# Crime Analysis

**Karl Attard**

Supervisor: Dr Joel Azzopardi

May, 2022

*Submitted in partial fulfilment of the requirements for the degree of Bachelors of Science in Information Technology (Honours) (Artificial Intelligence).*

L-Università ta' Malta
Faculty of Information &
Communication Technology

L-Università ta' Malta

## FACULTY/~~INSTITUTE/CENTRE/SCHOOL~~ ICT

## DECLARATIONS BY UNDERGRADUATE STUDENTS

Student's I.D. /Code 203501L

Student's Name & Surname Karl Attard

Course B.Sc in Information Technology (Hons.) (Artificial Intelligence)

Title of ~~Long Essay~~/Dissertation

Crime Analysis

Word Count 12,934

**(a) Authenticity of ~~Long Essay~~/Dissertation**

I hereby declare that I am the legitimate author of this ~~Long Essay~~/Dissertation and that it is my original work.

No portion of this work has been submitted in support of an application for another degree or qualification of this or any other university or institution of higher education.

I hold the University of Malta harmless against any third party claims with regard to copyright violation, breach of confidentiality, defamation and any other third party right infringement.

**(b) Research Code of Practice and Ethics Review Procedures**

I declare that I have abided by the University's Research Ethics Review Procedures.

K. Attard
Signature of Student

KARL ATTARD
Name of Student (in Caps)

27-05-2022
Date

# Acknowledgements

First and foremost, I would like to acknowledge my supervisor, Dr Joel Azzopardi who made this work possible. His continuous help and guidance carried me throughout every aspect of this dissertation, and for that I am immensely grateful.
I would also like to give my special thanks to my family and friends, particularly to my girlfriend Carla Scicluna who have listened to my concerns on a near-daily basis, in relation to this project. They have supported and encouraged me to keep going on through difficult moments. Without them, this would not have been possible.
Finally, I would like to individually mention my classmates, Malcolm, Mikea, Benjamin and Paolo, whom we have shared this journey together through thick and thin. Thank you everyone for sharing this ride together.

# Abstract

This dissertation focuses on using data mining techniques within the crime analysis domain, which is becoming very popular since many datasets are being publicly available to researchers. We focus on three main objectives; predicting the type of crime, predicting the future crime rate and the investigation of how much data is required to train our models. Two distinct American public datasets are chosen for this study, one focusing on the city of San Francisco (SF) and the other on New York City (NYC).

Our first objective concerns crime type prediction, where we implemented several models, ranging from classical Machine Learning (ML) to Deep Learning (DL) methods, that are able to classify the crime category based on the inputted data, mainly being, the time of crime occurrence, its location and how the crime was resolved. The results obtained are then directly compared to [31], where our main goal was to replicate and improve on their findings.

The second objective covers crime rate prediction. We implemented both ML and DL regression-based models, meaning that we predict a continuous value (the expected future crimes), as opposed to the classification-based approach above (which predicts a category). Additionally, we integrated external data (population, unemployment rate and median income data) to help us create more accurate models. Similar to the first objective, we replicate and improve on the results obtained by [9].

In our last objective, we investigated how much data the models require to learn from and yet still manage to produce decent results. Hence, for each prior objective, we diminished the original size of the training set whilst keeping the test set unchanged to check how much data is needed for all the models developed to retain their optimal performance.

Our classification and regression models managed to produce better results than [31] and [9] respectively. We inferred that the DL classification model created was not more effective than the other ML models built. We also observed that the unemployment rate feature is the most effective external data when predicting the future crime rate. Finally, diminishing the training sets of Objectives 1 and 2 respectively by approximately half allowed the models to retain a decent performance in their predictions.

# Contents

# List of Figures

# List of Tables

# List of Abbreviations

# 1 Introduction

In this chapter, we start off by establishing our research problem and solution. Then, we briefly explain what this problem entails, and how we solve it. Moving on, we highlight the motivation why such a domain has been chosen and why our research problem is a non-trivial task. Next, we provide a summary of our findings in this study together with formally defining our aims and objectives. Finally, we conclude by outlining the structure of this document dissertation.

## 1.1 Problem Definition and Motivation

Crimes in urban spaces are a concern for every person's own welfare. Crime has become a socio-economic problem that is continuously affecting the quality of life and economic growth [5]. Law enforcement agencies continuously strive to understand the nature of crimes which are occurring in their responsible city with the aim of identifying who is committing them; where and when they are most likely to occur; what is the likelihood that a certain crime takes place; and how much will the rate vary along the years.

Nowadays, there are numerous crime datasets which are being publicly available to researches. Criminology is also one of the most important fields when applying data mining techniques, as these can produce effective results [38]. Providing accurate and reliable crime predictions help law enforcement and other entities to effectively prevent crime from happening again as well as to handle them effectively when they occur [23].

Hence, in our research we focus on tackling crime prediction from an Artificial Intelligence (AI) point-of-view, by using both classification and regression techniques to predict the type of crime as well as the future crime rate respectively. Furthermore, we also study to see how much data is required by our models to train on such that their performance and effectiveness remain on acceptable levels.

## 1.2 Why the Problem is Non-Trivial

This problem comes up with some challenges which can be foreseen in advance. First of all, finding the right dataset/s which is/are suitable for our use case is an element of concern. Various countries collect data differently since crime information is considered as a sensitive data, and this collection differs between other countries. Additionally, not all countries publish their crime data, and thus, makes it more difficult for us to find the most suitable data. For instance, up to our knowledge, no such dataset exists for the Maltese islands within the public domain. It is also vital to ensure that any available datasets are not only comprehensive but are also ideal for our research topic objectives.

Another notable problem is the fact that datasets can be very large in size. Hence, dealing with these datasets can be challenging at times since creating both classical machine learning models as well as more advanced deep learning models can be computationally expensive to execute on a normal computer machine.

## 1.3 Proposed Solution & Summary of Results

As part of this study, we aim to replicate and improve on the findings of two research papers that perform crime type prediction [31] and crime rate prediction [9] respectively. We utilise the datasets that these research papers use, San Francisco (SF) crime dataset[1] and New York City (NYC) crime data[2] respectively. Furthermore, the same pre-processing techniques, as those described by Sardana et al. [31] and Catlett et al. [9] have been applied on both datasets to ensure a fair experimentation when it comes to replicating their findings, but then proceed to improve the performances of their results, achieved by balancing the SF dataset (grouping some feature's values into categories, for instance, converting the month column to season) as well as integrating additional data to the NYC crime dataset (including, population and unemployment rate data), which is further described in Section 3.1 below.

In both objectives of crime type prediction and crime rate prediction, we develop traditional machine learning models and deep learning models, including decision trees and multi-layer perceptron neural networks respectively. Parameter tuning optimisation was also performed manually on every model implemented, meaning that we executed each model a number of times with different permutations of parameters, and choosing that combination which produced the most accurate predictions. In this research, when it

[1] https://www.kaggle.com/competitions/sf-crime/data; [Last Accessed: 6th April 2022]

[2] https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i?ref=hackernoon.com; [Last Accessed: 2nd May 2022]

comes to crime type prediction, we decided to retain the *resolution* feature (how the crime was resolved, for example, if an arrest was made or not) in our dataset, as opposed to what Sardana et al. did in [31]. Additionally, we predict the type of crime when provided with temporal, spatial and historical crime information. Identifying the type of crime or law category as early as possible are found useful by police tasks forces when planning law enforcement and crime prevention duties [31].

On the other hand, for crime rate prediction, we compare different AI models with each other and analyse how the inclusion of new features (not part of the original dataset), including *population*, *median income* and *unemployment rate*, reflect on the performances of these created models. Hence, this research does not only directly compares our results with those in [9], but also analyses which is the most important integrated feature that lead to the highest improved model performance. These regression-based models predict the expected number of crimes for the years of 2014, 2015 and 2016 respectively (the test set), given a training set ranging from the periods of 2006 up to 2013.

For our last objective, that of investigating how much data our classification and regression models require to train on, we diminish the size of this training set and re-run our Machine Learning (ML) and Deep Learning (DL) models. After running these models, we tabulate the obtained results and see how the models fared. This process is repeated until we find the minimum training size required such that the performances of the models remain effective in predicting the crime type and forecasting the number of future crimes.

All classification-based models developed are evaluated using *k*-cross validation method. The evaluation metrics used are the weighted F1 score for the classification task and Mean Average Percentage Error (MAPE) for the regression task, similar to [31] and [9] respectively. For crime type prediction objective, we have found that when including the *resolution* data as part of the training features, the models produce more accurate results. Furthermore, we observed that the Multi-Layer Perceptron (MLP) DL model implemented was not more effective than the other ML models built, such as the Logistical Regression Classifier. The highest f1-score obtained after performing the *k*-cross validation process was 0.71, which was produced by the Logistical Regression Classifier, Decision Tree Classifier as well as the MLP neural network.

On the other hand, for crime rate prediction, we have inferred that all our classical ML models, including the Linear Regression, produced the best results when having crime data only as training features but then started predicting with lower accuracy as soon as additional features, such as *population* data, was included as training features. To the contrary, our MLP DL model performed better when adding external data as features, and hence, has managed to produce the best overall MAPE score of 0.74 when using crime and unemployment rate data. Thus, the *unemployment rate* data feature is the most effective external feature that is positively impacting the predictions. In both crime type prediction as well as crime rate prediction, we managed to improve on the performances of the results obtained by [31] and [9] respectively. Furthermore, with regards to the investigation of how much training data our built models require, we inferred that 40% rows of the whole dataset is sufficient (equivalent to 351219 records of crime data) for the classification models to train on whilst training data from 2010 to 2013 is enough for the regressor models. Finally, the highest f1-score obtained is 70% for the crime type prediction part whilst an MAPE value of 0.79% for the crime rate prediction part, as part of this investigation objective. These results obtained will be further discussed in detail in Chapter 4 below.

## 1.4 Aims and Objectives

The primary aim for this research is to employ different data mining techniques on urbanised city's open crime datasets to predict the type of crime and forecast the expected number of future crimes, using classification and regression methods on SF and NYC datasets respectively. We also aim to investigate the amount of data our classification and regression models built require to train on, such that they remain effective in their predictions.

This research aim is formalised through the following objectives:

- **Objective 1**: Identifying the best classification models, ranging from classical machine learning to deep learning methods, which best predict the type of crime given initial crime observations - namely temporal and spatial information, and whether an arrest was made or not.
- **Objective 2**: Finding the best regression models, also ranging from classical machine learning to deep learning methods, in order to forecast the expected crime rate along future years.
- **Objective 3**: Investigating the amount of training data all implemented models require such that their performance is not hindered, and thus, retain their effectiveness.

## 1.5 Document Structure

The rest of this document is organised in the following manner. Chapter 2 below discusses in depth what other researches have carried out within the crime analysis domain, and how it served as an inspiration to conduct this research. Chapter 3 then highlights the methodology of how we intend to solve our objectives, delving more into what datasets have been used, how they were pre-processed and how we went about implementing our solution. This is then followed by Chapter 4, where we explain our evaluation strategy, present our results and compare them with those in literature. Finally, Chapter 5 concludes this document by providing a summary of our research findings, and some interesting future work which can be carried out.

