International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Homicide Prediction Model in Bogotá Using the Decision Tree Regression Algorithm

Simanca H. Fredys A.[1], Abuchar Porras Alexandra[2], Anzola John[3], Palacios Jairo Jamith[4], Suarez Roldán Carolina[5], Lugo Manuel Barbosa Guerrero[6]

Submitted: 23/09/2023 **Revised**: 22/10/2023 **Accepted**: 10/11/2023

Abstract: Bogota is the capital of Colombia, and like many capitals in the world it faces challenges related to security and specifically homicides. Throughout the city's existence, the homicide rate has varied due to multiple political, social, and economic factors. These indicators have always been high and quite significant and a constant concern for the city authorities. However, the use of Machine Learning algorithms to predict homicides is a controversial application, but one of growing interest for authorities and experts in Data Mining. For this reason, the development of a Regression algorithm is proposed, specifically the Decision Tree algorithm that predicts the number of homicides in the city of Bogotá, applicable to any city in Colombia, seeking to identify the potential that this tool may have in the planning of prevention strategies. The design and validation of the algorithm yielded an accuracy between 70% and 75%, which is not a desired percentage, but neither can be ruled out in the framework of the use of prediction algorithms. Finally, it is important to point out that this issue should be approached with caution and responsibility and not fall into the promotion of profiles based on stereotypes or the reinforcement of negative stereotypes.

Keywords: *Machine Learning; Homicides; Bogota; Decision Tree*

1. Introduction

Data Mining is part of the field of statistics and computer science and according to it, it arises with the aim of processing information from a set of data, and achieving a mathematical analysis to be able to decide patterns and trends; Achieve conclusions and actions that can contribute to organizational improvement and growth. [1]

The information contained in the data, whether in an organizational context or at the level of people's digital footprint as it relates to social networks, is a fundamental source for the creation and elaboration of new products and services. The behavior of users on these networks is the main basis for the identification of a diversity of people's tastes, such as: personal data, identification data, profession, work environment, musical tastes, gastronomy, sports, fashion, intellect, travel, events, among others. This data is processed by prediction algorithms that allow the analysis of behavior, giving a result that is indispensable for classifying data and performing specific tasks. For this, it is important to know and have different methods to obtain the expected result. [2] [3]

During the last few years, Bogotá has presented an increase in the perception of insecurity, this is ratified by the report issued by the Metropolitan Police of Bogotá where in 2019 1,052 cases were registered, in 2021 1,126 cases, an increase of 7%. The same report shows that the localities with the most homicide cases are: La Candelaria, Teusaquillo and Sumapaz, with increases of between 100% and 200% compared to 2020, Los Mártires with 45%, Tunjuelito with 50%. On the other hand, Ciudad Bolívar and Kennedy continued to be the most violent localities in the city, concentrating more than a third of the homicides.

Universidad Cooperativa de Colombia; fredys.simanca@campusucc.edu.co
Universidad Distrital Francisco José de Caldas; aabucharp@udistrital.edu.co
Fundación Universitaria Los Libertadores; jpanzolaa@libertadores.edu.co
Docente de Planta, Colegio Mayor de Cundinamarca; jjpalacios@unicolmayor.edu.co
Universidad Cooperativa de Colombia; carolina.suarez@campusucc.edu.co
Docente de Planta, Colegio Mayor de Cundinamarca; lmbarbosa@unicolmayor.edu.co
Correspondence: fredys.simanca@campusucc.edu.co

Since homicide averages are changing, it is difficult to identify an exact or detailed cause, so some of this data can be lost and are not found in a classification, but simply remain as undefined, which is why it would be pertinent to expand the field of these deaths and include them in a category in order to have a true and consolidated figure of deaths from homicides.

Over time, Machine Learning has become the most important pillar of artificial intelligence, as it is based on the ability to recognize a large amount of information through different means, which are essential for classifying data and performing quite specific tasks. For this, it is important to know and have different methods to obtain the expected result. Even this tool can be better exploited, since there are different tools that adapt to a graphical environment, making it a powerful and effective element in scientific and technological advancement. In the case of homicides, it becomes interesting and striking to be able to predict the number of people who will be killed by place, date, time, among other variables. [2] [3] [6] [4]

That is why the algorithm that is intended to be designed is of great impact, it presents a prototype of a homicide prediction model in the city of Bogotá based on data from the national police through techniques of pre-processing, integration, cleaning and prediction of the statistical data of homicides of the national police for the year 2022.

