Banks in England Trust by Utilizing Data Science and Computational Machine Learning Algorithms

Author : Hamzah Asghar Farooqi

Student ID: 13191038

farooqih@coventry.uni.ac.uk

Coursework Title: Project Proposal

Degree Title: MSc Data Science And Computational Intelligence

Supervisor By: Peter Every

Coventry University

Faculty of Engineering, Environment and Computing _
School of Computing, Mathematics and Data Science_
7000CEM: Individual Research Project Presentation

Statement of Originality

Hamzah Asghar Farooqi
Student of Coventry University
21/07/2023

Statement of Originality

I, Hamzah Asghar Farooqi , hereby declare that the work presented in this Proposal and Literature Review, titled "Banks in England Trust by Utilizing Data Science and Computational Machine Learning Algorithms," is entirely my original work, except where otherwise indicated. All sources used in this work have been duly acknowledged and cited. This Proposal and Literature Review has not been previously submitted, in part or in whole, for any academic credit or evaluation.

Any contributions from other individuals or sources have been appropriately referenced and acknowledged. I have not engaged in any form of plagiarism or academic misconduct, and I take full responsibility for the content and authenticity of this proposal and Literature review.

I understand the significance of academic integrity and the ethical implications of presenting someone else work as my own. Therefore, I affirm that all ideas, data, and concepts presented in this Proposal and Literature Review are a result of my independent research, analysis, and critical thinking.

Hamzah Asghar Farooqi

Table of Contents

Component 1: Project Proposal 1

Project Title: 1

Project Topic: 1

Introduction 1

Aims , objective and motivations: 2

Client and Audience 2

Data Sample 2

Methodology 3

Potential Risks 4

Tools and Software 4

Ethical Considerations 5

Component 2: Initial Literature Review 5

Bibliography and References 6

Table of Figures

Figure 1 :Descriptive Statistics 3

Figure 2 : Time Management Gaint Chart 4

Component 1: Project Proposal

Project Title:

Banks in England Trust by Utilizing Data Science and Computational Machine Learning Algorithms

Project Topic:

"The Role of Trust in Maintaining Sustainability & Resilience of the Banking Sector in England: Empirical Investigation Using Data from the World Values Survey and Machine Learning Algorithms"

Abstract--Trust is a critical foundation of the financial system, impact both economic stability and customer confidence. The main aim of this study is to conduct a comprehensive survey of the factors affecting bank confidence across England. To achieve this, we leverage data from the World Values Survey (WVS) and apply advanced machine learning algorithms to empirically investigate key trust factors. This research approach uses various models such as linear regression, multiple linear regression, logistic regression, random forest, and decision tree to investigate the relationship between trust in banks and independent variables such as age, financial satisfaction, income size, life satisfaction, general trust, and cash shortage. We use specific predictive analytic techniques to develop the Bank Confidence Variable Index to measure the level of confidence in UK banks. Our research results show that there is a significant correlation between the above variables and bank confidence. In addition, this study comprehensively represents the spectrum of variables using numerical, statistical and visual means such as effect plots and graphs. This study provides valuable insight for both the banking industry and policy makers by clarifying the determinants of trust in England's banks. By understanding the factors that affect trust, banks can develop customer-centric strategies that increase trust, foster customer loyalty and create a resilient financial environment in the England.

Keywords: **World Values Survey (WVS), Logistic regression, Predictive analytic techniques, Financial environment**

Introduction

The significance of trust in banks has been highlighted by the worldwide economic downturn of 2007–2009 or domestic financial crisis of 2014–2017 (which resulted in the bankruptcy of ninety banks, tax relief, nationalization, devaluation of the UAH, skyrocketing inflation, loss of savings, and a significant increase in non–performing loans). The most crucial element for the banking industry is trust, which also provides the fundamental support for the entire financial system. Building trust with clients and stakeholders is crucial given the banking industry's crucial role in facilitating the distribution of capital, supporting economic activity, and fostering growth. customer confidence. Because keeping solid customer connections depends on trust. In order to sustain the financial health of the banking industry, confidence is necessary. The market standing and reputation of a bank are closely related to its credibility. Regulatory compliance is closely related to a bank's duty to follow laws and uphold moral principles. The degree of lending determines how readily available credit is for both borrowing and lending in an economy. Investor trust in banks as prospective investment opportunities is significantly influenced by investor confidence. Systemic risk management entails a connection between trust and the financial system's ability to effectively handle systemic risk.

This study's major objective is to examine the crucial role that trust plays in preserving a nation's banking system's stability and resilience, with an emphasis on public trust as an essential component of banks' ability to operate effectively. This study acknowledges that an absence of trust can result in withdrawals from banks and bank overdrafts and emphasizes the need of comprehending the elements that affect bank trust. Understanding these issues can help financial regulators build resilience. The report seeks to provide answers to important concerns regarding the UK financial crisis including the Deposit Guaranty Scheme (DGS), including:

A) Identify the reasons why individuals dislike banking organizations; B) Find out why people stop trusting banks. c) Recognize the elements that affect the public's confidence in UK banks.

