cisdiJournal

Computing, Information Systems, Development Informatics & Allied Research Journal
Vol. 15 No. 3, 2024 - www.cisdijournal.net

Predictive System for Loan Approvals

Otaren Fredrick Enoma, Akpojotor Ogagaoghene Paul & Efevberha-Ogodo Oghenevbaire
Lecturer, Department of computer Science, Benson Idahosa University, Benin City Nigeria
Graduate Student, Department of Computer Science, Benson Idahosa University, Benin City Nigeria
Lecturer, Western Delta University Oghara, Delta State, Nigeria
E-mails: fotaren@biu.edu.ng, fredrick.otaren@physci.uniben.edu, akporjotorpaul@gmail.com, oe.ogodo@wdu.edu.ng
Phone: +2348109161345

ABSTRACT

Abstract: In Nigeria, access to credit is a crucial factor for economic development, particularly in promoting entrepreneurship and supporting small and medium-sized enterprises (SMEs). However, the loan approval process in Nigerian financial institutions remains highly subjective and often time-consuming, leading to inefficiencies and limited access to credit for eligible applicants. This paper proposes the development of a predictive system for loan approval tailored to the Nigerian context. By leveraging machine learning models and historical loan data from Nigerian banks, this system aims to predict loan approval outcomes based on key applicant features such as income, employment status, credit history, and financial obligations. The system utilizes a range of algorithms, including logistic regression, decision trees, and gradient boosting machines, to optimize predictive accuracy while ensuring fairness and transparency in decision-making. To address common challenges in the Nigerian financial sector, such as incomplete credit records and socioeconomic disparities, the model incorporates local contextual factors, such as informal income sources and regional economic conditions. The proposed predictive system offers several benefits, including faster loan processing times, reduced subjectivity in decision-making and improved access to credit for underserved populations. Additionally, the model can help financial institutions in Nigeria enhance their risk management strategies by identifying high-risk applicants more accurately.

Keywords: Predictive, Loan, Approval, Machine Learning, Model

CISDI Journal Reference Format
Otaren, F.E., Akpojotor, O.P. & Efevberha-Ogodo, O. (2024): Predictive System for Loan Approvals. Computing, Information Systems, Development Informatics & Allied Research Journal. Vol 15 No 3, Pp 29-44.
Available online at www.isteams.net/cisdijournal. dx.doi.org/10.22624/AIMS/CISDI/V15N3P3

1. BACKGROUND OF THE STUDY

The term loan refers to a type of credit vehicle in which a sum of money is lent to another party in exchange for future repayment of the value or principal amount. In many cases, the lender also adds interest or finance charges to the principal value, which the borrower must repay in addition to the principal balance. Loans may be for a specific, one-time amount, or they may be available as an open-ended line of credit up to a specified limit. Loans come in many different forms including secured, unsecured, commercial, and personal loans. Lenders will consider a prospective borrower's income, credit score, and debt levels before deciding to offer them a loan.

A loan may be secured by collateral, such as a mortgage, or it may be unsecured, such as a credit card. In some cases, the lender may require collateral to secure the loan and ensure repayment. Loans may also take the form of bonds and certificates of deposit (CDs). Based on the applicant's creditworthiness, the lender either denies or approves the application. The lender must provide a reason should the loan application be denied. If the application is approved, both parties sign a contract that outlines the details of the agreement. The lender advances the proceeds of the loan, after which the borrower must repay the amount including any additional charges, such as interest (Kagan, 2023).

Kenton (2023), Observed that the International Monetary Fund (IMF) is an international organization that promotes global economic growth and financial stability, encourages international trade, and reduces poverty. The IMF's mission is to promote global economic growth and financial stability, encourage international trade, and reduce poverty around the world. One of the IMF's most important functions is to make loans to countries that are experiencing economic distress to prevent or mitigate financial crises. The International Monetary Fund (IMF) is based in Washington, D.C. The organization is currently composed of 190 member countries, each of which has representation on the IMF's executive board in proportion to its financial importance. The IMF's primary methods for achieving these goals are monitoring capacity building and lending. The IMF collects massive amounts of data on national economies, international trade, and the global economy in aggregate.

The organization also provides regularly updated economic forecasts at the national and international levels. These forecasts, published in the World Economic Outlook, are accompanied by lengthy discussions on the effect of fiscal, monetary, and trade policies on growth prospects and financial stability. The IMF makes loans to countries that are experiencing economic distress to prevent or mitigate financial crises. Members contribute the funds for this lending to a pool based on a quota system. In 2019, loan resources in the amount of SDR 11.4 billion (SDR 0.4 billion above target) were secured to support the IMF’s concessional lending activities into the next decade. IMF funds are often conditional on recipients making reforms to increase their growth potential and financial stability. Structural adjustment programs, as these conditional loans are known, have attracted criticism for exacerbating poverty.

