UCL DEPARTMENT OF GEOGRAPHY

YEAR 2024-25

EXAM CANDIDATE ID:	**LTKB8**
MODULE CODE:	**GEOG0115**
MODULE NAME:	**Introduction to Social Data Science**
COURSE PAPER TITLE:	**UK 2024 General Election Prediction Analysis Using Machine Learning Models: A Case Study of England and Wales**
WORD COUNT:	**1996 (excluding tables & figures contents and references)**

I acknowledge the use of Artificial Intelligence (AI) of ChatGPT (https://chatgpt.com/) to assist in improving my writing and troubleshooting code. Specifically, I used ChatGPT to make my writing more concise and clear after drafting my ideas. I reviewed the grammar and ensured the core message and ideas that I want to convey in my work remained intact. Additionally, I used ChatGPT to help resolve errors and improve my code. For instance, I encountered difficulties using GridSearchCV to define appropriate hyperparameters for various models. To resolve this, I used the prompts of "*i want to compare the prediction performance of RF, SVM KNN*" and *how to do gridsearch "param_grid = { 'n_estimators': [50, 75, 100], 'max_depth' for SVM*".Subsequently, the suggestions provided by ChatGPT were cross-referenced with the book *Hands-On Machine Learning with Scikit-Learn & TensorFlow* by Géron (2017), as well as data science related online forums and website, including Kaggle, Analytics Vidhya, and Geeks for Geeks. To maintain transparency, I documented and credited all ChatGPT-assisted code sections directly within the Jupyter Notebook submitted for this coursework. My use of ChatGPT was primarily aimed at enhancing the readability of my writing and addressing specific coding challenges, with all outputs verified for accuracy against trusted resources.

Your essay, appropriately anonymised, may be used to help future students prepare for assessment. Double click this box to opt out of this ☐

Introduction

Elections are essential for democracy, enabling citizens to shape governance and policy indirectly through their elected representatives and leaders (Sinha et al., 2024). Given the significant power held by electoral winners, predicting election outcomes has become a critical tool for anticipating and preparing for the potential shifts in governance and policy that may follow (Lewis-Beck and Stegmaier, 2014). Predicting elections is important for at least reasons like assessing impacts on financial markets, political stability, and economic development, and mitigating potential social crises arising from electoral trends (Kazadi and Musimwa, 2023). This is especially vital for elections in globally influential nations, such as the UK general election.

The UK traditionally conducts elections using the plurality system ("first-past-the-post") in single-member constituencies, with 650 Members of Parliament (MPs) elected from constituencies averaging 60,000 electors. The 2024 UK parliamentary general election, held on Thursday, 4 July 2024, saw the Labour Party, led by Sir Keir Starmer, winning the largest number of seats and votes. This election was particularly significant as the first "post-Brexit" election, both legally and politically, and the first since the COVID-19 pandemic—two major events that reshaped the UK and influenced the global economy (Cracknell & Baker, 2024; Surridge, 2024; Prosser, 2024; Hassan et al., 2019).

Several studies have explored predicting UK general elections and major votes, such as Brexit, using statistical methods and machine learning. Social media data have been used to predict UK elections (Usher & Dondio, 2022; Kaminska et al., 2017; Burnap et al., 2015), as well as polling data (Levene & Fenner, 2021; Ford et al., 2016; Rallings et al., 2016), and citizen forecasting (Murr, 2016; Lewis-Beck & Stegmaier, 2011; Johnston et al., 2018). Notably, some studies have incorporated socioeconomic indicators with machine learning models to predict election outcomes, as summarized in Table 1.

Table 1: Studies using Machine Learning Models with Socioeconomic Independent Variables

Scholarly References	Machine Learning Methods	Dependent Variables	Independent Variables
(Owens and Wade, 1988)	Ordinary Least Squares (OLS)	Conservative share of constituency vote	Middle class (%), Constituency distance from London (miles), Change in unemployment (%), Change in household income, Household income, Unemployment (%), Location of constituency, Weekly wages of employed males, Change in weekly wages (%)
(Clark, Morris and Lomax, 2018)	Support Vector Machine (SVM), K-Nearest Neighbour (kNN), Artificial Neural Networks (ANN), Self-Organising Maps (SOM), Gradient Boot Machines (GBM), Multivariate Adaptive Regression Splines (MARS/Bag Earth), and Cubist.	Leave/Remain in EU Votes	Number of individuals in an area who signed each e-petition (25 studied e-petition topics)
(Jensen, Lacombe and McIntyre, 2012)	Spatial Autoregressive (SAR), Spatial Error Model (SEM), Spatial Durbin Model (SDM)	The percentage of the vote obtained by Conservatives	% Won by Conservative, Incumbent dummy variable, Marginal constituency vote, Sex dummy variable, Race dummy variable, Senior occupation (%), Married (%

			population), Lone parent (%), White (% population), Student (% population), % Claiming unemployment for more than 1 year
(Borisyuk et al., 2005)	Neural Network	Winning Party	The input signal comprised "expert" responses to a set of 11 questions.

This study aims to contribute to UK election prediction research by applying machine learning models—Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbor (KNN)—to predict the winning party in each constituency for the 2024 UK general election in England and Wales, using relevant socioeconomic variables. This study compares the prediction accuracy of both base and GridSearchCV-tuned versions of these models, totaling six models. Furthermore, the study focuses on England and Wales only due to limited constituency data for Scotland and Northern Ireland (Payne, 2003).

The three machine learning models were selected for their proven effectiveness in prediction and classification tasks. Random Forests excel in classification by leveraging random inputs and features, yielding robust results (Sinha et al., 2024; Breiman, 2001; Fazil et al., 2023). Support Vector Machines (SVMs) are well-suited for classifying complex, small-to medium-sized datasets, offering robustness, strong generalization, and unique global optimum solutions (Géron, 2017; Awad & Khanna, 2015; Fazil et al., 2023). The k-Nearest Neighbors (kNN) classifier assigns unlabeled observations to the class of the most similarly labeled examples (Bhandari, 2024) and has been previously applied in predicting election outcomes (Clark, Morris & Lomax, 2018).

Data and Methods

The primary data for this study, including the 2024 General Election results and most independent variables, were sourced from the House of Commons Library. Most of the socioeconomic independent variables are derived from the 2021 UK Census, with additional feature-engineered variables, such as creation type. Details are provided in Table 2.

