



**ANÁLISIS POR MINERÍA DE DATOS DEL IMPACTO DE LOS  
SISTEMAS DE CALIDAD DE LAS INSTITUCIONES DE EDUCACIÓN  
SUPERIOR EN LOS RESULTADOS DE LAS PRUEBAS SABER  
PRO ENFOCADO A LOS PROGRAMAS DE INGENIERÍA  
INDUSTRIAL.**

**CAMILA GONZÁLEZ MONTES  
SERGIO ANDRÉS GUILLEN IBARRA**

UNIVERSIDAD TECNOLÓGICA DE BOLÍVAR  
FACULTAD DE INGENIERÍAS  
PROGRAMA DE INGENIERÍA INDUSTRIAL  
CARTAGENA DE INDIAS D.T Y C.  
Febrero 2019



ANÁLISIS POR MINERÍA DE DATOS DEL IMPACTO DE LOS  
SISTEMAS DE CALIDAD DE LAS INSTITUCIONES DE EDUCACIÓN  
SUPERIOR EN LOS RESULTADOS DE LAS PRUEBAS SABER PRO  
ENFOCADO A LOS PROGRAMAS DE INGENIERÍA INDUSTRIAL.

**AUTORES:**

CAMILA GONZÁLEZ MONTES  
SERGIO ANDRÉS GUILLEN IBARRA

**Trabajo de Grado para Optar al Título de Ingeniero Industrial**

**DIRECTOR**

ENRIQUE JOSÉ DE LA HOZ  
**MSc investigación de operaciones**

UNIVERSIDAD TECNOLÓGICA DE BOLÍVAR  
FACULTAD DE INGENIERÍAS  
PROGRAMA DE INGENIERÍA INDUSTRIAL  
CARTAGENA DE INDIAS D.T Y C.

Febrero 2019



Cartagena de Indias, 15, Febrero de 2019

**Señores:**  
**COMITÉ CURRICULAR**  
**Programa de Ingeniería Industrial**  
**Facultad de Ingeniería**  
**Universidad Tecnológica de Bolívar**  
**Ciudad**

Respetados Señores:

Por medio de la presente me permito someter para estudio, consideración y aprobación el anteproyecto de trabajo de grado titulado “Análisis por minería de datos del impacto de los sistemas de calidad de las instituciones de educación superior en los resultados de las pruebas Saber Pro enfocado a los programas de ingeniería industrial”, desarrollada por los estudiantes Camila González Montes y Sergio Andrés Guillen Ibarra, en el marco de su formación como Ingeniero Industrial.

Como director del proyecto considero que el anteproyecto cumple con el alcance y los requisitos exigidos para tal propósito, por lo que amerita ser presentado formalmente para su evaluación.

Cordialmente,

---

**Enrique José De La Hoz**  
Director de Trabajo de Grado



Cartagena de Indias, 15 Febrero de 2019

**Señores:**

**COMITÉ CURRICULAR**

**Programa de Ingeniería Industrial**

**Facultad de Ingeniería**

**Universidad Tecnológica de Bolívar**

**Ciudad**

Respetados Señores:

Por medio de la presente nos permitimos someter para estudio, consideración y aprobación el trabajo de grado titulado “Análisis por minería de datos del impacto de los sistemas de calidad de las instituciones de educación superior en los resultados de las pruebas Saber Pro enfocado a los programas de ingeniería industrial”, desarrollado en el marco de nuestra formación como Ingeniero Industrial.

Cordialmente,

---

**Camila González Montes**

Investigadora

---

**Sergio Andrés Guillen Ibarra**

Investigador



## RESUMEN

La acreditación institucional tiene como objetivo principal fomentar el mejoramiento continuo por medio de la evaluación y seguimiento de los procesos internos de una institución de educación superior. Este proceso de acreditación es voluntario con el fin de certificar ante la sociedad la calidad de sus servicios, razón por la cual las instituciones de educación superior están en la constante búsqueda de estas certificaciones en calidad [1].

En este proyecto de investigación se busca conocer si estas acreditaciones generan un impacto sobre el resultado de las instituciones de educación superior, medición que se realizara con base a los resultados de las pruebas Saber Pro, enfocado en los programas de ingeniería industrial en Colombia, a través de la información tomada de la base de datos del ICFES.

Para la obtención de resultados y el desarrollo del estudio se realizaron dos tipos de análisis, en primer lugar se planteó un enfoque cualitativo, el cual se fundamentó en la ejecución de entrevistas a profundidad a personas pertenecientes a los departamentos de la universidad Tecnológica de Bolívar, posteriormente se procedió a la obtención de resultados mediante un enfoque cuantitativo, para lo que fue necesario, la utilización de técnicas de minería de datos y de aprendizaje automático como los métodos de aprendizaje supervisado para la predicción de si una institución es acreditada o no según sus resultados en las pruebas Saber Pro, al igual que los métodos de aprendizaje no supervisado para la clasificación de las instituciones en grupos que cuenten con características similares.

Luego del análisis realizado, se obtuvieron resultados acertados, encontrando grandes diferencias entre los desempeños obtenidos por las instituciones de educación superior que cuentan con una acreditación institucional y aquellas que no cuentan con esta distinción, por lo que se pudo identificar la influencia que tienen las certificaciones de calidad en el desempeño de los estudiantes en este tipo de pruebas.

Dentro del marco científico se espera que esta investigación contribuya a futuros trabajos que se puedan desarrollar en el campo de la minería de datos enfocado a la educación superior en Colombia.

**Palabras claves:** Instituciones de educación superior, Saber Pro, Certificaciones de calidad, acreditación, minería de datos, aprendizaje automático, clúster, exactitud.



## ABSTRACT

The main purpose of institutional accreditation is to encourage continuous improvement through the evaluation and monitoring of the internal processes of a higher education institution. This accreditation process is voluntary in order to certify to society the quality of its services, which is why institutions of higher education are constantly searching for these quality certifications [1].

This research project seeks to know if these accreditations generate an impact on the results of higher education institutions, a measurement that will be made based on the results of the Saber Pro tests, focused on industrial engineering programs in Colombia, through the information taken from the ICFES database.

To obtain the results and the development of the study, two types of analysis were carried out. Firstly, a qualitative approach was considered, which was based on the execution of in-depth interviews with people belonging to the departments of the Technological University of Bolívar. Subsequently, results were obtained through a quantitative approach, for which it was necessary to use data mining and machine learning techniques such as supervised learning methods for the prediction of whether an institution is accredited or not based on its results. in Saber Pro tests, as well as unsupervised learning methods for the classification of institutions into groups with similar characteristics.

After the analysis, successful results were obtained, finding large differences between the performances obtained by higher education institutions that have an institutional accreditation and those that do not have this distinction, so that the influence of the certifications could be identified of quality in the performance of students in this type of tests.

Within the scientific framework it is expected that this research will contribute to future work that can be developed in the field of data mining focused on higher education in Colombia.

**Keywords:** Institutions of higher education, Saber Pro, Quality Certifications, accreditation, data extraction, machine learning, cluster, accuracy.



*Quiero dedicar esta tesis principalmente a mis padres, que con su esfuerzo y dedicación han permitido mi paso hasta aquí.*

*Agradecer a Dios por brindarme la oportunidad de crecer como persona, de culminar mis estudios satisfactoriamente y darme la fuerza para lograr mis objetivos.*

*A mis padres por confiar en mi y apoyarme en todo momento para la realización de mi sueños y darme fortaleza en los momentos que se tornaron difíciles durante este proceso, por ser mi motivación para todo.*

*A mi tutor Enrique de la Hoz por ser una guía y un apoyo fundamental, ya que nos brindó el tiempo y paciencia necesaria para culminar con este proyecto satisfactoriamente.*

*A mi compañero Sergio Guillen por poner su grano de arena y su esfuerzo para que este proyecto se llevara a cabo.*

**Camila González**

*Primero quiero agradecer a Dios por brindarme la sabiduría y el conocimiento para realizar este proyecto y poder cumplir este gran logro.*

*A mis padres por brindarme la oportunidad de formarme como profesional y crecer personalmente, siendo siempre pieza fundamental en el desarrollo de mis actividades, guiándome y dándome voz de aliento en los momentos de adversidad.*

*A mis hermanos y mi abuela Amira Olivo por siempre ser un apoyo para mí en los momentos buenos y malos.*

*A nuestro tutor Enrique de la Hoz por guiarnos en todo el camino y ser un apoyo fundamental en el desarrollo de este proyecto.*

*A mi compañera Camila González por su gran aporte a este proyecto.*

*Por ultimo pero no menos importante quiero dedicar este proyecto a mi abuelo Wilfrido Ibarra quien desde el cielo me cuida y me guía hacia el éxito.*

**Sergio Guillen**



## LISTA DE CONTENIDO

1	DESCRIPCIÓN DEL PROBLEMA .....	3
1.1	Contexto global .....	3
1.2	Contexto local .....	4
2	FORMULACIÓN DEL PROBLEMA.....	6
3	OBJETIVOS .....	7
3.1	Objetivo general .....	7
3.2	Objetivos específicos .....	7
4	REQUERIMIENTOS DE SOLUCIÓN.....	8
5	JUSTIFICACIÓN .....	9
6	MARCO TEORICO .....	10
6.1	Introducción .....	10
6.2	Calidad.....	10
6.2.1	Modelos de calidad:.....	10
6.2.1.1	Modelo Deming: .....	10
6.2.1.2	Modelo Malcolm Baldrige: .....	11
6.2.1.3	Modelo EFQM de excelencia:.....	12
6.2.1.4	Normas ISO 9000:.....	13
6.2.1.5	Normas ISO 9001 de 2015: .....	13
6.2.1.6	Modelo ABET (Accreditation Board of Engineering and Technology): 14	
6.2.1.7	Sistema nacional de acreditación: .....	15
6.3	Minería de datos y técnicas.....	16
6.3.1	Inteligencia artificial .....	16
6.3.2	Machine learning .....	17
6.3.2.1	Aprendizaje supervisado .....	17
6.3.2.1.1	Regresión logística.....	17
6.3.2.1.2	Arboles de decisión .....	18
6.3.2.1.3	Random Forest .....	18
6.3.2.2	Aprendizaje no supervisado .....	19
6.3.2.2.1	Análisis de componentes principales.....	19

6.3.2.2.2	Análisis de Clúster.....	19
6.3.3	Método jerárquico .....	20
6.3.4	Método no jerárquico .....	20
6.3.5	Calculo de Distancia .....	20
6.4	Examen Saber Pro.....	20
6.5	Estado del arte.....	21
6.5.1	Estudios de carácter cuantitativo: .....	21
6.5.1.1	Dentro del campo educativo .....	22
6.5.1.2	Fuera del campo educativo .....	23
6.5.2	Estudios de carácter cualitativo: .....	26
6.5.2.1	Educación superior.....	26
6.5.2.2	Minería de datos.....	27
7	METODOLOGIA DE LA INVESTIGACIÓN .....	30
7.1	Enfoque de la investigación .....	30
7.2	Tipo de investigación .....	30
7.3	Diseño de la investigación.....	30
8	ENTREVISTAS.....	32
8.1	Ficha técnica.....	32
8.2	Preguntas:.....	32
8.3	Resultados .....	33
9	ANÁLISIS Y RESULTADOS.....	36
9.1	Análisis estadístico de datos .....	41
9.2	Proceso de clasificación.....	43
9.2.1	Clúster jerárquico .....	44
9.2.2	Clúster no jerárquico.....	46
9.3	Modelos de predicción .....	48
9.3.1	Regresión logística .....	49
9.3.2	Random Forest.....	51
10	ENTREGABLES DEL PROYECTO .....	53
11	CONCLUSIÓN .....	54
	REFERENCIAS BIBLIOGRAFICAS .....	55

## LISTA DE FIGURAS

Figura 1. Ciclo PHVA. ....	11
Figura 2. Diagrama del modelo Malcolm Baldrige. ....	12
Figura 3. Modelo EFQM. ....	13
Figura 4. Estructura de los principios del 4 al 10 en el ciclo PHVA. ....	14
Figura 5. Estructura de un árbol de decisión. ....	18
Figura 6. Esquema del algoritmo Radom Forest. ....	19
Figura 7: Metodología utilizada para el desarrollo del análisis. ....	37
Figura 8. Scree test para el análisis PCA. ....	39
Figura 9. PCA Biplot. ....	41
Figura 10. Dendrograma obtenido por el cluster jerárquico. ....	44
Figura 11: Diagrama de grupos creados por K – Medias. ....	47
Figura 12: Curva ROC para el modelo de regresión logística. ....	51
Figura 13. Curva ROC para el modelo de Random Forest. ....	53

## LISTA DE TABLAS

Tabla 1. Porcentaje de accesibilidad a educación terciaria. ....	3
Tabla 2. Requerimientos de solución para la investigación. ....	8
Tabla 3. Estado del arte de investigaciones con enfoque cuantitativo. ....	26
Tabla 4. Estado del arte de investigaciones con enfoque cualitativo. ....	29
Tabla 5. Ficha técnica de entrevistas a profundidad. ....	32
Tabla 6. Competencias genéricas y específicas evaluadas en las pruebas Saber Pro .....	36
Tabla 7: Resultados de PCA para cada componente. ....	39
Tabla 8: Variabilidad explicada para las componentes 1 y 2. ....	40
Tabla 9. Resultados de prueba T-Student para instituciones acreditadas y no acreditadas. ....	42
Tabla 10. Resultados de prueba T-Student para instituciones públicas y privadas. ....	42
Tabla 11. Resultados de prueba T-Student para instituciones de la costa caribe .....	43
Tabla 12. Accuracy de los diferentes modelos de clúster utilizados. ....	44
Tabla 13: Medias por componente para las instituciones según la clasificación del clúster jerárquico. ....	46
Tabla 14: Medias por componente para las instituciones según la clasificación del clúster jerárquico. ....	48
Tabla 15: Matriz de confusión del modelo de regresión logística. ....	49
Tabla 16: Medidas de desempeño del modelo de regresión logística. ....	50
Tabla 17: Matriz de confusión del modelo Random Forest .....	51
Tabla 18: Medidas de desempeño del modelo Random Forest .....	52

## INTRODUCCIÓN

El uso progresivo de la tecnología ha dado paso al registro de grandes cantidades de datos dentro de las organizaciones e instituciones de esta era, lo que se ha convertido en una tarea rutinaria para los sistemas de información empresariales, ya que estos representan un historial de sus procesos internos y de su funcionamiento en general. Este almacenamiento de datos no es útil si no se le efectúa un tratamiento, por lo que es necesario la utilización de herramientas y técnicas que permitan la extracción de los mismos, es ahí donde se da lugar a la aplicación de minería de datos, la cual a menudo es utilizada para la resolución de problemas de todo en tipo en cualquier área, ya que a través de sus técnicas se hace posible el análisis de bases de datos de gran volumen. En el marketing por ejemplo, la minería de datos ha sido de clave para el logro de la alta satisfacción de los clientes con el fin de mejorar la rentabilidad de las empresas.[2]

Dentro de estas instituciones se encuentra el instituto colombiano para la evaluación de la educación en Colombia (ICFES) el cual se encarga de evaluar la educación en todos sus niveles, con el fin de la obtención de información que permita mejorar la calidad educativa[3], esta información se ve representada en bases de datos de gran tamaño. Esta calidad educativa antes mencionada es un tema con alta subjetividad, pero a pesar de eso gracias a que estamos en un mundo globalizado, se ha visto normalizada a través de requisitos y estándares estructurados en modelos de acreditación, con el fin de lograr satisfacer a la sociedad.[4]

El principal propósito de estudio de este proyecto de investigación es conocer la influencia que tienen las certificaciones y acreditaciones en calidad sobre los resultados institucionales de las pruebas Saber Pro en los programas de ingeniería industrial en Colombia desde un enfoque cuantitativo. Para lo anterior, se utilizaron las bases de datos del ICFES para la obtención de la información y posteriormente la aplicación de técnicas de minerías de datos de aprendizaje no supervisado y aprendizaje supervisado, con el fin de realizar los análisis correspondientes para la obtención de resultados. De este modo, el desarrollo de esta tesis se fundamenta en la siguiente pregunta de investigación ¿Qué impacto tienen las certificaciones de calidad sobre los resultados de las pruebas Saber Pro en los programas de ingeniería industrial?