With great external debt comes great responsibility. Countries resort to foreign borrowing to maintain financial liquidity and stimulate growth. For rich nations facing a downturn, taking a loan at low interest rates can be more desirable the raising taxes. For emerging nations, this kind of financing is even more essential to cover for domestic resource gaps and pay for programs that can help reduce poverty and foster longer-term growth. There is, however, a well-known problem with debts: borrowing money is easier than paying it back. “Debt is like any other trap,” the 19th-century American author Josh Billings has said: “Easy enough to get into, but hard enough to get out of.” How hard? To the extent that there is no such thing as zero external debt, and both the most and the least-developed countries in the world alike (and all those in between) today struggle more than ever under the burden of what they owe. According to estimates of the Institute of International Finance (IIF), the Washington-based global association of the financial industry, overall international borrowing decreased for the first time in more than two years during the first quarter of the year—by $1.7 trillion to $289 trillion. Behind these figures lies a complex and troubling scenario: the fall was largely driven by developed markets (meaning those that can pay back debt more easily), where overall debt dropped $2.3 trillion to below $203 trillion; whereas the debt level across emerging economies surged to $86 trillion (+$0.6 trillion), a new record.

Also, despite the Q1 overall dip, total global debt was still up $30 trillion (12%) since the end 2019. Meanwhile, the global debt-to-GDP ratio has jumped too—to 360%. Simply put, the world borrows over three times more than it. Generally, loan prediction involves the lender looking at various background information about the applicant and deciding whether the bank should grant the loan. Parameters like credit score, loan amount, lifestyle, career, and assets are the deciding factors in getting the loan approved. If, in the past, people with parameters similar to yours have paid their dues timely, it is more likely that your loan would be granted as well. Machine learning algorithms can exploit this dependency on experiences and comparisons with other applicants and formulate a data science problem to predict the loan status of a new applicant using similar rules. Several collections of data from past loan applicants use different features to decide the loan status. A machine learning model can look at this data, which could be static or time-series, and give a probability estimate of whether this loan will be approved. Let's look at some of these datasets.

Traditional processes determine the risk by manually looking at the applicant's income, credit history, and several other dynamic parameters and creating a data-driven risk model. Despite using data science in this process, there is still a large amount of manual work involved. Researchers have recently explored the possibility of using deep learning in various aspects of this process. For example, credit score and credit history are essential parameters for assessing the applicant's lending risk. DL-based approaches such as Embedding Transactional Recurrent Neural Network (E.T.-RNN) compute the credit scores of applicants by looking at the history of their credit and debit card transactions. Such an approach eliminates the high dependency on manual intervention, extensive domain knowledge, and human bias in loan approval prediction (ProjectPro, 2023). Bhandari (2020), observed that loans are the core business of banks. The main profit comes directly from the loan's interest. The loan companies grant a loan after an intensive process of verification and validation. However, they still don't have assurance if the applicant is able to repay the loan with no difficulties.

1.2 Statement of the Problem

The traditional loan approval process in financial institutions is often marred by inefficiencies and subjectivity, leading to prolonged decision-making timelines and potential biases. Human-based assessments may not fully utilize available data, resulting in suboptimal lending decisions. The need to expedite and enhance the loan approval process while ensuring fairness and accuracy necessitates the application of data mining algorithms in predictive modeling.

1.3 Aim and objectives of the study

The aim of this study is to contribute to knowledge by utilizing data mining algorithms to predict the likelihood of loan approval based on historical data.

The objectives of this study are as follows:

a) Design a data mining application for loan approval prediction.
b) Implement a data mining application for loan approval prediction.
c) Test the data mining application for loan approval prediction.

1.4 Scope of the study

Nigeria's financial sector plays a crucial role in the country's economic development by providing access to credit for individuals and businesses. However, the loan approval process in Nigeria faces challenges such as inefficiencies, manual reviews, and potential biases. Developing a predictive loan approval system tailored to Nigeria's context can improve access to credit, enhance decision-making accuracy, and promote financial inclusion.

1.6. Significance of the study

The significance of the study on developing a predictive loan approval system in Nigeria is multifaceted and impactful:

1. Improved Access to Credit: By streamlining the loan approval process through predictive analytics, the study can enhance access to credit for individuals and businesses in Nigeria. This is particularly significant in a country where access to finance remains a challenge for many, especially those in underserved communities and rural areas.
2. 2.Enhanced Financial Inclusion: The study has the potential to promote financial inclusion by making credit more accessible to previously marginalized groups, including women, youths, and smallholder farmers. By leveraging data-driven approaches, financial institutions can better assess the creditworthiness of individuals who may have limited or no traditional credit history.
3. Reduced Loan Default Rates: A predictive loan approval system can help mitigate the risk of default by accurately assessing the repayment ability of borrowers. By leveraging machine learning algorithms to analyze historical data and predict future behavior, financial institutions can make more informed lending decisions, ultimately reducing the incidence of loan defaults.
4. Efficiency Gains for Financial Institutions: Implementing a predictive loan approval system can lead to significant efficiency gains for financial institutions in Nigeria. By automating the decision-making process and reducing the need for manual reviews, banks and other lenders can lower operational costs, accelerate loan processing times, and improve overall customer experience.
5. Mitigation of Biases and Discrimination: The study can contribute to mitigating biases and discrimination in the loan approval process. By using fairness-aware machine learning techniques and transparent decision-making algorithms, the predictive system can help ensure that loan decisions are based on objective criteria rather than subjective factors such as race, gender, or ethnicity.
6. Policy and Regulatory Implications: The findings of the study can inform policy and regulatory discussions surrounding lending practices and financial inclusion in Nigeria. By demonstrating the feasibility and benefits of predictive analytics in the lending industry, policymakers and regulators can explore ways to encourage its adoption while safeguarding consumer rights and data privacy.

