5-2024

Crime Rate Analysis and E-crime prevention in Dubai using machine learning.

Mohammad Ahmad Hatam
mh7489@rit.edu

Crime Rate Analysis and E-crime prevention in Dubai using machine learning.

by

Mohammad Ahmad Hatam

A Thesis 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
Dubai
May 2024

RIT Dubai

Master of Science in Professional Studies:

Data Analytics

Graduate Thesis Approval

Student Name**: Mohammad Ahmad Hatam**

Graduate Capstone Title**: Crime Rate Analysis and E-crime prevention in Dubai using machine learning.**

Graduate Thesis Committee:

Name: Dr. Sanjay Modak Date:

Chair of committee

Name: Dr. Hammou Messatfa Date:

Member of committee

ACKNOWLEDGMENTS

I hereby cordially thank the many people who helped me so much in completing this research project. First let me express my thanks to Dr. Sanjay Modak (the Department Chair) for all the encouragement, guidance and suggestions he has given me over many years of working together on various projects in recognition capitalism's growing popularity. His guidance and encouragement have had a great influence on the direction this research has taken.

Therefore, I also would like to express my great respectful gratitude to the advisor of our thesis Dr. Hammou Messatfa, for all his excellent counsel and practical suggestions on how exactly we can perform the research work as well as support during this process itself. No doubt this work is also a contribution of his knowledge and time. Besides, I would like honorably to express thanks and sincere appreciation for all faculty members of the department whose cooperative efforts have created a climate conducive to academic thinking.

Thank you.

ABSTRACT:

Globally increasing criminal activity, such as E Crimes, presents a serious threat to both economic growth and social welfare. Nevertheless, Dubai is a safety haven in the middle of this trend, as seen by the significant decline in major and non-alarming crimes during the first quarter of 2023. This study uses an examination of the large crime dataset (about 34,567 events) from the Dubai Police Department that spans 2019 to 2021. Several main goals are to be achieved by this work by using geospatial methods, predictive modelling, correlation investigations, and exploratory analysis. First of all, it looks for complex trends in the socioeconomic, geographical, and temporal aspects of Dubai's criminal scene. Later, working with legislators and law enforcement organisations, strategic interventions will be developed to reduce e-crime rates via use of predictive intelligence. The increase in e-crime, driven by the widespread use of smartphones and the internet, which has given rise to cyber risks like hacking, identity theft, and online fraud, is of special concern to the Dubai Police. Understanding that contemporary crime is changing, initiatives are being made to raise cybersecurity knowledge and monitor regional threat trends. The main objectives of this work are to develop machine learning classifier models to accurately predict e-crime behaviour, analyse demographic, temporal, and economic trends in e-crime statistics, and offer practical suggestions for resource allocation and crime reduction techniques. It is projected that e-crime rates in Dubai will drop by 20% by 2025 by leveraging the potential of data-driven regulations, therefore creating a safer and more secure environment for its residents. The study approach uses the CRISP-DM analytics paradigm and includes stages including business understanding, data understanding, data preparation, modelling, assessment, and implementation. Even with their inherent drawbacks—depending on official datasets, unpredictable social and economic forces, and the haziness of projections—modern analytics have enormous potential to support Dubai's safety efforts. In the study, the Logistic Regression model, utilizing 46 predictor fields, achieved an impressive accuracy of 85.871%, with an AUC of 0.932 and precision, recall, and F1-measure scores of 0.855. More detailed statistics must be included, model alarms must be integrated with surveillance systems, and models must be updated often to identify new patterns. A safer Dubai is possible by a coordinated effort to get beyond these limitations and significantly improve the effectiveness and efficiency of crime prevention methods.

Keywords: Crime, data analytics, predictive analytics, E-crime rates

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION AND PROBLEM STATEMENT

1.1 Background:

The background information provided discusses the global increase in criminal activity and its negative impact on economic development and general well-being. It then highlights Dubai's remarkable progress in reducing crime rates, with a 25% decline in serious crimes and a 7.1% decline in non-alarming crimes in the first quarter of 2023, according to statistics from the Dubai Police. Despite these promising indicators, the background emphasizes the need for persistent efforts to make Dubai one of the safest cities in the world, and the potential usefulness of advanced analytical approaches for making strategic decisions to further increase safety.

1.2 Problem Statement:

The problem statement outlines the availability of the Dubai Police Department's crime dataset, covering the years 2019–2021, with more than 34,567 occurrences. The main objectives of this research are to achieve a thorough understanding of Dubai's crime trends across time, geography, and socioeconomic status through geospatial methods, predictive modelling, correlation research, and exploratory analysis. Additionally, the research aims to devise a strategy to decrease crime rates based on predictive intelligence in collaboration with policymakers and law enforcement. The types of crimes covered in the dataset include e-crime, physical violence, robbery, and traffic violations, with a particular emphasis on e-crime as a growing concern for the Dubai Police.

1.3 Project Goals

The goals driving this project are:

- Analysing demographic, temporal, economic, and hotspot patterns in Dubai's E-crime statistics might inform law enforcement strategies.
- Develop ML classifier models, such as Random Forests and Support Vector Machines, th at can accurately predict E-crime behaviour 82% of the time.

.

1.4 Aims and Objectives:

Research Aim and Objectives:

The scope of our project involves the research of the subject matter in detail by using online available materials like research papers and previous analysis and deep-dives, which aims to provide us through understanding of the subject at hand. Researchers plan to determine the following throughout the course of this project:

- Determine factors that affect the E-crime rates of a neighbourhood.
- Understand the relationship between these factors and their impact on the predictor variable.
- Determine severity of crimes based on predictor variables in the dataset. Improve the living conditions and economic betterment of Dubai by reducing crimes.
- Developing a machine learning model capable of effectively detecting illicit goods involved in E-crime using diverse factors.

1.5 The Study's Limitations

Despite the major limitations, modern analytics have the potential to enhance the Dubai's safety initiatives:

- This search can only find results from official datasets collected by the Dubai Police from 2019 to 2021.
- Economic and social factors, such the unemployment rate and the level of education, are beyond our control.
- There is an inherent degree of uncertainty in forecasts, and they cannot predict when trends will change.
- Anonymized data may make criminal connection analysis, a tool that might assist investigations, difficult.
- Due to a lack of contact with surveillance systems, it is inadequate for real-time crime alerts.

Improving efficiency can be achieved by resolving the certain constraints:

- Including more detailed datasets, covering the health, income, transit, etc.
- Integrating the model alerts with the camera's networks in the high-risk areas.
- Adding fresh data to models so they can detect the emerging trends.
- E-crime and the other more generic types of the crime can be better predicted by the automation.

1.6 Structure of Thesis:

Chapter1: Introduction and Problem Statement: In the opening, the thesis statement and problem statement will be made, outlining the background information on the increase in criminal activities globally and its effects on the economic development and general life of a society. The project aims and objectives are stipulated, with the research design and the study's constraints also highlighted.

Chapter2: Literature Review The section of the literature review brings up the topic of e-crime in Dubai then sheds light on high rate of the drivers of crime and employing analytics for crime prevention. It is worth noting that the key messages from the study are summarized.

Chapter3: Research Methodology The research methodology chapter is an introduction to the research approach: it explains the use of CRISP-DM methodology and machine learning algorithms. Alongside other procedures, data splitting and evaluation protocols are explained.

Chapter4: Data Analysis and Modelling The following chapter is a data analysis and modelling chapter that gives the details about the data set used, its data dictionary and includes some exploratory data analysis. These involves doing data profiling, summary statistics, cleaning data, statistical analysis, feature importance analysis and visualization of key features. The process of developing machine learning model is also described, which involves selecting input features, choosing algorithms for machine learning, tuning and optimizing the model, and finally evaluating the system performance.

Chapter5: Discussion The discussion chapter is characterized as a re-statement of the research aim, objectives, and questions, along with findings congruent with the research questions. The report is evaluated against the wider research literature and reasoning that follows it. The way analysis is carried is also provided.

Chapter6: Conclusion: The conclusion chapter summarizes the research, presents the results and implications for practice which are a value for the knowledge. The recommendations for further work are given, together with the final remarks.

CHAPTER 2: LITERATURE REVIEW:

2.1 Literature Review:

In his paper, (Ashton, 2018) has reported that crime rates can vary based on the time and place that the convict lives in because it's dependent on many reasons which have been proven by social scientists. But it has been attributed through numerous research that one of the significant factors to impact crime rates is low IQ. Based on studies, it has been proved that convicted persons in general have a standard deviation unit below average on the tests to gauge mental ability. An important hypothesis is also proven along with it that the person who commits the crime has lower IQ as opposed to the fact that the person who gets caught for criminal activities has lower IQ.

In their paper (Lewis, 2013), having been employed at the Home Office for 14 years in the 1980s and 1990s wants to better the understanding of the public in terms of the reported crime figures across different resources. The author also accounts for the factor that crimes change over time because of the development of technology, goods and services and many other factors. The main concerns portrayed in this paper are the core reasons for the initiation of criminal activities; are they political agenda or are they because of how news outlets report them for viewership's and many other reasons. Getting into the depths of the reasons behind these published reports and studying them thoroughly is what the author has also focused on in this research material.

The research work by (McMahon, 2010) has shown that the rise in education has a deterrent effect on the crime rates in a given region. Hence, the author wants to put emphasis on the need to educate the masses. In the entire relationship of education and types of crimes it was also observed that low to no education leads to petty crimes as well as in some cases serious crimes while in the case of highly educated individuals, crimes are more inclined towards white collar crimes or very low crime rates in general.

In their paper (Wang, 2016), has discussed the importance of being able to determine the different factors that impact the crime rates of different geographies. These factors ultimately can be used by policy makers to reduce crimes and thereby reduce the safety index of the region. A different approach to understanding the crime rates is taken in his paper which talks about the crime rates at neighbourhood level and its impact. Generally, the crime rates are studied using demographics and geographies, with the presence of modern tagging technologies it is possible to dive deeper at a granular level to understand more about criminal activities with a holistic picture in mind. With such an approach, Wang and his team were able to show some novel features which were highly significant in determining the crime rates in different regions.

In the article written by (Komiya, 1999), the author has attempted to showcase the significant gap between Japan and the Western world in terms of crime rates due to the legal culture setup between both these countries. The locality setup in the Japanese cities provides a sense of security, which can lead to lower crime rates ultimately. With such a setup of neighbourhoods, there is usually a restraining effect implied by localities which leads to lower crime rates in general. (McCue, 2015) With the individualistic approach in the Western world, there is a constant need to be expressive about people and their attributes and views due to which the crime rate prevention framework does not always apply. In this way, Japan has a more effective measure in place with the help of a locality-based setup of the population as opposed to the Western world.

In his article (Crutchfield, 1982) looks into the effect of the mobility in the residential setups on the crime rates constituted by the 65 biggest metropolitan statistical areas in the entire country. The interpretation of the mobility vs the crime rate relationship and its collinear dependence on each other is also mentioned in the article. Due to the higher motility rates, it is observed that it has a negative effect on the social integration factor of the population due to the constant motion which leads to lower control mechanisms on the individuals, leading to higher crime rates.

The factors that lead to the differences between the crime rates across different countries is analysed **in the paper by (Soares, 2004)** which primarily considers the reporting rates along with the development of the region. It was observed in the study that the reporting rates are quite related to the development, i..e, the rich countries report high crime rates while it is the opposite for the poor countries. There is a positive relationship between the development and crime rate which was determined in earlier studies, and this showed a correlation to substantiate the above effect. Ultimately, as also corroborated in some other research, inequality reduction and growth increase ultimately lead to lower crime rates.

During the COVID-19 pandemic, various regions across the globe were asked to make stay-at-home orders mandatory to stop the spread of the virus. Though this was intended for one reason, it also led to some other complications for the governments. Through their **study (Boman, 2020),** it has been attempted to depict the impact of government rules due to COVID-19 that had an impact on the crime rates in different regions. Even though the overall crime rates of the United States for instance saw a decline, there were chances that domestic homicides and batteries between intimate partners might have increased with such lockdown measures. Through their research, they have tried to identify the actual trends which were not evident during the lockdown phase across the world.

In his paper (Tonry, 2014) has mentioned about the decline in the homicide rates from 20 to 100 per 100,000 population in the west side of Europe since the late Middle Ages. These trends in the crime rates over long periods of times have been identified in his research work with the research also emphasising on the major cities where the average crime rates also fell overall until the middle of the twentieth. However, there was a steep increase in the crime rates in terms of property and violence from the 1960s to 1990s which was observed in the wealthy countries especially.

In their paper, Cohen proposed a routine approach to analyse crime rates and different trends associated with the same. Their research is centred more towards the circumstances that lead to criminal acts by the perpetrator. It was also studied that with the negative social structure in place, this allows illegal activities to increase in a society or an individual's daily life. Many data points were presented corroborating the hypothesis proposed by the authors which help in explaining the crime rates in the US between 1947 to 1974. (Cohen, 1979).

In another paper by Tseloni, the team examined crime trend variation between the years 1988 and 2004 for 26 different countries across five main crime categories using the International Crime victims survey. The main trends were identified with the help of multilevel statistical analyses. There were major criminal activity drops observed from the mid-1990s through the analysis that was performed. Through the analysis, it was also estimated that between 1995 and 2004 there was a decrease in different types of criminal activities; 77.1% for car thefts, 60.3% for theft from persons, 26% from burglary and so on. The study finally recommended that all criminal activities fell at the same rate across different countries.

It was determined for a crime rate study for **England and Wales between 1857 and 1892**. It was seen through the study that the overall crime rate increased for these countries and then gradually decreased. It has been interpreted that such criminal activities cannot be explained with the support of economics of the region in general. Through their study, it was observed that crime rates responded to pursuing legal and illegal activities. Economic growth and educational improvement led to decline in crime rates for the second half of the nineteenth century.

(Wong, 1994) Through this research it has been seen that there are various factors that lead to the increase or decrease in the crime rates for different cities based on many factors. Through our study we want to leverage a novel approach to also identify further 10 factors that impact crime rates in the city of Dubai which can then be reduced through advanced recommendations.

The research work by Sugie, (2018) argues that the macroeconomic indicators have received much attention, further investigation into detailed data at the home level is necessary to identify micro-level pressures. When faced with negative economic shocks like job loss or medical emergencies, individuals without social safety nets or with a history of drug addiction may turn to criminal behavior as a coping mechanism.