Table 2: Independent Variables Details

Variable Name	Scholarly References	Feature Engineering	Data Source
turnout_rate (Election participation rate (%))	(Nadeau, Lewis-Beck and Bélanger, 2009; Denver and Hands, 1974)	Number of Valid Votes per Valid Electorate	(Baker, Pollock and Cracknell, 2024)
majority (Vote difference between the winning and second-ranking parties)	(Denver and Hands, 1974; Jensen, Lacombe and McIntyre, 2012)	N/A	(Baker, Pollock and Cracknell, 2024)
top_ad_spending (Campaign ad expenditure in GBP (June 27 - July 4, 2024))	(Sinha et al., 2024; Jensen, Lacombe and McIntyre, 2012)	N/A	(Who Targets Me, 2024)

ad_associated_party (Party Name in ad spending)			
youth_population (Population aged 18–34 (%))	(Whiteley, 2023; Richardson and Hougen, 2020)	N/A	(Barton, 2024e)
white_pop_pct (White population (%))	(Jensen, Lacombe and McIntyre, 2012; Richardson and Hougen, 2020)	N/A	(Barton, 2024a)
Christian_pct (Christian population (%))	(Clements and Spencer , 2014)	N/A	(Zayed, 2024)
deprived_pct (Proportion of LSOAs classified as highly deprived (%))	(Bulutgil and Prasad, 2019; Richardson and Hougen, 2020)	N/A	(Francis-Devine, 2024b)
median_weekly_wage (Median weekly wage (GBP))	(Owens and Wade, 1988; Richardson and Hougen, 2020)	N/A	(Francis-Devine, 2024d)
children_poverty (Children (0–15) in low-income households (%))	(Richardson and Hougen, 2020)	N/A	(Francis-Devine, 2024a)
crime_rate (Crime per total population)	(Sinha et al., 2020)	Number of Crimes per Total Population	(UKCrimeStats, 2024)
employment _rate (Employment rate (%))	(Owens and Wade, 1988; Jensen, Lacombe and McIntyre, 2012)	N/A	(Office for National Statistics, 2024)
unemployment_claim_rate (People aged 16–64 claiming unemployment benefits (%))		N/A	(Ward, 2024)
public_sector_rate (Population working in the public sector (%))	(James, 2019)	N/A	(Francis-Devine, 2024c)
higher_qualifications (Population aged 16+ with higher education qualifications (%))	(Richardson and Hougen, 2020)	N/A	(Bolton, 2024)
non_disabled_pct (Non-disabled population (%))	(Obradovic, Gentile and Bruter, 2023)	N/A	(Esme Kirk-Wade, 2024)
families_children_households (Households with families and children (%))	(Fieldhouse and Cutts, 2012)	N/A	(Barton, 2024d)
house_own_mortgage (Population owning a house through a mortgage (%))	(Owens and Wade, 1988)	N/A	(Barton, 2024b)
median_house_price (Median house price (GBP))		N/A	(Barton, 2024c)
houseprice_change (House price change compared to the previous year (%))		N/A	
population_density (Population density (people per square kilometer))	(Mansley and Demšar, 2015)	Total Population per km²	(Barton, 2024e)

After data cleaning, merging, and variable selection, exploratory analysis was conducted. While normality is a common assumption for machine learning models, it is not strictly required (Barai, 2020). Therefore, variables were not transformed in this study, even though some did not exhibit a normal distribution. The choropleth map (Figure 3) highlights that deprivation percentage and child poverty percentage are predominantly concentrated in the Midlands region. In addition, the correlation matrix reveals a strong negative correlation (around -0.7) between turnout rate and variables such as deprivation percentage, child poverty percentage, and unemployment benefits claim rate, suggesting lower election participation in economically deprived areas. Additionally, several variables exhibit strong correlations, indicating potential multicollinearity as shown in Figure 4. However, multicollinearity is irrelevant in this case as the study focuses on prediction accuracy of the machine learning models (Lieberman & Morris, 2014).

The Labour Party's majority win in the 2024 UK general election, as shown in Figure 2, highlights the challenge of imbalanced datasets in machine learning. Imbalanced datasets can negatively impact classification accuracy (Tyagi & Mittal, 2020; Sinha et al., 2024; Kazadi & Musimwa, 2023; Sun et al., 2009). This issue is expected when predicting constituency winners in the UK due to the presence of multiple significant parties (Payne, 2003). Despite this fact, the study trains models on both the imbalanced dataset and a more balanced dataset, comprising constituencies won by the top three parties (Labour, Conservative, and Liberal Democrats), to evaluate model performance under different conditions.

