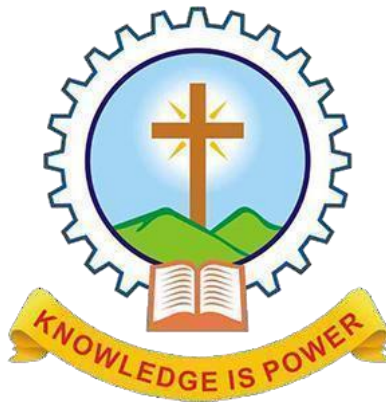


MAR ATHANASIUS COLLEGE OF ENGINEERING

KOTHAMANGALAM

(Affiliated to APJ Abdul Kalam Technological University, TVM)



Department of Computer Applications

Initial Project Report

On

**Credit Management System Integrated With Machine
Learning**

Submitted By:

Delwin Kuriakose

MAC23MCA-2023

Faculty Guide:

Prof. Manu John

Associate Professor

MCA Dept, MACE

Project Coordinator:

Prof. Sonia Abraham

Assistant Professor

MCA Dept, MACE

1. Introduction

The Credit Management System is an advanced platform designed to streamline the process of credit assessment, loan approval, and customer credit management for financial institutions. This system uses cutting-edge machine learning techniques, specifically XGBoost, to evaluate the eligibility of applicants based on various factors such as CIBIL score, income, loan amount, employment status, loan term, and other financial parameters. The system aims to provide quick, accurate, and data-driven decisions while reducing the manual effort involved in the traditional loan approval process.

The Credit Management System operates with two main user roles: customer and administrator. Customers can input their financial details into the system and receive personalized loan eligibility assessments based on their financial profile. The system evaluates the customer's creditworthiness using a trained XGBoost model, which helps determine the likelihood of loan approval and repayment capacity. It provides customers with feedback on their loan application status in real-time, offering clear insights into whether they are eligible for a loan and under what terms.

For administrators, the system provides full control over the management of loan applications and customer accounts. Administrators can view and review loan applications, assess risks, and approve or reject loan requests based on the eligibility prediction made by the system. The platform also includes tools for managing loan terms, loan limits, and repayment schedules. Furthermore, administrators can monitor and analyze trends in credit score data, loan balances, repayment histories, and other financial indicators through a real-time analytics dashboard. This ensures better decision-making and allows for timely interventions in case of rising risks or trends.

The system integrates a machine learning model using XGBoost, trained on historical loan data, to predict loan eligibility and assess the risk associated with applicants. By leveraging customer financial data, the system not only evaluates whether an applicant qualifies for a loan but also predicts the likelihood of repayment, enabling lenders to make informed decisions.

Additionally, the Credit Management System ensures secure management of sensitive customer data, with real-time data processing to keep the platform up-to-date. The system enables quick credit decision-making, reducing manual intervention and improving overall efficiency.

In terms of technology, the system uses a combination of HTML, CSS, and JavaScript for the front end to provide an intuitive and responsive user interface for both customers and administrators. On the back end, Python and Django are employed to handle the logic, data processing, and integration with the predictive model. The XGBoost library is used for building the machine learning models,

while Pandas and NumPy help with data manipulation and analysis. Customer data and loan information are securely stored in mySQL database

With its seamless integration of machine learning models, real-time data processing, and secure management of financial data, the Credit Management System offers a powerful solution for financial institutions looking to improve their loan approval processes and minimize credit risk. It provides a highly efficient, data-driven approach to credit assessment that benefits both customers and lenders, ensuring faster, smarter, and more reliable loan decisions.

1. Actors and Their Roles

1. Customer

The primary user of the system who interacts with the platform for loan and credit card services.

- **Registration & Login:** Create an account or log into the platform to access features.
- **Loan Application:** Submit loan applications by providing necessary details such as income, employment status, and loan amount.
- **Credit Card Application:** Apply for credit cards with different features and benefits tailored to their preferences.
- **Eligibility Prediction:** View machine learning-based predictions for loan or credit card eligibility.
- **Document Upload:** Upload required documents securely (e.g., ID proofs, income statements).
- **Track Application Status:** Monitor the progress of their applications in real time.
- **Profile Management:** Create and manage profiles to store personal and financial information for future applications.
- **Feedback and Ratings:** Provide feedback on the application process or rate the service, helping improve the system.
- **Receive pop-up Notifications:** Application status updates, Payment reminders

2. Administrator

Oversees the backend operations of the system to ensure smooth functionality, security, and efficiency.

