los datos respecto a los dominios, también indicadores resumidos de calidad y de completitud de los datos.
02 - Definición de deserción	Ficha que ayuda a la institución a definir la deserción para su caso particular; tiene una lista de opciones para marcar las que sean pertinentes.
03 - Sistemas de origen	En las columnas de esta matriz primero se agrupa por dominio, luego se ubican allí las variables definidas para el modelo, y por último el sistema de origen donde se encuentra la información. Es un inventario y permite hacer fácilmente el requerimiento al área de tecnología.
04 - Variables y tipología	Muestra la naturaleza de la variable, deja también marcar las que deben ser convertidas en los procesos de limpieza de los datos.
05 - Listas de valores	Listado de los valores recomendados por la metodología en cada campo.
06 – Indicadores	Esta hoja calcula indicadores de error y de completitud, de

	manera individual, por cada variable planteada en el modelo.
07 - Análisis univariado	Presenta el formato estándar para evaluar el poblamiento del atributo a validar, tiene en primer lugar una tabla de frecuencias, adicionalmente, como apoyo, un gráfico de barras agrupadas para mejor visualización, y presenta una vista inicial de la proporción de los datos en el total del poblamiento, mediante una barra 100% apilada.
08 - Análisis bivariado	Muestra la relación entre dos variables, cómo se distribuyen los datos; cada cruce tiene dos cuadros, el primer cuadro es la frecuencia de las dos cruces, el segundo es una tabla de frecuencia cruzada, donde la primera fila y columna tienen una cantidad de registros cruzados.
09 - Diccionario de datos	Se genera una hoja de diccionario de datos basada en la definición de campos de la metodología.

Fuente: elaboración propia

7.2 Fase No. 2 - Caso de estudio

Después de haber explicado la metodología propuesta en las IES para interpretar el problema de la deserción de estudiantes, en la modalidad de educación virtual y a distancia, es necesario examinar los diferentes actores educativos en quienes se puedan promover espacios para el análisis, el acuerdo sobre planes de acción, y la ejecución de dichos planes dirigidos a la permanencia y graduación de los estudiantes de educación superior.

Es por ello que la segunda fase está totalmente enfocada en la ejecución del caso de estudio, donde se integra, de modo totalmente activo, la IES que tiene una alta concentración de estudiantes bajo la modalidad virtual y a distancia.

7.2.1 Institución Educativa: Universidad Minuto de Dios (Rectoría de Educación Virtual y a Distancia)

La Universitaria Minuto de Dios (Uniminuto) se constituyó en 1990 y empezó a operar en 1992, en la localidad de Engativá (occidente de Bogotá), con el fin de facilitarle el acceso a la educación superior de calidad a una población que de otro modo no tendría posibilidades de educarse.

Desde sus inicios, orientó sus esfuerzos hacia la creación de un proyecto educativo innovador que ofreciera educación de calidad, integral, flexible, de fácil acceso, y que formara ciudadanos íntegros y con grandes calidades espirituales, profesionales competentes, éticamente orientados y con motivación hacia la transformación social, quienes estarán en pro de la construcción de un país justo, reconciliado y en paz. La Uniminuto se ha consolidado como una de las universidades con mayor número de estudiantes a nivel nacional en el territorio colombiano, actualmente cuentan con 10 rectorías y con más de 73 sedes.

Ahora bien, es clave aludir a la visión estratégica de la universidad, puesto que uno de los ejes de la sostenibilidad financiera y del impacto misional de Uniminuto son los procesos de la unidad de Educación Virtual y a Distancia, donde tienen como retos fortalecer y profundizar un modelo de educación virtual que apoye directrices clave de acceso, cobertura y calidad. Estos elementos se condensan y se explican en la Figura 17.

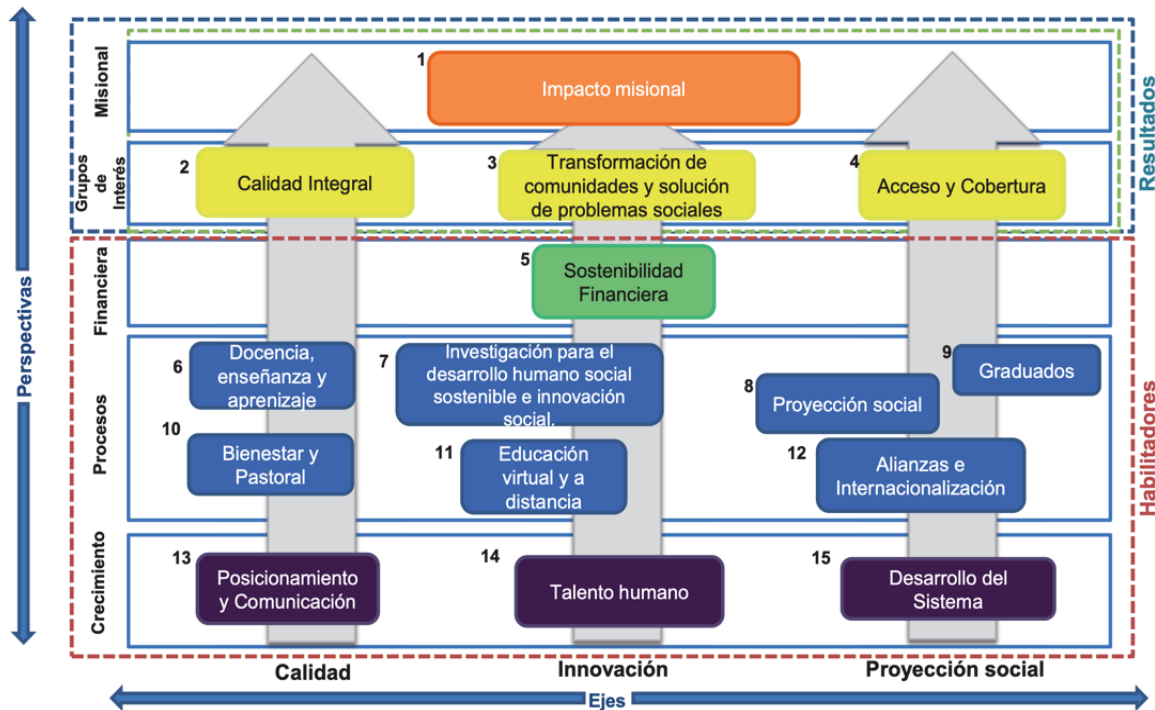


Figura 15. Mapa estratégico - Uniminuto

Fuente: (UNIMINUTO, Plan de desarrollo 2013-2019, 2014)

Por otra parte, un dato de igual relevancia es que el sistema de la Uniminuto actualmente cuenta con una población estudiantil que supera los 120 000 estudiantes en sus diferentes niveles de formación, como lo son:

- Pregrado (presencial, virtual y a distancia)
- Postgrado (presencial, virtual y a distancia)
- Educación continua
- Articulación con la educación media.

Si se revisa en detalle la participación de los estudiantes en programas de metodología a distancia (tradicional y virtual), esta fue de 71.2%, mientras que a la metodología presencial le correspondió un 28.8% de la población estudiantil, tal y como se muestra en la Figura 18. Cabe resaltar que, desde el segundo semestre académico del año 2013, se observa una tendencia en el

aumento de los estudiantes que optan por estudiar en la modalidad de educación virtual y a distancia.

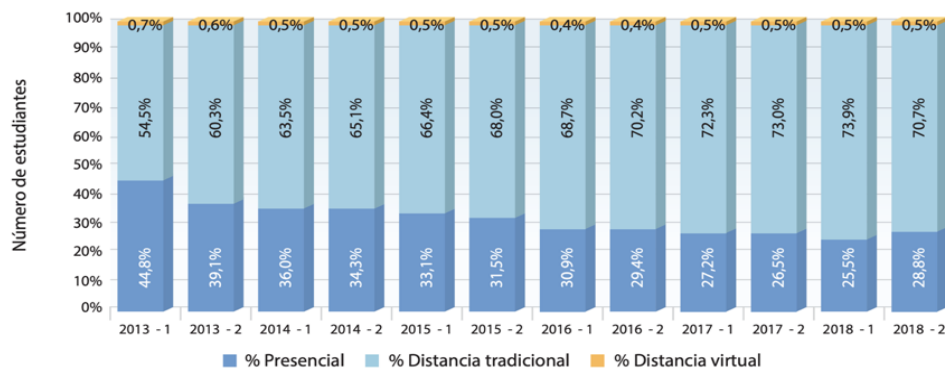


Figura 16. Participación de los estudiantes por modalidad de estudios

Fuente: (UNIMINUTO, 2018)

Por su parte, el Instituto de Educación Virtual y a Distancia (IEVD) inicia en el 2008, amparado con el proyecto de Fortalecimiento de la Educación Técnica y Tecnológica del MEN, cuya finalidad es diseñar programas por competencias y en ciclos propedéuticos en los niveles técnico y tecnológico, en alianzas estratégicas con los sectores productivos. A la par, se ha venido consolidando el modelo de articulación educativa con la educación media técnica, estrategia que fomenta la integración de la educación media con la educación superior.

