

|  |
| --- |
| **Universidad Internacional de La Rioja**  **Escuela Superior de Ingeniería y Tecnología**  **Máster Universitario en Análisis y Visualización de Datos Masivos** |
|  |
| Desarrollo de Modelos Analíticos para medir atención a Jóvenes con sanciones penales |

**Trabajo Fin de Máster**

**Tipo de trabajo:** Piloto Experimental

**Presentado por:** Celis Pineda, Diana Paola

**Director/a:** Garay Gallastegui, Luis Miguel

**Resumen**

Desarrollo de modelos analíticos que permitan mediante aplicación de estadistica descriptiva y predictiva mejorar la atención al programa Atención de la sanción Semicerrado, Externado Media Jornada de Adolescentes y Jóvenes en Conflicto con la Ley Penal (Santander – Colombia), permitiendo detectar patrones y resolver las inquietudes que surgen en el equipo de trabajo de la corporación (Corpoadases) y mediante el uso de Machine Learning detectar el porcentaje de reincidencia, Cluster para definir los grupos de atención bases por el nivel de consumo y desarrollo de reglas que permitan a la corporación detectar los niveles de atención que requiere un joven por las áreas de terapia ocupacional y psicología,< con los resultados que se obtengan permitir a la corporación tomar decisiones y realizar planes de trabajo para garantizar la mejora continua en la calidad del servicio ofertado.

**Palabras Clave**: Inteligencia Artificial, Machine Learning, Reincidencia Delito, Nivel Consumo Drogas, Power BI

**Abstract**

Development of analytical models that allow, through the application of descriptive and predictive statistics, to improve the attention of the Semi-Closed Health Care Program, Teens and Youth part-time non-residential program related with Conflict with Criminal Law (Santander - Colombia), allowing to detect patterns and resolve concerns that arise in the work team of the corporation (Corpoadases) and through the use of Machine Learning to detect the recidivism percentage, Cluster to define the care groups based on the level of consumption of psychoactive substances and development of rules that allow the corporation to detect the levels of care required by a young person in the areas of occupational therapy and psychology, with the results obtained allow the corporation to make decisions and make work plans to ensure continuous improvement in the quality of service offered.

**Keywords:** Artificial intelligence, Machine Learning, Crime Recidivism, Drug Use Level, Power BI

**Índice de contenidos**

[1. Introducción 10](#_Toc95224373)

[1.1 Justificación 10](#_Toc95224374)

[1.2 Planteamiento del trabajo 11](#_Toc95224375)

[1.3 Estructura de la memoria 12](#_Toc95224376)

[2. Contexto y estado del arte 12](#_Toc95224377)

[2.1 La Frenología 12](#_Toc95224378)

[2.2 Aplicaciones Existentes 13](#_Toc95224379)

[2.3 Análisis de información para medir atención a Jóvenes con sanciones penales 14](#_Toc95224380)

[2.3.1 Necesidad del negocio 14](#_Toc95224381)

[2.3.1 Implementación Herramientas Tecnológicas 15](#_Toc95224382)

[2.3.1.1 Desarrollo ETL 15](#_Toc95224383)

[2.3.1.2 Análisis Información 16](#_Toc95224384)

[2.3.1.2.1 Random Forest 16](#_Toc95224385)

[2.3.1.2.2 Regresiones 16](#_Toc95224386)

[2.3.1.2.3 Clúster 17](#_Toc95224387)

[2.3.1.2.4 Lógica Difusa 17](#_Toc95224388)

[2.3.1.2.5 SVM 18](#_Toc95224389)

[2.3.1.3 Visualización de Datos 18](#_Toc95224390)

[2.3.1.3.1 Herramientas para la Visualizacion de Datos 19](#_Toc95224391)

[3. Objetivos concretos y metodología de trabajo 20](#_Toc95224392)

[3.1. Objetivo general 20](#_Toc95224393)

[3.2. Objetivos específicos 20](#_Toc95224394)

[3.3. Metodología del trabajo 21](#_Toc95224395)

[3.3.1 CRISP-DM 21](#_Toc95224396)

[3.3.2 Cronograma del Proyecto 24](#_Toc95224397)

[3.3.3 Conjunto de Datos Iniciales 24](#_Toc95224398)

[3.3.4 Procesamiento de los Datos 26](#_Toc95224399)

[3.3.5 DESARROLLO PREGUNTAS 37](#_Toc95224400)

[3.3.5.1 Cuáles son los motivos de Ingreso de adolescentes y jóvenes con sanciones penales. 37](#_Toc95224401)

[3.3.5.2 Cuál es la Edad Promedio de ingreso de Jóvenes con sanciones penales 39](#_Toc95224402)

[3.3.5.3 Cuál es la tipología Familiar de los Jóvenes con sanciones penales 40](#_Toc95224403)

[3.3.5.4 Nivel de Consumo de los Jóvenes con sanciones penales. 42](#_Toc95224404)

[3.3.5.5 Nivel de Escolaridad de los Jóvenes con sanciones penales. 43](#_Toc95224405)

[3.3.5.6 ¿Existe alguna relación entre tipología Familiar Vs Reincidencia en Delito? 45](#_Toc95224406)

[3.3.5.7 Está asociado el Nivel de consumo con la reincidencia en un delito 47](#_Toc95224407)

[3.3.5.8 Cuál es la probabilidad de reincidencia de un Joven que tiene sanciones penales. 48](#_Toc95224408)

[Definicion Variables 48](#_Toc95224409)

[Random Forest 48](#_Toc95224410)

[Regresión Logística 50](#_Toc95224411)

[SVM 52](#_Toc95224412)

[Comparación Métricas 53](#_Toc95224413)

[Prueba Modelo Random Forest (impmentancion) 54](#_Toc95224414)

[3.3.5.9 Clasificacion del tipo de tratamiento que requiere un Joven con sanciones penales por el nivel de consumo de drogas. 56](#_Toc95224415)

[Definicion Variables 56](#_Toc95224416)

[3.3.5.10 Definicion de áreas de refuerzo que requiera un joven que ha ingresado a la corporación. 60](#_Toc95224417)

[Prueba Lógica Difusa 63](#_Toc95224418)

[4. Desarrollo específico de la contribución 70](#_Toc95224419)

[5. Conclusiones y trabajo futuro 71](#_Toc95224420)

[5.1. Conclusiones 71](#_Toc95224421)

[5.2. Líneas de trabajo futuro 72](#_Toc95224422)

[6. Bibliografía 73](#_Toc95224423)

[Anexos 74](#_Toc95224424)

[Anexo I. Título anexo I 74](#_Toc95224425)

[Anexo II. Título anexo II 74](#_Toc95224426)

**Índice de tablas**

[Tabla 13 Variables de Salida. 22](#_Toc95224427)

[Tabla 1 Diccionario de Datos 26](#_Toc95224428)

[Tabla 2 Cargar Información 28](#_Toc95224429)

[Tabla 3 ETL Dimensiones 30](#_Toc95224430)

[Tabla 4 STG1\_HECHOS 31](#_Toc95224431)

[Tabla 5 STG2\_Hechos 31](#_Toc95224432)

[Tabla 6 STG3\_Hechos 32](#_Toc95224433)

[Tabla 7 STG4\_Hechos 33](#_Toc95224434)

[Tabla 8 STG5\_Hechos 34](#_Toc95224435)

[Tabla 9 STG6\_Hechos 34](#_Toc95224436)

[Tabla 10 STG7\_Hechos 35](#_Toc95224437)

[Tabla 11 Fact\_corp 36](#_Toc95224438)

[Tabla 12 Estructura Final tabla FACT\_CORP 37](#_Toc95224439)

[Tabla 14 Documentación Visualizacion Motivo Ingreso al programa de Responsabilidad Penal 38](#_Toc95224440)

[Tabla 15 Visualizacion Edad Promedio de ingreso histograma 39](#_Toc95224441)

[Tabla 16 Visualizacion Edad por Delito 40](#_Toc95224442)

[Tabla 17 Visualizacion Recuento por TipologiaFamiliar 41](#_Toc95224443)

[Tabla 18 Visualizacion Nivel de Consumo de Jóvenes con sanciones penales. 43](#_Toc95224444)

[Tabla 19 Relación tipología familia - Reincidencia en Delito 46](#_Toc95224445)

[Tabla 20 Relación Niv\_consumo- Reincidencia en Delito 47](#_Toc95224446)

[Tabla 21 Métricas Modelos 53](#_Toc95224447)

[Tabla 22 Prueba RF 1 54](#_Toc95224448)

[Tabla 23 Prueba RF 2 55](#_Toc95224449)

[Tabla 24 Clasificacion por Tipo de Tratamiento Grupo 1 57](#_Toc95224450)

[Tabla 25 Clasificacion por Tipo de Tratamiento Grupo 2 58](#_Toc95224451)

[Tabla 26 Clasificacion por Tipo de Tratamiento Grupo 3 59](#_Toc95224452)

[Tabla 27 Definicion Antecedente Variable Delito 61](#_Toc95224453)

[Tabla 28 Definicion Antecedente Variable Nivel Consumo 62](#_Toc95224454)

[Tabla 29 Definicion Valores Consecuentes 62](#_Toc95224455)

[Tabla 30 Definicion Reglas 62](#_Toc95224456)

[Tabla 31 Prueba 1 Lógica Difusa 63](#_Toc95224457)

[Tabla 32 Prueba 2 Lógica Difusa 63](#_Toc95224458)

[Tabla 33 Prueba 3 Lógica Difusa 64](#_Toc95224459)

[Tabla 34 Definicion Antecedente Variable Escolaridad 66](#_Toc95224460)

[Tabla 35 Definicion Antecedente Variable Ocupación 66](#_Toc95224461)

[Tabla 36 Definicion Valores Consecuentes 66](#_Toc95224462)

[Tabla 37 Definicion Reglas 67](#_Toc95224463)

[Tabla 38 Prueba 1 Lógica Difusa 67](#_Toc95224464)

[Tabla 40 Prueba 2 Lógica Difusa 68](#_Toc95224465)

[Tabla 42 Prueba 3 Lógica Difusa 69](#_Toc95224466)

**Índice de figuras**

[Ilustración 1 Cronograma de actividades 29](#_Toc95224467)

[Ilustración 2 Consentimiento tratamiento de datos 30](#_Toc95224468)

[Ilustración 3 Job Inicial. 32](#_Toc95224469)

[Ilustración 4 DWH Carga Información 32](#_Toc95224470)

[Ilustración 5 Validación Archivos 33](#_Toc95224471)

[Ilustración 6 ETL Validación Archivos 34](#_Toc95224472)

[Ilustración 7 ETL Dimensiones 35](#_Toc95224473)

[Ilustración 8 ETL STG1\_HECHOS 36](#_Toc95224474)

[Ilustración 9 ETL STG2\_HECHOS 37](#_Toc95224475)

[Ilustración 10 ETL STG3\_HECHOS 37](#_Toc95224476)

[Ilustración 11 ETL STG4\_HECHOS 38](#_Toc95224477)

[Ilustración 12 ETL STG5\_HECHOS 39](#_Toc95224478)

[Ilustración 13 ETL STG6\_HECHOS 40](#_Toc95224479)

[Ilustración 14 ETL STG7\_HECHOS 40](#_Toc95224480)

[Ilustración 15 ETL FACT\_CORP 41](#_Toc95224481)

[Ilustración 16 Visualizacion Motivo de Ingreso al Programa de Responsabilidad Penal 43](#_Toc95224482)

[Ilustración 17 Edad Promedio de ingreso de Jóvenes con sanciones penales 44](#_Toc95224483)

[Ilustración 18 Tipología Familiar de los Jóvenes con sanciones penales 46](#_Toc95224484)

[Ilustración 19 Nivel de Consumo de los Jóvenes con sanciones penales 47](#_Toc95224485)

[lustración 20 Nivel de Escolaridad de los Jóvenes Con sanciones penales 49](#_Toc95224486)

[lustración 21 Relación Tipología Familiar vrs Reincidencia en Delito 50](#_Toc95224487)

[Ilustración 22 Relación Nivel de Consumo vrs Reincidencia de Delito 52](#_Toc95224488)

[Ilustración 23 Potencial Predictivo Random Forest 54](#_Toc95224489)

[Ilustración 24 Matriz de Confusion Random Forest 54](#_Toc95224490)

[Ilustración 25 Entrenamiento Modelo Random Forest 55](#_Toc95224491)

[Ilustración 26 Potencial Predictivo Regresión Logística 55](#_Toc95224492)

[Ilustración 27 Matriz de Confusion Regresión Logística 56](#_Toc95224493)

[Ilustración 28 Entrenamiento Modelo Regresión Logística 56](#_Toc95224494)

[Ilustración 29 Potencial Predictivo SVM 57](#_Toc95224495)

[Ilustración 30 Comparación Modelos Predictivos SVM 57](#_Toc95224496)

[Ilustración 31 Entrenamiento SVM 58](#_Toc95224497)

[Ilustración 32 Probabilidad de Reincidencia 59](#_Toc95224498)

[Ilustración 33 Probabilidad de Reincidencia II 60](#_Toc95224499)

[Ilustración 34 Clasificacion por Tipo de Tratamiento Grupo 1 62](#_Toc95224500)

[Ilustración 35 Clasificacion por Tipo de Tratamiento Grupo 2 63](#_Toc95224501)

[Ilustración 36 Clasificacion por Tipo de Tratamiento Grupo 3 64](#_Toc95224502)

[Ilustración 37 Porcentaje Atención Psicológica Baja 68](#_Toc95224503)

[Ilustración 38 Porcentaje Atención Psicológica Media 69](#_Toc95224504)

[Ilustración 39 Porcentaje Atención Psicológica Prioritaria 69](#_Toc95224505)

[Ilustración 40 Prueba 1 Lógica Difusa TO Bajo 73](#_Toc95224506)

[Ilustración 41 Prueba 2 Lógica Difusa TO Medio 73](#_Toc95224507)

[Ilustración 42 Prueba 3 Lógica Difusa TO Medio 74](#_Toc95224508)

# Introducción

El siguiente TFM (Trabajo Fin de Máster) pretende implementar modelos descriptivos, predictivos y la generación de visualizaciones que permitan medir la atención en los adolescentes y jóvenes que ingresan al programa "Atención de la sanción Semicerrado, Externado Media Jornada de Adolescentes y Jóvenes en Conflicto con la Ley Penal (Santander - Colombia)”.