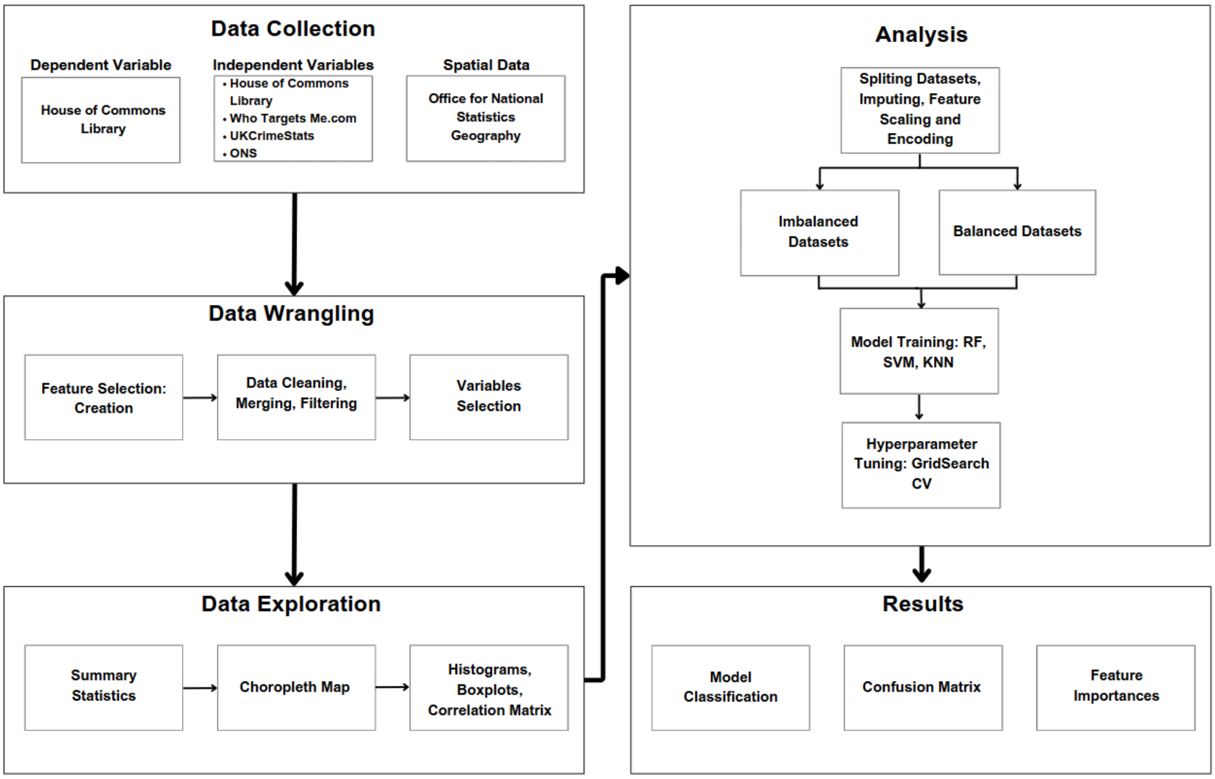

Figure 1: Research Methodology Workflow

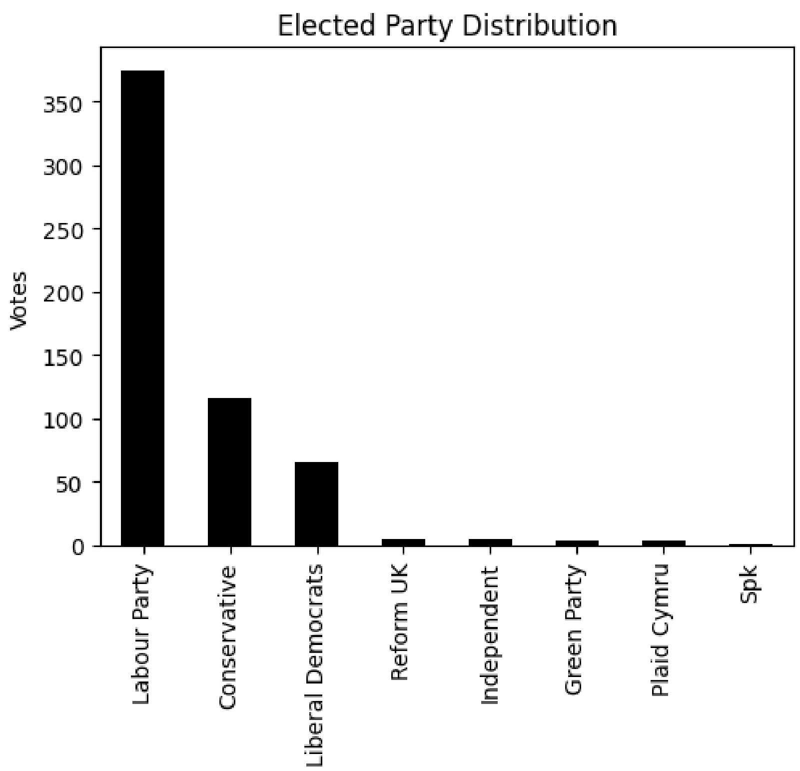

Figure 2: Election Result

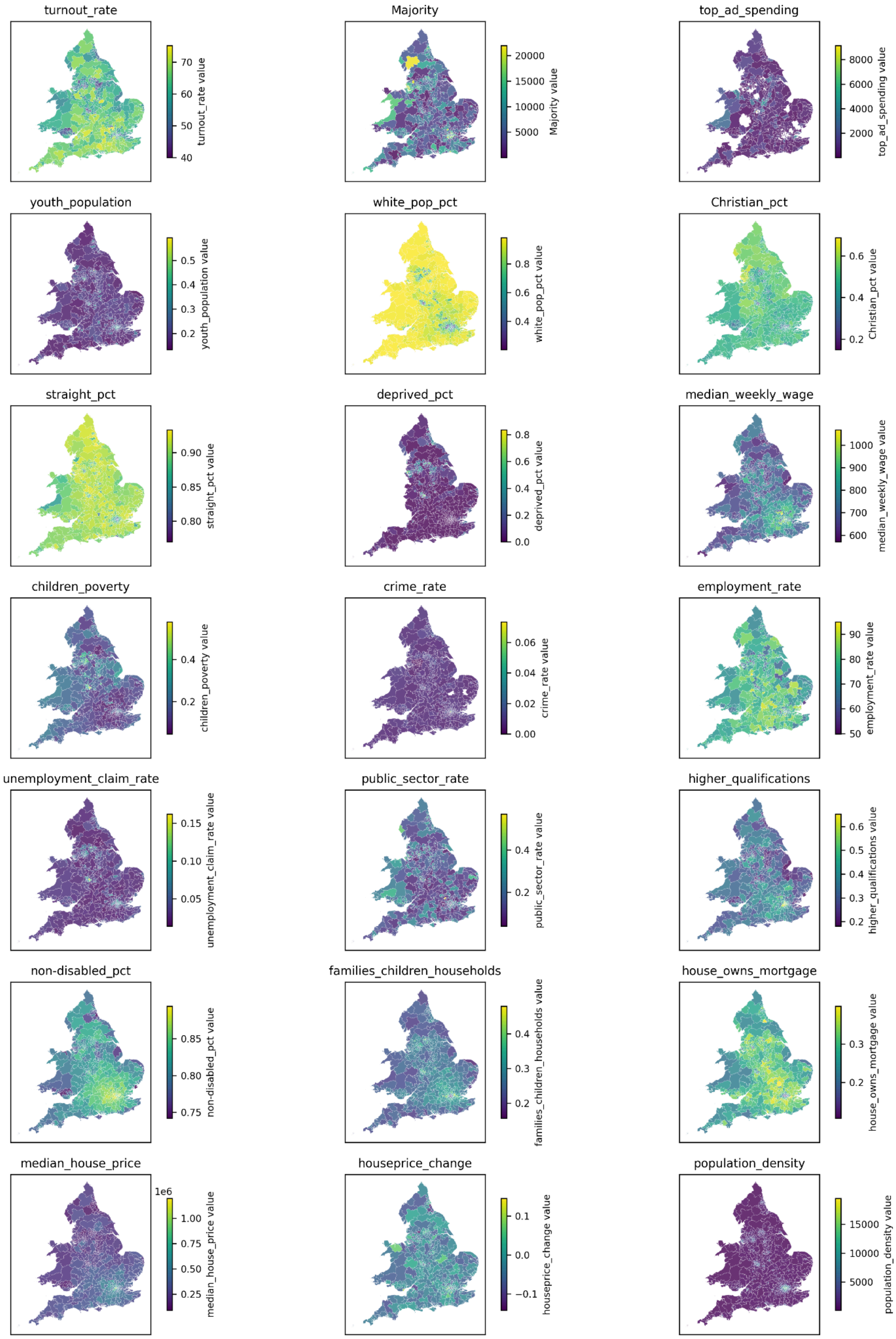

Figure 3: Choropleth Maps of the Independent Variables

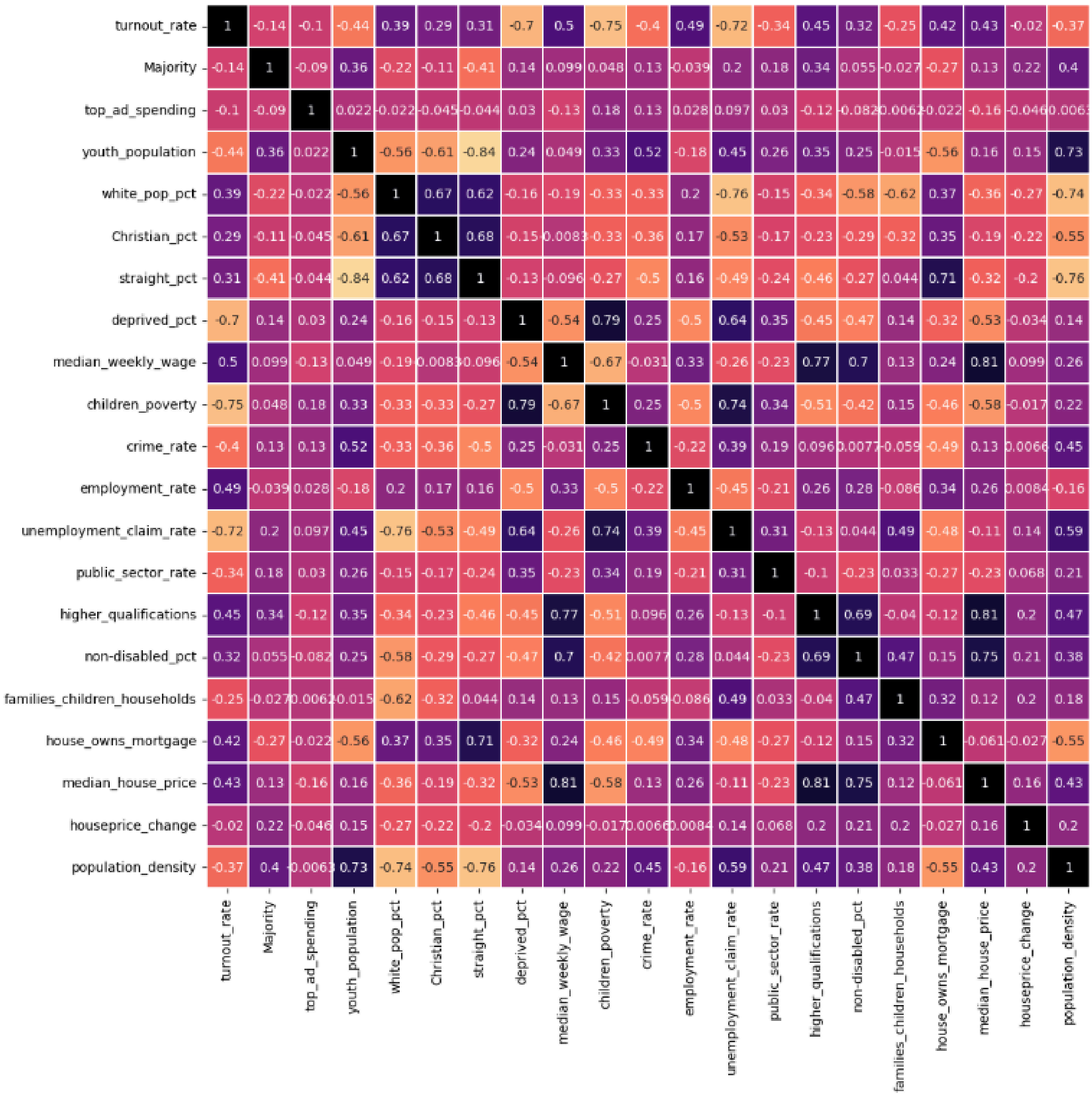

Figure 4: Correlation Matrix

The data is split into training and testing sets using an 80-20 split, performed prior to imputation, feature scaling, and encoding to prevent data leakage (Cross Validated, 2014). Missing numerical values are imputed with the column median, while categorical values are replaced with the most frequent category (Géron, 2017). Since distance-based algorithms like KNN and SVM are sensitive to feature scale, data is standardized to ensure comparability and stability during training. Categorical variables are one-hot encoded to convert them into numeric format (Bhandari, 2024).

Results

From the baseline models, the results are as follows: Random Forest achieved a Training Accuracy of 1.0 and Test Accuracy of 74.78%, Support Vector Machine (SVM) achieved a Training Accuracy of 88.91% and Test Accuracy of 75.65%, and K-Nearest Neighbors (KNN) achieved a Training Accuracy of 83.26% and Test Accuracy of 71.30%. Among these models, SVM showed the highest test accuracy at 75.65%. KNN demonstrated the lowest accuracy overall but was the least overfit, with a discrepancy of approximately 12% between training and test accuracy. Conversely, Random Forest was the most overfit, with a 16% difference between training and test accuracy.

Due to the imbalanced nature of the dataset, all models tend to predict most of the labels as the Labour Party. This bias arises from the overwhelming number of constituencies where Labour secured the majority of votes. In classification problems, imbalanced datasets can negatively affect the accuracy of predictions, as the model becomes skewed toward the majority class. When additional data is available, particularly for the minority classes, it provides more information that helps the model better distinguish rare samples from the majority. Addressing this imbalance is critical for improving the model's ability to generalize across all classes (Sun, Wong and Kamel, 2009; Tyagi and Mittal, 2020).

