Spatiotemporal forecasts of London's crime hotspots

MSc Research Project
MSc. Data Analytics

Andrew McMorrow
Student ID: 20216220

School of Computing
National College of Ireland

Supervisor: Jorge Basilio

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Andrew McMorrow
Student ID:	20216220
Programme:	MSc. Data Analytics
Year:	2022
Module:	MSc Research Project
Supervisor:	Jorge Basilio
Submission Due Date:	19/09/2022
Project Title:	Spatiotemporal forecasts of London's crime hotspots
Word Count:	10462
Page Count:	27

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	16th September 2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Spatiotemporal forecasts of London's crime hotspots

Andrew McMorrow
20216220

Abstract

This study investigates and predicts future crime hotspots of London via spatiotemporal modelling. A model was sought which provided ethical, unbiased and accurate predictions of forecasted crime hotspots in London, with the widespread societal benefits of reduced crime rates serving as motivation. This was successfully achieved via construction of the Extreme Gradient Boosting (XGB) model to predict future crime rates, before crime hotspot forecasts over the period of one year were made on these via the use of quantile statistics. This model design was deemed the most appropriate in terms of the aforementioned requirements due to its added interpretability and favourable results achieved in comparison to literature. The data used consisted of crime data in the Greater London Area (GLA) from 2012-2021, along with relevant spatial such as the number of food outlets, nightlife spots, entertainment venues etc. in a given area. Crime data in the GLA from 2012-2021 was used for this modelling, along with relevant spatial such as the number of food outlets, nightlife spots etc. in a given area. The GLA was partitioned into equi-sized grid cells, with the data per cell aggregated over a given year into each cell before implementation into the model. The *Prediction Accuracy Index* (PAI) (formulated by Chainey et al. (2008)) was deemed the most appropriate performance metric to base results upon. Predictions for hotspots were made on unseen data for both before and during the COVID-19 pandemic, with the most appropriate model returning PAI scores of 2.84 and 2.97. This offered an improvement of 11%/22% on pre/mid-pandemic data over the highest performing model which did not incorporate the data on spatial features.

This consistent performance despite wholesale changes in crime rates further indicates that this model is fit for use to accurately and ethically predict future crime hotspots in London.

1 Introduction

1.1 Motivation

Crime is an issue which is extremely prevalent in London, and indeed throughout the world today, leading to a detrimental effect on societies. The Guidelines for the Prevention of Crime ECOSOC Resolution 2002/13 as set out in *Action to promote effective crime prevention* (n.d.), describes crime prevention as measures which seek a reduction in the risk of crimes occurring and their potential harmful effects on individuals and society. Reports such as *Action to promote effective crime prevention* (n.d.) illustrate that well-planned crime prevention strategies and successful identification of the underlying causes not only prevent crime and victimization but also promotes community safety and contributes to the sustainable development of countries. In addition to the above, the reduction of crime has further benefits to a society with a lower cost on taxpayers who usually bear the financial burden of maintaining increased police personnel and operations, courts, jails and prisons directed towards these crimes and their perpetrators. Hence the reduction of crime can bring wide-ranging benefits to a society.

A wide range of literature [Owusu and Frimpong (2020)] has shown that crime events are not random in space distribution, with their distribution tending to concentrate in geographic clusters (often referred to in literature as hotspots) at some points, while appearing sparse in others. This feature of crime events distribution was described as an 'inherent geographical quality' by [Chainey and Ratcliffe (2005)].

The ability of public safety agencies to accurately identify and clearly visualize crime hotspots along with understanding their relation to crime related factors can bring significant benefit to crime prevention by providing a solid basis for threat visualization, police resource allocation and crime prediction [Wang et al. (2013)]

Due to the above, the proposed scope of this project is to utilize artificial intelligence (AI) and suitable machine learning (ML) techniques to provide long-term predictions of crime at micro places relative to other popular techniques. This will implement an area of data analytics known as spatiotemporal predictive modelling (STPM) (also known as predictive policing *PP*). Spatiotemporal data mining is defined as the extraction of latent knowledge, relationships or patterns from spatial-temporal datasets [Yuan et al. (2004)]. This field can be extended to Exploratory Spatiotemporal Data Mining (ESTDM), which is the extraction of relationships, patterns or latent knowledge through analysis. This is intrinsically a challenging task due to the complex nature of the associated data, data types and structures accompany a lack of efficient methods for its exploration [Yuan et al. (2004)]

Although the implementation of AI in law enforcement has been shown to make law enforcement less time-consuming, less prone to human errors and more cost-effective [Raaijmakers (2019)], it has not been without its critics with many raising the argument of how ethical it is. Jenkins and Purves (2020) argue that it has the power to stigmatise and discriminate when used for spatial analysis to predict at-risk areas.

Due to the aforementioned, the central question of this paper is ***"Can spatiotemporal predictive modelling be utilized to accurately and ethically predict crime hotspots in London?"*** with historical, publicly available data sets utilized.

The aforementioned benefits throughout society from the reduction of crime serve as motivation behind this study. As predictive policing is a relatively new phenomenon in the UK, room for improvement exists on current models in both accuracy and ethical procedure. Hence this paper serves to provide new and thought-provoking insights into the ability of STPM to predict areas of future crime hotspots across London, with results of various STPM methods compared upon both their performance and ethicality. This report will proceed by providing details of related work, before describing details of the model's provided data sets, architecture, design and development before evaluating and discussing it's results.

2 Related Work

2.1 Spatiotemporal analysis

Spatiotemporal analysis describes analysis of data collected across both space and time. According to Ibrahim et al. (2019), spatiotemporal crime analysis is *"an approach to analyze and identify different crime patterns, relations and trends in crime with identification of highly concentrated crime areas"*. According to [Chainey and Ratcliffe (2005)], a hotspot is an area that has a greater than average number of criminal events. Crime hotspot identification techniques aim to identify spatial and/or temporal clusters of an event with the clusters varying depending on both the geographic and temporal scales.

Crime hotspot mapping (HM) is a practice which has increased in popularity in recent years both with academics and crime prevention practitioners leading to some law enforcement agencies adopting these practices on behalf of their country [Maguire (2000)], [Ratcliffe (2002)]. One hypothesis for this increase is attributed by [Fan (2014)] to optimizing the allocation of limited resources/manpower to areas where crime is more likely to occur. In parallel to this, academic research and interest in HM theories and techniques have seen a similarly drastic increase, with theories ranging from routine activiees and repeat victimization to the social ecology of crime [Anselin et al. (2000)].

2.2 Traditional methods for analyzing crime patterns

Traditionally regression models have often been used to explore the relationships between crime and influential factors. Studies by [Sampson and Groves (1989)], [Stucky and Ottensmann (2009)], [Liu et al. (2016)] suggest that the most common factors are represented by socioeconomic, demographical and land use characteristics across different analytical units. Often data on mapping problems such as that proposed is available for small areas (neighbourhoods for example), which can result in biased parameter estimates from autocorrelation effects if not considered [Law and Chan (2012)]. Additionally, traditional regression models have been shown to be incompetent of addressing the small number problem, arising when unstable estimates occur due to low counts of events and high sampling variation [Law et al. (2014)].

2.2.1 Common spatiotemporal modelling methods

When spatiotemporal data is broken down to it's spatial component, Koperski and Han (1995) describes the mining of this element's data as a knowledge discovery technique used for the "extraction of implicit knowledge, spatial relations, or other patterns not explicitly stored in spatial databases". ST mapping methods have experienced increased development alongside that of the innovations in computational power. Some spatial techniques focus on subdivisions inside a studied area (e.g. census areas, electoral areas, villages, boroughs etc). When a spatial division such as the above is used, thematic mapping is the simplest mapping method available (this method can be used upon any spatial arrangement). However one issue which can occur when using spatial subdivisions such as electoral areas (or others as described above) is that these subdivisions often have different spatial arrangements. The use of differing spatial arrangements result in HM results in maps which differ from each other due to a difference in area meaning that measurements such as density can be deceptive. This area of statistical bias is known as the modifiable areal unit problem ($MAUP$) and cannot be ignored when HM techniques are applied to areas of differing spatial arrangements [Chainey et al. (2008)].

One popular technique to avoid MAUP is to utilize a grid which partitions the studied area into equisized sections of the same shape. An example of a method used to produce such grid maps is grid mapping. Values for the dependent and independant variable(s) can thus be assigned to this area, with areal density values calculated without any shape-related bias.

There exists a plethora of previous studies into the domain of geospatial mapping methods with an abundance of methods being utilized.]Hirschfield and Bowers (2001)] have argued that these methods can be partitioned into three main areas: *discrete point mapping* (PM), *choropleth mapping* (CM), and *kernal density estimation* (KDE). One of the initial methods on the scene of hotspot mapping is the Spatial and Temporal Analysis of Crime (STAC) program [Bates (1987)]. This method was based on point mapping, displaying crime hotspots using "standard deviational ellipses" without pre-defining any spatial boundaries. Criticisms exist for this work however [Eck et al. (2005)] with indications that STAC may mislead as they do not naturally follow the ellipses shape. Furthermore, this process can prove computationally intensive, and difficult to aggregate features.

According to [Hirschfield and Bowers (2001)] CM uses defined geographic boundaries (such as electoral areas, etc.) as it's basic mapping elements with data aggregated and hotspots identities restricted to the shape of the area in question. However this technique once more faces the statistical bias of MAUP which must be taken into consideration.

Grid mapping The utilization of grid mapping has many advantages, particularly with combatting problems associated with different sizes and shapes in a geographical region. A uniform grid map can be constructed over the studied area, and thematically shaded to indicate the levels of studied activity within a given grid. The utilization of equi-sized grids negates the aforementioned MAUP problem faced by aforementioned methods. Further advantages of grid mapping are that it can be utilized to aggregate data into defined boundaries, allowing for insights to be drawn into the impact of features on the levels of a studied activity within an area. This grid mapping technique can be utilized with a variety of ML models, along with implementation with KDE.

Critiques and limitations of this method include the impact of the usage of grid size and the restriction on how the hotspot can be displayed. Spatial detail within and across each grid is correspondingly lost, as the studied activity must conform to a specific grid Chainey et al. (2008). Chainey and Ratcliffe (2005) Eck et al. (2005) have commented on the "blocky" appearance of the resultant maps of these techniques. This can be countered by reducing the size of each cell, however care must be taken with this approach to ensure resolution or information is not lost. Guidance exist for this cell selection size, with [Gorr et al. (2003)] deeming a grid size of 370m^2 (4,000 square foot, 0.00037km^2) to be the smallest viable grid size in order to provide accurate forecasts. [Chainey et al. (2008)] builds upon this guidance, with further recommendations that the grid cell size should be approximately the longest distance in the studied area, divided by 50.

Kernal Density Estimation The KDE [Wand and Jones (1994)] method imposes a regular grid onto a studied area utilizing a three-dimensional kernal function to calculate a density value assigned to each cell [Eck et al. (2005)]. The point data is aggregated into an area of a user-specified radius/size, generating a continuous surface to represent the density of points. This method can be advantageous in overcoming the limitation of geometric shapes. Comparison work undertaken by several researchers have lead them to view KDE as the method most suitable for both visualization [Chainey and Ratcliffe (2005)] and accurate prediction of future crime incidents [Chainey et al. (2008)]. However this study is not without criticism. The inequality of technique comparison was highlighted by [Pezzuchi (2008)]. This was as the quantile divisions of results do not allow the same statistical precision, hence lacking standardisation between techniques. Further criticism was attached to the choice of techniques compared, with well performing techniques such as Nearest Neighbour Hierarchical (NNH) clustering, absent from the study [Levine (2004)],[Swain (2012)]. Due to this criticism, other models shall also be explored in tandem with KDE.

2.3 Hotspot definitions

When dealing with hotspot mapping, techniques which utilize ellipses or similar (such as convex hulls) define hotspots directly. However when dealing with geographical data such as polygons, continuous surfaces or grid techniques, the threshold for a hotspot (known as a critical factor) must be pre-defined.

Although crime hotspots have increased in popularity within academia and the crime analyst discipline, surprisingly there remains no established benchmark for defining a hot spot [Harries (1999)]. [Wu et al. (2020)] defines a hotspot as a region which has a greater concentration of incidences compared to the expected number of random distribution, these can be used to indicate the form of clustering in spatial distribution. One main differential between density and hotspots are that although both analyse the cluster, the density is not a statistically significant figure, whereas a hotspot utilizes features to identify statistically significant locations in a data set.

Figure (4) shows a map of crime density throughout the studied period, as can be visualized here minimal useful information can be drawn from such results.