Este proyectó se realizará como un plan Piloto en la Corporación “Alianza para el Desarrollo Ambiental Social y Económico Sostenible” (Corpoadases).

Con los resultados que se obtengan en el análisis de la información se podrán desarrollar estrategias entre las áreas transversales, (Especialista de Área (profesional que lidera proyecto de vida), Psicología, Nutrición, Trabajo Social, y Formadores (académico, Artes, Cultura y Deporte) permitiendo agregar a la institución valor al trabajo que realizan con los jóvenes y adolescentes mejorando así la calidad del servicio.

## Justificación

Actualmente la Corporación no tiene como medir el efecto en la atención en los Adolescentes y Jóvenes que ingresan al programa "Atención de la sanción Semicerrado, Externado media Jornada de Adolescentes y Jóvenes en Conflicto con la Ley Penal (Santander – Colombia)”, por qué la información no la tienen almacenada en un repositorio de datos, esta se encuentra en archivos de Excel y documentos físicos que están expuestos a daños y a pesar que se deben entregar reportes a los entes de control colombianos (ICBF Instituto colombiano del Bienestar Familiar – Procuraduría, Contraloría, Personería) para sus informes mensuales, trimestrales y anuales, toma tiempo y sobreesfuerzos en el equipo de trabajo para recompilar la información total de un Adolescente o Joven que se encuentran tratando.

La corporación quiere impactar en la vida de los adolescentes y jóvenes que ingresan al programa y con este proyecto además de almacenar la información en un repositorio de datos busca mediante el análisis de los datos un insumo que permita detectar puntos de mejora donde pueda definir: que paso, como paso, porque paso y que podría pasar con los adolescentes y jóvenes que se ingresa al programa.

El resultado del análisis de la información permitirá diseñar e implementación estrategias en la atención individual y grupal de los adolescentes y jóvenes que ingresan a la corporación, para que estos cambien de estilo de vida y puedan tener mejores oportunidades, disminuyendo de esta manera las posibilidades de reincidencia en el delito; también, Corpoadases quiere destacarse como una institución que se preocupa y ataca los problemas que pueda tener un Adolescente o Joven.

## 1.2 Planteamiento del trabajo

Lo que se quiere es aprender de los datos, por eso se tomara la información histórica del programa para procesarla y mediante modelos de datos estructurados e implementando técnicas para análisis de datos y así poder describir, predecir y generar visualizaciones que respondan las siguientes preguntas:

- Cuál es el Motivo de Ingreso de mayor prevalencia en los Jóvenes con sanciones penales.

- Cuál es la Edad promedio de ingreso de Jóvenes con sanciones penales

- Cuál es la tipología Familiar de los Jóvenes con sanciones penales

- Nivel de Consumo de los Jóvenes con sanciones penales.

- Nivel de Escolaridad de los Jóvenes con sanciones penales.

- ¿Existe alguna relación entre tipología Familiar Vs Reincidencia en Delito?

- Está asociado el Nivel de consumo con la reincidencia en un delito

- Cual es la probabilidad de reincidencia de un Joven que tiene sanciones penales.

- Clasificacion del tipo de tratamiento que requiere un Joven con sanciones penales por el nivel de consumo de drogas.

- Definicion de áreas de refuerzo que requiera un joven que ha ingresado a la corporación.

¿Qué es lo que se propone?

**Mediante el análisis de la información y el diseño de los modelos la Corporación podrá:**

Detectar el mayor motivo de ingreso al programa de Responsabilidad penal.

Conocer la tipología familiar de los Adolescentes y Jóvenes que ingresan al programa.

Detectar el nivel de consumo de los adolescentes y jóvenes que ingresan al programa.

Conocer el nivel de escolaridad.

Evaluar el nivel de escolaridad de los adolescentes y jóvenes que ingresan egresan del programa.

Detectar si un adolescente o joven puede reincidir en un delito.

Definir áreas de refuerzo que requiera un joven que ha ingresado a la corporación, para reestructurar el plan de atención ofertado a cada joven y con ello garantizar la mejora continua en la calidad del servicio ofertado.

## 1.3 Estructura de la memoria

El resto de la Memoria se estructura de la siguiente manera:

En el capítulo 2 Se sitúa el lector en el contexto y estado del arte que se ha tomado como base para el desarrollo de este proyecto el cual abarca temas técnicos relacionados con la definición de análisis de datos, técnicas de machine Learning, proyectos y aplicaciones basadas en IA para la prevención y predicciones de delitos y consumo de drogas, además una definición de visualización de datos y las herramientas más populares en el mercado.

En el capítulo 3 se abarca los Objetivos generales y específicos del alcance del proyecto y la metodología del trabajo usada para el desarrollo del proyecto.

En el Capítulo 4 se describe en detalle el análisis y planeación la definición de la necesidad de la corporación, análisis del negocio, las técnicas para el tratamiento de la información, limpieza de datos, desarrollo de los algoritmos y las visualizaciones.

En el Capítulo 5 Conclusiones y trabajo futuro se medirá si el proyecto Responde a la totalidad de las preguntas planteadas, se puedan generar visualización de información en Línea y si los modelos de predicción tienen una precisión del 70%, además se dejará planteado posibles líneas de trabajo a futuro para complementar el proyecto.

En el Capítulo 6 Bibliografía se listan las diferentes fuentes tomadas como referencia para el desarrollo de este proyecto.

Finalizado se encuentra el apartado de anexos donde se referencia anexos e información adicional el proyecto.

# 2. Contexto y estado del arte

## 2.1 La Frenología

Descifrar el comportamiento de una persona ha sido tema de trabajo en el trascurso del tiempo, de hecho existe la pseudociencia conocida como: La Frenología, que se conocía como la doctrina del cerebro la cual fue desarrollada por Franz Joseph Gall en el siglo XVIII que buscaban mediante la forma del cráneo descifrar como era una persona y sus comportamientos[1], basado en esto Cesare Lombroso que se considera el padre de la Criminología desarrolla teorías para definir los tipos de delincuentes que de acuerdo a su fisonomía, genero, entre otros factores, permitiera clasificar a una persona en los 6 tipos de criminales (Criminal nato, Delincuente loco moral, delincuente epiléptico, delincuente pasional, delincuente loco, delincuente ocasional ). [2]

Sin embargo, la frenología ha dejado de ser una disciplina de estudio pues para el desarrollo de las teorías no se usó un método científico, cuando se realizó pruebas experimentales se validaron que algunas de sus afirmaciones no podrían ser respaldadas ya que se basaba en observaciones y en evidencias anecdótica. [3]

En la actualidad se ha pedido no realizar publicaciones sobre algoritmos que mediante el uso de reconocimiento facial pueda detectar si una persona es criminal por la forma de su cara,[4] pues se consideran que llegan a ser racistas y generar injusticias sociales.[5].

## 2.2 Aplicaciones Existentes

Actualmente Existen aplicaciones que usan análisis de datos para predecir crímenes, consumo de drogas, reincidencia de un delito, a continuación, se mencionaran las más relevantes:

**Key Crime, “**la tecnología incursiona en la lucha contra los delitos en serie” [6].

Es una aplicación que mediante la recopilación de información de los delitos que se han cometido permite a la policía de Milán (Italia) realizar perfiles de criminales y mediante el cruce de información implementar modelos para predecir, prevenir y combatir delitos.

“Este sistema de “predicción delictiva”, ha sido replicado a otros países, se inspiró en un modelo utilizado para los análisis de riesgo sísmico, el cual pretende predecir réplicas de temblores. En el ámbito criminal funciona igual, pues analiza las conductas reincidentes del delincuente” [6]

Este modelo sirvió de inspiración para “**El PredPol** (Predict Crime in Real Time) creado en el año 2011 por George Mohler. “Analiza lugares y horarios de crímenes cometidos recientemente y los combina con información sociológica y con datos de los patrones de comportamiento criminal, para prevenir delitos cometidos en tiempo real”

Este aplicativo tienen detractores quienes indican que el uso de estas tecnologías podría ocasionar “detenciones arbitrarias” [6].

Basado en esto en la ciudad de Bogotá (Colombia) existe un proyecto que busca predecir los crímenes con inteligencia artificial, “Según indicadores de la Secretaría de Seguridad, Convivencia y Justicia, quienes viven en Bogotá se sienten inseguros y esta percepción puede estar influenciada por los 14 hurtos que se producen cada hora, los más de 60 casos de lesiones personales denunciados y los 2,7 homicidios registrados a diario. “[7]**.**

Inspirados en las aplicaciones Key Crime y El PredPol, “Un equipo de matemáticos de la Universidad Nacional desarrolla un modelo de predicción de crímenes en Bogotá. Un sistema pionero en América Latina que impulsará la formación de talento humano en este campo.” [8]

Se recompilará la información que suministra la Policía Nacional, la Fiscalía, Línea de atención de emergencias (123), cámaras de vigilancia y buscará crear patrones para diseñar modelos predictivos “no ira detrás de las personas” [8].

La buena calidad y correcto procesamiento de la información será el éxito de este aplicativo, evitar sesgos es el mayor reto, pues en las aplicaciones Key Crime y El PredPol el uso de la variable Raza incidió en los resultados del análisis de la información por tal motivo se acusó falsamente a delincuentes de raza negra con mayor probabilidad de reincidir en un delito. [8].

EL uso de técnicas de IA (inteligencia Artificial) también se está implementando para atacar otros problemas de salud pública como lo es el consumo de drogas para esto en Ecuador se está desarrollando “algoritmos emergentes de inteligencia artificial para la predicción del consumo problemático y controlado de drogas a partir de información sociodemográfica, psicosocial y de salud; y la creación de un “observatorio para el estudio del alcohol y otras drogas” que garantice el acceso a información actualizada y la difusión periódica de los resultados” esto para desarrollar programas preventivos y generar políticas públicas para el consumo de drogas. [9]

## 2.3 Análisis de información para medir atención a Jóvenes con sanciones penales

## 2.3.1 Necesidad del negocio

Actualmente cuando un joven ingresa a la corporación es remitido por el juzgado y dependiendo del grado del delito se define el tiempo de sanción.

Al ingresar a la corporación se les realiza: revisiones médicas, nutricionales, psicológicas, atención en terapia ocupacional y social para determinar su estado actual.

Todos los jóvenes participan en las mismas actividades grupales, seguimiento individual y Familiar, a medida que avanza su sanción van determinando los planes de tratamiento, sin embargo, la corporación quiere desde el inicio tener claro el plan de trabajo por cada adolescente y Joven que ingresan para: determinar el refuerzo en las áreas de atención prioritarias, identificar por su consumo cual debería ser el grupo de orientación en el plan de atención, detectar una posible reincidencia, tomando las medidas preventivas en el tratamiento del joven, garantizando una mejora en la atención de los jóvenes con sanciones penales.

En la ciudad de Bucaramanga (Santander – Colombia), las instituciones que trabajan con Jóvenes Infractores no cuentan con herramientas tecnológicas para el análisis de su información por tal motivo este Piloto permitirá a Corpoadases convertirse en una institución pionera en uso de análisis de datos para medir el servicio en la atención de los Jóvenes y Adolescentes que ingresan al programa de "Atención de la sanción semicerrado, Externado media Jornada de Jóvenes en conflicto con la Ley Penal.

No se busca Juzgar a los jóvenes, lo que se quiere es analizar la información y con los resultados cuantificar las relaciones entre cada una de las variables y entender comportamientos de los jóvenes implementando con modelos preventivos y buscar patrones que ayuden a determinar si un joven puede reincidir en delitos, clasificar por los niveles de consumo el tipo de atención que requiere y ayudar a detectar las áreas de refuerzo.

Para evitar caer en algún tipo de sesgo social como ha sucedido en las aplicaciones mencionadas anteriormente se usarán en los algoritmos predictivos variables que describen características psicosociales descartando variables como: etnia y raza para el entrenamiento de los modelos.

## 2.3.1 Implementación Herramientas Tecnológicas

## 2.3.1.1 Desarrollo ETL

La Corporación suministrara la data mediante la carga histórica de archivos en formato Excel y Se cargara en base de datos mediante procesos ETL.

