IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AI-Driven Loan Eligibility Prediction And Document Verification

BATTA CHAITANYA SRI*[1], DONTIBOINA NAGA LAKSHMI[2], BELLAM RAVINDRA[3], KALAGOTLA PEERAIAH[4], DWIJA DESIREDDY[5]

1Student, Department of CSE(AIML), Bapatla Engineering College, Bapatla 522101, AP, India.
2Student, Department of CSE(AIML), Bapatla Engineering College, Bapatla 522101, AP, India
3Student, Department of CSE(AIML), Bapatla Engineering College, Bapatla 522101, AP, India
4Student, Department of CSE(AIML), Bapatla Engineering College, Bapatla 522101, AP, India
5Assistant Professor, Department of CSE(AIML), Bapatla Engineering College, Bapatla 522101, AP, India

Abstract: Bank loan processing entails assessing the financial information of a loan applicant and verification of the information that has been presented. This check is in most instances performed manually and hence can slow down the process of approval and in some instances it may invoke some error. This paper proposes a hybrid model, which combines document validation and AI-based loan assessment and monitoring tools. The system is a fast-tracked clarification of the decision, which introduces transparency and enhances the efficiency of the process. The candidates complete their financial information and attach their documents, such as identity evidence, paychecks, and credit card reports. These documents are analyzed by the system to identify the relevant information, and compared with the information that the applicant has filled in; any mismatch should cause a red flag which is achieved via LlamaIndex. Once we have verified the data, a LoRA fine-tuned Qwen model is used in the prediction of the loan status either approved or rejected based on the financial profile of the applicant. The process is also more transparent and efficient because a brief explanation of the decision is presented by the system.

Index Terms - Loan Eligibility Prediction, Qwen Model, Document Verification, LlamaIndex, Financial Data Processing, Automated Decision System

INTRODUCTION

Granting of loans is a very important aspect in the operations of banks and financial institutions. Organizations will have to examine the financial history of an applicant, their credit record, and other supporting documents to determine whether an applicant has the ability to repay the loan before they are granted a loan. This assessment has customarily been conducted through manually authenticating documents with predetermined rules of finances. Despite being in practice over an extended period of time, these methods are also known to be very consuming in human labor and result in delays in loan application processing.

Over the past few years, financial institutions have found it difficult to physically check the loans being brought by different individuals as the number of loan applications keeps rising. Employees will have to run through numerous forms including ID proofs, salary slips and credit reports as well as ensure the information submitted by the applicants is accurate. This strategy can be time consuming and even lead to errors or other problems when the data located is not the same as the one located in the uploaded files.

Nevertheless, the recent developments in the field of artificial intelligence and machine learning have offered new possibilities of automatizing the decision-making process in the financial market. Different predictive models may adopt financial characteristics (i.e. income, credit score, employment status and liabilities) to forecast the chances of loan approval. The models can help the financial institutions to make faster and more uniform decision than the traditional rule-based systems.

At the same time, the advancements in the technologies related to processing documents have rendered the process of automatic extraction of the information contained in digital documents a possibility. There is a possibility of coming up with systems that do not only process financial information but also check the authenticity of the documents submitted by through the use of document information extraction coupled with predictive models depending on intelligence.

This paper suggests a loan eligibility prediction system that is an integration of an automated document verification system and a prediction model based on AI. The system allows quick and efficient decision making.

I. LITERATURE SURVEY

Bansal et al. (2020) suggested a loan prediction model based on the machine learning models, including Logistic Regression, Decision Tree, and Random Forest. Their work revealed that machine learning models are successful in the analysis of features of applicants like income, credit history, and loan amount, to ascertain eligibility. The findings revealed that they are capable of lowering the human touch work and enhancing the speed of loan processing. The disadvantage of such system is that it totally depends on structured input information whereas in real life applicants must submit the documents like identity evidences, pay slips. This is not the system to use when dealing with the data of documents [1].

Naik (2021) recommended a credit-risks framework based on machine-learning. Past data of borrowers was employed to forecast whether a borrower will default or not in the loan. The researcher tried different classification models to make forecasts which discovered LightGBM to possess the highest rate and efficiency of prediction. The discoveries indicate that machine-learning algorithms can assist banks that have enhanced credit-risk and lending decision-making accuracy. Structured financial attributes were the only ones that were used in this study. This prediction model failed to serve the purpose of document verification or the inconsistency between what the applicants presented and the documents they presented. [2].