2. RELATED WORKS

Viswanatha and Ramachandra (2023) observed that due to significant technology advancements, people's needs have expanded. As a result, there have been more requests for loan approval in the banking sector. A few qualities, taken for consideration, when choosing a candidate for loan approval in order to, determine loan's status. Banks face a major challenge; when it, comes to assessing loan applications and lowering the risks associated with potential borrower defaults. Since they must thoroughly evaluate each borrower's eligibility for a loan, banks find this process to be particularly challenging. This research proposes combining machine learning (ML) models and ensemble learning approaches to find the probability of accepting individual loan requests. This tactic can increase the accuracy with which qualified candidates are selected from a pool of applicants. As a result, this method can be used to address the problems with loan approval processes outlined above. Both the loan applicants and the bank employees profit from the strategy's dramatic reduction in sanctioning time.

As a result of the banking industry's expansion, more people were applying to loans at banks. In order to predict the accuracy of loan approval status for applied person, this used four different algorithms namely Random Forest, Naive Bayes, Decision Tree, and KNN. By using these, we obtained better accuracy of 83.73% with Naïve Bayes algorithm as best one.Das et al (2023), noted that in our daily life, it is difficult to meet financial demand while in crisis. This financial crisis may be solved with financial assistance from the banks. The financial assistance is nothing but availing loan from the bank with proper agreement to repay the amount including calculated interest within the loan approved tenure. The customer can only avail loans against the submission of some valid and important supportive documents. However, although the customer is aware of the whole process of repayment and installment along with loan approval tenure, most of the time it is hard to get the approved loan within a shorter period.

This study therefore automates this manual and long process by predicting the chance of approval of the loan by applying machine learning techniques and classification algorithms to predict loan eligibility through an automatic online loan application process On the issue of Credit using Machine Learning (ML) for credit risk prediction, Noriega et al (2023) raised the need for financial institutions to use AI and ML to assess credit risk, analyzing large volumes of information. This study posed research questions about algorithms, metrics, results, data sets, variables, and related limitations in predicting credit risk. The study utilized renowned databases to answer them and identified 52 relevant studies with the credit industry microfinance. Challenges and approaches in credit risk prediction using ML models we identified, difficulties with the implemented models such as the black box model, the need for explanatory artificial intelligence, the importance of selecting relevant features, addressing multicollinearity, and the problem of the imbalance in the input data.

By answering the questions, the identified that the Boosted Category is the most researched family of ML models; the most commonly used metrics for evaluation are Area Under Curve (AUC), Accuracy (ACC), Recall, precision measure F1 (F1), and Precision; Research mainly uses public data sets to compare models, and private ones to generate new knowledge when applied to the real world. The most significant limitation identified is the representativeness of reality, and the variables primarily used in the microcredit industry are related data to the demographic, the operation, and payment behavior. This study aims to guide the developers of credit risk management tools and software towards the existing offer of ML methods, metrics, and techniques used to forecast it, thereby minimizing possible losses due to default and guiding risk appetite. Kumari et al (2023) oberved that processes in numerous industries, including finance, real estate, security, and genomics, are being transformed by machine learning (ML) algorithms.

One of the major impediments in the banking sector is the loan approval process. Modern tools like ML models help accelerate, streamline, and increase the precision of loan approval procedures. It will benefit both the client and the bank in terms of time and manpower required for loan eligibility prediction. The entire work is cantered on a classification problem and is a form of supervised learning in which it is important to determine whether the loan will be approved or not. Also, it is a predictive modeling problem where a class label is predicted from the input data for a specific sample of input data. This study deployed various ML algorithms to identify the loan approval status and compare the performance of implemented models.

2.2 Emerging Trends and Innovations in Predictive Loan Approval Systems

In addition to traditional machine learning techniques, emerging trends and innovations in predictive loan approval systems offer novel approaches to addressing the unique challenges faced in Nigeria.