- **User Management:** Approve and manage customer accounts.
- **Application Review:** Review flagged loan or credit card applications and make final approval or rejection decisions.
- **Document Verification:** Validate the uploaded documents for authenticity.

- **Loan and Credit Card Product Management:** Define and update loan terms, interest rates, and credit card benefits.
- **Machine Learning Management:** Monitor the performance of machine learning models for eligibility predictions and risk assessments.
- **Data Security:** Ensure the secure storage of sensitive customer data with encryption and compliance with privacy laws.
- **System Maintenance:** Maintain databases containing customer profiles, application statuses, and system logs.
- **Reporting and Analytics:** Generate reports on system performance, user activity, and financial trends.
- **Feedback Monitoring:** Review customer feedback and make adjustments to improve the user experience.

3. Description of Modules

1. User Management Module

- **Customer Registration:** Allows customers to create an account by providing necessary personal and financial details.
- **Login and Authentication:** Secure login and access management for customers and admins.
- **Profile Management:** Enables customers to update their profiles, contact details, and reset their passwords.

2. Loan Application Module

- **Loan Application Form:** Customers submit loan applications by providing required documents and financial data.
- **Loan Application Status:** Customers can track the status of their loan applications (approved, rejected, under review).

3. Loan Eligibility Prediction Module (ML Integration)

- **Loan Eligibility Prediction:** Uses machine learning models to assess and predict whether a customer is eligible for a loan based on input data such as income, credit score, employment status, etc.
- **Feature Engineering and Model Training:** Admins can train and refine models for loan eligibility prediction based on historical loan data.

- **Result Display:** Shows the eligibility prediction result for the customer (approved, rejected, conditions).

4. Loan & Credit Management Module

- **Loan Disbursement:** Once approved, loans are disbursed to the customers.
- **Credit Limit Management:** Admins assign credit limits to customers, based on their financial profile and eligibility.
- **Loan Repayment Schedules:** Defines repayment plans, including due dates, installment amounts, and payment methods.

5. Payment Management Module

- **Online Payment:** Customers can make loan repayments through integrated online payment systems.
- **Payment History:** Customers can view a history of all their payments made and any outstanding dues.
- **Overdue Payments:** Tracks overdue payments and issues reminders.

6. Admin Dashboard Module

- **Loan Application Review:** Admins can review loan applications, manually override eligibility predictions if needed, and approve or reject applications.
- **Customer Overview:** Provides an overview of customer loans, credit limits, repayment status, etc.
- **Analytics and Reporting:** Generates reports on loan approval rates, payment status, and predictive model performance.

7. Credit Monitoring & Limits Module

- **Credit Score Tracking:** Allows customers to view and monitor their credit scores and how it affects loan eligibility.
- **Credit Limit Management:** Admins can adjust customers' credit limits based on payment history and financial behaviour.

4. Business Rules

1. Eligibility Criteria: Customers must meet basic eligibility criteria (e.g., minimum credit score, income level) to apply for loans or credit cards.
2. Loan Approval: Loan applications are approved based on customer creditworthiness, income, and other predefined factors.
3. Interest Rates: The interest rate applied to loans is based on the customer's credit history and loan amount.
4. Repayment Terms: Loan repayment schedules, including due dates and amounts, are set based on the loan type and customer profile.
5. Credit Limits: Credit limits are assigned based on the customer's credit score and payment history, with periodic reviews for increases or decreases.
6. Late Fees: Late payments incur penalties, and repeated late payments may lead to default status.
7. Loan Disbursement: Once a loan is approved, the funds are disbursed to the customer's account or via other channels.
8. Credit Score Updates: Customer credit scores are updated regularly based on their repayment behavior..
9. Fraud Prevention: Applications flagged for fraudulent activity are reviewed and may be rejected.
10. Account Monitoring: Regular monitoring of accounts helps detect and prevent overdue payments, defaults, and financial distress.
11. Debt Collection: If a loan defaults, collection procedures are initiated, including payment reminders and legal actions if necessary.
12. Data Privacy: Customer data is encrypted and stored securely, in compliance with applicable data protection regulations.
13. Loan Restructuring: Customers facing financial hardship may apply to modify the loan terms to make repayment more manageable.