2. Materials and Methods

[5]For the design of the algorithm, a methodology based on phases has been proposed, then 4 steps are designed to reach the result of the code elaborated in Python, which allows predicting homicides in the city of Bogotá. The data to be worked on will be using the open data of the Colombian National Police, in the institutional liaison of the Police (Figure 1). [6]

Fig 1. View National Police website homicide statistics [6]

2.1 Data collection

The National Police website has available some of the statistical reports related to homicide over the years, for example, for the year 2023, the homicide report, you have the possibility to download the document in Excel format through the download link https://www.policia.gov.co/grupo-informacion-criminalidad/estadistica-delictiva, and thus have access to the homicides reported for that year. The data needs to be transformed and adapted into a *.csv* file for later reading from the Python programming language with the help of the Pandas library.

2.2 Data pre-processing, integration and cleansing

In this phase, methods from the Pandas library are applied to make the pertinent filters to the information, corresponding to the need that is had, in this case, it is necessary to filter the data to obtain scalable variables to be able to apply, subsequently, the algorithm for predicting homicides.

2.3 Algorithm selection and libraries

[5] [8]For this process, the supervised algorithm is taken with the regression decision tree technique, mainly for accuracy and effectiveness, taking into account that the choice of algorithms depends on the data, since all the data are different.

Table 1 lists the libraries used for the development of the proposed model.

Table 2. Algorithm and libraries of the proposed model

Bookshop	Detail
Pandas	It is a fast, powerful, flexible, and easy-to-use open-source data analysis and manipulation tool, built on the Python programming language.(Pandas, 2020)
Numpy	It is a Python package that stands for "Numerical Python", it is the main library for scientific computing, it provides powerful data structures, implementing multidimensional matrices and matrices. These data structures ensure efficient calculations with matrices. (AprendeIA, 2020)
Matplotlib	It is used to visualize the graph of the analyzed data, in other words, "it is a complete library for creating static, animated and interactive visualizations in Python". (Matplotlib, 2012)
Sklearn	It is a machine learning library in Python, a basic tool to start programming and structuring data analysis and statistical modeling systems. Scikit-Learn's algorithms are combined and debugged with other data structures and external applications such as Pandas or PyBrain. (Alcalá, 2020)

2.4 Model Implementation

In this last phase, the data is already prepared, processed, integrated and cleaned to be able to apply the code that will lead to the results of homicide prediction.

3. Results

After obtaining the data from the Colombian National Police (Figure 2), the algorithm is developed. You must first delete the first 9 rows of the file and proceed to save the file as a csv (***homicides.csv).***

