

|  |
| --- |
| **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**

**Palabras Clave:**

**Abstract**

**Keywords:**

**Índice de contenidos**

[1. Introducción 7](#_Toc52650792)

[1.1 Justificación 7](#_Toc52650795)

[1.2 Planteamiento del trabajo 8](#_Toc52650796)

[1.3 Estructura de la memoria 9](#_Toc52650797)

[2. Contexto y estado del arte 10](#_Toc52650798)

[3. Objetivos concretos y metodología de trabajo 15](#_Toc52650799)

[3.1. Objetivo general 15](#_Toc52650800)

[3.2. Objetivos específicos 15](#_Toc52650801)

[3.3. Metodología del trabajo 16](#_Toc52650802)

[4. Desarrollo específico de la contribución 16](#_Toc52650803)

[5. Conclusiones y trabajo futuro 17](#_Toc52650804)

[5.1. Conclusiones 17](#_Toc52650805)

[5.2. Líneas de trabajo futuro 17](#_Toc52650806)

[6. Bibliografía 17](#_Toc52650807)

[Anexos 19](#_Toc52650808)

[Anexo I. Título anexo I 19](#_Toc52650809)

[Anexo II. Título anexo II 19](#_Toc52650810)

**Índice de tablas**

[Tabla 1. Título de la tabla 11](#_Toc8213227)

**Índice de figuras**

[Figura 1. Ramas del Data Mining [2] 10](#_Toc8213228)

[Figura 2. Figura 2. Los Tipos de Aprendizaje [7]](#_Toc8213228)………………………………………………...12

# Introducción

## El siguiente TFM (Trabajo Fin de Master) 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” y mediante la implementación de herramientas y análisis de información que puedan detectar y realizar acciones que midan el impacto en la gestión realizada en los adolescentes y jóvenes y la relación que existe entre el tipo de delito, situación socio-familiar, nivel de estudio entre otras variables y con los resultados que se obtengan ayuden a la corporación a tomar acciones y desarrollar estrategias entre las áreas transversales (Terapia Ocupacional, Psicología, Nutrición, Trabajo Social, Artes, Cultura y Deportes) para trabajar con los Adolescentes y Jóvenes que ingresan al programa y como institución agregar valor al trabajo que realizan mejorando 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, y con el resultado del análisis de la información 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 poder describir, predecir y generar visualizaciones que responsan las siguientes preguntas:

- Cuál es el Motivo de Ingreso al programa de Responsabilidad Penal.[[1]](#footnote-1)

- Cuál es la Edad de ingreso al programa de Responsabilidad Penal. 1

- Cuánto es el Numero de reincidencias en el trimestre y la proyección semestral.

- Cuál es la tipología Familiar de los Adolescentes y Jóvenes que Ingresan al programa de Responsabilidad Penal. 1

- Nivel de Consumo de los Adolescentes y Jóvenes que Ingresan al programa de Responsabilidad Penal.1

- Nivel de Escolaridad de los Adolescentes y Jóvenes que han ingresado al programa de Responsabilidad Penal.1

- Número de desertores del programa de Responsabilidad Penal.1

- Comparación el estrato Socio-Familiar Vs Delito.

- Comparación el nivel escolaridad Ingreso Vs nivel escolaridad Egreso.

- Porcentaje de Adolescentes y Jóvenes que pasan de trabajo Informal a trabajo Formal.

- Cual es la probabilidad de reincidencias de los adolescentes y jóvenes que Consumen drogas.

- Cual es la relación del Motivo de Ingreso al programa con el Consumo de drogas.

- Cual es la relación del Nivel de escolaridad con Motivo de ingreso.

- Cual es la probabilidad de reincidencia de un Adolescente o Joven que Egresa del programa con el mismo nivel de estudio con el que Ingresa.

- Cual es la relación entre la tipología familiar y el consumo de drogas.

- Cual es la probabilidad de deserción de un Adolescente o Joven con el tipo de familia disfuncional.

¿Qué es lo que se propone?

**Lo cual le permitirá a la Corporación:**

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

Cuantificar el número de reincidencias.

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.

Identificar los adolescentes y jóvenes que desertan del programa.

Comparar el estrato socio-familiar y el delito

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.

Reestructurar su plan de acción buscando la mejora constante 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 mas populares en el mercado.

En el capitulo 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 Capitulo 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.

