

# Análisis del terrorismo: Predicción del perpetrador

## Proyecto de Tesis II

Ingrid F. Ipanaqué C.<sup>1</sup>

<sup>1</sup>Facultad de Ciencias  
Universidad Nacional de Ingeniería

*Asesor: Juan Carlos Espejo Delzo*

August 13, 2021



# Tabla de Contenido

- 1 Introducción
- 2 Estado del Arte
- 3 Recursos y Herramientas
- 4 Metodología de desarrollo
  - Preprocesamiento de datos
  - Modelos de aprendizaje automático
- 5 Resultados
  - Pre-procesamiento de datos
  - Modelos de aprendizaje automático
  - Comparación de los modelos
- 6 Conclusiones y Trabajo a Futuro
  - Conclusiones
  - Trabajo a Futuro



# Introducción

## Motivación

- Estudio de las actividades terroristas.
- Interés por el estudio de modelos de machine learning.
- Interés por el el manejo de macrodatos.



# Introducción

## Objetivos

- Entender el funcionamiento de los modelos de máquinas de vectores de soporte, Árboles de decisión, Bosques aleatorios y Máquina de aumento de gradiente.
- Implementar los cuatro modelos predictivos mencionados anteriormente.
- Mejorar los resultados esperados aplicando distintas técnicas en el preprocesamiento de datos, y diferentes atributos de entrada mediante el análisis factorial datos mixtos.
- Evaluar el rendimiento del modelo implementado.



# Tabla de Contenido

- 1 Introducción
- 2 Estado del Arte
- 3 Recursos y Herramientas
- 4 Metodología de desarrollo
  - Preprocesamiento de datos
  - Modelos de aprendizaje automático
- 5 Resultados
  - Pre-procesamiento de datos
  - Modelos de aprendizaje automático
  - Comparación de los modelos
- 6 Conclusiones y Trabajo a Futuro
  - Conclusiones
  - Trabajo a Futuro



# Estado del Arte

## Conceptos previos

- Aprendizaje automático
  - Máquinas de vectores de soporte
  - Árboles de decisión
  - Bosques aleatorios
  - Máquina de aumento de gradiente
- Métricas de evaluación



# Estado del Arte

## Trabajos relacionados

- Terrorism Analytics, Learning to Predict the Perpetrator. *Disha Talreja, et al. (2017)*.
- TGPM: Terrorist Group Prediction Model for Counter Terrorism. *Abhishek Sachan, et al. (2012)*.
- An Experimental Study of Classification Algorithms for Terrorism Prediction. *Ghada M. Tolan, et al. (2015)*.



# Tabla de Contenido

- 1 Introducción
- 2 Estado del Arte
- 3 Recursos y Herramientas
- 4 Metodología de desarrollo
  - Preprocesamiento de datos
  - Modelos de aprendizaje automático
- 5 Resultados
  - Pre-procesamiento de datos
  - Modelos de aprendizaje automático
  - Comparación de los modelos
- 6 Conclusiones y Trabajo a Futuro
  - Conclusiones
  - Trabajo a Futuro





# Recursos y Herramientas

## Lenguaje y librerías

Para el desarrollo de este proyecto se utilizó el lenguaje de programación Python.

Librerías principales usadas:

- Pandas
- Scikit-learn (sklearn)
- XGBoost



## Recursos y Herramientas

### Global Terrorism Database (GTD)

Se hizo uso de la fuente desarrollada por el Consorcio Nacional para el Estudio del Terrorismo y las Respuestas al Terrorismo (START). GTD es la base de datos más completa de ataques terroristas del mundo. Es de código abierto y proporciona información sobre ataques terroristas nacionales e internacionales desde 1970 y a la fecha incluye más de 190,000 eventos.



# Tabla de Contenido

- 1 Introducción
- 2 Estado del Arte
- 3 Recursos y Herramientas
- 4 Metodología de desarrollo
  - Preprocesamiento de datos
  - Modelos de aprendizaje automático
- 5 Resultados
  - Pre-procesamiento de datos
  - Modelos de aprendizaje automático
  - Comparación de los modelos
- 6 Conclusiones y Trabajo a Futuro
  - Conclusiones
  - Trabajo a Futuro



# Preprocesamiento de datos

Orden seguido para el preprocesamiento de los datos.

**1 Selección de atributos:**

105,203 instancias y 17 atributos.

**2 Limpieza de datos:**

Eliminación de datos 'Desconocidos' o 'Unknown'.

**3 Codificación de etiquetas:**

Uso de la clase LabelEncoder de la librería sklearn.

**4 Manejo de datos faltantes:**

Uso de la clase SimpleImputer de la librería sklearn.

Reemplaza los datos con la media.

**5 Normalización:**

Uso de la clase MinMaxScalar de la librería sklearn.



# Preprocesamiento de datos

## Selección de atributos

Figure: Registro típico en el conjunto de datos.

iyear	2015
imonth	1 (Enero)
iday	1
extended	0
country	113 (Libya)
region	10 (Middle East & North Africa)
latitude	32.069287
longitude	20.151145
vicinity	0
crit1	1
crit1	1
crit3	1 (Outside international humanitarian law)
doubtterr	0
attacktype1	3 (Bombing/Explosion)
targetype1	17 (Terrorists/Non-State Militias)
weaptype1	6 (Explosives)
gname	Shura Council of Benghazi Revolutionaries



## Modelos de aprendizaje automático

- **Máquina de vectores de soporte:**  
Eliminación de grupos con menos de 10 registros. Uso de la librería SVC de Scikit-learn.
- **Árboles de decisión:**  
Uso de la librería DecisionTreeClassifier.
- **Bosques aleatorios:**  
Uso de la librería RandomForestClassifier de Scikit-learn.
- **Máquina de aumento de gradiente:**  
Uso de la librería XGBClassifier de XGBoost.



