Slide 1

Welcome to this research proposal on the topic of “Crime reduction with machine learning: a comparative analysis”

Slide 2

This project fulfils the requirements of an MSc research project because, as will be described in the presentation:

- It will demonstrate deep understanding of the challenges of applying machine learning models in real-world scenarios where consideration of standard performance metrics is not enough.
- It will apply many techniques including research, data gathering, feature engineering and machine learning.
- It is original, building on a previous literature review where gaps in methodology and application were identified.
- It will deal with complexity in terms of data, feature engineering and the development of multiple machine learning models, building upon models developed by previous scholars.
- And the entire project will be self-directed and will include self-evaluation throughout and in the final report.

Slide 3

This research proposal contributes to the discipline of machine learning, notably in the practical applications of machine learning in the domain of crime prediction.

There is a lot of published research on using machine learning for crime prediction, but to date there is low adoption. A previous literature review (Feaviour, 2024), found that many papers have explored crime prediction use cases, but the focus has been on the performance of the models to the detriment of the usefulness of the insights created. In other words, the academic standards were met to prove that the models were providing accurate predictions, but the predictions themselves were not necessarily useful. The main shortcomings were that in many cases the insights, whilst accurate, provided nothing actionable, and they were often “black box” solutions, meaning that the reasoning behind the predictions was unknown and hence would be challenging to adopt in a real-world situation. There was also very little consideration to the computational efficiency of the solutions, which would also be important for real-world applications to manage time and cost.

This study will build upon the existing research by using consistent performance metrics across various previously tested machine learning models and adding interpretability and computational efficiency metrics to derive an overall “usefulness” score for real-world application of a model. This approach would potentially be applicable to all domains using machine learning.

Slide 4

Two research questions are considered in this proposal. The first concerns the methodology to be defined and the second concerns the results of an assessment of machine learning models using that methodology.

The research questions are:

1. Which metrics should be used to measure machine learning model performance, interpretability, and computational efficiency, to provide an overall blended score?

2. Which machine learning models provide the best overall performance, interpretability and computational efficiency when predicting locations of crime?

Slide 5

As previously mentioned, there have been many studies on crime prediction using machine learning, using a variety of different approaches and machine learning models, but the performance metrics and datasets vary between the studies making direct comparison difficult. The aim of this study is therefore to define a methodology to compare different solutions, and then to use that methodology to compare the performance of three previously published approaches.

So, the objectives are:

Firstly, one of the challenges with the existing literature is little to no attention to interpretability and computational efficiency, as previously mentioned, so best practices for interpretability and computational efficiency will be reviewed, after which a methodology for measuring model usefulness, including those metrics plus the most appropriate metrics for model performance will be proposed.

Secondly, the proposed methodology will be used to compare three published approaches, using the same dataset, as far as possible, to 1) prove that the methodology to measure overall usefulness works, and 2) to provide a recommendation as to the most useful model from the three approaches tested.

Slide 6

A literature review (Feaviour, 2024) has already been undertaken to identify different approaches to using machine learning for crime prediction, with the papers examined summarised in this table.

The challenges identified during the literature review, and already highlighted in the aims and objectives of this project, included:

- Different papers used different metrics to measure the performance of their solutions, so making comparisons between papers is difficult. They also used different datasets, as would be expected, which compounds this problem.
- The papers focused on proving the performance of the models, but not necessarily on whether the insights that the models generated would be useable. If a model is 100% accurate at predicting something, but the prediction is unusable, then the model has no practical value.
- The papers overlooked interpretability and computational efficiency, both of which are important when looking at real-world applications.

The three approaches to be built and tested in this research project, Hajela et al. (2020), Cichosz (2020), and Rummens et al. (2017), as highlighted in the red box, were selected because they were considered to be the most useful from the literature review. The final solution in the table, Stalidis et al. (2021), was excluded simply because it would be too complicated to implement, and its complexity would almost certainly mean that it would score low on computational efficiency making it an unlikely candidate to be recommended anyway.

Slide 7

This research will use existing crime data to assess the three approaches already mentioned from the literature review (Feaviour, 2024), so primary research will not be required because crime datasets already exist, and indeed creating a new dataset would be impracticable. The solutions will be tested using standard quantitative machine learning performance metrics, but more typical quantitative and qualitative analysis will not be required.

Collecting primary data through interviews, such as police resource planners, or a public questionnaire to assess public appetite for predicting crime using machine learning and the potential concerns and ramifications were considered, which would have required qualitative and quantitative analysis respectively. Whilst this analysis could provide some direction for potential next steps, it was determined that it would not assist with answering the specific research questions and so would be excluded from this research. They might form the basis of a recommendation for future work at the conclusion of this project.