# Contexto y estado del arte

Para el desarrollo del TFM se utilizará los algoritmos de Aprendizaje Supervisados y se desarrollan visualizaciones que permitan tener una descripción de las variables que le permita a la Corporación Corpoadases (Alianza para el Desarrollo Ambiental Social y Económico Sostenible)” responder las preguntas con las que se inicia este proyecto además mediante la data procesadas se pueda entender el por Qué paso , Como paso , porque Paso y lo que puede pasar según el resultado del análisis de la información y que le permitirá tomar decisiones y acciones para la atención de los Adolescentes y Jóvenes que ingresan al programa “Atención de la sanción semicerrado, Externado media Jornada de Jóvenes en conflicto con la Ley Penal (Santander-Colombia)”.

Actualmente existen aplicaciones que usan análisis de datos para predecir crímenes, consumo de drogas, reincidencia de un delito sin embargo en la ciudad de Bucaramanga, (Santander – Colombia) las instituciones que trabajan con Jóvenes no cuentan con herramientas tecnológicas para el análisis de su información, con este Piloto Corpoadases se convierte en una institución pionera en uso de análisis de datos puedan 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.

**2.1 Analítica**

La analítica es un campo incluyente y multidimensional que utiliza matemáticas, estadística, modelos predictivos y técnicas de aprendizaje basado en máquina para hallar patrones y conocimientos significativos en datos grabados. [1] permitiendo mediante el análisis de la información responder a las preguntas del negocio y la resolución de problemas para con esto optimizar recursos, mejorar la calidad del servicio, encontrar riesgo que puedan afectar a la organización y ayude a la toma de decisiones.

**2.2 Machine Learning**

Es una derivación de inteligencia artificial que crea sistemas que aprenden de manera automatizada, es decir, identificar patrones complejos en millones de datos, predecir comportamientos todo mediante un algoritmo y que además de todo son capaces de mejorarse de manera independiente con el tiempo.[4]

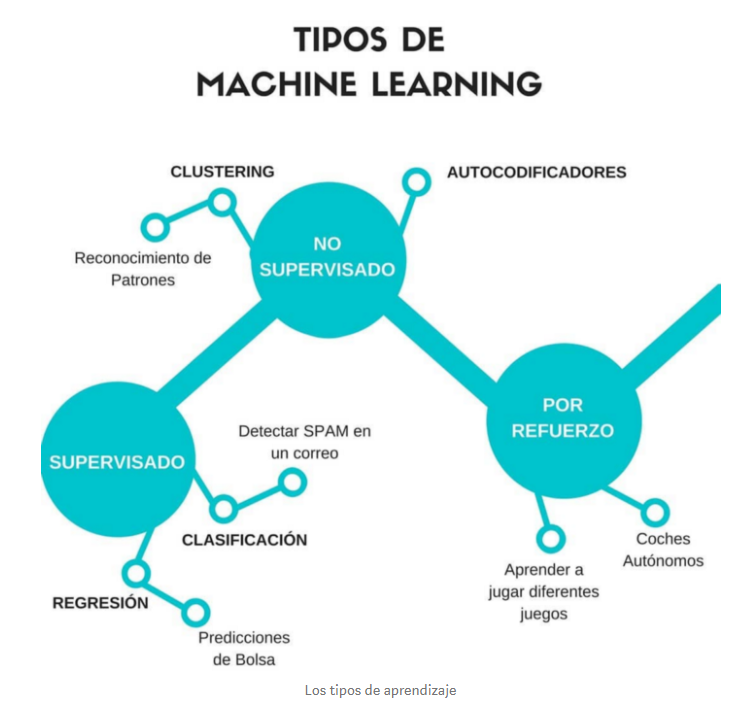


Figura 1. Los Tipos de Aprendizaje [7]

**El Machine Learning se puede diferenciar en tres Categorías:**

Aprendizaje supervisado (Supervised Learning)

Aprendizaje no supervisado (Supervised Leartning)

Reforzamiento de Aprendizaje (Reinforcement Learning)

**2.2.2.1 Aprendizaje Supervisados**

En el aprendizaje supervisado, los algoritmos trabajan con datos “etiquetados” (labeled data), intentado encontrar una función que, dadas las variables de entrada (input data), les asigne la etiqueta de salida adecuada. El algoritmo se entrena con un “histórico” de datos y así “aprende” a asignar la etiqueta de salida adecuada a un nuevo valor, es decir, predice el valor de salida.[5]

**2.4 Visualización de Datos**

La visualización de datos describe la presentación de la información abstracta en forma gráfica que permite detectar patrones, tendencias y correlaciones que de otro modo podrían pasar desapercibidos en los informes, tablas o hojas de cálculo tradicionales.