This research project seeks to provide useful insights and give assistance to lawmakers and regulators to strengthen the trust and resiliency of the UK banking system through empirical analysis and modelling methodologies.

Aims , objective and motivations:

This study's primary goal is to perform a thorough investigation to identify the variables that affect England's banks' dependability. The researchers are interested in determining the relationships between bank trust and a number of independent variables, including age, financial contentment, salary size, happiness in life, general trust, or cash shortages. This study attempts to experimentally analyse the main variables that affect trust in the UK banking sector using the strength of World Values Survey (WVS) data and cutting-edge machine learning methods. The Ojective is to clarify the variables that affect people's faith in UK-based banks. Learning why some people have unfavourable opinions of banking institutions as well as how bank trust may be eroding are the main topics of discussion. The study also tries to pinpoint particular elements that significantly affect the public's faith in UK banks. By achieving these objectives, the banking sector and policy makers will gain insightful information that will enhance consumer loyalty, trust, and financial resilience throughout the UK. The worldwide economic recession of 2007–2009 and the local financial crisis of 2014–2017 provide evidence of the important part that trust play in the stability of the financial system and economic confidence, which serves as the driving force behind this study. To forge enduring bonds with clients and stakeholders, trust is seen as a crucial cornerstone of the financial system as a whole. In the banking industry, confidence also impacts lending, investor confidence, regulatory compliance, market status, reputation, and systemic risk management.

Client and Audience

The Client and Audience would likely be individuals and institutions interested in banking, finance, and data science. Specifically, the paper targets researchers, academics, and professionals in the fields of banking, economics, data science, and machine learning.

Data Sample

The World Values Survey - WVS database, containing responses from 1,500 individuals worldwide, served as the source of the data for this study. This extensive sociological survey collects information on people's values and views. The trust respondents have in their banks was the subject of a separate survey question (V121). Using the literature sources mentioned

above, we determined the most pertinent factors after getting the dataset. Additionally, we converted her four response alternatives, which ranged in confidence from the lowest to the greatest degree, into a format appropriate for effect evaluation (positive/negative). With '1' denoting the lowest level of confidence and '4' denoting the highest level, this pre-processing step is crucial in determining the level of confidence. The data preparation job was successfully finished with the aforementioned stages.

Variables	n	mean	sd	median	min	max	range
V121ConfBank	1 500	2,09	0,82	2,00	1,00	4,00	3,00
*V121Dummy_ConfBank**	1 500	0,33	0,47	0,00	0,00	1,00	1,00
V10_Happiness	1 500	2,77	0,81	3,00	1,00	4,00	3,00
V17_ThriftMoney	1 500	0,45	0,50	0,00	0,00	1,00	1,00
V23_LifeSatisf	1 500	5,90	2,46	6,00	1,00	10,00	9,00
*V24_DummyPeopleTrust**	1 500	0,26	0,44	0,00	0,00	1,00	1,00
*V57_Marital**	1 500	0,57	0,50	1,00	0,00	1,00	1,00
V58_NOfChild	1 500	1,35	1,02	1,00	0,00	8,00	8,00
V59_FinSatisf	1 500	4,54	2,46	5,00	1,00	10,00	9,00
V71_ImprtToBeRich	1 500	3,02	1,35	3,00	1,00	6,00	5,00
V103_NeighbTrust	1 500	2,97	0,76	3,00	1,00	4,00	3,00
V104_KnowPersTrust	1 500	3,03	0,62	3,00	1,00	4,00	3,00
V105_UknownPersTrust	1 500	2,08	0,71	2,00	1,00	4,00	3,00
V191_LackOfCash	1 500	2,80	1,04	3,00	1,00	4,00	3,00
*V237_DummySavings**	626	0,15	0,35	0,00	0,00	1,00	1,00
V217NewsPaper	1 500	3,40	1,31	4,00	1,00	5,00	4,00
V219_NewsTV	1 500	4,66	0,83	5,00	1,00	5,00	4,00
V223_NewsInet	1 500	2,40	1,75	1,00	1,00	5,00	4,00
V224_NewsFr.	1 500	4,36	1,10	5,00	1,00	5,00	4,00
V238_SocClass	1 500	3,44	0,89	4,00	1,00	5,00	4,00
V239_ScaleIncome	1 500	4,32	1,88	4,00	1,00	10,00	9,00
*V240_SexFem.**	1 500	0,60	0,49	1,00	0,00	1,00	1,00
V242_Age	1 500	47,23	18,25	47,00	18,00	89,00	71,00
V248_Educ.	1 500	4,08	1,46	4,00	1,00	6,00	5,00

Source: Author`s own work

Figure 1:Descriptive Statistics (Pic Taken from: https://ieeexplore.ieee.org/document/8779974)

Methodology

Here we use the following methodology

1. Data gathering: The World Values Survey (WVS) database, a worldwide sociological survey with data on respondents' values and beliefs, was used by the researchers. The dependent variable was the survey question (V121ConfBank) about bank confidence.

