**Universidad Internacional de La Rioja**

**Escuela Superior de Ingeniería y Tecnología**

**Máster Universitario en Análisis y Visualización de Datos Masivos**

# Factores que afectan la matrícula en la IES en Colombia

**Trabajo Fin de Máster**

**Presentado por:** Galeano Camacho, Erika Gisela

**Director/a:** Baldiris Navarro, Silvia Margarita PhD

**Co -Director/a:** Diaz Piraquive, Flor Nancy PhD

# Resumen

El presente documento describe los pasos realizados en la búsqueda de patrones de algunos datos recolectados por el ministerio de educación de Colombia, en el periodo de 2014 a 2018.

El objetivo principal consiste en analizar la situación actual del país en materia de educación superior, permitiendo proponer estrategias que sirvan de guía para la creación de políticas públicas en el país.

La metodología utilizada es la CRISP – DM, un proceso que permite comprender tanto el entorno educativo como los datos que serán procesados y modelados en las técnicas aprendizaje automático; dando forma a los resultados y las propuestas.

Finalmente, el desarrollo de la metodología anterior permite evidenciar que el incremento en cobertura de educación superior está siendo impulsada por las ciudades principales del país, reduciendo la oportunidad de crecimiento en otras regiones; por lo que se hace necesario la creación de políticas de acceso que promuevan el aumento de la cobertura a través de estrategias como los programas en modalidad virtual.

**Palabras Clave:** Instituciones de Educación superior, Matrícula en IES , Minería de datos, Matricula, Educación.

# Abstract

This document describes the steps realized in the search for patterns of some data collected by the Ministerio de Educación Nacional of Colombia, in the period from 2014 to 2018.

The main objective is to analyze the current situation in the country regarding superior education, allowing to propose strategies that serve as a guide for the creation of public policies in the country.

The methodology used is CRISP - DM, a process that allows understanding both the educational environment and the data that will be processed and modeled in automatic learning techniques; establishing the results and proposals.

Finally, the development of the previous methodology shows that the increase in coverage of superior education is being promoted by the country main cities, reducing the opportunity for growth in other regions; therefore, it is necessary to make policies that promote strategies such as virtual programs.

**Keywords:** Higher education Institutions, Data mining, Enrollment, Education, Tuition in Higher education.

# Índice de contenidos

# Índice de tablas

# Índice de figuras

# 1. Introducción

La educación es una actividad que permite potencializar las capacidades humanas, permitiendo a los individuos que componen la sociedad aportar en el crecimiento social de la misma (Celli et al., 2008); este derecho se convierte en un motor de desarrollo para los países que lo utilizan como un instrumento de reducción de pobreza e igualdad de género. (Banco Mundial, 2017)

En ese sentido los beneficios que generan las acciones derivadas de la educación son tanto individuales como colectivas, el Banco Mundial estima que se incrementan directamente los ingresos hasta un 9% por año de escolarización adicional, promoviendo el desarrollo económico, la innovación y la cohesión social. (Banco Mundial, 2017)

En Colombia la educación es un derecho fundamental del ciudadano. El actual sistema de educación superior cuenta con aproximadamente 298 instituciones, divididos en 4 tipos: el 28% pertenecen a universidades, 42% a instituciones universitarias, 18 % a instituciones tecnológicas y el 13% restante corresponde a instituciones técnicas profesionales, de las cuales únicamente 52 instituciones se encuentran acreditas en alta calidad, es decir que realizaron un proceso adicional en el que se garantiza autorregulación y aseguramiento de la calidad avalado por el Ministerio de Educación y el Consejo Nacional de Acreditación –CNA- (Mineducación, 2020).

Adicionalmente, se estima que la tasa de cobertura en educación superior en Colombia es de un poco más del 50%, pero únicamente la mitad de los estudiantes llegan a graduarse; sumado a las estimaciones del Banco Mundial el cual estima que las personas en Colombia tardan 36% más tiempo que otros países en completar programas académicos (Banco Mundial, 2017), complicando el panorama de desarrollo en la educación superior colombiana.

En este trabajo de fin de master (TFM) se describe el proceso de descubrimiento del conocimiento en diversas bases de datos, desde la selección, limpieza y transformación de datos, que permitirá aplicar técnicas de minería de datos que apoye la generación de conocimiento útil para el diseño de políticas en la educación superior colombiana.

## 1.1. Justificación

La educación superior en Colombia está caracterizada por tener baja cobertura, capacidad subutilizada en entidades privadas y tasas de deserción altas (Bernal et al., 2009); generando una brecha educativa con otros países, que cada año incrementa.

En ese sentido, es de resaltar la inscripción escolar en el nivel terciario reportado por el Banco Mundial; en el que se observa como al pasar los años Colombia incrementa la diferencia educativa con otros países de la región. Para el año 1971, el número total de estudiantes matriculados en educación superior en Colombia, Chile y Argentina eran del 5%, 10% y 15% respectivamente; años más tarde, en 1994 los indicadores reportan inscripciones escolares de 17%, 27% y 37% respecto a los países mencionados anteriormente.

Continuando con el comparativo, para el 2017 la tasa de colombianos inscriptos en educación superior ascendió a un 56%, frente al 88% y 90% reportados en Chile y Argentina respectivamente; es decir que en cuarenta años la brecha en educación ascendió de un 10% a un poco más de un 30% con respecto a estos dos países de la región (Banco Mundial, 2020), ver figura1.

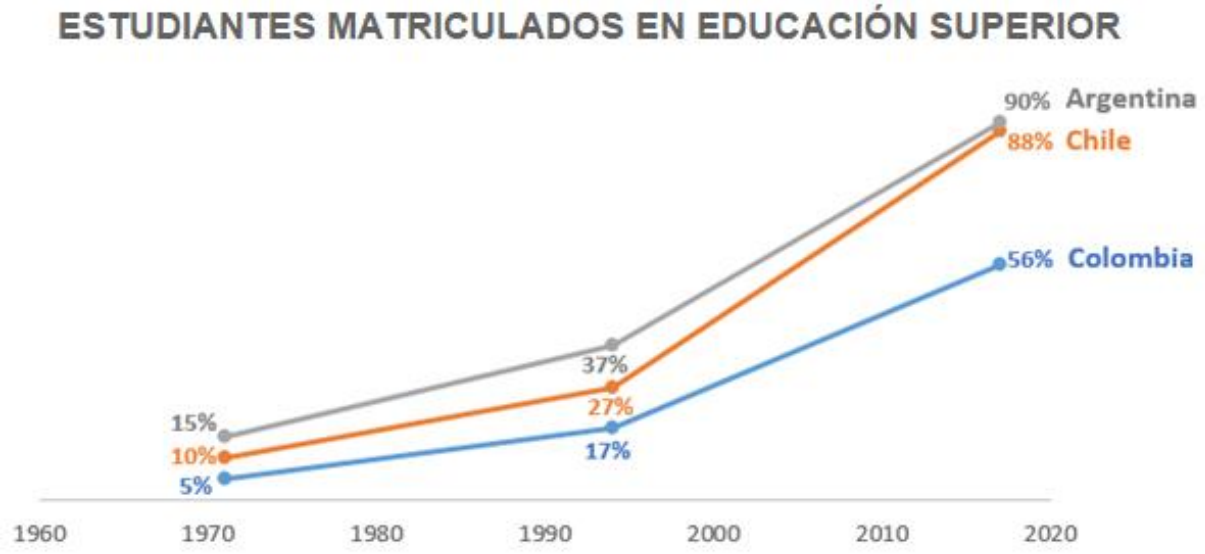



*Figura 1 Comparativo de matriculas*
Fuente: (Banco Mundia, 2020)

Las cifras anteriores recobran aun mayor importancia si se observa la tasa de deserción, dado que la permanencia de los estudiantes colombianos en los programas que se matriculan no es la esperada en número de graduados (Guzmán Ruiz et al., 2009, p. 60). El ministerio de Educación en el país estima que de cada dos estudiantes matriculados tan solo uno culmina el programa académico (Guzmán Ruiz et al., 2009, p.67).

De acuerdo con Porto (2001) las causas de deserción se dividen en de tipo: individual, académico institucionales y socioeconómicos; las causas individuales están relacionadas con variables como la edad, el género, la posición entere hermanos, el embarazo entre otras; las causas académicas están relacionadas con la orientación profesional, los resultados del examen o el rendimiento académico entre otros; en cuanto a las causas institucionales las variables relacionadas con los recursos universitarios, las becas, la normalidad académica

entre otros y en cuanto a las causas socioeconómicas variables como el estrato, la situación económica o la dependencia económica son las variables que influencian las matriculas en las universidades (Guzmán Ruiz et al., 2009).

La situación antes descrita disminuye la competitividad del país en el mundo globalizado en el que la importancia de la sociedad de conocimiento exige que se aproveche la ausencia de fronteras del conocimiento y la facilidad de adquirirlo (Gómez, 2010). Adicionalmente, la necesidad de subsistencia y la competencia entre las Instituciones de Educación Superior–IES-, por captar estudiantes, es cada vez mayor debido al número creciente de IES en el país.

Por lo anterior, es pertinente que se aproveche los datos que el Ministerio de educación recopila, para identificar factores que afectan la matriculas en las IES de Colombia y de esta manera poder identificar factores relevantes necesarios para apoyar la definición de políticas públicas que contribuyan con el crecimiento de matrículas en la educación superior en Colombia.

## 1.2. Planteamiento del trabajo

La competitividad del país disminuye frente a otros, debido a las dificultades que enfrenta la educación superior, por lo que se hace necesario que el gobierno colombiano aproveche los datos históricos de carácter público alrededor de la educación superior, específicamente el Sistema Nacional de Información de la Educación Superior –SNIES- es el sistema encargado de recopilar información de instituciones y programas académicos aprobados, los cuales se caracterizan por ser relevantes, confiables y oportunos (SNIES, 2020) debido a la fuente y tiempo en el que se divulgan, convirtiéndose en una oportunidad para caracterizar la educación superior en el país.

En ese sentido, con el fin de descubrir conocimiento específico que permita identificar factores relevantes en las matriculas de las universidades se propone aplicar técnicas de minería de datos. Una vez identificados los factores relevantes y los que no lo son, se realizara un análisis que sirva de guía para la formulación de políticas públicas que mejoren la competitividad del país en educación superior.

## 1.3. Estructura de la memoria

El presente documento está dividido en cinco capítulos, el capítulo I es la introducción a la investigación; seguido por el capítulo II, en el que se encuentran los conceptos claves de la minería de datos, la caracterización de la educación superior en Colombia y las

investigaciones en las que se utilizaron técnicas de minería de datos en temas relacionados con la educación, permitiendo al lector conceptualizarse en la temática a trabajar.

Seguido, en el capítulo III, se especifica el objetivo de la investigación y la metodóloga utilizada denominada CRISP – DM, definiendo las herramientas a utilizar en cada fase.

Continuando, el capítulo IV, el de mayor extensión describe cada una de las actividades realizadas con los datos, que permitieron encontrar y evaluar los patrones finales.

Finalmente, el capítulo V, de conclusiones y trabajos futuros permite identificar los resultados y las oportunidades presentes para profundizar en la investigación más adelante.

# 2. Contexto y estado del arte

El presente capítulo se encuentra dividió en tres partes, inicialmente un marco conceptual alrededor de la minería de datos, seguido de la caracterización del estado actual de la educación superior en Colombia, que se complementa con el listado de estudios relacionados a la temática de la investigación.

## 2.1. Minería de datos

La gran cantidad de datos que se recopilan, crea la necesidad de utilizar procesos que utilicen técnicas con el propósito de descubrir y describir patrones en los datos que se estén estudiando, para la generación de conocimiento y en la industria actual es de gran importancia el análisis efectivo y eficiente de los datos, (Jiawei Han, Micheline Kamber, 2014, p. 1-4) para ello la minería de datos utiliza técnicas de inteligencia artificial, siendo esta una fase en el descubrimiento del conocimiento (Jiawei Han, Micheline Kamber, 2014, p.6), ver figura 2.

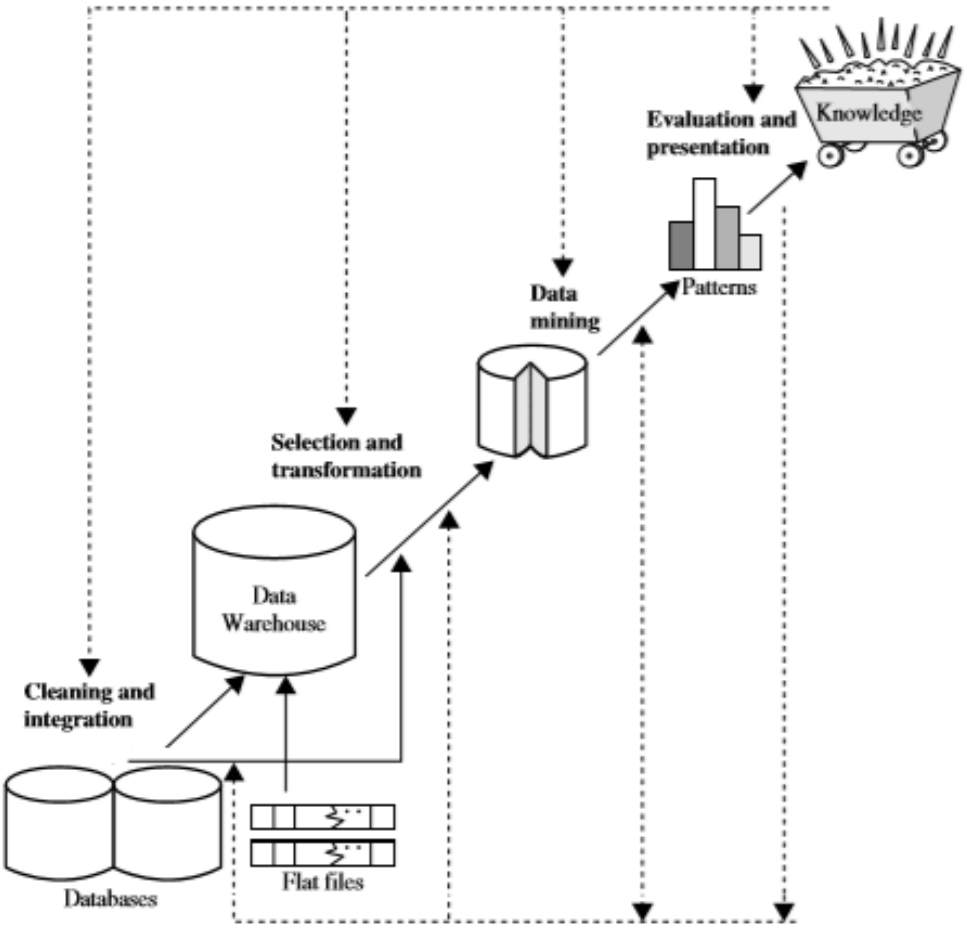



*Figura 2 La minería de datos como parte del descubrimiento del conocimiento*
(Jiawei Han, Micheline Kamber, 2014, p.7)

### 2.1.1. Descubrimiento de conocimiento

El proceso de descubrimiento de conocimiento, descrito en la figura 2, inicia con la limpieza de datos, seguido de la integración de diferentes fuentes, que permitirán seleccionar los datos relevantes para el análisis, en muchos casos esto implica que debe transformar o consolidar resúmenes de datos con lo que se puede realizar la minería de datos. (Jiawei Han, Micheline Kamber, 2014, p.7)

La minería implica el uso de métodos inteligentes para la extracción de patrones, que serán evaluados con ciertas medidas de interés para finalmente presentar conocimiento en muchas ocasiones con técnicas de visualización. (Jiawei Han, Micheline Kamber, 2014, p.7)

### 2.1.2. Análisis descriptivo de los datos

La primera tarea en la minería de datos es identificar qué tipos de atributos contiene los datos, los valores que asumen, como se distribuyen, la existencia de valores atípicos, es decir identificar en general las medidas de tendencia centra y de posición que nuestros valores asumen (Jiawei Han, Micheline Kamber, 2014, p.40).

### 2.1.3. Pre procesamiento de datos

Debido al múltiple origen de fuente, los datos suelen contener cifras susceptibles o inconsistentes que reducen su calidad; para corregir estos errores existen varias técnicas de pre procesamiento, como: la limpieza de datos usada en la eliminación de ruido e inconsistencias, la integración de datos de varias fuentes, la reducción de datos y las transformaciones de los mismos con procesos como la normalización. En cualquiera de estos procesos es fundamental prestar atención a los valores atípicos, para que no altere los patrones, manteniendo las características de calidad de los datos, la precisión, la integridad, la consistencia, la oportunidad, la interpretabildad y la credibilidad (Jiawei Han, Micheline Kamber, 2014, p.83-84).

### 2.1.4. Esquema de datos multidimensionales

El esquema de datos tiene como propósito facilitar el análisis de datos, existen diferentes tipos de bosquejos a la hora de representar los datos, como lo son: estrella, copo de nieve o constelación (Jiawei Han, Micheline Kamber, 2014, p.139-142). El modelo de estrella es el más usado, consiste en una tabla de hechos central relacionada con varias tablas (Bernabeu, 2007), ver figura 3.

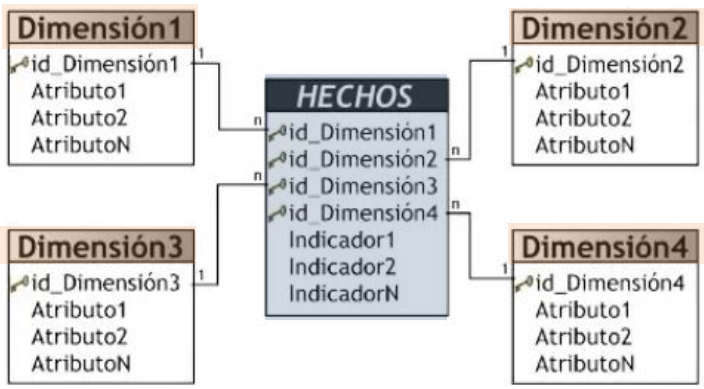



*Figura 3 Esquema en estrella*
Fuente: (Bernabeu, 2007)

Continuando con los tipos de bosquejo el copo de nieve, consiste en un diagrama con una única tabla de hechos, la cual está relacionada con diferentes tablas las cuales están jerarquizadas por las dimensiones entre ellas (Bernabeu, 2007), ver figura 4.

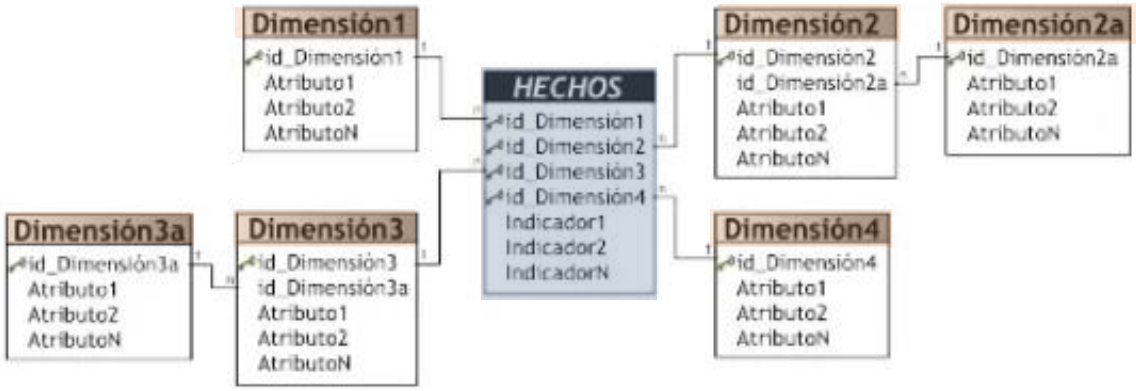



*Figura 4 Esquema en Copo de nieve*
Fuente: (Bernabeu, 2007)

Finalmente, el esquema de constelación, el cual considera varias tablas de hechos, y la jerarquización de las mismas (Bernabeu, 2007), ver figura 5.

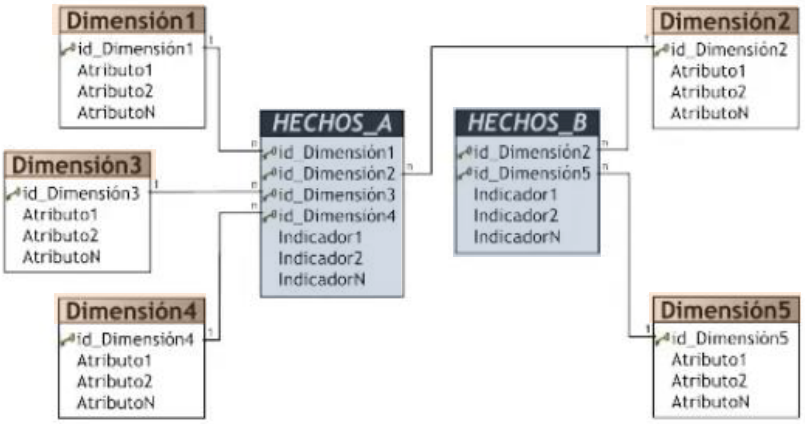



*Figura 5 Esquema en Constelación*
Fuente: (Bernabeu, 2007)

## 2.1.5. Minería de datos.

El objetivo de la minería de datos, es la búsqueda de relaciones recurrentes en los datos, utilizando técnicas de aprendizaje automático las cuales pueden ser catalogadas como aprendizaje supervisado cuando se conoce el concepto, en el caso contrario es denominado aprendizaje no supervisado (Jiawei Han, Micheline Kamber, 2014, p.549), por otro lado se puede hablar de subgrupos de algoritmos; como los son los arboles de decisión, las técnica de agrupamiento, las reglas, las redes neuronales artificiales y los algoritmos genéticos, entre otros, ver figura 6.

En primer lugar, los modelos que generan arboles de decisión son estructuras en donde cada atributo es representado en las ramas hasta llegar a las hojas que son las clases, algunos de los algoritmos más conocidos son el ID3 y el C4.5, el cual se diferencia por la realización de una poda adicional al modelo generado por parte de C4.5 (Jiawei Han, Micheline Kamber, 2014, p.331-332).

El segundo subgrupo, son las técnicas de agrupamiento también conocidas como clustering, son algoritmo de aprendizaje no supervisado que se dividen en: los exclusivos, que se caracteriza por que los objetos se agrupan de modo exclusivo en cada clúster; los jerárquicos, en una primera iteración contiene todos los objetos y estos van formando clústeres adicionales a medida que avanzan el número de iteraciones; los probabilistas, utilizan probabilidades para genera los subgrupos y los solapados, en el que los conjuntos se agrupan en subconjunto difusos, permitiendo que cada objeto pertenezca a varios grupos (Jiawei Han, Micheline Kamber, 2014, p.331-332).

En tercer lugar, se encuentran la clasificación basada en reglas, la cual se dividen en asociación, que dividen los pares de atributos valor y las de clasificación que predicen la clase; estas utilizan las expresiones de condicionalidad IF-THEN en las reglas, con las cuales se pueden generar arboles posteriormente (Jiawei Han, Micheline Kamber, 2014, p.355-356).

En cuarto lugar, se encuentran las redes neuronales artificiales las cuales utiliza conexiones entre neuronas artificiales; el perceptrón es la red más sencilla con una única neurona, las redes multicapa que contiene una serie de neuronas que se conectan entre ellas y las recurrentes que se retroalimentan automáticamente para aprender de los resultados generados (Jiawei Han, Micheline Kamber, 2014, p.398-400).

Finalmente, los algoritmos genéticos son parte de computación evolutiva basada en la idea de evolución natural, buscan la optimización utilizando la función de fitness que simula la calidad de cromosomas, los operadores genéticos como la selección, el cruce o la mutación y la técnica de diversidad en el que se utilizan diferentes métodos para la búsqueda.

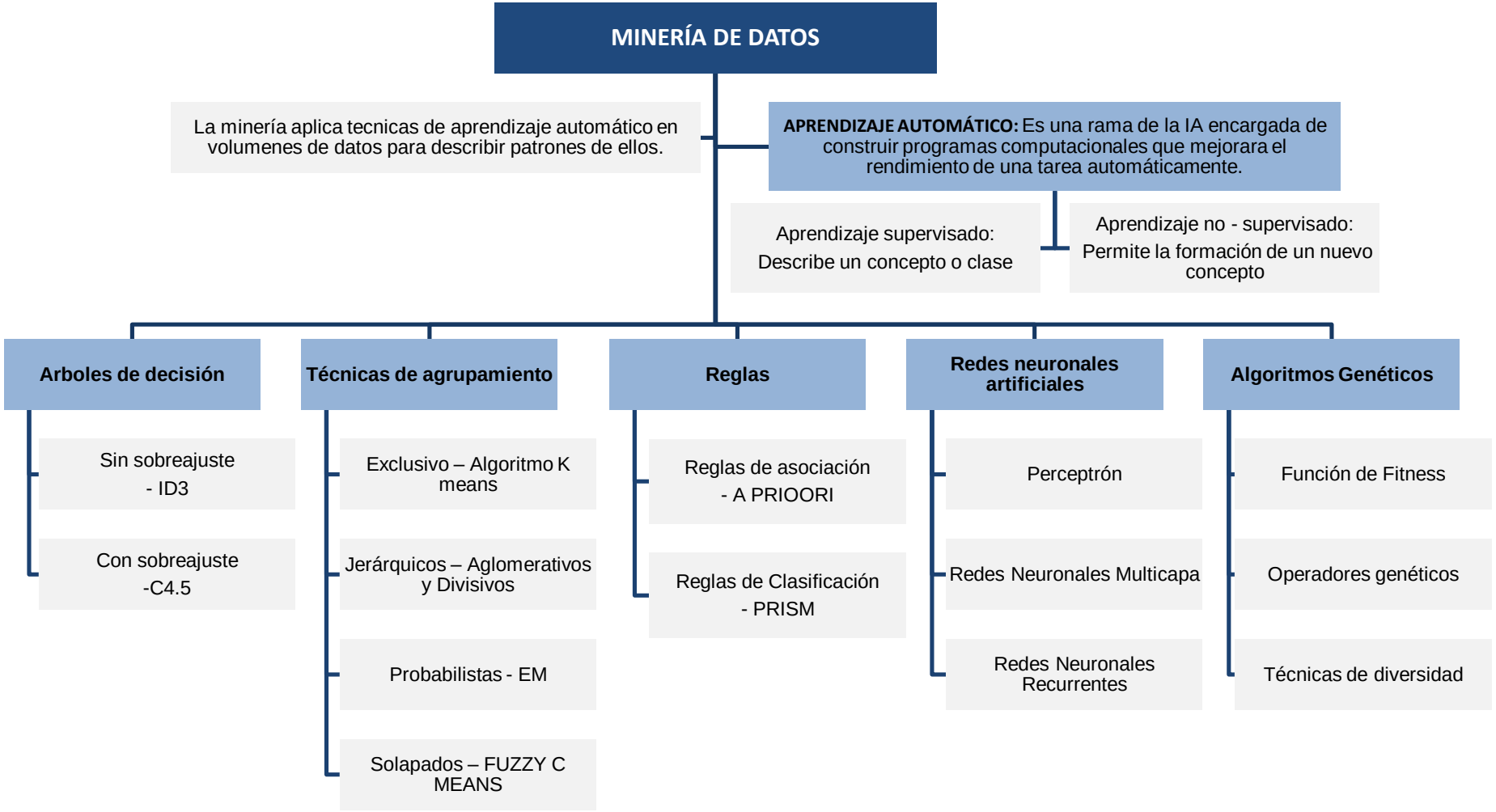



Figura 6 Aprendizaje automático
Fuente: Adaptado de (Jiawei Han, Micheline Kamber, 2014)

### 2.1.6. Evaluación de patrones

Existen diferentes métricas para evaluar los resultados generados de los diferentes algoritmo de aprendizaje automático, es de destacar que no siempre las reglas fuertes representan hallazgos significativos, pero en este punto la decisión es del analista, (Jiawei Han, Micheline Kamber, 2014, p. 272) pero estos siempre deben estar influenciadas por las medidas de evaluación ver tabla 1 y el contexto estudiado. (Jiawei Han, Micheline Kamber, 2014, p. 365)

Tabla 1 Medidas de evaluación

| *Measure* | *Formula* |
|---|---|
| accuracy, recognition rate | $\frac{TP+TN}{P+N}$ |
| error rate, misclassification rate | $\frac{FP+FN}{P+N}$ |
| sensitivity, true positive rate, recall | $\frac{TP}{P}$ |
| specificity, true negative rate | $\frac{TN}{N}$ |
| precision | $\frac{TP}{TP+FP}$ |
| $F$, $F_1$, $F$-score, harmonic mean of precision and recall | $\frac{2 \times precision \times recall}{precision + recall}$ |
| $F_\beta$, where $\beta$ is a non-negative real number | $\frac{(1+\beta^2) \times precision \times recall}{\beta^2 \times precision + recall}$ |

Fuente: (Jiawei Han, Micheline Kamber, 2014, p.365)

## 2.2. Educación superior en Colombia

La educación superior en Colombia nace en 1254 con la normativa del Código de las siete partidas que tenía como objetivo formar personal para el sector público como los abogados y para el sector religioso como los sacerdotes, aprobado por el Rey y el Papa (Rodríguez Valero & Gutiérrez Rodríguez, 2019, p. 163), más adelante en Bogotá se fundaron las Universidades Santo Tomás, San Francisco Javier y el Colegio Nuestra Señora del Rosario; las cuales impartían programas de teología, filosofía y medicina principalmente. (Melo-Becerra et al., 2017)

Así pues, la educación en el país era impartida hasta el momento como un privilegio para personas pertenecientes a órdenes religiosas y/o familias de clase social altas. En los años de la república, con la disolución de la Gran Colombia, se fundaron universidades centrales y públicas en 1826. (Rodríguez Valero & Gutiérrez Rodríguez, 2019, p. 163)

El tiempo trascurrió entre la fundación de nuevas universidades y guerras internas en el país, sin dar la importancia necesaria a la educación, se crea Ministerio de Educación Nacional con la Ley 7 de agosto 25 de 1886; hasta que en 1934. Con la presidencia de Alfonso López Pumarejo, se intenta fortalecer la educación en el país, asignando recursos a actividades que mejoraran la calidad educativa en el país; y se direcciona la creación de la Ciudad Universitaria, direccionamiento que se mantuvo dos gobiernos después, y que permitieron ampliar el número de facultades de la Universidad Nacional. En 1948 con el asesinato de Jorge Eliécer Gaitán, el país ingresa a una fase de violencia en la que se limitó la autonomía universitaria dando prioridad a la educación técnica, creado instituciones como el Servicio Nacional de Aprendizaje –SENA- y el Instituto Colombiano de Crédito Educativo y estudios Técnicos en el Exterior –ICETEX-. (Melo-Becerra et al., 2017, p.64-66)

A pesar de las advertencias de la misión Lebret, en el que advertían el riesgo de la expansión de universidades de baja calidad, los años después del gobierno de Rojas Pinilla, se caracterizó por un crecimiento de la demanda académica y un aumento de matrículas. En 1991, la Constitución Política de Colombia concibe la educación como un derecho ciudadano, y adjudica al estado el deber de prestar el servicio garantizando la calidad de programas académicos ofertados, adicionalmente se da autonomía a las entidades universitarias de expedir sus propios estatutos. (Melo-Becerra et al., 2017, p.67)

A continuación se expide la Ley 30 de 1992, por la cual se organiza la educación superior en el país; la cual define como IES las Instituciones Técnicas Profesionales, Instituciones Universitarias o Escuelas Tecnológicas y las Universidades (Dee & Heineman, 2016); entidades que deben ser avaladas por el ministerio de educación y acreditadas por el Sistema Nacional de acreditación –SNA-, actividades que busca incrementar las exigencias en la comunidad académica (Rodríguez Valero & Gutiérrez Rodríguez, 2019, p.167), algunos de los características de la educación superior en Colombia se describen en los siguientes apartados.

### 2.2.1. Tasa de cobertura

Es la relación entre los alumnos matriculados en pregrado y la población proyectada entre 17 y 21 años, los valores de matrícula es calculada con base en los datos que proporcionan las instituciones a SNIES y la proyección de población es realizada por el Departamento Administrativo Nacional de Estadística -DANE-, en la figura 7, a continuación, describe la tasa de cobertura en Colombia y por departamento para los años 2010, 2014 y 2018.

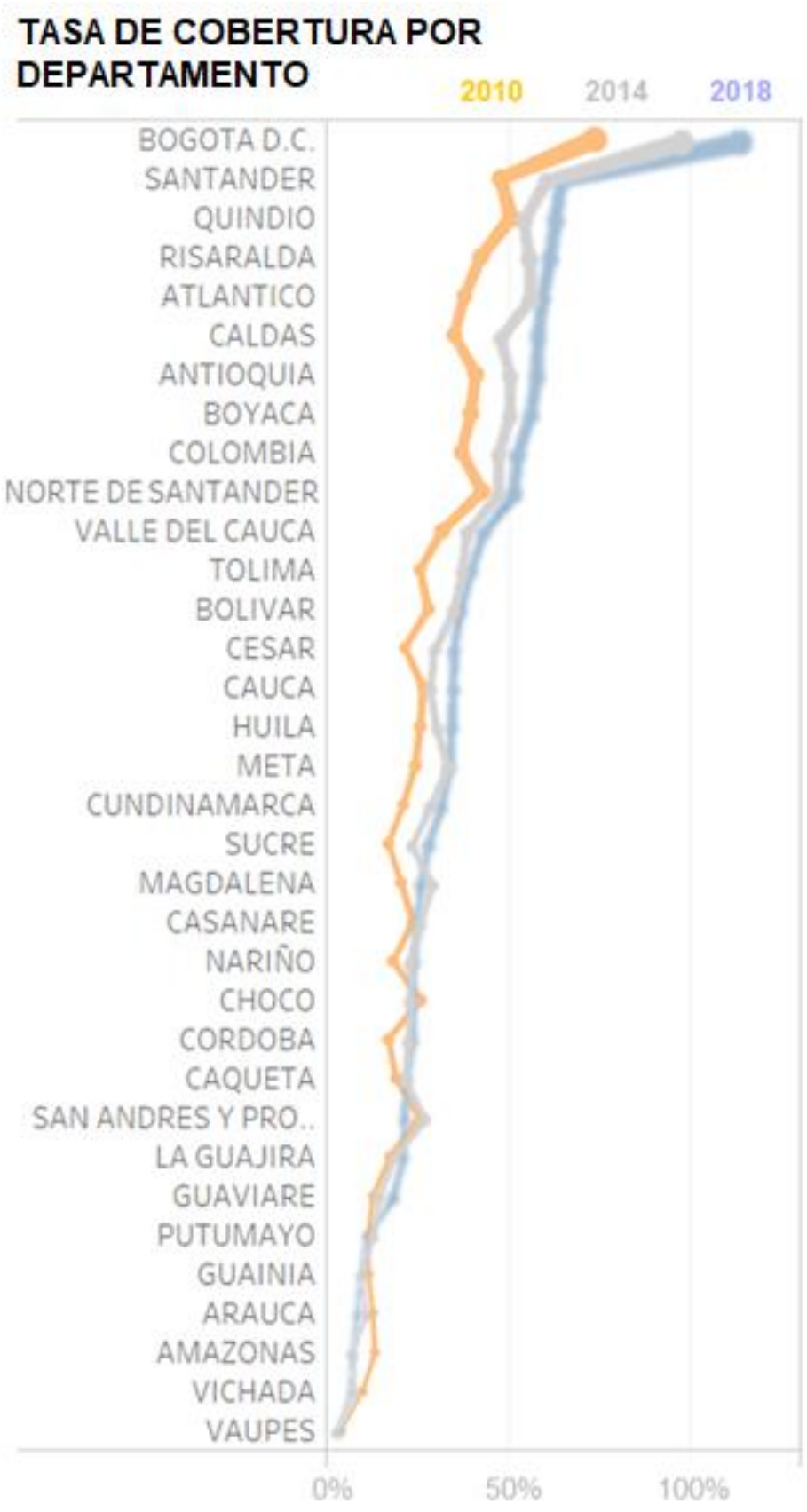



*Figura 7 Tasa* de cobertura por departam*ento en los años 2010, 2014 y 2018*
Fuente: Adaptado de Sistema Nacional de Información de Educación Superior – SNIES

En ese sentido, se observa una tendencia creciente en el valor reportado como país, valor que es impulsado por los principales departamentos, y por Bogotá especialmente el cual reporta para el año 2018, 114% debido a la migración constante de jóvenes de otras regiones a la capital; otros departamentos que se encuentra por arriba de la cifra nacional son: Santander, Quindío, Risaralda, Atlántico, Caldas, Antioquia y Boyacá con 65%, 63%, 62%, 60%, 58%, 58% y 56% para el 2018.

En ese mismo sentido, los departamentos que presentan menores cifras y en el que no se observan incrementos al pasar los años son: Putumayo, Guainía, Arauca, Amazonas, Vichada y Vaupés con 12%, 10%, 9%, 8%, 7% y 4%, para el año 2018.

## 2.2.2. Instituciones

Los sistemas de educación superior según el nivel académico se clasifican en Instituciones técnicas profesionales, instituciones tecnológicas, Instituciones universitarias y Universidades, en la actualidad Colombia cuenta con 30, 48, 134 y 86 respectivamente, concentrándose la mayor proporción de instituciones en las escuelas tecnológicas; del total de instituciones, 52 se encuentran acreditadas en alta calidad las cuales se encuentran agrupadas especialmente en el nivel universitario (SNIES, 2020), ver figura 8.