Liu et al. (2022) suggested a neural network-based system of credit evaluation based on deep learning. This system processed complex links among various financial qualities and displayed better prediction accuracy than the traditional machine learning models. The system consumed more computational resources and large training datasets. Moreover, the model was only concerned with the financial attributes and failed to cover automated document parsing and verification methods [3].

An additional paper suggested a multi-agent system, which is driven by large language models. Different agents in this system would have joined together to examine the various aspects of financial information of the borrower. This system demonstrated that credit scoring performance is enhanced by the multi-agent approach. The system primarily addressed the structured financial data analysis in credit scoring. It lacks document validation or automatic scuttling of financial documents uploaded in it like salary slips, identity proofs, and credit reports [4].

II. PROPOSED SOLUTION

This project presents an AI-based Loan Eligibility Prediction and Document Verification System that can help financial institutions to review loan applications effectively. The system enables the applicants to give their personal information and financial records via an online platform. With the aid of artificial intelligence and automated document processing the system analyses the information of the applicants and gives a loan eligibility decision with the explanation. The strategy is useful in minimizing the manual labor, enhancing accuracy of decisions and accelerating the loan approval process.

1. Smart Loaning Application Suite

The platform gives the users a digital platform on which they can post loan applications with ease. The details provided by the applicants include personal and financial details like income, employment information, amount of loan and credit score. This data assists the system to comprehend the financial history of the applicant and put the data to be further considered.

2. Document Processing Automation

The system enables the applicants to post the supporting documents like identity proofs, salary slips and credit reports. These are automatically processed through automated document analysis module which extracts vital details of the applicant such as name of applicant, income details, credit score, etc. The step transforms unstructured document data into structured information to analyze it correctly.

3. Data Checking and Data Verification

The system matches the information grabbed in the written document and the details that the applicant has provided. This comparison will be used to detect any mismatch or misinformation. The system minimizes the possibilities of mistakes by automatically verifying the data and enhances the accuracy of the process of loan evaluation.

4. AI-Based Loan Evaluation

Once the applicant data is verified, the system will analyze the financial profile with the help of an artificial intelligence model. Some of the factors that the model examines include income level, credit record, repayment capacity and loan value. According to this analysis, the system indicates the loan application to be approved or rejected.

5. Clear Decision Output

The user is shown the final result in a simple and easily comprehensible form. The system displays the eligibility status of loans and brief description of the decision. It is a transparent output that is beneficial to the applicants and financial institutions to know why the loan was approved or denied.

III. METHODOLOGY

The suggested system will combine automated document verification and a loan eligibility prediction model with the use of artificial intelligence. The methodology will be divided into a few steps such as data collection, document processing, data verification, and prediction. The system accepts the user inputs, and compares the uploaded documents to verify the documents and then Qwen model predicts the loan status that was either approved or rejected along with the reasons.

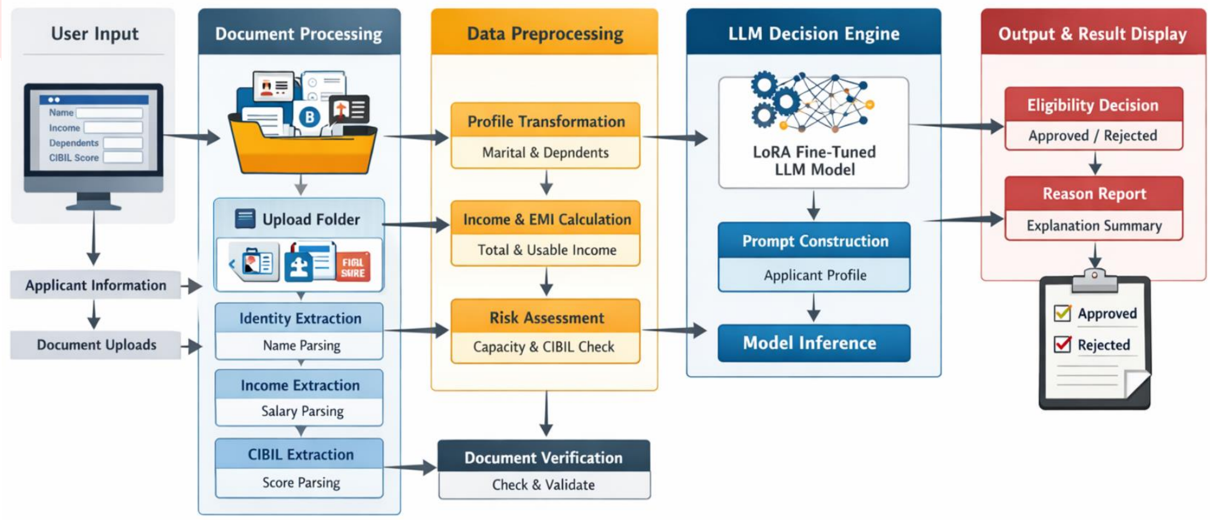

Figure 1: System Architecture of AI-Driven Loan Eligibility Prediction and Document Verification

4.1 User Input Module

This phase gathers the required details of the loan applicant. Personal and financial attributes are both gathered via the system interface.