Welsh and Farrington, (2008) Evidence suggests that drug restriction does have a dynamic influence on crime rates, since the murder rate in the US increased dramatically following its repeal, despite having fallen throughout the era of alcohol prohibition.

According to (Owens, 2011) research during peak hours, several police agencies place community patrol squads in crime hotspots as a situational preventive strategy to reduce the likelihood of victims becoming repeat victims. Seasonal changes in crime rates are caused by several variables. As an example, according to Jacob et al. (2007) violent crimes such as rape, assault, and murder tend to spike during the hot summer months, even when looking at yearly patterns.

Gundogdu et al., 2018 research shows that the system dynamics modelling using time series analysis has shown promise in mechanically mimicking these cyclical patterns and providing suitable crime rate estimates based on month, public events, vacations, etc.

Lazer et al., (2021) research shows that the difficulty in making reliable quantitative predictions about human conduct stems from the fact that statistics don't necessarily reflect reality. Asking incarcerated criminals about their views on temporal triggers and deterrents is one example of qualitative approaches that can help law enforcement authorities (Willis, 2018).

The research work by Caplan and Kennedy, 2016 shows that the predictive analytics, which use data mining and machine learning, have helped law enforcement organizations throughout the world get important information about criminal trends, projections, and patterns. Investigations can be substantially aided by this information. Using socioeconomic variables and criminal history to carefully profile and categories areas into categories like violent-property, cluster analysis can optimize resource allocation.

Wang et al., (2016) research shows that the process of risk landscape modelling involves plotting potential areas where criminal activities could take place, for instance, schools, pubs, and transport terminuses on the maps. This helps ascertain the regions that must adopt improved measures of security in order to prevent such cases from occurring again (Hunter, 2017).

According to Stalidis et al., (2021) these algorithms are capable to forecast the crime scenes, suspects and trend with some degree of accuracy. Audits, citizen oversight committees, and the upkeep of accountability standards are vital ways to do statistical checks, which helps keep the public's faith and allows us to use technology to reduce crime more quickly. Responsible regulation, not total withdrawal, is necessary in the face of such complex situations (Chong et al., 2020).

The research work by Mohler et al., 2020 shows that the worldwide lockdowns and travel restrictions implemented in reaction to the COVID-19 pandemic had unintended consequences, one of which was changes in criminal behavior. Situational crime prevention theory states that people's shift towards more sedentary lifestyles and the cessation of routine activities led to a decline in traditional property and violent crimes like shoplifting, burglaries, and vehicle thefts. There was a dramatic spike in reports of domestic violence during household quarantines, which may have exacerbated already violent relationship conflicts.

Heerden & Hemphill, (2016) research shows that the underreporting became considerably worse as a result of the reallocation of police resources to address the epidemic, which allowed several offenders to evade detection. Along with the meteoric increase of digital transactions for work, school, and basic needs, the prevalence of E-crime such email phishing, online identity theft, hacking attempts, etc., has also been on the rise. To avoid physical target hardening measures, several types of E-crime and harassment have migrated online, where they may more readily target vulnerable and isolated populations. Dark web markets saw an increase in the trading of firearms and narcotics, which goes against travel regulations.

According to Liggett et al., (2020) The overall trend in E-crime decreased in several countries, but this varied substantially among nations and types of offences. Researchers had to account for potential confounds such participants' ages, genders, and medical histories in order to draw conclusions about the pandemic's effects. In the realm of criminal justice, may do experiments, surveys, and ethnographic observations to gain a better understanding of these processes and to evaluate theoretical frameworks. Responsible models and declarations are essential for avoiding hasty judgements (Obia, 2021).

Komatsu et al., (2021) research describe that there is in contrast to the 1980s and 1960s, when E-crime rates were on the rise. Stronger institutional controls and socioeconomic advancement, according to academics, tend to reduce property crimes, which are exacerbated by inequality and instability in the early phases of industrialization. The "crime-development" relationship is multi-faceted and non-linear, contrary to popular belief. While temporary issues are caused by fast urbanization and migrant influx, balanced integration between traditional country life and new city cultures is necessary.

According to Aitkenhead et al., (2022) research, the gap between the "haves" and the "have-nots" is less pronounced now than it was in the past, and inequality is more pervasive within groups. It is crucial to use robust statistical modelling that controls major confounds in order to uncover the real dynamics underlying criminal behaviors in both emerging democracies and established economies.

Qualitive research of Robinson et al., (2020) shows the comparative research utilizing field interviews might be employed to assess cultural attitudes, community relationships, legal validity, etc., in order to augment macro data with sociological insights (Bennett et al., 2017). It is critical to assess data in a responsible manner; making ill-informed policy judgements based on inflated or understated numbers without a whole picture might cause biases (Sunstein, 2015).

According to research conducted by (Lo'ai et al., (2016) that the maps might be utilized for specific patrols, which reveals that the multiple randomized controlled studies found that a mix of more police presence and AI-based predictions significantly reduced crime rates compared to traditional enforcement techniques without AI in their projections (Mohler et al., 2020). Machine learning systems adapt to new information by constantly improving their knowledge. Detectives can use virtual suspect identification skills to search through photos, drawings, and surveillance footage for matches using algorithms that detect faces.

According to the research of World Health Organization, (2022) fair regulatory norms that balance public safety and civil freedoms are necessary for these new tools to be fully utilized. Stakeholder engagement is also crucial in order to develop confidence Responsible legislation that prioritizes accountability over abandonment is advocated as a way forward as communities struggle to move advanced analytics from the lab to the field.

Domenic Antonucci, (2017) book, The Cyber Risk Handbook: Creating and Measuring Effective Cybersecurity Capabilities (Wiley Finance)" helps organisations assess and manage cybersecurity risks, including e-crime. Due to the intricacy and frequency of cyber-attacks, Dupont believes organisations must be proactive and comprehensive about cybersecurity. The book discusses NIST and ISO 27000 cybersecurity risk management frameworks. Dupont prioritises risk assessments, access controls, data security, and incident response. The author suggests anomaly detection and behavioural analysis using machine learning to identify and prevent cyber threats. Organisations must assess the merits and cons of new technology, says Dupont.

2.2 Main keyTakeaways

- E-crime levels are impacted by socioeconomic factors like income inequality, unemployment as well as demographic attributes like gender, age, and population density.

- Temporal crime patterns exist - certain hotspots see spikes during evenings or seasons.
- Machine learning models can predict E-crime locations and trends fairly accurately to aid investigators.
- COVID caused unexpected shifts in E-crime rates across countries.

CHAPTER 3 RESEARCH METHODOLOGY:

This study uses information from hidden data using the CRISP-DM method. Organized and structured observations made possible by this approach simplify the study process and improve chances of success. Within the CRISP-DM framework are the following phases:

1. Business Understanding: Defining the research goals and needs, such as problem comprehension, corporate background, and the prediction of Dubai's E-crime rates to guide safety promotion initiatives.

2. Data Understanding: Obtaining pertinent information from official sources such the Dubai Statistics Centre's socioeconomic datasets and Dubai Police. Completion, correctness, and relevance of data quality assessment with descriptive statistics and SPSS visualization tools.

3. Data Preparation: In order to get data ready for analysis, SPSS preprocesses and cleans data by merging datasets, managing missing values, resolving outliers, and changing skewed variables. dividing the dataset for training and validation of algorithms into 70% training and 30% testing data.

4. Modeling: using computer learning and data mining methods to precisely classify criminal incidents, including Random Forest models and Support Vector Machines (SVM). Scaling data to increase machine learning performance and finding attribute correlations with regression matrix.

5. Evaluation: Using test data splits and cross-validation techniques, model performance versus project goals is evaluated. Use of key performance indicators (KPIs) such as accuracy, recall, and F1 scores is combined with feature significance evaluation in Random Forest models.

To improve the accuracy in crime prediction and local risk assessment, the work combines machine learning methods, geographical analysis, and E-crime trend monitoring. Following privacy laws and government rules directs the use and access of data. Using best-performing models for criminal act predictions, such as Random Forest, SVM, Decision Trees, and Logistic Regressions, the project seeks to create efficient safety promotion programs for the Dubai Police Department.

CHAPTER 4: DATA ANALYSIS AND MODELING

4.1 Description of the Dataset Used

For this project, we will meet with Dubai Police authorities to determine the optimal dataset format to properly provide criminal activity information to users. The goal is to find a collection with explainability to better understand crimes. Two main sources will provide the dataset: Based on Dubai Police reports and local incidences.

The proposed dataset includes 9,712 Dubai e-crime reports from 2019 to 2023. Those episodes will be displayed as attributes with factors that may affect larceny. The characteristics are the offender's nationality, gender, age, crime description, arrest method, time of day, residency status, year, country, type of identification document, employment status, police station juris.

Our dataset will also include CCTV camera locations, the offender's community welfare program participation, the crime's time and day, the presence of weapons, the victim's relationship with the offender, the court decision (arrested, released), the crime's alleged motives, the time between the crime and its reporting, and the victim's relationship with the offender. This analysis adds variables to the data set in addition to existing characteristics.

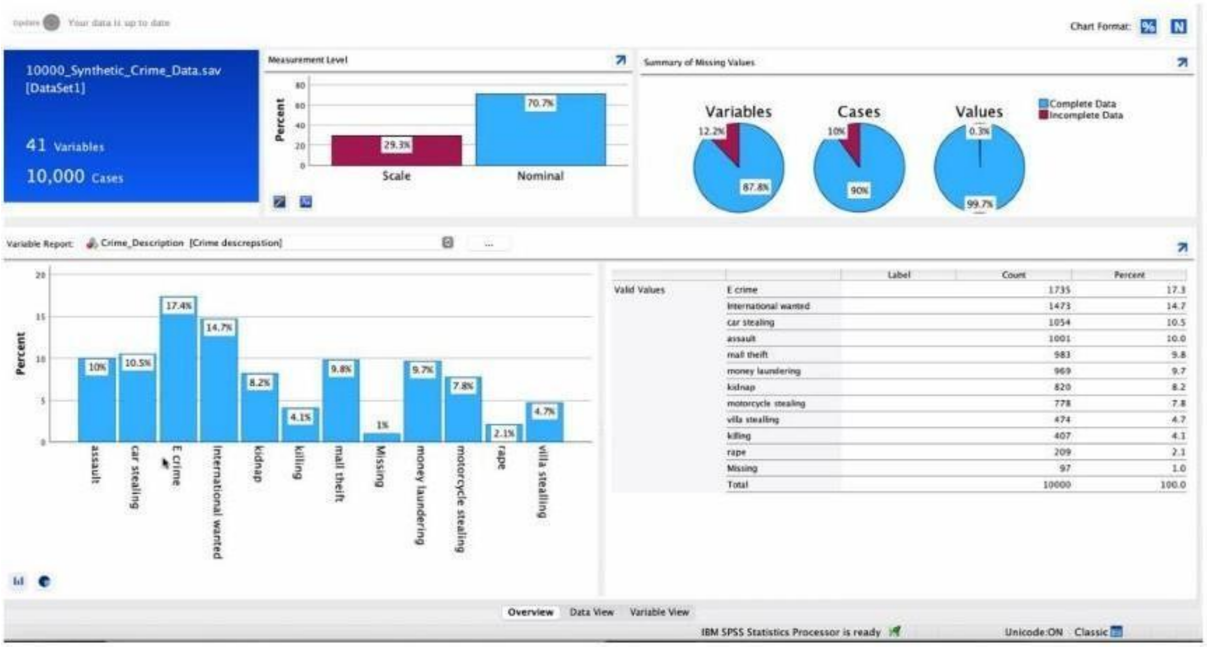


Figure 1: Data Description Dashboard.

Dubai Police representatives will collaborate on data collection for this initiative. Their experience and wisdom will help determine what information should be included in the dataset and how to adjust its format and features to provide a comprehensive understanding of crime, identify its root causes, and explain it. By combining Dubai Police criminal incident data, press stories, and local incident data, we will create a powerful and useful dataset for future project analyses and insights.

4.2 Data Dictionary:

The main core dataset used for analysis is from Dubai Police which represents reported E-crime in Dubai that run from the 2019 early 2023 period with the more than 19,292 entries. Each observation captures details related to a specific E-crime event. Relevant attributes are defined in the data dictionary below:

Attribute	Description	Data Type
Nationality	Nationality of the accused criminal	Categorical
Gender	Gender of the accused criminal	Categorical
E-Crime_Description	Type of E-crime committed	Categorical
Method_of_Arrest	How the accused was apprehended	Categorical
Time_of_Day	Timing when E-crime occurred - Day/Night	Categorical
Age	Age of the accused criminal	Numerical
Resident_Status	Residency status of the accused - Visitor, Resident etc	Categorical
Year	Year E-crime was committed	Numerical
Entry_Country	Entry method through which the accused reached UAE	Categorical
Document_Type	Visa legitimacy of the accused	Categorical
Employed_Status	Employment status of the accused	Categorical
Police_Station	Police station jurisdiction where E-crime occurred	Categorical
Area	Specific neighborhood where E-crime occurred	Categorical
Education_Level	Highest education level of the accused	Categorical
Criminal_History	Number of previous E-crimes committed by accused	Numerical
Community_Engagement_Index	Level of community policing present	Numerical

Tourist_Density_Index	Tourism activity density where crime happened	Numerical
E-crime_Rate	Underlying E-crime rate around the location	Numerical
x_coordinate	Longitude of E-crime location	Numerical
y_coordinate	Latitude of E-crime location	Numerical

Table 1: Data Dictionary.

The dataset that was gathered during this research gives a detailed description of the parameters and features used for online crimes in the UAE. Characteristics of the data collection include demographic information of the offender, e.g. gender, age, nationality, and the details about the crime, such as the type of cybercrime, time of the day, the method of arrest and the location of the event. Besides, the data set comprises contextual elements which have the potential to impact E-crime rates, like community engagement, tourist footfall, and the underlying E-crime rates. The crime statistics include both categorical and numerical variables and thus permit a multisided investigation of the causes of E-crimes in the region. This data is important in a way that it can aid law enforcement agencies, policymakers, and researchers to create specific strategies in order to prevent and counter E-crimes.

4.3 Exploratory Data Analysis

4.3.1 Data Profiling and Summary Statistics

Table 1 presents the descriptive statistics for the variables in the dataset.