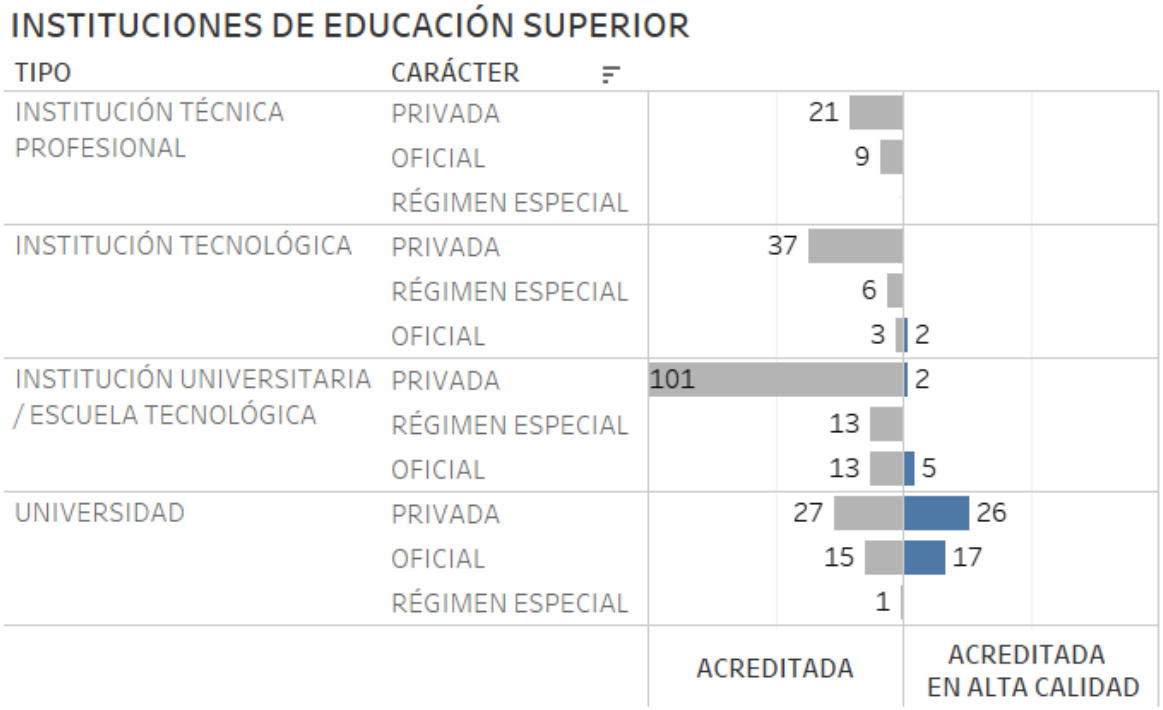



Figura 8 *Instituciones de educación superior en Colombia*
Fuente: Adaptado de Sistema Nacional de Información de Educación Superior - SNIES

Por otro lado, según la naturaleza jurídica las entidades están clasificadas como oficiales, privadas o de régimen especial los cuales tienen un manejo presupuestal diferente con el Gobierno Nacional sin llegar a ser oficiales. El sistema de educación superior cuenta con 64 entidades oficiales frente a 214 privadas (SNIES, 2020), ver figura 8; otro aspecto relevante a analizar es la ubicación geográfica de las instituciones de Educación Superior debido a que estas se encuentran concentradas en las ciudades principales, ver figura 9.

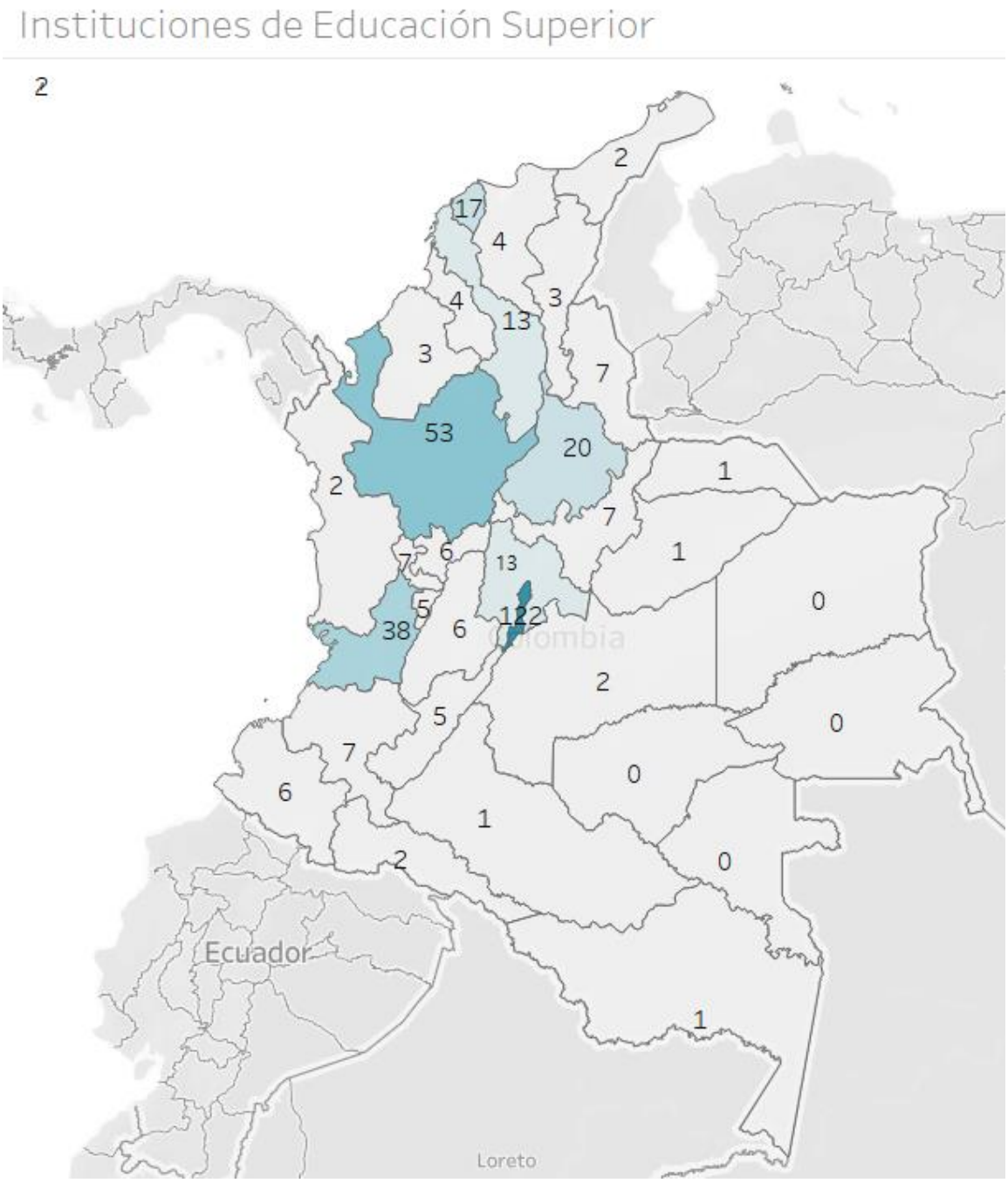



*Figura 9 Ubicación geográfica de las IES*
Fuente: Adaptado del Ministerio de Educación

### 2.2.3. Formación docentes

El número total de docentes en la IES incremento en un poco más de 50% en 8 años trascurridos, en los años 2010 y 2018 pasaron de tener un total de docentes de 102.727 a 158.951 respectivamente. Adicionalmente es de resaltar que la proporción de docentes en nivel pregrado y especialización disminuyo en el 2018, los docentes con nivel de pregrado pasando de representar 37% a 24%, del 2010 al 2018 y los docentes con nivel especialización, pasaron de representar 34% a 29%, en el mismo intervalo de tiempo descrito anteriormente (SNIES, 2020), ver figura 10.

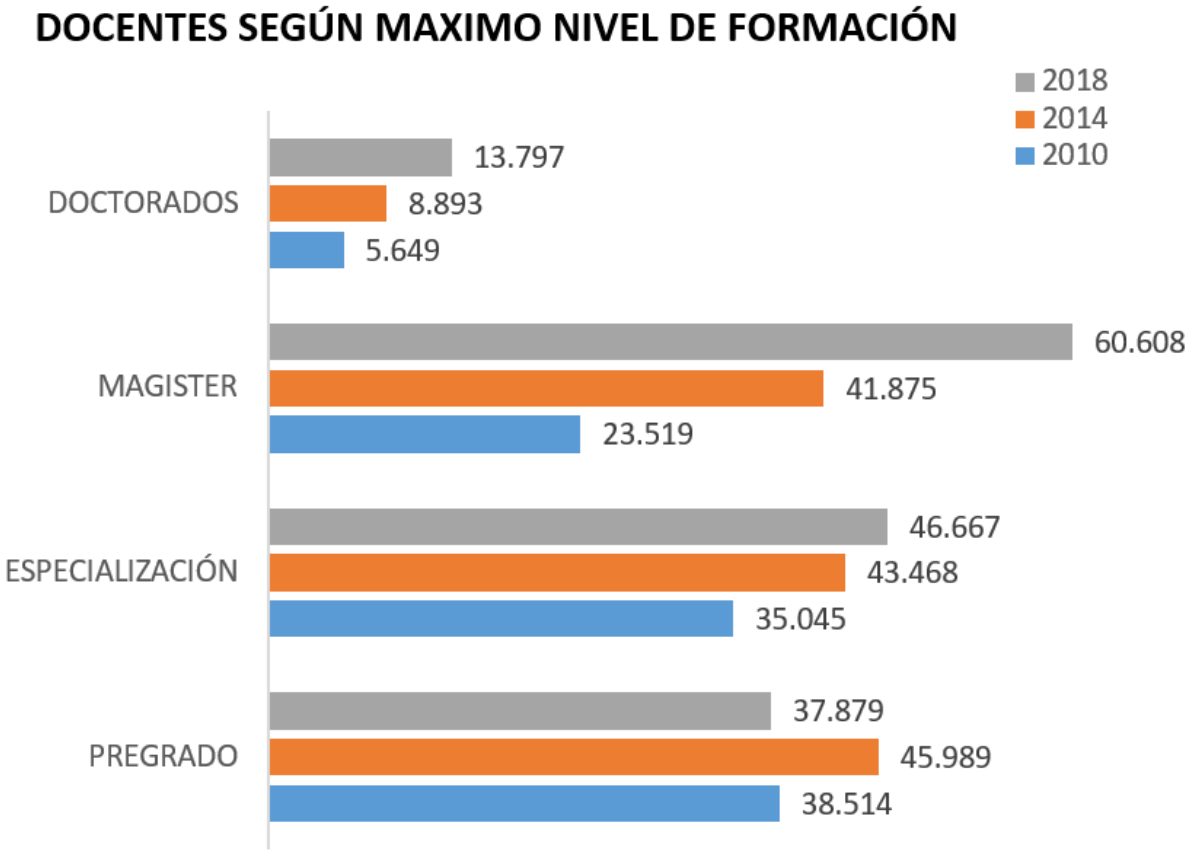



Figura 10 *Docentes según máximo novel de formación*
Adaptado de Sistema Nacional de Información de Educación Superior - SNIES

En concordancia, aumento el nivel docente de maestría y doctorado, pasando de 23% a 38% en el nivel de magister y de 5% al 9% en el nivel doctorado, en los periodos de 2010 y 2018 (SNIES, 2020), ver figura 10.

### 2.2.4. Nivel de graduados

El contraste del número de docentes, el número de graduados en Colombia se duplico en los 8 años trascurridos, pasando de 227.378 a 482.122, de 2010 a 2018. Pero la proporción de graduados por nivel de formación se mantuvo similar en los tres periodos de tiempo 2010, 2014 y 2018. En el nivel de Técnico profesional en promedio se gradúan 4%, en el Tecnológico 22%, en el Universitario 50%, es la especialización 19%, en la Maestría 4% y en el doctorado tan solo un 0.2% (SNIES, 2020), ver figura 11.

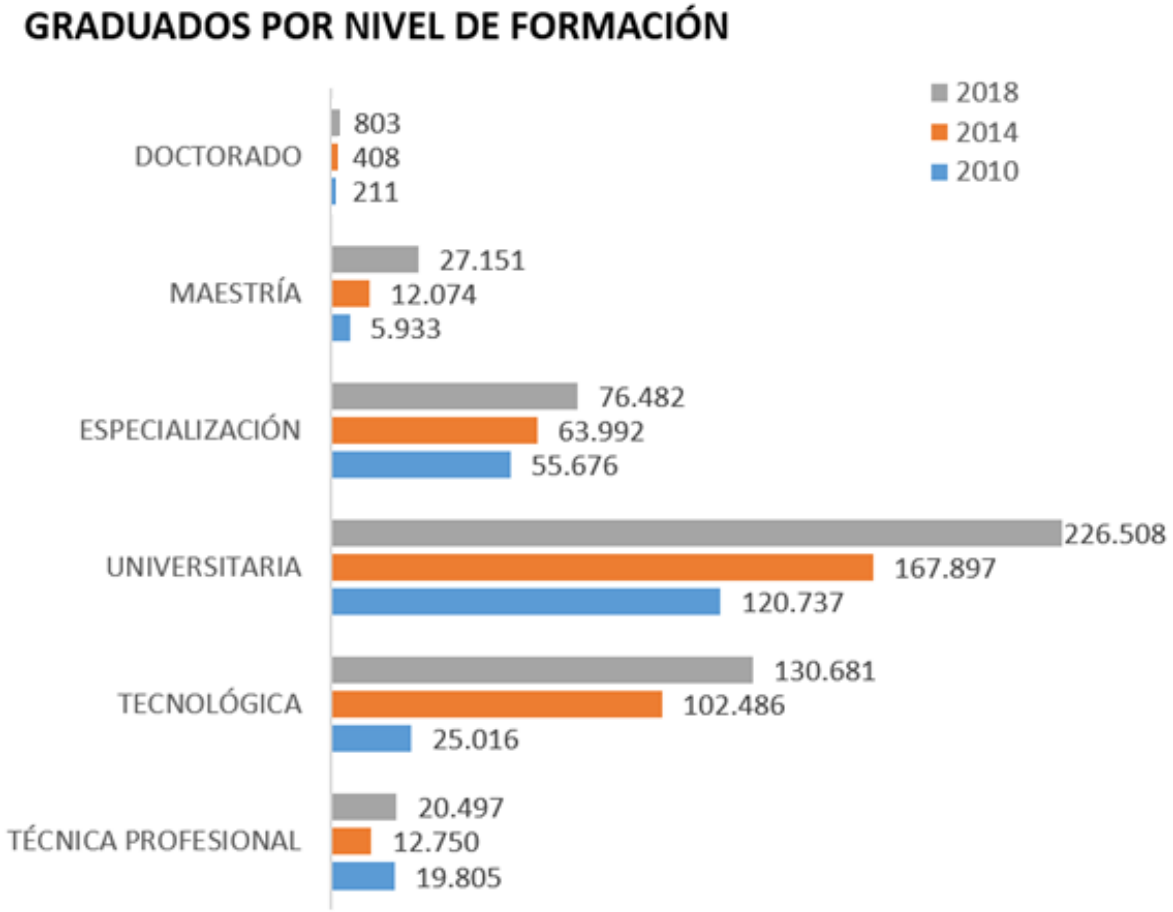



*Figura 11 Graduados por nivel de formación*
Adaptado de Sistema Nacional de Información de Educación Superior - SNIES

## 2.3. Estudios actuales

La educación superior es uno de los campos en los que se ha aplicado las técnicas de minería de datos para estudiar diferentes factores, en la tabla 2 se describe algunas investigaciones encontradas relacionadas con la temática, en ella se encuentra el tema de investigación, las técnicas y fuentes utilizadas, así como el año, la universidad en que se desarrolló, las principales conclusiones y las diferencias de estos artículos con la actual investigación.

Tabla 2 Estudios de Educación Superior y la Minería de datos

| TITULO | TEMA DE INVESTIGACIÓN | TÉCNICA UTILIZADA | AÑO / UNIVERSIDAD | CONCLUSIONES | DIFERENCIAS ESTUDIO ACTUAL |
|---|---|---|---|---|---|
| Modelo exploratorio de calidad en la educación superior (Villanueva Vázquez, 2019) | La investigación analiza los componentes que afectan la calidad en la educación superior, con valores de la Consejo Nacional de Acreditación –CNA-, Observatorio Laboral para la Educación – OLE- y SNIES. | Creación de ecuaciones estructuradas para explicar la calidad del programa estudiado. | 2019 / Universidad Autónoma del Caribe. Barranquilla, Colombia, | El reconocimiento social con las becas entregadas y los factores como el número de docentes tiempo completo en un programa son las variables que más impactan en la percepción de calidad de las IES | El presente estudio aborda los factores que afectan las matriculas en las IES, a diferencia de la calidad de los programas, |
| Características de aspirantes para el planteamiento de estrategias de captación de las instituciones de educación superior (Parra, 2019) | Conocer las herramientas existentes que permitan conocer las características de los graduados en la educación media. | Análisis descriptivo de fuentes de información en Colombia, SNIES, Sistema para la Prevención de la Deserción de la Educación Superior -SPADIES-, Instituto Colombiano para la Evaluación de la Educación -ICFES-. | 2019/ Escuela Colombiana de Ingeniería Julio Garavito. Bogotá, Colombia | Se entrega estrategias de captación basados fuentes de información públicas en el país a tener en cuenta por las IES. | Se realiza un análisis descriptivos |
| Análisis multivariado a los factores relacionados con el aprendizaje móvil en la educación superior en Colombia (Estrada-Villa & Boude-Figueredo, 2018) | Busca identificar los factores relacionados con el aprendizaje móvil en la educación superior. | Análisis de componentes principales basado en datos de 30 investigadores universitarios. | 2018 / Universidad Nacional CIDE. Bogotá, Colombia. | Para el desarrollo de políticas en la educación superior es clave tener en cuenta los factores de capacitación, características de estudiantes y falta de gestión administrativa. | La fuente de información es histórica y no subjetiva y el objetivo de la investigación |

Fuente: Adaptado de artículos citados

Tabla 2 (Continuación)

| TITULO | TEMA DE INVESTIGACIÓN | TÉCNICA UTILIZADA | AÑO / UNIVERSIDAD | CONCLUSIONES | DIFERENCIAS ESTUDIO ACTUAL |
|---|---|---|---|---|---|
| Implementación de minería de datos en la gestión académica de las instituciones de educación superior. (Harold Elbert Escobar Terán, Maritza Alcívar Saltos, Carlos Marquez de la Plata, 2017) | Revisión sobre las técnicas de minería de datos aplicables para el análisis de los problemas concernientes a los actores académicos y el centro educativo | Realiza una descripción de actores e investigaciones realizadas anteriormente | 2017/ Publicación cooperada entre CEDUT- Las Tunas y CEdEG-Granma, CUBA. | Las técnicas de minería de datos, contribuyen a la mejora del entendimiento del sistema educativo, | Se realiza un análisis descriptivos |
| Minería de datos y una aplicación en la educación superior (Beguerí & Malberti, 2017) | Investigar problemas de deserción, rezago y abandono de estudiantes de algunas facultades de la Universidad Nacional san Juan. | Tratamiento y análisis de satos con weka, R y RapidMiner | 2017 / Universidad Nacional de San Juan. San Juan, Argentina | La deserción preocupa a las universidades y aumenta estadísticamente. Se encuentran posibles relaciones entre el tipo de personalidad y la situación académica. | Fuente de información, objeto de la investigación. |
| Predicción de la eficiencia de las instituciones de educación superior colombianas con análisis envolvente de datos y minería de datos. (Cadavid & Mendoza, 2017) | Evaluar la eficiencia técnica de las instituciones de educación superior en Colombia de 2011 a 2013. | Uso de algoritmos en datos del ministerio para calcular la eficiencia de 32 universidades de Colombia | 2017 / Universidad del Norte Colombia. | El modelo valida que la variable más importante en la eficiencia de la IES, es el número de alumnos matriculados. | El Tamaño de la fuente de datos, y el periodo de años estudiado. |

Fuente: Adaptado de artículos citados

Tabla 2 (Continuación)

| TITULO | TEMA DE INVESTIGACIÓN | TÉCNICA UTILIZADA | AÑO / UNIVERSIDAD | CONCLUSIONES | DIFERENCIAS ESTUDIO ACTUAL |
|---|---|---|---|---|---|
| Aportaciones desde la minería de datos al proceso de captación de matrícula en instituciones de educación superior particulares (Estrada-Danell et al., 2016) | Diseño de un modelo predictivo de gestión de matrícula para las IES particulares de México. | Técnicas de correlación con minería de datos, para la creación de un árbol de decisión. En el software Rapid Miner. con datos de prospectos ficticios. | 2016 / Universidad Nacional de Costa Rica. Heredia, Costa Rica | El desarrollo de las nuevas tecnologías implica que las IES apoyen sus procesos en las técnicas de minería de datos. | La fuente de información y el lugar de estudio un análisis de universidades de México, el estudio actual se centra en Colombia. |
| Aplicación de la minería de datos en la extracción de perfiles de deserción estudiantil (RICARDO TIMARÁN PEREIRA, ANDRÉS CALDERÓN ROMERO, 2013) | Detectar patrones de deserción con técnicas de árboles de decisión. | Técnica de minería de datos, arboles de decisión datos de pregrado | 2013 / Universidad de Nariño e Institución Universitaria IUCESMAG. Pasto, Colombia | Se determinaron variables socioeconómicas y academias como las principales variables que afectan la deserción de la Universidad<br>La mala calidad de los datos dificulta observar todas la varíales deseadas. | La fuente de información y el objeto de estudio. |
| Minería de datos: Predicción de la deserción escolar mediante algoritmo de árboles de decisión y el algoritmo de k vecinos más cercanos. (Valero Orea et al., 2010) | Predecir la deserción escolar mediante algoritmos | Técnicas de minería de datos, información de la Universidad Tecnológica de Azúcar de Matamoros | 2010/ Universidad Tecnológica de Izúcar de Matamoros. Puebla, México. | Los alumnos de la UTIM desertan por factores como; la edad, los ingresos familiares y el nivel de inglés. | La fuente de información, el lugar de estudio y el objeto de investigación. |

Fuente: Adaptado de artículos citados

De las anteriores publicaciones, se puede observar como la minería de datos ha aportado importantes patrones en los campos investigados alrededor de la educación superior, como los son la predicción de la deserción, la gestión de matrículas en universidades privadas y el análisis de eficiencias de instituciones.

Por otro lado, las particularidades que diferencian este proyecto de las descritas anteriormente, son: la fuente de datos, al ser la primera investigación en la que aplicara minería de datos al histórico de indicadores –SNIES- y las relacionara con otras bases del gobierno; brindando información específica de los patrones que definen la Educación Superior en Colombia.

Es de destacar que, en la actualidad con el desarrollo de las nuevas tecnologías, se sabe que la educación enfrenta un reto de adaptación, y algunos investigadores afirman que la educación superior en Colombia, necesita de un nuevo modelo en las universidades que incluya las nuevas tecnologías y que configure el currículo de los programas académicos (Rodríguez Valero & Gutiérrez Rodríguez, 2019, p.177).

En ese sentido, es fundamental la integración con variables del sector de las telecomunicaciones, permitiendo analizar el estado del país para enfrentar los cambios tecnológicos proyectados en las investigaciones.

# 3. Objetivos concretos y metodología de trabajo

## 3.1. Objetivo general

Identificar los factores que afectan las matriculas en las IES de Colombia, mediante la aplicación de la minería de datos para proponer estrategias que sirvan como guía para la construcción de políticas públicas en el país.

## 3.2. Objetivos específicos

- Describir el estado actual de la educación superior en Colombia.
- Identificar y analizar descriptivamente los datos a utilizar en el estudio y pre procesarlos de acuerdo a las necesidades.
- Aplicar técnicas de minería de datos en las variables identificadas.
- Evaluar los patrones generados.
- Proponer recomendaciones basados en los patrones encontrados que sirvan de guía para la construcción de políticas públicas que promuevan la competitividad en educación superior del país.

## 3.3. Metodología del trabajo

De acuerdo con el problema de investigación, el enfoque es cuantitativo, por tanto, es secuencial, cada etapa precede a la siguiente y no se pueden eludir pasos, “utiliza la recolección de datos para probar hipótesis con base en la medición numérica y el análisis estadístico, con el fin establecer pautas de comportamiento y probar teorías (Hernández Sampieri Roberto, Fernández Carlos, Baptista María, 2014); las fuentes principales de información son secundarias, debido a que la información es proporcionada por cada una de las entidades al Ministerio de Educación.

Por otro lado, desde hace varios años se han propuesto diferentes metodologías pensadas específicamente para la minería de datos, algunas de las más tradicionales son la KDD, la CRISP – DM, la SEMMA y la CATALYST, debido a que el proyecto necesita una fase de entendimiento del negocio, etapas interactivas y una evaluación del resultado basado tanto en el modelo como en los objetivos del proyecto se seleccionó la metodología Cross Industry Standard Process or Data Mining -CRISP DM- (Aquino, Aldair. Molero, Gillermo. Rojano, 2015).

El CRISP DM, es una metodología que orienta la minería de datos en seis fases, el modelo es flexible permitiendo que se avance, se retroceda y se personalice a las necesidades de la investigación (IBM, 2020), a continuación una descripción de lo que se realizara en cada fase observada, lo cual se esquematiza en la figura 12:

**Fase 1:** *La Comprensión del negocio* en el que se determinara el objetivo del negocio, identificando los requisitos, los supuestos y las restricciones necesarias a tener en cuenta en el proyecto para cumplir el objetivo de la minería de datos.

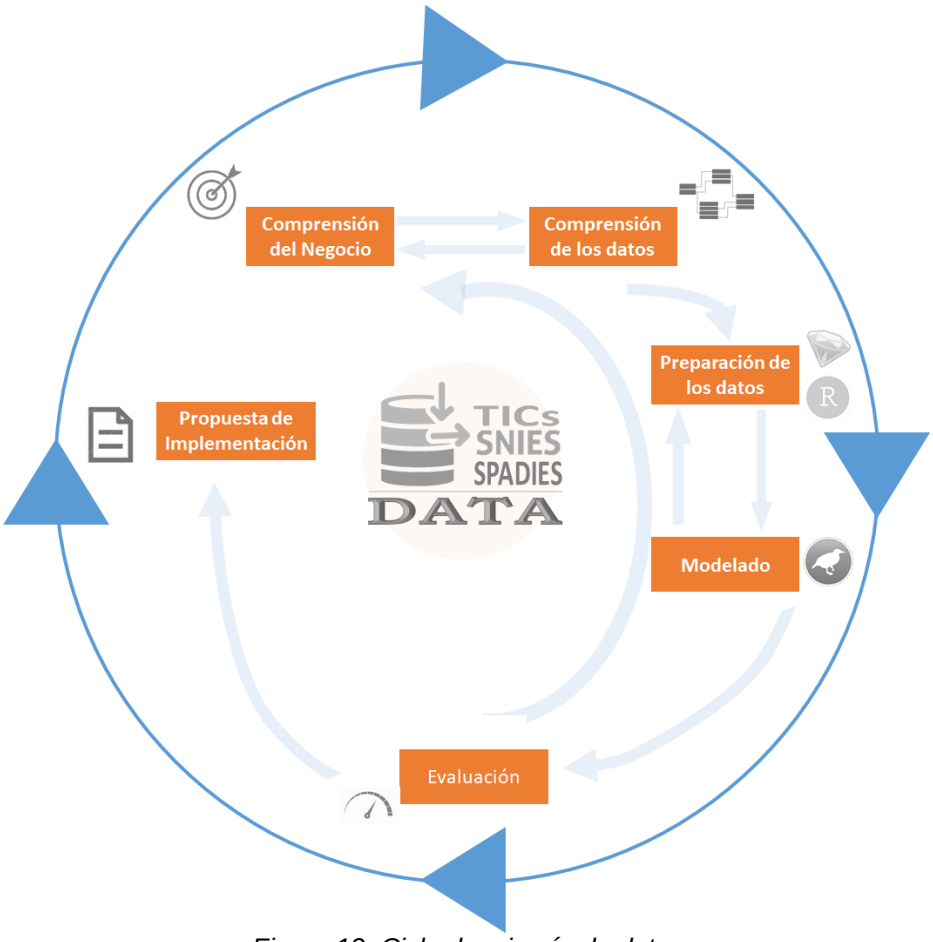



*Figura 12 Ciclo de minería de datos*
Adaptado de (IBM, 2020)

**Fase 2:** *La Comprensión de los datos* consiste en la descripción, la exploración y la calidad de los datos. La descripción de los datos incluye la identificación de los campos y el recuento del volumen existente. La exploración es la aplicación de la estadística básica, creación de tablas de frecuencia y tablas de dispersión en los casos que sea necesarios. La calidad de los datos realiza una verificación de datos nulos o fuera de rango, asegurando la completitud de los mismos.

**Fase 3:** *La Preparación de los datos* incluye la creación de subconjuntos en los datos apoyados en criterios descritos, así como en los casos que lo amerite la normalización de los campos, permitiendo crear nuevos atributos y nuevos registros que serán procesados.

Para el desarrollo de las fases dos y tres, se usarán herramientas como lo son R Studio y Open Refine.

**Fase 4:** *El Modelado* solicita la selección de la técnica de modelado y la ejecución de pruebas, esta fase suele exigir retrocesos a la preparación hasta poder llegar la construcción de un modelo elegido como el evaluado, se utilizara la herramienta weka para el desarrollo.

**Fase 5:** *La Evaluación* tiene en cuenta las métricas, los criterios del problema y el objetivo del proyecto, para aprobar el modelo que será estudiado.

**Fase 6:** *La Propuesta de implementación* o recomendaciones generadas del modelo, es la transformación a conocimiento, con una serie de recomendaciones de acciones basadas en la observación de resultados.

# 4. Desarrollo específico de la contribución

En este capítulo se desarrollará la contribución, se dividirán en secciones, una por cada fase propuestas en la metodología anteriormente.

## 4.1. Fase 1: Comprensión del negocio

Las políticas públicas educativas buscan cerrar las brechas existentes en el país, garantizando el derecho a la educación de calidad que impulsara el desarrollo de los individuos y la sociedad (Mineducación, 2020); en cuanto al conocimiento de la situación actual en la sección 2.2 caracteriza la educación superior en Colombia, la tasa de cobertura, las instituciones, el nivel de formación de los docentes y los grados realizados por nivel educativo en los últimos años, evidenciando deficiencias en la tasa de cobertura y altos índices de deserción, es por ello que para identificar los principales determinantes de deserción en

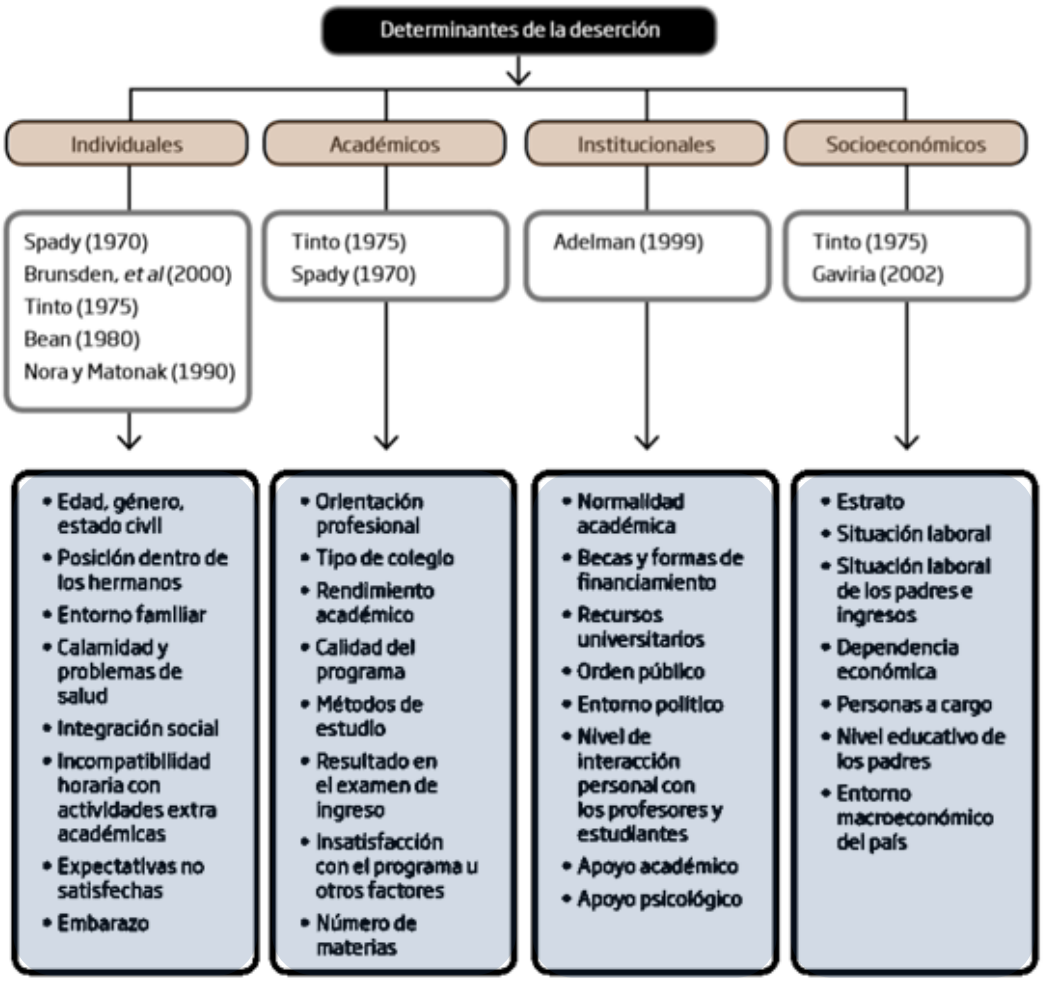



Figura *13 Estado del arte de los determinantes de la deserción estudiantil*
Fuente: Libro de la deserción estudiantil del Ministerio de Educación

Colombia se utilizará como guía la clasificación proporcionada en el libro de deserción estudiantil publicado por el Ministerio de Educación, figura 13; basado en ese esquema se clasificarán las variables obtenidas de SNIES y SPADIES con relación al perfil de las personas que se matriculan en primer semestre de las diferentes entidades, en la figura 14, se observan las variables clasificadas en los cuatro subconjuntos

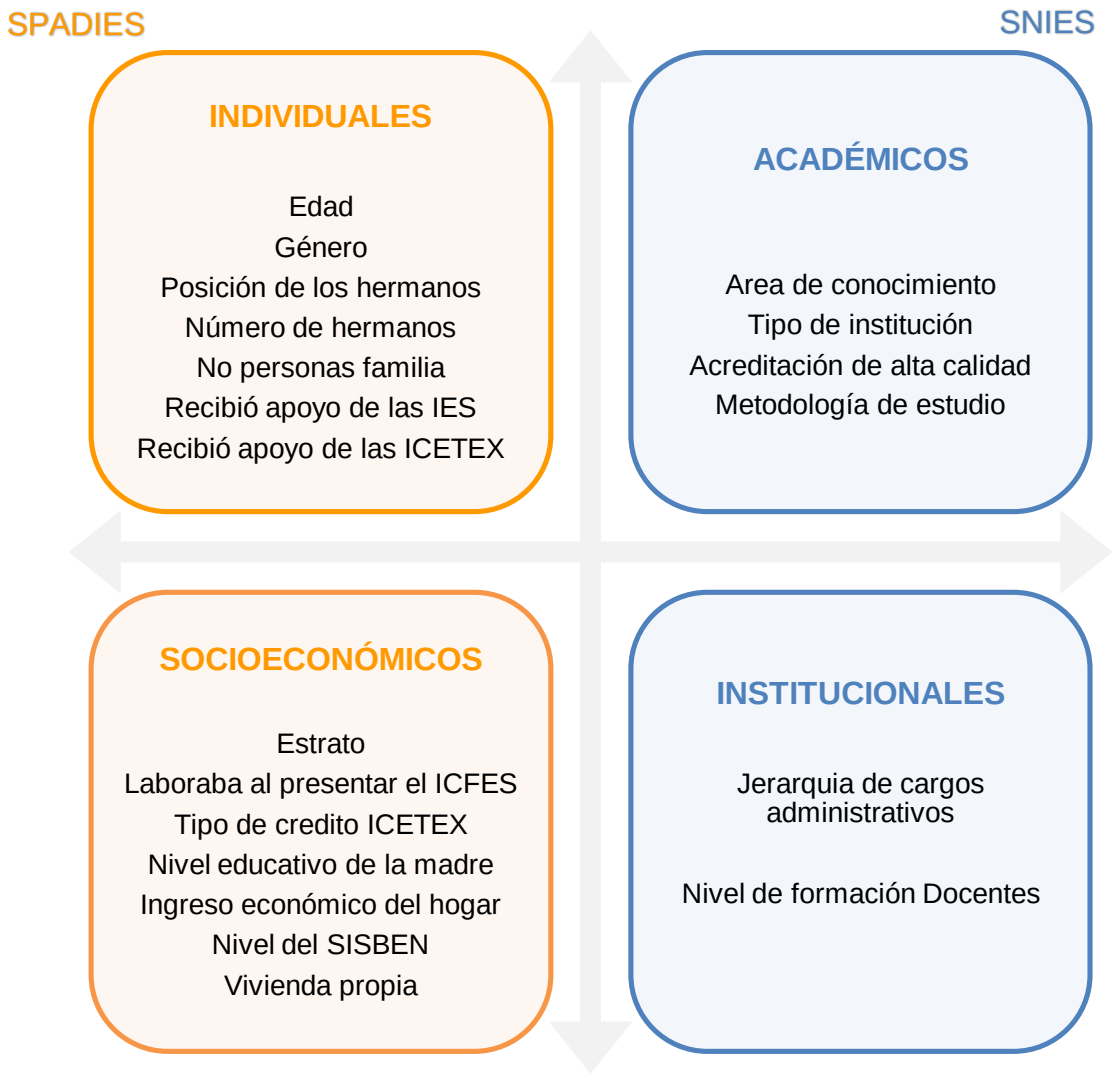



Figura *14 Clasificación general de variables*
Fuente: Elaboración propia adaptado de Min Educación

En la figura 12, se observan las variables SPADIES, clasificadas en dos subconjuntos, Individuales y socioeconómicos; adicionalmente se observan las variables SNIES, clasificadas en dos subconjuntos, académico e institucionales.

