School of Computing
FACULTY OF ENGINEERING AND
PHYSICAL SCIENCES

UNIVERSITY OF LEEDS

Final Report

Time and relative dimensions in spatial tensors: a study on the impacts of covariates and deep learning techniques on time series forecasting

Imogen Emily Fleur Mackrell

Submitted in accordance with the requirements for the degree of
Computer Science (Digital & Technology Solutions) BSc

2022/23

COMP3932 Synoptic Project

The candidate confirms that the following have been submitted.

Items	Format	Recipient(s) and Date
Final Report	PDF file	Uploaded to Minerva (16/05/23)
Web address of externally hosted supplementary code & data repository	URL	Sent to supervisor and assessor (16/05/23)

The candidate confirms that the work submitted is their own and the appropriate credit has been given where reference has been made to the work of others.

I understand that failure to attribute material which is obtained from another source may be considered as plagiarism.

E. MACKRELL.

Summary

In this study, performed in the domain of spatiotemporal crime forecasting, two binary variables are isolated: the presence within input data tensor T of crime-agnostic social covariate data; and the use of deep learning techniques in the modelling process. The focus of the study is upon the interaction between the values of these two variables, and how careful choices thereof can influence both the accuracy and mean bias—with ideal values of which dictated by the desired behaviour of the resultant model.

A priority is also held throughout the study for making careful and responsible choices in design in order to minimise the risk of bias within created models, as mitigation of the potential social and ethical complications of spatial crime forecasting is deemed paramount.

Acknowledgements

I'd like to thank so many for their kindness and support
And all their help with this great project—listed here, in short:
Professor Netta Cohen, for her frequent kind advice
and many patient meetings, be they lengthy or concise;
My wonderful dear family, who for my whole degree
Have given me such love, support, and endless cups of tea;
And little fluffpants Murphy of the dynasty McScragg
Who brings us joy and rodents (or whatever he can snag);
My wonderful friend Anna, who through all this kept me sane
And gave me such encouragement in this lasting campaign;
The right inspiring Daniel, all the chats and help from whom
Made such a daily difference in the Bragg SoC Study Room;
And all my lovely colleagues back at PwC
From whom I learned so very much when placed there in Year Three.
I hope that my acknowledgements sufficiently convey
My gratitude for your support you've given all the way.

Contents

Chapter 1

Introduction

Machine learning and its applications have, in recent decades, been the subject of a great deal of advanced research across a wide array of fields in previous decades. It is not infrequent that such applications, being based upon datasets formed on many observations of variables over time, find themselves classified as *time series forecasting*: the task of a model, given a series of past observations over time, to predict future observations directly therefollowing. It is the aim of this paper to explore facets of this technique within a criminological application.

Many constabularies and police forces across the world have, within roughly the past fifteen years, developed or adopted "predictive policing" techniques, in which past crime data is used to identify trends and therefrom forecast potential sources of crime in the near future (1). This can involve the prediction of either likelihoods of particular individuals offending in the future (the *individual* model) or future crime rates within geographical regions (the *spatial* model). While both the individual and the spatial have found use within the U.K. (1), this paper examines strictly the latter of these types. Both the spatial and individual models are controversial concepts, however: in fact, shortly before submission of this paper, the European Union voted in favour of legislation to restrict and prohibit such techniques as a part of its "AI Act" regulations (2). It is consequently a goal of this study to, identify the sources of such ethical concerns and address them by minimising aspects that pose ethical dangers.

Crime forecasting models may use a variety of datasets on which to train (*id est* use as a basis for making predictions). This almost always involves a set of already recorded crimes for the target area, and in some cases may additionally involve additional "crime-agnostic" covariate data (for example age and sex in an individual model) in order to make the prediction algorithm better informed and, ideally, more accurate (3). As well as the presence of such data, another varying factor across crime forecasting applications is the type of model itself. Developments in machine learning over the past decade have made deep learning—the use of *artificial neural network* models that emulate biological brains—increasingly viable in the field of crime forecasting (4). Little literature exists, however, investigating the interactions between inclusion of this agnostic data and the differing machine learning techniques in spatial predictive policing.

The aim of this project is hence laid out. These two variables—covariate presence and deep learning—have a potentially fascinating relationship, and within this study it is hoped that this relationship may be explored within the context of spatial crime forecasting. All the while, efforts are made to approach the issue with a degree of responsibility and care such that any forms of bias that could pose potential material harm in application are minimised. In examining metrics from a practical investigation of introducing geographical socioëconomic features to crime forecasting models, as well as contemporary neural network techniques, new insights and suggestions to future researchers are hoped to be formed, alongside conclusions drawn from this study's Question: *just how do these two variables interact?*

Chapter 2

Police cadets and neural nets: the state of the art and the Question

In this chapter, alongside definitions of the core concepts of this study, a literature view is presented across the field of crime forecasting. For this paper, such an overview of undertaken background research takes two forms: firstly, the frontier of understanding of time series forecasting is explored, from its theoretical foundations to its applications; and finally, the specific application of crime forecasting is then investigated, from its current adoption to its ethical facets.

2.1 Time series forecasting

A *time series* is any data structure for which one axis has been explicitly assigned a temporal context. Hence, to *forecast* a time series is to learn from its data and predict future values outside of its original definition. Time series also have variant attributes: if a time series is *multiple* then it contains a time series structure for the same variate of each of multiple observations, and if it is *multivariate* then an observation can contain a time series of data for each of multiple component variates (which, if independent, may be described as *covariates*). This study utilises a multiple multivariate time series, in which each time series describes data from a dependent variate and a set of covariates, and such a time series exists for every geographical unit used in the study. This is visualised in Figure 3.2.
In this section, the state of research into the process of time series forecasting and the optimisation thereof is discussed.

2.1.1 Temporal data characteristics

By adding an explicit order to data, as occurs in the addition of a temporal axis in definition of a time series, several unique characteristics can arise. One such characteristic is the *trend*: should the mean of fixed-length "windows" of values drift in one direction as the starting point of such windows increases, then the time series can be described as having a trend. Another unique characteristic of time series is *seasonality*: if the context by which the time series is defined is, say, weeks, then one may expect to observe, for example, that purchases of pancake ingredients temporarily spike with a regular frequency—a season—with a period of roughly 52 weeks. Existence of such a pattern denotes seasonality in a time series. Finally, alongside trends and seasons, time series may also possess *cycles*: instances of rising and falling means not unlike seasons, only without the regularity thereof. These qualities can all have an adverse effect on the accuracy of models trained on time series, and so the ability to identify such patterns in time series can be vital. (5)
All three of these qualities relate to the concept of *stationarity*—more specifically, *strict*

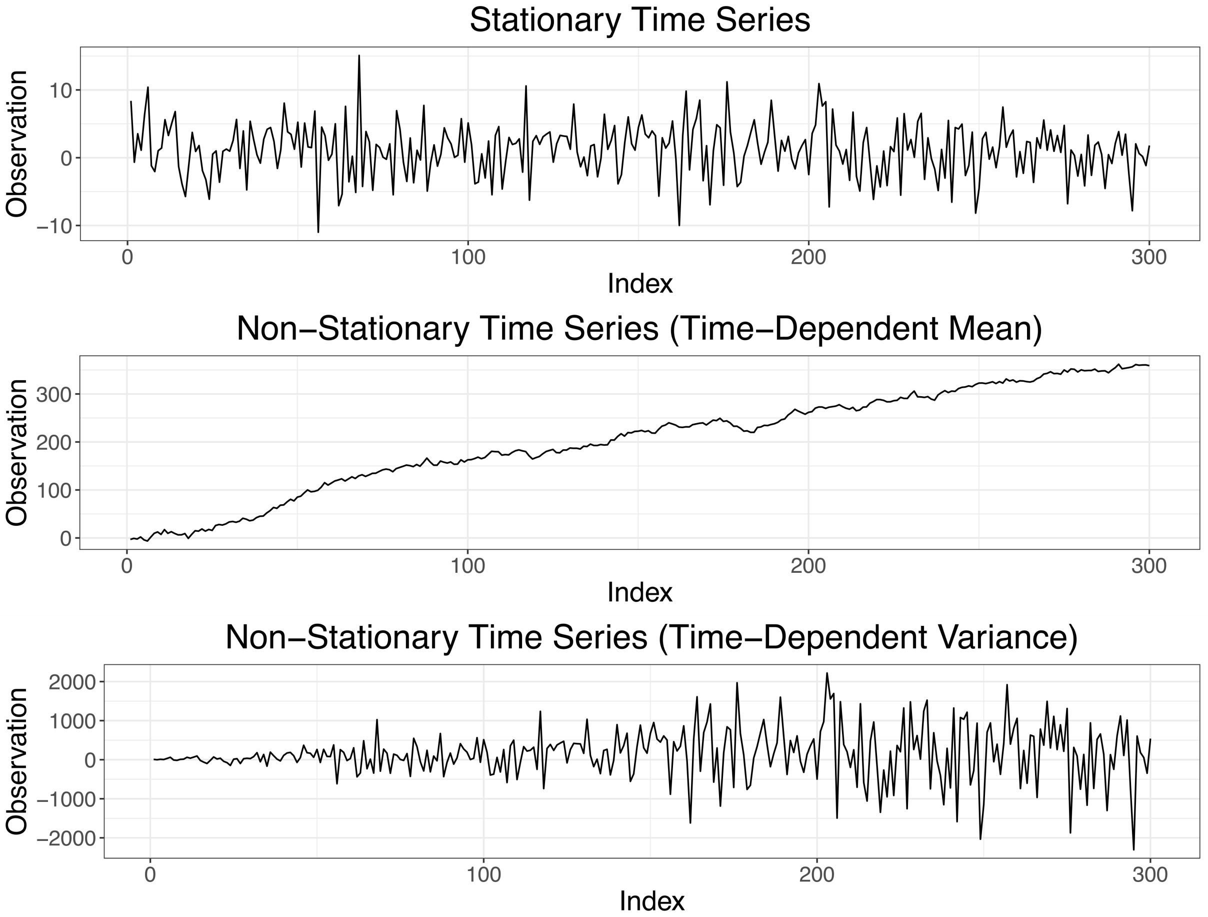

Figure 2.1: A visualisation of stationarity in time series. Note the "white noise stripe" appearance of the stationary set when compared to its neighbours, as a consequence of its static mean and variance. Source: (6).

stationarity[1]. This is the quality of a time series that denotes the inability to express values therein as a function of their time—*i.e.* data with a static mean and standard deviation with respect to time. (7) When visualised, as in Figure 2.1, stationary time series can be considered to appear at a glace like a stripe of white noise values. While nuances of performance depend on models, with some designed with a mitigating resilience to nonstationarity (8), it is usually the case that time series forecasting models perform better on stationary datasets, with, say, decision trees benefitting from stationarity as a result of their simple "one variable, one constant" conditions (9), and deep learning models similarly being defined on an assumption of stationarity (10). It is hence important for any time series in this study to be treated in the case of nonstationarity until such a quality has been removed. This process is described in §3.3.13. Before stationarisation may be fully addressed, it is perhaps essential to distinguish between *strict* stationarity, *trend* stationarity, and *difference* stationarity. While strict stationarity is the goal of this process for the aim of optimal model performance, mere trend and difference stationarity are both forms of *non*-stationarity and are undesirable for the purposes of this study. "Trend stationarity" implies both a lack of unit roots and, in the absence of strict stationarity, the existence within the process of a function of time underlying each observation—*i.e.* of a trend. Conversely, "difference stationarity" implies the presence of unit

[1] Henceforth, "stationarity" without further clarification is used as an abbreviation for the strict variety thereof.

roots[2] within the process, and herewith the necessity of differencing to remove this feature (12). It is for the reason of this distinction that it was concluded that stationarisation should occur using two parallel tests, allowing for not just the identification of time series that are likely stationary, but the specific identification of the *type* of nonstationarity as a means of informing the best transformation to remedy this fact.
For testing for the presence of unit roots, the Augmented Dickey-Fuller (ADF) may be employed (13), and for testing for trend stationarity, the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test (14) offers potentially useful insights. The two tests have opposing null hypotheses: ADF tests the hypothesis, denoted in this paper as H_0^A, of the presence of a unit root, and KPSS that, H_0^K, of trend stationarity. A time series could potentially be treated to make static a varying distribution thereunderlying in one of two ways: *differencing* replaces each value with its difference from its preceding neighbour, while *detrending* calculates and removes a trend—a process without any one standard method (7).

2.1.2 Application of models

Of course, time series forecasting by its very nature operates on regression: data, as a series, are used in order to inform future values of the same type. This limitation immediately narrows the field of potential models, however there exist many forms of regression model, and one perhaps must distinguish between those techniques used for time series forecasting.

- *Exponential smoothing:* posited originally by Poisson as an extension of his window function for smoothing data (15), exponential smoothing methods specifically in the domain of forecasting predict future values of a time series based on past ("lagged") values of the same series. This is performed by modelling the series as a sum of weighted means as the titular window moves across values of the series. The model is temporal in the way that it prioritises more recent observations by logarithmically reducing values as they move away from the window (16). However, as such methods use only lagged data from the same series to inform predictions, it follows that they are solely univariate, and so are not useful in this study in which covariate presence is a dependent variable, and not considered for use.

- *Decision tree:* despite being one of the most basic and commonly used forms of machine learning, decision trees are not a traditionally utilised model for time series forecasting (17). As series of rules comparing a single variable to a constant at every node from the root to decision leaves, they are relatively cheaply created and very easy for humans to understand and interpret after fitting, which is performed usually through the *iterative dichotomiser* (ID3) algorithm, its popular extension C4.5, or the *classification and regression trees* (CART) algorithm (18). While the former pair are strictly for classification problems and so not applicable to time series forecasting, the latter may be used. Of note also is the *random forest* ensemble approach, in which a number of differing

[2] Unit roots are a feature of such processes that are undesirable for this study as a consequence of their implications for stationarity. While mathematical derivation thereof is outside of the scope of this paper, they may be understood, for any process as expressed as a series of monomials, as any root of a monomial thereof that is equal to 1 (11).

decision trees independently forecast the same data, with, in the case of regression, the mean output of constituent trees being that of the ensemble forest as a whole.

- *Autoregressive integrated moving average (ARIMA):* this technique revolves around an *autoregression* model—intuitively, a time series' regression with its lagged self—as a component alongside a *moving average* model—the series' regression instead against lagged forecasting errors—in order to construct a model able to combine the two forms of regression and make a prediction ideally more accurate than either component model's capability (19). While ARIMA itself is a univariate approach by nature, one of many extensions to this quality is *ARIMA with exogenous variables* (ARIMAX), which integrates additional independent covariate time series into models, making ARIMAX potentially applicable to this study.
- *Deep learning:* as a subset of machine learning as a whole, deep learning is a significant field. Involving the application of artificial neural networks to learning models, such an approach can potentially bring increased accuracy when compared to aforementioned alternatives (20), heretofore referred to as "traditional" methods. However, the inherently opaque "black box" nature of deep learning means that such potential increases in accuracy come with losses in interpretability, which for many applications is a significant disadvantage (21). As the presence of deep learning is a dependent variable in this study, it is accepted immediately as a technique, however the specific deep learning model is not decided at this stage of the process.