Current practices recommended by [Chainey et al. (2008)] include the use of the top class of a five quantile classification system in order to classify hotspots (quantiles classification). This has the benefits of resulting in a visually balanced map [Monmonier (1996)], however criticisms of this method argue that the quantiles method sets arbitrary value ranges, thereby ignoring the significance of the spatial distribution. Furthermore, as this threshold value can vary significantly between techniques it prevents the fair comparison of hotspot mapping techniques.

Arguments for other approaches exist, with [Swain (2012)] arguing that "If the conceptual definition of a hotspot is a region containing a significantly high density of points, then the classification of a hotspot should be from the results of clustering analysis and be by a statistical significance threshold." This is in tandem with many other reports who utilize *Local Moran*'s statistics for this purpose. These statistics are described further in section (3.4.2).

2.4 Temporal features

As illustrated by the routine activities and crime pattern theories studied by [Cohen and Felson (1979)] and [Brantingham et al. (1981)] respectively, the identification of any crime pattern is reliant on both space and time. The impact of spatial features on the pattern of crime has drawn continuous attention from researchers in recent years [Andresen (2005)], [Chainey and Ratcliffe (2005)], [Weisburd (2015)]. However analysis of the temporal feature of crime patterns has not kept pace with the expansion of it's spatial analysis [Newton and Felson (2015)] & [Ratcliffe (2002)], with studies considering both space and time rarer still [Andresen and Malleson (2015)]. This area has seen an increased number of studies over the past decade, with integrations of both space and time leading researchers to find that different types of crime patterns vary by hour, day, month, season and year [Liu et al. (2021)] [Linning (2015)], [Andresen and Malleson (2015)], [Ceccato and Uittenbogaard (2014)], [Andresen and Malleson (2013)].

That being said, a great deal of study has been done in the area of long term crime hotspot stability. Studies such as that taken by [Shaw and McKay (1970)] on crime in Chicago found that it's crime pattern remained relatively stable. More recently, an interesting study on the topic of trajectory analysis was introduced by [Weisburd (2015)], investigated the crime pattern changes over a 14-year time period in Seattle by grouping the street segments according to their developmental trajectories, with this entire period yielding a high degree of stability. An alternative method of analysing the changing patterns of crime was proposed by [Andresen (2009)]. This was called the spatial point pattern test (SPPT) and was designed for studies of crime over a long-term period. The results of this test allows researchers to investigate the stability and changes in spatial patterns of crimes over the years [Hodgkinson et al. (2016)]. However, the SPPT is only suitable for measuring the similarity of two datasets at the area scale and is incapable of accommodating covariates. [Liu et al. (2021)].

The above studies of crime stability are in line with one of the most popular assumptions employed by police departments in crime forecasting - that hotspots of today are the hotspots of tomorrow Groff and La Vigne (2002). Surprisingly little research is available which tests this theory, however the few studies which have been identified such as that undertaken by Spelman (1995) suggest that this approach's effectiveness is dependant on the time period employed. Their study indicated that one month of data provides hardly further indication of hotspot locations than chance, however one year of data could predict these hotspots with up to 90% accuracy [1]. This is in line with research by [Walczak (2021a)]

[1] Note - This assessment was based upon hotspot identification at specific location types (schools, public transport stations etc.) rather than all potential hotspots distributed across a studied area

and [Zhu et al. (2019)], who has shown that there is "no basis for short term crime forecasting due to evidence from data showing that spatial heterogeneity and time lag cannot accurately be reflected in short-term prediction".

2.5 Summarize Lit review

As explained throughout this section, the topic of spatial crime analysis is one which has seen increased growth throughout the last number of decades and will continue to do so in tandem with the continued growth of computational power.

However research into spatiotemporal crime analysis still remains limited, with debates persisting on the most suitable techniques to use for both crime mapping and modelling. Furthermore, the lack of established hotspot definition benchmark means that discussions on suitable definitions will continue.

3 Methodology

3.1 High level overview

For improved clarity, a high level overview of the proposed solution is illustrated in figure (1).

3.2 Data selection, pre-processing and feature engineering

The crime data sets were obtained from the *UK Police open Data website*. The data provided is made available for public use under the Open Government License. The UK police open data contains information on crime and police data across the UK hence these terms are used interchangeably to describe this data set throughout this study. The data set on the spatial features of an area were obtained through the utilization of the *Foursquare API caller*. *Foursquare* members have the ability to use data obtained through their API for studies such as the undertaken provided crediting is given on any imagery produced.

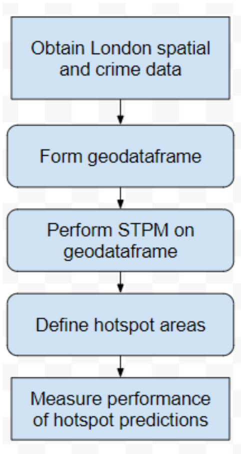


Figure 1: High level overview of proposed solution

3.2.1 Spatial data

The spatial data set pulled from FourSquareAPI contains information on the names of nearby features within a pre-determined radius of a given point. These nearby features include the individual names of facilities such as - entertainment, colleges, food venues, nightlife, shops, transport, outdoor recreation, professional work places and residences. The data available on this API is updated on a monthly basis.

3.2.2 Crime data

The crime data set described in section (3) provides information on the monthly outcomes of arrests, reported crimes, along with daily updates of "stop and search" exercises undertaken by police officers. This report shall focus on the "street crimes" which shall be referred to as "crime data" throughout the remainder of this report. This crime data included information on the latitude/longitude coordinates of where a crime occurred, the crime type along with the year and month in which the crime occurred.

For the purposes of anonymity to crime data, the latitude and longitude locations of crime is anonymised to represent only the approximate location of a crime - not it's exact location. This anonymization process includes placing a map point over the center point of a street, or above public places such as a park or a commercial premise such as a shopping center. It is therefore acknowledged that the spatial accuracy of this point data is reduced. This may have a small impact on utilizing distinct point mapping, however the impact of these small location changes to analysis performed on aggregated point data is thought to be negligible.

3.3 Ethical concerns

There exist exist certain factors which may effect the spatial distribution of crime, e.g. areas with a larger population density are more likely to have a higher crime count. Research by [Chainey and

Ratcliffe (2005)] have illustrated that areas of high population density which have a large concentration of theft crimes leads to motivated crime offenders being more likely to find potential crime victims. Further studies analysing the variables associated with crime have found relationships with poverty and socio-economic variables such as unemployment [Freeman (1999)], race [Braithwaite (1989)] and level of education [Ehrlich (1975)].

However, there exists ethical concerns over incorporating these aforementioned socio-economical factors. For example, a report released by Stop watch coalition in collaboration with London School of Economics [2] found that Black Britons are now "nine times more likely to be stopped and searched for drugs than white people". This is despite data quoted in the same report that black britons were using the illegal substances at a lower rate. This shows the negative effect PP model's can have on society if biased predictors and results are not taken into account.

As ML depends on the quality, size and objectivity of training data, if biased data is presented for training it will result in biased predictions. According to US department of Justice figures [3], a black person is twice as likely to be arrested than a white person, and five times as likely to be stopped without cause as a white person. Including such a biased variable such as race into any ML training data will result in racially biased results. It is illegal in countries such as the US to use race as a predictor for PP model's [4] however other socio-economic variables such as education and unemployment levels can act as proxies for race. For example, according to the UK government's ethnicity facts and figures, BAME people are nearly twice as likely to be unemployed as white people (7% vs 4%). Hence utilizing data on these aforementioned socio-economic variables will likely lead to biased results. Due to the above, great care shall be taken with the selection of variables to be used to minimize the bias in the model proposed. Hence, the aforementioned socio-economic variables shall not be considered.

3.4 Modelling and feature discussion

3.4.1 Hotspot mapping techniques

Due to the anonymization feature employed to protect the crime data (described in section (3.2.2), it was decided that only mapping methods which utilized aggregated data would be considered and not point data. These aggregated methods are described below.

Grid mapping As described in section (2.2.1), grid mapping can reliably produce forecasts into crime hotspots. Upon consideration of numerous previous studies (in particular a comparison study by [Swain (2012)]), along with the desire to avoid statistical bias during density calculations such as the MAUP discussed in section 2.2.1 it was deemed deemed that grid thematic mapping was an appropriate for the data in question. Hence, the GLA was divided into square grid cells. Following the guidance offered in the grid mapping section of (2.2), a grid size of approximately 1km^2 in area was chosen, resulting in 1,583 grid squares. This grid size was chosen as it will allow granular data to be aggregated over the GLA, whilst also following the aforementioned recommendations.

Thematic mapping of spatial subdivisions Although grid mapping will aide in resolving the MAUP of variable areal size problem, one issue is the extreme difficulty in obtaining the population of each of these cells. One solution to obtain relatively distributed population data is to utilize *Output areas (OAs)*. This mapping technique is used on other studies of English areas such as [Chainey et al. (2008)]. These have an average population of 130 residents (Lower threshold of 100/40 residents/households, upper threshold of 625/250 residents/households). [5]. As the areas will vary in size, the MAUP will need to be taken into consideration when implementing this mapping method. Utilizing numerous mapping methods will allow for comparisons to be made to their relative performance.

3.4.2 Hotspot definition

As described in section (2.3), there exist no established benchmark for the definition of a hotspot, hence multiple definitions shall be considered below.

[2] https://www.stop-watch.org/what-we-do/research/the-colour-of-injustice-race-drugs-and-law-enforcement-in-england-and-wales/

[3] (https://www.ojjdp.gov/ojstatbb/crime/ucr.asp?table_in=2)

[4] https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/

[5] https://ocsi.uk/2019/03/18/lsoas-leps-and-lookups-a-beginners-guide-to-statistical-geographies/

Gi* statistics It is often preferable to study patterns with data at a more local scale. Amongst others, [Swain (2012)] has cited arguments for local hotspots (areas with a relatively high crime rate compared to their surroundings) needing consideration along with global hotspots. The *Local Moran's* value (*Gi**) is recommended for this approach. This is a local spatial autocorrelation statistic based upon the *Moran's* statistic. The *Gi** can enable to detection of pockets of spatial association which may not be evident when using global statistics. First introduced by Getis and Ord (1992), it is a local indicator of spatial autocorrelation (LISA) statistics. For the purpose of this study, LISA stats can determine whether the strength/rate in which values are concentrated is unusually high/low, relative to what would be expected from pure chance. In this case, it can identify cases in which the value of an observation and the average of it's surroundings is either more similar or dissimilar than we would be expected from pure chance. This local Moran's I is the application of the global Moran's I to each observation. This is represented mathematically as:

$$I_i = \frac{z_i}{m_2} \sum_j w_{ij} z_j ; m_2 = \frac{\sum i z_i^2}{n} \tag{1}$$

where m_2 is the variance of the distribution of values in the data, $z_i = y_i - \tilde{y}$, $w_{i,j}$ is the spatial weight for the pair of observations i and j and n is the number of observations. Upon obtaining the local Moran's I value for each cell in question, LISA statistics shall be applied in order to obtain those which are spatially statistically significant. Each statistically significant grid should be classified according to the following criteria, with the remaining being classified as "not statistically significant":

- **HH** - Areas of unusually high concentration surrounded by other areas of unusually high concentration.
- **LL** - Areas of unusually low concentration surrounded by other areas of unusually low concentration.
- **HL** - Areas of unusually high concentration surrounded by areas of unusually low concentration.
- **LH** - Areas of unusually low concentration surrounded by areas of unusually high concentrations.

Furthermore, by utilizing equi-sized grid mappings along with Moran statistics for this study, it was possible to negate the effect of population density/distribution, as this study placed an equal emphasis on local hotspots to global hotspots.

This Gi* test also requires a threshold distance which is defined as the minimum radius required by a cell's centroid to encapsulate all of it's adjacent centroids. In the grid size of 1km^2 proposed, this would result in a threshold distance of 1.42km^2.

These Gi* statistics will be used to classify hotspots with two distinct tests - Two distinct hotspot definitions utilizing Gi* statistics will be tested in order to classify the hotspots -

- Classifying any grids deemed to be "hot" by the LISA statistics (areas described as either HH or HL) as hotspots. This will be referred to as "standard Gi* statistics".
- Classifying only grids deemed hot, surrounded by other areas deemed hot (areas described as HH only) as hotspots. This will be referred to as "Gi* HH only statistics".

