Initial Plan: Crime Visualization and Forecasting Application

CARDIFF UNIVERSITY

PRIFYSGOL CAERDYᴅᴅ

Author: Hugo Truman (1827047)

Project Supervisor: Federico Liberatore

Project Moderator: Hiroyuki Kido

Module: CM3203 One Semester Individual Project (40 Credits)

Project Description

Crime is inevitable. Police enforcement, despite their best efforts, will struggle to eliminate crime completely and thus instead focus on catching the perpetrators of said crimes. Preventative measures often play second fiddle to the real demands and pressures of solving those offences already committed.

Predictive policing refers to the use of mathematics and predictive analytics within the policing world to identify potential criminal offenses. An example of this is the West Midlands and West Yorkshire police started working with the home office to create a National Data Analytics Solution, which included the 'Most Serious Violence' system that would calculate a score for people the two forces had previously come into contact with. In theory, the higher this score the more likely that person was to commit knife or gun offences. However concerns over the inherent bias and marginalisation of such a system halted its development[1].

My project is crime forecasting

For my project I intend to use an idea of a crime forecasting algorithm but instead develop it to focus on geographic areas and determine how likely it is for crime to occur there. The decision to focus on geographical location as opposed to singular people should prevent the ethical dilemma associated with the MSV system. The statistics will be calculated using data collected by the police (available at data.police.uk) and will map crimes to specific locations thus determining the likelihood of further crime in an area. This information can then be used by law enforcement to regulate the allocation of resources in proportion to the feasibility of crime at certain locations and times. More generally the information would be made accessible to the public who could then use it for example when looking to buy a house in a new area to ensure there is a low crime rate en route to work or schools.

The system will come in the form of a web application that allows anyone with access to the internet to view it easily, whether on a smartphone or computer. It will offer a wide variety of ways to view and interpret the data including a map with crime hotspots pinned. However crime statistics are not always truly representative of how 'dangerous' an area is if we do not consider the type of crime, therefore a graph representing what crimes are prevalent in a certain area is equally useful, especially to the general public.

The algorithm I intend to use for this application is K-clustering, also known as Lloyd's algorithm. It is a method of vector quantization that arranges *n* observations into *k* clusters. Each observation belongs to the cluster with the nearest mean resulting in a partitioning of the data space into Voronoi cells[2]. Standard time series analysis will be applied to each cell to forecast future levels of crime, "A time series is simply a series of data points ordered in time"[3].

Project Aims and Objectives

Aims: The aim of this project is to create an easily accessible application that can be used by both the police and the general public to help forecast crime rates based on geographical

location. It will also incorporate different crime statistics to form general geographical impressions of crime. Python will process existing police data which will then be used in the application made through React Native.

Objectives:

Research & Planning:

- ❏ How to use react native
- ❏ The most effective way to forecast crime levels
- ❏ How maps can be implemented in react native
- ❏ How to use python scripts in react native
- ❏ Find other existing solutions and see what improvements can be made
- ❏ How to use machine learning libraries in python

Identify & Define Requirements:

- ❏ Look in depth at other solutions and see what requirements I should take from them
- ❏ Think about the target audience and what they would want from an application like this. How would it improve their lives, as well as how can the application be made easier to use.

Design & Implementation:

- ❏ Create wireframes and mock up designs of the application and think about what the best user interface would be for the target audience
- ❏ Think about the structure of the database and how it will be implemented
- ❏ Think about what methodology I will use to develop this application
- ❏ What will the structure of the code be
- ❏ The implementations of requirements
- ❏ The user interface for the application and how it will look in react native

Test & Evaluate:

- ❏ Test with sample target audience
- ❏ Get testers to complete a survey about the applications
- ❏ Test application against requirements
- ❏ Testing and Validating forecasting model

Documentation:

- ❏ Will be completed and added to throughout the continuations of the module until complete.
- ❏ Will consist of: Introduction, Background, Approach, Implementation, Results and Evaluation, Future Work, Conclusions, Reflection on Learning.

