

Abstract

Can a borough's name and the year alone predict how much crime it records? Using Metropolitan Police data for 2008–2016, annual crime counts for each London borough were modelled with linear, tree-based, nearest-neighbour, neural-network and ensemble methods, and compared with a mean-value baseline using MSE, MAE and MAPE.

Non-linear models, especially Decision Trees and ensembles, clearly beat the baseline. Borough-level crime variation can be explained by minimal spatiotemporal information, which points to stable differences between boroughs and gradual change over time. These forecasts describe recorded crime, not inherent features of any borough.

5×

lower error than the baseline (MAE 86 vs 427 crimes)

20%

lower error than the best linear model (MAE 86 vs 108)

7.2%

Decision Tree test MAPE, the best of all models

Introduction

Crime clusters in particular places rather than occurring at random [1], and machine learning is increasingly used to forecast it [2]. Most studies add many socio-economic variables [3]. This study asks a simpler question: **can borough and year alone predict recorded crime in London?**

Hypothesis: borough-level crime is stable over time, so models trained on borough and year will predict it accurately.

Ethics: police records reflect where police patrol as well as where crime happens, and predictive systems can reproduce bias against minority communities [4], [5]. Results are therefore forecasts of recorded crime, not of any borough's character.

Methods

Data: Metropolitan Police recorded crime (London Datastore), 2008–2016, over 1,000,000 records. Counts were summed to one total per borough per year (32 boroughs plus the City of London).

Features: borough (one-hot encoded to avoid false ordering [6]) and year (numeric). Inputs were standardised, fitted on training data only, for KNN and neural networks [7].

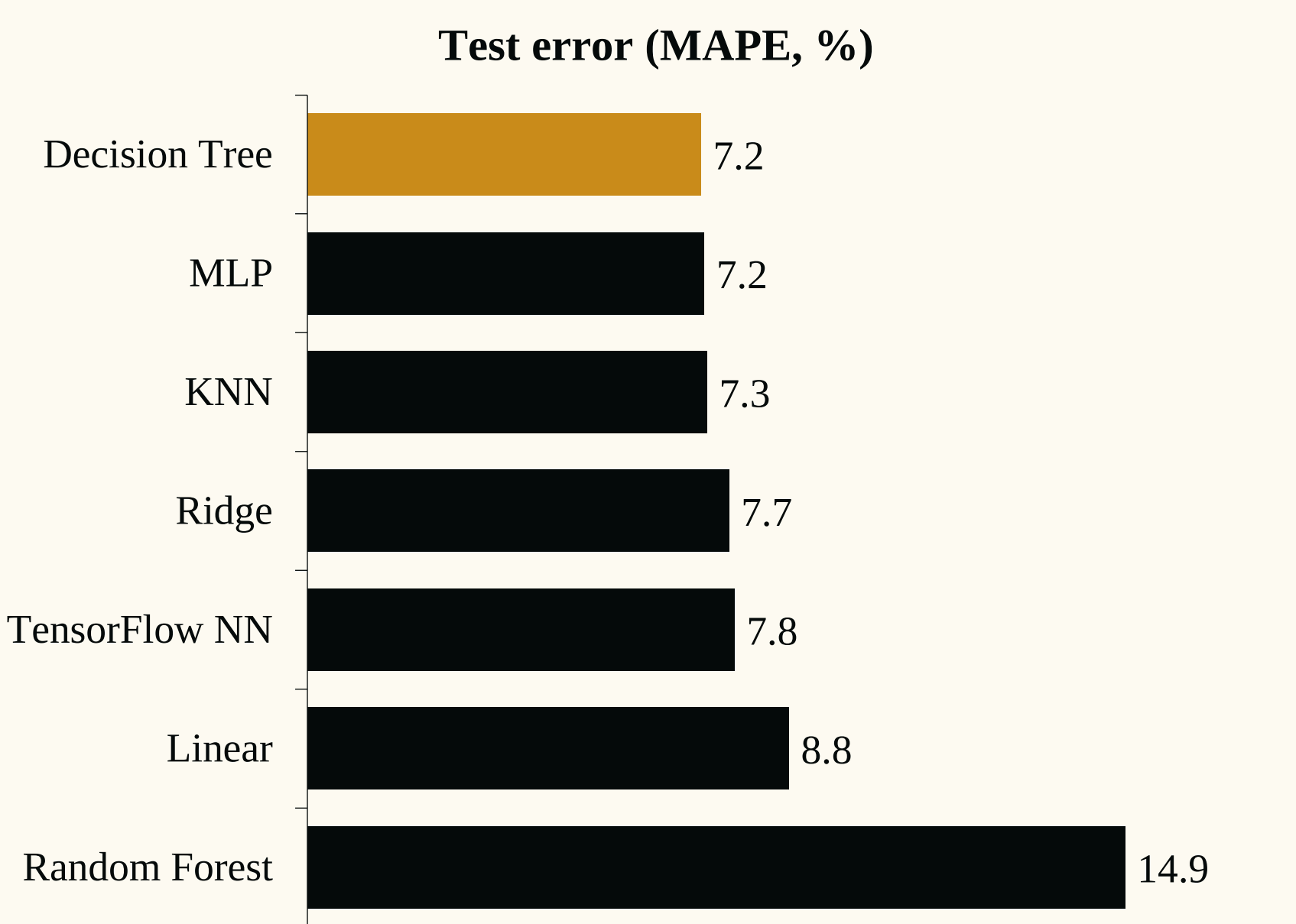
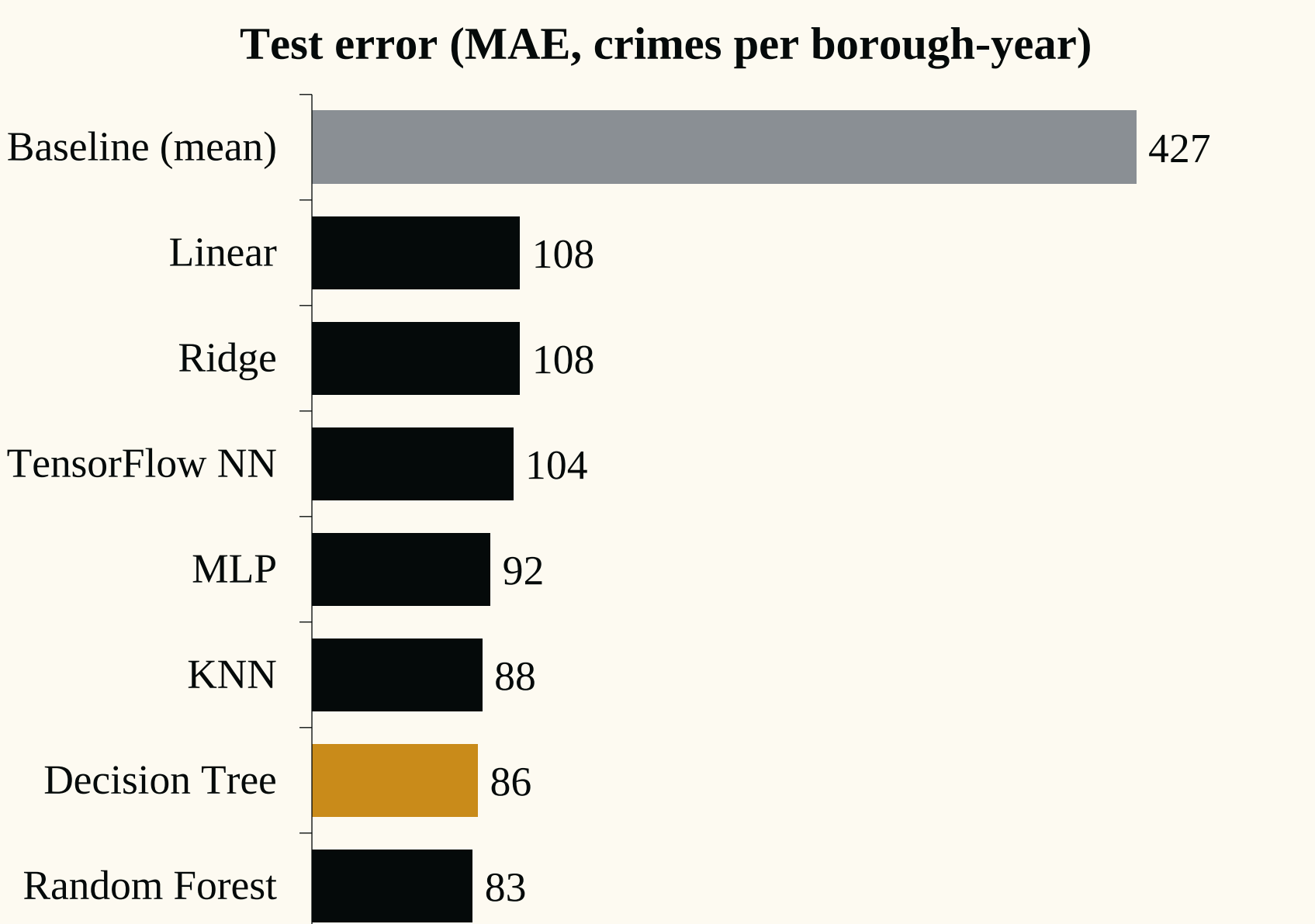
Split by time: validation on 2015, testing on 2016.