1. **Blockchain-Based Credit Scoring:** Blockchain technology has gained attention as a promising solution for improving credit scoring accuracy and transparency. Ogunleye et al. (2020) propose a blockchain-based credit scoring framework that utilizes immutable transaction records to assess creditworthiness, particularly for individuals with limited credit history. By leveraging decentralized ledger technology, this approach enhances data integrity and reduces the risk of fraud, thereby increasing lenders' confidence in extending credit to underserved populations.
2. **Explainable AI (XAI) in Lending:** Addressing concerns about algorithmic transparency and interpretability, explainable AI (XAI) techniques aim to provide insights into the decision-making process of predictive loan approval systems. A study by Adekunle and Adejimi (2021) explores the application of XAI methods, such as feature importance analysis and model-agnostic explanations, to enhance the interpretability of machine learning models in credit scoring. By generating explanations for loan approval decisions, XAI promotes trust and accountability in automated lending processes, facilitating regulatory compliance and consumer empowerment.
3. **Peer-to-Peer (P2P) Lending Platforms:** Peer-to-peer lending platforms have emerged as alternative sources of financing, connecting borrowers directly with individual investors. Adeleke et al. (2021) investigate the role of P2P lending in expanding access to credit for small businesses in Nigeria, particularly in sectors underserved by traditional financial institutions. By leveraging digital technologies and alternative credit assessment methods, P2P lending platforms offer faster loan processing times, lower transaction costs, and greater flexibility in loan terms, catering to the diverse needs of borrowers in the Nigerian market.
4. **Risk-Based Pricing Strategies:** Risk-based pricing strategies involve assessing individual borrowers' risk profiles and offering personalized loan terms based on their creditworthiness. A study by Onwuka et al. (2022) examines the implementation of risk-based pricing models by Nigerian banks to optimize loan pricing and mitigate credit risk. By segmenting borrowers into risk categories and adjusting interest rates accordingly, financial institutions can better align loan pricing with the level of risk associated with each borrower, thereby maximizing profitability while maintaining prudent lending practices.

2.3 Analysis of Existing system

The existing systems for predictive loan approval utilize a combination of data analysis, statistical methods, and machine learning algorithms to assess the risk associated with granting loans. These systems aim to streamline the loan approval process by automating decision-making, thereby reducing the time and effort required for manual evaluations.

2.3.1 Existing loan approval systems generally consist of the following components:

- **Data Collection:** Gathering relevant data from various sources such as credit bureaus, banking records, and applicant-provided information.
- **Data Preprocessing:** Cleaning and transforming the collected data to prepare it for analysis.
- **Feature Engineering:** Creating new features or selecting relevant features from the existing data to improve the model's performance.
- **Model Training:** Using statistical and machine learning algorithms to build predictive models.

- **Model Evaluation**: Assessing the performance of the models using various metrics and validating their effectiveness.
- **Decision Making**: Integrating the trained models into the loan approval workflow to make automated or assisted decisions.

There are a few well-known existing systems and platforms used for predictive loan approval. Prominent among them are;

2.4 FICO® Origination Manager:

Loan origination systems (LOS) are essential for automating and streamlining the loan application, approval, and disbursement process. One prominent example is **FICO Origination Manager**, a widely used platform for managing loan applications from submission to decision. This system, like other existing loan origination platforms, leverages data-driven decision-making, risk assessment, and compliance tools to help financial institutions manage loans efficiently. Here's an in-depth analysis of such systems with a focus on the Nigerian financial landscape. FICO® Origination Manager is a widely used system that integrates predictive analytics into the loan origination process, facilitating automated credit decisions. This system by FICO integrates predictive analytics with loan origination to automate credit decisions and streamline the loan approval process while it offers significant benefits.

2.4.1 Key Features of FICO Origination Manager

FICO Origination Manager provides a range of functionalities that help banks and other financial institutions assess and process loan applications:

- **Automated Decisioning**: The system uses predictive models to assess the likelihood of loan repayment and recommend approvals or denials based on historical data and credit scoring systems like FICO.
- **Risk Management**: It integrates credit scoring and risk assessment models to analyze borrowers' creditworthiness, helping lenders reduce exposure to high-risk customers.
- **Compliance and Regulatory Management**: FICO Origination Manager ensures that loan decisions comply with regulations such as consumer protection laws, anti-money laundering (AML) standards, and credit risk guidelines.
- **Scalability**: The platform is scalable and can handle large volumes of applications, making it suitable for institutions with diverse portfolios.
- **Customization**: Lenders can customize the decision models and workflows to suit their specific business rules and strategies.
- **Real-Time Analytics**: Provides real-time insights and analytics into application trends, risk patterns, and loan portfolio performance, enabling data-driven decision-making.
- 2. **Challenges and Limitations in the Nigerian Context**
- While FICO Origination Manager and similar systems are robust and effective in many markets, they face unique challenges when applied in countries like Nigeria. The Nigerian financial ecosystem presents several barriers to the direct application of such systems:
- **Incomplete Credit Information**: Credit scoring systems like FICO rely heavily on comprehensive credit data, which may not be widely available in Nigeria. The Nigerian credit reporting system is still maturing, and a significant portion of the population lacks formal credit histories, particularly in rural areas. This limits the system's ability to make accurate decisions.