Challenges:

- ❏ A potential problem of my project could be the large amount of data provided from the police API. The limitations of my laptop may struggle to process the capacity of data required for such a task. If this is the case I will have to reduce the data set size and provide examples from a few counties.
- ❏ A personal challenge for me in this project is my lack of experience with the software. Having never used React Native before the bulk of my research and preparation will entail me learning how to use and implement them into my application. To combat this problem I have scheduled 2 weeks to focus on

conducting research and figuring out these softwares, especially in relation to my specific project.
- ❏ Finally, there is the issue of K-clustering and my lack of familiarity with the algorithm. Consequently I will dedicate some time to researching and practicing implementing the algorithm to ensure I fully understand it.
- ❏ Learning machine algorithms like k clustering and time series modelling

Ethics:

- ❏ Talk about ethics of personal data when testing
- ❏ Crime data is anonymous

Work Plan

Week 1

Aims:

- ❏ Set up regular meetings with Supervisor
- ❏ Complete Initial Report

Milestones:

- ❏ Complete Initial Report

Deliverables:

- ❏ Initial Report

Week 2

Aims:

- ❏ Research and learn how to use React Native
- ❏ Research how to pull in data from Police API
- ❏ Learn how k-clustering works

Milestones:

- ❏ Understanding how police API works
- ❏ Understanding how react native works

Week 3

- ❏ Learn how to use Machine learning python libraries
- ❏ Learn how to use maps in react native
- ❏ Mock-up wire frames for application
- ❏ Start python code for visualizing and forecasting police data

Milestones:

- ❏ Understanding how python machine learning libraries work works and how I will implement it

Deliverables:

- ❏ Mock design for application

Week 4

Aims:

- ❏ Plan structure of database
- ❏ Decide what software to use for the database
- ❏ make UML diagrams for classes
- ❏ Make Use cases and acceptance criteria

Deliverables:

- ❏ UML diagrams
- ❏ Use cases and acceptance criteria

Week 5

Aims:

- ❏ Finish up crime visualizing and forecasting prototype
- ❏ Complete React Native research

Milestones:

- ❏ Learnt React Native to a capable standard

Deliverables:

- ❏ Python code for visualizing and forecasting crimes

Week 6

Aims:

- ❏ Meet with supervisor to check requirements
- ❏ Make any changes needed
- ❏ Start initial application prototype (pulling together all code completed so far)

Milestones:

- ❏ Approved crime visualizing and forecasting python code
- ❏ Start building application

Week 7

Aims:

- ❏ Further the development of application
- ❏ Incorporate maps with pinpoint locations
- ❏ Start python code for other county crime statistics (i.e. type of crime)
- ❏ Validation of Forecasting algorithms

Week 8

Aims:

- ❏ Pull in python code to the application to find hotspots of forecasted crime
- ❏ Complete and implement python code for other county crime statistics

Milestones:

- ❏ First completed prototype

Deliverables:

- ❏ Prototype application

Week 9

Aims:

- ❏ Meeting with supervisor to discuss further prototyping once core functionalities are in place
- ❏ Development of further prototyping

Week 10

Aims:

- ❏ Work on improvements from supervisor meeting
- ❏ Optimise the code
- ❏ Finishing touches

Deliverables:

❏ Completed application for user testing/requirement testing

Week 11

Aims:

❏ Start testing
❏ Evaluate results
❏ Make alterations

Milestones:

❏ Completed application

Deliverables:

❏ Final iteration of completed application

Week 12

Aims:

❏ Proofread and final adjustments to report

Deliverables:

❏ Final project
❏ Final application

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12
Research and Planning												
How to use react native		■	■	■	■							
The most eefective way to forecast crime	■	■										
how maps can be implimented in react		■	■	■	■							
how to use python scripts in react		■										
how to use machine learning libaries in python			■									
Identify & Define Requirements												
Look in depth at other solutions and see what requirements I should take from them				■	■	■						
Look at target audiecnce and what they would want				■	■	■						
Design & Implimentation												
Wireframes/Mockups			■									
Database design			■	■								
K-clustering code			■	■	■							
Implimentation of requirements						■	■	■	■			
refine layout of application									■	■		
Test & Evaluate												
Test with target Audience											■	
Get testers to complete surveys											■	
Test requirements											■	
Test Validating forcasting model							■					
Documentation												
Initial Plan	■	■										
Final Report		■	■	■	■	■	■	■	■	■	■	■

References

(1) Burgess, Matt. 2020. The Police built an AI to predict crime. It was seriously flawed. *Wired* 6 August https://www.wired.co.uk/article/police-violence-prediction-ndas [date accessed: 04/02/21]

(2) Hamerly, G and Drake, J. 2015. Accelerating Lloyd's Algorithm for k-Means Clustering. In: Celebi, M, E. ed. *Partitional Clustering Algorithms.* Cham: Springer, pp. 41-78

(3) Peixeiro, Marco. 2019. The Complete Guide to Time Series Analysis and Forecasting. Towards data science 7 august https://towardsdatascience.com/the-complete-guide-to-time-series-analysis-and-forecasting-70d476bfe775 [date accessed: 08/02/21]