**ETL: “**Como sus siglas indican, consiste en la extracción, transformación y carga de los datos en el **datawarehouse** (almacén de datos), de modo que se puede afirmar que es una parte fundamental de este. Antes de guardar los datos, deben ser transformados, limpiados, filtrados y redefinidos.”[9]

Para este TFM se desarrolla una Etl con la herramienta Pentaho para la carga de información suministrada por la corporación, con relación a la limpieza de datos se diseñan tablas de dimensiones que permitirán comparar cada variable y con esto se garantiza que los datos sean homogéneos.

La información procesada se insertará en una tabla Core llamada Fact\_corp y será el insumo para el desarrollo de las visualizaciones y los modelos predictivos.

El motor de base de datos es postgrest ya que permite el trabajo de datos masivos, se puede instalar en cualquier plataforma ilimitadamente y de manera gratuita adicional se puede implementar funciones de seguridad para resguardar la información de los adolescentes y jóvenes que ingresan a la corporación además les permite tener un Backup de la información en un repositorio de datos pues actualmente manejan archivos fiscos y Hoja de Excel que se encuentran vulnerables a posibles daños o pérdidas.

## 2.3.1.2 Análisis Información

Procesada ya la información se realizará **análisis descriptivo** de los datos y basados en las preguntas realizadas por el equipo de trabajo de la corporación se usarán algunos modelos de Aprendizaje supervisado y no supervisados como:

## 2.3.1.2.1 Random Forest

“El bosque aleatorio es un algoritmo de aprendizaje supervisado que se utiliza tanto para la clasificacion y regresión. Sin embargo, se utiliza principalmente para cuestiones de clasificacion. Como sabemos, un bosque está formado por árboles y más arboles significa un bosque más robusto. Asimismo, el algoritmo de bosque aleatorio crea arboles de decisión sobre muestras de data, luego obtiene la predicción de cada uno de ellos y finalmente selecciona la mejor solución mediante votación. Es un método de conjunto que es mejor que un árbol de decisión único porque reduce el sobreajuste al promediar el resultado".[10]

## 2.3.1.2.2 Regresiones

“Dos variables están relacionadas de una forma determinista, es decir, dado un valor de una variable, el valor de la otra variable se determina automáticamente sin error.

Mediante las regresiones se podrá no solo predecir el valor de una variable, también encontrar relación entre X (llamada variable explicativa, variable de predicción o variable independiente) y (llamada variable respuesta o variable dependiente). “[11]

para este TFM usáremos regresiones para validar si existe relación entre las variables que se han recolectado con la información brindada por la corporación

Hay diferentes tipos de modelos para hacer regresiones:

a. Regresión lineal simple: Variable dependiente continúa e independiente continúa.

b. Regresión lineal múltiple: Variable dependiente continua y variables múltiples independientes.

c. Regresión logística nominal: Variable dependiente binaria, independientes continuas o categóricas.

d. Regresión logística multinomial: Variable dependiente categórica, independientes continuas o categóricas.

e. Regresión logística ordinal: Variable dependiente categórica ordinal, independientes continuas o categóricas.

f. Regresión de poisson: Variable dependiente discreta, independientes continuas o categóricas.

g. Regresión de cox: Variable dependiente nominal y el tiempo como independiente

Dependiendo de los tipos de datos se podrá aplicar algunos de los modelos de regresión mencionados.

## 2.3.1.2.3 SVM Maquina de Vectores

Conjunto de algoritmos de aprendizaje supervisado utilizados generalmente para resolver problemas de clasificacion y regresión.

Una SVM representa los puntos de una muestra en el espacio, y separa las clases en dos espacios lo más amplios posibles, mediante un hiperplano de separación que se define por los 2 puntos de las 2 clases más cercanas, y que se llama vector soporte. Cuando utilizamos una nueva muestra sobre este modelo, podemos saber a qué clase pertenece, según si queda clasificada en un parte u en otra del hiperplano. [19]

Para la realización de este modelo es impórtate definir las variables con las cuales se va a entrenar el modelo pues no todas las variables que se creen puedan inferir en la reincidencia del delito lo son realmente.

Se han seleccionado estos tres algoritmos para realizar una comparación con el entrenamiento de los modelos analíticos a modo de clasificacion, con esto se espera definir cual podría ayudar a predecir más casos de reincidencia en un delito.

Se evaluará la matriz de confusión y el potencial predictivo que tenga cada modelo para seleccionar el mejor.

## 2.3.1.2.4 Clúster

“El clustering es uno de los métodos de aprendizaje no supervisado más importantes y, como cualquier otro método de aprendizaje no supervisado, busca caracterizar conceptos desconocidos a partir de instancias de estos. En este tipo de problemas de aprendizaje no supervisado la clase es desconocida y es, precisamente el descubrimiento de esta clase, el objetivo, a través de la agrupación de instancias en base a un esquema de similitud”. [12]

Esta técnica se aplicará en el proyecto ya que permitirá agrupar a los jóvenes e identificar patrones que ayudaren a definir los grupos para los tipos de atención que requiere un joven dependiendo el nivel de consumo, edad y el tipo de tratamiento que se impuesto por el área de psicología.

## 2.3.1.2.5 Lógica Difusa

“La lógica difusa o borrosa, también conocida como lógica fuzzy, es una lógica que nos permite procesar datos inciertos. Al contrario que la lógica booleana que sigue el principio bivalente o binario en el que la lógica solo permite dos estados posibles (verdadero y falso), la lógica difusa consigue tener infinitas degradaciones entre el valor verdadero y falso.” [13]

Este algoritmo se aplicará con el objetivo de detectar las áreas de refuerzo de los jóvenes en terapia ocupacional y psicología, se ha conversado con el grupo de trabajo de cada área y han definido reglas para indicar porcentaje de atención que requiere los adolescente y jóvenes cuando ingresan a la corporacion.

## 2.3.1.3 Visualización de Datos

“La visualización de datos es la presentación de datos en un formato pictórico o gráfico. Permite a los tomadores de decisiones ver los análisis presentados visualmente, para que puedan captar conceptos difíciles o identificar nuevos patrones.” [14]

Mediante gráficos se permitirá a las áreas trasversales de la corporación entender mejor los resultados del análisis de información además de tomar las mejores decisiones para la prestación del servicio a los jóvenes con sanciones penales que ingresan a la corporación.

Para el desarrollo de este TFM se tomara la información que se procesara en la Etl y se generaran mediante consultas de base de datos, el procesamiento de información mediante Python para finalmente generar archivos planos con la extensión .csv que se cargaran en Power Bi y permitirá la realización de las visualizaciones con el fin de responder las preguntas que se han planteado el equipo de trabajo y que se espera resolver permitiendo que visualmente se puedan encontrar comportamientos y patrones para definir por parte del equipo de trabajo estrategias para mejorar la calidad del servicio que prestan.

Aprovechar la integración de Power BI con Python para cargar los modelos de machine learning y así lograr visualizar los resultados de estos modelos en tiempo real y que permitan a la corporación en el momento de realizar un ingreso a un adolescente y joven definir patrones y comportamientos para adelantase en la atención individual.

## 2.3.1.3.1 Herramientas para la Visualizacion de Datos

Las 8 herramientas de visualización de datos más populares son:

1. Tableu
2. Qlik
3. Ploty
4. CartoDB
5. Power BI
6. DataWrapper
7. DataStudio
8. Power BI

Para este piloto la Herramienta seleccionada es Power BI.

“Microsoft Power BI, [o por sus siglas en inglés “Business Intelligence”](https://inbest.solutions/que-es-power-bi-de-office-365/), es un servicio de análisis de negocio basado en la nube que proporciona una vista única de los datos más críticos del negocio. [15]

Power BI es la herramienta de visualización que se usara para este TFM ya que ofrece las siguientes ventajas:

La Corporación tiene afiliación con Microsoft 365 por lo cual podría contar ya con la versión de escritorio de esta herramienta lo que significa un ahorro en costo, aunque la compra de su licencia PRO se realiza por usuarios el precio es mucho más económico que otras herramientas de visualización que se encuentra en el mercado.

El equipo de trabajo de la corporación no son personas técnicas por lo cual es una herramienta que es fácil de usar y se integra con Excel, en línea existe documentación útil, pero si se quiere explorar más los usos de Power Bi se recomienda realizar capacitación al equipo de trabajo para que puedan sacar el mejor provecho de esta herramienta,

Es colaborativa por lo cual se podrá tener la información actualizada en tiempo real, en el caso de que ingrese nuevos adolescentes y jóvenes realizando el cargue de la información por la Etl se podrá actualizar los gráficos y modelos que se han implementado.

# 3. Objetivos concretos y metodología de trabajo

## 3.1. Objetivo general

Desarrollar visualizaciones, modelos descriptivos y Predictivos que permitan medir la atención de los Jóvenes y Adolescentes que ingresan al programa de "Atención de la sanción semicerrado, Externado media Jornada de Jóvenes en conflicto con la Ley Penal (Santander - Colombia)”.

## 3.2. Objetivos específicos

- Recopilar la información histórica de la “Corporación (Alianza para el Desarrollo Ambiental Social y Económico Sostenible)” (Corpoadases) para ser disponibilizada en el insumo de datos para los modelos.

- Explicar mediante técnicas de Análisis de datos y Generación de visualizaciones los comportamientos de las variables recolectadas y sus características que permitan detectar tendencias.

- Implementar modelos que permita predecir de acuerdo con las variables situación sociofamiliar, nivel de estudio, consumo de drogas si un Joven puede reincidir en un delito.

- Determinar el nivel de atención Psicológica y Terapia Ocupacional que requiera un Joven dependiendo de las variables nivel de estudio,

ocupación, consumo de drogas y delito.

- Agrupar por el nivel de consumo a los jóvenes que ingresan a la corporación para determinar los tratamientos que deben ser suministrados por las entidades promotoras de salud.

## 3.3. Metodología del trabajo

## 3.3.1 CRISP-DM

Cross-Industry Standard Process for Data Mining (CRISP-DM), es una metodología para dar organización a los proyectos analíticos, consta de los siguientes 6 pasos los cuales se espera que el desarrollo del proyecto sea efectivo:

**Comprensión del Negocio**

En esta fase se exploran las necesidades que tienen en la corporación, recopilando la información sobre la situación actual y como se puede disponibilizar para el desarrollo de los modelos de datos, permitiendo establecer los objetivos y el alcance para este Piloto.

**Comprensión de datos**

Durante esta fase se le solicita a la corporación la información que tengan del programa Atención de la sanción semicerrado, Externado media Jornada de Jóvenes en conflicto con la Ley Penal (Santander - Colombia)”, la información entregada corresponde a archivos de Excel, para tratar los datos se cuenta con el vbo del director.

****

Ilustración 2 Consentimiento tratamiento de datos

No se aplicarán técnicas de anonimización y disociaciones de los datos personales ya que la información que se recompile quedara disponible para que el equipo de trabajo pueda generar informes, sin embargo, para los entrenamientos de los modelos no se usaran variables como el Nombre, Historia y Numero Documento de los jóvenes.

Analizado los datos más relevantes, se define las tablas de dimensiones y Hechos para la creación del DWH(DataWareHouse) para que la información quede disponibilizada en los modelos analíticos.

**Preparación de datos**

Para el desarrollo de esta fase se desarrollará un Etl en la herramienta Pentaho y base de datos postgrest la cual permitirá: limpiar los datos, cargarlos dentro de la base de datos para la carga de la información.

Durante esta fase se define la Etl llamada Inicio Malla que contiene 2 pasos que son:

**DWH**

Contiene todos los Job requeridos para el procesamiento de los datos desde la carga del archivo Excel hasta llevarlos a la tabla fact\_corp que será el repositorio oficial de la corporación para el insumo de los procesos de modelado y además disponibilizar los datos Validando cada una de las variables con tablas dimensiones que permiten validar y limpiar registros descartando los que no se encuentre dentro de estas dimensiones en tabla de log el cual el equipo de trabajo podrá revisar y corregir y de ser necesario volver a reprocesar para que los datos no se queden en log,

**Generación Modelos**

En ese paso se desarrollarán los insumos requeridos y la generación de los archivos en la extensión .csv que contiene la información requerida para el desarrollo de las visualizaciones y el entrenamiento de los modelos de lógica Difusa, Random Forest y el Clúster que se cargaran en Power BI.

**Modelado**

En estaFase se seleccionarán y aplicarán las técnicas de modelado mencionados (métodos estadísticos, aplicación de técnicas de Machine Learning) para resolver las preguntas planteadas por el equipo de trabajo.

Se identifican las siguientes variables como salidas para los modelos:

modelos:

|  |  |
| --- | --- |
| TIPO\_TRATAMIENTO | Text |
| POR\_ATEN\_TOCU | Int |
| POR\_ATEN\_PSICO | Int |
| REIN\_DELITO | Int |

Tabla 13 Variables de Salida.

**TIPO\_TRATAMIENTO**

Se define un modelo para realizar la clasificacion del tipo de tratamiento que requiere el joven basados por el nivel de consumo.