5. Dataset Exploration

Loan Approval Prediction

The dataset is taken from the Kaggle repository. The dataset contains 4269 sample observations and has 13 columns including 1 identifier, 1 class variable. The dataset contains numeric values and Boolean values. There are no missing values in the dataset.

The identifier is the Loan_id. The features are No of dependents, education, Self-employed, Income annum, Loan Amount, Loan Term, Cibil Score, Residential assets value, Commercial assets value, Luxury_assets_value, Bank asset value. The class label consists of Loan Status.

There are two class labels which includes Approved or Rejected which indicates whether the person is eligible or not to receive the loan

| loan_id | no_of_de | education | self_empl | income_a | loan_amc | loan_term | cibil_score | residential | commercial | luxury_as | bank_asset | loan_status |
|---------|----------|-----------|-----------|----------|----------|-----------|-------------|-------------|------------|-----------|------------|-------------|
| 1 | 2 | Graduate | No | 9600000 | 29900000 | 12 | 778 | 2400000 | 17600000 | 22700000 | 8000000 | Approved |
| 2 | 0 | Not Gradu | Yes | 4100000 | 12200000 | 8 | 417 | 2700000 | 2200000 | 8800000 | 3300000 | Rejected |
| 3 | 3 | Graduate | No | 9100000 | 29700000 | 20 | 506 | 7100000 | 4500000 | 33300000 | 12800000 | Rejected |
| 4 | 3 | Graduate | No | 8200000 | 30700000 | 8 | 467 | 18200000 | 3300000 | 23300000 | 7900000 | Rejected |
| 5 | 5 | Not Gradu | Yes | 9800000 | 24200000 | 20 | 382 | 12400000 | 8200000 | 29400000 | 5000000 | Rejected |
| 6 | 0 | Graduate | Yes | 4800000 | 13500000 | 10 | 319 | 6800000 | 8300000 | 13700000 | 5100000 | Rejected |
| 7 | 5 | Graduate | No | 8700000 | 33000000 | 4 | 678 | 22500000 | 14800000 | 29200000 | 4300000 | Approved |
| 8 | 2 | Graduate | Yes | 5700000 | 15000000 | 20 | 382 | 13200000 | 5700000 | 11800000 | 6000000 | Rejected |
| 9 | 0 | Graduate | Yes | 800000 | 2200000 | 20 | 782 | 1300000 | 800000 | 2800000 | 600000 | Approved |
| 10 | 5 | Not Gradu | No | 1100000 | 4300000 | 10 | 388 | 3200000 | 1400000 | 3300000 | 1600000 | Rejected |
| 11 | 4 | Graduate | Yes | 2900000 | 11200000 | 2 | 547 | 8100000 | 4700000 | 9500000 | 3100000 | Approved |
| 12 | 2 | Not Gradu | Yes | 6700000 | 22700000 | 18 | 538 | 15300000 | 5800000 | 20400000 | 6400000 | Rejected |
| 13 | 3 | Not Gradu | Yes | 5000000 | 11600000 | 16 | 311 | 6400000 | 9600000 | 14600000 | 4300000 | Rejected |
| 14 | 2 | Graduate | Yes | 9100000 | 31500000 | 14 | 679 | 10800000 | 16600000 | 20900000 | 5000000 | Approved |
| 15 | 1 | Not Gradu | No | 1900000 | 7400000 | 6 | 469 | 1900000 | 1200000 | 5900000 | 1900000 | Rejected |
| 16 | 5 | Not Gradu | No | 4700000 | 10700000 | 10 | 794 | 5700000 | 3900000 | 16400000 | 4400000 | Approved |
| 17 | 2 | Graduate | Yes | 500000 | 1600000 | 4 | 663 | 1300000 | 100000 | 1300000 | 700000 | Approved |
| 18 | 4 | Not Gradu | Yes | 2900000 | 9400000 | 14 | 780 | 2900000 | 2800000 | 6700000 | 4300000 | Approved |
| 19 | 2 | Graduate | No | 2700000 | 10300000 | 10 | 736 | 1000000 | 0 | 6200000 | 3300000 | Approved |
| 20 | 5 | Graduate | No | 6300000 | 14600000 | 12 | 652 | 10300000 | 3500000 | 23500000 | 5900000 | Approved |
| 21 | 2 | Graduate | No | 5000000 | 19400000 | 12 | 315 | 9500000 | 1600000 | 18000000 | 6100000 | Rejected |
| 22 | 4 | Graduate | No | 5800000 | 14000000 | 16 | 530 | 3800000 | 11300000 | 22200000 | 5400000 | Rejected |
| 23 | 4 | Graduate | Yes | 6500000 | 25700000 | 18 | 311 | 13100000 | 1700000 | 19500000 | 8500000 | Rejected |
| 24 | 0 | Not Gradu | Yes | 500000 | 1400000 | 2 | 551 | 900000 | 600000 | 1100000 | 300000 | Approved |
| 25 | 0 | Not Gradu | No | 4900000 | 9800000 | 16 | 324 | 3800000 | 8700000 | 10000000 | 3300000 | Rejected |
| 26 | 5 | Not Gradu | No | 3100000 | 9500000 | 20 | 514 | 7900000 | 3100000 | 6600000 | 2600000 | Rejected |