It is perhaps naïve to select models from this information alone. The context for this study is spatial crime forecasting, and so this context is examined and researched in order to inform a decision on this topic.

2.2 Spatial crime forecasting

Predictive policing is the process of using prior data to inform forecasts of potential future sources of crime. As a given region, under the spatial model, may in itself be modelled as a series of data—number of crime occurrences *per* time unit—across a period of time, it is perhaps intuitive to approach the problem by modelling each region as a time series, and by extension modelling a set of regions as a multiple time series.

In this section, the intersection of time series forecasting and criminology is explored by investigating the use of crime forecasting worldwide and the techniques adopted by differing researchers and constabularies.

2.2.1 Background and contemporary implementations

Since the field of criminology first identified the determinism of crime and its dependence on societal variables, from Becker's formalised establishment of links with economic factors in 1968 (22) to the wider suggestion of links across society by Brantingham *et al.* in 1984 (23), the concept of predictive policing has been an attractive avenue of thought for those considering means of proactive prevention as opposed to reactive resolution, as evidenced by the here discussed examples of its contemporary use. Spatial predictive policing as a concept could perhaps be argued to have been birthed in 1989, following the study of Sherman *et al.* (24) and its finding that in its subject city over one year, approximately half of all police incidents occurred in 3% of discretised locations. This concept of “hot spots” of crime would be repeated across future studies (25), especially with the proto-predictive strategy of using such hotspots as static, if reactive, indicators to inform future police resource allocation (26).

Predictive policing as it is currently understood—*i.e.* using previous data to dynamically forecast rates of crime—developed as a field in the 2000s, with Gorr *et al.* utilising regression models for this end in 2003 (27) and Chen *et al.* comparing both regression and clustering methods in 2004 (28). By 2008, the term “predictive policing” had become popular and the tools themselves were beginning to be frequently adopted by police forces across the world; however, the focus remained on hotspots, classification, or regression (29)—all so-called “traditional” (3) methods.

In the United Kingdom, predictive policing during this timespan has seen quick adoption by nationwide constabularies (1). At least 13 police forces, as of January 2019, responded affirmatively to questions of use of spatial prediction techniques this form as part of research by the political organisation Liberty (30), however precise details regarding these techniques are not publicly available. Kent Police, the first to implement predictive policing techniques, operated under a contract from 2016 to 2018 with U.S.A.-based firm PredPol (31). Elsewhere, West Yorkshire Police worked with University College London in creation of the “Patrol-Wise” tool (32) and answered, in a request under the Freedom of Information Act 2000, that this tool operates as a classifier for regions at risk of elevated crime rates (33). Unfortunately, the types of models used are unknown, and any assumption that traditional machine learning methods are used are informed only by the scarcity of deep learning techniques in the field of crime forecasting in this time period (3).

Of course, no intention is given of portraying the concept that any traditional method is, in itself, outdated or inferior to their counterparts. As of 2020 they remain the most commonly researched approach, and one variant thereof, the random forest method, is frequently found to be the optimal means in this context (3). However, an increasingly popular approach is to utilise deep learning models in the process (4). By introducing such techniques as, say, neural networks, observed differences in the efficacy thereof (34) may be studied, allowing a conclusion to be made regarding which of the two may be more applicable to the context of a given study. Kounadi *et al.* (3) found that the modal choices of the best performing model in crime forecasting studies as were the random forest, multilayer perceptron, and kernel density estimation-based approaches, all of which falling into the category of traditional machine learning. However, this mode is likely a consequence of the composition of the metaänalysis itself, with only 3 out of the 32 papers analysed utilising deep learning. While this perhaps reflects the relative scarcity of such approaches in crime forecasting applications, the fact that it was found that all three of these papers found deep learning approaches to perform better than more traditional counterparts demonstrates that this could potentially be a source for improvement in this study.

2.2.2 Application to the problem

Consequently, one traditional model and one deep learning model are chosen for use. Random forest is chosen for the former category, as its high frequency of use within this field (3) as well as the aforementioned status as a modal choice for best performer suggest a model that is appropriate, well-supported, and highly performing for crime forecasting. Regarding deep learning, the neural basis expansion analysis for interpretable time series forecasting (N-BEATS) model is chosen for use (8). This neural network model, fulfilling the aim of investigating the performance of deep learning models in a crime forecasting context, was chosen for multiple reasons. For one, its explicit purpose as a time series model suits it for he requirements of this study, and its design around mitigation of the problems surrounding nonstationary processes reduces the risk of any of the chosen variates lowering the accuracy of the model. N-BEATS additionally offers support for multivariate time series, suiting the covariate focus of this study, as well as multiple time series, suiting the application of forecasting on many regions.

However, regions as geographical units may take any form, and so it is next vital to precisely define what is meant by a "geographical unit" for the purpose of this study.

2.2.3 Social and ethical concerns

The controversy of crime forecasting, as a part of predictive policing, is perhaps difficult to understate, frequently being discussed in the United Kingdom as a point of concern by popular media (35), civil liberty organisations (30), and Parliament itself (1). While concerns regarding the aforementioned individual model are outside of the scope of this study, it is considered critical by the author to deeply consider and work to mitigate ethical concerns that may surround spatiotemporal crime forecasting.

The concern in question is the fact that since forecasting models exist to identify and replicate patterns that they learn from existing data, an implicit assumption is made in the usage of such techniques that all of these patterns are desired to be replicated. In practice, namely here in the application to crime, there exist patterns of implicit bias in police activity *apropos* to actual crime, for example in regions of differing ethnic makeup (36). Because direct crime data does not exist, only an indirect proxy through police activity, any biases in this proxy's ability to reflect crime will be reflected by the model. This gives the consequence of, for example, a region with levels of police activity disproportionately higher than actual levels of crime as a result of implicit racial bias having this bias identified in the learning process of a model trained on this police activity—a phenomenon observed by Lum *et al.* (37) in the aforementioned U.S.A.-developed product PredPol, later implemented by Kent Police. When such models are used to inform police activity such as resource allocation, this causes the model to act as a truss underlying these biases, identifying them and by nature perpetuating them in their predictions (1). As models in this study utilise covariate social data, there is a risk of, should data selection and preparation be careless, regions of particular economic or ethnic makeups being predicted to experience disproportionate levels of crime, which in some applications could lead to increased police activity therein in the future, in turn causing material consequences for residents. Consequently, it is considered critical for this study for

		OA	LSOA	MSOA	Ward
Quantity		175,434	35,672	7,264	7,666
(%) Relative difference	μ	62.2	62.2	62.2	62.2
	σ	3.3	1.4	0.8	0.6
	min	48.7	57.2	60.7	61.3
	max	74.1	65.9	64.6	63.5
(%) Relative bias	μ	-62.2	-62.2	-62.2	-62.2
	σ	3.3	1.4	0.8	0.6
	min	-74.1	-65.9	-64.6	-63.5
	max	-48.7	-57.2	-60.7	-61.3

Table 2.1: Relative levels of difference and bias in the "Moretti model" between actual crimes and police activity (38) in output areas, lower and middle layer super output areas, and wards, alongside the quantity of each in England & Wales (40).

such risks to be mitigated, as creation of such a tool with the risk of causing material harm to communities affected thereby is not a desirable outcome.

Moretti *et al.* (38) modelled actual crime using data from the Crime Survey for England and Wales, wholly independent from policing, against observed crime, by applying police awareness as a separate factor to this model. Through these means, a model of police bias was able to be created, through which a study of correlation with spatial resolution in mapping studies could be made. Spatial resolutions studied were the use of, listed from smallest to largest, output areas; lower layer super output areas; middle layer super output areas; and wards (39). The study concluded that, while choice of geographical unit does not effect the mean level of bias μ in crime awareness, it has a significant effect on the standard deviation σ thereof. Namely, a negative correlation was found between the size of a model's geographical unit and the standard deviation of its bias, with actual figures available in 2.1 alongside respective counts of each division. Because of the nature of bias in this form being more damaging in cases of volatility (high values of σ) than correspondingly high values of μ, these findings informed the suggestion to researchers that models using police activity as a source of data can reflect actual crime much more accurately and with less bias by decreasing spatial resolution.

The suggestions of Moretti *et al.* were useful to this study, and informed the choice of geographical unit to be used in the assembly of data. While spatial resolution was aimed to be maximised in order to ensure a sufficiently large dataset for training accurate models, it was also deemed important to minimise the variance of bias. Informed by the quantities of each, it was decided that the significantly higher resolution of LSOAs as units than MSOAs and wards, alongside their significantly lower σ-value than OAs, allowed an acceptable compromise between a sufficiently large dataset and sufficiently minimised bias. Herefrom, the unit of space for this study was decided to be the lower layer super output area.

2.3 The Question

Hence, the problem that this study aims to solve is isolated. Both in the United Kingdom and worldwide, crime forecasting—specifically the spatial model—is applied with inconsistent data sources, with no literature quantifying the efficacy of the presence of covariates in such a model. Additionally, with so few published studies utilising the interaction of this criminological context of time series forecasting with deep learning techniques, it may be additionally useful to future researchers should such techniques be applied in conjunction with and without covariate presence in order to gain understanding of the interaction of these two variables. Thus, the Question is posed: *in an arbitrary spatial crime forecasting model, how do crime-agnostic societal features interact with the application of deep learning techniques?*

Chapter 3

The present tensor: design and implementation

The forecasting process includes the selection and collection of data sources, the transformation thereof into a tridimensional tensor, T, and finally the use of T to train four machine learning models and make predictions of future crime data. The aim of this chapter is to elaborate on this process, as well as justifying each decision made in the design and execution thereof.

All work and data described in this chapter and §4.1 is available in the supplementary Git repository that accompanies this paper. This repository is used throughout the process for the purposes of version control and as a backup.

3.1 The process model

It is beneficial before engaging in work for a model to be chosen to act as a guide through structuring the tensor construction process. Two such models that exist are the *cross-industry standard process for data mining*, henceforth known as CRISP-DM (41), and the *knowledge discovery in databases* process, henceforth known as KDD (42). The models have many largely equivalent subprocesses, with the practical portion of the process directly involving the data being in very similar: Azevedo *et al.* (43) argued that CRISP-DM may be viewed as an implementation of its counterpart. This implementation adds the stages of business understanding and of deployment. The former describes the process of acquiring an understanding of the context, requirements, and plan for the project at hand—a process followed throughout Chapter 2. The latter refers to processes of deploying in production environments, which are not planned in this study and so may be ignored. Because of these benefits of the applied context of CRISP-DM over KDD, this study hence adopts the former model, adapted to discard the final deployment stage.

With this modified CRISP-DM model chosen, the study, divided into five phases, may be expressed as such:

1. *Business understanding:* to identify goals, the academic context, and a practical plan;
2. *Data understanding:* to identify, collect, and inspect data sources;
3. *Data preparation:* to clean, process, and transform the source data into a desirable output;
4. *Modelling:* to train and apply chosen machine learning models in order to achieve predictions and, therefrom, metrics; and
5. *Evaluation:* to assess both results and the prior process to identify conclusions.

Stage 1 is addressed in Chapter 2. This chapter covers stages 2–3 of this process. Stages 4–5 are addressed in Chapter 4.

Type	Dataset	Provenance	Licence	Spatial resolution[1]	Temporal resolution	Consistency[2]	Outcome
Dependent	A (45)	Constabularies	OGL v3.0	$r < 1$ (≈ 10cm; co-ordinate)	monthly	—	Included
Covariate	B (46)	FSA (47) ONS (48)	OGL v3.0	$r < 1$ (≈ 10cm; co-ordinate) $r < 1$ (≈ 15 addresses; postcode)	biannual	0.917	Included
	C (49)	Census 2021	OGL v3.0	$r = 1$ (LSOA)	static	0.433; 0.692[3]	Discarded
	D (50)	CDRC	Safeguarded	$r = 1$ (LSOA)	static	1.000	Discarded
	E (51)	ONS	OGL v3.0	$r = 1$ (LSOA)	quarterly	0.921; 0.636[4]	Included
	F (52)	EC	OGL v3.0	$r > 1$ (constituency) $r > 1$ (ward)	2015; 2017 annual	—	Discarded
	G (53)	ONS	OGL v3.0	$r = 1$ (LSOA)	annual	0.900; 1.000[5]	Included
Lookup	X (54)	ONS	OGL v3.0	$r = 1$ (LSOA)	static	—	Included
	Y (55)	ONS (48)	OGL v3.0	$r < 1$ (≈ 1m; co-ordinate)	static	—	Included

Table 3.1: A summary of the sources considered for inclusion within the final tensor.

[1] "Resolution" in spatial scope refers to value r in accuracy value $\frac{L}{r}$, with L representing the size of a given LSOA. $r < 1$ denotes that a variable is more accurate than LSOA level, $r = 1$ denotes that a variable is already in terms of LSOAs, and $r > 1$ denotes a variable being less accurate, or on a higher level, than that of LSOAs. By this definition, resolution is inversely proportional to accuracy, and so lower values are considered to be more desirable. Co-ordinate accuracy is inferred from the number of decimal places; postcode accuracy is inferred from an estimated size of postcode units (56). In these two cases, r can be interpreted as a margin of error for placement within the correct LSOA, as described in §3.2.9.
[2] "Consistency" in covariate sources refers to the strength of the correlation with crime in terms of the consistency score given by Ellis *et al.* (44) where such a score was made.
[3] Referring to "teenagers & young adults percentage" (subtable 10.1.3a) and "education of residents, average level" (subtable 10.1.3b) respectively (44).
[4] Referring to "neighborhood conditions" (subtable 10.1.3b) and "income of residents, median level" (subtable 10.1.3c) respectively (44).
[5] Referring to "city (or county) population size" (subtable 10.1.3b) and "age" (subtable 10.1.2a) respectively. Gender may be measured as an aggregate of entries within subtable 10.1.2a, in which case a mean may be taken from 9 variables to be ≈ 0.909. (44)

3.2 Data understanding

In this section, datasets, after collection, are considered for inclusion within T. Collected datasets are within exactly one of three groups: *dependent* variates (of which only A, crime occurrences, is a member); *covariate*s (B–F); and geospatial *lookup* tables (X and Y). Choices of covariates are informed by those identified by Ellis *et al.* (44) as likely to hold a relationship with rates of crime, as well as the availability of such data at an LSOA (or lower) level and the target of reflecting a diverse range of correlates.