- Socioeconomic Disparities: Nigeria has a large informal economy, where many individuals and SMEs operate without formal financial documentation. Traditional loan origination systems like FICO may fail to account for informal income sources or non-conventional employment arrangements, which are common in Nigeria.
- Limited Financial Inclusion: A large percentage of Nigeria's population remains unbanked or underbanked. Existing origination systems may not adequately capture the creditworthiness of these individuals due to limited access to conventional financial services.
- Infrastructure Gaps: Many automated systems, including FICO Origination Manager, rely on stable internet access, strong IT infrastructure, and digital records. However, Nigeria's infrastructure, especially in rural areas, may pose challenges to the consistent operation of such systems.
- Regulatory Compliance: While FICO Origination Manager is built to adhere to international regulations, Nigerian financial regulations have specific nuances, such as compliance with Central Bank of Nigeria (CBN) guidelines, local anti-fraud frameworks, and financial inclusion policies. Customizing an international system to meet local regulations can be resource-intensive.

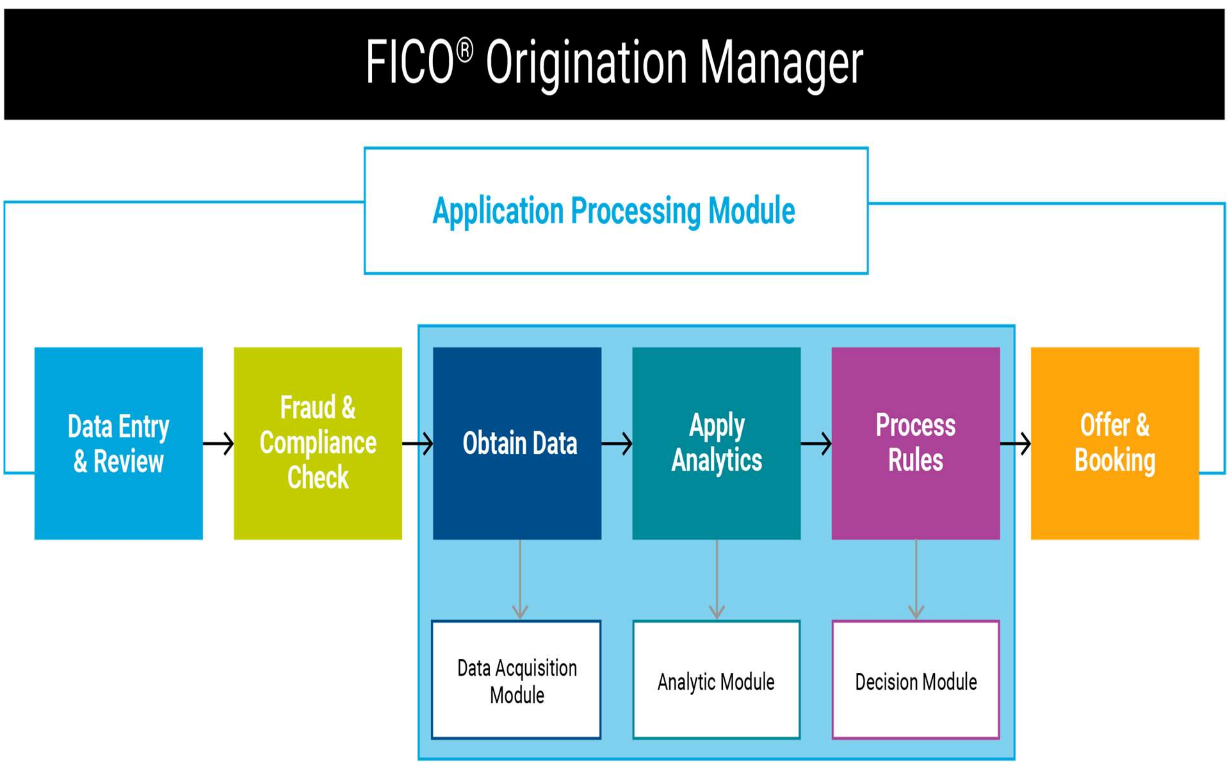

Fig 1: Architectural diagram of FICO® Origination Manager

3. RESEARCH METHODOLOGY

3.1 Research Design

The loan approval process is critical for financial institutions to manage risk and ensure profitability. Predictive models can automate this process, improving efficiency and decision accuracy. The overall activities constituting the study are represented as follows:

Quantitative research involving:

- Data collection
- Preprocessing
- Feature Selection
- Model Development
- Model Evaluation
- Model Interpretation
- Deployment and
- Maintenance.

3.2 Data Collection

Data collection is a critical phase in the development of a predictive system for loan approval. It involves gathering relevant information that can be used to train machine learning models, enabling accurate and reliable predictions. This section outlines the strategies, sources, and steps involved in the data collection process.

3.3 Data Preprocessing

Data preprocessing is a crucial step in the development of a predictive system for loan approval. It involves transforming raw data into a clean and usable format, ensuring that the machine learning models can learn effectively from it. The main reasons for data preprocessing are:

- Clean and prepare the data for analysis and modeling.
- To handle missing values, outliers, and inconsistencies in the data.
- Encode categorical variables and scale numerical features.
- To ensure the data is suitable for training machine learning models.