Selecting the most appropriate performance metrics is part of one of the aims of this research. A non-exhaustive list of performance metrics to be considered is:

Accuracy, which is the proportion of predictions that were correct, but it can be misleading with imbalanced datasets and so unlikely to be useful on its own in this scenario.

Precision, which is a measure of how accurate the positive predictions were, scoring higher with fewer false positives but not penalising false negatives.

Recall, which is the proportion of positive predictions out of all actual positive instances without penalising false positives.

F1 score, which is a combination of precision and recall to balance the impacts of false negatives and false positives.

And area under the curve, or AUC, which plots true positives against false positives at various threshold settings to measure overall performance.

Of the three approaches to be assessed with this research, classification of crime hotspots after clustering (Hajela et al., 2020) used accuracy, precision, recall and AUC. Classification of hotspots by points of interest (Cichosz, 2020) used AUC. And Classification of hotspots by grid and time of day (Rummens et al., 2017) used precision, who noted that precision, by penalising false positives, measured the need for efficiency, which seems appropriate when taking the solution from academic proof into real-world application. That kind of reasoning is exactly what will be explored further in this part of the research to finalise the set of measures to be used.

In additional a further literature review will be undertaken to determine the best way to measure interpretability and computational efficiency so that those metrics can be included in the methodology to be proposed for objective one.

Slide 8

The three approaches to be tested each use two different solutions, and in the case of Hajela et al. (2020), the two solutions require three machine learning models between them, so a total of seven models will be required for six solutions, as illustrated here.

In order to measure the relative performance of the six solutions in the most comparable way the source data should be the same for all six. However, the three approaches all use slightly different features, so a search will be undertaken to find a dataset that has all of the required features to be able to train the models for all three approaches after selecting the required features for each of them from the same dataset. If a suitable single dataset cannot be found it will still be possible to proceed with the comparison using different datasets, such as the datasets used in the original publications, however the comparative evidence will be less strong and this will be noted in the final report, so this is not a preferred option. In either case, feature engineering will be undertaken to optimise the features for each model to ensure the best outcome for each and therefore the most robust comparison possible, so even if the original datasets are chosen as a last resort, the precise features used could still differ.

Once the data has been sourced, either from a single dataset or multiple, the models will be built and trained in Python using Jupyter Notebooks on GitHub running on Google Colab. Python is a standard, and highly accessibly language for this kind of task. Putting the code in Jupyter Notebooks makes the output easier to publish and read, and putting it all on GitHub will help to facilitate sharing the source code and the results. Running the code on Google Colab will ensure standardised processing to make measuring the computational efficiency more reliable.

All of the results will be gathered and processed according to the methodology to be proposed for objective one so that a like-for-like comparison and a final recommendation of the best solution for real-world application can be made.

Slide 9

The data required for the three approaches will be publicly available, including geographical and historic crime features, so no personal information will be required. Since no additional research will be conducted, there are no concerns about the use of personal data.

The second objective of the research is to recommend a machine learning solution based upon the real-world usefulness metrics to be defined for the first objective, so the recommendation is intended to include an ethical component in the form of an explainable model already built in, which goes some way to address the ethical concern from the original literature review (Feaviour, 2024) concerning the deployment of a crime prediction model and its influence on police presence in certain locations.

Slide 10

A number of artefacts will be created for this research project.

Firstly, a separate dataset for each of the three approaches containing the features required for that solution.

Secondly, the six solutions will require seven trained and tested machine learning models, all of which will be available, alongside the datasets, on GitHub.

Finally, the research report will provide the methodology to measure real-world usefulness, the results of the six solutions measured using that methodology, and a recommendation for the best solution to potentially investigate further.

Slide 11

The timeline of activities to complete this research project are illustrated in this Gantt chart.

After getting approval for the project in February, the project will start with a literature review in March to gather evidence of best practice and emerging trends for measuring interpretability, computational efficiency and appropriate machine learning performance metrics. Once the various metrics to be used have been determined they will be combined and weighted to form a proposed methodology.

From mid-March, in parallel to completing the research, a search for an appropriate dataset will start, with a key objective being for the dataset to provide the required features for each of the selected models, so data segmentation will start in parallel to completing the selection of the dataset and will continue to the end of April, handing over to feature engineering to extract the required data from the dataset for each model.

In May, the building, training and testing of the machine learning models will commence, taking approximately two months. In reality this will also include feature engineering to optimise the models, not shown on this high-level plan.

In July, the previously proposed methodology will be used to measure the usefulness of the three models, followed by a write-up of the report in August ready for a defence of the report and its findings in September, which will mark the conclusion of this project.

Slide 12

These are the references used within this presentation.

Thank you.