1	Field	Measurem	Min	Max	Mean	Std. D	Skewne	Uniq	Valid
2	TR_Age	Continuous	19.083	41.935	27.194	5.278	0.571	--	9712
3	TR_Police_Response_Time_MN	Continuous	5.001	20	12.565	4.315	-0.03	--	9712
4	TR_Security_Camera_Density_SquareKM	Continuous	20.001	59.159	39.474	11.36	0	--	9516
5	TR_Tourist_Density_Index	Continuous	415	546	498.69	21.017	-0.217	--	9504
6	TR_Luxury_Vehicle_Density	Continuous	50.003	73.437	57.736	5.47	0.666	--	9518
7	TR_Cybercrime_Incident_Rate	Continuous	140	228	199.05	13.246	-0.144	--	9490
8	TR_Seasonal_Event_Density_Index	Continuous	20.019	135.09	57.383	26.827	0.694	--	9523
9	TR_Happiness_Index	Continuous	70	95.416	74.696	3.567	1.015	--	9712
10	L_PSLongitude	Continuous	25.003	25.273	25.107	0.089	0.294	--	9712
11	L_PSLatitude	Continuous	55.148	55.392	55.197	0.073	1.662	--	9712
12	TR_Distance_Area_Police	Continuous	0.249	30.276	10.445	6.797	0.65	--	9712
13	L_Nationalty	Nominal	--	--	--	--	--	18	9712
14	L_Gender	Nominal	--	--	--	--	--	2	9712
15	L_Crime_Description	Nominal	--	--	--	--	--	12	9712
16	L_Method_of_Arrest	Nominal	--	--	--	--	--	4	9712
17	L_Time	Nominal	--	--	--	--	--	2	9712
18	L_Age	Nominal	--	--	--	--	--	7	9712
19	L_Resident_Status	Nominal	--	--	--	--	--	3	9712
20	L_Crime_Year	Nominal	--	--	--	--	--	4	9712
21	L_Entry_Country	Nominal	--	--	--	--	--	2	9712
22	L_Document_Type	Nominal	--	--	--	--	--	2	9712
23	L_Employed_Status	Nominal	--	--	--	--	--	2	9712
24	L_Police_Station	Nominal	--	--	--	--	--	9	9712
25	L_Area	Nominal	--	--	--	--	--	4	9712
26	L_Educational_Level	Nominal	--	--	--	--	--	3	9712
27	L_Crime_Area	Nominal	--	--	--	--	--	39	9712
28	L_Criminal_History	Nominal	1	6	--	--	--	6	9712
29	TR_Presence_of_CCTV_Cameras	Nominal	--	--	--	--	--	2	9712
30	TR_Community_Engagement_Program	Nominal	--	--	--	--	--	2	9712
31	TR_Time_of_Day_of_Crime_Occurrence	Nominal	--	--	--	--	--	4	9712
32	TR_Day_of_the_Week_of_Crime_Occurrence	Nominal	--	--	--	--	--	7	9712
33	TR_Presence_of_Weapons	Nominal	--	--	--	--	--	2	9712
34	TR_Victim_Offender_Relationship	Nominal	--	--	--	--	--	4	9712
35	TR_Arrest_Outcome	Nominal	--	--	--	--	--	2	9712
36	TR_Motive_of_Crime	Nominal	--	--	--	--	--	6	9712
37	TR_Time_Interval_Crime_Occurrence_and_Repo	Nominal	--	--	--	--	--	4	9712
38	TR_Victim_Dem_Profile	Nominal	--	--	--	--	--	2	9712
39	TR_Presence_Witness	Nominal	--	--	--	--	--	2	9712
40	L_Longitude	Nominal	24.984	25.296	--	--	--	39	9712
41	L_Latitude	Nominal	55.125	55.544	--	--	--	39	9712
42	Corrected_TR_Ecrime	Nominal	--	--	--	--	--	2	9712

Fig 2: Summary Statistics.

The table above offers both category information and numerical statistics related to crime. It contains, for example, measures of the average age (19.083 with a standard deviation of 5.278) and the average police reaction time to criminal events (5.001 minutes with a standard variation of 4.315). It also provides statistics like the mean of 39.474 security camera density per square kilometer and the mean of 498.69 tourist density. Among the many categories of categorical data are nationality, gender, description of the crime, arrest procedure, and CCTV camera presence. The table attempts to provide scholars a thorough grasp of the many elements—from environmental conditions to law enforcement actions and demographics—that contribute to crime in a particular place.

The dataset includes numerical variables representing various aspects of the reported crimes, such as the age of the accused, police response time, security camera density, high-end vehicle density, event density, reported incident rates, happiness index, geographical coordinates, distance from the crime area to the police station, and derived variables like the Mahalanobis distance.

1	Field	Measurement	Outliers	Extremes	% Complete	Valid Records
2	IMP_Age1	Continuous	0	0	100	9712
3	IMP_Distance_Area_Police	Continuous	0	0	100	9712
4	IMP_Happiness_Index	Continuous	85	2	100	9712
5	IMP_Seasonal_Event_Density_Index	Continuous	0	0	100	9712
6	IMP_Luxury_Vehicle_Density	Continuous	0	0	100	9712
7	IMP_Cybercrime_Incident_Rate	Continuous	14	0	100	9712
8	IMP_Tourist_Density_Index	Continuous	23	0	100	9712
9	IMP_Police_Response_Time_MN	Continuous	0	0	100	9712
10	IMP_Security_Camera_Density_SquareKM	Continuous	0	0	100	9712
11	I_Nationalty	Nominal	--	--	100	9712
12	I_Method_of_Arrest	Nominal	--	--	100	9712
13	I_Police_Station	Nominal	--	--	100	9712
14	I_Crime_Area	Nominal	--	--	100	9712
15	TR_Day_of_the_Week_of_Crime_Occurrence	Nominal	--	--	100	9712
16	TR_Motive_of_Crime	Nominal	--	--	100	9712
17	TR_Criminal_History	Nominal	--	--	100	9712
18	Corrected_TR_Ecrime	Nominal	--	--	100	9712

Fig 3: Outliers.

Table above lists crime factors including age, police distance, happiness index, and more. The age measurement data has 0 outliers, 0 extremes, 100% completeness, and 9712 valid records. Distance to police areas is 100% complete with 9712 verified records. The happiness index evaluates happiness using 9712 valid data, 85 outliers, 2 extremes, and 100% completeness. The Seasonal Event Density Index measures seasonal event density with 100% completeness and 9712 valid data. Luxury Vehicle Density is 100% complete with 9712 verified records. With 0 outliers and 100% completeness, the cybercrime incident rate includes 9712 valid data with 14 outliners. It shows tourist density with 23 outliers and 100% completeness from 9712 reliable records. Comprehensive police response time with 9712 verified recordings. Security camera density square displays cameras per square kilometer with 100% completeness and 9712 valid records. Final categorical data includes nationality, arrest method, police station, and more, 100% complete and 9712 valid records.

4.3.2 Data Cleaning

The dataset underwent rigorous data cleaning procedures to ensure the reliability and accuracy of the analysis. This process involved the following techniques:

Audit Quality Annotations

Complete fields (%): 87.8% Complete records (%): 90.05%

Field	Measurement	Outliers	Extremes	Action	Impute Missing	Method	% Complete	Valid Records	Null Value
TR_Cybercrime_Incident_Rate	Continuous	14	0	None	Never	Fixed	97.714	9490	222
TR_Tourist_Density_Index	Continuous	21	0	None	Never	Fixed	97.858	9504	208
TR_Security_Camera_Density_S...	Continuous	0	0	None	Never	Fixed	97.982	9516	196
TR_Luxury_Vehicle_Density	Continuous	0	0	None	Never	Fixed	98.002	9518	194
TR_Seasonal_Event_Density_Ind...	Continuous	0	0	None	Never	Fixed	98.054	9523	189
I_Nationalty	Nominal	--	--	--	Never	Fixed	100	9712	0
I_Gender	Nominal	--	--	--	Never	Fixed	100	9712	0
I_Crime_Description	Nominal	--	--	--	Never	Fixed	100	9712	0

Fig 4: Data Cleaning.

The table provides insights into completion rates, outliers, missing data, and field variables. With a completion rate of 87.8%, about 12.2% of fields lack information, while the overall completion rate is 90.05%, indicating 9.95% of records have incomplete data. Notably, the Tourist Density Index has the highest number of outliers at 21, suggesting potential data anomalies. Fields like Cybercrime Incident Rate and Tourist Density Index with null values require attention during data cleaning. While completion rates vary among indicators, maintaining data consistency is crucial for effective analysis and decision-making. Some features like nationality, gender, and E-Crime exhibit 100% completion rates, enhancing their reliability for analysis.

Completion rates vary among the first 5 indicators. For example, Cybercrime Incident Rate has a 97.714% accuracy rate, meaning 2.286% of records are incomplete. Similarly, Tourist Density Index has a 97.858% completion rate, with 2.142% incomplete records. Security Camera Density

Score and Luxury Vehicle Density have completion rates of about 97.982% and 98.002%, respectively, with just a small percentage of incomplete records. Lastly, Seasonal Event Density Index has a 98.054% completion rate, indicating only 1.946% incomplete records. Addressing incomplete data is important for reliable analysis and decision-making.

4.3.3.1 Techniques for Handling Missing Data

The hypothesis had initially some empty cells for certain columns. To avoid this problem, suitable imputation strategies was used. It was the median of the continuous variables for which a case of the missing data imputation was created.

Complete fields (%): 100% Complete records (%): 100%

Field	Measurement	Outliers	Extremes	Action	Impute Missing	Method	% Complete	Valid Records	Null Value
I_Nationalty	Nominal	--	--	--	Never	Fixed	100	9712	
I_Gender	Nominal	--	--	--	Never	Fixed	100	9712	
I_Crime_De...	Nominal	--	--	--	Never	Fixed	100	9712	
I_Method_of...	Nominal	--	--	--	Never	Fixed	100	9712	
I_Time	Nominal	--	--	--	Never	Fixed	100	9712	
I_Resident_...	Nominal	--	--	--	Never	Fixed	100	9712	
I_Crime_Year	Nominal	--	--	--	Never	Fixed	100	9712	
I_Entry_Cou...	Nominal	--	--	--	Never	Fixed	100	9712	
I_Document...	Nominal	--	--	--	Never	Fixed	100	9712	
I_Employed_...	Nominal	--	--	--	Never	Fixed	100	9712	
I_Police_Stat...	Nominal	--	--	--	Never	Fixed	100	9712	
I_Area	Nominal	--	--	--	Never	Fixed	100	9712	
I_Educationa...	Nominal	--	--	--	Never	Fixed	100	9712	
TR_Age	Continuous	0	0	None	Never	Fixed	100	9712	
I_Crime_Area	Nominal	--	--	--	Never	Fixed	100	9712	
I_Criminal_H...	Nominal	--	--	--	Never	Fixed	100	9712	
TR_Presenc...	Nominal	--	--	--	Never	Fixed	100	9712	
TR_Commu...	Nominal	--	--	--	Never	Fixed	100	9712	
TR_Time_of...	Nominal	--	--	--	Never	Fixed	100	9712	
TR_Day_of_t...	Nominal	--	--	--	Never	Fixed	100	9712	
TR_Presenc...	Nominal	--	--	--	Never	Fixed	100	9712	
TR_Victim_O...	Nominal	--	--	--	Never	Fixed	100	9712	
TR_Arrest_O...	Nominal	--	--	--	Never	Fixed	100	9712	
TR_Motive_o...	Nominal	--	--	--	Never	Fixed	100	9712	
TR_Time_Int...	Nominal	--	--	--	Never	Fixed	100	9712	
TR_Victim_D...	Nominal	--	--	--	Never	Fixed	100	9712	

OK

Figure 5: Handling of Missing Data.

Accordingly, missing data was dealt with by using the appropriate imputation strategies, particularly for continuous variables, where the median value of the variables was taken as the imputation method. With this technique, the statistical features of the dataset were preserved, and the values were incomplete were successfully handled.

The wholeness of the data set is indeed a proof of the process of data scrubbing and imputation being taken as a step to ensure quality. Dataset completeness is a key factor for successful analysis, as it will eliminate the need for additional data work during the modeling or analysis stage. It provides the basis for the fact that all data is considered, and hence, more systematic and deep conclusions are reached.

The expenditure of time and resources in order to generate a fully functioning data ecosystem exemplifies the importance of detail, as well as the dedication to data quality and reliability. An integrity lost the dataset is the missing values compacted, which is ready to be used for the later analysis and modeling tasks, so that a strong basis is gained for drawing insightful information and making decisions.

4.3.3.2 Approaches for Detecting Outliers

The table provides insights into outliers detected in three fields: Happiness Index, Tourist Density Index, and Cybercrime Incident Rate. For instance, the Happiness Index had 85 outliers and 2 extreme values, indicating data points significantly diverging from the general distribution.

Field	Measurement	Outliers	Extremes
TR_Happiness_Index	Continuous	85	2
TR_Tourist_Density_Index	Continuous	21	0
TR_Cybercrime_Incident_Rate	Continuous	14	0

Table 2: Outliers

To detect outliers, Standard Deviation (SD) was employed. It measures the degree of variation in a dataset. Values beyond the mean ± 3SD are considered outliers, while those beyond mean ± 5SD are extreme. Using this method, 85 outliers and 2 extreme values were identified for the Happiness Index.

Additionally, the study utilized the SO Anomaly algorithm with SPSS Modeler to automatically detect anomalies in the dataset. This method, combined with descriptive analysis, uncovered a total of 6 anomalies, highlighting data points that deviate from expected patterns or behaviors (Anomaly Detection Node, 2023).

Value	Proportion	%	Count
0.000		99.94	9706
1.000		0.06	6

Figure 6: Outlier Geography Distribution.

The illustration (Figure 5) visually represents the outlier detection process using SPSS Modeler. It involves identifying missing values, detecting outliers using SO Anomaly, retaining non-outlier observations, and selecting significant factors using random trees and feature importance. This comprehensive approach ensures data cleanliness and reliability, ultimately improving the accuracy of statistical analysis and modeling results (IBM Documentation, 2023).

4.3.4 Statistical Analysis:

4.2.4.1 Cross tab:

TR_Motive_of_Crime * Crime descrepstion Crosstabulation

Count

		Crime descrepstion												
		assault	car stealing	E crime	International wanted	kidnap	killing	mall theift	money laundering	moto	motorcycle stealing	rape	villa stealling	Total
TR_Motive_of_Crime	Finacial Gain	338	375	598	510	266	127	354	351	0	242	71	160	3392
	Ideological	92	108	176	145	80	50	89	92	0	74	20	48	974
	Opportunistic	190	172	303	262	162	66	176	169	0	146	39	81	1766
	Others	39	49	88	80	38	18	37	50	0	45	10	26	480
	Personal Conflict	129	119	195	165	93	52	125	108	0	96	20	51	1153
	Revenge	195	214	335	283	167	88	183	183	1	158	43	97	1947
Total		983	1037	1695	1445	806	401	964	953	1	761	203	463	9712

Fig 7: Motive of Crime; Crime Description.