	A	B	C	D	E	F	G	H
1	MINISTERIO DE DEFENSA NACIONAL							
2	POLICÍA NACIONAL							
3								
4								
5								
6	DIRECCIÓN DE INVESTIGACIÓN CRIMINAL E INTERPOL							
7	GRUPO DE INFORMACIÓN DE CRIMINALIDAD							
8	HOMICIDIO INTENCIONAL							
9	PERÍODO DEL 01 DE ENERO AL 30 DE JUNIO DEL 2023							
10	ARMA MEDIO	DEPARTAMENTO	MUNICIPIO	FECHA HECHO	GENERO	*AGRUPA_EDAD_PERSONA	CODIGO DANE	CANTIDAD
11	ARMA BLANCA / C	AMAZONAS	Leticia (CT)	24/03/2023	MASCULINO	ADULTOS	91001000	1
12	ARMA BLANCA / C	AMAZONAS	Leticia (CT)	16/04/2023	MASCULINO	ADULTOS	91001000	1
13	ARMA BLANCA / C	AMAZONAS	Leticia (CT)	18/05/2023	FEMENINO	ADULTOS	91001000	1
14	ARMA BLANCA / C	AMAZONAS	Leticia (CT)	30/06/2023	MASCULINO	ADULTOS	91001000	1
15	ARMA BLANCA / C	ANTIOQUIA	Amagá	01/01/2023	MASCULINO	ADULTOS	05030000	1
16	ARMA BLANCA / C	ANTIOQUIA	Amagá	14/01/2023	MASCULINO	ADULTOS	05030000	1
17	ARMA BLANCA / C	ANTIOQUIA	Amagá	12/06/2023	MASCULINO	ADULTOS	05030000	1
18	ARMA BLANCA / C	ANTIOQUIA	Amalfi	27/05/2023	MASCULINO	ADULTOS	05031000	1
19	ARMA BLANCA / C	ANTIOQUIA	Amalfi	29/05/2023	MASCULINO	ADULTOS	05031000	1
20	ARMA BLANCA / C	ANTIOQUIA	Andes	08/01/2023	MASCULINO	ADULTOS	05034000	1
21	ARMA BLANCA / C	ANTIOQUIA	Andes	12/02/2023	MASCULINO	ADULTOS	05034000	1
22	ARMA BLANCA / C	ANTIOQUIA	Andes	30/03/2023	MASCULINO	ADULTOS	05034000	1
23	ARMA BLANCA / C	ANTIOQUIA	Andes	16/06/2023	MASCULINO	ADULTOS	05034000	1
24	ARMA BLANCA / C	ANTIOQUIA	Angelópolis	25/02/2023	MASCULINO	ADULTOS	05036000	1

Fig 2. Excel visualization of the downloaded data

3.1 Code development

The code built in Python for the implementation of the algorithm is detailed and explained below.

First, the *Pandas* library is imported to read the Excel file. The initial contents of this file are printed (See Figure 3).

```
Import Pandas as PD

homicides = pd.read_excel("homicides.xlsx")

Homicides
```

	ARMA MEDIO	DEPARTAMENTO	MUNICIPIO	FECHA HECHO	GENERO	*AGRUPA_EDAD_PERSONA	CODIGO DANE	CANTIDAD
0	ARMA BLANCA / CORTOPUNZANTE	AMAZONAS	Leticia (CT)	2023-03-24	MASCULINO	ADULTOS	91001000	1
1	ARMA BLANCA / CORTOPUNZANTE	AMAZONAS	Leticia (CT)	2023-04-16	MASCULINO	ADULTOS	91001000	1
2	ARMA BLANCA / CORTOPUNZANTE	AMAZONAS	Leticia (CT)	2023-05-18	FEMENINO	ADULTOS	91001000	1
3	ARMA BLANCA / CORTOPUNZANTE	AMAZONAS	Leticia (CT)	2023-06-30	MASCULINO	ADULTOS	91001000	1
4	ARMA BLANCA / CORTOPUNZANTE	ANTIOQUIA	Amagá	2023-01-01	MASCULINO	ADULTOS	5030000	1
...	...	...	...	...	...	...	...	...
5359	CONTUNDENTES	VALLE	Roldanillo	2023-05-14	MASCULINO	ADULTOS	76622000	1
5360	CONTUNDENTES	VALLE	San Pedro	2023-06-10	FEMENINO	ADULTOS	76670000	1
5361	CONTUNDENTES	VALLE	Sevilla	2023-03-15	FEMENINO	ADULTOS	76736000	1
5362	CONTUNDENTES	VALLE	Trujillo	2023-02-11	MASCULINO	ADULTOS	76828000	1
5363	CONTUNDENTES	VALLE	Yumbo	2023-05-12	MASCULINO	ADULTOS	76892000	1

5364 rows × 8 columns

Fig 3. Initial view of downloaded police data

In the first line of this section, the three columns of interest to be analyzed are selected: The date of the event, DANE Code and Quantity. In the second line, the name of the columns is changed, then in the third line the data is filtered to select only the city of Bogotá, the data index is restarted and finally in line 5 an ID name is given to that index. When printing the data (line 6), Figure 5 shows the result, ready to be analyzed.

```
data = homicides.iloc[:, [3, 6, 7]]
data.columns = ["Date", "City",  "Amount"]
data = data[data['City'] == 11001000]
data.reset_index(drop=True, inplace=True)
data.index.names = ["Id"]
date
```