Su clasificacion es la siguiente:

* Consumo abusivo - requiere tratamiento intramural
* Consumo disfuncional - tratamiento ambulatorio
* Consumo experimental/social - citas de seguimiento con psicología
* No consume - participación en actividades de prevención
* Abstinencia - participación en grupo de apoyo

**REIN\_DELITO**

Este modelo se desarrollará para clasificar si un joven reincidirá en un delito basados en las siguientes variables: delito, tipo familia, nivel consumo, escolaridad aprobada y Ocupación.

Las salidas de modelo será una clasificacion:

* 1: Si
* 0: No

**AREAS DE REFUERZO**

Se requiere desarrollar un modelo el cual se pueda por Joven definir de acuerdo con las variables: Delito, Programa, Edad, Escolaridad Aprobada, Escolaridad Actual, Ocupación, Tiempo Sanción, Tipo Familia, Nivel de consumo los porcentajes de atención que requiera de cada una de las áreas de trabajo (esto basado en las reglas que definen en el equipo de trabajo de la corporación):

**Terapia Ocupacional**

**Psicología**

**Evaluación**

Durante esta etapaya se han construido los modelos por lo cual nos va a permitir evaluar los resultados revisando de los datos para trabajos futuros.

**Despliegue**

En esta fase se entregarán los resultados del análisis de datos mediante la generación de las visualizaciones en POWER BI.

La ejecución de la Etl se realizará mediante una .BAT llamado EjecutarEtls.bat.

El cual procesa los archivos y los carga en la base de datos (DWH) generando el insumo para los modelos y las visualizaciones en Power BI.

## 3.3.2 Cronograma del Proyecto

Se realizaron sesiones de trabajo con el equipo de trabajo de la corporación proyectando el siguiente cronograma de actividades.

Tabla

Descripción generada automáticamente

Ilustración 1 Cronograma de actividades

## 3.3.3 Conjunto de Datos Iniciales

A continuación, se describe cada uno los campos del archivo de Excel BASE:

|  |  |
| --- | --- |
| CAMPO | DEFINICIÓN |
| Nombre | Nombre del joven |
| Edad | Edad con la que el Joven ingresa al programa. |
| Historia | Número que el ICBF[[1]](#footnote-1) da al caso del Joven. |
| Defensor | Nombre defensor del Joven. |
| Equipo Psicosocial | Equipo asignado por el ICBF para la atención para el caso del Joven. |
| Juzgado | Nombre del Juzgado donde está el caso del Joven. |
| Asistente Social | Profesional que delega el juzgado para hacerle seguimiento a los casos de los chicos sancionados. |
| Delito | Nombre del delito con el que el Joven ingresa al programa. |
| Tiempo de sanción | Tiempo de sanción del Joven de acuerdo con el delito. |
| Documento | Numero de documento del Joven. |
| Ocupación | Ocupación actual del Joven. |
| Escolaridad Aprobada | Nivel escolaridad con la que ingresa el Joven. |
| Colegio | Nombre institución educativa. |
| Barrio | Nombre del barrio donde se encuentra el domicilio del Joven. |
| Sector | Nombre del Sector donde se encuentra el domicilio del Joven. |
| Jornada Programa | Jornada en la que el joven debe asistir a la corporación. (Mañana y/o Tarde) |
| Fecha Ingreso | Fecha de ingreso al programa. |
| Dirección | Dirección del acudiente. |
| Teléfono | Teléfono del acudiente |
| Responsable | Nombre del acudiente. |
| Numero Documento | Numero de documento del acudiente |
| Salud | EPS en la que se encuentra afiliado el joven. |
| Tutorías |  |
| Empleo | Indicador si el joven trabaja o no trabaja. |

Tabla 1 Diccionario de Datos

## 3.3.4 Procesamiento de los Datos

Para el desarrollo del procesamiento de la información se realizará usando el programa PENTAHO y para el almacenamiento el motor de base de datos POSTGRESQL.

**Inicio:** Job Inicial con el que arranca el procesamiento de información el cual Tiene como primer paso el arranque del proceso Star, continuando con el procesamiento de la información y finalmente la generación de los modelos Analíticos.

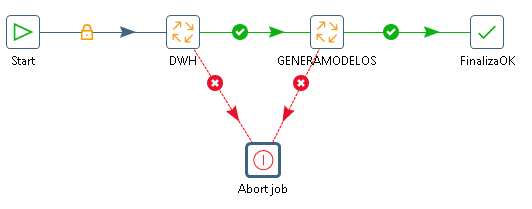


Ilustración 3 Job Inicial.

**TFM\_DWH:** Este paso se incluye el procesamiento de la información, se inicia con la validación de los archivos, continuando con la carga de las dimensiones y continua con la limpieza de datos.

Gráfico

Descripción generada automáticamente

Ilustración 4 DWH Carga Información

**Validación Archivos:** En este paso se realiza la validación del archivo de Excel entregado por la corporación.

Se busca un archivo con el Nombre BASE\_[enumerado].

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Ilustración 5 Validación Archivos

**Cargar Información**

|  |  |
| --- | --- |
| **Proceso:** CargaInformacion.ktr | |
| **Origen** | Archivo Excel Base\_[Enumerado] |
| **Destino** | Public.COR\_IN\_BASE  Public.log\_estrcuturaarchivos |
| **Descripción** | Carga la información del archivo plano Base validando lo siguiente: **Numero Documento que sea numérico.**  **Tipo de sanción contenga los valores Meses, Días.**  Si alguno de los dos (2) campos no cumple con las reglas se enviará la tabla log\_estrcuturaarchivos, de lo contrario pasaran los registros en la tabla **Public. COR\_IN\_BASE.** |

Tabla 2 Cargar Información

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Ilustración 6 ETL Validación Archivos

**Dimensiones**

|  |  |
| --- | --- |
| **Proceso:** Dimensiones.ktr | |
| **Origen** | dim\_delito.csv  dim\_tip\_familiar.csv  dim\_niv\_consumo.csv  dim\_escolaridad.csv  dim\_sit\_joven.csv  dim\_genero.csv  dim\_ocupacion.csv |
| **Destino** | Public.DIM\_DELITOS  Public. DIM\_TIP\_FAMILIAR  Public.DIM\_NIV\_CONSUMO  Public.DIM\_ESCOLARIDAD  Public.DIM\_SIT\_JOVEN  Public.DIM\_GENERO  Public.DIM\_OCUPACION  Public.LOG\_DIM\_DELITOS  Public.LOG\_DIM\_TIPOfAMILIAR  Public.LOG\_ DIM\_NIVELCONSUMO  Public.LOG\_DIM\_ESCOLAPR  Public.LOG\_DIM\_SITJOVEN  Public.LOG\_DIM\_GENERO  Public.LOG\_DIM\_OCUPACION |
| **Descripción** | En este proceso se crean archivos paramétricos de extensión .CSV, que contiene el campo id y descripción de las variables: Delitos, Tipo\_Familia, Nivel\_Consumo, Escolaridad, Situación\_Joven, Genero, Ocupación. Esto usado para los procesos de homologación y limpieza de los datos que se cargan desde los archivos Excel que suministra la corporación, finalizado el paso los registros se insertaran en las tablas Destino mencionados en este apartado. |

Tabla 3 ETL Dimensiones

Imagen que contiene Calendario

Descripción generada automáticamenteIlustración 7 ETL Dimensiones

**STG1\_HECHOS**

|  |  |
| --- | --- |
| **Proceso:** STG1\_Hechos.ktr | |
| **Origen** | Public.cor\_in\_base  public.dim\_delitos |
| **Destino** | Public.STG1\_Hechos  Public.log\_stg1\_hechos |
| **Descripción** | Este proceso cruza la información de la tabla **COR\_IN\_BASE**, limpiando y validando el campo Nombre del Delito con la tabla **DIM\_DELITO,** los registros que no crucen se envían a una tabla(**log\_stg1\_hechos**) para que el equipo de trabajo pueda revisarlos y de ser necesario arreglarlos para volver a procesar la información, (Se tiene un control para evitar la inserción de duplicados), de lo contrato se insertan los registros en la tabla **STG1\_Hechos.** |

Tabla 4 STG1\_HECHOS

Escala de tiempo

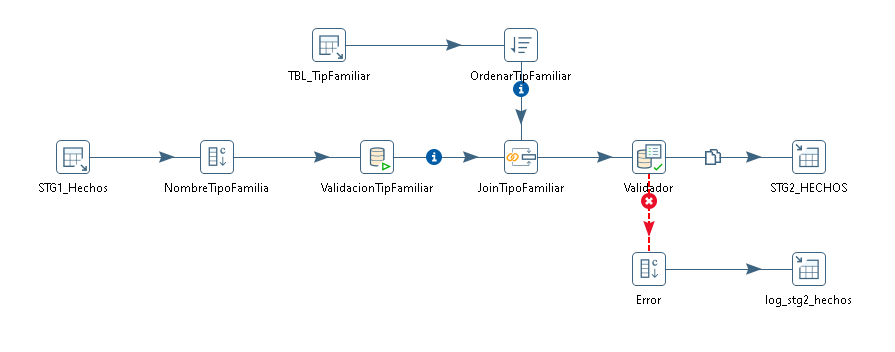
Descripción generada automáticamente

Ilustración 8 ETL STG1\_HECHOS

**STG2\_HECHOS**

|  |  |
| --- | --- |
| **Proceso:** STG2\_Hechos.ktr | |
| **Origen** | Public.STG1\_Hechos  public.dim\_tip\_familiar |
| **Destino** | Public.STG2\_HECHOS  Public.log\_stg2\_hechos |
| **Descripción** | Este proceso cruza la información de la tabla **STG1\_Hechos**, limpiando y validando el campo Nombre Tipo Familia con la tabla **DIM\_TIP\_FAMILIAR,** los registros que no crucen se envían a una tabla(**log\_stg2\_hechos**) que el equipo de trabajo pueda revisarlos y de ser necesario arreglarlos para volver a procesar la información, de lo contrato se insertan los registros en la tabla **STG2\_HECHOS.** |

Tabla 5 STG2\_Hechos

Ilustración 9 ETL STG2\_HECHOS

**STG3\_HECHOS**

|  |  |
| --- | --- |
| **Proceso:** STG3\_Hechos.ktr | |
| **Origen** | Public.STG2\_HECHOS  public.dim\_niv\_consumo |
| **Destino** | Public.STG3\_HECHOS  Public.log\_stg3\_hechos |
| **Descripción** | Este proceso cruza la información de la tabla **STG2\_Hechos**, limpiando y validando el campo nivel Consumo con la tabla **DIM\_NIV\_CONSUMO,** los registros que no crucen se envían a una tabla(**log\_stg3\_hechos**) que el equipo de trabajo pueda revisarlos y de ser necesario arreglarlos para volver a procesar la información, de lo contrato se insertan los registros en la tabla **STG3\_HECHOS.** |

Tabla 6 STG3\_Hechos

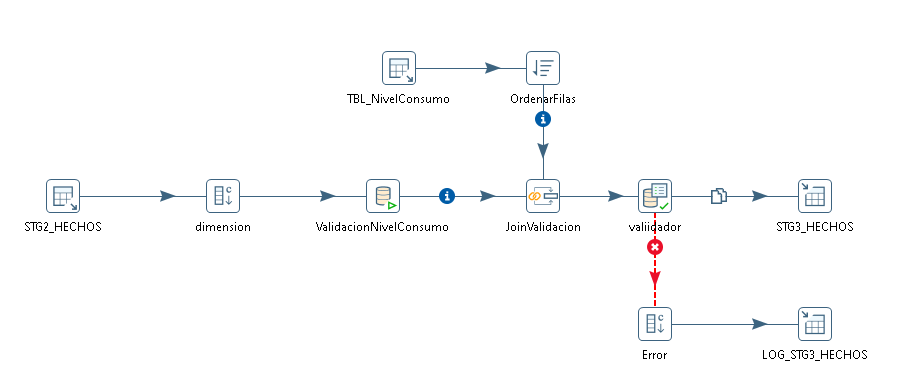


Ilustración 10 ETL STG3\_HECHOS

**STG4\_HECHOS**

|  |  |
| --- | --- |
| **Proceso:** STG4\_Hechos.ktr | |
| **Origen** | Public.dim\_sit\_joven  Public.STG3\_HECHOS |
| **Destino** | Public.STG4\_HECHOS  Public.log\_stg4\_hechos |
| **Descripción** | Este proceso cruza la información de la tabla **STG3\_Hechos**, limpiando y validando el campo situación Joven con la tabla **DIM\_SIT\_JOVEN,** los registros que no crucen se envían a una tabla(**log\_stg4\_hechos**) que el equipo de trabajo pueda revisarlos y de ser necesario arreglarlos para volver a procesar la información, de lo contrato se insertan los registros en la tabla **STG4\_HECHOS.** |