## 4.2. Fase 2: Comprensión de los datos

La presente sección describe los datos del conjunto total de variables, se divide en dos subcapítulos según la procedencia de los datos.

### 4.2.1. Comprensión de datos - Variables académicos e institucionales

El primer conjunto de datos fue exportado del sitio web del Ministerio de educación de Colombia, en el sistema de SNIES, en la tabla 3 se encuentra la relación de las bases de datos utilizadas detallando el nombre, la descripción, el intervalo de años a estudiar y el número total de registros que cada una de las bases de datos contiene.

Tabla 3 Descripción de bases de datos

| Nombre | Descripción | Años | No. de registros |
|---|---|---|---|
| Estudiantes Matriculados 1er Curso | Personas naturales que formalizan el primer curso en el programa que fue admitido. | 2014 a 2018 | 197.797 |
| Recurso Humano Administrativo | Personas que desarrollaron las actividades organizacionales relacionadas con la prestación del servicio de educación superior. | 2014 a 2018 | 2.899 |
| Recurso Humano Docente | Personas naturales que orientan la formación de los estudiantes de educación superior. | 2014 a 2018 | 78.757 |
| IES | Listado de instituciones de educación superior. | Actualización 2019 | 570 |

Fuente: Adaptado de Sistema Nacional de Información de Educación Superior – SNIES

La primera base de datos reporta las matriculas en primer curso de los estudiantes, la cual detalla el programa académico, el nivel académico, la metodología, área del conocimiento, genero entre otros; seguido del recurso administrativo el cual cuenta con la cantidad de docentes por nivel; la tercera base de datos es el recurso docente el cual especifica el género, el tipo de contrato y la cantidad reportada; y finalmente las IES esta base registra el total de instituciones y diferencia aquellas que realizaron el proceso voluntario de implementación de mecanismos eficaces de autoevaluación y aseguramiento de la calidad los cuales permiten certificarse en alta calidad, ver figura 15.

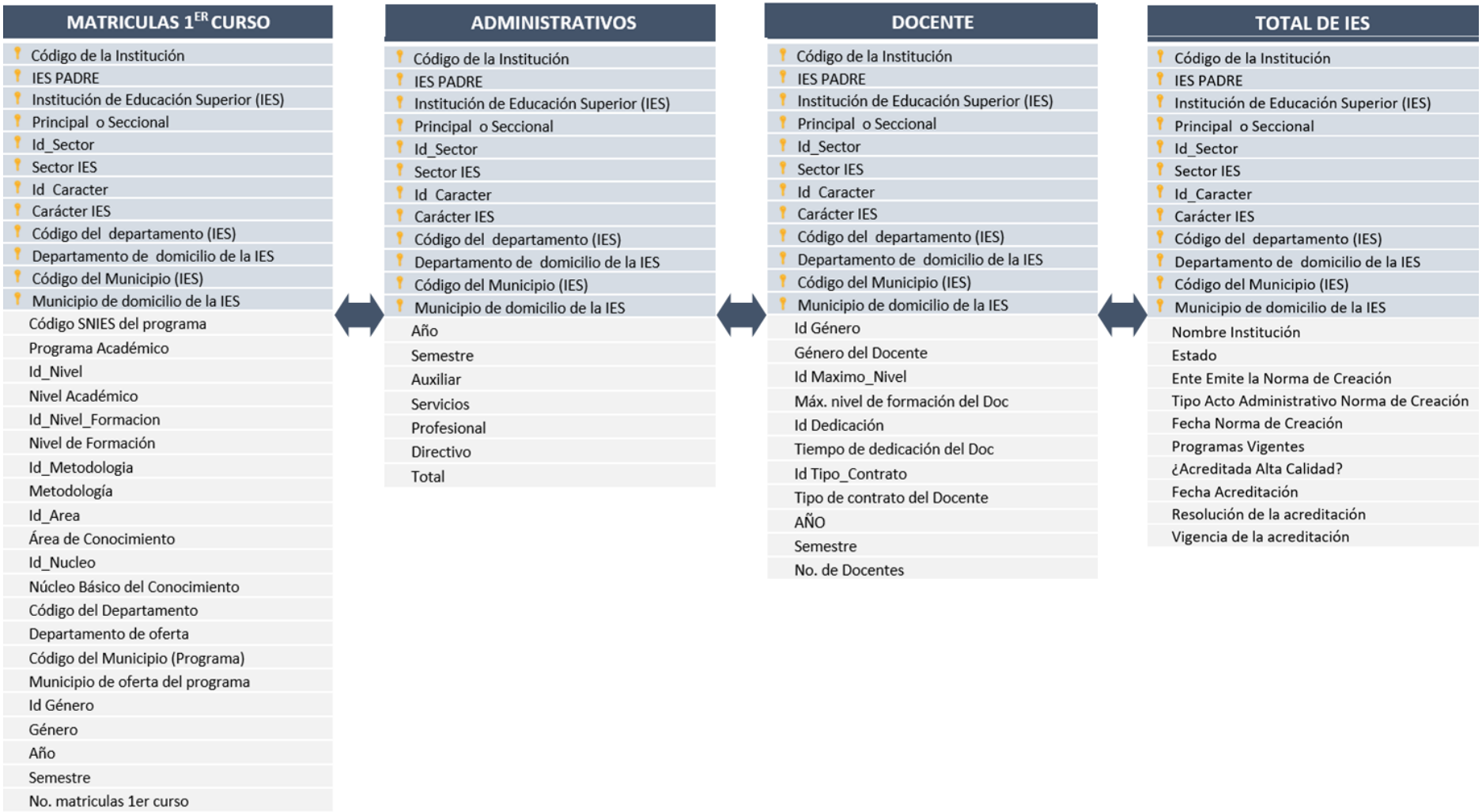



Figura *15 Estructura de las variables académicas e institucionales*
Fuente: Elaboración propia

Es necesario describir cada una de las variables, iniciando con las variables que tienen en comun las cuatro bases de datos, ver tabla 4.

Tabla 4 Descripción de variables comunes SNIES

| Variable | Tipo de variable | Valores |
|---|---|---|
| Código de la Institución | Numérica | [1000-9999] |
| IES PADRE | Numérica | [1000-9999] |
| Institución de Educación Superior (IES) | Texto | Razón Social de las entidades |
| Principal o Seccional | Texto | Principal<br>Seccional |
| Id_Sector<br>Sector IES | Numérica<br>Texto | 1 = Oficial<br>2 = Privada |
| Id_Caracter<br>Caracter IES | Numérica<br>Texto | 1 = Institución técnica<br>2 = Institución tecnológica<br>3 = Institución Universitaria<br>4 = Universidad |
| Código del departamento (IES) | Numérica | [1-100] |
| Departamento de domicilio de la IES | Texto | Nombre del Departamento |
| Código del Municipio (IES) | Numérica | [0000-99999] |
| Municipio de domicilio de la IES | Texto | Nombre del Municipio |

Fuente: Elaboración propia

En ese sentido se describen las características específicas del Total de las IES, es de resaltar la variable de Alta calidad en este subconjunto, debido a que se utilizara más adelante en la integración con las otras bases de datos, ver tabla 5.

Tabla 5 Descripción de variables IES

| Variable | Tipo de variable | Valores |
|---|---|---|
| Nombre Institución | Texto | Nombre entidades |
| Estado | Texto | ACTIVA<br>INACTIVA |
| Ente emite la norma de creación | Texto | Entidad gubernamental |
| Tipo Acto Administrativo Norma de Creación | Acto Administrativo | Texto |
| Fecha Norma de Creación | Fecha | dd/mm/aaaa |
| Programas Vigentes | Numérica | Natural |
| ¿Acreditada Alta Calidad? | Texto | SI<br>NO |
| Fecha Acreditación | Fecha | dd/mm/aaaa |
| Resolución de la acreditación | Numérica | Natural |
| Vigencia de la acreditación | Numérica | Natural |

Fuente: Elaboración propia

Igualmente, la tabla 6 describe las variables asociadas al perfil de los estudiantes matriculados en 1er semestre.

Tabla 6 Descripción de variables estudiantes

| Variable | Tipo de variable | Valores |
|---|---|---|
| Código SNIES del programa | Numérica | Natural |
| Programa Académico | Texto | Nombre del programa |
| Id_Nivel<br>Nivel Académico | Numérica<br>Texto | 0 = No aplica<br>1 = Pregrado<br>2 = Posgrado |
| Id_Nivel_Formacion<br>Nivel de Formación | Numérico<br>Texto | 0 = No aplica<br>1 = Especialización<br>2 = Maestría<br>3 = Doctorado<br>4 = Formación técnica profesional<br>5 = Tecnológica<br>6 = Universitaria<br>7 = Especialización Técnico Profesional<br>8 = Especialización Tecnológica<br>10 = Especialización Médico Quirúrgico<br>No aplica |
| Id_Metodología<br>Metodología | Numérico<br>Texto | 1 = Presencial<br>2 = Distancia Tradicional<br>3 = Virtual<br>No aplica |
| Id_Area<br>Área de Conocimiento | Numérico<br>Texto | 1 = Agronomía veterinaria y afines<br>2 = Bellas Artes<br>3 = Ciencias de la educación<br>4 = Ciencias de la salud<br>5 = Ciencias Sociales y Humanas<br>6 = Contabilidad, administración, contaduría y afines.<br>8 = Ingeniería, arquitectura, urbanismo y afines<br>9 = Matemáticas y ciencias naturales<br>0 = No aplica |
| Id_Nucleo<br>Núcleo Básico del Conocimiento (NBC) | Numérico<br>Texto | Numérico<br>Nombre del núcleo de conocimiento |
| Código del Departamento (Programa)<br>Departamento de oferta del programa | Numérico<br>Texto | Código del Departamento<br>Nombre del departamento |
| Código del Municipio (Programa)<br>Municipio de oferta del programa | Numérico<br>Texto | Código del municipio<br>Nombre del municipio |

Fuente: Elaboración propia

Tabla 6 (Continuación)

| Variable | Tipo de variable | Valores |
|---|---|---|
| Id Género<br>Género | Numérico<br>Texto | 1 = Masculino<br>2 = Femenino |
| Año | Numérico | 2014 al 2018 |
| Semestre | Numérico | 1,2 |
| Matriculados 1er Curso | Numérico | Natural |

Fuente: Elaboración propia

En la tabla 7, se describe las variables que especifican el perfil docente para cada entidad.

Tabla 7 Descripción de variables docente

| Variable | Tipo de variable | Valores |
|---|---|---|
| Id Género<br>Género del Docente | Numérico<br>Texto | 1 = Masculino<br>2 = Femenino |
| Id Maximo_Nivel<br>Máximo nivel de formación del docente | Numérico<br>Texto | 0 = No Informa<br>1 = Posdoctorado<br>2 = Doctorado<br>3 = Maestría<br>4 = Especialización<br>5 = Profesional<br>6 = Licenciatura<br>7 = Tecnólogo<br>8 = Técnico<br>9 = Formación técnica profesional<br>10 = Estudiante pregrado<br>11 = Especialización medico Quirúrgica |
| Id Dedicación<br>Tiempo de dedicación del Docente | Numérico<br>Texto | 1 = Tiempo Completo<br>2 = Medio tiempo<br>3 = Parcial<br>4 = Catedra<br>Sin información |
| Id Tipo_Contrato<br>Tipo de contrato del Docente | Numérico<br>Texto | 0 = Sin información<br>1 = Término indefinido<br>2 = Termino Fijo<br>3 = Horas<br>4 = Ocasional<br>5 = Ad honorem |
| AÑO | Numérico | 2014 a 2018 |
| Semestre | Numérico | 1<br>2 |
| No. de Docentes | Numérico | Natural |

Fuente: Elaboración propia

Por último, en la tabla 8, corresponde a los atributos relacionados con el nivel jerárquico del personal administrativo.

Tabla 8 Descripción de variables administrativas

| Variable | Tipo de variable | Valores |
|---|---|---|
| Año | Numérico | 2014 a 2018 |
| Semestre | | 1 2 |
| Auxiliar<br>Servicios<br>Profesional<br>Directivo<br>Total | | Natural |

Fuente: Elaboración propia

Para la verificación de la calidad de los datos se utilizó Open Refine, herramienta de código abierto gratuita que permite entre varias opciones la importación, la exploración y transformación de los datos.

En ese sentido se encontró la existencia de valores que correspondían a un valor numérico, registrados como cadena de caracteres y variables que asumían valores negativos, cuando por definición solo son posibles los números naturales, ver figura 16. Así que se realizó las transformaciones de cadenas de caracteres a variables numéricas y se transformaron los valores negativos a valores positivos.


Figura 16 Open Refine Valores de las variables - SNIES
Fuente: Elaboración propia

Adicionalmente, se encontraron valores no numéricos en variables definidas con numéricas, estrictamente, ver figura 17. Por lo que se hicieron las transformaciones a variables numéricas.

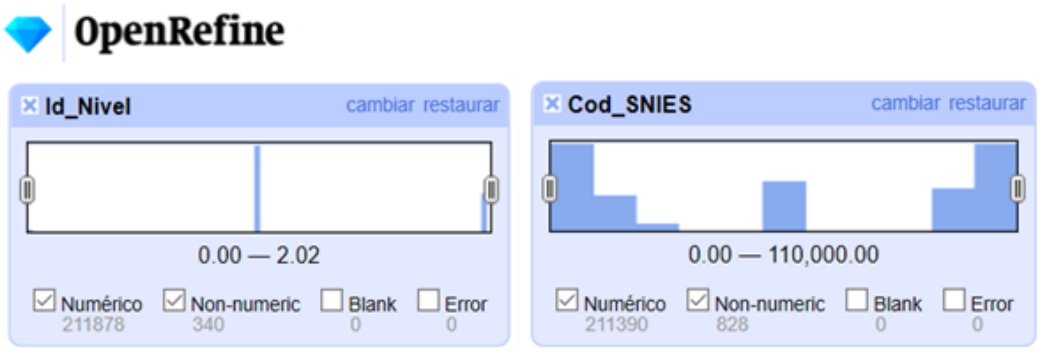



Figura 17 Open Refine Valores no numéricos - SNIES
Fuente: Elaboración propia.

Finalmente, existían diferentes valores no catalogados, que reportaron como sin clasificar, si información, no aplica, #¡REF! ver figura 18, los cuales fueron unificados con el valor de cero "0" en todos los campos.


Figura 18 Open Refine Valores nulos - SNIES
Fuente: Elaboración propia

Por otra parte, iniciando con la exploración estadística de las variables académicas las cuales se componen por el tipo de institución, la acreditación en alta calidad, las áreas de conocimiento y la metodología de estudio de las entidades. En lo que corresponde a instituciones acreditadas con alta calidad y resumiendo lo descrito en capítulos anteriores es de destacar que tan solo un 22% de las entidades se encuentran acreditadas en alta calidad.

En cuanto a las áreas de conocimiento, desde el año 2014 al 2018 se registra un mayor número de matrículas en el primer curso en las economías y las ingenierías, ver figura 19; aunque en general se registra un incremento en el año 2016 se observa una caída posterior manteniendo en los valores en las cifras registradas para el 2014.

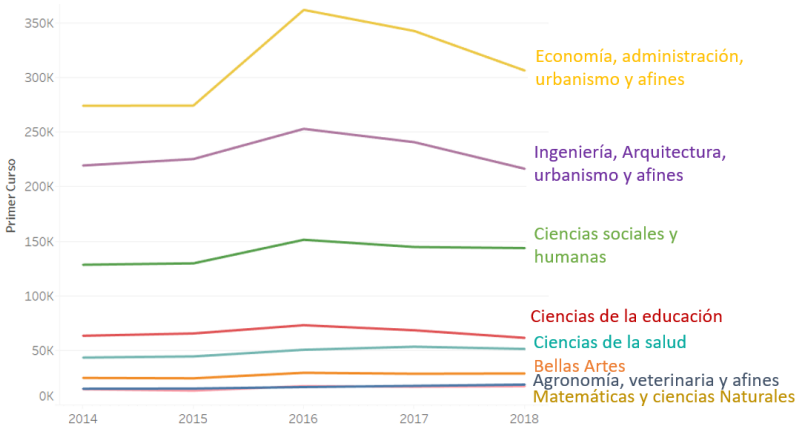



*Figura 19 Áreas de conocimiento - No. Matriculados en Primer Curso*
Fuente: Elaboración propia

En cuanto a la metodología de educación de las instituciones el Ministerio de educación clasifica tres modalidades; 1 Presencial, 2 Distancia y 3 Virtual; es evidente que la mayor proporción de metodología se encuentra enfocada a la modalidad presencial, ver figura 20.

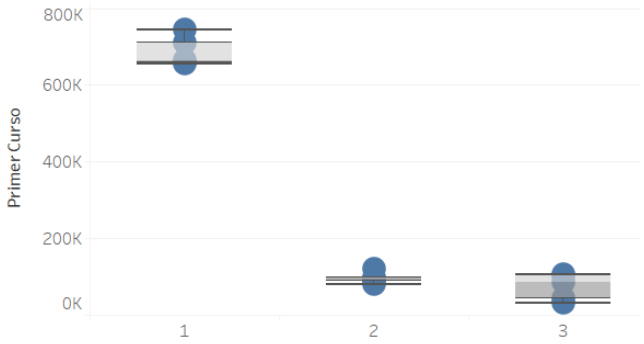



*Figura 20 Medidas de posición de Metodología - No. Matriculados en Primer Curso*
Fuente: Elaboración propia

Finalmente, el número de matriculas en Primer curso diferenciándolo por el sector al que pertenece la institución, en la figura 21, se evidencia una proporción parecida en los dos grupos. Adicionalmente se evidencia las agrupaciones del base de datos matriculados, desde un Matriculado como valor mínimo hasta un poco más de cuatro mil como valor máximo de agrupación.

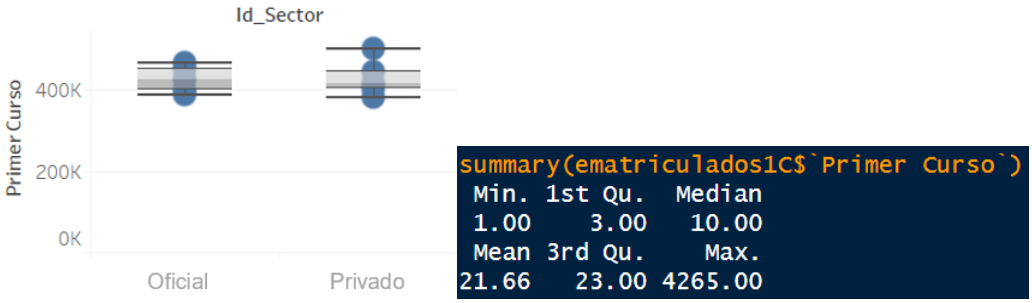



*Figura 21 Medidas de posición de Sector - No. Matriculados en Primer Curso*
Fuente: Elaboración propia

Por otro lado, las variables institucionales se componen por la caracterización del nivel de formación del personal administrativo y docente. Debido a que estos valores pueden variar según el sector al que pertenecen las instituciones, se describen teniendo en cuenta esta variable.

Iniciando con personal administrativo, es de resaltar que en general el sector privado registra un mayor número de personas den la totalidad de niveles, eso incluye el auxiliar, servicios, profesional y directivo, ver figura 22.

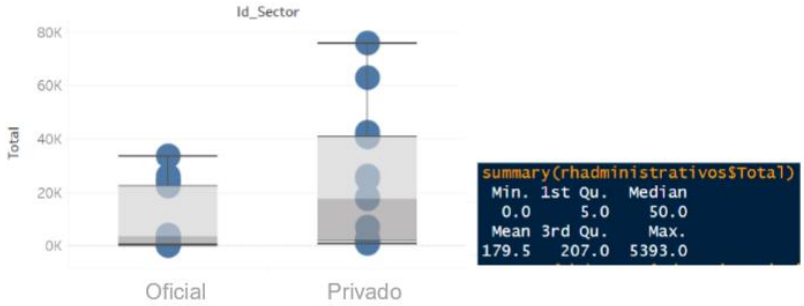



*Figura 22 Medidas de posición de Personal administrativo*
Fuente: Elaboración propia

En cuanto a la caracterización docente se observa en promedio anual, mayor número de docente con Maestría como nivel máximo de formación, ver figura 23, siendo el postdoctorado el nivel con menor promedio anual de docentes contratados.

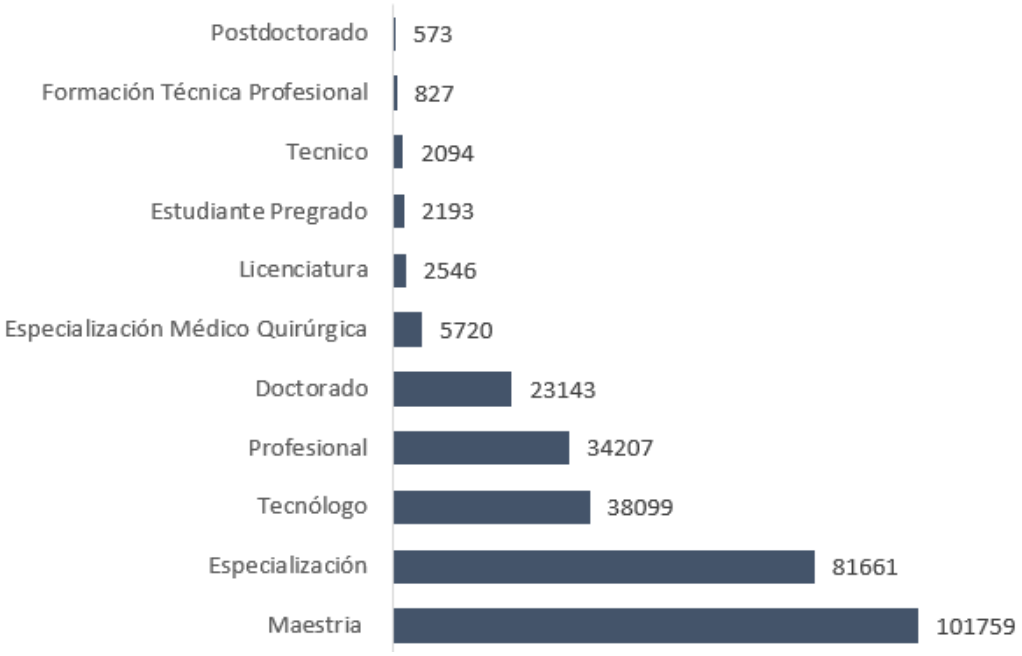



*Figura 23 Nivel Máximo de formación Promedio anual*
Fuente: Elaboración propia

Adicionalmente se puede observar mayor utilización docente en el sector privado que en el oficial, replicando el comportamiento del personal administrativo, ver figura 24.

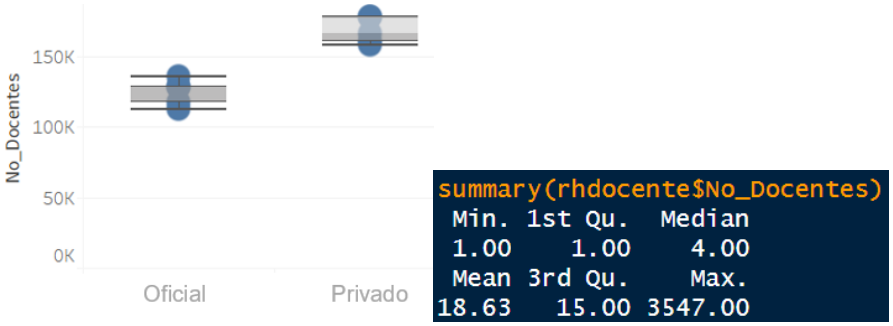



*Figura 24 Medidas de posición de Personal Docente*
Fuente: Elaboración propia

## 4.2.2. Comprensión de datos - Variables individuales y socioeconómicas

La presente sección describe el tratamiento realizado con las variables individuales y socioeconómicas, cuya fuente de información es el SPADIES. Este segundo conjunto de datos es el resultado de una serie de consultas realizadas en el sistema SPADIES del ministerio de educación superior, ver en la figura 25; el cual recolecta información de los créditos entregados por el Instituto Colombiano de Crédito Educativo y Estudios Técnicos en el Exterior -ICETEX-, la información de las pruebas de estado y situación socioeconómica registradas en los formularios de los exámenes realizados por el Instituto Colombiano para la Evaluación de la Educación –ICFES- y la información reportada por las IES.


*Figura 25 Consulta SPADIES*
Fuente: Elaboración propia

Las consultas realizadas mantienen el tipo de cálculo en IES; recorriendo tanto las variables que caracterizan a los estudiantes de primer semestre y los apoyos entregados el primer semestre; la información contiene la cantidad de estudiantes en cada una de las categorías como se describe la tabla 9 y 10 respectivamente.

Tabla 9 Descripción de variable caracterización de primer semestre

| Variable | Tipo de variable | Valores |
|---|---|---|
| Sexo | Mujer<br>Hombre<br>Sin información | Natural |
| Trabajaba al presentar el ICFES | No trabaja<br>Si trabaja<br>Sin información | Natural |
| Edad de presentación del ICFES | 15 ó menos años<br>16 a 20 años<br>21 a 25 años<br>26 ó más años<br>Sin información | Natural |
| Ingreso Hogar | Menos de un salario mínimo<br>Entre 1 y menos de 2 salarios mínimos<br>Entre 2 y menos de 3 salarios mínimos<br>Entre 3 y menos de 5 salarios mínimos<br>Entre 5 y menos de 7 salarios mínimos<br>Entre 7 y menos de 10 salarios mínimos<br>10 o más salarios mínimos<br>Sin información | Natural |
| Nivel del SISBEN | Nivel 1<br>Nivel 2<br>Nivel 3<br>Clasifica en otro nivel<br>No clasificado por SISBEN<br>Sin información | Natural |
| Estrato | Estrato 1 Estrato 2<br>Estrato 3 Estrato 4<br>Estrato 5 Estrato 6<br>Hogares no clasificados<br>Sin Información | Natural |
| Número de hermanos | Ninguno<br>1 2 3 4<br>Más de cuatro<br>Sin información | Natural |
| Posición entre sus hermanos | Primero<br>Segundo<br>Tercero<br>Cuarto<br>Quinto<br>Posterior al quinto<br>Sin información | Natural |

Fuente: Elaboración propia

Tabla 9 (Continuación)

| Variable | Tipo de variable | Valores |
|---|---|---|
| Ingreso de la familia del estudiante | Menos de un salario mínimo<br>Entre 1 y menos de 2 salarios mínimos<br>Entre 2 y menos de 3 salarios mínimos<br>Entre 3 y menos de 5 salarios mínimos<br>Entre 5 y menos de 7 salarios mínimos<br>Entre 7 y menos de 9 salarios mínimos<br>Entre 9 y menos de 11 salarios mínimos<br>Entre 11 y menos de 13 salarios mínimos<br>Entre 13 y menos de 15 salarios mínimos<br>15 o más salarios mínimos<br>Sin información | Natural |
| Vivienda propia | Posee Carece Sin información | Natural |
| Número de personas familia | 1 2 3<br>4 5 6<br>7 8 9<br>10 11 12<br>13 Sin información | Natural |
| Clasificación examen de estado | Bajo Medio Alto<br>Sin información | Natural |
| Nivel educativo de la madre | Nivel educativo primaria<br>Nivel educativo secundaria<br>Nivel educativo técnico o tecnológico<br>Nivel educativo superior y posgrado<br>Sin información | Natural |
| Área | Agronomía veterinaria y afines<br>Bellas artes<br>Ciencias de la educación<br>Ciencias de la salud<br>Ciencias sociales y humanas<br>Economía, administración, contaduría y afines<br>Ingeniería, arquitectura, urbanismo y afines<br>Matemáticas y ciencias naturales<br>Sin información | Natural |

Fuente: Elaboración propia

Tabla 10 Descripción de variables apoyos entregados por semestre

| Variable | Tipo de variable | Valores |
|---|---|---|
| Recibió/no recibió apoyo del ICETEX | Recibió<br>No recibió | Natural |
| Tipo de crédito ICETEX recibido | No recibí crédito<br>Largo plazo<br>Mediano Plazo<br>ACCES<br>Otro | Natural |
| Recibió/no recibió apoyo de las IES | Recibió<br>No recibió | Natural |

Fuente: Elaboración propia

Para la verificación de la calidad de los datos se utilizó Open Refine, herramienta de código abierto gratuita que permite entre varias opciones la importación, la exploración y transformación de los datos, al no encontrar errores en la consolidación de las consultas realizadas se procedió a la exploración de datos.

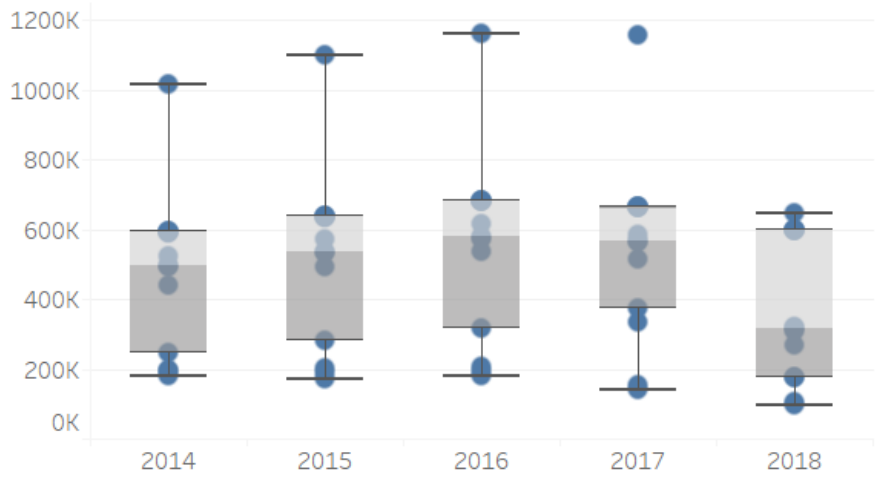


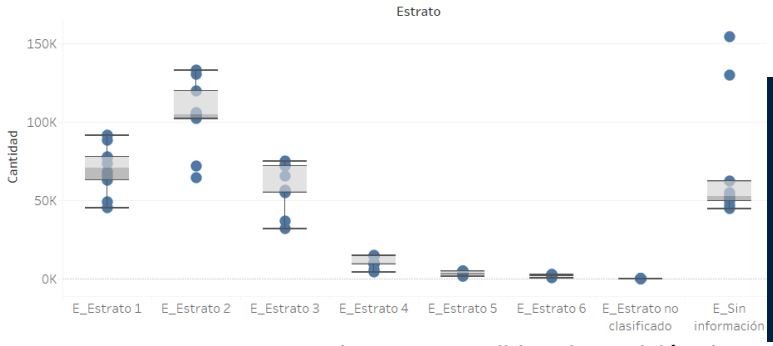

*Figura 26 Medidas de posición de variables SPADIES*
Fuente: Elaboración propia

La exploración de datos de las variables SPADIES componen el grupo socio económico y el individual de los alumnos de primer semestre del año 2014 al 2018. En la figura 26, se observa el diagrama de caja de la totalidad de las variables, así como el valor mínimo de estudiantes agrupados por cada una de ella que es 1, el valor máximo es de 360.750, a continuación, se desglosan los dos grupos y se describe las medidas de posicionamiento de cada una de las variables.

Iniciando con las variables socio económicas, las cuales corresponden a la caracterización de los alumnos de primer semestre en el país en el intervalo de tiempo de 2014 a 2018. En este grupo se encuentran el estrato, la situación laboral en el momento de presentar el examen de estado ICFES, el tipo de crédito ICETEX solicitado para iniciar sus estudios, el nivel educativo de la madre, los ingresos económicos totales del hogar, el nivel del SISBEN y la posesión o no de vivienda propia.

A continuación, se describe cada una de las variables, logrando observar de forma gráfica el comportamiento de cada uno de los valores que asume dicha variable; adicionalmente se identifica el valor mínimo, la media, la moda y los principales cuantiles de cada variable. Iniciando por el estrato, en la figura 27, la cual puede asumir desde el estrato 1 hasta el 6, puede no clasificarse o puede no contener información; en los diagramas de caja se observa como la mayor proporción corresponde al estrato 1 y 2, con una media de 39651 teniendo en cuenta la cantidad sin información y media de 41246 excluyéndolas, debido a que los valores

varían al tener en cuenta los campos sin información se calcula las medidas de posicionamiento en XEstrato, excluyendo los campos sin información.




```
> summary(Estrato$Cantidad)
   Min. 1st Qu.  Median
      1    2293   23290
   Mean 3rd Qu.    Max.
  39651   66278  154198
> summary(XEstrato$Cantidad)
   Min. 1st Qu.  Median
    545    3852   23290
   Mean 3rd Qu.    Max.
  41246   71753  133259
```

*Figura 27 Medidas de posición de Estrato*
Fuente: Elaboración propia

En segundo lugar, en la figura 28, se encuentra los ingresos en el hogar, la cual se encuentra por intervalos en Salarios Mínimos Legales Vigentes –SMLV-; iniciando por los hogares cuyos ingresos son inferiores a 1 SMLV, de 1 a 2, de 2 a 3, de 3 a 5, de 5 a 7, de 7 a 10, y los candidatos sin información. La mayor proporción de estudiantes de primer semestre reciben de 1 a 2 SMLV y una cantidad considerable no se tiene información de los ingresos familiares, por lo que también se recalcula los valores sin tener en cuenta los campos sin información. Al excluir los campos sin información valores máximos pasan de 153.270 a 143.357, debido a que los valores varían al tener en cuenta los campos sin información se calcula las medidas de posicionamiento en XIngHogar, excluyendo los campos sin información.

```
> summary(IngHogar$Cantidad)
   Min. 1st Qu.  Median
   1225    4834   29146
   Mean 3rd Qu.    Max.
  39651   55985  153270
> summary(XIngHogar$Cantidad)
   Min. 1st Qu.  Median
   1225    4652   20338
   Mean 3rd Qu.    Max.
  35534   55836  143357
```

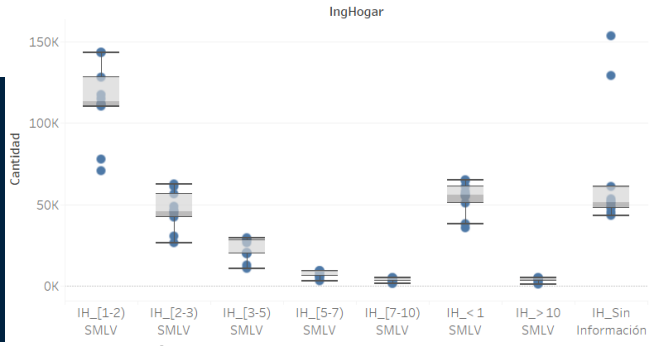
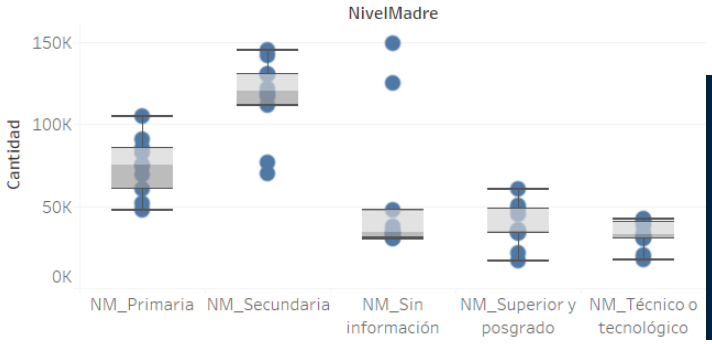



*Figura 28 Medidas de posición de Ingresos de Hogar*
Fuente: Elaboración propia

El nivel máximo educativo de las madres asume las variables por nivel; Primaria, Secundaria, técnico, superior y sin información; en general la mayor proporción de estudiantes de primer semestre se clasifica en el nivel educativo secundaria para sus madres; debido a que los valores varían al tener en cuenta los campos sin información se calcula las medidas de posicionamiento en XNivelMadre, excluyendo los campos sin información, ver figura 29.




```
> summary(NivelMadre$Cantidad)
   Min. 1st Qu.   Median
  16699   33851    47953
   Mean 3rd Qu.     Max.
  63442   84770   149224
> summary(XNivelMadre$Cantidad)
   Min. 1st Qu.   Median
  16699   34504    51284
   Mean 3rd Qu.     Max.
  65441   86920   145350
```

Figura 29 Medidas de posición de Nivel educativo de la madre
Fuente: Elaboración propia

El nivel de SISBEN asume valores; Nivel 1, Nivel 2, Nivel 3, clasificado en otro nivel, no clasificado y sin información; en general la mayor proporción de estudiantes de primer semestre se clasifica en el nivel 1 y sin información; debido a que los valores varían al tener en cuenta los campos sin información se calcula las medidas de posicionamiento en XNivelSISBEN, excluyendo los campos sin información, ver figura 30.

```
> summary(NivelSISBEN$Cantidad)
   Min. 1st Qu.   Median
   1114    9888    50924
   Mean 3rd Qu.     Max.
  52869   83785   185750
> summary(XNivelSISBEN$Cantidad)
   Min. 1st Qu.   Median
   1114    5086    21910
   Mean 3rd Qu.     Max.
  33840   58797   104166
```

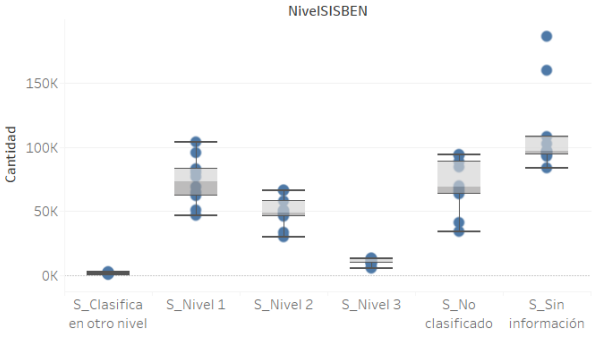



*Figura 30 Medidas de posición de Nivel de SISBEN*
Fuente: Elaboración propia

El tipo de crédito ICETEX, solicitado por los estudiantes de primer semestre para iniciar sus estudios se clasifica como tipo ACCES, largo plazo, mediano plazo, no recibió crédito, o recibió crédito; el tipo ACCES es un crédito a largo plazo, que cuenta con un periodo de gracia posterior a la terminación de estudios, y un posterior periodo de pago del crédito. En general la mayor proporción de estudiantes que inician sus estudios no reciben créditos y el tipo de crédito más usado es el ACCES en forma general, debido a que los valores varían al tener en cuenta los campos sin información se calcula las medidas de posicionamiento en XTipoCreditoICETEX, excluyendo los campos sin información, ver figura 31.