The size of T, which is tridimensional, is expressed as a three-value product quantity of floating point values lmv, in which l represents the number of LSOAs included therewithin, m that of months, and v the number of variates accepted for the dataset. This section will find the value of v as well as initial values for l and m that are liable to be reduced (but not increased) during the data preparation stage in §3.3, after which the dimensions of T are $l' \times m' \times v$.

3.2.1 Dependent variate A: number of crime occurrences

Archives of street-level crime information in England, Wales, and Northern Ireland have been provided by His Majesty's Government since December 2013 through the service officially entitled *data.police.uk* (45). Updated archives are published monthly, and each archive spans 36 months of crime data, with the exception of those published before May 2017 which contained all data theretofore available, *i.e.* since 2010. Each archive is made publicly available *via* both application programming interface (API) access and as a ZIP archive of comma-separated value (CSV) files for each permutation of month, constabulary, and type of crime data included. Archives include three types of crime data: street-level crime, "stop and search" incidents, and crime outcomes. As only the former relates to this study, the rest are discarded.

To assemble the dependent variable dataset, the complementary archives from February 2023, February 2020, and February 2017 are downloaded from the service and extracted into the same directory. The dataset does not directly address the variate, however to achieve the number of crime occurrences *per* LSOA and month it would be possible to infer the LSOA for each crime, and sum each match. LSOA information, however, is not universally addressed, with many present null values. To address this problem, instead of discarding such records, the need forms to establish a lookup table to calculate this information for each record, as addressed in §3.2.8.

There do exist problems with using this dataset in its entire form, however. Firstly, the computational complexity of many data transformation and model fitting operations performed in this study entails that it would be infeasible to work upon the dataset at this size using the hardware available for this study. Additionally, inspection of the dataset makes it clear that different constabularies, when submitting their data for inclusion within the set, apply different standards of data quality thereto, with some constabularies missing columns entirely. This could cause rifts in prediction quality along constabulary borders. To mitigate both of these problems, the size of T is limited from the beginning of the process by selecting only data falling under the jurisdiction of West Yorkshire Police. The two dimension size variables l and m are initially dictated by this decision: the number of LSOAs in T is represented by the number of LSOAs reported by this constabulary, $l = 1,514$, while the number of months therein is dictated by the length of time that West Yorkshire Police has data available: $m = 135$.

Data from *data.police.uk* are anonymised by the constabularies that submit them, and cover street-level—not domestic—incidents. The set is published under the Open Government Licence (OGL) v3.0, allowing use of the data for this study.

3.2.2 Covariate B: number of public houses

As the correlation between density of establishments that serve alcohol and (particularly violent) crime in the environs thereof is globally established (57), it is hypothesised that including temporal data regarding this density may improve the accuracy of models based thereupon. In assembling this dataset, the Food Standards Agency (FSA) keeps records of all such establishments publicly available in its hygiene ratings. The FSA updates its ratings as frequently as local authorities publish the relevant data (47), however historical records are not retained, making direct formation of a time series impossible. However, the GetTheData project has, since 2016, biannually collected and published this data in the scope of public houses themselves as *Open Pubs* (46), thus allowing for temporal analysis. Like variate A, the variate is not directly addressed in the dataset, but may instead be inferred by a similar counting procedure.

The dataset has flaws: many public house records (including all from the first 2016 edition) are without co-ordinate data, and the resolution—biannual without any data from the first four years of the dataset—is very low. However, the latter issue is partially aided by relaxing the integer requirement for pub quantities and assuming linear trends and static quantities prior to 2016, therefrom interpolating. The former issue is solved by looking up any records without co-ordinate data by the `postcode` column (discarding the small minority therewithout) using Source Y (§3.2.9).

The FSA source for this dataset is published under OGL v3.0, as is *Open Pubs.*

3.2.3 Covariate C: population in full-time education

The Census 2021 project is examined as a source for information regarding rates of individuals in full-time education, as there exist arguments that this factor may be a correlate of crime (44). The results of this study were published as *S01 Census 2021: Usual resident population in full-time education by age 18 to 30 years, local authorities in England and Wales* (49). However, as within the time scope of this study such a measure has only been observed on one occasion, as well as the relatively weak consistency score seen in Table 3.1 making it difficult to justify inclusion for lack of evidence as a correlate, this dataset is discarded for the study.

3.2.4 Covariate D: ethnicity proportion

With ethnicity being a frequently discussed subject within criminology (44), proportions of individuals self-identifying as each ethnicity within each LSOA are compiled by the Consumer Data Research Centre (50). However, in addition to the static nature of the dataset making it of little use to this temporally-driven study and the restrictive licence for use (precise terms of which being obfuscated), this dataset is rejected for a pragmatic end: existing ethical concerns relating to the material consequences of training predictive policing models on ethnicity data, as discussed in §2.2.3, are deemed to outweigh the potential benefit in accuracy of the model.

3.2.5 Covariate E: mean house price

Mean house prices are investigated as a useful and frequently measured proxy of the wealth and neighbourhood conditions of a given area, with both being identified as likely negatively correlated with crime rates (44). Indeed, the ONS publishes quarterly reports of this very nature for every LSOA, with dataset #47 of the House Price Statistics for Small Areas (HPSSA) series satisfying this requirement (51). As such, with the assumption of linearity within the data, the dataset could be added to the tensor relatively simply through extraction from the ONS' spreadsheet and interpolation between the quarterly observations in order to align with the monthly equivalent in source A.

As an ONS dataset, this information is published under OGL v3.0 and so may be used for this study.

3.2.6 Covariate set F: electoral success *per* political party

Attitudes, a great focus of Ellis *et al.* (44), are difficult to measure in a format compatible with this study. Perhaps an obvious choice of source for this information is opinion polls, which are often taken frequently and interpreted as time series (58). However, these are not suitable sources of data for this purpose as they operate on samples and not censuses of at least every LSOA in scope. The alternative therefore must be such a census. However, the ONS' Census do not address the issue of attitudes, making the only remaining alternative that of general elections, local government elections, and referenda, despite political data being unaddressed by Ellis (44), with all of this information published by the Electoral Commission (EC) under the "Elections and referendums" header (52).

The latter complicate the issue yet further: there have only been two referenda in England and Wales since 2010—the Alternative Vote referendum of 2011 and the European Union membership referendum of 2016—both of which addressing greatly different attitudes, making this an infeasible source. Meanwhile, general and local government elections, with spatial resolutions of parliamentary constituencies and wards, operate on a much higher geographical level than the LSOA scope of this study (39). Not only is this spatial accuracy low, but variable also: with constituency and ward boundary and electorate sizes frequently changing upon review and with historical data unavailable (59), inference of LSOA for historical data is impossible. For these reasons, as well as the low temporal accuracy, it is deemed infeasible to use election data and hence attitude measurement is discarded as a potential covariate.

3.2.7 Covariate set G: population estimates

Demographics, specifically population estimates (the counting of members of particular demographics within the population), are perhaps a particularly useful subject for this study, being regularly surveyed and published in time series format by the ONS (60). Population estimates in particular are chosen as a means of approaching some of the most strongly supported correlates in Ellis *et al.* (44): gender and age (specifically, positive correlations with population density, maleness, and youth). It is decided to calculate these two variables as proportions of the local population, lest they effectively become proxies for population as a whole. For this reason, the ONS dataset entitled "Lower layer Super Output Area population estimates" (53) is selected as a source. With observations taken on 30 June annually, interpolation (assuming linearity) would trivially align the data with the target's monthly resolution, and with data gathered by LSOA being an ideal spatial resolution and an OGL v3.0 publication allowing use without difficulty, these variates—precisely, sex ratio and proportion of individuals aged 16–29—are easily included within the tensor alongside total population.

With G being the final variate source considered for inclusion, the final number of variates, including the three originating from this source, can be said to take the value of $v = 6$.

3.2.8 Geospatial lookup X: Lower layer Super Output Areas

Sources A and B, while accounting for spatial data, lack a universal discrete column for LSOA data, which must be inferred through a lookup table. The ONS offers an Open Geography Portal (OGP) service, through which boundaries are published openly under OGL v3.0. This includes the publication of LSOA boundaries, albeit in a variety of forms, dependent on the variables of *resolution* and *geographic extent* (61). As, for the former, the options of “intermediate” or “generalised” sets prioritise file size over accuracy, for this study it is opted to use a “full” set, since non-negligible margin of error in postcode lookups (Table 3.1) is of concern. Regarding geographic extent, it is opted to use “extent of the realm” boundaries (extending to Mean Low Water) over “clipped to the coastline” (extending to Mean High Water), to avoid the necessity of assuming that constabularies would not report crimes in the gap therebetween. Consequently, the dataset *Lower Layer Super Output Area (2021) EW BFE* (54) is chosen, covering both of these requirements for England and Wales. Regarding format options thereof, GeoJSON was chosen over the shapefile standard, as the relative simplicity thereof (62) could potentially be critical when performing expensive operations such as intersection calculations on a large *per*-crime scale. Therefrom, using the contained LSOA boundary data, it would be feasible to calculate geospatial co-ordinates’ corresponding LSOA through the intersection therebetween. As this process is deemed worthwhile for its enablement of LSOA lookup for given co-ordinates, this dataset is accepted into the tensor.

3.2.9 Geospatial lookup Y: Postcodes

A problem posed by covariate B in §3.2.2 is the frequent presence of postcode data but lack of an equivalent in co-ordinate data. For this end, a lookup table must be identified for co-ordinate inference. *Postcode Directory* is published by the OGP service (48), and provides a list of current (and terminated) postcodes alongside spatial information, such as northing/easting, latitude/longitude, LSOA, and ward. This study, in cases of present postcodes and no other information, used the latitude/longitude co-ordinates in order to infer LSOA information. Among all other columns bar the postcodes themselves, the LSOA data themselves from this dataset are discarded and re-inferred, as they are incomplete. This study furthermore uses specifically the variant of the *Postcode Directory* published by the London Borough of Camden, *National Statistics Postcode Lookup UK Coordinates* (55), that combines the former directory into a single CSV file.

A flaw with postcode lookups stands in the fact that there is no basis to support the assumption that postcode units align with LSOA boundaries, and as a consequence it is likely that some buildings used in the process lay in the intersection of a postcode and LSOA boundary, causing the building to be placed in the incorrect LSOA. As postcode lookups were only used in this study in places where no more precise means of locating a building existed, this problem can not be mitigated. It is assumed that the number of buildings in this study that fall into this category is negligible.

The source ONS directory is, as with all other ONS data, published under OGL v3.0, as is the Camden-published variant, and so may be used for this study.

3.3 Data preparation

The accepted sources are compiled into tensor T using an R script, `tensor.R`. This file receives its six inputs (sources A, B, E, G, X, and Y) and sanitises, filters, and transforms each of them before binding into a tensor, transforming for standardisation and stationarity, and writing to files. This process is summarised in Figure 3.1, and is briefly elucidated in this section.

3.3.1 Preparation

Preparation involves the loading in of the dataset by R and initial operations to prepare data for transformation. Perhaps most simple is the transformation of the population estimate dataset in source G: in extraction of the relevant columns of each spreadsheet, specifically the total, female, and 16-29 variables can be selected and stored separately in the process, effectively splitting the source set into three variates. By performing successive full outer joins on observations of population estimates, the sheets for each variate can be effectively compiled into one, with gaps being able to be interpolated later in the process.

This join process is unnecessary in cases of CSV sources, however. By using the `plyr` package's `ldply` tool to load in these cases of CSV data, for datasets consisting of multiple files, a union is automatically formed by simply passing a list of files in the dataset. Likewise, for the case of multiple spreadsheets as with G, such a list can be generated using a regular expression before using the base `lapply` functionality to apply the titular function of the `read_excel` package over this list in order to form a union. This technique is especially useful when preparing B, which consists of four CSV files, each representing a different observation, with largely identical schemata: by adding a column to the table containing the corresponding file path for each record, the month can easily be inferred from the string filename.

3.3.2 Cleaning

While Table 3.1 could perhaps suggest that cleaning was a single process that occurred once for each source, in practice this process occurs not just at this stage but frequently, to an extent, thereafter: whenever an object or column or record is unnecessarily included, it is removed and freed from memory at the earliest possible stage. This is a necessary stage due to the scale of the present data: to avoid operations taking longer than would be an ideal amount of time to process, it is necessary to regularly prune the dataset at each stage possible.

This can take several forms. Most frequently, columns or rows are dropped for efficiency of renamed for convenience: often, this takes the form of trimming the domains of LSOAs (from size $l = 1,514$ to eventually $l' = 1,475$) for which insufficient data exists, through joining, intersection, or selection. Thereafter, further joins are applied with this reduced domain for other sources, thus ensuring consistency across the LSOA domain. Duplicate records of crimes, present perhaps as a result of human error by the constabularies submitting them, also must be removed through cleaning, hence the inclusion of crime IDs until (but not after) this point. Finally, R's dynamic typing necessitates particular focus to be paid to the data type of any variable throughout the process, and so cleaning often involves conversion between such types, be it parsing strings to the `Date` format for month inference or using tuples of floats as co-ordinates in order to generate spatial point data for LSOA inference.