3.4 Feature Selection

Feature selection is a crucial step in developing a predictive model for loan approval. It involves identifying and selecting the most relevant features from the dataset that contribute significantly to the prediction task. Effective feature selection improves model performance, reduces overfitting, and enhances interpretability. Reasons for feature selection;

- To improve the accuracy of the predictive model.
- Reduce the computational cost by minimizing the number of features.
- To enhance model interpretability by focusing on the most important features.
- Mitigate the curse of dimensionality and overfitting

3.5 The proposed system

The proposed loan approval system leveraging machine learning (ML) algorithms such as **Random Forest, Decision Tree**, and **Logistic Regression.** The system is designed to enhance the efficiency and accuracy of loan approval decisions in Nigerian financial institutions by analyzing applicant data and

predicting the likelihood of loan repayment.
This system uses historical loan data to predict whether a loan application should be approved or rejected. It applies various ML models that can automatically learn patterns in the data to make data-driven decisions. Three models—**Logistic Regression**, **Decision Tree**, and **Random Forest**—are utilized, offering a robust combination of explainability, interpretability, and predictive power. The system is trained on past loan data to understand the relationships between applicant features (e.g., income, credit score) and loan approval outcomes.

3.6 Key Algorithms

a) **Logistic Regression:** Logistic regression is used to model the probability of a binary outcome—in this case, whether a loan will be approved or denied. It is highly interpretable, making it suitable for understanding the influence of each feature on the decision-making process. The algorithm models the relationship between the dependent variable (loan approval) and one or more independent variables (income, credit score, etc.) using a logistic function. The output is a probability score between 0 and 1, which can be thresholded to predict approval or rejection.

Advantages:

- Simple and easy to implement.
- Provides clear insights into feature importance.
- Well-suited for binary classification tasks.

b) **Decision Tree:** A decision tree is used to model decisions in a tree-like structure, where each internal node represents a decision based on a feature, and each leaf node represents an outcome (loan approval or denial). It's highly interpretable and easy to visualize. The decision tree algorithm splits the dataset into smaller subsets based on feature values. At each node, the algorithm selects the feature that results in the largest information gain (or lowest impurity) and recursively creates branches until the tree reaches the terminal nodes (outcomes).

Advantages:

- Handles both numerical and categorical data well.
- Interpretable, as the decision process can be visualized and understood by non-experts.
- Captures non-linear relationships between features.

c) **Random Forest:** Random Forest is an ensemble method that builds multiple decision trees and aggregates their predictions to enhance accuracy and reduce overfitting. The algorithm builds a "forest" of decision trees on random subsets of the training data and features. Each tree produces an individual prediction, and the final output is determined by majority voting (classification) or averaging (regression) across all trees.

Advantages:

- Reduces overfitting compared to a single decision tree by averaging the results of many trees.
- Handles both categorical and numerical data.
- More accurate and robust than decision trees alone.

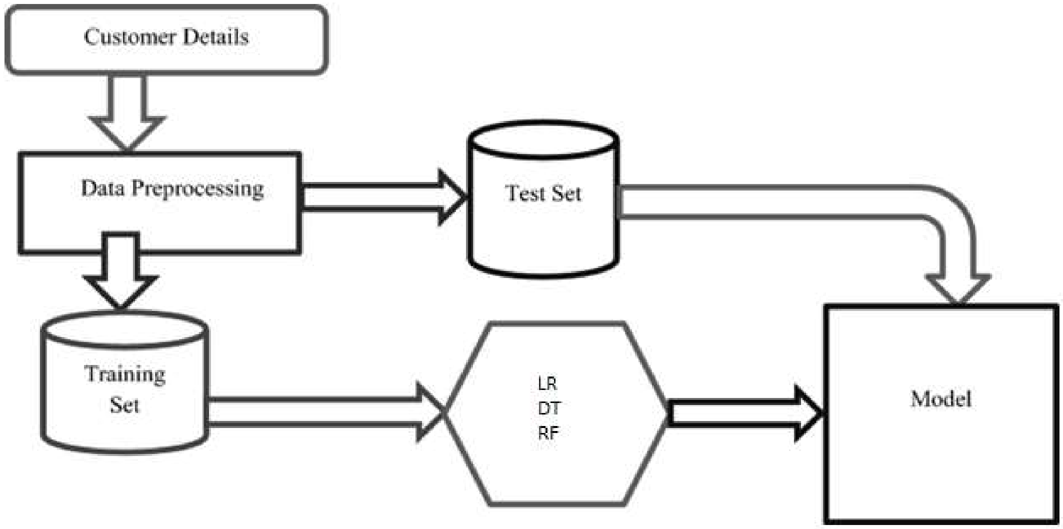

Fig.2 Architecture of the Propose System

The proposed loan approval predictive system architecture integrates logistic regression, decision tree, and random forest models to create a robust, efficient, and fair decision-making process. By addressing challenges related to data quality, model interpretability, bias, scalability, and regulatory compliance, the system aims to enhance operational efficiency while ensuring accurate and fair loan approval decisions. Continuous monitoring and improvement will ensure the system remains effective and relevant in dynamic market conditions. In the proposed system in figure2, we introduced Logistic regression as one of the classification algorithm for loan approval prediction along with Decision Tree and Random forest. We used an updated dataset obtain from a public repository (kaggle.com). The purpose of using the updated dataset is to bring about improvement in the training and prediction of loan approval prediction.