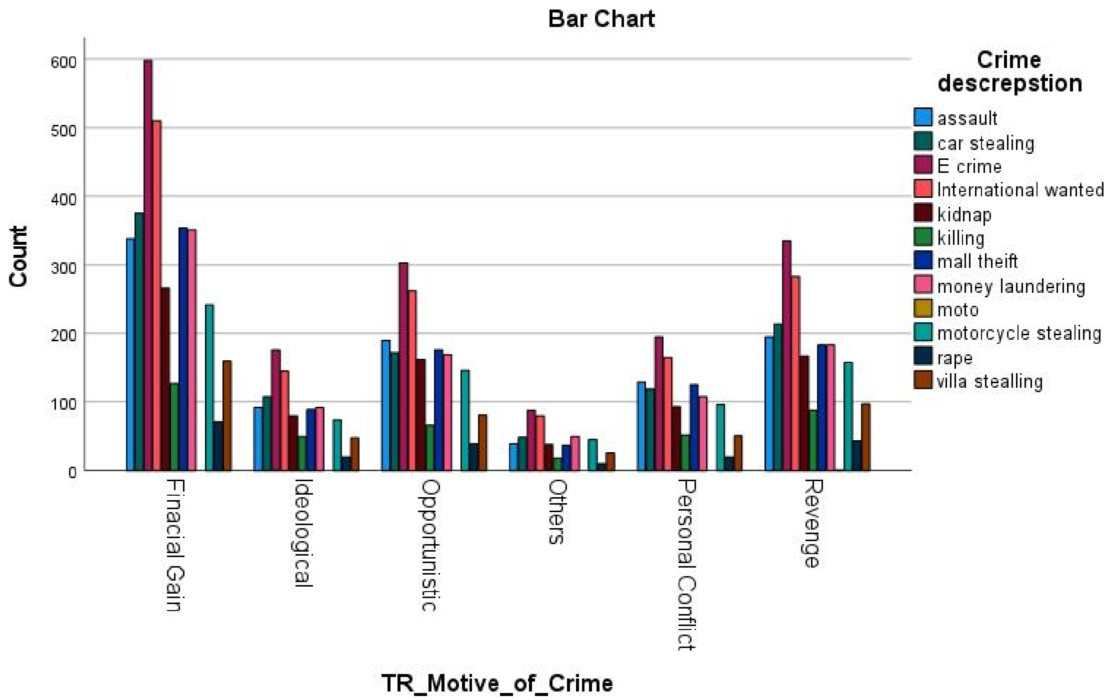


Fig 8: Crime counts; Crime Description.

TR_Motive_of_Crime * Police Station Crosstabulation

Count

		Police Station									
		barsha	bur dubai	jabal ali	khawaneej	mrqbat	naif	qusais	rashdya	Refaa	Total
TR_Motive_of_Crime	Finacial Gain	128	1384	1186	75	255	50	93	137	84	3392
	Ideological	36	397	345	16	55	24	40	42	19	974
	Opportunistic	69	703	594	43	126	51	47	84	49	1766
	Others	15	195	160	6	39	8	19	21	17	480
	Personal Conflict	41	462	384	22	95	26	40	47	36	1153
	Revenge	90	744	733	34	124	33	55	72	62	1947
Total		379	3885	3402	196	694	192	294	403	267	9712

Figure 9: Motive of Crime; Police Station.

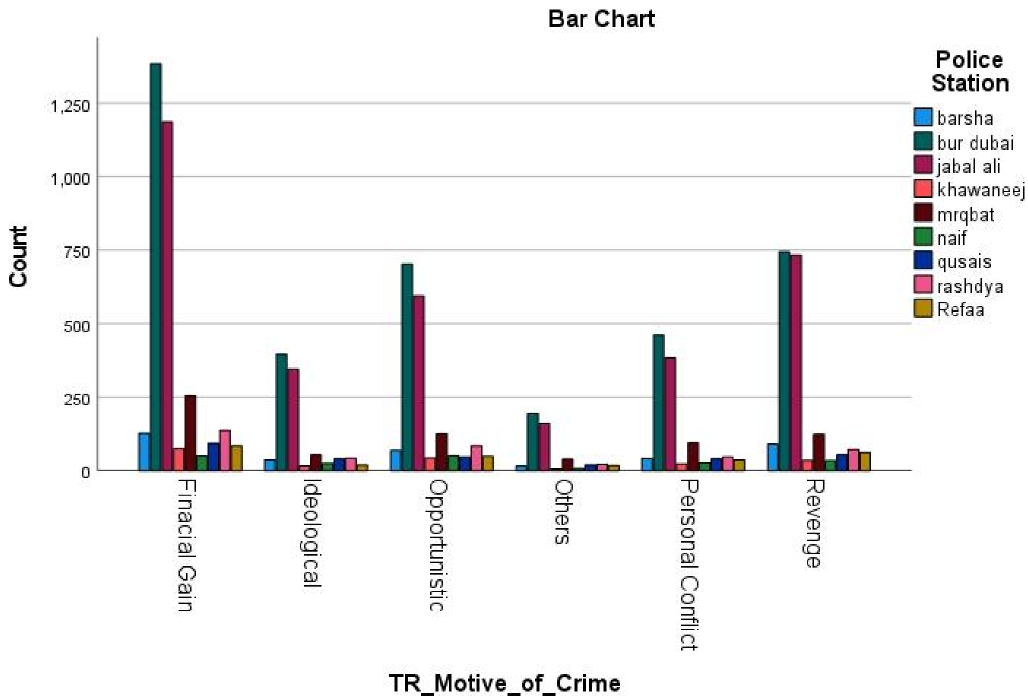


Fig 10: Crime Counts; Police Station.

The cross-tabulation analysis of the Dubai E-crime dataset reveals a strong correlation between crime descriptions, motives, and police stations, providing valuable insights into criminal activity distribution. The highest number of cases is observed at Bur Dubai and Jabal Ali police stations, with 1504 and 1315 cases respectively. E-crime constitutes the majority of reported incidents (1,695), followed by international wanted (1,445), car stealing (1,037), and assault (983). The distribution of guilty individuals across age groups is skewed towards the young population, particularly in the 21-25 age range.

4.2.4.2 Comparing mean:

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
Nationality * TR_Age	9712	100.0%	0	0.0%	9712	100.0%
Police Station * TR_Age	9712	100.0%	0	0.0%	9712	100.0%
Nationality * TR_Police_Response_Time_MN	9712	100.0%	0	0.0%	9712	100.0%
Police Station * TR_Police_Response_Time_MN	9712	100.0%	0	0.0%	9712	100.0%
Nationality * TR_Happiness_Index	9712	100.0%	0	0.0%	9712	100.0%
Police Station * TR_Happiness_Index	9712	100.0%	0	0.0%	9712	100.0%

Figure 11: Summary.

In the categorical variables part of the summary, this research investigates the relationships between nationality and police station using mean comparison tests and the relationships between the continuous variables (TR_Age, TR_Police_Response_Time_MN, and TR_Happiness_Index) using the same tests. It is implied that all 9,712 cases in the sample were involved in the analysis, and that there were no cases that were excluded by the research. This implies the mean comparison tests for categorical and continuous variables to be performed on the entire dataset, which guarantees that all the interactions between them be included in the final output. Through the analysis of the mean-differences among different categories, which can obtain information of possible inequity or the commonalities among age, police response time, and happiness index according to countries and police station authority. The results of these studies can help the resource allocation, policy making, and development of specific purposes to solve the problems and reach the equal outcomes for different social and geographical groups.

4.3.5 Feature Selection Analysis:

In these section two techniques were employed to analyze crime data: random tree and feature selection:

Fig 12: Random Tree Predictor Importance

Firstly, the random tree algorithm was utilized to rank the most influential predictors. The Random Tree node can be seen as a fitted tree-based classification and prediction tool, which is built on Classification and Regression Tree techniques. Similar to the C&R Tree's procedure, these prediction methods apply recursive partitioning to split the training records into segments with similar target field values (Dataplatform.cloud, 2024). The top ten important predictors identified included Crime Area, Crime Description, Nationality, Day of the Week of Crime Occurrence, Motive of Crime, Age, Area, Cybercrime Incident Rate, and Method of Area. Notably, the Crime Area emerged as the most critical predictor.

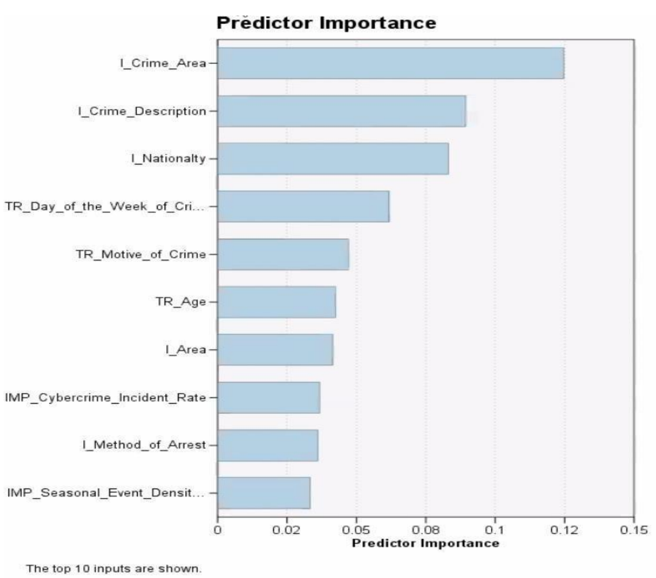


Fig 13: Feature Selection.

Secondly, feature selection was implemented to select the most important features from a larger set, aiming to improve machine learning model performance and efficiency. Feature selection is the technique of data mining which enhances the quality of the data by finding the most significant input variables that can be used for predicting a target variable. It implies eliminating the troublesome or unimportant inputs, sorting out the rest of the inputs based on their importance, and then picking up the most essential subset of them to be used for modeling in the next stage, thus the modeling process is accelerated while being simplified (Ibm.com, 2024). Out of 41 available bio-clinical fields, nine were identified as possible predictors. The five most substantial determinants were Crime Area, Day of the Week of Crime Occurrence, Nationality, Motive of Crime, and Police Station, each with an importance of 1.0. This process helped filter relevant features and determine their relative importance. Through comparison of the results from both methods it was discovered that Crime Area, Nationality, Day of the Week of Crime Occurrence, and Motive of Crime were reproducibly found to be key predictors.

Rank

Rank	Field	Measurement	Importance	Value
1	I_Area	Nominal	Important	1.0
2	I_Crime_Area	Nominal	Important	1.0
3	TR_Day_of_the_Week_of_Crime_Occurrence	Nominal	Important	1.0
4	I_Nationalty	Nominal	Important	1.0
5	TR_Motive_of_Crime	Nominal	Important	1.0
6	I_Police_Station	Nominal	Important	1.0
7	I_Method_of_Arrest	Nominal	Important	1.0
8	I_Crime_Description	Nominal	Important	1.0
9	IMP_Distance_Area_Police	Continuous	Important	1.0
10	TR_Time_of_Day_of_Crime_Occurrence	Nominal	Marginal	0.943
11	TR_Presence_Witness	Nominal	Marginal	0.931
12	IMP_Police_Response_Time_MN	Continuous	Marginal	0.926
13	I_Crime_Year	Nominal	Unimportant	0.896
14	TR_Victim_Offender_Relationship	Nominal	Unimportant	0.89
15	I_Employed_Status	Nominal	Unimportant	0.863
16	TR_Time_Interval_Crime_Occurrence_and_Reporting	Nominal	Unimportant	0.748
17	TR_Age	Continuous	Unimportant	0.55
18	IMP_Age1	Continuous	Unimportant	0.55
19	TR_Presence_of_Weapons	Nominal	Unimportant	0.538
20	I_Criminal_History	Nominal	Unimportant	0.502

Selected fields: 9 Total fields available: 41

> 0.95 <= 0.95 < 0.9

	Field	Measurement	
	IMP_Tour...	Continuous	Coefficient of variation below threshold
	IMP_Lux...	Continuous	Coefficient of variation below threshold
	IMP_Hap...	Continuous	Coefficient of variation below threshold
	IMP_Cyb...	Continuous	Coefficient of variation below threshold
	I_PSLon...	Continuous	Coefficient of variation below threshold
	I_PSLatit...	Continuous	Coefficient of variation below threshold
	I_Longitu...	Nominal	Single category too large
	I_Latitude	Nominal	Single category too large
	I_Gender	Nominal	Single category too large
	I_Docum...	Nominal	Single category too large
	$O-Anom...	Flag	Single category too large

Fig 14: Non important features selected by feature selection algorithm.