3.3.3 Inference

This process applies only to covariate B, for which a proportion of records lack co-ordinate data but not postcode data, requiring the latter to be used to infer the former. This is the

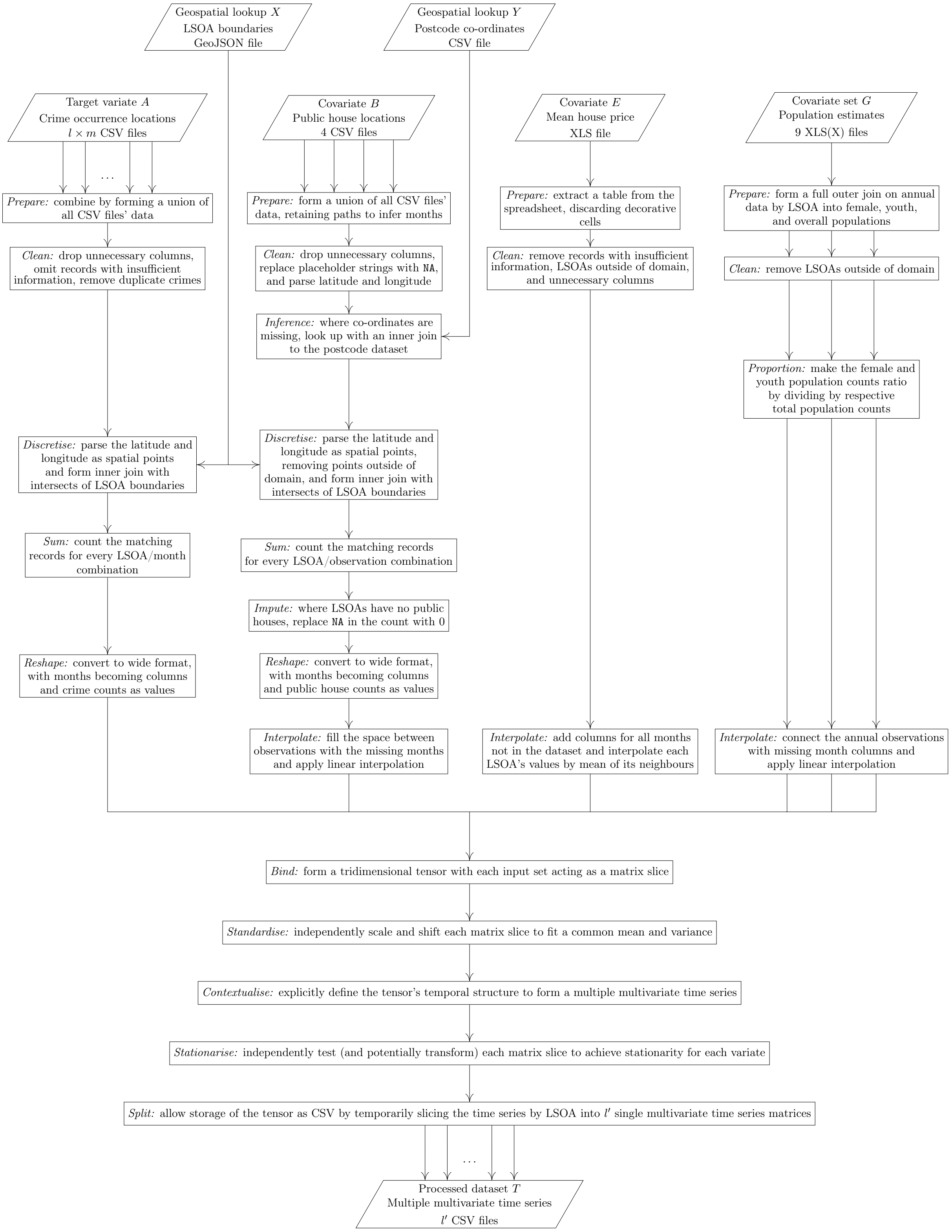
Geospatial lookup X
LSOA boundaries
GeoJSON file
Geospatial lookup Y
Postcode co-ordinates
CSV file
Target variate A
Crime occurrence locations
$l \times m$ CSV files
Covariate B
Public house locations
4 CSV files
Covariate E
Mean house price
XLS file
Covariate set G
Population estimates
9 XLS(X) files
...
Prepare: combine by forming a union of all CSV files' data
Prepare: form a union of all CSV files' data, retaining paths to infer months
Prepare: extract a table from the spreadsheet, discarding decorative cells
Prepare: form a full outer join on annual data by LSOA into female, youth, and overall populations
Clean: drop unnecessary columns, omit records with insufficient information, remove duplicate crimes
Clean: drop unnecessary columns, replace placeholder strings with NA, and parse latitude and longitude
Clean: remove records with insufficient information, LSOAs outside of domain, and unnecessary columns
Clean: remove LSOAs outside of domain
Inference: where co-ordinates are missing, look up with an inner join to the postcode dataset
Proportion: make the female and youth population counts ratio by dividing by respective total population counts
Discretise: parse the latitude and longitude as spatial points and form inner join with intersects of LSOA boundaries
Discretise: parse the latitude and longitude as spatial points, removing points outside of domain, and form inner join with intersects of LSOA boundaries
Sum: count the matching records for every LSOA/month combination
Sum: count the matching records for every LSOA/observation combination
Impute: where LSOAs have no public houses, replace NA in the count with 0
Reshape: convert to wide format, with months becoming columns and crime counts as values
Reshape: convert to wide format, with months becoming columns and public house counts as values
Interpolate: fill the space between observations with the missing months and apply linear interpolation
Interpolate: add columns for all months not in the dataset and interpolate each LSOA's values by mean of its neighbours
Interpolate: connect the annual observations with missing month columns and apply linear interpolation
Bind: form a tridimensional tensor with each input set acting as a matrix slice
Standardise: independently scale and shift each matrix slice to fit a common mean and variance
Contextualise: explicitly define the tensor's temporal structure to form a multiple multivariate time series
Stationarise: independently test (and potentially transform) each matrix slice to achieve stationarity for each variate
Split: allow storage of the tensor as CSV by temporarily slicing the time series by LSOA into l' single multivariate time series matrices
...
Processed dataset T
Multiple multivariate time series
l' CSV files

Figure 3.1: An abstracted summary of the processing pipeline performed on the source data to form T. Horizontally aligned (and identically titled) processes serve similar functions.

justification for the inclusion of source Y within the dataset: by performing an inner join between the set of public houses without co-ordinates and the postcode lookup table, irrelevant postcodes are discarded, public houses without a valid postcode (and hence no spatial information whatsoever, rendering the record without use) are discarded, and public house postcodes are otherwise matched with co-ordinate data.

3.3.4 Proportionment

In this stage, the sets of total population counts, female population counts, and youth population counts are combined to give the latter pair context by proportioning them as ratios of the former. This is simply done by dividing the sum of youths and women in each LSOA by the corresponding total population, and in doing so they gain their own significance as opposed to merely acting as proxies for total population information.

3.3.5 Discretisation

In this stage, co-ordinate data, which is continuous and unsuitable for a multiple time series by definition, is discretised through calculation. Lookup table X includes detailed definitions of the boundaries of every used LSOA, and from the spatial point of each crime and public house, itself inferred from a tuple of longitude and latitude floats, a list is returned by the `sf` package's `ST_Intersects` function of each LSOA with which it intersects. This list necessarily cannot have more than one member, as LSOAs, as divisions, are complementary, however they may be empty in cases of crimes or public houses listed, for whatever reason, as being outside of the extent of the realm (§3.2.8). This list is stored as a column with crime and public house data, which in turn is the subject of a full outer join with the LSOA IDs of X and a selection in order to clean spatial information and only leave behind the ID of each record's corresponding LSOA, thus completing the conversion from continuous co-ordinates to discrete LSOAs.

3.3.6 Summation

As sources A and B both store, for each observation, *occurrences* of their subject, they must be transformed into a tabular format with axes of months and LSOAs. This process is subdivided between summation and reshaping (§3.3.8). Firstly, the sources are grouped and summed: in other words, the number of instances of each combination of months and LSOAs, as matched by `dplyr`'s `group_by`, is summed by `plyr`'s `summarize` function, leaving a long format table with three columns, respectively for months, LSOAs, and sums.

3.3.7 Imputation

Public houses, after summation, must also be imputed: as LSOAs without any public houses throughout the whole time period had no matches under `group_by`, their month was summarised as `NA`. For these cases, a record for a single arbitrary month is added with a sum of 0 is added before reshaping to allow them to be included. After reshaping, in these cases, imputation is completed by directly adding 0 values for the four observation months once these columns exist.

3.3.8 Reshaping

As summed (and potentially also imputed) datasets contain three columns instead of $m + 1$, it is necessary to complete the process of transforming occurrence-based datasets into tabular time series by reshaping. As the latter requires a *wide* dataset, the current *long* format must be corrected, and this can be swiftly done using the `spread` function, found in the `tidyr` package. This allows a transposition of sorts within the table with months acting as a key and the sums as values, thus turning the former into column headers and the latter into values for corresponding months.

3.3.9 Interpolation

Before the tensor may be initially assembled, one problem remains across the component datasets: while each has a length of $l' = 1,475$, *i.e.* spatial resolution is consistent, the widths vary greatly as a consequence of inconsistent temporal resolution[1]. This can be rectified by adjusting other sources to match that of the target variate, giving each a length of $m = 135$. As this target is annual, the most regular of the four, the process may simply involve interpolating the other three to match. This involves adding as columns vectors filled with `NA` values of length l' to each, corresponding to those months that are missing relative to A. With these "gaps" now filled with `NA` values, interpolation may commence. This is calculated as such: when interpolating time series S, given two numeric (*i.e.* not `NA`) points S_a and $S_{a+\delta}$, where every point in set $\{S_{a+1}, \ldots, S_{\delta-1}\}$ is `NA`, interpolated value $S_{a+i} = S_a + \frac{i(S_\delta - S_a)}{\delta - a}$. It is important to note from this definition that each instance of interpolation within this section assumes linearity between data: it is acknowledged that this is a simplification without evidence to support that it accurately reflects the time series upon which it is applied. It is, however, used with the hope that it can sufficiently transform resolution without adding outlier values by virtue of the fact that it follows from this definition that interpolated values must be within the range of their neighbours, or more precisely that for all numeric values $S_a, S_{a+\delta}$ with only `NA` values therebetween, where integer $i < \delta - a$, it must be the case that $S_a \leq S_{a+i} \leq S_{a+\delta}$.

3.3.10 Binding

The binding stage represents the initial construction of the tridimensional tensor from its bidimensional component matrices: the merging. In R, this is initially done by forming an array, equivalent to a tensor: by using the base `sapply` function to apply the identity function to a vector of datasets the former function can be effectively used as a proxy to return its data untouched, however restructured as an array through the `simplify = "array"` argument prompting internal use of its adjacent `simplify2array` function. After this has been executed once, creating the array, the `abind` function for array concatenation can be used to add further bidimensional matrices.

The output of this process is a tensor, visualised by Figure 3.2, with dimensions $l' \times m \times v = 1,475 \times 135 \times 6$.

[1] actual temporal resolutions may be found in Table 3.1.

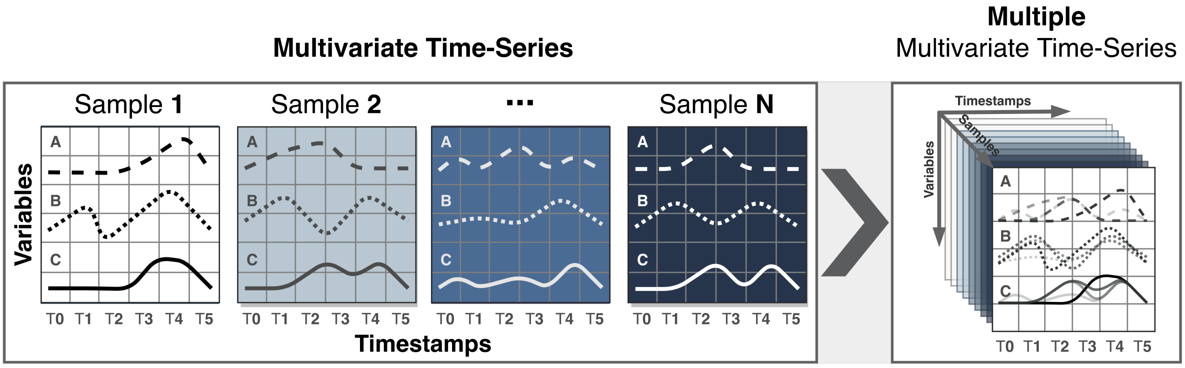

Figure 3.2: A visual representation of the tensor structure comprising the multiple multivariate time series. Applied to this study, each sample is a unique LSOA, TN represents N months after the first in the domain, and A, B, C, &c. represent the variates of crime, public houses, mean house prices, &c. Source: (63).

3.3.11 Standardisation

In the modelling stage, data are used to fit random forest models and neural network models. The latter are accepted to benefit from data standardisation (64), while no consensus on the former could be found, with suggestions that in the case of specifically *regressive* random forest models over classifiers, standardisation has either no or a slight negative effect on accuracy. Weighing these factors, it was concluded that data standardisation was worthwhile in maximising accuracy of models. This involves, for each variate slice V of the tensor, finding the mean value V_μ and standard distribution V_σ across the entire matrix (as standardisation takes place on matrices as a whole, as opposed to individually on records thereof)[2], and applying a standardisation function S on every matrix value x such that $S(x) = \frac{x-V_\mu}{V_\sigma}$. By executing this transformation, each variate's distribution is shifted and scaled to have a common mean and variance, allowing weights across variates to be balanced relative to one another without outweighing.

3.3.12 Contextualisation

Perhaps the simplest and briefest stage of tensor construction, contextualisation involves simply casting each variate as a multiple *univariate* time series in preparation for analysis, using the `ts` casting function of the `stats` package. In doing so, stationarisation can be made more accessible by explicitly defining context such as the month range involved, the frequency of observations, and defining a "unit of time" for the casting function as twelve observations, or one year, allowing annual seasonality to be more easily identified.

3.3.13 Stationarisation

Contrary to the previous operation, stationarisation is perhaps the most complex element of the process. First, ADF and KPSS tests are applied to each variate, before treatment in the form of differencing or detrending is applied. For the purpose of this study, *detrending* as a process computes a linear regression of a time series and from each point subtracts the

[2] These values are stored for later use in reversing this process in §4.1.3.

		Augmented Dickey-Fuller test		Kwiatkowski–Phillips–Schmidt–Shin test	
		Test statistic	p_A	Test statistic	p_K
Total of crimes	$D = 0$:	−2.103	0.534	2.45	0.01
	$D = 1$:	−6.18	0.01	0.052	0.1
	$D = 2$:	−7.926	0.01	0.02	0.1
Mean house price	$D = 0$:	−2.571	0.534	2.445	0.01
	$D = 1$:	−2.475	0.379	0.548	0.031
	$D = 2$:	−5.24	0.01	0.05	0.1
Total population	$D = 0$:	0.322	0.99	2.728	0.01
	$D = 1$:	−4.181	0.01	0.955	0.01
	$D = 2$:	−6.784	0.01	0.102	0.1
Gender ratio	$D = 0$:	−3.362	0.064	1.166	0.01
	$D = 1$:	−3.082	0.126	0.197	0.1
	$D = 2$:	−5.343	0.01	0.033	0.1
Youth ratio	$D = 0$:	−0.17	0.99	2.63	0.01
	$D = 1$:	−6.584	0.01	1.256	0.01
	$D = 2$:	−7.042	0.01	0.107	0.1
Total of public houses	$D = 0$:	−2.613	0.322	0.727	0.011
	$D = 1$:	−2.18	0.502	0.257	0.1
	$D = 2$:	−5.843	0.01	0.034	0.1

Table 3.2: Full results of variate analysis across three levels of differencing, using the Augmented Dickey-Fuller test and the Kwiatkowski–Phillips–Schmidt–Shin test.

corresponding residual (effectively the error of the model at that point), or in other words makes the series a function of each point's distance from the overall trend.

