5-10-2022

Predictive Policing

Amna Ali Al Boom
aaa9028@rit.edu

Shaikha Khalid Bin Thani

Predictive Policing

by

Amna Ali Al Boom

Shaikha Khalid Bin Thani

A Capstone Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science in Professional Studies: Data Analytics

Department of Graduate Programs & Research

Rochester Institute of Technology

RIT Dubai

10 May 2022

RIT

Master of Science in Professional Studies:

Data Analytics

Graduate Capstone Approval

Students Names**: Amna Ali Al Boom**

Shaikha Khalid Bin Thani

Graduate Capstone Title**: Predictive Policing**

Graduate Capstone Committee:

Name: Dr. Sanjay Modak Date:

Chair of committee

Name: Dr. Boutheina Tlili Date:

Member of committee

Acknowledgments

We would like to thank our mentor Dr. Boutheina Tllili for her support and encouragement through our project. Working with Dr. Boutheina on the capstone project has been very insightful and helpful in our journey. We would also like to show appreciation to the chair committee Dr. Sanjay Moudak for his ongoing support and patience through the journey. There were some challenges that we as students faced, however having Dr. Boutheina Tllili as a mentor and Dr. Sanjay as our chair committee of the data analytics program has sure made our journey better and useful. I would also like to thank Dr. Yousef Assaf the president of RIT Dubai for giving us a chance in learning and experiencing this amazing field . Finally, I would like to thank Colonel Rashid Dhabooi and Major Khalid Abdulrahman from the criminal investigation department at Dubai Police for their amazing help and fast response in giving us access to the data and supporting us through our capstone project.

Abstract

UAE is one of the safest countries to live in, but that does not indicate that the country does not witness crimes, During the COVID-19 pandemic, the country saw an increase in cyber and digital crimes.

Apart from cybercrime, there are other types of crimes, such as street crimes and violent crimes.

Data analytics aids Dubai Police to predict crimes. Criminal investigation is one of the fields that is very interesting and is taught in colleges and academies. Data analytics opens the door for studying the details of each crime.

Data mining tools consist of a variety of techniques that can help solve a problem or indicate a cause or an effect of something.

Data analysts use data mining tools through a lot of software that allow the user to analyze data easily and fluently. SAS (statistical analysis system) is one of the reputable software that is used especially for visualizing and analyzing data. In this capstone, we will use SAS since it is a software that is accredited from Dubai Police and we use it already in our workplace.

Prediction techniques supports to interpret and facilitate Dubai Police to develop strategies to reduce the crime rate. Hence, it allows UAE to sustain its position as the "safest" country. The capstone idea will actually help us develop what we do at work and stop or reduce crime which is one of the main pillars in Dubai Police. The crime related data will be collected from CID in Dubai police.

Link analysis and predictive analysis will be performed in this project to forecast any crime. We will build a predictive model using SAS to predict crime. This proposed project will help to identify the trends of historical crime data. Project timeline has been provided in this writing to have a better outline.

The first step is to collect the data from the source which is in our case, the criminal investigation department in Dubai Police. Meeting with the department; they have agreed on giving us datasets of specific crimes that Dubai Police finds critical and needs further analysis from five years. Thus, the data that we will be analyzing will be from the years 2017 to the year 2021. After collecting the data ; the processing took place which is the cleaning part of the data. Since the data is in Arabic and it is old as mentioned earlier that the data of the past five years are collected; there are some missing fields, some inconsistencies and some redundant data. After cleaning the dataset which took 70% of the time working on this project. Now the dataset is ready and can be analyzed in SAS. Importing the dataset through SAS was the first step. Then, we started analyzing the criminals first as we wanted to build a portfolio of the criminals and observe of any patterns found. The highest nationality of the criminals was India. We tried to see if there are higher nationalities in certain years, but in all five years the analysis showed that India was the number one nationality in criminals. Then we wanted to observe the criminals' education level; the highest education level was unemployed meaning they do not have any degree that supports them. The education level

part was very interesting because we found out that even though university degrees did not come first in the highest education level. however there is a sample of the criminals that hold very high level degrees such as PhDs and Masters degrees and this shows us that the stereotype of how uneducated people are bad or are the only people that commit crimes should be disregarded. Next , we analyzed the criminals' age group and the outcome was that 30 – 45 age groups are the ones that commit crimes the most in Dubai. Finally, we have analyzed the criminals' gender to see which gender commits most crimes in Dubai and from our analysis; the outcome showed that men are the most that commit crimes in Dubai. After analyzing the criminals' profiles ; we have moved on to analyzing the crimes in the past five years. The type of crime was the first thing we wanted to analyze to observe what is the most crime committed in Dubai in the last five years. Fraud was the most crime committed in Dubai and this was not a huge shock to us since Dubai is considered a business city and it attracts some people to do their business in it. Dubai has always been interested in building the city financially in the best , legal way possible, however there will always be people that see it as a city to commit fraud in since it has a large population and has many tourists visiting the city. Next, we analyzed the crime replotting per year. 2019 has scored the highest in crime reporting in Dubai; right before the pandemic. We analyzed the police stations that had the most reporting in the past five years in order to observe the locations that are considered crime appealing to criminals. This analysis is very important since every area has a police station assigned to it and the outcome of this analysis was that Bur Dubai police station had the highest number of incidents in the last five years. Lastly, we wanted to analyze what time was the crime committed and the result was that most crimes have been committed in the morning between 9AM and 11AM and that was very shocking and interesting to us because it is know globally that most crimes are committed at night in the dark where no one can see the criminal , but this is due to the type of crime as well , and as we have observed that fraud is the most committed crime, then the morning is the best time to commit this crime since people are awake and willing to do business with other people whether it was online or offline. Finally, the purpose of this whole project is to forecast the crime rates; thus, we built a forecasting model in SAS and it showed us that in the upcoming years, the crime rates in Dubai will decrease dramatically based on the pattern of crimes in the historical data. This is a positive result; however this does not mean that Dubai Police should neglect the surveillance and monitoring of the city due to this forecasting as it is not always accurate.

Keywords: Forecasting, Crime, Prediction , Criminals, SAS, Preprocessing data

Table of Contents

List of Figures

List of Tables

Chapter 1

1.1 Introduction

Various types of crimes are affecting the whole world. Due to this problem, many organizations, including individuals, suffer the most. Data mining is an important way to extract data from a large set of data. On the other hand, it can be said that this is a type of approach that can be used to discover any hidden data from the dataset. However, there are different types of data mining tools and applications that have been found in that market that have been used for detecting crime. As per the views of David (2017), criminology is an essential field to apply data mining. In other words, it can be said that data mining helps to forecast crime before it happens. It prevents such activities, and it allows the government and other security employees to be alert. In addition, it can be said that the identification of crime characteristics is the main approach for the prediction. Every analyst in this field tries to identify the characteristics of the crime by collecting historical data.

Therefore, it can be said that data mining approaches help to gain proper knowledge and information for the crime that helps the security forces. This allows preventing any big accidents that tend to occur. As per the comment of Ye *et al.* (2017), entity extraction, clustering technique, deviation detection, and classification are the main data mining techniques that have been used for crime detection. Depending on the situation and the circumstances, these data mining techniques have been used to detect different types of crimes. Many law enforcement agencies collect different information about various crimes, and data mining techniques have been used by the analysis to forecast the crime scenes (Madni *et al*., 2017). Depending on this scenario, this project will focus on identifying the crimes with the help of data mining tools.

Global Database Number for 2020 on safety ranks shows the UAE has ranked the third position as the safest country in the world. It gained 84.55 % on the safety index score and 15.45% on the crime index score. The crime levels that are less than 20 are very low, between 20 and 40 is low and crime levels between 40 and 60 are moderate. But if the score is between 60 and 80 it is high, and over 80 indicates very high according to (Gamboni, 2020).

There are different types of crimes such as antisocial behavior, burglary, child abuse, cybercrime, and online fraud, domestic abuse, fraud, hate crime, modern slavery, murder, rape, and sexual assault, robbery, sexual harassment, stalking and harassment, terrorism, violent crime (Victim Support, 2020).

According to 'UAE 2020 Crime and Safety Report: Dubai', Dubai is a low-threat location. The report points out street crimes like shoplifting, pickpocketing occurs, but sparse use of weapons.

Violent criminal activities are rare and mostly takes place within third-country nationalities. But cybercrimes are prevalent across UAE (OSAC, 2020). The crime rate statistics of UAE shows there is a fluctuation every year by increasing or decreasing of crime rates.

In 2019, the cybercrime rate doubled; emphasizes the need to address the issue on priority (Bitar, 2019). According to Blue Voyant (2021), the top five cybercrimes are phishing scams, website spoofing, ransomware, malware, and Internet of Things hacking. According to the Institute 'Cost of Cybercrime

Study' 2019, cybercrime is dynamic and everchanging. Moreover, it takes a long time to resolve and is expensive for organizations. By increasing cybersecurity protection, it is possible to reduce cybercrime costs.

Due to the pandemic in UAE, businesses expect to see considerable changes in the coming year as phishing scams (61%), email scamming (54%), online scams (49%), security breaches (44%), and ransomware (42%). Hence to combat the cybercrime incidents, the UAE Banks Federation, the Central Bank of UAE, Abu Dhabi Police, and Dubai Police launched UAE's first national fraud awareness campaign to educate the public and protect the consumers from financial cybercrime and fraud. (KPMG, 2021). In Dubai, during pandemic cyber-attacks increased and the top are phishing and ransomware incidents (Clarke, 2021). Hence, to sustain the position as safest country, UAE must predict crimes ahead and prevent it.

1.2 Project goals

The main goal of this project will be to detect different types of crimes and predict a crime before it happens in order to prevent it from happening. In other words, it can be said that this is the aim of this project. Depending on this goal, the deliverables of this project will be provided later. The other goal of this project is to use the CRISP – DM methodology as the main process of our project. Then, implement data mining tools via SAS in order to build a predictive model. Different Types of data mining tools are available in the market, such as Oracle data mining, Python, Rapid miner, SPSS modeler etc. However, in this project, SAS will be used as the main data-mining tool to predict crime. Therefore, it can be said that this project will define the way of crime prediction with the help of this data-mining tool. It will help to gather the knowledge related to the application of SAS as the data-mining tool by implementing link analysis as the data mining technique. The objectives of this project will be:

- Use the CRISP – DM methodology as a pillar for our project.
- Identify the critical crimes that need to be predicted.
- To collect historical data related to crime.
- To investigate the data mining process and tools in crime detection.
- To implement a predictive model using SAS program to predict crime rates.