Id	Date	City	Amount
0	2023-01-01	11001000	1
1	2023-01-01	11001000	7
2	2023-01-02	11001000	1
3	2023-01-08	11001000	1
4	2023-01-10	11001000	3
...	...	...	...
313	2023-06-16	11001000	1
314	2023-06-16	11001000	1
315	2023-06-19	11001000	1
316	2023-06-23	11001000	1
317	2023-06-23	11001000	2

318 rows × 3 columns

Fig 4. View of the selected data for the application of the algorithm

Continuing with the development of the code, in the next segment the Date field is converted to the datetime type. Subsequently, the data is filtered for the dates between January 1, 2023 and May 30 of the same year. Finally, the result data is printed, leaving 264 records (Figure 5).

```
data["Date"] = pd.to_datetime(data["Date"], dayfirst=True)
data = data.loc[(data["Date"]>='2023-01-01') & (data["Date"]<='2023-05-30')]
date
```

Id	Date	City	Amount
0	2023-01-01	11001000	1
1	2023-01-01	11001000	7
2	2023-01-02	11001000	1
3	2023-01-08	11001000	1
4	2023-01-10	11001000	3
...	...	...	...
305	2023-05-16	11001000	1
306	2023-05-18	11001000	2
307	2023-05-24	11001000	1
308	2023-05-28	11001000	1
309	2023-05-28	11001000	1

264 rows × 3 columns

Fig 5. Filter result data

A *data* variable is defined to create indexes by date of daily frequency, and *the number of cases for each of the indices (*Day) is counted in the *for cycle*. Note that in Figure 6, only one record appears for each day of the month and for each record the total number of homicides. For the case of January 1, 2023, 10 cases were registered in the city of Bogotá.

```
data = pd. DataFrame(columns=["Day", "Date", "Number"])

num_dias = pd.period_range(start=data["Date"].min(), end=data["Date"].max(), freq="D")

size = len(num_dias)

for i in range(size):

 num_casos = data["Amount"].loc[data["Date"]==str(num_dias[i])].sum()

data = data.append({"Day": i+1, "Date": num_dias[i], "Number": num_casos}, ignore_index=True)

data
```

	Day	Date	Number
0	1	2023-01-01	10
1	2	2023-01-02	3
2	3	2023-01-03	0
3	4	2023-01-04	1
4	5	2023-01-05	0
...	...	...	...
145	146	2023-05-26	2
146	147	2023-05-27	4
147	148	2023-05-28	12
148	149	2023-05-29	3
149	150	2023-05-30	3

150 rows × 3 columns

Fig 6. Data ready for model application

After having the data ready, the model is created. First, from the *scikit Learn library*, *the* train_test_split *class is imported* to separate the data in training and testing and the *DecisionTreeRegression* class to make the prediction with the decision tree.

```
Import Numpy as NP

Import Pandas as PD

import matplotlib.pyplot as plt

from sklearn.model_selection import train_test_split

from sklearn.tree import DecisionTreeRegressor
```

The independent variable *x is established with* the number of the day and the dependent variable *y*

with the number of homicides. A histogram of the data is also created (See Figure 7).

```
plt.figure(figsize=(15,6))
plt.scatter(x=data["Day"], y=data["Number"])
plt.title("Number of homicides as of May 30")
plt.xlabel("Day");
plt.ylabel("Number of Homicides")
plt.show()
```