Differencing is implemented by `timeDate`'s `diff` function, while the detrending is implemented manually, with assistance from the `stats` package's `lm` (linear model) and `residuals` tools. It should be noted that as differencing operates on differences between data, a differenced time series will necessarily cover a time range one observation smaller than its source. This implies that a time series differenced multiple times, with D representing the number of iterations of differencing applied, will span the range of $m - D$ observations, and so the tensor must be trimmed from the earliest observations to ensure that each component spans the same range, giving it a final range of $m' = m - D_{\text{max}}$. For the purposes of this analysis, each test was interpreted with a level of significance of 5%. Table 3.2 shows the result of applying each of the two tests to each component slice after D iterations of differencing, and transformations for stationarity for each component variate occur as follows.

- *Total of crimes:* ADF suggests acceptance of H_0^A with $p_A = 0.534 > 0.05$, and KPSS suggests rejection of H_0^K (and hence acceptance of alternative hypothesis H_1^K) with $p_K = 0.01 < 0.05$. As the two tests have complementary null hypothesis, these conclusions agree and strongly suggest the presence of a unit root causing nonstationarity. Hence, the variate is differenced. After this operation has been applied, making $D = 1$, application of ADF suggests rejection of H_0^A with $p_A = 0.01 < 0.05$ (and with it acceptance of H_1^A) and acceptance of H_0^K with $p_K = 0.1 > 0.05$. These conclusions mutually agree that this variate is now very likely to be stationary.

- *Mean house price:* initial test values of $p_A = 0.534 > 0.05$ and $p_K = 0.01 < 0.05$ suggest acceptance of H_0^A and H_1^K, suggesting a unit root's existence, so differencing is applied. However, at $D = 1$ both tests still maintain likelihood of H_0^A and H_1^K, and so another iteration of differencing is applied. Finally, at $D = 2$, conclusions agree at H_1^A and H_0^K, and so stationarity is accepted as likely.

- *Total population:* with initial concurrent conclusions of H_0^A and H_1^K, differencing is applied. At $D = 1$, however, the tests suggest something perhaps more complex: with $p_A = 0.01 < 0.05$, ADF suggests H_1^A and the lack of a unit root, whereas with $p_K = 0.01 < 0.05$, KPSS also rejects its null hypothesis with H_1^K, suggesting nonstationarity. While potentially seeming contradictory, these conclusions may be interpreted as difference nonstationarity (65), and so differencing is applied again to reach $D = 2$. At this stage, the conclusions of H_1^A and H_0^K are reached, indicating stationarity.

- *Gender ratio:* after reaching H_0^A and H_1^K at $D = 0$ and applying differencing, $D = 1$ brings a conversely complex case. This time, both tests accept their null hypotheses, with $p_A = 0.126 > 0.05$ suggesting H_0^A and the presence of a unit root, and $p_K = 0.1 > 0.05$ suggesting H_0^K and trend stationarity. From this, an interpretation of trend stationarity may be made (65). To remedy this, detrending was applied to the time series.

- *Youth ratio:* $D = 1$ is reached after original conclusions of H_0^A and H_1^K, after which a similar dissonance between conclusions as with the total population variate, with $p_A = 0.01 < 0.05$ suggesting H_1^A and $p_K = 0.01 < 0.05$ suggesting H_1^K. With an interpretation therefrom of difference stationarity, the time series is differenced again. At $D = 2$, conclusions are reached of H_1^A and H_0^K, implying that stationarity has likely been achieved.

- *Total of public houses:* finally, and similarly to the gender ratio variate, after $D = 0$ giving H_0^A and H_1^K, $D = 1$ gives H_0^A and H_0^K. Interpreting as trend stationarity, the time series was detrended to achieve strict stationarity.

As the maximum value of D that was used was 2, the tensor's final range after the stationarisation process is equal to $m' = m - D_{\max} = 135 - 2 = 133$. Having stationarised the data, the assembly and transformation of T is complete. With the now-reduced timespan, the final size of T can be found to be $l' \times m' \times v = 1,475 \times 133 \times 6 = 1,177,050$ data.

3.3.14 Splitting

Having constructed T in its final form, it must be written to disk for later use. CSV is chosen as a format for storage due to its simplicity and native compatibility with `Darts`, the Python library used for modelling. A problem is faced here in the fact that CSV by nature stores data up to bidimensionally, and T is of a higher dimension. In order to circumvent this issue, T is stored as a set of bidimensional slices, *i.e.* a single multivariate CSV file for every LSOA. This is a brief process: for each of the l' LSOAs in T, a new file was created and written containing that LSOA's variates and time series as a table, using the LSOA's ID as a filename to allow for retrieval of this information by Python later, for example giving a path for LSOA #E01010779 as `Tensor/E01010779.csv`. Having achieved this, T has been written to disk, and data preparation is hence complete.

With this, design and preparation have been complete. The established implementation may now be loaded and used in order to train models and form predictions on provided test data.

Chapter 4

Look, cover, write, check: training, prediction, and evaluation

In this section, stages 4–5 of the modified CRISP-DM model (§3.1) are undertaken. First, the implemented T is loaded and applied to models in order to form predictions and metrics thereon, before evaluation and discussion may occur.

4.1 Modelling

The process of defining the models for the implemented tensor, fitting them, and applying them for predictions is elected to be executed on local hardware as opposed to using remote web services such as Google's CoLab or Microsoft's Azure Notebooks. This is a consequence of the size of T having the potential to cause these service's processing timeouts to be reached while fitting is occurring, causing difficulty in training models and losses of time. Additionally, the local approach allows for the modelling process to be integrated with the study's version control system for the purposes of reversion and backups. However, the lack of the specialist hardware available with such cloud services means that it becomes additionally essential to optimise the modelling process for the local hardware available.

4.1.1 Loading T

Once writing of the constructed T is complete, work is therefrom performed in Python. This is a consequence of the reduced need for data transformation operations in which R specialises and the increased need for optimised multiple multivariate time series forecasting operations due to scale, which were performed using the Python package `Darts`. The script responsible for this portion of the process, `model.py`, after loading libraries[1], asserts the presence of an ARM processor and available GPU thread, as the code has been optimised for training on a MacBook Pro equipped with an M2 Pro processor, with full utilisation of the ARM architecture-enabled GPU acceleration. Having completed the initialisation process, the main function is called and execution may begin.

Firstly, T is iteratively loaded file-by-file into memory in a dictionary structure, with filenames being used as a proxy for LSOA information. For easier separation in training later, at this point crimes are permanently divorced from the covariates, leaving for every LSOA a univariate time series for crime and a separate covariate tensor (*i.e.* multivariate time series) for other variables. After this, LSOAs are pseudorandomly divided into train and test sets by `sklearn`'s `train_test_split` function, from which the split LSOA set is mapped to respective time series for access by models.

[1] This is ensured by wrapping initialising code and the `main()` function call in a `__name__ == "__main__"` check at the end of the code, so that such processes are not run until the conditional is parsed, which necessarily occurs after libraries have been completely loaded.

A respective 80%/20% split[2] for training and testing is chosen for several reasons. For one, a large overall dataset affords a slightly larger training set and smaller test set than may be used for a smaller overall dataset as training set size is desired to be maximised in order to increase accuracy. Additionally, no dedicated validation set was generated, with the test set being used for this purpose when applicable. While evidence suggests that this may negatively affect the accuracy of the N-BEATS model (66), this was because the random forest model does not require a validation set and so in doing so the same sets may be passed to all models for consistency in analysis, eliminating the potential caveat in evaluation that differences in performance may be a product of random set assignment. This process effectively loads T into one multiple *univariate* time series for crime and one multiple multivariate time series for covariates.

4.1.2 Model definition and fitting

Having loaded T, models must now be defined. Before this stage, it is perhaps important to discuss hyperparameter tuning. This term refers to the heuristic search for optimal parameters in the definition of a learning model, and is accepted to be a critical part of the process of maximising the accuracy of such a model (67). However, due to the extended length of time necessary to mitigate compatibility issues that occurred between the tuning implementation and the the developed models, a decision is made to instead apply arbitrary informed estimates of parameter values supported by similar use cases (68) in order to meet the time constraints required for this study. Although it is likely that these definitions may have produced results less accurate than the optimal examples potentially given by hyperparameter tuning, it is the hope of the author that the test set, used for validation, is sufficiently large to allow the learning algorithm's internal tuning to be optimised to an acceptable extent, and thus make this a valid compromise.

Additional vital points of clarification lie with variables and time divisions. As the two binary independent variables—type of model and presence of covariates—are the subject of this study, four models in total are trained for each combination of the two. The former covers the two types of models selected for the study, random forest and N-BEATS, while the latter either includes solely the crime time series or additionally the covariate time series. *Apropos* to time divisions, prediction takes place at the *present*, with the *past*, which includes it, referring to all data theretofore. The *future* refers to all time after the present in which all predictions are made. In this study, models in fitting and prediction are given no data for the future, meaning that predictions must be made using solely the past, both in the case of the dependent crime variate (which must be the case lest it not be a prediction) and also of the covariates[3].

In applying these time divisions to code, `Darts` operates time series forecasting models with the concept of input and output chunks. The former of which refers to a bounded segment of the past: input chunks must have a finite length as opposed to extension backwards from the present *ad infinitum*, and no data from outside this chunk will be considered by the model on

[2] Explicitly, of 1,475 total LSOAs, 1,180 LSOA time series are assigned to the training set and 295 to the test set.

[3] This is a consequence of this study's focus upon the practical case of forecasting data using everything available at the time of prediction and not afterwards. In real applications, it is perhaps unlikely that the user of such a tool would have access to covariate data from after the time of prediction.

any given prediction. Likewise, output chunks refer to the bounded segment of the future on which predictions will be made: for every element e_i in this chunk, a prediction will be made for the observation taking place i elements after the present. As this study focuses only on predictions for the next month after the end of the dataset, output chunk lengths are set for training in all models to be 1. Regarding input chunk lengths, while such concepts are not valid for this implementation of random forests, N-BEATS models are set to have a length of 60 (which, in the context of T of monthly observations, corresponds to a five-year window). This window is chosen as a balance between optimising for feasible fitting time on the equipment available for this study and a suitable high resolution (3) for desirable accuracy in a crime forecasting application.

Having chosen the parameters for each model, they are defined and trained upon the training and test set. Each model is trained on the training set of crime data, with one each of the types of models also being given the training set of covariate data and N-BEATS models, as required for their architecture, also being provided with equivalent test sets for crime and covariate data for the purposes of validation in the training process. The training process, with GPU acceleration enabled, took approximately one day for all models to be trained, with the neural networks training for 100 epoch iterations each. Using the function of `Darts`, checkpoints of each neural network model are saved after completion of every epoch, so difficulties from any extraneous failure during the training process can be mitigated by resuming training from the last epoch. After each model has been trained, it is written as a Python pickle by `PyTorch`'s `save` capability. These models are available in the provided Git repository for use or inspection by the reader.

4.1.3 Applying the models

Prediction is now possible. First, the test set must be transformed: as no data after the endpoint of T, the present, with respect to training, it is necessary to separate the final observation from the test set before predicting to move the present to one month therebefore, making the formerly final observation period the new target for crime predictions and the data separated a means of checking these predictions against actual values, allowing metrics to be collected. The transformed test sets are passed to each model along with the parameter of prediction for one single observation into the future (and, in the case of the N-BEATS models, the enablement of parallelisation). Henceforth, crime data is the only relevant information and covariates may be discarded, with their applicability for prediction already used. Once predictions have been gathered, they are destandardised for the sake of human interpretability, as standardised values perhaps obfuscate the practical implications of each prediction. This is simply done by using the values previously calculated in §3.3.11 for the crime variate slice V_μ, V_σ to apply $S^{-1}(x) = V_\sigma x + V_\mu$ to each value in each model's prediction set. These destandardised predictions are stored for writing to disk.

Before writing to disk, however, metrics are gathered. For their applicability to regression models and time series, as well as their differing insights into their input, the root-mean-square error (RMSE), useful for assessing the overall accuracy of a given model, and the mean bias error (MBE) as defined by Fox (69), useful for assessing a tendency of a given model to over- or under-estimate, are chosen as metrics. These are calculated by a created function `getMetrics`

	Random forest		N-BEATS	
	Univariate	Multivariate	Univariate	Multivariate
MBE	−0.225	−0.078	0.946	−0.040
RMSE	6.615	3.803	4.209	3.477

Table 4.1: The mean bias and root-mean-square error values, to three decimal places, for the performance of each of the four models on prediction of crime rates in February 2023, across the variables of model type and presence of covariates. $N = 295$.

that returns an array of these respective calculations given predicted data from each model and the corresponding actual, or observed, values. For vectors P and A of predicted and actual results, RMSE was calculated as $\sqrt{\mu((P-A)^2)}$ and MBE as $\mu(P-A)$, with $\mu(R)$ representing the mean value of a results set R, or formally $\mu(R) = \frac{1}{l'}(\sum_{i=1}^{l'} R_i)$.

Finally, these metrics are added to a table along with predicted and actual values and written to `predictions.csv`, thus completing the process.

4.2 Evaluating model performance

After writing of predictions and metrics, they may now be extracted from `predictions.csv` and evaluated. These metrics are compiled in Table 4.1 as a summary of model performance[4]. In terms of accuracy, the calculated root-mean-squared metrics suggest that both variables have simple relationships with accuracy. The presence of covariates, on average, is suggested to reduce the mean error quite significantly (reducing by $\approx 42.5\%$ for random forest models and by $\approx 17.4\%$ for N-BEATS models), making it by a good margin the most impactful variable of the two on minimising error. Similarly, albeit on a smaller scale, the type of model also is suggested to have an effect that similarly agrees across presence of covariates (reducing by $\approx 36.4\%$ for univariate models and by $\approx 8.6\%$ when covariates are included). From this, it is perhaps reasonable to suggest that, in the domain of this study and its constraints, N-BEATS models are slightly more accurate than random forest models and the presence of covariates in supplied datasets is significantly beneficial thereto.

However, this is also perhaps a naïve interpretation. It is the case that different applications of such tools may have different requirements and as such there exists no “best” model with only accuracy as a priority. This is the justification for the presence of MBE calculations in metrics for this study: as MBE is calculated identically to the mean absolute error (MAE) metric, only with each calculated error therein not made absolute, both overestimated predictions and underestimated predictions are taken into account in the calculation of the mean, thereby indicating a model’s tendency to do one or the other.