Figure 5.1 Snapshot of loan_approval_dataset.csv



```
df.shape
```

```
(4269, 13)
```

Figure 5.2 Shape of loan_approval_dataset.csv

The info () method provides a concise summary of a Data Frame, including the number of non-null values, the data type of each column, and the memory usage of the Data Frame. It also lists the total number of columns, and the number of entries in each column. Dataset has 4269 rows and 13 columns

| # | Column | Non-Null Count | Dtype |
|----|--------------------------|----------------|--------|
| 0 | loan_id | 4269 non-null | int64 |
| 1 | no_of_dependents | 4269 non-null | int64 |
| 2 | education | 4269 non-null | object |
| 3 | self_employed | 4269 non-null | object |
| 4 | income_annum | 4269 non-null | int64 |
| 5 | loan_amount | 4269 non-null | int64 |
| 6 | loan_term | 4269 non-null | int64 |
| 7 | cibil_score | 4269 non-null | int64 |
| 8 | residential_assets_value | 4269 non-null | int64 |
| 9 | commercial_assets_value | 4269 non-null | int64 |
| 10 | luxury_assets_value | 4269 non-null | int64 |
| 11 | bank_asset_value | 4269 non-null | int64 |
| 12 | loan_status | 4269 non-null | object |

dtypes: int64(10), object(3)
memory usage: 433.7+ KB

Figure 5.3 Datatypes of loan_approval_dataset.csv

Dataset contains 10 numerical attributes and 3 non numerical attributes. The target class contains two variables that is the approved and rejected.

```
print(df['loan_status'].head())
```

| | |
|---|----------|
| 0 | Approved |
| 1 | Rejected |
| 2 | Rejected |
| 3 | Rejected |
| 4 | Rejected |

Name: loan_status, dtype: object

Figure 5.4 Class labels

Target Variable

- **Target (Predicted Variable):** The dataset's target variable is Loan Status, which indicates if a loan application is approved (Y) or rejected (N).

The Loan Eligibility Prediction dataset is a collection of financial and personal data used to assess an individual's or entity's eligibility for loan approval. It includes key features like CIBIL score, income, employment status, loan term, loan amount, assets value, and loan status. This dataset is widely used in machine learning and data analysis to build predictive models and algorithms that determine the probability of loan approval based on various factors. By analyzing these features,

financial institutions can automate and optimize the loan approval process, improving decision-making and minimizing risk.

6. Analysis of Algorithm

XGBoost (Extreme Gradient Boosting)

XGBoost (Extreme Gradient Boosting) is a powerful and efficient supervised learning technique used for both classification and regression problems. It belongs to the family of boosting algorithms, which are ensemble learning techniques that combine the outputs of multiple weak learners to create a strong predictive model. XGBoost is known for its speed and performance, making it a popular choice for structured/tabular data tasks. XGBoost builds decision trees sequentially, where each new tree corrects the errors made by the previous trees. Instead of taking majority votes (as in Random Forest), XGBoost optimizes the model using gradient descent to minimize the loss function. The final output is a weighted sum of the predictions from all the trees.

Working

1. Initialize the model by making a single prediction, often the mean value of the target variable for regression or a uniform probability distribution for classification.
2. Calculate the residual errors (for regression) or gradients (for classification) between the predicted and actual values.
3. Fit a new decision tree to predict the residuals or gradients. This tree aims to correct the errors made by the previous trees.
4. Update the model's predictions by adding the predictions from the new tree, scaled by a learning rate.
5. Repeat Steps 2-4 for a predetermined number of trees or until the model converges.
6. For a new data point, combine the predictions from all the trees to make a final prediction

Pseudocode

1. Input:
 - Training data with N samples and M features.
 - Number of boosting rounds: T.
 - Learning rate: η (eta).
 - Maximum depth of each tree: D.
- 2.