1.3 Aims and Objectives

Crime has been one of the main pillars for tourists and citizens when asked whether they are happy in the city or not. As Dubai always aims to be the happiest and safest city, in this study we aim to focus on the crimes and use the data analytics techniques we learned in our studies in order to predict the future of crime in Dubai. The objective of the study is to detect different types of crimes and predict crime rates in the upcoming years. The aim of the study is to predict crimes using data

mining tools. Our main aim is to find out whether crimes will increase or decrease in the future and what are the factors that affect this prediction. Using CRISP – DM methodology, we aim to reach our goal which is to accurately forecast the crime rates in Dubai.

The objectives of the process of this study are as follows:

- Collect data from Criminal investigation department.
- Clean the data and ensure the dataset is ready for analysis.
- Use the data analytics tools and techniques to analyse the data.
- Implement data visualization.

Research Methodology

Data collection is a very important process in research work. As per the comment of Archibald et al. (2019), depending on the data, the success of the whole researcher is dependent. Therefore, in this research, data will be collected from CID (criminal investigation department) in Dubai police. In this case, all the data of several crimes will be documented in this data so that the data mining will be performed. As per our understanding of how prediction in data mining occurs, we have identified that we need at least the crimes that occurred in the last five years. We will identify and tackle specific critical crimes and not random or all crimes, because we realized that this might occur a kind of confusion and might mess up the model. We will meet with the CID in Dubai police in order to understand what are the critical crimes that need to be tackled in this project and need to be predicted. This will allow performing predictive analysis to prevent any crimes. In this process, all the historical data will be collected regarding critical crimes in Dubai. This will help the CID of this country to prevent any big losses.

The CRISP – DM methodology will be followed as:

Business understanding:

This project is very intense and serious as it tackles a problem that every person in society is affected by. Hence, understanding the needs of Dubai police and what they aim to reach alongside their mission and vision is crucial.

Data understanding:

The project will be a total failure if the analysts do not understand the data and what it represents. Clear understanding of each variable, attribute, feature of the data gives the analysts the vision of how to fix the issue and problem.

Data preparation:

As mentioned above, the data we asked from CID is specific critical crimes and not all of the crimes that occur, because we want to prevent any confusion and distraction. The data needs to be prepared well for building an accurate predictive model.

Modeling:

SAS will be our main tool for analysis and predicting techniques will be applied for this project.

Evaluation:

We will evaluate the data and decide whether the predictive model used in SAS meets the objective of our project which is in fact predicting crimes before it happens.

Deployment:

We hope if the project is a success, that we deploy and execute it to Dubai Police and present it to CID and the commander chief to help reduce crime and make Dubai a safer place.

1.4 Limitations of the Study

This study has some limitations due to the sensitivity of the data and the topic such as prediction and forecasting methods in data mining are not always accurate and the measurements or results might be wrong.

The cleaning process of the data as also it was in Arabic and we faced a bit of difficulty in translating the data with this size issue so, after searching we found a way by using google forums commend which was so useful and time saver.

The limitation of the confidentiality of the data and the level it needs to be secure and not used in open-source programs could be an issue, because some more analysis can be done using these open – source programs to have an accurate estimation of the results.

Another limitation is that the data is imported from a system in Dubai police and they are all real – life cases and crimes, however sometimes there are crimes that do not get recorded in this system due to a negligence from an employee or even that the crime has been disregarded by the person who filed a report so it gets deleted from the system or not recorded.

This will give us an inaccurate estimation since we are using forecasting and prediction tools which might affect the accuracy of the study.

Chapter 2 – Literature Review

2.1 Introduction

The literature review has been done extensively throughout the study in order to observe and benchmark what other organizations have done in regard to crime prediction and forecasting. We found a lot of sources and information that helped us in the journey of the study.

2.2 Data mining applications:

Data mining applied in various fields; there are various commercial data mining system available in the market to meet specific needs in the following fields.

a) Financial data analysis:

In the financial sector such as banks and financial services, data mining provides high-quality and reliable data analysis. As a result, the data mining tool facilitates to forecast of loan payment and customer credit policy, categorizes and cultures customers to support targeted marketing, detects money laundering and other financial frauds and crimes (Tutorial Point, 2021).

b) Telecom industry:

The Telecom industry provides a wide range of products for individuals, homes, and businesses. The emerging technologies and advancement of communication and IT technologies brought drastic changes in the telecom industry. The data mining application in the telecom industry supports identifying the telecom patterns, fraudulent activities, utilizing the resources, enhance quality. For example, it performs multidimensional analysis of telecom data, identifies usual patterns, fraudulent patterns, multidimensional association, sequential pattern analyses, mobile telecom services, and utilization of visualization tools (Tutorial Point, 2021).

c) Life science – Biological field:

Genomics, functional genomics, proteomics, biomedical research is the gained more attention in recent years. Data mining contributes more to the biological field as it facilitates enhancing data analysis. It ensures semantic integration of heterogeneous, proteomic databases and distributed genomic. Furthermore, it performs aligning, indexing, comparative analysis, similarity identification, multiple nucleotide sequences. Likewise, functions identify the structural patterns, examine the genetic networks and protein pathways and find links and path analysis. The visualization tools in genetic data analysis support the interpretation of the data efficiently (Tutorial Point, 2021).

d) Application in other scientific areas:

Like biological field, data mining application is applicable for geosciences, astronomy, and other fields that have large data sets. As a result, it performs numerical simulations in various areas like ecosystem modelling, climate, fluid dynamics, chemical engineering. Thus, data mining facilitates to performance of data warehouses, data processing, graph-based mining, visualization, and domain-specific knowledge.

e) Retail Industry:

Data mining supports the retail industry to gather data concerned with sales, goods, purchasing history, transportation, services, and consumption. In the retail industry, data mining facilitates recognizing customer purchasing behavior patterns and trends. Consequently, it allows enhancing customer service excellence and customer retention, and customer satisfaction. Thus, data mining brings multidimensional analysis of customers, products, sales, time, and region. Also, it helps to recommend products and cross-reference items related to personalized ads, customer retention, and assess the success of sales campaigns (Tutorial Point, 2021)

f) Intrusion detection:

In the digital era, security is a fundamental and critical issue that requires attention. Therefore, intrusion detection in the network is important in network administration. Thus, data mining supports in creating data mining algorithms specific to intrusion detection, correlation analysis, aggregation, and discriminating attributes, examining stream data, distributed data mining, and visualization and query tools (Tutorial Point, 2021).

g) In police sectors (crimes) :

The crime sector is rich of data, as long as you gather, ore you will be getting a good result, the methodology, which has been used is formulating by asking questions, main question generating parts under it which helps mining the data.

The question state as first the subject, why do we want to explain? Then the interest domains and then why inquires in order to be closer to the matter as I explained it below in the figure.

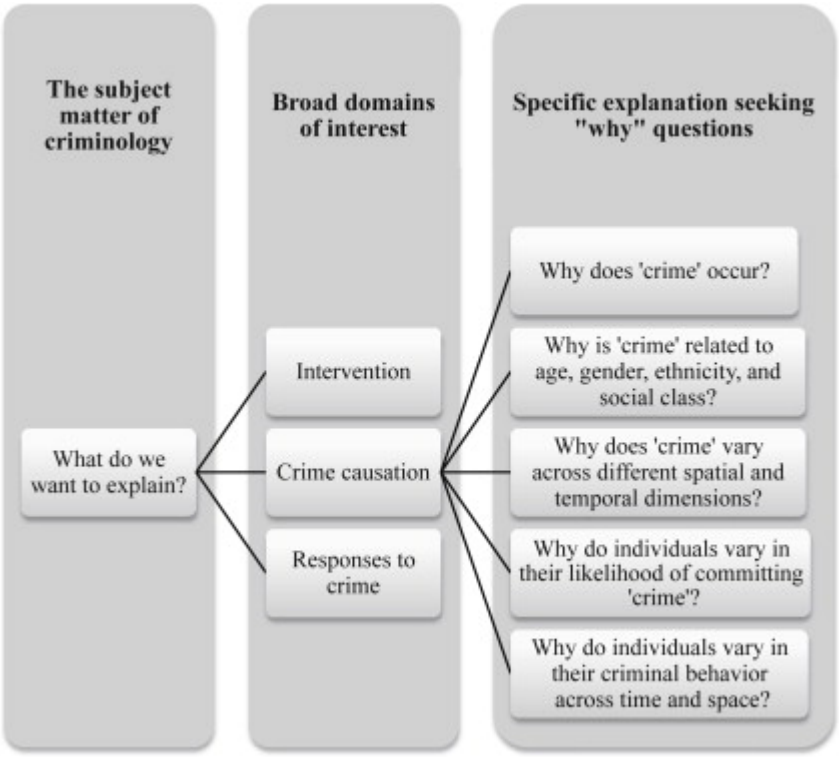


Figure 2: The questions
(Source: Self-created)

h) Selection of data mining system:

Data mining system selection depends on the following factors.

- Data types
- System issues
- Data sources
- Data mining functions and methodologies
- Coupling data mining with database/ data warehouse systems
- Scalability
- Visualization tools
- Data mining query language and graphical user interface (Tutorial Point, 2021).

2.3 Importance of data mining in crime detection

Data on crime is very important for justice professionals as well as for security purposes. Therefore, data mining is considered the most crucial technique in crime detection. Data mining is a type of data analysis. According to the vision of Zhao (2018), different law enforcement agencies depend on data analytics in order to find connections. In addition, different governments also initiated this data-driven approach to predict crimes. In addition, it can be said that data mining helps to gather all the large data related to varieties of crimes. Data mining clustering techniques have been used in most cases. It mainly helps to deal with huge amounts of data, including any missing data as well.