En este escenario, la Rectoría Uniminuto Virtual y a Distancia (UVD), como unidad responsable del desarrollo y de la promoción y prestación de servicios de educación superior, ofrece como métodos de enseñanza-aprendizaje-evaluación el desarrollo de habilidades y actitudes de aprendizaje autónomo, con acompañamiento virtual. Con ello se ha logrado que la Uniminuto hoy esté presente en 20 departamentos y 79 municipios, a través de centros de tutoría en las diferentes sedes, convenios y Ceres (Centros Regionales de Educación Superior) administrados por el sistema.

En ese sentido, es justo destacar el número de estudiantes con los que cuenta la Rectoría UVD, 11 289 de los 83 780 estudiantes, se forman en la metodología a distancia en Uniminuto, al igual que la proporción de estudiantes por género, cuya relación es de tres mujeres por un hombre, y que el 45% de la población estudiantil se encuentra en el rango de los 27 a 36 años.

7.2.2 Alineamiento conceptual (Rectoría de Educación Virtual y a Distancia)

Para empezar, un dato clave es que el nivel de deserción que se presenta en la Rectoría UVD, el cual se ubica en el 8.3%, frente al Promedio Nacional Colombia del SPADIES, ubicado para el 2016-1 en 12.36%, implica una cifra que es baja, y con ello se resaltan las estrategias de Bienestar Universitario para la retención de estudiantes, sobre todo en cuanto a acompañamiento psicológico, orientación vocacional, acompañamiento psicosocial y apoyo socioeconómico.

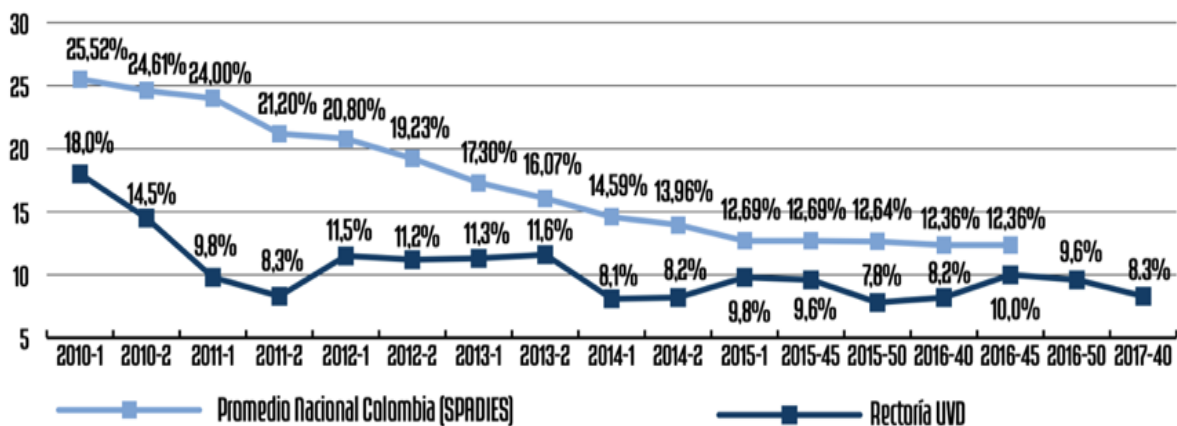


Figura 17. Cifras de deserción de Rectoría IEVD - Uniminuto

Fuente: (UNIMINUTO, UVD en cifras , 2017)

A continuación, se presenta el análisis de capacidades que hace parte de la metodología propuesta, pero antes, es importante mencionar que un modelo de capacidad describe el conjunto completo de capacidades que la institución necesita para ejecutar su modelo de negocio y desarrollar su planeación estratégica. Con esta claridad, se ha diligenciado al artefacto sugerido

para las cuatro unidades de negocio que hacen parte de la Red de Dependencia de Beneficios.

Estos procesos de negocio son:

NOMBRE DE LA CAPACIDAD		Gestión de Comunicaciones	
DESCRIPCIÓN			
Desarrollar estrategias e iniciativas de comunicaciones para los grupos de interés de la universidad para toda la comunidad académica, haciendo uso de los diferentes canales de comunicación a fin de fortalecer la imagen institucional de la Rectoría de Educación Virtual y A Distancia			
PROBLEMÁTICA		OPORTUNIDADES	
<ul style="list-style-type: none"> No se tiene definido un plan de Comunicación y de acción para los estudiantes que son ausentes. No existen Campañas de Comunicación enfocados a la gestión que realizan los tutores a los estudiantes. No se cuenta con un plan de comunicaciones para gestionar asesorías del estudiante Alta dependencia de TI para construcción de campañas específicas de comunicación a grupo o poblaciones objetivos muy específicos. 		<ul style="list-style-type: none"> Definir un plan de comunicación que cubra desde los procesos de Admisión, hasta los procesos de Graduación. Gestionar un plan de comunicación de acuerdo con lo establecido en cada una de las áreas de negocio de la institución. Definir métricas y flujos de trabajo que gestione alertas tempranas para aquellos estudiantes que presenten inasistencia a clase virtual o tutoría presencial. Diseñar y ejecutar planes de comunicaciones dinámicos, personalizados, oportunos y multicanal de acuerdo a los procesos de la gestión académica 	
SUB-PROCESOS	INFORMACIÓN	PERSONAS (ROLES)	TECNOLOGÍA
<ul style="list-style-type: none"> Definir Plan de Comunicación a las áreas académicas y de apoyo Ejecutar el Plan de Acción de Comunicaciones a las áreas académicas y de apoyo 	<ul style="list-style-type: none"> Admitidos Inscritos Graduados Docentes Finanzas del Estudiante Publicaciones (Material Audiovisual) Publicidad Impresa 	<ul style="list-style-type: none"> Cordinador de Comunicaciones Auxiliar de Marketing Auxiliar de CRM para Admisiones e Inscripciones 	<ul style="list-style-type: none"> CRM SIS ERP MAIL SERVICES
INICIATIVAS			
<ul style="list-style-type: none"> Identificar necesidades para fortalecer procesos de comunicaciones en todos los procesos que hacen parte del Ciclo de Vida del Estudiante de Educación Virtual y a Distancia. Definir flujos de trabajo automáticos para envío de comunicaciones a la comunidad académica Definición y monitoreo de los indicadores de eficiencia y eficacia del proceso de la gestión académica específicamente para estudiantes que sean ausentes, inasistencia a clases, 			
RESULTADOS DEL AREA / INDICADORES			
<ul style="list-style-type: none"> Apoyo de formación Postgrados Plan de capacitación docente Mantenimientos Inscripciones efectivas Cumplimiento del Calendario Académico 			

Figura 18. Artefacto - Análisis de capacidades en Gestión de Comunicaciones - Uniminuto

Fuente: elaboración propia

NOMBRE DE LA CAPACIDAD		Gestión de Bienestar Estudiantil	
DESCRIPCIÓN			
Planear, promover y ejecutar proyectos que busquen el mejoramiento de la calidad de vida de la comunidad académica, contribuyendo a la formación integral, a través de eventos deportivos, recreativos, culturales, y de promoción y prevención, así como estrategias que busquen fortalecer la permanencia de los estudiantes.			
PROBLEMÁTICA		OPORTUNIDADES	
<ul style="list-style-type: none"> No contar con información a nivel de movilidad académica de los participantes Gestionar encuestas oportunas a fin de identificar necesidades a poblaciones Contar con una plataforma que permita administrar los eventos de la institución, permitiendo con esto un mayor control de las actividades programadas. No contar con calendario académicos exclusivamente para la modalidad de educación virtual y a distancia. Contar con una vista 360 del estudiante (desde el punto de vista académico) 		<ul style="list-style-type: none"> Gestionar eventos para el fortalecimiento de la formación integral del estudiante. Realizar talleres, campañas y asesorías profesionales acompañados por las unidades académicas Generar planes de Atracción y acompañamiento a los nuevos alumnos Liderar programas de refuerzos para estudiantes en riesgo académico 	
SUB-PROCESOS	INFORMACIÓN	PERSONAS (ROLES)	TECNOLOGÍA
<ul style="list-style-type: none"> Fortalecer aptitudes y actitudes culturales, deportivas y de aprovechamiento libre Permanencia de los estudiantes 	<ul style="list-style-type: none"> Estudiantes Matriculados Graduados Docentes Eventos Personas 	<ul style="list-style-type: none"> Director de Bienestar Estudiantil Coordinador Líder de Campañas Auxiliares de Eventos Recreativos Auxiliar Académicos Equipo de Logística 	<ul style="list-style-type: none"> LMS SIS ERP
INICIATIVAS			
<ul style="list-style-type: none"> Liderar políticas de permanencia y bienestar, como parte de una estrategia integral de mejoramiento de la calidad académica y el fomento a la graduación estudiantil. Reestructurar procesos que este enfocados a convertirse en un área de Éxito Estudiantil 			
RESULTADOS DEL AREA / INDICADORES			
<ul style="list-style-type: none"> Mejoramiento de la Calidad de Vida en la comunidad académica Aumento de la participación en los grupos culturales y deportivos 			

