Machine Learning for Financial Inclusion and Safety: Empowering Women Against Violence

MSc Research Project
MSc FinTech

Adedoyin Fatobi
Student ID: X23111569

School of Computing
National College of Ireland

Supervisors: Noel Cosgrave and Faithful Chiagoziem Onwuegbuche

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: ……. Adedoyin Fatobi……………………………………

Student ID: …………x23111569……………………………….……

Programme: …………MSc in FinTech…………………… **Year:** …2023/2024…..

Module: ………………MSc Research Project…………..………

Supervisor: …. Noel Cosgrave and Faithful Chiagoziem Onwuegbuche ……

Submission Due Date: ………………September 16, 2024……………………………………..………

Project Title: ……………… Machine Learning for Financial Inclusion and Safety: Empowering Women Against Violence ………..………

Word Count: ………6171……………… **Page Count**……………23…………….……..

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: …………………Adedoyin Fatobi…………………………………………

Date: …………………September 16, 2024…………………………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Machine Learning for Financial Inclusion and Safety: Empowering Women Against Violence

Adedoyin Fatobi
x23111569

Abstract

This study examines how digital financial inclusion could help in reducing the gender-based inequalities that exists with access to financial technology. For the study, we explore the use of machine learning algorithms applied to dataset sources from the World Bank Demographic and Health Survey. The study utilized a number of machine learning models from the Random Forest classifier to Principal Component Analysis (PCA), and the Random Forest Regressor along with parametric tuning process using the RandomizedSearchCV to optimize the model parameters. Amongst other things, the study was able to show that the chosen models were appropriate for the designed tasks as model performance R squared values of 0.888 implying 88.8% accuracy, 0.9996 implying 99.9% accuracy. Extents of digital financial inclusion showed Sierra Leone, Russian Federation, Mongolia, El-Salvadore lacking behind in the post- Covid Era. Violence Hotspots against women were mainly concentrated in countries like Guinea, Mali, Sierra Leone, Ethiopia and Chad. The study further adapted the models to the data to create composite indices viz- the Financial Exclusion Index and the Vulnerability Index. The Financial Exclusion Index was used to plot geographical patterns and disparities in financial exclusion across countries telling the different areas of the extents of financial exclusion. The combination of the indices – Financial Exclusion Index and the Vulnerability Index also helped in distinguishing cluster variations for high, low and medium financial exclusion across the world.

1 Introduction

In Machine learning models have in recent times been applied in various sectors from healthcare to finance. One area that has also been affected is the promotion of financial inclusion and safety especially for vulnerable demographic like women (Nwaimo, et al., 2024). By making use of alternate data sources such as those from mobile phone usage, social media interactions, and transactional records, machine learning models can provide tailored solutions for individuals with limited traditional financial histories. (Nwaimo, et al., 2024).

Financial inclusion refers to the procedures whereby it is ensured that all individuals have access to financial services especially the populations that are often marginalised by the regular banking systems, while safety is concerned with physical and financial security which are important for the empowerment and well-being of women.

This demographic - women especially has been noted to be among the least served in terms of access to financial services with evidence depicting how several types of financial inclusion projects when well designed and implemented can benefit them (Department for International Development, 2013). This in spite of the advancements of brought about using sophisticated tools like machine learning models. According to (The World Bank Group, 2021), there are approximately one billion women globally that are excluded from the formal financial system. In addition to this, about one out of three women will experience sexual or physical violence in their lifetime (United Nations, 2020). This data shows an urgent need to leverage novel solutions like machine learning models to address issues like this in a holistic way.

Machine leaning, which is a subdivision of artificial intelligence usually involves developing systems or algorithms that allow computers learn from available data and then make decisions or predictions (Thyago, et al., 2019). It consists of a range of applications from natural language processing to predictive analytics and is currently being largely used in the financial sector to analyse copious data amounts, detection of fraud and the personalisation of financial services. In the same vein machine learning models has shown applicability in the domain of safety especially in the prediction and prevention of violence through pattern recognition and predictive modelling.

1.1Justification for the Research

In connecting machine learning, financial inclusion and women's safety holds a unique chance to address two essential challenges at the same time. While lots of research has been conducted in each of these areas individually, there still exists a considerable gap in the literature exploring their convergence. Most of the available studies focus mainly on more generalised applications of machine learning in finances or its deployment in predicting and preventing violence (Jenga, et al., 2023; Negre, et al., 2024; Rodriguez-Rodriguez, et al., 2020). There are however a few literatures that have explored how machine learning technologies can be integrated especially while empowering women by equipping them with financial tools and ensuring their safety (Levantesi & Zacchia, 2021; Smith & Rushtagi, 2021; Vinuesa, et al., 2020).

This study attempts to fill this gap by exploring how marching learning can be leveraged to promote financial inclusion among women and protect them from violence.

1.2Research Questions and Hypothesis

The main research question guiding this study is: *How can we use machine learning to enhance financial inclusion and safety for women, especially in the context of preventing violence?*

To address this question, the following sub-questions will be considered:

1. What are the existing machine learning models employed in financial inclusion?
2. How can these models be adapted or enhanced to address the specific needs of women?
3. What machine learning techniques are effective in predicting and preventing violence against women?
4. How can financial inclusion and safety be integrated into a cohesive framework using machine learning?

The hypotheses for this research are:

1. Machine learning models can be tailored to improve financial inclusion for women via improved financial services.
2. Predictive analytics using machine learning can effectively identify and prevent potential violence against women.

2 Related Work

In this section, we review previous related work to the study topic and provide a critical review exploring the contributions, strengths, and weaknesses of each paper with a view to justifying the need for this study.

Bitetto, et al., (2024) in their work investigated the ability of machine learning to predict credit risk ratings for small firms. The study rightfully showed that there was a reduction in the rates of default and increased access to loans. The study however was not gender-specific in its inquiry, dealing mainly with the business as an entity. There were also limitations mentioned about data privacy and the potential for algorithm bias. While the machine learning models could make access to credit better, there was also the risk of reinforcing existing biases if the training data did not represent the domain of enquiry well. The paper however produced the bright suggestion that different data sources should be integrated and there should be an implementation of robust privacy measures to lessen the negative effects.

Siddiqui (2024) in their work explored the use of AI and IoT services while enhancing women’ s safety. The study shed light on smart surveillance systems, wearable IoT devices, and AI-driven threat analysis to provide an environment that is more secure. While this was considered as largely feasible there were the expressed limitations of privacy concerns and ensuring that such systems are deployed in an ethical manner.

With a focus on financial inclusion, Rahman, et al. (2023) examined the use of machine-learning algorithms in the prediction of vulnerability to domestic violence among women in Liberia. Techniques used included as ANN, KNN, RF, DT, XGBoost, LightGBM, and CatBoost. The study observed that LightGBM and RF models were the most effective in attaining the highest accuracy in predicting women’s vulnerability. The study showed the importance factors like emotional violence as playing a significant role in the risk of domestic violence. It also showed the need for all-encompassing methods that would include social and financial dimensions. The major weakness of this study is its selective focus on machine learning methods.

Mabrouk, et al. (2023) examined the role of digital financial services in boosting financial inclusion for women focusing on how such services can help further gender equality and make women economically empowered. Among other things, the study concluded that digital financial services are important for empowering women giving them access to financial services.

2.1 Digital Financial Inclusion

2.1.1 Evidence Supporting Women's Empowerment through Digital Financial Inclusion

Digital financial inclusion has become one of the most significant ways that women are empowered economically. The advent of technologies like mobile digital wallets, and financial services that can be accessed on the web, access has increased for women who were before removed from the typical banking systems (Jedi, 2022). This section looks at the available empirical evidence for the advantages of digital financial inclusion for the empowerment of women.

2.1.2 Case Studies and Empirical Evidence

There are a number of studies that show the transformative effect of digital inclusion in finance in the lives of women. According to Jack and Suri (2016), the Kenya mobile money product – M-Pesa has since its advent brought betterment to women's economic status in the country. The study observed that women using this service tended to save and put their money into small businesses that lessened their dependence on informal savings. In the same vein the (International Financial Corporation, 2017) observed in their study that providing digital financial services to women in Bangladesh gave the women motivation to assert themselves on financial issues and have a voice in making household decisions.