Attention should potentially be drawn, however, to the fact that bias is not being intrinsically measured, but *mean* bias. Hence, while it may be desired for MBE values to be as close to 0 as possible, suggesting that a model has no tendency in one direction or the other, such a value quite critically does *not* inform conclusions that a model neither overestimates nor underestimates and as such is accurate. Instead, it simply supports the claim that the model

[4] Full predictions for each model for February 2023 alongside actual values are available in Appendix C. Values therein are rounded for conciseness: raw data is, however, available in `predictions.csv`, found in the repository accompanying this paper.

overestimates at approximately the same frequency as it underestimates: it is possible that bias does not exist, but it is also possible that positive and negative bias cancel one another. However, in conjunction with a minimal RMSE, a good combined picture of accuracy and mean bias can potentially be conjured. If the user of such a tool as this desires a priority of avoiding underestimated values—for example, if a constabulary prioritises minimisation of the number of "missed" crimes in allocating resources to LSOAs and cares less for wasted budget—then a suitably positive bias may be desired. This is a relatively similar approach to that of sensitivity and specificity in cases of Type I and II errors in classification models. With these models, inclusion of covariates within the dataset significantly reduced absolute bias levels while maintaining a slight negative mean bias in the case of random forest models. With N-BEATS models, mean bias was also greatly reduced, from a relatively strong positive to a very slightly negative mean bias. However, results were more conflicting on the variable of covariate presence. For univariate models, random forest gave a significantly lower mean level of bias however on the side of underestimating rather than N-BEATS' overestimation tendencies. For multivariate models, mean bias consistently remained lower in magnitude, however being slightly smaller on the same side of underestimating with N-BEATS.

4.3 The Question, reprised

In §2.3, the Question was outlined as such: in an arbitrary spatial crime forecasting model, how do crime-agnostic societal features interact with the application of deep learning techniques? It is the hope of the author that this study provides an answer that at least somewhat elucidates this interaction. Thereon, recommendations may be made to future researchers on this topic. Firstly, a conclusion can be made that presence of covariates as an independent variable in itself can, with proper selection of variates and effective tensor preparation, have the effect of significantly increasing the accuracy of a model. Regarding the use of deep learning techniques, evidence in this study suggests that such techniques may demonstrate an improvement over so-called "traditional" machine learning methods, with N-BEATS outperforming the frequently-cited "ideal" crime forecasting method of random forest, despite the former implementation being unable to be ideally tuned in its hyperparameters in this study. Regarding the interaction itself between these two variables, target levels of mean bias are significant. If a mean bias value is desired to be close to zero, then a recommendation can be made for the use of covariates, which in this model consistently moved this value b to the range $-0.1 < b < 0$. However, if a preference in a particular application for over- or under-estimating crime exists, then the interaction is more complex. Univariate ("pure crime") models were found to be less accurate in this study, however if this tradeoff is deemed worthwhile for the desired level of bias then such a model using the random forest method yielded a negative (underestimating) bias, whereas N-BEATS yielded a stronger positive (overestimating) bias. With these variables' effects in mind, future researchers may find that both traditional machine learning and deep learning methods are worthwhile with their own advantages depending on priorities of the application, and that in a good proportion of cases the inclusion of covariates can be worthwhile. The author hopes that, with this study, deep learning and covariates are justified in their advantages to the goal of crime forecasting.

4.4 Roads ahead

Of course, this study poses many potential avenues for further research. It would be, in the opinion of the author, a wonderful thing for the reader to pursue any such avenues; some suggestions are made in this section.

- *Analysis of covariate importance:* while a static set of covariates are chosen for inclusion in T in this study, this set is, after assembly, treated as atomic for the purposes of practical handling. It is very much possible that the covariates chosen may yield no use to the models applied, thus producing a very different conclusion, or potentially slightly different choices of covariates may have proved significantly more useful. Additionally, intercovariate phenomena are not observed in this study: one may hypothesise that covariates could, say, cancel one another in their use to models. A researcher may find some useful conclusions by investigating the effects of individual covariates and their relationships with one another; perhaps "covariate tuning" through such analysis could be a future possibility in crime forecasting with more research.

- *Application to classification modelling:* many crime forecasting applications in the U.K. remain focused upon "hotspot"-based classification methods (1) that identify given areas of having binary levels of *risk* of crime, as introduced in §2.2.1. As both random forest and N-BEATS models are designed with compatibility for classification methods, perhaps a researcher could expand the scope of this study to such an approach in order to find whether the concluded interactions between covariate presence and model type apply similarly in cases of classification.

- *Introducing accountability:* perhaps a disadvantage of both the random forest model and deep learning models in general (21) is a relative lack of interpretability in exchange for ideally increased accuracy. With a topic of the sensitivity and controversy of predictive policing, it is perhaps important to work to ensure interpretability of results in order to add an element of accountability to the process in the event that material harm is caused by the effects of the model. A special feature of the N-BEATS model is its interpretable architecture (8), which allows outputting of the data of its internal *trend* and *seasonality* stacks. A researcher may find helpful conclusions in using covariate interpretations in their work, which could help both in accountability as well as the previously suggested refinement of covariate preparation and selection.

- *Comparison with human bias:* of course, the role of spatial crime forecasting to inform police resource allocation was formerly performed by humans and remains so in a large proportion of constabularies[5]. Some informative insights may be drawn by interviewing members of relevant constabularies with experience of making such insights regarding the relevance of particular potential correlates of crime and comparing the results thereof with found correlate importance by machine learning models.

[5] This assumes that those that did not respond affirmatively to Freedom of Information requests regarding use of such tools (30) instead are mostly performing this task manually.

Bibliography

(1) Justice and Home Affairs Committee. *Technology rules? The advent of new technologies in the justice system.* HL paper 180. House of Lords, 58th session, 2022.

(2) Fair Trials. *EU Parliament votes for landmark ban on "discriminatory and unjust" predictive policing and criminal prediction systems.* 2023. URL: `https://www.fairtrials.org/articles/news/eu-parliament-votes-for-landmark-ban/` (visited on 05/12/2023).

(3) Ourania Kounadi et al. "A systematic review on spatial crime forecasting". In: *Crime Science* 9 (2020). DOI: `10.1186/s40163-020-00116-7`.

(4) Fatima Dakalbab et al. "Artificial intelligence & crime prediction: A systematic literature review". In: *Social Sciences & Humanities Open* 6.1 (2022), pp. 100–342. DOI: `10.1016/j.ssaho.2022.100342`.

(5) Rob J. Hyndman and George Athanasopoulos. "Forecasting: Principles and Practice". In: 2nd ed. OTexts, 2018. Chap. Time series patterns.

(6) André Bauer. "Automated Hybrid Time Series Forecasting: Design, Benchmarking, and Use Cases". PhD thesis. 2021. DOI: `10.25972/OPUS-22025`.

(7) Rob J. Hyndman and George Athanasopoulos. "Forecasting: Principles and Practice". In: 2nd ed. OTexts, 2018. Chap. Stationarity and differencing.

(8) Boris N. Oreshkin et al. "N-BEATS: Neural basis expansion analysis for interpretable time series forecasting". In: *International Conference on Learning Representations.* 2020.

(9) Aman Arora. *Why Random Forests can't predict trends and how to overcome this problem?* 2018. URL: `https://medium.datadriveninvestor.com/why-wont-time-series-data-and-random-forests-work-very-well-together-3c9f7b271631` (visited on 05/06/2023).

(10) Bohdan Pavlyshenko. "Forecasting of Non-Stationary Sales Time Series Using Deep Learning". 2022. DOI: `10.48550/arXiv.2205.11636`.

(11) James Stock. "Unit roots, structural breaks and trends". In: *Handbook of Econometrics.* Ed. by R. F. Engle and D. McFadden. 1st ed. Vol. 4. Elsevier, 1986. Chap. 46, pp. 2739–2841.

(12) Michio Hatanaka. *Time-Series-Based Econometrics: Unit Roots and Co-integrations.* Oxford University Press, 1996. DOI: `10.1093/0198773536.001.0001`.

(13) David A. Dickey and Wayne A. Fuller. "Distribution of the Estimators for Autoregressive Time Series With a Unit Root". In: *Journal of the American Statistical Association* 74 (1979). DOI: `10.2307/2286348`.

(14) Denis Kwiatkowski et al. "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?" In: *Journal of econometrics* 54.1-3 (1992), pp. 159–178.

(15) Alan V. Oppenheim and Ronald W. Schafer. *Digital Signal Processing.* Prentice Hall, 1975.

(16) Rob J. Hyndman and George Athanasopoulos. "Forecasting: Principles and Practice". In: 2nd ed. OTexts, 2018. Chap. Exponential smoothing.

(17) El-Houssainy A. Rady, Haitham Fawzy, and Amal Mohamed Abdel Fattah. "Time Series Forecasting Using Tree Based Methods". In: *Journal of Statistics Applications & Probability* (2021). DOI: `10.18576/jsap/100121`.

(18) Sonia Singh and Priyanka Gupta. "Comparative study ID3, cart and C4. 5 decision tree algorithm: a survey". In: *International Journal of Advanced Information Science and Technology* 27.27 (2014), pp. 97–103.

(19) Rob J. Hyndman and George Athanasopoulos. "Forecasting: Principles and Practice". In: 2nd ed. OTexts, 2018. Chap. Non-seasonal ARIMA models.

(20) Ahmed Tealab. "Time series forecasting using artificial neural networks methodologies: A systematic review". In: *Future Computing and Informatics Journal* 3.2 (2018), pp. 334–340. DOI: `https://doi.org/10.1016/j.fcij.2018.10.003`.

(21) Supriyo Chakraborty et al. "Interpretability of deep learning models: A survey of results". In: *2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation.* IEEE. 2017, pp. 1–6.

(22) Gary S. Becker. "Crime and Punishment: An Economic Approach". In: *Journal of Political Economy* 76.2 (1968), pp. 169–217.

(23) Paul J. Brantingham and Patricia L. Brantingham. *Patterns in crime.* Macmillan, 1984.

(24) Lawrence W. Sherman, Patrick R. Gartin, and Michael E. Bürger. "Hot spots of predatory crime: Routine activities and the criminology of place". In: *Criminology* 27.1 (1989), pp. 27–56.

(25) Zhuang Yong et al. "Crime Hot Spot Forecasting: A Recurrent Model with Spatial and Temporal Information". In: *IEEE International Conference on Big Knowledge*, pp. 143–150. DOI: `10.1109/ICBK.2017.3`.

(26) Joel Hunt. "From crime mapping to crime forecasting: The evolution of place-based policing". In: *National Institute of Justice* (2019).

(27) Wilpen Gorr, Andreas Olligschlaeger, and Yvonne Thompson. "Short-term forecasting of crime". In: *International Journal of Forecasting* 19.4 (2003), pp. 579–594. DOI: `10.1016/S0169-2070(03)00092-X`.

(28) Hsinchun Chen et al. "Crime data mining: a general framework and some examples". In: *Computer* 37.4 (2004), pp. 50–56.

(29) Walter L. Perry et al. *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations.* RAND Corporation, 2013. DOI: `10.7249/RR233`.

(30) Hannah Couchman and Alessandra Prezepiorski Lemos. *Policing by machine.* Report. 2019.

(31) Patricia Nilsson. “First UK police force to try predictive policing ends contract”. In: *Financial Times, The* (2018).

(32) The Yorkshire Post. *I predict a break-in: Yorkshire police use cutting-edge technology to deter burglars.* Newspaper Article. 2017.

(33) West Yorkshire Police. *Use of AI.* Government Document. Response to a request under the Freedom of Information Act 2000. 2020.

(34) Petra Perner, Uwe Zscherpel, and Carsten Jacobsen. “A comparison between neural networks and decision trees based on data from industrial radiographic testing”. In: *Pattern Recognition Letters* 22.1 (2001), pp. 47–54.

(35) Michael Cogley. “Experts urge ban on Minority Report-style police technology to 'predict crime'”. In: *Telegraph, The* (2020).

(36) Jules Holroyd. “Implicit racial bias and the anatomy of institutional racism”. In: *Criminal Justice Matters* 101 (2015).

(37) Kristian Lum and William Isaac. “To predict and serve?” In: *Significance* 13 (2016), pp. 14–19.

(38) Angelo Moretti and David Buil-Gil. *Mapping the bias of police records.* Report. The University of Manchester, 2021.

(39) Tower Hamlets Council. *A Guide to Census Geography.* Government Document. 2013.

(40) Office for National Statistics. *Census 2021 geographies.* Government Document. 2022.

(41) Colin Shearer. “The CRISP-DM model: the new blueprint for data mining”. In: *Journal of data warehousing* 5.4 (2000), pp. 13–22.

(42) Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. “From data mining to knowledge discovery in databases”. In: *AI magazine* 17.3 (1996), pp. 37–37. DOI: `10.1609/aimag.v17i3.1230`.

(43) Ana Azevedo and Manuel Filipe Santos. “KDD, SEMMA and CRISP-DM: a parallel overview”. In: *IADIS European Conf. Data Mining.* 2008.

(44) Lee Ellis, Kevin Beaver, and John Wright. *Handbook of Crime Correlates.* 1st ed. Academic Press, 2009.

(45) The National Archives. *data.police.uk.* 2023. URL: `https://data.police.uk/`.

(46) GetTheData Publishing Limited. *Open Pubs.* 2022. URL: `https://www.getthedata.com/open-pubs`.

(47) Food Standards Agency. *Food hygiene rating data.* 2023. URL: `https://ratings.food.gov.uk/open-data/`.

(48) Office for National Statistics. *ONS Postcode Directory.* Version 2. 2023. URL: `https://geoportal.statistics.gov.uk/datasets/ons-postcode-directory-february-2023-version-2/about`.

(49) Office for National Statistics. *Usual resident population in full-time education by age 18 to 30 years.* 2021. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/usualresidentpopulationinfulltimeeducationbyage18to30yearsenglandandwalescensus2021.

(50) Justin van Dijk et al. *CDRC Modelled Ethnicity Proportions (LSOA Geography).* 2022. URL: https://data.cdrc.ac.uk/dataset/cdrc-modelled-ethnicity-proportions-lsoa-geography.

(51) Aimee North. *Mean price paid for residential properties by LSOA.* Office for National Statistics, 2023. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/housing/datasets/meanpricepaidbylowerlayersuperoutputareahpssadataset47.

(52) The Electoral Commission. *Past elections and referendums.* 2019. URL: https://www.electoralcommission.org.uk/who-we-are-and-what-we-do/elections-and-referendums/past-elections-and-referendums.

(53) Neil Park. Office for National Statistics, 2021. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/lowersuperoutputareamidyearpopulationestimatesnationalstatistics.

(54) Office for National Statistics. *Lower Layer Super Output Area (2021) EW BFE.* 2023. URL: https://geoportal.statistics.gov.uk/datasets/ons::lower-layer-super-output-area-2021-ew-bfe/about.

(55) Office for National Statistics. *National Statistics Postcode Lookup UK Coordinates.* London Borough of Camden, 2023. URL: https://opendata.camden.gov.uk/Maps/National-Statistics-Postcode-Lookup-UK-Coordinates/77ra-mbbn.