Tabla 7 STG4\_Hechos

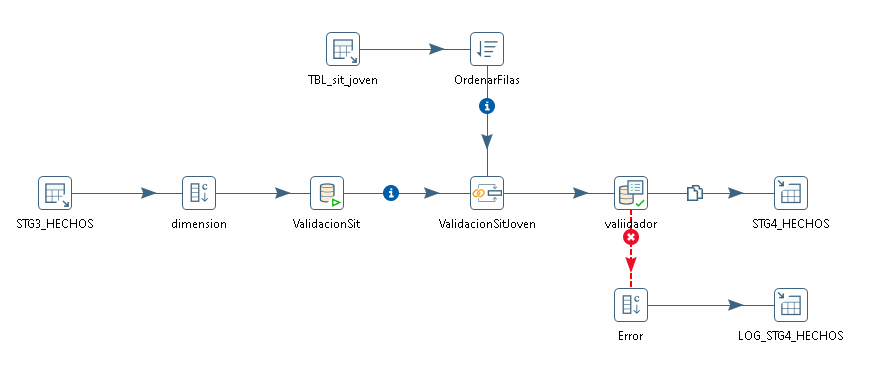


Ilustración 11 ETL STG4\_HECHOS

**STG5\_HECHOS**

|  |  |
| --- | --- |
| **Proceso:** STG5\_Hechos.ktr | |
| **Origen** | Public.dim\_genero  Public.STG4\_HECHOS |
| **Destino** | Public.STG5\_HECHOS  Public.log\_stg5\_hechos |
| **Descripción** | Este proceso cruza la información de la tabla **STG4\_Hechos**, limpiando y validando el campo genero con la tabla **DIM\_GENERO,** los registros que no crucen se envían a una tabla(**log\_stg5\_hechos**) que el equipo de trabajo pueda revisarlos y de ser necesario arreglarlos para volver a procesar la información, de lo contrato se insertan los registros en la tabla **STG5\_HECHOS.** |

Tabla 8 STG5\_Hechos

Diagrama

Descripción generada automáticamente

Ilustración 12 ETL STG5\_HECHOS

**STG6\_HECHOS**

|  |  |
| --- | --- |
| **Proceso:** STG6\_Hechos.ktr | |
| **Origen** | Public.dim\_escolaridad  Public.STG5\_HECHOS |
| **Destino** | Public.STG6\_HECHOS  Public.log\_stg6\_hechos |
| **Descripción** | Este proceso cruza la información de la tabla **STG5\_Hechos**, limpiando y validando el campo escolaridad con la tabla **DIM\_ESCOLARIDAD,** los registros que no crucen se envían a una tabla(**log\_stg6\_hechos**) que el equipo de trabajo pueda revisarlos y de ser necesario arreglarlos para volver a procesar la información, de lo contrato se insertan los registros en la tabla **STG6\_HECHOS.** |

Tabla 9 STG6\_Hechos

Diagrama

Descripción generada automáticamente

Ilustración 13 ETL STG6\_HECHOS

**STG7\_HECHOS**

|  |  |
| --- | --- |
| **Proceso:** STG7\_Hechos.ktr | |
| **Origen** | Public.dim\_ocupacion  Public.STG6\_HECHOS |
| **Destino** | Public.STG7\_HECHOS  Public.log\_stg7\_hechos |
| **Descripción** | Este proceso cruza la información de la tabla **STG6\_Hechos**, limpiando y validando el campo ocupación con la tabla **DIM\_OCUPACION,** los registros que no crucen se envían a una tabla(**log\_stg7\_hechos**) que el equipo de trabajo pueda revisarlos y de ser necesario arreglarlos para volver a procesar la información, de lo contrato se insertan los registros en la tabla **STG7\_HECHOS.** |

Tabla 10 STG7\_Hechos

Diagrama

Descripción generada automáticamente

Ilustración 14 ETL STG7\_HECHOS

**FACT\_CORP**

|  |  |
| --- | --- |
| **Proceso:** FACT\_CORP.ktr | |
| **Origen** | Public.STG7\_HECHOS  Public.fact\_corp |
| **Destino** | Public.fact\_corp |
| **Descripción** | Este proceso inserta los registros con la información transformada y homologada en la tabla **FACT\_CORP**, antes se realiza una validación para controlar los duplicados. |

Tabla 11 Fact\_corp

Diagrama

Descripción generada automáticamente

Ilustración 15 ETL FACT\_CORP

**Tabla Final**

|  |  |  |  |
| --- | --- | --- | --- |
| **ID** | **Nombre Columna** | **Tipo** | **Tabla Asociada** |
| 1 | ID | Double | Número de registro |
| 2 | DELITO | Text | public.dim\_delitos |
| 3 | NOM\_APE | Text |  |
| 4 | NUM\_DOC\_ID | Text |  |
| 5 | PROGRAMA | Text |  |
| 6 | ANNO\_INGRESO | Text |  |
| 7 | MES\_INGRESO | Text |  |
| 8 | GENERO | Text | public.dim\_genero |
| 9 | EDAD | Text |  |
| 10 | ESCO\_APRO | Text | public.dim\_escolaridad |
| 11 | ESCO\_ACT | Text | public.dim\_escolaridad |
| 12 | OCUPACION | Text | public.dim\_ocupacion |
| 13 | AUTORIDAD | Text |  |
| 14 | NUM\_JUEZ | Text |  |
| 15 | TIEMPO\_SAN | Text |  |
| 16 | SALUD | Text |  |
| 17 | REGIMEN | Text |  |
| 18 | TIP\_FAMILIAR | Text | public.dim\_tip\_familiar |
| 19 | BARRIO | Text |  |
| 20 | MUNICIPIO | Text |  |
| 21 | NIV\_CONSU | Text | public.dim\_niv\_consumo |
| 22 | SIT\_ATUAL | Text | public.dim\_sit\_joven |
| 23 | ARCHIVO | Text | Nombre Del Archivo para control. |
| 24 | FEC\_DATA | Text | Fecha Procesamiento |

Tabla 12 Estructura Final tabla FACT\_CORP

## 3.3.5 DESARROLLO PREGUNTAS

En este apartado se responderán las preguntas que han sido planteadas por el equipo de trabajo y que busca mediante los datos procesados encontrar la respuesta, para la solución de las respuestas costa de 2 partes la primera información técnica del uso de los gráficos y procesos realizados para la carga de información, otro ítem con el análisis funcional por parte de la corporación.

## 3.3.5.1 Cuáles son los motivos de Ingreso de adolescentes y jóvenes con sanciones penales.

La Corporación quiere conocer cuál es el motivo de ingreso más recurrente en los adolescentes y jóvenes, esto para desarrollar planes de trabajo que impacten positivamente a cada usuario.

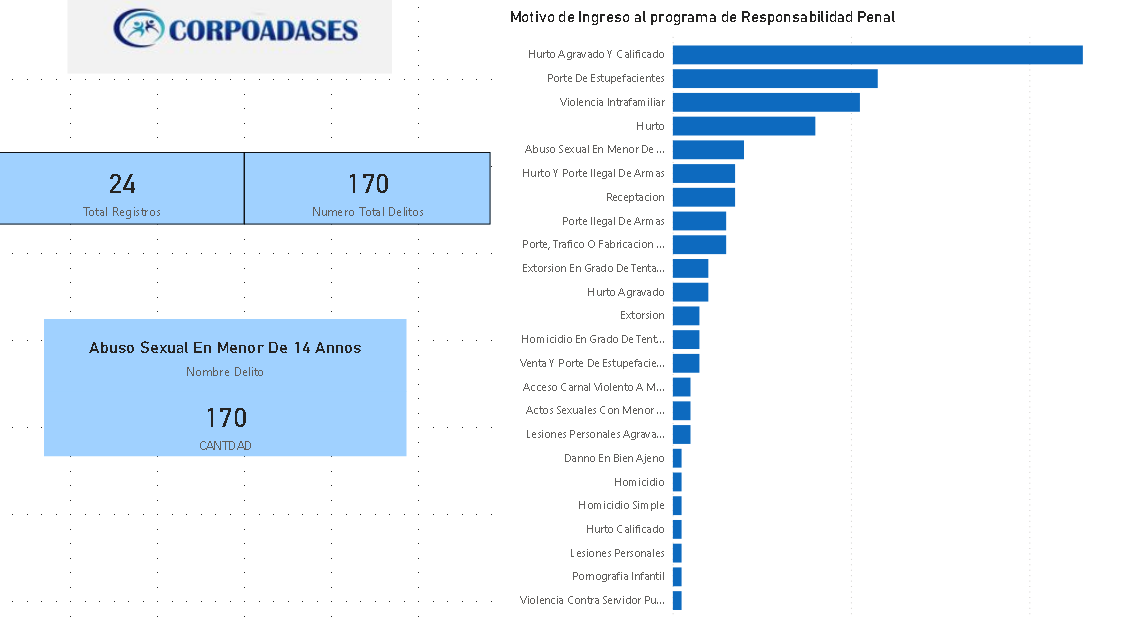


Ilustración 16 Visualizacion Motivo de Ingreso al Programa de Responsabilidad Penal

|  |  |
| --- | --- |
| **Visualizacion:** Motivo de Ingreso al Programa de Responsabilidad Penal | |
| **Tabla Origen** | fact\_corp  dim\_delito |
| **Variables** | Delito |
| **Tipo grafico** | Barras |
| **Objetivo** | Representar, Comparar y diferenciar cada uno de los delitos. |

Tabla 14 Documentación Visualizacion Motivo Ingreso al programa de Responsabilidad Penal

Teniendo en cuenta los datos analizados el delito de mayor prevalencia es el Hurto Agravado y Calificado, lo que infiere que pudieron encontrarse bajo los efectos de alguna sustancia psicoactivas, existiendo una relación directa en la comisión del delito.

Esta información genera insumo a la corporación para orientar el plan de atención enfatizando concientización al delito y el daño causado, buscando la interiorización de valores que logren modificar su conducta reorientando su proyecto de vida.

Otro aspecto para considerar se orienta hacia la identificación del nivel de consumo de drogas, vinculándose en el tratamiento que oferte el sistema de salud, acorde con la particularidad de cada caso.

## 3.3.5.2 Cuál es la Edad Promedio de ingreso de Jóvenes con sanciones penales

Se requiere identificar por parte de la corporación cual es la edad promedio de ingreso a los programas ofertados, esto porque existe una tendencia a minimizar la responsabilidad del delito por parte de los adolescentes y Jóvenes.

Gráfico

Descripción generada automáticamente Ilustración 17 Edad Promedio de ingreso de Jóvenes con sanciones penales

Esta visualización se dividió en 2 partes:

|  |  |
| --- | --- |
| **Visualizacion:** Edad Promedio de ingreso de Jóvenes con sanciones penales | |
| **Tabla Origen** | fact\_corp |
| **Variables** | Edad |
| **Tipo grafico** | Histograma |
| **Objetivo** | Visualizar la distribución de la edad de los adolescentes y jóvenes que ingresan a la corporación resaltando la media. |

Tabla 15 Visualizacion Edad Promedio de ingreso histograma

|  |  |
| --- | --- |
| **Visualizacion:** Edad por Delito | |
| **Tabla Origen** | fact\_corp |
| **Variables** | Delito  Edad |
| **Tipo grafico** | Gráfico de Tarta  Tabla |
| **Objetivo** | Comparar el porcentaje de edad de los adolescentes y jóvenes e identificar la cantidad de jóvenes agrupados por la edad y el delito. |

Tabla 16 Visualizacion Edad por Delito

La edad es un factor impórtate por que permite identificar las características del desarrollo inherentes al usuario que se va a intervenir haciendo énfasis principalmente en todo lo relacionado en el desarrollo de la personalidad.

Se identifica como para el caso del programa operado por Corpoadases la edad promedio de ingreso es de 17 años, esta es una etapa en la cual los jóvenes se encuentran consolidando su identidad, donde el establecimiento de las relaciones con el grupo de pares sigue generando un nivel de influencia alto en este aspecto, teniendo mayor incidencia los comportamientos anti sociales que los haría permanecer en contextos vulnerables frente a las actividades de alto riesgo psicosocial (delitos, consumo de sustancias, vinculación en grupos al margen de la ley, entre otros).

Por esta razón, las acciones generadas desde el proceso tanto a nivel individual, grupal, familiar y en el contexto contemplan el afianzamiento de la autonomía como factor fundamental para el desarrollo y mantenimiento de la conciencia social; sin embargo, este trabajo exige mayor compromiso del contexto familiar quienes al tener mayor tiempo de interacción con los adolescentes deben fortalecer los aspectos abordados desde el programa.

## 3.3.5.3 Cuál es la tipología Familiar de los Jóvenes con sanciones penales

La corporación quiere Identificar la tipología familiar de los adolescentes y Jóvenes que ingresan a la corporación para determinar si en algún caso se pueda relacionar con la reincidencia en un delito.