Quantile classification The quantile classification method recommended by Chainey et al. (2008) will also be utilized due to it's favourable results historically. Arguments for it's inclusion include allowing further comparison methods for our models to a larger scale of studies, along with it's relatively simplistic implementation relative to Gi* statistics. This method will split the predicted crime rates into 5 equi-sized bins, with areas featuring crime rates in the top tier being considered as "hotspots".

Due to the aforementioned arguments, results for both *Gi** statistics and quantile classification will be formulated. These hotspot classification methods shall be referred to as *G*, *G HH* and *bins* throughout this report.

3.5 Aims of the study

This study will focus on forecasting crime hotspots, as this has ample studies and results available for comparison. The research undertaken by Groff and La Vigne (2002), Spelman (1995), Walczak (2021a) and Zhu et al. (2019) (described in section (2.4)) detailing that there is "no basis for short term crime forecasting due to evidence from data showing that spatial heterogeneity and time lag cannot accurately be reflected in short-term prediction" were influential in the author's decision to focus on a long term forecasting approach, with the aim of the study being to predict hotspots over an annual period, rather than focusing on shorter-term predictions. Due to the above, it is the author's aim to successfully generate models which have the power to accurately and ethically predict crime hotspots over the period of one year for a variety of crime types.

3.5.1 Impact of crime pattern changes on results

Differing police operations and crime reduction initiatives may potentially have been in place throughout the studied time period of 2012-2021 which impacted on crime levels. However for this study we are comparing analytical techniques over the same data, hence any changes in crime patterns would similarly be applied to each technique in question. This means internal comparisons can be made amongst the techniques used on the data studied seamlessly, however external influences such as those noted should be taken into account during comparisons with results on external data sets. It is likely that the advent of COVID-19 and temporary lockdown laws in 2020 will have a significant impact on results and should be taken into account before external comparisons are drawn on data encompassing this time period.

3.6 Machine learning techniques

Suitable ML models are required to be trained on the data for each crime mapping methods described in section (2.2.1) to produce crime rate forecasts. Hotspot predictions will then be made on this forecasted data based on the hotspot definitions defined in section (3.4.2). Due to this requirement, regression models were deemed the most appropriate to obtain this continuous variable output.

Research into the comparative performance of regression models was undertaken, with notable studies such as (Fernández-Delgado et al. (2019)) proving influential to the author. Both traditional and more recently developed (also commonly referred to as "deep regression") regression methods were considered.

3.6.1 Traditional methods

Linear regression As regression models are being considered, it is appropriate to consider perhaps the most basic and commonly used type of predictive analysis - linear regression (LR). Three common uses and strengths of LR are determining the strength of predictors, forecasting an effect and trend forecasting [6]. This trait of trend forecasting in particular is of benefit to the application in predicting crime trends, with further benefits such as the interpretability of the model being of particular benefit in preventing model's used being considered "black boxes".

Random forest Random forest (RF) regression is a supervised learning algorithm which uses ensemble learning method for regression. Ensemble learning method is a technique which combines predictions from multiple ML algorithms to make a more accurate prediction than that of a single model. Attributes which make RF attractive to this application are it's ability to use averaging to improve predictive accuracy and control over-fitting [7].

extraTrees The extraTrees (ET) regressor model is an ensemble of extremely randomized regressor trees. Described as a class which "implements a meta estimator that fits a number of randomized decision trees (a.k.a. extra-trees) on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting"[8]. Although similar to RFs, there are two main differences to these models which is why both are considered. RF uses bootstrap replicas, subsampling the input data with replacement whilst ET uses the whole original sample. Additionally the models differ in their selection of cut points in order to split nodes. RF chooses the optimum split, whilst ET chooses it randomly. Upon splitting, both algorithms choose the best of all feature subsets, hence ET adds randomization whilst still retaining optimization [9]. ET also saves time computationally. During comparison studies by Fernández-Delgado et al. (2019), ET achieved the highest R^2 value of the 77 models studied.

[6] https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/what-is-linear-regression/
[7] https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html
[8] https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.ExtraTreesRegressor.html
[9] https://quantdare.com/what-is-the-difference-between-extra-trees-and-random-forest/

Extreme Gradient Boosting Extreme Gradient Boosting (XGB) is an efficient implementation of gradient boosting which can XGB builds upon RF, although in this case each successive tree planted is weighted in such a way as to compensate for any weakness (residual errors) in the previous tree. This is what give's it it's "boosting" characteristic. It is both highly computationally efficient and effective, with it's two main attributes being execution speed and model performance. It dominates regression predictive modelling on structured or tabular data sets on competitive data science platforms such as *Kaggle*[10]. XGB generally fits training data to an improved standard than linear regression, however this can also lead to overfitting. Further criticisms of this method are it's lack of interpretability versus other regression methods such as linear regression.

Support Vector Machine Regressor Support Vector Machines (SVM)s are supervised learning models with associated learning algorithms for regression analysis. It works by finding a hyperplane in an n-dimensional space which distinctly classifies the data points. The points on either side of the hyperplane closest to this plane are known as support vectors [11]. It possesses several advantages which makes it's implementation attractive for this study. These are it's robustness to outliers, it's easily updated decision model, it's impressive generalization capability along with it's relatively straightforward implementation.

K Nearest Neighbours K Nearest Neighbours (KNN) is a clustering learning method which can be utilized for both classification and regression problems. The algorithm works by retrieving the K datapoints that are nearest in distance to the original point. It uses the *feature similarity* of items in the training data to predict the values of new data points. It is used in supervised learning settings where the target variable is a continuous variable. Advantages of using KNN in this application are [12] -

- It can learn non-linear decision boundaries.
- No explicit training step required - this is as all of the work happens during the prediction.
- Constantly evolves with new data As there exists no explicit training step, the prediction can be adjusted as new data is added to the dataset, without the need to retrain the model.
- Single hyperparameter - As the only hyperparameter is the value of K, this simplifies the tuning.

The KNN model possesses several disadvantages also however. These include the high prediction complexity for large datasets, it's assumption of equal importance to all features and it's sensitivity to outliers. This sensitivity to outliers can be minimized by scaling the input data of both the training and test data sets before training and running predictions.

3.6.2 Deep learning

Deep learning (DL) is a machine learning (ML) methodology which excel at classification and prediction problems [Walczak (2021b)], and are highly capable of solving complex problems [Hornik (1991)]. It is a subfield of ML which utilizes a set of neurons organized in layers. A DL model consists of three layers - the input layer, the hidden layers and the output layer. Advantages of utilizing DL models over ML algorithms are that they can create new features from a limited set of information and perform advanced analysis, learning more complex features than their ML counterparts. Although possessing many advantages, challenges include the need for a large data set in order for successful training to occur.

Research conducted by [Walczak (2021b)] into the usage of NNs in previous studies have indicated that approximately 52% (36 of the 69 NN crime studies reviewed) utilize NNs to predict crime hotspots. Various types of NNs are used, with Recurrent NNs (RNNs) and Convolutional NNs (CNNs) making up the bulk of the study. Other popular usage of NNs in crime analysis is in data mining, mining large quantities of police data in particular. This usage of NNs will not be relevant for our study in question however. As mentioned previously, this aspect of the study will be a regression problem, only those NNs which excel at regression were considered.

[10] https://machinelearningmastery.com/xgboost-for-regression/
[11] https://towardsdatascience.com/unlocking-the-true-power-of-support-vector-regression-847fd123a4a0
[12] https://machinelearninginterview.com/topics/machine-learning/how-does-knn-algorithm-work-what-are-the-advantages-and-disadvantages-of-knn/

NN configuration [Walczak and Cerpa (1999)] provide research guidelines when establishing architecture of NN systems. These include using one to two hidden layers along with varation in the number of NN nodes within each hidden layer. These guidelines further indicate that the number of nodes in a given hidden layer should start at half of that number in the previous layer. These should be increased incrementally until performance begins to decrease [Walczak (2021b)]. This technique can aide in preventing over-fitting. Regarding the node numbers of hidden layers - this is recommended to use an incremental increase of two nodes at a time for the first hidden layer. If a second hidden layer is used, this layer should be increased a single node at a time.

Popular NN training methods Backpropagation (BP) is the most common NN training method hence it's utilization enables comparable cross examinations with many similar studies. Whilst fitting an NN, BP efficiently computes the gradient of the loss unfciton with respect to the weights of the network. The high efficiency of this approach makes it feasible for using other gradient methods in the training of multilayer networks [13]. Other training methods considered include radial basis functions (RBFs). This has been shown to outperform BP when data extrapolation is required, or limited data is available [Walczak and Cerpa (1999)]. As sufficient data is available for the task in question, BP shall be considered to be the primary method used to train NNs.

Activation function An activation function/layer (AF) in a NN defines how the weighted sum of the inputs is transformed into an output from the nodes in the network's previous layer. AFs are used for both the hidden layer and output layer. Non-linear AFs are often preferred for the hidden layer in order to gain access to a much richer hypothesis space that would benefit from deep representations [Chollet (2021)]. It is most unusual to vary the AF through the hidden layers of a model, with the most common hidden layer AFs being $ReLU$, *Tanh* and *sigmoid* AFs. Both the sigmoid and tanh functions are known to make the model more susceptible to the vanishing gradients problem during training, hence the $ReLU$ AF is the default recommendation for modern NN [Goodfellow et al. (2017)]. The application at hand dictates the choice of AF for the output layer. As this is a regression task, the linear AF is recommended for this use case.

Network optimization Optimizers for training NNs are responsible for finding the weights of a cost function (such as a performance measure, also known as a loss function). This allows the quality of a set of weights to be assessed. Typically in regression functions, the Mean Squared Error (MSE) is the default loss used due to it's favourable results. Mean Absolute Error (MAE) is deemed to be the appropriate loss function in the case of the target variable possessing gaussian distribution. As our model displays behaviour which can be considered gaussian, both MAE and MSE will be considered as our loss functions.

Artificial Neural Networks Artificial Neural Networks (ANN) are one of the deep learning (DL) algorithms which simulate the workings of neurons in the human brain. The attractiveness of ANN for regression is it's ability to learn the complex non-linear relationship between the features and target variable due to the presence of activation function in each layer.There exist many types of ANN such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN). Vanilla Vanilla Neural Networks (VNN) have the ability to handle structured data only, which is of benefit to this application. The vanilla (standard) configuration will be utilized and referred to as "ANN" throughout this task.

3.7 Kernel Density Estimation

Due to the favourable results achieved by KDE in studies cited in section (2.2.1), KDE was also considered as a model to predict hotspots. KDE is a model which is growing in popularity for applications such as this study. This is due to the aesthetic nature of the resultant hotspot map, along with it's high performance.

Point data are aggregated within a user-specified area (in this case grid cells and OAs), before a curve is distribution is created of this data. This curve is calculated by weighing the distance of all the points in each specific location along the distribution. If there are more points grouped locally, the estimation is higher as the probability of seeing a point at that location increases [14]. This results in

[13]https://en.wikipedia.org/wiki/Backpropagation
[14]https://deepai.org/machine-learning-glossary-and-terms/kernel-density-estimation

the creation of a smooth surface map, illustrating the variation of crime density across the studied area. Two parameters of cell size and bandwidth are available for user variation. The cell size chosen here will be the pre-determined grid cell and OA size, hence the KDE will use these pre-aggregated data sets. The default bandwidth value is optimal for normal distribution. Analysts who are not experts in spatial analysis are encouraged to utilize the default values for the bandwidth parameter Chainey et al. (2008).

4 Design Specification

4.1 Data storage and architecture

This study was performed in Google Colab [15]. This web IDE for python enables ML with cloud storage. This was chosen for many reasons, with the collusion with google cloud providing additional security in terms of data loss prevention, as well as the increased data storage ability relative to the author's local storage influential in the decision. The data sets can be easily pulled via a mount with google drive, with a total storage size of 166GB, and a RAM size of 25GB allowing for a high computational processing ability.

4.2 Data mining, cleaning and exploration

The Knowledge Discovery in Databases (KDD) data mining methodology was utilized in this study. This methodology enabled the successful extraction of relevant information from large database from the sources described in section 3.2. It was chosen due to it's popular usage in forecast predictions. Upon pulling the relevant data from the aforementioned sources, it was ingested into a geodataframe in python. The *Geopandas* library was utilized to transform this data into it's appropriate geodataframe.The KDD methodology was applied in mining, cleaning, transforming and data selection of this data. These steps are described in further detail below.