Figura 19. Artefacto - Análisis de capacidades en bienestar estudiantil - Uniminuto

Fuente: elaboración propia

NOMBRE DE LA CAPACIDAD		Asesores Académicos	
DESCRIPCIÓN			
Establecer una interacción e integración recíproca con estudiantes, teniendo en cuenta las necesidades y expectativas de la comunidad en general, mediante un intercambio dinámico de información para el control y seguimiento de estudiantes			
PROBLEMÁTICA		OPORTUNIDADES	
<ul style="list-style-type: none"> No cuenta con perfilamiento de estudiantes de educación virtual y a distancia en la Universidad. (Solamente lo tiene para estudiantes presenciales) No contar con una vista 360 del estudiante de educación virtual y a distancia (desde el punto de vista académico) de tal manera que los asesores puedan consultar de manera centralizada toda la información personal y académica de los alumnos que tienen a cargo. 		<ul style="list-style-type: none"> Manejar información del proceso de asesoría clara y oportuna que sirva para la toma de decisiones a través de dashboard y reportes. Flujos de Procesos y calendarización de las asesorías que se les debe brindar a los estudiantes en riesgo. 	
SUB-PROCESOS	INFORMACIÓN	PERSONAS (ROLES)	TECNOLOGÍA
<ul style="list-style-type: none"> Gestionar Poblaciones Objetivo Gestionar el perfil de Riesgo de las poblaciones Gestionar las Alertas Tempranas Gestionar las Intervenciones Gestionar actividades de asesoramiento 	<ul style="list-style-type: none"> Estudiantes Matriculados Docentes Notas Detalle Notas Históricas Eventos Personas 	<ul style="list-style-type: none"> Director de Académico Directores de Programas Académicos de Modalidad Virtual Coordinador de Programas Auxiliares Académicos 	<ul style="list-style-type: none"> LMS SIS ERP CRM
INICIATIVAS			
<ul style="list-style-type: none"> Consolidar modelo de seguimiento de asesorías para los estudiantes de educación virtual y a distancia. Desarrollar programas de asesoría (Coaching) para las poblaciones en riesgo de ausentismo o deserción Crear programas de orientación para los nuevos estudiantes. 			
RESULTADOS DEL AREA / INDICADORES			
<ul style="list-style-type: none"> Monitorias académicas Uso de recursos electrónicos por parte de la comunidad académica Medición de satisfacción de los servicios académicos que se le brindan al estudiante 			

Figura 20. Artefacto - Análisis de capacidades en asesores académicos - Uniminuto

Fuente: elaboración propia

NOMBRE DE LA CAPACIDAD		Gestión Académica	
DESCRIPCIÓN			
Asegurar la calidad de la formación, la gestión del conocimiento y el aprendizaje continuo que le permita al estudiante ser creador de oportunidades, crítico y propositivo para dar respuesta efectiva y de impacto al desarrollo social.			
PROBLEMÁTICA		OPORTUNIDADES	
<ul style="list-style-type: none"> No se cuenta con información del manejo y gestión de las cargas académicas asociadas a los facilitadores. Lo que impide una correcta gestión y asignación de los facilitadores a los cursos. El control de asistencia es llevado por el facilitador en una lista física entregada por la Universidad (cursos virtuales y a distancia). Lo cual implica que la información no esté en línea y se tenga que duplicar el ingreso o por un tercero. Contar con una vista 360 del estudiante (desde el punto de vista académico) La lista de curso no se encuentra en línea, lo que impide al facilitador contar con la información de los alumnos efectivamente activos en el curso. 		<ul style="list-style-type: none"> Gestionar procesos de calificaciones para la modalidad de Educación Virtual y a Distancia, a través del autoservicio de los docentes o tutores (Los docentes registrarían las calificaciones) Apoyar el cumplimiento del reglamento estudiantil a través de un sistema que permita registrar la asistencia del estudiante Gestión de los índices académicos de manera certera y automática Configurar la visualización de los reportes del avance del plan académico desde la parte administrativa. 	
SUB-PROCESOS	INFORMACIÓN	PERSONAS (ROLES)	TECNOLOGÍA
<ul style="list-style-type: none"> Gestionar la asistencia a clase Gestionar el ingreso de notas calificaciones Gestionar el avance académico 	<ul style="list-style-type: none"> Estudiantes Personas Graduados Docentes Cursos Notas Parciales y Finales de Cursos 	<ul style="list-style-type: none"> Vice Rector Académico (General) Vice Rector Financiero (General) Rector de Educación Virtual y a Distancia Directores de Programas de IEVD Coordinadores de Programas Docentes y Tutores 	<ul style="list-style-type: none"> LMS SIS ERP
INICIATIVAS			
<ul style="list-style-type: none"> Mejora en los índices de retención académica Aumentar la información para la toma de decisiones. Aumenta la Información precisa del avance curricular del alumno para procesos como graduación y asesoría 			
RESULTADOS DEL AREA / INDICADORES			
<ul style="list-style-type: none"> Evaluación de las actividades desarrolladas Cumplimiento del calendario académico Solicitudes de Servicios Académicos 			

Figura 21. Artefacto - Análisis de capacidades en asesores académicos - Uniminuto

Fuente: elaboración propia

En última instancia, se ha utilizado el artefacto de alineación conceptual para la deserción, junto con el equipo que participa en la definición de deserción estudiantil para los programas de Educación Virtual y a Distancia, con el objetivo de entender cuáles son los principios generales, las directrices y las problemáticas que desde la parte estratégica de la institución se requiere resolver.

Adicionalmente, se pretende tener una alineación con las capacidades o procesos de negocio, para que en la parte de construcción del modelo se pueden considerar elementos clave como la población objetivo, beneficios que se esperan con el modelo propuesto, entre otros elementos. A continuación, se presenta el artefacto de alineamiento conceptual.



Resumen – Alineamiento Conceptual








Definición  La rectoría de IEVD (Educación Virtual y A distancia) de la Universidad Minuto de Dios se basa en los principios del Ministerio de Educación Nacional en donde por medio del Sistema de Prevención y Análisis de la Deserción (SPADIES) donde la denominada "primera deserción" (first drop out), es decir, el abandono de un programa académico por dos semestres consecutivos.		Stakeholder (Área de Negocio)  Rectoría IEVD Vicerector Académico Bienestar Estudiantil (Éxito Estudiantil)
Problemática Identificada  <ol style="list-style-type: none"> 1. Problemas Financieros de los Estudiantes. 2. Deficiencias Académicas de la Secundaria 3. Las expectativas del estudiante no se alinean con el programa académico. 4. Pérdida constante de Materias y de créditos académicos 5. Imposibilidad de detectar y monitorear oportunamente los estudiantes en riesgo de deserción 6. Incapacidad para definir, recolectar y analizar los posibles riesgos asociados (Alertas) 7. Efectividad en el uso de los recursos destinados a la asesoría estudiantil 8. Efectividad en el uso de los recursos destinados a la asesoría 	Beneficios Esperados  <ol style="list-style-type: none"> 1. Cumplimiento de tasas de graduación (eficiencia terminal) 2. Aumentar el nivel de persistencia de los estudiantes 3. Mejoramiento de la calidad de los resultados académicos producidos por la institución 4. Aumentar el nivel de satisfacción de los estudiantes con los servicios prestados en la institución. 5. Planes de comunicaciones enfocados en fortalecer el compromiso de los estudiantes 	
Población Objetivo  Para el ejercicio la población objetivo es todos los estudiantes que están inscritos con materias activas desde el año 2010 en todos los programas académicos que son ofrecidos por la rectoría de Educación Virtual y a Distancia	Beneficios Esperado  <ol style="list-style-type: none"> 1. Cumplimiento en tasas de retención 2. Aumentar a nivel de motivación 3. Apoyar las habilidades académicas 4. Incrementar los resultados académicos 5. Disminuir el ausentismo de los estudiantes 	
Factores Críticos de Éxito  <ol style="list-style-type: none"> 1. Cargue Oportuna de la información por periodo académico. 2. Análisis periódico de la información 3. Compromiso de las áreas de apoyo (Tecnologías de Información) 4. Realizar auditorías internas de procesos e información 	Factores Críticos de Riesgo  <ol style="list-style-type: none"> 1. Áreas institucionales no comprometidas. 2. Deficiencias en la recolección de información 3. Falta de validación de la información del estudiante 4. Cambiar criterios en el modelo definido 5. Ausencia de registros de seguimiento de ausentismo y permanencia 	

Figura 22. Artefacto - Alineamiento conceptual de deserción estudiantil - Uniminuto

Fuente: elaboración propia

7.2.3 Desarrollo del trabajo

7.2.3.1 Definición de deserción

Para la Uniminuto, la deserción estudiantil se define como el abandono de un programa académico por dos periodos consecutivos, las problemáticas que lo generan pueden ser las siguientes:

- Problemas financieros del estudiante.
- Deficiencias académicas de educación media.
- Expectativas del estudiante no se alinean con el programa académico.
- Pérdida constante de materias y de créditos académicos.
- Posibilidad de detectar y monitorear a los estudiantes en riesgo de deserción o ausentismo.