When comparing the base models to the tuned models, all the tuned models either failed to improve or performed worse than the base models. The F1 macro average and F1 weighted average were also lower for the tuned models, indicating they were not effective at handling the imbalance between the minority and majority classes. The reason the tuned model does not show improved accuracy is that, when working with imbalanced datasets and using the Grid Search method, the balanced accuracy metric provides a more realistic assessment of the model's performance and capabilities. This metric accounts for both the majority and minority classes, which may result in lower overall accuracy but a more equitable evaluation (Kaneva et al., 2024; Mathew et al., 2023).

Among all the base and tuned models with imbalanced dataset, the best-performing model was the base SVM, with an accuracy of 76%, an F1 score (macro average) of 29%, an F1 score (weighted average) of 73%, and a test-train accuracy discrepancy of 13%. These results suggest that the base SVM model is better suited for this imbalanced dataset compared to the other models.

Table 3: Machine Learning Results (Imbalanced Datasets)

Machine Learning Model	Hyperparameters	F1 Accuracy	Macro Average F1 Score	Weighted Average F1 Score
Base Random Forest	max_depth=None, min_samples_leaf= 1, min_samples_split= 2, n_estimators= 100	0.75	0.26	0.71
Tuned Random Forest with GridSearchCV	max_depth= 10, min_samples_leaf= 4, min_samples_split= 2, n_estimators= 75	0.74	0.23	0.68

Based SVM	kernel=linear, C=1	0.76	0.29	0.73
Tuned SVM with GridSearchCV	kernel=rbf, gamma = auto, C=10	0.73	0.28	0.70
Based KNN	n_neighbors=5, weights='uniform', metric='minkowski'	0.71	0.26	0.68
Tuned KNN with GridSearchCV	metric: manhattan, n_neighbors:19, weights: distance	0.71	0.25	0.67

Note: hyperparameters source: (OpenAI, 2024)

For the imbalanced dataset, the models achieved an average accuracy of 73%, an F1 score macro average of 26%, and an F1 score weighted average of 73%. In contrast, with the balanced dataset, the models performed with an average accuracy of 81%, an F1 score macro average of 72%, and an F1 score weighted average of 80%. This indicates that the models handled the balance between minority and majority classes much better, as reflected by the significant improvement in F1 scores. However, some imbalance still remains, as evidenced by the limited impact of GridSearchCV on the performance of the Random Forest and SVM models.

Table 4: Machine Learning Results (Balanced Datasets)

Machine Learning Model	Hyperparameters	F1 Accuracy	Macro Average F1 Score	Weighted Average F1 Score
Base Random Forest	max_depth=None, min_samples_leaf= 1, min_samples_split= 2, n_estimators= 100	0.82	0.72	0.81
Tuned Random Forest with GridSearchCV	max_depth= 10, min_samples_leaf= 2, min_samples_split= 5, n_estimators= 100	0.81	0.70	0.81
Based SVM	kernel=linear, C=1	0.83	0.77	0.83
Tuned SVM with GridSearchCV	kernel=rbf, C=10	0.82	0.77	0.82
Based KNN	n_neighbors=5, weights='uniform', metric='minkowski'	0.75	0.62	0.74
Tuned KNN with GridSearchCV	Metric: euclidean, n_neighbors:25, weights: distance	0.81	0.72	0.81

Note: hyperparameters source: (OpenAI, 2024)

The base Random Forest model remains overfit, achieving 100% accuracy on the training set but only 82% on the test set (18% discrepancy). The limited data likely reduce the efficiency of Random Forest (RF), preventing it from identifying robust patterns and complicating hyperparameter tuning (Sinha et al., 2024). The base KNN model achieved a training accuracy of 86% and a test accuracy of 75% (11% discrepancy). The tuned KNN model improved significantly, achieving 81% test accuracy and reducing the train-test discrepancy to just 5%. The base SVM model is the best-performing model in the balanced dataset, achieving a test accuracy of 83%, an F1 score macro average of 77%, an F1 score weighted average of 83%, and a train-test accuracy discrepancy of only 5%. This indicates no significant overfitting. The SVM confusion matrix for the balanced dataset shows that it only makes incorrect predictions 10% of the time on average.

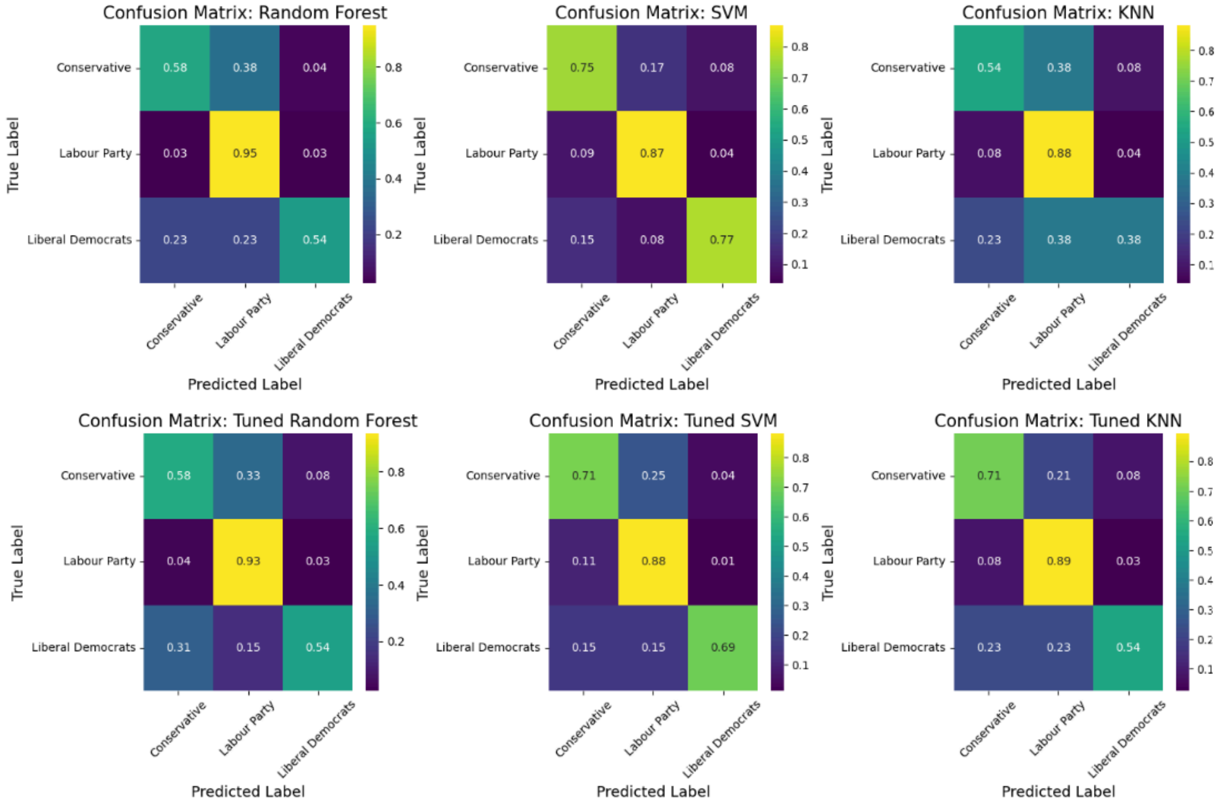

Figure 5: Confusion Matrices (Top row: Imbalanced datasets; Bottom row: Balanced datasets)