En primer lugar se pretende contextualizar el objeto de estudio a través de la formulación y definición del problema en cuestión, para luego dar paso a la presentación de objetivos y a la justificación del mismo. Luego, se plantea una referencia teórica sobre las temáticas abordadas en la presente investigación, abarcando desde el concepto de calidad hasta la aproximación de las técnicas

aplicadas, con el fin de dar una correcta introducción para la comprensión de la posterior presentación de los análisis y resultados a nivel cuantitativo que fueron obtenidos en el desarrollo de la misma. También se presentan resultados a nivel cualitativo, los cuales se fundamentan en entrevistas a profundidad realizadas a personas representativas dentro del programa de ingeniería industrial en la universidad Tecnológica de Bolívar.

# ANÁLISIS POR MINERÍA DE DATOS DEL IMPACTO DE LOS SISTEMAS DE CALIDAD DE LAS INSTITUCIONES DE EDUCACIÓN SUPERIOR EN LOS RESULTADOS DE LAS PRUEBAS SABER PRO ENFOCADO A LOS PROGRAMAS DE INGENIERÍA INDUSTRIAL

## 1 DESCRIPCIÓN DEL PROBLEMA

### 1.1 Contexto global

La educación superior representa para la sociedad hoy en día, los pilares de su evolución, debido a la toma de conciencia de los países sobre la influencia de la misma sobre el desarrollo de los sistemas económicos, culturales, productivos y competitivos de cualquier economía. Según la UNESCO, durante la segunda mitad del siglo actual, se presentó el fenómeno más importante en la historia de la educación superior, un incremento realmente significativo sobre el número de estudiantes matriculados a nivel mundial, lo que dio paso al comienzo de una nueva era, aunque esto trajo consigo una serie de problemáticas que representaron nuevos retos[5].

Una de las principales problemáticas desatadas es la accesibilidad a la educación superior en igualdad de condiciones a nivel global, esto a raíz de factores como la poca inversión y la dependencia de las instituciones de educación superior públicas con el estado. Esto repercute en la creación de instituciones de educación superior privadas, para las cuales solo un porcentaje de la población cuenta con el acceso por disponibilidad de recursos, lo que genera que se presente una tasa de movilidad social muy baja[6].

Según los datos recolectados por la UNESCO de la tasa de matrículas a nivel mundial de educación terciaria, los países con más accesibilidad se encuentran mayormente en américa del norte (83,8%) principalmente EE.UU[7] (Ver tabla 1).

África (subsahariana)	8,40%
Asia (central y meridional)	24,90%
África (septentrional)	32,80%
Asia (oriental y sudoriental)	43,40%
América latina y el caribe	48,40%
Europa	70,30%
Oceanía	78,60%
América del norte	83,80%

Tabla 1. Porcentaje de accesibilidad a educación terciaria.  
Fuente: UNESCO[7].

Con base en los datos anteriores se puede concluir que, en América Latina, la tasa de matrículas es una de las más bajas nivel mundial. Fenómenos como la globalización, que se denotan como una oportunidad de desarrollo, también representan una amenaza, por lo que han generado un reto y una presión constante por alcanzar los mismos niveles de competitividad con la educación superior en países industrializados[8].

Aunque, gracias a la internacionalización, se ha logrado una integración intercultural, lo que permite que se den intercambios de conocimientos con el fin de que haya una disminución en la desigualdad que se presenta[9].

Según la UNESCO, las instituciones de educación superior deben articular sus métodos de enseñanza con los avances tecnológicos, la investigación, la calidad y la autonomía institucional, esto con el fin hacer más eficiente su pedagogía y de esta manera desarrollar en los estudiantes habilidades y conocimientos de calidad que generen un gran aporte socioeconómico.

El concepto de calidad es importante para las instituciones de educación superior ya que, contribuye a la mejora de sus niveles de competitividad y así mismo, garantizan el valor que tienen ante la sociedad, asegurando su eficiente participación con la misma. Esta se mide mediante una evaluación en la cual se tienen en cuenta diferentes aspectos como, los estudiantes, la infraestructura, los docentes y todo el personal encargado de funciones administrativas y gestión universitaria. Es decir, debe ser un análisis sistemático del conjunto de factores que componen las instituciones de ES, lo que representa para estas un esfuerzo continuo por llegar a estándares y exigencias que establecen los organismos de acreditación[9].

## **1.2 Contexto local**

En Colombia la educación superior juega un papel muy importante en el desarrollo económico, social y político con el cual está comprometido el gobierno nacional, con el fin de formar y capacitar a las futuras generaciones de manera integral para que estas posean las competencias necesarias para asumir las responsabilidades y problemas de la sociedad actual.

El ente encargado a nivel nacional de formular la política de educación y fomentar el desarrollo de una educación de calidad y competitiva que contribuya a disminuir las brechas de inequidad y mejorar la participación social, es el Ministerio de Educación Nacional de Colombia (MinEducación).

En el año 2014 se creó el programa Ser Pilo Paga, el cual es un programa del gobierno nacional que busca beneficiar a los mejores estudiantes de educación secundaria del país en las pruebas saber 11, que posean menores recursos

económicos, para que accedan a instituciones de educación superior acreditadas de alta calidad. Este programa funciona a través de créditos 100% condonables, los cuales sirvan para satisfacer el valor de la matrícula académica y adicionalmente brindar un apoyo para garantizar la sostenibilidad durante el periodo de estudios [10].

Para este programa se tuvo en cuenta solamente las instituciones de educación superior que cuenten con certificados de alta calidad, las cuales son en total más de 40 instituciones de educación superior, ubicadas en 15 departamentos: Antioquia, Atlántico, Bolívar, Boyacá, Caldas, Cauca, Cundinamarca, Huila, Magdalena, Meta, Norte de Santander, Quindío, Risaralda, Santander y Valle del Cauca. Es decir, pilos de 17 departamentos deberán desplazarse a otras ciudades, por lo que cuentan con un subsidio mayor para su sostenibilidad [11]. Un factor clave a destacar es que dentro de las carreras profesionales más escogidas por los pilos se encuentra Ingeniería Industrial. Si bien este programa ha ayudado a ingresar a instituciones de educación superior a casi 40.000 estudiantes (10.000 por año desde 2015 hasta el año vigente) el ministerio de educación ha emitido un comunicado en el mes de septiembre en el cual informa que el programa de Ser Pilo Paga se acaba y que en el mes de octubre el vigente año se presentara un nuevo programa de acceso a la educación superior que vincula criterios de equidad, mérito y que buscara fortalecer las instituciones públicas [12].

Por otra parte es pertinente hablar sobre el tema de las acreditaciones a nivel de Colombia, en este aspecto encontramos el sistema nacional de acreditación en Colombia (SNA) el cual es el conjunto de políticas, estrategias, procesos y organismos cuyo objetivo fundamental es garantizar a la sociedad que las instituciones de educación superior que hacen parte del sistema cumplen con los más altos requisitos de calidad y que realizan sus propósitos y objetivos [13]. La Acreditación es un testimonio que da el Estado sobre la calidad de un programa o institución con base en un proceso previo de evaluación en el cual intervienen la institución, las comunidades académicas y el Consejo Nacional de Acreditación el cual es un organismo de naturaleza académica que hace parte del SNA y depende del consejo nacional de educación superior (CESU), el cual define el reglamento, funciones e integración del CNA [14].

Cabe resaltar que el proceso de acreditación es voluntario para las instituciones de educación superior, y este certificado de alta calidad es temporal, por lo que luego de cumplir este tiempo si se desea mantener la acreditación la institución debe realizar procesos de re acreditación.

## **2 FORMULACIÓN DEL PROBLEMA**

Día a Día la mayoría de las instituciones de educación superior a nivel mundial, se encuentran en una búsqueda constante por conseguir certificaciones de alta calidad, con el fin de validar que la educación que están brindando es de excelencia, pero muchas veces esto no garantiza cumplir con las exigencias de la sociedad actual.

En la actualidad son aproximadamente 40, las instituciones de educación superior a nivel nacional, que cuentan con acreditación de alta calidad [15] y muchas otras están en busca de esta. En Colombia se realiza anualmente el examen de estado de la calidad de la educación superior “Saber Pro” que tiene como objetivo principal evaluar a los estudiantes de programas universitarios de nivel profesional, el cual tiene como fin evaluar y brindar resultados del grado de desarrollo de habilidades y conocimientos generales de aquellos estudiantes que han superado el 75% de los créditos de sus programas de formación profesional [16]. Este examen es aplicado tanto a instituciones de educación profesional acreditadas, como no acreditadas, motivo por el cual nos preguntamos si existe influencia de las certificaciones de calidad en las instituciones de educación superior, sobre los resultados obtenidos en las pruebas Saber Pro, fundamentalmente en los resultados obtenidos por los programas de Ingeniería Industrial, realizando un análisis de estos resultados por medio de minería de datos.

### **3 OBJETIVOS**

#### **3.1 Objetivo general**

Crear un modelo para la caracterización y análisis de los resultados obtenidos en las pruebas Saber Pro de los programas de ingeniería en Colombia, mediante la aplicación de minería de datos, para identificar el impacto de la calidad en el desempeño de las instituciones de educación superior.

#### **3.2 Objetivos específicos**

- Identificar los resultados obtenidos por los programas de ingeniería industrial a nivel nacional en las pruebas Saber Pro.
- Diseñar una metodología para la caracterización de los resultados obtenidos.
- Analizar los resultados obtenidos, dependiendo de varios criterios:
  - Instituciones de educación superior acreditadas vs Instituciones de educación superior no acreditadas
  - Instituciones de educación superior privadas vs Instituciones de educación superior públicas.
  
- Determinar el impacto la relación que existe entre las certificaciones de alta calidad y el desempeño de las instituciones de educación superior en la prueba Saber Pro.

#### 4 REQUERIMIENTOS DE SOLUCIÓN

Los requerimientos para la solución del presente proyecto de investigación se muestran en la siguiente tabla:

<b>Objetivo</b>	<b>¿Cómo?</b>	<b>Recursos</b>
Identificar los resultados obtenidos por los programas de ingeniería industrial a nivel nacional en las pruebas Saber Pro.	<ul style="list-style-type: none"><li>- Obtención de datos.</li><li>- Revisión bibliográfica.</li></ul>	<ul style="list-style-type: none"><li>- Bases de datos de Saber Pro</li></ul>
Diseñar una metodología para la caracterización de los resultados obtenidos.	<ul style="list-style-type: none"><li>- Uso de herramientas computacionales</li></ul>	<ul style="list-style-type: none"><li>- R Studio</li></ul>
Analizar los resultados obtenidos, dependiendo de varios criterios	<ul style="list-style-type: none"><li>- Uso de herramientas computacionales</li></ul>	<ul style="list-style-type: none"><li>- R Studio</li></ul>
Determinar el impacto la relación que existe entre las certificaciones de alta calidad y el desempeño de las instituciones de educación superior.	<ul style="list-style-type: none"><li>- Uso de herramientas computacionales</li><li>- Análisis de resultados</li></ul>	<ul style="list-style-type: none"><li>- R Studio</li></ul>

Tabla 2. Requerimientos de solución para la investigación.

## 5 JUSTIFICACIÓN

Las certificaciones de calidad juegan un papel fundamental hoy en día en toda la sociedad, cada día el interés por esta temática aumenta más y más, teniendo un gran peso e impacto en el sector académico, sobre todo en la educación superior, y es una herramienta fundamental para plantear estrategias que contribuyan a la mejora de la educación y calidad de la misma. Por otro lado existen instituciones como el ICFES que ayuda por medio de sus servicios a evaluar la educación en Colombia y que además, realiza investigaciones sobre los factores que afectan la calidad de la educación con el fin de obtener resultados que permitan realizar una mejora continua de esta. Si bien es cierto que obtener una certificación de educación superior requiere de muchos requisitos por parte del Consejo Nacional de Acreditación (CNA), los resultados obtenidos por las instituciones de educación superior en las pruebas “Saber Pro” son tenidos en cuenta como criterio para la asignación de este tipo de certificaciones. Por lo que es interesante e importante la elaboración de un modelo para caracterizar los resultados obtenidos por las instituciones de educación superior en la prueba, para identificar si existen tendencias que nos permitan asociar el comportamiento de los datos con la clasificación de las instituciones como acreditadas o no acreditadas con ayuda de la minería de datos.

Este modelo podría llegar a servir de mucha ayuda para instituciones de educación superior que busquen mejorar la calidad de la educación que les brindan a sus estudiantes, a su vez podría ser un punto de partida para futuros análisis en el campo del análisis de datos y la calidad de la educación superior en Colombia.

## **6 MARCO TEORICO**

### **6.1 Introducción**

Con el objetivo de alcanzar una óptima comprensión de los temas de investigación abordados en el presente proyecto, se referenciará la base teórica necesaria para el reconocimiento de los conceptos y aspectos que fueron necesarios para el desarrollo del mismo. De esta manera, se aborda desde el tema principal “la calidad” hasta los modelos representativos de esta, algunos de los cuales fueron de gran aporte para la consecución del objetivo de la investigación.

En segunda instancia, se presenta a través de un estado del arte, investigaciones y/o artículos que han tenido como objeto de estudio las temáticas abordadas, con el fin, de recopilar información relevante para el análisis de los resultados ya obtenidos en investigaciones similares y de esta manera, lograr articularlos con la investigación en curso.

### **6.2 Calidad**

El concepto de “calidad” ha sido un tema controversial desde el siglo XX. Algunos filósofos denominados “Filósofos de la calidad” presentaron sus argumentos y razonamientos a cerca de lo que para ellos es la calidad, dando lugar a discrepancias y acuerdos en puntos específicos. Según las definiciones de Philip B. Crosby, W. Edwards Deming, Armand V. Feigenbaum, Kaoru Ishikawa, Joseph M. Juran, Robert M. Pirsig, Walter A. Shewhart y Genishi Tagushi, se da lugar a similitudes, las cuales se pueden tomar como base para llegar a una sola definición de calidad. Según esto, la calidad es el grado en que una serie de características y requerimientos establecidos, satisfacen las necesidades del cliente.[17]

La calidad es vista como una necesidad en cualquier organización, por lo que se ha dado lugar a la creación de modelos internacionales de gestión de calidad, que sirven como un una referencia para el camino a la excelencia de estas. Estos modelos son:

#### **6.2.1 Modelos de calidad:**

##### **6.2.1.1 Modelo Deming:**

Fue desarrollado en el año 1951 en Japón por la Unión Japonesa de Científicos e Ingenieros (JUSE) en colaboración con el americano William Edwards Deming. (Ver figura 1).

Los 10 criterios que recoge este modelo para la evaluación de la gestión de la calidad en una organización son:

1. Políticas y objetivos.
2. Organización y operativa.

3. Educación y su diseminación.
4. Flujo de información y su utilización.
5. Calidad de productos y procesos.
6. Estandarización.
7. Gestión y control.
8. Garantía de calidad de funciones, sistemas y métodos.
9. Resultados.
10. Planes para el futuro.



Figura 1. Ciclo PHVA. .

#### **6.2.1.2 Modelo Malcolm Baldrige:**

Este modelo nace en Estados Unidos en 1987, como una necesidad de fomentar y sensibilizar al país a cerca de la importancia que tiene la gestión de la calidad total en las organizaciones, a raíz, de la invasión de productos japoneses en el mercado. (Ver figura 2).

Los criterios en los que se basa el modelo son:

1. Liderazgo.
2. Planificación.
3. Enfoque hacia el cliente y el mercado.
4. Medida, análisis y gestión del conocimiento.
5. Gestión de los recursos humanos.
6. Gestión de los procesos.
7. Resultados.