- Imposibilidad actual de detectar ausentismo académico y en ocasiones ejecución de análisis cuando el estudiante ya ha desertado.
- Trámite de desvinculación académica realizado por el estudiante.

7.2.3.2 Entendimiento de los datos

- Colección de datos iniciales

En la Uniminuto se llevó a cabo el diligenciamiento de la plantilla de orígenes de datos, el resultado de ese ejercicio fue el siguiente:

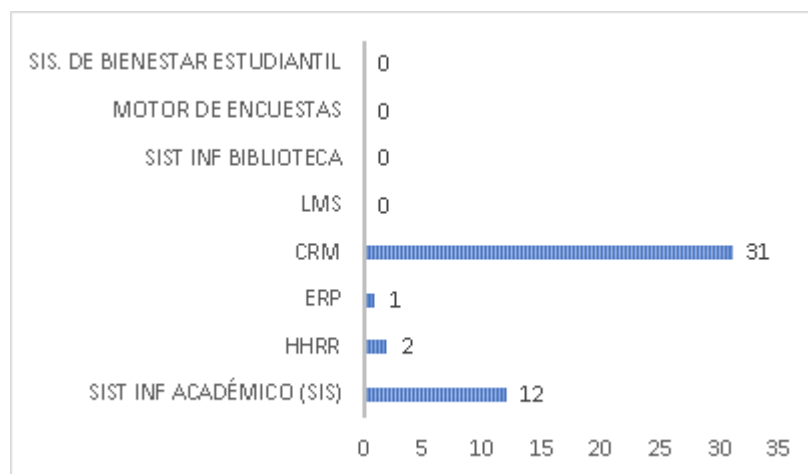


Figura 23. Validación orígenes de datos del caso de estudio

Fuente: elaboración propia

La cantidad de datos de la Uniminuto se ve afectada debido a que no está disponible en los sistemas de información actuales, hay muchos archivos de Excel que los contienen, y a causa de esta situación, la universidad emprendió un proyecto en el año 2017 para cambiar los sistemas actuales. Una vez se determinó dónde estaba la información, la universidad procedió al proceso de extracción, y entregó a los autores de esta investigación los siguientes conjuntos de datos:

- Admitidos: en esta colección de datos se visualiza a las personas que iniciaron un proceso de admisión para la educación virtual y a distancia.

- Bienestar: se recolecta la información de los eventos que realiza el área de Bienestar Universitario y donde se registra información de los estudiantes que participan en estos.
- Carga académica: la composición de las actividades académicas de cada docente, tales como información de horas de docencia, horas de investigación, horas en actividades de bienestar, etc.
- Catálogo cursos: en este se encuentran en detalle las características esenciales del curso como el nivel, el responsable de la materia, entre otras, incluso la intensidad horaria semanal.
- Demográficos: contiene información biográfica de las personas que actualmente tienen el rol de estudiante, quienes se encuentran registradas dentro del sistema de información académica.
- Deserción: aquí se evidencian los datos de los estudiantes que han tomado materias y se demuestra continuidad dentro de los periodos académicos y a lo largo de los años. Es importante afirmar que se deben calcular y estimar criterios de ausentismo y deserción.
- Docentes: listado de los profesores con atributos como categoría, tipo de contrato, facultad a la cual pertenece, periodo académico.
- Graduados: en esta colección de datos se halla información de los estudiantes que han iniciado su proceso de grado y de cuál fue su promedio histórico. Cabe agregar que se encuentra un nivel de granularidad bastante alto por el manejo de los periodos académicos, dado que se pueden dar varios escenarios y uno de ellos es que los estudiantes hayan terminado todas sus materias, y posteriormente (un año después) hayan iniciado su proceso de grado.
- Materias: dentro de este conjunto de datos hay información descriptiva, y se abordan las características de las materias registradas por estudiante dentro de la universidad, bajo la modalidad de educación virtual y a distancia.

- Notas detalle: allí se reconoce información en detalle del estudiante por todas las materias vistas en cada periodo académico del año. Es oportuno agregar que este conjunto de datos contiene registros desde el año 2010 hasta el año 2016, y que el objetivo es cubrir dos cohortes académicas en la modalidad de educación virtual y a distancia.

- Descripción de los datos

Para hacer la descripción de los datos se hizo uso de la hoja del artefacto 09 - Diccionario de datos, esta hoja se diligencia con base en las entidades identificadas en las capacidades de negocio de la Uniminuto.

- Exploración de los datos

La exploración de los datos fue llevada a cabo mediante la aplicación de las hojas del artefacto 07 - Análisis univariado y 08 - Análisis bivariado. Sobre esta actividad se considera necesario resaltar los siguientes puntos:

- La ciudad que predomina en los alumnos es Bogotá, con casi el 79% de concentración, lo cual permite desde el comienzo descartar este atributo como importante para detectar tendencias de deserción. Asimismo, hay un porcentaje alto de municipios muy cercanos a la capital que ratifican que no tiene sentido usar este dato para la predicción.
- El estrato está altamente concentrado en alumnos que pertenecen al 2 y al 3, con casi el 89% de los estudiantes entre los dos, este dato coincide con el público objetivo al que le apunta la universidad.
- Sobre el género de los estudiantes, el 73% de ellos son mujeres.
- La edad que predomina es de 26 a 30 años, esta representa el 26%, seguida por el rango de edad de 31 a 35 años con un 25%.
- El estado civil mayoritario es soltero en el 56% de los estudiantes.

7.2.3.3 Preparación de los datos

- Limpieza de los datos

Se procedió a limpiar los dataset, estos se analizaron mediante el análisis univariado, se hicieron correcciones por errores, y también se realizaron las conversiones de datos a que hubo lugar. Entre los principales cambios hechos están los siguientes:

Tabla 3. Limpieza de atributos de DataSet

Campo	Valores vacíos o nulos	% Vacíos	Valores erróneos
Género	72	1%	1%
Icfes	60705	36,58%	16.49%
Horas aulas virtuales por semestre	108	0.11%	
Estrato	490	0.51%	0.54%
Ciudad	460	0.48%	

Fuente: elaboración propia

- Dataset análisis de los datos

El proceso de preparación de datos llevó demasiado tiempo, en primer lugar, debido a la demora por parte de la universidad en la entrega de los datos, en segundo lugar, a que hubo tres revisiones de datos que fueron fallidas a causa de la poca cantidad de información entregada por

la entidad, en tercer lugar, a las diversas fuentes de datos que tiene la entidad y a la gran cantidad de tablas que fue necesario analizar y depurar.

Con el análisis ya realizado se hizo la integración de datos al generar una tabla que los contenía, esta tenía un total de 95 719 datos de semestres académicos por estudiante, y en ella se encontraron problemas como datos vacíos y valores incorrectos. Una vez realizada la limpieza de datos, quedó un total de 35 014 registros, de los cuales 6 093 corresponden a registros de semestres de estudiantes que desertaron. El proceso tuvo la siguiente secuencia de pasos en su construcción.

Tabla 4. *Descripción pasos de Preparación de Tabla Resumen*

Paso	Descripción
1	Se crea la vista agrupada de los datos demográficos por año de cada estudiante.
2	Se cruzan la vista de demográficos por año contra las notas detalladas de cada estudiante.
3	El catálogo de cursos se cruza con la tabla del paso anterior para agregar información, y se guarda en una vista.
4	Con base en la vista anterior se crea la tabla de resumen, cada registro consiste en un periodo académico del estudiante.
5	Se calcula la cantidad de materias cursadas por semestre y se agrega en un campo nuevo.
6	Se crea una columna con el cálculo del promedio de ese semestre.
7	Se calculan en dos columnas nuevas las materias aprobadas y pérdidas por semestre.