**2.4.2 Herramientas para la Visualizacion de Datos**

Las 7 herramientas de visualización de datos más populares son:

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

**2.4.2.1 Tableu**

[Tableau](https://www.tableau.com/es-es) es una de las herramientas de visualización de datos más populares y completas. Su interfaz permite generar visualizaciones sobre grandes volúmenes de datos. Cuenta con un funcionamiento sencillo, además de poder personalizar las informaciones. Tiene una versión gratuita.[11]

**2.4.2.2 Qlik**

[Qlik](http://www.qlik.com/es-es) es el mayor competidor de Tableau, la empresa QlikTech ofrece dos productos. QlikView es el más popular de los dos, destaca por su sencillez, con una configuración muy personalizable, permitiendo a los usuarios tomar decisiones basadas en datos. La aplicación viene acompañada de QlikSense, una versión más sencilla para crear visualizaciones flexibles e interactivas.[11].

**2.4.2.3 Ploty**

[Plotly](https://plot.ly/feed/) este software flexible ofrece visualizaciones más complejas y sofisticadas. Está integrado con lenguajes de programación como Python, JavaScript o Matlab.[11]

**2.4.2.4 Carto**

[CartoDB](https://cartodb.com/) es un servicio que almacena los datos geográficos en una base de datos en la “nube” y ofrece una variedad de herramientas para acceder, analizar y visualizar los datos. La plataforma está construida usando una base de datos [PostGIS](http://postgis.net/) de código abierto y se ejecuta en una infraestructura a gran escala de Amazon Web Services.

**2.4.2.5 Power BI**

Esta herramienta permite llevar las grillas de Excel y tablas dinámicas a un nuevo nivel. El mayor atractivo de [Power Bi](https://powerbi.microsoft.com/en-us/) es que permite realizar presentaciones interactivas y ajustar las visualizaciones para obtener un mayor detalle según la pregunta que queremos que nuestros datos respondan con la opción “drill in”-lo cual es una gran ventaja para las clásicas preguntas imprevistas en presentaciones..[10]

**2.4.2.6 DataWrapper**

La popularidad de [DataWrapper](https://www.datawrapper.de/) va en aumento. Al igual que las otras herramientas, su interfaz es sencilla y clara. Esta herramienta ofrece la creación de gráficas e infografías de forma intuitiva.[11]

**2.4.2.7 DataStudio (Google)**

Es una herramienta gratuita y sencilla de utilizar de Google para desarrollo de visualizaciones y creación de cuadros de mando El objetivo de la herramienta es permitir un análisis de los datos de manera visual de forma que sea más fácil e inmediato obtener resultados.[12] [Se accede a través del navegador](https://datastudio.google.com/navigation/reporting) sin necesidad de instalar software adicional.

**2.4 Aplicaciones Actuales de Predicción**

En este apartado se presentarán algunas aplicaciones que usan métodos de predicción.

**Key Crime**

Un software que predice los crímenes.

Los ladrones son víctimas de la tecnología. Mario Venturi, sargento de policía de Milán (Italia) inventó en 2007 un sofisticado programa informático, llamado “Key Crime”, destinado a las agencias policiales. Sirve para predecir, prevenir y combatir los delitos en serie.  Atracos, robos en almacenes, en bancos y en las calles, entre otros, son el blanco de la persecución.

“Key Crime”, analiza por computador información sobre comisión de delitos, y se enfatiza en el perfil del delincuente; a partir de esto, construye una base de datos, de la cual obtiene, tras un refinado y matemático cruce de información, claves para calcular y predecir la eventual realización del crimen. En seguida, la policía envía patrullas a los ‘puntos calientes’ o lugares de delincuencia local más sensibles, y si la precisión no falla, sorprende in fraganti a los malhechores y los arresta.[13]

**En Bogotá se predecirán los crímenes con inteligencia artificial**

Proyecto que se desarrolla en Bogotá desde el año 2019 en conjunto con la secretaria Distrital de seguridad y el departamento de matemáticas de la universidad Nacional, su principal tarea en este proyecto es tratar de predecir el crimen en Bogotá mediante el uso de la inteligencia Artificial.[14]

**Técnicas de procesamiento automático aplicadas al análisis y predicción del**

**consumo de drogas.**

El consumo de alcohol y otras drogas sigue constituyendo uno de los principales problemas de salud pública por sus consecuencias para el consumidor y para su entorno, incluyendo problemas de salud, violencia intrafamiliar y pérdida de productividad en Ecuador. Sin embargo, los programas de prevención y/o las políticas públicas de actuación eficaces siguen siendo un reto de acuerdo con los datos ofrecidos por los diferentes Observatorios Nacionales de Drogas, la Comisión Nacional para el Desarrollo y Vida Sin Drogas (DEVIDA), y la Comisión Interamericana para el Control del Abuso de Drogas (CICAD).