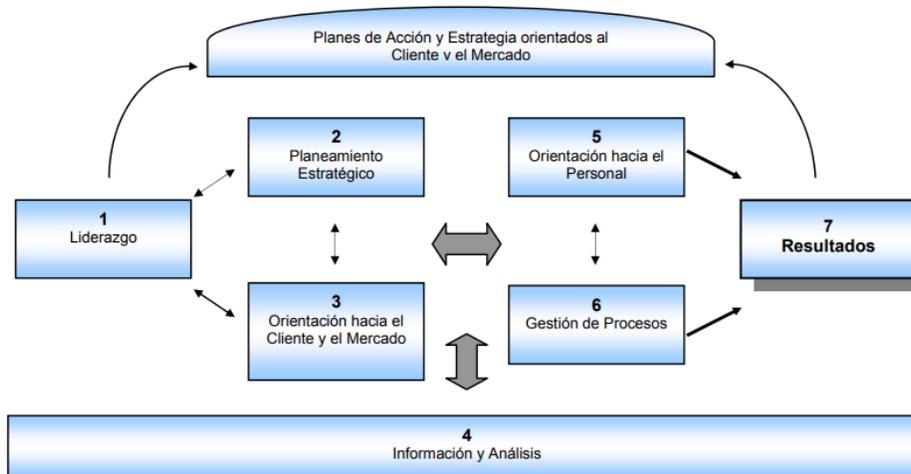


Figura 2. Diagrama del modelo Malcolm Baldrige.  
Fuente: Modelo de excelencia en la gestión malcolm baldrige [18].

### 6.2.1.3 Modelo EFQM de excelencia:

Europa no se quedó atrás en la aplicación de la calidad total, por lo que 14 importantes empresas crearon la Fundación Europea Para La Gestión De Calidad (EFQM) en 1988, el cual se basa en la siguiente premisa: *“Los resultados excelentes en el rendimiento general de una organización, en sus clientes, personas y en la sociedad en la que actúa se logran mediante un liderazgo que dirija e impulsa la Política y Estrategia, que se hará realidad a través de las personas de la organización, las alianzas y recursos y los procesos”* [19]. (Ver figura 3).

Este modelo consta de 9 criterios:

1. Liderazgo.
2. Personas.
3. Políticas y estrategias.
4. Alianzas y Recursos.
5. Procesos.
6. Resultados en las personas.
7. Resultados en los clientes.
8. Resultados el sociedad.
9. Resultados clave.



Figura 3. Modelo EFQM.  
Fuente: Modelo de excelencia y calidad EFQM [20]

#### 6.2.1.4 Normas ISO 9000:

Estas son un conjunto de normas internacionales de gestión de calidad que fueron establecidas por la Organización Internacional de Normalización (ISO).

Los criterios que maneja esta norma son:

1. Enfoque del cliente.
2. Liderazgo.
3. Participación del personal.
4. Enfoque basado en procesos.
5. Mejora continua.
6. Enfoque de sistemas para la gestión.
7. Enfoque basado en hechos para la toma de decisiones.
8. Relaciones mutuamente beneficiosas con proveedores[21].

#### 6.2.1.5 Normas ISO 9001 de 2015:

La ISO 9001 es una norma internacional cuyo propósito es mejorar la eficacia de un SGC dentro de una organización, con el fin de garantizar la satisfacción del cliente, mejorando sus productos o servicios, basados en: un enfoque por procesos, el pensamiento basado en riesgos y el ciclo PHVA. (Ver figura 4).

Los principios que maneja son:[22]

1. Objeto y campo de aplicación.
2. Referencias normativas.
3. Términos y definiciones.
4. Contexto de la organización.
5. Liderazgo.
6. Planificación.

7. Apoyo.
8. Operación.
9. Evaluación del desempeño.
10. Mejora.

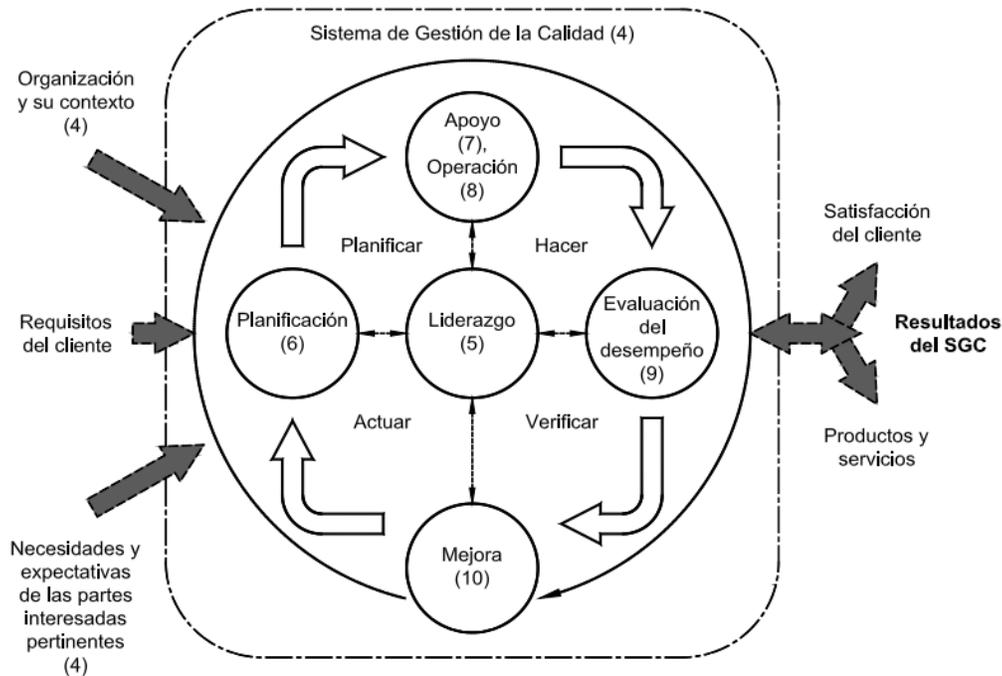


Figura 4. Estructura de los principios del 4 al 10 en el ciclo PHVA.  
Fuente: Norma técnica ntc-iso colombiana 9001," 2015 [22].

#### 6.2.1.6 Modelo ABET (Accreditation Board of Engineering and Technology):

ABET es una agencia internacional no gubernamental y sin ánimo de lucro conformada por 35 sociedades, cuyo fin es brindar acreditación a programas de educación superior en ciencias aplicadas, ingeniería, tecnología de ingenierías y ciencias de computación. Los programas universitarios acreditados por ABET garantizan cumplir con los criterios de calidad, otorgando a los estudiantes una óptima formación[23].

La acreditación ABET es un proceso voluntario, está a cargo de pares evaluadores, los cuales realizan una revisión y evaluación de la formación que ofrecen los respectivos programas[24]. El proceso básicamente consiste en:

- Evaluación de un informe de autoestudio presentado por el programa que se encuentra en el proceso de acreditación, el cual debe cumplir con los estándares y especificaciones requeridos.
- Visita al campus del respectivo programa.
- Revisión periódica, con el fin de garantizar si el programa sigue cumpliendo con la calidad exigida y de esta manera renovar la acreditación.

- Publicación de los programas acreditados.

#### **6.2.1.7 Sistema nacional de acreditación:**

El sistema nacional de acreditación fue creado por la Ley 30 del 28 de diciembre 1992 (Artículo 53) y reglamentado por el Decreto 2904 de diciembre 31 de 1994[25]. En Colombia, para la educación superior, existe un ente acreditador: el Consejo Nacional De Acreditación (CNA) el cual se fundamenta en el Sistema Nacional de Acreditación (SNA) para otorgar a las instituciones educativas las certificaciones y acreditaciones en calidad. El modelo del CNA tiene en cuenta los siguientes criterios [26]:

1. Misión y Proyecto Institucional
2. Profesores
3. Estudiantes
4. Procesos Académicos
5. Investigación
6. Pertinencia e Impacto Social
7. Proceso de Autoevaluación y Autorregulación
8. Bienestar Institucional
9. Organización, Gestión y Administración
10. Planta física y Recursos de Apoyo Académico
11. Recursos Financieros

La acreditación institucional es un proceso sistemático en el cual se evalúan diversos factores con el fin de definir si la institución cumple o no con los criterios para obtener dicha acreditación [27]. El lineamiento es el siguiente:

1. **Apreciación de condiciones iniciales de acreditación institucional:**  
El CNA cuenta con unas condiciones iniciales como requisito mínimo que deben ser cumplidas por parte de la institución que este en busca de la acreditación. Estas condiciones iniciales deben ser solicitadas por la IES para luego ser evaluadas por el CNA y determinar si cumple o no para continuar con el proceso [28].
2. **Autoevaluación institucional.**  
La IES realiza un proceso de autoevaluación según los criterios del CNA, los cuales deben ser solicitados según área de conocimiento o tipo de institución.
3. **Evaluación externa.**  
Esta evaluación es realizada por pares académicos externos de diversas instituciones, nombrados por el CNA, con el fin de verificar que la autoevaluación realizada por la IES cumpla con todos los lineamientos y requisitos establecidos. Estos pares emiten un informe al CNA con el fin de informar sobre los resultados obtenidos y emisión de recomendaciones a la institución cuando sea necesario.

4. Se obtiene una respuesta del informe dado a la IES.
5. Si la IES cumplió satisfactoriamente con lo establecido, el CNA envía un informe final al Ministerio de Educación que es el encargado de dar la certificación de acreditación.

### **6.3 Minería de datos y técnicas**

La minería de datos es un proceso que permite el manejo y la extracción de grandes cantidades de datos para la búsqueda de patrones, tendencias o perfiles con el fin de obtener información importante para la toma de decisiones, mediante el análisis de los mismos. [29]

Durante el desarrollo de la investigación, se aplicaron técnicas de minería de datos para la obtención de resultados, por lo que es de vital importancia conocer en que consiste cada una y asimismo algunos conceptos relacionados. Estas se explican a continuación:

#### **6.3.1 Inteligencia artificial**

La inteligencia artificial consiste en la creación de máquinas con capacidades iguales a las del ser humano a través de la exploración e interpretación de datos, con el fin de obtener conocimientos que permitan realizar actividades determinadas.

La IA es una disciplina en la que se ha venido pensando desde antes de Cristo, filósofos como Aristóteles idearon los primeros pasos para la IA. Al paso de unos años, científicos como Ramon Llull y Alan Turing llegaron a conclusiones importantes para el desarrollo de esta doctrina en las que determinaron el efectivo uso de máquinas, pero fue en 1943 cuando Warren McCulloch y Walter Pitts presentaron el primer trabajo de campo para esta disciplina denominado “El primer modelo de neuronas artificiales” Alan Turing fue quien dio paso a los primeros avances con la publicación de su artículo “Computing machinery and intelligence” en 1950 y la creación años más tarde del primer programa capaz de jugar ajedrez. El termino Inteligencia artificial fue creado por John McCarthy en 1958 quien invento el lenguaje de la inteligencia artificial LISP. Dado que la IA no daba los resultados esperados, estuvo en incógnita y con poca atención por muchos años, ya que requería tiempo y dinero que las personas no estaban dispuestas a invertir, por lo que fue a finales de los 80 y principio de los 90 en los que se retomaron las investigaciones y se produjeron nuevos avances como las redes ocultas de Markov, las redes probabilísticas, entre otras técnicas innovadoras que permitieron la mayor aplicación de esta ciencia hasta el tiempo presente, en el que las investigaciones avanzan cada vez más y se logra ver su utilidad en la industria y cualquier otro entorno.[30]

Ejemplos de aplicaciones de esta ciencia son: en las ciencias de la computación, donde se ha utilizado para resolver los problemas más complejos utilizando técnicas entre las cuales se encuentra la minería de datos, en las empresas financieras para la organización de sus operaciones, en la industria donde ha sido de gran utilidad para el desarrollo de actividades consideradas peligrosas para los humanos o de poca ergonomía. Existen en la actualidad infinidad de usos de la Inteligencia Artificial, lo que ha contribuido a la facilitación de diversas tareas para el hombre[31].

### **6.3.2 Machine learning**

El machine learning es una disciplina que es considerada una rama de la inteligencia artificial, la cual tiene como objetivo la invención de algoritmos que generen el aprendizaje automático de sistemas, con el fin de que estos realicen tareas sin ninguna instrucción determinada. Este se basa en el reconocimiento de datos y la predicción de comportamientos futuros de los mismos. Tiene extensas aplicaciones, por ejemplo en detección de fraudes en el uso de tarjetas de crédito, predecir fallas en equipos tecnológicos, seleccionar clientes potenciales, en diagnósticos médicos sobre enfermedades, minería de datos, etc. [32].

Este tipo de aprendizaje cuenta con diversos modelos de algoritmos, entre los cuales se encuentran el aprendizaje supervisado y el aprendizaje no supervisado.

#### **6.3.2.1 Aprendizaje supervisado**

Se basa en el aprendizaje con datos ya existentes, por lo que se necesita un entrenamiento del sistema para el reconocimiento y la predicción de datos[33]. Dentro de este tipo de aprendizaje se encuentran técnicas como:

##### **6.3.2.1.1 Regresión logística**

Consiste básicamente en el análisis del comportamiento de una variable dependiente sobre la variación de otra u otras independiente, con el fin de predecir la probabilidad de que ocurra un evento determinado[34]. Esta se usa en su gran mayoría en la medicina, por ejemplo, se ha utilizado para determinar el riesgo que tiene un paciente de padecer una enfermedad o no, si se va a presentar una muerte o no antes de alta, si se produce un reingreso o no, etc.[35]. Existen aplicaciones en otras áreas como en la psicología y las ciencias sociales para predecir una variable categórica dicotómica[36], en la educación para determinar factores determinantes en la demanda de estudios universitarios[37], entre muchas otras.

### 6.3.2.1.2 Árboles de decisión

Un árbol de decisión es una técnica de minería de datos que permite la clasificación, segmentación y predicciones de datos con el fin de analizar una serie de decisiones organizadas de manera secuencial y construidas lógicamente, basadas en reglas para resolver un problema y tomar la decisión más acertada. Estos son útiles en cualquier proceso que requiera toma de decisiones, por lo que son aplicados en diversas organizaciones de distintos campos, como la medicina, el marketing, finanzas y la industrial en general. [38] la representación gráfica de este modelo se presenta en la siguiente figura:

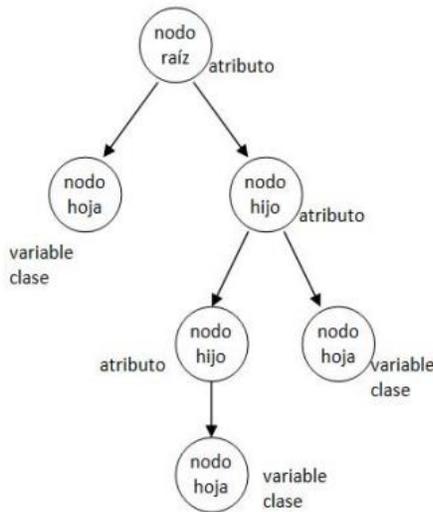


Figura 5. Estructura de un árbol de decisión.  
Fuente: Artículo original[39].

### 6.3.2.1.3 Random Forest

Es un algoritmo desarrollado por Leo Breiman, el cual se basa en la técnica de aprendizaje automático “árboles de decisión” para la creación de aleatoriedad mediante la combinación entre varios de estos con igual distribución, con el fin de posteriormente realizar un promedio y la predicción de nuevos datos[40]. (Ver figura 6).