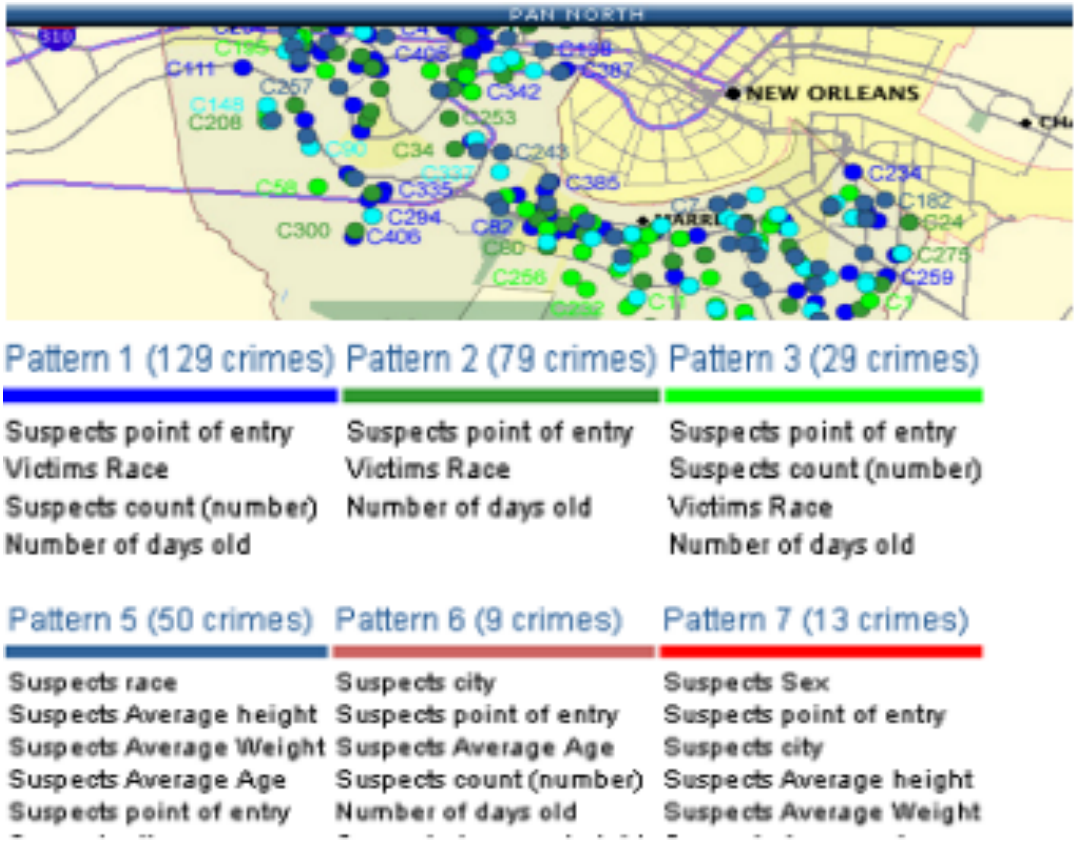


Figure 2: K-means clustering
(Source: Ghorbani *et al*. 2018)

This figure shows the application of this clustering technique in order to identify the crime patterns. This particular study is only based on this technique to predict crime (Ghorbani et al., 2018). In this case, all the crime data have been collected from the crime reports and based on that the analysis has been made. Similarly, there are many techniques that have been used in criminology. However, it can be said that data mining is very important in this particular field. Hence, data mining helps to provide several convenient and practical ways to analyze hidden information about criminal cases.

2.4 Application of several data mining tools in crime detection

Different types of data mining tools have been used in crime detection. It is a type of activity that has been performed with the help of some algorithms. Therefore, it can be said that this is a type of scientific and intelligent technique for detecting crimes. It has been identified that there are different types of data mining tools found in the market, and those are Oracle data mining, Python, SPSS modeller, Rapid miner etc. (Al, 2021). Some of the commercial data mining system available are Rapid Miner, Oracle Data Mining, IBM SPSS Modeler, Python, Orange, Knime, Kaggle, Rattle, Weka, Teradta, Apache Spark, H2O, Sisence and Xplenty (Sarangam, 2022). All the descriptions of these tools are provided in this section that is found in this study.

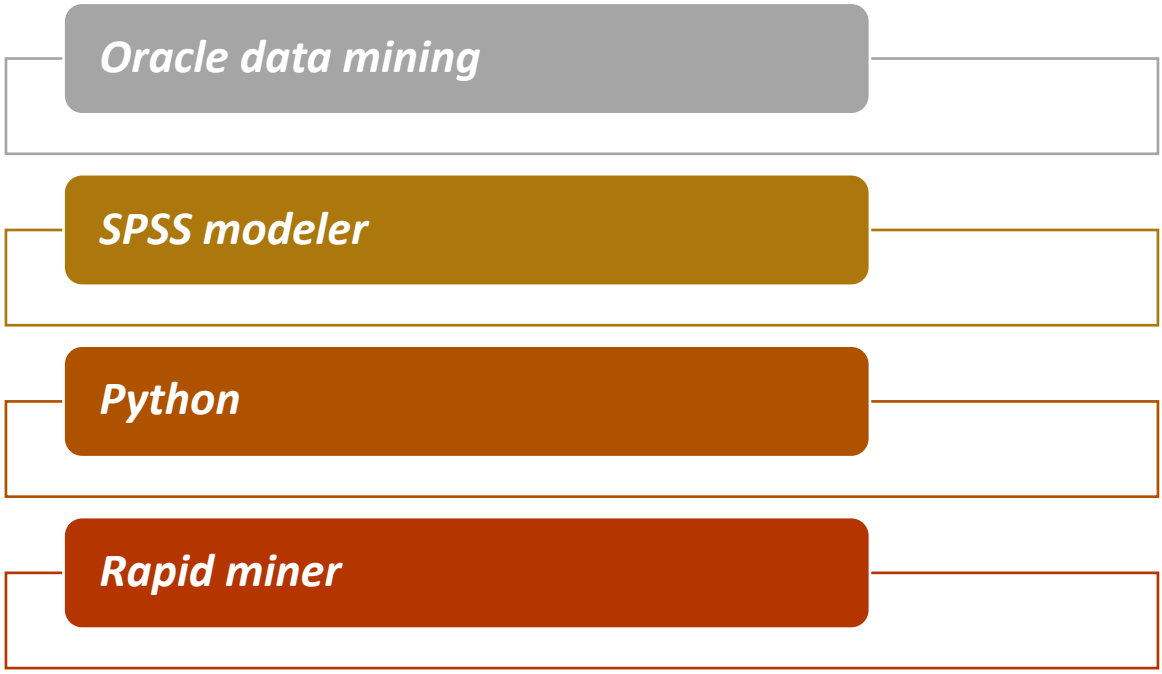


Figure 3: Application of several data mining tools

(Source: Self-created)

Oracle data mining

Oracle is popular software in the IT world. It mainly helps to improve the database technology in this field. As mentioned by Walia (2020), the analytical tool of oracle helps to make it powerful for data analysis and data mining. Therefore, it has been identified that all; the algorithms in this tool are directly integrated with the database kernel. In addition, it also helps to extract data from standalone servers. However, this tool provides GUI and different testing models to predict crime.

SPSS modeler

SPSS modeler is also a popular name when it comes to data mining. This tool also has many data mining algorithms that can be used in the crime detection process (Zhuang *et al.*, 2018). This particular tool provides drag and drops facilities to the user so that they can easily handle large sources of data related to any crime. Therefore, this software has been used to prepare the data for predictive analysis with the help of proper deployment.

Python

Python is a well-known programming language, and it is an open-source language. Python has been used both in development and analysis works. This language consists of a large set of packages that can be used for several purposes (Tang *et al.*, 2019). In this matter, it has been found that python has wide acceptability in the field of data mining. This is very powerful in the field of predictive analysis and visualization.

Rapid miner

Rapid manner is a type of data science software, and it provides an integrated environment for handling large datasets. Different functionalities like data visualization, cleaning, and preparation are found in this software. In addition, it also helps in data mining and predictive analysis.

2.5 Successful Application of Data Mining in Crime Prevention:

One of the recent studies explored data mining for crime prevention. The study acknowledges that data mining facilitates crime prevention. The study focused on the crime dataset of Gujrat Police, India. The authors developed a predictive model to examine the crime pattern by using both supervised and unsupervised data mining systems. Thus, they applied Multiple Linear Regression, K-means Clustering, and Association Rules Analysis (Singh et al., 2018).

Norouzi & Ateie (2021) used data mining to identify and discover hidden theft patterns. The study proposed a data mining model by applying two coupling rules with the existing probability algorithm and clustering with the Chi-mean algorithm to identify and discover the hidden patterns.

Another study explored the use of data mining techniques in crime prediction and prevention. The authors underscore the application of machine learning and data mining techniques like neural networks and clustering in predicting or preventing crimes. The findings confirm that data mining plays a crucial role in predicting or preventing crimes that differ in severity levels such as cybercrimes, city crime, fraud, and terrorist attacks (al Saidi & Zeki, 2019).

Domdouzis et al. (2016) focused on crime prevention during a crisis like natural disasters or pandemics where society faces the loss of lives and rise in crime rate. The research proposed an automated social media and crowdsourcing data mining system to support law enforcement agencies and police to prevent criminal activities.

Researchers emphasized the ATHENA crisis management system as it facilitates the application of various data mining techniques to gather and examine crisis-related information from social media. The findings confirm that social media and crowdsourcing data mining techniques significantly provide desired results by supporting crime prediction during a crisis. Thus, in the post-covid-19 pandemic, the ATHENA system supports Dubai Police to predict crime based on social media and crowdsourcing data.

Hassani et al. (2016) reviewed more than 100 studies available on the application of data mining in crime. The study provides widely used data mining techniques in the literature. The study emphasizes entity extraction, clustering, association rule mining, decision trees, support vector machines, naïve Bayes rule, neural networks, and social network analysis. Thus, applying these data mining techniques in crime detection facilitates predicting crime.

Yu et. al. (2011) developed architecting datasets based on criminal records as categorized based on the US Police department, such as a number of crimes and crime-related events. Initially, consider the location and time of events, spatial and temporal features. According to that applied the data mining techniques to predict crime. The study applied various classifiers such as One Nearest Neighbour (INN), Decision Tree, Support Vector Machine, Neural Network, Naive Bayes, location constrained variation,

It is interesting to note that instead of focusing on causes of crime such as political animosity, criminal background, a study by Sathyadevan et al. (2014) considered the crime factors for each day. The findings reveal that classification based on the Bayes theorem provides more than 90% accuracy in predicting crimes. At the same time, stresses the need to identify more crime attributes to enhance the accuracy of crime prediction.