Another study conducted by the (Consultative Group to Assist the Poor, 2019) sought to unravel the ascendance of digital financial indicators in India. The study observed that women who could access digital financial services tended to work beyond the home in paid jobs and they could also put money into educating their children and enjoy a higher quality of life. This evidence suggests that there is a positive effect in including women in digital financial services and gives impetus to study how to do this using advanced methods like machine learning.

2.1.3 Effects on Economic Empowerment of Women

Digital financial inclusion has a wider spillover effect and goes beyond the mere access to financial services, leading digital payment facilities as an entry point for women economic empowerment. This includes an increase in financial independence, better opportunities to start a business and an all-round boost of their social status (Garz, et al., 2020). Another advantage is that digital financial inclusion can enhance women's financial literacy, by introducing them to financial products and services through the use of digital means (Jedi, 2022).

Furthermore, we find evidence that digital financial inclusion increases women's participation in the formal economy. This study by the Global Partnership for Financial Inclusion (2020) found that women using digital financial services were greater users of their bank accounts, more likely to have formalized a business, accessed credit and diversified entrepreneurship activities. This subsequently, leads to wider economic growth and development.

2.1.4 Challenges and Limitations

Digital financial inclusion also has some challenges and problems for the effectiveness despite this promising evidence. Gender norms and low levels of digital literacy pose socio-cultural barriers that limit women's access to use basic API-based services, such as an electronic account-centric approach. Women in low- and middle-income countries are 10% less likely than men to own a mobile phone, according to the GSMA (2019)-thus affecting their ability acquire mobile financial services.

In addition, the digital divide is still a large hindrance to success. In a report by the International Telecommunication Union (2020), it was stated that women have less access to internet than men, and this is even truer for developing countries. This digital divide reinforces inequalities, constraining the effect of programs aimed at achieving financial inclusion through digitization.

Interventions are required to be targeted in order to address these barriers. To make digital financial inclusion effective, initiatives are required to strengthen literacy on digitization, address socio-cultural standards and support affordable accessibility of all kinds and forms of technology. Creating an environment that empowers women financially will require coordinated actions between governments, financial institutions, and technology providers.

2.2Violence

2.2.1 Insight on How AI Can Predict Violence-prone Zones

The technologies of artificial intelligence (AI) can be leveraged to improve security conditions for women. With the use of machine learning, big data form several sources can be analysed to discover patterns and trends that are not easily observable. This can help in the prevention of violence against women. In this section, we investigate the predictive potential of AI in this regard.

2.2.2 Background of Predictive Analytics and AI

Predictive analytics is conducted by searching for existing patterns within data that may be used for predicting outcomes and trends (Farayola, et al., 2024). In the prevention of violence especially against women AI can be used to analyse big data like social media content, crime reports or logs of emergency calls to unveil patterns that indicate violence hotspots (Dakalbab, et al., 2022).

2.2.3 AI for Violence prediction: Case-studies and Empirical

A few studies show AI to be effective in the prediction and prevention of violence (Varun, et al., 2023) in their study examined 150 articles to provide evidence for effective predictive policing methods using machine learning and deep learning methods. The authors observed increased levels of effectiveness that led to better violence prevention and increased security.

In studies conducted examining live scenarios the Chicago Police Department implemented a predictive policing program using AI to identify individuals and locations at higher risk for violent crime. According to a study by Saunders, et al. (2016), this initiative led to a significant reduction in gun violence in the targeted areas. Similarly, a project in Los Angeles utilized AI to analyze social media data and identify potential gang-related violence hotspots, resulting in a decrease in violent incidents (Saunders, et al., 2016).

In India, an AI-based platform named "SAFE" (Situational Awareness for Enhanced Security) has been used to predict and prevent crimes against women. According to a report by the World Economic Forum (2020), SAFE analyses data from various sources, including emergency calls, social media, and public transportation data, to identify high-risk areas and times for potential violence. This information is then used to deploy security resources more effectively and to inform public awareness campaigns.

Furthermore, the city of New York has leveraged AI to enhance its crime prevention strategies. By utilizing machine learning algorithms, the NYPD has been able to predict and prevent violent crimes in certain neighborhoods. A study by the RAND Corporation (2013) found that the implementation of AI-driven predictive policing in New York resulted in a 15% reduction in violent crimes over a two-year period.

2.2.4 Challenges and Limitations

However, the application of AI for predictive modeling has its limitations. Of course, there is an important constraint for the availability and quality of data. In many areas, the data of violent incidents can be obscured or wrong which then would make AI predictions not as useful. In addition, human behavior is dynamic and requires significant effort to keep predictive models up-to-date and accurate (Lum & Isaac, 2016).

Another obstacle is how to deliver AI-driven interventions. Although AI can pinpoint those areas most likely to experience future outbreaks of violence, transforming those predictions into successful responses will require the combined efforts of law enforcement, community organizations and policy makers (Cockerill 2024). Meaningfully rooted in communities, not to endure increased surveillance and criminalization.

Furthermore, many regions lack the resources needed to deploy costly AI systems with even higher complexity - it is especially so in low- and middle-income countries. To ensure AI-driven violence prevention initiatives are viable for the long term, Hunt et al (2020) argue investment in infrastructure is necessary to support them with training and maintenance.

3 Research Methodology

The methodology integrates different data sources with a common theme and used machine learning models to explore issues of financial inclusion and safety for women. This method has been used in previous work. The World Bank developed the Global Findex Database which has a store of comprehensive data on the transaction activities of adults around the word including how they borrow money, make payments, and manage risk. The database contains integrated data from over 140 countries and makes use of statistical and machine-learning methods to analyse financial inclusion metrics (Demirguc-Kunt, et al., 2018).

Overall, the methods will comprise data preprocessing to indicator filtering and preparation for principal component analysis which will be used to create important indices like the financial exclusion and vulnerability indices. There would be exploratory data analysis to uncover trends and hotspots, model building using Random Forest to predict violence and financial exclusion levels followed by clustering to categorise countries based on created indices. The full scheme is shown below:

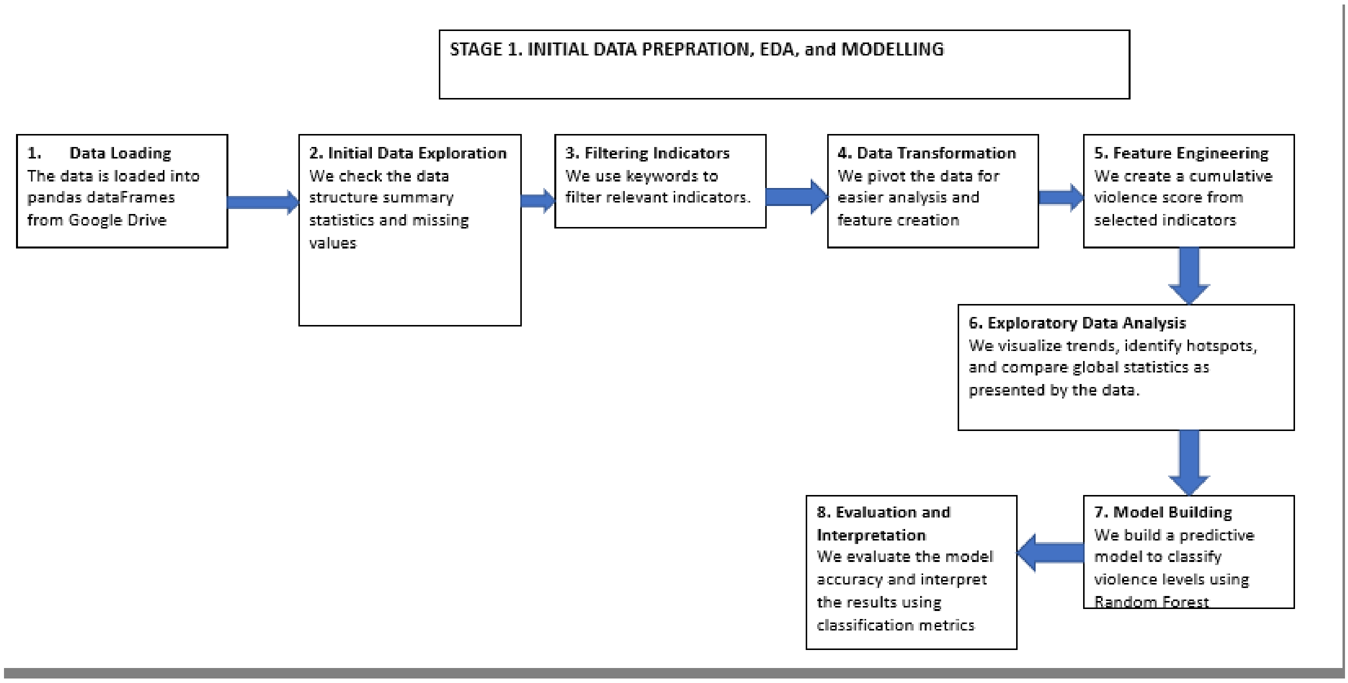

Figure 1. Methodolgy Flow Diagram 1

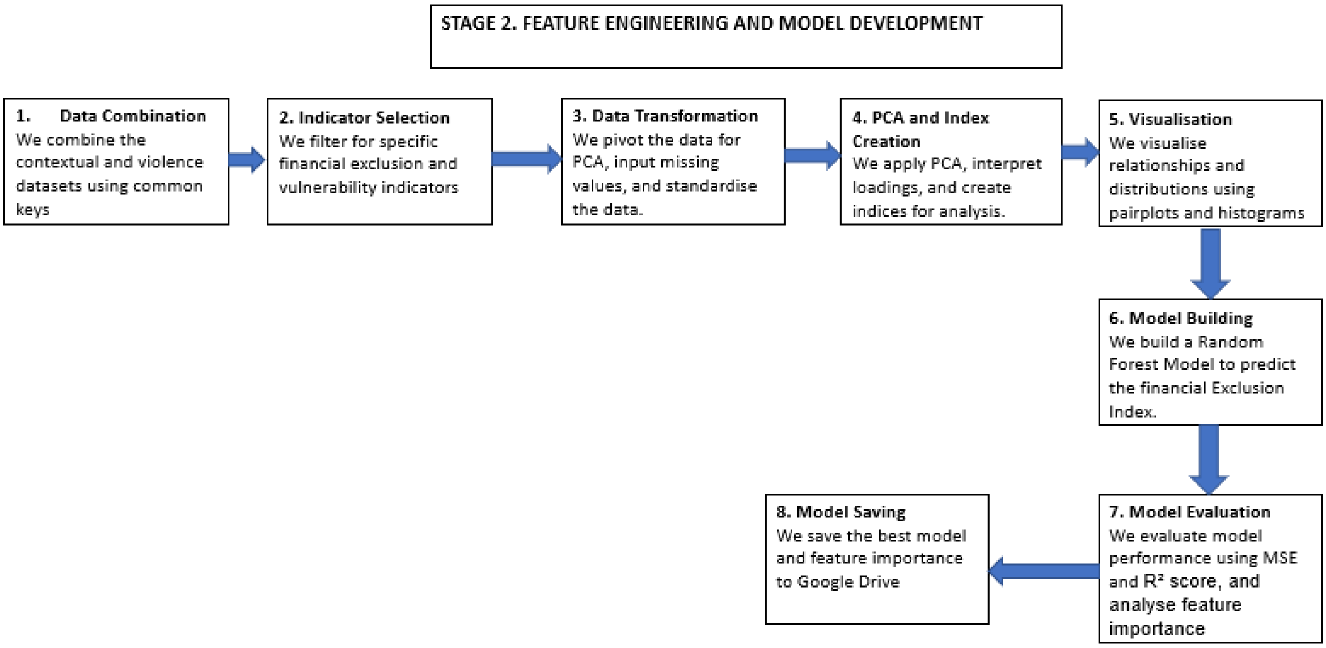

Figure 2:Methodology Flow Diagram 2

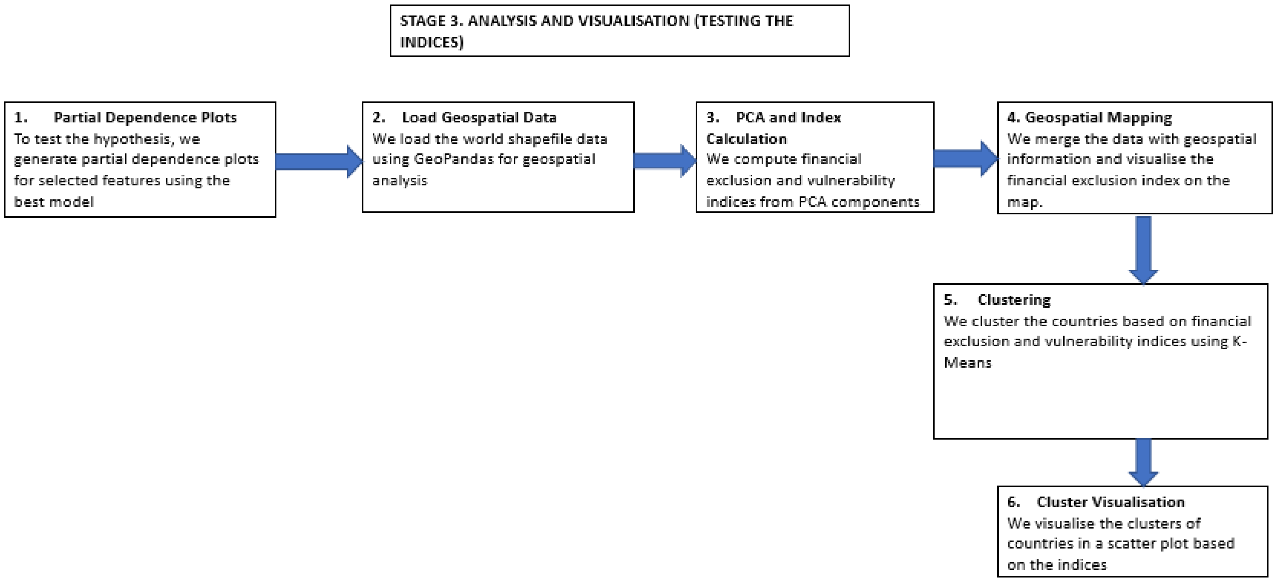

Figure 3: Methodology Flow diagram 3

3.1Data Description

3.1.1 Data Structure and Sources

The data used for the study was obtained from the World Bank and Demographic and Health Survey and it contains the following indicators relevant to the study.

Contextual Indicators: This dataset is made up of socio-economic and demographic data, demarcated by income levels, education, and geographical distribution.

Violence Indicators: This contains data related to gender-based violence, such as incidence rates, types of violence, and availability of support services.

Financial Inclusion Indicators: this includes information on financial access and usage, like account ownership, savings behavior, and credit access.

3.2Part 1: Exploratory Data Analysis

This was first conducted to observe the trend on two counts – the trend by country for mobile money usage around the world post COVID-19. Secondly, there was exploratory data analysis for violence hotspots for women and the observable trends for violence.

3.2.1 Financial Inclusion

This is the first part of the exploratory data analysis and it focusses on digital financial inclusion indices within the data. From the start we initiate the EDA process by mounting Google Drive to access the contextual_indicators.csv, which is loaded into a panda DataFrame. In the initial part, the data was verified with the display of the first few rows to ensure that the data has been read correctly.

3.2.2 Mobile Money Indicators

This part of the analysis focused on the indicators related to mobile money usage from the contextual_indicators dataset. The data was filtered as appropriate to conduct a Post-COVID

Trends Analysis for general use of mobile money and then more specifically, independent use of mobile money.