-
- 8 En un nuevo campo se totalizan la cantidad de créditos académicos por semestre.
- 9 Se crea una columna nueva en la cual se calcula el promedio acumulado de la carrera en ese semestre.
-

Fuente: elaboración propia

Adicionalmente, la tabla resumen de datos quedó con la siguiente estructura:

Tabla 5. *Descripción de atributos de Tabla Resumen*

Atributo	Observación
Programa	Código del programa académico.
Id	Código del estudiante.
Ciudad	Ciudad de residencia del estudiante.
Estrato	1, 2, 3, 4, 5, 6
Edad	Edad del estudiante en el periodo académico.
Género	Femenino, masculino.
Est. Civil	Casado(a), Soltero, U. Libre, Separado(a), Divorciado(a), Viudo(a), Religioso.
Periodo por estudiante	Número de semestre cursado en ese registro por el estudiante.
Horas aula virtual	Horas de clase virtual.
Horas talleres	Horas de actividades grupales en línea.

Cantidad materias por semestre	Cantidad materias por semestre tomadas por el estudiante.
Promedio semestre	Promedio de ese semestre cursado.
Cantidad aprobadas por semestre	Cantidad materias aprobadas.
Cantidad perdidas por semestre	Cantidad materias perdidas.
Cantidad créditos por semestre	Total de créditos cursados en ese semestre.
Promedio carrera acumulada en ese semestre	Promedio carrera a ese semestre.
Cantidad participación eventos de bienestar	Cantidad actividades promovidas por la universidad para mejorar el bienestar.
Número grupo	Número de grupo de estudiantes (estudiantes con los que se matriculó al comienzo de la carrera).
Periodos antes de matrícula	Número de meses desde que fue admitido vs. primer periodo de matrícula.
Ausentismo	0,1 (0 no se ausentó, 1 se ausentó)
ICFES	A, B, C, D, E
Desertó	0, 1 (0 no desertó, 1 desertó)

7.2.3.4 Modelamiento

- Técnicas de modelado a usar

En esta parte del proceso, para el desarrollo del proyecto aquí descrito, se planteó el uso de tres métodos:

- Regresión Logit
- KMeans
- Árbol de decisión

En esa medida, se estableció cuál de los tres es más funcional para la institución, pues cada modelo tiene un enfoque diferente. Para probar cada técnica se usó la misma metodología, se dividieron los registros en dos grupos, el 70% se utilizó para crear el modelo y el 30% restante para probarlo. A continuación, se explica el cálculo de cada modelo.

- Construcción del modelo

Regresión Logit

Los modelos de regresión incluyen técnicas estadísticas que tratan de explicar cómo se comporta un resultado cuando cambian otras variables, denominadas predictores.

A su vez, la función logística halla para cada individuo, basada en los valores de variables, la probabilidad de que presente el efecto. Así, al aplicar transformación logarítmica a la ecuación (denominada logit), convierte la probabilidad en posibilidades.

Este modelo se implementó en R, usa solo variables numéricas, y hace posible la creación de un modelo de regresión logarítmico multivariable para dar un resultado, en este caso, es la probabilidad de que un estudiante deserte. Cabe aclarar que este no es un modelo de minería de datos, es un modelo matemático y estadístico. Los pasos para su aplicación fueron los siguientes:

Paso 1. Validación de variables significativas para el cálculo

Este proceso consiste en detectar qué variables son las que realmente tienen relación con la deserción, para eso, el modelo calcula la desviación estándar para verificar la dispersión vs. la media, el valor Z. El resultado de la ejecución es este:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	2.1004934	0.0849483	24.727	< 2e-16	***
deser\$ESTRATO	0.0212997	0.0119060	1.789	0.0736	.
deser\$GENEROMASCULINO	0.1016394	0.0185307	5.485	4.14e-08	***
deser\$GENERONO_DISPONIBLE	0.8815562	0.3280418	2.687	0.0072	**
deser\$EST_CIVILDivorciado(a)	0.0053702	0.0816670	0.066	0.9476	
deser\$EST_CIVILReligioso	-0.2978853	0.2374498	-1.255	0.2097	
deser\$EST_CIVILSeparado(a)	-0.0009951	0.0523771	-0.019	0.9848	
deser\$EST_CIVILSoltero	-0.0041380	0.0203124	-0.204	0.8386	
deser\$EST_CIVILU. Libre	0.0009193	0.0253060	0.036	0.9710	
deser\$EST_CIVILViudo(a)	0.1339626	0.1694819	0.790	0.4293	
deser\$Horas_Salon	-0.0113609	0.0008051	-14.111	< 2e-16	***
deser\$Horas_Laboratorio	-0.0944879	0.0105212	-8.981	< 2e-16	***
deser\$Cantidad_Materias_Semestre	-0.0096167	0.0078442	-1.226	0.2202	
deser\$Promedio_Semestre	-0.2406825	0.0194363	-12.383	< 2e-16	***
deser\$Cantidad_Aprobadas_Semestre	-0.0526205	0.0076301	-6.896	5.33e-12	***
deser\$Cantidad_Creditos_Semestre	-0.0396117	0.0034084	-11.622	< 2e-16	***
deser\$Intensidad_Horaria_Semestre	0.0113495	0.0007922	14.327	< 2e-16	***
deser\$Promedio_Carrera_en_ese_Semestre	-0.3700241	0.0226100	-16.366	< 2e-16	***
deser\$Ausentismo	-0.8627369	0.0192419	-44.836	< 2e-16	***

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Figura 24. Variables Significativas

En síntesis, el resultado de la corrida marca con asteriscos las columnas que son buenas para estimar el comportamiento de la variable deserción; con ese propósito se evalúa la desviación con respecto a la media aritmética en unidades de desviación típica, representada en la columna Z, de este modo, entre menor desviación es más útil en la predicción.

Paso 2. Ejecutar el modelo logit

En este paso se desarrolla la ejecución del modelo, se descartan las variables que no son significativas, y los resultados son los siguientes:

```

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      2.0873480  0.0714981  29.194 <2e-16 ***
deser$ESTRATO     0.0253393  0.0117489   2.157  0.031 *
deser$Horas_Salon -0.0109154  0.0007915 -13.790 <2e-16 ***
deser$Horas_Laboratorio -0.0933761  0.0104678  -8.920 <2e-16 ***
deser$Promedio_Semestre -0.2319700  0.0180763 -12.833 <2e-16 ***
deser$Cantidad_Aprobadas_Semestre -0.0583232  0.0051545 -11.315 <2e-16 ***
deser$Cantidad_Creditos_Semestre -0.0414632  0.0032188 -12.882 <2e-16 ***
deser$Intensidad_Horaria_Semestre  0.0109670  0.0007789  14.080 <2e-16 ***
deser$Promedio_Carrera_en_ese_Semestre -0.3717384  0.0223174 -16.657 <2e-16 ***
deser$Ausentismo  -0.8633984  0.0191383 -45.114 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Figura 25. Variables Definitivas

La ejecución da como variables importantes: estrato, horas aula virtual, horas talleres, promedio semestre, cantidad de créditos aprobados, cantidad de créditos por semestre, intensidad horaria, promedio de carrera en ese semestre, ausentismo.

Paso 3. Análisis de residuos

En él se deben validar la desviación de los residuos; en este caso, para que la ejecución de este modelo sea buena, la desviación debe ser baja.

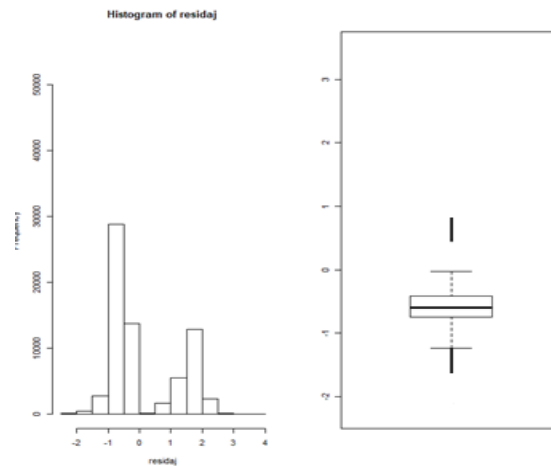


Figura 26. Histograma Residaj

En el *box plot* no se ven tantos puntos fuera de la caja, sin embargo, los residuos si están muy desviados, esta misma situación se ve en la línea de tendencia del tercer gráfico.

Paso 6. Calcular la función objetivo

Esta función permite, mediante la aplicación de la ecuación obtenida del modelo de regresión logit, calcular la probabilidad de desertar, y solo usa las variables del paso 2. En el ejemplo se calcula la probabilidad de un solo estudiante:

```
> Probab_deser1=predict(modelo.logit2, newdata = ndatos1,type = "response")
> Probab_deser1[1]
      1
0.1725952
```

La función en este ejemplo da como resultado que el estudiante tiene una probabilidad de desertar de un 17%. Además, en la aplicación de la función a los registros de prueba, se halló el siguiente resultado:

Desertó ↓ / Predicción →	No	Sí
No	7329	1184
Sí	1224	767

Accuracy: 77% - Error 23%

KMeans

Este es un algoritmo de *clustering* que se usa para detectar patrones, no busca una variable objetivo, a su vez, el modelo tiene múltiples variables de entrada.

K-means requiere que se defina el número de grupos en los que se quieren segmentar los datos. Con la definición de la cantidad de *clusters*, la técnica ubica primero los centroides, y luego asigna a los puntos las muestras con las menores distancias.

Posteriormente el punto se mueve a la media de las muestras más cercanas, y se reasignan las muestras para corregir la cercanía a otros centroides. La repetición iterativa hace que los grupos

se ajusten hasta que la asignación no cambia más moviendo los puntos. Al final, el ajuste maximiza la distancia entre grupos y minimiza la distancia dentro del grupo.

Este método permite clasificar un estudiante en un grupo que tenga comportamiento similar, y con base en eso tiene una probabilidad promedio de desertar o no. De igual modo, trabaja solo con variables numéricas, por lo cual se usan las mismas variables que resultaron significativas en la aplicación del modelo logit.

Por consiguiente, para poder encontrar la mejor distribución de los grupos, se llevó a cabo una iteración de 2 a 6 grupos, y se buscó que se distribuyeran homogéneamente. El resumen de la iteración se muestra en la siguiente tabla.

Tabla 6. *Descripción de Iteraciones de Grupo*

Número de grupos	Porcentajes	Observación
1	96%, 3%	El primer clúster o segmento tiene un peso de casi el 100%, lo cual hace necesario aumentar el número de grupos.
2	8%, 3%, 88%	El tercer grupo tiene un porcentaje cercano al 90%, aún hay desbalance.
3	40%, 50%, 6%, 4%	La mayor parte del peso se distribuye en los dos primeros clústeres.
4	14%, 31%, 44%, 6%, 4%	La distribución es mejor, el 70% se concentra en dos grupos.
5	0.5%, 5%, 48%, 43%, 3%, 0.2%, 0.14%	La mayor concentración se da en dos clústeres con un 91%.

Fuente: elaboración propia

Al revisar los resultados, la mejor combinación es la de 5 grupos, por ese motivo se procede a revisar las características de cada clúster y la probabilidad de deserción.

Tabla 7. *Descripción de Segmentos Detectados - kmeans*

Clúster	Características
1 (14.3%)	<p>Segmento 1: el 67.6% de los estudiantes no deserta, mientras el 32.4% sí deserta (es el grupo con deserción más alta). Se puede ver que las horas en aula virtual están entre 4 y 30 horas por semana. Además, la mayoría de los estudiantes tiene cero horas de taller virtual (81%). La dispersión en estos estudiantes es baja, el promedio del semestre va de 0 a 5, con mediana 2.8. El 75% de estos estudiantes han aprobado máximo 3 asignaturas, y han aprobado de 0 a 9 asignaturas. La mediana de asignaturas perdidas de estos estudiantes es de 3 y el 75% de ellos tienen máximo 5 asignaturas perdidas. Los estudiantes en este grupo tienen mediana promedio de la carrera en 3.4, el 75% tiene máximo 3.9 de promedio de la carrera. El 52.2% de ellos no presenta ausentismo y el 42.8% sí lo presenta.</p>
2 (31.4%)	<p>Segmento 2: el 74.5% de los estudiantes no deserta, mientras el 25.5% sí deserta, teniendo características descritas a continuación. Se evidencia que las horas en el salón están entre 0 y 15 horas en aula virtual. El 68.7% de los estudiantes tienen cero horas de laboratorio y el 31.2% horas de laboratorio. La dispersión en estos estudiantes es alta, el promedio del semestre va de 0 a 5, con mediana 4. Los estudiantes han aprobado de 0 a 20 asignaturas. La mediana de asignaturas perdidas por ellos es de 0, el 75% tienen máximo una asignatura perdida, y han perdido de 0 a 6 asignaturas. Los estudiantes en este grupo tienen mediana promedio de la carrera en 4, el 75% tiene máximo 4.2 de promedio de la carrera. El 75.1% de ellos no presenta ausentismo y el 24.9% sí lo presenta.</p>

-
- 3 (44.8%) Segmento 3: el 82.8% de los estudiantes no deserta, mientras el 17.2% sí deserta, con las características descritas a continuación. Las horas en el salón están entre 13 y 29 horas. La mayoría de los estudiantes se sitúa en cero horas de laboratorio (97.6%). La dispersión en ellos es alta, el promedio del semestre va de 0 a 5, con mediana 4. A su vez, la mediana de asignaturas perdidas por ellos es de 0, y el 75% tiene máximo una asignatura perdida. Los estudiantes en este grupo tienen mediana promedio de la carrera en 4, el 75% tiene máximo 4.2 de promedio de la carrera, y el 100% tienen entre 0 y 4.9 de promedio de la carrera. El 59.3% de estos estudiantes no presenta ausentismo y el 40.7% sí lo presenta.
- 4 (5.7%) Segmento 4: el 67% de los estudiantes no deserta, mientras el 33% sí deserta (es el segundo grupo con más deserción y muy cercano al grupo 1); las características en torno a estas cifras se describen a continuación. Se observa que las horas en el aula virtual están cerca a las 30 horas en salón. La mayoría de los estudiantes se sitúa en cero horas de laboratorio (96.1%). La dispersión en ellos es alta, el promedio del semestre tiene mediana 4.1. Asimismo, han perdido de 0 a 6 asignaturas. Los estudiantes en este grupo tienen mediana promedio de la carrera en 4, el 75% tiene máximo 4.2 de promedio de la carrera y el 100% tienen entre 0 y 4.9 de promedio de la carrera. El 73.1% no presenta ausentismo y el 26.9% sí lo presenta.
- 5 (3.8%) Segmento 5: el 72.8% de los estudiantes no deserta, mientras el 27.2% sí deserta, con las características descritas a continuación. La mayoría de los estudiantes se sitúa en cero horas de laboratorio (100%). Su dispersión es alta, y el promedio del semestre va de 0 a 5, con mediana 4.5. El 75% de ellos han aprobado máximo dos asignaturas, y han aprobado de 0 a 21 asignaturas. La mediana de asignaturas perdidas es de 30, el 75% de los estudiantes tienen máximo 0 asignaturas perdidas, y han perdido de 0 a 5 asignaturas. Los estudiantes en este grupo tienen mediana promedio de la carrera en 4.1, el 75% tiene máximo 4.3 de promedio de la carrera y el 100% tienen entre 0 y 4.9 de promedio de la carrera. El 69.4% de estos estudiantes no presenta ausentismo y el 30.6 sí lo presenta.