4.1.1 Applicant Information

The applicants would update the necessary input data in terms of name, gender, marital status, number of dependents, applicant income, co-applicant income, and credit score. These are the most significant parameters of the financial capability of the applicant. This information is stored in a session object of the system to be processed further.

4.1.2 Document Upload

Once the necessary information is entered, the applicant places the necessary documents such as identity proof, salary slips, credit reports among others. These documents have significant information that is used to confirm the information in the application form. These writings are stored in a folder with a uniquely generated directory with a UUID-based folder structure to avoid conflicts with files.

4.2 Document Parsing Module

The system checks documents uploaded to the system through automated document parsing. System extracts valuable content of uploads with LlamaParse, it turns the text of documents into a machine-readable one. The uploaded documents are individually processed in order to extract textual information. The identity document is read to get the name, the credit report is read to get cibil score, the salary slips are read to get income details. The data extracted is validated and the result of the verification sent to the prediction model.

4.3 Preprocessing and Feature Transformation of Data.

The system has a preprocessing stage that turns the raw inputs into meaningful financial indicators after which the data is sent to the prediction model. The content entered by the user is converted to well-organized financial features. The marital status input is put under the category of either Married or Unmarried. Dependents input is divided into three levels namely Low, Moderate, and High.

Income and Financial Capacity Calculation.

Calculation of Total Income: Total income is obtained by adding the amount of applicant income and co-applicant income.

Usable Income Estimation: The system is based on the assumption that half the total income can be used to repay the loans after taking up the cost of living.

EMI Repayment Capacity Analysis

The system is used to calculate the capability of the applicant to pay off the loan on the basis of the EMI ratio.

EMI Ratio Formula:

EMI Ratio = (Loan Amount/Loan Period) / Usable Income.

According to this ratio, repayment capacity is divided into: High, Moderate, Low. This is one of the main parameters that are employed by the prediction model.

4.4 Prediction of Loan Eligibility

The system then predicts loan eligibility using a language model based on transformers which has been fined tuned after preprocessing and verification. The prediction engine is a pretrained Qwen model with LoRA fine-tuning that is used to tune it using loan evaluation tasks. LoRA is an adaptation method that allows fine-tuning large language models without changing all model parameters.

4.4.1 Prompt Construction

The system builds a programmed prompt with the financial profile of the applicant. Some details in this prompt are: Personal information, Financial capacity, Quality of credit score, Document verification status. This hint is then transferred to the language model to produce a rational decision.

4.4.2 Model Inference

The model works on the tokenized point to come up with a response. This reply has arguments and a conclusion. The parameters of the controlled sampling used in the generation process include: Temperature, Top-p sampling, Repetition penalty These parameters enhance the quality of response with consistency.

4.5 Result Generation

The final output includes:

1. Rejection/Approval of loan.
2. AI-generated reasoning

This output enables the applicant to know the factors that are involved in the decision.

IV. RESULTS AND DISCUSSION

The proposed System was submitted to various loan application scenarios having financial profiles and supporting documents. The system verifies the applicant, verifies the documents provided and analyzes the financial capability and is then decided based on AI. The outcome of the experiments demonstrates that the platform manages to combine the three parts of document verification, financial analysis, and intelligent decision generation in order to help with the process of loan evaluation.

Input Interface Applicant Details:

The initial phase of the system will enable the applicant to fill in personal and financial information in a web-based form. This interface gathers information such as name of applicant, earnings, marital status, number of dependents, value of loan, loan repayment time, employment, and the place of property. In the process of testing, the interface was able to capture the data of the applicants and send the data into the processing module.