Mohamad Noor et al. (2014) reviewed literature published between 2000 and 2015 related to Supporting Decision Making (SDM) in crime prediction. Researchers identified forty-four journals based on the classification of index crimes and data mining techniques for six classes. The crime index includes violent crimes and property crimes.

The data mining techniques are relevant to prediction, classification, visualization, clustering, and outlier detection. The findings indicate that prediction and clustering techniques were more supporting in predicting both violent crimes and property crimes. Also, in the context of SDM in crime prevention, the neural network and nearest neighbor data mining techniques were successfully applied in diverse studies.

Gupta et. al. (2008) proposed a crime detection model relevant to the Indian context. The tool comprised of three modules – Data Extraction Module, Crime Analysis Module, and Data Information and Comparison Model. Data mining technique such as clustering was applied to predict crime.

Ying (2016) focused on predicting a crime potential risk with mild and normal categories. The study applied data mining methods such as association analysis, analysis and prediction, cluster analysis, outlier mining. The findings support data mining application for crime detection is successful.

Yet, the author points out to consider the risks associated with data mining such as dealing with large-scale databases and multi-dimensional data, structured data issues, and personal privacy issues.

Xu et al. (2020) considered crime behavior observation in real-time by combining cloud computing image processing and data mining. Therefore, the researchers collected data related to a wide range of criminal behavior characteristics and integrated the rapid response capability of cloud computing to enact data processing.

Furthermore, to increase the accuracy of identifying criminal behavior by indexing criminal behaviors based on population samples. Then, included data processing technology to identify individual crimes per population segmentation.

The findings reveal that the algorithm identifies abnormal behavior with a high accuracy rate.

Bachner (2013) emphasizes that law enforcement departments can successfully deploy predictive methodologies to predict crime by including three categories of analysis techniques. Three segments of analysis are the analysis of time and space, analysis of social networks, and analysis of space.

Furthermore, it is necessary to ensure that crime prevention considers the three pillars of crime prediction such as primary prevention strategies, criminal justice strategies, and law enforcement strategies.

Prado et al. (2020) focused on identifying patterns based on criminal incident data. Researchers considered the identification, characterization, and meta-analyze methods and intelligent algorithms in identifying crime patterns.

The findings show that unsupervised machine learning, supervised machine learning and association rules are effective. Concerning the algorithms aspect, the study suggests K-Means, K-Nearest Neighbors, and Apriori techniques provide the desired outcome.

In the Fortaleza-Brazil context, Damasceno et al. (2012) utilized crime and socioeconomic data for crime prediction. The study confirms that data mining techniques based on crime and socioeconomic data facilitate the prediction of crimes.

Crime analysis domain

data mining tools and techniques are most probably applied on the data of repository which are named Data safehouse which integrates the data extracted from several sources on a many various integrals, that link of the data may come in different formats such as. Databases, spreadsheets, systems downloaded files and other different formats.

Normally this data is cleaned up and has been converted to the required formats and then moved to the data safehouse, so the sequenced used is collecting the data through the police stations and the re collection in house (between the police stations) then goes to the analytics department in order to process the data and then to model it finally prepare the report to present it to the diction makers. The below figure shows the steps.

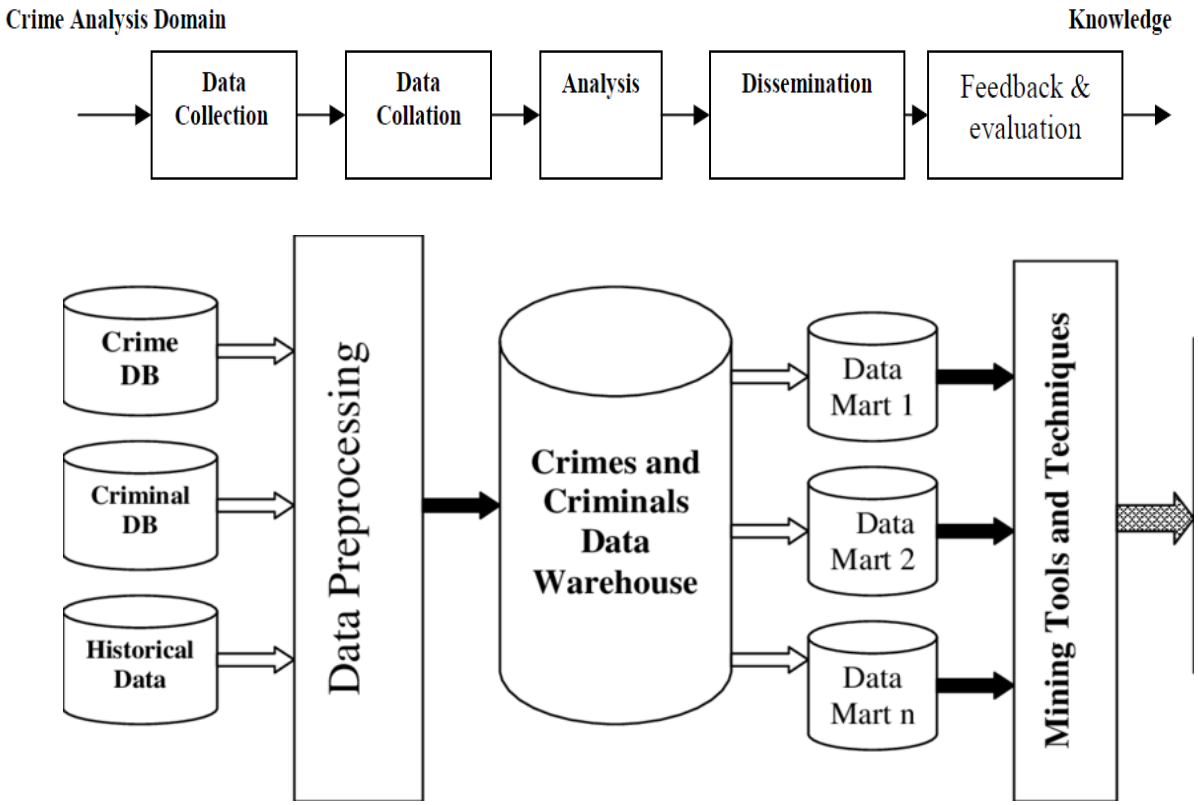


Figure 4 Crime Analysis Steps

(Source: *researchgate.net*. (n.d.). Retrieved April 23, 2022,)

Chapter 3- Project Description

3.1 Project description

In this particular case, SAS will be used as the main tool for data mining. In addition, link analysis will be performed as the main data mining technique to predict crime from the historical data that is provided by CID. In order to perform this prediction, some processes will be performed in SAS.

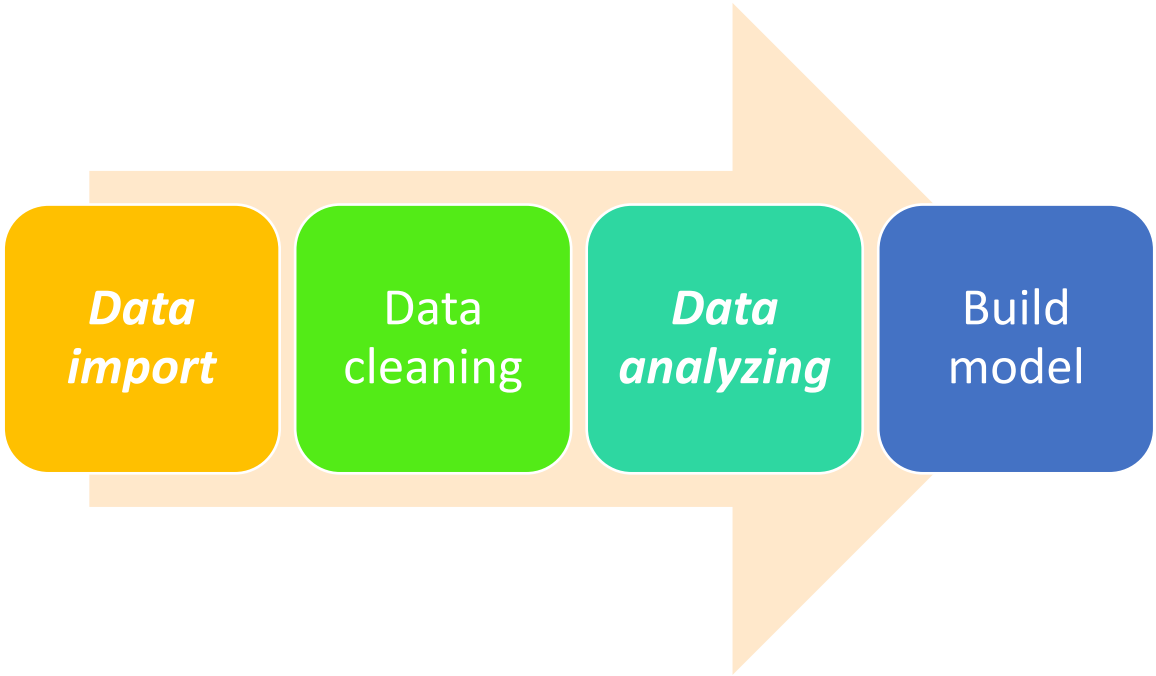


Figure 5: Steps to perform this task
(Source: Self-created)

- Data import

The collected data needs to be imported first in SAS. This will allow working with the objectives of data mining and prediction. Data will be imported from the file that will be collected from the CID in Dubai police. After importing the data into the SAS, then the rest of the work will be processed.

- Data cleaning

Data cleaning also will be performed if there are any missing values in the collected data. The advantage that SAS offers is that it can clean the data on its own so, no coding or manual work is necessary to do so which will cut a huge amount of time since cleaning the data takes a lot of time. Finally, it can be said that this will be the process to initiate this project for predicting crimes.

- Data analyzing

In this step, we will use the forecast tool in SAS, in order to give us a glimpse of how the data will be analyzed. Then, we will start analyzing the data via SAS , however we can also manually use excel in order to support the analysis of SAS . This step is crucial to know the most committed crime, the frequent time

that the crimes are committed at , the nationalities that mostly commit certain crimes, the gender and the demographics of the criminals .