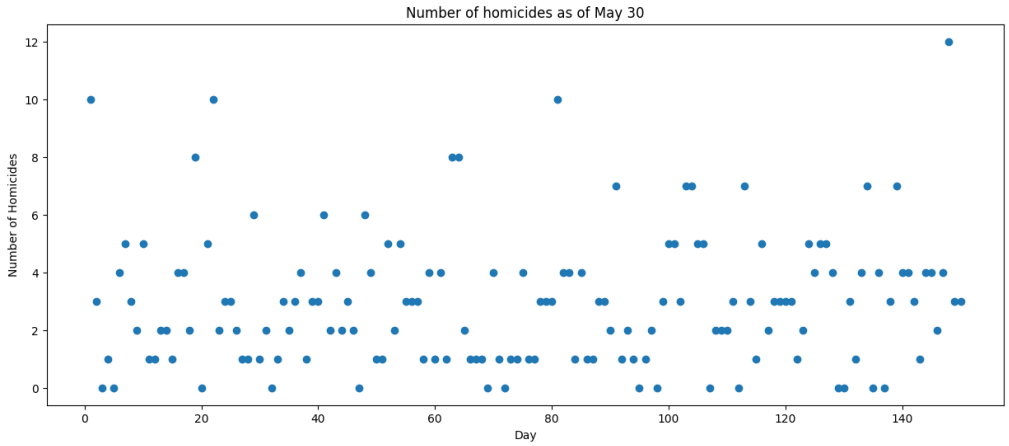

Fig 7. Chart Day of the Year Vs Number of Homicides

On lines 1 and 2 of the segment, the *X and* y variables are determined , followed by the separation of the data into training and testing. Subsequently, the *DecisionTreeRegression class is instantiated*, which will represent the prediction model. And finally, the model is trained and starts predicting.

```
X = data.drop(["Date", "Number"], axis=1)
y = data["Number"]
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=44)
model = DecisionTreeRegressor(max_depth=10)
model.fit(X_train, y_train)
predictions = model.predict(X_test)
predictions
```

Remember that the data is made up of 150 records (150 days), of which 80% was separated for training and 30% for testing the accuracy of the model. There are 30 y_test data, which are shown in Figure 8.

```
array([3.5       , 3.        , 3.        , 0.        , 7.        ,
       2.83333333, 2.85714286, 2.        , 7.        , 2.        ,
       3.        , 2.83333333, 2.83333333, 5.        , 1.5       ,
       0.        , 7.        , 2.        , 1.5       , 2.        ,
       1.        , 7.        , 3.        , 0.        , 3.5       ,
       2.        , 1.5       , 2.83333333, 2.83333333, 1.        ])
```

Fig 8. Predicted data from the 30 test values

To validate the accuracy of the model, it is calculated. It gives us a result between 70% and 75%, values given between the different executions that were made of the model. In the case presented in this article, the accuracy was 74.52%.

```
errors = abs(predictions - y_test)
MAPE = 100 * (errors / y_test)
accuracy = 100 - np.mean(mape)
print("Model accuracy: ", round(accuracy, 2))
```

Model accuracy: 74.52

And finally, a graph is printed to compare the real values (*y_test*) versus the predicted values (predictions*), to have a graphical view of the real values, versus the values predicted by the model based on the Decision Tree (Figure 9).*

```
plt.figure(figsize=(10,5))

plt.scatter(range(30), y_test, label="Real value")

plt.plot(range(30), predictions, label="Predicted value", c= 'red')

plt.title("Regression Decision Tree")

plt.xlabel("Number")

plt.ylabel("Value")

plt.legend()

plt.show()
```