3.2.3 Violence Indicators

This was done with the violence.csv dataset. A list of keywords relevant to violence on women was used to filter the dataset. The filtered dataFrame was then transformed into a pivot table using 'Country Name' and 'Year' as index. A 'Violence Score' column was added to the filtered dataFrame with values derived from the cumulative sum of the violence indicator score for each country. This data was then visualized to show the top ten (10) violence hotspots.

The analysis also involved a trend analysis showing the global mean and median violence values in the course of time (Seen in Figure 5). A trend analysis was also conducted for the identified violent hotspots over time (Seen in Figure 6). The last part of the exploratory analysis here involved a predictive model building used to predict high (above median violence levels) as against low (below median violence levels). The performance of the model is also evaluated.

3.3Part 2: Combined Analysis of Contextual and Violence Indicators

The second part of the analyses combined the two datasets contextual_indicators.csv, and violence.csv to conduct an analysis across countries.

3.3.1 Data Loading and Merging

The datasets contained in csv files contextual_indicators.csv, and violence.csv were loaded into google colab from google drive and read using pandas package "read_csv." These were then merged on the basis of common columns – Country Code and Year.

3.3.2 Data Preprocessing

Indicator Filtering: To ensure we were touching on the right issues; we filtered the relevant indicators that have to do with financial exclusion and vulnerability by searching for key terms in the indicator names in the combined dataset. The filters contained the keywords 'financial,' 'school,' 'education' for financial indicators and 'poverty,' 'literacy,' 'violence,' 'harassment' or vulnerability indicators.

Creating Composite Indices:

Financial Exclusion Index and Vulnerability Index

Through the integration of contextual indicators and violence-related data, we created two key indices using Principal Component Analysis (PCA).

Financial Exclusion Index: This index captures the principal components of the financial exclusion indicators by summarising the multifaceted barriers and limitations women face in accessing financial services.

Vulnerability Index: This index encapsulates the dimensions of vulnerability related to exposure to violence and socioeconomic instability.

3.3.3 Machine Learning Analysis: Data Preparation for Machine Learning

A. Supervised Learning

For the model, we use supervised learning which learns a function that maps an input with an output on the basis of an example input-output match (Vankov & Gadhoumi, 2024).
The Financial_Exclusion_Index is set as the target variable for the model. The principal components and the Vulnerability_Index are used to predict the target variable. By predicting the Financial_Exclusion_Index, the model will seek to provide insights into the factors that promote financial exclusion. The included Vulnerability_Index helps to examine the effect of vulnerability and violence on financial exclusion.

B. Clustering Analysis

Data is scaled and clustered using the K-means algorithm to uncover patterns and segment the population based on financial inclusion and vulnerability indicators.

C. Evaluation Methodology

The evaluation involves cross-validation to ensure the robustness of the machine-learning models. To optimize the parameters for the machine learning model, we used the RandomizedSearchCV which assists in the finding of the best set of hyperparameters for the random forest model to enhance the model performance. Performance metrics are also compared across different models to select the best-performing one.
Sensitivity analysis was conducted to assess the impact of different predictors (the strongest) on the models' performance in a bid to ensure the reliability of the results.

4 Design Specification

This section outlines the techniques and framework underpinning the implementation, along with the associated requirements. A detailed description of the model's functionality is included.

4.1Techniques and Architecture

Data Integration Framework: for the first part of the analysis – exploratory data analysis, the datasets were examined individually and assessed for their unique features to guide the exploratory process.
Subsequently, in the second section of the analysis, we imported, merged, and preprocessed the datasets using python. The framework ensures that data from various sources is harmonised and prepared for analysis.
Principal Component Analysis: PCA is applied to reduce the dimensionality of the data, capturing the most significant variance in fewer dimensions – we used a total of 5 principal components.
Visualization and Analysis of PCA Results: we explore the explained variance ratio to understand the importance of each principal component. This would be seen in the next section 5.
K-means Clustering: this system works by partitioning the dataset into K clusters. Each observation would then belong to the cluster bearing the nearest mean. This approach helps

with identifying basic patterns and segments within the data. This was applied for unsupervised learning to highlight the patterns and segments.

4.2 Machine Learning Model

4.2.1 Random Forest Model

This is a collective learning method that constructs multiple decision trees during training and outputs the mean prediction of the individual trees. It is known to handle large datasets and maintain high accuracy (Nnanna, et al., 2024). For this project, it is used for supervised learning to predict the Financial Exclusion Index. The model is chosen for its robustness and ability to handle large datasets with numerous predictors.

4.2.2 Random Grid Search

To optimize the parameters for the machine learning model, we used the RandomizedSearchCV which assists in the finding of the best set of hyperparameters for the random forest model to enhance the model performance.

5 Implementation

This section discusses the implementation of the proposed solution, including the final stage of implementation and the outputs produced.

Transformed Data: The raw data is transformed through preprocessing steps, including imputation of missing values and creation of composite indices. The transformed data is stored in a structured format suitable for analysis.

Machine Learning Models: the following models are used:

Random Forest Model: This is developed and trained to predict the Financial Exclusion Index. The model's output will include predictions and performance metrics such as RMSE and R-squared.

Visualizations: Various visualizations are created to highlight key relationships and findings, including bar charts, heat maps, line graphs, scatter plots, and geographic heat maps.

5.1.1 Tools and Languages Used

Python: this is used for data loading, pre-processing, and analysis. Packages such as pandas for data manipulation, `RandomForest` for machine learning, and `matplotlib` for visualization.

Microsoft Excel: This is used for initial data exploration and basic pre-processing steps.

6 Evaluation

This chapter details the findings and analysis with discussions of the results in relation to the stated research questions and hypothesis.

6.1Part 1: Exploratory Data Analysis (EDA)

Financial Inclusion

This section sets a context for the analysis as we explore the indicators relating to financial inclusion and violence. Figure 1 below shows the EDA result on the average mobile use by country.

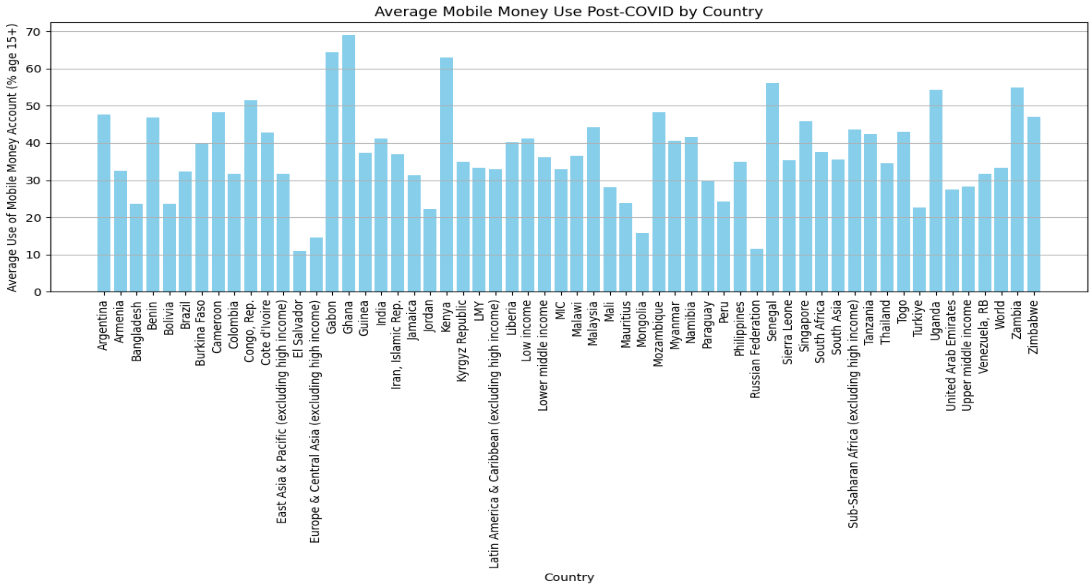

Figure 4. Average Mobile Money by Country

Figure 2 below shows the EDA result on the top 5 countries with the Highest use of Mobile Money Post-COVID.