- Build model

Finally, we will start with building the predictive model which is our main purpose and that will give us the accuracy of each prediction and the prediction of certain crimes.

Chapter 4- Data Analysis

This chapter is supposed to show the results of the data analysis done in SAS program. Initially, we have done a descriptive analysis on the criminals' profile and their crimes' profile. Lastly, we have used the forecasting method in SAS to build a prediction model for the crimes' rates. Data sources are all about the platform from which data has been collected. As mentioned here, the data related to several crimes will be collected from the CID in Dubai police. In addition, some additional information and similar experiences that have been done across the world will be collected to gain some understanding of this work. However, the main data related to this work is collected from the Dubai police.

4.0 Translating the dataset

When we first got access to the dataset, it was in Arabic exported from a system in CID called the criminal unified system. It is a system where all crimes get recorded there with every detail such as location of the crime , date of the crime , tools used in the crime , nationality of the criminal, gender of the criminal , education level of the criminal ..etc. There was also a space where the employee could write a brief summary of what happened in the crime according to the reporter's words. The figure below is the dataset before we cleaned it and translated it :

	نوع التهمة	تاريخ البلاغ	وقت البلاغ	مركز الشرطة	الدور	العمر	الجنس	الجنسية	المستوى التعليمي	الحالة الاجتماعية	الأدوات المستخدمة	الأساليب المستخدمة
1												
2	الشروع فى السرقة	12/31/2017	22:27:00	مركز شرطه المرقبات	متهم	24	أنثى	ايران	-	أعزب	بدون أداة	-
3	جرة خدمات وسيلة نقل مع	12/31/2017	03:14:00	مركز شرطه بردبي	متهم	33	ذكر	الهند	شهادة الثانويه العامه	متزوج	بدون أداة	-
4	المغفى إلى مرض أو عجز:	12/31/2017	18:39:00	مركز شرطه بردبي	متهم	24	ذكر	الاردن	الصف الاول الثانوى	أعزب	قوة جسدية	-
5	المغفى إلى مرض أو عجز:	12/31/2017	18:39:00	مركز شرطه بردبي	متهم	24	ذكر	الاردن	يقرا ويكتب	أعزب	قوة جسدية	-
6	المغفى إلى مرض أو عجز:	12/31/2017	18:39:00	مركز شرطه بردبي	متهم	29	ذكر	الاردن	شهادة الثانويه العامه	أعزب	قوة جسدية	-
7	المغفى إلى مرض أو عجز:	12/31/2017	03:06:00	مركز شرطه بردبي	متهم	29	ذكر	الامارات	غير محدد	أعزب	قوة جسدية	-
8	لماء شيك بدون رصيد بسوء	12/31/2017	19:24:00	مركز شرطه بردبي	متهم	28	أنثى	نيبال	-	أعزب	ادعاء الحاجة والفقر	-
9	لماء شيك بدون رصيد بسوء	12/31/2017	23:03:00	مركز شرطه نايف	متهم	33	ذكر	بنغلادش	شهادة الثانويه العامه	أعزب	الإحتيال العقاري	-
10	الشروع فى السرقة	12/31/2017	22:27:00	مركز شرطه المرقبات	متهم	24	أنثى	ايران	-	أعزب	الحيلة	-
11	الشروع فى السرقة	12/31/2017	22:27:00	مركز شرطه المرقبات	متهم	24	أنثى	ايران	-	أعزب	السلب	-
12	المغفى إلى مرض أو عجز:	12/31/2017	18:39:00	مركز شرطه بردبي	متهم	24	ذكر	الاردن	الصف الاول الثانوى	أعزب	الضرب/ الركل	-
13	المغفى إلى مرض أو عجز:	12/31/2017	18:39:00	مركز شرطه بردبي	متهم	24	ذكر	الاردن	يقرا ويكتب	أعزب	الضرب/ الركل	-
14	المغفى إلى مرض أو عجز:	12/31/2017	18:39:00	مركز شرطه بردبي	متهم	29	ذكر	الاردن	شهادة الثانويه العامه	أعزب	الضرب/ الركل	-
15	المغفى إلى مرض أو عجز:	12/31/2017	03:06:00	مركز شرطه بردبي	متهم	29	ذكر	الامارات	غير محدد	أعزب	الضرب/ الركل	-
16	بلاء على مال الغير بدون وج	12/31/2017	21:47:00	مركز شرطه القصيص	متهم	28	ذكر	مصر	0	أعزب	ايهام المجنى عليه بتقديم خدمة	-
17	جرة خدمات وسيلة نقل مع	12/31/2017	03:14:00	مركز شرطه بردبي	متهم	33	ذكر	الهند	شهادة الثانويه العامه	متزوج	بدون أسلوب	-
18	لماء شيك بدون رصيد بسوء	12/31/2017	19:24:00	مركز شرطه بردبي	متهم	28	أنثى	نيبال	-	أعزب	-	-
19	لماء شيك بدون رصيد بسوء	12/31/2017	23:03:00	مركز شرطه نايف	متهم	33	ذكر	بنغلادش	شهادة الثانويه العامه	أعزب	-	-
20	بلاء على مال الغير بدون وج	12/31/2017	21:47:00	مركز شرطه القصيص	متهم	28	ذكر	مصر	0	أعزب	-	-
21	المغفى إلى مرض أو عجز:	12/31/2017	18:39:00	مركز شرطه بردبي	متهم	24	ذكر	الاردن	الصف الاول الثانوى	أعزب	-	-
22	المغفى إلى مرض أو عجز:	12/31/2017	18:39:00	مركز شرطه بردبي	متهم	24	ذكر	الاردن	يقرا ويكتب	أعزب	-	-
23	المغفى إلى مرض أو عجز:	12/31/2017	18:39:00	مركز شرطه بردبي	متهم	29	ذكر	الاردن	شهادة الثانويه العامه	أعزب	-	-
24	المغفى إلى مرض أو عجز:	12/31/2017	03:06:00	مركز شرطه بردبي	متهم	29	ذكر	الامارات	غير محدد	أعزب	-	-
25	مروعه دون التجديد أومغاد	12/31/2017	12:16:00	مركز شرطه نايف	متهم	28	ذكر	باكستان	0	أعزب	-	-
26	مروعه دون التجديد أومغاد	12/31/2017	13:17:00	مركز شرطه الراشديه	متهم	28	ذكر	باكستان	0	أعزب	-	-
27	مروعه دون التجديد أومغاد	12/31/2017	16:18:00	مركز شرطه الرفاعة	متهم	32	ذكر	بنغلادش	ثانويه علمى	أعزب	-	-
28	مروعه دون التجديد أومغاد	12/31/2017	10:58:00	مركز شرطه الرفاعة	متهم	33	أنثى	بنغلادش	0	متزوج	-	-
29	مروعه دون التجديد أومغاد	12/31/2017	09:34:00	مركز شرطة البرشاء	متهم	35	ذكر	باكستان	اعداديه عامه	أعزب	-	-
30	مروعه دون التجديد أومغاد	12/31/2017	08:35:00	مركز شرطه الرفاعة	متهم	35	ذكر	بنغلادش	ثانويه علمى	أعزب	-	-
31	الشروع فى السرقة	12/31/2017	22:27:00	مركز شرطه المرقبات	متهم	24	أنثى	ايران	-	أعزب	-	-
32	تاجرة بالمشروبات الكحول	12/31/2017	00:08:00	مركز شرطه الرفاعة	متهم	25	ذكر	نيبال	ابتدائيه عامه	أعزب	-	-
33	تعاطى المشروبات الكحولية	12/31/2017	23:53:00	مركز شرطه جبل على	متهم	27	ذكر	الهند	-	أعزب	-	-
34	تعاطى المشروبات الكحولية	12/31/2017	23:53:00	مركز شرطه جبل على	متهم	32	ذكر	الهند	0	أعزب	-	-
35	تعاطى المشروبات الكحولية	12/31/2017	00:24:00	مركز شرطه جبل على	متهم	33	ذكر	نيبال	-	أعزب	-	-
36	عد أتلاف أو أخفاء جواز الس	12/31/2017	01:23:00	مركز شرطه نايف	متهم	34	ذكر	كينيا	ابى	متزوج	-	-
37	او تزوير العملة او السندات	12/31/2017	22:51:00	مركز شرطه نايف	متهم	31	ذكر	الهند	شهادة الثانويه العامه	أعزب	-	-
38	خيانة الأمانة بالاختلاس	12/31/2017	23:41:00	مركز شرطه بردبي	متهم	32	ذكر	باكستان	-	أعزب	-	-

We have met in order to put a plan on how we can clean and translate the dataset ,since the data is huge and it has over 3,000 fields ; we did extensive research , and we have found out that there is a formula that we can use a formula in google forms which is : ***=GOOGLETRANSLATE(B2,"ar","en")*** it was the easiest and fastest way to translate such large dataset. Thus, we have agreed on using google forms to translate the dataset.