3.7 Addressing Challenges in the Nigerian Context

The proposed system takes into account challenges specific to Nigeria:

- Handling Incomplete Credit Data: The system can be designed to include alternative data sources such as mobile payment history, utility payments, or transaction history with microfinance institutions to assess creditworthiness.
- Informal Employment and Income: Feature engineering can incorporate variables that account for informal sector income, which is common in Nigeria.
- Bias and Fairness: The model will be tested for bias based on factors like gender, ethnicity, or geographic region, ensuring that the system adheres to fairness standards and local regulations.

Expected Benefits

- **Improved Efficiency**: Automating the loan approval process with machine learning will significantly reduce the time and resources spent on manual reviews.
- **Better Risk Management**: ML algorithms, particularly Random Forest, can enhance risk assessment by identifying high-risk applicants more accurately.
- **Increased Financial Inclusion**: By integrating alternative data and accommodating informal sector incomes, the system will help banks approve more loans for underserved populations in Nigeria.
- **Regulatory Compliance**: The system will be designed to comply with the Central Bank of Nigeria (CBN) lending regulations, ensuring fairness and transparency in decision-making.

User Documentation

The following steps enables the user to test run the application.
Step i. Download and install anaconda3.
Step ii. Code the application.
Step iii. Run the application.

4. RESULTS AND DISCUSSION

Decision Tree Prediction

Training Data Set Accuracy	:	1.0
Training Data F1 Score	:	1.0
Validation Mean F1 Score	:	0.6574515951536865
Validation Mean Accuracy	:	0.6965574108431252

Decision Tree classification

Test Accuracy	:	0.8536585365853658
Test F1 Score	:	0.903225806451613

Random Forest Classification

Train F1 Score	:	0.8699080157687253
Train Accuracy	:	0.7983706720977597
Validation Mean F1 Score	:	0.7021300922943281
Validation Mean Accuracy	:	0.7983714698000413
Test Accuracy	:	0.8536585365853658
Test F1 Score	:	0.903225806451613

Logistic Regression Algorithm

Test Accuracy	:	0.8617886178861789
Test F1 Score	:	0.9081081081081082

4.1 Evaluation Metrics

There are a number of metrics which can be used to evaluate a binary classification model, and accuracy is one of the simplest to understand. Accuracy is defined as simply the number of correctly categorized examples divided by the total number of examples. Accuracy can be useful but does not take into account the subtleties of class imbalances, or differing costs of false negatives and false positives.

The F1-score is useful:
Where there are either differing costs of false positives or false negatives, such as in the mammogram example or where there is a large class imbalance, such as if 10% of apples on trees tend to be unripe. In this case the accuracy would be misleading, since a classifier that classifies all apples as ripe would automatically get 90% accuracy but would be useless for real-life applications.
The accuracy has the advantage that it is very easily interpretable, but the disadvantage that it is not robust when the data is unevenly distributed, or where there is a higher cost associated with a particular type of error.

Machine Learning algorithm	Accuracy	f1-score
Logistic Regression	86%	90%
Random Forest	85%	90%
Decision Tree	85%	90%

The result of test run shows that Logistic regression has the highest evaluation as accuracy is 86% and f1-score is 90%. Random forest and decision tree have each 85% accuracy and 90% accuracy respectively.

5. SUMMARY

The proposed loan approval system aims to leverage machine learning algorithms—**Logistic Regression**, **Decision Tree**, and **Random Forest**—to enhance the loan decision-making process in Nigerian financial institutions. By utilizing historical loan data, these models are trained to predict whether a loan application should be approved or denied based on features like income, employment status, credit history, and loan amount. The system automates and streamlines the traditionally manual loan approval process, offering faster, more accurate, and data-driven decisions.

Each algorithm brings unique strengths to the system:

- **Logistic Regression** provides simplicity and interpretability, making it easier to understand how features affect loan outcomes.
- **Decision Trees** offer clear visual representations of decisions but are prone to overfitting.
- **Random Forest**, being an ensemble model, reduces overfitting and improves predictive accuracy by combining the output of multiple decision trees.

The system also addresses challenges specific to Nigeria, including incomplete credit histories, the prevalence of informal employment, and the need for alternative data sources. Additionally, it ensures compliance with local financial regulations and fairness in decision-making, promoting broader financial inclusion.

6. CONCLUSION

The implementation of this predictive loan approval system will significantly improve the efficiency, accuracy, and fairness of loan processing in Nigerian banks and microfinance institutions. By automating the loan evaluation process and leveraging machine learning techniques, the system can reduce the manual workload, minimize errors, and ensure better risk management for lenders. Logistic Regression offers transparency in decision-making, Decision Trees provide ease of interpretation, and Random Forest delivers the highest predictive power by mitigating overfitting. Together, these algorithms form a well-rounded solution that can handle both the complexities of loan applications and the socioeconomic challenges of the Nigerian financial landscape. This system is particularly valuable for enhancing financial inclusion by supporting underserved populations, such as those in the informal sector, through the incorporation of alternative credit data. As Nigerian financial institutions continue to grow, such a predictive system is essential for meeting increasing demand while managing credit risk effectively.