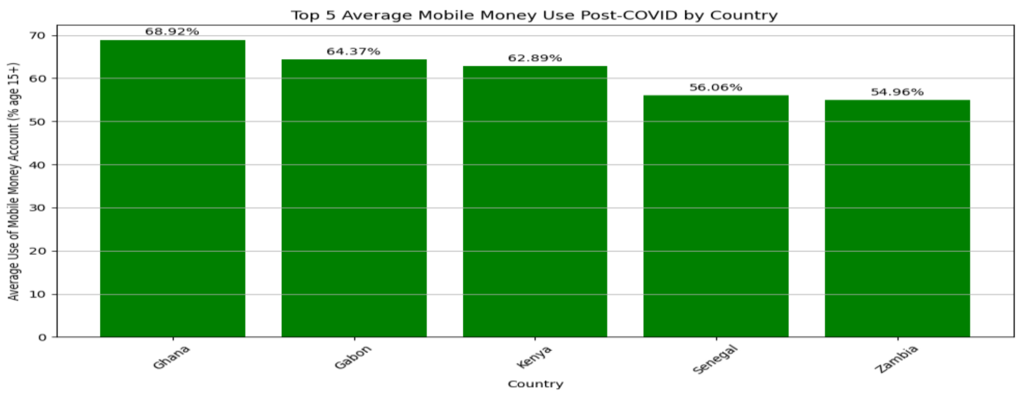

Figure 5. Highest Use of Mobile Money Post-COVID

Figure 3 below shows the EDA result on the top 5 countries with the Lowest use of Mobile Money Post-COVID with El Salvador being the country with the lowest use.

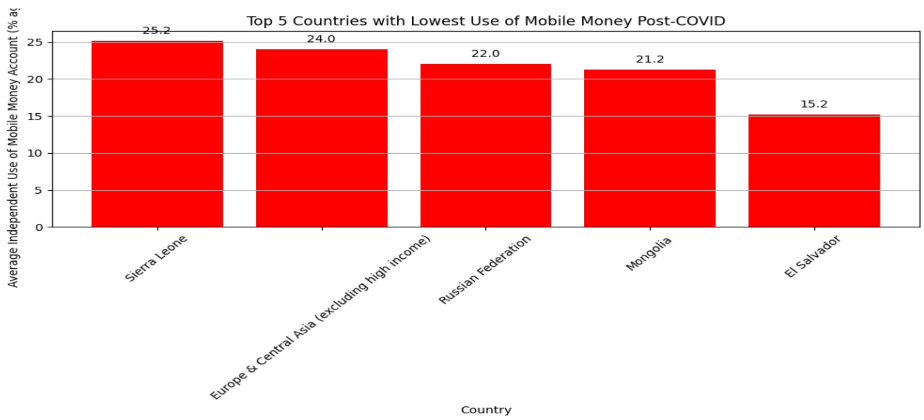

Figure 6. Lowest use of Mobile Money Post-COVID

Violence

This section continues with the exploration of the indicators relating to violence. Figure 4 below shows the EDA result on the top 5 violence hotspots by countries.

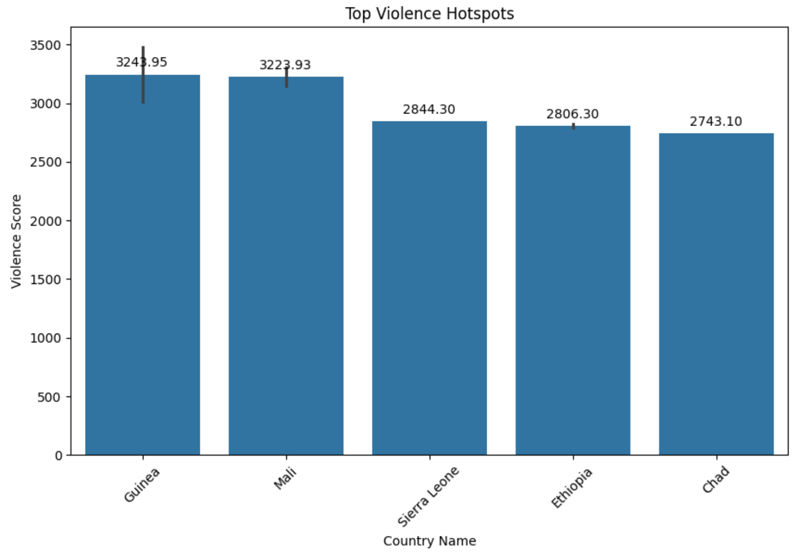

Figure 7: Top 5 Violence Hotspots

Trend Analysis of Violence Incidents

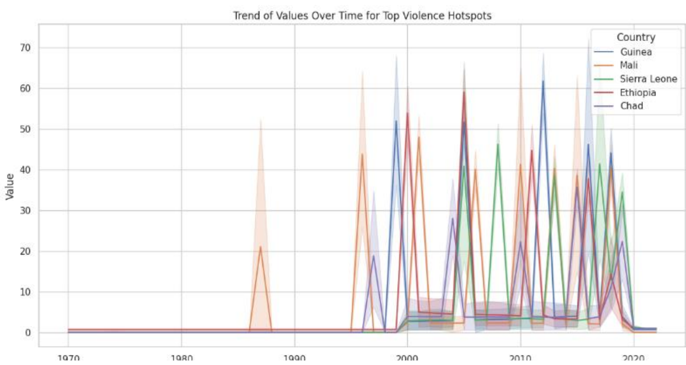

Figure 8: Values over time for Top Violence Hotspots

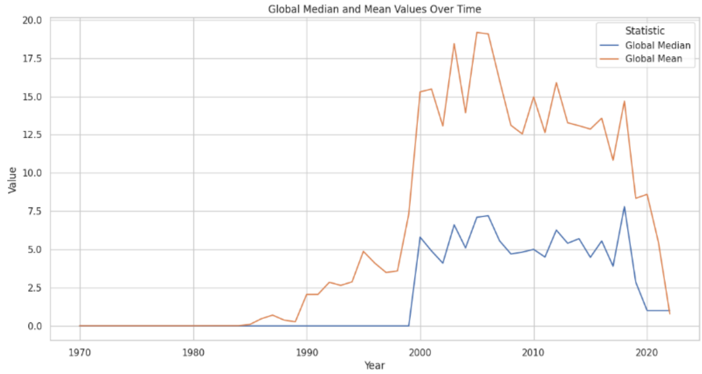

Figure 9: Global Mean an Median Values over Time (Violence Hotspots)

6.2Part 2: Combined Analysis of Contextual and Violence Indicators

6.2.1 Target Variable

In the analysis, we create a target variable 'Financial_Exclusion_Index' used in the machine learning model. This involved a combination of steps like data merging, filtering, transformation, and principal component analysis (PCA).

6.2.2 Combination of Datasets

Combining the datasets allowed for merging contextual indicators and violence datasets which is important for understanding the complex nature of financial inclusion and gender-based violence.

The data head for the combined datasets are shown in Figure 7 below:

	Indicator Name_contextual	Indicator Code_contextual	Country Name_contextual	Country Code	Year	Value_contextual	Indicator Name_violence	Indicator Code_violence	Country Name_violence	Value_violence
0	Children out of school (% of primary school age)	SE.PRM.UNER.ZS	Afghanistan	AFG	1993	73.234001	Criminal penalties or civil remedies exist for sexual harassment in employment (1=yes; 0=no)	SG.PEN.SXHR.EM	Afghanistan	0.000000
1	Children out of school (% of primary school age)	SE.PRM.UNER.ZS	Afghanistan	AFG	1993	73.234001	There is legislation on sexual harassment in employment (1=yes; 0=no)	SG.LEG.SXHR.EM	Afghanistan	0.000000
2	Children out of school (% of primary school age)	SE.PRM.UNER.ZS	Afghanistan	AFG	1993	73.234001	There is legislation specifically addressing domestic violence (1=yes; 0=no)	SG.LEG.DVAW	Afghanistan	0.000000
3	Children out of school (% of primary school age)	SE.PRM.UNER.ZS	Afghanistan	AFG	1974	73.177788	Criminal penalties or civil remedies exist for sexual harassment in employment (1=yes; 0=no)	SG.PEN.SXHR.EM	Afghanistan	0.000000
4	Children out of school (% of primary school age)	SE.PRM.UNER.ZS	Afghanistan	AFG	1974	73.177788	There is legislation on sexual harassment in employment (1=yes; 0=no)	SG.LEG.SXHR.EM	Afghanistan	0.000000

Figure 10. Data head for combined datasets

6.2.3 Identification of Relevant Indicators

The data was filtered for indictors relating to financial exclusion and vulnerability to ensure that the analysis focuses on the most important variables. Keywords - 'financial', 'exclusion', 'banking', 'income', and 'poverty' were used to filter for financial inclusion. Vulnerability was filtered with - 'vulnerability', 'risk', 'violence', 'crime', and 'safety'.