The data we got was not clean and it had a lot of data inconsistencies. We had to disregard these points , because they did not make any sense to analyse blank fields. The data had also some extra information such as a brief of each crime and the name of the reporter, after many meetings with the criminal investigation department; they have informed us that they were not interested in such information for the analysis thus, they have been disregarded as well. As shown in the figure below , the missing information :

تقرير إحصائي بالتهم (معلوم / مجهول)

مؤشرات عامة عن البلاغ

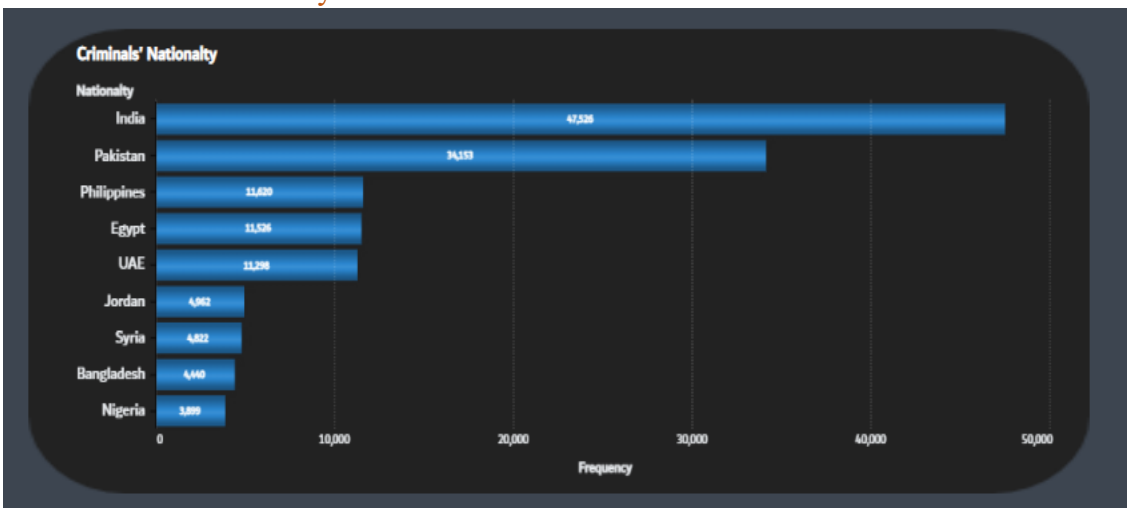
نوع البلاغ : جنائي	**طريقة التبليغ :**		**وضع البلاغ :**	**فئة الشخص :**
حدث\بالغ :	**وضعية الحادث :**		**دور الشخص :**	**وضع الشخص :**
حصر النتائج في	**مـــــن**			**إلـى**
تاريخ البلاغ :	01\01\2018			31\12\2018
جهة تلقي البلاغ :	112000000	وزاره الداخليه / القياده العامه لشرطه دبى		
التصنيف الرئيسي للتهمة :				
التصنيف الفرعي للتهمة :				
التهمة :	(6001,6002,600	بإستثناء التهم : تعاطي المواد المخدرة او ا لمؤثرات العقلية و تعاطي المواد المخد		
الجنسية :				
مكان الحادث				
الإمارة :			**المدينة\البلد :**	
المنطقة :			**الشارع :**	

After manually cleaning each dataset and deleting all the missing values and rows which took a lot of time, we have finally reached a clean and ready dataset.

The figure below is the dataset cleaned and translated:

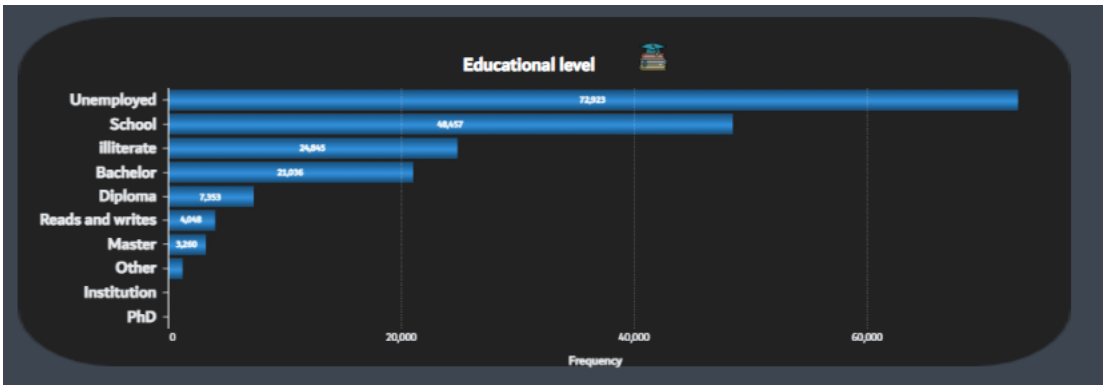
Tools used	Social Status	Educational level	Nationalty	Gander	Age	Police Station	eporting time	eporting Date	Type of charge
Without Tools	Unmarried	Undefined	Iran	Female	24	ALMuraqabat	22:27:00	12/31/2017	Stealing
Without Tools	Married	High School certificate	India	Male	33	Bur Dubai	03:14:00	12/31/2017	Non-payment of services
Physical Force	Unmarried	First grade secondary	Jordan	Male	24	Bur Dubai	18:39:00	12/31/2017	Abuse on body safety
Physical Force	Unmarried	Reads and writes	Jordan	Male	24	Bur Dubai	18:39:00	12/31/2017	Abuse on body safety
Physical Force	Unmarried	High School certificate	Jordan	Male	29	Bur Dubai	18:39:00	12/31/2017	Abuse on body safety
Physical Force	Unmarried	Undefined	UAE	Male	29	Bur Dubai	03:06:00	12/31/2017	Abuse on body safety
Without Tools	Unmarried	Undefined	Nepal	Female	28	Bur Dubai	19:24:00	12/31/2017	Producing financial issues
Without Tools	Unmarried	High School certificate	Bangladesh	Male	33	Naif	23:03:00	12/31/2017	Producing financial issues
Without Tools	Unmarried	Undefined	Iran	Female	24	ALMuraqabat	22:27:00	12/31/2017	Stealing
Without Tools	Unmarried	Undefined	Iran	Female	24	ALMuraqabat	22:27:00	12/31/2017	Stealing
Without Tools	Unmarried	First grade secondary	Jordan	Male	24	Bur Dubai	18:39:00	12/31/2017	Abuse on body safety
Without Tools	Unmarried	Reads and writes	Jordan	Male	24	Bur Dubai	18:39:00	12/31/2017	Abuse on body safety
Without Tools	Unmarried	High School certificate	Jordan	Male	29	Bur Dubai	18:39:00	12/31/2017	Abuse on body safety
Without Tools	Unmarried	Undefined	UAE	Male	29	Bur Dubai	03:06:00	12/31/2017	Abuse on body safety
Without Tools	Unmarried	Undefined	Egypt	Male	28	ALQusais	21:47:00	12/31/2017	Grab others
Without Tools	Married	High School certificate	India	Male	33	Bur Dubai	03:14:00	12/31/2017	Non-payment of services
Without Tools	Unmarried	Undefined	Nepal	Female	28	Bur Dubai	19:24:00	12/31/2017	Producing financial issues
Without Tools	Unmarried	High School certificate	Bangladesh	Male	33	Naif	23:03:00	12/31/2017	Producing financial issues
Without Tools	Unmarried	Undefined	Egypt	Male	28	ALQusais	21:47:00	12/31/2017	Grab others
Without Tools	Unmarried	First grade secondary	Jordan	Male	24	Bur Dubai	18:39:00	12/31/2017	Abuse on body safety
Without Tools	Unmarried	Reads and writes	Jordan	Male	24	Bur Dubai	18:39:00	12/31/2017	Abuse on body safety
Without Tools	Unmarried	High School certificate	Jordan	Male	29	Bur Dubai	18:39:00	12/31/2017	Abuse on body safety
Without Tools	Unmarried	Undefined	UAE	Male	29	Bur Dubai	03:06:00	12/31/2017	Abuse on body safety
Without Tools	Unmarried	Undefined	Pakistan	Male	28	Naif	12:16:00	12/31/2017	Legal residence
Without Tools	Unmarried	Undefined	Pakistan	Male	28	ALRashideya	13:17:00	12/31/2017	Legal residence
Without Tools	Unmarried	Secondary secondary	Bangladesh	Male	32	ALRafaah	16:18:00	12/31/2017	Legal residence
Without Tools	Married	Undefined	Bangladesh	Female	33	ALRafaah	10:58:00	12/31/2017	Legal residence
Without Tools	Unmarried	Preparation of a year	Pakistan	Male	35	ALBursha	09:34:00	12/31/2017	Legal residence
Without Tools	Unmarried	Secondary secondary	Bangladesh	Male	35	ALRafaah	08:35:00	12/31/2017	Legal residence
Without Tools	Unmarried	Undefined	Iran	Female	24	ALMuraqabat	22:27:00	12/31/2017	Stealing
Without Tools	Unmarried	Primary General	Nepal	Male	25	ALRafaah	00:08:00	12/31/2017	Issues of alcoholic beverages
Without Tools	Unmarried	Undefined	India	Male	27	Jebel Ali	23:53:00	12/31/2017	Issues of alcoholic beverages
Without Tools	Unmarried	Undefined	India	Male	32	Jebel Ali	23:53:00	12/31/2017	Issues of alcoholic beverages
Without Tools	Unmarried	Undefined	Nepal	Male	33	Jebel Ali	00:24:00	12/31/2017	Issues of alcoholic beverages
Without Tools	Married	My mom	Kenya	Male	34	Naif	01:23:00	12/31/2017	Destruction of public property
Without Tools	Unmarried	High School certificate	India	Male	31	Naif	22:51:00	12/31/2017	Forging financial documents

4.1 Criminals' nationality



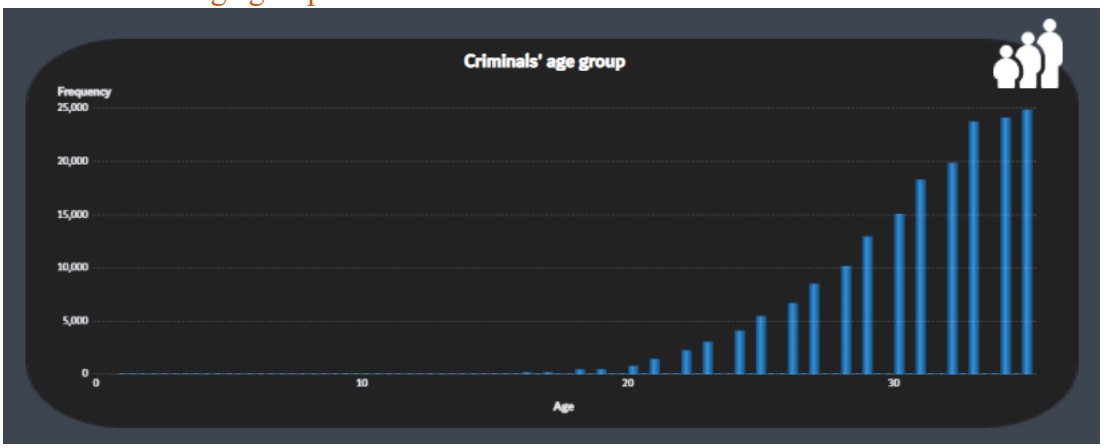
First, we have analyzed the nationality of the criminals, this is important because Dubai is one of the top-rated tourist attractions and the nationalities need to be observed in order to observe the cultural values of the particular nationality. As shown in the figure above, India has the major rate in the criminals' nationality through all the years 2017 – 2021; which is predictable due to their high population. The second highest nationality is Pakistan scoring over 35,000 cases and crimes have been done by this nationality throughout all years from 2017 – 2021.