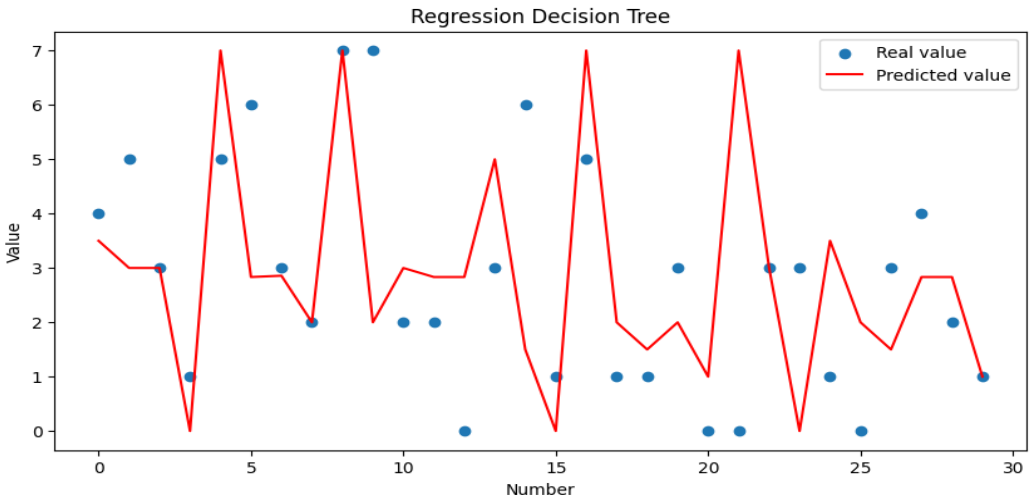

Figure 9. Actual Values vs. Predicted Values

4. Discussion

According to the results obtained, which gives us an accuracy close to 75%, this percentage can allow us to predict possible future values in homicide cases in the city of Bogota....

Within the contextualization of security in the city of Bogota, it can be seen in Figure 7 that the days with the highest number of homicides occurred between Saturdays and Sundays, representing 38% in the visualized data. These data are in line with the quarterly follow-up reports of the comprehensive plan for citizen security, coexistence and justice carried out by the Bogotá mayor's office, which records that 60% of homicides have occurred at night and early in the morning. It is important to note that weekdays (Monday to Thursday) in the afternoon hours is when this type of crime has increased the most.

The proposed prediction model has a remarkably high accuracy, with a specific value of 74.52%. This accuracy, located between a range of 70% to 75%, is indicative of the robustness and reliability of the model in predicting homicide incidences in Bogotá, highlighting that, despite the inherent variability in crime data, which are influenced by a multitude of political, economic and social factors, the model has achieved a significant success rate.

A large part of the usefulness of the model is reflected in the corresponding days from Monday to Thursday, where an accuracy close to 75% was cumulatively obtained, representing most of the values predicted by the model, leaving as outliers the homicides presented on holidays and weekends that correspond to Fridays. Saturday and Sunday. It has been found that the model for the data corresponding to the weekends mostly represents the complementary error rate that corresponds to 25.48%.

The outliers and the error rate are contextually correlated in cultural behavior, since the citizen security reports provided by the Mayor's Office of Bogotá, reflect days such as the first of January (holiday at the beginning of the year) and the

weekends of celebration of Mother's Day, have become days of more violent celebrations in the city of Bogotá, reporting the highest number of homicides being 30% above the average value in terms of annual homicide rates.

The proposed model was presented with a 150-day window, since when expanding a dataset larger than 300, 600, 900 or 1200 days or more, the results obtained did not represent having greater relevance or influence on current predictions. This suggests that the analysis of crime data presents an independence between the most recent events in relation to past events. Contextually, this is because the conditions or factors that affected homicide rates in the past have changed, making homicide data from previous years not indicative of current trends. Therefore, by considering only the last 150 days, we achieved an accuracy of 75%, indicating that the most recent information is more relevant to current predictions.

5. Conclusions

The Classification Tree Regression Algorithm to predict homicide cases in the city of Bogota, according to the results obtained in the validation of the model, is feasible to generate optimal effects.

The possibility of using this type of tool to predict future events that affect society in general allows us to open up a culture of data conservation at all levels, which can provide information and achieve the detection of patterns that in a certain way are capable of preventing or predicting social, economic and political events at large levels. Given that, the acquisition of data or use of Big Data has developed in an accelerated way in the last decade, being reliable and relevant, therefore, it allows the tool to be effective and with high precision if it is handled with the indicated variables.

The city of Bogotá has shown specific patterns in the occurrence of homicides, with notable increases during weekends and certain holidays, such as the beginning of the year and Mother's Day. These patterns reflect cultural behaviors and are supported by citizen safety reports. The proposed prediction model, focused on a period of 150 days, has shown a high accuracy, approximately 75%, in the prediction of homicide incidences, particularly on the days from Monday to Thursday. However, its effectiveness decreases when incorporating data older than 300 days, indicating the importance of focusing on more recent data due to changes in conditions and factors affecting homicide rates. In general, the model has been robust and reliable, although the need to consider the variability and influence of external factors in the prediction of crimes in Bogotá is highlighted.