4.3.6 Visualization of Key Features

To gain a visual understanding of the dataset and uncover potential patterns or relationships, various plots and graphs were generated

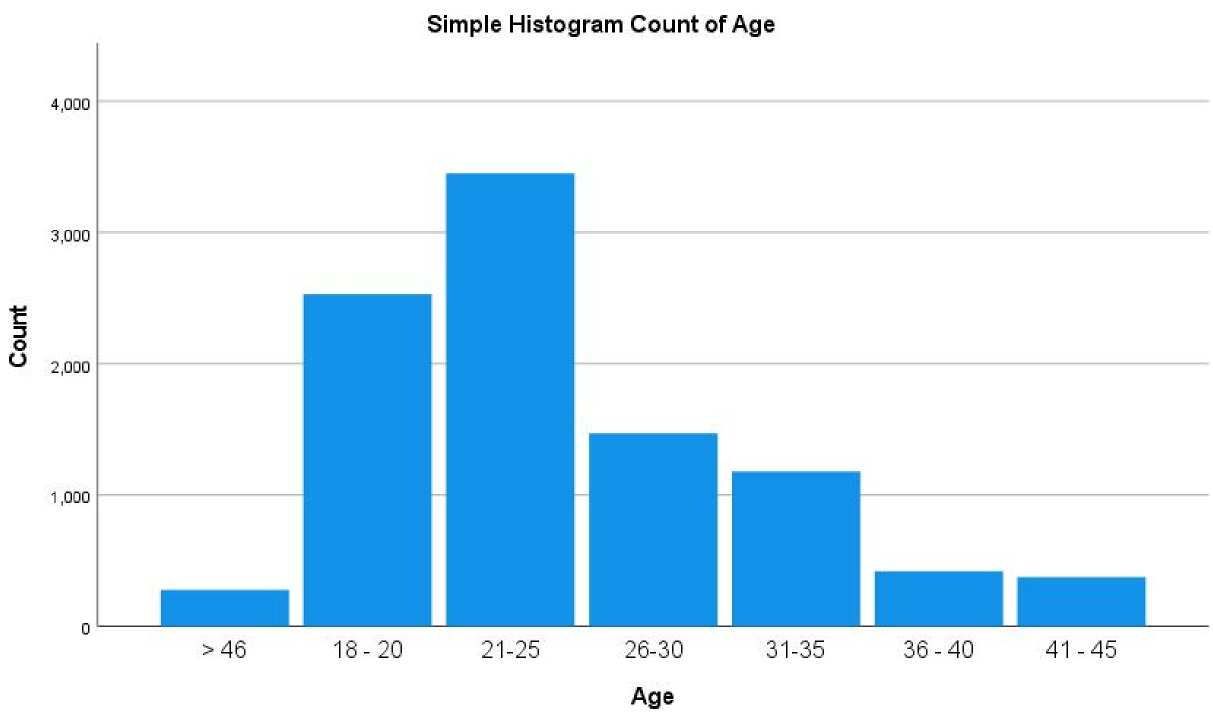


Fig 15: Histogram of Age Distribution.

The survey data set shows the histogram shows that the age distribution is slightly skewed to the right, with the majority of the accused individuals falling within the age range of 21 to 25 years.

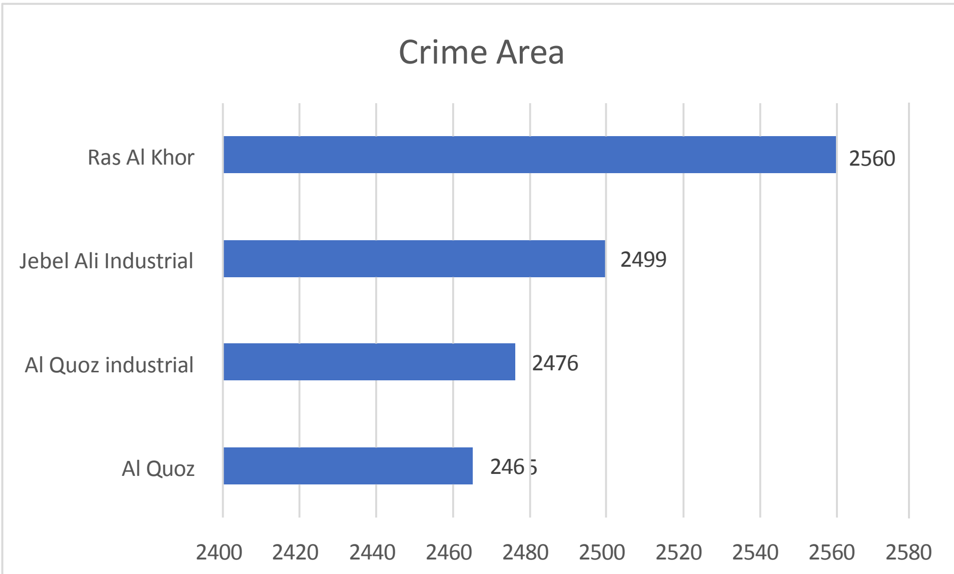


Figure 16: Histogram of Crime area

The data obtained from the survey shows that the data indicates that Ras Al Khor clocked in with the most cases (2,560) and Jebel Ali Industrial was close behind with 2,499 incidents. For Al Quoz Industrial and Al Quoz, there are only 2,476 and 2,465 cases, respectively, making the counts in these areas relatively low. The fact that the incidents are clustered in Ras Al Khor and Jebel Ali Industrial areas implies that the places share some attributes or factors that resulted to the same crime incident. The numbers of Al Quoz Industrial and Al Quoz Author: Artificial Intelligence

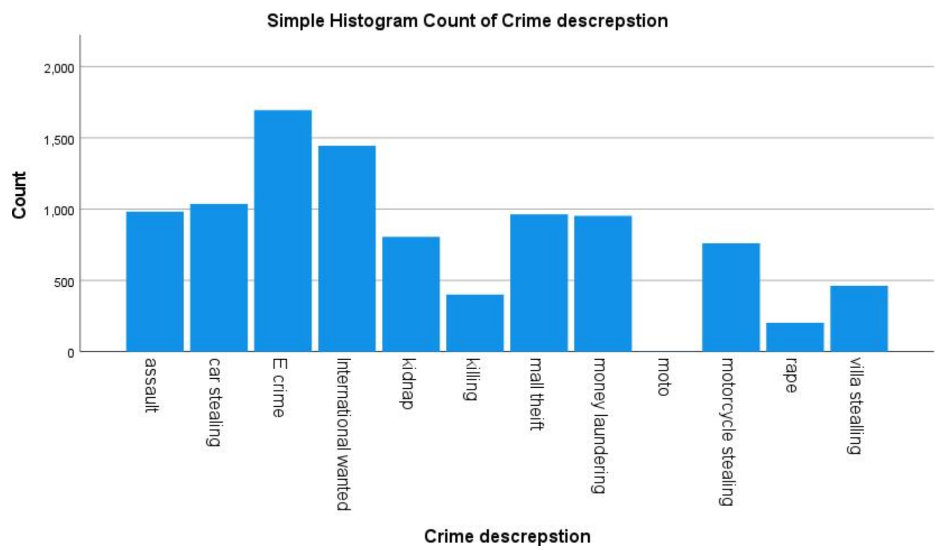


Figure 17: Bar Chart of Crime Type Distribution.

The collected data used for the bar chart which clearly shows that e-crime is the most prevalent crime type, followed by international wanted, car stealing, and assault.

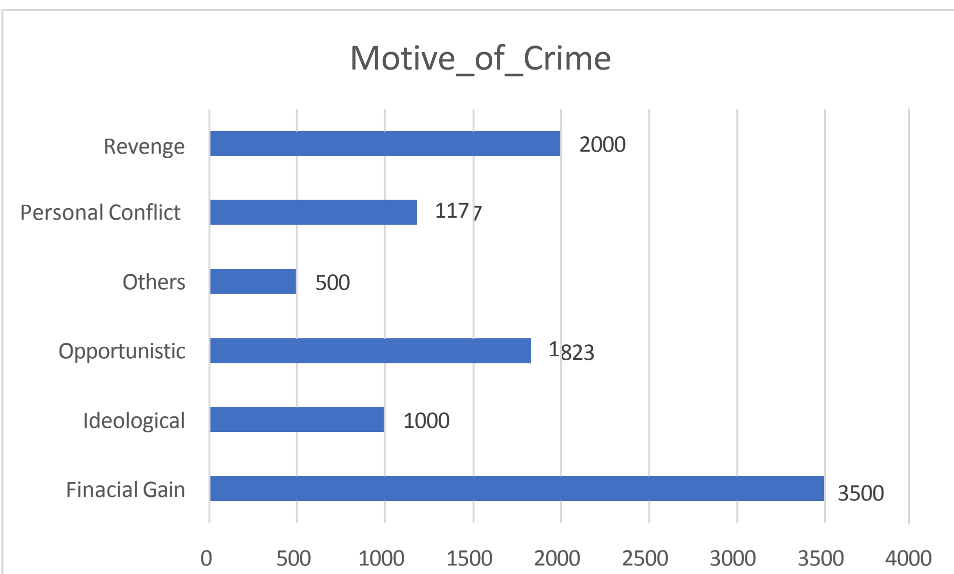


Figure 18: Motive of Crime:

The graph made from the data collection from the survey results, shows motive-based crime. Most of the 3,500 incidents include financial gain, showing that money motivates most crime.

Personal vendettas and retaliation motivate crime, with revenge being the second most common motive (2,000). Opportunistic crimes rank third (1,823), illustrating how often criminals exploit opportunities. Interpersonal conflicts (1,177) often cause crime. The 1,000 ideological cases demonstrate that politics, religion, and other ideologies can motivate crime. The 500-item "Others" category reflects non-category purposes.

4.4 Machine Learning Model Development

4.4.1 Selection Regarding Random tree and Feature Selection:

The accompanying figure visually demonstrates the automatic selection process, highlighting the importance of advanced techniques like random forest and feature selection. Through this approach, crucial input features for understanding crime dynamics and output variables for predictive modeling were determined. This integrated approach enhances researchers' ability to derive actionable insights from complex datasets, aiding informed decision-making in crime analysis and prevention efforts.

Field	Measurement	Values	Missing	Check	
I_Nationalty	Nominal	UK,UAE,Iran,India,China,Chile,Nepal,Spain,Egypt,Nigeria,Ire...		None	Input
I_Crime_Description	Nominal	rape,moto,kidnap,"E crime",killing,assault,"car stealing","mo...		None	Input
I_Method_of_Arrest	Nominal	Camera,"by call",investigation,suspicion		None	Input
I_Police_Station	Nominal	naif,Refaa,barsha,qusais,mrqbat,rashdya,"bur dubai",khawa...		None	Input
I_Area	Nominal	"Al Quoz","Al Quoz industrial","Ras Al Khor","Jebel Ali Industri...		None	Input
TR_Age	Continuous	[19.082507245658945,41.9348082257308]		None	Input
I_Crime_Area	Nominal	Liwan,Warsan,"Al Ras","Al Tay","Nadd Al Hamar","Al Quoz","A...		None	Input
TR_Day_of_the_Week_of_Crime_Occurrence	Nominal	Friday,Monday,Sunday,Tuesday,Wednesday,Saturday,Thurs...		None	Input
TR_Motive_of_Crime	Nominal	Others,Revenge,Ideological,"Finacial Gain","Personal Confli...		None	Input
IMP_Distance_Area_Police	Continuous	[0.24866711399702354,30.275815724824312]		None	Input
IMP_Seasonal_Event_Density_Index	Continuous	[20.01905066794339,135.0897081613033]		None	Input
IMP_Cybercrime_Incident_Rate	Continuous	[140.0,228.0]		None	Input
Corrected_TR_Ecrime	Flag	"E crime"/other		None	Target

Figure 19: Automatic selection process.

4.4.2 Detailed Explanations of the Chosen Machine Learning Algorithms:

In our study on container inspection prediction, we've opted for five machine learning algorithms to tackle the task.

- **Logistic Regression** stands out for its simplicity and effectiveness in binary classification, making it easy to comprehend and computationally efficient.
- **Linear Support Vector Machine (LSVM)** is chosen for its adeptness in handling linearly separable data, offering efficiency in classifying instances into two classes.
- **Neural Networks** are included to capture intricate, nonlinear relationships within the data, though they may require substantial data for training due to their complexity.
- **C&R Trees** are selected for their simplicity and interpretability, making them suitable for identifying significant predictors and capturing non-linear relationships.
- **Decision Trees** are chosen for their ability to partition the feature space and interpret feature hierarchies effectively.

These algorithms collectively provide a comprehensive toolkit for analyzing container inspection data, each offering unique strengths and capabilities tailored to the predictive task at hand.

4.4.3 Validation and testing procedures:

4.4.3.1 Data Partitioning

This section discusses how the researcher validated and tested machine learning models. First, split the dataset was split into three parts: training, validation, and testing sets. A common technique called train-test split was used, where 70% of the data is used for training the models, and the remaining 30% is kept aside for testing. This ensures that the models learn from most of the data while still having unseen data to test their performance. Additionally, the use of cross-validation, which divides the data into multiple subsets, or folds, and trains the model iteratively on different combinations of these subsets was also picked to ensure robustness in our evaluation process.

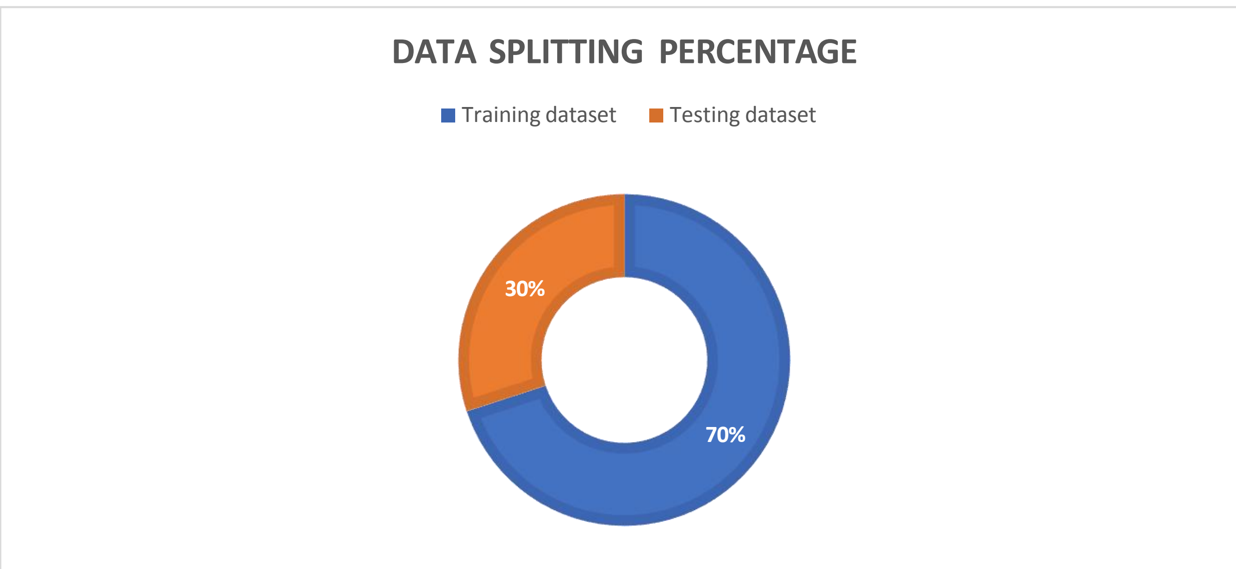


Fig 20: Data Partition.