4.2 Criminals' education level



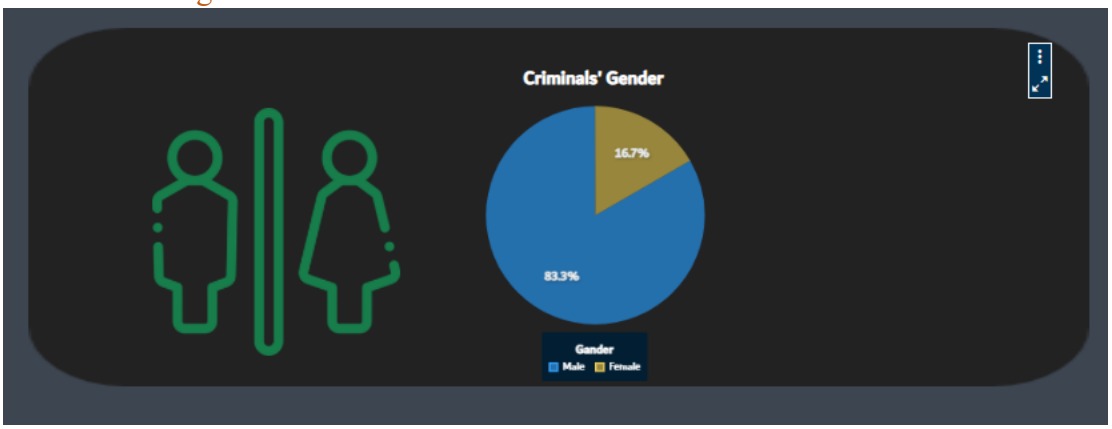
Next, we have analyzed the education level of the criminals. As shown in the figure, the majority of the criminals were unemployed and this shows that unemployment aids in crimes such as theft or financial fraud due to lack of money. The next educational level of the criminals would be school level, which means that they haven't finished or graduated from school.

4.3 Criminals' age group



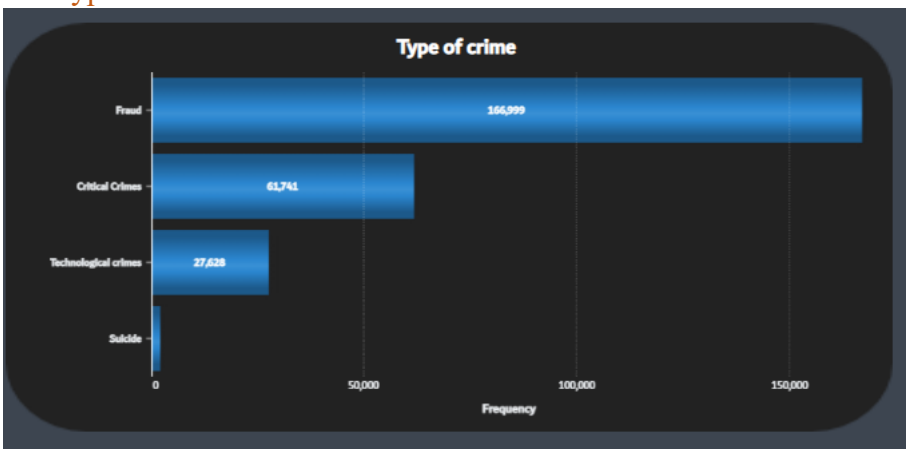
We have also analyzed the criminals' age group , to observe whether the criminals are teenagers, young , young adults or adults. The analyses shows a histogram which is clearly left skewed which means that the age groups of the criminals range from 30 – 45 which are considered adults and they are the ones that have the highest rates of committing crimes.

4.4 Criminals' gender



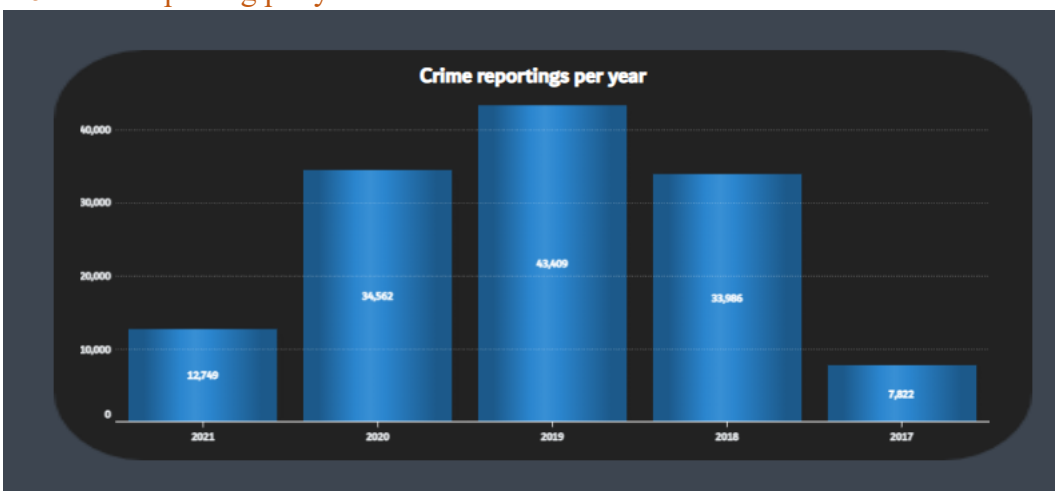
After analyzing the age groups, we wanted to observe whether most criminals are male or female and as shown above, the majority of the criminals were males scoring a 83% whereas female criminals scored 16% of the study.

4.5 Type of crime



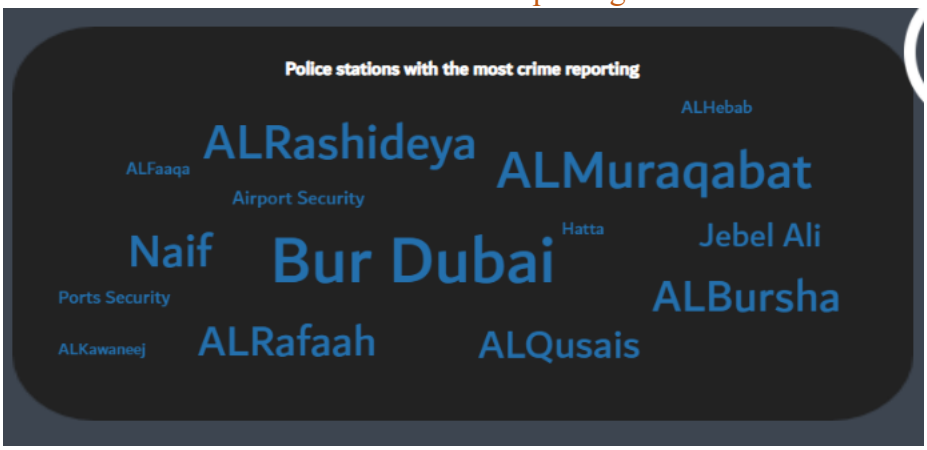
After analyzing the criminals' profiles, we have moved on to analyzing the crimes and the type of crimes that have been committed in the last five years. Fraud has scored the most crime committed throughout these five years followed by critical crimes. We have grouped some crimes such as of murder, kidnapping, rape..etc. and classified them as critical crimes as this was instructed to us by the criminal investigation department at Dubai Police. A positive insight is that suicide has decreased in the last five years as shown below which is a great sign of how aware people in Dubai have been regarding mental illnesses and depression that relate to suicide.

4.6 Crime reporting per year



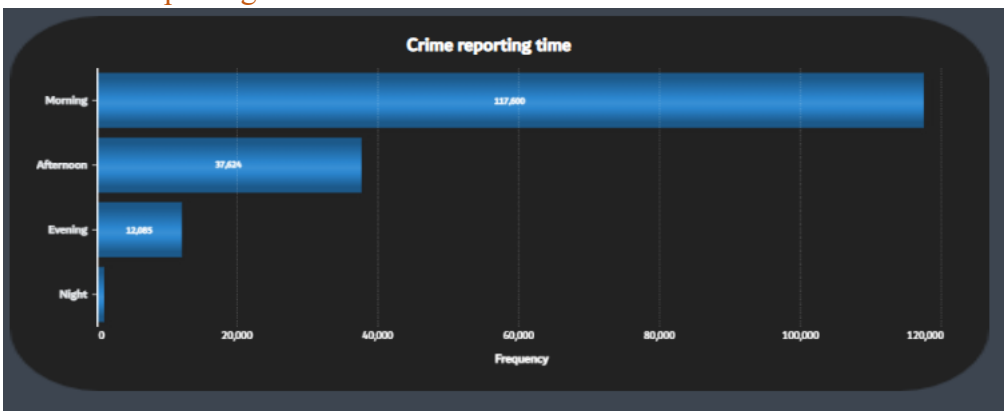
The crime reporting per year has reached its peak in 2019 right before the pandemic , however even though there has been a global pandemic in 2020 , it still scored as the second highest year in regards to crime reporting and the major crime related to 2020 was fraud since the quarantine has been going on in that year , fraud was the main crime for criminals to gain money from since most people has been laid off their work and have been needing money at that period of time. 2017 has scored the least year for crime reporting.