2. Data cleaning and processing: To prepare the data for analysis, they were extracted from the WVS database. From lowest to highest level of confidence, the researchers sorted the responses on bank reliability (V121ConfBank) into his four categories. To determine and prioritise the most crucial elements that might have an impact on bank confidence, they also looked to the body of existing literature.

3. Exploratory Data Analysis (EDA): To visualize and comprehend distributions and correlations among diverse variables, researchers used exploratory data analysis. To understand their data, they employed a variety of visualization approaches, including box plots, histograms, and density graphs.

4. Regression analysis: The researchers utilized linear regression to model the link between the independent factors (age, financial contentment, income, and confidence in people) and the dependent variable (degree of trust in banks). The most crucial predictors in their regression model were found through forward selection.

5. Logistic regression: To assess the robustness of their findings, the researchers used logistic regression to examine the binary dependent variable V121Dummy_ConfBank (1 for bank confidence and 0 for no confidence).

6. Random Forest: They developed a predictive model for banking trust levels using the Random Forest algorithm. This machine learning method makes predictions by combining a number of decision trees and identifies the most significant predictors.

7. Model evaluation: A number of methods, including as cross-validating, confusion matrices, ROC curves, accuracy estimates, along with other pertinent metrics, were used to assess the performance of machine learning models. 8. Interpretation of results: To determine the variables that have a major impact on trust in England banks, the researchers analyses the regression and machine learning model results.

Potential Risks

There are potential risks and limitations associated with it. Some of these risks are Data Limitations, Sample Bias, Model Over-fitting, Causality vs. Correlation, External Validity, Lack of Domain-specific Factors, Interpret-ability of Machine Learning Models and Time Sensitivity.

Tools and Software

- Python is a popular programming language for data analysis and machine learning.
- Jupyter Notebook allows for an interactive and reproducible computational environment, which is well-suited for data exploration, analysis, and visualization.
- Pandas is a powerful Python library used for data manipulation and analysis.
- Scikit-learn is a popular machine learning library in Python.
- Matplotlib and Seaborn are libraries used for data visualization in Python.
- Random Forest Implementation, Data Cleaning and Preprocessing, Notebook Environment, Latex. Etc

Time Management

Here are time management plan for our project

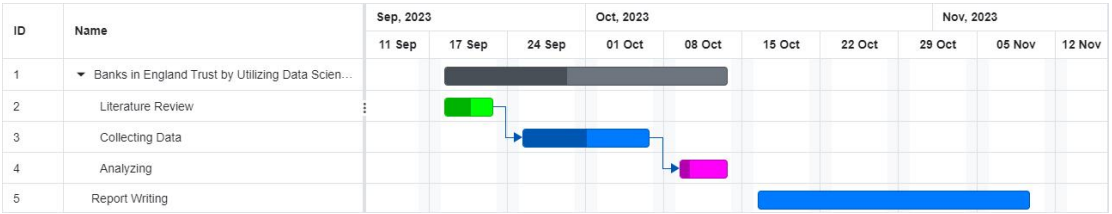

Figure 2: Time Management Gaint Chart

Ethical Considerations

The suggested project on determining the influence of different factors on the level of trust in the Banks in England Trust by Utilizing Data Science and Computational Machine Learning Algorithms tools may raise several legal and ethical considerations I.e. Data Privacy and Consent, Data Bias, Transparency and Explainable, Model Fairness, Intellectual Property, Responsible Use of Results, Data Security, Ethical Review.

These legal and ethical considerations will help ensure the research integrity, validity, and responsible use of data science and machine learning techniques in the study of trust in the England Bank.

Component 2: Initial Literature Review

The literature review aims to provide an extensive overview of existing academic studies on the significance of trust in preserving the sustainability and resilience of the UK banking system. Trust is a critical factor that underpins the entire financial system, establishing strong connections between financial institutions, clients, and stakeholders. Economic transactions heavily rely on trust, facilitating efficient capital allocation, stimulating economic activity, and fostering overall growth.

Confidence is highly correlated with bank resilience, which is a bank's capacity to absorb external shocks and systemic hazards. The degree of trust that the public has in banks affects how resilient they are to financial crises and downturns. In times of turbulence, resilient banks can manage systemic risk, uphold regulatory compliance, and meet client needs. Therefore, trust is crucial to a bank's ability to run efficiently and sustainably.