-

Cabe aclarar que este método es de agrupamiento, y se probó mediante el uso de la misma metodología, hasta que dio los siguientes resultados:

Desertó ↓ / Predicción →	No	Sí
No	7260	1426
Sí	1293	525

Accuracy: 74% - Error 26%

Árbol de decisión

Es un enfoque que usa técnicas matemáticas para describir la categorización de un conjunto de datos, por medio de la utilización de subdivisiones, y basado en las variables introducidas, clasifica los resultados en ramas. Esta técnica se usa para predecir en qué nodo se puede clasificar un conjunto de registros; tiene como ventajas que los datos no tienen restricción, pueden ser numéricos o categóricos, y no es necesario que tengan tratamiento previo de *outliers* o de datos vacíos, puesto que estos no influyen el resultado del modelo. En este caso, se aplicó un árbol de clasificación a los datos, para lo cual se usó el índice gini sin poda; la composición del árbol quedó de la siguiente manera:

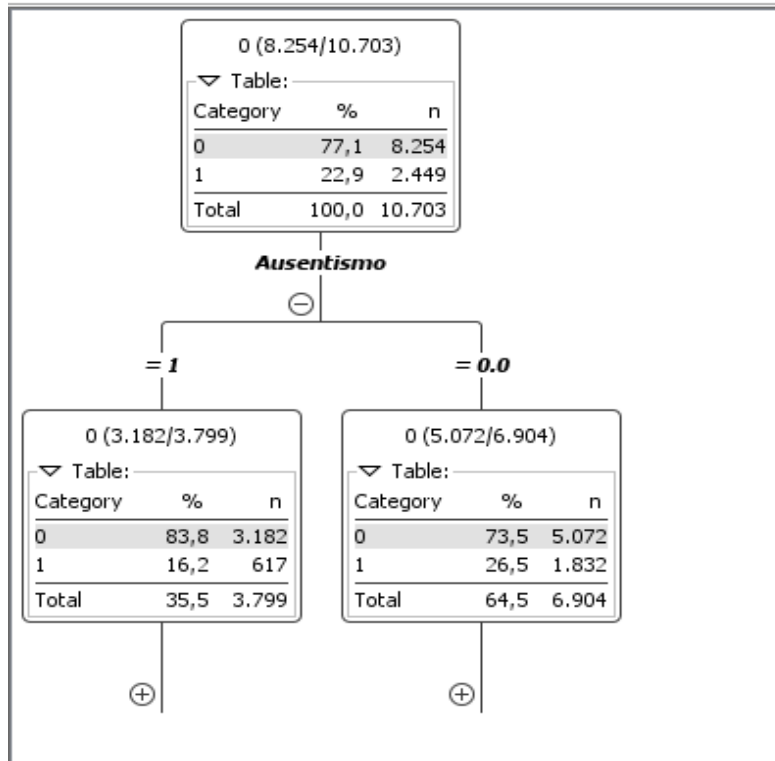


Figura 27. División por ausentismo

Fuente: elaboración propia

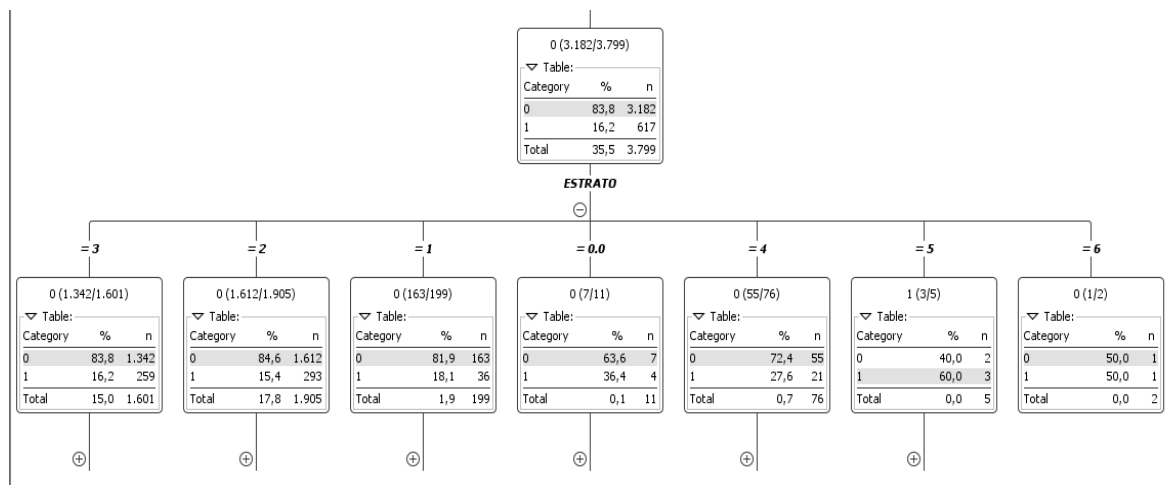


Figura 28. División por estrato socioeconómico

Fuente: elaboración propia

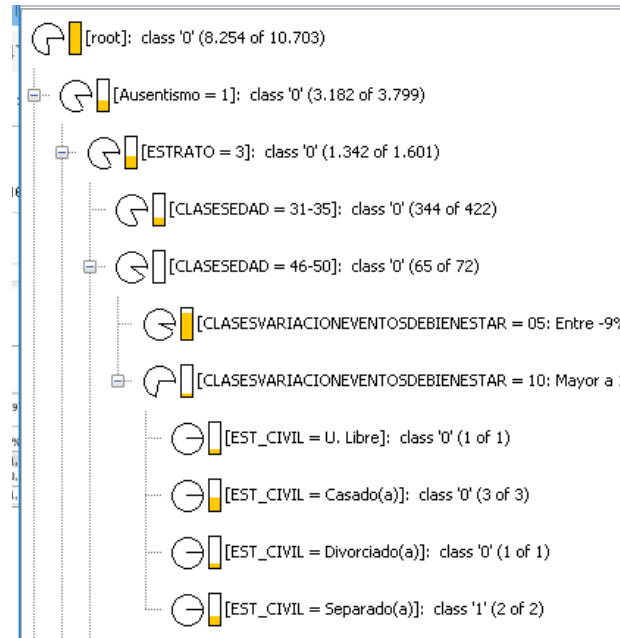


Figura 29. Ramas del árbol

Fuente: elaboración propia

El árbol quedó con 5 ramas, y se utilizaron como campos ausentismo, estrato, edad, eventos de bienestar y estado civil. Por consiguiente, los resultados dados por el mejor árbol construido son los siguientes:

Desertó ↓ / Predicción →	No	Sí
No	6790	1500
Sí	1763	451

Accuracy: 69% - Error 31%

Al comparar los resultados de los tres métodos en el caso de estudio, se seleccionó la regresión logit, pues fue la que mejor resultado dio (menor margen de error). Se consideró también que la información disponible se basa más en el rendimiento académico, y estos datos, sumados al enfoque de registro basado en el semestre, permiten actuar más rápidamente.

7.3 Despliegue de componentes tecnológicos

Para el desarrollo de los pasos, la dependencia entre las actividades de entendimiento del marco metodológico propuesto y el caso de estudio es clave, dado que posibilita definir un modelo a construir, acompañado de componentes y funcionalidades de aplicaciones de Data Science y demás aplicaciones satélites. Por ello, y como parte solamente para el desarrollo del caso, se desarrolló un modelo de despliegue acorde con la solución planteada, a fin de garantizar funcionalidad. A continuación, se relaciona un diagrama de componentes tecnológicos que evidencian de manera más clara y desglosada los elementos que son parte de la solución propuesta.

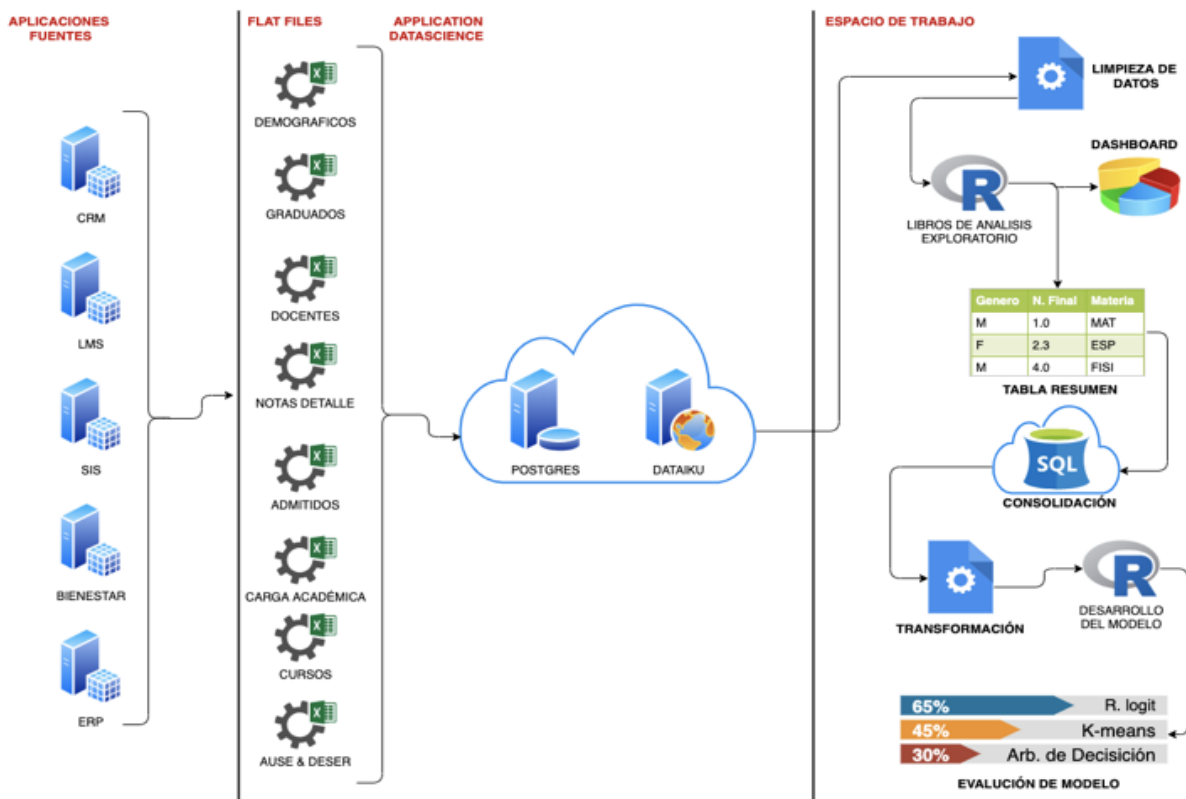


Figura 30. Diagrama de despliegue de componente de tecnología

Fuente: elaboración propia

7.4 Conclusiones del modelo

Los datos que brindó la universidad Uniminuto no fueron suficientes para hacer un mejor ejercicio, sin embargo, es claro que permiten hacer una estimación por diferentes medios, basada en la metodología planteada, al dar un enfoque de probabilidad. En síntesis, a pesar de las distorsiones, se pueden usar estos métodos para una buena estimación.

8 Conclusiones y Recomendaciones

8.1 Conclusiones

De acuerdo con los objetivos trazados y mencionados en el documento, se concluye lo siguiente:

- Para llevar a cabo un entendimiento de la deserción en las IES que tienen programas de educación virtual y a distancia fue indispensable y necesario realizar un diagnóstico a fin de buscar un alineamiento conceptual de deserción, además de realizar el análisis de capacidades en las áreas de negocio que monitorean la deserción estudiantil.
- En relación directa con los artefactos utilizados en la fase de diagnóstico de la deserción en con el análisis de capacidades en las cuatro en las áreas de negocio, permitió identificar los sistemas de información o fuentes de información que sirven como apoyo a los procesos operativos de las IES, además permitió identificar cuáles son las entidades y datos de negocio que participan directamente en los procesos de gestión de comunicaciones, gestión académica, asesores académicos y bienestar universitario. Finalmente, los artefactos propuestos de análisis de capacidades pueden ser utilizados por el EIS si se requiere analizar y extender a otros procesos académicos.

- Para lograr construir el modelo predictivo usando técnicas estadísticas y de minería de datos que permitiera identificar la deserción estudiantil para los programas de educación virtual y a distancia, fue necesario comprobar y clasificar los datos o variables en cada uno de los dominios (individuales, socioeconómicos, académicos, institucionales) y los cuales son determinantes en la deserción.
- Finalmente, para comprender el fenómeno del abandono estudiantil en la educación superior para los programas de modalidad virtual, no hay nada mejor que la aplicación de la metodología con datos del mundo real, donde junto con la rectoría Virtual y A distancia de la Universidad Minuto de Dios, se logro realizar el análisis de información de deserción a aplicando técnicas estadísticas y de minería de datos, logrando predecir con una exactitud del 75% la deserción de los estudiantes por medio de la técnica estadística de regresión logística binaria, además se logra identificar los factores y variables que inciden en la deserción.

8.2 Trabajos Futuros

Durante el desarrollo de la tesis han surgido algunas líneas futuras de investigación que se han dejado abiertas y que pueden ser de interés para estudiantes de la maestría que deseen profundizar; ya que están directamente relacionadas con este trabajo de tesis y son el resultado de cuestionamiento e interrogantes que han ido surgiendo durante la realización de esta. A continuación, se presentan algunos trabajos futuros que pueden desarrollarse como resultado de esta investigación de profundización. Entre los posibles trabajos futuros se destacan:

- Realizar análisis de texto y/o sentimientos en los foros, evaluaciones, grupos, actividades y demás elementos que son usados en las aulas virtuales (LMS). Estas herramientas se usan