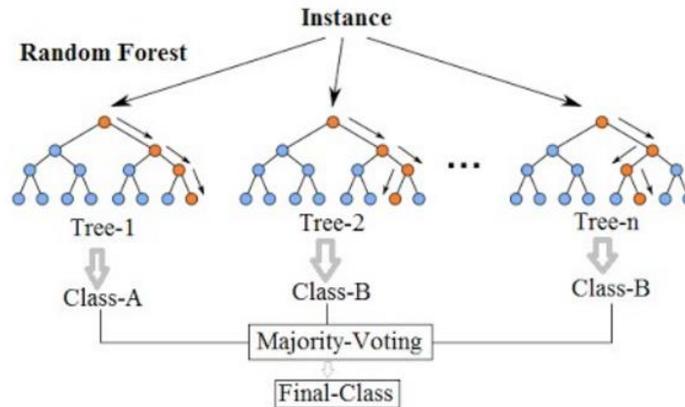


Figura 6. Esquema del algoritmo Random Forest  
Fuente: Random Forest Simple Explanation[41].

### 6.3.2.2 Aprendizaje no supervisado

En el aprendizaje no supervisado no se cuenta con un registro de datos iniciales, por lo que el algoritmo debe descifrar la información por sí mismo, detectando patrones que le permitan generar una clasificación de los datos basándose en sus atributos. Algunas técnicas de este tipo de aprendizaje son:

#### 6.3.2.2.1 Análisis de componentes principales.

El análisis de componentes principales (PCA) es un método multivariable que nos permite estudiar y evaluar un grupo de variables cuantitativas con el fin de reducir la dimensión de un conjunto de datos y a la vez que conservar la mayor variabilidad posible. Esto se logra por la identificación de direcciones llamadas componentes principales, dentro de las que la variación en los datos es máxima. Una de las ventajas es que tanto las observaciones como las variables pueden ser representadas gráficamente, haciendo posible evaluar visualmente similitudes y diferencias entre las muestras y determinar si es posible realizar procesos de agrupación.[42]

#### 6.3.2.2.2 Análisis de Clúster

El análisis de clúster es una técnica de aprendizaje no supervisado en el área de machine learning, el cual nos permite agrupar datos teniendo en cuenta la distancia que existe entre cada una de estas observaciones, por lo que los datos pertenecientes a un mismo clúster se encuentran a distancias muy cortas entre ellas y a una mayor distancia de los datos ubicados en otro clúster.

Existen dos métodos fundamentales para este tipo de análisis, el método jerárquico y el método no jerárquico.

### 6.3.3 Método jerárquico

Este tipo de método busca generar una jerarquía de grupos ya sea de manera ascendente o descendente.

#### Criterios de enlace:

Existen tres criterios de enlace utilizados para el modelo jerárquico:

- Enlace completo: Distancia máxima entre dos conjuntos
- Enlace simple: Distancia mínima entre dos conjuntos
- Enlace promedio: Distancia promedia entre dos conjuntos

### 6.3.4 Método no jerárquico

En el método no jerárquico el modelo más usado es el algoritmo de k – medias, el cual consiste en la partición de un conjunto de datos en k grupos teniendo como criterio el valor medio más cercano.

### 6.3.5 Calculo de Distancia

En el proceso de aprendizaje no supervisado es importante establecer una función de distancia entre los resultados de las instituciones de educación superior , debido a que entre más similares sean los resultados de dos instituciones de educación superior más cercanas van a ser. Para el cálculo de distancia en las técnicas de aprendizaje no supervisado existen los siguientes métodos en el software R:

1. Euclidiana
2. Máxima
3. Manhattan
4. Canberra
5. Binaria
6. Minkowski

Para nuestro estudio se utilizará la distancia Euclidiana debido a que fue la que mejor resultados nos brindó al momento de realizar el análisis, este método realiza el cálculo de distancia entre cada punto de la siguiente manera:

$$Distancia\ Euclidiana = \sqrt{\sum_{n=1}^N (x_{in} - x_{jn})^2}$$

## 6.4 Examen Saber Pro

Las pruebas Saber Pro son el examen de estado de la calidad de la educación superior, el cual es realizado por el instituto Colombiano para la evaluación de la educación (ICFES) cuyo objetivo es la aplicación de los módulos de competencias

genéricas y específicas para cada programa educativo con el fin de evaluar y proporcionar un reporte de desarrollo de las habilidades y conocimientos generales de estudiantes que han aprobado el 75% de los créditos de sus respectivos programas.[16]

### **¿Qué se evalúa en las pruebas Saber Pro?[16]**

El examen Saber Pro cuenta con una primera sesión obligatoria para todos los que presentan el examen, que se compone de 5 módulos que evalúan competencias genéricas:

- Lectura crítica;
- Razonamiento cuantitativo;
- Competencias ciudadanas;
- Comunicación escrita;
- Inglés.

El puntaje global de las pruebas Saber Pro se obtiene de la media simple de los resultados para cada competencia genérica, debido a que estas hacen parte de los módulos obligatorias en el examen [43].

## **6.5 Estado del arte**

La minería de datos es utilizada para el análisis y la caracterización de bases de datos a gran escala, por lo que puede ser útil dentro de cualquier ámbito. Esta ha sido aplicada en múltiples investigaciones y trabajos de campo, tanto de carácter cuantitativo como cualitativo, con el fin de obtener resultados más certeros y concretos sobre el objeto de estudio en cuestión.

A continuación se hace una recopilación de las investigaciones y/o artículos más relevantes dentro de la temática abordada. En primer lugar, se presentan los artículos de carácter cuantitativo, en los que se aplican diversas técnicas de análisis y minería de datos. Luego, se plantea la recopilación de investigaciones en las que se abordan las temáticas estudiadas cualitativamente.

### **6.5.1 Estudios de carácter cuantitativo:**

A través de la recopilación de información, se pudo evidenciar la aplicabilidad de la minería y análisis de datos (Ver tabla 3). En este caso, se abordaran, revisaran y clasificaran los estudios relacionados de acuerdo al campo de acción, tomando en cuenta la siguiente clasificación:

- Estudios dentro del campo educativo.
- Estudios fuera del campo educativo.

### **6.5.1.1 Dentro del campo educativo**

En relación con la minería de datos dentro del ámbito educativo, se realizó una aplicación de técnicas de minería de datos distribuida para la mejora de cursos e-learning, con el fin de aportar mejoras a los sistemas de aprendizaje online mediante la aplicabilidad de una metodología para el mantenimiento y análisis de los datos obtenidos, en la que se logró evidenciar una mejora sobre el rendimiento de los estudiantes debido a las modificaciones realizadas[44]; en otro estudio, se aplica minería de datos utilizando técnicas de “caja blanca” con el fin de predecir el fracaso escolar en estudiantes de media o secundaria, a través del diagnóstico de aquellos factores determinantes para que los estudiantes deserten o fracasen de las escuelas, con el objetivo de obtener una clasificación o un estándar de estudiantes que podrían ser propicios a esto, y de esta manera, tomar las acciones necesarias para evitarlo[45].

Unas de las técnicas más utilizadas son el análisis clúster y PCA, las cuales han sido aplicadas en el presente estudio para el análisis de los datos obtenidos. Dentro de algunos artículos relacionados con estas mismas técnicas en el ámbito educativo, se encuentra su aplicación para la clasificación de programas de ingeniería industrial acreditados en Colombia, estudio que tiene como objetivo conocer cuáles son aquellos aspectos en los que se relacionan o diferencian dichos programas y de qué manera, lo que contribuye al reconocimiento de los diferentes enfoques que existen para la ingeniería industrial en Colombia, este se encuentra relacionado de manera muy directa con el objeto de estudio actual, por lo que ha sido una importante referencia[42].

Entre otros estudios relacionados con la técnica de análisis clúster, se encuentra su aplicación para la caracterización de los diferentes perfiles de aprendizaje, tomando en cuenta una población estudiantil en la que se logran determinar estilos de aprendizaje dominantes, con el fin de adecuar a estos, los métodos de enseñanza existentes y lograr una mejora en el rendimiento estudiantil[46]. Del mismo modo, en el estudio de los procesos de internacionalización y globalización en la educación superior en los países OCDE; en este caso, utilizando esta técnica para el análisis del flujo estudiantil internacional de los países de la OCDE[47]. Se presenta también, una aplicación en el análisis de los perfiles motivacionales en una muestra de estudiantes adolescentes y su relación con la importancia de la educación física en secundaria, tomando en cuenta una serie de factores motivacionales, en el cual se obtuvo a través de esta técnica 3 perfiles diferenciados [48].

Gracias a estos estudios, se puede dimensionar la aplicabilidad que tiene la minería de datos en la educación, utilizando diversas técnicas en diferentes aspectos, por lo que se concluye que son técnicas bastante útiles para la presente investigación.

Para el análisis de datos a nivel educativo, no solo existe la minería de datos, por lo que se optó también por el reconocimiento de otras técnicas en el campo de estudio. En el desarrollo de estudios relacionados con la educación superior, se han aplicado técnicas de análisis envolvente de datos (DEA), para la evaluación de la eficiencia de las escuelas profesionales de ingeniería industrial en el Perú con enfoque de desarrollo sostenible y la medición de la eficiencia universitaria, estudio que se basa en la técnica DEA para determinar el grado de eficiencia con el que se manejan los recursos universitarios con el fin de lograr una mejor asignación de estos[49]. Otras técnicas presentes son las de frontera estocástica, aplicadas en este caso a un estudio para analizar la eficiencia de los programas y las instituciones de educación superior en Colombia, con el fin de conocer el estado actual de la misma[50].

### 6.5.1.2 Fuera del campo educativo

Con el fin de evidenciar la aplicación de minería de datos en diferentes entornos, se presentan estudios fuera del contexto educativo en los que estas técnicas han sido de gran utilidad. En primer lugar, en el sistema cubano de farmacovigilancia, aplicado con el fin de diseñar y desarrollar sistemas de información y bases de datos, con el objetivo de transformar la base de datos nacional[51]. Se aplicó también minería de datos, para la Identificación y detección de patrones delictivos basados en los homicidios dolosos que han sido cometidos en Argentina[52].

AÑO	ARTICULO	AUTOR	OBJETIVO	TÉCNICA
2018	Aplicación de Minería de Datos para la Clasificación de Programas Universitarios de Ingeniería Industrial Acreditados en Alta Calidad en Colombia. [42]	Tomás J Fontalvo-Herrera, Enrique J Delahoz, Adel A Mendoza-Mendoza	Identificar diferencias entre perfiles de programas de ingeniería industrial acreditados y cuáles son los factores principalmente influyentes.	Análisis De Componentes Principales (PCA) Y Análisis de clúster.
2007	Minería de datos para descubrir estilos de aprendizaje. [46]	Elena Durán-Rosanna Costaguta	Determinar características de los perfiles de aprendizaje en una población de estudiantes y obtener un perfil	Análisis clúster.

			de aprendizaje dominante.	
2002	Los procesos de internacionalización Y globalización en la educación superior: Un análisis de los países ocde.[47]	Sara fernández López-Emilio ruzo sanmartín.	Analizar el sistema de internacionalización, globalización y regionalización de la educación superior en los países OCDE.	Análisis clúster.
2007	Aplicación de la minería de datos al Sistema Cubano de Farmacovigilancia. [51]	Omar Calzadilla Fernández de Castro- Giset Jiménez López- Blanca Estela González Delgado- Jenny Ávila Pérez	Definir, diseñar y desarrollar los sistemas de tratamiento de la información y administrar la base de datos nacional "VigiBaseCuba".	CRISP-DM (CRoss-Industry Standard Process for Data Mining)
2008	Sistema recomendador colaborativo usando minería de datos distribuida para la mejora continua de cursos e-learning.[44]	Enrique García Salcines, Cristóbal Romero Morales, Sebastián Ventura Soto y Carlos de Castro Lozano.	Contribuir a la mejora continua del rendimiento de los sistemas de educación on-line.	Minería de datos distribuida
2012	La Medición de la Eficiencia Universitaria: Una Aplicación del Análisis Envlovente de Datos.[53]	Raquel Martín	Analizar la eficiencia técnica en los departamentos universitarios.	DEA.
2016	Modelo de análisis envolvente de datos (DEA) para evaluar la eficiencia de las escuelas profesionales de ingeniería industrial en el Perú con enfoque de desarrollo sostenible.[49]	Gloria Huamani-Salutiniano Huamani-Joaquin Salcedo- Cesar Fernandez-	Proponer un modelo DEA para evaluar la eficiencia de los programas de ingeniera industrial en el Perú con enfoque a	DEA

			desarrollo sostenible.	
2012	Predicción del Fracaso Escolar mediante técnicas de Minería de Datos.[45]	Carlos Márquez Vera, Cristóbal Romero Morales y Sebastián Ventura Soto.	Aplicar minería de datos para conocer cuáles son las causas más influyentes por las que los estudiantes de media o secundaria fracasan.	Minería de datos.
2012	Análisis de los perfiles motivacionales y su relación con la importancia de la educación física en secundaria.[48]	Manuel Gómez López- Antonio Granero Gallegos- Antonio Baena Extremera y J. Arturo Abrales.	Basados en una muestra de estudiantes de educación secundaria, lograr la identificación de los diferentes perfiles motivacionales en función de la importancia de la educación física.	Análisis clúster.
2007	Identificación y detección de patrones delictivos basada en minería de datos. [52]	Perversi, I. - Valenga, F. - Fernández, E. - Britos, Paola Verónica- García Martínez, Ramón	Presentar resultados obtenidos de un estudio de identificación y detección de patrones delictivos en argentina mediante minería de datos.	Minería de datos.
2014	La Educación Superior en Colombia: Situación Actual y Análisis de Eficiencia.[50]	Ligia Alba Melo B. Jorge Enrique Ramos F. Pedro Oswaldo Hernández S.	Analizar de manera global el estado de la educación superior en Colombia, realizando una evaluación de la	Técnicas de frontera estocástica.

			eficiencia de los programas y las instituciones del país.	
--	--	--	---	--

Tabla 3. Estado del arte de investigaciones con enfoque cuantitativo.

### 6.5.2 Estudios de carácter cualitativo:

Existen diferentes investigaciones respecto a la educación superior englobando diferentes aspectos, al igual que de la aplicabilidad de la minería de datos, por lo que se analizaran respectivamente (Ver tabla 4).

#### 6.5.2.1 Educación superior

Inicialmente se abordaran distintas temáticas relacionadas con el objeto de estudio. En primer lugar, la acreditación; a través de análisis cualitativo y entrevistas, se logró determinar la percepción de un grupo de personas sobre la acreditación en calidad de los programas educativos en México y su impacto sobre la gestión académica, concluyendo así, que la acreditación no necesariamente contribuye a una mejora sobre la calidad educativa, ya que esta debe ir en paralelo con las políticas institucionales con el fin de un óptimo aprovechamiento de las herramientas de gestión que esta brinda[54]. Del mismo modo en este país, se realizó una investigación sobre los procesos de acreditación en calidad en las IES, con el fin de conocer a fondo los parámetros y procedimientos que llevan a cabo sus entes acreditadores[55].

El proceso de acreditación presenta variaciones, razón por la cual se realizó un estudio en los diferentes países de Centroamérica, con el fin de efectuar una comparación entre estos procesos, conocerlos más a fondo y determinar las diferencias y similitudes en cada uno. Esto, en el artículo “La calidad de la educación superior y su acreditación: la experiencia centroamericana”[56].

Por otro lado, las certificaciones y acreditaciones en calidad se han vuelto una necesidad para las instituciones de educación superior, ya que deben ser competitivas en el mercado y de esta manera contar con reconocimiento en calidad, pero existen algunos efectos negativos que se producen en este proceso, generando resultados no deseados para algunos, lo que es abordado a mayor profundidad en el artículo “La acreditación y certificación en las instituciones de educación superior. Hacia la conformación de circuitos académicos de calidad: ¿Exclusión o Integración?”[57].

En segunda instancia, la calidad en la educación superior. Esta cuenta con muchas investigaciones, las cuales tienen como finalidad encontrar parámetros y factores determinantes sobre la misma, un ejemplo claro es el artículo “Calidad de la

educación superior: un análisis de sus principales determinantes”[58]; al igual que el artículo “Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde la calidad de la educación superior pública” que de igual manera busca determinar factores influyentes en calidad, en este caso, enfocado al rendimiento de los estudiantes en las IES públicas[59].

