



# **Evaluación de algoritmos de aprendizaje y método Wenner en tomografía eléctrica para la detección automática de fosas simuladas**

**Nataly Ipuz Paloma**

Universidad Antonio Nariño  
Facultad de Ingeniería Mecánica, Electrónica y Biomédica  
Bogotá, Colombia  
Año 2021

# **Evaluación de algoritmos de aprendizaje y método Wenner en tomografía eléctrica para la detección automática de fosas simuladas**

**Nataly Ipuz Paloma**

Proyecto de grado presentado como requisito parcial para optar al título de:  
**Especialista Instrumentación Industrial y Biomédica**

Director:

PhD. Ingeniero Mario Enrique Duarte

Co-Directora:

Docente Alejandra Juliette Baena

Universidad Antonio Nariño

Facultad de Ingeniería Mecánica, Electrónica y Biomédica

Bogotá, Colombia

Año 2021

*Puede que no podamos controlar lo que nos pasa, pero podemos controlar lo que ocurre en nuestro interior (Benjamín Franklin)*

*Este libro está dedicado a Dios, por ser la luz que guía mi camino a pesar de las adversidades. A mi familia e hija que son mi motivación, motor y apoyo incondicional.*

## **Agradecimientos**

El tiempo de Dios es perfecto. No ha sido un año fácil, la pandemia nos mostró lo vulnerables que somos y nos exigió medir nuestra capacidad para superar las adversidades y obtener lo mejor de ellas.

El modelo de educación virtual me permitió no solo fortalecer mis conocimientos, sino conocer un grupo de docentes comprometidos que aportaron sabiduría y experiencia a mi vida profesional y académica.

Sin embargo, quiero destacar la labor y hacer un reconocimiento especial a mis directores de trabajo de grado, se trata del Ing. Mario Enrique Duarte González, PhD y la profesora Alejandra Baena, por su entrega, compromiso y apoyo incondicional, por mostrar y compartir su conocimiento con generosidad, amor, pero en especial, por inspirarme a ser mejor profesional cada día.

Qué Dios ilumine y siga llenando de bendiciones sus vidas. Igualmente, agradezco infinitamente a mis padres, hermanas e hija por estar siempre ahí, siendo mi apoyo incondicional, por alentarme a seguir y no desfallecer, Dios en su infinita sabiduría me ha premiado con una gran y hermosa familia.

## Resumen

La desaparición forzada en Colombia constituye un crimen de lesa humanidad que viola los derechos humanos y quebranta el bienestar de las familias. En el país más de 80.500 personas han sido víctimas de este hecho que solo genera zozobra, incertidumbre y temor.

Aunque el Congreso de la República aprobó la Ley 1408 para redimir a las víctimas y familiares de la desaparición forzada, además, de contribuir en la búsqueda de los desaparecidos; los resultados no han sido efectivos. El Grupo Investigativo de Justicia y Paz de la Dirección de Investigación Criminal ha fallado en la implementación de las técnicas que demandan tiempo, son costosas y poco acertadas.

En ese sentido y en aras de contribuir en la mitigación del sufrimiento de las víctimas, este trabajo de investigación experimental enmarcado en la geofísica forense, presenta la implementación del Método Wenner y la validación de tres modelos de Machine Learning.

A partir de los modelos, se implementaron algoritmos de aprendizaje automático como Bagged Trees; Boosted Trees, Rusboosted y Subspace KNN, los cuales se aplicaron con etiquetas y sin etiquetas. En cada uno de los análisis, se utilizaron herramientas de aprendizaje de clasificación de Matlab, en algunos casos, los datos se procesaron mediante la búsqueda y el llenado de valores atípicos o normalizados.

No obstante, es el modelo Bageed Trees, el que registró los mayores índices de precisión y rendimiento. De acuerdo a la matriz de confusión, los porcentajes oscilaron entre el 59 hasta 75.8% en comparación con los otros algoritmos.

Estos resultados se convierten en una herramienta ideal para contribuir en la detección automática de fosas simuladas en el país y una esperanza para miles de familiares que aún viven con el dolor y la ausencia en sus corazones.

**Palabras clave:** Resistividad eléctrica, Fosas simuladas, método de Wenner, algoritmos de aprendizaje, base de datos, modelo de validación.

## **Abstract**

Forced disappearance in Colombia constitutes a crime against humanity that violates human rights and undermines the well-being of families. In the country more than 80,500 people have been victims of this fact that only generates anxiety, uncertainty and fear.

Although the Congress of the Republic approved Law 1408 to redeem the victims and relatives of the forced disappearance, in addition, to contribute in the search for the disappeared; the results have not been effective. The Justice and Peace Investigative Group of the Criminal Investigation Directorate has failed to implement the techniques that are time-consuming, expensive, and not very successful.

In this sense and in order to contribute to the mitigation of the suffering of the victims, this experimental research work framed in forensic geophysics, presents the implementation of the Wenner Method and the validation of three Machine Learning models.

From the models, machine learning algorithms such as Bagged Trees were implemented; Boosted Trees, Rusboosted, and Subspace KNN, which were applied with and without tags. In each of the analyzes, Matlab classification learning tools were used, in some cases, the data was processed by searching and filling in outliers or normalized values.

However, it is the Bageed Trees model that had the highest accuracy and performance rates. According to the confusion matrix, the percentages ranged from 59 to 75.8%

The results of the investigation become an ideal tool to contribute to the automatic detection of simulated graves in the country and a hope for thousands of relatives who still live with pain and absence in their hearts.

**Keywords:** Electrical resistivity, simulated pits, Wenner method, learning algorithms,



# Contenido

<b>Lista de figuras.....</b>	<b>XI</b>
<b>Lista de tablas .....</b>	<b>XII</b>
<b>Lista de Ecuaciones.....</b>	<b>XIII</b>
<b>Introducción .....</b>	<b>1</b>
<b>2. Marco teórico.....</b>	<b>7</b>
2.1 Desaparición forzada.....	7
2.2 Drama familiar .....	8
2.1.1 Identificación de cuerpos .....	9
2.1.2 Herramientas de búsqueda.....	10
2.2 Disciplinas del conocimiento .....	12
2.2.1 Métodos geofísicos .....	12
2.3 Método de detección Wenner.....	13
2.4 Metodología Machine Learning .....	15
2.5 Modelos de clasificación.....	16
2.5.1 Support Vector Machines.....	16
2.5.2 KNN .....	17
2.7 Métricas de evaluación .....	18
2.8 Base de datos.....	21
2.8.1 Minería de datos .....	22
2.8.2 Modelo de referencia CRISP-DM .....	22
<b>3. Metodología CRISP – DM aplicada a la detección de fosas .....</b>	<b>26</b>
○ 3.1 Fase 1: Entendimiento del proceso.....	26
○ 3.2 Fase 2: Entendimiento de los datos .....	26
▪ 3.2.1 Análisis a los datos y su método de recolección.....	26
▪ 3.2.3 Bases de datos .....	28
▪ 3.2.4 Creación de base de datos .....	30
○ 3.3 Preparación de los datos.....	33
○ 3.4 Modelado.....	36
○ 3.5 Evaluación del Modelo .....	37
<b>4. Resultados.....</b>	<b>38</b>
○ 4.1 Experimento 1: Impacto de la normalización en el tratamiento de los datos ...	38
○ 4.2 Experimento 2: Impacto de la corrección de valores atípicos en el procesamiento de datos. ....	40
○ 4.3 Experimento 3 Impacto cuando hay una tumba de control (con etiqueta 0). .	41
○ 4.4 Experimento 4: Impacto en los algoritmos de Machine Learning se confundan cuando se agregan las etiquetas 2, 4,5 y una tumba de control (Grave 3). ....	43
○ 4.5 Experimento 5: Impacto a través del tiempo .....	45
<b>5. Conclusiones y perspectivas futuras. ....</b>	<b>49</b>

---

- **Anexos ..... 51**
- **Bibliografía ..... 53**

## Lista de figuras

<b>Figura 1.</b> Método Wenner.....	14
<b>Figura 2 -</b> Matriz de Confusión Binaria.....	19
<b>Figura 3.</b> Ciclo de vida del modelo CRISP - DM.....	22
<b>Figura 4.</b> Diagrama de disposición de fosas.....	27
<b>Figura 5.</b> Implementación método Wenner.....	28
<b>Figura 6.</b> Diseño experimental.....	29
<b>Figura 7.</b> Medición equipo de tomografía.....	31
<b>Figura 8.</b> Ubicación de las muestras de fosas en un mapa 2D.....	32
<b>Figura 9.</b> Extracción de predictores para el entrenamiento de los modelos de machine learning.....	33
<b>Figura 10..</b> Ejemplo de construcción y organización de los datos.....	33
<b>Figura 11.</b> Matriz de confusión Rusboosted Trees sin normalización.....	38
<b>Figura 12.</b> Matriz de confusión Rusboosted Trees con normalización.....	39
<b>Figura 13.</b> Matriz de confusión – Bagged Trees sin corrección de datos atípicos.....	40
<b>Figura 14.</b> Matriz de confusión - Bagged Trees con llenado de datos atípicos.....	41
<b>Figura 15.</b> Matriz de confusión – Bagged Tree sin fosa de control.....	41
<b>Figura 16.</b> Matriz de confusión - Bagged Tree con fosa de control.....	42
<b>Figura 17.</b> Matriz de confusión Subspace KNN sin fosa de control.....	43
<b>Figura 18.</b> Matriz de confusión Boostd Trees con fosa de control con etiqueta 3.....	44
<b>Figura 19.</b> Matriz de confusión modelo de predicción mes de febrero 2020.....	45
<b>Figura 20.</b> Matriz de confusión modelo de predicción mes de Octubre 2020.....	45
<b>Figura 21.</b> Matriz de confusión modelo de predicción mes de Diciembre 2020.....	46
<b>Figura 22.</b> Matriz de confusión modelo de predicción mes de Febrero del 2021.....	47

## Lista de tablas

<b>Tabla 1.</b> Designación de fosas .....	27
<b>Tabla 2.</b> Datos obtenidos método Wenner .....	29
<b>Tabla 3.</b> Modelo de base de datos .....	34

## Lista de Ecuaciones

<b>Ecuación 1</b> - Fórmula para calcular la Exactitud.....	20
<b>Ecuación 2</b> - Cálculo de Precisión .....	20
<b>Ecuación 3</b> - Fórmula para calcular la Sensibilidad .....	21
<b>Ecuación 4</b> - Fórmula para calcular la especificidad.....	23
<b>Ecuación 5</b> - Ecuación de profundidad dada por el fabricante.....	35



# Introducción

En Colombia el flagelo de la desaparición forzada ha sido uno de los hechos victimizantes que más ha afectado a las familias. Las numerosas cifras de desaparecidos que hoy ascienden más de 80.000 personas, sin distinción de género, edad y condición social, siguen creciendo con el paso de los años.

Aunque el Gobierno ha liderado acciones para resarcir la pérdida, no son suficientes para una familia que clama justicia y anhela despedir a su ser querido.

Durante muchos años la búsqueda de cuerpos enterrados fue una tarea bastante compleja para los investigadores forenses, ya que las técnicas implementadas eran directamente invasivas y se corría el riesgo de la pérdida o destrucción de pruebas incriminatorias, consideradas de mucha importancia en una causa penal.

En los años noventa se implementó el uso de técnicas que permitieron fortalecer la búsqueda de fosas a partir de la interpretación de anomalías en los datos obtenidos mediante la aplicación geofísica de alta resolución. Y actualmente, se aplica la geofísica incluso se estudió la profundidad de la resistividad del suelo permitiendo identificar la fosa clandestina.

En ese sentido y con el ánimo de contribuir en la mitigación de la incertidumbre de las familias, se presenta un estudio para evaluar el uso de algoritmos de aprendizaje y método Wenner en tomografía eléctrica para la detección automática de fosas simuladas.

Para ello, se organizarán perfiles geoelectricos de fosas simuladas obtenidos por un tomógrafo en configuración Wenner; se procesarán los perfiles para su uso en máquinas

de aprendizaje; se aplicarán algoritmos de clasificación y se comparará el desempeño de los diferentes algoritmos de aprendizaje utilizando métricas de desempeño.

### **Planteamiento del problema**

Los métodos geofísicos actuales existentes para el hallazgo de fosas comunes son de gran complejidad, ya que requieren de personal capacitado en el manejo de la tecnología aplicada específicamente en técnicas de exploración para la ubicación de entierros clandestinos.

Hoy, en Colombia se emplean dos métodos para la ubicación de fosas común. El primer método implementado y más frecuente para la búsqueda de fosas clandestinas basa su funcionamiento en dos fases; la primera fase, realiza indagaciones e investigaciones preliminares efectuando una recolección y comparación de información de los desaparecidos con el fin, de tener un estudio claro y preciso que permita cotejar los datos existentes.

La segunda fase, consiste en el traslado al lugar donde se encuentran los posibles restos para ser exhumados, se determina un área aproximada y se realizan excavaciones, luego, empleando una sonda de acero inoxidable se hace una exploración por el área delimitada.

El segundo método, radica en la implementación de dispositivos de alta tecnología con el que se mide la variabilidad de la resistencia en el subsuelo cuyos resultados solo pueden ser interpretado por un geofísico, dificultando el hallazgo de las personas desaparecidas.

Las investigaciones judiciales que utilizan los métodos anteriormente mencionado, muchas veces son ineficaces ya que no se cuenta con una metodología con los protocolos adecuados. Labor que se vuelve imperiosa y dolorosa para las familias que piden respuesta y reparación a los daños ocasionados por décadas de violencia.