como apoyo para el estudiante en las modalidades de educación presencial y virtual en la a la educación superior.

- Realizar el análisis de deserción con respecto al tiempo, específicamente con la deserción precoz en donde la persona que ha sido admitida por la institución no logra o no culmina su proceso de matricula, allí se lograría evidenciar otros factores que pueden estar ligados a factores institucionales como por ejemplo procesos administrativos densos, entre otros.
- Definir e implementar procesos de un modelo de éxito estudiantil en las instituciones de educación superior apoyado en metodologías de arquitectura empresarial.
- Desarrollar maquinas de aprendizaje (Machine Learning) para la optimización de horarios, espacios físicos, programación de los cursos entre otros procesos de apoyo que tienen las instituciones de educación superior.

9 Referencias

- Bennett, G., Seashore, H., & Wesman, A. (1992). *Tests de Aptitudes Diferenciales, DAT. Manual Forma T.* . Buenos Aires: Paidós.
- Burn, A., Feickert, D., Newby, M., & Winterton, J. (1983). The miners and new technology. *Industrial Relations Journal*, 14(4), 7-20.
- Hernández, A. (2009). *Deserción en las instituciones de educación superior a distancia en América Latina y el Caribe*. República Dominicana: Ediciones UAPA.
- Herrero, V., Merlino, A., Ayllón, S., & Escanés, G. (2013). Aplicación de un modelo de duración en programas de prevención de deserción universitaria. *REDIE. Revista Electrónica de Investigación Educativa*, 15(3), 38-52.
- Kember, D. (1989). A Longitudinal-Process Model of Drop-Out from Distance Education. *The Journal of Higher Education*, 278-301.
- Ministerio de Educación. (2014). *Informe mensual sobre el soporte técnico y avance del contrato para garantizar la alimentación, consolidación, validación y uso de la información del SPADIES*. From https://www.mineduccion.gov.co/sistemasdeinformacion/1735/articles-254702_Informe_determinantes_desercion.pdf
- Ministerio de Educación Nacional. (2008). *Elementos para su diagnóstico y tratamiento*. From https://www.mineduccion.gov.co/sistemasdeinformacion/1735/articles-254702_diagnostico_desercion.pdf
- Ministerio de Educación Nacional. (2011). *Template* . From https://www.mineduccion.gov.co/1759/articles-341922_archivo_pdf.pdf

Peppard, J. (2016). *A Tool to Map Your Next Digital Initiative*. From Harvard Business Review:

<https://hbr.org/2016/06/a-tool-to-map-your-next-digital-initiative>

Salas-Velasco, M. (2017). The transition from higher education to employment in Europe: the analysis of the time to obtain the first job. *Higher Education*, 54(3), 333-360.

Santiago, C. (2013). *Propuesta de arquitectura empresarial para una institución de educación superior*. From

<https://repositorio.escuelaing.edu.co/bitstream/001/460/1/Santiago%20Cely%2c%20Claudia%20Patricia%20-%202013.pdf>

Spady, W. (1971). Dropouts from higher education: toward an empirical model. *Interchange*, 2(3), 1573-1790.

Tinto, V. (1975). Dropout from Higher Education: A Theoretical Synthesis of Recent Research. *Review of Educational Research*, 89-125.

UNIMINUTO. (2014). *Plan de desarrollo 2013-2019*. Bogotá: Editorial UNIMINUTO.

UNIMINUTO. (2017). *UVD en cifras* . From

http://www.uniminuto.edu/documents/968618/10842084/uvd_cifras_201740_digital.pdf/0d27e31f-e27f-4860-adb9-a90a6192495c

UNIMINUTO. (2018). *Informe de gestión*. From

<http://www.uniminuto.edu/documents/10181/0/Informe+de+Gesti%C3%B3n+UNIMINUTO+2018/1f2b5589-3ac2-4adf-89f1-4ce54d673851>

Glosario

ADMISIÓN: Proceso académico realizado por la institución de educación superior donde las personas se convertirán en estudiantes para el programa académico en el que se inscribió previamente.

ALERTA TEMPRANA: Permite identificar comportamientos o cambios estado en los diferentes pasos en el ciclo de vida del estudiante dentro de las instituciones de educación superior esto con el fin de detectar a tiempo posibles riesgos de ausentismo y de deserción por parte del estudiante.

AYUDA FINANCIERA: Son estrategias que brinda la institución o alguna entidad del gobierno, en donde principalmente que busca apoyar a estudiantes con dificultades económicas para continuar sus estudios en la educación superior a fin de garantizar la permanencia del estudiante en su programa académico.

CNA: El Consejo Nacional de Acreditación, es un organismo de naturaleza académica y el cual hace parte del Ministerio de Educación en Colombia.

CARACTERIZACIÓN: Es la aplicación de métodos de análisis de datos que permite identificar características y perfiles de estudiantes a fin de entender factores e evidenciar necesidades particulares del estudiante.

COHORTE: Grupo de estudiantes que ingresan en un primer periodo académico en una institución de educación superior.

DESERCIÓN: Es el abandono que hace el estudiante de manera forzada o voluntaria del proceso educativo formán antes de finalizar el ciclo académico.

IES: Es la abreviación del termino que se les da a Instituciones o Institutos de Educación Superior.

LMS: Es la abreviación del termino de Learning Management System – Sistema de Gestión de Aprendizaje. Básicamente es un sistema de información como apoyo a la educación que permite administrar, gestionar y controlar las actividades de formación de los estudiantes bajo la modalidad de educación virtual y presencial en las instituciones de educación superior.

MEN: Es la abreviación del termino de Ministerio de Educación Nacional de Colombia.

PERSISTENCIA: Es la capacidad académica, además de características sociales y la motivación del estudiante para alcanzar sus propias metas educativas.

SPADIES: Es la abreviación del termino de Sistema para la Prevención de la Deserción de la Educación Superior. Este servicio que esta a cargo del Ministerio de Educación Superior de Colombia, consolida la información que reportar las instituciones de educación superior a fin de conocer información estadística de aspectos demográficos, académicos y financieros de los

estudiantes, lo cual permite diseñar planes de acción a las instituciones para disminuir el riesgo de deserción.

SNIES: Es la abreviación del termino de Sistema Nacional de Información de Educación Superior en donde las instituciones de educación superior realizan tramites administrativos e institucionales.