# 2 Literature Review

From the background research conducted on other researchers' work, the application of different data mining techniques is vast within the crime analysis domain and can take many different forms. These include analysing criminal networks, calculating the likelihood that an individual becomes a victim, creating patrol routes as well as forecasting where a perpetrator is likely to commit a future crime based on their previous activity and their previous crimes [40]. However, as already highlighted in Section 1.4 above, this dissertation focuses on crime rate prediction and crime type prediction. Subsequently, relevant existing research have been identified after carrying out our literature review which has motivated and inspired us to conduct such a research.

The rest of this chapter is split into two sections; Section 2.1 below describes research papers which solve crime pattern discovery and profiling whilst Section 2.2 explains different techniques other research authors used to solve the crime prediction objective within the crime analysis domain. In the former section, an overview of common approaches are discussed, ranging from clustering algorithms to association rule mining methods. On the other hand, the latter section discusses both classification and regression approaches to solve that respective objective.

## 2.1 Crime Pattern Discovery and Profiling

Crime pattern discovery and profiling can be implemented in several ways. This objective is usually handled either using clustering techniques or else using association rule mining techniques. The former technique is more often used for profiling crimes, victims or criminals whilst the latter technique is mainly used for deducing frequent occurring crime patterns.

### 2.1.1 An Overview of Common Approaches

**k-Means, SOM & GSOM Clustering**

Clustering algorithms used in similar research range from simple k-Means clustering [15] to Self Organising Maps (SOMs) [1, 21], and Growing Self Organising Maps (GSOMs) [6]. Gera and Vohra uses k-Means clustering algorithms to create crime profiles for the city of Delhi, India by analysing the results from the clusters generated [15]. On the other hand, Adderley and Musgrove aimed to identify the typical profile of a sexual offender by applying a SOM that performs clustering [1]. Similarly, the authors in [21] also makes use of SOMs to cluster their data. However, their main objective is to profile crimes with criminals to unsolved crime incidents from the clustering results. A MLP neural network is included on top of the clustering as they believe that it will benefit in crime matching, i.e., classifying crime to criminals and vice-versa. The main reason why the above mentioned authors are using SOMs is because this implementation is useful in reducing high-dimensional data to a lower dimensional [25, 21]. Moreover, it creates a 2D topological map which can show links between neighbouring clusters.

Boo and Alahakoon uses GSOMs - a hierarchical clustering technique, an extension to a SOM [6]. Their main aim is to explore hidden patterns at different granularity of the data. Hence, a GSOM fits perfectly to their implementation as it allows them to highlight clusters with its respective details with higher granularity. Additionally, GSOMs are a great way to visualise your data and understand it well. In [6], an example is shown where they are showing the different types of firearms which they found in the clusters generated.

**Association Rule Mining using Apriori Algorithm**

Crime patterns can also be identified using association rule mining methods [26, 32]. These unsupervised learning algorithms are basically used to find frequently occurring relationships between hidden features in the dataset. Marzan et al. utilise the Apriori algorithm as an association rule mining technique to find patterns between crime type, occurrence time, weather, holiday, the crime location and the district [26]. This algorithm generates a set of rules according to some predefined minimum support and minimum confidence having the district variable as the consequent. The authors only output those generated rules having their lift value greater than 1, meaning there is a positive correlation. Hence, using this technique, Marzan et al. found hidden relationships which can be used by law enforcement to prevent crime according to the above rules. The same technique is used in [32]. Here, the Apriori algorithm is used the find the patterns and trends of crimes happening in specific locations. Hence, these rules generated for each location, it can help authorities in deducing that there is a probability that a crime might occur based on that localities previous trends.

## 2.2 Crime Prediction

Crime prediction is perhaps one of the most popular objectives which researchers aim to attain within crime analysis domain [20, 22, 2, 31, 26, 13, 34, 32]. This objective is often handled through supervised machine learning algorithms, and are applied into two forms, either as a classification task or as a regression task. A classification task is one which uses a model that classifies a discrete value, meaning that this model approximates a mapping function which is responsible to predict discrete output variables, which can be either labels or categories. An example includes Elluri et al. research which attempts to study whether adding weather data to existing crime data increases the performance of classification models, such as Support Vector Machine (SVM) and Deep Neural Network (DNN), which predicts the type of crime given crime data, weather data as well as victim's and suspect's data [13]. On the other hand, a regression task is one that utilises a model to predict some continuous value, as opposed to discrete values. This means that a regression model is a mapping function that maps input variables to a continuous-valued target variable, such as a quantity variable or a probability value. For instance, Marzan et al. used time-series forecasting methods, like Linear Regression and SMOreg, to predict the crime trends for the future days and weeks in each district of Manila's City [26]. Hence, this objective is divided into two separate sub-sections, which are discussed in-depth in Section 2.2.2 and 2.2.3 below.

### 2.2.1 Datasets Utilised in Research

In order to solve this objective, it is imperative to have data, which can either be collected manually [24, 26, 35] or publicly available on the internet [34, 33, 27, 31, 22, 13, 2, 3, 29, 37, 9, 30]. In this section, we will be highlighting the different types of datasets used from the read literature, for both classification and regression approaches.

Pradhan et al. in [27] used an open-source dataset consisting of crimes together with the temporal and spatial features, collected by the San Francisco Police Department (SFPD) and available through the Open Data initiative [12]. Furthermore, it is worth noting that Sardana et al. [31] used a subset of [12], which is found on Kaggle[1]. Babakura et al. gathered their data from the UCI machine learning repository website [28], a website con-

[1]https://www.kaggle.com/competitions/sf-crime/data; [Last Accessed: 6$^{th}$ April 2022]

taining publicly available datasets free to download. This data combines socio-economic data from the 1990 US Census, law enforcement data as well as crime data, and has many variables, such as the percent of population and the median family income data [3]. On the other hand, both Catlett et al. [9] and Elluri et al. [13] decided to use NYC crime data respectively. This dataset includes all misdemeanor, violation and felony crimes which were reported to the New York Police Department (NYPD) for the years of 2003 up until 2019[2]. Essentially, this dataset is an extension of [12] in terms of feature variables available since it also includes data related to the victim and suspect (age group, race, gender respectively). Moreover, Elluri et al. integrated weather data as well to this current NYC dataset.

Similarly, both Yuki et al. and Wang et al. also used a publicly available dataset for the city of Chicago[3] [37, 34], which has the same exact features as [12]. However, Wang et al. and Rumi et al. [29] extended their research by moving on from the traditional use of demographic and geographic data. Hence, they make use of Point-of-Interest (POI) and Taxi Flow data in addition to the aforementioned traditional data. POI data is data that provide venue information such as coordinates, category, popularity and reviews. Recent studies have highlighted that using such POIs data are useful to profile neighbourhood functions [36] which could have a correlation with the number of crimes within such region. Taxi flow data reflect how people commute in the city and subsequently, can highlight frequently locations travelled to which in turn can be correlated to the location being examined. The main reason why including these type of data is to improve the overall accuracy of the developed models.

As for data collected manually, Llaha collected data from law enforcement agencies, and their dataset (which is in CSV format) includes 100 examples of crimes having 6 distinct features, ranging from the crime place to the age of perpetrator [24]. Similarly, Marzan et al. collected their data manually from Manila Police District (MPD) Office, a law enforcement agency in Philippians [26]. Their dataset contains crime information data from 16 different districts of Manila City during the years of 2012 up to 2016. Furthermore, this dataset included a total of 7108 rows of crime data prior to data pre-processing, having temporal, spatial and type of crime data only. Marzan et al. also decided to integrate additional data to their dataset to add even more features that can help them predict with better performance. Such examples of newly integrated data includes whether it was raining or not when the crime occurred as well as the weather conditions. Yu et al. gathered their data from a police department in a northeastern city in the USA, having data

[2]https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i?ref=hackernoon.com; [Last Accessed: 2nd May 2022]

[3]https://data.cityofchicago.org/Public-Safety/clear/6bgw-6s23; [Last Accessed: 6th April 2022]

from January 2006 up to December 2009. On the other hand, Stalidis et al. collected an unstructured dataset from news, websites, social media and other means [33].

**KDE-Based Methods**

A method implemented by several authors from literature in this crime prediction objective is the Kernel Density Estimation (KDE) [16, 19, 17, 26], where a probability density function is fitted to historical crime records. This method is popular since it represents spatial distribution of crime data better as well as it is considered to be the most accurate method for crime hotspot detection [10], leading to more accurate predictive models. Gerber [16] focused their research on the city of Chicago, and their main aim was to highlight that adding the data retrieved from Twitter can improve predictions of crimes when implemented with a KDE method. Furthermore, hotspots maps are deemed as being helpful for crime prediction due to the fact that potential future crimes tend to occur in the neighbourhood of historical crimes.

Likewise, Hu et al. [17] and Junior et al. [19] also utilise KDE approach in their research. Junior et al. utilise a proposed algorithm called Polygon Hotspots Approximated to Road (PHAR), which increases the accuracy of predicting crime incidents in certain areas of a city when applied on the hotspot maps generated through KDE. Moreover, Hu et al. gave weight to temporal features attributed to crime, similarly to Yu et al. in [35], since Hu et al. found out that the majority of KDE-based approaches do not even consider them. Subsequently, these authors proposed their own spatio-temporal KDE-based framework. This proposed framework only makes use of the spatial component as opposed to the traditional KDE-based approach mentioned above. In their research, Yu et al. inferred that their framework surpassed other methods commonly used in crime hotspot prediction research studies, such as ProMap [7].

### 2.2.2 Classification Approach

Various authors from the read literature use different methods to predict the type of crime or location, ranging from traditional ML algorithms to DL Architectures [20, 27, 31, 22, 13, 2, 3, 29, 24, 37, 33].

Pradhan et al. in [27] uses Naïve Bayesian classifier and Decision Tree classifier as algorithms. Their purpose is to predict the type of crime in a particular location at specific times, using the SF dataset [12]. Similarly, Yuki et al. in [37] also performs crime type prediction using temporal and spatial features but on the city of Chicago. Multiple ML algorithms and Ensemble methods are implemented, for instance, Random Forest Tree and Bagging respectively. Similar to the works of [27], Sardana et al. [31] implemented similar solutions. The main difference between these two papers is that the latter enhances their model's scores by grouping crime types into three super classes according to the corresponding United States of America (USA)'s law category (Misdemeanour, Infraction and Felony). Sardana et al. implements traditional ML models, such as Logistical Regression Classifier, Decision Tree Classifier and much more, to predict this new categorical variable. Hence, a comparative study analysis was carried out to find the best model which is predicting the most accurate for their research problem. It is worth noting that in [31], a subset of the dataset in [12] is used, having the same features and date range, which is publicly available on Kaggle[4]. Similarly to the works of Sardana et al., Babakura et al. [3] implements a Naïve Bayes Classifier and Back Propagation (BP) to also predict the crime category for diverse states in the USA, using a socio-economic dataset [28]. From their results obtained, they found out that the former model outperforms the other model and thus, is supportive for the prediction in different states in USA.