4.2.1 Spatial data

In order to obtain the spatial data features, the centroid of each of the square grid cells described in section [3.4.1] was calculated. The Foursquare API caller was then utilized in order to search for venues within a 710m radius of the centroid of each grid map. 710m was chosen as the radius to allow features from the edge of neighbouring grids be incorporated, along with being the necessary value to ensure all corners of the $1km^2$ grid square is accounted for. Values returned as *NaN/null* were replaced with zeroes. This value is also in line with the recommended grid size described in section (2.2.1). This API call returned the individual business names of the features in question. In order to summarize and make sense of this data, the venues were all collected and categorized by their umbrella categorical term as specified by [16]. The tabular data on this webpage was scraped using pandas *read_html* package before being utilized to categorize the venues found per grid. See figure (2) for a sample of this data.

4.2.2 Crime data

Upon examining a heatmap of the crime data coordinates, it could be seen that several points were located outside of London. These were removed through a spatial join with the *GLA boundary data set* [17]. Prior to searching for outliers, entries who did not have a numeric longitude/latitude value were removed. The crimes are listed chronologically in year/month format. This was extracted to separate variables for year and month individually in order to explore the dependence of certain crime types on their year committed.

Crime filtering The crime data was then filtered by type. In total 16 unique crime types were provided by the data set. These were *Anti-social behavior, Criminal damage and arson, Public disorder and weapons, Vehicle crime, Violent crime, Other theft, other crime, burglary, robbery, shoplifting, theft from the person, violence and sexual offences, public order, bicycle theft, possession of weapons.*

In order to obtain sufficient data per cell along with conserving memory usage, it was decided to filter these crime types into the below sets -

[15]https://colab.research.google.com
[16]https://developer.foursquare.com/docs/legacy-foursquare-category-mapping
[17]https://mapit.mysociety.org/area/2247.geojson

- **Anti social behaviour** - Anti-social behaviour.
- **Property crimes** - Burglary, criminal damage and arson, vehicle crime.
- **Violent crimes** - Robbery, violence and sexual offences, possession of weapons, public disorder and weapons, violent crime.
- **Theft crimes** - Theft from the person, bicycle theft, shoplifting, other theft.

By utilizing this grouping of data, it was possible to cut the memory of the data set in use from 3.5GB to 70Mb, increasing computational speed exponentially.

4.2.3 Data set joining

The final step of the data set processing involved joining the crime and geofeatures data sets. It was decided to aggregate crimes per grid on each year with no dependence on the month that the crime occurred (described further in section (4.5))

Initially the location of crimes were collected into the grids in which they occurred. This data was then further aggregated over the total number of crimes per type per grid per year (aggregating over all occurrences of a distinct crime type (e.g. anti social behaviour) which occurred per year over the studied period). Upon aggregation of this crime data into crimes per grid per year, this data set was then spatially joined to the spatial features data set. This gave a data set containing the geographic grid location, the spatial features of said grid location and the newly calculated aggregated crime types for a given year in said grid, along with the year in which this calculation is valid.. A sample of this data is provided in figure (2) for a visual aide in comprehending the data sets nature.

Figure 2: Sample of cleaned and processed data set

Year	geometry	totalCrimes	theft	propertyCrimes	antiSocial	Arts & Entertainment	College & University	Food	Nightlife Spot	Residence	Outdoors & Recreation	Professional & Other Places	Shop & Service	Travel & Transport
2012	POLYGON ((-0.01061 51.32810, -0.01061 51.33637...	67.0	12.0	23.0	16.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
2012	POLYGON ((-0.01061 51.33637, -0.01061 51.34463...	657.0	88.0	165.0	275.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	2.0

4.3 Data exploration

Upon cleaning the data, it was explored in order to get a better understanding of nature of the data in question along with understanding if a class imbalance existed. The crime sets described in section (4.2.2) had their crime count aggregated across all grid maps over the course of the studied period, with the results displayed in figure (3). As can be seen, the data remains relatively evenly distributed hence there is negligible need for imbalanced data set considerations.

Furthermore, a crime density grid map was produced of the aggregated crimes per grid map over the studied period in order to understand the hotspots observed over historical data. As can be seen, crime appears to be only concentrated in the center of the GLA. Visualising crime density in such a way is not overly useful, as it does not take into account statistics which dictate whether a grid can be deemed a hotspot or not.

Figure 3: Crime occurrences bar chart

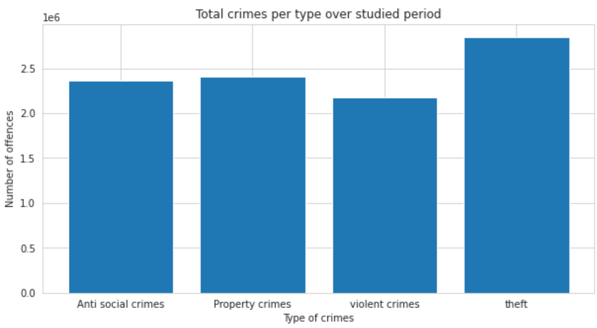


Figure 4: Crime density map of total crime per grid

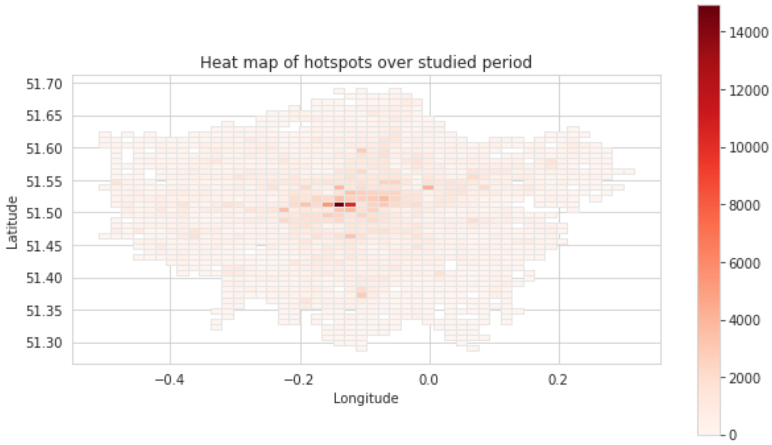
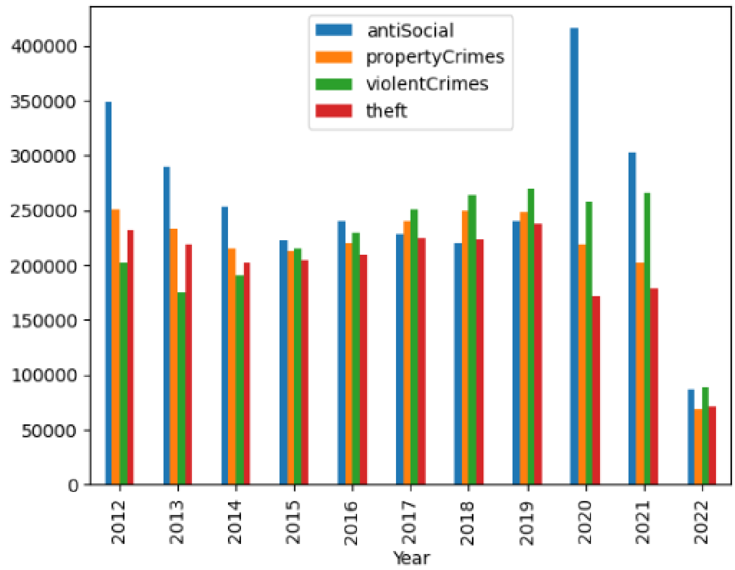


Figure 5: Aggregated crimes per year across studied period

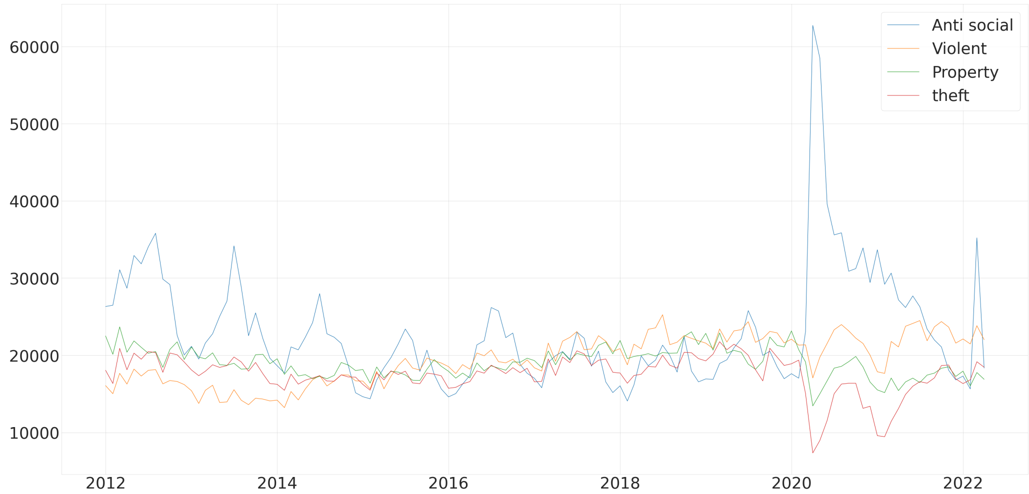


The variation in annual crime rates aggregated per grid is shown in figure (5). This is broken down into individual crime components for visualization purposes. It can be visually validated that the aggregated crime rates do not display any outlier values for the most part. This trend is broken by the data shown in 2020 and 2022. 2022 is due to data only being pulled as far as April, whilst a huge increase in anti social behaviour crimes occurred in 2020. As 2022 only has data for part of the year, it will be removed from the study. Upon investigating the change of crime rates in 2020, it can be seen that this coincided with the start of the COVID-19 pandemic in London and is likely due to the enforcement of new emergency lockdown rules. A timeline illustrating the aggregate crimes per month across the studied period can be seen in figure (6). This variation in data distribution must be taken into account, and is discussed further in section (4.5).

4.4 Performance evaluation

The performance metric used by researchers to assess technique performance has evolved over the years. One commonly used measure is the *hit rate*. This is the percentage of new crimes which occur within the areas predicted, over the total number of new crimes within the studied area. Although having it's uses and being easy to understand, one downfall with this metric is that this does not take into account the hotspot areal size where the crimes are predicted to occur. This can lead to misleading results. For example a hit rate could be 100%, but the predicted hotspot areas may cover the entire studied. This

Figure 6: Aggregated crime timeline per month



would lead to little use to analysts who seek to identify where to target resources [Chainey et al. (2008)]. Another metric used in past studies is the *Search Efficiency Rate*. First proposed by Bowers et al. (2004), it measures the number of events/crimes per square km in the areas where crimes are predicted to occur. This can provide useful statistics when a single study area is considered, but unfortunately does not allow for easy comparisons across studies. [Chainey et al. (2008)] introduced the *Prediction Accuracy Index (PAI)*. This considers the hit rate against the areas whereby crimes are predicted to occur with respect to the size of the studied area. It is calculated by dividing the hit rate by the Area Percentage (the percentage of the total areas designated as predicted hotspots. This is depicted below mathematically:

$$\frac{\frac{n}{N} * 100}{\frac{a}{A} * 100} = \frac{HitRate}{AreaPercentage} = PredictionAccuracyIndex \tag{2}$$

$$\begin{aligned} \text{where } n &= \text{Number of crimes in predicted hotspot areas,} \\ N &= \text{Total number of crimes in study area} \\ a &= \text{Total area size of predicted hotspot areas} \\ A &= \text{Total size of studied area} \end{aligned}$$

The benefits of this metric are that it considers both the number of crimes falling into the predicted hotspot against the size of the hotspot. This metric can be used as a measure against any study size, or crime collection data (point data or otherwise) and for any analysis which predicts spatial patterns of crime [Chainey et al. (2008)]. As many studies still utilize the hit rate as a measure of performance, it shall be considered in tandem for the purposes of performance comparison.

4.5 Training and test data used

Due to the reasons discussed in section (3.5), it was decided to not attempt predictions over a shortened time scale such as the year and month scale provided by the data. Instead it was decided to predict hotspots over a yearly period.

Upon investigation of the aggregated crime plots per month and per year throughout the studied period (figures 6 and 5 respectively), it is evident that the COVID-19 pandemic has been a drastic external influence on crime rates. Prior to this, crime rates remained roughly steady with relatively minimal variation. Due to this unprecedented event and variation, the data illustrated in figure (2) is partitioned into training data consisting of all crimes from 2012-2017 inclusive, and testing data from 2018 to 2021 inclusive. Although this is a 60:40 split of data into training and testing data, this is as it was desired to have two years of both pre-covid testing data, and two years of mid-covid testing data in order to compare the data equitably.

5 Implementation

5.1 Model design and architecture

As described in section (3.4.1), both grid mapping of a 1km^2 cell size and thematic mapping of OA shall be utilized for this study.