Surgiendo de esta manera, la necesidad de implementar a partir de los métodos geofísicos como el método Wenner, cuyo funcionamiento se basa en la circulación de corriente la cual ingresa por el electrodo A y circula hasta el electrodo B, se realiza una diferencia de



potencial entre el electrodo M y N, el dato de resistividad esta dado en el medio de los electrodos M y N. Los datos obtenidos son almacenados y organizados en una imagen de resistividad, a diferentes niveles de profundidad. A partir de esta imagen en tres dimensiones, se pretende utilizar algoritmos de aprendizaje para predecir la existencia de una fosa simulada.

Teniendo en cuenta las dificultades con las que cuenta el Gobierno Nacional para esclarecer y dar reparación a las víctimas del conflicto, se plantea la evaluación de algoritmos de aprendizaje y Método Wenner en tomografía eléctrica para la detección automática de fosas simuladas, con el que se planea sufragar a resolver la problemática social que vive el país.

### **Justificación**

Es importante implementar metodologías que permitan acelerar el proceso de búsqueda de fosas comunes. Informes realizados por el Observatorio de Memoria y Conflicto (OMC) del Centro Nacional de Memoria Histórica (CNMH) datan de 80.472 víctimas directas registradas durante el período 1958-2018, de igual manera existen 125,577 víctimas indirectas de desaparición forzada en el Registro Único de Víctimas (datos RUV a 1 de abril de 2019).

Desde la creación del Grupo de Investigativo de Justicia y Paz se ha realizado la recuperación de mil ciento cincuenta y siete cadáveres (1.157) de 976 fosas comunes encontradas hasta la fecha, datos desalentadores ante la cifra que se tiene de desapariciones forzadas de víctimas directa. Sólo 1 de 10 comisiones judiciales tiene éxito (Director Nacional de Justicia Transicional Carlos Villamil, Caracol radio, julio 2015). Dando así reparación y no olvido a todas las familias afectadas por el flagelo de la violencia, aplicando nuevas tecnologías para la aplicación de geofísica forense.

### **Objetivos**

#### **Objetivo general**

- Evaluar el uso de algoritmos de aprendizaje y método wenner en tomografía eléctrica para la detección automática de fosas simuladas.

**Objetivos específicos**

- Organizar los perfiles geoelectricos de fosas simuladas obtenidos por un tomógrafo en configuración wenner.
  - Procesar los perfiles geoelectricos para su utilización en las máquinas de aprendizaje.
  - Aplicar diferentes algoritmos de clasificación para la detección automática de fosas simuladas.
  - Comparar el desempeño de los diferentes algoritmos de aprendizaje utilizando métricas de desempeño.
- 
- **Alcances y limitaciones**

La investigación examinó la fase de Preparación de Datos tomando como referencia la metodología CRISP-DM; la investigación sólo se enfocó en aplicaciones de Minería de Datos en el área de ciencia forense, basado en los datos obtenidos en el laboratorio de la Universidad Antonio Nariño a través del Método de Wenner.

En este proyecto se desarrolló un algoritmo de aprendizaje ML para la detección de fosas común a partir de la información extraída por el método Wenner, utilizando los datos de impedancia circundante para la ubicación y predicción automática de la ubicación de fosas común.

Una de las limitantes presentadas en el desarrollo de la investigación está relacionada con el trabajo en fosas simuladas; teniendo en cuenta la geografía, precipitaciones, intensidad, radiación solar, temperatura, sistemas de vientos y humedad atmosférica del país imposibilita hallar un patrón base con las fosas simuladas.

- **Estructura**

El presente documento está dividido en cinco capítulos. Capítulo Uno, presenta un contexto amplio sobre la desaparición en Colombia, los antecedentes en aplicación de métodos geofísicos para la detección de fosas.

El capítulo Dos, esboza los referentes teóricos relacionados con métodos geofísicos implementados, Minería de Datos, Preparación de Datos, algoritmos de predicción y modelo de validación.

Entre tanto, Capítulo tres, se aplica la metodología CRISP –DM planteada para el desarrollo del proyecto, teniendo en cuenta las fases de proceso se realiza un análisis exhaustivo el cual se profundiza en el capítulo 1 con el fin de entender el proyecto, seguidamente se procede a realizar el entendimiento de los datos aplicando el método wenner y de los perfiles creados a partir de las fosas 6 implementadas, se realiza la aplicación de los parámetros de profundidad para la elaboración del mapa 2D donde se ubican las tomas de resistividad. En la preparación de los datos se toma los 4 valores más cercanos y se etiqueta de tal manera que le permita algoritmo predecir si existe fosa, en la etapa de modelamiento se realiza emplean las herramientas de machine learnig para realizar el entrenamiento de los modelos Bagged Trees , Ruboosted Trees y Subspace KNN , Por último en la etapa de evaluación del modelo se plantean 5 experimentos cuyo rendimiento se basa en las métricas como sensibilidad , especificidad , precisión y matriz de confusión .En el Capítulo Cuarto, se muestra los resultados obtenidos para este caso los mejores resultados fueron los obtenidos por el algoritmo Bagged Trees Exactitud = 75.8% Precision = 75.6% sensibilidad =40% especificidad=93.6%. Por ultimo se concluye los avances de la investigacion.

Finalmente, este trabajo cumple con su propósito de consolidarse en un aporte académico significativo, que no sólo contribuye en la disminución del sufrimiento de los familiares de las personas desaparecidas, sino que, demuestra la importancia de la implementación de algoritmos de aprendizaje automático en la detección segura de fosas comunes.

Los resultados de los estudios y pruebas aplicadas a cada uno de los modelos, son la garantía de la efectividad del mismo, abriendo una posibilidad para su uso bajo las variables comprobadas de precisión y rendimiento comparadas a través de una Matriz de Confusión para determinar el desempeño de los algoritmos implementados, lo que permitirá seleccionar el modelo más idóneo.

Es así como desde la academia se contribuye a mitigar las repercusiones psicológicas y sociales de un lamentable flagelo que viven las familias de los miles de desaparecidos que se han registrado en el país.

## **2. Marco teórico**

El propósito de este capítulo es presentar los fundamentos académicos que validan el desarrollo de la presente tesis de investigación resultado de una completa búsqueda de referentes teóricos.

### **2.1 Desaparición forzada**

La desaparición forzada de personas sigue siendo un atroz, lamentable y cruel fenómeno que vulnera los derechos humanos, (Ambos, y otros, 2009) y con el paso de los años, se ha convertido en parte de la cotidianidad en América Latina.

Esta repudiable experiencia surgió en los años sesenta, específicamente en Guatemala, en donde se ocultaba información sobre el paradero de las víctimas por parte de sus agentes.

Aunque su origen a nivel mundial se remonta a Alemania, específicamente, en el año 1941 con el Decreto Noche y Niebla de Hitler, donde miles de personas desaparecieron de manera clandestina por las autoridades alemanas, lo que significó un precedente histórico relacionado con la violación de los derechos humanos (Basaure Miranda, 2018).

Sin embargo, en América Latina, las desapariciones no son un rasgo exclusivo de las dictaduras militares del siglo XX, sino que ha estado presente en los territorios; por ejemplo, en países como México, Colombia y Perú, con gobiernos civiles electos no fue la excepción y han sido escenarios de este flagelo (Molina Theissen, 1996).

De acuerdo con los historiadores, dos décadas después, la cuestionada práctica desarrollada en Guatemala, se extendió a El Salvador, Chile, Uruguay, Argentina, Brasil, Colombia, Perú, Honduras, Bolivia, Haití y México (Ambos, y otros, 2009).

En (Gómez Camacho, 2017) la Convención Interamericana sobre Desaparición Forzada de Personas consideran este hecho, como una privación de la libertad a una o más personas, cometida por agentes del Estado, personas o grupos en su representación que se niegan a informar sobre el paradero de la persona y aceptar la privación de la libertad.

## 2.2 Drama familiar

De acuerdo con la ley colombiana, específicamente, la Ley 975 de 2005, las víctimas de las violaciones a los Derechos Humanos no solo es la persona a quien se le vulneran todos sus derechos, sino, también sus familiares; madres, padres, hijos, hermanas, abuelos, entre otros. Ellos viven un sufrimiento psicológico de manera permanente resultado de la incertidumbre de las amenazas, tortura, secuestro y muerte de su ser querido. (Alvis - Rizzo, Duque Sierra, & Rodríguez Bustamante, 2015)

Sin embargo, el drama de la desaparición también lo viven los niños, niñas y jóvenes que se enfrentan a pérdidas que difícilmente se asumen psicológicamente, generando marginación social, vulneración de igual forma en sus derechos, procesos de revictimización; una desatención por parte del Estado y hasta su propia familia; sumado a otras condiciones que pueden generarse (Guglielmucci, 2017).

Las cicatrices intangibles y difíciles de sanar producto de los hechos violentos y aún más en los casos de desaparición forzada, donde se genera un alto grado de incertidumbre y dolor. (Rodríguez Valderrama & Bastidas Robayo, 2017).

Las personas víctimas de este violento delito ponen en riesgo su estabilidad física y psicológica, es decir, se forja un detrimento, pérdida y daño psicológico, lesiones que limitan el desarrollo y bienestar de la persona.

En Colombia se ha identificado que en las personas víctimas de este tipo de hechos violentos, hasta en un 63%, según la Organización Mundial de la Salud, presentan algún tipo de sintomatología clínica relacionada con trastornos de la alimentación, problemas del sueño, trastornos del estado de ánimo, síntomas de estrés postraumático, psicoticismo,

aumento problemáticas familiares y comunitaria, entre otras. Lo más alarmante es que el 33% presenta los criterios de diagnóstico de trastorno mental. (Moncada Buelvas, Nisperuza Guerra, Simancas Fernández, & Celedón Rivero, 2020)

### **2.1.1 Identificación de cuerpos**

Para las familias de los desaparecidos, la falta de un cadáver que evidencie la pérdida del ser amado, es fundamental para iniciar su proceso de duelo; aunque en ocasiones pasan años y décadas sin noticias del paradero de su familiar o sin rastros de su cuerpo.

En ese sentido, poder hallar el cuerpo, se convierte en un alivio que, aunque no repara la pérdida sí permite hacer el duelo (Díaz Facio Lince, 2008).

Desde finales de los años 80, el Instituto Nacional de Medicina Legal y Ciencias Forenses junto a la Procuraduría General de la Nación inician con la búsqueda de herramientas para establecer la identificación de las personas que se encontraban asesinadas y sin identidad; aunque la prioridad en el país era el estudio y diagnóstico de muertes naturales, ya que la patología forense esa la principal disciplina que orientaba las investigaciones (Muñoz Marín, 2015).

Los cuerpos que se analizaban eran aquellos en los que la muerte era resultado de síntomas que permitían diagnosticar enfermedades. Por lo tanto, la revisión de cuerpos violentados y sin identidad no aportaba nada en términos de avances médicos, razón por la cual, estas tareas las desarrollaba la epidemiología o medicina social. Es decir, que, en los años 80, los estudios para establecer la identidad de las personas asesinadas fueron nulos. (Muñoz Marín, 2015)

Según (Muñoz Marín, 2016) hacia el año de 1987, la desaparición forzada se había extendido por todo el territorio nacional, convirtiendo espacios cotidianos en lugares de la muerte. Es en ese año, cuando se organiza la primera jornada oficial de búsqueda de personas desaparecidas liderada por el Comité Permanente de Derechos Humanos y la Procuraduría General de la Nación.

El desarrollo de la jornada de búsqueda permitió establecer las principales problemáticas de identificación de los desaparecidos, entre los que se destacan las inadecuadas prácticas de levantamiento de cuerpos; falta de mecanismos de recolección de información; complicidad de las autoridades en la práctica masiva de enterramiento en fosas comunes. (Muñoz Marín, 2016)

La jornada se cumplió en un antecedente para el avance de los procesos de búsqueda e identificación de personas desaparecidas durante la década de los noventa; lo que se convirtió en un insumo para lograr la tecnificación de los procesos. (Muñoz Marín, 2016).

Sin embargo, es la violencia generalizada y cotidiana que se vivía en el país, el detonante que llevó a ciencia forense a plantear su objeto de investigación, ya no era urgente establecer la causa de las enfermedades que llevaban a la persona a morir, sino, intentar establecer una identidad de las personas afectadas con artefactos explosivos, y de esta manera poder entregar a sus seres queridos. (Muñoz Marín, 2016), conservando el mismo estilo de ésta, a lo largo de todo el documento.

### **2.1.2 Herramientas de búsqueda**

Entre tanto, con la Ley 589 de 2000 que tipifica la desaparición forzada, genocidio, desplazamiento forzado y tortura, se crean dispositivos de búsqueda e identificación de personas desaparecidas como son el Mecanismo de Búsqueda Urgente, el Plan Nacional de Búsqueda, Registro Nacional de Desaparecidos, Comisión Nacional de Búsqueda de Personas Desaparecidas y Registro de detenidos y capturados y administración de bienes (Muñoz Marín, 2016).

De acuerdo con la Fiscalía General de la Nación, el Mecanismo de Búsqueda Urgente, es una herramienta que se activa de manera gratuita para ubicar a las personas que se presumen como desaparecidas y con el que se ordena de manera inmediata a las autoridades judicial el desarrollo del inicio de las diligencias necesarias para lograr su localización.