The table illustrates the partition sizes for the training and testing sets. In the "Train and test" row, the training partition size is 70%, and the testing partition size is 30%. The labels for these partitions are "Training" and "Testing," respectively. The corresponding values are "1_Training" and "2_Testing." The second row, "Train, test and validation," has both partition sizes set to 0, indicating that this partitioning scheme was not used in the current study. By employing these validation and testing techniques, the researcher aims to ensure that the machine learning models are properly evaluated, and their performance is assessed on unseen data, providing a more reliable indication of their generalization capabilities.

4.4.4 Evaluation Metrics

This section examines machine learning models with ROC curves and AUC. ROC curves show a model's classification abilities by comparing TPR and FPR at various thresholds. FPR measurements misclassified negative events as positive, whilst TPR measures correctly identified positive events (TP). TN and FN are true and false negatives, respectively. High TPR and low FPR indicate a superior model. From 0 to 1, AUC assesses model performance. Higher values indicate better performance. Through the FPR range of 0 to 1, AUC = ∫(TPR d(FPR)). This Dubai crime rate and e-crime prevention study compares two ROC curve-based machine learning methods. The E1F-corrected model detected e-crime better than the E1F-Crime model in AUC, TPR, and FPR. The best Dubai crime prediction and prevention model is chosen by comparing ROC curves and AUC values. To evaluate machine learning models, confusion matrix, precision, recall, and F1 Score are examined.

4.4.4.1 Confusion Matrix:

A confusion matrix evaluates a classification model by comparing expected and actual labels. It lists the model's TP, TN, FP, and FN forecasts. The E1F-Corrected model confusion matrix may look like this in this research of Dubai crime rate and e-crime prevention studies. Using the confusion matrix, we can calculate accuracy, precision, recall, and F1 score.

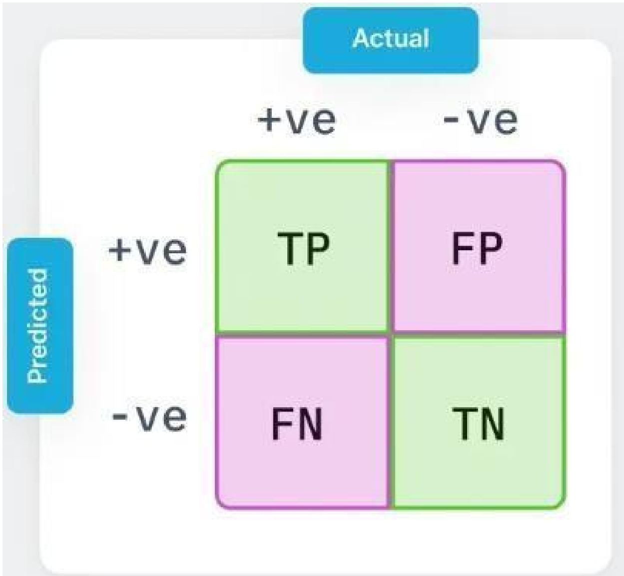


Fig 21: Figure illustrating confusion matrix.

4.4.4.2 Precision:

Precision quantifies accurate positive predictions out of all positive forecasts. It shows how often e-crime predictions are correct.

Precision = TP / (TP + FP)

High accuracy implies the model seldom labels non-e-crime as e-crime thanks to its low false positive rate.

4.4.4.3 Recall:

Recall, sensitivity, or true positive rate (TPR) estimates the percentage of genuine positive predictions among all positive cases. It shows how many positive e-crime occurrences the programmed accurately identifies.

Recall = TP / (TP + FN)

A high recall means that the model has a low false negative rate, i.e., it rarely misses actual e-crime instances.

4.4.4.4 F1 Score:

The harmonic means of accuracy and memory, the F1 score, balances both metrics. It is especially effective in unbalanced datasets when one class (e.g., e-crime) is much less common than the other.

F1 Score = 2 * (Precision * Recall) / (Precision + Recall)

A model with a high F1 score balances false positives and negatives with good accuracy and recall. Our work uses confusion matrices to calculate accuracy, recall, and F1 score for the E1F-Corrected and E1F-Crime models. We may compare these measures to discover which model identifies e-crime more reliably and with fewer false positives and negatives. Since it balances precision and recall better, the E1F-Corrected model may be better for crime prediction and prevention in Dubai if it has a higher F1 score. Analyzing the confusion matrix, accuracy, recall, and F1 score can help us identify the optimum machine learning model for our use case.

4.3.4 Result

4.3.4.1 Presentation of the Experimental Results

Model	NO. Fields used	Overall Accuracy (%)	Accumulated Accuracy (%)	Area Under Curve	Accumulated AUC	Precision	Recall	F1-Measure
Logistic Regression	12	85.871	85.871	0.932	0.932	0.855	0.855	0.855
LSVM	12	85.46	85.46	0.93	0.93	0.851	0.852	0.852
Neural Networks	12	82.819	82.819	0.909	0.909	0.809	0.848	0.848
C&R Trees	11	72.942	72.942	0.784	0.784	0.747	0.674	0.674
Decision List	4	66.975	66.975	0.689	0.689	0.67	0.639	0.639

Figure 22: Performance comparison.

Key Insights:

1. Logistic Regression Model:

- Utilizes 12 predictors fields for prediction.
- Achieves an overall accuracy of 85.871% and an accumulated accuracy of 85.871%.
- Exhibits an Area Under Curve (AUC) of 0.932 and an accumulated AUC of 0.932.
- Precision, recall, and F1-measure are 0.855, 0.855, and 0.855 respectively.

2. LSVM Model:

- Utilizes 12 predictor fields, similar to Logistic Regression.
- Achieves an overall accuracy of 85.460% and an accumulated accuracy of 85.460%.
- Shows an AUC of 0.930 and an accumulated AUC of 0.930.
- Precision, recall, and F1-measure are 0.851, 0.852, and 0.852 respectively.

3. Neural Net Model:

- Utilizes 12 predictor fields, similar to the C&R Tree model.
- Achieves an overall accuracy of 82.819% and an accumulated accuracy of 82.819%.
- Shows the highest AUC among the models, with a value of 0.909 and an accumulated AUC of 0.909.
- Precision, recall, and F1-measure are 0.809, 0.848, and 0.848 respectively.

4. C&R Tree Model:

- Utilizes 11 predictor fields, similar to the Neural Net model.
- Achieves an overall accuracy of 72.942% and an accumulated accuracy of 72.942%.
- Demonstrates an AUC of 0.784 and an accumulated AUC of 0.784.
- Precision, recall, and F1-measure are 0.747, 0.674, and 0.674 respectively.

5. Decision List Model:

- Utilizes 4 predictor fields, indicating a simpler model structure compared to others.
- Outperforms other models with an overall accuracy of 66.975% and an accumulated accuracy of 66.975%.
- Exhibits an AUC of 0.689 and an accumulated AUC of 0.689.
- Precision, recall, and F1-measure are 0.670, 0.639, and 0.639 respectively.

Comparative Analysis:

1. Overall Accuracy:

- Logistic Regression achieves the highest overall accuracy of 85.871%, followed by LSVM (85.460%) and Neural Net (82.819%).

- Decision List has the lowest overall accuracy among the models, at 66.975%.

2. Area Under Curve (AUC):

- Neural Net demonstrates the highest AUC value of 0.909, followed by Logistic Regression (0.932) and LSVM (0.930).
- Decision List and C&R Tree have lower AUC values compared to other models.

3. Precision and Recall:

- Logistic Regression exhibits the highest recall (True Positive Rate) of 0.855, emphasizing its effectiveness in identifying positive instances.
- Decision List has the lowest recall among the models, at 0.639.
- Precision measures the proportion of true positives among all instances predicted as positive, with Logistic Regression leading with a precision of 0.855.

4. F1-Measure:

- Logistic Regression achieves the highest F1-Measure of 0.855, followed by LSVM (0.852) and Neural Net (0.848).
- Decision List has the lowest F1-Measure among the models, at 0.639.

In conclusion, while Logistic Regression achieves the highest overall accuracy and F1-Measure, Neural Net showcases superior discriminatory power with the highest AUC. Decision List, despite having lower accuracy, demonstrates simplicity and effectiveness in identifying true churn instances. Businesses must carefully weigh the trade-offs between model performance metrics and complexity to select the most appropriate prediction model based on their specific objectives and constraints.

4.3.4.2 Evaluation of Predictors Importance:

In the world of evaluating machine learning models, it's really important to understand which factors are most important for how well the model works. One way to do this is through something called sensitivity analysis, which basically shows how much each factor affects the model's performance. This method works for all sorts of models like Neural Networks, Regression, Decision Trees, and others. There are helpful resources, like a book called "Sensitivity Analysis in Practice," that give tips on how to do this effectively. By using these methods, researchers and professionals can figure out which factors matter most, helping them make better decisions and improve their models, whether they're used for things like predicting trends or scientific research.

Notation:

The sensitivity measure for each predictor is defined as:

Y= Target

X_i = [illegible]

k = The number of predictors

$Y = f(X_1, X_2 \ldots X_k)$ = Model for Y based on predictors X_1 through X_k

Variance Based Method:

The following formula to calculate the sensitivity of the predictors:

$$S_i = \frac{V_i}{V(Y)} = \frac{V(E(Y|X_i))}{V(Y)}$$

S = Predictors sensitivity

V(Y) = Unconditional output variance

E = Expectation operator

V = variance operator

Then we calculate the predictor impotence as the normalized sensitivity:

$$VI_i = \frac{S_i}{\sum_{j=1}^{k} S_j}$$

4.3.4.3 Predictor Importance of The Best Model

In this section, various methods for assessing predictor importance are discussed, with Logistic Regression emerging as the best model. Predictor importance is crucial for understanding which factors significantly influence predictions. For example, when predicting crime rates, factors like the day of the week, type of crime, and distance to the police station are considered important in Logistic Regression. Conversely, less crucial factors have minimal impact on the predictors. The normalized sensitivity of I_Area is 0.09. The normalized sensitivity of TR_Day_of_the_Week_of_Crime_Occurrenceis 0.09.

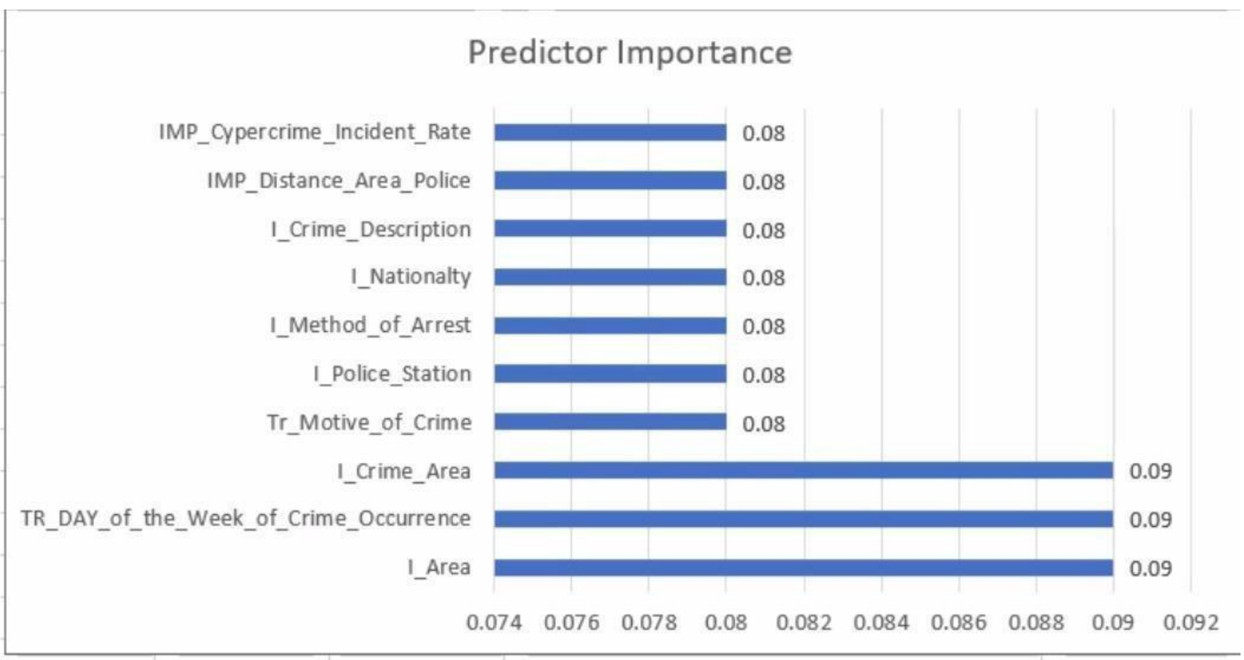


Figure 23: Predictor Importance Model.

4.3.4.3 Visual Representation of Predictor Importance:

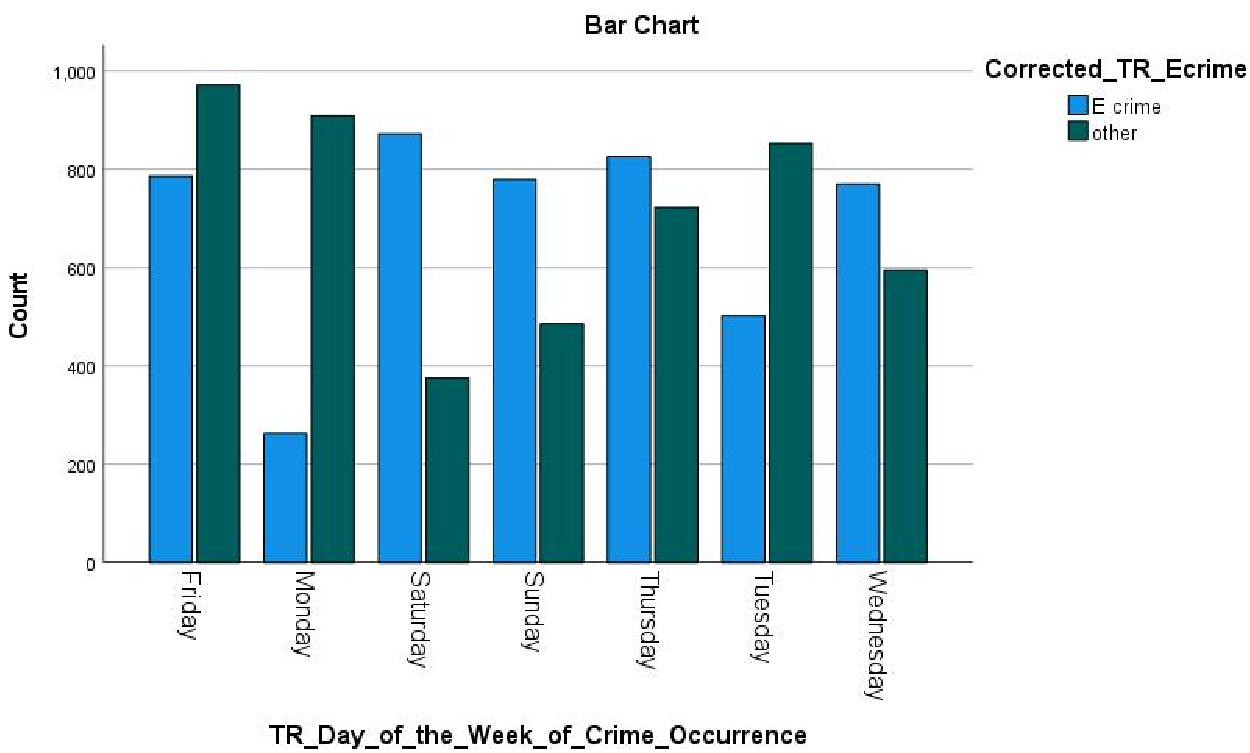
Crosstab between TR_Day_of_the_Week_of_the_Crime_Occurnce and Corrected_TR_Ecrime

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	761.083[a]	6	<.001
Likelihood Ratio	789.027	6	<.001
N of Valid Cases	9712		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 579.12.