On the other hand, Elluri et al. [13] implemented similar solutions as well, but decided to integrate additional data to crime. Apart from crime, temporal and spatial data, Elluri et al. also decided to include weather data. They focused their work on NYC crime data and their main aim is to study if weather data can help in predicting the type of crime. Hence, they did a comparative study between predicting crime types with and without weather data. Additionally, they went deeper in their implementations by implementing more machine learning algorithms like SVMs and deep learning methods, including Neural Networks (NNs), Convolution Neural Networks (CNNs) and Recurrent Neural Networks (RNNs).

Furthermore, Rumi et al. identifies several dynamic features which are obtained from Foursquare check-ins based on research findings within the Criminology domain [29]. The main purpose of their research is to study if dynamic features (for example, Region Popularity) help in predicting crime events more accurately, as opposed to static features used in [27, 31, 3, 13]. Rumi et al. implemented several models, including SVMs, Logistical Regression Classifier and NNs, each being fed the same data (historical features, geographical features and demographical features). Every model was executed twice, one containing dynamic features and the other without. Hence, a comparative study could be conducted accordingly in order to answer their research problem in question.

Another research paper tackling this objective is [24]. Llaha implemented several ML models, such as the MLP neural network, using the Waikato Environment for Knowledge Analysis (WEKA) framework [14]. After conducting multiple experiments on their manually collected data, the Decision Tree Classifier turned out to produce the best results from

[4]https://www.kaggle.com/competitions/sf-crime/data; [Last Accessed: 6$^{th}$ April 2022]

the rest. On the other hand, Stalidis et al. aimed at examining DL architectures for crime classification prediction [33]. These authors compare state-of-the-art ML models with DL models, having different configurations. Furthermore, their experiments are conducted on 5 different open datasets and from the results obtained in [24], Llaha conclude that these DL architectures outperform the existing best performing methods. Finally, these authors also evaluate how different parameters configurations can directly improve the overall performance of these architectures in crime prediction.

Yu et al. construct socio-temporal patterns since they identify temporal features as being crucial in crime analysis [35]. These patterns outlines the time, location as well as correlated incidents which are all used as features fed to the model, and thus, can be used for crime prediction. Yu et al. propose an algorithm called Cluster-Confidence-Rate-Boosting (CCRBoost) and compares its performance on other classifiers, such as SVMs and Naïve Bayes. When these models are applied on the USA dataset, they concluded that the CCRBoost performed best overall.

Sathyadevan et al. used the Naïve Bayesian classifier to predict the type of crime given an unstructured document obtained from the unstructured dataset [32]. This is very interesting to note since they managed to build a 90% accurate model that is able to classify crime type from such data. Apart from this, they also implement a predictive model which is able to classify whether a crime will likely occur in a given place according to some attributes. In this final task, the Decision Tree Classifier is used since it is very easy to interpret with a high level of accuracy.

### 2.2.3 Regression Approach

Marzan et al.'s main aim is to predict the future crime trends [26]. They focus on predicting the number of crimes in each district of Manila City's crime data for the future days and weeks. This is achieved by using Time Series Forecasting methods, a data mining technique which is dependent of time. Subsequently, these authors develop several machine learning and deep learning methods that predict future crime rates, which include Linear Regression, Gaussian Processes, Multi-Layer Perceptron and SMOreg. Similarly, Catlett et al. also tackled this objective by predicting the crime rate of Manhattan borough in NYC as well as predicting the crime rate in Chicago [9]. To achieve this, these authors initially implement a clustering algorithm (in their case, DBSCAN) to find regions having crime densely hotspots. Subsequently, every $k$ cluster corresponds to a region, and Catlett et al. then chooses the top 3 most dense regions to predict their future number of crimes individually as well as overall (for whole city). Several ML algorithms are implemented, including Decision Tree Regression and Auto-Regressive Integrated Moving

Average (ARIMA), with the latter model performing the best across both datasets.

On the other hand, Wang et al. [34] and Rumi et al. [30] focus their work on applying linear regression and negative binomial regression on their dataset where they combined POI and taxi flow data to the traditional data, such as geographic data. In their research, they deduced that the addition of such data impacts positively the performance of their developed models. Similar to the works of [34], Zhao and Tang [39] produce their own framework which is able to forecast the rate of crime occurring in the next day and next 7 days. In their experiments, they make use of already existing regression models, such as Linear Regression and Auto-Regressive Moving Average (ARMA). This is then evaluated by experimenting on real crime incidents to observe how accurate this proposed framework is as well as how temporal and spatial patterns affect its performance.

Huang et al. developed a crime prediction framework called DeepCrime [18]. This framework is built on a Deep Neural Network architecture which uncovers dynamic crime patterns and captures time-evolving dependencies between crimes as well as ubiquitous data in urban space. In their research study, each individual crime occurrence is given an input weight vector, and thus, this weight vector denotes the importance of each incident in their respective crime category (types of crime, for instance, felony assaults and robberies) which leads to better prediction of crimes. Furthermore, a MLP is appended in their proposed framework, and hence, they can assign an importance weight to each category. Like Wang et al. [34], POI data is used and combined with other data, i.e., crime data and public service complaint data. Huang et al. then aggregated all this data to predict where crime is expected to occur according to their category. Finally, these authors compare their framework with other existing baseline models (for example, ARIMA) used in [35, 39, 8, 9] in order to evaluate the performance of their framework, where they have concluded that their framework performs better than these baseline models.

As opposed to the already mentioned research papers, Carton et al. [8] took a different approach for this objective. Rather than predicting the number of crimes in the future, these authors are predicting the probability that a police officer will be involved in an adverse event [8]. Therefore, they implement various models, such as Logistical Regression and Random Forest Regressor, which generate a risk score. Hence, this will allow the necessary actions to be taken by law enforcements to further protect these officers having the highest score. Furthermore, their research study also shed a light on how officer's characteristics, situational factors and neighbourhood factors affect the predictive models of adverse events. Examples of such feature's data include the officer's frequent weapon use, number of police misconduct, and if they have been previously suspended. From their research, they concluded that the Random Forest model produced the best results overall.

### 2.2.4 Evaluation Strategies

From the reviewed literature, different evaluation strategies are applied according to the problem in question, meaning if it is either a classification or regression task. In the classification task, the most common approach is outputting metrics after performing k-cross validation, such as accuracy and F1-scores for each model implemented, as done in [20, 22, 35, 2, 31, 13, 27, 3, 29, 24, 37, 33], whilst for the regression task, Mean Absolute Error (MAE), Mean Average Percentage Error (MAPE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) are the most common metrics used across the reviewed literature [26, 34, 39, 30, 9].

Apart from the mentioned evaluation metrics, some authors also calculate the Log Loss score function and Area Under Curve (AUC) scores in their classification problem [31, 13, 27, 29, 33]. Wang et al. [34] also apply the leave-one-out method and calculate the MAE and Mean Relative Error (MRE). The researches who decided to focus their work on crime hotspot prediction (detecting regions with high volumes of crime) [17, 19] evaluated their results obtained using the Predictive Accuracy Index (IPA), which was introduced by Chainey et al. in [11].

## 2.3 Conclusion

From the read literature discussed above, we note very interesting aspects which several authors have focused on within the crime analysis domain, and this has served to us as an inspiration for our work. In the following section, we highlight our methodology towards solving crime type prediction and crime rate prediction respectively, explaining the approaches taken for each objective. In essence, we implement both machine learning and deep learning models, ranging from Decision Trees to Multi-Layer Perceptron neural networks to solve our objectives. Finally, for each objective, we study the amount of training data our developed models requires in order to retain their optimal performances.

# 3 Methodology

In this chapter of our research, we outline the approach taken towards the objectives of crime type prediction and crime rate prediction. Subsequently, we explain the process and decisions taken in our research, highlight any pre-processing steps taken as well as document any observations made. Figure 3.1 illustrated below outlines an overview of our research study.

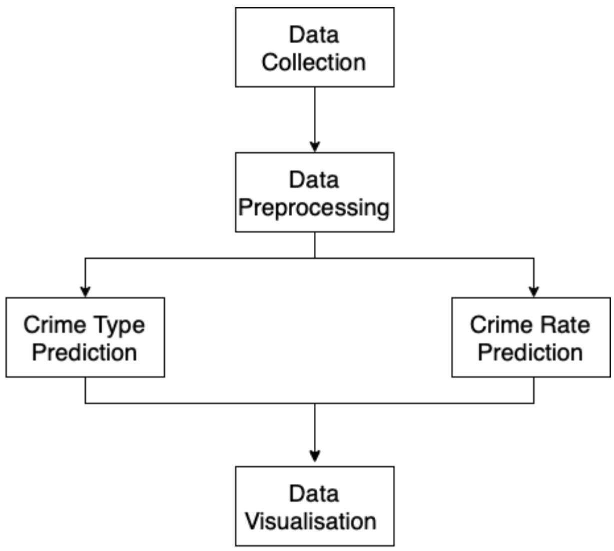


Figure 3.1: Overview of Solution

## 3.1 Datasets Used

In our research, we intend to compare our results obtained with those in literature [31, 9]. For this reason, we decided to utilise the same datasets which the respective authors used in their research study in order to provide this comparative analysis. The dataset used for Objective 1 is SF's dataset found on Kaggle, as used by Sardana et al. [31], and contains 878049 rows of crime incidents between the years of 2003 to 2015. On the other hand, for Objective 2 NYC crime data was used, as utilised by Catlett et al. [9], and contains a total of 1328134 historical crimes between January 2006 to December 2016.

Before implementing our models, it was vital to prepare the respective datasets. Therefore, pre-processing techniques commonly used in literature [32, 26, 3] were applied to the data:

1. **Data Reduction**: reducing the data, for instance, removing unneeded columns.
2. **Data Cleaning**: dealing with missing and inconsistent values. Fixing column values which are misspelled or uneasy to understand.
3. **Data Integration**: integrating new attributes from current features, or from external sources into the dataset.
4. **Data Transformation and Discretisation**: transforming feature values into groups and encoding categorical features before implementing the models to solve our objectives.

Since our research study consists of two different datasets, then different pre-processing steps had to be applied for the respective data. Furthermore, to guarantee a fair comparative analysis, any pre-processing steps done by the authors, were also done from our end.

When it comes to the NYC crime data used in the crime rate prediction objective, we followed the same pre-processing methodology done by Catlett et al. [9]. The first step included reducing this dataset by removing any features which will not be used for this particular objective [9]. Examples of such features includes the suspect's age as well as the victim's ethnicity. Furthermore, these authors focused solely on predicting the number of crimes in Manhattan borough only between the period of 2006 to 2016 [9], and thus, our dataset was further reduced to filter out data where the borough is not Manhattan as well as where the date is not between the stipulated period. The reason why Catlett et al. did this is because Manhattan is the epicenter of crimes within NYC and thus, predicting the future number of crimes is beneficial for law enforcement to plan their task force more effectively. Figure 3.2 illustrated below shows how the rate of crime fluctuates in Manhattan borough between the years of 2006 and 2016.