Entre tanto, el Plan Nacional de Búsqueda, es la hoja de ruta para encontrar las personas desaparecidas que se construyó de manera colectiva a través de encuentros en Bogotá, Cali, Cundinamarca, Bucaramanga, Barranquilla y Neiva; propone una búsqueda humanitaria y extrajudicial basada en tres pilares; información que se analiza, actualiza y se usa en la búsqueda; localización, recuperación e identificación de las personas desaparecidas, y garantía, en la participación de personas y organizaciones. (Unidad de Búsqueda de personas dadas por desaparecidas, 2020)

El Registro Nacional de Desaparecidos hace referencia a un sistema de información referencial de datos suministrados por las entidades que hacen parte de la búsqueda e identificación de personas desaparecidas. Este procedimiento hace posible la identificación de cadáveres sometidos a necropsia médico legal en el territorio nacional y orienta la búsqueda de las personas reportadas como víctimas. (Muñoz Marín, 2016)

Finalmente, la Comisión Nacional de Búsqueda de Personas Desaparecidas y Registro de detenidos, tiene como principal tarea, apoyar y promover la investigación del delito de desaparición forzada, con pleno respeto de las competencias institucionales y de las facultades de los sujetos procesales.

Después del año 2005, con la aplicación de la Ley 975 -Ley de Justicia y Paz- empezaron a circular publicaciones de tipo científico que funcionan como guía de procesos y prácticas médicas, además de marcos jurídicos que establecen las funciones y responsabilidades de cada una de las autoridades públicas en materia de búsqueda e identificación de personas desaparecidas.

Uno de los resultados más importantes de esta ley fue el hallazgo de más de 6.400 cuerpos de desaparecidos que se encontraban en estado de osamenta y los procesos de identificación han resultado ser uno de los principales retos de los últimos diez años para el Estado colombiano (Muñoz Marín, 2016).

## **2.2 Disciplinas del conocimiento**

En Latinoamérica, la antropología forense empezó a implementarse resultado de los múltiples reportes de violaciones de los derechos humanos; específicamente, en lo relacionado con la desaparición forzada de personas.

Establecer la identidad de una víctima requiere de expertos en disciplinas como Medicina, antropología forense, odontología, dactiloscopia, ciencias sociales, toxicología y otras ramas con aplicación forense (Parra & Palma, 2005).

En ese sentido, de acuerdo a lo que se requiera analizar e identificar, así mismo, se selecciona la técnica adecuada que puede ir desde difracción de rayos X, fluorescencia de rayos X, catodoluminiscencia, espectrometría de masas con plasma inductivamente acoplado, microscopia electrónica de barrido y difracción láser hasta imágenes de satélite y técnicas como radar de penetración del terreno, métodos geo eléctricos, inducción electromagnética, magnetometría y gravimetría (Molina Gallego, 2012).

### **2.2.1 Métodos geofísicos**

Las técnicas geofísicas iniciaron su implementación en el sector de hidrocarburos, específicamente, en la exploración de petróleo y minería en los años 1910 a 1930. Los siguientes diez años fueron parte de la arqueología; se destacan la detección de depósitos ilegales de basura tóxica con GRP y tomografía; la detección de minas personales; exploración de yacimientos minerales (Molina Gallego, Carlos Martin, 2016).

Sin embargo, es a partir de los años setenta que se aplica en investigaciones forenses.

Aunque en un buen número de países latinoamericanos, incluyendo Colombia, la geofísica forense no se aplica para la búsqueda de personas desaparecidas; sí se utilizan los conocimientos de las Ciencias de la Tierra y el uso de tecnologías para evaluar su utilidad en múltiples fosas clandestinas.

Por otra parte, factores como el tiempo del entierro, el estilo, el tipo de suelo, vegetación y clima, son variables que hace parte de los métodos de detección óptima.

Teniendo en cuenta que en América del Sur se carece de investigación con experimentos de prueba controlados para determinar las técnicas geofísicas óptimas de búsqueda y configuración de los equipos; es Argentina donde se han realizado algunas búsquedas de personas con tomografía y georradar. Entre tanto, en Colombia se usa de manera empírica y espontánea el GPR en algunos casos judiciales.

### **2.2.2 Penetración del terreno**

El método electromagnético denominado radar de penetración del terreno (GPR) utiliza la propagación y reflexión de ondas de radio en un rango de frecuencia entre 25 y 1.200 MHz; que se mide a través de una antena transmisora que emite impulsos electromagnéticos desde la superficie hacia el subsuelo.

Si hay suelos alterados de origen antrópico, el método de GPR entregará información sobre la intensidad, el tamaño y la forma general de la anomalía, la cual puede ser causada por la presencia de restos humanos y objetos artificialmente enterrados.

Desde hace varias décadas, se valora y destaca la eficiencia de este método en la búsqueda de evidencias enterradas en el subsuelo y es uno de los más utilizados por las autoridades del mundo.

cuadradas).

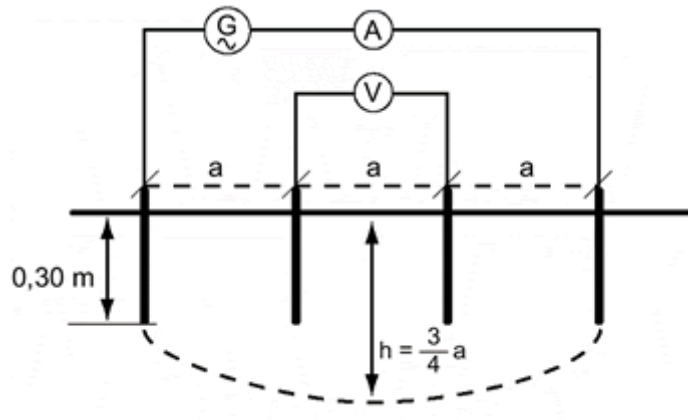
## **2.3 Método de detección Wenner**

Frank Wenner en 1915 introdujo el concepto de resistividad aparente e impulsó un método para calcular la resistividad aparente en un terreno. Con las mediciones de resistividad se conocen las características, constitución y estructura del subsuelo.

Para ello, se colocan cuatro electrodos en el suelo, ubicados en línea recta, con la misma distancia entre cada uno y enterrados en una profundidad igual y menor al 5% del espaciamiento de los electrodos. Luego se conectan a los extremos del telurómetro o Megger (Cardenas Valencia, 2011).

Es decir, se crea un campo eléctrico mediante dos electrodos puntuales, que podrían ser A y B, que se denominan de emisión, a través de estos se inyecta en el terreno una corriente eléctrica continua de intensidad  $I$ . Entre tanto, otros dos puntos del terreno con ayuda de los electrodos M y N de medida, ubicados en estos puntos y mediante el correspondiente instrumento de medida, se calculan las diferencias del potencial que se ha generado (Quispilaya Marmolejo, Robinson; Cruz Chumpitaz, William Jason, 2018).

**Figura 1.** Método Wenner



El principio básico de este método, - uno de los de mayor uso-, es la inyección a través de la tierra de una corriente directa o de baja frecuencia para evitar el problema de polarización que se tiene con la corriente directa. Permite diferenciar tipos de materiales, según, la capacidad que tiene cada uno de resistir el paso de corriente eléctrica. En el caso específico forense es utilizado para determinar la presencia o ausencia de entierros y vertederos de basura ilegal. (Cardenas Valencia, 2011)

Este método tiene una gran importancia; el grado de variabilidad está ligado a los cambios en el subsuelo, resultado de los diferentes tipos de litologías presentes. Determinar los diferentes valores de resistividad eléctrica del subsuelo permitirá, identificar unidades

litológicas con distinta textura o grado de alteración, aspectos estructurales (fallas) y geomorfológicos (cuevas y rellenos), etc. (Cardenas Valencia, 2011).

Entre tanto, técnicas como la magnetometría todo un éxito en ambientes simulados y con objetivos enterrados, y el electromagnetismo que mide la conducta del suelo, se aplica con mayor intensidad en casos ambientales, sin embargo, ha tenido resultados en casos penales (Cardenas Valencia, 2011).

## **2.4 Metodología Machine Learning**

La metodología de desarrollo de software denominada Machine Learning o Aprendizaje Automático está relacionada con la construcción de programas informáticos que mejoran automáticamente una determinada experiencia o proceso. (Sanchez Galan, 2020).

Es un subcampo de las ciencias de la computación especializada en el reconocimiento de patrones complejos en conjunto de datos y su principal característica consiste en extraer de forma autónoma información relevante en los datos que se están procesando. De esta manera, el programa mejora la ejecución de la tarea para la que ha sido programado (Gonzalez Garcia, 2018).

Así que el modelo, suele aplicarse en problemas con un gran conjunto de datos; en aquellos que requieren de un razonamiento numérico, o donde se registren relaciones complejas entre las variables. Su aplicación se registra en cualquier área que requiera de un computador, pero que es difícil de programar (Sanchez Galan, 2020).

Cabe anotar, que dependiendo de la tarea que se requiere, se utilizan distintos tipos de aprendizaje que existen dentro del Machine Learning, se trata de aprendizaje supervisado, no-supervisado y por refuerzo. (Sanchez Galan, 2020)

Por otro lado, según las áreas del conocimiento en el que se aplique, se tiene una utilidad específica; por ejemplo, en bioinformática apoya la determinación espectral-molecular de especies de mosquitos, entre tanto, en ecología contribuye a la determinación y clasificación de vocalizaciones de manatíes, mientras que en agricultura ayuda en la

determinación automática de variedades de plantas arroz y para la clasificación automática de sandías para exportación (Sanchez Galan, 2020).

## 2.5 Modelos de clasificación

### 2.5.1 Support Vector Machines

Support Vector Machines (SVM), es el conjunto de algoritmos de aprendizaje supervisado y desarrollado por Vladimir Vapnik en compañía de su equipo en los laboratorios AT&T (Valencia Valencia & Galicia Haro, 2016).

SVM es una técnica de reconocimiento de patrones basada en la metodología de aprendizaje, considerada una herramienta sólida para regresión y clasificación en dominios complejos; aunque fueron pensados para resolver problemas de clasificación binaria. Su diferencia con otras técnicas radica, en que no son afectadas por el problema de los mínimos locales, resultado, de su entrenamiento basado en problemas de optimización convexa (Montt, Castro, & Rodriguez).

Esta técnica que se caracteriza por el reconocimiento a sus sólidos fundamentos teóricos, ha sido utilizada con éxito en diversos campos, entre los que se destacan, visión artificial, reconocimiento de caracteres, categorización de texto e hipertexto, clasificación de proteínas, procesamiento de lenguaje natural, análisis de series temporales (Ramírez Pérez, Laguna Estrada, & Rubín Ramírez, 2017)

Para este caso, un proceso de clasificación supervisado, se deben aplicar dos fases: entrenamiento y prueba. En la fase de entrenamiento o training, un conjunto de datos inicial es usado para concluir que parámetros deberán ser ponderados y combinados. Entre tanto, en la fase de prueba o testing, el clasificador determinado – *conexión entre las variables dadas y las que se predicen*- en la fase de entrenamiento, es aplicado a un conjunto de datos u objetos. De esta manera, se clasifican elementos y se comparan con los datos reales, para establecer la efectividad del modelo (Montt, Castro, & Rodriguez).

### 2.5.2 KNN

Es un método de aprendizaje inductivo supervisado que permite obtener nuevos conceptos a partir de ejemplos. No existe un modelo global asociado a los conocimientos a aprender, por el contrario, las predicciones se basan en modelos parecidos al que se quiera predecir (García Cambroner & Gómez Moreno, S.F).

KNN es usado por ejemplo para detectar y clasificar las etapas de sueño; diagnostica enfermedades del corazón, entre otros múltiples usos en distintas áreas del conocimiento. Este algoritmo, clasifica un nuevo vector aplicando un conjunto de entrenamiento, sin necesidad de aprender parámetros; caso contrario al uso de otros algoritmos que aplicando un conjunto de entrenamientos aprende algunos parámetros (Rosado Muñoz, Bataller Mompean, Frances Villora, & Guerrero Martinez, 2018).

### 2.5.3 Ensamble

El ensemble se constituye por un conjunto de modelos de machine learning. Cada uno de los modelos produce una predicción diferente, al combinarse, los errores de cada tipo tienden a compensarse, lo que hace posible una única predicción y de esta manera resulta un mejor error de generalización (Martines Hera, 2019).

La cantidad numerosa de algoritmos permite la construcción de ensambles de modelos y muchas aplicaciones a diversos problemas (Rosati, 2017)

Los ensambles de clasificadores son conocidos por ser una buena alternativa para mejorar el rendimiento de un único clasificador utilizando una combinación de varios clasificadores. Inicialmente, los clasificadores individuales que forman el ensemble deben ser diversos para que en el momento de combinarlos se mejoren los resultados (Uriz & et al, 2018).

Cuando se construye un ensemble para un determinado problema, se debe tener presente dos aspectos, en primer lugar, cómo generar diversidad entre clasificadores y, en segundo lugar, cómo combinar la decisión de todos los clasificadores en una única salida. (Uriz & et al, 2018)

Existen cuatro tipos de ensambles considerados los más usados, se trata de votación por mayoría, bagging, boosting y stacking.

- **Votación por mayoría**

Hace referencia a ensayar varios modelos de aprendizaje automático con los mismos datos. Si los datos son nuevos se obtiene una predicción para cada modelo, y cada uno tendrá asociado un voto; de esta manera, la predicción final será el resultado de lo que voten la mayoría de los modelos (Martines Hera, 2019).

Entre tanto, cuando los modelos de machine learning dan una probabilidad de utilizar el soft – voiting; que da más importancia a los resultados en los que algún modelo este muy seguro. Por ejemplo, cuando la predicción está muy cercana a la probabilidad 0 ó a 1, se le da más peso a la predicción de ese modelo. (Martines Hera, 2019).

- **Bagging**

El segundo modelo es el denominado Bagging, en el que también se combina distintos modelos de machine learning. La forma de conseguir que los errores se reparen entre sí, es que cada modelo se entrena con subconjuntos, que se eligen de muestras aleatoriamente del conjunto de entrenamiento. (Martines Hera, 2019). Los resultados se combinan; sin embargo, para problemas de regresión, normalmente se utiliza la media aritmética.