Several model types will be constructed. These are described in detail below -

5.1.1 Grid mapping of spatial features with ML methods

This model type will utilize the data set construction described in section (4.2.3). It will consider all of the ML techniques described in section (3.6) to produce predicted crime rates per grid. The hotspot definitions described in section (3.4.2) will then be applied to this predicted crime rates per grid cell in order to classify the predicted hotspot grids.

Performance evaluation for the PAI will then occur on the unseen testing data containing both pre and mid-pandemic data. The flow of this model is illustrated in figure (7b).

5.1.2 Grid mapping of baseline statistics

These models will not utilize any of the spatial features described in section (4.2.3). Instead it will focus on purely the crime rates per year along with the specific grid where these crimes occurred. Hotspot predictions on unseen test data will be made based purely on historical crime data.

Models such as these will be referred to as "baseline statistics" as they do not contain any of the spatial features data utilized in section (5.1.1). This model class will be subdivided into two further model types -

Grid mapping baseline hotspot This method will obtain the hotspots of the training data defined by the hotspot definitions in section (3.4.2). No crime rate predictions will be made. Instead, these hotspots obtained from historical crime data (or baseline statistics) will be applied to the unseen testing data, with their PAI performance measured. These results will be referred to as "baseline grid quantile stats", "baseline grid G stats" and "baseline grid G HH only" stats for the hotspot previously described hotspot definitions of quantile statistics, G statistics and G HH only statistics respectively. This flow is described in figure (7b).

KDE baseline hotspot Similar to above, this method will obtain the hotspots of the training data, but only the quantiles hotspot definition will be considered for this model type. No crime rate predictions will be made. Instead, KDE shall be performed on this crime and geographical data, formulating hotspots on this trained data. These hotpots will then be applied to the unseen data, with the performance of these hotspots on unseen testing data measured. This method will be utilized to both KDE with crimes aggregated per grid, and KDE with crimes aggregated per OA. These will be referred to as "KDE grid quants" and "KDE OA quants" respectively. This model flow is similar to the above described grid mapping baseline hotspot, illustrated in figure (7b).

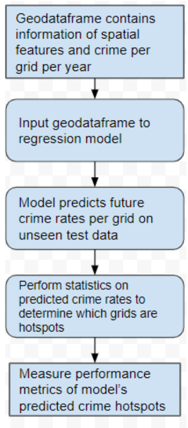
5.2 Optimization

The model design described in section (5.1) were all implemented individually. Modelling was conducted on each crime category individually, as well as the aggregated total crime data. Feature scaling was performed prior to the implementation of the LR, NN and KNN models in a bid to reduce the negative impact of outliers and optimize the results. This transformation was not necessary for the ensemble methods however. Python's ***scikit-learn***[18] module was utilized in the running of the RF, KNN, ET and LR models, with the ***open-sourced Xgboost*** being utilized for the XGB model and the ***tensorflow keras library***[19] utilized likewise for the NN models.

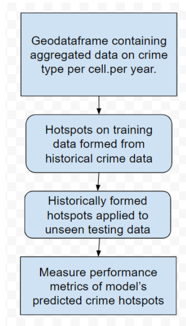
The sole hyperparameter for KNN of the number of nearest neighbours to consider was found via construction of elbow plots for each crime type. Each crime type's elbow plot indicated optimal values when a K=1 evaluation was used.

[18] https://scikit-learn.org/stable/

[19] https://www.tensorflow.org/api$_docs/python/tf/keras/Model$



(a) Grid mapping of spatial features model design



(b) Grid mapping of baseline statistics

Figure 7

The hyperparameters for the ensemble model's were optimized via Random Search Cross-Validation (RSCV) with respect to RMSE. Although PAI is viewed as the primary performance metric, unfortunately due to it's complex geospatial nature, it was not possible to utilize this as a scoring metric for training purposes.

Similar to the ensemble method, the NN model was optimized with respect to the RMSE. A normal kernal initializer and the rectified linear activation function were deemed most appropriate due to the nature of the problem at hand. The selection of further hyperparameters such as the number of hidden layers, number of neurons and learning rate were found via the recommended best practise approach of experimentation [20].

5.3 Hotspots output

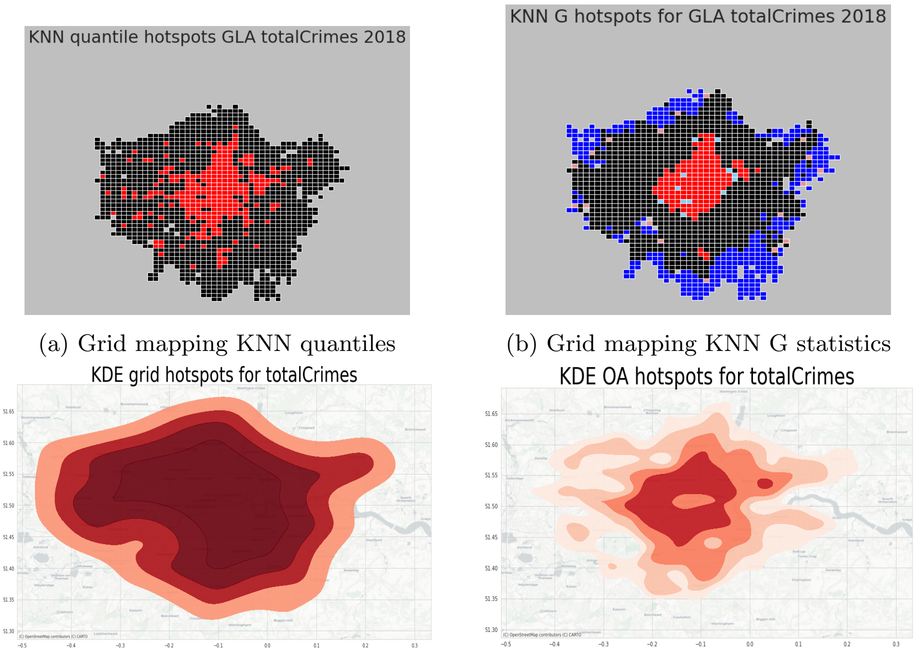
The visual aide of a hotspot map is deemed a useful tool in displaying findings in hotspot analysis. A sample map is shown below for each of the model's previously described. The red represents the hotspots in each of the quantiles/G statistic predictions. Black is seen to represent insignificant grids. Blue represents LL areas, whilst pink represents HL areas (the presence of these in calculations is the difference between the G and G HH only statistics). As can be seen, the KDE plots appear more aesthetically pleasing versus the "blocky" appearance of the grid mapping models. In the KDE plots, the levels of crime increase with the darkness of red used. The areas we classified as "hotspots" for KDE are the darkest red shown. The G hotspot plot offers more information than the quantiles, with data on statistically significantly HH, LL, LH and HL grids. This is compared to the KNN quantiles which do not account for spatial autocorrelation, only showing information on the cells which are in the top 80+ percentile.

6 Evaluation

As described previously in section (4.5), data for both pre-pandemic (2018,2019) and midst pandemic (2020,2021) data was utilized as unseen testing data for the models. The models were ran for each of these years individually, before combining and sectioning the results into pre and midst pandemic results respectively. In order to obtain a reliable result for this data, the geometric mean was calculated for the results obtained by the models for 2018 and 2019 for pre pandemic, with likewise being completed for years 2020 and 2021 to obtain midst pandemic data.

The benefit of using this averaging technique means fairer results can be obtained. The equal weighting being used across both years for the pre and midst pandemic data signifies that is important for any model to be consistent across multiple years, with no great variance in performance from one year to the next. In particular, the pre pandemic results will be used to provide comparisons with previous studies undertaken into non-pandemic data.

[20] https://machinelearningmastery.com/how-to-configure-the-number-of-layers-and-nodes-in-a-neural-network/



(a) Grid mapping KNN quantiles

(b) Grid mapping KNN G statistics

(c) Grid mapping KDE quantiles

(d) OA mapping KDE quantiles

Figure 8: Hotspot maps produced by respective models - crimeForecastsGLA coding notebook - Andrew McMorrow

The results for each crime type were measured for each model, and each hot spot classification tool. They were also compared with so-called "baseline statistics". These are the *PAI* scores achieved if the hotspot classification tool used utilized historical hotspots on the future, unseen testing data with no models incorporated. The KDE model's can also likewise be viewed as "baseline statistics" as they do not incorporate spatial data. KDE baseline statistics exist for both the grid mapping and OA data. Utilizing these baseline statistics will provide fair comparison across the models.

The results for both pre-pandemic and mid-pandemic models for each crime type are shown below in figures (9) and (11). In

6.1 Results evaluation

Due to the differing data available for each crime type, different model's will be considered for each crime.

Total crime Starting with the models for total crime, notable performances are achieved by the KNN model, followed both the RF and XGB models. Despite the dramatic change in crime rates during the pandemic, it is somewhat surprising to see relatively little deviation in the results between pre and mid pandemic performances. It is seen that the quantile statistics are achieve the leading performance across all models used, with pre pandemic PAI results of 2.85 achieved by the KNN's utilization of quantile statistics being the highest performance, followed by 2.73 and 2.77 achieved by the RF and XGB model. This offers an improvement in predictive performance of 12% compared to results achieved (2.53) when utilizing baselines statistics with no spatial information, with 2.53 achieved by both the quantiles baseline statistic, and the quantiles KDE model. All of the models incorporating grid mapping offered far superior predictive ability to the OA mapped KDE quantiles model performance of 1.28. Performance on post-pandemic model's took a slight dip as expected, but still stood up well. Once again the quantile statistics method proved incisive, with it's utilization by the KNN model obtained the highest result, with a PAI of 2.77 usurping the other impressive results of 2.66, 2.63 and 2.70 achieved by RF, ET and XGB respectively. All of these model offered higher predictive values than the highest performing baseline statistics of 2.58 achieved by the baseline quantile statistics, with performance increases of 8%, 3%, 2% and 4.5% achieved by the respective aforementioned models. The fact that performance increases are seen both pre and mid-pandemic offers encouragement that these models can be utilized during both future precedented and unprecedented changes in crime patterns. As the quantile statistics do not take