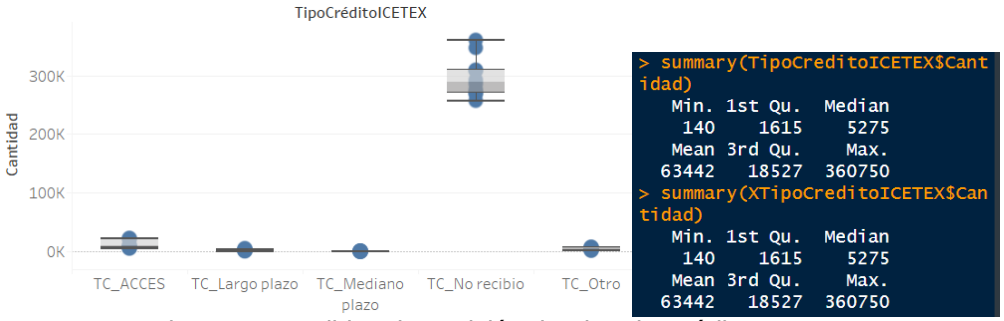



*Figura 31 Medidas de posición de Tipo de Crédito ICETEX*
Fuente: Elaboración propia

En cuanto a la situación laboral a la hora de presentar el examen de estado ICFES para el acceso a la educación superior, las variables que asumen son Si, para los que laboraban; No, para los que no lo hacían y los campos sin información. En general la mayor proporción de estudiantes, si laboraban al presentar el examen de estado, lo que podría explicar la relación de créditos solicitados; debido a que los valores varían al tener en cuenta los campos sin información se calcula las medidas de posicionamiento en XTrabajoPresICFES, excluyendo los campos sin información, ver figura 32.

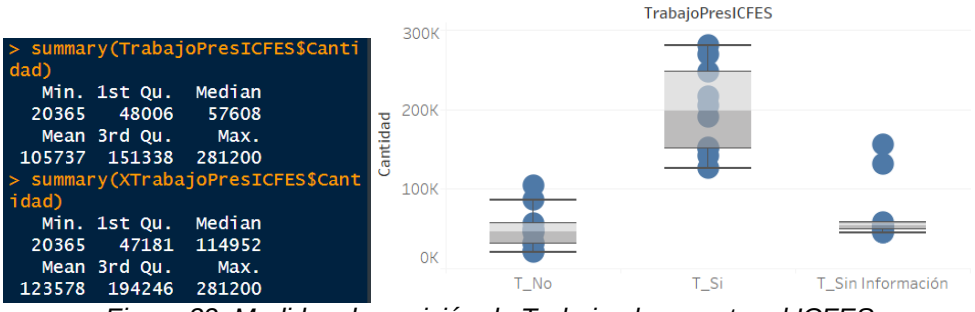



*Figura 32 Medidas de posición de Trabajo al presentar el ICFES*
Fuente: Elaboración propia

En cuanto a la posesión de vivienda, la variable puede asumir el hecho de poseer la vivienda, carecer de ella o no tener información de la variable. En general no se tiene información de la mayor proporción de estudiantes de primer semestre; debido a que los valores varían al tener en cuenta los campos sin información se calcula las medidas de posicionamiento en XVivienda, excluyendo los campos sin información, ver figura 33.

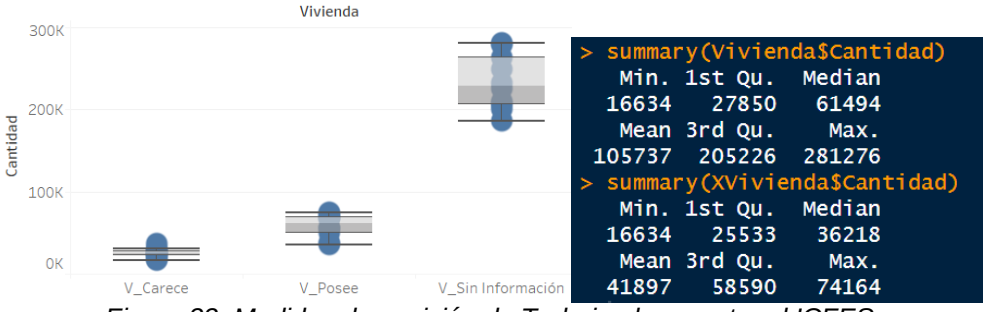
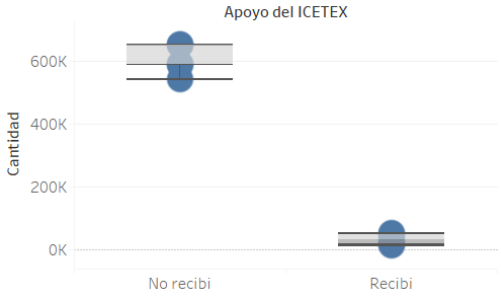



*Figura 33 Medidas de posición de Trabajo al presentar el ICFES*
Fuente: Elaboración propia

Por otro lado, la exploración de datos cuenta con las variables individuales, corresponden a la caracterización de los alumnos de primer semestre en el país de 2014 a 2018. En este grupo se encuentran el apoyo de entidades externas para los estudios, como el ICTEX o las IES, la edad, el género, el número de hermanos, la posición de hermanos y el número de personas en la familia. En cuanto a los apoyos recibidos por el ICETEX, se encuentran las opciones recibí y no recibí. En general la mayor proporción de estudiantes, no recibieron apoyo del ICETEX, con una media de 158.606 para el grupo en general, ver figura 34.




```
> summary(ApoyoICETEX$Cantidad)
   Min. 1st Qu.  Median
   5202   15260  146748
   Mean 3rd Qu.    Max.
 158606  292979  360031
```

*Figura 34 Medidas de Posición de Apoyo del ICETEX*
Fuente: Elaboración propia

En cuanto a los apoyos recibidos por las IES, se encuentran las opciones recibí y no recibí. En general la mayor proporción de estudiantes, no recibieron apoyo de las IES, con una media de 158.606 para el grupo en general, ver figura 35.

```
> summary(ApoyoIES$Cantidad)
   Min. 1st Qu.  Median
  96519  140285  168320
   Mean 3rd Qu.    Max.
 158606  178644  193649
```

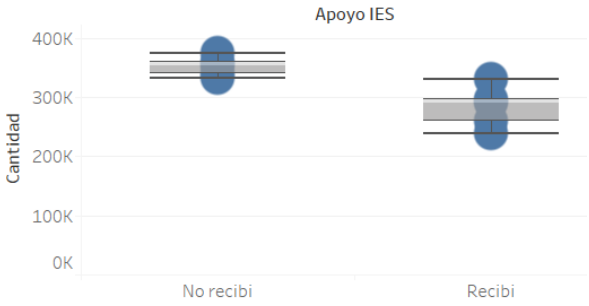



*Figura 35 Medidas de Posición de Apoyo del IES*
Fuente: Elaboración propia

En cuanto a la Edad de los estudiantes de primer semestre, las variables que asumen rangos de edad de 15 o menos, de 16 a 20, de 21 a 25, de 26 o más y los campos sin información; en general la mayor proporción de estudiantes de primer semestre se encuentra en el rango de edad de 16 a 20 años; debido a que los valores varían al tener en cuenta los campos sin información se calcula las medidas de posicionamiento en XEdad, excluyendo los campos sin información, ver figura 36.

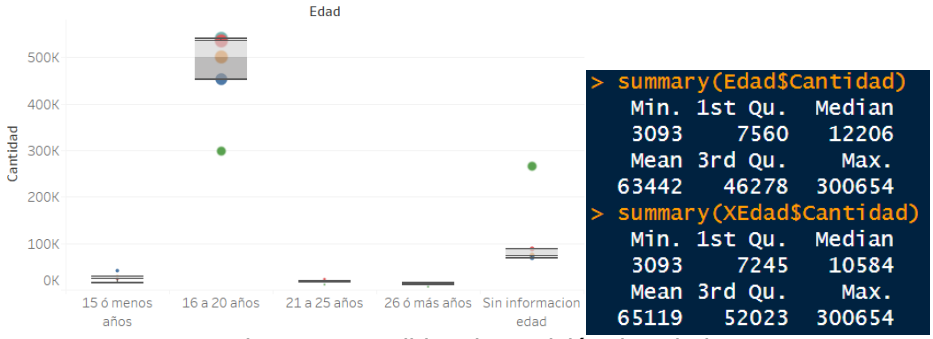



*Figura 36 Medidas de Posición de Edad*
Fuente: Elaboración propia

En cuanto al Género, las variables que asumen son hombre, mujer y sin información; en general la mayor proporción de estudiantes de primer semestre son mujeres; debido a que los valores varían al tener en cuenta los campos sin información se calcula las medidas de posicionamiento en XGenero, excluyendo los campos sin información, ver figura 37.

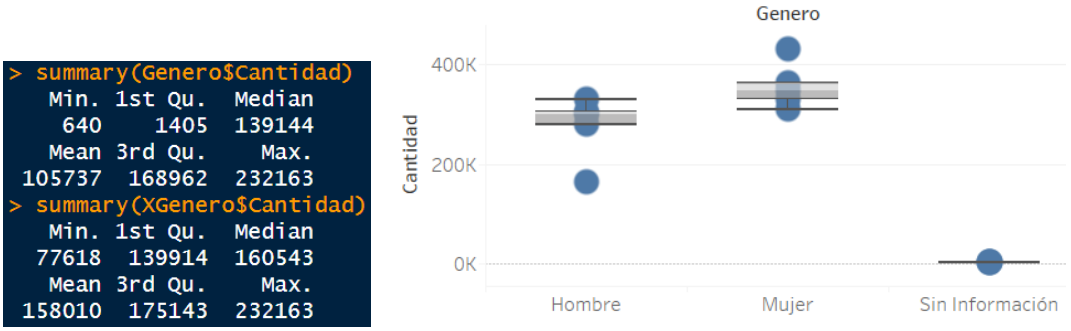



*Figura 37 Medidas de Posición de Genero*
Fuente: Elaboración propia

En cuanto al número de hermanos, las variables que asumen son 1, 2, 3, 4, más de cuatro, ninguno y sin información. En general la mayor proporción de estudiantes no cuenta con información; se calcula las medidas de posicionamiento en XHermanos, excluyendo los campos sin información, ver figura 38.

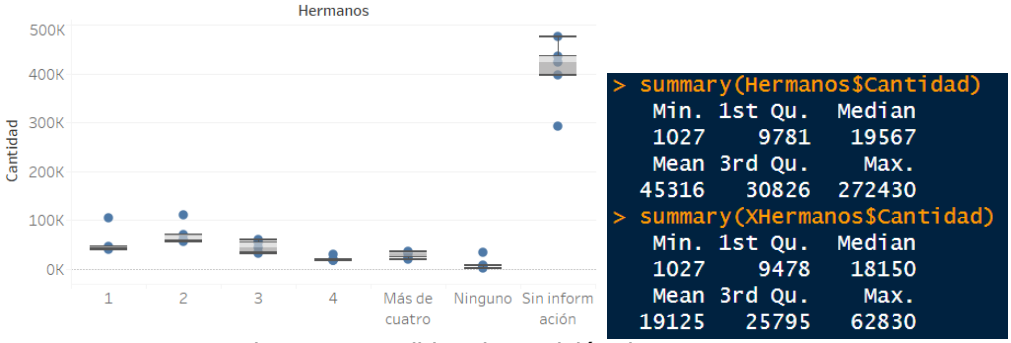



*Figura 38 Medidas de Posición de Hermanos*
Fuente: Elaboración propia

En cuanto a la posición entre los hermanos, los estudiantes de primer semestre pueden asumir variables en las posiciones del primero al quinto, posterior al quinto y sin información. En

general no se tiene información de la mayor proporción de estudiantes, se calcula las medidas de posicionamiento en XPosHermanos, excluyendo los campos sin información, ver figura 39.

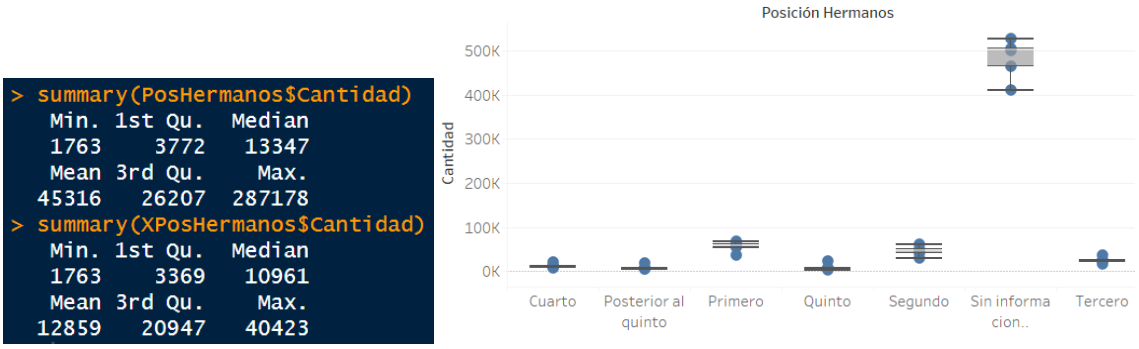



*Figura 39 Medidas de Posición de Posición de Hermanos*
Fuente: Elaboración propia

En cuanto al número de personas en la familia, los valores que pueden asumir las variables se encuentran en el rango de 1 a 13 y sin información. En general la mayor proporción de estudiantes, se clasifica en cuatro personas por familia; debido a que los valores varían al tener en cuenta los campos sin información se calcula las medidas de posicionamiento en XNo_personas_familia, excluyendo los campos sin información, ver figura 40.

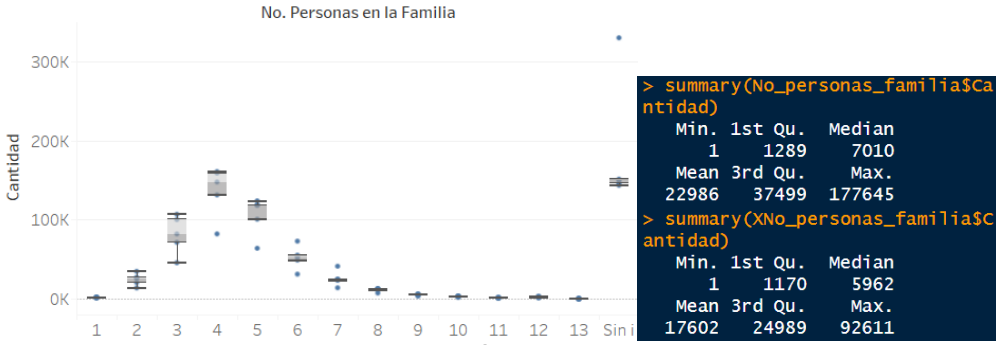



*Figura 40 Medidas de Posición de No. personas en la Familia*
Fuente: Elaboración propia

## 4.3. Fase 3: Preparación de los datos

Esta sección pretende preparar los valores para el proceso de modelado, con la creación de nuevos atributos que faciliten el proceso.

### 4.3.1. Preparación de datos - Variables académicas e institucionales

La preparación de las variables académicas e institucionales, inicia con la variación en la estructura actual de los datos, iniciando con una integración de la base de datos IES con Matriculados Primer Curso, Administrativos y Docentes por medio del campo código institución se determina si las instituciones de cada registro están acreditadas o no en alta calidad.

Continuando con el cambio de estructura se totalizan el número de matriculados en primer curso, docentes y administrativos según la demás especificación existentes en cada base de datos, en las que se incluye los departamentos, debido a la brecha existente entre los mismos, la cual fue observada en capítulos anteriores.

Finalmente se realiza una discretización de los campos numéricos con base en los cuartiles obtenidos de los nuevos subconjuntos de datos, ver figura 41, con el fin de definir la clase de los subconjuntos en bajo, medio, alto y muy alto.

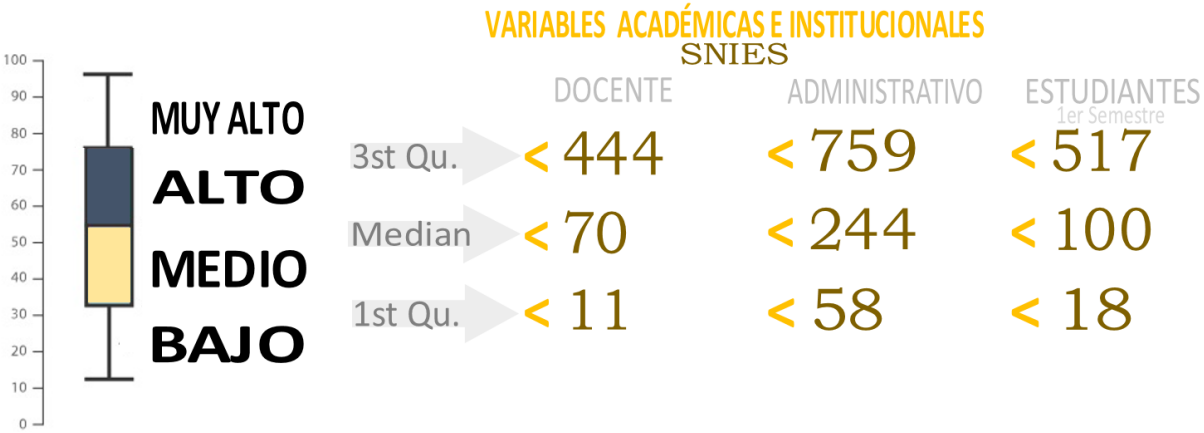



*Figura 41 Discretización variables académicas e institucionales*
Fuente: Elaboración propia

En ese sentido se deben correr tres modelados de datos; en el primer subconjunto docentes se tendrá en cuenta los atributos alta calidad, sector, máximo nivel, dedicación, código departamento del programa y la clase; el segundo subconjunto del personal administrativo tendrá en cuenta la alta calidad, sector, carácter, código departamento del programa, nivel administrativo y la clase; el ultimo es la caracterización de estudiantes de primer semestre el cual tiene como atributos la alta calidad, sector, carácter, nivel, nivel de formación, metodología, área, código departamento del programa y la clase.

### 4.3.2. Preparación de datos -Variables individuales y socioeconómicas

La preparación de las variables individuales y socioeconómicas consiste en una discretización de los campos numéricos con base en los cuartiles obtenidos en el conjunto total de variables SPADIES, según la figura 42. En el modelado de las variables SPADIES, se tienen en cuenta los atributos año, periodo, variable, sub variable y clase.

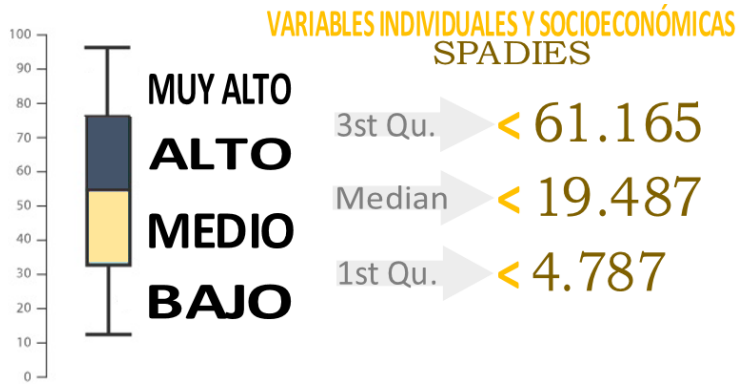



*Figura 42 Discretización variables individuales y socioeconómicas*
Fuente: Elaboración propia

## 4.4. Fase 4: Modelado de los datos

Las fases de modelado de datos ejecutaran algoritmos en weka, para identificar posibles patrones en los subconjuntos de datos.

### 4.4.1. Modelado de datos - Variables académicos e institucionales

Iniciando con las variables institucionales los cuales corresponden a formación o cantidad de los docentes y personal administrativo contratados en las entidades, el problema consiste en determinar cómo influyen las instancias en la contratación del personal en las IES; respondiendo a la pregunta de si ¿existe alguna relación en la cantidad de personal contratado con respecto a las variables institucionales?

El proceso de modelado inicia identificando en diferentes tipos de algoritmos, cual contiene mayor número de instancias correctamente clasificadas para cada uno de los grupos. Para el modelado de los docentes y administrativos se corrieron cuatro algoritmos de árbol y tres algoritmos de reglas, los que clasificaron mayores instancias correctamente fueron J48 y JRIP como se ve en la tabla 11, en los dos casos.

*Tabla 11 Modelado variable académica / Sector Oficial*

| Tipo de algoritmos | Nombre | Correctly Classified Instances | |
|---|---|---|---|
| | | Administrativo | Docente |
| Árbol | DecisionStump | 31% | 29% |
| | HoeffdingTree | 35% | 48% |
| | Id3 | 35% | 47% |
| | J48 | 36% | 49% |
| Reglas | ZeroR | 26% | 25% |
| | JRip | 30% | 38% |
| | Prism | 24% | 28% |

Fuente: Elaboración propia

En el Anexo I y II corresponden a las salidas JRip y J48 del personal administrativo respectivamente, así como los Anexos III y IV las cuales corresponden a las salidas J48 y JRip del personal docente respectivamente.

Por otro lado, en cuanto al análisis de variables académicas se utilizó un modelado de las matriculas en primer semestre, se tiene en cuenta el sector de la universidad, la acreditación en alta calidad, el nivel de formación en el que se inscribieron, la metodología de estudio, el área de estudio y el departamento. Los que clasificaron mayores instancias correctamente fueron J48 y JRIP como se ve en la tabla 12, en los dos casos.

*Tabla 12 Modelado variable académica / Sector Privado*

| Tipo de algoritmos | Nombre | Correctly Classified Instances Estudiantes |
|---|---|---|
| Árbol | DecisionStump | 34% |
| | HoeffdingTree | 41% |
| | Id3 | 44% |
| | J48 | 42% |
| Reglas | ZeroR | 25% |
| | JRip | 37% |
| | Prism | 30% |

Fuente: Elaboración propia

En el Anexo V y VI se encuentran las salidas weka correspondiente a personal estudiantil con los algoritmos J48 y JRip, respectivamente.

## 4.4.2. Modelado de datos - Variables individuales y socioeconómicas

En la fase de modelado de datos, se ejecutó los datos en diferentes tipos de algoritmos tanto de árbol, como en reglas de decisión, el porcentaje que arrojaron de instancias correctamente clasificada se encuentra a continuación en la tabla 13.

*Tabla 13 Modelado variables individuales y socioeconómicas*

| Tipo de algoritmos | Nombre | Correctly Classified Instances SPADIES |
|---|---|---|
| Árbol | DecisionStump | 27% |
| | HoeffdingTree | 83% |
| | Id3 | 84% |
| | **J48** | **84%** |
| Reglas | ZeroR | 25% |
| | **Jrip** | **82%** |
| | Prism | 70% |

Fuente: Elaboración propia

En el Anexo VII y VIII se adjuntan las salidas weka J48 y JRip correspondientes al modelado de variables individuales y socioeconómicas.

# 4.5. Fase 5: Evaluación de los datos

Basados en las salidas de los anexos anteriormente mencionados, se especifica a continuación los resultados teniendo en cuenta los criterios del problema anteriormente mencionados.

## 4.5.1. Evaluación de datos - Variables académicos e institucionales

Iniciando con las variables institucionales las cuales buscan describir el comportamiento en el nivel educativo del personal docente y administrativo de las universidades en el país; en cuanto a la contratación del personal administrativo en las IES en Colombia, ver figura 43, se observan las siguientes tendencias:

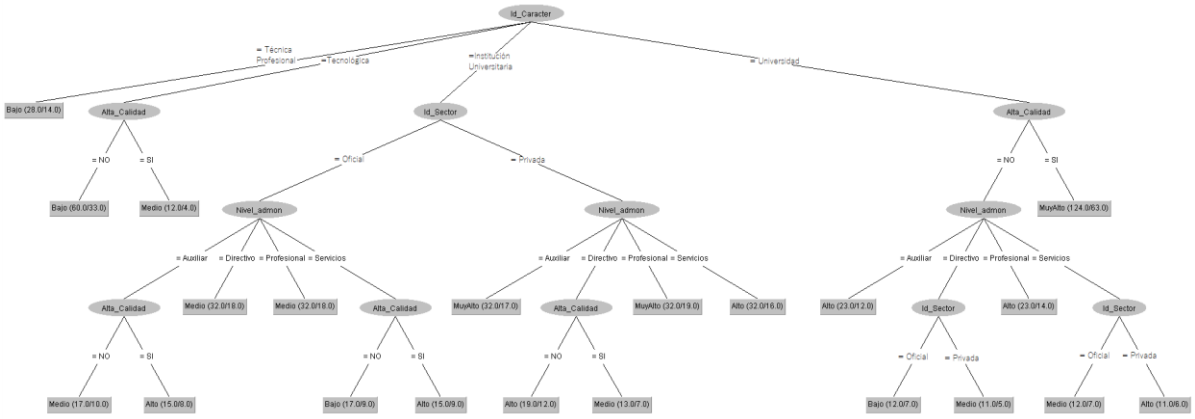



*Figura 43 Diagrama de árbol – Institucionales administrativos*
Fuente: Elaboración propia

- Las instituciones técnicas profesionales utilizan baja cantidad de personal administrativo; tanto en el nivel auxiliar, profesional, servicio y directivo.
- Las instituciones tecnológicas acreditadas en alta calidad utilizan mayor proporción de personal administrativo en todos los niveles, con respecto a las que no se encuentran acreditadas del mismo tipo.
- Las instituciones universitarias del sector oficial usan menor proporción de personal administrativo con respecto a las instituciones privadas.
- Las instituciones universitarias del sector oficial acreditadas en alta calidad tienden a utilizar mayor número de personal auxiliar y de servicios.
- Las instituciones universitarias del sector privado tienden a utilizar mayor número de personal auxiliar y profesional.

- Las instituciones de educación superior clasificadas como universidades, acreditadas en alta calidad tienden a utilizar la proporción mayor de personal administrativo; tanto en el nivel auxiliar, profesional, servicio y directivo.
- Las universidades no acreditadas en alta calidad utilizan proporción alta en el nivel auxiliar y profesional.
- Las universidades privadas no acreditadas en alta calidad utilizan mayor proporción de personal administrativo directivo y de servicios, en comparación al sector oficial.

En cuanto a la contratación del personal docente en las IES en Colombia, en la figura 44, se observan algunas tendencias para docentes con máximo nivel de formación posdoctorado, doctorado, maestría y especialización:

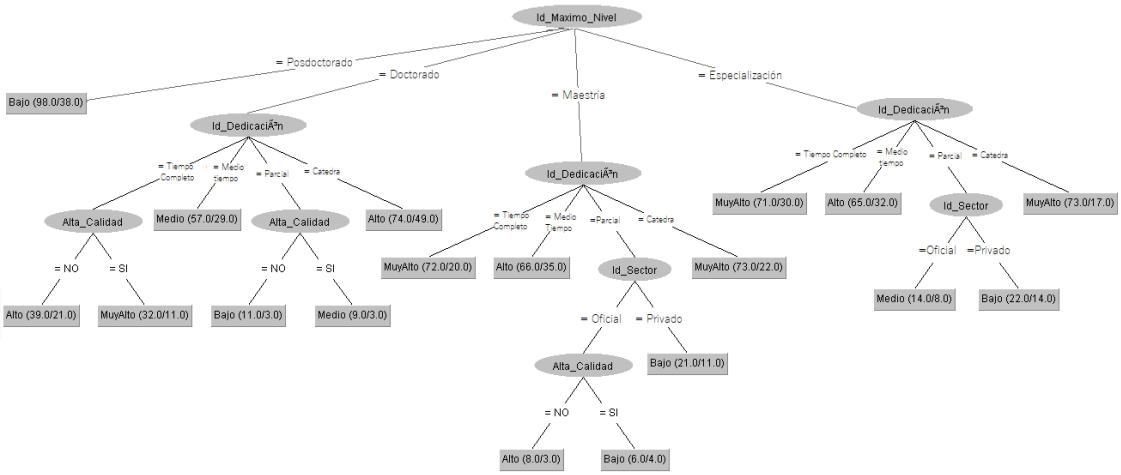



*Figura 44 Diagrama de árbol – Institucionales docentes parte I*
Fuente: Elaboración propia

- La contratación baja en docentes con posdoctorado como máximo nivel de formación.
- Las contrataciones altas en tiempo completo para docentes con doctorado como máximo nivel de formación, siendo mayores en las entidades acreditadas en alta calidad.
- Las contrataciones altas en catedra para docentes con doctorado.
- Al igual que en el doctorado, los docentes con maestría como máximo nivel de formación tiene mayores tasas de contratación en el tiempo completo y en catedra.

- Una de las particularidades es la contratación alta en la modalidad parcial para docentes con maestría en instituciones del sector oficial no acreditadas en alta calidad.

- Las contrataciones son igualmente altas en modalidad tiempo completo y catedra para los docentes con formación de especialización.

En cuanto a los docentes con máximo nivel de formación profesional, licenciatura y tecnólogo; ver figura 45, se observan las siguientes tendencias:

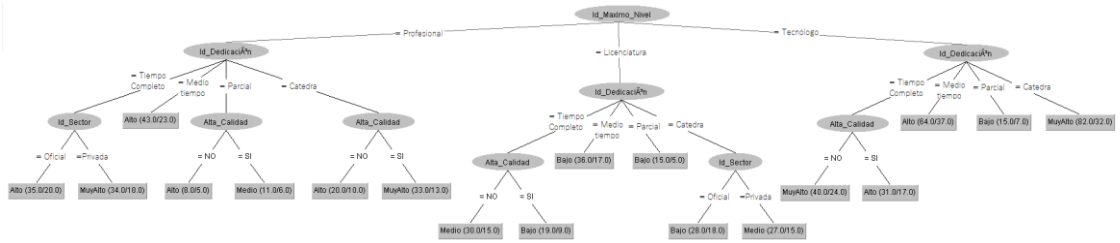



*Figura 45 Diagrama de árbol – Institucionales docentes parte II*
Fuente: Elaboración propia

- Las contrataciones son altas en cualquiera de las modalidades de contratación para los docentes con el profesional como máximo nivel de formación; siendo puntualmente menor las contrataciones parciales en instituciones acreditadas en alta calidad.

- Las contrataciones son bajas en cualquiera de las modalidades de contratación para los docentes con licenciatura, siendo menores en las acreditadas en alta calidad y las de sector oficial.

- Particularmente el nivel tecnólogo también cuenta con altas tasas de contratación docente en las modalidades de tiempo completo, medio tiempo y catedra. Bajas contrataciones en modalidad parcial.

Para finalizar con las variables institucionales, en la figura 46 se observan los docentes con máximo nivel de educación técnico, técnica, estudiantes de pregrado y especialización medico quirúrgica, en las cuales se observan siguientes tendencias:

- Los docentes con máximo nivel de formación técnico, en general presentan bajas tasas de contratación en todas las modalidades, aunque el contrato en tiempo completo es menor en las intuiciones acreditadas en alta calidad.

- El docente en formación técnica profesional es baja en todas las modalidades de contratación.

- En general los estudiantes de pregrado también presentan bajas contrataciones en la mayoría de las modalidades de contratación, solo se incrementa en instituciones de alta calidad con modalidad catedra.

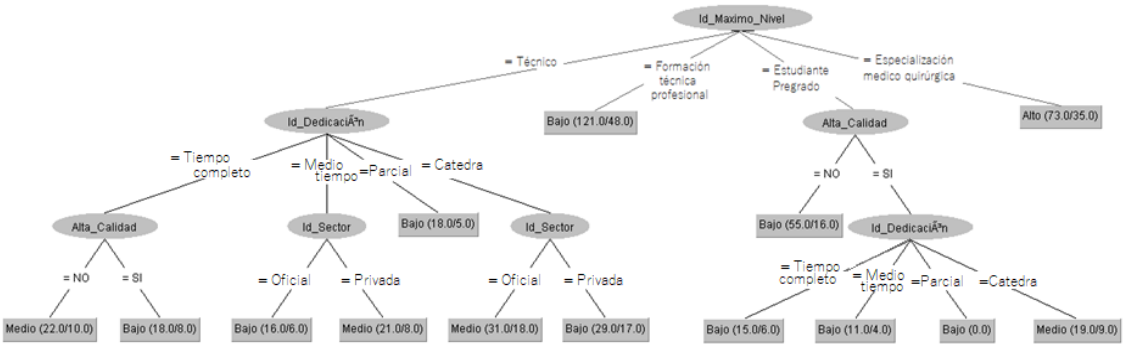



*Figura 46 Diagrama de árbol – Institucionales docentes parte III*
Fuente: Elaboración propia

Por otro lado, esta sección considera las variables académicas, como el tipo de institución, la acreditación en alta calidad, la metodología de estudio y el área de conocimiento; relacionado directamente con el número de matrículas de estudiantes de primer semestre. Como cada departamento se comporta diferente, se caracterizará cada uno en los siguientes apartados.

El departamento de Antioquia, paso de tener una tasa de cobertura de educación superior de 51% al 58%, del 2014 al 2018 respectivamente. En la figura 47, se evidencia el comportamiento de las matriculas de primer semestre en el mismo transcurso de tiempo; algunas de las tendencias que se observan son:

- La tendencia en el número de matrículas en primer semestre en el departamento de Antioquia tanto en el nivel pregrado como en el posgrado en la modalidad presencial.
- En el nivel de pregrado, en la metodología presencial se evidencia mayor proporción de matrículas en las instituciones no acreditadas en alta calidad.
- En el nivel de pregrado, en la metodología a distancia tradicional se evidencia mayor proporción de matrículas en las instituciones no acreditadas en alta calidad; patrón que se repite en la modalidad virtual de las universidades.
- En el nivel de posgrado, tanto la metodología virtual y la distancia tradicional presenta proporción media de matrículas, siendo mayor los posgrados en modalidad presencial de las universidades privadas.

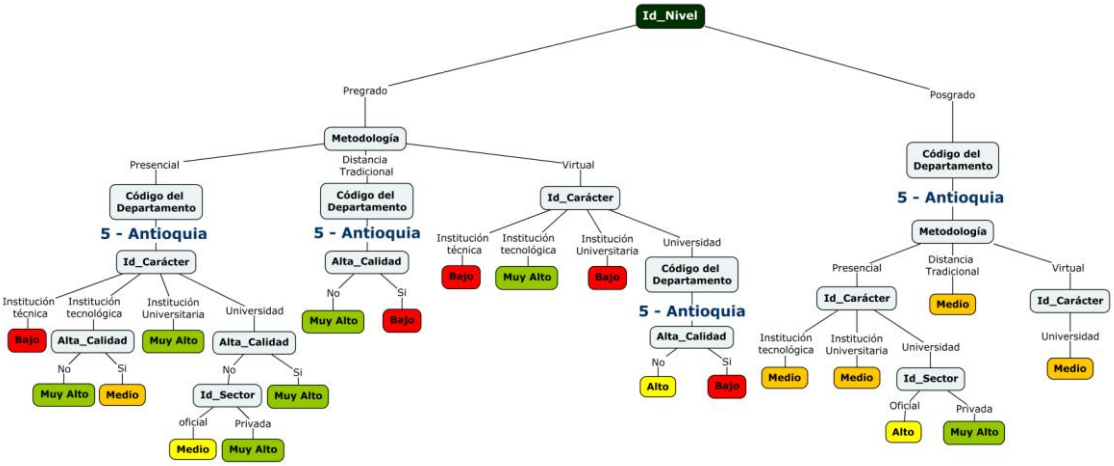



*Figura 47 Diagrama de árbol – Académicas Antioquia*
Fuente: Elaboración propia

El departamento de Atlántico, paso de tener una tasa de cobertura de educación superior de 56% al 59%, del 2014 al 2018 respectivamente. En la figura 48, se evidencia el comportamiento de las matriculas de primer semestre en el mismo transcurso de tiempo; algunas de las tendencias que se observan son:

- En el nivel de pregrado, el mayor número de matrículas en modalidad presencial; pero en la modalidad de virtual y en la distancia tradición las matriculas son medias.
- En el nivel de pregrado, las instituciones universitarias acreditadas en alta calidad tienden a tener mayor número de matrículas que las que no.
- En el nivel de posgrado, el mayor número de matrículas se encuentran en las instituciones privadas de metodología presencial.