New opportunities for trust analysis in the banking industry have emerged as a result of recent developments in data science and machine learning. The connection between trust in banks and independent variables has been studied by researchers using predictive analytics approaches such logistic regression, random forests, and decision trees (Bengio et al., 2020). These techniques will make it possible to study trust patterns in the banking industry in a more thorough and data-driven manner.

In the literature, a number of factors that affect trust in banks have been found. Customer satisfaction, financial contentment, income level, overall trust, and life satisfaction are some of these drivers (Molander & Wikstrom, 2018). Demographics and age also affect trust in banks because different age groups may have varying levels of trust (Hannan & McDowell, 1987). In order to design strategies to boost client loyalty and trust in the banking industry, it is essential to understand these elements.

This research project specifically focuses on the European Central Bank (ECB) and aims to examine the factors influencing trust in the ECB using data science and machine learning approaches. The data for this study is obtained from the Eurobarometer Survey 89 by the European Commission. The study employs predictive analytics to forecast ECB confidence levels and uses various data visualization charts, such as ECB TrustMap graphs, correlation heat map matrices, and delta plots, to illustrate the impact of respondent replies and decision-making behavior on network structure changes.

The methodology utilizes multiple models, including logistic regression, decision trees, random forests, and neural networks, to generate the central bank confidence index. The results are validated through cross-validation, confusion matrices, ROC curves, accuracy

estimations, and logistic model variance inflation factor (VIF) analyses. The findings reveal that several factors significantly influence public trust in the European Central Bank, including belief in the euro, concerns about inflation, hopes for the future of the EU, well-being indices, and other relevant elements. This study provides valuable insights into the critical elements affecting public trust in the ECB.

Furthermore, the literature review highlights the potential impact of data science technologies, such as artificial intelligence (AI), Internet of Things (IoT), big data, behavioral/predictive analytics, and blockchain, in transforming government operations and giving rise to GovTech firms. The "smartening" of public services and national infrastructure in government is expected to have a greater impact due to its essential role and significance to institutions and individuals. The review outlines promising GovTech systems, including robo-advisors to assist officials, real-time IoT and blockchain management of national assets, automated compliance/regulation, secure storage of public records on blockchain distributed ledgers, online constitutional and dispute resolution structures, and laws encoded as blockchain smart contracts.

In conclusion, trust plays a crucial role in the long-term viability and adaptability of the UK banking industry. The banking system's foundation relies on trust, which affects client confidence, economic stability, and the overall health of the industry. This literature review establishes a theoretical framework for ongoing research by establishing connections with previous work and identifying areas for further investigation. The integration of data science and machine learning into trust analytics holds immense promise in providing valuable insights to support the development of customer-centric strategies by the banking industry and policymakers. These strategies aim to strengthen the UK banking industry's financial resilience, foster loyalty, and enhance consumer trust.

Bibliography and References

[1] Peter Eccles, Paul Grout, Paolo Siciliani, Anna (Ania) Zalewska “The impact of machine learning and big data on credit markets” Bank of England, July 9, 2021.

[2] Priyanka C Hiremath, Raju G T “Recommender System for Geo-Social Access Control Framework” International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-2S, December 2019.

[3] Priyanka C Hiremath ,Dr. G T Raju “Analogy of Trust Models for Information Security” Proceedings of IEEE Third International Conference on Multimedia Processing, Communication & Information Technology – MPCIT 2020.

[4] Andrii Skirka, Bogdan Adamyk, Oksana Adamyk, Mariana Valytska “Trust in the European Central Bank: Using Data Science and predictive Machine Learning Algorithms”10th International Conference, 2020.

[5] B. Adamyk, A. Skirka, K. Snihur and O. Adamyk, "Analysis of Trust in Ukrainian banks based on Machine Learning Algorithms," 2019 9th International Conference on Advanced Computer Information Technologies (ACIT), Ceske Budejovice, Czech Republic, 2019.

[6] Z. Engin and P. Treleaven, "Algorithmic Government: Automating Public Services and Supporting Civil Servants in using Data Science Technologies," in The Computer Journal, vol. 62, no. 3, pp. 448-460, March 2019

[7] David Bholat,Mohammed Gharbawi, Oliver Thew “The Impact of Covid on Machine Learning and Data Science in UK Banking” Bank of England, 19 Oct 2021

[8] Joel Suss, Henry Treitel, “Predicting Bank Distress in the UK with Machine Learning ” Bank of England Working Paper No. 831, Octdober 2019.

[9] Yang Hong-tao “Research on Cost Decision of Specialized-Automobile Manufacturing Enterprise Based on the Theory of Decision Tree” Wuhan University of Technology, Wuhan, 2010.