### 6.5.2.2 Minería de datos

Dentro del contexto de las técnicas a aplicar en el presente estudio, se logran encontrar investigaciones relevantes para el reconocimiento de las mismas y su aplicación. En este caso se presentan tres artículos referentes, en primer lugar la investigación sobre el estado actual de la aplicación de minería de datos a los sistemas de enseñanza basada en web, que tiene como objetivo conocer la tarea de estas técnicas sobre los sistemas e-learning y la importancia de las mismas[60], también “Minería de datos educativa: Una herramienta para la investigación de patrones de aprendizaje sobre un contexto educativo” donde se presenta un análisis de la importancia que representa esta herramienta sobre el estudio y análisis de datos e información a nivel educativo[61].

De igual manera, fuera del ámbito educativo se presenta una investigación realizada para la identificación de la utilidad que representa la aplicación de las técnicas de minería de datos sobre el marketing en las organizaciones llamado “Aplicaciones de Minería de Datos en Marketing”[62], que tiene como objetivo determinar de qué manera la implementación de estas técnicas podría ayudar a la mejora de la calidad y de la fidelización del cliente a través de la toma de mejores decisiones gracias a la información obtenida.

AÑO	ARTICULO	AUTOR	OBJETIVO	TÉCNICA
2013	La acreditación de la calidad educativa y la percepción de su impacto en la gestión académica: el caso de una institución del sector no universitario en México. [54]	Iván Alejandro Salas Durazo.	Determinar la percepción del impacto de la acreditación en calidad en una IES sobre el sistema de gestión educativo con el que ya cuentan.	Análisis cualitativo y Entrevistas.
2008	La calidad de la educación superior y su acreditación: la experiencia centroamericana. [56]	Carlos Tünnermann Bernheim	Realizar una evaluación y análisis de los procesos de acreditación de Centroamérica.	Investigativo

2014	Calidad de la educación superior: un análisis de sus principales determinantes. [58]	Diana Lago de Vergara-Audin Aloiso Gamoba Suárez-Alexander Javier Montes Miranda	Analizar de acuerdo a factores determinados, la influencia que tienen estos en la calidad en la educación superior.	Investigativo
2007	Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde la calidad de la educación superior pública. [59]	Guiselle María Garbanzo Vargas	Analizar y determinar factores influyentes en el rendimiento de los estudiantes en la educación superior pública desde la investigación.	Investigativo
2006	La acreditación y certificación en las instituciones de educación superior. Hacia la conformación de circuitos académicos de calidad: ¿Exclusión o Integración?[57]	Alma Rosa Hernández Mondragón	Análisis de los efectos de la necesidad de acreditación y certificación en la educación superior.	Investigativo
2011	Organismos y procesos de certificación de calidad de la educación superior privada en México. [55]	Yazmín Cuevas	Conocer y evaluar los procedimientos de los entes certificadores en calidad de la IES en México.	Investigativo
2005	Estado actual de la aplicación de la minería de datos a los Sistemas de enseñanza basada en web. [60]	Cristóbal Romero Morales, Sebastián Ventura Soto, Cesar Hervás Martínez	Dar a conocer el estado de la investigación acerca de la aplicación de minería de datos a los sistemas de enseñanza por web.	Investigativo
2013	Minería de datos educativa: Una herramienta para	Alejandro Ballesteros	Dar a conocer un estudio	Investigativo

	la investigación de patrones de aprendizaje sobre un contexto educativo.[61]	Román, Daniel Sánchez-Guzmán, Ricardo García Salcedo	acerca de los métodos y técnicas de Minería de Datos Educativa.	
2016	Aplicaciones de Minería de Datos en Marketing. [62]	H. Escobar, M. Alcivar, A. Puris.	Presentar un análisis de la utilidad de la aplicación de técnicas de minería de datos en las organizaciones.	Investigativo

Tabla 4. Estado del arte de investigaciones con enfoque cualitativo.

## **7 METODOLOGIA DE LA INVESTIGACIÓN**

A continuación, se pretende realizar una descripción de la metodología del presente estudio, dando a conocer el tipo de investigación, el diseño, instrumentos, técnicas y procedimientos utilizados durante el desarrollo del mismo.

### **7.1 Enfoque de la investigación**

El objetivo principal de este proyecto de investigación, es dar respuesta a través un modelo de aprendizaje no supervisado de datos a la problemática particular planteada inicialmente, la cual tiene como foco central conocer la influencia que tienen las certificaciones y acreditaciones de calidad sobre los resultados de las pruebas Saber Pro en el programa de ingeniería industrial, esto desde un punto de vista cuantitativo, tomando como base datos existentes para su análisis a través de la aplicación de técnicas de minería de datos. Por lo que este estudio presenta un enfoque metodológico empírico analítico, ya que busca a través de la observación y experimentación realizar análisis cuantitativos de los datos obtenidos para verificar o rechazar una hipótesis[63].

### **7.2 Tipo de investigación**

La presente investigación es de tipo aplicativo, ya que se busca inferir a través de un modelo estadístico basado en minería de datos, la relación que existe entre los resultados de las pruebas Saber Pro en los programas de ingeniería industrial en Colombia y las certificaciones y acreditaciones de calidad con las que cuentan estos o en su defecto las instituciones de educación superior correspondientes.

### **7.3 Diseño de la investigación**

La investigación se encuentra estructurada por dos etapas principales:

- La recolección, observación y clasificación de los datos.
- El análisis y estudio de los datos obtenidos mediante las técnicas de minería de datos y posteriormente interpretación y generación de conocimiento.

Par la obtención de los datos de los resultados de las pruebas Saber Pro en el programa de ingeniería industrial a nivel nacional, se realizó una búsqueda en la base de datos del ICFES, donde posteriormente se efectuó una depuración y clasificación de los mismos con el fin de obtener la información más relevante para la investigación. Luego, se procedió a la aplicación de técnicas de minería de datos como PCA y análisis clúster con el fin de realizar el análisis y obtener los resultados para determinar si la hipótesis planteada es verdadera o falsa. A demás, se utilizaron técnicas investigativas como las entrevistas a profundidad, las cuales se

le realizaron a directivos o personas influyentes sobre el programa de ingeniería industrial en la universidad tecnológica de bolívar.

## 8 ENTREVISTAS

Para el análisis del objeto de estudio del presente proyecto de investigación, se utilizaron herramientas con enfoque cualitativo, en este caso, entrevistas a profundidad. Estas se realizaron en la Universidad tecnológica de bolívar, en dos departamentos diferentes: el departamento de planeación y el departamento del programa de ingeniería industrial, en los cuales se eligió 1 persona por cada uno para la ejecución de la entrevista. Llevada a cabo la citación con las personas, se procedió a la realización de las preguntas con el fin de obtener información importante para la investigación. En la ficha técnica plasmada a continuación se pueden apreciar los aspectos generales de dichas entrevistas.

### 8.1 Ficha técnica.

Participantes	Directivos o responsables relacionados con el rendimiento académico de los estudiantes en el programa de ingeniería industrial. <ul style="list-style-type: none"><li>• Entrevistado 1: Persona vinculada con el departamento de planeación de la UTB.</li><li>• Entrevistado 2: Persona influyente dentro del programa de ingeniería industrial de la UTB.</li></ul>
Objetivo	Conocer la percepción de estas personas acerca de la influencia de las certificaciones y acreditaciones de calidad sobre los resultados de los estudiantes en las pruebas Saber Pro en el programa de ingeniería industrial.
Justificación	Obtener información relevante para el análisis del estudio que se quiere realizar.
Fecha de aplicación	Enero del 2019.
Técnica de recolección	Entrevista a profundidad.

Tabla 5. Ficha técnica de entrevistas a profundidad.

### 8.2 Preguntas:

1. ¿Considera usted que es importante medir el desempeño de los estudiantes?
2. ¿Considera usted pertinente la medición que se efectúa a través de las pruebas Saber Pro?
3. ¿Considera usted que si una institución posee certificaciones y acreditaciones en calidad produce estudiantes con mejor desempeño académico que una institución que no cuente con estas?
4. ¿Qué tan relevante encuentra usted para el estudiante el realizar las pruebas Saber Pro?

5. ¿Considera usted que el resultado de las pruebas Saber Pro es determinante para evaluar el nivel académico de los estudiantes y por ende el de la institución?
6. ¿Considera usted que una institución que no cuente con certificaciones y acreditaciones de calidad brinda un nivel de educación bajo?
7. ¿Considera que el programa de ingeniería industrial de la UTB cumple con todos los estándares y otorga a los estudiantes una formación de calidad?
8. ¿Considera usted que las certificaciones y acreditaciones con las que cuenta la UTB actualmente impactan sobre el resultado de los estudiantes en las pruebas Saber Pro?
9. ¿Considera usted relevante que la UTB obtenga otras certificaciones y acreditaciones de calidad para mejorar el desempeño de sus estudiantes?
10. ¿Cuál es su percepción acerca del impacto que tienen las certificaciones y acreditaciones de calidad sobre el resultado de las pruebas Saber Pro?

### **8.3 Resultados**

La presentación de los resultados de la encuesta se llevara a cabo mediante la agrupación de preguntas relacionadas entre sí, dando paso al análisis de las respuestas de las personas entrevistadas.

#### **Preguntas 1-2-4-5**

**Objetivo:** Conocer la consideración que se tiene sobre el proceso de medición que se efectúa a través de las pruebas Saber Pro.

**Justificación:** La educación ha evolucionado y los procesos de medición tradicionales que realiza el estado aún se consideran relevantes para la medición del desempeño de las instituciones y de los estudiantes, por lo que se desea saber si estas personas están de acuerdo o no con esto.

Los entrevistados coinciden en que el desempeño de los estudiantes debe ser medido, ya que el servicio educativo debe ser visto como un proceso donde se miden unas entradas y unas salidas. En este caso, el estado en el que son entregados los profesionales a la sociedad respecto a las competencias que fueron adquiridas por el estudiante en la institución de educación secundaria y al proceso de aprendizaje que se le brinda durante la etapa de educación universitaria, esto con el fin del aseguramiento de la calidad de las competencias de los egresados y de la eficiencia del servicio prestado. A pesar de esto, consideran que a través de las pruebas Saber Pro no se puede medir la formación integral impartida, ya que se están midiendo conocimientos técnicos específicos de una manera genérica, por lo que no se puede afirmar de manera tajante que se puede verificar a través de esta el desempeño del estudiante como futuro egresado. Encuentran también irrelevante las pruebas para el estudiante, ya que no se ha hecho una apropiación del por qué

estas se realizan, lo que conlleva al desinterés por parte de estos sobre las mismas. En conclusión, consideran que se debe medir el desempeño pero teniendo en cuenta otras fuentes, ya que las pruebas Saber Pro no pueden ser las únicas determinantes para la evaluación de este.

### **Preguntas 3-6**

**Objetivo:** Conocer la percepción del impacto de las certificaciones y acreditaciones en calidad sobre la calidad del servicio educativo.

**Justificación:** Las certificaciones y acreditaciones de calidad son vistas por las instituciones de educación superior como un objetivo por alcanzar, con el fin de mejorar la calidad de sus procesos y su formación académica, por lo que se desea conocer la percepción sobre la influencia de estas en las instituciones de educación superior.

El entrevistado 1 considera que no se puede afirmar que si una institución cuenta con certificaciones y acreditaciones de calidad esto se retribuye en el desempeño de un egresado, ya que se deben tener en cuenta varios factores del estudiante como persona, por lo que no quiere decir que el desempeño del estudiante egresado de una universidad acreditada va a ser mayor que el de una que no lo esté o viceversa ya que este está amarrado a muchas variables. Para el entrevistado, las certificaciones y acreditaciones sirven como un respaldo de calidad ante la sociedad, garantizando una alta calidad en los procesos internos y un buen nivel competitivo. El entrevistado 2 por su parte difiere un poco de esta idea, ya que considera que el hecho de que una institución de educación superior cuente con certificaciones y acreditaciones en calidad, debería otorgar a los estudiantes mejores condiciones a nivel educativo que promuevan un mejor desempeño que el de las instituciones que no cuenten con estas.

### **Pregunta 7-8-9**

**Objetivo:** Conocer la apreciación que se tiene acerca de las certificaciones y acreditaciones de calidad con las que cuenta la UTB y el programa de ingeniería industrial sobre la formación académica y el resultado en las pruebas Saber Pro.

#### **Justificación:**

La Universidad Tecnológica de Bolívar cuenta con acreditación institucional y su programa de ingeniería industrial se encuentra doblemente acreditado, lo que es considerado una ventaja competitiva frente a otras instituciones de educación superior ya que esto sugiere que el servicio prestado por la UTB es de alta calidad y produce egresados integrales y de buen desempeño, por esto se busca conocer la opinión de las personas respecto a esto.

Para el entrevistado 1 las certificaciones y acreditaciones podrían afinar algunos procesos internos que la universidad desarrolla como competencias de razonamiento escrito, matemática y lectoescritura, pero no es un parámetro decisivo sobre el desempeño del egresado en las pruebas Saber Pro, por lo que el hecho de que la universidad cuente con acreditación institucional, no va a garantizar un excelente desempeño de todos sus egresados sobre estas. El entrevistado 2 tiene otra concepción sobre esto, ya que considera que de alguna manera estas certificaciones y acreditaciones impactan sobre los resultados, siendo estos superiores a los de instituciones que no cuenten con estas.

Coinciden en la idea de que el programa de ingeniería industrial otorga a sus estudiantes una formación de calidad, que se ha visto reflejada en un desempeño adecuado de los egresados en el medio, otorgando a estos cierta preferencia sobre egresados de otras instituciones. De igual manera están de acuerdo en que se debe estar siempre en mejora continua, por lo que consideran relevante que la UTB obtenga otras certificaciones y acreditaciones que le permitan estar a la vanguardia, pero no con el fin de mejorar el desempeño ya que esto no lo garantiza.

#### **Pregunta 10.**

**Objetivo:** Conocer la percepción general del impacto de las certificaciones y acreditaciones de calidad sobre los resultados de las pruebas Saber Pro.

**Justificación:** El objetivo principal de este estudio se basa en determinar si las certificaciones y acreditaciones de calidad influyen sobre los resultados de las pruebas Saber Pro en el programa de ingeniería industrial, por lo que es importante conocer la percepción que tienen estas personas sobre esta temática.

En general, las consideraciones coinciden en que no tiene una incidencia directa, en el que se pueda afirmar que el contar con certificaciones y acreditaciones va a garantizar que los resultados de las pruebas Saber Pro sean excelentes, ya que el estudiante se ve afecto por diferentes variables, pero si sirve para detectar falencias en el sistema educativo y contribuir a la mejora de estas, lo que aporta a la evolución de la formación académica.