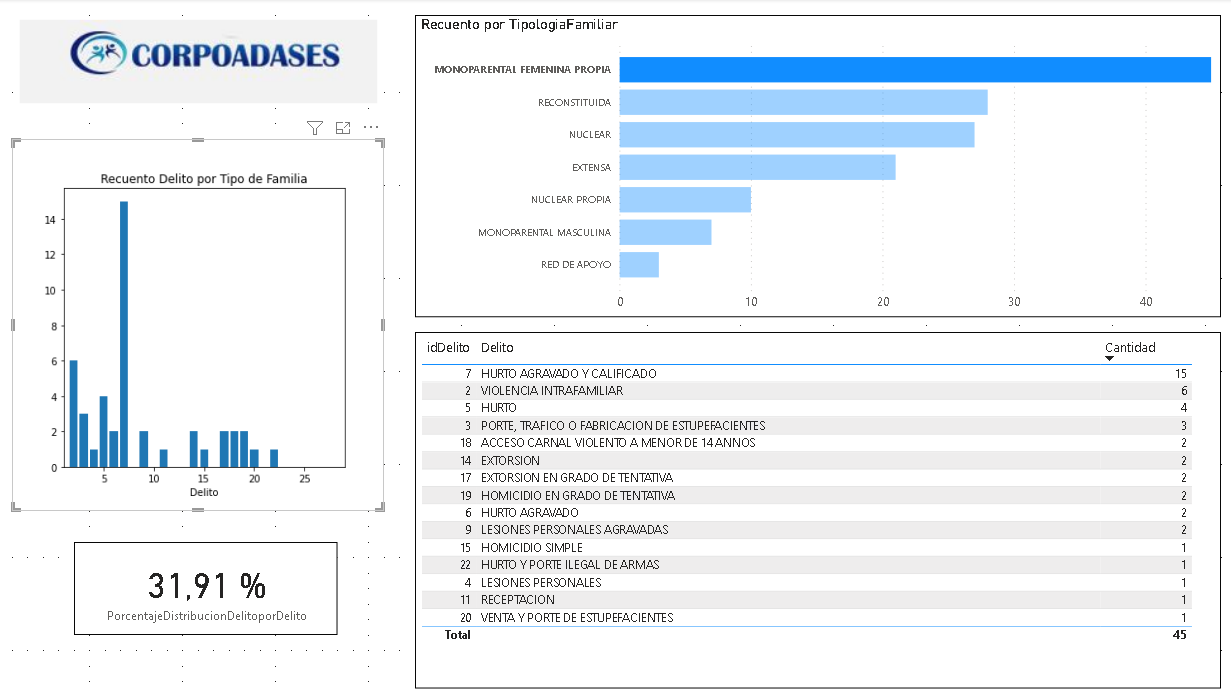


Ilustración 18 Tipología Familiar de los Jóvenes con sanciones penales

|  |  |
| --- | --- |
| **Visualizacion:** Tipología Familiar de los Jóvenes con sanciones penales | |
| **Tabla Origen** | fact\_corp  dim\_tip\_familiar |
| **Variables** | Tip\_familiar  Delito |
| **Tipo grafico** | Gráfico de Barras Apiladas  Tabla  Tarjeta |
| **Objetivo** | Identificar el porcentaje de distribución de la tipología familiar y la agrupación por delito. |

Tabla 17 Visualizacion Recuento por TipologiaFamiliar

La tipología familia Monoparental Femenina es la más recurrente y representa el 31.91%.

Acorde a lo evidenciado en el gráfico se mantiene la prevalencia en los casos de familias de tipología familia monoparental con jefatura femenina, siendo característica principal de estos casos la presentación de mujeres cabezas de familia que por sus responsabilidades económicas deben permanecer la mayor parte del tiempo fuera de casa por lo que delegan el cuidado de sus hijos a terceros incrementando los factores de riesgo que pudieron incidir en el ingreso al Sistema de Responsabilidad Penal para Adolescentes.

De esta manera se destaca a nivel generalizado que este aspecto es importante para la operativización del programa, desde el punto de vista de que el apoyo que ejerce la familia para el cumplimiento de los compromisos del programa y el acompañamiento que realizan desde el medio familiar son los que permiten generar un cambio significativo en la transformación de su estilo de vida.

## 3.3.5.4 Nivel de Consumo de los Jóvenes con sanciones penales.

Se quiere identificar el nivel de consumo y la distribución de edades de los jóvenes que tienen sanciones penales, esto para determinar si podría relacionar esta variable con la reincidencia en delitos.

Interfaz de usuario gráfica, Gráfico, Aplicación

Descripción generada automáticamente

Ilustración 19 Nivel de Consumo de los Jóvenes con sanciones penales

|  |  |
| --- | --- |
| **Visualizacion:** Visualizacion Nivel de Consumo de Jóvenes con sanciones penales | |
| **Tabla Origen** | fact\_corp  dim\_niv\_consumo |
| **Variables** | Niv\_consumo  Edad |
| **Tipo grafico** | Histograma  Tabla  Gráfico Barras Apiladas  Gráfico de Anillos |
| **Objetivo** | Representar, comparar la proporción y distribución de la variable nivel de consumo y edad.  Agrupar los jóvenes por delito y Nivel de consumo. |

Tabla 18 Visualizacion Nivel de Consumo de Jóvenes con sanciones penales.

El nivel de consumo Problemático es el más alto y representa el 34.6%.

Acorde con los gráficos se evidencia que prevalece el nivel de consumo problemático en un alto porcentaje de adolescentes y jóvenes vinculados en los programas de Corpoadases, destacando frente a esta variable que para el caso de estos usuarios las drogas referidas son las que son clasificadas como sustancias psicoactivas ilícitas (marihuana, perico, pópper, bóxer, benzodiacepinas, entre otras).

El hacer análisis de este aspecto se identifica que la sustancia por sí sola no genera riesgo frente a la reincidencia de la conducta delictiva, sino lo que es fundamental en el momento de analizarla es dentro de la historia de consumo identificar las estrategias de afrontamiento de cada adolescente o joven al momento de “regular” el consumo; encontrando entonces que casi la totalidad de los usuarios han estado bajo efecto de consumo de drogas y que entre más aguda es esta problemática mayor es el riesgo de reincidencia al delito.

## 3.3.5.5 Nivel de Escolaridad de los Jóvenes con sanciones penales.

La Corporación quiere conocer si los jóvenes y adolescentes que ingresan a la corporación continúan con sus estudios, esto para determinar estrategias de motivación y desarrollo de actividades que logren reorientación para el desarrollo del proyecto de vida.

Gráfico

Descripción generada automáticamente

lustración 20 Nivel de Escolaridad de los Jóvenes Con sanciones penales

El porcentaje de jóvenes que no continúan con sus estudios representan el 73.53%.

|  |  |
| --- | --- |
| **Visualizacion:** Nivel de Escolaridad de los Jóvenes Con sanciones penales | |
| **Tabla Origen** | fact\_corp |
| **Variables** | Delito  EscoApro  EscoAct |
| **Tipo grafico** | Gráfico de Tarta  Gráfico de Barras Apiladas  Tabla |
| **Objetivo** | Representar el porcentaje de los jóvenes que continúan y no continúan con sus estudios.  Identificar por delito el nivel de escolaridad Aprobado vrs Actual.  Agrupar los jóvenes por nivel de escolaridad aprobada, actual y el delito. |

Acorde a lo evidenciado en el gráfico, el mayor porcentaje de los adolescentes que permanecen vinculados al programa no continúan con sus estudios académicos; sin embargo, se destaca que el curso hasta el que han llegado en su mayoría son los grados sexto, séptimo y octavo; siendo este un alto limitante para vincularse en procesos de formación ocupacional sobre todo en el Sena (carreras técnicas y\o tecnológicas), afectando así directamente la estructuración de su proyecto de vida.

De esta manera, al analizar esta variable Corpoadases podrá buscar estrategias de motivación para que retomen su interés por continuar con sus estudios llevándoles a que se permitan la posibilidad de proyectarse en actividades formales y estables que les garanticen así disminuir los niveles de reincidencia en el delito; destacando que es un aspecto que poco asimilan pues los estilos de vida que han naturalizado están directamente relacionados con actividades delictivas: vinculación en pandillas, consumo de droga, participación en actividades ilícitas, entre otros.

Del mismo modo se destaca que de los adolescentes y jóvenes bachilleres, le representan a la corporación la posibilidad de reorientarlos para aprovechar sus habilidades y ubicarlos en procesos de educación que favorezcan su calidad de vida.

## 3.3.5.6 ¿Existe alguna relación entre tipología Familiar Vs Reincidencia en Delito?

Gráfico, Gráfico de líneas

Descripción generada automáticamenteI

lustración 21 Relación Tipología Familiar vrs Reincidencia en Delito

Las variables tipología familia y la reincidencia en el delito no se relacionan

|  |  |
| --- | --- |
| **Visualizacion:** Relación Tipología Familiar vrs Reincidencia en Delito | |
| **Tabla Origen** | fact\_corp |
| **Variables** | Tip\_familiar  Rein\_delito |
| **Tipo grafico** | Gráfico de Barras Agrupadas  Distribución Chi-Cuadrado |
| **Objetivo** | Comparar la variable Tip\_familiar entre los jóvenes que reinciden y no Reinciden para determinar si existe o no relación entre estas 2 variables usando el método chi-cuadrado para determinar la relación entre las variables. |

Tabla 19 Relación tipología familia - Reincidencia en Delito

No existe relación, para analizar la incidencia de la tipología familiar con la reincidencia al delito, es necesario también tener en cuenta este aspecto junto con los factores de vulnerabilidad sociofamiliares que se identifican en la particularidad de cada caso, destacando que de este aspecto se encuentra poco sustento teórico, generando entonces la hipótesis que la ausencia de las figuras de autoridad en el núcleo familiar genera en los adolescentes y jóvenes un incremento en el tiempo que pasan sin supervisión lo que los hace más vulnerables a la reincidencia al delito.

## 3.3.5.7 Está asociado el Nivel de consumo con la reincidencia en un delito

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Ilustración 22 Relación Nivel de Consumo vrs Reincidencia de Delito

Las variables nivel consumo y la reincidencia en el delito se relacionan.

|  |  |
| --- | --- |
| **Visualizacion:** Relación Nivel de Consumo vrs Reincidencia de Delito | |
| **Tabla Origen** | fact\_corp |
| **Variables** | Niv\_consumo  Rein\_delito |
| **Tipo grafico** | Gráfico de Barras Agrupadas  Distribución Chi-Cuadrado |
| **Objetivo** | Comparar la variable Niv\_consumo entre los jóvenes que reinciden y no Reinciden para determinar si existe o no relación entre estas 2 variables usando el método chi-cuadrado para determinar la relación. |

Tabla 20 Relación Niv\_consumo- Reincidencia en Delito

Las variables si se relacionan y al hacer análisis de este aspecto se identifica que la sustancia por sí sola no genera riesgo frente a la reincidencia de la conducta delictiva, sino lo que es fundamental en el momento de analizarla es dentro de la historia de consumo identificar las estrategias de afrontamiento de cada adolescente o joven al momento de “regular” el consumo; encontrando entonces que casi la totalidad de los usuarios han estado bajo efecto de consumo de drogas y que entre más aguda es esta problemática mayor es el riesgo de reincidencia al delito.

## 3.3.5.8 Cuál es la probabilidad de reincidencia de un Joven que tiene sanciones penales.

La Corporación quiere conocer el porcentaje en la probabilidad de reincidencia de un adolescente o joven que ingresa a la corporación, esto como una herramienta de apoyo y alerta temprana para mediante actividades grupales e individuales desarrollar redes de apoyo para la estructuración de su proyecto de vida.

## Definicion Variables

Para el entrenamiento se han seleccionado las siguientes variables por recomendación del equipo de trabajo de la corporación:

Delito

Nivel de Consumo

Ocupación

Se descargar la tipología Familiar ya que esta no se relaciona con la reincidencia del delito.

Teniendo en cuenta la clase salida: 0 – No Reincide, 1 – Reincide se aplica los siguientes Modelos:

Random Forest

Regresión Logística

SVM

## Random Forest

Para la ejecución de este modelo se realiza los siguientes pasos:

1. Selección de las columnas: delito, niv\_consumo, ocupación como variables regresaras.
2. Selección de la variable rein\_delito como variable respuesta.
3. Definicion del test\_size de 0.20, n\_estimators=50 (50 árboles).
4. Ejecución modelo RandomForestClassifier (Árbol de clasificacion).
5. Se obtiene el potencial predictivo:

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Ilustración 23 Potencial Predictivo Random Forest

1. Impresión de la matriz de confusión.

Gráfico

Descripción generada automáticamente

Ilustración 24 Matriz de Confusion Random Forest

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Ilustración 25 Entrenamiento Modelo Random Forest

## Regresión Logística

Para la ejecución de este modelo se realiza los siguientes pasos:

1. Selección de las columnas: delito, niv\_consumo, ocupación como variables regresaras.
2. Selección de la variable rein\_delito como variable respuesta.
3. Definicion del test\_size=0.20, random\_state=0
4. Ejecución modelo LogisticRegression.
5. Se obtiene el potencial predictivo:

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Ilustración 26 Potencial Predictivo Regresión Logística

1. Impresión de la matriz de confusión.

Gráfico, Gráfico de rectángulos

Descripción generada automáticamente

Ilustración 27 Matriz de Confusion Regresión Logística

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Ilustración 28 Entrenamiento Modelo Regresión Logística

## SVM

1. Selección de las columnas: delito, niv\_consumo, ocupación como variables regresaras.
2. Selección de la variable rein\_delito como variable respuesta.
3. Definicion del test\_size=0.20, random\_state=0
4. Ejecución modelo SVM.
5. Se obtiene el potencial predictivo:

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Ilustración 29 Potencial Predictivo SVM

1. Impresión de la matriz de confusión.

Gráfico, Gráfico de rectángulos

Descripción generada automáticamente

Ilustración 30 Comparación Modelos Predictivos SVM

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Ilustración 31 Entrenamiento SVM

## Comparación Métricas

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Modelo** | **Accuracy** | **Precision** | **Recall** | **F1** |
| RandonForest | 0.75675 | 0.78571 | 0.6470 | 0.70967 |
| RegresionLogistica | 0.59459 | 0.58333 | 0.41176 | 0.48275 |
| SVM | 0.75675 | 0.875 | 0.4666 | 0.60869 |

Tabla 21 Métricas Modelos

**Accuracy:** Porcentaje de predicciones correctas frente al total, en este caso los modelos Random Forest y SVM superen el 70% la exactitud es aceptable con un 75% en ambos casos, La regresión logística tiene un porcentaje debajo de 60%.