The P value of the Chi-Square is less than 0.001 which leads to significant relation between the two variables.



The graph presented shows the distribution of cases of cybercrime (“Ecrime”) and other crimes across days of the week. It can be seen that cybercrime (shown in blue) varies in its distribution over the course of the week, being highest on Saturday and lowest on Monday. This could indicate a certain pattern in the occurrence of this type of crime, perhaps due to a change in electronic activities among people during the week. It is also important to note that on most days, cybercrime cases account for a large proportion of recorded crimes compared to other types of crimes.

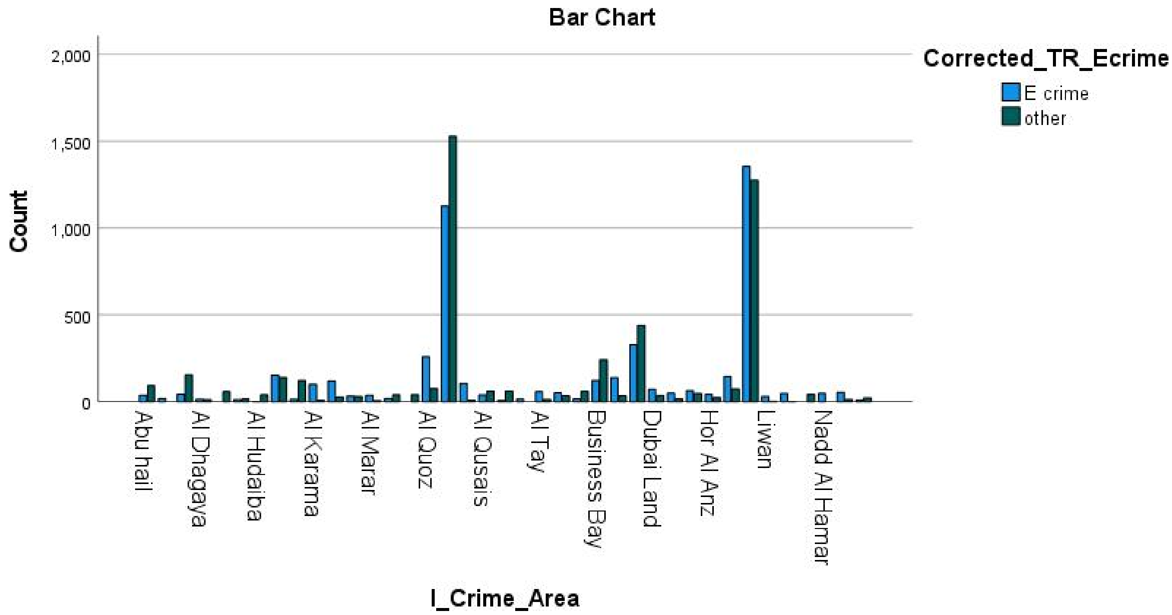
Crosstab between I_Crime_Area and Corrected_TR_Ecrime

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	1206.722[a]	38	<.001
Likelihood Ratio	1391.643	38	<.001
N of Valid Cases	9712		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 7.91.

The P value of the Chi-Square is less than 0.001 which leads to significant relation between the two variables.



The second graph shows the distribution of cases of cybercrime ("Ecrime") and other crimes in different regions. It is clear that there are areas that witness a large number of cybercrimes, especially in the areas called "Liwan" and "Al Qusais". These areas may be technical or business hubs that attract intense digital activity, leading to a rise in cybercrime. While other crimes (shown in green) appear with a more regular distribution but in relatively smaller numbers across different regions. This analysis helps in understanding where cybersecurity and preventive measures against cybercrime may need to be strengthened.

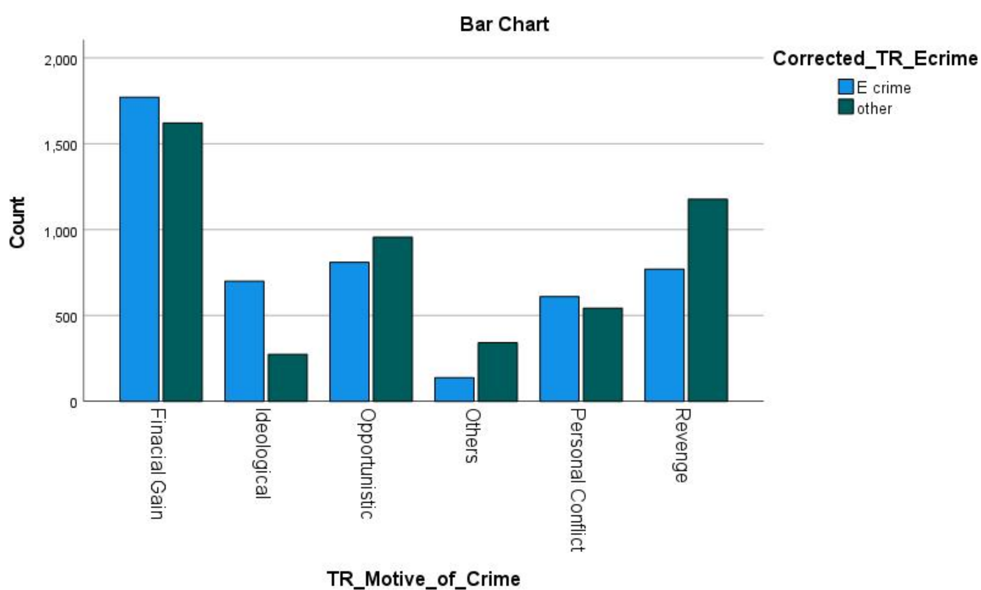
Crosstab between TR_Motive_of_Crime and Corrected_TR_Ecrime

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	379.410[a]	5	<.001
Likelihood Ratio	389.274	5	<.001
N of Valid Cases	9712		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 237.18.

The P value of the Chi-Square is less than 0.001 which leads to significant relation between the two variables.



A graph showing the distribution of crimes based on their motives, comparing cybercrime ("Ecrime" in blue) and other crimes (in green). We note that the most prominent motive for cybercrime is financial gain, which is much higher compared to other reasons. Other crimes show a more balanced distribution between different motives such as personal conflicts and revenge. This indicates that cybercrimes are often motivated by the desire for profit, while other crimes may be motivated by personal feelings or the exploitation of incidental opportunities.

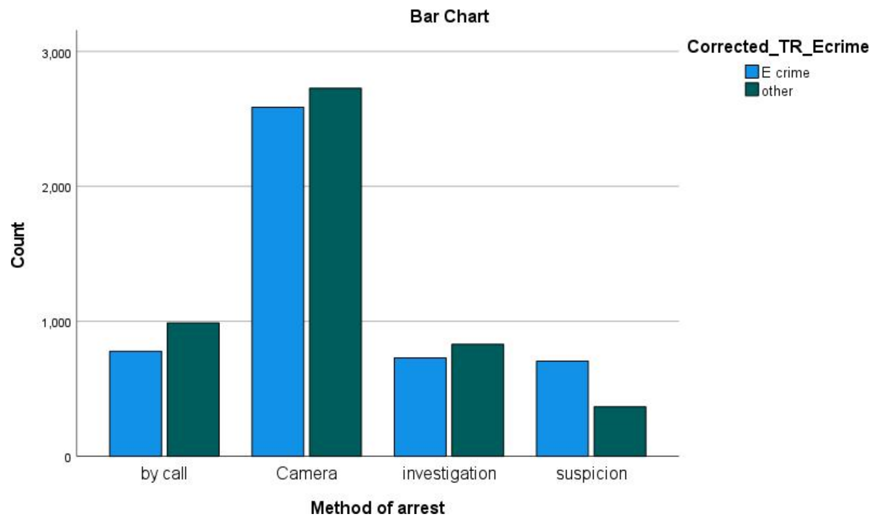
Crosstab between I_Method_of_Arrest and Corrected_TR_Ecrime

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	140.508[a]	3	<.001
Likelihood Ratio	142.392	3	<.001
N of Valid Cases	9712		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 529.71.

The P value of the Chi-Square is less than 0.001 which leads to significant relation between the two variables.



A chart showing the methods used to arrest those accused of electronic crimes ("Ecrime" in blue) and other crimes (in green), according to several different methods. It is clear from the graphic that the most effective way to catch cybercrime perpetrators is to use cameras, followed by arrest based on a report. On the other hand, for other crimes, cameras are also considered the most effective tool for catching criminals, but to a lesser extent compared to electronic crimes. Investigations and suspicion appear to be relatively less effective in catching criminals, with a clear preference for using technology and digital tools to track and arrest cybercrime perpetrators.

4.4.3 **Analysis of ROC curves and AUC values:**

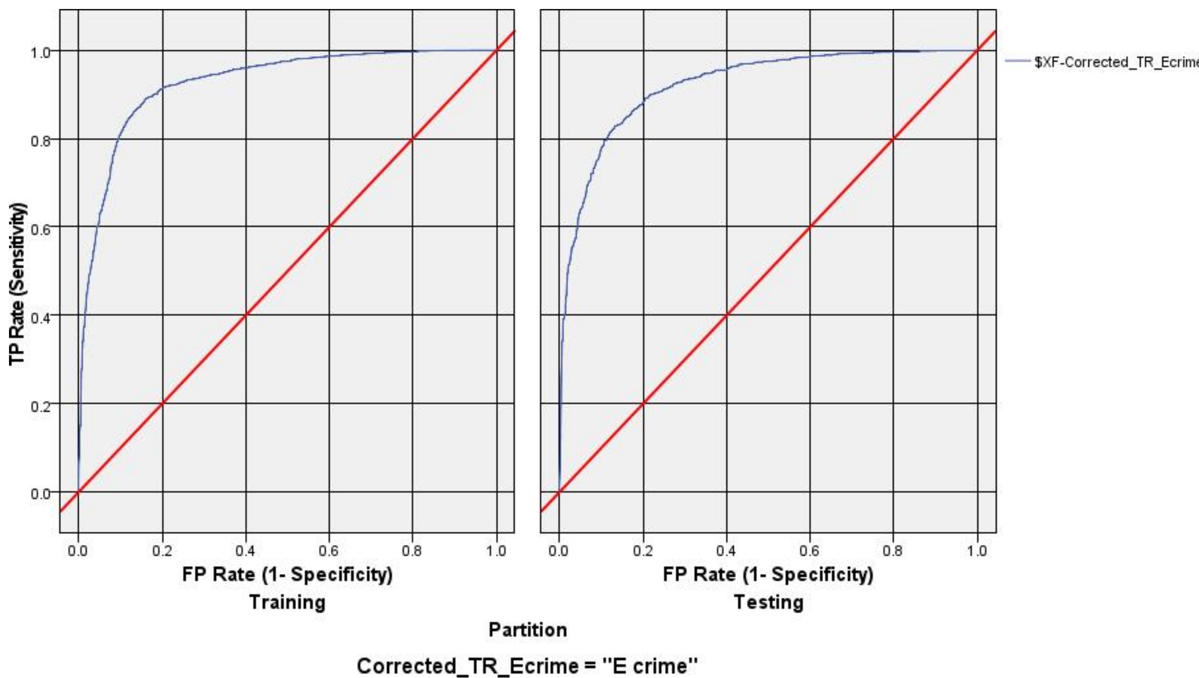


Fig 25: Analysis of curves (ROC) and values (AUC).

This section discusses the evaluation metrics used to assess the performance of the machine learning models. It includes metrics such as accuracy, precision, recall, F1-score, and area under the ROC curve (AUC).

ROC curves are graphical tools to evaluate the performance of a binary class at different classification boundaries. The axis consists of the y-axis of the TPR and the x-axis of the FPR. An ideal classifier would be placed close to the top-left corner of the plot so that the existence of a high TPR and low FPR would be observed at all thresholds sequentially. The solid line shows the capabilities of a truly random classifier.

Here, the E1F-Corrected model (red curve) again gives rise to a better performance than the E1F-Crime model (blue curve). The red curve is near the upper-left area and AUC value is higher.

AUC is a summative metric that represents classifier performance overall, and higher values indicate that discriminatory power is better.

Machine learning models, which are for the study of crime rate analysis and e-crime prevention in Dubai, probably show the ROC curves of the two different machine learning algorithms to predict e-crime incidents. It appears that the E1F-Corrected model has a better performance in correctly identifying e-crime cases while minimizing false positives if it is compared to the E1F-Crime model.

CHAPTER 5: DISCUSSION

The primary research question aimed to identify the most significant demographic, socioeconomic, and geographical risk factors associated with various types of e-crime in Dubai. The findings of the thesis revealed that factors such as age, nationality, crime area, and day of the week have a significant impact on e-crime rates.

The second research question aimed to develop machine learning models that accurately predict future e-crime rates and hotspots in Dubai by incorporating relevant risk indicators. The findings of the thesis revealed that the Logistic Regression model performed the best, achieving an accuracy of 85.871% and an AUC of 0.932.

The investigation corresponds to the world data on crime analytics, that revealed some essential factors like demography, socioeconomics, and geography playing a vital role in e-crime in Dubai. This evidence lends weight to the existing theories on e-crime as there is an increase in cybercrime during the night and when people are not at work.