# Tabla de Contenido

- 1 Introducción
- 2 Estado del Arte
- 3 Recursos y Herramientas
- 4 Metodología de desarrollo
  - Preprocesamiento de datos
  - Modelos de aprendizaje automático
- 5 Resultados**
  - Pre-procesamiento de datos
  - Modelos de aprendizaje automático
  - Comparación de los modelos
- 6 Conclusiones y Trabajo a Futuro
  - Conclusiones
  - Trabajo a Futuro



# Pre-procesamiento de datos

**Table:** Comparación de los resultados para los distintos modelos

Conjunto de datos	Cantidad de atributos	Cantidad de instancias
GTD	135	201,183
GTD procesado	17	105,203

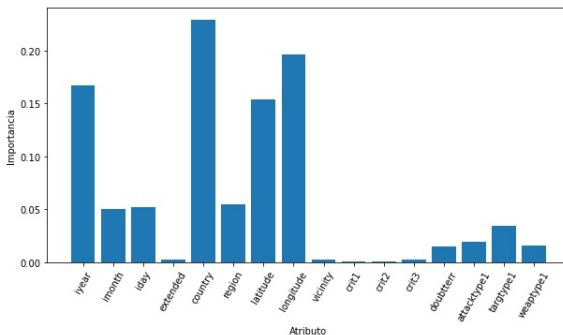




# Modelos de aprendizaje automático

## Árboles de decisión

**Figure:** Importancia de las características del modelo de Árboles de decisión.



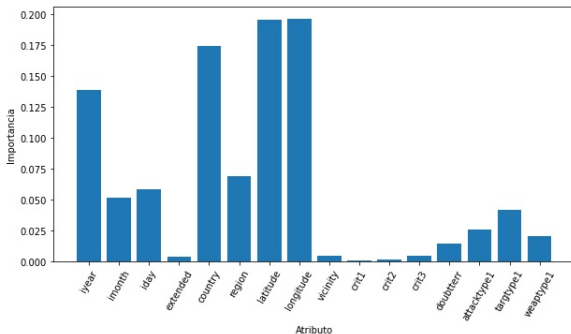
(Fuente: Elaboración propia)



# Modelos de aprendizaje automático

## Bosques aleatorios

**Figure:** Importancia de las características del modelo de Bosques aleatorios.

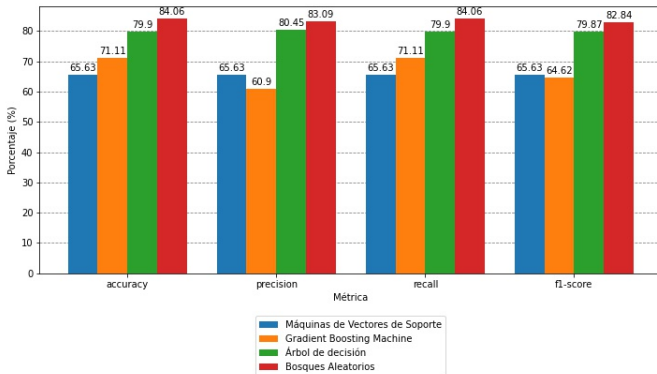


(Fuente: Elaboración propia)



## Comparación de los modelos

Figure: Comparación de las métricas resultantes.



(Fuente: Elaboración propia)



## Comparación de los modelos

**Table:** Comparación de los resultados para los distintos modelos

Algoritmo	Accuracy (%)	Precisión (%)	Recall (%)	F1-Score (%)
MVS	65.63	65.63	65.63	65.63
GBM	71.11	60.90	71.11	64.62
Árbol de decisión	79.9	80.45	79.9	79.87
Bosques aleatorios	84.06	83.09	84.06	82.84



# Tabla de Contenido

- 1 Introducción
- 2 Estado del Arte
- 3 Recursos y Herramientas
- 4 Metodología de desarrollo
  - Preprocesamiento de datos
  - Modelos de aprendizaje automático
- 5 Resultados
  - Pre-procesamiento de datos
  - Modelos de aprendizaje automático
  - Comparación de los modelos
- 6 Conclusiones y Trabajo a Futuro
  - Conclusiones
  - Trabajo a Futuro



# Conclusiones

- El análisis de datos y el aprendizaje automático ofrecen un enfoque prometedor para ayudar a los investigadores a determinar rápidamente el autor más probable de un ataque terrorista.
- A comparación de otros modelos, como Máquinas de vectores de soporte, los modelos de Árboles de decisión no se vieron muy influenciados por *outliers*.



## Conclusiones

- En comparación con los análisis y resultados de los trabajos previos, el modelo desarrollado de Bosques aleatorios y el preprocesamientos de datos seguido durante el proceso de investigación demostraron ser los más adecuados para la predicción del grupo terrorista usando el Conjunto de Datos Mundial sobre el Terrorismo.
- El Conjunto de Datos Mundial sobre el Terrorismo es un ejemplo para macrodatos, por esta razón se demuestra que el uso de Árboles es adecuado, y su rendimiento mejora, para procesar macrodatos.



## Trabajo Futuro

- mejorar el rendimiento del modelo de Bosques aleatorios mediante técnicas para lidiar con clases desbalanceadas en el conjunto de datos de entrada.
- Probar otros métodos, como los clasificadores de conjuntos (ensemble classifiers) y el aprendizaje profundo, para mejorar aún más la precisión de la predicción de grupos terroristas
- Posible aplicación en otros campos.





**Gracias**