**Precision:** Porcentaje de predicciones positivas fueron correctamente identificadas, En este caso el modelo de SVM el porcentaje de precisión es del 87% mientras que Random Forest es del 78%, El modelo SVM es más preciso en predecir si puede reincidir en un delito, mientras que la regresión lógica tiene un porcentaje menor al 60%.

**Recall:** Porcentaje casos identificados del total posible. En este caso el modelo de Random Forest predice casi el 65% de los casos reales positivos correctamente mientras que los modelos SVN y Regresión Logística está por debajo del 50%.

SVM predecirá menos casos de posible reincidencia a comparación del Random Forest que detectaría 18.04% más.

La corporación busca una herramienta de ayuda que le permita tomar acciones ante una posible alerta permitiendo crear estrategia y planes trabajo individuales con los adolescentes y jóvenes que ingresan a la corporación, el modelo que mejor resultado aporta es RandonForest con un potencial de predicción del 75.68%.

La probabilidad de reincidencia es del 40%.

La implementación de Modelos Predictivos dentro de los programas sociales de Corpoadases, le permite sistematizar y analizar patrones de comportamientos para organizar el proceso de atención institucional basado en la toma de decisiones oportuna que favorezca a los grupos de atención encontrados en el presente trabajo de maestría.

El desarrollo del modelo predictivo con precisión de 0.755 le permite a la corporación tener un punto de partida para predecir si puede o no reincidir un joven, sin embargo, el modelo debe seguir en entrenamiento para mejorar su score.

## 3.3.5.9 Clasificacion del tipo de tratamiento que requiere un Joven con sanciones penales por el nivel de consumo de drogas.

La Corporación quiere definir los grupos de jóvenes que por su Nivel de consumo, edad y tipo de tratamiento se puedan separar y de esta forma realizar trabajos enfocados en actividades de prevención al consumo, orientación sobre el daño que genere el consumo de drogas o si es necesario inicio de tratamientos con apoyo de otras instituciones.

## Definicion Variables

Para el entrenamiento se han seleccionado las siguientes variables por recomendación del equipo de trabajo de la corporación:

**Tipo Tratamiento**

Esta variable se define como:

* Consumo abusivo - requiere tratamiento intramural
* Consumo disfuncional - tratamiento ambulatorio
* Consumo experimental/social - citas de seguimiento con psicología
* No consume - participación en actividades de prevención
* Abstinencia - participación en grupo de apoyo

**Edad**

**Nivel de Consumo**

Con la definicion de las variables se realiza el entrenamiento de los Clúster, el equipo de trabajo de la corporación se definen 3 grupos los cuales puedan diferenciar las áreas

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

El resultado de la ejecución de este Modelo se exportará en n archivo csv el cual será el insumo para el desarrollo de las visualizaciones en Power BI.

**Grupo 1**

Gráfico, Gráfico de barras

Descripción generada automáticamenteIlustración 34 Clasificacion por Tipo de Tratamiento Grupo 1

Resultado:

|  |  |
| --- | --- |
| Tipo tratamiento | Seguimiento con psicología  Tratamiento ambulatorio |
| Edad | 17  18  19  20 |
| Nivel consumo | Habitual  Problemático  Social |

Tabla 24 Clasificacion por Tipo de Tratamiento Grupo 1

Esta visualización le permitirá a la corporación frente a la toma de decisiones, que los adolescentes y jóvenes que el modelo ubique es este grupo se oriente el plan de atención hacia la concientización del daño generado por el consumo de drogas y, como esto a su vez incide dentro de la estructuración de su proyecto de vida, teniendo en cuenta que por la edad es prioritaria su atención toda vez que una reincidencia al delito en ellos le representaría el ingreso al sistema judicial de adultos y todas las desventajas que este aspecto conlleva.

**Grupo 2**

Gráfico, Gráfico de barras, Histograma

Descripción generada automáticamenteIlustración 35 Clasificacion por Tipo de Tratamiento Grupo 2

Resultado:

|  |  |
| --- | --- |
| Tipo tratamiento | Intramural  Tratamiento ambulatorio |
| Edad | 14  15  16  17 |
| Nivel consumo | Problemático  Habitual |

Tabla 25 Clasificacion por Tipo de Tratamiento Grupo 2

Con este grupo de adolescentes y jóvenes el programa debe orientarse principalmente a la articulación de acciones interinstitucionales, toda vez que los adolescentes y jóvenes clasificados en esta visualización requieren ser apartados de su contexto para que la agudización de la problemática de consumo no afecte su relación con el entorno.

**Grupo 3**

Gráfico, Gráfico de barras, Histograma

Descripción generada automáticamenteIlustración 36 Clasificacion por Tipo de Tratamiento Grupo 3

Resultado:

|  |  |
| --- | --- |
| Tipo tratamiento | Grupo apoyo |
| Edad | 14  15  16  17  18  19  20  21 |
| Nivel consumo | Abstinencia |

Tabla 26 Clasificacion por Tipo de Tratamiento Grupo 3

Finalmente, la ubicación de adolescentes y jóvenes en este grupo identifica la necesidad de mantener acciones preventivas que se orienten hacia el mantenimiento del no consumo como la mejor estrategia de prevención a la reincidencia al delito. De esta manera, el equipo técnico fortalecerá las estrategias de afrontamiento y toma de decisiones en cada uno de los adolescentes y jóvenes vinculados a la corporación.

## 3.3.5.10 Definicion de áreas de refuerzo que requiera un joven que ha ingresado a la corporación.

La corporación quiere enfocarse en atender a los jóvenes y adolescentes que ingresan a la corporación y conocer cuál sería el porcentaje de atención en las áreas de Psicología y Terapia Ocupacional.

**Atención en el Área de Psicología**

El profesional en psicología promueve y orienta procesos integrales, dinámicos y continuos para fortalecer los recursos psicológicos y sociales de los y las adolescentes y jóvenes y sus familias. La atención y acompañamiento del psicólogo debe ser asumida desde una postura ética que pasa de instruir a construir, considerando así al ser humano como sujeto creativo y proactivo, por lo que los procesos reflexivos que se derivan de las atenciones deben fortalecer la capacidad de respuesta de los y las adolescentes y jóvenes ante las adversidades por las que atraviesa durante su curso de vida.

Teniendo en cuenta el liderazgo que ejerce el profesional de psicología dentro del plan de intervención que se estipula con cada adolescente o joven vinculado en Corpoadases, dentro del presente trabajo final de máster y acorde con la orientación recibida por la coordinadora de la modalidad de atención, el área de psicología para generar avances debe liderar todos los aspectos relacionados con la determinación del nivel de consumo de cada usuario, remisión a los servicios de salud para el tratamiento del mismo e intervención para el seguimiento individual donde identifica como este consumo ha incidido o incide en la comisión de los delitos.

**Variables de Entrada:**

Delito

Nivel de Consumo

**Salida**

Porcentaje de atención que requiere el adolescente y Joven que ingresa a la corporación por su Delito y Nivel de Consumo.

**Delito**

Descripción del motivo de ingreso del adolescente y Joven a la corporación,Se divide el delito en 3 categorías y cada uno se le asigna un peso que servirán de antecedentes para el desarrollo de la lógica difusa:

|  |  |  |
| --- | --- | --- |
| **Definicion** | **Grado Delito** | **Descripción Delitos** |
| Menores | Delitos del grado 1 al 8 | Aborto  Violencia Contra Servidor Público  Daño Ambiental  Receptación  Lesiones Personales  Daño en bien ajeno  Hurto y Porte Ilegal de Armas  Venta y Porte de estupefacientes |
| Leves | Delitos del grado 9 al 16 | Violencia Intrafamiliar  Lesiones Personales Agravadas  Hurto Agravado  Porte, Trafico o Fabricación de estupefacientes.  Extorsión  Extorsión en grado de tentativa. |
| Agravados | Delitos del grado 17 al 29 | Abuso Sexual en menor de 14 Años  Hurto Agravado y Calificado  Homicidio  Actos Sexuales con Menor de 14 Años  Pornografía Infantil  Acceso carnal violento a menor de 14 años  Homicidio en grado de tentativa  Acceso Carnal abusivo con menor de 14 años  Hurto Calificado. |

Tabla 27 Definicion Antecedente Variable Delito

**Nivel de Consumo**

Descripción del nivel de consumo del adolescente y Joven, esta variable se divide en 3 categorías y cada uno se le asigna un peso que servirán de antecedentes para el desarrollo de la lógica difusa:

|  |  |  |
| --- | --- | --- |
| **Definicion** | **Peso Nivel de consumo** | **Descripción Delitos** |
| Bajo | Nivel de consumo 1 al 2 | No Consume  Abstinencia |
| Medio | Nivel de consumo 3 al 4 | Social  Experimental |
| Alto | Nivel de consumo 5 al 6 | Habitual  Problemático |

Tabla 28 Definicion Antecedente Variable Nivel Consumo

**Atención en Psicología**

Porcentaje de atención en esta área el cual se calcula en un rango de 0 a 100.

|  |  |
| --- | --- |
| **Definicion** | **Valor** |
| Baja | 0, 0, 50 |
| Media | 0, 50, 100 |
| Prioritaria | 50,100,100 |

Tabla 29 Definicion Valores Consecuentes

Para el desarrollo de este Modelo se requiere las siguientes Librarías:

Numpy: Librería Matemática de Python

scikit-fuzzy (skfuzzy): Librería usada para lógica difusa

pandas: Manejo Estructuras de datos.

matplotlib: Librería que ayuda a pintar funciones.

Se configuran las siguientes Reglas para la definicion de los porcentajes de atención en Psicología.

|  |  |  |
| --- | --- | --- |
| **Delito** | **Nivel Consumo** | **Resultado**  **Atención Psicológica** |
| Menores | Bajo | Baja |
| Menores | Medio | Baja |
| Menores | Alto | Media |
| Leves | Bajo | Media |
| Leves | Medio | Prioritaria |
| Leves | Alto | Prioritaria |
| Agravados | Bajo | Media |
| Agravados | Medio | Media |
| Agravados | Alto | Prioritaria |

Tabla 30 Definicion Reglas

**Atención en Terapia Ocupacional**

Por otra parte, el profesional en terapia ocupacional asume dentro del talento humano el perfil del especialista de área, siendo transversal dentro de sus acciones: Realizar seguimiento y acompañamiento permanente a la población del SRPA, durante el desarrollo de sus ABC (actividades básicas cotidianas), y las AVD (actividades de la vida diaria), con el fin de mejorar el desempeño funcional en cada una de sus actividades. Promover en la población del SRPA la construcción de un proyecto de vida productivo y sostenible, orientándolos hacia procesos de inclusión social reales y duraderos.

Es por ello, que dentro del Modelo Predictivo que se está creando en este Trabajo Final de Maestría este profesional asume el liderazgo en todos los aspectos relacionados con la escolaridad y la ocupación desempeñada por cada usuario, toda vez que estos aspectos son fundamentales al momento de reorientar sus proyectos de vida.

**Se definen las variables:**

Escolaridad Aprobada

Escolaridad Actual

Ocupación

**Salida**

Se divide la Escolaridad Aprobada y Escolaridad Actual se divide en 3 categorías y cada uno se le asigna un peso que servirán de antecedentes para el desarrollo de la lógica difusa:

|  |  |  |
| --- | --- | --- |
| **Definicion** | **Grado Escolaridad** | **Descripción Escolaridad** |
| Primaria | Escolaridad grados 1 al 5 | Primero  Segundo  Tercero  Cuarto  Quinto  Ciclo III |
| Secundaria | Escolaridad grados 6 al 11 | Sexto  Séptimo  Octavo  Noveno  Ciclo IV  Decimo  Once  Ciclo V |
| Superior | Escolaridad grados 12 al 15 | Bachiller  Técnico  Tecnología  Universitarios |

Tabla 34 Definicion Antecedente Variable Escolaridad

Se divide la ocupación en 2 categorías y cada uno se le asigna un peso que servirán de antecedentes para el desarrollo de la lógica difusa:

|  |  |  |
| --- | --- | --- |
| **Definicion** | **Peso Ocupación** | **Descripción Ocupación** |
| RealizaActividad | Ocupación grado 2 al 5 | Estudia  Estudia y trabaja  trabaja |
| NoRealizaActividad | Ocupación grado 1 | Cesante |

Tabla 35 Definicion Antecedente Variable Ocupación

Consecuente

Se define la variable AtencionTerapiaOcupacional como consecuente con los siguientes valores.

|  |  |
| --- | --- |
| **Definicion** | **Valor** |
| Baja | 0, 0, 50 |
| Media | 0, 50, 100 |
| Prioritaria | 50,100,100 |

Tabla 36 Definicion Valores Consecuentes

Se configuran las siguientes Reglas.

|  |  |  |  |
| --- | --- | --- | --- |
| **Escolaridad Actual** | **Escolaridad Aprobada** | **Ocupación** | **Resultado**  **Atención Terapia Ocupacional** |
| primaria | primaria | RealizaActividad | Media |
| primaria | primaria | NoRealizaActividad | Prioritaria |
| primaria | secundaria | RealizaActividad | Baja |
| primaria | secundaria | NoRealizaActividad | Prioritaria |
| secundaria | secundaria | RealizaActividad | Baja |
| secundaria | secundaria | NoRealizaActividad | Prioritaria |
| secundaria | superior | RealizaActividad | Baja |
| secundaria | superior | NoRealizaActividad | Media |
| superior | superior | RealizaActividad | Baja |
| superior | superior | NoRealizaActividad | Media |

Tabla 37 Definicion Reglas

Mediante la aplicación de reglas se puede establecer los porcentajes en la atención que requiere un adolescente y joven que ingresa a la corporación en el área de psicología y terapia ocupacional, y con la agrupación realizada en el paso anterior se puede personalizar cada joven y esto permitir enfatizar en las áreas que tenga alguna dificultad para mejorar tanto su calidad como proyecto de vida, es importante involucrar más áreas como trabajo social y pedagogía.