From the analysis of feature importance, youth population percentage emerged as the most influential variable in determining classification outcomes in the SVM model. Other important variables according to SVM include: households with children, non disabled population percentage, median wage, and high educational qualification percentages. Interestingly, SVM uniquely highlighted the significance of campaign ad spending during the week leading up to the election. This finding suggests potential areas for future research into how these variables shape election outcomes. However, this study did not account for multicollinearity, which may impact the interpretability of variable importance (Lieberman & Morris, 2014).

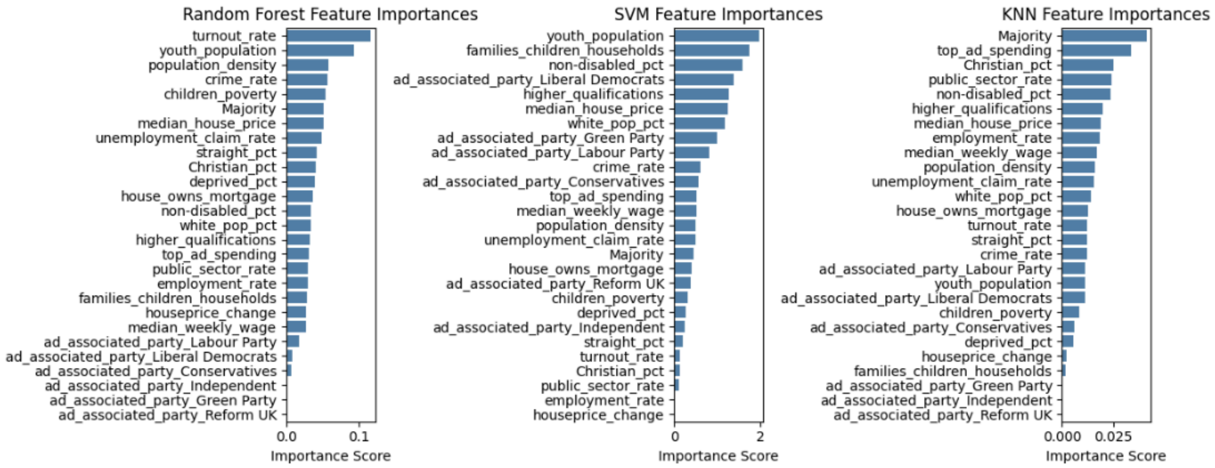

Figure 6: Feature Importance (Balanced datasets)

Conclusions

Random Forest, SVM, and KNN demonstrated strong performance in predicting the 2024 general election outcomes in England and Wales. On the imbalanced dataset, the models achieved an average accuracy of 73%, though the average macro F1 score was only 26%. When trained on the balanced dataset, performance improved to an average accuracy of 81%, a macro-average F1 score of 72%, and a weighted-average F1 score of 80%. Among all models, the base SVM model performed best, achieving test accuracies of 76% on the imbalanced dataset and 83% on the balanced dataset.

The primary limitation of this research is the imbalanced dataset. While limiting the dataset to the top three parties improved model performance, future research could explore alternative approaches, such as predicting party coalitions. This would allow for a more balanced dataset while still incorporating smaller parties. Other methods to address imbalances, such as oversampling minority class examples or undersampling majority classes, could also be explored (Tyagi and Mittal, 2020).

This research only utilized a few machine learning models. Future studies could expand to include algorithms such as neural networks or XGBoost, which may better capture the complexities of the data. Additionally, multicollinearity was not addressed, potentially affecting variable interpretability, thus making feature importance largely unexplored deeply in this study.

Future research could also explore variables that are critical topics in elections, such as migration rates and NHS waiting times, as highlighted in reports like (Ipsos, 2024). Additionally, well-established predictors of election outcomes, such as incumbency and approval ratings, should be considered to provide a more comprehensive analysis (Sinha et al., 2024). Overall, this research enhances the validity of machine learning models such as Random Forest, SVM, and KNN in providing valuable insights into election predictions. By leveraging these models, the public can better prepare for potential political and socioeconomic changes, as well as the ripple effects that may follow after an election.

References

Awad, M. and Khanna, R. (2015). Support vector machines for classification. *Efficient learning machines: Theories, concepts, and applications for engineers and system designers*, pp.39–66.

Baker, C., Pollock, L. and Cracknell, R. (2024). *General election 2024 results*. [online] House of Commons Library. Available at: https://commonslibrary.parliament.uk/research-briefings/cbp-10009/.

Barai, A. (2020). *Normal Distribution and Machine Learning*. [online] Medium. Available at: https://medium.com/analytics-vidhya/normal-distribution-and-machine-learning-ec9d3ca05070 [Accessed 28 Dec. 2024].

Barton, C. (2024a). *Constituency data: Ehnic Groups*. [online] House of Commons Library. Available at: https://commonslibrary.parliament.uk/constituency-statistics-ethnicity/ [Accessed 2024].

Barton, C. (2024b). *Constituency data: Home ownership and renting*. [online] House of Commons Library. Available at: https://commonslibrary.parliament.uk/constituency-data-housing-tenure/ [Accessed 2024].

Barton, C. (2024c). *Constituency data: house prices*. [online] House of Commons Library. Available at: https://commonslibrary.parliament.uk/constituency-data-house-prices/ [Accessed 2024].

Barton, C. (2024d). *Constituency data: Household composition*. [online] House of Commons Library. Available at: https://commonslibrary.parliament.uk/constitunecy-data-household-composition-2021-census/ [Accessed 2024].

Barton, C. (2024e). *Constituency data: population, by age*. [online] House of Commons Library. Available at: https://commonslibrary.parliament.uk/constituency-statistics-population-by-age/ [Accessed 2024].