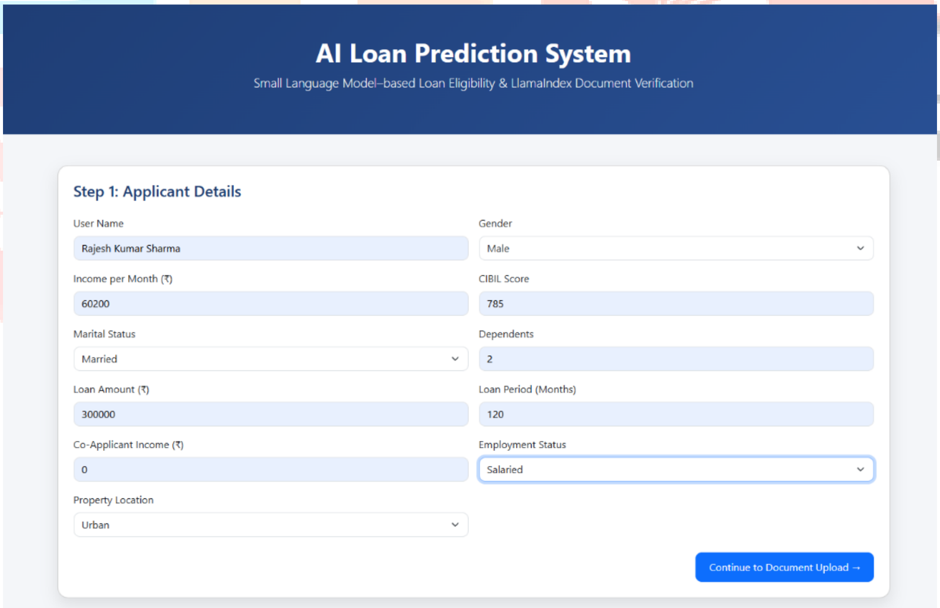

Figure 2: Applicant Details Interface

Uploading document:

The second step is the one that permits the applicants to post the necessary documents including government identity proof, salary slip and credit score report. These papers will be necessary in order to verify the information given by the applicant. Once the documents are submitted, the document parsing module will extract the important information in the files that are uploaded by the applicant. This retrieved information is contrasted to the details given by the user. This check-up procedure assists in revealing discrepancy and enhances the effectiveness of the loan assessment procedure.

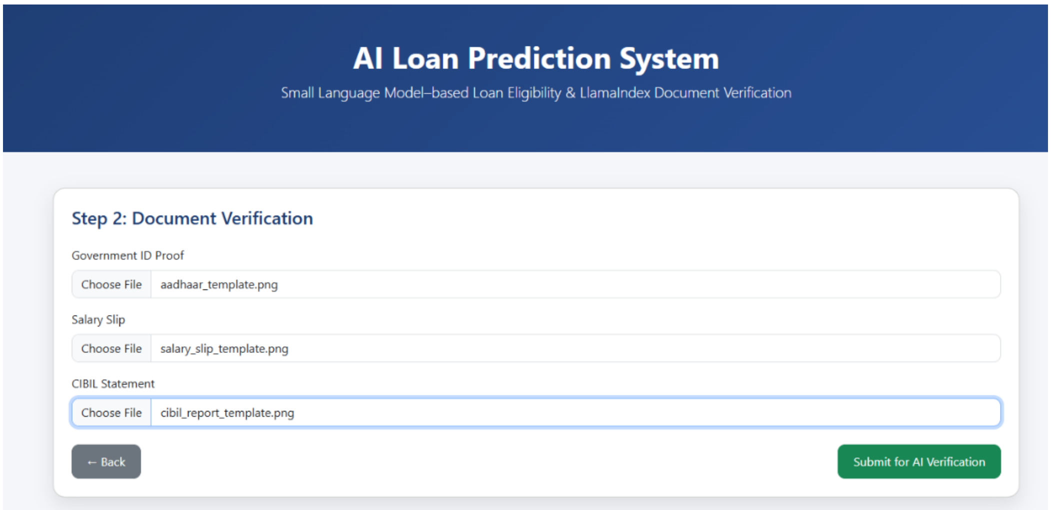

Figure 3: Document Verification Upload Interface

AI-Generated Loan Decision and Reasoning:

Once the document verification and financial analysis was complete, the LoRA fine-tuned Qwen model analyzed the profile of the applicant and produces a loan eligibility decision and the reasoning of the AI. It will give the reason behind the loan eligibility status.