(56) Office for National Statistics. *Postal geographies.* 2022. URL: https://www.ons.gov.uk/methodology/geography/ukgeographies/postalgeography (visited on 01/06/2023).

(57) Traci L. Toomey et al. "The association between density of alcohol establishments and violent crime within urban neighborhoods". In: *Alcoholism: clinical and experimental research* 36 (2012), pp. 1468–73. DOI: 10.1111/j.1530-0277.2012.01753.x.

(58) Politico. *Poll of Polls.* 2023. URL: https://www.politico.eu/europe-poll-of-polls/united-kingdom/ (visited on 05/04/2023).

(59) Boundary Commission for England. *Current constituencies and electorate changes.* 2023. URL: https://boundarycommissionforengland.independent.gov.uk/data-and-resources/.

(60) Office of National Statistics. *Population estimates.* 2023. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates (visited on 05/04/2023).

(61) Office for National Statistics. *Boundary Dataset Guidance: 2011 to 2021.* Government Document. 2011.

(62) Mapscaping. *Converting shapefiles to Geojson.* 2023. URL: https://mapscaping.com/converting-shapefiles-to-geojson/ (visited on 05/04/2023).

(63) Gabriel Spadon et al. "Pay Attention to Evolution: Time Series Forecasting With Deep Graph-Evolution Learning". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.9 (2022), pp. 5368–5384. DOI: 10.1109/TPAMI.2021.3076155.

(64) M. Shanker, M. Y. Hu, and M. S. Hung. "Effect of data standardization on neural network training". In: *Omega* 24.4 (1996), pp. 385–397. DOI: 10.1016/0305-0483(96)00010-2.

(65) Josef Perktold, Skipper Seabold, and Jonathan Taylor. *Stationarity and detrending (ADF/KPSS).* 2023. URL: https://www.statsmodels.org/dev/examples/notebooks/generated/stationarity_detrending_adf_kpss.html (visited on 05/08/2023).

(66) T. Hastie, R. Tibshirani, and J.H. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer series in statistics. Springer, 2009.

(67) Hilde J. P. Weerts, Andreas C. Mueller, and Joaquin Vanschoren. "Importance of Tuning Hyperparameters of Machine Learning Algorithms". In: *Computing Research Repository* (2020).

(68) Unit8 SA. *Hyper-parameters Optimization for Electricity Load Forecasting.* URL: https://unit8co.github.io/darts/examples/17-hyperparameter-optimization.html (visited on 05/09/2023).

(69) Douglas G. Fox. "Judging Air Quality Model Performance: A Summary of the AMS Workshop on Dispersion Model Performance, Woods Hole, Mass., 8-11 September 1980". In: *Bulletin of the American Meteorological Society* 62.5 (1981), pp. 599–609.

(70) *Code of Conduct for BCS Members.* Tech. rep. British Computing Society, 2022.

Appendix A

Self-appraisal

A.1 Reflections

To speak personally, I consider this project a great success. As I shall elaborate, the eventual study borne of my efforts was unlike my initial imagined version thereof, with some elements about which I was excited eventually dropped, but also with some of my favourite parts of the final work not a part of its original conception. Through this project I have learned lifelong skills, connected with truly interesting and passionate individuals, and, while I acknowledge the often sarcastic nature of the term “character-building”, I do truly believe that a long-term project as disciplined and large in scale as this study has developed my perception of what I can do.

On the question of what informed this project, answers are diverse. Of course, knowledge given to me by the School of Computing is invaluable, extending from the joy in chasing an insightful observation in a modelling process with integrity and curiosity as shown to me by my supervisor, Professor Netta Cohen, to the academic discipline so passionately taught by Professor Eric Atwell, who also instilled a spark of excitement into the process of using data mining and machine learning to affect a cause about which one cares. As I complete this project, and with it this degree programme, I have an earnest appreciation not just for the knowledge given to me in processes of data science and machine learning, but for the passion paired therewith by those who taught it.

I also have a great deal of gratitude for my experiences working with PricewaterhouseCoopers. My time working in a data scientist role at the company taught me, through practical experience in a team, a significant proportion of what I know about not just data science problems and machine learning facets, but also the nature of how long-term projects should be approached, from methodology to foresight and planning. While such work of course was collaborative by nature, my own personal projects have also informed my work in this study. Long-term projects that involved great amounts of time spent designing and constructing databases on the live histories of Godspeed You! Black Emperor and black midi were invaluable to this project, with lessons learned on thorough planning, formalised methodology, and, of course, practical data science and research skills.

Finally, perhaps the most significance in information of this project lies with the project itself. By undertaking this project, I was forced to immerse myself in entirely new fields and skills in order to bring my understanding to a level sufficient for this project. For example, before this project I had no familiarity or understanding whatsoever of R or deep learning, and no practical experience of stationarising data. By undertaking a project of this scope, an excellent opportunity for deep research arose, from which I gained skills that I believe were invaluable to the project and will continue to prove invaluable to me through my career. Organically watching my R script finally outputting T was a moment of true joy that may only appear after extended fumbling through a new skill, and the first time `predictions.csv` was

generated so I could see predicted values and metrics was genuinely rewarding. The long and difficult process of often blind learning involved with this study was so very worthwhile!
The skill of academic research was also a great and unfamiliar abyss for me: I particularly found great improvement after finding that metaänalysis papers acted as a wonderful nexus of ideas across the world for comparatively branching through research. Through research such as this, this study provided a wonderful opportunity to develop the core abilities of topic decomposition, effective reading, and healthy sceptical scrutiny.

A.2 Improvements

Of course, this study, like all other studies, was not ideal. For example, the final suggestion in §4.4, relating to a comparison with human bias, was originally a significant part of this project after being posited by my supervisor, Professor Netta Cohen. Sadly, my enthusiasm for investigating a human element of this study's Question was insufficient in making it feasible: as time progressed, it became clear that the process of ethical approval of questionnaires had been insufficiently considered in preparing for the study, and I had underestimated the amount of time that should have been dedicated thereto. Consequently, the plan was dropped from the study.
In fact, time was a common factor in many difficulties that arose. I had, in the planning stages of the study, created a Gantt chart of the tasks required to complete the project, included here in Figure A.1. However, while I had decomposed the study into corresponding lengths of time, such lengths were perhaps naïve. Data selection and preparation, in the end, took significantly more time than the three allotted weeks. Additionally, by allocating too much time to some other tasks (such as one week for prediction or a fortnight for model training) I had reduced the relative amount of time assigned to more lengthy processes, making my ability to work to the schedule impaired.
A consequence of this was the lack of explicit hyperparameter tuning in N-BEATS models. As a result of compatibility issues between my implementation thereof and the Python installation on my computer, the tuning process could not succeed. Because of the increasingly pressing time constraints arising as a result of aforementioned errors, after a period of attempting to remedy the problem it became infeasible to continue and the study had to continue without hyperparameter tuning. While thankfully some conclusions that I consider at least somewhat insightful were able to be made, the efficacy of the models was impacted by this, and with that I was unable to represent deep learning models in a fair light, but instead with arbitrary and likely suboptimal hyperparameters.
Through this experience, I have learned that even time allocation deserves its own time for research to ensure that proper periods are allotted, and therewith that work may be more efficiently performed to schedule. I would like to revisit this project and apply hyperparameter tuning properly; to see the deep learning models performing optimally and the effect on my conclusions of such performance would very much interest me.

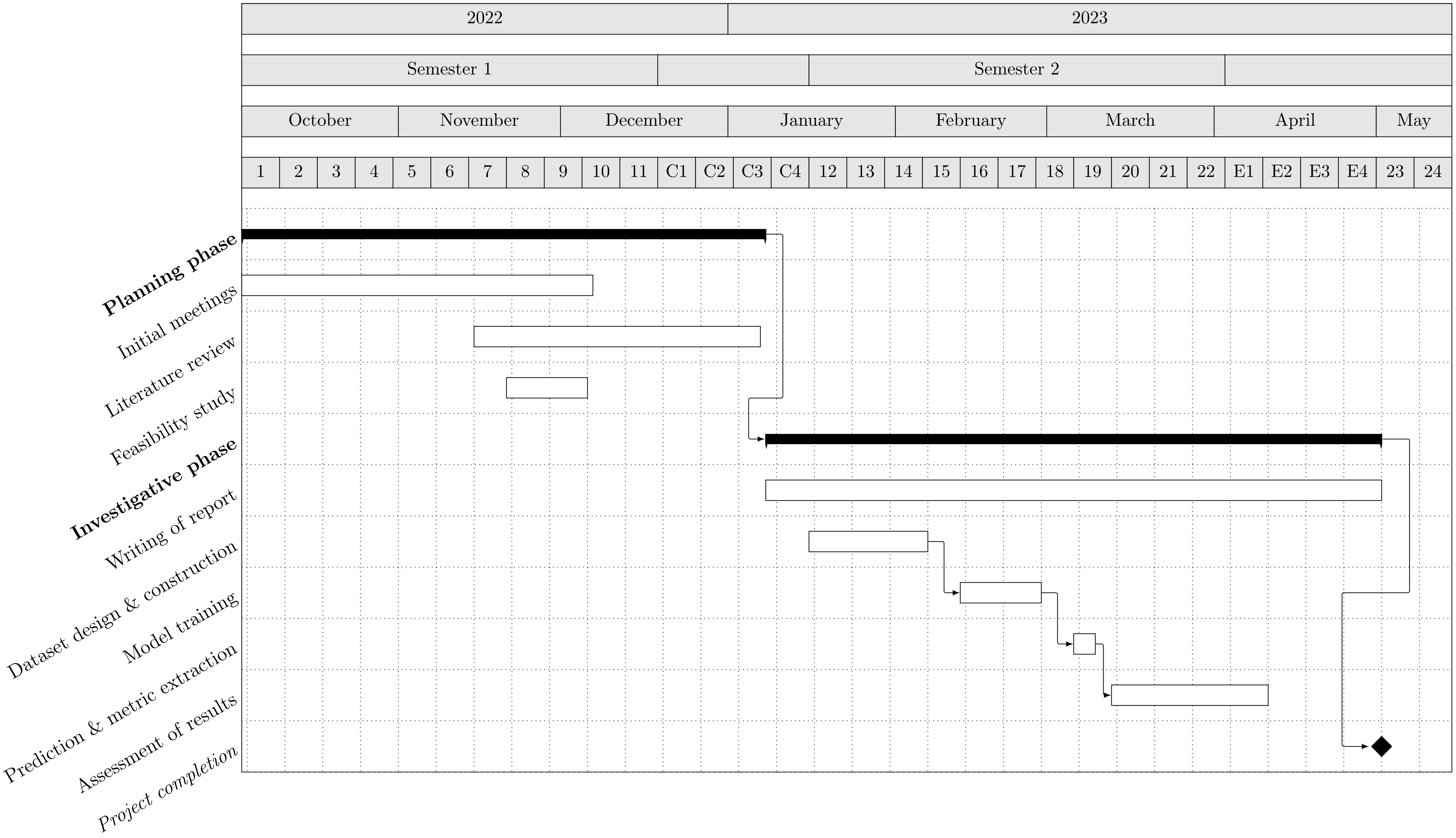

Figure A.1: The Gantt chart created in the early stages of the project decomposing the tasks thereof.

A.3 Implications

A.3.1 Legal issues

In every instance of usage of external material, consideration was made regarding the right for use in this study. A list of such materials may be found in Appendix B.

Every dataset examined with the exception of source D was released under the Open Government Licence v3.0. This licence is delivered by The National Archives, and grants the right to publication, distribution, adaptation, and exploitation of all information published thereunder, as long as proper attribution is made. Such terms have been followed, with attribution given in the bibliography and §B.1, statement of the licence made where possible, and the data distributed in the accompanying repository for this study and exploited in the creation of tensor T. Source D used a proprietary "safeguarded" licence, however as the dataset was rejected for inclusion in T in §3.2.4, no action needs to be taken.

In terms of libraries, each that was used in the study is declared in §B.2, with a justification for inclusion and URL for source information correct at the time of submission. Every library used is released under a free-use licence.

A.3.2 Social issues

The only social issues in this study are the societal implications of crime forecasting, discussed in §A.3.3, and the issue of personal data being potentially included in the data either used as a part of, or published alongside, this study. Every data source distributed in the accompanying Git repository of this study has been published by His Majesty's Government in an anonymised form. Additionally, in order to further reduce the feasibility of inference of personal information therefrom, all information was removed as soon as it was no longer absolutely required.

Additionally, as information was grouped by LSOA, information about, say, crimes could be narrowed down to no more than, on average, $1,500$ people or 650 households (40), thus making the possibility of models being trained on personalised data infeasible. Additionally, as discussed in §3.2.1, as only street-level crimes were considered for this study, potentially more sensitive data such as that included in crime outcome and "stop-and-search" archives including ethnicity and gender were never involved in source datasets at all.

A.3.3 Ethical issues

Due to the sensitive nature of crime forecasting, ethical issues occupied a significant proportion of the research for this study. The subject of models having climacteric material consequences of perpetuating, or potentially deepening, trends of discriminatory bias in police activity as a result of such trends' presence in the datasets used in training was of critical concern throughout the project. Due to this being a delicate matter, it is considered paramount for the subject to be approached responsibly, with care taken to avoid these consequences. This approach is detailed in §2.2.3 in full, with the outcome being that selection of LSOAs as a geographical unit allowed an acceptable compromise between accuracy and volatility of bias.

A.3.4 Professional issues

To ensure that the study, in its output and its proceedings, is of a nature both responsible and proficient, this project was undertaken in total accordance with two codes of professional guidelines throughout. Firstly, as detailed in §3.1, the study was structured around a modified form the professionally standard CRISP-DM process model for analytics. This ensured that, from both a practical and professional perspective, no vital stage was compromised and a high degree of quality was assured to the foundations of the study's process. Finally, the Code of Conduct of the British Computing Society (70) was followed throughout work; in doing so, professional standards of work were upheld through the whole process, and no aspect of responsibility was compromised.

Appendix B

External Material

B.1 Dataset bibliography

(45) The National Archives. *data.police.uk*. 2023. URL: `https://data.police.uk/`.

(46) GetTheData Publishing Limited. *Open Pubs*. 2022. URL: `https://www.getthedata.com/open-pubs`.

(47) Food Standards Agency. *Food hygiene rating data*. 2023. URL: `https://ratings.food.gov.uk/open-data/`.

(48) Office for National Statistics. *ONS Postcode Directory*. Version 2. 2023. URL: `https://geoportal.statistics.gov.uk/datasets/ons-postcode-directory-february-2023-version-2/about`.

(49) Office for National Statistics. *Usual resident population in full-time education by age 18 to 30 years*. 2021. URL: `https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/usualresidentpopulationinfulltimeeducationbyage18to30yearsenglandandwalescensus2021`.