4.7 Police stations with the most crime reporting



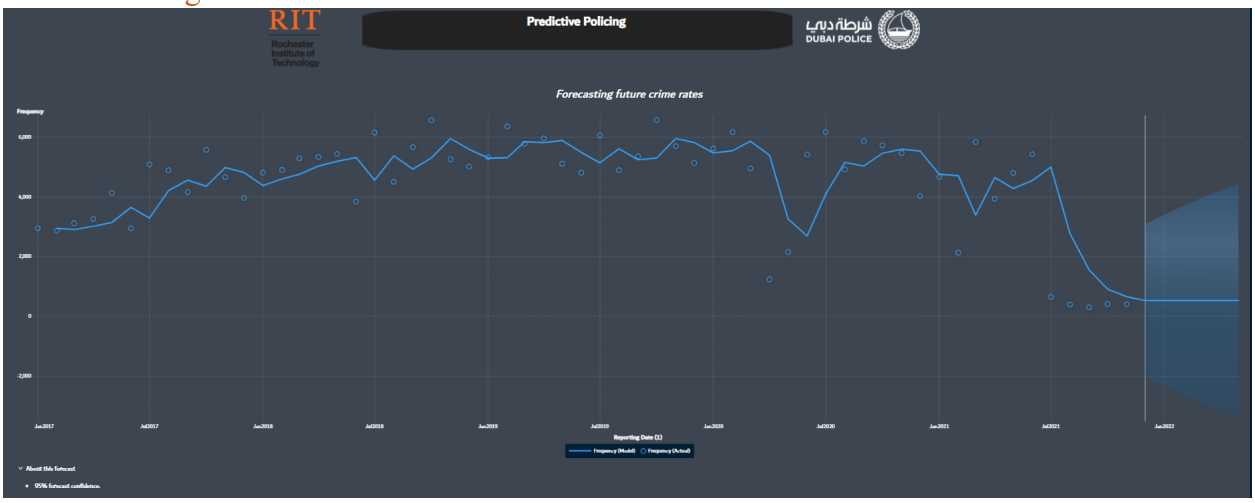
As shown in the word cloud above, we wanted to observe what police stations have had the most crime reporting, this is very useful in order to have an insight of which areas in Dubai do crimes take place on. The most police station with crime reporting was Bur Dubai which is not a surprise since Bur Dubai is a very attractive tourist location and it is always very crowded and has most of the attention in Dubai, next will be Al Muraqabat scoring a 38,000 reporting in the last five years.

4.8 Crime reporting time



We wanted to know more of when the crimes happen in order to have an accurate forecasting later on ; and as shown in the bar chart above, morning time has scored the highest time the crimes are committed which is surprising since it is known that night time is the most time where crime happens since its dark and quiet. That was very surprising and astonishing at the same time.

4.9 Forecasting future crime rates



Finally , we started with building our forecasting model. According to the figure viewed in the dashboard of the crime rate prediction the crime rates will decrease dramatically in the upcoming years. As per the observation, in July 2021, the crime rate shave drastically decreased which will be based on other factors. We have met with the CID to ask them what they think is the reason to this sudden decrease (was there a seasonal trend that month?) We have found that due to the coronavirus surge in the summer, we have also observed the variables within the dataset that are dependent on this decrease of crime rates and we have seen that people have mostly made financial electronic crimes and all the criminals were illiterate in their education level in July 2021. Although the forecast also predicts that crimes might go as high as 4,000 but the model insists that if no outer factors happen during these years, the crime rates would decrease to approximately 500 crimes in a day.

Chapter 5 Conclusion

5.1 Conclusion

In conclusion, the proposed data mining method facilitates the identification of the tendencies of historical crime data. Also, it permits to forecasting of critical crimes in a particular area. The data mining approach provides better insight into the collected crime dataset to find relationships between every hidden point. Additionally, the SAS program permits the exploration of individual activities related to any crime scenarios based on the gathered data. Consequently, the data mining supports the prediction of crimes ahead and aligns with the goal of the Dubai Police, which is to make Dubai a safer place. As the results have shown, the crime rate in Dubai has decreased dramatically in the recent years which shows the extensive work His Excellency Abdulla Khalifa Al Marri commander – in – chief of Dubai police is doing to ensure the safety of the city and the well-being of the citizens. This project has made us use our data mining knowledge that we have done throughout our masters' journey in a way where we can now understand how data analytics is implemented in organizations. Throughout the whole process starting with the project problem till the evaluation and results, the journey has been amazing and very insightful. Data analytics can save lives and can help improve the efficiency of organizations. Data is the oil that keeps organizations working.

5.2 Recommendations

The selection of appropriate data mining is crucial for the success of data mining in crime prediction. Therefore, it is essential to identify the available data types, system issues, and data sources. Also, clarity on the data mining functions and integration with the existing database is crucial to gaining access to the data. For crime prediction in real-time, it is necessary to utilize emerging technologies such as cloud and image capturing to reap more benefits. Therefore, the selection of data mining systems and the integration of emerging technologies deliver better crime prediction.

The understanding of the problem is the main pillar of the success of the project, it is the factor that can actually measure whether the project will succeed or fail. Understanding the problem is the main step.

Analytical maturity in organizations is a must in this period of time. Being a data – driven organization will help aid in the success of the organization and will help you get on track with everything going on in life. Therefore , having an analytical mentality is crucial in an organization.

References

al Saidi, W., & Zeki, A. (2019). The Use of Data Mining Techniques in Crime Prevention and Prediction. *2nd Smart Cities Symposium (SCS 2019)*. https://doi.org/10.1049/cp.2019.0225

Al-Shamiri, A.Y.R., 2021. Artificial intelligence and pattern recognition using data mining algorithms. *International Journal of Computer Science & Network Security*, *21*(7), pp.221-232.

Archibald, M.M., Ambagtsheer, R.C., Casey, M.G. and Lawless, M., 2019. Using zoom videoconferencing for qualitative data collection: perceptions and experiences of researchers and participants. *International Journal of Qualitative Methods*, *18*, p.1609406919874596.

Bachner, J. (2013). Predictive policing: preventing crime with data and analytics. Washington, DC: IBM Center for the Business of Government.

Bommarito, E. and Bommarito, M., 2019. An empirical analysis of the python package index (pypi). *arXiv preprint arXiv:1907.11073*.

Damasceno, M., Teixeira, J., & Campos, G. (2012). A prediction model for criminal levels using socio-criminal data. *International Journal of Electronic Security and Digital Forensics*, *4*(2/3), 201. https://doi.org/10.1504/ijesdf.2012.048414

David, H. and Suruliandi, A., 2017. SURVEY ON CRIME ANALYSIS AND PREDICTION USING DATA MINING TECHNIQUES. *ICTACT journal on soft computing*, *7*(3).

Domdouzis, K., Akhgar, B., Andrews, S., Gibson, H., & Hirsch, L. (2016). A social media and crowdsourcing data mining system for crime prevention during and post-crisis situations. *Journal of Systems and Information Technology*, *18*(4), 364–382. https://doi.org/10.1108/jsit-06-2016-0039

Ghorbani, A. and Farzai, S., 2018. Fraud detection in automobile insurance using a data mining based approach. *International Journal of Mechatronics, Elektrical and Computer Technology (IJMEC)*, *8*(27), pp.3764-3771.

Gupta, M., Chandra, B., & Gupta, M. P. (2008). Crime data mining for Indian police information system. Computer society of India, 40(1), 388-397.

Hassani, H., Huang, X., Silva, E. S., & Ghodsi, M. (2016). A review of data mining applications in crime. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *9*(3), 139–154. https://doi.org/10.1002/sam.11312

Madni, H.A., Anwar, Z. and Shah, M.A., 2017, September. Data mining techniques and applications—a decade review. In *2017 23rd International Conference on Automation and Computing (ICAC)* (pp. 1-7). IEEE.

Mohamad Noor, N. M., Ab Hamid, S. H., Mohemad, R., Abdul Jali, M. M., & Hitam, M. S. (2014). A Review on a Classification Framework for Supporting Decision Making in Crime Prevention. *Journal of Artificial Intelligence*, *8*(1), 17–34. https://doi.org/10.3923/jai.2015.17.34

Norouzi, N., & Ataei,(2021). Application of data mining in identifying and discovering hidden patterns of theft. International Journal of Innovation Research in the Humanities, 1 (1), 29 – 42.

Prado, K. H. J., Santos Souza, L., Dias De Jesus Junior, I., & Colaço Júnior, M. (2020). Applied Intelligent Data Analysis to Government Data Related to Criminal Incident: A Systematic Review. *Journal of Applied Security Research*, *15*(3), 297–331. https://doi.org/10.1080/19361610.2020.1716511

Sarangam, A. (2022, February 15). *Top 14 Data Mining Tools*. Jigsaw Academy. Retrieved March 20, 2022, from https://www.jigsawacademy.com/blogs/data-science/data-mining-tools/

Sathyadevan, S., Devan, M., & Gangadharan, S. S. (2014). Crime analysis and prediction using data mining. *2014 First International Conference on Networks & Soft Computing (ICNSC2014)*. https://doi.org/10.1109/cnsc.2014.6906719

Singh, N., Bellathanda Kaverappa, C., & Joshi, J. D. (2018). Data Mining for Prevention of Crimes. *Human Interface and the Management of Information. Interaction, Visualization, and Analytics*, 705–717. https://doi.org/10.1007/978-3-319-92043-6_55

Tang, Y.P., Li, G.X. and Huang, S.J., 2019. ALiPy: active learning in python. *arXiv preprint arXiv:1901.03802*.

Figure2. architecture for proposed ... - researchgate.net. (n.d.). Retrieved April 23, 2022, from https://www.researchgate.net/figure/Figure2-Architecture-for-proposed-technique_fig1_355334763

Tutorial Point. (2021). *Data Mining - Applications & Trends*. Retrieved March 15, 2022, from https://www.tutorialspoint.com/data_mining/dm_applications_trends.htm

Walia, N. and Kalia, A., 2020. Tools in Data Mining A Comparative Analysis. *vol*, *12*, p.6.

Xu, Z., Cheng, C., & Sugumaran, V. (2020). Big data analytics of crime prevention and control based on image processing upon cloud computing. *Journal of Surveillance, Security and Safety*. https://doi.org/10.20517/jsss.2020.04

Ye, Y., Li, T., Adjeroh, D. and Iyengar, S.S., 2017. A survey on malware detection using data mining techniques. *ACM Computing Surveys (CSUR)*, *50*(3), pp.1-40.

Ying, Z. (2016). Analysis of Crime Factors Correlation Based on Data Mining Technology. *2016 International Conference on Robots & Intelligent System (ICRIS)*. https://doi.org/10.1109/icris.2016.21

Yu, C. H., Ward, M. W., Morabito, M., & Ding, W. (2011, December). Crime forecasting using data mining techniques. In 2011 IEEE 11th international conference on data mining workshops (pp. 779-786). IEEE.

Zhao, X. and Tang, J., 2018. Crime in urban areas: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, *20*(1), pp.1-12.

Zhuang, Q.R., Yao, Y.W. and Liu, O., 2018. Application of data mining in term deposit marketing. In *Proceedings of the International MultiConference of Engineers and Computer Scientists* (Vol. 2).