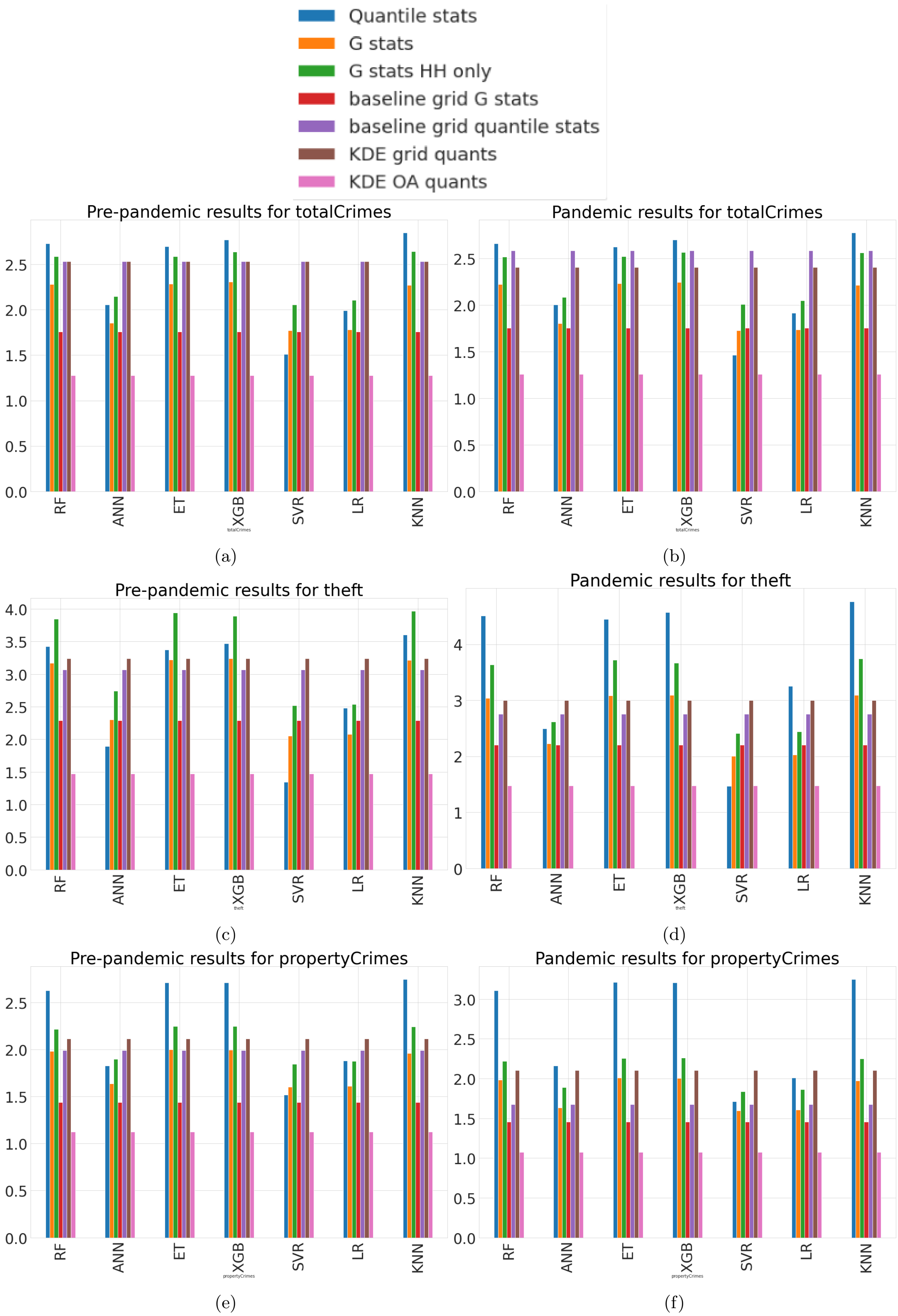
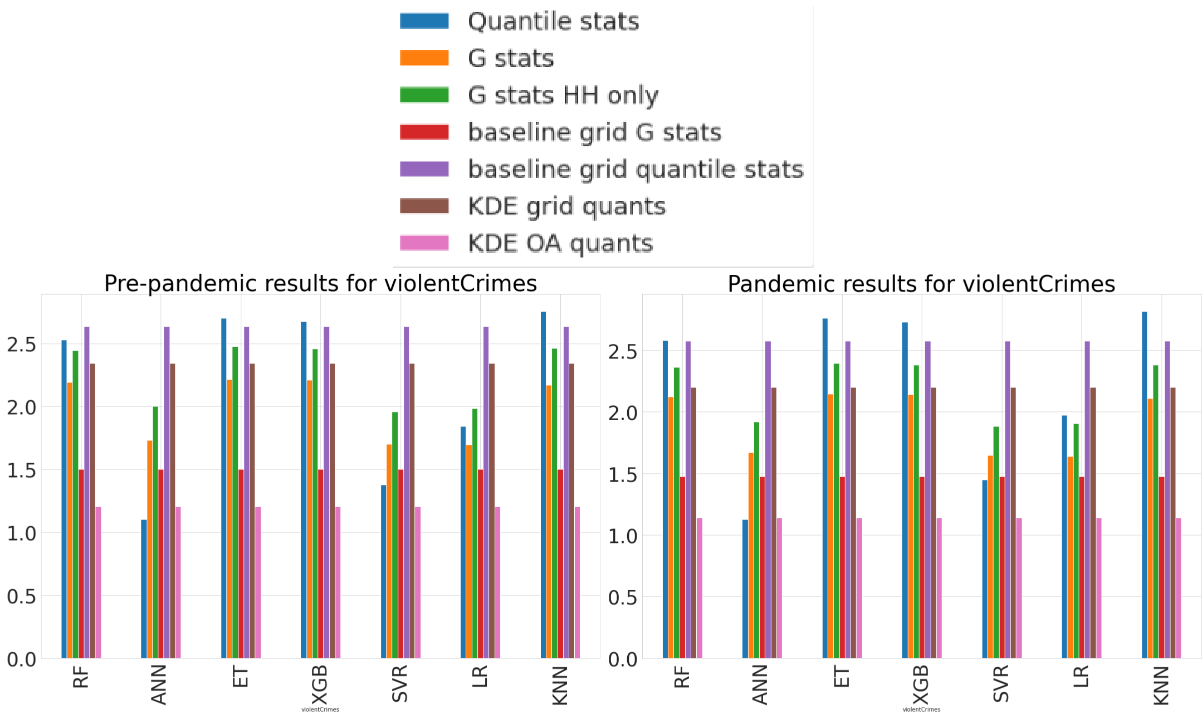
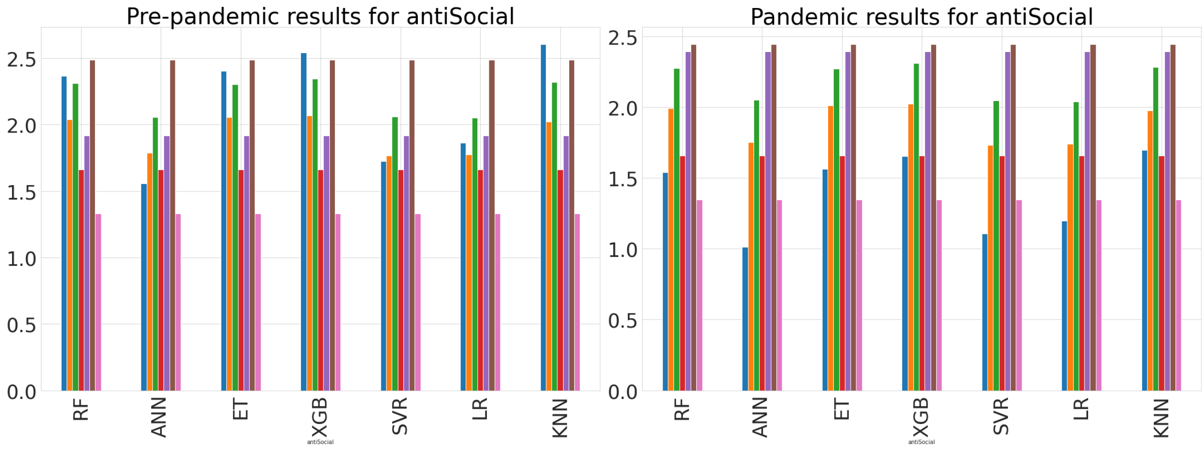


Figure 9: Results for both pre and mid-pandemic data across total crimes, theft and property crimes



(a) (b)



(c) (d)

Figure 10: Results for both pre and mid-pandemic data across violent and anti-social crimes

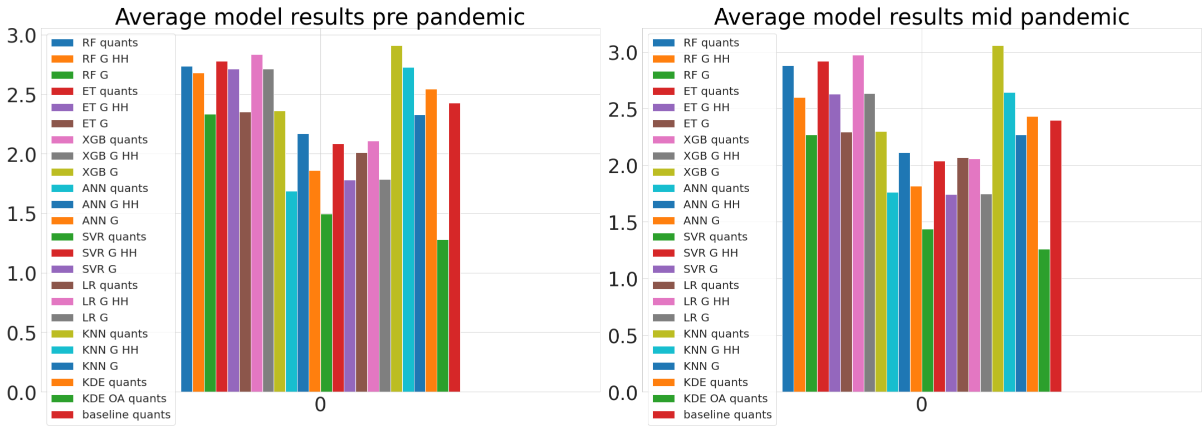


Figure 11: Mean model results across both pre and mid-pandemic data

into account spatial autocorrelation, it is interesting to see these outperform the G statistics. This suggests that certain grids high crime rates could be dependant on neighbouring grids, which is not entirely surprising.

Theft In contrast to the relatively similar performances seen pre and mid-pandemic for the total crimes, large dissimilarities are evident for these performances with theft crime. Pre-pandemic, the G HH only statistics achieved excellent PAI results of 3.97, 3.89, 3.95 and 3.85 when utilized with the KNN, XGB, ET and RF model's respectively. The quantile statistics utilized with the above models offer the next highest performance with impressive results of 3.62, 3.47, 3.38 and 3.42 for the KNN/ET/XGB/RF models respectively. For the first time we see impressive results of the baseline KDE quantiles model, offering a PAI of 3.24 on pre-pandemic theft data.

Although these model's are also the best performing when used during mid-pandemic data, it is through the utilization of quantile statistics which achieve the leading results of 4.78/4.510/4.45/4.57 for KNN/RF/ET/XGB models vs the 3.75/3.64/3.72/3.67 achieved with G HH only statistics. Notably, this is still an increase in performance of 25% when utilizing the G statistics versus the best performing baseline statistics (quantiles KDE), but nothing compared to the increase in predictive power of nearly 58% for the KNN model utilizing quantile stats versus that achieved by the KDE quantile's baseline. This increase in predictive power versus the best performing baseline statistics compares with just 12% found on pre-pandemic data. This degree of randomness is not entirely unexpected, as figures (5) and (6) have illustrated a huge change in theft crime rates pre and mid pandemic, meaning vastly different unseen testing data sets are being modelled upon. These results indicate that KNN, XGB, RF and ET model's incorporating either quantile or G HH statistics for hotspot classification methods are both fit to be used in levels of both precedented and unprecedented crime rates.

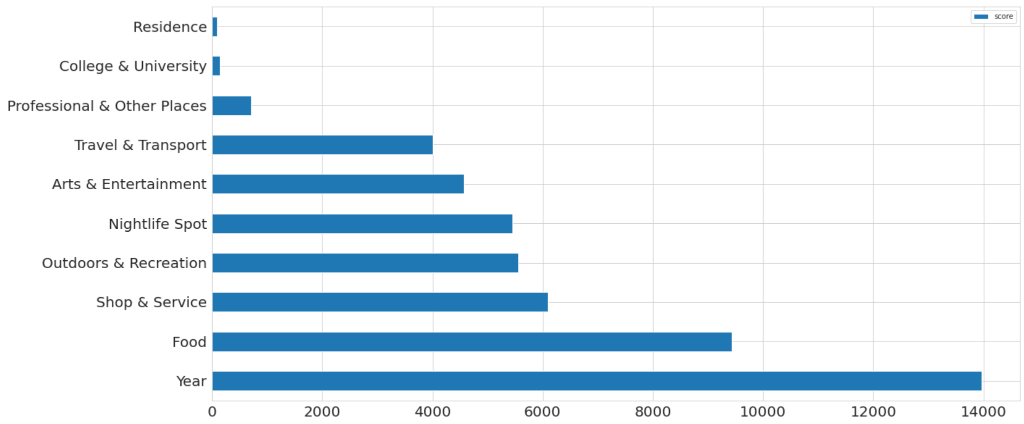
Property crimes As is the case with the two previously mentioned crime types, the use of quantile statistics once again performs excellently in property-related crimes. There are large similarities with the previous crime types, with KNN, XGB, RF and ET models all producing results with much higher PAI scores than others. It can be noted in figure (6) that property crimes decreased slightly with the advent of the pandemic, with higher scores being found by the KNN/XGB/ET/RF models when studied on the mid-pandemic data (scores of 3.25/3.21/3.21/3.1 and 2.75/2.71/2.71/2.63 for mid and pre-pandemic XGB/ET/RF model's respectively). This continues a trend first indicated by the theft crimes - that the model's are better able to predict the crime hotspots for lower levels of overall crime. The spatial features aspect of our model is particularly influential whilst studying this crime type, offering improvements of 30%/54% compared to forecasts undertaken by the best performing baseline method of KDE quantiles for pre/mid pandemic data.

Violent crimes Unlike the property and theft crimes, the best performing model's incorporating spatial data only marginally outperform the baseline quantiles statistics. The KNN model coupled with quantile statistics offers an increase in predicting power of merely 4.5% (2.75 vs 2.63 PAI of baseline grid quantiles) for pre-pandemic data. This increases slightly to 9% when modelled on mid-pandemic data (2.82 vs 2.58). In both the pre and mid pandemic data, the KNN model coupled with quantile statistics achieved the highest performance. Although these results indicate limitations of the model's when forecasting on violent crimes, it highlights the power of the baseline quantile statistics on this crime type that they are able to compete with minimal knowledge of surrounding data.