- **Boosting**

En este ensemble, el modelo intenta arreglar los errores de los modelos anteriores, se consigue otorgando más peso a las muestras mal clasificadas y menos peso a las muestras bien clasificadas. Para problemas de regresión, las predicciones con un mayor [error cuadrático medio](#) tendrán más peso para el siguiente modelo. Hay muchas implementaciones de ensembles que usan boosting. El primero fue el [AdaBoost](#). Los más usados actualmente son [xgboost](#), [CatBoost](#) y [LightGBM](#) (Martines Hera, 2019).

## 2.7 Métricas de evaluación

Una matriz de confusión o matriz de error es una tabla de contingencia que sirve como herramienta estadística para el análisis de observaciones emparejadas; permite informar



la exactitud temática de cualquier producto de datos derivados de la teledetección y puede ser utilizada para evaluar la calidad temática de cualquier tipo de dato espacial. (Ariza Lopez, Rodriguez Avi, & Fernandez, 2018)

Una alternativa que permita verificar la bondad del clasificador es la matriz de confusión, que hace posible la visualización mediante una tabla de contingencia a la distribución de errores cometidos por un clasificador (Corso, Cynthia Lorena).

Este modelo facilita la detección de errores, el análisis descriptivo y analítico; la visión general de las asignaciones correctas y de las equivocaciones. Es, además, la forma más usual de informar la exactitud temática de productos derivados de la clasificación de datos procedentes de imágenes.

En la matriz pueden existir dos tipos de errores, el primero, denominado errores de omisión que hace referencia a elementos que perteneciendo a esa clase no aparecen en ella por estar incluidos erróneamente en otra; es decir, datos por debajo de la diagonal principal de la matriz de confusión. Y el segundo, está relacionado con errores de comisión; elementos que no perteneciendo a una clase aparecen en ella y están por encima de la diagonal principal de la matriz de confusión (Sanchez Muñoz, 2016).

En conclusión, la Matriz de confusión permite percibir el desempeño de un algoritmo de aprendizaje específico. Para ello, cada columna de la matriz representa los aciertos y errores que tiene el modelo durante el proceso de aprendizaje a través de los datos.

Es decir, al seleccionar el algoritmo se realiza una clasificación determinada y se establece el porcentaje de acierto en las pruebas, luego, se comparan los resultados del modelo real con el modelo predictivo y se conforma la Matriz de confusión con las probabilidades VP – Verdadero positivo; VN – Verdadero Negativo; FN – Falso Negativo y FP – Falso Positivo. Estas métricas de rendimiento desempeñan son fundamentales en los problemas de clasificación donde discriminan los algoritmos de Machine Learning.

**Figura 2** - Matriz de Confusión Binaria



La Matriz de confusión tiene unas métricas específicas, se trata de exactitud y precisión; sensibilidad y especificidad. (Barrios Arce, 2019)

- **Exactitud y Precisión**

La Exactitud hace referencia al resultado de una medición del valor verdadero. Es decir, está relacionada con el sesgo de una estimación y es conocida como Verdadero Positivo. Está representada por la proporción entre los positivos reales pronosticados por el algoritmo y todos los casos positivos. Se puede determinar finalmente, que la Exactitud es la cantidad de predicciones positivas que fueron correctas.

Es decir, la Exactitud es la cantidad de predicciones positivas que fueron correctas.

**Ecuación 1** - Formula para calcular la Exactitud

$$(VP + VN)/(VP + FP + FN + VN)$$

Entre tanto, la Precisión se refiere a la dispersión del conjunto de valores obtenidos a partir de mediciones repetidas de una magnitud. A menor dispersión mayor Precisión. Para este caso, se representa la proporción entre el número de predicciones correctas positivas como negativas y el total de predicciones. Para calcular el porcentaje de los casos positivos detectados se debe realizar la ecuación que se presenta a continuación,

**Ecuación 2** - Cálculo de Precisión

$$VP/(VP + FP)$$

- **Sensibilidad y Especificidad**

Estos dos valores indican la capacidad del algoritmo para diferenciar los casos positivos de los negativos. Mientras la sensibilidad puede definirse como la fracción de verdaderos positivos, la especificidad, es la fracción de los verdaderos negativos.

La sensibilidad es conocida como Verdadero Positivo, que representa los casos positivos que fueron correctamente identificados por el algoritmo y se calcula de la siguiente manera:

**Ecuación 3** - Fórmula para calcular la Sensibilidad

$$VP/(VP + FN)$$

La especificidad nos permite medir la fracción de verdaderos negativos, es decir cuántos de los casos negativos el algoritmo ha clasificado correctamente.

**Ecuación 4** - Fórmula para calcular la Especificidad

$$VN/(VN + FNP)$$

## 2.8 Base de datos

Una base de datos puede definirse como el conjunto de información estructurada en registros y almacenada en un soporte electrónico legible. Cada registro constituye una unidad autónoma de información (Rodríguez Yunta, 2001).

Hoy, las bases de datos se consolidan como un activo más y quizás el más importante de las empresas, entidades e instituciones, por convertirse en la memoria y registro de toda su información.

Este cumulo de información puede clasificarse en tres grades grupos; las bases de datos con información factual, es decir, datos estadísticos, series históricas, resultados de encuestas, convocatorias de becas o subvenciones, convocatorias de eventos, ofertas de empleo, entre otros. En el segundo grupo se encuentran los llamados Directorios con información especializada sobre personas o instituciones, y finalmente, las bases de datos

documentales que hacen referencia a una publicación impresa, un documento audiovisual, etc. (Rodríguez Yunta, 2001)

En conclusión, las bases de datos permiten resolver necesidades de información y marcar tendencias, por ejemplo, el actual Ministerio de Salud y Protección Social de Colombia registra diariamente la información de las bases de datos enviadas por las regiones sobre el número de personas contagiadas con COVID 19, lo que permita estudiar el comportamiento del virus a través de distintas variables.

### **2.8.1 Minería de datos**

La minería de datos es conocida como Data Mining, en inglés, tiene sus orígenes desde los años 60's, considerada como una ciencia para extraer la información de grandes volúmenes de datos (Camana, 2016).

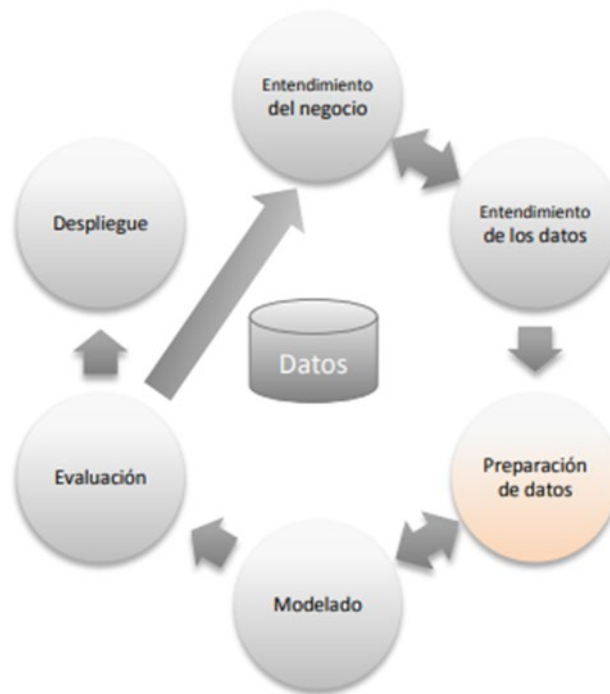
La minería de datos (MD) es un intento de buscarle sentido a la explosión de información que actualmente puede ser almacenada y que está presente en todas las actividades del ser humano, incluso en las áreas del comercio electrónico y la bioinformática (Riquelme, Ruiz, & Gilbert, 2006).

La minería de datos es fundamental en el proceso para aplicar algoritmos específicos y extraer patrones (modelos) de los datos que ajustado a unos datos proporciona conocimiento.

### **2.8.2 Modelo de referencia CRISP-DM**

El CRISP-DM plantea una metodología de Minería de Datos estandarizada, la cual es bastante utilizada por su flexibilidad y capacidad de caracterizarse para la aplicación en diferentes dominios con gran facilidad.

**Figura 3.** Ciclo de vida del modelo CRISP - DM



La metodología CRISP – DM como se puede observar en la figura 3 está compuesta por varias fases de desarrollo:

- **Entendimiento del negocio:** En esta fase se comprende los objetivos y requerimientos del proyecto desde la perspectiva del negocio.
- **Entendimiento de los datos:** Etapa para la recolección de datos y el análisis con el fin de ser acondicionados a las necesidades del proyecto para su posterior procesamiento.
- **Preparación de Datos:** En esta fase se realiza una selección de datos para efectuar la respectiva limpieza de datos quedando preparado para el debido modelamiento.
- **Modelado:** En esta fase se elabora el modelo de predicción con base a los datos previamente procesados.

- **Evaluación del modelo:** En esta fase de evaluación del modelo, se observa, si los resultados son los esperados; de lo contrario, se procede a verificar los pasos anteriores con el fin de encontrar mejores resultados.
  
- **Despliegue del modelo:** En la implementación de esta fase, se definen las estrategias para poner en marcha el proyecto, teniendo en cuenta la supervisión y mantenimiento del algoritmo de detección



## ▪ **3. Metodología CRISP – DM aplicada a la detección de fosas**

En este capítulo se aplica el CRISP – DM como se explica en el *capítulo 2 sección 1.7.2*, para el desarrollo del proyecto; se registra el paso a paso de la metodología aplicada al tratamiento de los datos obtenido a través del método Wenner en el Laboratorio de Fosa Común ubicado en Usme.

### ○ **3.1 Fase 1: Entendimiento del proceso**

En esta fase se ahonda la problemática social que vive el país y el drama que afrontan los familiares producto del flagelo de la desaparición forzada. Sumado a ello, se muestra un esbozo de los métodos aplicados en la actualidad para la ubicación de fosas comunes y los mínimos resultados que registran como se describe a profundidad en el *capítulo 1*.

### ○ **3.2 Fase 2: Entendimiento de los datos**

A continuación se muestra el análisis para la recolección de datos a partir de la instalación de 6 fosas implementadas en la universidad Antonio Nariño sede Usme , a través del método Wenner se toman los datos de resistividad los cuales arrojaron 3 líneas de profundidad, estos son organizados en un mapa 2D permitiendo una mejor visualización.

#### ▪ **3.2.1 Análisis a los datos y su método de recolección**

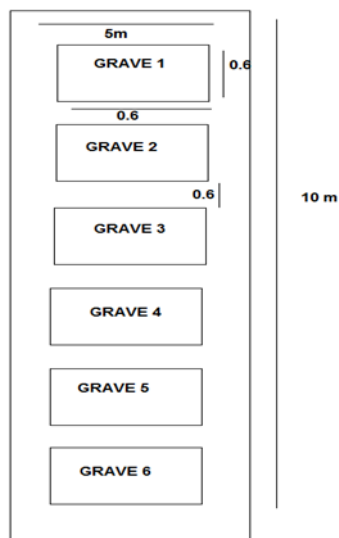
La recopilación de los datos es la base fundamental para obtener los resultados esperados; por ello, se debe conocer y entender, el origen de los datos para establecer de esta manera, sus características fundamentales.



Inicialmente este proyecto usa la aplicación de diferentes técnicas de prospección no invasivas para estudiar las propiedades del subsuelo de una zona y determinar, si esa zona tiene enterramientos o no.

A partir de esta información, se crean fosas experimentales controladas que incluyen cerdos muertos. El procedimiento se basó en el modus operandi de los paramilitares, quienes realizaban enterramientos pequeños ya que los cuerpos eran desmembrados, dificultando de esta forma su detección.

**Figura 4.** Diagrama de disposición de fosas



Como se observa en la Figura 4 sobre la disposición de las fosas, cada una tiene una dimensión de 0.6 m x 0.6 m con una profundidad de 50 cm y una distancia entre cada fosa de 0.6 m. Cada fosa contiene piezas de cerdo como se explica en la Tabla 1.

**Tabla 1.** Designación de fosas

DESIGNACIÓN DE LAS FOSAS	
FOSA 1	4 piezas de cerdo con ropa que corresponden a la cabeza, extremidades superiores, y torso superior

FOSA 2	3 piezas de cerdo que representan las extremidades inferiores y la parte baja del torso.
FOSA 3	Fosa de control. Se realizó la excavación y se llena nuevamente con tierra.
FOSA 4	4 piezas de cerdo con ropa, sin carne, que corresponden a la cabeza, extremidades superiores y torso superior.
FOSA 5	3 piezas de cerdo con extremidades inferiores y la parte baja del torso
FOSA 6	Técnica penetrollogger

### ▪ 3.2.3 Bases de datos

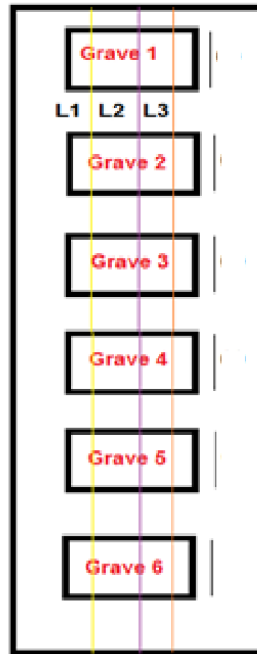
Luego de aclarar el proceso para la generación de datos, se procede a analizar el equipo de tomografía implementando el método Wenner. Como se indica en la figura 5, se dispone de una línea 32 electrodos de acuerdo a la configuración del dispositivo, así se debe implementar.

**Figura 5.** Implementación método Wenner



Para lograr una prospección de toda la zona, teniendo claro que se debe realizar por línea de 32 electrodos, se configura un diseño experimental como se muestra en la figura 6.