In terms of data cleaning, there was not much to perform since every row in the dataset is a crime in itself. Subsequently, it was ensured that any records with missing values within the *date* or *borough* feature values were eliminated from the dataset. The reason why this was required is due to the next pre-processing step involved, i.e., data integration and this will be highlighted below.

We decided to integrate additional attributes to help us in our regression-based prediction task, similarly to [13, 4]. Therefore, we included population data, unemployment rate data as well as median income data, which were downloaded from Citizens' Committee for Children (CCC) of New York[1], where this data is sourced from the official US Census[2]. This downloaded data was merged into our current dataset using the *date* column and *borough* column, and this is the reason why these mentioned attributes could not contain any missing values. In addition, we also calculated the number of crimes which occurred in each year. Finally, there was no need to perform any data transformation and discretisation since our dataset only includes numerical values. A sample of the dataset can be seen in Table 3.1 below.

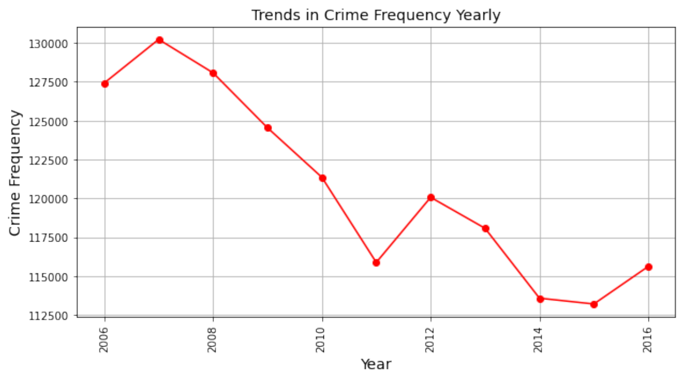


Figure 3.2: Line Graph for Crime Rate in Manhattan

Table 3.1: NYC Pre-Processed Dataset Sample

| Year | Population | Unemployment Rate | Median Incomes | Crimes |
|---|---|---|---|---|
| 2010 | 1586698 | 0.09173 | 75009 | 121364 |
| 2011 | 1601948 | 0.08969 | 75527 | 115889 |
| 2012 | 1619090 | 0.08816 | 74853 | 120076 |

On the other hand, similar pre-processing steps, as done by Sardana et al. [31], were carried out on the SF crime data which is used for crime type prediction. Initially, our downloaded dataset contained some features which are not needed for our research, for instance, the *address* and *coordinates*, which were all removed. It is important to note that Sardana et al. used the *coordinates* features to derive a new feature called *zipcode*.

[1] https://data.cccnewyork.org/data; [Last Accessed: 14$^{th}$ April 2022]

[2] https://www.census.gov/programs-surveys/popest/technical-documentation/research/evaluation-estimates/2020-evaluation-estimates/2010s-state-total.html; [Last Accessed: 14$^{th}$ April 2022]

However, we inferred that such a new feature is not needed since it does not really improve the models' accuracy, and thus, that is the main reason why this feature has been eliminated from the SF dataset. The data cleaning process was then performed on the dataset by eliminating the small percentage of rows which contains missing values from this dataset.

The next pre-processing step done is data integration. In this step, the *date* column was split into 3 separate columns to derive the *year*, *month* and *hour* features, like [31]. As opposed to [31], we decided to categorise these newly added feature columns since they contain multiple distinct values. This is because we deduced that transforming such feature's values into categories results in better overall model accuracies. For instance, the *month* feature contains 12 different values, each value corresponding to a month. Hence, we decided to convert this feature into a new feature called *Season*, where we mapped the values within the *month* attribute into their respective category group, in this case the season. The same process have been done for *hour* and *day*, creating *DayGroup* and *WeekGroup* features respectively. This mentioned mapping of feature's values to categories is further highlighted in Tables 3.2, 3.3 and 3.4 below. On the other hand, Figures 3.3, 3.4 and 3.5 below show the distribution of these integrated feature's values for *DayGroup*, *WeekGroup* and *Season* respectively. Finally, all these categorical features were encoded using the one-hot-encoded method prior to creating the model for prediction, as done by Sardana et al. [31]. A sample of SF dataset can also be seen in Table 3.5 below.

Table 3.2: SF *DayGroup* Super Class

| **Super Class** | **Super Class Types** |
|---|---|
| Morning | 05, 06, 07, 08, 09, 10, 11 |
| Afternoon | 12, 13, 14, 15, 16 |
| Evening | 17, 18, 19, 20 |
| Night | 21, 22, 23, 00, 01, 02, 03, 04 |

Table 3.3: SF *WeekGroup* Super Class

| **Super Class** | **Super Class Types** |
|---|---|
| Weekday | Monday, Tuesday, Wednesday, Thursday, Friday |
| Weekend | Saturday, Sunday |

Table 3.4: SF *Season* Super Class

| Super Class | Super Class Types |
|---|---|
| Spring | March, April, May |
| Summer | June, July, August |
| Winter | December, January, February |
| Autumn | September, October, November |

Table 3.5: SF Pre-Processed Dataset Sample

| Category | PdDistrict | Resolution | DayGroup | Season | WeekGroup |
|---|---|---|---|---|---|
| Misdemeanor | NORTHERN | ARREST, BOOKED | Night | Spring | Weekday |
| Felony | PARK | NONE | Night | Spring | Weekday |
| Felony | CENTRAL | NONE | Night | Winter | Weekday |

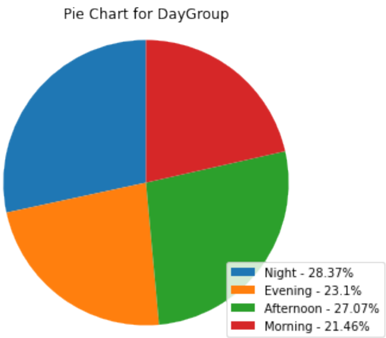


Figure 3.3: Pie Chart - *DayGroup*

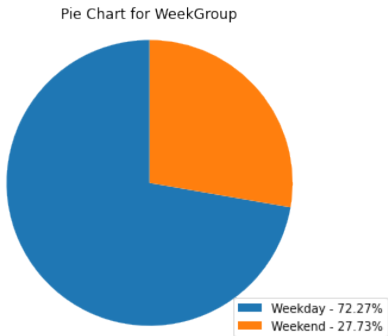


Figure 3.4: Pie Chart - *WeekGroup*

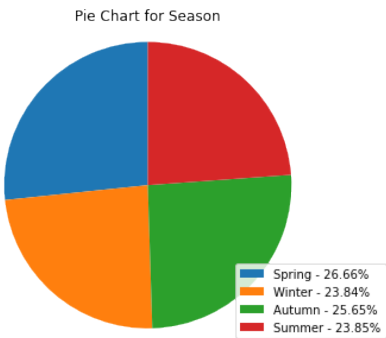


Figure 3.5: Pie Chart - *Season*

## 3.2 Crime Type Prediction

In this section we describe the first objective carried out in our dissertation, which focuses on predicting the type of crime category given temporal and spatial data. Our main aim in this objective is to replicate [31] in such a way to implement models that perform with a higher level of accuracy than those of Sardana et al., leading to a direct comparative analysis with this research paper. Therefore, as already explained in Section 3.1, the same dataset will be used and we will be implementing both classical ML models as well as a DL model, which will be further outlined in Section 3.2.2 below. The methodology process of this objective is illustrated in Figure 3.6 below.

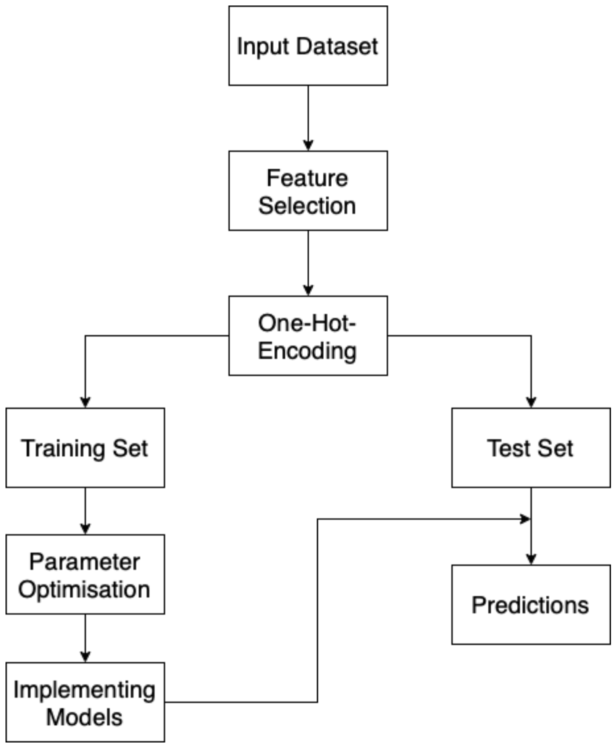


Figure 3.6: Crime Type Prediction Objective Methodology

### 3.2.1 Dataset Structure and Feature Selection

Our SF dataset structure differs from some datasets seen in the read literature [29, 32]. Rumi et al. identifies several dynamic features from Foursquare check-ins, such as region popularity, to help them achieve better performances in their model implementations [29], as explained in Section 2 above. On the contrary, our dataset only has static data, ranging from temporal and spatial data to the outcome of arrest data. Additionally, their dataset included demographic data too which were being fed to their created models, which again differs from ours. As opposed to Sathyadevan et al., our dataset is a structured one rather than an unstructured one like [32], since Sathyadevan et al. gathered their data from several different means, including news, social media as well as websites.

As already mentioned, Sardana et al. started off their work by implementing models which predict all the possible crime types, having a total of 39 distinct classes [31]. Since this SF dataset is imbalanced, they did not manage to obtain a good overall score (their best model produced an F1 score of 0.19). Therefore, they decided to optimise their models by grouping these crime types into three crime super classes, called Infraction, Misdemeanour and Felony respectively. This resulted into a more balanced dataset overall and lead to an improvement in the performance of the models (their best model now produced an F1 score of 0.59). A sample of the dataset in question can be seen in Table 3.5 above, however, Sardana et al. eliminated the *resolution* feature in their work.

Now, our goal is to start off by replicating [31], and then expand their work in order to improve on their existing research. To do so, we first made sure that we were getting similar results when using the same data which Sardana et al. fed into our models. For Objective 1, we decided to use various classical ML algorithms and a DL method, such as the Decision Tree Classifier and MLP neural network respectively, which will be further discussed in Section 3.2.2 below. Hence, our study will make use of a DL method as opposed to [31], and thus, we will be able to analyse if such a more advanced models produces better results or not for this objective. When we obtained almost identical results (approximately 2% difference between our models and theirs), we moved on to decide an action plan to try and improve these models performances. To achieve this, we begin by integrating new attribute features (for example, *Season*), as described in Section 3.1 above, to further continue balancing our dataset. Furthermore, we noticed that these authors eliminated *resolution* feature (highlights if an arrest was made or not) from their dataset, which could be a very beneficial for our models to predict with a higher degree of accuracy. Thus, we decide to retain this feature and feed it as input to our developed models, and compare our findings with those in [31].