**3.3.6 Resultado Final**

Modelo Random Forest

CASO 1

|  |  |
| --- | --- |
| **Delito** | Hurto Agravado y Calificado |
| **Ocupación** | Estudia |
| **Nivel de consumo** | Abstinencia |

Tabla 22 Prueba RF 1

Diagrama

Descripción generada automáticamente

Ilustración 32 Probabilidad de Reincidencia

La probabilidad de reincidencia es del 17%.

Caso 2

|  |  |
| --- | --- |
| **Delito** | Danno Ambiental |
| **Ocupación** | Estudia y Trabaja |
| **Nivel de consumo** | Problemático |

Tabla 23 Prueba RF 2

La probabilidad de reincidencia es del 40%.

Diagrama

Descripción generada automáticamente

Ilustración 33 Probabilidad de Reincidencia II

Lógica Difusa

**Prueba 1**

|  |  |
| --- | --- |
| **Delito** | Violencia contra Servidor Publico |
| **Nivel Consumo** | No Consumo |
| **Porcentaje Atención Psicológica** | 28 % |

Tabla 31 Prueba 1 Lógica Difusa

Gráfico, Gráfico de líneas

Descripción generada automáticamenteIlustración 37 Porcentaje Atención Psicológica Baja

**Prueba 2**

|  |  |
| --- | --- |
| **Delito** | Violencia Intrafamiliar |
| **Nivel Consumo** | No Consumo |
| **Porcentaje Atención Psicológica** | 50% |

Tabla 32 Prueba 2 Lógica Difusa

Diagrama

Descripción generada automáticamenteIlustración 38 Porcentaje Atención Psicológica Media

**Prueba 3**

|  |  |
| --- | --- |
| **Delito** | Hurto y Porte Ilegal de Armas |
| **Nivel Consumo** | Problemático |
| **Porcentaje Atención Psicológica** | 78% |

Tabla 33 Prueba 3 Lógica Difusa

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Ilustración 39 Porcentaje Atención Psicológica Prioritaria

**Prueba 1**

|  |  |
| --- | --- |
| **Escolaridad Actual** | Tecnología |
| **Escolaridad Aprobada** | Tecnología |
| **Ocupación** | Trabaja |
| **Porcentaje Atención en Terapia Ocupacional** | 37% |

Tabla 38 Prueba 1 Lógica Difusa

Diagrama

Descripción generada automáticamente

Ilustración 40 Prueba 1 Lógica Difusa TO Bajo

**Prueba 2**

|  |  |
| --- | --- |
| **Escolaridad Actual** | Técnico |
| **Escolaridad Aprobada** | Ciclo V |
| **Ocupación** | Cesante |
| **Porcentaje Atención en Terapia Ocupacional** | 50% |

Tabla 40 Prueba 2 Lógica Difusa

Diagrama

Descripción generada automáticamente

Ilustración 41 Prueba 2 Lógica Difusa TO Medio

**Prueba 3**

|  |  |
| --- | --- |
| **Escolaridad Actual** | Primero |
| **Escolaridad Aprobada** | Primero |
| **Ocupación** | Cesante |
| **Porcentaje Atención en Terapia Ocupacional** | 82% |

Tabla 42 Prueba 3 Lógica Difusa

Gráfico

Descripción generada automáticamente

Ilustración 42 Prueba 3 Lógica Difusa TO Medio

Con la información de los jóvenes que ingresan a la corporación y ya almacenada en la base de datos se propone la generación del siguiente reporte

Interfaz de usuario gráfica, Aplicación, Word

Descripción generada automáticamente

El cual permitirá a las áreas transversales de la corporación definir el plan de trabajo teniendo en cuenta:

-Porcentaje de reincidencia

-Clasificarlo según su niel de consumo en el tratamiento que requiera.

-Porcentaje de refuerzo entre las áreas de Psicología y Terapia Ocupacional.

Mejorando con esto la calidad en el servicio ofertado a los adolescentes y Jóvenes que ingresan a la corporación.

# 4. Desarrollo específico de la contribución

Con el desarrollo de este piloto la corporación ha identificado cada una de las áreas de trabajo para reforzar y realizar personalización en la atención de los jóvenes que ingresan a la corporación.

Con los resultados observados se estarán por parte del equipo de trabajo evaluando el desarrollo de estrategias individuales para la atención de los adolescentes y Jóvenes involucrando entes de control y entidades estatales.

Mediante el uso de herramientas tecnológicas la corporación ha encontrado nuevas formas para almacenar la información esto les permitirá obtener en menor tiempo los reportes que deben entregar a los entes de control además con la disponibilidad de la información se les permite encontrar patrones que permitirá la reestructuración del plan de atención ofertado a cada joven garantizando la mejora continua en la calidad del servicio ofertado.

Se requiere por parte del equipo de trabajo compromiso en la actualización de los archivos que se vayan a procesar y cumplir con un estándar para que la información pueda ser procesada por la ETL, también tendrán que revisar cada una de las tablas de log en donde los datos si no pueden cruzar con las tablas de dimensiones puedan revisarlas y volver a generar la carga de data.

# 5. Conclusiones y trabajo futuro

## 5.1. Conclusiones

Se realizo la recopilación de la información histórica de la “Corporación (Alianza para el Desarrollo Ambiental Social y Económico Sostenible)” (Corpoadases), Con la data entregada se realizaron procesos de extracción, transformación y carga de información para generar insumo para el desarrollo de los modelos.

Con la información ya almacenada en un repositorio de datos se aplicaron técnicas de análisis de datos para responder cada una de las preguntas planteadas por el equipo de trabajo y con el uso de Power Bi visualizar los resultados, se han encontrado tendencias y se logrando con esto la mejora continua en la calidad del servicio ofertado a cada uno de los jóvenes que ingresan a la corporación.

Mediante el análisis de la información se ha podido determinar que: delito, nivel consumo de drogas y la ocupación son las variables que podrían predecir si un joven puede o no reincidir en un delito con una precisión del 75.68%., sin embargo, la variable tipología familiar no tiene ninguna relación con la reincidencia por lo cual se descarta del modelo.

El uso de la lógica difusa se aplicó para obtener los porcentajes de atención que requiera un adolescente o joven que ingresan con sanciones penales a la corporación, las variables: delito, nivel de consumo, escolaridad Actual, escolaridad Aprobada y ocupación fueron las que mediante configuración de reglas permitieron encontrar un porcentaje aproximado para la definicion de las áreas de refuerzo en psicología y terapia Ocupacional.

Se logro determinar que existe 3 grupos de jóvenes por su nivel de consumo, tipo de tratamiento y edad lo cual permite definir los pasos a seguir por parte de la corporación pues algunos de los jóvenes que presentan niveles de consumo problemático y abusivo deben ser intervenidos por las entidades de salud.

## 5.2. Líneas de trabajo futuro

Como trabajo futuro Implementar el desarrollo de este TFM no solo para el programa "Atención de la sanción Semicerrado, Externado media Jornada de Adolescentes y Jóvenes en Conflicto con la Ley Penal (Santander – Colombia)” también llevarlo a todos los programas de la corporación que se encuentran en la ciudad de Bucaramanga y en las ciudades de San gil y Barrancabermeja donde la corporación hoy en día está haciendo presencia con los mismos programas.

Es necesario continuar con la recopilación de la información para seguir alimentando los procesos y mejorar el resultado de los modelos.

Encontrar nuevas variables para alimentar los modelos predictivos y que logren determinar si puede o no un joven reincidir en un delito pues ahora solo se encontraron 3 variables: Delito, Nivel de Consumo y Ocupación.

Desarrollar nuevas reglas para definir con más certeza los niveles de atención que requieren los adolescentes y jóvenes que ingresan a la corporación, también involucrar las áreas de trabajo Social y Pedagogías para áreas de refuerzo.

Publicar las visualizaciones diseñadas en Power BI, para que todo el personal de trabajo de la corporación pueda acceder a estas en tiempo real y les permita tomar decisiones.

Implementar el reporte Formato De Ingreso, para que el equipo de trabajo de la corporación pueda tener un documento base con el cual darían inicio a la atención de los jóvenes y adolescentes que ingresan al programa y con esto mejorar la prestación del servicio, logrando buena calificación con los entre de control territorial lo cual podría significar para la corporación la posibilidad de crecer con más programas de atención.

# 6. Bibliografía

[1] Arias G, Walter L. (10/02/2018), *La frenología y sus implicancias: un poco de historia sobre un tema olvidado.*

https://www.scielo.cl/scielo.php?script=sci\_arttext&pid=S0717-92272018000100036

[2] Universidad Internacional de la Rioja, *Tipos de delincuentes, según Lombroso*

<https://www.unir.net/derecho/revista/tipos-de-delincuentes-criminales/>

[3] Rodríguez Puerta Alejandro, *Frenología: Historia y Frenólogos Destacados*

<https://www.lifeder.com/frenologia/>

[4] Diaz Raquel, (29/01/2020), *El polémico programa de ordenador que predice si eres un criminal por tu aspecto: "Tienes cara de malo"*

<https://www.elmundo.es/tecnologia/2020/06/29/5ef9243721efa04e418b459e.html>

[5] Powell Alekzan. (Madrid 24 Jun), *1.700 expertos condenan la IA que predice crímenes basándose en el rostro por ser socialmente injusta*

<https://www.europapress.es/portaltic/sector/noticia-1700-expertos-condenan-ia-predice-crimenes-basandose-rostro-ser-socialmente-injusta-20200624135813.html>

[6] Revista Semana (14/11/2013), *‘Key Crime’, un software que predice los crímenes*

<https://www.semana.com/mundo/articulo/key-crime-un-software-que-predice-previene-crimenes/364708-3/>

[7] Castiblanco Cindy, *En Bogotá se predecirán los crímenes con inteligencia artificial*

<https://bogota.gov.co/mi-ciudad/seguridad/modelo-de-prediccion-de-crimenes-en-bogota>

[8] Revista Semana (17/04/2019), *La Universidad Nacional ayudará a predecir crímenes en Bogotá*

<https://www.semana.com/educacion/articulo/la-universidad-nacional-y-la-alcaldia-de-bogota-trabajan-en-un-sistema-de-prediccion-de-crimenes/609920/>

[9] Toan C. Ong, 2017

[10] *Tecnicas-de-procesamiento-automatico-aplicadas-al-analisis-y-prediccion-del-consumo-de-drogas*

<https://www.cedia.edu.ec/es/proyectos-ganadores/cepra-xii/tecnicas-de-procesamiento-automatico-aplicadas-al-analisis-y-prediccion-del-consumo-de-drogas>

[11] Mario F. Triola, Correlación y Regresión.

[12] Universidad Internacional de La Rioja, Clustering: Agrupamiento o clasificación no supervisada

[13] Shakaran (03/12/2008), *Fundamentos de la lógica difusa*

<https://shakaran.net/blog/2008/12/fundamentos-de-la-logica-difusa-fuzzy-logic/>

[14] *AI en 3 minutos: Tipos de Machine Learning*

<https://www.sas.com/es_mx/insights/big-data/data-visualization.html>

[15] Castro Alondra (12/06/2017),*10 ventajas de Microsoft Power BI que unificarán datos de tu empresa*

<https://inbest.solutions/10-ventajas-de-microsoft-power-bi-que-unificaran-datos-de-tu-empresa/>

[16] Pedregosa et al. (Abril de 2021), *Scikit-learn: Machine Learning in Python version 0.24.2.*

https://scikit-learn.org/stable/modules/generated/ sklearn.metrics.silhouette\_score.html

[17] Zadeh, L. Fuzzy logic, IEEE Computer, 1:83, 1988.

[18] Martínez Heras, José. (18/09/2020), Random Forest (Bosque Aleatorio): combinando árboles

<https://www.iartificial.net/random-forest-bosque-aleatorio/>

[19] Mendoza Jaiver, (13/11/2020), Modelo SVM

<https://estadisticamente.com/que-es-un-modelo-svm/>

query\_aclcontrolm2016contra

# Anexos

## Anexo I. Título anexo I

## Anexo II. Título anexo II

1. ICBF (Instituto colombiano de Bienestar Familiar) [↑](#footnote-ref-1)