2. Initialize:

- Start with initial predictions, often the mean of the target variable (for regression) or Equal probabilities for all classes (for classification).

3. For each boosting round (from 1 to T):

- Compute the gradient of the loss function with respect to the predictions.
- Fit a decision tree to predict the negative gradient (residuals).
- Scale the predictions from the tree by the learning rate η .
- Update the model by adding the scaled predictions to the current predictions.

4. Final Model:

- The final prediction for a new sample x is the sum of the predictions from all trees.

Key Features of XGBoost:

- Regularization: Helps prevent overfitting by penalizing complex models.
- Handling Missing Data: XGBoost can handle missing data internally, making it robust.
- Parallelization: XGBoost can train models faster by using parallel process.
 - Sparsity Awareness: Efficient handling of sparse data (i.e., data with many zeros)

XGBoost is highly flexible, capable of working with different loss functions, and has become a go-to algorithm for winning machine learning competitions due to its high accuracy and efficiency.

7. Project Feasibility

A project feasibility study evaluates the viability of a proposed project, considering its technical, economic, and operational aspects. For the Credit Management System, this analysis ensures that the project can be implemented successfully within its defined scope, budget, timeline, and available resources, while addressing any potential challenges before significant resources are committed.

i. Technical Feasibility

The Credit Management System is technically feasible, leveraging modern, reliable, and widely supported technologies:

- **Front-end:** The user interface is built using HTML, CSS, and JavaScript, providing a responsive and intuitive experience for customers and administrators. These technologies ensure smooth interaction on a variety of devices, such as desktops, tablets, and smartphones.
- **Back-end:** The Python Django framework is used for the back-end, handling the business logic, processing loan applications, managing workflows, and ensuring secure data handling.
- **Database:** SQL and NoSQL databases are employed for storing and managing customer and loan data, ensuring reliability, scalability, and data integrity.
- **Machine Learning Models:** XGBoost powers the credit risk prediction engine, evaluating applicants' creditworthiness based on historical data, financial parameters, and predicted repayment likelihood.

The choice of widely adopted open-source technologies enhances compatibility and scalability while offering strong community support and regular updates. These technologies ensure the system can be easily integrated with various devices and is capable of handling future growth.

ii. Economic Feasibility

The Credit Management System is designed to be cost-effective by utilizing open-source tools and cloud-based infrastructure:

- **Open-Source Tools:** The use of Python, Django, and XGBoost eliminates licensing fees, lowering software acquisition costs. Additionally, libraries such as Pandas and NumPy aid in data processing and model building efficiently.
- **Automation:** Features like automated credit scoring, loan eligibility prediction, and loan approval workflows reduce manual effort, cutting down on operational labor costs.
- **Infrastructure:** The system can be deployed on standard web servers or cloud infrastructure, avoiding the need for costly on-premise hardware. This makes the solution both scalable and cost-effective.
- **Routine maintenance, hosting, and occasional updates** will incur some ongoing costs. However, these costs are far outweighed by the long-term benefits, including improved decision-making, faster loan approvals, reduced operational errors, and increased customer satisfaction, contributing to a solid return on investment (ROI).

iii. Operational Feasibility

The Credit Management System enhances operational efficiency by automating key processes, improving decision-making, and streamlining loan management:

- **Frontend Usability:** The HTML, CSS, and JavaScript front-end offer a user-friendly design, allowing customers to easily input their financial details and receive loan eligibility results, while administrators can efficiently review and approve loan applications.
- **Backend Efficiency:** Python Django enables seamless backend processing, including loan application handling, data management, and integration with the machine learning models for credit risk prediction.
- **Database Management:** The use of SQL and NoSQL databases ensures secure, accurate, and reliable storage of customer data, loan details, credit scores, and historical records.
- **Machine Learning Integration:** The XGBoost model analyzes historical loan data to predict applicants' creditworthiness, improving the accuracy and speed of credit decisions. It enables real-time risk assessment, reducing manual review times and errors.

The Credit Management System is a highly feasible solution that integrates advanced technologies to streamline credit management, reduce risks, and improve operational efficiency. Its combination of technical robustness, cost-effectiveness, and automation ensures it can transform how financial institutions assess credit risk and manage loan approvals.