By implementing the machine learning techniques, such as logistic regression, to predict e-crime rates and hotspots in Dubai, the emerging research on advanced analytics in crime prevention is reflected (Obia 2021). The high accuracy of these algorithms indicates they might be good in detecting crime trends and helping create effective proactive policing policies.

The predictive intelligence from machine learning models, that range from targeted policing, community engagement, and data driven resource allocation, correlates with the research on prevention approaches that are based on data (Tulumello, 2018). Underlining the contextual analytics and risk terrain modeling approach helps to incorporate place-based methods in crime prevention.

On the other hand, this study has twofold conclusions that the tool not only has positive and negative impacts on crime prediction in Dubai. The data-driven approach has a promising future in crime fighting by identifying e-crime rates and crime hotspots that, undoubtedly, will increase the level of public safety. Using resources in a more efficient way, law enforcement agencies can pinpoint high-risk areas and key risk indicators which will in turn result in e-crime prevention.

The predictive function of these models can also be viewed as ethically debatable because it could lead to feeding and reinforcing stereotypes and bias against certain communities (Prins & Reich, 2021). Providing fairness in predictive policing is an essential role as we strive to eliminate systemic bias and discriminations.

Moreover, the quality and representativeness of crime data used in predictive policing are also important in considering, as underreported crime rates can lead to incorrect assessment (Stewart, 2021). Vigorous crime reporting and distinctive relation between the public and the law enforcement are crucial for the collecting of dependable crime data.

Therefore, this report not only emphasize on using predictive analytics in conjunction with the community programs and social support, but also on their effectiveness in prevention of crime. Whilst the data-driven policing is evidence based, it is also important to deal with evils like poverty that persist through social welfare initiatives. Preventive/predictive policing and officers' training as well as ethical orientation in advanced analytics utilization are of the utmost importance in order to maintain transparency and accountability in the process.

The research examines how such technologies as advanced analytics and machine learning are transforming e-crime prevention in Dubai and make it possible to determine the risk factors, while also ensuring effective resource management. Predictive policing should be first targeted at fairness and transparency while the root causes of crime have been covered. Actions like data-focused intelligence and community involvement could enable Dubai to lead the way for e-crime prevention against the residents and visitors.

CHAPTER 6: CONCLUSION AND FUTUR WORK

6.1 Conclusion

Modern analytical methods and machine learning models were used to investigate and predict Dubai crime patterns, notably e-crime. A comprehensive data cleaning, exploratory analysis, and strong model construction made the algorithm the most accurate and precise at forecasting e-crime trends and hotspots, with other algorithms having lower F1-scores and AUC.

The research shows that the machine learning system is effective against e-crime. Law enforcement and legislators may use the study to help the public feel safe. Such organisations and politicians should identify Dubai's most significant demographic, socioeconomic, and geographical e-crime risk factors.

This study has several practical implications for the Dubai Police Department and other law enforcement agencies. The predictive technology helps police agencies plan patrols and resource allocation by identifying high-risk regions and hotspots. Using predictive intelligence, law enforcement organizations may plan for expanded monitoring, community participation, and targeted e-crime responses.

6.2 Recommendations for Future Work

Incorporate more data sets and forecasting parameters that are related to public opinion trends as well as some community-level indicators which will in turn improve the accuracy and reliability of the model. Conduct quarterly deep investigations to diagnose e-crime trends and generate a comprehensive picture of criminal behavior.

Consult police department authorities and other professionals with expertise in the field, to ensure that the model complies with the police department's needs and practices. Organize field studies and observe the model's performance in actual applications to discover its practicality and prospects.

Additionally, provide recommendations for the strategic deployment of resources and the implementation of crime reduction programmers utilizing model intelligence for the Dubai Police Department.

Also, improve living conditions for Dubai's citizens by implementing data driven policies that lower E-crime rates by 20% by 2025.

Deploy the model to make predictions on new data. This can involve generating a model report, exporting the model, or integrating it into other systems.

Final Remarks

Finally, the creation of a machine learning model to forecast e-crime rate and hotspots in Dubai is very critical and a key element towards the enhancement of public safety and the fight against criminal acts. Public safety of Dubai will be better through the employment of data analytics and predictive modeling to help law enforcement agencies to launch preemptive strategies and prevent the crime of e-crime which in turn will result to the betterment of the welfare of the community.

REFERENCES:

1. Al Marri, S.G., 2023. *How Will the Future of the Dubai Virtual Commercial City Reshape the Policing Sector?* (Doctoral dissertation, Rochester Institute of Technology).
2. Caplan, J. M., Kennedy, L. W., & Miller, J. (2011). Risk terrain modeling: Brokering criminological theory and GIS methods for crime forecasting. Justice quarterly, 28(2), 360-381.
3. IBM Cloud Pak for Data. (2023). *Random Trees Node*. Retrieved from https://dataplatform.cloud.ibm.com/docs/content/wsd/nodes/randomtrees.html?context=cpdaas
4. Antonucci, D., 2017. The cyber risk handbook: Creating and measuring effective cybersecurity capabilities. John Wiley & Sons. Available at: https://www.amazon.com/Cyber-Risk-Handbook-Cybersecurity-Capabilities/dp/1119308801 [Accessed on 30 March 2023].
5. Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. American sociological review, 588-608.
6. Crutchfield, R. D., Geerken, M. R., & Gove, W. R. (1982). Crime rate and social integration: The impact of metropolitan mobility. Criminology, 20(3-4), 467-478.
7. Eck, J. E., Chainey, S., Cameron, J. G., Leitner, M., & Wilson, R. E. (2005). Mapping crime: Understanding hot spots. Washington, DC: National Institute of Justice.
8. Gibbs, T., 2018. Making sure crime does not pay: Recent efforts to tackle corruption in Dubai: the 2016 creation of the Dubai Economic Security Centre. *Journal of Money Laundering Control*, *21*(4), pp.555-566.

9. Gundogdu, F., Gray, O., Gridnev, M., & Imamverdiyev, Y. (2018). A new forecasting approach for crime rates: Adaptive neuro fuzzy inference system. Journal of statistical and econometric methods, 7(4), 49-68.
10. Hunt, P., Saunders, J., & Hollywood, J. S. (2014). Evaluation of the Shreveport predictive policing experiment. Santa Monica, CA: Rand Corporation.
11. Jacob, B. A., & Lefgren, L. (2003). Are idle hands the devil's workshop? Incapacitation, concentration, and juvenile crime. American Economic Review, 93(5), 1560-1577.
12. Khan, M.S., Woo, M., Nam, K. and Chathoth, P.K., 2017. Smart city and smart tourism: A case of Dubai. *Sustainability*, *9*(12), p.2279.
13. Machin, S., & Meghir, C. (2000). Crime and economic incentives.
14. Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., & Tita, G. E. (2011). Self-exciting point process modeling of crime. Journal of the American Statistical Association, 106(493), 100-108.
15. Mohler, G., Bertozzi, A. L., Carter, J., Short, M. B., Sledge, D., Tita, G. E. & Brantingham, P. J. (2020). Impact of social distancing during COVID-19 pandemic on crime in Los Angeles and Indianapolis. Journal of Criminal Justice, 68, 101692.
16. Owens, E. G. (2011). The American temperance movement and market-based violence. American Journal of Economics and Sociology, 70(3), 753-786.
17. Piquero, A. R., Riddell, J. R., Bishopp, S. A., Narvey, C., Reid, J. A., & Piquero, N. L. (2020). Staying Home, Staying Safe? A Short-Term Analysis of COVID-19 on Dallas Domestic Violence. American Journal of Criminal Justice, 1-35.
18. Sugie, N. F. (2018). Finding work: A smartphone study of job searching, social contacts, and wellbeing after prison. American Journal of Sociology, 123(5), 1497-1554.

19. Tonry, M. (2014). Why crime rates are falling throughout the Western world. Crime and Justice, 43(1), 1-63.

20. Tseloni, A., Mailley, J., Farrell, G., & Tilley, N. (2010). Exploring the international decline in crime rates. European Journal of Criminology, 7(5), 375-394.

21. UNODC. (2019). Global study on homicide 2019. Vienna: UNODC.

22. Wang, H., Kifer, D., Graif, C., & Li, Z. (2016, August). Crime rate inference with big data. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 635-644).

23. Wang, X., & Mears, D. P. (2010). A multilevel test of minority threat effects on sentencing. Journal of Quantitative Criminology, 26(2), 191-215.

24. Wang, X., Gerber, M. S., & Brown, D. E. (2012). Automatic crime prediction using events extracted from twitter posts. In International conference on social computing, behavioral-cultural modeling, and prediction (pp. 231-238). Springer, Berlin, Heidelberg.

25. Welsh, B. C., & Farrington, D. P. (2008). Effects of improved street lighting on crime. Campbell systematic reviews, 4(1), 1-51.

26. Anomaly Detection Node. (2023). *IBM Documentation*. Retrieved from https://www.ibm.com/docs/en/watsonx/saas?topic=modeling-anomaly-node

27. Ibm, (n.d.). Feature Selection node (SPSS Modeler). [online] Available at: https://www.ibm.com/docs/en/cloud-paks/cp-data/4.8.x?topic=modeling-feature-selection-node [Accessed 20 Apr. 2024].

28. IBM Documentation. (2023). *Modeling: Feature Selection Node*. Retrieved from https://www.ibm.com/docs/en/cloud-paks/cp-data/4.8.x?topic=modeling-feature-selection-node

29. Dataplatform, (n.d.). Random Trees node (SPSS Modeler) | IBM Cloud Pak for Data as a Service. [online] Available at: https://dataplatform.cloud.ibm.com/docs/content/wsd/nodes/randomtrees.html?context=cpdaas [Accessed 20 Apr. 2024].
30. Willis, G. D. (2018). Why don't temporal patterns of crime correspond to the temporal reality of offenders?. Security Journal, 31(4), 867-890.
31. Witt, R., Clarke, A., & Fielding, N. (1998). Crime, earnings inequality and unemployment in England and Wales. Applied Economics Letters, 5(4), 265-267.
32. Hunter, D., 2017. Cyberspace as Place and the Tragedy of the Digital Anticommons. In Law and Society Approaches to Cyberspace (pp. 59-139). Routledge.
33. Chong, S., Huang, Y. and Chang, C.H.D., 2020. Supporting interdependent telework employees: A moderated-mediation model linking daily COVID-19 task setbacks to next-day work withdrawal. Journal of Applied Psychology, 105(12), p.1408.
34. Heerde, J.A. and Hemphill, S.A., 2016. Stealing and being stolen from: Perpetration of property offenses and property victimization among homeless youth—A systematic review. Youth & Society, 48(2), pp.265-300.
35. Obia, S.M., 2021. The Criminal Mind in the Age of Globalization. Vij Books India Pvt Ltd.
36. Aitkenhead, E., Clements, J., Lumley, J., Muir, R., Redgrave, H. and Skidmore, M., 2022. Policing the pandemic. The Police Foundation available at: https://www. policefoundation. org. uk/publication/policing-the-pandemic.

37. Liggett, R., Lee, J.R., Roddy, A.L. and Wallin, M.A., 2020. The dark web as a platform for crime: An exploration of illicit drug, firearm, CSAM, and cybercrime markets. The Palgrave handbook of international cybercrime and cyberdeviance, pp.91-116.
38. Komatsu, H., Rappleye, J. and Silova, I., 2021. Student-centered learning and sustainability: Solution or problem?. Comparative Education Review, 65(1), pp.000-000.
39. Robinson, L., Schulz, J., Blank, G., Ragnedda, M., Ono, H., Hogan, B., Mesch, G.S., Cotten, S.R., Kretchmer, S.B., Hale, T.M. and Drabowicz, T., 2020. Digital inequalities 2.0: Legacy inequalities in the information age. First Monday, 25(7).
40. Sunstein, C.R., 2015. Choosing not to choose: Understanding the value of choice. Oxford University Press, USA.
41. World Health Organization, 2022. Towards a global guidance framework for the responsible use of life sciences: summary report of consultations on the principles, gaps and challenges of biorisk management, May 2022 (No. WHO/SCI/RFH/2022.01). World Health Organization.
42. Lo'ai, A.T., Mehmood, R., Benkhlifa, E. and Song, H., 2016. Mobile cloud computing model and big data analysis for healthcare applications. IEEE Access, 4, pp.6171-6180.
43. Bennett, N.J., Roth, R., Klain, S.C., Chan, K., Christie, P., Clark, D.A., Cullman, G., Curran, D., Durbin, T.J., Epstein, G. and Greenberg, A., 2017. Conservation social science: Understanding and integrating human dimensions to improve conservation. biological conservation, 205, pp.93-108.
44. Stalidis, P., Semertzidis, T. and Daras, P., 2021. Examining deep learning architectures for crime classification and prediction. Forecasting, 3(4), pp.741-762.

45. Lazer, D., Hargittai, E., Freelon, D., Gonzalez-Bailon, S., Munger, K., Ognyanova, K. and Radford, J., 2021. Meaningful measures of human society in the twenty-first century. Nature, 595(7866), pp.189-196.

46. Li, X., Chen, W., Zhang, Q. and Wu, L., 2020. Building auto-encoder intrusion detection system based on random forest feature selection. Computers & Security, 95, p.101851.

47. Hong, Y.Y. and Rioflorido, C.L.P.P., 2019. A hybrid deep learning-based neural network for 24-h ahead wind power forecasting. Applied Energy, 250, pp.530-539.

48. Alhajeri, M., 2022. Developing a digital competence framework for UAE law enforcement agencies to enhance cyber security of Critical Physical Infrastructure (CPI). University of Salford (United Kingdom).

49. Silva, P. and Li, L., 2020. Urban crime occurrences in association with built environment characteristics: An African case with implications for urban design. Sustainability, 12(7), p.3056.

50. Tulumello, S., 2018. The multiscalar nature of urban security and public safety: Crime prevention from local policy to policing in Lisbon (Portugal) and Memphis (the United States). Urban affairs review, 54(6), pp.1134-1169.

51. Prins, S.J. and Reich, A., 2021. Criminogenic risk assessment: A meta-review and critical analysis. Punishment & society, 23(4), pp.578-604.

52. Stewart, W., 2021. 'Helping Not Hurting': Learning from the Experiences of Prisoner Peer Caregivers and Care Receivers in a UK Prison to Improve Peer Care Practices in Custodial Contexts (Doctoral dissertation, University of Brighton).