Anti-social crimes In contrast to the results of violent crimes, the KDE grid quantiles baseline has predictive abilities to rival the highest performing model's of KNN and XGB incorporating spatial data. The KNN and XGB model's incorporating quantile statistics marginally outperform the KDE grid quantiles baseline (4%/2.5% increase for KNN/XGB vs KDE quantiles) in pre-pandemic data. Interestingly, there is a large shift in the performance of the model's on post-pandemic data. For the first time, a baseline model (both KDE grid quantiles, and baseline grid quantiles) outperforms model's incorporating spatial data. They offer an increase in predictive power of 5.7% and 3.5% for KDE grid quantiles and baseline grid quantiles respectively over the highest performing model incorporating spatial data of ET incorporating G HH statistics.

This change in performance of the model's is not entirely unexpected, with the radical change in the trend of anti-social crime rates as indicated in figures (5) and (6) meaning the mid-pandemic model's were being tested on drastically different data. These results indicate that the model's incorporating spatial data are not fit for forecasting anti-social crimes during a pandemic, but it is notable to see

Figure 12: Feature importance ranking for XGB model on total crimes



that the post-pandemic KDE grid quantile statistic result of 2.45 offering only marginally less predictive performance than the best performing pre-pandemic model - the 2.60 achieved by the KNN model with quantile statistics.

6.1.1 Feature importance

Due to the ethical issues discussed in section (3.3), it was important to include some levels of interpretability in the model. Although it can be seen in figure (11) that KNN offers the highest performance in terms of PAI results, obtaining the features which the model deem most important is not a straightforward task for this model type. The XGB as the next highest performing model (with a decrease in PAI performance of only 2.6%/2.8% for pre/mid-pandemic data) on the other hand offers this required feature importance. The results of which are demonstrated in figure (12). It is clear to see the effect the year has on the crime rates, which is not unsurprising upon examining the crime rates trend in figures (5) and (6). This indicates a level of increasing crime rates with time. This increase in crime rates may correlate with the increase in London's population levels over time [21], however investigation of this relationship in detail is outside the scope of this study. Furthermore, it is interesting to see the impact of food outlets on crime rates, possible suggestions for this are that areas with increased levels of food outlets may also reside higher population levels. This theory could be applied to several of the features, however similar to the population increase, it's investigation is outside the scope of this study.

6.2 Summary of optimal models per crime types

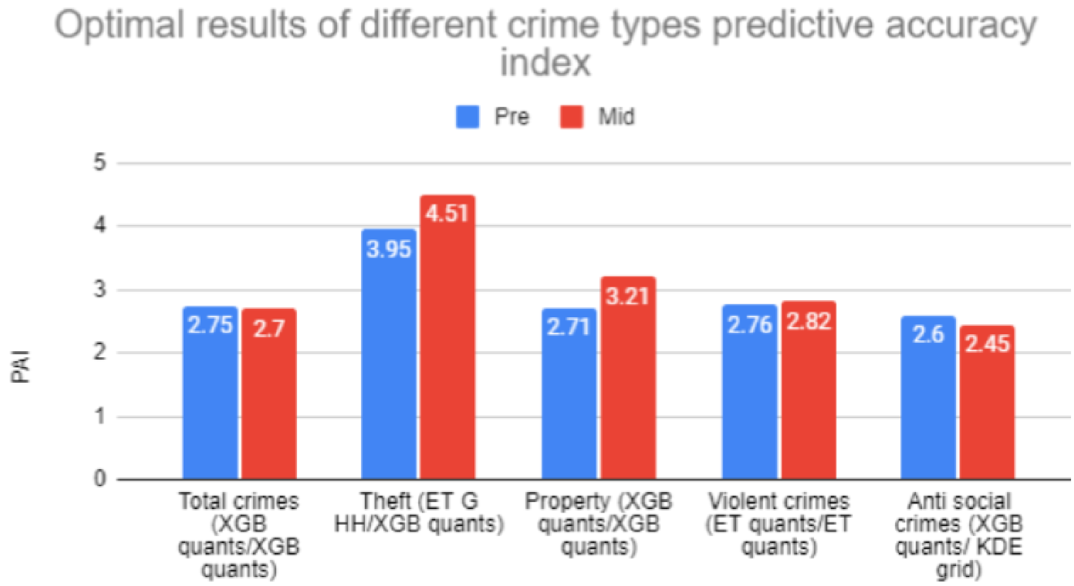
For ease of comparison, the optimal results achieved for different crime types along with the model utilized is illustrated in figure (13). As can be seen, different models achieve optimal values for differing crime types with the predictive accuracy index differing widely across crime types. The majority of the results indicate performance consistency across models (total crimes, property and violent crimes models all indicated optimal results for XGB, XGB and ET models respectively across both pre and mid-pandemic data). This indicates a fit for purpose use of these models across both standard crime data trends (pre-pandemic data) and unprecedented crime data trends (mid-pandemic data). This figure also illustrates the effect of different crime types on hotspot prediction. It can be seen that theft crime hotspots achieve the highest predictive accuracy, followed property, violent and anti social crimes. The most popular, efficient and optimal model across these crime types is the XGB model.

6.3 Comparison with external studies

It is difficult to compare of crime forecasting like for like, due to a variety of external factors such as the areas studied, time frame investigated, population levels, crime policies in place etc. Furthermore, many similar studies do not focus on the same crime types. For example, Swain (2012)'s study of differing modelling and mapping techniques categorised the targeted crime types into four categories "Burglary", "From moving vehicle", "of moving vehicle" and "from Person". It can be difficult to draw too many

[21] https://www.macrotrends.net/cities/22860/london/population

Figure 13: Different crime types and their models which achieved optimal results for pre/mid pandemic



conclusions between comparisons such as these due to the factors mentioned above. An average PAI result per model across all crime types is available in this paper, with a similar measure performed in figure (11). In both the pre and mid-pandemic studies, the KNN, RF, XGB and ET quants, along with their G statistic HH counterparts all receive a promising PAI over 2.5. This compares favourably with the optimal average FPI received of 2.32 by the "STAC Convex Hulls" technique in Swain (2012)'s study. Chainey et al. (2008) also compared different hotspot mapping techniques, with this study taking place across the London Borough of Camden. They also compared average PAI values across crime types, with the 2.90 achieved by the Kernel Density Estimation method the highest value achieved. This value compares positively with the overall average of 2.82 received by the KNN model utilizing quantile statistics (2.85 pre pandemic, 2.78 post pandemic) over total crime. It is interesting to note their improvement in performance of KDE results over the 2.49 (2.55/2.43 for pre/mid pandemic) and 1.27 (1.28/1.26 for pre/mid pandemic) for grid and OA quantiles respectively. This indicates room for improvement is available for the KDE model's. The OA model scores particularly poorly considering that 's study focuses also on OAs, however as their focus is on the much smaller area of the borough of Camden, this may go some way to explaining the increase in predictive ability.

A study ran by Adepeju et al. (2016) built upon Chainey et al. (2008)'s investigation of crime in Camden, London proposing several different hotspot classification methods in tandem with KDE for shoplifting, violence and burglary. These different prediction methods include self-exciting point process (SEPP), prospective hotspot (PHotspot), prospective KDE (PKDE) and Prospective space-time scan statistic (PSTSS). The optimal PAI results achieved were 2.99/4.58/2.37 for violence/shoplifting/burglary respectively. These values compare with the 2.79 (2.76/2.82 for pre/mid pandemic), 4.18 (3.61/4.76 for pre/mid pandemic) and 3.00 (2.75/3.25 for pre/mid pandemic) for violent crimes (including violence), theft (including shoplifting) and property crimes (including burglary) respectively achieved by the model proposed in this study. Notably, the study by Adepeju et al. (2016) offers a much narrower focus into specific crime types of shoplifting, violence and burglary versus the wider range of overall property and theft crimes examined in this study. It is therefore highly encouraging to see the results of the model's proposed in this paper compare favourably with studies which focus on a much more granular window of crime types. The optimal quoted results were achieved by the KNN model, although comparable results exist for each of the XGB, RF and ET model's incorporating quantile statistics.

A further study by Wheeler and Reuter (2021) presented novel DBSCAN (Density Based SCAN) clustering technique to create hotspot of crime, focusing only on violent and theft crimes. This novel technique achieved PAI results ranging from 2.6 to 3.2 for crime which are sub categories of the "theft" category utilized in this study, and 2.9 to 3.1 for crimes deemed as violent crimes in this report. These results indicate that the PAI results achieved by the KNN model of 2.79 and 4.18 for violent and theft crimes respectively once again compare favourably with values found from historical studies.

It is very encouraging to see the values achieved by the method proposed in this paper compare in a favourable manner with an array of differing hotspot mapping techniques. These findings indicate that the model proposed in this paper is fit for it's intended purpose of predicting future crime hotspots in London.

6.4 Discussion

The mean result of each model across the distinct crime types are illustrated in figure (11). In both pre and mid-pandemic data, the KNN model utilizing quantile statistics achieved the optimal performance with PAI scores of 2.92/3.06 followed by XGB's 2.84/2.97 on pre/mid-pandemic data. These results offer an increase in performance of 15% and 26% (2.54/2.43 for pre/mid) over the best performing baseline statistic (KDE quantiles). These results indicate that significantly further predictive power can be obtained by incorporating the spatial features utilized in this study, versus using purely statistics on historical data. The results obtained suggest that KNN utilizing quantile statistics is a reliable and suitable way to predict crime hotspots over a variety of crime types in London. It would be interesting to study the performance of introducing more sophisticated clustering methods to the data in question, with unsupervised learning methods of particular interest. Combining K means clustering with a regression model (such as SVR) would be a suggested model to incorporate if further work was undertaken on this task.

Another model's performance worth commenting on is that of the ANN. It achieved a low PAI score of approximately 2.1 for both pre and mid-pandemic data in comparison to the impressive results by the KNN, XGB, RF and ET models. There are several suggested reasons for this performance. One is that NN models typically require much more data than traditional ML algorithms in order to compute reliable results. Furthermore, due to the recommended experimentation method used in order to find the hyperparameters such as number of neurons, learning rate etc., along with their relatively long computational time, these are a very computationally expensive method to be used in the (relatively) straightforward regression aspect of this task.

Areas of improvement exist for this model however. The KDE results achieved by grid quantiles are comparable with literature, in some ways justifying the grid size choice, however the KDE model on OA data in particular achieved compartively poor results. This indicates room for improvement exist on the hyperparameter settings employed by the KDE, with the novel approaches of SEPP, PHotspot, PKDE, PSTSS offered by Adepeju et al. (2016) to be considered if further work is to be completed in this area.

Another item to consider is that this model considers the spatial features of a given grid to be static over time, as it does not take into account circumstances such as a new restaurant opening, or a cafe closing. Currently it assumes that the data pulled during the data-processing aspect, is the same data in place for all of the years used for the training and testing data. Whilst understandable to make such an assumption, improvements for the model exist in compiling the spatial features data dynamically.

It would have been preferable to utilize a crime data source which possessed the time and date of the given crime. This would have allowed studies on the impact of the time of day, hours of daylight, weather and full moon cycles to be considered as features to be considered for a model. For example, studies such as [COHN (1990)] have highlighted the correlation between the weather and crime occurrences in the UK, indicating that violent crimes against civilians increase linearly with heat until a threshold of approximately 32°. Additionally, this data would have allowed the implementation of the sophisticated LSTM model on this more granular temporal data. This would have allowed comparisons on further studies to be made, due to it's popularity in literature of similar studies.

Furthermore, it would have been desirable to consider and study the effects of introducing the features considered unethical discussed in section (3.3) to understand their effect on model predictive ability. A comparison on the effects of this data introduction to the model's performance would have been insightful. Finally, in order to assess the model's suitability to external data further scope should be considered to test the performance of this model on an external city (e.g. Birmingham). This would require data scraping and processing to the same level of intensity as performed for this investigation, hence is outside of the scope of this study. Nevertheless it should be considered were the scope of this study extended.

7 Conclusion and Future Work

This study aimed to address the question *"Can spatiotemporal predictive modelling be utilized to accurately and ethically predict crime hotspots in London?"*. Several models have been produced using ethically, unbiased data which produce results which agree with (and in some cases improve upon) literature, hence this study can be deemed a success.