It is important to note that our dataset has been split into a training set and test set in an 80:20 ratio split. This was done in order to allow sufficient data for our models to train on and enough data for evaluation purposes. In total, our data contains 878049 rows, where each row corresponds to a single crime. The structure of both the training set and test set is the same as Table 3.5 above, where *category* is the target column whilst the rest are solely the explanatory features. Moreover, since ML and DL models cannot work with categorical data, then it is imperative to choose the right approach that converts these variables from categorical to numerical form. Hence, all the categorical inputted features are encoded using the one-hot-encoding method, like Sardana et al. used in [31]. This method essentially represents categorical variables as binary vectors and thus, does not assume there is a numerical ordinal between the categories of a variable. This is the main reason why such a pre-processing method was used which leads to better performance of the developed models overall. For instance, the *WeekGroup* variable has two distinct categories (Weekday and Weekend), and therefore, two binary variables are needed respectively. A value of one is placed in the binary variable for the value of *Weekday* and a zero for *Weekend*.

### 3.2.2 Classification-based Models

To solve this objective, several models were implemented in order to classify the crime category type. These include k-Nearest Neighbours (kNNs), Logistical Regression Classifier, Decision Tree Classifier and a MLP Classifier. To the contrary of [31], we decided to develop a DNN for this research objective to study whether more advanced and complex models perform better than traditional ML algorithms.

Prior to implementing the aforementioned models, we performed parameter tuning optimisation on each of the models created accordingly. The approach taken for this technique was finding the optimal parameters for each model manually. This was achieved by experimenting with different permutations of the parameters and choosing the one which gave us the best results. For instance, we inferred that a $k$ value of 11 for the kNN model produced best results whilst a 5 hidden layer architecture with 100 units each performed best for the MLP model.

## 3.3 Crime Rate Prediction

The other objective focuses on forecasting the number of crimes likely to happen in the next three years individually, based on previous crime data. We use regression-based traditional ML models as well as DL models to predict the crime rate for the years 2014, 2015 and 2016 separately, similar to Catlett et al. [9]. In this way, we will be able to directly compare our obtained results with those found in [9]. The process of this objective methodology is depicted in Figure 3.7 below.

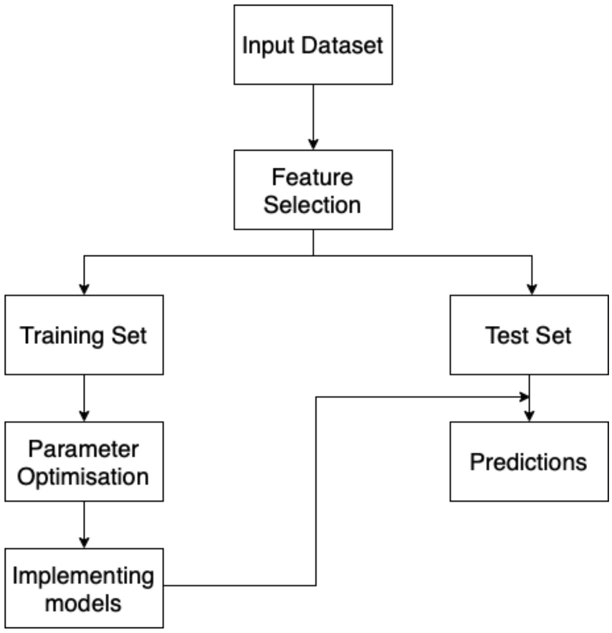

Figure 3.7: Crime Rate Prediction Objective Methodology

### 3.3.1 Dataset Structure and Feature Selection

Due to the fact that we did not follow a KDE-based approach nor a binary classification approach in Objective 2, our selected features differ from numerous researches in the read literature that followed either of the aforementioned approaches [16, 19, 17, 26, 18]. For instance, Gerber uses Twitter and KDE-based features [16], similar to [19, 17, 26] whilst Zhao and Tang uses attributes from POI data in addition to human mobility [39], and Huang et al. uses public service complaint data amongst others [18]. Our chosen features are different than these mentioned studies, and such features used in our study will be discussed next.

Our methodology also uses external integrated data as well as historical crime data to predict the number of crimes in Manhattan borough. We begin by generating yearly predictions on the test set using past crime data. Then, we start selecting additional features (either *population*, *unemployment rate* or *median income*), add them to our training features and keep on predicting the target years separately, each time choosing different features to see how results are faring. For instance, we start predicting crime for the years of 2014, 2015 and 2016 by using crime data and *population* data, then we add *unemployment rate* data, re-create the models, and repeat this process by choosing other feature combinations. At the end of this process, we end up with results for every feature combinations and for each model created. Subsequently, we are now in a position to observe how every model performed when using different features combinations, and thus, we are able to deduce which models performed best with what feature combinations.

Furthermore, our dataset is split into a training set and a test set, with the former having the period from 2006-2013 whilst the latter includes the already mentioned target years, and thus, we are replicating [9] for a fair comparative analysis. With this split, we ensure that we have sufficient input data for the models to train on. The structure of our training set and test set is identical to Table 3.1 above, however, since Table 3.1 is a sample, the training set actually consists of data ranging from 2006 up to 2013 whilst the test set contains data for 2014, 2015 and 2016 respectively since one year ahead predictions will be performed. Additionally, when we select the external feature columns, we add/remove them accordingly from the input dataframe in order to end up with the desired training features.

### 3.3.2 Regression-based Models

To achieve the goal of this objective, two traditional ML algorithms and one DL method were implemented. These models include the Linear Regression, Random Forest Regressor and MLP Regressor. The two ML models were chosen because these were the most commonly implemented in the read literature (as described in Section 2 above) whilst the last model was implemented in order to extend our research using more modern techniques that are able to provide great results.

Each of the above models were implemented for every combination of features chosen. Additionally, parameter tuning optimisation was carried out throughout the whole implementation of each model. Parameter tuning optimisation was done manually by experimenting with different parameters for each model created. Hence, the best parameter combination could be identified which yields the best evaluation metrics on the test set. It is important to note that the same models can have different parameter values for different combinations of features. For instance, the MLP models built can have different architectures depending on the features used in training. Furthermore, examples of parameters combinations derived from this process include the following two; the first concerns the number of hidden layers and units needed for the MLP model whilst the second includes the number of trees required in the Random Forest regressor.

Furthermore, we decided to fit both a simple Linear Regression as well as a Multiple Linear Regression since this objective focuses on predicting a continuous value. Subsequently, we observed how multiple independent variables, being the population, unemployment rate and median income, can lead to better overall predictions. Hence, we implement both types of models in our research, and thus, consider the aforementioned explanatory variables to deduce whether they are indeed helping our models in predicting the forecasted crime rate more accurately.

## 3.4 Training Data Required for Models Implemented

In this section, we will be discussing how we went about in solving Objective 3. In this objective, we investigate the amount of training data our classification and regression models require such that their optimal performance is retained. Thus, we split this section into two sub-sections in order to highlight the methodology for our classification and regressor models respectively.

### 3.4.1 Classification-based Models

The first part of this objective is analysing how much training data our built classification models require to retain their performances effectiveness. Thus, we will be using every model built in Objective 1 and perform multiple experiments. The scope of these experiments is to determine the amount of training data required in order to start performing classification prediction. In each experiment, we first diminish the training set gradually and then re-run the models accordingly. The process of diminishing the training set is repeated until the minimum training size is derived such that the models retain their optimal performances. It is important to note that the test set and the model's parameters remain the same throughout these experiments and experiments are only conducted on classifying crime super classes using temporal, spatial as well as *resolution* features. Furthermore, in these experiments we also perform the process of one-hot-encoding method on the features in order to ensure fairness. For example, the whole SF dataset is stored within a dataframe, and the first 80% rows belong to the training set whilst the last 20% pertains to the test set. The results obtained after performing these experiments are found in Section 4.3.

The experiment process conducted is the following. Training features are first passed through the one-hot-encoding process, then all models are re-run and finally results are noted down. Then, this process is reiterated again until the model's remain effective, meaning that the training set is reduced from 80% to for example, 60% whilst the last 20% rows of the whole dataset always pertains to the test set and models are executed again. At the end of this process, we would have found the minimum data size our ML and DL models require to produce effective results.

### 3.4.2 Regression-based Models

The second and last part of Objective 3 is similar to the first part and includes the investigation of how much data our regressor models need to train on to keep on predicting the number of future crimes with great performance. Hence, to guarantee fairness throughout this whole experimentation, we decided to use the same ML and DL models with the exact same feature combinations (as described in Section 3.3), and each respective model has identical parameters as performed in Objective 2 previously. For instance, if in Objective 2 a Linear Regression model was used having crime data and population data as features, then the same Linear Regression model having the exact same parameter values will be used in Objective 3 with the corresponding features.

Then, we start off by splitting our NYC dataset into a training set and a test set, and move on to diminish the training set while keeping the test set unchanged. As mentioned in Section 4.3 above, our dataset contains data from 2006 to 2016, and the models implemented are performing one-year ahead predictions for the years of 2014, 2015 and 2016 respectively (test set) and are being trained on data from 2006 up to 2013. Hence, this training set is diminished gradually, each time re-running the models and checking their performance. This process is repeated until the minimum size of training data is found such that all models retain their effectiveness.

## 3.5 Conclusion

In this section, we have given an overview of what our dissertation objectives includes, and how we intend to solve them. Specifically, we utilise classification-based techniques to predict the type of crime as well as utilised regression-based techniques to predict the number of future crimes. Several different models have been implemented for each objective, ranging from traditional ML, such as decision trees and linear regression algorithms respectively, to DL models, including the MLP neural network. In Section 4 below, we explain how our developed models will be evaluated and note down interesting observations inferred from the results obtained.

# 4 Evaluation

In this chapter, we explain our evaluation strategy, present our results and evaluate all objectives. In Section 4.1 and 4.2, which tackle Objectives 1 and 2 respectively, we analyse the results for the models built and evaluate them. Then, in Section 4.3, we assess our final Objective 3 by also presenting our findings and discussing them accordingly. Furthermore, we highlight any similarities between the read literature's evaluation strategies and our own.

## 4.1 Evaluation of Crime Type Prediction

For Objective 1, we aim to classify the type of crime based on temporal, spatial and whether an arrest was made or not, using classical ML and DL models, being Logistical Regression Classifier, kNN, Decision Tree Classifier as well as MLP classifier neural network. As highlighted in Section 3.2 above, we begin by replicating the findings of Sardana et al. [31], and then proceed to improve their results. Hence, a total of three experiments were made, two of which concerns replicating Sardana et al.'s system using the same original features as used there, and the last one is an extension to Sardana et al.'s system to expand on their work. Since the San Francisco (SF) dataset has an imbalanced class distribution, we used the weighted F1 score as it is the weighted harmonic mean between precision and recall values, as utilised by [31]. Furthermore, 10-fold cross validation was also performed on every model implemented in order to avoid possible overfitting.