.

6.3Predictive Modelling

The Financial_Exclusion_Index was the target variable for the model, a Vulnerability_Index was also incorporated into the model to help analyse the impact of vulnerability on financial exclusion and violence. A Random Forest Regressor was used to ensure predictive accuracy and modulate overfitting. This was used to predict the target variable based on the principal components.

6.3.1 Model Performance Metrics

Mean Squared Error (MSE): 0.021566873071589818

For an MSE value of approximately 0.022 the model's predictions are quite close to the actual values.

R^2 Score: 0.888413700181317

For an R^2 score of approximately 0.871, about 88.8% of the variance in the target variable is explained by the model. This indicates a strong model fit.

6.3.2. Possible Limitations of the Model

There may however be limitations to these metrics. For instance, the R^2 score is typically high and this might indicate that the model may have captured some noise in the data along with the underlying patterns. There is also the possibility that the feature selection process included shades of redundancy in the selected features which may interfere with model accuracy. Given the broad dimensionality of the raw data, this cannot be ruled out.

6.3.2 Feature Importance

Feature importance indicates the contribution of each feature to the model's predictions. Higher values suggest that the feature is more influential in predicting the target variable.

Feature	**Importance**
PC1	0.445169
PC2	0.217461
PC5	0.104365
PC3	0.095931
Vulnerability_Index	0.063625
PC4	0.05402
Year	0.019429

Table 1. Feature Importance Values

The loadings with the highest values for the indicators for the strongest feature PC1 are shown below:

Indicator Name_contextual	PC1
Reason for not having a mobile money account: don't have the necessary documentation (% age 15+)	0.267664
Reason for not having a mobile money account: available mobile money products are too expensive (% age 15+)	0.259897
Reason for not having a mobile money account: mobile money agents are too far away (% age 15+)	0.254372
No account because of a lack of necessary documentation (% age 15+)	0.234838
Reason for not having a mobile money account: don't have enough money to use a mobile money account (% without an account, age 15+)	0.23469
Reason for not having a mobile money account: do not have their own mobile phone (% age 15+)	0.232251
No account because financial institutions are too far away (% age 15+)	0.231964

Table 2: Description of Indicator Loadings

PC1 appears to be the most important feature which contributes 44.5% of the prediction of the model. This is closely followed by PC2. The 'Vulnerability_Index' and 'Year' are less important but still contribute to the model's predictions.

6.4Test of Hypothesis

6.4.1 Hypothesis 1

Machine learning models can be tailored to improve financial inclusion for women via improved financial services.

This is broken down into the following null and alternative hypothesis.

Null hypothesis

Machine leaning models cannot tailored to improve financial inclusion for women via improved financial services.

Alternative Hypothesis

Machine leaning models can be tailored to improve financial inclusion for women via improved financial services.

This was done using the residual analysis. The plot is shown below:

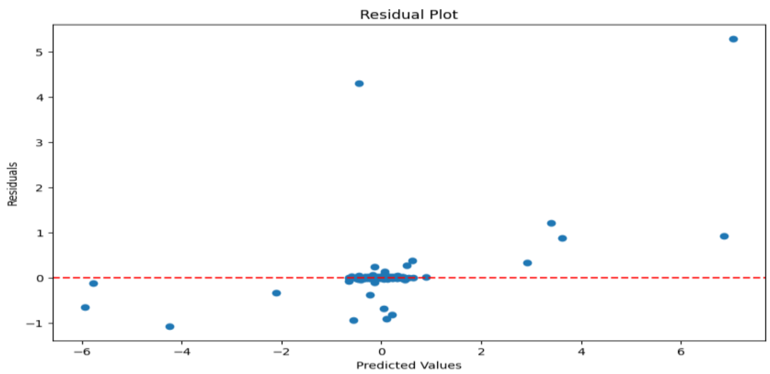

Figure 11: Residual Plot for the Random Forest Regressor

From the residual plot we see that the Random Forest Regressor performs well for most data points. There is also no clear pattern with the residuals which implies a generally good fit, however, there are some outliers. From this evidence we can reject the null hypothesis.

6.4.2 Hypothesis 2

Predictive analytics using machine learning can effectively identify and prevent potential violence against women.

This is broken down into the following null and alternative hypothesis.

Null hypothesis

Predictive analytics using machine learning cannot effectively identify and prevent potential violence against women.

Alternative Hypothesis

Predictive analytics using machine learning can effectively identify and prevent potential violence against women.

For this, we use the sensitivity analysis. The plot is shown below:

Sensitivity Analysis

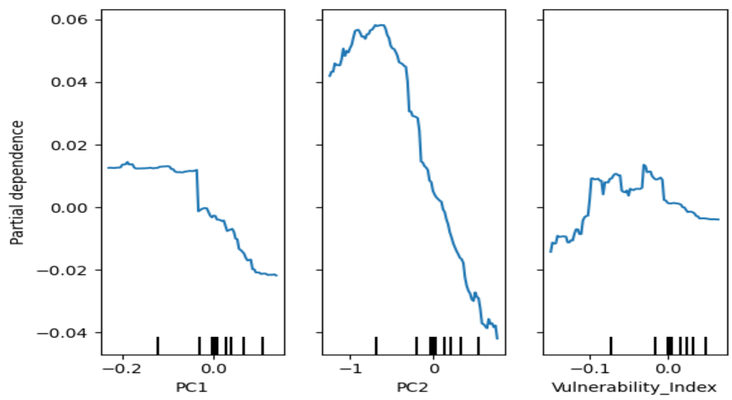

Figure12: Sensitivity Analysis Plots

The plot shows how changes in components PCI and PC2 (the strongest features from the analysis) and the Vulnerability_Index determine output productions related to financial exclusion sowing that improvement in economic factors and infrastructural factors can reduce exclusion. Both features show a positive relationship with the Vulnerability_Index which implies that reducing vulnerabilities to factors like poverty and violence is important.

With this analysis, we can reject the null hypothesis.

6.4.3 Other Insights from the Analysis

Here, we try to examine specifics that come with the data based on our analysis.

6.4.4 Geospatial Distribution of the Target Variable (Financial Exclusion Index)

The output of the plot is shown below:

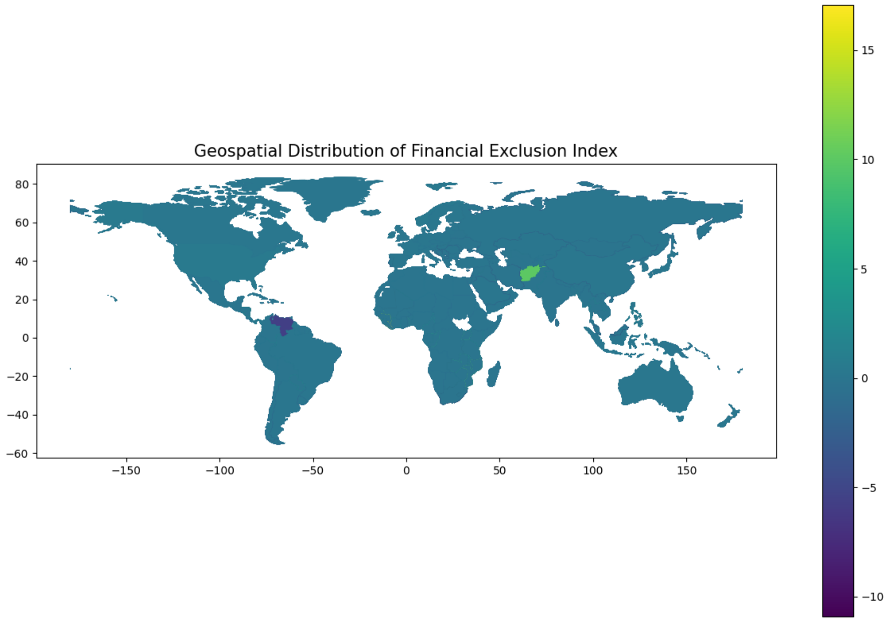

Figure 13: Geospatial Distribution of the Financial Exclusion Index

Cluster of Countries Based on Indices

This was done using K-Means Clustering. The output is shown in the plot below:

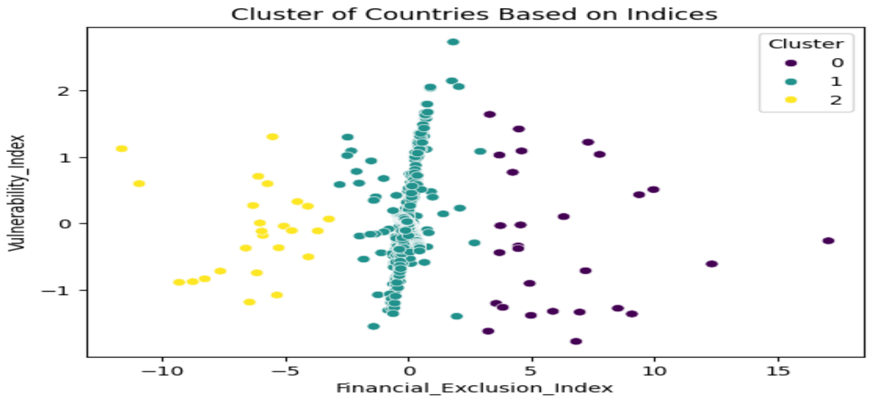

Figure 14: Cluster of Countries based on Indices

6.5. Discussion

The analysis began from the context of a measure of the level of financial inclusivity and levels of violence experienced in the countries as presented by the data. The exploratory data analysis conducted revealed those countries like Kenya, Ghana, Gabon, Senegal, and Zambia had high mobile money adoption rates. Ghana led the pack with a 68.92% rate of adoption. The data thus showed a high progression of financial inclusion especially in the regions where the traditional banking systems tend to exclude vulnerable demographic like women.

The data depicted locations like Guinea, Mali, Sierra Leone, Ethiopia, and Chad as top violence hotspots. Guinea being the highest with about 3,244 incidents.

A deeper look into the violence data showed an increasing trend over time with clear spikes in some countries in times that have historically noted for upheaval in such countries. An example here is the late 1990s in Sierra Leone

Haven gotten this initial outlook on the data. The analysis was tailored to answer the research questions.

How financial inclusion and safety can be integrated into a cohesive framework using machine learning.

To answer this question, the two datasets on financial inclusion indictors and violence indictors were combined. After this, we reduced the dimensionality of the combined data using PCA to create new variables - the principal components that capture a summary of the variance of the original data. The Financial Exclusion Index was determined from this using the weighted sum of the first 3 components while the Vulnerability Index was determined using the last 2 principal components capturing aspects of the data that border on vulnerability. We then used the Financial Exclusion Index and the Vulnerability Index as inputs in a RandomForestRegressor model to make predictions. This uses indices from the PCA to make accurate predictions.

Model performance metrics gave an MSE value of 0.022 and R2 score of 0.888413. This was a strong model fit showing that our attempt to integrate details on financial inclusion and safety into a cohesive model for prediction using machine learning was successful.

How machine learning models can be adapted or enhanced to address the specific needs of women

We sought to see how machine models could be adapted to address specific needs of women. In the light of this a predictive model was built to predict violence hotspots. The model gave an accuracy of 0.9997 indicating a 99.97% accuracy of the predictions made. This was backed up by a precision value of 1.00 indicating that there are no false positives and a recall value of 1.00, also indicating that there are no false negatives.

The model gave an F1 score of 1.00 indicating a perfect precision and recall showing a well-performing model. This shows how machine learning could be adapted to predict violence. This can be useful for women.

Another way this question was answered was that we took our created Index – the Financial Exclusion Index and plotted its geospatial distribution. This immediately gave plots that showed high financial exclusion areas – Central Asia. Moderate financial exclusion areas – a global spread involving most parts of the world and low financial exclusion areas – North America, Western Europe. The depiction which was shown on a map help confirm the accuracy of the Financial Exclusion Index created by this study. This is supported by data from the literature. For instance, in a study conducted by the World Bank – as of 2017 – over 116 million adults in Europe and Central Asia had no traditional bank accounts (Demirguc-Kunt & Muller, 2019). There report also shows that women in the area had 25% lesser opportunities than men to have formal accounts (Demirguc-Kunt & Muller, 2019). The United States was also shown by a World Bank Report to have a banked to unbaked rate of 21% as at 2019 showing a low rate of financial exclusion as that period compared to other developing regions (World Bank Group, 2022). The spatial distribution of our Financial Exclusion Index shows United States as an area with low financial exclusion further making our created index for this study valid. An index like this becomes useful in demarcating areas of high vulnerability and low vulnerability which can be useful when trying to cater to the women demographic.

Our K-Means clustering was also used to demarcate countries into different clusters. This was based on our 2 indices -Financial_Exclusion_Index and Vulnerability_Index. This also helped to depict countries where there could be prominent levels of financial exclusion and vulnerability form the lower ones which can become useful when catering to women.

7 Conclusion and Future Work

In this study, we sought to determine existing machine learning models employed in financial inclusion and how can these models be adapted or enhanced to address the specific needs of women. We also sought to determine what machine learning techniques are effective in predicting and preventing violence against women and how can financial inclusion and safety be integrated into a cohesive framework using machine learning. From the literature review, we saw that a number of models were deployed in this regard some of these includes ANN, KNN, RF, DT, XGBoost, LightGBM, and CatBoost. This study was able to use the RandomForestRegressor, Principal Component Analysis amongst others with the models giving performance metrics that showed prominent levels of accuracy. We were also able to merge financial inclusion and safety into a cohesive framework using machine learning. Not only were the indices use for prediction, but their predictions could also be verified as accurate when compared to available data from the literature.

This study has thus demonstrated the critical role that digital financial inclusion can play in reducing gender inequalities and enhancing the safety and security of women and girls in developing nations. By leveraging advanced machine learning techniques, we have identified key predictors of financial exclusion and vulnerability, providing actionable insights for policymakers and stakeholders. As we move forward, it is essential to continue building on these findings, fostering collaboration, and innovatively addressing the challenges and opportunities in this vital area. Also, nothing is stopping doing research on the other genders as now gender classes are being expanded to various classes e.g., men, transgender people etc.