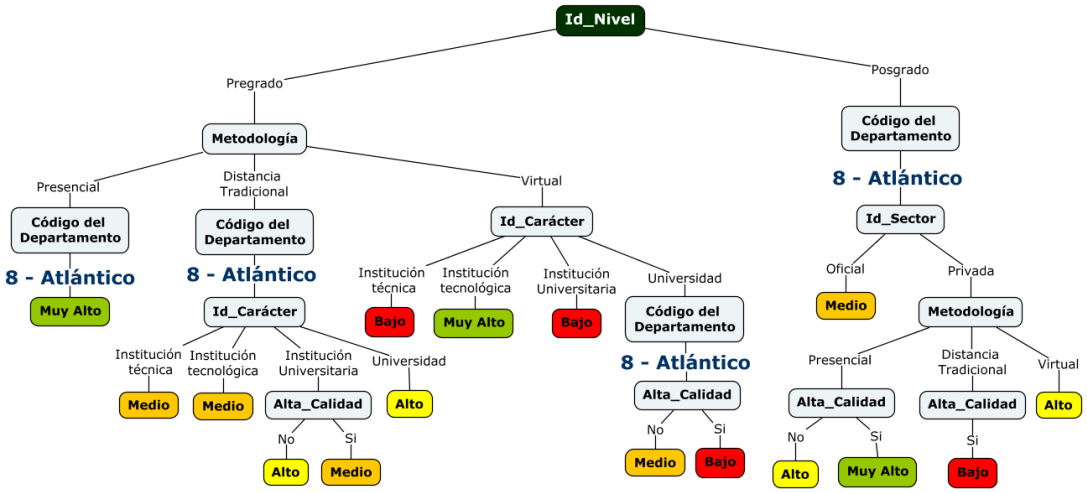



*Figura 48 Diagrama de árbol – Académicas Atlántico*
Fuente: Elaboración propia

El departamento de Bogotá, paso de tener una tasa de cobertura de educación superior de 98% al 113%, del 2014 al 2018 respectivamente. En la figura 49, se evidencia el comportamiento de las matriculas de primer semestre en el mismo transcurso de tiempo; algunas de las tendencias que se observan son:

- En el nivel de pregrado, tanto la modalidad presencial como virtual tiene aceptación alta; al igual que los programas con metodologías de distancia tradicional en instituciones técnicas y universidades.

- El nivel de posgrados, es mayor en instituciones universitarias y universidades; teniendo mayor aceptación en las entidades privadas.

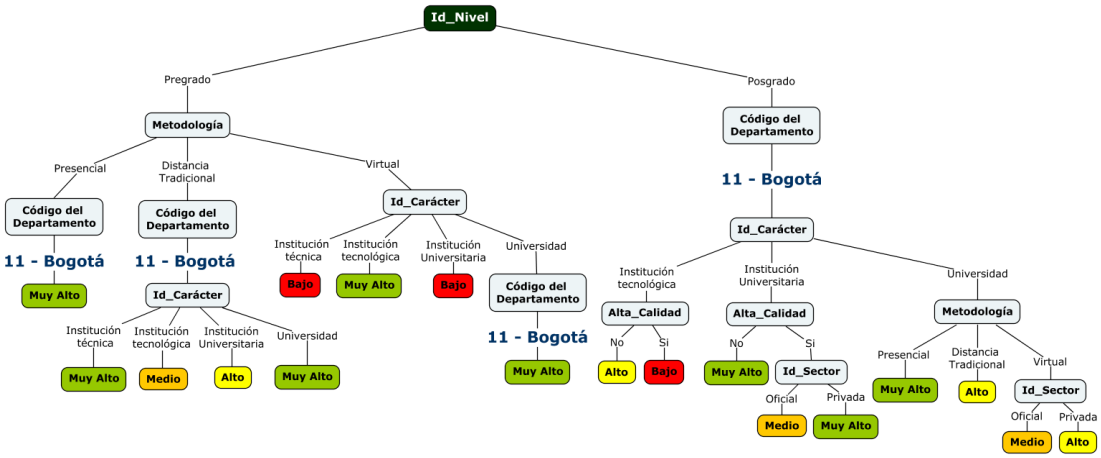



*Figura 49 Diagrama de árbol – Académicas Bogotá*
Fuente: Elaboración propia

El departamento de Bolívar, paso de tener una tasa de cobertura de educación superior de 35% al 36%, del 2014 al 2018 respectivamente. En la figura 50, se evidencia el comportamiento de las matriculas de primer semestre en el mismo transcurso de tiempo; algunas de las tendencias que se observan son:

- La tendencia en el número de matrículas en primer semestre en el departamento de Bolívar tanto en el nivel pregrado como en el posgrado en la modalidad presencial, siendo mayor la población de pregrados.

- En el nivel de pregrado, en la metodología presencial y la distancia tradicional se evidencia mayor proporción de matrículas.

- En el nivel de pregrado, en la metodología a distancia tradicional se evidencia mayor proporción de matrículas en las instituciones no acreditadas en alta calidad.

- En el nivel de posgrado, en la metodología presencial con instituciones de carácter universitario llevan la mayor proporción de matrículas.

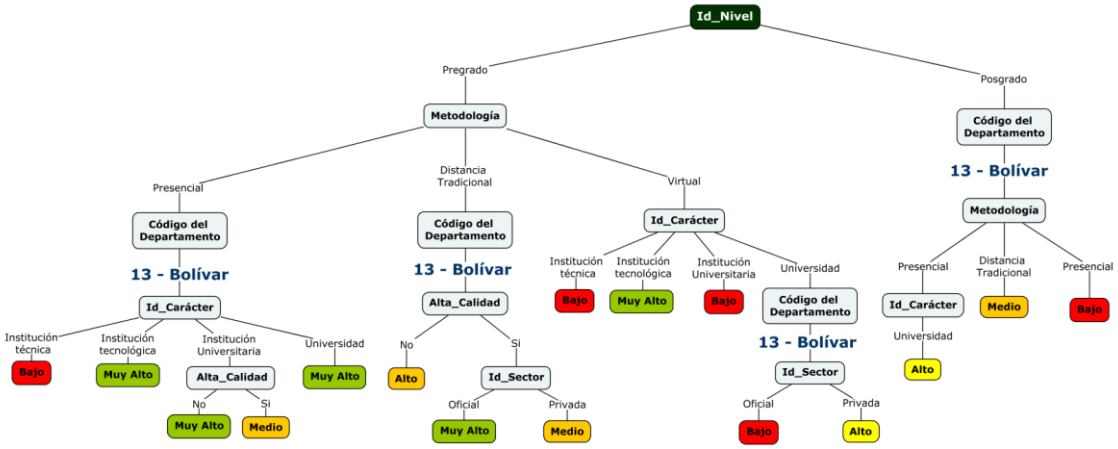



*Figura 50 Diagrama de árbol – Académicas Bolívar*
Fuente: Elaboración propia

El departamento de Boyacá, paso de tener una tasa de cobertura de educación superior de 50% al 56%, del 2014 al 2018 respectivamente. En la figura 51, se evidencia el comportamiento de las matriculas de primer semestre en el mismo transcurso de tiempo; algunas de las tendencias que se observan son:

- La tendencia en el número de matrículas en primer semestre en el departamento de Boyacá en el nivel pregrado es la modalidad presencial.
- En el nivel de pregrado, en la metodología a distancia tradicional se evidencia mayor proporción de matrículas en las instituciones oficiales, acreditadas en alta calidad. Pero aun así la modalidad virtual de entidades privadas tiene índices altos de matrículas.
- En el nivel de posgrado, tiene mayor número de matrículas en la modalidad presencial.

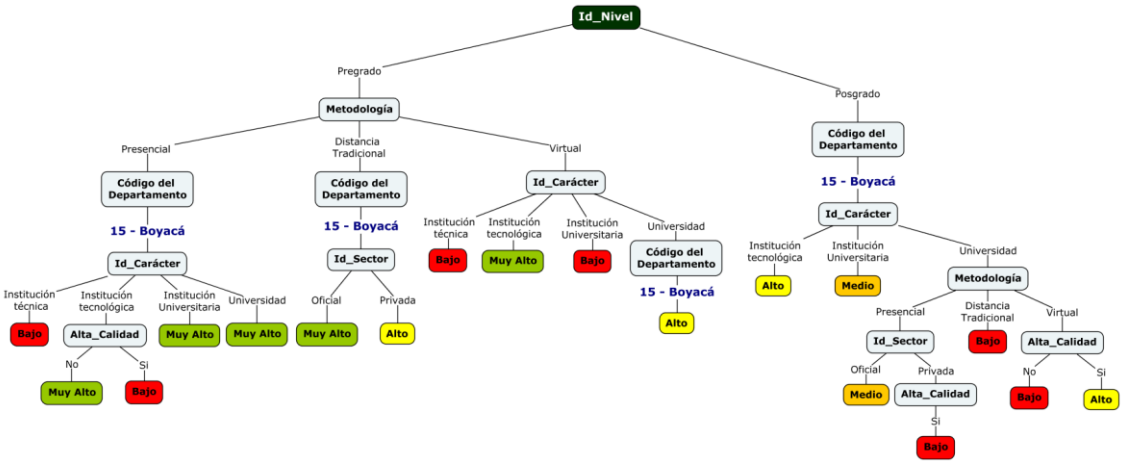



*Figura 51 Diagrama de árbol – Académicas Boyacá*
Fuente: Elaboración propia

El departamento de Caldas, paso de tener una tasa de cobertura de educación superior de 48% al 58%, del 2014 al 2018 respectivamente. En la figura 52, se evidencia el comportamiento de las matriculas de primer semestre en el mismo transcurso de tiempo; algunas de las tendencias que se observan son:

- La tendencia en el número de matrículas en primer semestre en el departamento de Caldas es mayor en entidades acreditadas de alta calidad y entidades privadas virtuales.
- En el nivel de posgrado, el número mayor de matrículas se enfoca en las instituciones acreditadas en alta calidad.

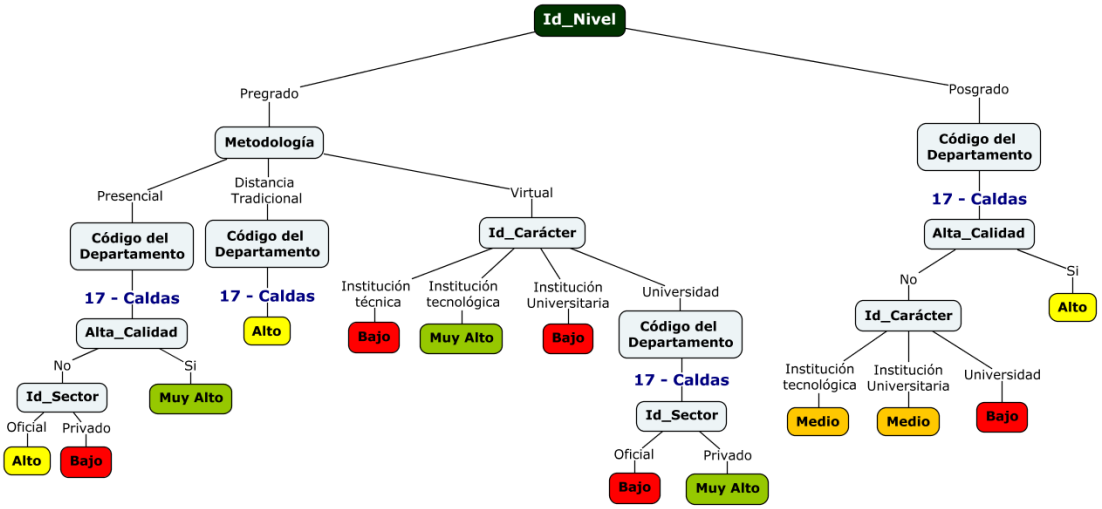



*Figura 52 Diagrama de árbol – Académicas Caldas*
Fuente: Elaboración propia

El departamento de Caquetá, paso de tener una tasa de cobertura de educación superior de 22% al 22%, del 2014 al 2018 respectivamente. En la figura 53, se evidencia el comportamiento de las matriculas de primer semestre en el mismo transcurso de tiempo; algunas de las tendencias que se observan son:

- La tendencia en el número de matrículas en primer semestre en el departamento de Caquetá, tanto en el nivel pregrado son las entidades no acreditadas en alta calidad del sector oficial.
- En el nivel de pregrado, en la metodología virtual se caracteriza por las pocas matriculas en entidades universitarias.
- En el nivel de posgrado, tanto la metodología presencial como la distancia tradicional presenta proporción media de matrículas.

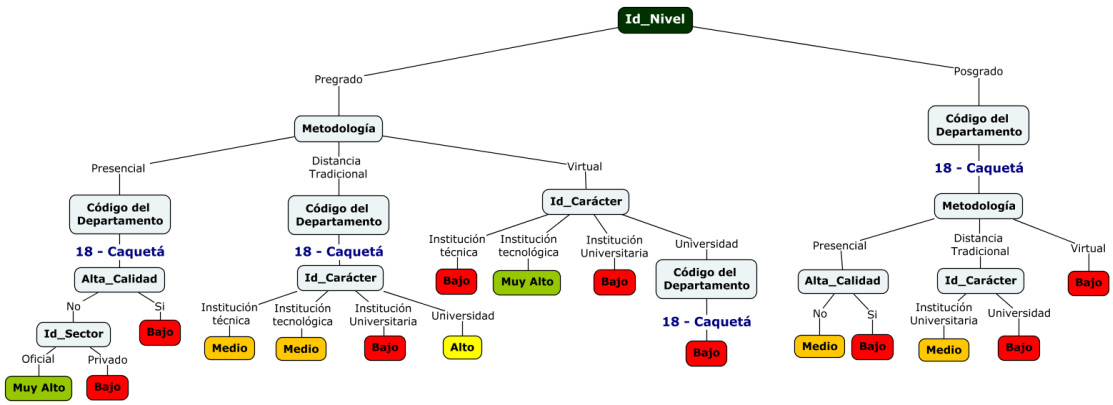



*Figura 53 Diagrama de árbol – Académicas Caquetá*
Fuente: Elaboración propia

El departamento de Cauca, paso de tener una tasa de cobertura de educación superior de 29% al 35%, del 2014 al 2018 respectivamente. En la figura 54, se evidencia el comportamiento de las matriculas de primer semestre en el mismo transcurso de tiempo; algunas de las tendencias que se observan son:

- La tendencia en el número de matrículas en primer semestre en el departamento de Cauca en el nivel pregrado es mayor en entidades oficiales, no acreditadas en alta calidad.

- En el nivel de posgrado, es alta en la metodología de distancia tradicional, en entidades acreditadas en alta calidad del sector privado.

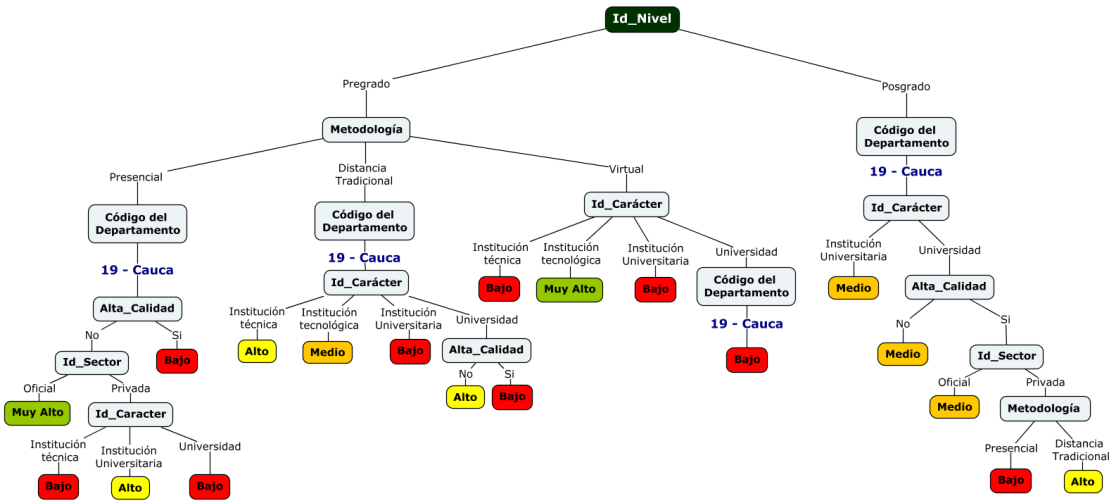



*Figura 54 Diagrama de árbol – Académicas Cauca*
Fuente: Elaboración propia

El departamento de Cesar, paso de tener una tasa de cobertura de educación superior de 30% al 35%, del 2014 al 2018 respectivamente. En la figura 55, se evidencia el comportamiento de las matriculas de primer semestre en el mismo transcurso de tiempo; algunas de las tendencias que se observan son:

- La tendencia en el número de matrículas en primer semestre en el departamento de Cesar en el nivel pregrado es alto en la modalidad presencial, son las entidades del sector oficial no acreditadas en alta calidad.
- En el nivel de pregrado, en la metodología a distancia tradicional y la virtual cuenta con una cantidad media de casos.
- En el nivel de posgrado, tiene número de matrículas baja en general.

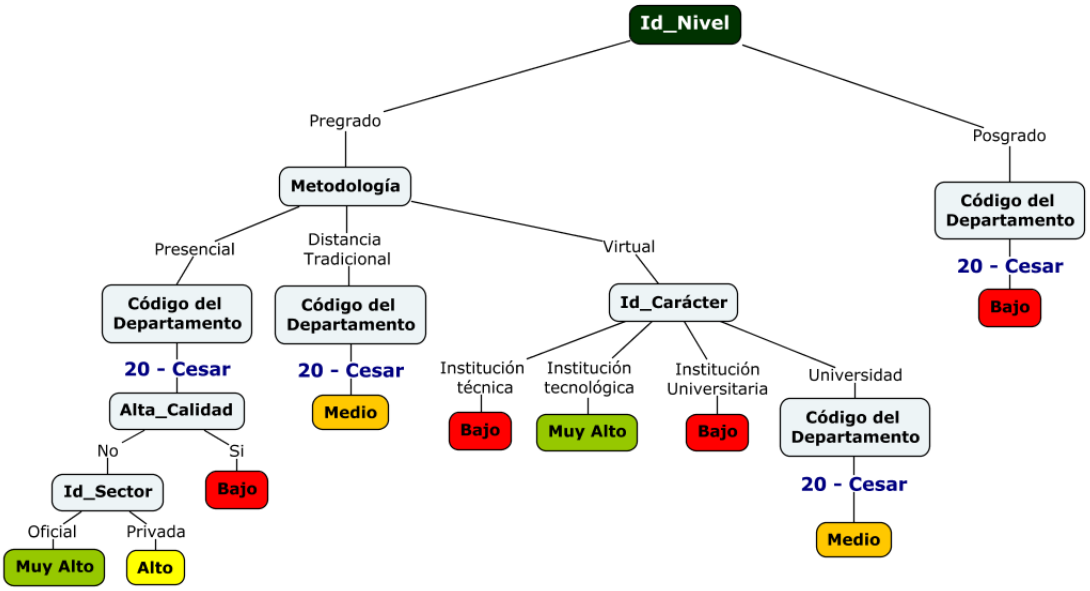



*Figura 55 Diagrama de árbol – Académicas Cesar*
Fuente: Elaboración propia

El departamento de Córdoba, paso de tener una tasa de cobertura de educación superior de 23% al 24%, del 2014 al 2018 respectivamente. En la figura 56, se evidencia el comportamiento de las matriculas de primer semestre en el mismo transcurso de tiempo; algunas de las tendencias que se observan son:

- La tendencia en el número de matrículas en primer semestre en el departamento de Córdoba en el nivel pregrado es alto en la modalidad presencial.
- En el nivel de pregrado, en la metodología a distancia tradicional es mayor en la institución técnica.

- En el nivel de posgrado, tiene número menor número de matrículas siendo mayor en instituciones privadas.

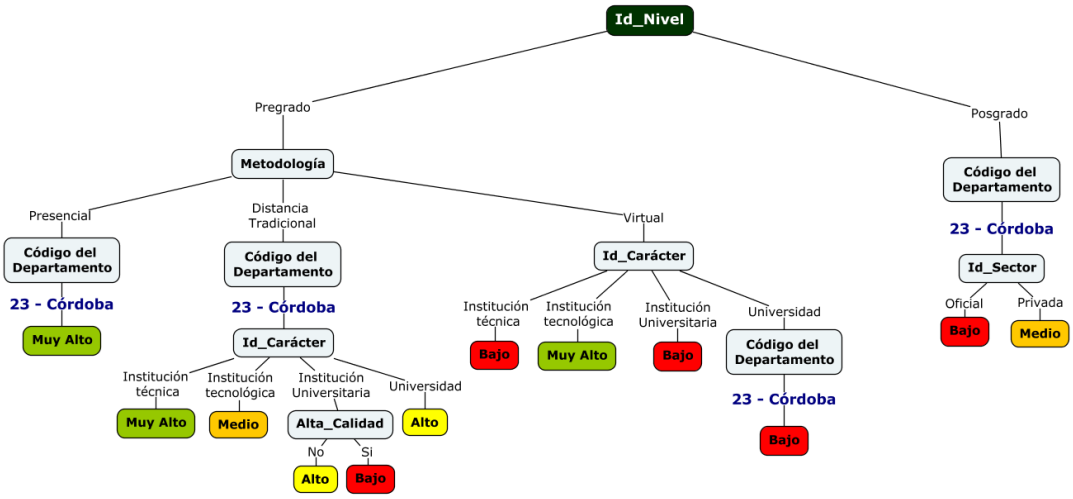



*Figura 56 Diagrama de árbol – Académicas Córdoba*
Fuente: Elaboración propia

El departamento de Cundinamarca, paso de tener una tasa de cobertura de educación superior de 29% al 31%, del 2014 al 2018 respectivamente. En la figura 57, se evidencia el comportamiento de las matriculas de primer semestre en el mismo transcurso de tiempo; algunas de las tendencias que se observan son:

- La tendencia en el número de matrículas en primer semestre en el departamento de Cundinamarca en el nivel pregrado es muy alto en la modalidad presencial, para las entidades universitarias y las instituciones tecnológicas del sector oficial.
- En el nivel de pregrado, en la metodología a distancia tradicional es muy alta en entidades oficiales.
- En el nivel de pregrado, en la metodología virtual es alta en instituciones acreditadas.
- En el nivel de posgrado, es menor al nivel de pregrado, enfocando la mayor proporción a las entidades de carácter universitario privadas.

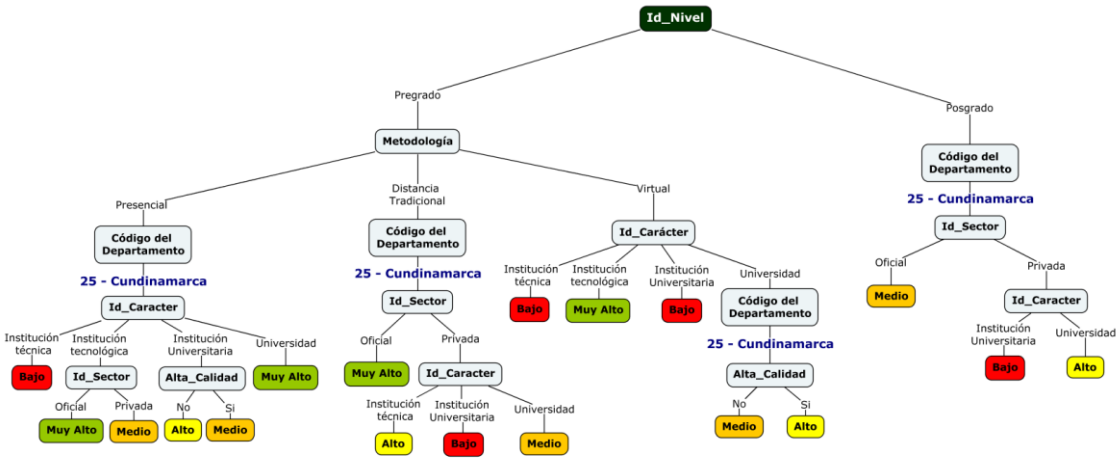



*Figura 57 Diagrama de árbol – Académicas Cundinamarca*
Fuente: Elaboración propia

El departamento de Chocó, paso de tener una tasa de cobertura de educación superior de 24% al 24%, del 2014 al 2018 respectivamente. En la figura 58, se evidencia el comportamiento de las matriculas de primer semestre en el mismo transcurso de tiempo; algunas de las tendencias que se observan son:

- En el nivel pregrado la mayor proporción de matrículas está enfocada en universidades del sector oficial.
- En el nivel pregrado, existen bajas matriculas en las modalidades virtuales y distancia tradicional.
- El nivel posgrado, tiene menores índices de matrículas en entidades privadas.

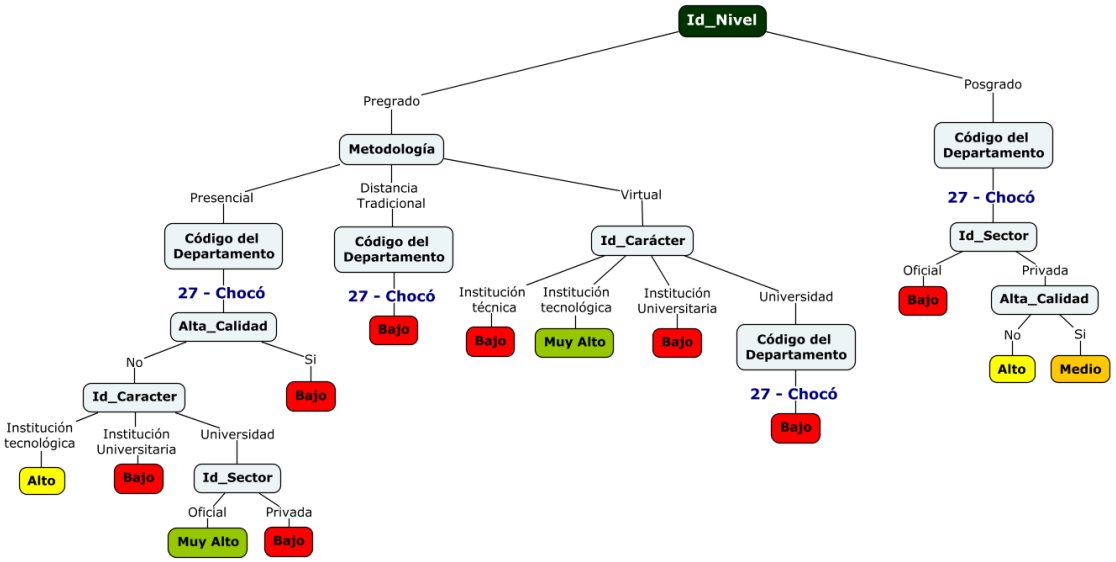



*Figura 58 Diagrama de árbol – Académicas Chocó*
Fuente: Elaboración propia

El departamento de Huila, paso de tener una tasa de cobertura de educación superior de 31% al 35%, del 2014 al 2018 respectivamente. En la figura 59, se evidencia el comportamiento de las matriculas de primer semestre en el mismo transcurso de tiempo; algunas de las tendencias que se observan son:

- La tendencia en el número de matrículas en primer semestre en el departamento de Huila en el nivel pregrado es muy alto en la modalidad presencial, para las entidades universitarias y las instituciones tecnológicas no acreditadas.
- En el nivel de pregrado, en la metodología a distancia tradicional es alta en entidades no acreditadas.
- En el nivel de posgrado, tiene una tendencia media en matriculas, enfocadas en instituciones no acreditadas en alta calidad y del sector oficial.

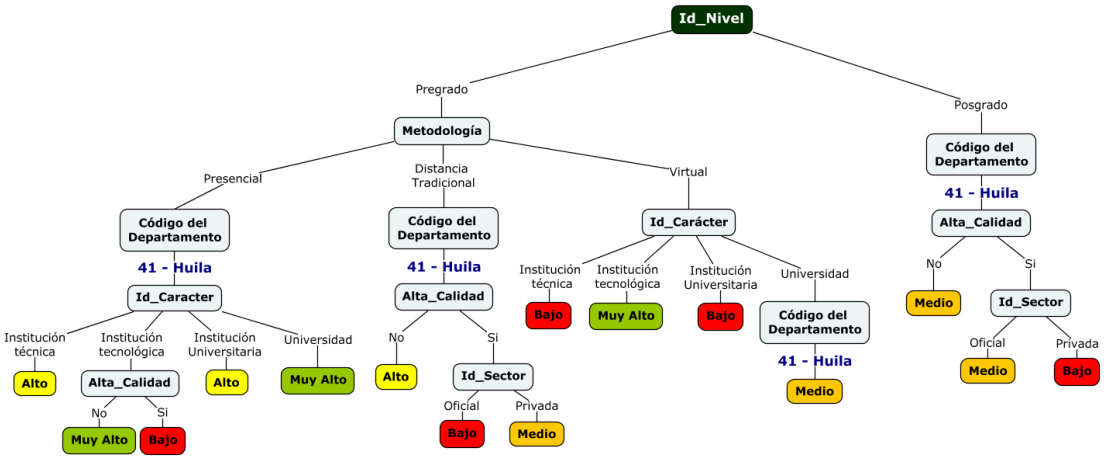



*Figura 59 Diagrama de árbol – Académicas Huila*
Fuente: Elaboración propia

El departamento de La guajira, paso de tener una tasa de cobertura de educación superior de 19% al 21%, del 2014 al 2018 respectivamente. En la figura 60, se evidencia el comportamiento de las matriculas de primer semestre en el mismo transcurso de tiempo; algunas de las tendencias que se observan son:

- La tendencia en el número de matrículas en primer semestre en el departamento de La guajira en el nivel pregrado es muy alto en la modalidad presencial, para las entidades oficiales no acreditas
- En el nivel de pregrado, en la metodología a distancia tradicional y virtual, es baja.
- En el nivel de posgrado, tiene una tendencia baja en general con el posgrado.

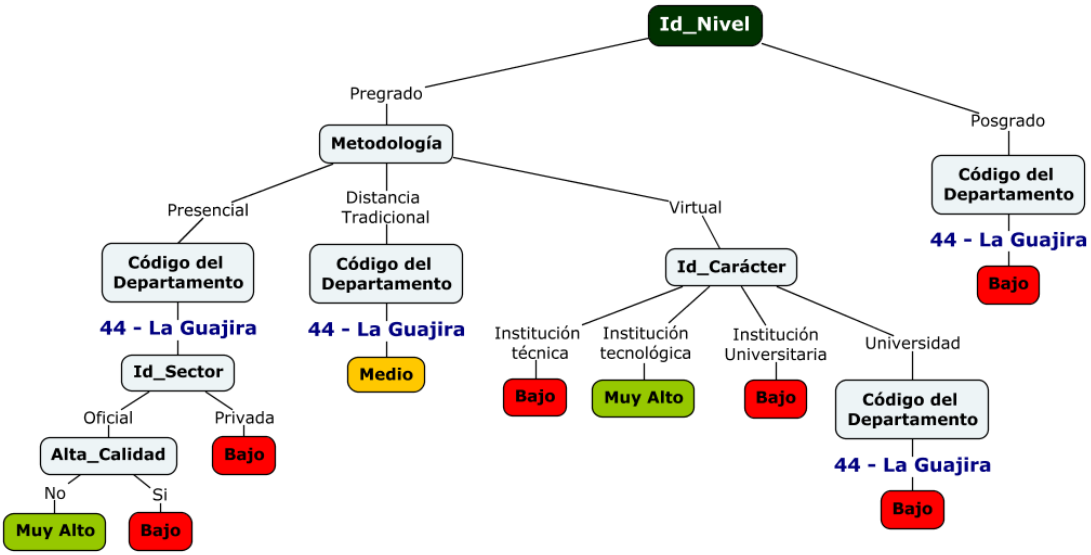



*Figura 60 Diagrama de árbol – Académicas La guajira*
Fuente: Elaboración propia

El departamento de Amazonas, paso de tener una tasa de cobertura de educación superior de 7% al 8%, del 2014 al 2018 respectivamente. En la figura 61, se evidencia el comportamiento de las matriculas de primer semestre en el mismo transcurso de tiempo; algunas de las tendencias que se observan son:

- La tendencia en el número de matrículas en primer semestre en el departamento de Amazonas en general es bajo, para los niveles de pregrado y posgrado. Tanto en el nivel de pregrado como en el de posgrado, se evidencia un leve incremento de matrículas en el sector oficial.

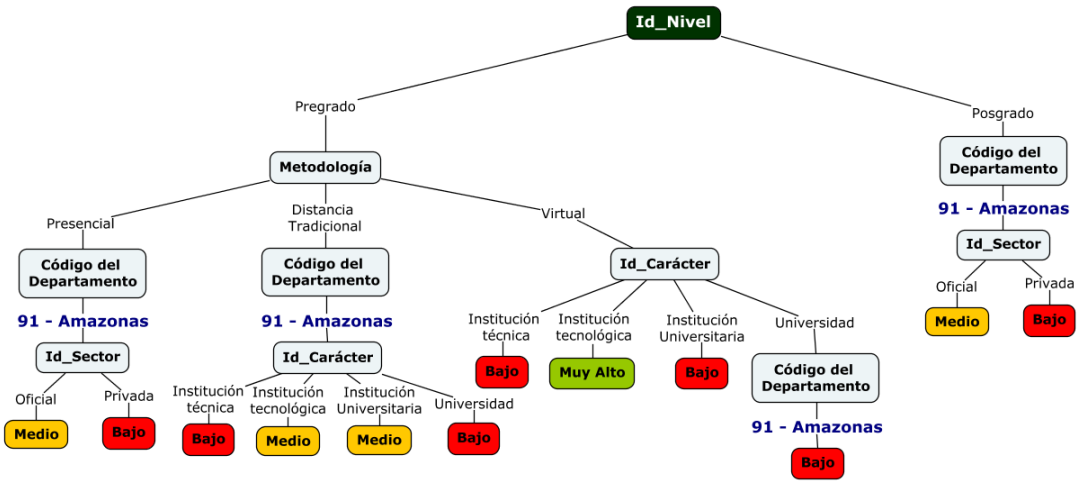



*Figura 61 Diagrama de árbol – Académicas Amazonas*
Fuente: Elaboración propia

El departamento de San Andrés y Providencia, paso de tener una tasa de cobertura de educación superior de 27% al 21%, del 2014 al 2018 respectivamente. En la figura 62, se evidencia el comportamiento de las matriculas de primer semestre en el mismo transcurso de tiempo; algunas de las tendencias que se observan son:

- Uno de los casos más preocupantes a nivel departamental es la tendencia decreciente en la tasa de cobertura en San Andrés y providencia, tanto para el nivel de pregrado como el de posgrado.

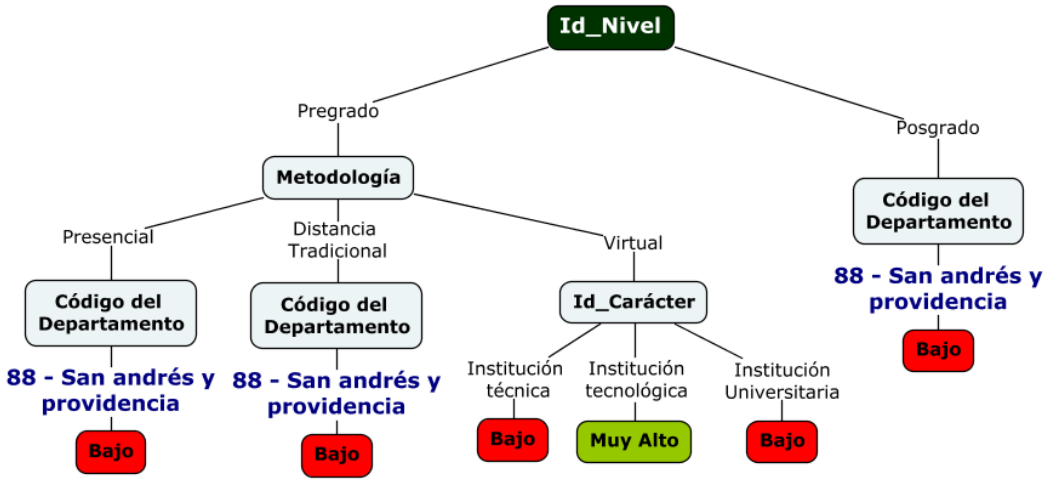



*Figura 62 Diagrama de árbol – Académicas San Andrés y providencia*
Fuente: Elaboración propia

El departamento de Putumayo, paso de tener una tasa de cobertura de educación superior de 13% al 11%, del 2014 al 2018 respectivamente. En la figura 63, se evidencia el comportamiento de las matriculas de primer semestre en el mismo transcurso de tiempo; algunas de las tendencias que se observan son:

- La tendencia en el número de matrículas en primer semestre en el departamento de putumayo, es bajo en la mayoría de metodologías. Al igual que en el nivel de posgrado, el cual tiene un crecimiento leve para entidades oficiales.