In order to ensure this study remains as ethical as possible, only data on the spatial features of an area (such as the number of food outlets, shops, entertainment venues etc.) were used, with data which has previously been associated with racial bias such as race, employment status or education levels ignored. The highest performing models on the aforementioned data utilize quantile statistics on predicted crime rates to denote hotspots, although this method does not take into account spatial autocorrelation like that of G statistics, it is a method which has been used throughout literature achieving impressive, comparable results.

The highest PAI result was achieved by the KNN model in tandem with these quantile statistics, however in order to provide interpretability to the model (which aids in assessing ethical issues), the next highest performing model of XGB in tandem with these aforementioned statistics is deemed the most appropriate for usage.

A miniscule drop off in PAI performance of 2.6%/2.8% for pre/mid-pandemic data is deemed an appropriate trade-off for this increase in interpretability, reducing the PAI results from 2.92/3.05 to 2.84/2.97 for pre-mid-pandemic data. This offered an increase in predictive ability of 11%/22% for pre/mid-pandemic data, compared to the highest performing (and recommended in literature) baseline model of KDE.

This increased interpretability indicated that the number of food outlets was the most important feature of predicting crime hotspots in a given year, with the crime rates also showing a strong positive correlation with the year studied.

This successful investigation could provide useful insights to law enforcement agencies, in their bid to reduce crime rates across the GLA in an ethical manner. Limitations exist, with the inability to predict crime rates within a shorter period of one year notable. There exists scope for improvement on this model however which could be addressed in future studies. As this model focused on grids of equal area, it is highly likely there is a large variation in population levels across these grids. Utilizing OAs of relatively similar population levels was attempted in a bid to account for this change in population, however the results of which were very poor in comparison to the grid mapping results. It would be interesting to see this model's performance compared to a model which accounted for population, along with the numerous spatial features presented here. Furthermore, the advent of the pandemic presented difficulties with this study, deeming it necessary to partition the unseen testing data into pre and mid-pandemic data to account for these drastic change in crime rates. It would be useful to be able to validate results on a wider scale of non-pandemic data were it available.

The model's performance should be tested on data from an external city to London in order to accurately assess it's suitability and scalability. Were sufficient performance on the aforementioned external study achieved, the model proposed could be utilized in order to decrease crime rates across the UK and beyond.

In conclusion, this study was successful in it's goal of accurately and ethically predicting London crime hotspots with spatiotemporal predictive modelling with the XGB ML model in association with quantile statistics providing the optimal results.

References

Action to promote effective crime prevention (n.d.). "Annex: Guidelines for the Prevention of Crime": p. 43-47.
URL: *http://digitallibrary.un.org/record/475865*

Adepeju, M., Rosser, G. and Cheng, T. (2016). Novel evaluation metrics for sparse spatio-temporal point process hotspot predictions-a crime case study, *International Journal of Geographical Information Science* **30**(11): 2133–2154.

Andresen, M. A. (2005). Crime Measures and the Spatial Analysis of Criminal Activity, *The British Journal of Criminology* **46**(2): 258–285.
URL: *https://doi.org/10.1093/bjc/azi054*

Andresen, M. A. (2009). Testing for similarity in area-based spatial patterns: A nonparametric monte carlo approach, *Applied Geography* **29**(3): 333–345.
URL: *https://www.sciencedirect.com/science/article/pii/S0143622809000046*

Andresen, M. A. and Malleson, N. (2013). Crime seasonality and its variations across space, *Applied Geography* **43**: 25–35.
URL: *https://www.sciencedirect.com/science/article/pii/S0143622813001410*

Andresen, M. and Malleson, N. (2015). Andresen, m.a., malleson, n. (2015). intra-week spatial-temporal patterns of crime. crime science, 4, article 12., **4**: Article 12.

Anselin, L., Cohen, J., Cook, D., Gorr, W. and Tita, G. (2000). *Spatial Analyses of Crime*, Vol. 4, pp. 213–262.

Bates, S. (1987). Spatial and temporal analysis of crime, *Research Bulletin, April* .

Bowers, K. J., Johnson, S. D. and Pease, K. (2004). Prospective hot-spotting: the future of crime mapping?, *British journal of criminology* **44**(5): 641–658.

Braithwaite, J. (1989). *Crime, Shame and Reintegration*, Cambridge University Press.

Brantingham, P. J., Brantingham, P. L. et al. (1981). *Environmental criminology*, Sage Publications Beverly Hills, CA.

Ceccato, V. and Uittenbogaard, A. C. (2014). Space–time dynamics of crime in transport nodes, *Annals of the Association of American Geographers* **104**(1): 131–150.
URL: *https://doi.org/10.1080/00045608.2013.846150*

Chainey, S. and Ratcliffe, J. (2005). *Spatial Theories of Crime*, John Wiley Sons, Ltd, chapter 4, pp. 79–113.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118685181.ch4*

Chainey, S., Tompson, L. and Uhlig, S. (2008). The utility of hotspot mapping for predicting spatial patterns of crime, *Security Journal* **21**: 4–28.

Chollet, F. (2021). *Deep learning with Python*, Simon and Schuster.

Cohen, L. E. and Felson, M. (1979). Social change and crime rate trends: A routine activity approach, *American Sociological Review* **44**(4): 588–608.
URL: *http://www.jstor.org/stable/2094589*

COHN, E. G. (1990). WEATHER AND CRIME, *The British Journal of Criminology* **30**(1): 51–64.
URL: *https://doi.org/10.1093/oxfordjournals.bjc.a047980*

Eck, J., Chainey, S., Cameron, J. and Wilson, R. (2005). Mapping crime: Understanding hotspots.

Ehrlich, I. (1975). On the relation between education and crime, *Education, income, and human behavior*, NBER, pp. 313–338.

Fan, S. (2014). The spatial-temporal prediction of various crime types in houston, tx based on hot-spot techniques.

Fernández-Delgado, M., Sirsat, M. S., Cernadas, E., Alawadi, S., Barro, S. and Febrero-Bande, M. (2019). An extensive experimental survey of regression methods, *Neural Networks* **111**: 11–34.

Freeman, R. B. (1999). The economics of crime, *Handbook of labor economics* **3**: 3529–3571.

Getis, A. and Ord, K. (1992). The analysis of spatial association by use of distance statistics, *Geographical Analysis* **24**: 189 – 206.

Goodfellow, I., Bengio, Y. and Courville, A. (2017). Deep learning (adaptive computation and machine learning series), *Cambridge Massachusetts* pp. 321–359.

Gorr, W., Olligschlaeger, A. and Thompson, Y. (2003). Short-term forecasting of crime, *International Journal of Forecasting* **19**(4): 579–594.
URL: *https://www.sciencedirect.com/science/article/pii/S016920700300092X*

Groff, E. and La Vigne, N. (2002). Forecasting the future of predictive crime mapping, *Crime Prevention Studies* **13**.

Harries, K. D. (1999). *Mapping crime: Principle and practice*, US Department of Justice, Office of Justice Programs, National Institute of ....

Hirschfield, A. and Bowers, K. (2001). *Mapping and analysing crime data: Lessons from research and practice*, CRC Press.

Hodgkinson, T., Andresen, M. A. and Farrell, G. (2016). The decline and locational shift of automotive theft: A local level analysis, *Journal of Criminal Justice* **44**: 49–57.
URL: *https://www.sciencedirect.com/science/article/pii/S0047235215300180*

Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks, *Neural Networks* **4**(2): 251–257.
URL: *https://www.sciencedirect.com/science/article/pii/089360809190009T*

Ibrahim, N., Wang, S. and Zhao, B. (2019). Spatiotemporal crime hotspots analysis and crime occurrence prediction, *in* J. Li, S. Wang, S. Qin, X. Li and S. Wang (eds), *Advanced Data Mining and Applications*, Springer International Publishing, Cham, pp. 579–588.

Jenkins, R. and Purves, D. (2020). Ai ethics and predictive policing: A roadmap for research.
URL: *http://www.aipolicing.org/year-1-report.pdf*

Koperski, K. and Han, J. (1995). Discovery of spatial association rules in geographic information databases, *in* M. J. Egenhofer and J. R. Herring (eds), *Advances in Spatial Databases*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 47–66.

Law, J. and Chan, P. (2012). Bayesian spatial random effect modelling for analysing burglary risks controlling for offender, socioeconomic, and unknown risk factors, *Applied Spatial Analysis and Policy* **5**.

Law, J., Quick, M. and Chan, P. (2014). Bayesian spatio-temporal modeling for analysing local patterns of crime over time at the small-area level, *Journal of Quantitative Criminology* **30**.

Levine, N. (2004). Journey to crime estimation, *CrimeStat III: A spatial statistics program for the analysis of crime incident locations* .

Linning, S. J. (2015). Crime seasonality and the micro-spatial patterns of property crime in vancouver, bc and ottawa, on, *Journal of Criminal Justice* **43**(6): 544–555.
URL: *https://www.sciencedirect.com/science/article/pii/S0047235215000598*

Liu, D., Song, W. and Xiu, C. (2016). Spatial patterns of violent crimes and neighborhood characteristics in changchun, china, *Australian & New Zealand Journal of Criminology* **49**(1): 53–72.
URL: *https://doi.org/10.1177/0004865814547133*

Liu, D., Song, W., Xiu, C. and Xu, J. (2021). Understanding the spatiotemporal pattern of crimes in changchun, china: A bayesian modeling approach, *Sustainability* **13**: 10500.

Maguire, M. (2000). Policing by risks and targets: Some dimensions and implications of intelligence-led crime control, *Policing and Society: An International Journal* **9**: 315–336.

Monmonier, M. (1996). How to lie with maps, *The American Statistician* **51**.

Newton, A. and Felson, M. (2015). Editorial: crime patterns in time and space: the dynamics of crime opportunities in urban areas, *Crime Science* **4**: 11.

Owusu, G. and Frimpong, L. (2020). Crime geography.

Pezzuchi, G. (2008). A brief commentary on "the utility of hotspot mapping for predicting spatial patterns of crime", *Security journal* **21**(4): 291–292.

Raaijmakers, S. (2019). Artificial intelligence for law enforcement: challenges and opportunities, *IEEE security & privacy* **17**(5): 74–77.

Ratcliffe, J. (2002). Aoristic signatures and the spatio-temporal analysis of high volume crime patterns, *Journal of Quantitative Criminology* **18**: 23–43.

Sampson, R. J. and Groves, W. B. (1989). Community structure and crime: Testing social-disorganization theory, *American Journal of Sociology* **94**(4): 774–802.
URL: *http://www.jstor.org/stable/2780858*

Shaw, C. and McKay, H. (1970). Juvenile delinquency in urban areas, *American Sociological Review* **35**.

Spelman, W. (1995). Criminal careers of public places, *Crime and Place* **4**.

Stucky, T. and Ottensmann, J. (2009). Land use and violent crime, *Criminology* **47**: 1223 – 1264.

Swain, A. W. (2012). A comparison of hotspot mapping for crime prediction major.

Walczak, S. (2021a). Predicting crime and other uses of neural networks in police decision making, *Frontiers in Psychology* **12**.
URL: *https://www.frontiersin.org/article/10.3389/fpsyg.2021.587943*

Walczak, S. (2021b). Predicting crime and other uses of neural networks in police decision making, *Frontiers in Psychology* **12**.
URL: *https://www.frontiersin.org/article/10.3389/fpsyg.2021.587943*

Walczak, S. and Cerpa, N. (1999). Heuristic principles for the design of artificial neural networks, *Information and Software Technology* **41**(2): 107–117.
URL: *https://www.sciencedirect.com/science/article/pii/S0950584998001165*

Wand, M. P. and Jones, M. C. (1994). *Kernel smoothing*, CRC press.

Wang, D., Ding, W., Lo, H., Morabito, M., Chen, P., Salazar, J. and Stepinski, T. (2013). Understanding the spatial distribution of crime based on its related variables using geospatial discriminative patterns, *Computers, Environment and Urban Systems* **39**: 93–106.
URL: *https://www.sciencedirect.com/science/article/pii/S0198971513000185*

Weisburd, D. (2015). The law of crime concentration and the criminology of place, *Criminology* **53**.

Wheeler, A. P. and Reuter, S. (2021). Redrawing hot spots of crime in dallas, texas, *Police Quarterly* **24**(2): 159–184.

Wu, J. T., Leung, K. and Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study, *The Lancet* **395**(10225): 689–697.

Yuan, M., Buttenfield, B., Gahegan, M. and Miller, H. (2004). *Geospatial Data Mining and Knowledge Discovery*, pp. 365–388.

Zhu, Q., Zhang, F., Liu, S. and Li, Y. (2019). An anticrime information support system design: Application of k-means-vmd-bigru in the city of chicago, *Information & Management* p. 103247.