**Figura 6.**Diseño experimental



Se toman 3 líneas de medición, que corresponden a la parte lateral izquierda, medio y parte lateral derecha. Por cada línea, se realiza una medición de resistividad. El tomógrafo eléctrico empleando el método Wenner arroja los datos resultantes en 3 columnas; la columna 1 muestra la medida de la distancia que se genera a lo largo de la toma de la prospección; la columna 2 es la distancia entre los electrodos, y la columna 3 contiene los datos de las resistividades tal como se indica en la tabla 2.

**Tabla 2.** Datos obtenidos método Wenner

Usme-L1-previa_feb19_2020		
0.2		
1		
140		
1		
0		
Posicion	Distancia electrodos	Resistividad
0.3000	0.2000	103.4598
0.5000	0.2000	78.2003
0.7000	0.2000	73.4725
0.9000	0.2000	84.9520
1.1000	0.2000	64.8378
1.3000	0.2000	96.7772
1.5000	0.2000	100.0721
1.7000	0.2000	97.0146
1.9000	0.2000	109.3088
2.1000	0.2000	106.2633
2.3000	0.2000	84.1489
2.5000	0.2000	132.5610
2.7000	0.2000	114.7749
2.9000	0.2000	98.6754
3.1000	0.2000	117.1526
3.3000	0.2000	126.9614
3.5000	0.2000	84.2166
3.7000	0.2000	115.9278
3.9000	0.2000	90.9023
4.1000	0.2000	98.8088
4.3000	0.2000	99.2699
4.5000	0.2000	97.6217
4.7000	0.2000	107.6590
4.9000	0.2000	94.5511
5.1000	0.2000	96.2708
5.3000	0.2000	93.3864
5.5000	0.2000	95.1820
5.7000	0.2000	98.3017
5.9000	0.2000	64.7858
0.6000	0.4000	45.4720
0.8000	0.4000	57.9553
1.0000	0.4000	75.8411
1.2000	0.4000	97.9772

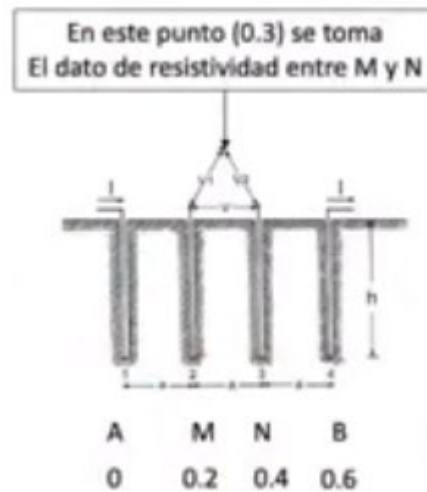
### ▪ 3.2.4 Creación de base de datos

Para profundizar los datos entregados por la metodología Wenner se debe tener claridad en la forma cómo se toman las medidas. De acuerdo con la tabla 2 y a los valores que corresponde, es importante comprender la medición en el equipo de tomografía eléctrica.

En la figura 7 se observa los primeros 4 electrodos posicionados y referenciados de la siguiente manera electrodo 1 = A, electrodo 2 = M, electrodo 3 = N y el electrodo 4 = B.

Un dato muy importante es que el primer electrodo A es de referencia y parte de un valor cero de acuerdo a los parámetros del tomógrafo; en el método Wenner la distancia entre cada electrodo es de 0.2 metros, quedando distribuidos como se muestra en la figura 7.

**Figura 7.** Medición equipo de tomografía



Cuando la corriente circula, ingresa por el electrodo A y circula hasta el electrodo B, se realiza una diferencia de potencial entre el electrodo M y N, el dato de resistividad está dado en el medio de los electrodos M y N. Para el caso de la figura 7, el primer dato de resistividad se encontraría en 0.3 y así sucesivamente para obtener la siguiente medida de resistividad, cada electrodo se mueve 0.2, y de esta manera se organizan las mediciones tomadas por el tomógrafo en cada columna.

Es importante resaltar que la distancia o configuración entre cada electrodo permite profundizar más a la señal eléctrica logrando establecer varios niveles de profundidad. A continuación se expresa la relación matemática para determinar la profundidad de los datos de resistividad dados por el equipo, cuya configuración y ecuación 4 está dada por el fabricante.

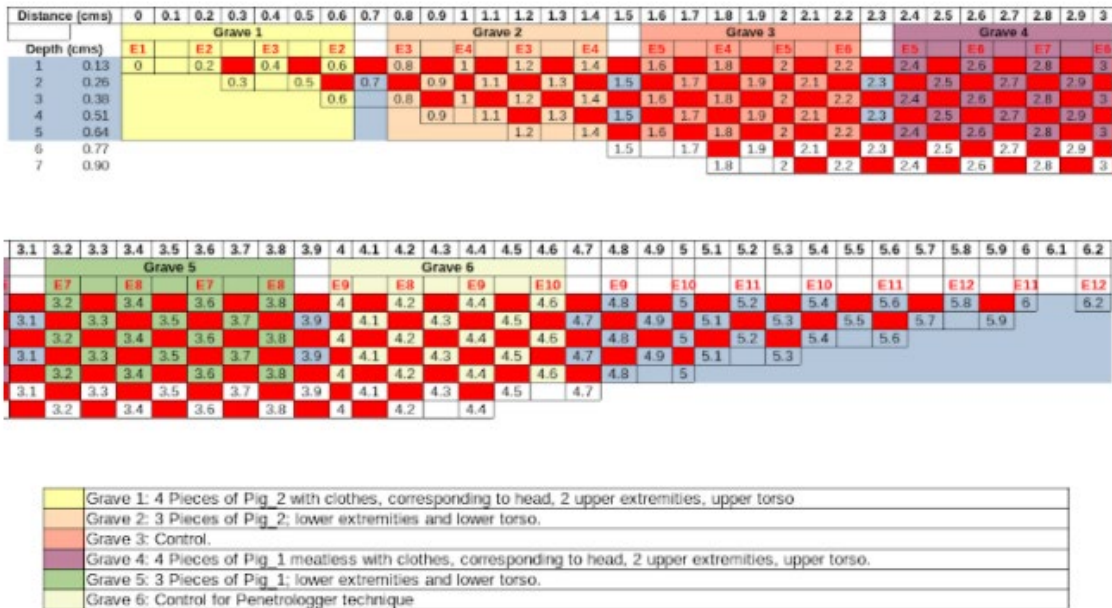
**Ecuación 4.** Ecuación de profundidad dada por el fabricante.

$$\text{Profundidad} = \frac{32}{50} a$$

Donde a = es la distancia entre los electrodos y la ecuación se toma de las especificaciones del fabricante del tomógrafo.

Por lo tanto, para comprender si una muestra pertenece a una tumba y el tipo de tumba, cada medición se ubica en un mapa 2D para cada línea de prospección tomográfica.

**Figura 8.** Ubicación de las muestras de fosas en un mapa 2D



Es significativo señalar que este gráfico 2D es aplicable a las tres líneas tomográficas. En el gráfico, los cuadrados rojos representan la ubicación estimada de las medidas correspondientes a la resistividad.

### 3.3 Preparación de los datos

Para la preparación de los datos se propone utilizar los cuatro datos alrededor del mapa 2D (figura 8) para estimar si en esa región hay o no una fosa como se ilustra en la figura 9.

**Figura 9.** Extracción de predictores para el entrenamiento de los modelos de machine learning.

Distance (cms)	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	
	Grave 1							Grave 2							Grave 3						
Depth (cms)	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14	E15	E16	E17	E18	E19	E20	
1	0.13	0	0.2	0.4	0.6	0.8	1	1.2	1.4	1.6	1.8	2	2.2	2.4	2.6	2.8	3	3.2	3.4	3.6	
2	0.26		0.3	0.5	0.7	0.9	1.1	1.3	1.5	1.7	1.9	2.1	2.3	2.5	2.7	2.9	3.1	3.3	3.5	3.7	
3	0.39			0.6	0.8	1	1.2	1.4	1.6	1.8	2	2.2	2.4	2.6	2.8	3	3.2	3.4	3.6	3.8	
4	0.51				0.9	1.1	1.3	1.5	1.7	1.9	2.1	2.3	2.5	2.7	2.9	3.1	3.3	3.5	3.7	3.9	
5	0.64					1.2	1.4	1.6	1.8	2	2.2	2.4	2.6	2.8	3	3.2	3.4	3.6	3.8	4	
6	0.77						1.5	1.7	1.9	2.1	2.3	2.5	2.7	2.9	3.1	3.3	3.5	3.7	3.9	4.1	
7	0.90							1.8	2	2.2	2.4	2.6	2.8	3	3.2	3.4	3.6	3.8	4	4.2	

A partir de la extracción de estos datos o predictores como se evidencia en la figura 9, se construye la tabla 3, en la que se incluye la etiqueta de las fosas, que indican lo que se espera que realice el algoritmo de predicción.

Por ejemplo, en la figura 10, la resistividad 1, 2, 3,4 se utiliza para estimar si en esa posición hay fosa y en este caso está etiquetado con 2 porque la fosa correspondiente es la 2.

**Figura 10..** Ejemplo de construcción y organización de los datos

1	2	3	4	FOSA
55.0827	85.4625	68.4589	72.8553	2

**Tabla 3.** Modelo de base de datos

PREDICTOR 1	PREDICTOR 2	PREDICTOR 3	PREDICTOR 4	FOSA
55.8297	38.5226	22.5847	38.6133	0
74.3524	55.8297	38.6133	48.0306	0
93.2396	74.3524	48.0306	72.0669	0
112.1113	72.0669	34.2453	75.9068	0
72.0669	48.0306	23.5358	34.2453	0
91.5947	75.9068	34.1050	76.4730	0
123.4248	70.8662	64.4560	118.7576	0
118.7576	57.6695	70.3305	85.4625	0
55.0827	85.4625	68.4589	72.8553	2
72.8553	68.4589	43.3275	53.6824	2
85.4625	70.3305	52.2454	68.4589	2
98.5105	72.8553	53.6824	53.9265	2
53.6824	43.3275	32.8150	43.0415	2
53.9265	53.6824	43.0415	43.0596	2
67.0105	53.9265	43.0596	58.4134	2
43.0596	43.0415	30.6201	37.6147	2
37.6147	30.6201	20.5972	25.5912	2
58.4134	43.0596	37.6147	45.0564	2
106.3286	58.4134	45.0564	62.8011	0
45.0564	37.6147	25.5912	34.3435	0
34.3435	25.5912	22.7504	47.4471	0
78.0142	62.8011	47.4471	81.9721	0
47.4471	34.3435	29.1905	34.1135	0
29.1905	22.7504	21.0500	25.8105	0
62.8011	45.0564	34.3435	29.1905	0
81.9721	47.4471	34.1135	60.6196	0
34.1135	29.1905	25.8105	27.8711	0
68.7940	81.9721	60.6196	77.4900	0
60.6196	34.1135	27.8711	36.6192	0
27.8711	25.8105	20.8951	30.6248	0
56.0725	49.3878	58.5216	72.1105	0
58.5216	39.4808	28.8445	37.7611	0
53.8229	56.0725	72.1105	86.3970	0
72.1105	58.5216	37.7611	48.6503	0
37.7611	28.8445	20.7936	22.8840	0
86.3970	72.1105	48.6503	66.3348	0
48.6503	37.7611	22.8840	32.2997	0
81.0630	86.3970	66.3348	89.1955	0



66.3348	48.6503	32.2997	40.9989	0
32.2997	22.8840	19.6569	26.7475	0
89.1955	66.3348	40.9989	59.2549	0
40.9989	32.2997	26.7475	63.9675	0
59.2549	89.1955	63.9675	69.6532	0
63.9675	40.9989	26.9851	38.9614	0
26.9851	26.7475	18.3454	20.1800	0
69.6532	63.9675	38.9614	45.8302	4
38.9614	26.9851	20.1800	29.2977	4
69.3790	69.6532	45.8302	55.2378	4
45.8302	38.9614	29.2977	28.4095	4
29.2977	20.1800	19.0142	22.4726	4
55.2378	45.8302	28.4095	40.2971	4
28.4095	29.2977	22.4726	22.5827	4
87.5722	55.2378	40.2971	61.5798	4
40.2971	28.4095	22.5827	45.9404	4
22.5827	22.4726	18.8667	23.3628	4
61.5798	40.2971	45.9404	61.5798	4
45.9404	22.5827	23.3628	45.9404	4
61.5798	64.5619	45.9404	77.8103	4
45.9404	36.4597	23.3628	33.9556	4
23.3628	17.5122	17.6739	21.4944	4
77.8103	45.9404	33.9556	66.3958	4
33.9556	23.3628	21.4944	29.2156	4
69.0939	77.8103	66.3958	73.0104	0
66.3958	33.9556	29.2156	40.0923	0
73.0104	66.3958	40.0923	71.8894	0
40.0923	29.2156	25.0105	31.7569	0
59.4850	73.0104	71.8894	85.6313	0
71.8894	40.0923	31.7569	57.9513	0
85.6313	71.8894	57.9513	57.9916	0
81.7866	85.6313	57.9916	75.6169	0
57.9916	57.9513	38.8555	52.9045	0
75.6169	57.9916	52.9045	58.3317	5
88.3746	52.9045	58.3317	75.0802	5
74.1722	75.0802	64.1179	60.4224	5
75.0802	58.3317	46.3911	64.1179	5

## ○ 3.4 Modelado

Para el modelamiento se emplea la herramienta `classification learner` de matlab el cual brinda los siguientes algoritmos de aprendizaje de máquina: `Bagged Trees`; su funcionamiento se basa inicialmente en el árbol de decisiones ya que se hace un muestreo de subconjuntos de los datos, seleccionando el mejor atributo o característica formando un nodo de decisión el cual se divide en subconjuntos más pequeños, este proceso se repite para cada atributo hasta formar el árbol de decisiones, el algoritmo `Bagged Trees` toma muestras adicionales es decir que los mismos datos se pueden incluir varias veces, se entrenan nuevos modelos y luego se promedian las predicciones. Un `Bagged Trees` o árbol de bolsa puede incluir una gran cantidad de árboles. (Martines Hera, 2019).

