

Identifying the Determinants of Financial Inclusion

A Data Mining Approach

Background

Financial inclusion refers to people's ability to access and use formal financial services such as bank accounts, savings, and digital payment systems. It plays an important role in economic participation and financial stability. Although high-income countries generally show high levels of financial inclusion, gaps still exist across different population groups. Understanding the factors associated with financial account ownership can help policymakers and institutions design more inclusive financial systems.

Data

This project uses data from the World Bank Global Findex Database (2025 edition), based on surveys conducted in 2024. The analysis focuses on a subset of 2,630 observations from 51 high-income countries, including a specific comparison involving the United Kingdom. The dataset contains demographic, digital access, and financial behaviour indicators relevant to financial inclusion.

Methodology

A data mining approach was applied using the R programming language. The analysis followed a structured workflow that included data cleaning, preprocessing, feature preparation, and supervised machine learning. Three classification models were implemented and compared:

- K-Nearest Neighbours (KNN)
- Decision Tree
- Random Forest

Model performance was evaluated using accuracy and Cohen's Kappa to assess predictive strength and reliability.

Key Findings

- Decision Tree and Random Forest models achieved very high predictive accuracy
- KNN also demonstrated strong classification performance
- Tree-based models provided useful interpretability through feature importance
- The United Kingdom showed higher financial account ownership compared to the average of other high-income economies
- Digital access and financial behaviour variables were among the most influential predictors

Conclusion

The results indicate that financial inclusion in high-income economies is strongly influenced by digital access, behavioural characteristics, and socio-economic factors. Interpretable machine learning models such as Decision Trees and Random Forests are particularly valuable for this type of analysis, as they combine strong predictive performance with explanatory insights. The findings highlight the importance of digital financial infrastructure in supporting inclusive financial systems.

Portfolio Note

This executive summary is provided for professional portfolio purposes. The full assessed coursework report and raw dataset are intentionally not included in the public repository.