Table 4.1 and 4.2 below outlines the metric scores obtained for the first and second experiments respectively. As mentioned, these experiments use the exact same original features, including *month* and *day*, as used by Sardana et al. in [31]. When these F1 scores are directly compared to [31], we deduce that we have managed to obtain similar scores. Sardana et al. obtained a maximum F1 score of 0.19 and 0.59 when classifying all crime classes and crime super classes respectively whilst we obtained a maximum F1 score of 0.17 and 0.58 respectively, which we consider to be almost identical.

Now, to improve on the findings of [31], we convert some input feature's values into their respective category and introduce the *resolution* feature (which was neglected by Sardana et al.), as explained in Section 3.2.1 above. Table 4.3 below tabulates the results obtained after optimising some features and adding the *resolution* feature (if an arrest was made or not).

Table 4.1: Model Evaluation F1 Scores for Classifying Crime Classes

| **Model** | **F1 Score (weighted average)** |
|---|---|
| Logistical Regression Classifier | 0.08 |
| kNN | 0.16 |
| Decision Tree Classifier | 0.17 |
| MLP | 0.17 |

Table 4.2: Model Evaluation F1 Scores for Classifying Crime Super Classes

| **Model** | **F1 Score (weighted average)** |
|---|---|
| Logistical Regression Classifier | 0.56 |
| kNN | 0.57 |
| Decision Tree Classifier | 0.57 |
| MLP | 0.58 |

Table 4.3: Model Evaluation F1 Scores for Classifying Crime Super Classes using *Resolution* Feature

| **Model** | **F1 Score (weighted average)** |
|---|---|
| Logistical Regression Classifier | 0.71 |
| kNN | 0.68 |
| Decision Tree Classifier | 0.71 |
| MLP | 0.71 |

Hence, comparing our classification techniques with each other, we can note that overall all models implemented fared quite similar. Furthermore, once the *resolution* feature was introduced as part of the training features, the results improved for every evaluation metric and thus, implies that such a feature impacts positively the performance of each developed model. When comparing our findings with those in [31], we observed that our models are producing better overall results, where the f1 score has increased from approximately 0.59 up to 0.71. From Table 4.3 above, the model which performed the worst is kNN whilst Logistical Regression Classifier, Decision Tree Classifier as well as MLP produced better and identical results.

## 4.2 Evaluation of Crime Rate Prediction

As explained in Section 3.3 above, in Objective 2 we predict the number of crimes that will take place in future years, by implementing classical ML (non-DL) and DL models on NYC's dataset. Hence, we have implemented Linear Regression, Random Forest Regressor and a MLP regressor neural network models. For each technique, we perform 5 experiments; the first experiment using only previous crime incidents, the second using crime and population data, the third using crime and median income data, the forth using crime and unemployment rate data and the last using crime, population and unemployment rate data. Then, for each experiment conducted, we provide the MAPE evaluation scores, a common evaluation metric used by [9, 26, 30], which will allow us to observe the quality of the developed models and directly compare our findings with [9].

Tables 4.4, 4.5, 4.6, 4.7 and 4.8 below outline the results obtained when conducting the mentioned experiments, meaning that models were trained using different data in each table. Furthermore, the values in these tables represents the MAPE scores, which are rounded to the nearest 2 decimal places, and are used to evaluate the quality of our implemented models on the test set. Each experiment utilised the training set from 2006 up till 2013, and we are predicting the yearly crime rate for the next 3 years (2014, 2015 and 2016 respectively), as done by Catlett et al. [9]. For the Random Forest Regressor, we present our metric scores when using 150 trees whilst for MLP, the number of hidden layers varies depending on which training data is used. We deduced these parameter's values since we found out that they produced the lowest MAPE metric scores across the different experiments.

From the results tabulated below, we can observe that we managed to obtain great metric scores throughout all experiments. When comparing each regression technique results, we can note that the MLP model fared best overall, generating the best score of 0.74 whilst the Random Forest Regressor produced the worst overall results across the 4 experiments. It is worth noting that both the Linear Regression and Random Forest models started producing worse results as soon as additional data (example, *population* data) was appended to the training set (included crime data only initially). For instance, in Table 4.4 we can observe that the average MAPE value for the Linear Regression is 1.72, however, worsened to 2.3, 2.06, 2.21 and 2.41 in Tables 4.5, 4.6, 4.7 and 4.8 respectively. On the contrary, when it comes to the MLP DL model, appending additional data results in better scores, with the exception when adding *median income* data. For example, in Table 4.4 above we can note that initially, we managed to obtain an average MAPE score of 8.24, but this improved to 1.2, 0.74 and 2.55 in the other experiments conducted (Tables 4.5, 4.6 and 4.8 respectively).

**Comparative Analysis with Catlett et al.'s Research Paper**

When directly comparing our findings with those in [9], we can infer that we have managed to obtain better scorings than Catlett et al.'s study. It is worth re-mentioning that we have used the exact same NYC crime dataset as well as performed all pre-processing steps which Catlett et al. did. The main differences between our study and [9] is that we integrated external data to the dataset, such as *unemployment rate* and also decided to implement other traditional ML models, such as Linear Regression as well as more advanced DL model, being the MLP neural network. In their study, Catlett et al. concluded that the ARIMA time-series forecast model fared best overall, with an average MAPE of 5.67. However, in our study we managed to implement a MLP model having an average MAPE score of 0.74, which is substantially less than the best score in [9]. This score is tabulated in Table 4.6 above, where we used past crime and unemployment rate data to perform the mentioned one-year-ahead forecasts. From these results obtained and discussed, it can evidently be seen that the *unemployment rate* feature is much more important than both the *population* and *median income* features when predicting crime rate since the MLP model produced the best overall result when using crime data and *unemployment rate* data as training features than any other combination of features. Furthermore, we also observed that when incorporating the *median income* feature, the models predicted less accurate results as opposed to *population* and *unemployment rate* data. Therefore, this feature is evidently the least important feature overall due to its impact on the model's performances.

Table 4.4: MAPE Scores for Crime Rate Prediction using Crime Data Only

| Year | Linear Regression | Random Forest Regression | MLP |
|---|---|---|---|
| 2014 | 1.11 | 4.28 | 8.71 |
| 2015 | 0.19 | 4.63 | 9.12 |
| 2016 | 3.88 | 2.44 | 6.89 |
| Average MAPE | **1.72** | 3.78 | 8.24 |

Table 4.5: MAPE Scores for Crime Rate Prediction using Crime and Population Data

| Year | Multiple Linear Regression | Random Forest Regression | MLP |
|---|---|---|---|
| 2014 | 3.14 | 2.63 | 1.13 |
| 2015 | 5.08 | 5.42 | 3.23 |
| 2016 | 1.37 | 2.21 | 0.03 |
| Average MAPE | 2.3 | 4.57 | **1.2** |

Table 4.6: MAPE Scores for Crime Rate Prediction using Crime and Unemployment Rate Data

| Year | Multiple Linear Regression | Random Forest Regression | MLP |
|---|---|---|---|
| 2014 | 3.12 | 2.4 | 0.68 |
| 2015 | 5.97 | 6.32 | 4.1 |
| 2016 | 0.18 | 0.56 | 1.48 |
| Average MAPE | 2.06 | 5.46 | **0.74** |

Table 4.7: MAPE Scores for Crime Rate Prediction using Crime and Median Income Data

| Year | Multiple Linear Regression | Random Forest Regression | MLP |
|---|---|---|---|
| 2014 | 3.47 | 5.93 | 10.08 |
| 2015 | 2.02 | 6.24 | 9.51 |
| 2016 | 1.14 | 4.06 | 8.63 |
| Average MAPE | **2.21** | 5.41 | 9.4 |

Table 4.8: MAPE Scores for Crime Rate Prediction using Crime, Population and Unemployment Rate Data

| Year | Multiple Linear Regression | Random Forest Regression | MLP |
|---|---|---|---|
| 2014 | 3.7 | 6.77 | 2.72 |
| 2015 | 3.34 | 7.12 | 3.58 |
| 2016 | 0.19 | 4.88 | 1.36 |
| Average MAPE | **2.41** | 6.25 | 2.55 |

## 4.3 Evaluation of Training Data Required

In this section, we evaluate Objective 3 where we analyse the minimum amount of data our built models require to train on such that their overall effectiveness is not hindered. To the best of our knowledge, no one has attempted to solve an objective similar to this, and thus, we cannot directly compare our findings with the read literature. This section is split into two sub-sections; the first sub-section refers to our classification models whilst the second one concerns our regressor models.

### 4.3.1 Classification-based Models

As previously explained in Section 3.4 above, we analyse the effectiveness of the classification models when their training data size is reduced. Furthermore, in Section 3.4 we also discuss the process of how this objective is achieved. Through this process, we have derived the size of the training set needed to train our models and the respective f1 scores obtained for each model. Hence, in this experiment, the training set consisted of the first 40% of crime records from the whole SF dataset, whilst the test set is always the last 20% rows. It is also worth re-mentioning that in this part of Objective 3, the *resolution* feature together with the other features, such as *DayGroup* and *Season* were utilised throughout the entire list of experiments carried out.

Table 4.9: Model Evaluation F1 Scores for Classifying Crime Super Classes using *Resolution* Feature on the full and the diminished Training Set

| Model | F1 Score (Obj. 1) | F1 Score (Obj. 3) |
|---|---|---|
| Logistical Regression Classifier | 0.71 | 0.70 |
| kNN | 0.68 | 0.67 |
| Decision Tree Classifier | 0.71 | 0.70 |
| MLP | 0.71 | 0.70 |

Table 4.9 above outlines the f1-scores obtained for each classification model implemented in Table 4.3 above for Objective 1 together with their corresponding result after having diminished the training set to 40% of the whole SF dataset (equivalent to 351219 rows of crime data). As can be observed in this table, the results obtained fared almost identical to each other, having the highest score of 0.7 for Logistical Regression Classifier, Decision Tree Classifier and Multi-Layer Perceptron (MLP) neural network models, as compared to the highest score of 0.71 in Section 4.1 above. It is also interesting to observe that there is a similarity between these obtained scores (with diminished training set) and those in Objective 1 (whole training set) because all the respective models which are trained on the minimal training size have f1-scores (Objective 3) that are 0.01 less than their corresponding model performance in Objective 1. For instance, the kNN model in Table 4.9 produced an f1 score of 0.67 whilst in Table 4.3, it produced the result of 0.68, which can be directly observed in 4.9 where these results are compared side by side. Therefore, a training set which is 40% of the whole dataset is enough for the models to train on and yet produce effective results.

### 4.3.2 Regression-based Models

As highlighted in Section 3.4 above, we aim to reduce the size of the training data our ML and DL models are using. After the process of identifying the minimum training size, we obtained the following results which are tabulated accordingly. In the tables highlighted below, we are directly comparing side by side the results obtained for Objective 2 in Section 4.2 above with their corresponding value obtained for Objective 3.