References

Bitetto, A. et al., 2024. Can we trust machine learning to predict the credit risk of small businesses?. *Review of Quantitative Finance Accounting,* pp. 1-30.
Bjorkegren, D. & Grissen, D., 2017. Behavior Revealed in Mobile Phone Usage Predicts Loan Repayment. *SSRN Electronic Journal.*
Cockerill, R. G., 2024. Ethics Implications of the Use of Artificial Intelligence in Violence Risk Assessment. *The Journal of the American Academy of Psychiatry and Law ,* 52(2), pp. 1-5.
Consultative Group to Assist the Poor , 2019. *Advancing Financial Inclusion to Improve the Lives of the Poor,* Washington, DC: Consultative Group to Assist the Poor .
Cui, J., Liu, J. & Liao, Z., 2013. *Research on K-Means Clusitering Algorithm and its Implementation.* s.l., Atlantis Press.
Dakalbab, F. et al., 2022. Artificial intelligence & crime prediction: A systematic literature review. *Social Sciences and Humanities Open ,* 6(1), pp. 1-23.
Damasevicius , R., 2023. Progress, Evolving Paradigms and Recent Trends in Economic Analysis. *Finacial Ecoomics Letters ,* 2(2), pp. 35-47.
Demirguc-Kunt, A. et al., 2018. *The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution.* [Online]
Available at: https://chooser.crossref.org/?doi=10.1596%2F978-1-4648-1259-0
[Accessed 19 July 2024].
Department for International Development, 2013. *Promoting women's Financial Inclusion: A Toolkit.* [Online]
Available at:
https://assets.publishing.service.gov.uk/media/5a7c6a8eed915d6969f44b53/promoting-womens-financial-inclusion-toolkit.pdf
[Accessed 3 August 2024].
Devineni, S. K., 2024. AI in Data Privacy and Security. *International Journal of Artificial Intelligence and Machine Learning,* 3(1), pp. 35-49.
Farayola, O. A. et al., 2024. Advancements in predictive analytics: A philosophical and practical overview. *World Journal of Advanced Research and Reviews,* 21(03), pp. 240-252.
Figeiredo, D., Junior, S. J. & Rocha, E. C., 2011. What is R2 all about?. *Leviathan – Cadernos de Pesquisa Política,* 3(3), pp. 60-68.
Garz, S., Heath, R., Kipchumba, E. K. & Sulaiman, M., 2020. *Evidence of digital financial services impacting women's economic empowerment: What explains the impacts and what is left to learn?,* Dhaka, Bangaldesh: Women's Economic Empowerment and Digital Finance .
Global Partnership for Financial Inclusion , 2020. *Advancing Women's Digial Financial Incusion ,* Saudi Arabia : Global Partnership for Financial Inclusion, World Bank Group.
Greenacre, M., Groenen, P. J. & Hastie, T., 2022. Principal Component Analysis. *Nature Reviews Methods Primers,* 2(1), p. 100.
GSMA, 2019. *Connected Women: The Mobile Gender Gap Report,* London: GSMA.
Hartley, S. E., 2010. *Kiva.org: Crowd-Sourced Microfinance & Cooperation in Group Lending,* s.l.: s.n.

Hodson, T. O., 2022. Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. *Geoscientific Model Development* , 15(14), pp. 5481-5487.
Hunt, X. et al., 2020. Artificial Intelligence, Big Data, and mHealth: The Frontiers of the Prevention of Violence Against Children. *Frontiers in Arificial Intelligence* , 12(1), pp. 1-16.
International Financial Corporation, 2017. *Merchant Payments and Digital Financial Services,* Washington, D.C: International Financial Corporation.
International Telecommunication Union , 2020. *Bridging the gender divide.* [Online] Available at: https://www.itu.int/en/mediacentre/backgrounders/Pages/bridging-the-gender-divide.aspx
[Accessed 11 July 2024].
Jack , W. & Suri, T., 2014. Risk Sharing and Transactions Costs: Evidence from Kenya's Mobile Money Revolution. *American Economic Review* , 104(1), pp. 183-233.
Jedi, F. F., 2022. The Relationship between Financial Inclusion and Women's Empowerment: Evidence from Iraq. *Journal of Business and Management Studies* , 4(3), pp. 104-120.
Jenga, K., Catal, C. & Kar, G., 2023. Machine learning in crime prediction. *Journal of Ambient Intelligence and Humanised Computing,* Volume 14, pp. 2887-2913.
Levantesi, S. & Zacchia, G., 2021. Machine Learning and Financial Literacy: An Exploration of Factors Influencing Financial Knowledge in Italy. *Journal of Risk and Financial Management* , 14(3), pp. 1-12.
Lu, L., 2018. How a Little Ant Challenges Giant Banks? The Rise of Ant Financial (Alipay)'s Fintech Empire and Relevant Regulatory Concerns. *Electronic Banking* , 28(1), pp. 12-30.
Lum, K. & Isaac , W., 2016. To Predict and Serve?. *Royal Statistics Society,* 13(5), pp. 14-19.
Mabrouk, F., Bousrih, J., Elhaj, M. & Binuswadan, J., 2023. Empowering Women through Digital Financial Inclusion: Comparative Study before and after COVID-19. *Sustainability* , 15(12), pp. 1-17.
Meijer , A. & Wessels, M., 2019. Predictive Policing: Review of Benefits and Drawbacks. *International Journal of Public Administration* , 42(12), pp. 1031-1039.
Negre, P. et al., 2024. Literature Review of Deep-Learning-Based Detection of Violence in Video. *Sensors,* 24(12), pp. 1-29.
Nnanna, C. E. et al., 2024. Development of Random Forest Model for Stroke Prediction. *Internaional Journal of Innovative Science and Research Technology* , 9(4), pp. 2783-2795.
Nwaimo, C. S., Adegbola, A. E. & Adegbola, M. D., 2024. Predictive analytics for financial inclusion: Using machine learning to. *Computer Science & IT Research Journal,* 5(6), pp. 1358-1373.
Rahman , R. et al., 2023. A comparative study of machine learning algorithms for predicting domestic violence vulnerability in Liberian women. *BMC Women's Health,* 23(542), pp. 1-15.
RAND Corporation, 2013. *Safety and Justice Program.* [Online]
Available at:
https://www.rand.org/dam/rand/pubs/research_reports/RR200/RR233/RAND_RR233.sum.pdf
[Accessed 7 August 2024].
Robeson , S. M. & Willmott, C. J., 2023. Decomposition of the Mean Absolute Error (MAE) into Systematic and Unsystematic Components. *PLOS ONE,* 18(2), p. e0279774.

Rodriguez-Rodriguez, I. et al., 2020. Modeling and Forecasting Gender-Based Violence through Machine Learning Techniques. *Applied Sciences* , 10(22), pp. 1-16.
Saunders , J., Hunt, P. & Hollywood, J. S., 2016. Predictions put into practice: a quasi-experimental evaluation of Chicago's predictive policing pilot. *Journal of Experimental Criminology,* 12(3), pp. 347-371.
Siddiqui, A. T., 2024. Reviewing Women's Safety Using AI and IoT Devices. In: 1st, ed. *AI Tools and Applications for Women's Safety* . Hershey, Pennsylvania: IGI Global, pp. 250-261.
Smith, G. & Rushtagi, I., 2021. When Good Algorithms Go Sexist: Why and How to Advance AI Gender Equity. *Stanford Social Innovation Review,* 32 March , p. Online.
The World Bank Group , 2021. *The Global Findex Database 2021: Financial Inclusion, Digital Payments, and Resilience in the Age of COVID-19,* Washington DC: The World Bank Group .
Thyago, C. P. et al., 2019. A systematic literature review of machine learning methods applied to predictive maintenance. *Computers and Industrial Engineering* , Volume 137, pp. 1-10.
UN Women, 2019. *Progress of the World's Women 2019–2020: Families in a Changing World.* [Online]
Available at: https://www.unwomen.org/en/digital-library/publications/2019/06/progress-of-the-worlds-women-2019-2020
[Accessed 19 July 2024].
United Nations , 2020. *Ending Violence Against Women* , Geneva: The United Nations .
Vankov, E. R. & Gadhoumi, K., 2024. Supervised Learning. In: S. Chiang , V. Rao & M. Vannucci, eds. *Statistical Methods in Epilepsy.* 1st ed. Boca Raton: Chapman and Hall/CRC.
Varun, M., Elluri, L., Vyas, P. & Roy, N., 2023. Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions. *IEEE Access,* 1(1), p. 99.
Vinuesa, R. et al., 2020. The role of artificial intelligence in achieving the Sustainable Development Goals. *Nature Communications,* 11(233), pp. 1-11.
World Bank, 2018. *Financial Inclusion on the Rise, But Gaps Remain, Global Findex Database Shows.* [Online]
Available at: https://www.worldbank.org/en/news/press-release/2018/04/19/financial-inclusion-on-the-rise-but-gaps-remain-global-findex-database-shows
[Accessed 11 July 2024].
World Economic Forum, 2020. *World Economic Forum on India: From Deliberation to ransformaion* , Geneva, Switzerland : World Economic Forum.