`Rusboosted Trees`, su funcionalidad se basa en un conjunto de árboles de decisión individuales, entrenados de forma secuencial donde el modelo intenta arreglar los errores de los modelos anteriores, esto se da otorgando más peso a las muestras mal clasificadas y menos peso a las muestras bien clasificadas.

`Subspace KNN`, El método del subespacio aleatorio se basa en un proceso probabilístico en que combina las predicciones de múltiples árboles de decisión entrenados en diferentes subconjuntos de columnas del conjunto de datos de entrenamiento. Modificar aleatoriamente las columnas en el entrenamiento de cada dato del conjunto y obtener nuevos conceptos a partir de ejemplos. No existe un modelo global asociado a los conocimientos a aprender, por el contrario, las predicciones se basan en modelos parecidos al que se quiera predecir (Rosado Muñoz, Bataller Mompean, Frances Villora, & Guerrero Martinez, 2018).

A estos algoritmos se les presentó los datos organizados como se muestra en la tabla 3. Para el entrenamiento de los algoritmos de predicción, se empleó un `cross validation` con un `k-Folds` de 10 con el fin de entrenar el modelo y probarlo varias veces para observar la variabilidad de sus resultados.

## ○ 3.5 Evaluación del Modelo

Para la evaluación del modelo se plantean 5 experimentos, los cuales se detallan a continuación.

Experimento 1, para determinar el impacto que tiene la normalización en el tratamiento de datos.

Experimento 2, se desarrolla con el fin de identificar la viabilidad de la corrección de valores atípicos el procesamiento de datos.

Experimento 3, se realiza con el fin de determinar cuál es el mejor algoritmo de aprendizaje automático para identificar la existencia de una tumba (incluyendo la fosa de control 0)

Experimento 4, se implementa para establecer la probabilidad de que los algoritmos de Machine Learning se confundan cuando hay etiquetas 2, 4,5 y una tumba de control (Grave 3).

Experimento 5, se quiere probar el grado de dificultad en el proceso de detección automática de las fosas a lo largo del tiempo.

Para evaluar cada uno de los experimentos como métrica se emplean la matriz de confusión figura 2 –capítulo sección 2.7 , la precisión, la especificidad ,exactitud y la sensibilidad cuyas ecuaciones se encuentran descritas en el capítulo 2 – sección 2.7 (ecuación 1 , 2 ,3 y 4). Los resultados de las métricas aplicadas en los experimentos se muestran en el capítulo siguiente.

## ▪ 4. Resultados

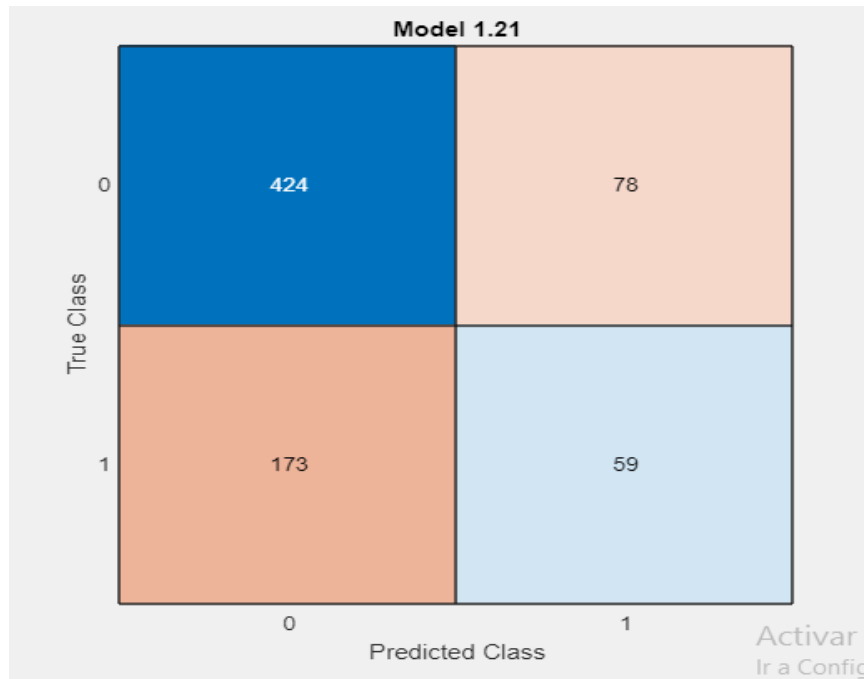
En este capítulo se presentan los resultados de las métricas obtenidas al aplicarlas en los diferentes experimentos, a continuación se detalla cada uno de los experimentos.

### ○ 4.1 Experimento 1: Impacto de la normalización en el tratamiento de los datos

Para el desarrollo de este experimento se realizó la preparación de los datos detallada en el capítulo 3 sección 3.3, se creó un código en Matlab el cual corresponde a corrección de valores atípicos y normalización (Anexo A), sin embargo en el modelamiento de este algoritmo de predicción se quiere analizar si la normalización es de utilidad en el tratamiento de datos, por ello se realiza un segundo entrenamiento de algoritmo sin la implementación de la normalización (Anexo B). Se entrenaron y probaron los algoritmos de Machine Learning con 2 tipos de datos: con y sin normalización. Los resultados obtenidos se muestran a continuación.

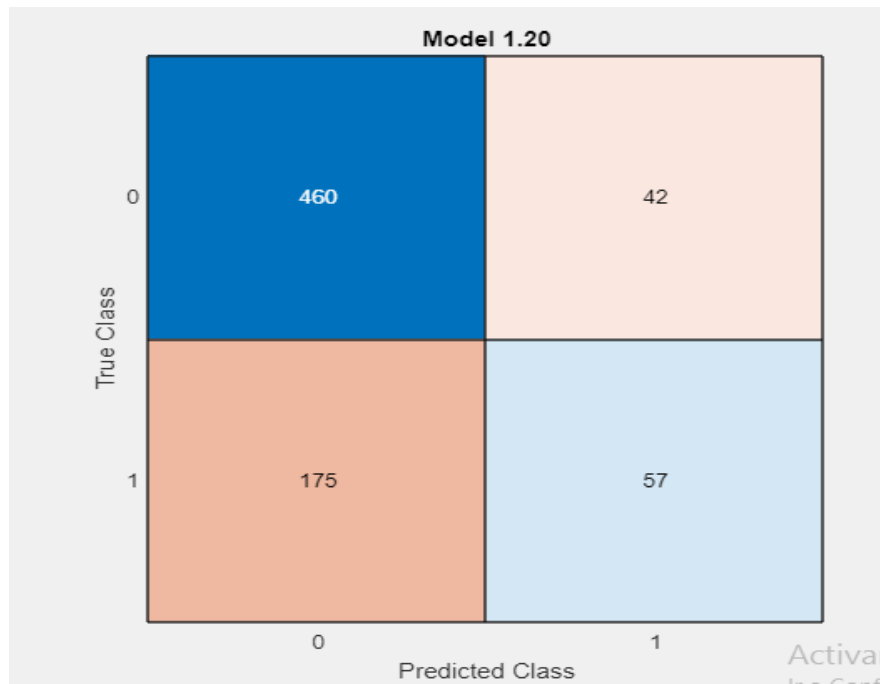
El algoritmo con mejor resultado para el primer caso es Rusboosted Trees

**Figura 11.** Matriz de confusión Rusboosted Trees sin normalización



Exactitud = 65.8% Precisión = 43% sensibilidad =25.4% especificidad=71%

**Figura 12.** Matriz de confusión Rusboosted Trees con normalización



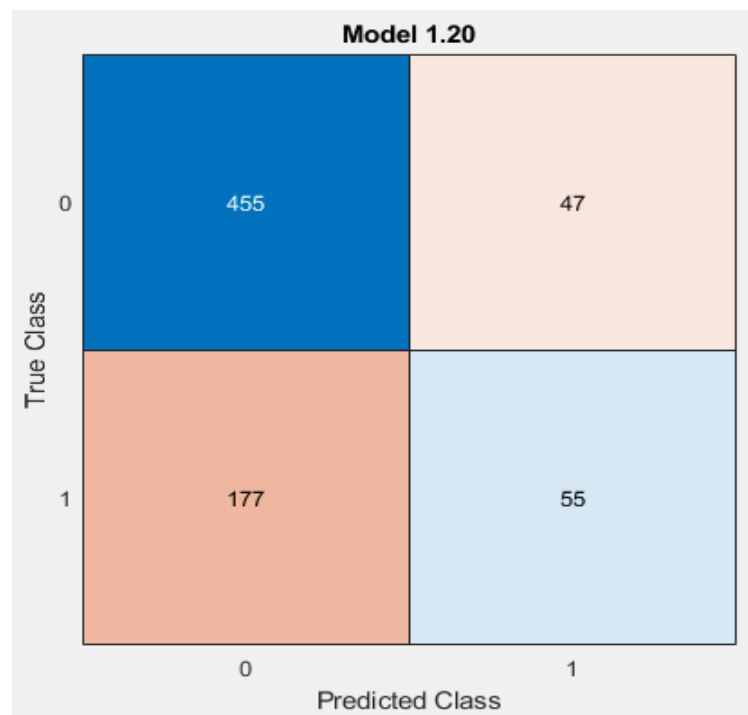
Exactitud = 70.4% Precisión = 55.5% sensibilidad =24.5% especificidad=91.63%

En el desarrollo del experimento 1, los resultados del modelo de predicción sin normalización a través de la matriz de confusión presenta unos datos muy parecidos a los arrojados por el modelo de predicción con normalización, sin embargo, se observa el impacto que genera el modelo con normalización respecto a su especificidad ya que la capacidad de predecir verdaderos negativos mejora implementando el algoritmo con normalización .

- **4.2 Experimento 2: Impacto de la corrección de valores atípicos en el procesamiento de datos.**

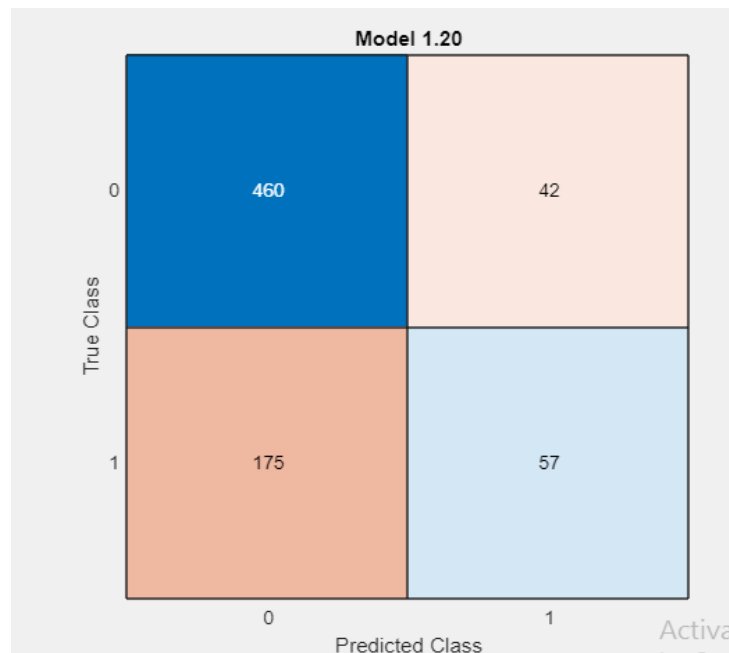
La realización de este experimento se hace con el fin de determinar el impacto de la corrección de los valores atípicos, en este experimento el algoritmo se entrenó con dos códigos en Matlab uno con corrección y el otro sin corrección de datos.

**Figura 13.** Matriz de confusión – Bagged Trees sin corrección de datos atípicos.



Exactitud = 69.4% Precision = 53.9% sensibilidad =23.7% especificidad=90.6%

**Figura 14.** Matriz de confusión - Bagged Trees con llenado de datos atípicos



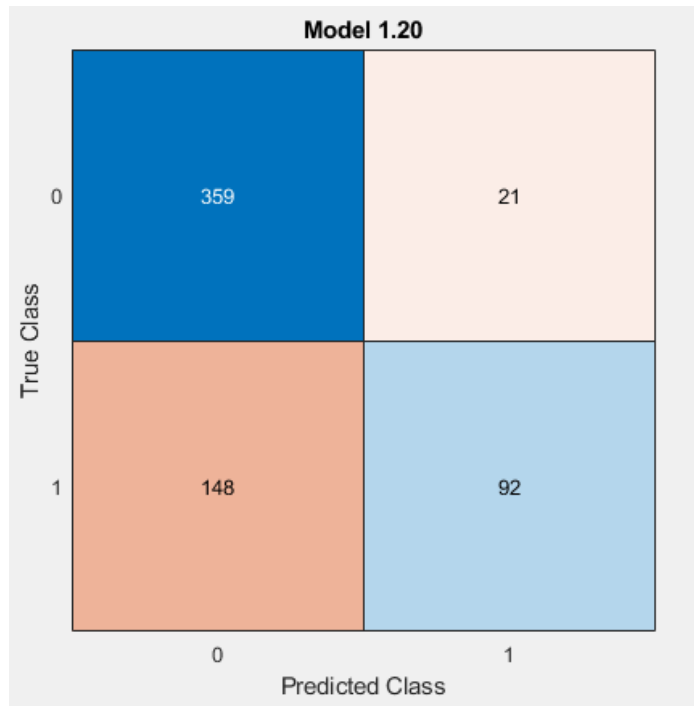
Exactitud = 70.4% Precisión = 55.5% sensibilidad =24.5% especificidad=91.63%

Analizando los resultados obtenidos de nuestro modelo de predicción sin corrección de datos atípicos, se concluye que la corrección de datos no genera un gran cambio en nuestro algoritmo, debido a que el modelo Bagged Trees maneja los valores atípicos y mantiene la precisión de estos datos.