El objetivo de este proyecto es aplicar algoritmos emergentes de inteligencia artificial a 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, y que sirva como apoyo al diseño de programas de prevención y políticas públicas en drogas, sin perjuicio de las competencias que otras instituciones de otros territorios puedan ejercer sobre esta materia. [15]

**Prisma**

El programa busca apoyar a los fiscales al momento de solicitar prisión preventiva en cárcel para los investigados. El propósito del programa es que los fiscales cuenten con más información para sustentar esa solicitud o evitar que personas que no representan peligro estén tras las rejas.[16]

# 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 la “Corporación Corpoadases (Alianza para el Desarrollo Ambiental Social y Económico Sostenible)” (Corpoadases) para ser almacenada en un modelo de datos.

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

- Relacionar el tipo de delito, situación socio-familiar, nivel de estudio entre otras variables que se encuentren con la reincidencia, deserciones, y el consumo de drogas de los Jóvenes.

- Implementar un modelo que permita predecir de acuerdo a las variables situación socio-familiar, nivel de estudio, consumo de drogas si un Joven puede reincidir en un delito.

## 3.3. Metodología del trabajo

# 4. Desarrollo específico de la contribución

# 5. Conclusiones y trabajo futuro

## 5.1. Conclusiones

## 5.2. Líneas de trabajo futuro

# 6. Bibliografía

[1] <<Analítica Qué es y por qué es importante>> [En Línea].

Available: <https://www.sas.com/es_co/insights/analytics/what-is-analytics.html>.

[2] T. Camilo, A. Oscar, Modelos de Analítica Aplicados a los Negocios, Universidad de la Sabana, 2020.

[3] <<Análisis predictivo, descriptivo y prescriptivo>> [En Línea].

Available:<https://www.analisiscientifico.com/single-post/An%C3%A1lisis-predictivo-descriptivo-y-prescriptivo>.

[4]<<que-es-machine-Learning-y-para-que-sirve>> [En Línea].

Available: <https://www.inbest.cloud/comunidad/que-es-machine-learning-y-para-que-sirve>

[5] S. Osvaldo, A Very Brief Introduction to Machine Learning with Applications to Communication Systems, IEEE, 2018.

[6]<<que-algoritmo-elegir-en-ml-aprendizaje>>[EnLínea].

Available: <https://empresas.blogthinkbig.com/que-algoritmo-elegir-en-ml-aprendizaje/>.

[7]<<algoritmos-del-machine-learning>> [En Línea].

Available: <https://www.apd.es/algoritmos-del-machine-learning/>.

[8] <<AI en 3 minutos: Tipos de Machine Learning>>[EnLínea].

Available: <https://medium.com/ai-learners/ai-en-3-minutos-tipos-de-machine-learning-945b708ac78>.

[9] <<Visualizacion>>[EnLínea].

Available: https://www.tecnologias-informacion.com/visualizacion.html

[10] <<6-herramientas-gratuitas-para-analizar-y-visualizar-datos>>[EnLínea].

Available: <https://blogs.iadb.org/conocimiento-abierto/es/6-herramientas-gratuitas-para-analizar-y-visualizar-datos/>.

[11] <<herramientas-para-la-visualizacion-de-datos>>[EnLínea].

Available: https://blogthinkbig.com/herramientas-para-la-visualizacion-de-datos.

[12] <<Google Data Studio>>[EnLínea].

Available: <https://www.metriplica.com/herramientas-analitica-web/google-data-studio>.

[13] <<key-crime-un-software-que-predice-previene-crimenes>>[EnLínea].

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

[14] <<modelo-de-prediccion-de-crimenes-en-bogota>>[EnLínea].

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

[15] <<tecnicas-de-procesamiento-automatico-aplicadas-al-analisis-y-prediccion-del-consumo-de-drogas>>[EnLínea].

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

[16] <<prisma-el-programa-de-la-fiscalia-para-predecir-la-reincidencia-criminal/>>[EnLínea].

Available: https://www.elespectador.com/noticias/judicial/prisma-el-programa-de-la-fiscalia-para-predecir-la-reincidencia-criminal/

# Anexos

## Anexo I. Título anexo I

## Anexo II. Título anexo II

1. Responsabilidad Penal hace Referencia al programa Atención de la sanción Semicerrado, Externado media Jornada de Adolescentes y Jóvenes en Conflicto con la Ley Penal (Santander – Colombia). [↑](#footnote-ref-1)