## 9 ANÁLISIS Y RESULTADOS

El objetivo principal de este trabajo fue identificar el impacto de la calidad en el desempeño de las instituciones de educación superior en las pruebas “Saber Pro” además de generar visualizaciones que nos permitan un fácil análisis e interpretación de los resultados obtenidos por el modelo. Los datos a usar en el estudio fueron obtenidos de la base de datos oficial del ICFES, para los resultados de las pruebas “Saber Pro” de los años 2016 y 2017, teniendo en cuenta tanto las competencias genéricas como específicas. Los módulos evaluados en cada competencia son los siguientes:

<b>Competencias genéricas</b>	<b>Competencias específicas</b>
<b>Comunicación escrita</b>	<b>Formulación de proyectos de ingeniería</b>
<b>Razonamiento cuantitativo</b>	<b>Diseño de sistemas logísticos y productivos</b>
<b>Inglés</b>	<b>Pensamiento científico matemática y estadística</b>
<b>Competencia ciudadana</b>	
<b>Lectura crítica</b>	

Tabla 6. Competencias genéricas y específicas evaluadas en las pruebas Saber Pro

La metodología utilizada para desarrollar el análisis de la investigación se encuentra representado en la siguiente figura:

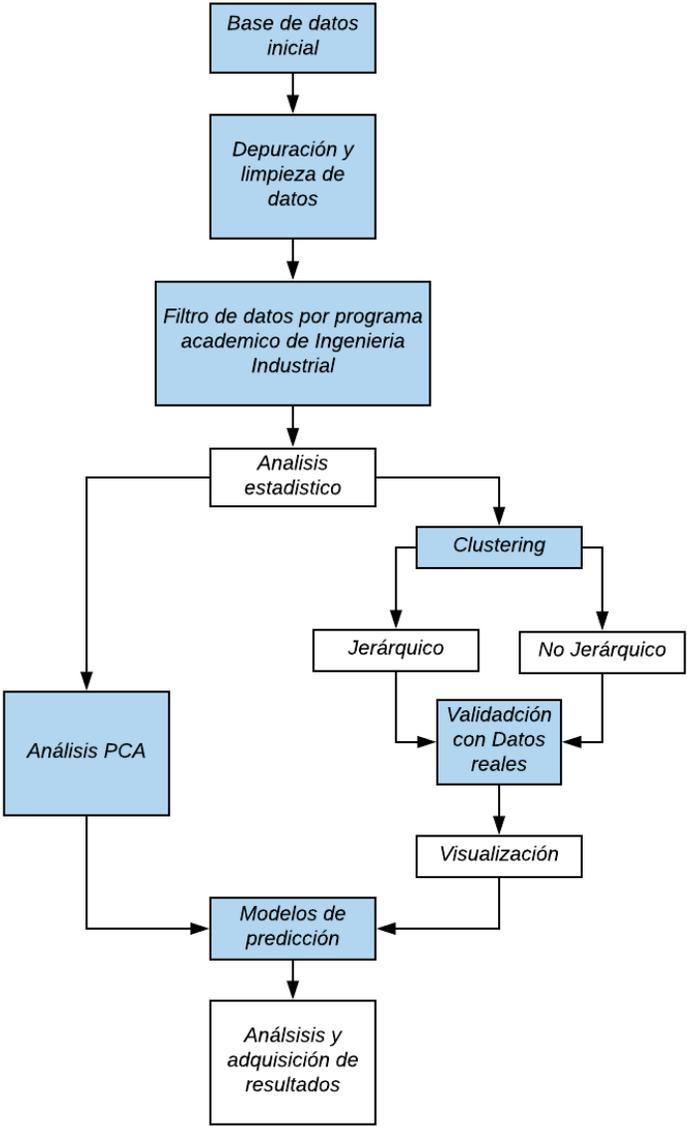


Figura 7: Metodología utilizada para el desarrollo del análisis.

Para el análisis del impacto de las certificaciones de calidad sobre los resultados en las pruebas se tuvo como foco de estudio los programas de Ingeniería Industrial a nivel nacional, por lo que se procedió a realizar una depuración de los datos en la cual nos encontramos con algunas inconsistencias en variables de importancia para el estudio, como lo son el nombre de la institución de educación superior y el programa educativo, teniendo cambios en la escritura en las bases de datos del año 2016 y el año 2017, por lo que previo al análisis se realizó una limpieza de estos, unificando los datos para un mejor manejo en el modelo y poder tener un modelo

que represente la realidad lo más preciso posible. Posterior a la depuración y limpieza de los datos se realizaron una serie de filtros de la información para tener solamente los datos de relevancia para el estudio, en primera instancia se realizó un filtro por el programa de Ingeniería Industrial, para el cual se encontró que 23.964 personas de las 488.222 que realizaron la prueba en 2016 y 2017 pertenecen al programa de Ingeniería Industrial, lo cual representa un 4,90% de los resultados. Posterior a este filtro se procedió a buscar los resultados de los módulos de competencias específicas de dichas personas pertenecientes al programa de Ingeniería Industrial, ya que los resultados de las competencias genéricas y específicas se encontraban en bases de datos diferentes y poseían formatos diferentes por lo que se tuvo que usar como variable de búsqueda el código asignado por el ICFES para cada estudiante, como último paso se agrupó toda la data en una sola base donde se tuvieran los resultados de todos los módulos tanto de competencias genéricas como específicas por lo que el número de personas con las que realizaría el análisis bajó a 22.714 personas, debido a que algunas personas no tenían resultados en todos los módulos.

Teniendo esta información, se procedió por medio del software R Studio [64] a organizar en una tabla los datos, en la cual se encontrarán la media de los resultados de cada uno de los módulos de competencias tanto genéricas como específicas para cada institución de educación superior que cuente con programas de ingeniería industrial, teniendo como premisa que los resultados proporcionados por el ICFES se encuentran en cifras enteras, se redondearon las medias de cada módulo a números enteros para tener uniformidad en los datos iniciales y los utilizados en el análisis. En la tabla realizada se encontraron 87 instituciones de educación superior que cuentan con resultados en las pruebas “Saber Pro”, las cuales se encuentran distribuidas en 20 de los 32 departamentos del territorio nacional.

Inicialmente se realizó un análisis de componentes principales para identificar que variables nos brindaban una mayor variabilidad a nuestro modelo, teniendo en cuenta que dichas variables son las medias de cada módulo de las pruebas para cada institución de educación superior. Del análisis PCA, se encontró que los datos cuentan con 8 componentes principales, donde el porcentaje de varianza acumulado de las 3 primeras componentes es 88,68%.

Los resultados de desviación estándar, porcentaje de varianza y su porcentaje acumulado para cada componente principal se pueden apreciar en la siguiente tabla:

Componente principal	Desv estándar	% de varianza	% acumulado
1	2,4713	76,34%	76,34%
2	0,7304	6,67%	83,01%
3	0,6732	5,67%	88,68%
4	0,5287	3,49%	92,17%
5	0,4671	2,73%	94,90%
6	0,4030	2,03%	96,93%
7	0,3887	1,89%	98,82%
8	0,3076	1,18%	100,00%

Tabla 7: Resultados de PCA para cada componente.

El análisis de componentes principales también es muy útil al momento de representar nuestros resultados visualmente, por lo que para determinar el número de componentes que genere una mejor visualización de los resultados se usó el scree test el cual se basa en generar una graficar de la desviación estándar de cada componente y buscar el punto de quiebre natural en los datos (ver figura 8), de esta manera se seleccionan 2 componentes principales para continuar la investigación.

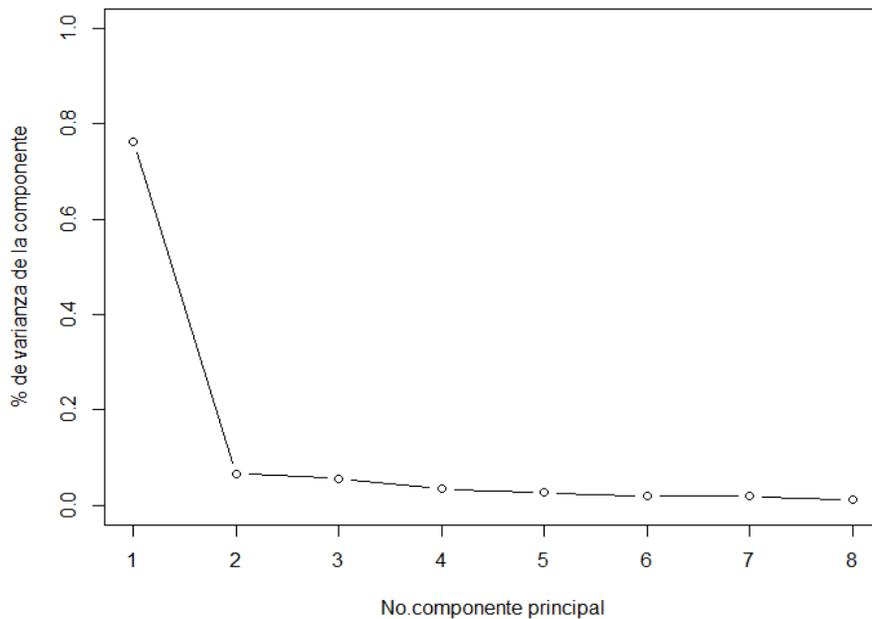


Figura 8. Scree test para el análisis PCA.

Tomando las componentes 1 y 2 procedemos a identificar que variables de nuestros datos se encuentra mayormente representadas en dichas componentes principales, por lo que se genera la siguiente tabla donde se encuentra la variabilidad explicada para cada variable en cada una de las componentes principales de estudio:

<b>Variable</b>	<b>Componente 1</b>	<b>Componente 2</b>
<b>Inglés</b>	0,35	-0,22
<b>Comunicación escrita</b>	0,30	-0,78
<b>Razonamiento cuantitativo</b>	0,36	-0,14
<b>Competencia ciudadana</b>	0,37	0,20
<b>Lectura crítica</b>	0,37	0,08
<b>Diseño de sistemas log y prod</b>	0,36	0,24
<b>Formulación de proyectos</b>	0,34	0,46
<b>Pensamiento científico</b>	0,37	0,05

Tabla 8: Variabilidad explicada para las componentes 1 y 2.

En la componente 1 se puede observar como los módulos de competencia ciudadana, lectura crítica y pensamiento científico son las que mayormente representa la variabilidad explicada de la componente, con valores de 0,37 respectivamente, sin embargo variables como los módulos de razonamiento cuantitativo y diseño de sistemas logísticos y productivos, e inglés que también cuentan con una variabilidad explicada cercana a la poseen las variables con mayor peso. Esta dimensión se puede interpretar como el puntaje global de la prueba.

En la componente 2 se evidencia que los módulos de comunicación escrita y formulación de proyectos de ingeniería son las variables que mejor representan la variabilidad explicada de la componente con valores de -0,78 y 0,46 respectivamente, generando una división entre los datos de instituciones con buen desempeño en comunicación escrita y formulación de proyecto, como se puede apreciar en la figura 9. Este análisis de componentes principales nos va a ayudar mucho al momento de realizar las visualizaciones de los resultados de los modelos de agrupamiento no jerárquico.

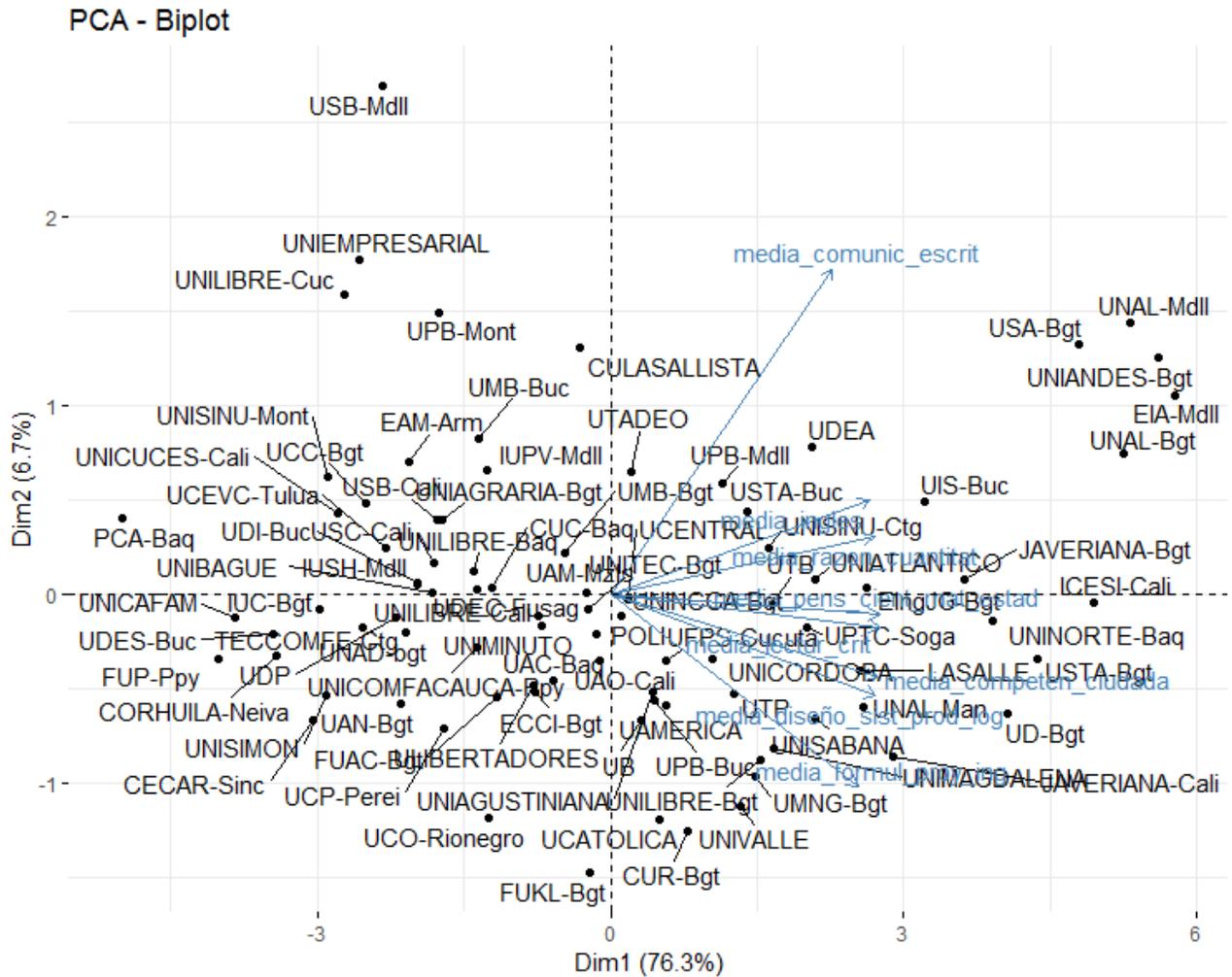


Figura 9. PCA Biplot

La anterior figura es la visualización de las instituciones en un plano formado por las componentes principales (Dimensiones) 1 y 2 las cuales logran explicar el 83,01% de la variabilidad de los datos, cada uno de los puntos en el plano corresponde a una instituciones de educación superior y se encuentran distribuidas de acuerdo al peso específico de cada una en los ejes de las dimensiones 1 y 2.

### 9.1 Análisis estadístico de datos

Como segunda instancia se realizaron ciertos análisis estadísticos de los datos teniendo en cuenta los puntajes globales con el fin de unificar los resultados en una sola variable.

Teniendo en cuenta que nuestro interrogante se centra en el impacto de los sistemas de calidad en la educación superior sobre los resultados obtenidos en las

pruebas Saber Pro se realizó un análisis estadístico con el fin de determinar si existe una diferencia entre las medias globales de las instituciones acreditadas y no acreditadas.

Por medio del software R Studio [64] se realizó una prueba T-Student para identificar si existe igualdad entre las medias de las instituciones acreditadas y no acreditadas, obteniendo los resultados que se presentan en la siguiente tabla:

<b>P-valor</b>	<b>2,35 E-09</b>
<b>Medias</b>	
<b>No acreditadas</b>	<b>149</b>
<b>Acreditadas</b>	<b>170</b>

Tabla 9. Resultados de prueba T-Student para instituciones acreditadas y no acreditadas.

Como se puede apreciar en la tabla 9 existe una diferencia de 21 puntos entre las medias de los resultados obtenidos por las instituciones acreditadas y no acreditadas, adicionalmente se cuenta con un p-valor menor a 0,05 por lo que se puede considerar que dicha diferencia es significativa desde un punto de vista estadístico.

Adicionalmente a la hipótesis planteada anteriormente, se quisieron plantear dos hipótesis más, la primera es si las instituciones de carácter privado que cuentan con acreditación institucional poseen un mejor desempeño que las instituciones públicas que poseen acreditación institucional, en las pruebas Saber Pro, con el fin de identificar si por medio de los resultados se puede evidenciar una diferencia en la educación brindada. Los resultados obtenidos de la prueba T-Student realizada se presentan en la siguiente tabla:

<b>P-valor</b>	<b>0,6261</b>
<b>Medias</b>	
<b>Publicas</b>	<b>168</b>
<b>Privadas</b>	<b>171</b>

Tabla 10. Resultados de prueba T-Student para instituciones públicas y privadas.