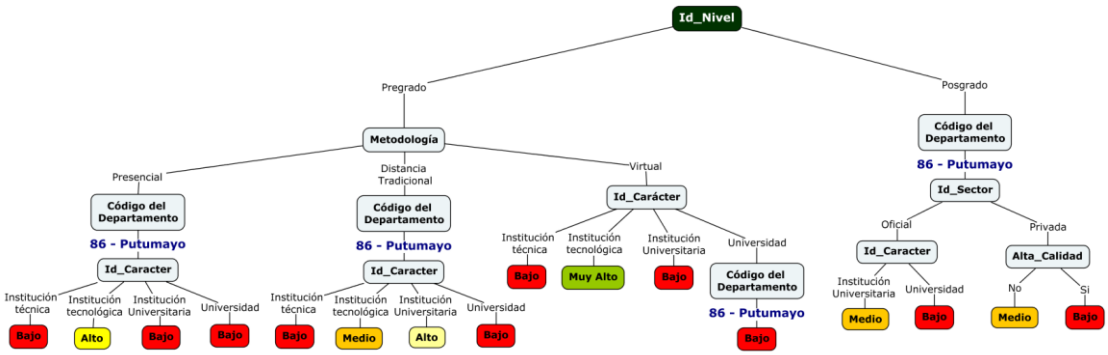



*Figura 63 Diagrama de árbol – Académicas Putumayo*
Fuente: Elaboración propia

El departamento de Casanare, paso de tener una tasa de cobertura de educación superior de 26% al 25%, del 2014 al 2018 respectivamente. En la figura 64, se evidencia el comportamiento de las matriculas de primer semestre en el mismo transcurso de tiempo; algunas de las tendencias que se observan son:

- En el nivel de pregrado, en el departamento de Casanare tiene una mayor aceptación en las entidades no acreditadas en alta calidad, en la modalidad presencial.
- En cuanto a la modalidad virtual, en el nivel de pregrado existe una tendencia de matrículas medias en instituciones universitarias. En cuanto a la distancia tradicional es un poco mayor en las entidades acreditadas en alta calidad.
- En el nivel de posgrado, se evidencia un bajo nivel de matrículas en general.

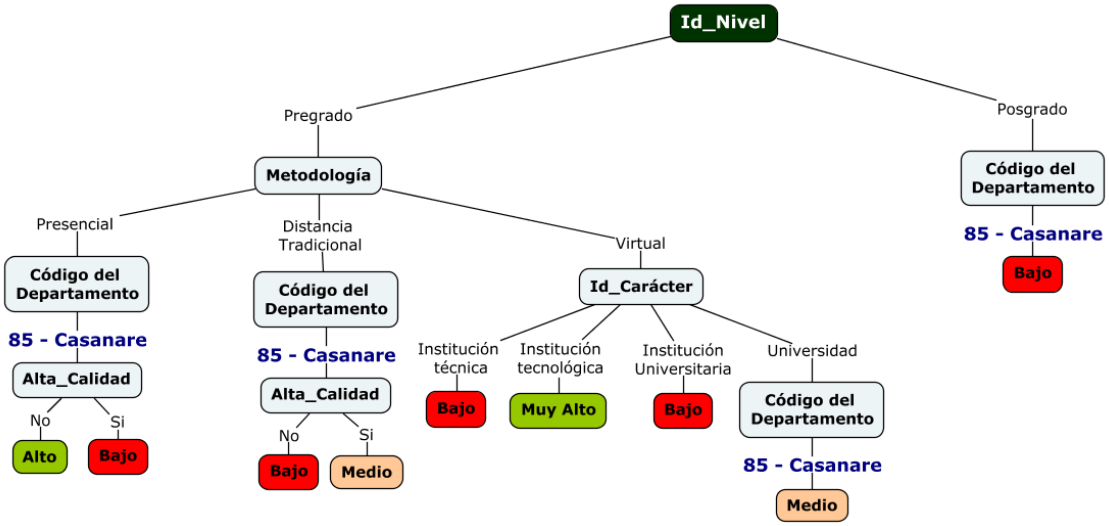



*Figura 64 Diagrama de árbol – Académicas Casanare*
Fuente: Elaboración propia

El departamento de Valle del Cauca, paso de tener una tasa de cobertura de educación superior de 39% al 43%, del 2014 al 2018 respectivamente. En la figura 65, se evidencia el comportamiento de las matriculas de primer semestre en el mismo transcurso de tiempo; algunas de las tendencias que se observan son:

- En el valle del cauca existe una aceptación muy alta a nivel de pregrado, en las modalidades presenciales y distancia tradicional. Siendo mayor en las instituciones no acreditadas en alta calidad.
- En el nivel de posgrado, se evidencia una tendencia muy alta para las instituciones universitarias privadas, de metodología presencial.

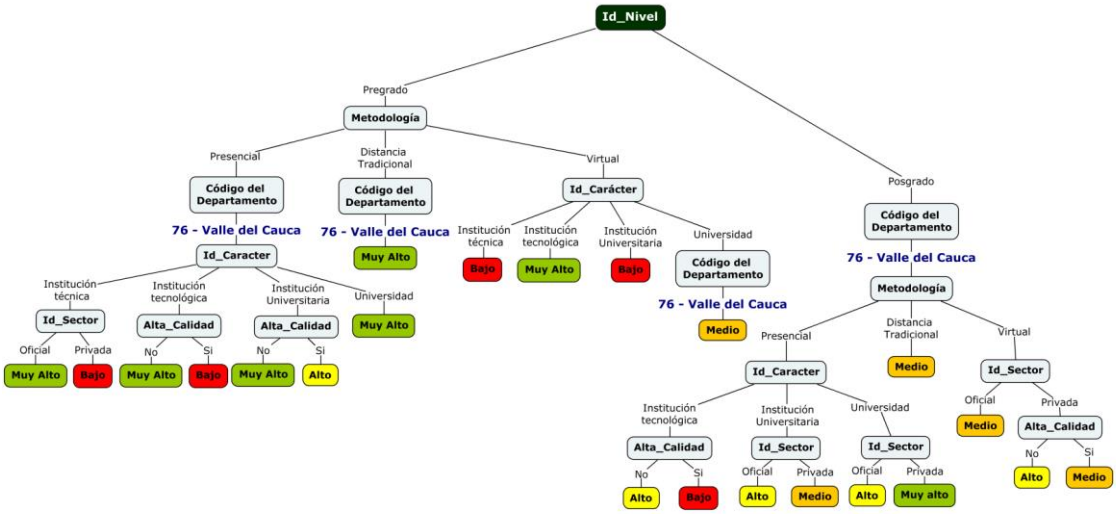



*Figura 65 Diagrama de árbol – Académicas Valle del Cauca*
Fuente: Elaboración propia

El departamento de Tolima, paso de tener una tasa de cobertura de educación superior de 37% al 39%, del 2014 al 2018 respectivamente. En la figura 66, se evidencia el comportamiento de las matriculas de primer semestre en el mismo transcurso de tiempo; algunas de las tendencias que se observan son:

- En el departamento de Tolima se evidencia mayor aceptación en la modalidad de las diferentes metodologías en el nivel pregrado. Siendo menor las matriculas en las instituciones caracterizadas como universidad.
- En el nivel de posgrado se evidencia un menor número de matrículas en el nivel de posgrado.

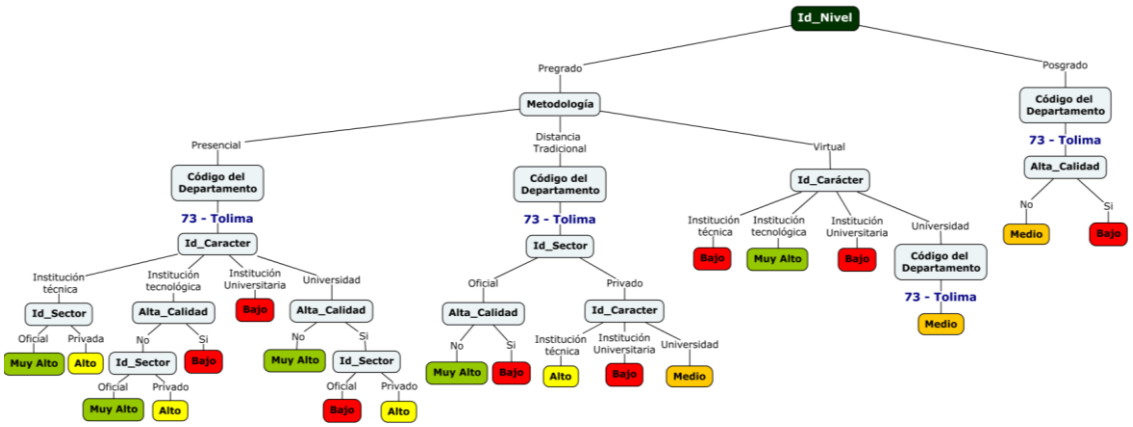



*Figura 66 Diagrama de árbol – Académicas Tolima*
Fuente: Elaboración propia

El departamento de Sucre, paso de tener una tasa de cobertura de educación superior de 24% al 28%, del 2014 al 2018 respectivamente. En la figura 67, se evidencia el comportamiento de las matriculas de primer semestre en el mismo transcurso de tiempo; algunas de las tendencias que se observan son:

- En el departamento de sucre en el nivel de pregrado, se evidencia una tendencia muy alta en las modalidades presenciales de instituciones no acreditadas en alta calidad y en la modalidad de distancia tradicional en instituciones de caracterizadas de instituciones técnicas.

- En el nivel de posgrado se evidencia un menor número de matrículas, siendo mayor en la modalidad privada.

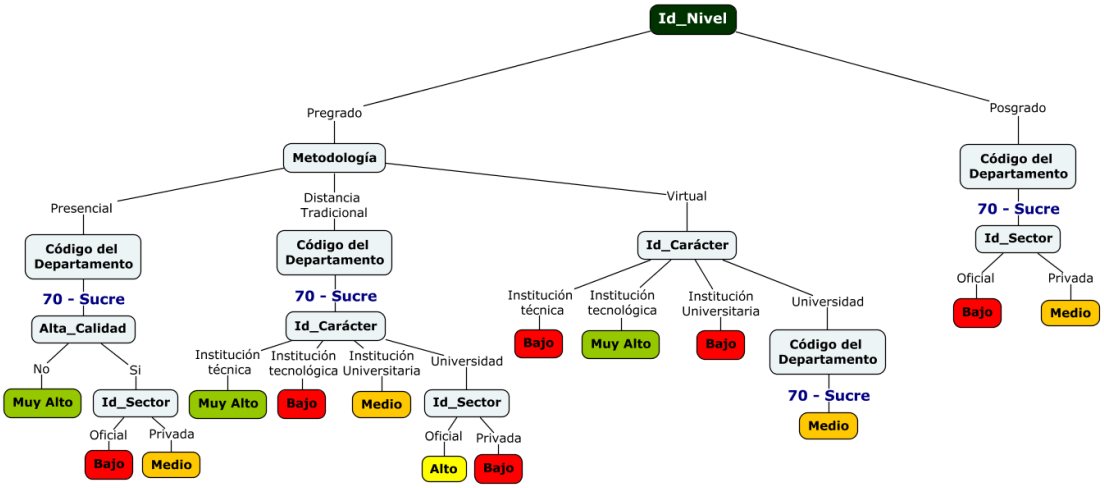



*Figura 67 Diagrama de árbol – Académicas Sucre*
Fuente: Elaboración propia

El departamento de Santander, paso de tener una tasa de cobertura de educación superior de 61% al 65%, del 2014 al 2018 respectivamente. En la figura 68, se evidencia el comportamiento de las matriculas de primer semestre en el mismo transcurso de tiempo; algunas de las tendencias que se observan son:

- En el nivel de pregrado existe mayor aceptación en la modalidad presencial, aunque en la distancia tradicional tiene un alto número de matrículas.

- En el departamento de posgrados existe un número medio de matrículas.

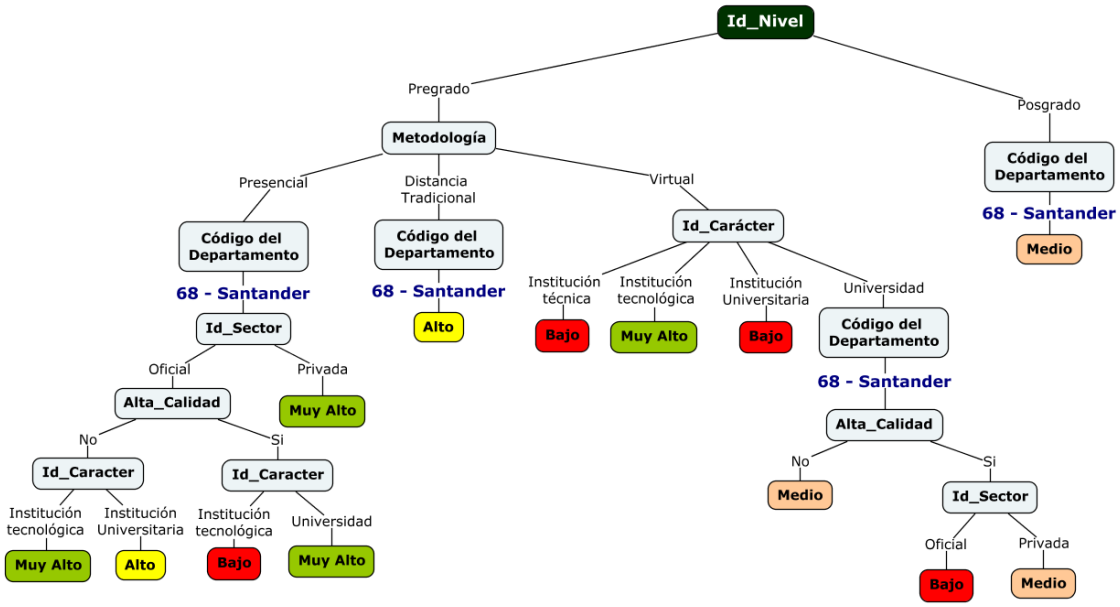



*Figura 68 Diagrama de árbol – Académicas Santander*
Fuente: Elaboración propia

El departamento de Quindío, paso de tener una tasa de cobertura de educación superior de 55% al 63%, del 2014 al 2018 respectivamente. En la figura 69, se evidencia el comportamiento de las matriculas de primer semestre en el mismo transcurso de tiempo; algunas de las tendencias que se observan son:

- En el nivel de pregrado existe una aceptación muy alta en las tres metodologías, lo que podría explicar el crecimiento exitoso en el periodo de tiempo.
- En el departamento de posgrados existe un número medio de matrículas.

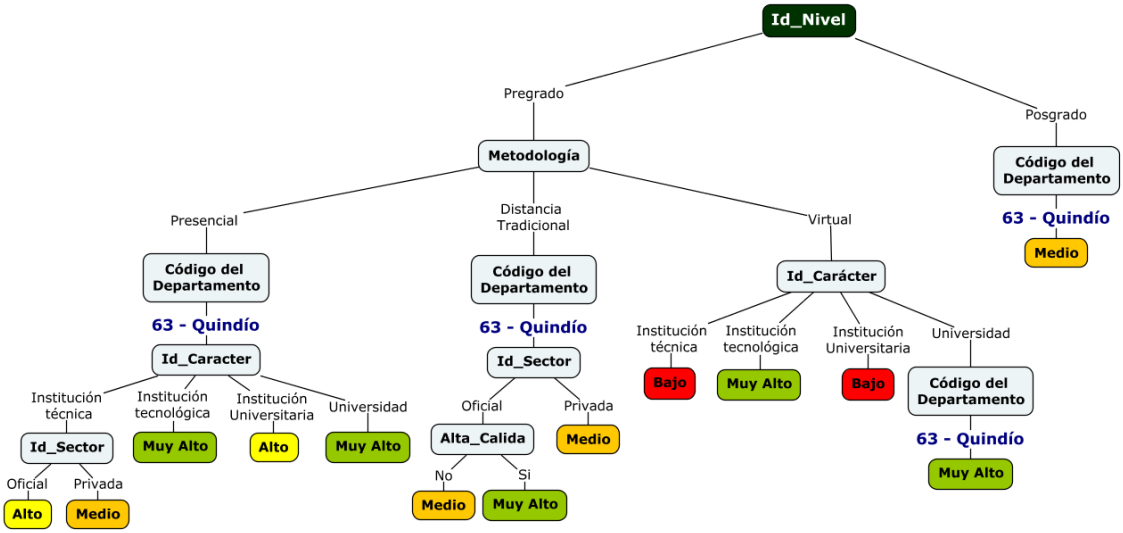



*Figura 69 Diagrama de árbol – Académicas Quindío*
Fuente: Elaboración propia

El departamento de Nariño, paso de tener una tasa de cobertura de educación superior de 24% al 24%, del 2014 al 2018 respectivamente. En la figura 70, se evidencia el comportamiento de las matriculas de primer semestre en el mismo transcurso de tiempo; algunas de las tendencias que se observan son:

- En el nivel de pregrado existe una aceptación muy alta en instituciones presenciales de carácter universitario.
- En la metodología de distancia tradicional a nivel de pregrado, existe mayor número de matrículas en programas no acreditados en alta calidad.
- En el departamento de posgrados existe un número medio de matrículas.

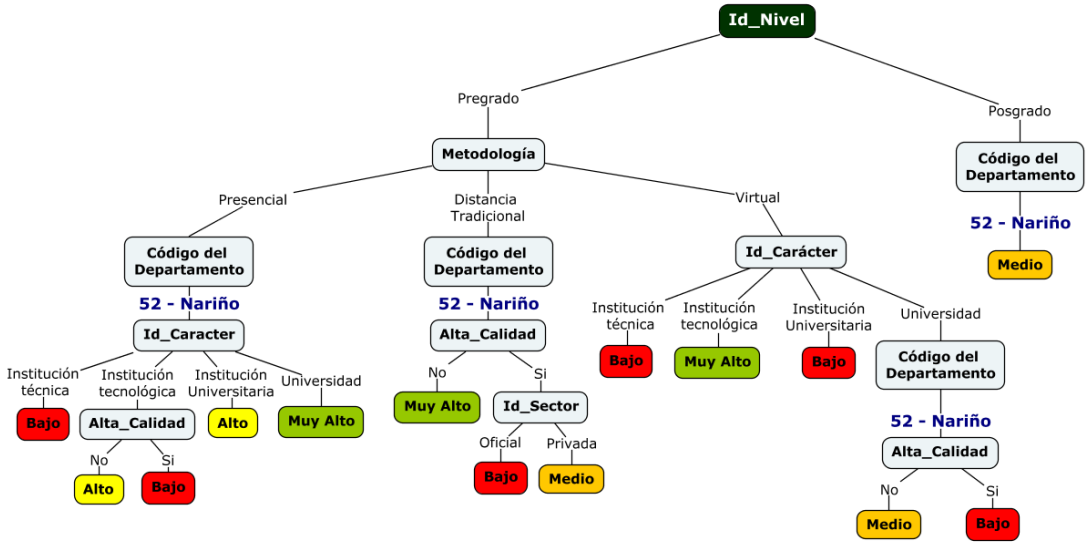



*Figura 70 Diagrama de árbol – Académicas Nariño*
Fuente: Elaboración propia

El departamento de Risaralda, paso de tener una tasa de cobertura de educación superior de 56% al 61%, del 2014 al 2018 respectivamente. En la figura 71, se evidencia el comportamiento de las matriculas de primer semestre en el mismo transcurso de tiempo; algunas de las tendencias que se observan son:

- En el nivel de pregrado existe mayor número de matrículas en las metodologías presenciales, en instituciones de caracterizadas como universidad.
- En el nivel de posgrado, existe menor número de matrículas especialmente en las modalidades a distancia tradicional y virtual.

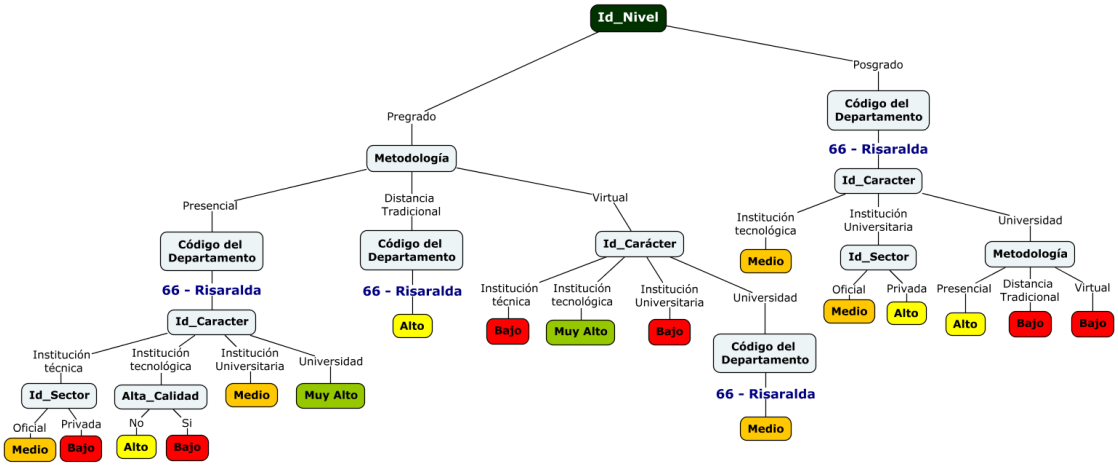



*Figura 71 Diagrama de árbol – Académicas Risaralda*
Fuente: Elaboración propia

El departamento de Meta, paso de tener una tasa de cobertura de educación superior de 34% al 34%, del 2014 al 2018 respectivamente. En la figura 72, se evidencia el comportamiento de las matriculas de primer semestre en el mismo transcurso de tiempo; algunas de las tendencias que se observan son:

- En el departamento del meta, en el nivel de pregrado se evidencia una evidencia tendencia muy alta en modalidades virtuales y distancia tradicional. Con una aceptación media instituciones universitarias con metodologías virtuales.
- En el departamento de posgrado existe una tendencia media en matriculas del sector oficial y baja en privadas.

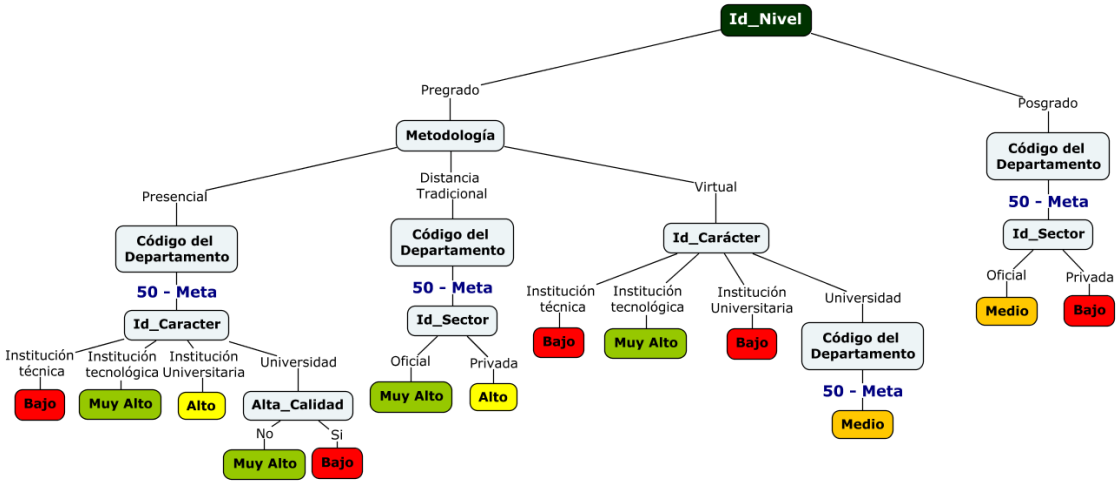



*Figura 72 Diagrama de árbol – Académicas Meta*
Fuente: Elaboración propia

El departamento de Norte de Santander, paso de tener una tasa de cobertura de educación superior de 47% al 51%, del 2014 al 2018 respectivamente. En la figura 73, se evidencia el comportamiento de las matriculas de primer semestre en el mismo transcurso de tiempo; algunas de las tendencias que se observan son:

- En el nivel de pregrado el departamento de Norte de Santander, evidencia altas tasas de matrículas en las modalidades presenciales y distancia tradicional. En el nivel de posgrado existe tendencia media de matrículas.

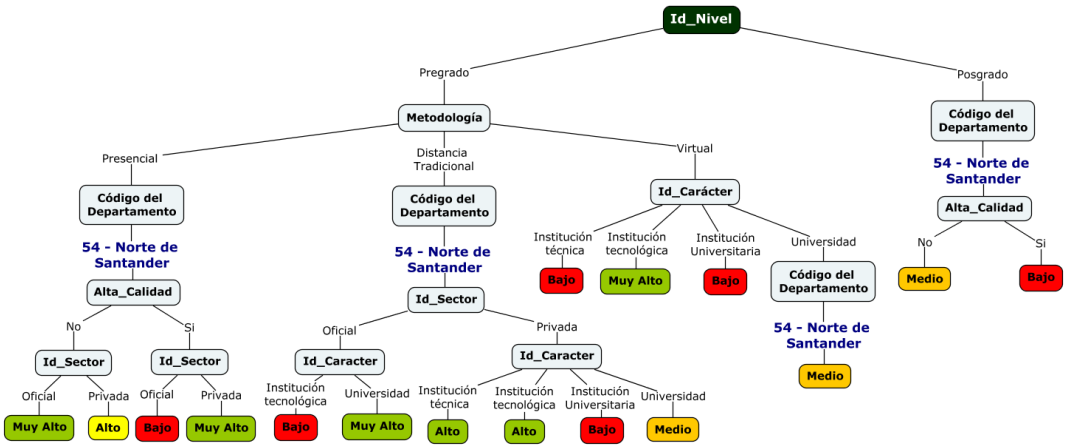



*Figura 73 Diagrama de árbol – Académicas Norte de Santander*
Fuente: Elaboración propia

El departamento de Magdalena, paso de tener una tasa de cobertura de educación superior de 30% al 26%, del 2014 al 2018 respectivamente. En la figura 74, se evidencia el comportamiento de las matriculas de primer semestre en el mismo transcurso de tiempo; algunas de las tendencias que se observan son:

- En el nivel de pregrado se evidencia tendencia muy alta en modalidad presencial y distancia tradicional, a diferencia del posgrado el cual es bajo en forma general.

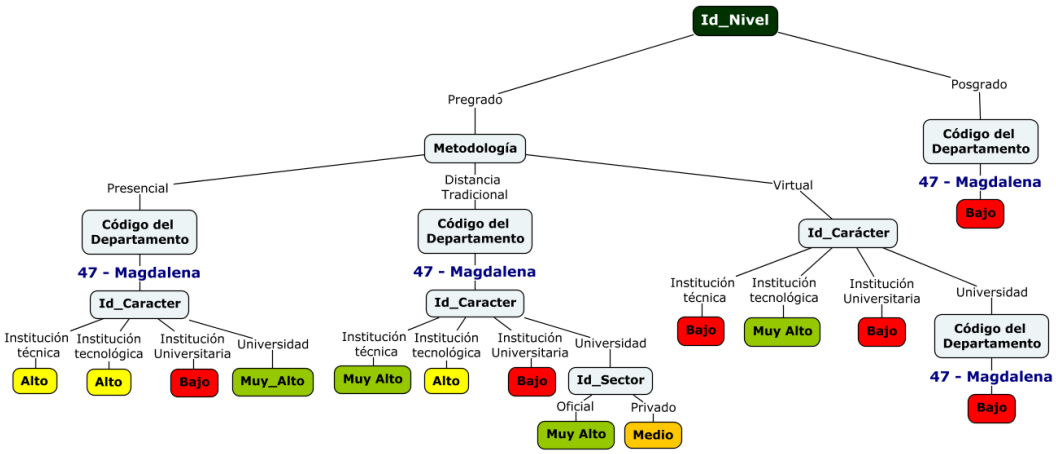



*Figura 74 Diagrama de árbol – Académicas Magdalena*
Fuente: Elaboración propia

## 4.5.2. Evaluación de datos - Variables individuales y socioeconómicas

En cuanto a las variables individuales los análisis de las salidas permiten definir las caracterizar de los estudiantes matriculados en primer semestre en el periodo de 2014 a 2018; la frecuencia más alta de matrículas indica un perfil de personas entre 16 a 25 años, que no suele recibir apoyo de las IES y el ICETEX; los perfiles más bajos de estudiantes son aquellos cuya posición de hermanos es posterior al quinto, viven solos o no tienen hermanos, ver figura 75.


*Figura 75 Diagrama de calor – Individuales*
Fuente: Elaboración propia

Por último, en cuanto a las variables socioeconómicas, el mayor número de matrículas de primer semestre es de personas en estrato 1 y 2, de sisen nivel 1, cuyo nivel educativo de la madre es secundaria, con ingresos económicos en el hogar menor a 2 salarios mínimos; los perfiles más bajos de estudiantes son personas de estratos altos, con ingresos superiores a los 7 salarios mínimos o con créditos académicos que suele ser a largo plazo o a mediano plazo, ver figura 76.

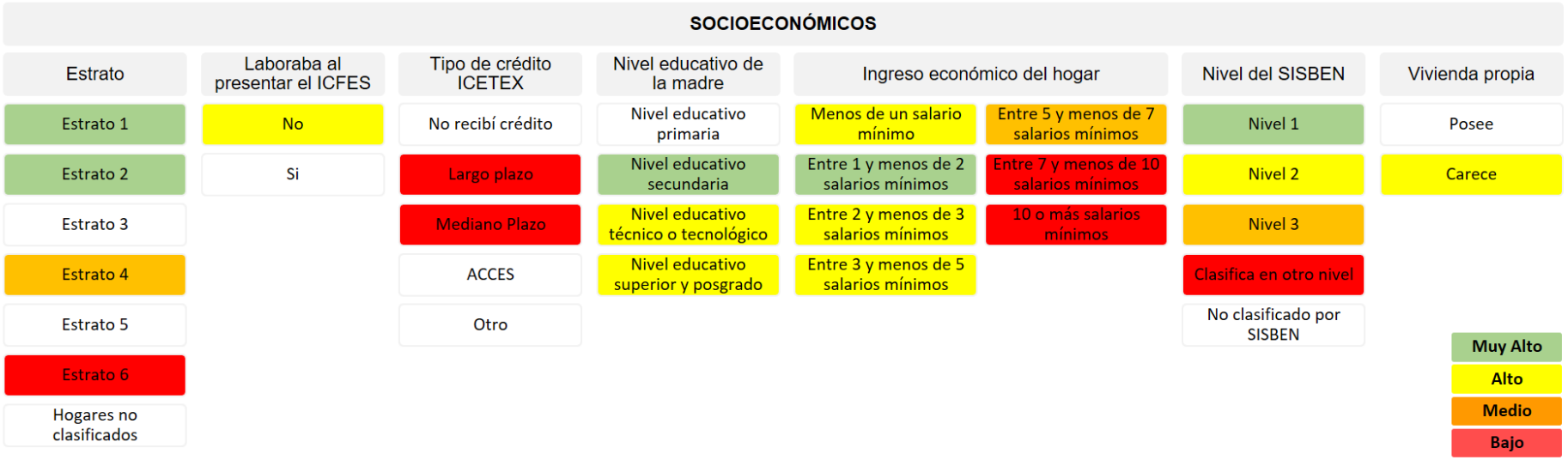



*Figura 76 Diagrama de calor – Socioeconómicos*
Fuente: Elaboración propia

## 4.6. Fase 6: Propuesta de implementación

Basados en la evaluación de los datos anteriores se propone las siguientes recomendaciones:

I. Es importante promover programas académicos diferenciados por departamento, considerando los ingresos económicos de las familias en la región; debido a que en la mayoría de matrículas de instituciones o programas acreditados en alta calidad se reducen el número de matrículas, posiblemente debido al incremento de costos resultado de la contratación de docentes con mayor nivel educativo.

II. A nivel departamental se debe promover la cobertura de instituciones que ofrezcan programas con metodologías variadas; considerando el crecimiento de departamentos como Caldas, Quindío y Boyacá en el periodo trascurrido de 2014 a 2018, ver figura 77, y evidenciando similitudes en sus resultados en cuanto a la variedad de las metodologías educativas que acogieron las regiones, sin importar si son entidades de carácter privado u oficial.

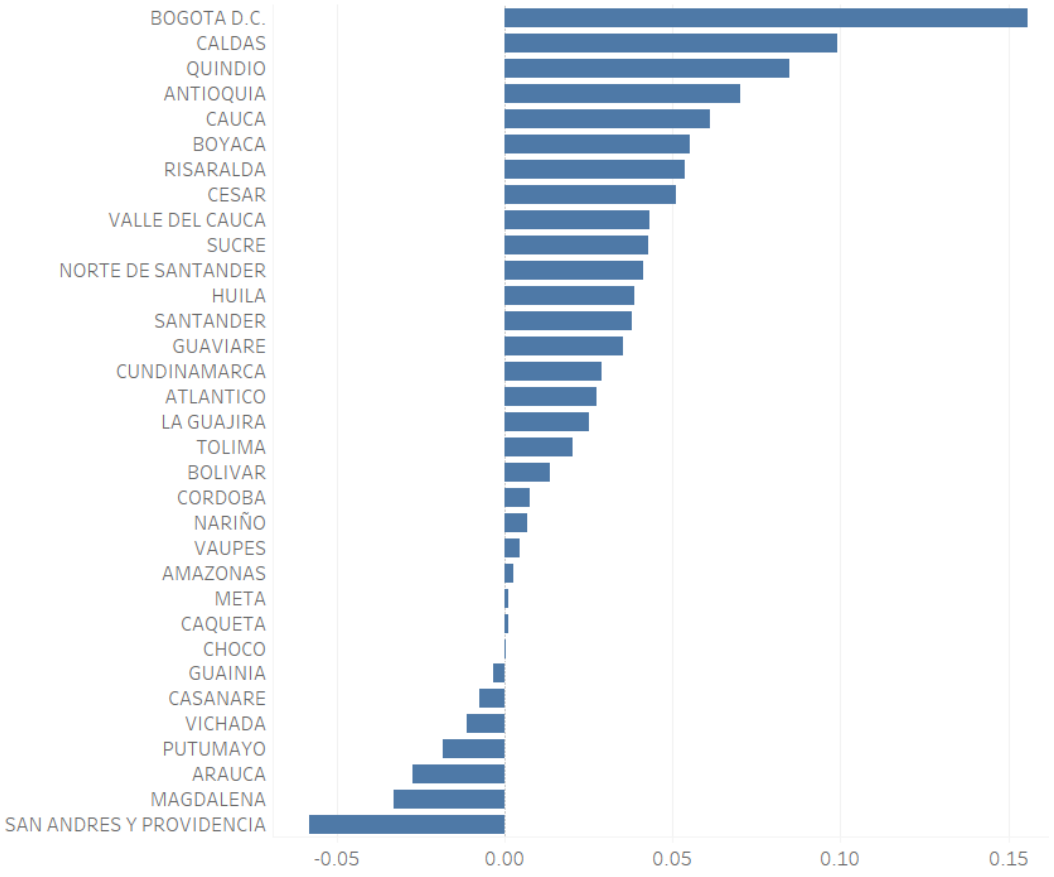



*Figura 77 Variación de la tasa de cobertura de educación superior de 2014 a 2018*
Fuente: Adaptado Ministerio de educación

III. Promover políticas de acceso a internet a zonas aledañas del país; en la figura 77 se observa una pérdida de cobertura de educación superior en lugares como San andrés y providencia, Magdalena, Arauca y Putumayo. Adicionalmente se observa un crecimiento nulo en regiones como el Meta, Caquetá y Choco; en estas regiones se evidencia una baja cobertura en la totalidad de clasificación, en parte debido a el difícil acceso a muchas de estas regiones.

   En ese sentido promover la educación virtual en estas regiones permitiría mitigar la migración de personas a las ciudades, así como un crecimiento en las comunidades donde se forma la población, pero para poder lograrlo es indispensable el acceso a internet a las diferentes zonas.

IV. Promover el ingreso a educación superior de nivel posgrado; aunque la contratación de docentes con niveles de formación máximo superiores al pregrado incremento en los últimos años, la formación en niveles de posgrado parece estancarse. En algunas regiones del país se evidencia un crecimiento en la formación virtual de programas posgrado, pero estos deben ser incentivados desde políticas que incentiven el ingreso hasta la generación de empleo para estos cargos.

# 5. Conclusiones y trabajo futuro

Esta sección describe las conclusiones y recomendaciones de trabajos futuros.

## 5.1. Conclusiones

La educación superior en Colombia, aunque está creciendo en los últimos años, presenta una serie de problemas que detiene el crecimiento en el país, algunas de ellas son las cifras impulsadas por ciudades principales, esconden la realidad de zonas como Choco, Guainía o vichada en las que se hace necesario llegar con sistemas de educación en la zona, que permita el crecimiento en la comunidad.