Table 4.10: MAPE Scores for Crime Rate Prediction using Crime data on the full and the diminished Training Set

| Year | Linear Regression | | Random Forest Regr. | | MLP | |
|---|---|---|---|---|---|---|
| | **Obj. 2** | **Obj. 3** | **Obj. 2** | **Obj. 3** | **Obj. 2** | **Obj. 3** |
| 2014 | 1.11 | 3.38 | 4.28 | 4.23 | 8.71 | 4.76 |
| 2015 | 0.19 | 3.22 | 4.63 | 4.57 | 9.12 | 5.15 |
| 2016 | 3.88 | 0.58 | 2.44 | 2.38 | 6.89 | 3.01 |
| Average MAPE | **1.72** | **2.39** | 3.78 | 3.72 | 8.24 | 4.3 |

Table 4.11: MAPE Scores for Crime Rate Prediction using Crime and Population data on the full and the diminished Training Set

| Year | Multiple Linear Regr. | | Random Forest Regr. | | MLP | |
|---|---|---|---|---|---|---|
| | **Obj. 2** | **Obj. 3** | **Obj. 2** | **Obj. 3** | **Obj. 2** | **Obj. 3** |
| 2014 | 3.14 | 2.86 | 2.63 | 4.23 | 1.13 | 9.53 |
| 2015 | 5.08 | 2.23 | 5.42 | 4.57 | 3.23 | 10.44 |
| 2016 | 1.37 | 1.61 | 2.21 | 2.39 | 0.03 | 8.09 |
| Average MAPE | 2.3 | **2.41** | 4.57 | 3.72 | **1.2** | 9.35 |

Table 4.12: MAPE Scores for Crime Rate Prediction using Crime and Unemployment Rate Data on the full and the diminished Training Set

| Year | Multiple Linear Regr. | | Random Forest Regr. | | MLP | |
|---|---|---|---|---|---|---|
| | **Obj. 2** | **Obj. 3** | **Obj. 2** | **Obj. 3** | **Obj. 2** | **Obj. 3** |
| 2014 | 3.12 | 5.43 | 2.4 | 3.64 | 0.68 | 1.62 |
| 2015 | 5.97 | 161.69 | 6.32 | 3.99 | 4.1 | 2.44 |
| 2016 | 0.18 | 400.18 | 0.56 | 1.81 | 1.48 | 0.28 |
| Average MAPE | 2.06 | 189.1 | 5.46 | 3.14 | **0.74** | **1.44** |

Table 4.13: MAPE Scores for Crime Rate Prediction using Crime and Median Income Data on the full and the diminished Training Set

| **Year** | **Multiple Linear Regr.** | | **Random Forest Regr.** | | **MLP** | |
|---|---|---|---|---|---|---|
| | **Obj. 2** | **Obj. 3** | **Obj. 2** | **Obj. 3** | **Obj. 2** | **Obj. 3** |
| 2014 | 3.47 | 2.7 | 5.93 | 5.93 | 10.08 | 10.08 |
| 2015 | 2.02 | 2.98 | 6.24 | 6.24 | 9.51 | 9.55 |
| 2016 | 1.14 | 0.38 | 4.06 | 4.06 | 8.63 | 8.63 |
| Average MAPE | **2.21** | **2.02** | 5.41 | 5.41 | 9.4 | 9.42 |

Table 4.14: MAPE Scores for Crime Rate Prediction using Crime, Population and Unemployment Rate Data on the full and the diminished Training Set

| **Year** | **Multiple Linear Regr.** | | **Random Forest Regr.** | | **MLP** | |
|---|---|---|---|---|---|---|
| | **Obj. 2** | **Obj. 3** | **Obj. 2** | **Obj. 3** | **Obj. 2** | **Obj. 3** |
| 2014 | 3.7 | 5.4 | 6.77 | 3.73 | 2.72 | 0.21 |
| 2015 | 3.34 | 160.71 | 7.12 | 4.07 | 3.58 | 0.65 |
| 2016 | 0.19 | 397.57 | 4.88 | 1.9 | 1.36 | 1.52 |
| Average MAPE | **2.41** | 187.89 | 6.25 | 3.23 | 2.55 | **0.79** |

From the results tabulated above, it can be observed that our models produced similar results to Tables 4.4, 4.5, 4.6, 4.7 and 4.8 respectively. The models were fed in training data between the years of 2010 and 2013, as opposed to crime data from 2006 to 2013 in Section 4.2 above. However, although the training size was reduced by half, the models still managed to produce great results overall. The lowest MAPE score obtained in this part of Objective 3 is 0.79, which is produced by the MLP model when using crime, population and unemployment rate data, as highlighted in Table 4.14 above, when compared to 0.74 in Objective 2. Additionally, the other ML models generated better accuracies, as opposed to the MLP DL model when crime data was only used as training features. However, the MLP model then started producing better the ML models as soon as external data was used as training features. It is worth noting that in this experiment, the *unemployment rate* feature was also the most important feature since it had the most positive impact on the model's predictions (when only one external data is combined with crime data). Furthermore, the MLP model is also the best overall model in terms of accurate predictions as it produced the lowest MAPE value when training it on crime, population and unemployment rate data.

## 4.4 Conclusion

For crime type prediction, we found that adding *resolution* feature as part of the training set increases the overall f1 scores of the models. This feature was neglected by Sardana et al. [31], however, we observed that such a feature can be highly beneficial when training the models, and thus, should be used when classifying crime. Additionally, we also note that the MLP DL model was not more effective when compared to the other ML models, such as the Random Forest Classifier because they produced the same results.

With regards to crime rate prediction, we observed that all classical ML models produced their best results when using historical crime data only, whilst the MLP neural network started producing better results than the rest as soon as additional data (for example, *population*) was integrated to the training set. Overall, the MLP model performed best when the *unemployment rate* feature was incorporated in the dataset. Subsequently, we also observe that the DL models can be very effective when predicting the future number of crimes since our MLP model performed the best.

When it comes to Objective 3, we inferred that diminishing the size of the training set, our implemented models can still produce similar results and thus, remain effective. With regards to the classification models, we found out that taking the first 40% records of the whole SF dataset as training set is enough for the models to train on. The f1 scores for the classification models were all 0.01 less than the the respective f1 scores in Objective 1. Furthermore, the classification models were fed training features ranging from temporal data, spatial data up to whether an arrest was made or not. On the other hand, for the regression part of this objective, we deduced that using training data from 2010 up to 2013 was sufficient to train our models without hindering their performances, thus retaining their optimal performances. Hence, similar MAPE values were obtained when compared to Objective 2. Additionally, we observed that the MLP model produced the best results overall when using crime, population and unemployment rate data as training features.

Given various challenges encountered throughout the implementation of these objectives, we have still managed to reach our research objectives. However, we still recognise that this project has room to grow. Hence, possible future work which can be done will be described in Section 5.2 below.

# 5 Conclusions

This chapter provides a summary of the whole project. We revisit the aims and objectives which were set up, and discuss how me managed to achieve them in Section 5.1. In Section 5.2 below, we then describe potential future work which can be done to this system and finalise this document in Section 5.3 where we provide our concluding remarks on this project.

## 5.1 Revisiting the Aims and Objectives

In Objective 1 of our research, we aimed to classify the types of crimes by implementing both ML and DL models, ranging from Decision Tree Classifier to Multi-Layer Perceptron neural network. Furthermore, our main goal was to improve on the findings of Sardana et al., and thus, we first replicated the results of [31] and then proceeded to extend on their work. From our conducted implementation, we observed that the MLP neural network model was not that effective when compared to the other ML models, such as the Decision Tree Classifier, because the produced results were very similar. In fact, the Logistical Regression Classifier, Decision Tree Classifier as well as the MLP classifier all produced identical results, which is that of a f1-score of 71%. On the other hand, the k-Nearest Neighbour model performed the worst overall, having a f1-score of 68%.

Through Objective 2, we forecasted the future number of crimes in Manhattan borough, which is achieved through the implementation of ML and DL models, including Random Forest Regression and MLP regressor. Similar to Objective 1, our target was to improve on the findings of Catlett et al. [9]. Hence, we decided to integrate additional data to our dataset, being *population*, *unemployment rate* and *median income* data. Therefore, we perform multiple experiments, and in each experiment, we train our models using different combinations of features. Subsequently, we observed which feature/s is/are impacting positively the predictions and compare our findings with [9]. From the results obtained, we inferred the MLP model as the most accurate model, which produced a Mean Average Percentage Error (MAPE) of 0.74 when using crime and unemployment rate data as training features. Hence, the *unemployment rate* feature is evidently the most important feature out of all of them since it was that external feature which produced the best overall result.

Last but not least, in Objective 3 we investigated the amount of training data required by our classification and regression models. Subsequently, we deduced that 40% of the whole San Francisco (SF) dataset is enough for our classification models to remain effective whilst training data from 2010 to 2013 is also sufficient (half the size of the training set in Objective 2) for our regressor models to make accurate predictions. For the classification models, the Logistical Regression Classifier, Decision Tree Classifier as well as the MLP classifier produced the highest f1-score of 70%, as opposed to the kNN which produced the lowest score of 67%, whilst for the regression-based models, the best MAPE value was 0.79 produced by the MLP model.

## 5.2 Future Work

Although we have managed to solve all objectives previously set up in this document despite of the challenges we have faced, we still believe that there is always room for improvement. As such, for future work, we would perform crime hotspot detection using clustering algorithms in order to identify crime dense regions and predict crime for these regions detected. This work can be further extended by studying the application of hierarchical spatial methods which are able to split clusters when their sizes are huge. Furthermore, for crime rate prediction, we can also study the correlation between the crime trends and other city events in order to identify a relationship amongst them.

On the other hand, for crime type prediction, it could be very interesting to include demographic data, such as victim and suspect race, to observe whether this has any effect of the performances of the models. Finally, further adding advanced state-of-the-art Deep Learning models, for example, Long Short-Term Memory (LSTM) and Recurrent Neural Network (RNN), would be interesting to see if the inclusion of these more complex models can help in predicting crime with higher accurate scores.

## 5.3 Final Remarks

Throughout this whole project, we came across many challenges, particularly when creating models which perform better than the read literature. However, we believe that we have successfully managed the solve this research objectives highlighted in Section 1.4 above. As explained in Section 1.1, providing accurate and reliable predictions can not only help law enforcement agencies to effectively prevent crime from re-happening but can also aid in handling them better when they occur. Hence, through this research, we believe that we have carried on providing further vital contributions to the crime analysis domain.

# References

[1] Adderley, R. and Musgrove, P. B. Data mining case study: Modeling the behavior of offenders who commit serious sexual assaults. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '01, page 215–220, New York, NY, USA, 2001. Association for Computing Machinery. ISBN 158113391X. doi: 10.1145/502512.502541. URL https://doi.org/10.1145/502512.502541.

[2] Almanie, T., Mirza, R., and Lor, E. Crime prediction based on crime types and using spatial and temporal criminal hotspots. *International Journal of Data Mining and Knowledge Management Process*, 5, 08 2015. doi: 10.5121/ijdkp.2015.5401.