Como se puede apreciar en la tabla 10 existe una diferencia de 9 puntos entre las medias de los resultados obtenidos por las instituciones públicas y privadas, adicionalmente se cuenta con un p-valor mayor a 0,05 por lo que se puede considerar que dicha diferencia no es significativa desde un punto de vista estadístico.

Por otro lado se planteó la hipótesis de si las instituciones de la costa caribe poseen un mejor desempeño a las instituciones del resto del país, y si no saber qué tan alejados se encuentran las medias de estas dos poblaciones.

Los resultados obtenidos por la prueba T-Student realizada para las medias de las instituciones de la costa caribe y del resto del país se aprecian en la siguiente tabla.

<b>P-valor</b>	<b>0,3487</b>
<b>Medias</b>	
<b>Resto del país</b>	<b>154</b>
<b>Costa caribe</b>	<b>150</b>

Tabla 11. Resultados de prueba T-Student para instituciones de la costa caribe y del resto del país.

Como se puede apreciar en la tabla 11 existe una diferencia de 4 puntos entre las medias de los resultados obtenidos por las instituciones públicas y privadas, adicionalmente se cuenta con un p-valor mayor a 0,05 por lo que no se puede considerar que dicha diferencia es significativa desde un punto de vista estadístico.

## 9.2 Proceso de clasificación

Como paso siguiente se generó un proceso de clasificación de las instituciones por medio de un análisis de clúster que hace parte a la rama del aprendizaje no supervisado, usando las técnicas de agrupamiento jerárquico y no jerárquico para identificar si teniendo como base del análisis las medias de los resultados obtenidos por cada institución de educación superior es posible agrupar aquellas instituciones que poseen certificaciones de calidad y aquellas que no cuentan con dicha certificación. Luego de tener estas clasificaciones se comparó cada técnica de agrupamiento con la clasificación real de las instituciones acreditadas y no acreditadas, para validar cuál de los dos métodos representa mejor la información real de las instituciones de educación superior.

Luego de consultar las páginas web del CNA [65] y del icetex [66] para encontrar el número de instituciones de educación superior con programas de ingeniería industrial a nivel nacional que cuenten con acreditaciones, se encontró que 23 instituciones de las 87 que presentaron el examen entre 2016 y 2017 cuentan con acreditación institucional.

Para la comparación de los modelos usados con la asignación real de las instituciones de educación superior se tuvieron en cuenta dos variables, el accuracy (precisión del modelo) del modelo del modelo jerárquico con las variables escaladas y sin escalar, así como el modelo no jerárquico con las variables escaladas y sin

escalar, teniendo en resumen 4 modelos diferentes para comparar con la asignación real y definir qué modelo representa mejor el comportamiento real de la variable que responde a si la institución cuenta con acreditación o no. Un resumen de lo mencionado anteriormente se puede evidenciar a continuación:

Cluster jerárquico sin escalar variables	Cluster no jerárquico sin escalar variables	Cluster jerárquico con variables escaladas	Cluster no jerárquico con variables escaladas
<b>Accuracy</b>	<b>Accuracy</b>	<b>Accuracy</b>	<b>Accuracy</b>
81,61%	78,16%	<b>83,91%</b>	<b>79,31%</b>

Tabla 12. Accuracy de los diferentes modelos de clúster utilizados.

Como se puede apreciar en la tabla anterior los dos modelos que mejor representan la asignación real de las certificaciones de calidad son el modelo jerárquico y no jerárquico con las variables escaladas ya que estos dos modelos poseen una precisión superior al de los modelos donde se usan las variables sin ser escalar previamente.

Los resultados obtenidos para los modelos jerárquicos y no jerárquicos con las variables escaladas fueron los siguientes:

### 9.2.1 Clúster jerárquico

Usando la distancia euclidiana y el método “complete” se obtuvo la clasificación representada en el siguiente dendrograma:



Figura 10. Dendrograma obtenido por el cluster jerárquico

El resultado del análisis jerárquico nos brinda de manera visual la existencia de dos grupos, por un lado tenemos el grupo de las instituciones que cuentan con certificaciones de alta calidad, las cuales se identifican con el color azul, y se encuentran 23 de las 87 instituciones de educación superior que cuentan con resultados en los años 2016 y 2017, y a su vez tenemos 64 instituciones que no poseen certificación con base a la clasificación del modelo jerárquico.

Luego del análisis jerárquico se procedió a identificar los perfiles de las instituciones de educación superior clasificadas como acreditadas y no acreditadas teniendo en cuenta los resultados de las medias por cada competencia. Las instituciones de educación superior que se clasifican como acreditadas cuentan con un buen desempeño en las competencias de inglés, razonamiento cuantitativo y lectura crítica especialmente con medias de 182, 187 y 176 puntos como se puede observar en la tabla 10, además es bueno mencionar que de manera general las instituciones de educación superior clasificadas por el análisis jerárquico como instituciones acreditadas cuentan con un mejor resultado en cada competencia de la prueba cuando los comparamos con los resultados de aquellas instituciones que no se clasificaron como acreditadas.

Luego del análisis jerárquico se procedió a identificar los perfiles de las instituciones de educación superior clasificadas como acreditadas y no acreditadas teniendo en cuenta los resultados de las medias por cada competencia. Las instituciones de educación superior que se clasifican como acreditadas cuentan con un buen desempeño en las competencias de inglés, razonamiento cuantitativo y lectura crítica especialmente con medias de 182, 187 y 176 puntos como se puede observar en la tabla 13, además es bueno mencionar que de manera general las instituciones de educación superior clasificadas por el análisis jerárquico como instituciones acreditadas cuentan con un mejor resultado en cada competencia de la prueba cuando los comparamos con los resultados de aquellas instituciones que no se clasificaron como acreditadas.

Al comparar, las medias globales de ambos grupos tenemos que las instituciones clasificadas como acreditadas poseen una media global de 176 y las instituciones clasificadas como no acreditadas un puntaje de 148 puntos, teniendo una diferencia de aproximadamente 19 %, lo que nos brinda una diferencia significativa entre ambos grupos de instituciones.

Componente	No acreditadas	Acreditadas
Ingles	149	182
Comunicación escrita	148	169
Razonamiento cuantitativo	160	187
Competencia ciudadana	135	164
Lectura critica	146	176
Diseño de sistemas log y prod	142	171
Formulación de proyectos	145	169
Pensamiento científico	142	168

Tabla 13: Medias por componente para las instituciones según la clasificación del clúster jerárquico.

Luego de hacer la comparación de la asignación del clúster jerárquico con la asignación real de las acreditaciones institucionales se identificaron 14 instituciones con diferencia entre la asignación real y la asignación del clúster, donde 7 de estas se encuentran ubicadas en el grupo de las instituciones que no cuentan con acreditación según el clúster jerárquico, pero que según la asignación real son instituciones que cuentan con acreditación, de igual manera existen 7 instituciones cuya asignación del clúster las ubica en el grupo de instituciones acreditadas, pero en realidad no cuentan con acreditación institucional.

### 9.2.2 Clúster no jerárquico

El resultado del análisis no jerárquico junto con el PCA nos brindan de manera visual la existencia de dos grupos, por un lado tenemos el grupo de las instituciones que cuentan con certificaciones de alta calidad, las cuales se identifican con el color azul, y se encuentran 35 de las 87 instituciones de educación superior que cuentan con resultados en los años 2016 y 2017, y a su vez tenemos 52 instituciones que no poseen certificación con base a la clasificación del modelo no jerárquico.



aproximadamente 17 %, lo que nos brinda una diferencia significativa entre ambos grupos de instituciones.

Componente	No acreditadas	Acreditadas
Ingles	146	175
Comunicación escrita	148	162
Razonamiento cuantitativo	156	182
Competencia ciudadana	131	160
Lectura critica	142	171
Diseño de sistemas log y prod	138	167
Formulación de proyectos	141	166
Pensamiento científico	138	164

Tabla 14: Medias por componente para las instituciones según la clasificación del clúster jerárquico.

Luego de hacer la comparación de la asignación del clúster no jerárquico con la asignación real de las acreditaciones institucionales se identificaron 18 instituciones con diferencia entre la asignación real y la asignación del clúster, donde 3 de estas se encuentran ubicadas en el grupo de las instituciones que no cuentan con acreditación según el clúster no jerárquico, pero que según la asignación real son instituciones que cuentan con acreditación, de igual manera existen 15 instituciones cuya asignación del clúster las ubica en el grupo de instituciones acreditadas, pero en realidad no cuentan con acreditación institucional.

### 9.3 Modelos de predicción

Luego de identificar la clasificación de las instituciones de educación superior por medio de los clúster jerárquico y no jerárquico en los grupos de instituciones acreditadas y no acreditadas, se decidió realizar un modelo de predicción que nos ayude a comprobar nuestra hipótesis de si las certificaciones de calidad tienen impacto en los resultados obtenidos por las instituciones de educación superior en las pruebas Saber Pro, para la predicción de si una institución es acreditada o no con respecto a los resultados de las pruebas se usaron dos modelos de la rama de aprendizaje supervisado como lo son el modelo de regresión logística y Random Forest (Bosques aleatorios).

### 9.3.1 Regresión logística

Para el modelo de regresión logística se usaron las medias para cada componente por institución y la asignación real de si son acreditadas o no, teniendo en cuenta que para la aplicación de este modelo en R la asignación real debe estar expresada de manera binaria, siendo 0 la asignación de las instituciones que no cuentan con acreditación y 1 aquellas que cuentan con acreditación institucional.

Luego de realizar la predicción se procedió a comparar esta con la asignación real para cada institución, generando la siguiente matriz de confusión, en la cual podemos apreciar que 19 instituciones tienen una predicción acertada siendo acreditadas al igual que 51 instituciones se encuentran con una predicción acertada siendo no acreditadas. (Ver tabla 15).

		Predicción	
		No acreditada	Acreditada
Asignación real	No acreditada	51	13
	Acreditada	4	19

Tabla 15: Matriz de confusión del modelo de regresión logística.

Es bueno identificar que 4 de las 13 instituciones restantes del estudio cuentan con acreditación institucional pero su desempeño en las pruebas no permite que el modelo los ubique como una universidad acreditada, estas instituciones son:

- Universidad Jorge Tadeo Lozano de Bogotá
- Universidad autónoma de Manizales
- Universidad autónoma de occidente – Cali
- Universidad pontificia Bolivariana – Medellín

De igual forma existen 13 instituciones que no cuentan con acreditación institucional pero sus resultados obtenidos permiten que el modelo los clasifique como instituciones acreditadas, estas instituciones son:

- Corporación universitaria Confacauca
- Fundación universitaria Konrad Lorenz
- Universidad católica de Colombia
- Universidad de Boyacá
- Universidad del atlántico
- Universidad del Sinú – Cartagena
- Universidad distrital Francisco Jose de Caldas
- Universidad EIA – Medellín
- Universidad militar nueva granada

- Universidad nacional de Colombia – Manizales
- Universidad nacional de Colombia – Medellín
- Universidad Santo Tomas – Bucaramanga
- Universidad Agustiniana

Las medidas de desempeño obtenidas para este modelo son las siguientes:

<b>Accuracy</b>	<b>80,46 %</b>
<b>Sensibilidad</b>	<b>82,60 %</b>
<b>Especificidad</b>	<b>79,68 %</b>
<b>AUC</b>	<b>81,15 %</b>

Tabla 16: Medidas de desempeño del modelo de regresión logística.

El modelo cuenta con una precisión del 80,46 %. Sin embargo tener un 80 % de precisión es bastante favorable ya que nos permite identificar que las predicciones realizadas por medio del modelo de regresión logística se acercan mucho a la información real.

Por otro lado tenemos la sensibilidad que nos indica la capacidad de nuestra predicción para dar como acreditadas las instituciones de educación superior que realmente cuentan con dicha distinción, caracterizando la capacidad que tiene nuestro modelo para identificar las instituciones acreditadas con una proporción del 82,60 %.

De igual manera la especificidad nos indica la capacidad de nuestro modelo para dar como instituciones no acreditadas a aquellas que realmente no cuenten con dicha distinción con una proporción del 79,68 %.

Por ultimo tenemos la curva ROC y el AUC (Área bajo la curva) del modelo que muestran el rendimiento de nuestra predicción en todos los umbrales de clasificación, la curva ROC se encuentra representada bajo los parámetros de sensibilidad y especificidad mencionados anteriormente y el AUC es simplemente la proporción de área que se encuentra bajo la curva ROC que se puede interpretar como la probabilidad del 81,15 % de que el modelo clasifique una institución acreditada de manera correcta al igual que una institución que no cuente acreditación (Ver figura 12).

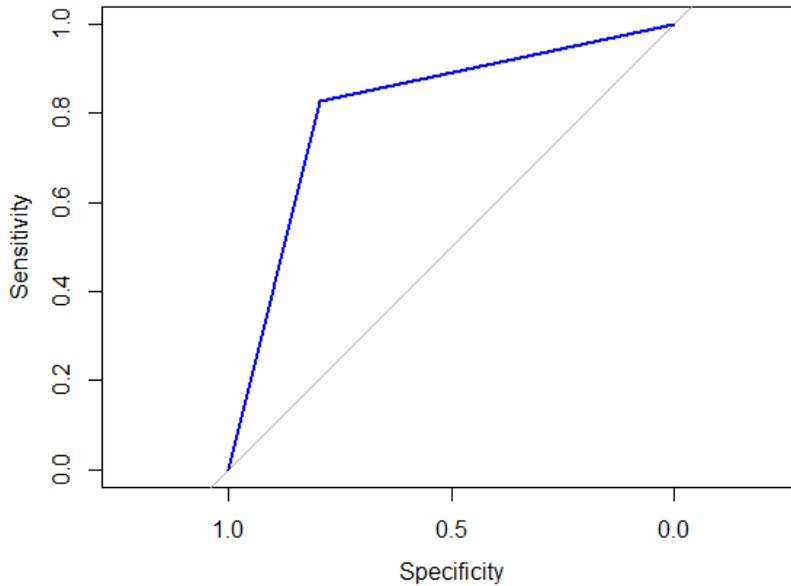


Figura 12: Curva ROC para el modelo de regresión logística.

### 9.3.2 Random Forest

Para el modelo de Random Forest se usaron las medias para cada componente por institución y la asignación real de si son acreditadas o no, teniendo en cuenta que para la aplicación de este modelo en R la asignación real debe estar expresada de manera binaria, siendo 0 la asignación de las instituciones que no cuentan con acreditación y 1 aquellas que cuentan con acreditación institucional. El número de árboles propuestos para nuestro modelo fue de 1000 árboles.

Luego de realizar la predicción se procedió a comparar esta con la asignación real para cada institución, generando la siguiente matriz de confusión, en la cual podemos apreciar que 18 instituciones tienen una predicción acertada siendo acreditadas al igual que 56 instituciones se encuentran con una predicción acertada siendo no acreditadas (Ver tabla 17).

		Predicción	
		No acreditada	Acreditada
Asignación real	No acreditada	56	8
	Acreditada	5	18

Tabla 17: Matriz de confusión del modelo Random Forest

Es bueno identificar que 5 de las 13 instituciones restantes del estudio cuentan con acreditación institucional pero su desempeño en las pruebas no permite que el modelo los ubique como una universidad acreditada, estas instituciones son:

- Universidad Jorge Tadeo Lozano de Bogotá
- Universidad autónoma de Manizales
- Universidad autónoma de occidente – Cali
- Universidad pontificia Bolivariana – Medellín
- Universidad tecnológica de Pereira

De igual forma existen las 8 instituciones que no cuentan con acreditación institucional pero sus resultados obtenidos permiten que el modelo los catalogue como instituciones acreditadas, estas instituciones son:

- Universidad de Córdoba – Montería
- Universidad de Atlántico – Barranquilla
- Universidad distrital Francisco Jose de Caldas
- Universidad EIA – Medellín
- Universidad militar Nueva Granada
- Universidad nacional de Colombia – Manizales
- Universidad nacional de Colombia – Medellín
- Universidad santo Tomas – Bucaramanga

Las medidas de desempeño obtenidas por el modelo pueden variar cada vez que se realice la réplica del modelo, para este modelo las mejores estimaciones se pueden apreciar en la Tabla 18.