Los datos analizados y a los que se le aplicaron las técnicas de minería de datos corresponde a bases de datos abiertas, del Ministerio de educación de Colombia, los indicadores –SNIES- y –SPADIES. Datos que fueron procesados y a los cuales se le realizaron aplicaciones de algoritmos de árbol y reglas de clasificación principalmente. Estos factores fueron clasificados como individuales, socioeconómicas, académicas e institucionales, de los cuales se puede resumir que:

- En cuanto a las factoras individuales es de resaltar que en Colombia las personas mayores a 25 años no tienden a acceder a programas de educación superior en ninguno de sus niveles.
- Por otro lado, en cuanto a los factores socioeconómicos, la mayor población de estudiantes de las IES en Colombia son de estrato 1 y 2; es de resalta que si los ingresos familiares son mayores a 7 salarios mínimos no acceden a educación superior en el país, y las personas que laborar a la hora de presentar los exámenes de acceso a educación superior tiene menores tasas de ingreso a las instituciones.
- En cuanto a las variables institucionales, se encontró que las instituciones acreditadas en alta calidad contrata mayor persona administrativo y docentes con mayor nivel de formación académica. Otro aspecto particular que se observo es la baja contratación de personal administrativo en las entidades del sector oficial y las tasas altas de contratos en tiempo parcial de los docentes de estas instituciones.
- En cuanto a las variables académicas las acreditaciones en alta calidad no parecen ser un tema diferenciador a nivel de instituciones tecnológicas; siendo levemente significativas a nivel de instituciones de carácter universitarias o posgrado. A nivel departamental se observó que el fomento de metodologías de enseñanza variadas incrementa positivamente la tasa de cobertura, pero que se debe garantizar el acceso a aspectos como el internet.

En cuanto a las propuestas se recomienda, realizar políticas de educación superior diferenciadas por región, debido a que en los factores académicos se puede identificar diferentes necesidades en las zonas, promover las metodologías variadas en todas las regiones garantizando suplir necesidades de movilidad o enseñanza, crear políticas que garanticen el acceso a zona internet en el país y promover el ingreso y contratación de personal con formación posgrado.

## 5.2. Líneas de trabajo futuro

Colombia cuenta con bases de datos en diferentes campos como lo son salud, educación, vivienda entre otros y mediante la minería de datos, se podría realizar estudios para identificar u abordar diferentes problemáticas. En cuanto a la problemática tratada en el presente documento, existe histórico de varios años y variables no utilizadas que pueden cambiar completamente el enfoque de futuras investigaciones realizadas con las mismas bases de datos.

## 5.3. Limitaciones del estudio

La limitación del estudio fue lograr encontrar una base de datos publica en las que se pudiese identificar la conexión a internet por departamento, indispensable para lograr determinar las oportunidades y necesidades específicas en políticas de educación.

# 6. Bibliografía

Aquino, Aldair. Molero, Gillermo. Rojano, R. (2015). Hacia un nuevo proceso de minería de datos centrado en el usuario. ISSN 1405-1249, 272–291.

Beguerí, G., & Malberti, A. (2017). Minería de datos y una aplicación en la educación superior. *XIX Workshop de Investigadores En Ciencias de La Computación*, 1276–1279. http://sedici.unlp.edu.ar/handle/10915/62964

Bernabeu, R. D. (2007). *DATA WAREHOUSING: Investigación y Sistematización de Conceptos*. 116.

Bernal, R., Camacho, A., Flórez, C. E., Gaviria, A., Jaramillo, C., Nupia, O., Peña, X., Rodríguez, C., & Sánchez, F. (2009). *Documentos CEDE*.

Cadavid, D. V., & Mendoza, A. M. (2017). Predicting the Efficiency of Colombian Higher Education Institutions with Data Envelopment Analysis and Data Mining. *Pensamiento y Gestión*, *42*(January), 140–162. https://doi.org/10.14482/pege.42.10467

Celli, J. F. B., Ledezma, N. A., Torrecilla, F. J. M., Díaz, H. D., Ludeña, A. F., & Yaselli, M. B. (2008). *Una mejor educación para una mejor sociedad*. 190.

Dee, J. R., & Heineman, W. A. (2016). Understanding the Organizational Context of Academic Program Development. *New Directions for Institutional Research*, *2015*(168), 9–35. https://doi.org/10.1002/ir.20158

Estrada-Danell, R. I., Zamarripa-Franco, R. A., Zúñiga-Garay, P. G., & Martínez-Trejo, I. (2016). Aportaciones desde la minería de datos al proceso de captación de matrícula en Instituciones de Educación Superior particulares. *Revista Electrónica Educare*, *20*(3), 1. https://doi.org/10.15359/ree.20-3.11

Estrada-Villa, E. J., & Boude-Figueredo, O. R. (2018). Multivariate analysis of elements related to mobile learning in higher education in Colombia. *Revista Electronica Educare*, *22*(3). https://doi.org/10.15359/ree.22-3.6

Gómez, O. S. (2010). Tendencias mundiales que afectan la educación superior. *Universidad & Empresa*, *7*(9), 42–65.

Guzmán Ruiz, C., Muriel Durán, D., Nacional, Franco Gallego, J., Castaño Velez, E., Gómez Portilla, K., & Vásquez Velásquez, J. (2009). *Educación Superior Colombiana*.

Harold Elbert Escobar Terán, Maritza Alcívar Saltos, Carlos Marquez de la Plata, C. E. E. T. (2017). *Implementación de minería de datos en la gestión académica de las instituciones de educación superior minería*. *VIII*, 203–212.

Hernández Sampieri Roberto, Fernández Collado Carlos, Baptista Lucio María del Pilar. (2014). Metodología de la investigación. McGRAW-HILL / INTERAMERICANA EDITORES, S.A. DE C.V. p. 634. ISBN: 978-1-4562-2396-0.

IBM. (2020). Cross-Industry Standard Process for Data Mining. *Knowledge Center*. https://www.ibm.com/support/knowledgecenter/es/SS3RA7_sub/modeler_crispdm_ddita/clementine/crisp_help/crisp_overview.html

Jiawei Han, Micheline Kamber, J. P. (2014). Data mining: Concepts and techniques. In *Proceedings - 2013 International Conference on Machine Intelligence Research and Advancement, ICMIRA 2013*. https://doi.org/10.1109/ICMIRA.2013.45

Melo-Becerra, L. A., Ramos-Forero, J. E., & Hernández-Santamaría, P. O. (2017). La educación superior en Colombia: situación actual y análisis de eficiencia. *Desarrollo y Sociedad*, *2017*(78), 59–111. https://doi.org/10.13043/DYS.78.2

Mineducación. (2020). *Educación Superior*. https://www.mineducacion.gov.co/portal/Educacion-superior/

Banco Mundial. (2017). *Graduarse: solo la mitad lo logra en América Latina*. https://www.bancomundial.org/es/news/feature/2017/05/17/graduating-only-half-of-latin-american-students-manage-to-do-so

Banco Mundial. (2020). *Inscripción escolar, nivel terciario (% bruto)*. Instituto de Estadística de La Organización de Las Naciones Unidas Para La Educación, La Ciencia y La Cultura ( UNESCO ). https://datos.bancomundial.org/indicador/SE.TER.ENRR

Parra, A. D. P. (2019). Caracterización de Aspirantes para el Planteamiento de Estrategias de Captación de las Instituciones de Educación Superior. *Society*, *1*(1), 52. https://doi.org/10.1017/CBO9781107415324.004

RICARDO TIMARÁN PEREIRA, ANDRÉS CALDERÓN ROMERO, J. J. T. (2013). Aplicación de la minería de datos en la extracción de perfiles de deserción estudiantil. *Ventana Informática*, *28*, 31–47. http://revistasum.umanizales.edu.co/ojs/index.php/ventanainformatica/article/view/181%0Ahttp://revistasum.umanizales.edu.co/ojs/index.php/ventanainformatica/article/viewFile/181/228%0Ahttp://revistasum.umanizales.edu.co/ojs/index.php/ventanainformatica/ar

Rodríguez Valero, L. C., & Gutiérrez Rodríguez, R. E. (2019). Estudio prospectivo de escenarios de la tecnología en el trabajo en Colombia al 2050. *Económicas Cuc*, *40*(2), 101–116. https://doi.org/10.17981/econcuc.40.2.2019.07

SNIES. (2020). *Resumen de indicadores de Educación Superior*. https://www.mineducacion.gov.co/sistemasdeinformacion/1735/w3-article-212350.html?_noredirect=1

Valero Orea, S., Salvador Vargas, A., & García Alonso, M. (2010). Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos. *Recursos Digitales Para La Educación y La Cultura*, 33–39. http://ccita2011.itsmotul.edu.mx/documentos/Recursos_digitales.pdf

Villanueva Vázquez, A. (2019). Modelo Exploratorio de Calidad en la Educación Superior. *Dimensión Empresarial*, *18*((1)). https://doi.org/10.15665/dem.v18i(1).2239

# Anexos

A continuación, se adjuntan las capturas de pantalla, resultado de los diferentes modelados de datos.

## Anexo I. Modelado J48 personal administrativo

```
=== Run information ===

Scheme:       weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     admonweka-weka.filters.unsupervised
Instances:    572                    .attribute.Remove-R4
Attributes:   5
              Alta_Calidad
              Id_Sector
              Id_Caracter
              Nivel_admon
              Clase
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
------------------

Id_Caracter = 1: Bajo (28.0/14.0)
Id_Caracter = 2
|   Alta_Calidad = NO: Bajo (60.0/33.0)
|   Alta_Calidad = SI: Medio (12.0/4.0)
Id_Caracter = 3
|   Id_Sector = 1
|   |   Nivel_admon = Auxiliar
|   |   |   Alta_Calidad = NO: Medio (17.0/10.0)
|   |   |   Alta_Calidad = SI: Alto (15.0/8.0)
|   |   Nivel_admon = Directivo: Medio (32.0/18.0)
|   |   Nivel_admon = Profesional: Medio (32.0/18.0)
|   |   Nivel_admon = Servicios
|   |   |   Alta_Calidad = NO: Bajo (17.0/9.0)
|   |   |   Alta_Calidad = SI: Alto (15.0/9.0)
|   Id_Sector = 2
|   |   Nivel_admon = Auxiliar: MuyAlto (32.0/17.0)
|   |   Nivel_admon = Directivo
|   |   |   Alta_Calidad = NO: Alto (19.0/12.0)
|   |   |   Alta_Calidad = SI: Medio (13.0/7.0)
|   |   Nivel_admon = Profesional: MuyAlto (32.0/19.0)
|   |   Nivel_admon = Servicios: Alto (32.0/16.0)
Id_Caracter = 4
|   Alta_Calidad = NO
|   |   Nivel_admon = Auxiliar: Alto (23.0/12.0)
|   |   Nivel_admon = Directivo
|   |   |   Id_Sector = 1: Bajo (12.0/7.0)
|   |   |   Id_Sector = 2: Medio (11.0/5.0)
|   |   Nivel_admon = Profesional: Alto (23.0/14.0)
|   |   Nivel_admon = Servicios
|   |   |   Id_Sector = 1: Medio (12.0/7.0)
|   |   |   Id_Sector = 2: Alto (11.0/6.0)
|   Alta_Calidad = SI: MuyAlto (124.0/63.0)

Number of Leaves  :     21
Size of the tree :      33

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         209               36.5385 %
Incorrectly Classified Instances       363               63.4615 %
Kappa statistic                          0.1534
Mean absolute error                      0.3523
Root mean squared error                  0.4329
Relative absolute error                 93.9681 %
Root relative squared error             99.9908 %
Total Number of Instances              572

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               0,348     0,189     0,369       0,348   0,358      0,162    0,589     0,298     Bajo
               0,318     0,229     0,326       0,318   0,322      0,090    0,595     0,315     Medio
               0,268     0,200     0,306       0,268   0,286      0,071    0,582     0,285     Alto
               0,528     0,229     0,437       0,528   0,478      0,282    0,675     0,365     MuyAlto
Weighted Avg.  0,365     0,212     0,360       0,365   0,361      0,151    0,611     0,316

=== Confusion Matrix ===

  a  b  c  d   <-- classified as
 48 41 23 26 |  a = Bajo
 47 47 28 26 |  b = Medio
 23 35 38 46 |  c = Alto
 12 21 35 76 |  d = MuyAlto
```

# Anexo II. Modelado JRip personal administrativo

```
=== Run information ===

Scheme:       weka.classifiers.rules.JRip -F 3 -N 2.0 -O 2 -S 1
Relation:     admonweka-weka.filters.unsupervised.attribute.Remove-R4
Instances:    572
Attributes:   5
              Alta_Calidad
              Id_Sector
              Id_Caracter
              Nivel_admon
              Clase
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

JRIP rules:
===========

(Id_Sector = 1) and (Id_Caracter = 2) and (Alta_Calidad = NO)
 and (Nivel_admon = Servicios) => Clase=Bajo (8.0/2.0)

                               => Clase=Medio (564.0/417.0)
Number of Rules : 2


Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         172               30.0699 %
Incorrectly Classified Instances       400               69.9301 %
Kappa statistic                          0.0595
Mean absolute error                      0.3666
Root mean squared error                  0.4326
Relative absolute error                 97.7711 %
Root relative squared error             99.9225 %
Total Number of Instances              572

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,014    0,021    0,182      0,014    0,027      -0,019   0,545     0,265     Bajo
                 0,764    0,691    0,278      0,764    0,408      0,070    0,562     0,302     Medio
                 0,021    0,019    0,273      0,021    0,039      0,008    0,526     0,266     Alto
                 0,375    0,210    0,375      0,375    0,375      0,165    0,624     0,370     MuyAlto
Weighted Avg.    0,301    0,241    0,278      0,301    0,216      0,057    0,564     0,301

=== Confusion Matrix ===

   a   b   c   d   <-- classified as
   2 107   3  26 |   a = Bajo
   7 113   2  26 |   b = Medio
   1 100   3  38 |   c = Alto
   1  86   3  54 |   d = MuyAlto
```

# Anexo III. Modelado J48 personal docente

```
=== Run information ===

Scheme:       weka.classifiers.trees.J48 -C 0.25-M 2
Relation:     docente-weka.filters.unsupervised
                           .attribute.Remove-R5

Instances:    1831

Attributes:   5
              Alta_Calidad
              Id_Sector
              Id_Maximo_Nivel
              Id_DedicaciÃ³n
              Clase
Test mode:    10-fold cross-validation


=== Classifier model (full training set) ===

J48 pruned tree
------------------

Id_Maximo_Nivel = 1: Bajo (98.0/38.0)
Id_Maximo_Nivel = 2
|   Id_DedicaciÃ³n = 1
|   |   Alta_Calidad = NO: Alto (39.0/21.0)
|   |   Alta_Calidad = SI: MuyAlto (32.0/11.0)
|   Id_DedicaciÃ³n = 2: Medio (57.0/29.0)
|   Id_DedicaciÃ³n = 3
|   |   Alta_Calidad = NO: Bajo (11.0/3.0)
|   |   Alta_Calidad = SI: Medio (9.0/3.0)
|   Id_DedicaciÃ³n = 4: Alto (74.0/49.0)
Id_Maximo_Nivel = 3
|   Id_DedicaciÃ³n = 1: MuyAlto (72.0/20.0)
|   Id_DedicaciÃ³n = 2: Alto (66.0/35.0)
|   Id_DedicaciÃ³n = 3
|   |   Id_Sector = 1
|   |   |   Alta_Calidad = NO: Alto (8.0/3.0)
|   |   |   Alta_Calidad = SI: Bajo (6.0/4.0)
|   |   Id_Sector = 2: Bajo (21.0/11.0)
|   Id_DedicaciÃ³n = 4: MuyAlto (73.0/22.0)
Id_Maximo_Nivel = 4
|   Id_DedicaciÃ³n = 1: MuyAlto (71.0/30.0)
|   Id_DedicaciÃ³n = 2: Alto (65.0/32.0)
|   Id_DedicaciÃ³n = 3
|   |   Id_Sector = 1: Medio (14.0/8.0)
|   |   Id_Sector = 2: Bajo (22.0/14.0)
|   Id_DedicaciÃ³n = 4: MuyAlto (73.0/17.0)
Id_Maximo_Nivel = 5
|   Id_DedicaciÃ³n = 1
|   |   Id_Sector = 1: Alto (35.0/20.0)
|   |   Id_Sector = 2: MuyAlto (34.0/18.0)
|   Id_DedicaciÃ³n = 2: Alto (43.0/23.0)
|   Id_DedicaciÃ³n = 3
|   |   Alta_Calidad = NO: Alto (8.0/5.0)
|   |   Alta_Calidad = SI: Medio (11.0/6.0)
|   Id_DedicaciÃ³n = 4
|   |   Alta_Calidad = NO: Alto (20.0/10.0)
|   |   Alta_Calidad = SI: MuyAlto (33.0/13.0)
Id_Maximo_Nivel = 6
|   Id_DedicaciÃ³n = 1
|   |   Alta_Calidad = NO: Medio (30.0/15.0)
|   |   Alta_Calidad = SI: Bajo (19.0/9.0)
|   Id_DedicaciÃ³n = 2: Bajo (36.0/17.0)
|   Id_DedicaciÃ³n = 3: Bajo (15.0/5.0)
|   Id_DedicaciÃ³n = 4
|   |   Id_Sector = 1: Bajo (28.0/18.0)
|   |   Id_Sector = 2: Medio (27.0/15.0)
Id_Maximo_Nivel = 7
|   Id_DedicaciÃ³n = 1
|   |   Alta_Calidad = NO: MuyAlto (40.0/24.0)
|   |   Alta_Calidad = SI: Alto (31.0/17.0)
|   Id_DedicaciÃ³n = 2: Alto (64.0/37.0)
|   Id_DedicaciÃ³n = 3: Bajo (15.0/7.0)
|   Id_DedicaciÃ³n = 4: MuyAlto (82.0/32.0)
Id_Maximo_Nivel = 8
|   Id_DedicaciÃ³n = 1
|   |   Alta_Calidad = NO: Medio (22.0/10.0)
|   |   Alta_Calidad = SI: Bajo (18.0/8.0)
|   Id_DedicaciÃ³n = 2
|   |   Id_Sector = 1: Bajo (16.0/6.0)
|   |   Id_Sector = 2: Medio (21.0/8.0)
|   Id_DedicaciÃ³n = 3: Bajo (18.0/5.0)
|   Id_DedicaciÃ³n = 4
|   |   Id_Sector = 1: Medio (31.0/18.0)
|   |   Id_Sector = 2: Bajo (29.0/17.0)
Id_Maximo_Nivel = 9: Bajo (121.0/48.0)
Id_Maximo_Nivel = 10
|   Alta_Calidad = NO: Bajo (55.0/16.0)
|   Alta_Calidad = SI
|   |   Id_DedicaciÃ³n = 1: Bajo (15.0/6.0)
|   |   Id_DedicaciÃ³n = 2: Bajo (11.0/4.0)
|   |   Id_DedicaciÃ³n = 3: Bajo (0.0)
|   |   Id_DedicaciÃ³n = 4: Medio (19.0/9.0)
Id_Maximo_Nivel = 11: Alto (73.0/35.0)

Number of Leaves  :     50

Size of the tree :      74

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         881               48.1158 %
Incorrectly Classified Instances       950               51.8842 %
Kappa statistic                          0.3082
Mean absolute error                      0.3031
Root mean squared error                  0.3987
Relative absolute error                 80.823  %
Root relative squared error             92.0695 %
Total Number of Instances             1831

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,560    0,168    0,526      0,560    0,542      0,384    0,799     0,525     Bajo
                 0,258    0,146    0,370      0,258    0,304      0,127    0,640     0,345     Medio
                 0,441    0,226    0,393      0,441    0,415      0,207    0,683     0,368     Alto
                 0,665    0,151    0,596      0,665    0,629      0,497    0,839     0,605     MuyAlto
Weighted Avg.    0,481    0,173    0,471      0,481    0,473      0,304    0,740     0,461

=== Confusion Matrix ===

   a   b   c   d   <-- classified as
 256 114  68  19 |   a = Bajo
 159 118 127  54 |   b = Medio
  56  65 201 134 |   c = Alto
  16  22 116 306 |   d = MuyAlto
```

# Anexo IV. Modelado JRip personal docente

```
=== Run information ===

Scheme:       weka.classifiers.rules.JRip -F 3 -N 2.0 -O 2 -S 1
Relation:     docente-weka.filters.unsupervised.attribute.Remove-R5
Instances:    1831
Attributes:   5
              Alta_Calidad
              Id_Sector
              Id_Maximo_Nivel
              Id_DedicaciÃ³n
              Clase
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

JRIP rules:
===========

(Id_DedicaciÃ³n = 2) and (Id_Maximo_Nivel = 4) and (Id_Sector = 1) => Clase=Alto (31.0/11.0)
(Id_Maximo_Nivel = 5) and (Id_DedicaciÃ³n = 2) and (Alta_Calidad = SI) => Clase=Alto (25.0/11.0)
(Id_Maximo_Nivel = 1) and (Alta_Calidad = NO) => Clase=Bajo (52.0/12.0)
(Id_Maximo_Nivel = 9) => Clase=Bajo (121.0/48.0)
(Id_DedicaciÃ³n = 3) and (Id_Maximo_Nivel = 8) => Clase=Bajo (18.0/5.0)
(Id_Maximo_Nivel = 10) => Clase=Bajo (100.0/41.0)
(Id_DedicaciÃ³n = 3) and (Id_Maximo_Nivel = 6) => Clase=Bajo (15.0/5.0)
(Id_Maximo_Nivel = 8) and (Id_DedicaciÃ³n = 2) and (Id_Sector = 1) => Clase=Bajo (16.0/6.0)
(Id_DedicaciÃ³n = 3) and (Alta_Calidad = NO) and (Id_Maximo_Nivel = 2) => Clase=Bajo (11.0/3.0)
(Id_Maximo_Nivel = 6) and (Id_DedicaciÃ³n = 2) => Clase=Bajo (36.0/17.0)
(Id_Maximo_Nivel = 6) and (Id_DedicaciÃ³n = 1) and (Alta_Calidad = SI) => Clase=Bajo (19.0/9.0)
(Id_Maximo_Nivel = 8) and (Alta_Calidad = SI) and (Id_DedicaciÃ³n = 1) => Clase=Bajo (18.0/8.0)
(Id_Maximo_Nivel = 1) and (Id_DedicaciÃ³n = 2) => Clase=Bajo (12.0/4.0)
(Id_Maximo_Nivel = 8) and (Id_DedicaciÃ³n = 1) => Clase=Medio (22.0/10.0)
(Id_Maximo_Nivel = 8) and (Id_DedicaciÃ³n = 2) => Clase=Medio (21.0/8.0)
 => Clase=MuyAlto (1314.0/870.0)

Number of Rules : 16

Time taken to build model: 0.06 seconds


=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         706               38.5582 %
Incorrectly Classified Instances      1125               61.4418 %
Kappa statistic                          0.18
Mean absolute error                      0.3443
Root mean squared error                  0.4192
Relative absolute error                 91.8244 %
Root relative squared error             96.8151 %
Total Number of Instances             1831

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,462    0,123    0,555      0,462    0,504      0,361    0,690     0,445     Bajo
                 0,039    0,022    0,375      0,039    0,071      0,047    0,539     0,294     Medio
                 0,090    0,035    0,461      0,090    0,150      0,111    0,607     0,317     Alto
                 0,948    0,640    0,332      0,948    0,492      0,296    0,652     0,327     MuyAlto
Weighted Avg.    0,386    0,206    0,430      0,386    0,305      0,204    0,622     0,346

=== Confusion Matrix ===

   a   b   c   d   <-- classified as
 211  21  11 214 |   a = Bajo
 119  18  27 294 |   b = Medio
  40   5  41 370 |   c = Alto
  10   4  10 436 |   d = MuyAlto
```