○ **4.3 Experimento 3 Impacto cuando hay una tumba de control (con etiqueta 0).**

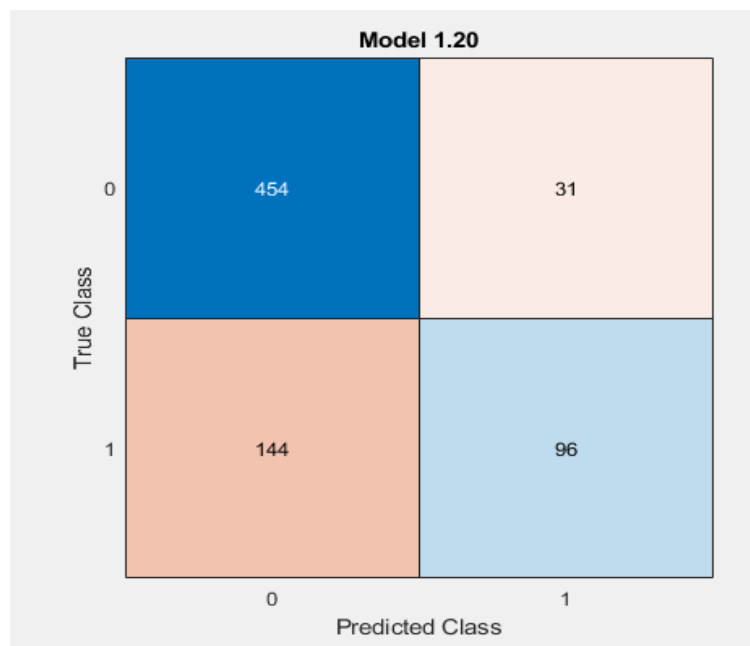
Para el desarrollo de este experimento se implementó en el entrenamiento de nuestro modelo la inclusión de la fosa de control etiquetada con 0, de igual forma se entrena el algoritmo sin fosa de control, para determinar qué tanto se confunde nuestro algoritmo de predicción.

**Figura 15.** Matriz de confusión – Bagged Tree sin fosa de control



Exactitud = 72.7% Precision = 81.4% sensibilidad =38.3% especificidad=94.47%

**Figura 16.** Matriz de confusión - Bagged Tree con fosa de control



Exactitud = 75.8% Precisión = 75.6% sensibilidad =40% especificidad=93.6%

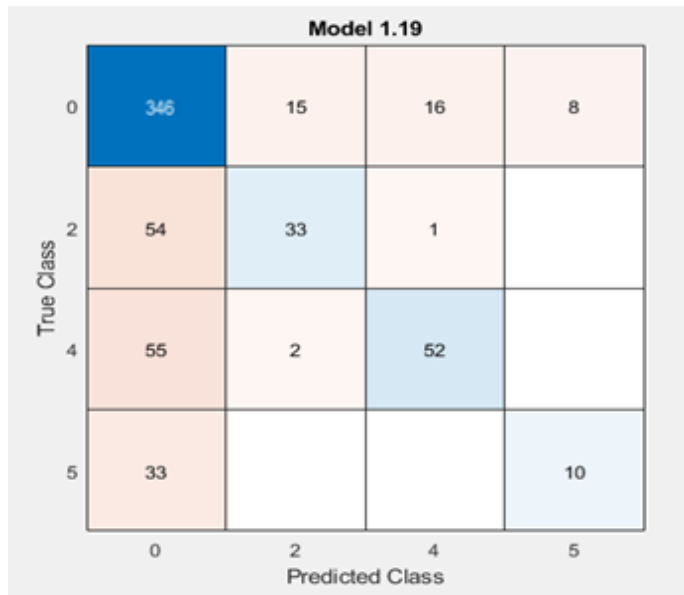


Se concluye de esta fase experimental varios aspectos, la exactitud es más alta en el modelo que se entrenó con fosa de control, sin embargo, como se puede observar el modelo sin fosa de control obtiene una mejor precisión esto se debe al impacto de confusión que genera la fosa de control en el algoritmo de predicción.

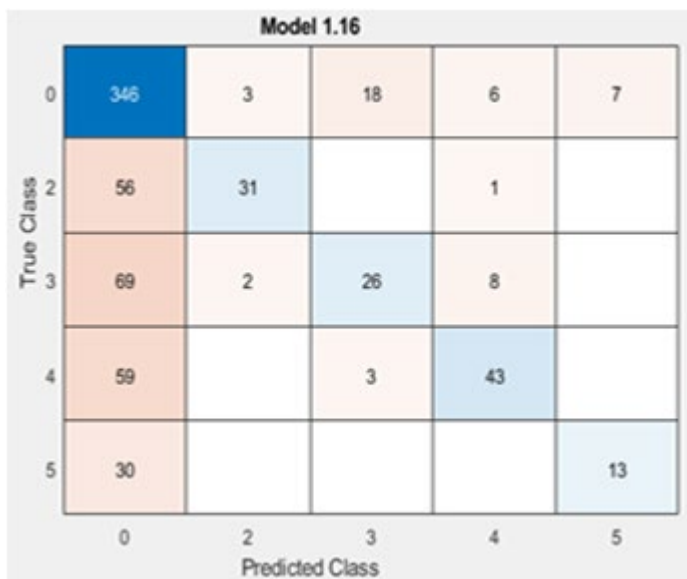
- **4.4 Experimento 4: Impacto en los algoritmos de Machine Learning se confundan cuando se agregan las etiquetas 2, 4,5 y una tumba de control (Grave 3).**

El desarrollo de este experimento consiste en verificar el desempeño del clasificador en identificar, no solamente la existencia de una fosa, si no, en la posibilidad que el clasificador, de manera automática, clasifique el tipo de fosa detectada. Para realizar lo anterior, se utilizan los datos etiquetados de acuerdo a la nomenclatura presentada en la tabla 1. Adicionalmente, en este experimento, se compara el desempeño del algoritmo cuando se considera y cuando no se considera la fosa de control.

**Figura 17.** Matriz de confusión Subspace KNN sin fosa de control



**Figura 18.** Matriz de confusión Boostd Trees con fosa de control con etiqueta 3



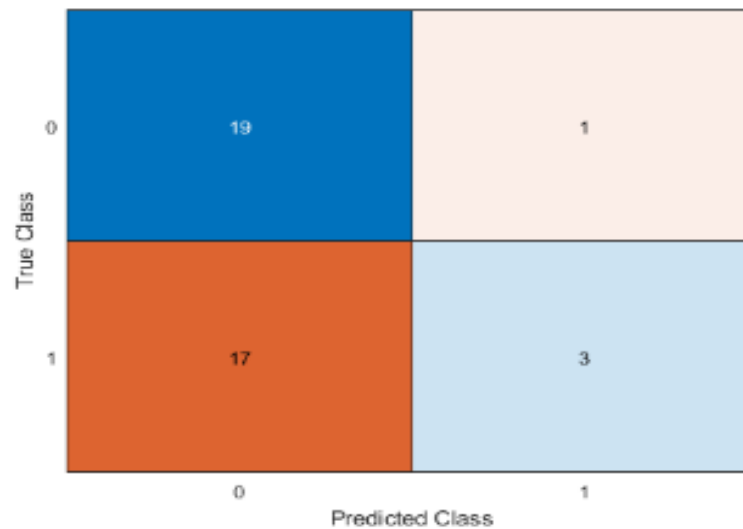
Analizando la matriz de confusión sin la fosa de control (etiqueta 3), donde cada fila de la matriz corresponde al caso en que al algoritmo se le presenta datos de la correspondiente etiqueta (clase verdadera). Se analizan las filas correspondientes a las clases verdaderas. La fosa 5 del modelo de ML multiclase tiene 43 datos de los cuales predice de manera correcta solo 10 datos. También podemos apreciar la confusión que tiene nuestro modelo

de predicción para identificar las fosas: 2, 3, 4, y 5, ya que la tasa de predicción de clases verdadera es menor que cuando con fosa de control: 31/88, 26/105, 43/105 y 13/43, para las clases 2, 3, 4 y 5. Sin fosa de control: 33/88, 52/109 y 10/43, para las clases 2, 4 y 5. Al revisar los resultados, se observa el mejor desempeño para la detección de clases verdades para la etiqueta 2, por lo tanto, concluye que la fosa de control etiquetada con 3 logra confundir más el modelo. Adicionalmente, contrario a lo esperado, los resultados con la fosa de control no indican grandes cambios para la etiqueta 0 (no fosa).

#### ○ 4.5 Experimento 5: Impacto a través del tiempo

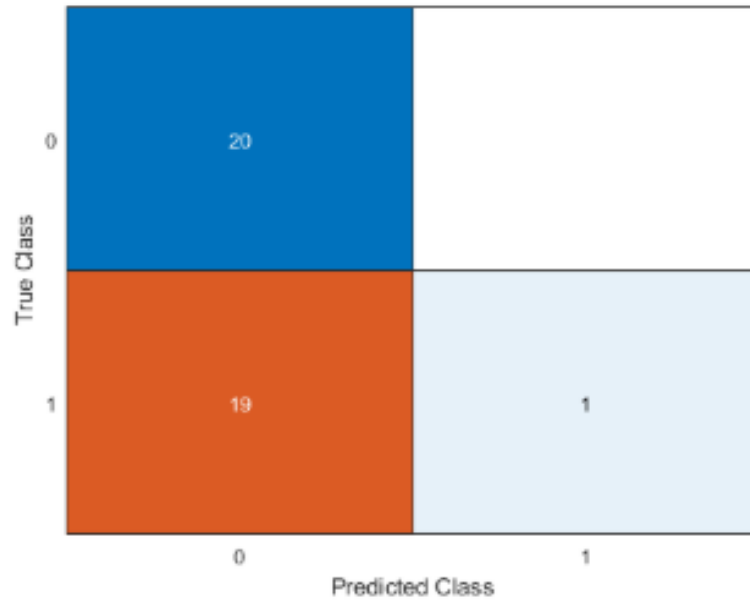
La metodología para el desarrollo de este experimento se basó en el retiro de 40 muestras de cada mes y con los datos restantes se realizó el entrenamiento. El algoritmo entrenado se probará con las muestras retiradas y el rendimiento de la clasificación se registrará a lo largo del tiempo.

**Figura 19.** Matriz de confusión modelo de predicción mes de febrero 2020



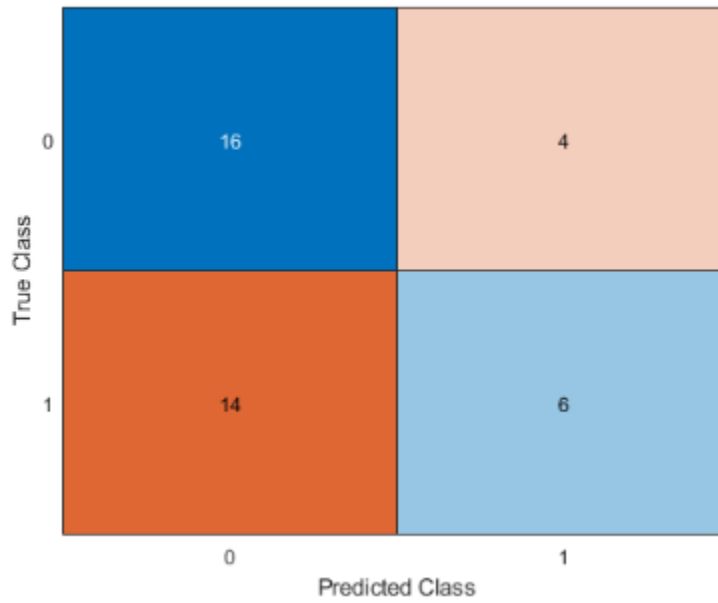
Exactitud = 55% Precisión = 75 % sensibilidad =15% especificidad=95%

**Figura 20.** Matriz de confusión modelo de predicción mes de Octubre 2020



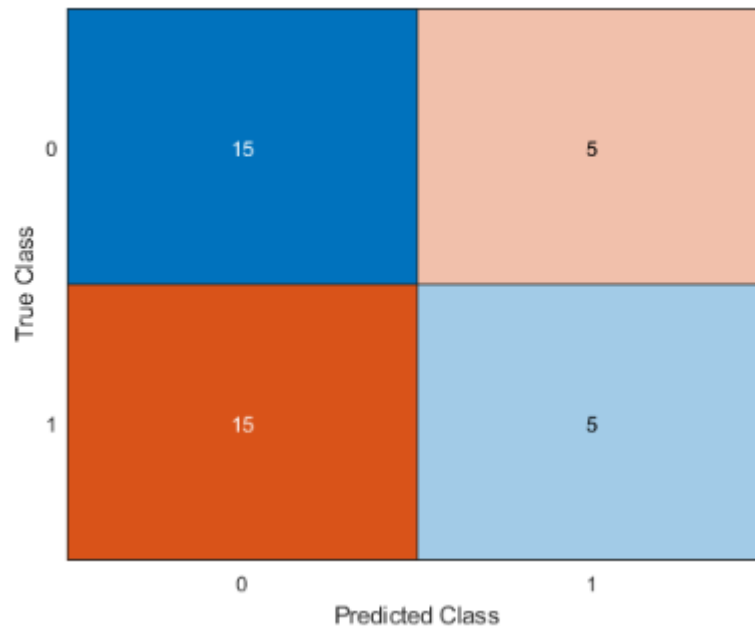
Exactitud = 52.5% Precision = 75.6% sensibilidad =40% especificidad=100%

**Figura 21.** Matriz de confusión modelo de predicción mes de Diciembre 2020



Exactitud = 55% Precisión = 60% sensibilidad =30% especificidad=80%

**Figura 22.** Matriz de confusión modelo de predicción mes de Febrero del 2021



Exactitud = 50% Precision = 50% sensibilidad =25% especificidad=75%

Al observar los resultados obtenidos a través del tiempo, se concluye que los datos registrados no son los esperados. Al culminar este proyecto se quería comprobar que, al pasar el tiempo, la exactitud de nuestro modelo fuera disminuyendo, es decir, a medida que pasa el tiempo, se hace más difícil detectar los enterramientos. Como se visualiza en cada matriz entregada, la precisión no varía estos resultados. Pueden estar asociados a las herramientas utilizadas para obtener los puntos de referencia vinculados a la localización de los electrodos con respecto a las fosas. Lo anterior se puede justificar de la siguiente manera: el desempeño del modelo empeora cuando se retiran 40 datos de cada mes porque se reducen los datos usados para el entrenamiento.