<b>Accuracy</b>	<b>85,05 %</b>
<b>Sensibilidad</b>	<b>78,26 %</b>
<b>Especificidad</b>	<b>87,50 %</b>
<b>AUC</b>	<b>82,88 %</b>

Tabla 18: Medidas de desempeño del modelo Random Forest

El modelo cuenta con una precisión del 85,05% un poco mejor que la precisión del modelo de regresión logística anteriormente propuesto. Sin embargo tener un 85 % de precisión es bastante favorable ya que nos permite identificar que las predicciones realizadas por medio del modelo de Random Forest se acercan mucho a la información real.

Por otro lado tenemos la sensibilidad que nos indica la capacidad de nuestra predicción para dar como acreditadas las instituciones de educación superior que realmente cuentan con dicha distinción, caracterizando la capacidad que tiene

nuestro modelo para identificar las instituciones acreditadas con una proporción del 78,26 %.

De igual manera la especificidad nos indica la capacidad de nuestro modelo para dar como instituciones no acreditadas a aquellas que realmente no cuenten con dicha distinción con una proporción del 87,50 %.

Por ultimo tenemos la curva ROC y el AUC (Área bajo la curva) del modelo que muestran el rendimiento de nuestra predicción en todos los umbrales de clasificación, la curva ROC se encuentra representada bajo los parámetros de sensibilidad y especificidad mencionados anteriormente y el AUC es simplemente la proporción de área que se encuentra bajo la curva ROC que se puede interpretar como la probabilidad del 82,88 % de que el modelo clasifique una institución acreditada de manera correcta al igual que una institución que no cuente acreditación (Ver Figura 13).

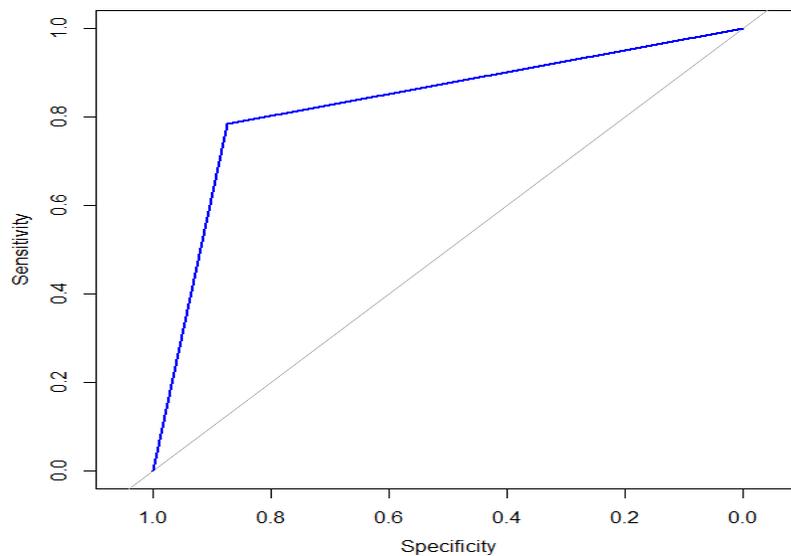


Figura 13. Curva ROC para el modelo de Random Forest

## 10 ENTREGABLES DEL PROYECTO

- Artículo Científico.
- Código en R.
- Documento físico.
- Documento digital.

## 11 CONCLUSIÓN

En este trabajo se desarrolló el diseño de una metodología articulada en la técnica de aprendizaje automático para la clasificación y predicción de si las instituciones de educación superior que poseen programas de ingeniería Industrial del país cuentan con certificaciones de calidad, teniendo como base de estudio los resultados y desempeño obtenido en cada componente tanto genérica como específica por dichas instituciones en las pruebas Saber Pro los años de 2016 y 2017.

De los resultados de la investigación se puede concluir que en Colombia las instituciones con programas de ingeniería industrial que cuentan con acreditación institucional presentan un mejor desempeño que aquellas que no cuentan con dicha distinción, a su vez se pudo evidenciar que las instituciones públicas cuentan con mejores resultados que las universidades privadas, y que las instituciones de la costa caribe no tienen una diferencia significativa con respecto a las demás, todo lo dicho anteriormente obtuvo teniendo un intervalo de confianza del 95%.

Por otra parte se identificó por varios métodos de aprendizaje no supervisado la capacidad de las instituciones a clasificarse entre ellas, con base a los resultados obtenidos, teniendo una precisión aproximada del 84% y 79% para los métodos jerárquico y no jerárquico respectivamente. De igual forma para los modelos de predicción implementados se obtuvieron buenos resultados y se logró identificar aquellas instituciones que hacen diferencia al estar ubicadas como acreditadas y no contar con acreditación al igual que el grupo de aquellas que realmente cuentan con acreditación institucional pero que sus resultados en las pruebas Saber Pro no permiten que se tenga una predicción correcta.

Todos estos resultados ayudaron a generar un conocimiento que puede servir para futuros proyectos en diferentes líneas de investigación.

## REFERENCIAS BIBLIOGRAFICAS

- [1] E. R. Borroto Cruz and R. S. Salas Perea, *Revista Cubana de educacion medica superior.*, vol. 18, no. 3. Centro Nacional de Informacion de Ciencias Medicas, 2004.
- [2] T. A.-Q. quaderns d'estadística i investigació and undefined 2001, "La minería de datos, entre la estadística y la inteligencia artificial," *raco.cat*.
- [3] "Misión y Visión del Icfes - Icfes Instituto Colombiano para la Evaluación de la Educación." [Online]. Available: <http://www2.icfes.gov.co/quienes-somos/mision-y-vision>. [Accessed: 10-Feb-2019].
- [4] C. E. Los Servicios Educativos, "ANDRÉS SENLLE NILDA GUTIÉRREZ."
- [5] "DECLARACION MUNDIAL SOBRE LA EDUCACION SUPERIOR EN EL SIGLO XXI: VISION Y ACCION." .
- [6] "TENDENCIAS MUNDIALES EN LA EDUCACION SUPERIOR: EL PAPEL DE LOS ORGANISMOS MULTILATERALES." .
- [7] "UNESCO UIS." [Online]. Available: <http://uis.unesco.org/>. [Accessed: 12-Feb-2019].
- [8] O. Salinas Gómez, "Tendencias mundiales que afectan la educación superior."
- [9] "Documento de política para el cambio y el desarrollo en la educación superior; 1995."
- [10] "Colombia Aprende | La red del conocimiento." .
- [11] "Qué es Ser Pilo Paga." .
- [12] "Se acaba el programa Ser Pilo Paga anunció la ministra de Educación María Victoria Angulo." .
- [13] "Ley 30 de Diciembre 28 de 1992."
- [14] "Sistema Nacional de Acreditación en Colombia - CNA." .
- [15] "¿Donde puedes estudiar? - Ser Pilo Paga." .
- [16] "Información general del examen de estado de la calidad de la educación superior, Saber Pro - Icfes Instituto Colombiano para la Evaluación de la Educación." .
- [17] R. W. Hoyer *et al.*, "¿Qué es calidad?"
- [18] "MODELO DE EXCELENCIA EN LA GESTIÓN MALCOLM BALDRIGE Por: José Antonio Villagra ¿Cómo nace el Modelo de Excelencia?"
- [19] M. Ferrando Sánchez and J. Granero Castro, *Calidad total: modelo EFQM de excelencia*. Fundación Confemetal, 2005.
- [20] "Modelo de excelencia y calidad EFQM." .

- [21] V. M. Nava Carbellido, A. R. Jiménez Valadez, and C. Corrie, *ISO 9000:2000 : estrategias para implantar la norma de calidad para la mejora continua*. Limusa, 2002.
- [22] “NORMA TÉCNICA NTC-ISO COLOMBIANA 9001,” 2015.
- [23] “Aspectos generales ABET 1.”
- [24] “Descripción del modelo ABET.”
- [25] “Consejo Nacional de Acreditación - CNA.” [Online]. Available: <https://www.cna.gov.co/1741/article-186382.html>. [Accessed: 12-Feb-2019].
- [26] “Acuerdo 03 de 2014.”
- [27] “CNA - ¿Cómo se lleva a cabo la acreditación?” .
- [28] D. C. Bogotá and C. J. De, “Ministerio de Educación Nacional Consejo Nacional de Acreditación CNA GUIA DE PROCEDIMIENTO 02 APRECIACIÓN DE CONDICIONES INICIALES PARA ACREDITACIÓN INSTITUCIONAL PRIMERA EDICIÓN.”
- [29] C. Pérez López and D. Santín González, *Minería de datos : técnicas y herramientas*. Thomson, 2007.
- [30] A. García, *Inteligencia artificial : fundamentos, práctica y aplicaciones*. RC Libros, 2012.
- [31] W. B. Rauch-Hindin, *Aplicaciones de la inteligencia artificial en la actividad empresarial, la ciencia y la industria : (fundamentos-aplicaciones)*. Díaz de Santos, 1989.
- [32] “¿Qué es Machine Learning?” [Online]. Available: <https://cleverdata.io/que-es-machine-learning-big-data/>. [Accessed: 14-Feb-2019].
- [33] “Which machine learning algorithm should I use? - The SAS Data Science Blog.” [Online]. Available: <https://blogs.sas.com/content/subconsciousmusings/2017/04/12/machine-learning-algorithm-use/>. [Accessed: 14-Feb-2019].
- [34] H. Chitarroni, “La regresión logística.”
- [35] M. D. Sociedad Española de Nefrología. and J. C. Rodríguez Pérez, *Nefrología : publicación oficial de la Sociedad Española de Nefrología.*, vol. 20, no. 6. Elsevier, 2000.
- [36] A. M. Universidad Nacional de Córdoba. Facultad de Psicología, *Evaluar.*, vol. 6, no. 1. 2006.
- [37] “La regresion logistica. Una aplicación a la demanda de estudios universitarios.”
- [38] “Cómo aplicar árboles de decisión en SPSS.”
- [39] “ARTÍCULO ORIGINAL.”

- [40] L. Breiman, "Random Forests," 2001.
- [41] "Random Forest Simple Explanation – Will Koehrsen – Medium." [Online]. Available: <https://medium.com/@williamkoehrsen/random-forest-simple-explanation-377895a60d2d>. [Accessed: 08-Feb-2019].
- [42] T. J. Fontalvo-Herrera, E. J. Delahoz, A. A. Mendoza-Mendoza, T. J. Fontalvo-Herrera, E. J. Delahoz, and A. A. Mendoza-Mendoza, "Aplicación de Minería de Datos para la Clasificación de Programas Universitarios de Ingeniería Industrial Acreditados en Alta Calidad en Colombia," *Inf. tecnológica*, vol. 29, no. 3, pp. 89–96, Jun. 2018.
- [43] ICFES, "Reporte de resultados institucional Guía de interpretación y uso de resultados del examen Saber Pro," p. 34, 2017.
- [44] E. García, C. Romero Morales es Profesor, S. Ventura Soto es Profesor, E. García Salcines, C. Romero Morales, and S. E. Ventura Soto Carlos de Castro Lozano, "Sistema recomendador colaborativo usando minería de datos distribuida para la mejora continua de cursos e-learning," vol. 3, no. 1, p. 19, 2008.
- [45] "Predicting of school failure using data mining techniques."
- [46] R. N. C. Elena B. Durán, "Minería de datos para descubrir estilos de aprendizaje."
- [47] "Los procesos de internacionalización y globalización en la educación superior: Un análisis de los países OCDE."
- [48] J. A. Asociación Iberoamericana de Diagnóstico y Evaluación Psicológica., Manuel; Granero-Gallegos, Antonio; Baena-Extremera, Antonio; Abrales, *Análisis de los perfiles motivacionales y su relación con la importancia de la educación física en secundaria*, vol. 2, no. 38. Paidós Ibérica, 1995.
- [49] G. T. Huamaní, S. A. Huamani, J. M. Salcedo, and C. A. Fernandez, "Modelo de Análisis Envoltante de Datos (DEA) para Evaluar la Eficiencia de las Escuelas Profesionales de Ingeniería Industrial en el Perú con Enfoque de Desarrollo Sostenible," *TECNIA*, vol. 26, no. 2, pp. 72–72, Dec. 2016.
- [50] Ligia Alba Melo B., Jorge Enrique Ramos F., and Pedro Oswaldo Hernández S., "La Educación Superior en Colombia: Situación Actual y Análisis de Eficiencia."
- [51] O. Calzadilla Fernández de Castro, G. Jiménez López, B. E. González Delgado, and J. Ávila Pérez, "Aplicación de la minería de datos al Sistema Cubano de Farmacovigilancia," *Rev. Cuba. Farm.*, vol. 41, no. 3, pp. 0–0, 2007.
- [52] I. Perversi, F. Valenga, E. Fernández, Britos P, and G. -Martínez, "IDENTIFICACIÓN Y DETECCIÓN DE PATRONES DELICTIVOS BASADA EN MINERIA DE DATOS."
- [53] R. Martín, "La Medición de la Eficiencia Universitaria: Una Aplicación del Análisis Envoltante de Datos," *Form. Univ.*, vol. 1, no. 2, pp. 17–26, 2008.
- [54] I. A. Salas Durazo, "La acreditación de la calidad educativa y la percepción de su impacto en la gestión académica: el caso de una institución del sector no universitario en México," *Calid. en la Educ.*, no. 38, pp. 305–333, Jun. 2013.

- [55] Y. Cuevas, "Organismos y procesos de certificación de calidad de la educación superior privada en México," *Calid. en la Educ.*, no. 35, pp. 231–254, Dec. 2011.
- [56] C. Tünnermann Bernheim, "La calidad de la educación superior y su acreditación: la experiencia centroamericana," *Avaliação Rev. da Avaliação da Educ. Super.*, vol. 13, no. 2, pp. 313–336, Jun. 2008.
- [57] A. R. Hernández Mondragón, *La acreditación y certificación en las instituciones de educación superior. Hacia la conformación de circuitos académicos de calidad: ¿Exclusión o Integración?*, vol. 7, no. 26. Universidad La Salle, 2006.
- [58] "CALIDAD DE LA EDUCACIÓN SUPERIOR: UN ANÁLISIS DE SUS PRINCIPALES DETERMINANTES."
- [59] Guiselle María Garbanzo, "FACTORES ASOCIADOS AL RENDIMIENTO ACADÉMICO EN ESTUDIANTES UNIVERSITARIOS, UNA REFLEXIÓN DESDE LA CALIDAD DE LA EDUCACIÓN SUPERIOR PÚBLICA," *Educación*, vol. 31, no. 1, 2007.
- [60] C. R. Morales, V. Soto, and C. Hervás Martínez, "Estado actual de la aplicación de la minería de datos a los sistemas de enseñanza basada en web," 2005.
- [61] A. B. Román, D. Sánchez-Guzmán, and R. G. Salcedo, "Minería de datos educativa: Una herramienta para la investigación de patrones de aprendizaje sobre un contexto educativo," 2014.
- [62] H. Escobar, M. Aleivar, and A. Puris, *Aplicaciones de minería de datos en Marketing*, vol. 3, no. 8. MISC, 2015.
- [63] D. Rodríguez, G. Jordi, and V. Roquet, "Metodología de la investigación."
- [64] The R Development Core Team, "R: A Language and Environment for Statistical Computing."
- [65] "Consultar Instituciones Acreditadas - CNA." .
- [66] "Instituciones acreditacion alta calidad." .