Models: Linear, Ridge, Decision Tree, Random Forest, KNN, MLP, TensorFlow neural network, and average and weighted ensembles, tuned with randomised search and 3-fold cross-validation.

Metrics: MSE, MAE and MAPE against a mean-value baseline.

Results

Every model beat the baseline (MAE 427). The **Decision Tree had the best MAPE (7.2%)** with an MAE of 86, while Random Forest had the lowest MAE (83). Non-linear models beat linear ones by about 20%. The average ensemble (MAE 88) did not beat the best single models.



Discussion

Non-linear models captured borough-year structure better than linear ones, supporting the hypothesis that borough-level patterns are stable and change gradually.

Different model types converged: predictions varied by about 39 crimes per borough-year (3% of the average).

Caution: Decision Tree training error (MSE 8,157) was far below validation error (106,277), a sign of overfitting. Random Forest over-predicted in about 73% of cases.

Conclusion

Borough and year alone explain a large share of the variation in recorded crime across London. Non-linear models, especially the Decision Tree, predict it best, about five times more accurately than a simple average.

Limits: one city, 2008–2016, annual borough totals, and police records that carry reporting and policing bias.

Next: smaller areas (LSOA), wealth, population and policing data, other cities, and borough-level error analysis.

References

[1] Brantingham, P. L. & Brantingham, P. J. (1984). Patterns in crime. Macmillan.

[2] Dakalbab, F. et al. (2022). Artificial intelligence and crime prediction: A systematic literature review. Social Sciences & Humanities Open, 6(1), 100342.

[3] Zhou, Y., Wang, F. & Zhou, S. (2023). The spatial patterns of the crime rate in London and its socio-economic influence factors. Social Sciences, 12(6), 340.

[4] Lum, K. & Isaac, W. (2016). To predict and serve? Significance, 13(5), 14–19.

[5] Richardson, R., Schultz, J. & Crawford, K. (2019). Dirty data, bad predictions. NYU Law Review Online, 94, 192–233.

[6] James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013). An introduction to statistical learning. Springer.

[7] Scikit-learn (2019). StandardScaler.