# Anexo V. Modelado J48 personal estudiante

```
=== Run information ===

Scheme:       weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     estudiantes-weka.filters.unsupervised.attribute.Remove-R5,7
Instances:    4549
Attributes:   7
              Alta_Calidad
              Id_Sector
              Id_Caracter
              Id_Nivel
              Id_Metodologia
              Cod_DepartamentoPrograma
              Clase
Test mode:    10-fold cross-validation


=== Classifier model (full training set) ===

J48 pruned tree
------------------

Id_Nivel = 1
|   Id_Metodologia = 1
|   |   Cod_DepartamentoPrograma = 5
|   |   |   Id_Caracter = 1: Bajo (7.0)
|   |   |   Id_Caracter = 2
|   |   |   |   Alta_Calidad = NO: MuyAlto (20.0/8.0)
|   |   |   |   Alta_Calidad = SI: Medio (3.0/1.0)
|   |   |   Id_Caracter = 3: MuyAlto (48.0/14.0)
|   |   |   Id_Caracter = 4
|   |   |   |   Alta_Calidad = NO
|   |   |   |   |   Id_Sector = 1: Medio (12.0/7.0)
|   |   |   |   |   Id_Sector = 2: MuyAlto (10.0/4.0)
|   |   |   |   Alta_Calidad = SI: MuyAlto (28.0/10.0)
|   |   Cod_DepartamentoPrograma = 8: MuyAlto (87.0/37.0)
|   |   Cod_DepartamentoPrograma = 11: MuyAlto (123.0/33.0)
|   |   Cod_DepartamentoPrograma = 13
|   |   |   Id_Caracter = 1: Bajo (4.0)
|   |   |   Id_Caracter = 2: MuyAlto (21.0/10.0)
|   |   |   Id_Caracter = 3
|   |   |   |   Alta_Calidad = NO: MuyAlto (12.0/3.0)
|   |   |   |   Alta_Calidad = SI: Medio (6.0/2.0)
|   |   |   Id_Caracter = 4: MuyAlto (28.0/14.0)
|   |   Cod_DepartamentoPrograma = 15
|   |   |   Id_Caracter = 1: Bajo (4.0)
|   |   |   Id_Caracter = 2
|   |   |   |   Alta_Calidad = NO: MuyAlto (8.0/2.0)
|   |   |   |   Alta_Calidad = SI: Bajo (2.0)
|   |   |   Id_Caracter = 3: MuyAlto (10.0/4.0)
|   |   |   Id_Caracter = 4: MuyAlto (24.0/10.0)
|   |   Cod_DepartamentoPrograma = 17
|   |   |   Alta_Calidad = NO
|   |   |   |   Id_Sector = 1: Alto (15.0/8.0)
|   |   |   |   Id_Sector = 2: Bajo (10.0/5.0)
|   |   |   Alta_Calidad = SI: MuyAlto (28.0/11.0)
|   |   Cod_DepartamentoPrograma = 18
|   |   |   Alta_Calidad = NO
|   |   |   |   Id_Sector = 1: MuyAlto (11.0/3.0)
|   |   |   |   Id_Sector = 2: Bajo (5.0)
|   |   |   Alta_Calidad = SI: Bajo (10.0)
|   |   Cod_DepartamentoPrograma = 19
|   |   |   Id_Sector = 1: MuyAlto (25.0/9.0)
|   |   |   Id_Sector = 2
|   |   |   |   Id_Caracter = 1: Bajo (2.0)
|   |   |   |   Id_Caracter = 2: Bajo (0.0)
|   |   |   |   Id_Caracter = 3: Alto (9.0/5.0)
|   |   |   |   Id_Caracter = 4: Bajo (12.0/6.0)
|   |   Cod_DepartamentoPrograma = 20
|   |   |   Alta_Calidad = NO
|   |   |   |   Id_Sector = 1: MuyAlto (18.0/8.0)
|   |   |   |   Id_Sector = 2: Alto (14.0/8.0)
|   |   |   Alta_Calidad = SI: Bajo (6.0)
|   |   Cod_DepartamentoPrograma = 23: MuyAlto (39.0/21.0)
|   |   Cod_DepartamentoPrograma = 25
|   |   |   Id_Caracter = 1: Bajo (13.0/5.0)
|   |   |   Id_Caracter = 2
|   |   |   |   Id_Sector = 1: MuyAlto (12.0/4.0)
|   |   |   |   Id_Sector = 2: Medio (4.0/2.0)
|   |   |   Id_Caracter = 3
|   |   |   |   Alta_Calidad = NO: Alto (12.0/7.0)
|   |   |   |   Alta_Calidad = SI: Medio (2.0/1.0)
|   |   |   Id_Caracter = 4: MuyAlto (30.0/16.0)
|   |   Cod_DepartamentoPrograma = 27
|   |   |   Alta_Calidad = NO
|   |   |   |   Id_Caracter = 1: MuyAlto (0.0)
|   |   |   |   Id_Caracter = 2: Alto (6.0/3.0)
|   |   |   |   Id_Caracter = 3: Bajo (4.0/2.0)
|   |   |   |   Id_Caracter = 4
|   |   |   |   |   Id_Sector = 1: MuyAlto (8.0/4.0)
|   |   |   |   |   Id_Sector = 2: Bajo (3.0/2.0)
|   |   |   Alta_Calidad = SI: Bajo (5.0/1.0)
|   |   Cod_DepartamentoPrograma = 41
|   |   |   Id_Caracter = 1: Alto (9.0/4.0)
|   |   |   Id_Caracter = 2
|   |   |   |   Alta_Calidad = NO: MuyAlto (7.0/3.0)
|   |   |   |   Alta_Calidad = SI: Bajo (2.0)
|   |   |   Id_Caracter = 3: Alto (9.0/4.0)
|   |   |   Id_Caracter = 4: MuyAlto (20.0/10.0)
|   |   Cod_DepartamentoPrograma = 44
|   |   |   Id_Sector = 1
|   |   |   |   Alta_Calidad = NO: MuyAlto (19.0/10.0)
|   |   |   |   Alta_Calidad = SI: Bajo (3.0/1.0)
|   |   |   Id_Sector = 2: Bajo (8.0)
|   |   Cod_DepartamentoPrograma = 47
|   |   |   Id_Caracter = 1: Alto (16.0/6.0)
|   |   |   Id_Caracter = 2: Alto (8.0/4.0)
|   |   |   Id_Caracter = 3: Bajo (2.0)
|   |   |   Id_Caracter = 4: MuyAlto (15.0/4.0)
|   |   Cod_DepartamentoPrograma = 50
|   |   |   Id_Caracter = 1: Bajo (5.0)
|   |   |   Id_Caracter = 2: MuyAlto (11.0/7.0)
|   |   |   Id_Caracter = 3: Alto (13.0/7.0)
|   |   |   Id_Caracter = 4
|   |   |   |   Alta_Calidad = NO: MuyAlto (13.0/5.0)
|   |   |   |   Alta_Calidad = SI: Bajo (12.0/5.0)
|   |   Cod_DepartamentoPrograma = 52
|   |   |   Id_Caracter = 1: Bajo (11.0/2.0)
|   |   |   Id_Caracter = 2
|   |   |   |   Alta_Calidad = NO: Alto (8.0/4.0)
|   |   |   |   Alta_Calidad = SI: Bajo (2.0)
|   |   |   Id_Caracter = 3: Alto (11.0/7.0)
|   |   |   Id_Caracter = 4: MuyAlto (27.0/15.0)
|   |   Cod_DepartamentoPrograma = 54
|   |   |   Alta_Calidad = NO
|   |   |   |   Id_Sector = 1: MuyAlto (22.0/7.0)
|   |   |   |   Id_Sector = 2: Alto (17.0/8.0)
|   |   |   Alta_Calidad = SI
|   |   |   |   Id_Sector = 1: Bajo (5.0)
|   |   |   |   Id_Sector = 2: MuyAlto (6.0/3.0)
|   |   Cod_DepartamentoPrograma = 63
|   |   |   Id_Caracter = 1
|   |   |   |   Id_Sector = 1: Alto (4.0/1.0)
|   |   |   |   Id_Sector = 2: Medio (8.0/5.0)
|   |   |   Id_Caracter = 2: MuyAlto (9.0/4.0)
|   |   |   Id_Caracter = 3: Alto (11.0/4.0)
|   |   |   Id_Caracter = 4: MuyAlto (20.0/11.0)
|   |   Cod_DepartamentoPrograma = 66
|   |   |   Id_Caracter = 1
|   |   |   |   Id_Sector = 1: Medio (3.0/1.0)
|   |   |   |   Id_Sector = 2: Bajo (4.0/2.0)
|   |   |   Id_Caracter = 2
|   |   |   |   Alta_Calidad = NO: Alto (11.0/5.0)
|   |   |   |   Alta_Calidad = SI: Bajo (2.0)
|   |   |   Id_Caracter = 3: Medio (14.0/9.0)
|   |   |   Id_Caracter = 4: MuyAlto (28.0/15.0)
```

```
|   |   Cod_DepartamentoPrograma = 68
|   |   |   Id_Sector = 1
|   |   |   |   Alta_Calidad = NO
|   |   |   |   |   Id_Caracter = 1: MuyAlto (0.0)
|   |   |   |   |   Id_Caracter = 2: MuyAlto (12.0/2.0)
|   |   |   |   |   Id_Caracter = 3: Alto (9.0/3.0)
|   |   |   |   |   Id_Caracter = 4: MuyAlto (0.0)
|   |   |   |   Alta_Calidad = SI
|   |   |   |   |   Id_Caracter = 1: MuyAlto (0.0)
|   |   |   |   |   Id_Caracter = 2: Bajo (3.0)
|   |   |   |   |   Id_Caracter = 3: MuyAlto (1.0)
|   |   |   |   |   Id_Caracter = 4: MuyAlto (9.0/3.0)
|   |   |   Id_Sector = 2: MuyAlto (49.0/28.0)
|   |   Cod_DepartamentoPrograma = 70
|   |   |   Alta_Calidad = NO: MuyAlto (31.0/18.0)
|   |   |   Alta_Calidad = SI
|   |   |   |   Id_Sector = 1: Bajo (5.0/1.0)
|   |   |   |   Id_Sector = 2: Medio (5.0/2.0)
|   |   Cod_DepartamentoPrograma = 73
|   |   |   Id_Caracter = 1
|   |   |   |   Id_Sector = 1: MuyAlto (10.0/3.0)
|   |   |   |   Id_Sector = 2: Alto (12.0/2.0)
|   |   |   Id_Caracter = 2
|   |   |   |   Alta_Calidad = NO
|   |   |   |   |   Id_Sector = 1: MuyAlto (8.0/3.0)
|   |   |   |   |   Id_Sector = 2: Alto (2.0)
|   |   |   |   Alta_Calidad = SI: Bajo (3.0)
|   |   |   Id_Caracter = 3: Bajo (11.0/5.0)
|   |   |   Id_Caracter = 4
|   |   |   |   Alta_Calidad = NO: MuyAlto (16.0/5.0)
|   |   |   |   Alta_Calidad = SI
|   |   |   |   |   Id_Sector = 1: Bajo (9.0/2.0)
|   |   |   |   |   Id_Sector = 2: Alto (9.0/5.0)
|   |   Cod_DepartamentoPrograma = 76
|   |   |   Id_Caracter = 1
|   |   |   |   Id_Sector = 1: MuyAlto (12.0/7.0)
|   |   |   |   Id_Sector = 2: Bajo (11.0/5.0)
|   |   |   Id_Caracter = 2
|   |   |   |   Alta_Calidad = NO: MuyAlto (15.0/4.0)
|   |   |   |   Alta_Calidad = SI: Bajo (2.0)
|   |   |   Id_Caracter = 3
|   |   |   |   Alta_Calidad = NO: MuyAlto (19.0/4.0)
|   |   |   |   Alta_Calidad = SI: Alto (4.0/1.0)
|   |   |   Id_Caracter = 4: MuyAlto (43.0/10.0)
|   |   Cod_DepartamentoPrograma = 85
|   |   |   Alta_Calidad = NO: Alto (19.0/11.0)
|   |   |   Alta_Calidad = SI: Bajo (12.0/3.0)
|   |   Cod_DepartamentoPrograma = 86
|   |   |   Id_Caracter = 1: Bajo (3.0)
|   |   |   Id_Caracter = 2: Alto (9.0/4.0)
|   |   |   Id_Caracter = 3: Bajo (3.0)
|   |   |   Id_Caracter = 4: Bajo (10.0/1.0)
|   |   Cod_DepartamentoPrograma = 88: Bajo (17.0/6.0)
|   |   Cod_DepartamentoPrograma = 91
|   |   |   Id_Sector = 1: Medio (6.0/3.0)
|   |   |   Id_Sector = 2: Bajo (7.0)
|   Id_Metodologia = 2
|   |   Cod_DepartamentoPrograma = 5
|   |   |   Alta_Calidad = NO: MuyAlto (31.0/17.0)
|   |   |   Alta_Calidad = SI: Bajo (13.0/8.0)
|   |   Cod_DepartamentoPrograma = 8
|   |   |   Id_Caracter = 1: Medio (3.0)
|   |   |   Id_Caracter = 2: Medio (3.0/1.0)
|   |   |   Id_Caracter = 3
|   |   |   |   Alta_Calidad = NO: Alto (7.0/3.0)
|   |   |   |   Alta_Calidad = SI: Medio (2.0)
|   |   |   Id_Caracter = 4: Alto (16.0/9.0)
|   |   Cod_DepartamentoPrograma = 11
|   |   |   Id_Caracter = 1: MuyAlto (6.0/3.0)
|   |   |   Id_Caracter = 2: Medio (6.0/2.0)
|   |   |   Id_Caracter = 3: Alto (14.0/8.0)
|   |   |   Id_Caracter = 4: MuyAlto (28.0/16.0)
|   |   Cod_DepartamentoPrograma = 13
|   |   |   Alta_Calidad = NO: Alto (22.0/12.0)
|   |   |   Alta_Calidad = SI
|   |   |   |   Id_Sector = 1: MuyAlto (4.0)
|   |   |   |   Id_Sector = 2: Medio (9.0/3.0)
|   |   Cod_DepartamentoPrograma = 15
|   |   |   Id_Sector = 1: MuyAlto (17.0/9.0)
|   |   |   Id_Sector = 2: Alto (13.0/8.0)
|   |   Cod_DepartamentoPrograma = 17: Alto (30.0/17.0)
|   |   Cod_DepartamentoPrograma = 18
|   |   |   Id_Caracter = 1: Medio (3.0/1.0)
|   |   |   Id_Caracter = 2: Medio (1.0)
|   |   |   Id_Caracter = 3: Bajo (5.0/1.0)
|   |   |   Id_Caracter = 4: Alto (11.0/6.0)
|   |   Cod_DepartamentoPrograma = 19
|   |   |   Id_Caracter = 1: Alto (4.0/2.0)
|   |   |   Id_Caracter = 2: Medio (1.0)
|   |   |   Id_Caracter = 3: Bajo (5.0/3.0)
|   |   |   Id_Caracter = 4
|   |   |   |   Alta_Calidad = NO: Alto (11.0/5.0)
|   |   |   |   Alta_Calidad = SI: Bajo (5.0/2.0)
|   |   Cod_DepartamentoPrograma = 20: Medio (26.0/16.0)
|   |   Cod_DepartamentoPrograma = 23
|   |   |   Id_Caracter = 1: MuyAlto (4.0/2.0)
|   |   |   Id_Caracter = 2: Medio (3.0)
|   |   |   Id_Caracter = 3
|   |   |   |   Alta_Calidad = NO: Alto (7.0/4.0)
|   |   |   |   Alta_Calidad = SI: Bajo (2.0/1.0)
|   |   |   Id_Caracter = 4: Alto (19.0/8.0)
|   |   Cod_DepartamentoPrograma = 25
|   |   |   Id_Sector = 1: MuyAlto (16.0/8.0)
|   |   |   Id_Sector = 2
|   |   |   |   Id_Caracter = 1: Alto (4.0/1.0)
|   |   |   |   Id_Caracter = 3: Bajo (7.0/2.0)
|   |   |   |   Id_Caracter = 4: Medio (4.0/1.0)
|   |   Cod_DepartamentoPrograma = 27: Bajo (22.0/12.0)
|   |   Cod_DepartamentoPrograma = 41
|   |   |   Alta_Calidad = NO: Alto (20.0/10.0)
|   |   |   Alta_Calidad = SI
|   |   |   |   Id_Sector = 1: Bajo (3.0/1.0)
|   |   |   |   Id_Sector = 2: Medio (5.0/2.0)
|   |   Cod_DepartamentoPrograma = 44: Medio (20.0/13.0)
|   |   Cod_DepartamentoPrograma = 47
|   |   |   Id_Caracter = 1: MuyAlto (5.0/2.0)
|   |   |   Id_Caracter = 2: Alto (4.0/2.0)
|   |   |   Id_Caracter = 3: Bajo (5.0/3.0)
|   |   |   Id_Caracter = 4
|   |   |   |   Id_Sector = 1: MuyAlto (15.0/9.0)
|   |   |   |   Id_Sector = 2: Medio (3.0)
|   |   Cod_DepartamentoPrograma = 50
|   |   |   Id_Sector = 1: MuyAlto (13.0/8.0)
|   |   |   Id_Sector = 2: Alto (16.0/7.0)
```

```
|   |   Cod_DepartamentoPrograma = 52
|   |   |   Alta_Calidad = NO: MuyAlto (22.0/13.0)
|   |   |   Alta_Calidad = SI
|   |   |   |   Id_Sector = 1: Bajo (4.0)
|   |   |   |   Id_Sector = 2: Medio (4.0/2.0)
|   |   Cod_DepartamentoPrograma = 54
|   |   |   Id_Sector = 1
|   |   |   |   Id_Caracter = 1: MuyAlto (0.0)
|   |   |   |   Id_Caracter = 2: Bajo (4.0/2.0)
|   |   |   |   Id_Caracter = 3: MuyAlto (1.0)
|   |   |   |   Id_Caracter = 4: MuyAlto (11.0/4.0)
|   |   |   Id_Sector = 2
|   |   |   |   Id_Caracter = 1: Alto (3.0/1.0)
|   |   |   |   Id_Caracter = 2: Alto (3.0/1.0)
|   |   |   |   Id_Caracter = 3: Bajo (3.0/2.0)
|   |   |   |   Id_Caracter = 4: Medio (7.0/3.0)
|   |   Cod_DepartamentoPrograma = 63
|   |   |   Id_Sector = 1
|   |   |   |   Alta_Calidad = NO: Medio (5.0/3.0)
|   |   |   |   Alta_Calidad = SI: MuyAlto (5.0/2.0)
|   |   |   Id_Sector = 2: Medio (13.0/6.0)
|   |   Cod_DepartamentoPrograma = 66: Alto (29.0/12.0)
|   |   Cod_DepartamentoPrograma = 68: Alto (36.0/21.0)
|   |   Cod_DepartamentoPrograma = 70
|   |   |   Id_Caracter = 1: MuyAlto (4.0/2.0)
|   |   |   Id_Caracter = 2: Bajo (2.0/1.0)
|   |   |   Id_Caracter = 3: Medio (10.0/5.0)
|   |   |   Id_Caracter = 4
|   |   |   |   Id_Sector = 1: Alto (11.0/6.0)
|   |   |   |   Id_Sector = 2: Bajo (5.0/2.0)
|   |   Cod_DepartamentoPrograma = 73
|   |   |   Id_Sector = 1
|   |   |   |   Alta_Calidad = NO: MuyAlto (13.0/6.0)
|   |   |   |   Alta_Calidad = SI: Bajo (3.0)
|   |   |   Id_Sector = 2
|   |   |   |   Id_Caracter = 1: Alto (5.0/2.0)
|   |   |   |   Id_Caracter = 2: Medio (0.0)
|   |   |   |   Id_Caracter = 3: Bajo (6.0/3.0)
|   |   |   |   Id_Caracter = 4: Medio (5.0/2.0)
|   |   Cod_DepartamentoPrograma = 76: MuyAlto (41.0/23.0)
|   |   Cod_DepartamentoPrograma = 85
|   |   |   Alta_Calidad = NO: Bajo (21.0/15.0)
|   |   |   Alta_Calidad = SI: Medio (7.0/3.0)
|   |   Cod_DepartamentoPrograma = 86
|   |   |   Id_Caracter = 1: Bajo (3.0/2.0)
|   |   |   Id_Caracter = 2: Medio (1.0)
|   |   |   Id_Caracter = 3: Alto (5.0/1.0)
|   |   |   Id_Caracter = 4: Bajo (14.0/8.0)
|   |   Cod_DepartamentoPrograma = 88: Bajo (2.0/1.0)
|   |   Cod_DepartamentoPrograma = 91
|   |   |   Id_Caracter = 1: Bajo (3.0/2.0)
|   |   |   Id_Caracter = 2: Medio (1.0)
|   |   |   Id_Caracter = 3: Medio (5.0/1.0)
|   |   |   Id_Caracter = 4: Bajo (13.0/4.0)
|   Id_Metodologia = 3
|   |   Id_Caracter = 4
|   |   |   Cod_DepartamentoPrograma = 1: Medio (0.0)
|   |   |   Cod_DepartamentoPrograma = 3: Medio (0.0)
|   |   |   Cod_DepartamentoPrograma = 5
|   |   |   |   Alta_Calidad = NO: Alto (9.0/3.0)
|   |   |   |   Alta_Calidad = SI: Bajo (5.0/3.0)
|   |   |   Cod_DepartamentoPrograma = 8
|   |   |   |   Alta_Calidad = NO: Medio (12.0/4.0)
|   |   |   |   Alta_Calidad = SI: Bajo (4.0/1.0)
|   |   |   Cod_DepartamentoPrograma = 11: MuyAlto (23.0/14.0
|   |   |   Cod_DepartamentoPrograma = 13
|   |   |   |   Id_Sector = 1: Bajo (10.0/4.0)
|   |   |   |   Id_Sector = 2: Alto (5.0/2.0)
|   |   |   Cod_DepartamentoPrograma = 15: Alto (16.0/10.0)
|   |   |   Cod_DepartamentoPrograma = 17
|   |   |   |   Id_Sector = 1: Bajo (10.0/3.0)
|   |   |   |   Id_Sector = 2: MuyAlto (4.0/2.0)
|   |   |   Cod_DepartamentoPrograma = 18: Bajo (10.0/5.0)
|   |   |   Cod_DepartamentoPrograma = 19: Bajo (10.0/6.0)
|   |   |   Cod_DepartamentoPrograma = 20: Medio (9.0/4.0)
|   |   |   Cod_DepartamentoPrograma = 23: Bajo (10.0/3.0)
|   |   |   Cod_DepartamentoPrograma = 25
|   |   |   |   Alta_Calidad = NO: Medio (9.0/4.0)
|   |   |   |   Alta_Calidad = SI: Alto (2.0)
|   |   |   Cod_DepartamentoPrograma = 27: Bajo (10.0/1.0)
|   |   |   Cod_DepartamentoPrograma = 41: Medio (10.0/5.0)
|   |   |   Cod_DepartamentoPrograma = 44: Bajo (8.0/4.0)
|   |   |   Cod_DepartamentoPrograma = 47: Bajo (9.0/4.0)
|   |   |   Cod_DepartamentoPrograma = 50: Medio (10.0/5.0)
|   |   |   Cod_DepartamentoPrograma = 52
|   |   |   |   Alta_Calidad = NO: Medio (9.0/3.0)
|   |   |   |   Alta_Calidad = SI: Bajo (3.0/1.0)
|   |   |   Cod_DepartamentoPrograma = 54: Medio (10.0/4.0)
|   |   |   Cod_DepartamentoPrograma = 63: MuyAlto (1.0)
|   |   |   Cod_DepartamentoPrograma = 66: Medio (10.0/4.0)
|   |   |   Cod_DepartamentoPrograma = 68
|   |   |   |   Alta_Calidad = NO: Medio (13.0/7.0)
|   |   |   |   Alta_Calidad = SI
|   |   |   |   |   Id_Sector = 1: Bajo (4.0/1.0)
|   |   |   |   |   Id_Sector = 2: Medio (8.0/5.0)
|   |   |   Cod_DepartamentoPrograma = 70: Medio (10.0/5.0)
|   |   |   Cod_DepartamentoPrograma = 73: Medio (11.0/5.0)
|   |   |   Cod_DepartamentoPrograma = 76: Medio (10.0/5.0)
|   |   |   Cod_DepartamentoPrograma = 85: Medio (12.0/5.0)
|   |   |   Cod_DepartamentoPrograma = 86: Bajo (9.0/2.0)
|   |   |   Cod_DepartamentoPrograma = 88: Medio (0.0)
|   |   |   Cod_DepartamentoPrograma = 91: Bajo (8.0/1.0)
```

```
Id_Nivel = 2
|   Cod_DepartamentoPrograma = 5
|   |   Id_Metodologia = 1
|   |   |   Id_Caracter = 1: Medio (0.0)
|   |   |   Id_Caracter = 2: Medio (7.0/4.0)
|   |   |   Id_Caracter = 3: Medio (30.0/13.0)
|   |   |   Id_Caracter = 4
|   |   |   |   Id_Sector = 1: Alto (24.0/14.0)
|   |   |   |   Id_Sector = 2: MuyAlto (32.0/20.0)
|   |   Id_Metodologia = 2: Medio (18.0/9.0)
|   |   Id_Metodologia = 3
|   |   |   Id_Caracter = 1: Medio (0.0)
|   |   |   Id_Caracter = 2: Alto (2.0)
|   |   |   Id_Caracter = 3
|   |   |   |   Id_Sector = 1: Bajo (2.0/1.0)
|   |   |   |   Id_Sector = 2: Alto (5.0/3.0)
|   |   |   Id_Caracter = 4: Medio (23.0/7.0)
|   Cod_DepartamentoPrograma = 8
|   |   Id_Sector = 1: Medio (36.0/18.0)
|   |   Id_Sector = 2
|   |   |   Id_Metodologia = 1
|   |   |   |   Alta_Calidad = NO: Alto (14.0/7.0)
|   |   |   |   Alta_Calidad = SI: MuyAlto (19.0/10.0)
|   |   |   Id_Metodologia = 2
|   |   |   |   Alta_Calidad = NO: Medio (3.0/2.0)
|   |   |   |   Alta_Calidad = SI: Bajo (2.0)
|   |   |   Id_Metodologia = 3: Alto (6.0/1.0)
|   Cod_DepartamentoPrograma = 11
|   |   Id_Caracter = 1: Alto (1.0)
|   |   Id_Caracter = 2
|   |   |   Alta_Calidad = NO: Alto (8.0/1.0)
|   |   |   Alta_Calidad = SI: Bajo (2.0)
|   |   Id_Caracter = 3
|   |   |   Alta_Calidad = NO: MuyAlto (45.0/28.0)
|   |   |   Alta_Calidad = SI
|   |   |   |   Id_Sector = 1: Medio (5.0/1.0)
|   |   |   |   Id_Sector = 2: MuyAlto (9.0/3.0)
|   |   Id_Caracter = 4
|   |   |   Id_Metodologia = 1: MuyAlto (63.0/30.0)
|   |   |   Id_Metodologia = 2: Alto (12.0/2.0)
|   |   |   Id_Metodologia = 3
|   |   |   |   Id_Sector = 1: Medio (9.0/3.0)
|   |   |   |   Id_Sector = 2: Alto (15.0/8.0)
|   Cod_DepartamentoPrograma = 13
|   |   Id_Metodologia = 1
|   |   |   Id_Caracter = 1: Alto (0.0)
|   |   |   Id_Caracter = 2: Bajo (2.0)
|   |   |   Id_Caracter = 3: Medio (4.0/2.0)
|   |   |   Id_Caracter = 4: Alto (36.0/23.0)
|   |   Id_Metodologia = 2: Medio (12.0/5.0)
|   |   Id_Metodologia = 3: Bajo (6.0/3.0)
|   Cod_DepartamentoPrograma = 15
|   |   Id_Caracter = 1: Bajo (0.0)
|   |   Id_Caracter = 2: Alto (4.0/2.0)
|   |   Id_Caracter = 3: Medio (10.0/5.0)
|   |   Id_Caracter = 4
|   |   |   Id_Metodologia = 1
|   |   |   |   Id_Sector = 1: Medio (19.0/12.0)
|   |   |   |   Id_Sector = 2
|   |   |   |   |   Alta_Calidad = NO: Medio (7.0/5.0)
|   |   |   |   |   Alta_Calidad = SI: Bajo (14.0/6.0)
|   |   |   Id_Metodologia = 2: Bajo (4.0/1.0)
|   |   |   Id_Metodologia = 3
|   |   |   |   Alta_Calidad = NO: Bajo (7.0/2.0)
|   |   |   |   Alta_Calidad = SI: Alto (4.0/1.0)
|   Cod_DepartamentoPrograma = 17
|   |   Alta_Calidad = NO
|   |   |   Id_Caracter = 1: Bajo (0.0)
|   |   |   Id_Caracter = 2: Medio (6.0/2.0)
|   |   |   Id_Caracter = 3: Medio (7.0/4.0)
|   |   Alta_Calidad = SI: Alto (47.0/28.0)
|   Cod_DepartamentoPrograma = 18
|   |   Id_Metodologia = 1
|   |   |   Alta_Calidad = NO: Medio (15.0/8.0)
|   |   |   Alta_Calidad = SI: Bajo (4.0)
|   |   Id_Metodologia = 2
|   |   |   Id_Caracter = 1: Bajo (0.0)
|   |   |   Id_Caracter = 2: Bajo (0.0)
|   |   |   Id_Caracter = 3: Medio (3.0/1.0)
|   |   |   Id_Caracter = 4: Bajo (3.0)
|   |   Id_Metodologia = 3: Bajo (6.0)
|   Cod_DepartamentoPrograma = 19
|   |   Id_Caracter = 1: Medio (0.0)
|   |   Id_Caracter = 2: Medio (3.0/1.0)
|   |   Id_Caracter = 3
|   |   |   Id_Sector = 1: Medio (6.0/1.0)
|   |   |   Id_Sector = 2
|   |   |   |   Id_Metodologia = 1: Alto (4.0)
|   |   |   |   Id_Metodologia = 2: Medio (2.0)
|   |   |   |   Id_Metodologia = 3: Alto (0.0)
|   |   Id_Caracter = 4
|   |   |   Alta_Calidad = NO: Medio (12.0/6.0)
|   |   |   Alta_Calidad = SI
|   |   |   |   Id_Sector = 1: Medio (18.0/8.0)
|   |   |   |   Id_Sector = 2
|   |   |   |   |   Id_Metodologia = 1: Bajo (6.0/2.0)
|   |   |   |   |   Id_Metodologia = 2: Alto (3.0/1.0)
|   |   |   |   |   Id_Metodologia = 3: Bajo (0.0)
|   Cod_DepartamentoPrograma = 20: Bajo (29.0/14.0)
|   Cod_DepartamentoPrograma = 23
|   |   Id_Sector = 1
|   |   |   Id_Caracter = 1: Bajo (0.0)
|   |   |   Id_Caracter = 2: Bajo (0.0)
|   |   |   Id_Caracter = 3: Medio (2.0)
|   |   |   Id_Caracter = 4: Bajo (19.0/8.0)
|   |   Id_Sector = 2
|   |   |   Id_Metodologia = 1
|   |   |   |   Alta_Calidad = NO: Alto (2.0)
|   |   |   |   Alta_Calidad = SI: Medio (8.0/5.0)
|   |   |   Id_Metodologia = 2: Medio (5.0/2.0)
|   |   |   Id_Metodologia = 3: Medio (0.0)
|   Cod_DepartamentoPrograma = 25
|   |   Id_Sector = 1: Medio (32.0/15.0)
|   |   Id_Sector = 2
|   |   |   Id_Caracter = 3: Bajo (2.0)
|   |   |   Id_Caracter = 4: Alto (18.0/9.0)
|   Cod_DepartamentoPrograma = 27
|   |   Id_Sector = 1: Bajo (16.0/5.0)
|   |   Id_Sector = 2
|   |   |   Alta_Calidad = NO: Alto (4.0/1.0)
|   |   |   Alta_Calidad = SI: Medio (8.0/4.0)
|   Cod_DepartamentoPrograma = 41
|   |   Alta_Calidad = NO: Medio (23.0/8.0)
|   |   Alta_Calidad = SI
|   |   |   Id_Sector = 1: Medio (14.0/9.0)
|   |   |   Id_Sector = 2: Bajo (11.0/4.0)
|   Cod_DepartamentoPrograma = 44: Bajo (22.0/11.0)
|   Cod_DepartamentoPrograma = 47: Bajo (41.0/24.0)
|   Cod_DepartamentoPrograma = 50
|   |   Id_Sector = 1: Medio (27.0/12.0)
|   |   Id_Sector = 2: Bajo (15.0/8.0)
|   Cod_DepartamentoPrograma = 52: Medio (58.0/29.0)
|   Cod_DepartamentoPrograma = 54
|   |   Alta_Calidad = NO: Medio (32.0/16.0)
|   |   Alta_Calidad = SI: Bajo (14.0/8.0)
|   Cod_DepartamentoPrograma = 63: Medio (35.0/17.0)
|   Cod_DepartamentoPrograma = 66
|   |   Id_Caracter = 1: Medio (0.0)
|   |   Id_Caracter = 2: Medio (6.0/1.0)
|   |   Id_Caracter = 3
```

```
|   |   |   Id_Sector = 2: Alto (8.0/4.0)
|   |   Id_Caracter = 4
|   |   |   Id_Metodologia = 1: Alto (38.0/24.0)
|   |   |   Id_Metodologia = 2: Bajo (3.0/1.0)
|   |   |   Id_Metodologia = 3: Bajo (7.0/3.0)
|   Cod_DepartamentoPrograma = 68: Medio (94.0/53.0)
|   Cod_DepartamentoPrograma = 70
|   |   Id_Sector = 1: Bajo (15.0/6.0)
|   |   Id_Sector = 2: Medio (13.0/5.0)
|   Cod_DepartamentoPrograma = 73
|   |   Alta_Calidad = NO: Medio (39.0/22.0)
|   |   Alta_Calidad = SI: Bajo (20.0/9.0)
|   Cod_DepartamentoPrograma = 76
|   |   Id_Metodologia = 1
|   |   |   Id_Caracter = 1: Alto (0.0)
|   |   |   Id_Caracter = 2
|   |   |   |   Alta_Calidad = NO: Alto (5.0/2.0)
|   |   |   |   Alta_Calidad = SI: Bajo (2.0)
|   |   |   Id_Caracter = 3
|   |   |   |   Id_Sector = 1: Alto (4.0/1.0)
|   |   |   |   Id_Sector = 2: Medio (4.0/1.0)
|   |   |   Id_Caracter = 4
|   |   |   |   Id_Sector = 1: Alto (20.0/8.0)
|   |   |   |   Id_Sector = 2: MuyAlto (26.0/15.0)
|   |   Id_Metodologia = 2: Medio (15.0/5.0)
|   |   Id_Metodologia = 3
|   |   |   Id_Sector = 1: Medio (10.0/5.0)
|   |   |   Id_Sector = 2
|   |   |   |   Alta_Calidad = NO: Alto (2.0)
|   |   |   |   Alta_Calidad = SI: Medio (3.0/1.0)
|   Cod_DepartamentoPrograma = 85
|   |   Id_Caracter = 3
|   |   |   Id_Sector = 1: Medio (2.0)
|   |   |   Id_Sector = 2: Bajo (5.0/3.0)
|   |   Id_Caracter = 4: Bajo (19.0/5.0)
|   Cod_DepartamentoPrograma = 86
|   |   Id_Sector = 1
|   |   |   Id_Caracter = 3: Medio (2.0/1.0)
|   |   |   Id_Caracter = 4: Bajo (7.0/2.0)
|   |   Id_Sector = 2
|   |   |   Alta_Calidad = NO: Medio (5.0/1.0)
|   |   |   Alta_Calidad = SI: Bajo (3.0/1.0)
|   Cod_DepartamentoPrograma = 88: Medio (5.0/2.0)
|   Cod_DepartamentoPrograma = 91
|   |   Id_Sector = 1
|   |   |   Id_Caracter = 3: Medio (2.0)
|   |   |   Id_Caracter = 4
|   |   |   |   Id_Metodologia = 1: Medio (5.0/1.0)
|   |   |   |   Id_Metodologia = 2: Bajo (2.0)
|   |   |   |   Id_Metodologia = 3: Bajo (4.0/1.0)
|   |   Id_Sector = 2: Bajo (2.0)

Number of Leaves  :     440

Size of the tree :      618

Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        2019               44.3834 %
Incorrectly Classified Instances      2530               55.6166 %
Kappa statistic                          0.2584
Mean absolute error                      0.3141
Root mean squared error                  0.4199
Relative absolute error                 83.7547 %
Root relative squared error             96.9793 %
Total Number of Instances             4549

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               0,465    0,142    0,517      0,465   0,490      0,335    0,724     0,520     Bajo
               0,397    0,212    0,386      0,397   0,391      0,183    0,648     0,355     Medio
               0,270    0,177    0,338      0,270   0,300      0,100    0,589     0,312     Alto
               0,645    0,210    0,507      0,645   0,567      0,405    0,794     0,512     MuyAlto
Weighted Avg.  0,444    0,186    0,436      0,444   0,437      0,255    0,688     0,424

=== Confusion Matrix ===

   a   b   c   d   <-- classified as
 522 299 155 146 |   a = Bajo
 279 454 230 181 |   b = Medio
 134 314 309 388 |   c = Alto
  75 110 219 734 |   d = MuyAlto
```

# Anexo VI. Modelado JRip personal estudiante

```
=== Run information ===

Scheme:       weka.classifiers.rules.JRip -F 3 -N 2.0 -O 2 -S 1
Relation:     estudiantes-weka.filters.unsupervised.attribute.Remove-R5,7
Instances:    4549
Attributes:   7
              Alta_Calidad
              Id_Sector
              Id_Caracter
              Id_Nivel
              Id_Metodologia
              Cod_DepartamentoPrograma
              Clase
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

JRIP rules:
===========

(Id_Metodologia = 3) and (Id_Nivel = 2) and (Id_Sector = 1) and (Alta_Calidad = NO) and (Id_Caracter = 4) => Clase=Bajo (157.0/60.0)
(Id_Metodologia = 3) and (Id_Caracter = 3) and (Id_Nivel = 1) => Clase=Bajo (99.0/40.0)
(Id_Caracter = 1) and (Id_Metodologia = 3) => Clase=Bajo (43.0/13.0)
(Alta_Calidad = SI) and (Id_Caracter = 2) and (Id_Sector = 1) => Clase=Bajo (58.0/11.0)
(Id_Caracter = 4) and (Id_Metodologia = 2) and (Id_Nivel = 2) and (Id_Sector = 1) and (Alta_Calidad = NO) => Clase=Bajo (80.0/31.0)
(Alta_Calidad = SI) and (Cod_DepartamentoPrograma = 18) => Clase=Bajo (16.0/0.0)
(Id_Caracter = 4) and (Alta_Calidad = SI) and (Id_Nivel = 1) and (Cod_DepartamentoPrograma = 27) => Clase=Bajo (10.0/1.0)
(Id_Caracter = 1) and (Id_Metodologia = 1) and (Id_Sector = 2) and (Cod_DepartamentoPrograma = 52) => Clase=Bajo (10.0/1.0)
(Alta_Calidad = SI) and (Id_Caracter = 4) and (Id_Nivel = 1) and (Cod_DepartamentoPrograma = 86) => Clase=Bajo (11.0/2.0)
(Id_Caracter = 1) and (Id_Metodologia = 1) and (Id_Sector = 2) and (Cod_DepartamentoPrograma = 5) => Clase=Bajo (7.0/0.0)
(Id_Caracter = 4) and (Id_Metodologia = 3) and (Id_Sector = 1) and (Alta_Calidad = SI) and (Id_Nivel = 1) => Clase=Bajo (43.0/19.0)
(Id_Caracter = 4) and (Alta_Calidad = SI) and (Cod_DepartamentoPrograma = 85) => Clase=Bajo (28.0/11.0)
(Id_Metodologia = 1) and (Id_Nivel = 1) and (Id_Sector = 1) => Clase=MuyAlto (718.0/338.0)
(Cod_DepartamentoPrograma = 11) and (Id_Metodologia = 1) and (Id_Sector = 2) => Clase=MuyAlto (139.0/49.0)
(Id_Nivel = 1) and (Id_Metodologia = 1) and (Id_Caracter = 4) and (Cod_DepartamentoPrograma = 76) => Clase=MuyAlto (21.0/5.0)
(Id_Nivel = 1) and (Id_Metodologia = 1) and (Id_Caracter = 3) and (Cod_DepartamentoPrograma = 5) => Clase=MuyAlto (16.0/4.0)
(Id_Nivel = 1) and (Id_Metodologia = 1) and (Id_Caracter = 4) and (Cod_DepartamentoPrograma = 5) => Clase=MuyAlto (22.0/8.0)
(Id_Nivel = 2) and (Alta_Calidad = NO) and (Id_Sector = 1) and (Id_Metodologia = 2) => Clase=Medio (77.0/21.0)
(Id_Nivel = 2) and (Id_Sector = 1) and (Id_Caracter = 2) => Clase=Medio (83.0/31.0)
(Id_Nivel = 2) and (Alta_Calidad = NO) and (Id_Metodologia = 1) and (Cod_DepartamentoPrograma = 68) => Clase=Medio (26.0/11.0)
(Id_Sector = 1) and (Id_Nivel = 2) and (Cod_DepartamentoPrograma = 19) => Clase=Medio (20.0/8.0)
 => Clase=Alto (2865.0/2034.0)

Number of Rules : 22
Time taken to build model: 0.19 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        1722               37.8545 %
Incorrectly Classified Instances      2827               62.1455 %
Kappa statistic                          0.1705
Mean absolute error                      0.3425
Root mean squared error                  0.4163
Relative absolute error                 91.3319 %
Root relative squared error             96.1431 %
Total Number of Instances             4549

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,320    0,062    0,628      0,320    0,424      0,335    0,679     0,446     Bajo
                 0,177    0,092    0,393      0,177    0,245      0,117    0,606     0,344     Medio
                 0,687    0,587    0,282      0,687    0,400      0,089    0,571     0,284     Alto
                 0,328    0,089    0,553      0,328    0,411      0,291    0,690     0,436     MuyAlto
Weighted Avg.    0,379    0,208    0,463      0,379    0,370      0,207    0,636     0,377

=== Confusion Matrix ===

   a   b   c   d   <-- classified as
 359  86 608  69 |   a = Bajo
 134 203 733  74 |   b = Medio
  48 151 787 159 |   c = Alto
  31  76 658 373 |   d = MuyAlto
```

# Anexo VII. Modelado J48 variables individuales y socioeconómicas

```
=== Run information ===

Scheme:       weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     SpadiesSI-weka.filters.unsupervised.attribute
              .Remove-R2-weka.filters.unsupervised.attribute
Instances:    648                                  .Remove-R1
Attributes:   3
              Variable
              SubVariable
              Clase
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
------------------

SubVariable = 1: Bajo (10.0)
SubVariable = 2: Medio (10.0/1.0)
SubVariable = 3: Alto (10.0/1.0)
SubVariable = 4: MuyAlto (10.0/3.0)
SubVariable = 5: Alto (10.0/3.0)
SubVariable = 6: Alto (10.0/2.0)
SubVariable = 7: Medio (10.0/2.0)
SubVariable = 8: Medio (10.0/2.0)
SubVariable = 9: Bajo (10.0)
SubVariable = 10: Bajo (10.0)
SubVariable = 11: Bajo (10.0)
SubVariable = 12: Bajo (10.0)
SubVariable = 13: Bajo (8.0)
SubVariable = 15: Medio (10.0/1.0)
SubVariable = 26: Medio (10.0/2.0)
SubVariable = 16a20: MuyAlto (10.0)
SubVariable = 1a2: MuyAlto (10.0)
SubVariable = 21a25: Medio (10.0)
SubVariable = 2a3: Alto (10.0/1.0)
SubVariable = 3a5: Alto (10.0/2.0)
SubVariable = 5a7: Medio (10.0/2.0)
SubVariable = 7a10: Bajo (10.0/2.0)
SubVariable = ACCES: Medio (10.0/3.0)
SubVariable = C: Medio (10.0/3.0)
SubVariable = Carece: Alto (10.0/2.0)
SubVariable = OTN: Bajo (10.0)
SubVariable = cuatro: Medio (10.0)
SubVariable = dos: Alto (10.0/1.0)
SubVariable = E1: MuyAlto (10.0/2.0)
SubVariable = E2: MuyAlto (10.0)
SubVariable = E3: Alto (10.0/4.0)
SubVariable = E4: Medio (10.0/1.0)
SubVariable = E5: Bajo (10.0/4.0)
SubVariable = E6: Bajo (10.0)
SubVariable = H: MuyAlto (10.0)
SubVariable = LP: Bajo (10.0/1.0)
SubVariable = M: MuyAlto (10.0)
SubVariable = m1: Alto (10.0/2.0)
SubVariable = m10: Bajo (10.0/1.0)
SubVariable = M4: Medio (10.0/2.0)
SubVariable = MP: Bajo (10.0)
SubVariable = N: Bajo (10.0/2.0)
SubVariable = N1: MuyAlto (10.0/2.0)
SubVariable = N2: Alto (10.0/2.0)
SubVariable = N3: Medio (10.0)
SubVariable = No: Alto (10.0/2.0)
SubVariable = nor: MuyAlto (10.0)
SubVariable = Nr: MuyAlto (20.0)
SubVariable = Otro: Medio (10.0/4.0)
SubVariable = P: Alto (10.0/2.0)
SubVariable = Posee: Alto (10.0/5.0)
SubVariable = PQ: Bajo (10.0/1.0)
SubVariable = Primaria: MuyAlto (10.0/3.0)
SubVariable = Q: Bajo (10.0/1.0)
SubVariable = R
|   Variable = AIcetex: Medio (10.0/4.0)
|   Variable = AIES: MuyAlto (10.0)
SubVariable = S: Alto (10.0/3.0)
SubVariable = Secundaria: MuyAlto (10.0)
SubVariable = Si: MuyAlto (10.0)
SubVariable = syp: Alto (10.0/1.0)
SubVariable = T: Medio (10.0/1.0)
SubVariable = tot: Alto (10.0/1.0)
SubVariable = tres: Medio (10.0/4.0)
SubVariable = uno: Alto (10.0/3.0)

Number of Leaves  :     75

Size of the tree :      77
Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         549               84.7222 %
Incorrectly Classified Instances        99               15.2778 %
Kappa statistic                          0.7963
Mean absolute error                      0.1134
Root mean squared error                  0.2518
Relative absolute error                 30.2345 %
Root relative squared error             58.1516 %
Total Number of Instances              648

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               0,913     0,025     0,924       0,913   0,918      0,892    0,983     0,935     Bajo
               0,823     0,066     0,808       0,823   0,816      0,753    0,926     0,797     Medio
               0,790     0,082     0,762       0,790   0,776      0,699    0,923     0,725     Alto
               0,864     0,031     0,903       0,864   0,883      0,846    0,969     0,942     MuyAlto
Weighted Avg.  0,847     0,051     0,849       0,847   0,848      0,797    0,950     0,849

=== Confusion Matrix ===

   a   b   c   d   <-- classified as
 146  14   0   0 |   a = Bajo
  11 135  18   0 |   b = Medio
   1  18 128  15 |   c = Alto
   0   0  22 140 |   d = MuyAlto
```

# Anexo VIII. Modelado JRip variables individuales y socioeconómicas

```
=== Run information ===

Scheme:       weka.classifiers.rules.JRip -F 3 -N 2.0 -O 2 -S 1
Relation:     SpadiesSI-weka.filters.unsupervised.attribute.Remove-R2-weka.filters.unsupervised.attribute.Remove-R1
Instances:    648
Attributes:   3
              Variable
              SubVariable
              Clase
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

JRIP rules:
===========

(Variable = NPF) and (SubVariable = 1) => Clase=Bajo (10.0/0.0)
(Variable = NPF) and (SubVariable = 9) => Clase=Bajo (10.0/0.0)
(Variable = TC) and (SubVariable = MP) => Clase=Bajo (10.0/0.0)
(Variable = NPF) and (SubVariable = 10 => Clase=Bajo (10.0/0.0)
(SubVariable = 13) => Clase=Bajo (8.0/0.0)
(SubVariable = Q) => Clase=Bajo (10.0/1.0)
(SubVariable = 11) => Clase=Bajo (10.0/0.0)
(SubVariable = 12) => Clase=Bajo (10.0/0.0)
(SubVariable = E6) => Clase=Bajo (10.0/0.0)
(SubVariable = 7a10) => Clase=Bajo (10.0/2.0)
(SubVariable = PQ) => Clase=Bajo (10.0/1.0)
(SubVariable = m10) => Clase=Bajo (10.0/1.0)
(SubVariable = OTN) => Clase=Bajo (10.0/0.0)

Number of Rules : 42
Time taken to build model: 0.17 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         533               82.2531 %
Incorrectly Classified Instances       115               17.7469 %
Kappa statistic                          0.7633
Mean absolute error                      0.131
Root mean squared error                  0.266
Relative absolute error                 34.9268 %
Root relative squared error             61.4259 %
Total Number of Instances              648

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,913    0,025    0,924      0,913    0,918      0,892    0,973     0,919     Bajo
                 0,817    0,097    0,740      0,817    0,777      0,698    0,913     0,719     Medio
                 0,716    0,086    0,734      0,716    0,725      0,635    0,901     0,707     Alto
                 0,846    0,029    0,907      0,846    0,875      0,837    0,971     0,938     MuyAlto
Weighted Avg.    0,823    0,059    0,826      0,823    0,823      0,765    0,939     0,820

=== Confusion Matrix ===

   a   b   c   d   <-- classified as
 146  13   1   0 |   a = Bajo
  11 134  19   0 |   b = Medio
   1  31 116  14 |   c = Alto
   0   3  22 137 |   d = MuyAlto
```

```
(SubVariable = E5) => Clase=Bajo (10.0/4.0)
(SubVariable = N) => Clase=Bajo (10.0/2.0)
(SubVariable = LP) => Clase=Bajo (10.0/1.0)
(SubVariable = Nr) => Clase=MuyAlto (20.0/0.0)
(Variable = G) => Clase=MuyAlto (20.0/0.0)
(SubVariable = E2) => Clase=MuyAlto (10.0/0.0)
(SubVariable = Si) => Clase=MuyAlto (10.0/0.0)
(SubVariable = 16a20) => Clase=MuyAlto (10.0/0.0)
(Variable = AIES) => Clase=MuyAlto (10.0/0.0)
(SubVariable = Secundaria) => Clase=MuyAlto (10.0/0.0)
(SubVariable = N1) => Clase=MuyAlto (10.0/2.0)
(SubVariable = 1a2) => Clase=MuyAlto (10.0/0.0)
(SubVariable = nor) => Clase=MuyAlto (10.0/0.0)
(SubVariable = Primaria) => Clase=MuyAlto (10.0/3.0)
(SubVariable = E1) => Clase=MuyAlto (10.0/2.0)
(SubVariable = 4) => Clase=MuyAlto (10.0/3.0)
(Variable = NM) => Clase=Alto (20.0/2.0)
(Variable = IH) => Clase=Alto (40.0/15.0)
(SubVariable = dos) => Clase=Alto (10.0/1.0)
(Variable = V) => Clase=Alto (20.0/7.0)
(SubVariable = N2) => Clase=Alto (10.0/2.0)
(Variable = T) => Clase=Alto (10.0/2.0)
(SubVariable = 3) => Clase=Alto (10.0/1.0)
(SubVariable = 6) => Clase=Alto (10.0/2.0)
(SubVariable = uno) => Clase=Alto (10.0/3.0)
(SubVariable = P) => Clase=Alto (10.0/2.0)
(SubVariable = 5) => Clase=Alto (10.0/3.0)
(SubVariable = S) => Clase=Alto (10.0/3.0)
 => Clase=Medio (170.0/40.0)
```