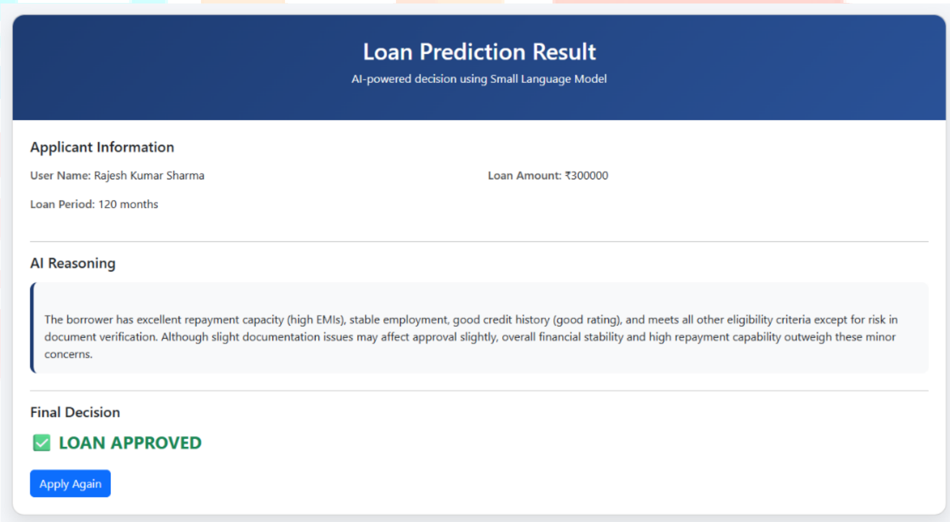

Figure 4: Loan Prediction Result

Table: System Performance Evaluation

Metric	Observation
Document Parsing Accuracy	High — LlamaParse extracted fields correctly from PDFs and images
Verification Consistency	Effective cross-matching of applicant-submitted vs. document data
Financial Analysis	EMI ratio, usable income, and credit band computed correctly
Decision Quality	Decisions aligned with standard credit evaluation criteria
Reasoning Clarity	Natural-language explanations were clear and informative
Processing Time	Significantly faster than manual evaluation (seconds vs. days)

Conclusion:

This paper presented an AI-based Loan Eligibility Prediction and Document Verification System that will automate the whole loan evaluation process by financial institutions. The suggested system solves the inherent weaknesses of the current loan assessment methods by adding intelligent document parsing through LlamaParse, structured financial analysis, and explainable AI-based decision-making with a LoRA fine-tuned Qwen LLM.

The given system has proven to be effective in gathering the information about applicants, processing the data automatically loaded with financial documents, and coming up with the necessary financial ratios and offering loan eligibility prediction with the possibility to explain the decision-making process with the help of AI. Experimental validation of the efficiency of the proposed system in every step of a pipeline of loan evaluation was carried out on sample loan applications.

The major contribution of the paper is that it gives a comprehensive assessment pipeline of loans that fills the void in the literature between automated document verification and AI-based loan eligibility forecast which to the best of our knowledge has been overlooked.

V. REFERENCES

[1] S. Bansal, A. Sharma, and R. Goel, “Loan Prediction using Machine Learning Techniques,” International Journal of Computer Applications, vol. 176, no. 38, pp. 1–5, 2020.

[2] K.S. Naik, “Predicting Credit Risk for Unsecured Lending: A Machine Learning Approach,” arXiv preprint arXiv:2110.02206, 2021.

[3] Y. Liu, H. Chen, and J. Zhang, “Credit Risk Assessment Using Deep Learning Models,” IEEE Access, vol. 10, pp. 12345–12356, 2022.

[4] G. Jajoo, P. A. Chitale, and S. Agarwal, “MASCA: LLM based-Multi Agents System for Credit Assessment,” arXiv preprint arXiv:2507.22758, 2025.

[5] I. Brown and C. Mues, "An experimental comparison of classification algorithms for imbalanced credit scoring data sets," Expert Systems with Applications, vol. 39, no. 3, pp. 3446–3453, 2012.

[6] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[7] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory, 1992, pp. 144–152.

[8] J. Schmidhuber, "Deep learning in neural networks: An overview," Neural Networks, vol. 61, pp. 85–117, 2015.

[9] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proceedings of NAACL-HLT 2019, Minneapolis, MN, pp. 4171–4186.
[10] LlamaIndex, "LlamaParse: Advanced document parsing for AI applications," LlamaIndex Documentation, 2024. [Online]. Available: https://docs.llamaindex.ai
[11] Alibaba Cloud, "Qwen: Large language model series," Technical Report, Alibaba Cloud, 2023.
[12] E. J. Hu et al., "LoRA: Low-rank adaptation of large language models," in International Conference on Learning Representations (ICLR), 2022.
[13] A. Vaswani et al., "Attention is all you need," in Advances in Neural Information Processing Systems, vol. 30, 2017.
[14] M. A. Paleologo, A. Elisseeff, and G. Antonini, "Subagging for credit scoring models," European Journal of Operational Research, vol. 201, no. 2, pp. 490–499, 2010.