As a future work, it is planned to evaluate different machine learning techniques to strengthen the predictive model, also incorporating the identification of a greater number of predictor variables, which, although the proposed model has proven its robustness and reliability in the prediction of homicides in the city, it is imperative to recognize that it is a dynamic problem. Therefore, it is necessary to link factors and conditions that affect homicide rates to the model. This means that analysis and predictions must be continuously adapted to the context and time frame, because in the field of data analysis, keeping the information current and relevant is as essential as the accuracy of the model itself.

References

[1] I. H. Witten, E. Frank, and M. Hall, Data Mining: Practical Machine Learning Tools and Techniques, Burlington: Elsevier Inc, 2011.

[2] S. H. Fredys A., B. G. Fabian, L. Jesser, T. C. Wilfred, P. R. Jairo, and B. G. Lugo, "Air Quality Index Prediction Model for the City of Bogotá, DC," Advances in Mechanics, vol. 9, no. 3, pp. 542-553, 2021.

[3] S. H. Fredys A., H. B. Miguel, A. T. Andrés, P. R. Jairo, B. G. Fabián, O. D. Camilo, and B. G. Lugo, "Application of the Polynomial Regression Algorithm to Predict Covid-19 Cases Per Day in Colombia," Advances in Mechanics, vol. 9, no. 3, pp. 49-61, 2021.

[4] R. F. R. Forradellas, S. L. Alonso Náñez, M. L. Ródriguez, and J. J. Vásquez, "Applied machine learning in social sciences: Neural networks and crime prediction," Social Sciences, vol. 10, no. 4, pp. 1-20, 2021.

[5] G. M. Campedelli, "Explainable machine learning for predicting homicide clearance in the United States," Journal of Criminal Justice, vol. 79, no. 1, pp. 1-10, 2022.

[6] F. A. Simanca H., J. A. Cortés Méndez, A. Abuchar Porras, F. Blanco Garrido, J. A. Páez Páez, and J. A. Páez Páez, "Algorithm for predicting the most frequent causes of mortality by analyzing age and gender variables.," Journal of Positive

Psychology & Wellbeing, vol. 6, no. 1, p. 1419 – 1429, 2022.

[7] H. A. Ordoñez Eraso, C. J. Pardo Calvache, and C. A. Cobos Lozada, "Detection of Homicide Trends in Colombia Using Machine Learning," Journal of the Faculty of Engineering, vol. 29, no. 54, pp. 1-20, 2020.

[8] J. Gironés Roig, J. Casas Roma, J. Minguillón Alfonso and R. Caihuelas Quiles, Data Mining: Models and Algorithms, Barcelona: Editorial UOC, 2017.

[9] National Police of Colombia, "National Police of Colombia," National Police of Colombia, 10 8 2023. [Online]. Available: https://www.policia.gov.co/grupo-informacion-criminalidad/estadistica-delictiva. [Accessed 10 8 2023].

[10] J. Z. Mohammed and M. Wagner, Data Mining and Analysis: Fundamental Concepts and Algorithms, Cambridge: Cambridge University, 2013.

[11] E. Russano and E. Ferreira Avelino, Fundamentals of Machine Learning using Python, Oakville: Arcler Press, 2020.

[12] Pandas, 5 October 2020. [Online]. Available: https://pandas.pydata.org/.

[13] "LearnAI," 2020. [Online]. Available: https://aprendeia.com/introduccion-a-numpy-python-1/.

[14] Matplotlib, "Matplotlib visualization with Python," 2012. [Online]. Available: https://matplotlib.org/.

[15] U. d. Alcalá, "SCIKIT-LEARN, A BASIC TOOL FOR DATA SCIENCE IN PYTHON," 2020. [Online]. Available: https://www.master-data-scientist.com/scikit-learn-data-science/.

[16] E. Ribas, «IBES,» 08 JANUARY 2018. [Online]. Available: https://www.iebschool.com/blog/data-mining-mineria-datos-big-data/#:~:text=El%20Data%20Mining%20es%20un,el%20comportamiento%20de%20estos%20datos..

[17] G. . E. Chanchí Golondrino, L. . M. Sierra Martinez and W. Y. Campo Muñoz, "Application of polynomial regression for the characterization of the COVID-19 curve, using machine learning techniques," Research and Innovation in Engineering, vol. 8, no. 2, pp. 87-105, 2020.