(50) Justin van Dijk et al. *CDRC Modelled Ethnicity Proportions (LSOA Geography)*. 2022. URL: `https://data.cdrc.ac.uk/dataset/cdrc-modelled-ethnicity-proportions-lsoa-geography`.

(51) Aimee North. *Mean price paid for residential properties by LSOA*. Office for National Statistics, 2023. URL: `https://www.ons.gov.uk/peoplepopulationandcommunity/housing/datasets/meanpricepaidbylowerlayersuperoutputareahpssadataset47`.

(52) The Electoral Commission. *Past elections and referendums*. 2019. URL: `https://www.electoralcommission.org.uk/who-we-are-and-what-we-do/elections-and-referendums/past-elections-and-referendums`.

(53) Neil Park. Office for National Statistics, 2021. URL: `https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/lowersuperoutputareamidyearpopulationestimatesnationalstatistics`.

(54) Office for National Statistics. *Lower Layer Super Output Area (2021) EW BFE*. 2023. URL: `https://geoportal.statistics.gov.uk/datasets/ons::lower-layer-super-output-area-2021-ew-bfe/about`.

(55) Office for National Statistics. *National Statistics Postcode Lookup UK Coordinates*. London Borough of Camden, 2023. URL: `https://opendata.camden.gov.uk/Maps/National-Statistics-Postcode-Lookup-UK-Coordinates/77ra-mbbn`.

B.2 External libraries

File	Library	Description of use	Location
tensor.R	data.table	Extension of dataframe functionality	https://cran.r-project.org/web/packages/data.table/
	magrittr	Addition of pipelining	https://cran.r-project.org/web/packages/magrittr/
	plyr	Splitting of CSV sources	https://cran.r-project.org/web/packages/plyr/
	dplyr	SQL-like dataframe joining	https://cran.r-project.org/web/packages/dplyr/
	sf	Spatial vector encoding and operations	https://cran.r-project.org/web/packages/sf/
	geojsonsf	Processing of GeoJSON data in X	https://cran.r-project.org/web/packages/geojsonsf/
	readxl	Processing of Microsoft Excel data in E and G	https://cran.r-project.org/web/packages/readxl/
	lubridate	Parsing of month strings	https://cran.r-project.org/web/packages/lubridate/
	imputeTS	Imputation of missing months	https://cran.r-project.org/web/packages/imputeTS/
	abind	Binding variates into T	https://cran.r-project.org/web/packages/abind/
	tidyr	Cleaning records with insufficient information	https://cran.r-project.org/web/packages/tidyr/
	tseries	Contextualisation of variates as time series	https://cran.r-project.org/web/packages/tseries/
model.py	platform	CPU architecture assertion for acceleration	https://docs.python.org/3/library/platform
	os	Directory scanning for loading of tensor	https://docs.python.org/3/library/os
	pandas	Data science operations	https://anaconda.org/anaconda/pandas
	numpy	Mathematical array operations	https://anaconda.org/anaconda/numpy
	scikit-learn	Training and test set splitting	https://anaconda.org/anaconda/scikit-learn
	Darts	Implementations of time series typing, random forest, and N-BEATS	https://anaconda.org/conda-forge/u8darts-all
	PyTorch	GPU acceleration and model dependency for Darts	https://anaconda.org/pytorch/pytorch
	Ray Tune	Hyperparameter tuning (unused in final implementation)	https://anaconda.org/conda-forge/ray-tune

Appendix C

Predictions

February 2023		Random forest		N-BEATS	
LSOA	Actual	Univariate	Multivariate	Univariate	Multivariate
E01011918	10	4	7	-6	10
E01005744	12	13	13	13	12
E01010931	12	10	12	11	10
E01011357	14	16	14	20	16
E01010688	17	10	16	16	17
E01006062	12	13	13	13	12
E01010695	17	16	16	18	17
E01010879	5	9	7	9	6
E01011496	7	11	8	10	9
E01010842	5	1	5	2	3
E01011613	5	13	12	6	7
E01033697	13	16	13	13	15
E01011565	18	15	18	16	17
E01010609	17	5	5	15	16
E01010970	3	13	13	4	5
E01010708	16	13	16	16	17
E01013394	12	13	12	13	12
E01011783	0	9	4	6	0
E01010858	9	13	10	12	12
E01010785	11	7	6	6	11
E01011147	9	12	10	14	9
E01011202	9	8	10	11	10
E01011042	23	12	19	15	23
E01010884	16	17	18	17	17
E01011665	20	7	14	30	19
E01011216	13	12	13	16	4
E01027603	12	13	12	13	12
E01010692	11	10	11	7	14
E01011731	11	11	11	15	11
E01011117	13	15	14	15	16
E01027583	12	13	14	13	12
E01011788	20	15	18	16	14
E01011207	12	10	12	11	11
E01010957	3	6	5	7	3
E01011383	10	15	15	14	12
E01011837	13	17	5	11	17
E01010904	15	13	15	15	15
E01011281	17	12	13	19	17
E01011263	18	16	19	20	16
E01011409	13	13	12	17	14
E01010760	13	13	13	16	14
E01011028	1	13	11	3	2
E01010939	18	16	16	18	17
E01011766	23	13	21	18	22
E01010770	11	13	12	12	11
E01011097	11	9	11	12	6
E01011246	10	15	13	14	14
E01011507	5	10	11	9	7
E01011705	12	8	8	14	13
E01011761	11	11	11	11	9
E01011068	10	12	11	17	9
E01010968	14	15	14	14	15
E01011359	8	10	10	13	10
E01011860	19	12	16	8	14
E01011302	9	12	9	14	11
E01011596	15	18	18	15	17
E01011770	18	12	16	13	17
E01011451	10	13	12	22	11
E01011903	-3	-5	1	0	-4
E01010768	10	12	11	12	11
E01011113	15	11	12	12	11
E01011170	-1	8	3	5	0
E01011692	6	9	9	6	4
E01010878	4	8	7	6	0
E01010748	4	4	4	8	6
E01011273	23	13	20	24	21
E01010783	17	15	12	16	18
E01011196	11	12	12	15	12
E01011316	-3	12	3	4	-4
E01011416	8	12	10	10	10
E01011245	14	13	14	15	13
E01011014	12	15	13	16	14
E01011418	20	13	18	20	14
E01007587	12	13	12	13	10
E01011709	15	15	15	14	11
E01011884	10	13	11	11	5
E01010874	14	9	12	13	10
E01011218	6	5	3	7	7
E01011881	19	15	17	17	21
E01010582	12	14	13	15	16
E01011033	3	9	9	6	5
E01010683	17	10	15	17	16
E01011240	17	15	17	17	15
E01011407	9	11	9	16	12
E01010751	12	10	11	12	9
E01032494	12	17	14	13	13
E01010779	10	7	9	9	9
E01011424	21	16	16	16	18
E01010633	17	13	16	20	19
E01011186	15	12	14	15	15
E01010798	19	11	17	20	18
E01011871	24	15	21	22	22
E01011517	19	15	19	18	17
E01011253	11	10	10	14	14
E01010585	9	13	12	9	4
E01010765	19	15	17	21	19
E01010722	19	16	16	18	15
E01011673	0	16	7	1	0
E01011152	4	9	6	3	3
E01010915	24	12	19	22	22
E01011093	13	13	14	24	15
E01011728	-1	10	4	12	9
E01011937	16	13	17	16	16
E01032500	13	10	12	16	12
E01011945	9	10	12	9	7
E01033002	6	10	9	6	7
E01010987	10	14	14	12	10
E01011142	6	13	13	13	8
E01010657	24	13	11	25	24
E01010620	12	14	14	13	11
E01010698	4	12	7	14	6
E01010687	19	15	17	18	16
E01010674	4	13	12	3	6
E01033013	-3	5	6	-5	-3
E01010836	13	13	14	9	13
E01011221	18	13	16	14	14
E01011687	12	14	14	15	10
E01011589	14	15	14	16	17
E01011753	2	10	5	9	5
E01010896	5	12	9	11	10
E01011477	16	10	14	20	17
E01010958	14	9	13	7	15
E01011511	15	13	15	12	13
E01010848	18	17	19	12	18
E01011602	14	8	11	8	13
E01011601	19	9	17	17	17
E01011557	11	12	11	14	17
E01011188	10	12	11	15	13

table continues on next column...

... continued from previous column

table continues on next column...

. . . continued from previous column

		Random forest		N-BEATS	
LSOA	Actual	Univariate	Multivariate	Univariate	Multivariate
E01010894	10	8	5	14	7
E01011706	13	14	15	16	16
E01010696	11	15	12	16	16
E01011714	10	12	12	10	10
E01010615	23	13	20	30	23
E01010947	6	9	8	13	11
E01010818	15	9	14	14	10
E01011420	15	17	18	17	18
E01032606	7	13	13	7	8
E01007987	12	13	13	13	12
E01011415	21	15	19	25	19
E01011422	17	15	16	16	16
E01011847	14	13	14	15	13
E01010666	20	12	12	20	-13
E01007359	10	12	12	12	10
E01027564	12	13	12	13	12
E01011713	10	14	12	12	11
E01011599	15	15	15	17	15
E01011054	6	8	7	12	7
E01011584	21	11	17	15	20
E01011886	11	14	12	15	12
E01011037	-4	5	-1	-3	-4
E01010613	3	5	4	5	5
E01011027	6	12	9	13	10
E01011638	11	14	13	16	13
E01011659	10	7	8	6	12
E01011771	15	12	14	14	17
E01011650	23	13	20	15	24
E01011763	6	7	7	7	6
E01011405	14	14	15	13	14
E01010903	2	12	12	13	6
E01010959	25	12	21	19	24
E01012656	12	13	13	13	13
E01010743	9	12	10	12	6
E01011319	8	14	13	12	7
E01011802	5	4	4	4	7
E01010636	-4	7	-1	9	-1
E01011299	-5	4	-1	-3	-7
E01011013	9	13	10	11	11
E01033696	8	9	10	9	16
E01011809	20	16	19	18	19
E01010805	7	16	10	8	7
E01010870	9	12	10	12	12
E01010855	20	16	20	17	14
E01011798	9	11	10	16	11
E01011547	8	13	13	12	13
E01011924	14	9	11	11	14
E01011609	6	9	7	10	4
E01011166	7	16	16	12	10
E01010614	15	16	15	16	12
E01011363	41	4	25	49	43
E01011047	16	14	16	14	14
E01011145	10	16	14	13	12
E01011337	12	7	10	12	14
E01027922	11	13	13	13	12
E01010834	21	0	10	5	24
E01011111	8	11	10	13	12
E01011554	13	12	12	14	13
E01010825	10	13	12	13	14
E01011794	9	11	11	14	9
E01011757	13	13	13	14	16
E01027725	11	13	12	12	12
E01011446	9	10	12	13	9
E01011642	6	15	10	9	7
E01010974	8	16	11	11	9
E01011660	9	14	9	13	9
E01010807	13	13	14	16	15
E01011223	-5	6	-1	-7	-4
E01011064	13	17	15	16	16
E01011846	6	13	16	12	10
E01011826	14	10	12	12	13
E01010993	11	5	10	14	11
E01010847	13	13	13	8	15
E01011286	12	14	13	14	18
E01011454	15	14	16	15	13
E01010917	12	13	13	12	12
E01011080	16	16	15	16	14
E01027924	12	13	13	13	13
E01033015	9	11	10	10	10
E01010889	-7	-2	-4	1	-9
E01007379	12	13	12	13	12
E01011099	6	15	10	11	6
E01011538	14	13	14	7	14
E01011863	6	14	8	10	17
E01010648	14	13	14	14	13
E01011445	17	13	15	19	18
E01011247	18	15	17	16	16
E01010732	18	16	17	19	19
E01011168	7	11	8	11	9
E01011339	24	16	20	29	24
E01011026	13	13	14	12	10
E01011568	12	13	11	14	16
E01011107	-28	13	-11	-24	-35
E01010618	13	16	15	12	13
E01011667	7	2	4	9	9
E01011612	41	17	25	55	39
E01010881	14	10	13	15	11
E01010756	0	5	4	-5	5
E01011330	-6	4	-1	-4	-5
E01011307	8	13	10	11	20
E01011469	5	8	7	10	5
E01011057	1	13	7	6	5
E01011153	8	13	10	13	10
E01027695	12	13	12	13	12
E01011813	14	16	14	16	12
E01011222	1	17	6	7	8
E01010916	13	11	13	13	6
E01010800	17	15	14	7	18
E01010953	8	9	9	10	10
E01011569	13	11	13	17	23
E01010691	8	13	11	14	12
E01011132	14	14	15	14	14
E01011371	21	7	5	25	21
E01010999	13	13	14	13	19
E01011744	1	8	8	11	9
E01011810	15	15	15	20	14
E01010625	-1	9	4	2	-3
E01010850	20	9	5	22	23
E01009204	12	13	12	13	12
E01011082	13	15	15	16	19
E01011160	18	14	17	15	17
E01011603	10	14	11	15	15
E01011571	13	14	14	16	14
E01011719	9	13	11	14	15
E01011431	8	9	8	5	5
E01010617	24	13	21	23	23
E01010901	10	9	12	10	8
E01010734	6	16	10	7	7
E01011287	18	8	11	6	21
E01010867	1	10	4	7	3
E01032493	16	13	15	15	16
E01011628	13	13	13	15	11
E01011442	10	13	12	17	10
E01011176	11	12	11	10	10
E01010832	0	6	3	3	2
E01027566	12	13	13	13	13
E01010862	12	12	12	12	13
E01011127	7	12	8	13	12
E01010658	14	12	13	21	13
E01011677	2	11	8	1	7

table continues on next column. . .

. . . continued from previous column

		Random forest		N-BEATS	
LSOA	Actual	Univariate	Multivariate	Univariate	Multivariate
E01011055	1	13	6	7	1
E01011314	6	12	9	13	5
E01011751	7	13	9	11	10
E01011926	13	15	14	16	13
E01011423	6	15	13	11	11
E01011226	13	13	14	16	16
E01007424	12	13	12	13	13
E01011946	16	15	17	16	16
E01010725	26	14	23	13	23
E01011197	13	10	12	12	12
E01011552	9	8	8	8	12
E01011624	13	10	11	13	12
E01010793	24	16	21	30	24
E01011748	17	15	17	14	17
E01011474	-1	2	0	-6	-2
E01011843	19	16	18	24	18
E01011165	13	13	12	18	15
E01011212	6	9	7	10	5
E01010584	4	10	6	8	3
E01027915	12	13	13	13	13
E01011662	16	16	16	16	18
E01011324	14	12	9	14	14
E01011268	22	15	20	29	22
E01011830	28	9	21	33	30
E01011232	7	8	9	8	6
E01011448	18	16	17	16	17
E01010761	5	13	8	6	16