Bhandari, A. (2024). *Feature Scaling: Engineering, Normalization, and Standardization (Updated 2025)*. [online] Analytics Vidhya. Available at: https://www.analyticsvidhya.com/blog/2020/04/feature-scaling-machine-learning-normalization-standardization/#:~:text=Distance%20algorithms%20like%20KNN%2C%20K,points%20to%20determine%20their%20similarity. [Accessed 27 Dec. 2024].

Bolton, P. (2024). *Constituency data: Educational qualifications*. [online] House of Commons Library. Available at: https://commonslibrary.parliament.uk/constituency-data-educational-qualifications-2021-census/ [Accessed 2024].

Borisyuk, R., Borisyuk, G., Rallings, C. and Thrasher, M. (2005). Forecasting the 2005 General Election: A Neural Network Approach. *The British Journal of Politics and International Relations*, 7(2), pp.199–209. doi:https://doi.org/10.1111/j.1467-856x.2005.00182.x.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45, pp.5–32.

Bulutgil, H.Z. and Prasad, N. (2019). Inequality and voting among deprived ethnic groups: evidence from India. *Journal of Elections, Public Opinion and Parties*, 30(2), pp.1–23. doi:https://doi.org/10.1080/17457289.2019.1624555.

Burnap, P., Gibson, R.K., Sloan, L., Southern, R. and Williams, M.L. (2015). 140 Characters to Victory?: Using Twitter to Predict the UK 2015 General Election. *SSRN Electronic Journal*, 41. doi:https://doi.org/10.2139/ssrn.2603433.

Clark, S.D., Morris, M.A. and Lomax, N. (2018). Estimating the outcome of UKs referendum on EU membership using e-petition data and machine learning algorithms. *Journal of Information Technology & Politics*, 15(4), pp.344–357. doi:https://doi.org/10.1080/19331681.2018.1491926.

Clements, B. and Spencer , N. (2014). *Voting and Values in Britain: Does religion count?* Theos.

Cracknell, R. and Baker, C. (2024). *General election 2024: Results and analysis*. House of Commons Library.

Cross Validated. (2014). *Imputation before or after splitting into train and test?* [online] Available at: https://stats.stackexchange.com/questions/95083/imputation-before-or-after-splitting-into-train-and-test [Accessed 25 Dec. 2024].

Denver, D.T. and Hands, H.T.G. (1974). Marginality and Turnout in British General Elections. *British Journal of Political Science*, 4(1), pp.17–35. doi:https://doi.org/10.1017/s0007123400009340.

Esme Kirk-Wade (2024). *Constituency data: Disability*. [online] House of Commons Library. Available at: https://commonslibrary.parliament.uk/constituency-data-disability/ [Accessed 2024].

Fazil, A.W., Hakimi, M., Akbari, R., Quchi, M.M. and Khaliqyar, K.Q. (2023). Comparative Analysis of Machine Learning Models for Data Classification: An In-Depth Exploration. *Journal of Computer Science and Technology Studies*.

Fieldhouse, E. and Cutts, D. (2012). The Companion Effect: Household and Local Context and the Turnout of Young People. *The Journal of Politics*, 74(3), pp.856–869. doi:https://doi.org/10.1017/s0022381612000345.

Ford, R., Jennings, W., Pickup, M. and Wlezien, C. (2016). From polls to votes to seats: Forecasting the 2015 British general election. *Electoral Studies*, 41, pp.244–249. doi:https://doi.org/10.1016/j.electstud.2015.11.013.

Francis-Devine, B. (2024a). *Constituency data: Child poverty*. [online] House of Commons Library. Available at: https://commonslibrary.parliament.uk/constituency-data-child-poverty/.

Francis-Devine, B. (2024b). *Constituency data: Indices of deprivation*. [online] House of Commons Library. Available at: https://commonslibrary.parliament.uk/constituency-data-indices-of-deprivation/ [Accessed 2024].

Francis-Devine, B. (2024c). *Constituency data: public sector employment*. [online] House of Commons Library. Available at: https://commonslibrary.parliament.uk/constituency-data-public-sector-employment/ [Accessed 2024].

Francis-Devine, B. (2024d). *Constituency data: wages*. [online] House of Commons Library. Available at: https://commonslibrary.parliament.uk/constituency-data-wages/ [Accessed 2024].

Géron, A. (2017). *Hands-on machine learning with Scikit-Learn and TensorFlow concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, Inc.

Hassan, T.A., Hollander, S., van Lent, L. and Tahoun, A. (2019). The Global Impact of Brexit Uncertainty. *The Journal of Finance*, 79(1).

Ipsos (2024). *IPSOS UK GENERAL ELECTION 2024 CAMPAIGN TRACKER*. [online] Ipsos. Available at:

https://www.ipsos.com/sites/default/files/ct/news/documents/2024-07/ipsos-general-election-campaign-tracker-july-2024-charts.pdf.

James, T.S. (2019). Better workers, better elections? Electoral management body workforces and electoral integrity worldwide. *International Political Science Review*, 40(3), pp.370–390. doi:https://doi.org/10.1177/0192512119829516.

Jensen, C.D., Lacombe, D.J. and McIntyre, S.G. (2012). A Bayesian spatial econometric analysis of the 2010 UK General Election. *Papers in Regional Science*, 92(3), p.no-no. doi:https://doi.org/10.1111/j.1435-5957.2012.00415.x.

Johnston, R., Hartman, T. and Pattie, C. (2018). Predicting general election outcomes: campaigns and changing voter knowledge at the 2017 general election in England. *Quality & Quantity*, [online] 53(3), pp.1369–1389. doi:https://doi.org/10.1007/s11135-018-0819-1.

Kaminska, M., Gallacher, J.D., Kollanyi, B., Yasseri, T. and Howard, P.N. (2017). Social Media and News Sources during the 2017 UK General Election. *comprop. oii. ox. ac. uk.*.

Kaneva, T., Evstatiev, B., Valova, I., Valov, N. and Gabrovska-Evstatieva, K. (2024). Comparing Different Evaluation Metrics with Grid Search Method for Classification of Highly Imbalanced Data. *2024 8th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*.

Kazadi, J. and Musimwa, I. (2023). Predicting election results: An improved popularity index. *SSRN Electronic Journal*. doi:https://doi.org/10.2139/ssrn.4646666.