## ▪ 5. Conclusiones y perspectivas futuras.

### .5.1 conclusiones

En este trabajo se evaluó el uso de algoritmos de aprendizaje y método wenner en tomografía eléctrica para la detección automática de fosas simuladas. La investigación permitió probar el rendimiento y la precisión de algoritmos de aprendizaje automático como Bagged Trees; Boosted Trees, Rusboosted y Subspace KNN, a través del procesamiento de datos.

Según los resultados obtenidos se organizaron los perfiles geoelectricos de fosas simuladas obtenidos por un tomógrafo en configuración wenner con el que se realizó el análisis de los datos a partir de los parámetros de profundidad.

En este trabajo se procesaron los perfiles geoelectricos para su utilización en las máquinas de aprendizaje. Los algoritmos fueron comparados entre sí en la Matriz de Confusión, aspecto que permitió establecer la efectividad de cada uno a la hora de ubicar fosas comunes, propósito de la presente investigación.

Según los resultados se aplicaron diferentes algoritmos de clasificación para la detección automática de fosas simuladas es así como luego de varias pruebas con distintos parámetros y en donde se evaluó el comportamiento entre los algoritmos. Por ejemplo, Rusboosted Trees sin normalización se comparó con Bagged Trees con normalización; Bagged Tree con llenado de datos atípicos con Bagged Tree sin llenado de datos atípicos; Bagged Tree con fosa de control con Bagged Tree sin fosa de control y finalmente, Subspace KNN con Boosted Trees. De esta manera se logró determinar que el algoritmo Bagged Trees registró el mejor rendimiento, y la mejor tasa de precisión para detectar los falsos negativos. De acuerdo con los resultados, es bastante acertada ya que de 386 datos donde se indica que no hay fosa, el modelo algorítmico predice 303.

Por sólo enumerar algunos resultados; el modelo de predicción Bagged Trees registró una precisión de 70.4%, mientras que el modelo de predicción Rusboosted Trees obtuvo

una precisión 65.8%. Así mismo, se destacan las métricas de evaluación la especificidad y la sensibilidad.

## **5.2 Perspectivas futuras**

A partir de los estudios y discusiones presentadas en esta tesis, las perspectivas de trabajos futuros se orientan en dos direcciones.

En un primer plano se podría modificar el uso de los datos e implementar las imágenes de tomografía, mediante nuevas técnicas de machine learning.

También sería interesante analizar datos de diferentes técnicas usadas para la prospección en el laboratorio forense, datos de tomografía eléctrica, GPR ground penetrating radar y sensores remotos.



## ▪ Anexos

### Anexo A Experimento 4.1

```
Dataset=readtable ('EXPERIMENTO 1.xlsx')
Columnname={'Var1', 'Var2' , 'Var3', 'Var4', 'Var5'};
for i=1:length(Columnname)-1
    for j=1:(height(Dataset))
        Data(j,i)= str2num(cell2mat(table2array(Dataset(j,Columnname{i})))));
    end
end
end
Datan=table2array(Dataset(:, {'Var5'}));
New_Dataset= Data

Predictores = New_Dataset
tam = size(Predictores);
Respuesta = Datan

One_column = Predictores (:);

PreProcessed predictors = reshape(Normalized, tam);
```

## Anexo B Experimento 4.2

```
Dataset=readtable ('EXPERIMENTO 1.xlsx')
Columnname={'Var1','Var2' , 'Var3','Var4','Var5'};
for i=1:length(Columnname)-1
    for j=1:(height(Dataset))
        Data(j,i)= str2num(cell2mat(table2array(Dataset(j,Columnname{i})))));
    end
end
end
Datas=table2array(Dataset(:,{'Var5'}));
New_Dataset= Data

Predictores = New_Dataset
tam = size(Predictores);
Respuesta = Datas

One_column = Predictores (:);

Normalized = normalize(One_column);

PreProcessed_predictors = reshape(Normalized, tam);
```

## Anexo C Experimento 4.2

```
Dataset=readtable('EXPERIMENTO 2.xlsx')
Columnname={'Var1','Var2' , 'Var3','Var4','Var5'};
for i=1:length(Columnname)-1
    for j=1:(height(Dataset))
        Data(j,i)= str2num(cell2mat(table2array(Dataset(j,Columnname{i})))));
    end
end
end
Datas=table2array(Dataset(:,{'Var5'}));
New_Dataset= Data

Predictores = New_Dataset
tam = size(Predictores);
Respuesta = Datas

One_column = Predictores (:);

Normalized = normalize(One_column);

PreProcessed_predictors = reshape(Normalized, tam);
```

## Anexo D: Experimento 4.5

```
Dataset=readtable('EXPERIMENTO 2.xlsx')
Columnname={'Var1','Var2','Var3','Var4','Var5'};
for i=1:length(Columnname)-1
    for j=1:(height(Dataset))
        Data(j,i)= str2num(cell2mat(table2array(Dataset(j,Columnname{i})))));
    end
end
end
Datas=table2array(Dataset(:,{'Var5'}));
New_Dataset= Data

Predictores = New_Dataset
tam = size(Predictores);
Respuesta = Datas

One_column = Predictores(:);

Normalized = normalize(One_column);

PreProcessed_predictors = reshape(Normalized, tam);

load mymodel.mat

DATOSPRUEBAOCTUBRE = importfile1("DATOS PRUEBA FEBRE 2021.xlsx", "Hoja1", [1, 40]);
ypred = trainedModel.predictFcn(DATOSPRUEBAOCTUBRE);
c=confusionchart(table2array(DATOSPRUEBAOCTUBRE(:,{'VarNames5'})), ypred);
```

## ▪ Bibliografía

- Ariza Lopez, F., Rodriguez Avi, J., & Fernandez, V. (2018). Control estricto de matrices de confusión por medio de distribuciones multinomiales. *Revista Internacional de Ciencia y Tecnología de la Información Geográfica*(21), 215-226.
- Avilés Ponce, L., Cañar Muñoz, M., Reyes Pozo, M., & Mullo Aimacaña, C. (Mayo de 2020). Geofísica aplicada a la búsqueda de restos óseos en el Ecuador. *Revista de Producción, Ciencias e Investigación*, 4(34).

- Barrios Arce, J. (26 de Julio de 2019). *Health Big Data*. Obtenido de La matriz de confusión y sus métricas: <https://www.juanbarrios.com/la-matriz-de-confusion-y-sus-metricas/>
- Camana, R. (2016). Potenciales Aplicaciones de la Minería de Datos en Ecuador. *Revista Tecnológica ESPOL – RTE*, 170 - 183.
- Cardenas Valencia, J. (2011). Manual para la interpretación del perfil de resistividad obtenido al realizar el estudio de la resistividad del suelo a partir de las configuraciones del método de Wenner. *Tesis*.
- Cardenas Valencia, Juan David; Galvis Garcia, Esteban. (2011). *Manual para la interpretación del perfil de resistividad obtenido al realizar el estudio de la resistividad del suelo a partir de las configuraciones del método de Wenner*. Pereira.
- Corso, Cynthia Lorena. (s.f.). Aplicación de algoritmos de clasificación supervisada usando Weka.
- Flores García, W. (Agosto de 2019). El radar de penetración terrestre como una herramienta de geofísica somera, aplicado a la exploración arqueológica. *Revista Geofísica*, 66, 77 -93.
- García Cambroner, C., & Gómez Moreno, I. (S.F). ALGORITMOS DE APRENDIZAJE: KNN & KMEANS. *Inteligencia en Redes de Telecomunicación*].
- Gonzalez Garcia, C. (2018). En qué consiste el aprendizaje automático (machine learning) y qué está aportando a la Neurociencia Cognitiva. *Ciencia Cognitiva*, 12, 48 - 50.
- Martines Hera, J. (2019). *Ensembles: voting, bagging, boosting, stacking*. Recuperado el 2021, de IArtificial.net: <https://www.iartificial.net/ensembles-voting-bagging-boosting-stacking/>
- Molina Gallego, C. (2012). AportEs dE la GEoloGía forEnsE En la invEstiGaCiÓN Criminal En Colombia. *Geología Colombiana*, 37, 171 - 178.
- Molina Gallego, Carlos Martin. (2016). Metodología para la búsqueda de fosas a partir de la interpretación de anomalías en los datos obtenidos mediante la aplicación geofísica de alta resolución. (179). Bogotá: Universidad Nacional de Colombia.
- Molina Gallego, Carlos Martin. (2016). *Metodología para la búsqueda de fosas a partir de la interpretación de anomalías en los datos obtenidos mediante la aplicación geofísica de alta resolución*.
- Montt, C., Castro, F., & Rodriguez, N. (s.f.). Análisis de Accidentes de Tránsito con Máquinas de Soporte Vectorial LS-SVM. *Ingeniería de Transporte*, 15(2), 7 - 14.

- Parra, R., & Palma, M. (2005). Desde el rincón de los muertos y la memoria de sus familiares. Aportes forenses de la antropología a los Derechos Humanos. *Revista de Antropología Iberoamericana*(39), 15.
- Pringle, J., & Jervis, J. (Octubre de 2010). Encuesta de resistividad eléctrica para buscar un entierro clandestino reciente de una víctima de homicidio, Reino Unido. *Forensic Science International*, 202, 1 - 7.
- Pringle, J., Ruffell, A., Donnelly, L., McKinley, J., Hansen, J., Morgan, R., . . . Harrison, M. (2012). El uso de métodos de geociencia para búsquedas forenses terrestres. *Reseñas de Ciencias de la Tierra*, 114(1 - 2), 108 - 123.
- Quispilaya Marmolejo, Robinson; Cruz Chumpitaz, William Jason. (2018). *Propuesta del método no convencional wenner de resistividad eléctrica para caracterización de suelos en institución educativa Libertador San Martín – recuay 2018*". Perú: Universidad Privada del Norte.
- Ramírez Pérez, N., Laguna Estrada, M., & Rubín Ramírez, N. (Octubre de 2017). CLASIFICACIÓN SVM-DIFUSO. *Pistas Educativa*(125).
- Riquelme, J., Ruiz, R., & Gilbert, K. (2006). Minería de Datos: Conceptos y Tendencias. *Revista Iberoamericana de Inteligencia Artificial*, 11 - 18.
- Rodriguez Yunta, L. (2001). Bases de datos documentales: estructura y principios de uso. *La información especializada en Internet*.
- Rojas, C. (2015). *Fortaleciendo el Liderazgo Local del siglo XXI para el trabajo frente al Cambio Climático*. Paraguay.
- Rosado Muñoz, A., Bataller Mompean, M., Frances Villora, J., & Guerrero Martínez, J. (2018). Detección de Fibrilación Ventricular Mediante Tiempo-Frecuencia y Clasificador KNN sin Extracción de Parámetros. *Revista Iberoamericana de Automática e Informática Industrial*, 124 - 132.
- Rosati, G. (2017). Construcción de un modelo de imputación para variables de ingreso con valores perdidos a partir de ensamble learning. Aplicación en la encuesta permanente de hogares. 9(1), 91- 111.
- Sagripanti, L., Villalba, D., Aguilera, D., & Giaccardi, A. (2017). Avances de la geología forense en Argentina: Búsqueda con métodos no invasivos de víctimas de desaparición forzada. *Boletín de Geología*, 39(3), 55 - 69.
- Sagripanti, G., Villalba, D., Aguilera, D., & Giaccardi, A. (2013). Geología forense: métodos aplicados en la búsqueda de desaparecidos en la región central de Argentina. *Revista de la Asociación Geológica Argentina*, 70(1).
- Sanchez Galan, J. (2020). MACHINE LEARNING Y SUS APLICACIONES. *Actas del V Congreso Investigación, Desarrollo e Innovación*.

- Sanchez Muñoz, J. (Octubre de 2016). Análisis de la calidad cartográfica mediante el estudio de la Matriz de Confusión. *Revista de Investigación Pensamiento Matemático*, 6(2), 09 - 26.
- Uriz, M., & et al. (2018). Medidas difusas para la combinación de ensembles: una primera aproximación utilizando el rendimiento en conjuntos de datos altamente desequilibrados. *XVIII Conferencia de la Asociación Española para la Inteligencia Artificial*, 1302- 1307.
- Valencia Valencia , A., & Galicia Haro, S. (2016). Detección de malware con modelo de lenguaje y su clasificación mediante SVM. *Research in Computing Science* 115 , 9 - 18.



▪

▪