7. RECOMMENDATIONS

1. Adapt the System for Local Conditions: To maximize the effectiveness of the system in Nigeria, it's essential to integrate alternative data sources, such as mobile payment histories, utility payments, and community-based financial activities. This will help assess the creditworthiness of individuals with incomplete credit histories.
2. Monitor and Improve Model Fairness: Financial institutions should ensure that the system remains unbiased in its decisions. Regular fairness audits should be conducted to detect and mitigate any bias related to sensitive attributes such as gender, ethnicity, or geographic region.
3. Continuous Model Updates: The system should be periodically updated with new loan data to reflect changing market conditions and borrower behaviors. Retraining the models with fresh data ensures that the system remains accurate and relevant over time.
4. Regulatory Alignment: Nigerian financial institutions should ensure that the system adheres to the Central Bank of Nigeria (CBN) guidelines on consumer protection and anti-discrimination, aligning the loan approval process with local laws and international best practices.
5. Expand Usage to Microfinance Institutions: The system can be scaled to benefit microfinance institutions, which play a crucial role in providing loans to Nigeria's low-income and underserved populations. This can help enhance financial inclusion efforts and improve credit access for small businesses.

By implementing these recommendations, the predictive loan approval system can serve as a powerful tool for Nigerian financial institutions, driving growth, improving customer satisfaction, and reducing risk in loan portfolios.

REFERENCES

1. Chaudhary, B., Sivakami, R., Sathiyabhama, B. and Raghava, S. C. (2023).Intelligent Lender – Prediction of Loan Applicant Credibility Based on the Machine Learning Model.
2. Dansana, D., Patro, G., K., Mishra, B. K., Prasad, V., Razak, A., Wodajo, A. W.(2023). Analyzing the impact of loan features on bank loan prediction using Random Forest algorithm.
3. Das, A., Agarwal, A. and Das, R. R. (2022). Machine Learning Techniques-Based Banking Loan Eligibility Prediction.
4. Gaspar, V., Medas, P. and Perrelli, R. (2021) . Global Debt Reaches a Record $226 Trillion
5. Policymakers must strike the right balance in the face of high debt and rising inflation.
6. GFMag (2021). Countries With The Most External Debt 2021.external-debt-in-countries-around-the-world/
7. IMF Annual Report (2022). Crisis Upon Crisis. https://www.imf.org/external/pubs/ft/ar/2022/
8. Kagan, J. (2023)What Is a Loan, How Does It Work, Types, and Tips on Getting One.
9. Kenton, W. (2023). What Is the International Monetary Fund (IMF)?
10. Kumari, S., Swapnesh, D., Swarnkar, T., Swapnesh, D. and Kumar Naya, K. (2023).LOAN ELIGIBILITY Prediction Using Machine Learning: A Comparative Approach.
11. CNoriega, J., Rivera, L., and José Herrera, J. (2023). Machine Learning for Credit Risk Prediction: A Systematic Literature Review.
12. T., Berhane, T., Mohammed, A. and Assaye Walelgn, A. (2023). Credit-Risk Prediction Model Using Hybrid Deep—Machine-Learning Based Algorithms.
13. Nwafor, C. N. and Nwafor, Z. C. (2023).Determinants of non-performing loans: Anexplainable ensemble and deep neural network approach.
14. Uddin, N.,Ahamed, K. U., Ray, S. (2024). Top 10 Machine Learning Algorithms to Use in 2024. Analytics
15. Vidhya.Said, I, and Qu, Y. (2023). Improving the Performance of Loan Risk Prediction based on Machine Learning via Applying Deep Neural Networks.
16. Saini, S., Bhatnagar, A. and Rani, L. (2023). Loan Approval Prediction using Machine Learning: A Comparative Analysis of Classification Algorithms,3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater
17. Noida, India, Shankar, S. (2016). Interesting things about History of Lending.
18. Uddin, A., Islam, M., Talukder, A. and Aryal, S. (2023). An ensemble machine learning based bank loan approval predictions system with a smart application.

19. Uwais, A. A. and Khaleghzadeh, H. (2022). Loan default prediction using spark machine learning algorithms. https://researchportal.port.ac. uk/en/publications/ loan-default-prediction-using-spark-machine-learning-algorithms
20. Viswanatha V. and Ramachandra A. C (2023). Prediction of Loan Approval in Banks using Machine Learning Approach.
21. Yan, X. (2023). Research on Financial Field Integrating Artificial Intelligence: Application Basis, Case Analysis, and SVR Model-Based Overnight.
22. Zhang, Q. and Zhang, E. (2023). Loan Risk Prediction based on Random Forest Model.