[3] Babakura, A., Sulaiman, M. N., and Yusuf, M. A. Improved method of classification algorithms for crime prediction. In *2014 International Symposium on Biometrics and Security Technologies (ISBAST)*, pages 250–255, 2014. doi: 10.1109/ISBAST.2014.7013130.

[4] Barbara, C. Analysis of police violence records through text mining techniques. Undergraduate dissertation, University of Malta, Msida, Malta, 2021.

[5] Bogomolov, A., Lepri, B., Staiano, J., Oliver, N., Pianesi, F., and Pentland, A. Once upon a crime: Towards crime prediction from demographics and mobile data. 09 2014. doi: 10.1145/2663204.2663254.

[6] Boo, Y. L. and Alahakoon, D. Mining multi-modal crime patterns at different levels of granularity using hierarchical clustering. *2008 International Conference on Computational Intelligence for Modelling Control & Automation*, pages 1268–1273, 2008.

[7] Bowers, K., Johnson, S., and Pease, K. Prospective hot-spotting: The future of crime mapping? *British Journal of Criminology*, 44:641–658, 09 2004. doi: 10.1093/bjc/azh036.

[8] Carton, S., Helsby, J., Joseph, K., Mahmud, A., Park, Y., Walsh, J., Cody, C., Patterson, C. E., Haynes, L., and Ghani, R. Identifying police officers at risk of adverse events. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 67–76, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450342322. doi: 10.1145/2939672.2939698. URL https://doi.org/10.1145/2939672.2939698.

[9] Catlett, C., Cesario, E., Talia, D., and Vinci, A. Spatio-temporal crime predictions in smart cities: A data-driven approach and experiments. *Pervasive and Mobile Computing*, 53:62–74, 2019. ISSN 1574-1192. doi: https://doi.org/10.1016/j.pmcj.2019.01.003.

[10] Chainey, S. Examining the influence of cell size and bandwidth size on kernel density estimation crime hotspot maps for predicting spatial patterns of crime. *Bulletin of the Geographical Society of Liege*, 60, 12 2013.

[11] Chainey, S., Tompson, L., and Uhlig, S. The utility of hotspot mapping for predicting spatial patterns of crime. *Security Journal*, 21:4–28, 02 2008. doi: 10.1057/palgrave.sj.8350066.

[12] City and of San Francisco, C. Police department incident reports: Historical 2003 to may 2018. 02 2012. URL https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-Historical-2003/tmnf-yvry.

[13] Elluri, L., Mandalapu, V., and Roy, N. Developing machine learning based predictive models for smart policing. In *2019 IEEE International Conference on Smart Computing (SMARTCOMP)*, pages 198–204, 2019. doi: 10.1109/SMARTCOMP.2019.00053.

[14] Frank, E., Hall, M., Holmes, G., Kirkby, R., Pfahringer, B., Witten, I., and Trigg, L. *Weka-A Machine Learning Workbench for Data Mining*, pages 1269–1277. 07 2010. doi: 10.1007/978-0-387-09823-4_66.

[15] Gera, P. and Vohra, R. City crime profiling using cluster analysis. volume 5, pages 5145–5148. Association for Computing Machinery, 2014.

[16] Gerber, M. S. Predicting crime using twitter and kernel density estimation. *Decision Support Systems*, 61:115–125, 2014. ISSN 0167-9236. doi: https://doi.org/10.1016/j.dss.2014.02.003.

[17] Hu, Y., Wang, F., Guin, C., and Zhu, H. A spatio-temporal kernel density estimation framework for predictive crime hotspot mapping and evaluation. *Applied Geography*, 99:89–97, 2018. ISSN 0143-6228. doi: https://doi.org/10.1016/j.apgeog.2018.08.001.

[18] Huang, C., Zhang, J., Zheng, Y., and Chawla, N. V. Deepcrime: Attentive hierarchical recurrent networks for crime prediction. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, CIKM '18, page 1423–1432, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450360142. doi: 10.1145/3269206.3271793. URL https://doi.org/10.1145/3269206.3271793.

[19] Junior, F. C. F. N., Silva, T. L. C. d., Neto, J. F. d. Q., Macedo, J. A. F. d., and Porcino, W. C. A novel approach to approximate crime hotspots to the road network. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Prediction of Human Mobility*, PredictGIS'19, page 53–61, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450369640. doi: 10.1145/3356995.3364538. URL https://doi.org/10.1145/3356995.3364538.

[20] Kanimozhi, N., Keerthana, N. V., Pavithra, G. S., Ranjitha, G., and Yuvarani, S. Crime type and occurrence prediction using machine learning algorithm. In *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, pages 266–273, 2021. doi: 10.1109/ICAIS50930.2021.9395953.

[21] Keyvanpour, M. R., Javideh, M., and Ebrahimi, M. R. Detecting and investigating crime by means of data mining: a general crime matching framework. *Procedia Computer Science*, 3:872–880, 2011. ISSN 1877-0509. doi: https://doi.org/10.1016/j.procs.2010.12.143. World Conference on Information Technology.

[22] Kim, S., Joshi, P., Kalsi, P. S., and Taheri, P. Crime analysis through machine learning. In *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pages 415–420, 2018. doi: 10.1109/IEMCON.2018.8614828.

[23] Lee, J. Y., Kang, U., Koutra, D., and Faloutsos, C. Fast anomaly detection despite the duplicates. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13 Companion, page 195–196, New York, NY, USA, 2013. Association for Computing Machinery. ISBN 9781450320382. doi: 10.1145/2487788.2487886. URL https://doi.org/10.1145/2487788.2487886.

[24] Llaha, O. Crime analysis and prediction using machine learning. pages 496–501, 09 2020. doi: 10.23919/MIPRO48935.2020.9245120.

[25] Maizura, N. and Ab Hamid, S. H. Visualization of crime data using self-organizing map (som) and improvement in som: A review and available tools. *Journal of Computer Science & Computational Mathematics*, 6:37–43, 06 2016. doi: 10.20967/jcscm.2016.02.003.

[26] Marzan, C. S., Baculo, M. J. C., de Dios Bulos, R., and Ruiz, C. Time series analysis and crime pattern forecasting of city crime data. In *Proceedings of the International Conference on Algorithms, Computing and Systems*, ICACS '17, page 113–118, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450352840. doi: 10.1145/3127942.3127959. URL https://doi.org/10.1145/3127942.3127959.

[27] Pradhan, I., Potika, K., Eirinaki, M., and Potikas, P. Exploratory data analysis and crime prediction for smart cities.

pages 1–9, 06 2019. ISBN 978-1-4503-6249-8. doi: 10.1145/3331076.3331114.

[28] Redmond, M. Communities and Crime. UCI Machine Learning Repository, https://archive-beta.ics.uci.edu/ml/datasets/communities+and+crime, 2009. Accessed: May, 2022.

[29] Rumi, S., Deng, K., and Salim, F. Crime event prediction with dynamic features. *EPJ Data Science*, 7, 12 2018. doi: 10.1140/epjds/s13688-018-0171-7.

[30] Rumi, S. K., Luong, P., and Salim, F. D. Crime rate prediction with region risk and movement patterns. *ArXiv*, abs/1908.02570, 2019.

[31] Sardana, D., Marwaha, S., and Bhatnagar, R. Supervised and unsupervised machine learning methodologies for crime pattern analysis. *International Journal of Artificial Intelligence & Applications*, 12:83–99, 01 2021. doi: 10.5121/ijaia.2021.12106.

[32] Sathyadevan, S., Devan, M., and Gangadharan, S. S. Crime analysis and prediction using data mining. In *2014 First International Conference on Networks Soft Computing (ICNSC2014)*, pages 406–412, Aug 2014. doi: 10.1109/CNSC.2014.6906719.

[33] Stalidis, P., Semertzidis, T., and Daras, P. Examining deep learning architectures for crime classification and prediction. *Forecasting*, 3(4):741–762, 2021. ISSN 2571-9394. doi: 10.3390/forecast3040046.

[34] Wang, H., Kifer, D., Graif, C., and Li, Z. Crime rate inference with big data. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 635–644, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450342322. doi: 10.1145/2939672.2939736. URL https://doi.org/10.1145/2939672.2939736.

[35] Yu, J., Ding, W., Chen, P., and Morabito, M. Crime forecasting using spatio-temporal pattern with ensemble learning. volume 8444, 05 2014. ISBN 978-3-319-06604-2. doi: 10.1007/978-3-319-06605-9_15.

[36] Yuan, J., Zheng, Y., and Xie, X. Discovering regions of different functions in a city using human mobility and pois. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '12, page 186–194, New York, NY, USA, 2012. Association for Computing Machinery. ISBN 9781450314626. doi: 10.1145/2339530.2339561. URL https://doi.org/10.1145/2339530.2339561.

[37] Yuki, J. Q., Sakib, M. M. Q., Zamal, Z., Habibullah, K. M., and Das, A. K. Predicting crime using time and location data. In *Proceedings of the 2019 7th International Conference on Computer and Communications Management*, ICCCM 2019, page 124–128, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450371957. doi: 10.1145/3348445.3348483. URL https://doi.org/10.1145/3348445.3348483.

[38] Zakir Hussain, K., Durairaj, M., and Farzana, G. R. J. Criminal behavior analysis by using data mining techniques. In *IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012)*, pages 656–658, 2012.

[39] Zhao, X. and Tang, J. Modeling temporal-spatial correlations for crime prediction. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, CIKM '17, page 497–506, New York, NY, USA, 2017. Association for Computing Machinery. ISBN 9781450349185. doi: 10.1145/3132847.3133024. URL https://doi.org/10.1145/3132847.3133024.

[40] Zhao, X. and Tang, J. Crime in urban areas: A data mining perspective. *SIGKDD Explor. Newsl.*, 20(1):1–12, may 2018. ISSN 1931-0145. doi: 10.1145/3229329.3229331.

# A Implementation Details

Our artefact is developed using Jupyter Notebooks and Python 3.8 as a programming language. All models are built and trained on an 8gb core i5 MacBook Pro. The main libraries used throughout the implementation of this dissertation can be found in Table A.1 below.

| Library Name | Usage |
|---|---|
| numpy[1] | Used to prepare data for input into algorithms and for plotting purposes |
| pandas[2] | Used to read data from CSV files as well as to generate dataframes and pre-process them |
| sci-kit[3] | Used to create Machine Learning and Deep Learning models for every objective in this project |
| matplotlib[4] | Used to generate any charts and plots |

Table A.1: Libraries used in our implementation

The source code for this project can be found in the following GitHub Link.

[1] `https://github.com/numpy/numpy`; [Last Accessed: 19th May 2022]
[2] `https://github.com/pandas-dev/pandas`; [Last Accessed: 19th May 2022]
[3] `https://github.com/scikit-learn/scikit-learn`; [Last Accessed: 19th May 2022]
[4] `https://github.com/matplotlib/matplotlib`; [Last Accessed: 19th May 2022]