Levene, M. and Fenner, T. (2021). A stochastic differential equation approach to the analysis of the 2017 and 2019 UK general election polls. *International Journal of Forecasting*, 37(3), pp.1227–1234. doi:https://doi.org/10.1016/j.ijforecast.2021.02.002.

Lewis-Beck, M. and Stegmaier, M. (2014). Election Forecasting, Scientific Approaches. In: *Encyclopedia of Social Network Analysis and Mining* . Springer Publishing Company.

Lewis-Beck, M.S. and Stegmaier, M. (2011). Citizen forecasting: Can UK voters see the future? *Electoral Studies*, 30(2), pp.264–268. doi:https://doi.org/10.1016/j.electstud.2010.09.012.

Lieberman, M.G. and Morris, J.D. (2014). The Precise Effect of Multicollinearity on Classification Prediction. *Multiple Linear Regression Viewpoints*, 40(1), pp.5–10.

Mansley, E. and Demšar, U. (2015). Space matters: Geographic variability of electoral turnout determinants in the 2012 London mayoral election. *Electoral Studies*, 40, pp.322–334. doi:https://doi.org/10.1016/j.electstud.2015.10.003.

Mathew, J., Kshirsagar, R., Abidin, D.Z., Grin, J., Kanarachos, S., James, J., Alamaniotis, M. and Fitzpatrick, M.E. (2023). A comparison of machine learning methods to classify radioactive elements using prompt-gammaray neutron activation data. *Scientific Reports*, 13(1).

Murr, A.E. (2016). The wisdom of crowds: What do citizens forecast for the 2015 British General Election? *Electoral Studies*, 41, pp.283–288. doi:https://doi.org/10.1016/j.electstud.2015.11.018.

Nadeau, R., Lewis‐Beck, M.S. and Bélanger, É. (2009). Election Forecasting in the United Kingdom: A Two‐Step Model. *Journal of Elections, Public Opinion and Parties*, 19(3), pp.333–358. doi:https://doi.org/10.1080/17457280903074276.

Obradovic, S., Gentile, G. and Bruter, M. (2023). *Study on participation of citizens with disabilities in elections*. European Commission.

Office for National Statistics Geography (2024). *Westminster Parliamentary Constituencies (July 2024) Boundaries UK BFC*. [online] Office for National Statistics. Available at: https://geoportal.statistics.gov.uk/datasets/ons::westminster-parliamentary-constituencies-july-2024-boundaries-uk-bfc-2/about [Accessed 2024].

Office for National Statistics. (2024). *LI02 Regional labour market: local indicators for Parliamentary constituencies*. [online] Available at: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/datasets/li02regionallabourmarketlocalindicatorsforparliamentaryconstituencies [Accessed 2024].

OpenAI (2024). *now can you create a comprehensive steps if i want to predict ‘elected party’ and i want to compare the prediction performance of RF, SVM KNN*. [online] ChatGPT(December 30, 2024 Version). Available at: https://chatgpt.com/c/67732696-a768-800a-91c4-ad0123101605 [Accessed 30 Dec. 2024].

Owens, J.R. and Wade, L.L. (1988). Economic Conditions and Constituency Voting in Great Britain. *Political Studies*, 36(1), pp.30–51. doi:https://doi.org/10.1111/j.1467-9248.1988.tb00215.x.

Payne, C. (2003). Election Forecasting in the UK: The BBC's Experience. *EURAMERICA*, 33(1), pp.193–234.

Prosser, C. (2024). Fragmentation revisited: the UK General Election of 2024. *West European Politics*, pp.1–13. doi:https://doi.org/10.1080/01402382.2024.2430915.

Rallings, C., Thrasher, M. and Borisyuk, G. (2016). Forecasting the 2015 general election using aggregate local election data. *Electoral Studies*, 41, pp.279–282. doi:https://doi.org/10.1016/j.electstud.2015.11.007.

Richardson, B. and Hougen, D.F. (2020). Districts by Demographics: Predicting U.S. House of Representative Elections using Machine Learning and Demographic Data. *19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp.833–838.

Sinha, P., Kumar, A., Gupta, C. and Biswas, S. (2024). Forecasting US Presidential Election 2024 using multiple machine learning algorithms. *Munich Personal RePEc Archive*.

Sinha, P., Verma, A., Shah, P., Singh, J. and Panwar, U. (2020). Prediction for the 2020 United States Presidential Election using Machine Learning Algorithm: Lasso Regression. *Munich Personal RePEc Archive*.

Sun, Y., Wong, A.K. and Kamel, M.S. (2009). Classification of imbalanced data: A review. *International journal of pattern recognition and artificial intelligence*, 23(4).

Surridge, P. (2024). Values in the Valence Election: Fragmentation and the 2024 General Election. *The Political Quarterly*. doi:https://doi.org/10.1111/1467-923x.13494.

Tyagi, S. and Mittal, S. (2020). Sampling Approaches for Imbalanced Data Classification Problem in Machine Learning. *Proceedings of ICRIC 2019: Recent innovations in computing*, pp.209–221.

UKCrimeStats. (2024). *All constituencies in England, Wales and Northern Ireland, ordered by highest total crime (October 2024)*. [online] Available at: https://www.ukcrimestats.com/Constituencies/?page=12.

Usher, J. and Pierpaolo Dondio (2022). BREXIT: Predicting the Brexit UK Election Results by Constituency using Twitter Location based Sentiment and Machine Learning. *CS & IT Conference Proceedings*, 12(11). doi:https://doi.org/10.5121/csit.2022.121101.

Ward, M. (2024). *Constituency data: people claiming unemployment benefits*. [online] House of Commons Library. Available at:

https://commonslibrary.parliament.uk/constituency-data-people-claiming-unemployment-benefits/ [Accessed 2024].

Whiteley, P. (2023). Has Age Replaced Class in British Elections?A Hierarchical age-period-cohort Analysis. *Electoral Studies*, [online] 86. doi:https://doi.org/10.1016/j.electstud.2023.102695.

Who Targets Me. (2024). *All UK Parties (Google)*. [online] Available at: https://trends.whotargets.me/countries/GB/explorations/ad-spending?report_id=fb9d3dd4-de33-4b19-bc90-5af0a0a0ec5e&start_date=2024-12-06&end_date=2025-01-04 [Accessed 24 Dec. 2024].

Zayed, Y. (2024). *Constituency data: Religion*. [online] House of Commons Library. Available at: https://commonslibrary.parliament.uk/constituency-data-religion/ [Accessed 2024].