WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Study on Machine Learning in Credit Risk Assessment for Startup Companies

Ms. J Disowja[1], Dr. S. Venkata Ramana[2]

[1] Student, MBA Department, KL Business School, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Guntur, AP, India
Email.Id: 2301510151mba@gmail.com
[2] Assoc. Professor, KL Business School, KLEF(Deemed to be University), Green Fields, Vaddeswaram, A.P. – 522302.
E-mail-Id: svramana@kluniversity.in,

ABSTRACT:

The paper discusses the transformative power of machine learning in the assessment of credit risk for startups. Based on static models and fixed datasets, traditional credit scoring systems usually fail to capture the distinct and complex risk factors surrounding startups, such as limited credit histories and volatile cash flows. Machine learning, instead, uses data-driven algorithms that can analyze large amounts of data and identify hidden correlations. This capacity helps financial institutions improve predictive accuracy in a dynamic environment, thus making better-informed lending decisions. The research focuses on developing state-of-the-art ML models, such as Random Forests, Support Vector Machines, and Neural Networks, which outperform traditional models because they learn from heterogeneously distributed datasets, including transaction histories, social media metrics, and psychometric evaluations. Key benefits to these models include enhanced risk mitigation, lower NPAs, and a holistic perspective of the creditworthiness of startups.

Despite these advancements, there are still challenges. These include a lack of data availability for startups, model interpretability, and regulatory compliance issues. Ethical considerations in addressing bias and ensuring transparency also come into play in implementing ML in financial decision-making. A research gap identified is in the tailoring of ML models specifically to startups and in the integration of unconventional data points such as founder experience and market conditions.

By focusing on banks,this study demonstrates how ML techniques can revolutionize credit risk assessment, paving the way for more inclusive and innovative lending practices while aligning with regulatory frameworks.

Keywords: Credit Risk Assessment, Machine Learning in Finance, Startup Loan Evaluation, Predictive Analytics, Credit Scoring Models, Risk Classification, Loan Approval Prediction, Alternative Data for Credit Risk, Financial Risk Modeling

Introduction:

The Increase in startup companies had made a need for effective credit assessment techniques. The traditional models are mostly based on the financial statistics which cannot capture the dynamics of startup risk. Machine learning models use different types of data with different datasets to understand the patterns that can increase the accuracy of risk assessment. This research is to understand the application of machine learning to assess the credit worthiness of the startups from 2020 to 2024 loan data in different sectors.
Machine learning (ML) provides a viable solution by analysing vast amounts of structured and unstructured data to identify patterns, predict credit risks, and make more informed lending decisions. By leveraging algorithms such as Random Forest, Logistic Regression, and Decision Trees, financial institutions can process complex relationships between variables that traditional scoring models might overlook. Additionally, ML-based approaches can incorporate alternative data sources such as market trends, industry performance, and even behavioural analytics to provide a more holistic risk assessment.

2. Literature Review :

J. M. G. de Almeida et al. (2020) research examines the use of supervised learning models such as Decision Trees, Random Forests, and Extremely Randomized Trees and concludes that tree-based models are most appropriate for enhancing credit scoring accuracy.

Machine Learning-Based Credit Risk Prediction: A Case Study (L. H. Zhang et al., 2019): In this, traditional methods are compared with ML models, and the way Random Forest and Neural Networks outperform logistic regression models is revealed.

A detailed analysis of machine learning techniques for credit risk assessment has been discussed by Sharma and Gupta (2021), which uses different ML techniques and shows the importance of deep learning compared to traditional models.

Enhancing Credit Risk Prediction through Ensemble Learning Techniques (A. T. Lee et al., 2022): Examines ensemble learning techniques and finds that they enhance predictability and interpretability of models.

Deep Learning for Credit Risk Evaluation: A Review and Framework (M. R. Patel & N. K. Singh, 2023): It describes deep learning models such as CNNs and RNNs, which provides higher accuracy while operating with large and complex data sets.

Feature Engineering for Credit Risk Models: A Machine Learning Perspective (P. J. Chen et al., 2021): Highlights feature selection as the key area to improve ML model performance.

The Application of Big Data to Credit Risk Assessment: Machine Learning Applications (R. D. Kumar & S. L. Yadav, 2020): Describes the application of big data during the credit risk assessment process by ML applications.

A Comparative Study of Machine Learning Algorithms for Credit Default Prediction (T. S. Nguyen et al., 2019): In this Research it compare various ML models for credit risk prediction and how we can select the optimal model.

F.Y. Wang et al. (2023) introduced an explainable AI framework in their paper that is able to bring in more transparency into credit rating models based on machine learning.

3. Objectives :

- To analyze how the current ML models affect the loan default prediction.
- To discuss why loan defaults.
- To analyze the effect of current ML models on NPA reduction.

4. Methodology :

- **Data Collection:** The research was based on secondary data supplied by banks, credit rating agency reports, and loan documents of startups.
- **Data Preprocessing:** One-hot encoding of categorical features, processing of missing values, and scaling of numerical features.
- Random Forest, Logistic Regression, and Decision Trees machine learning algorithms were used in prediction across different startup risk classes.
- **Evaluation Metrics:** Accuracy, precision and AUC-ROC scores were used to measure model performance.

5. Conceptual Model :

Conceptual Framework contains financial variables, ML-based prediction, and risk categorization.

- **Input Parameters**: Amount borrowed, income, debt-to-equity ratio, credit rating.
- **ML Algorithms**: Random Forest, Logistic Regression, Decision Trees.
- **Risk Assessment**: Startups as Low, Medium, or High risk.
- **Decision Output**: Granting or denying of loan based on ML forecasts.

6. Results and Data Analysis :

Exploratory data analysis (EDA) showed highest approval rates for technology and health care startups and rejections received by manufacturing and retail. Feature importance analysis identified revenue, debt-equity ratio, and credit score as extremely predictive. Random Forest model provided an accuracy rate of 85%, which was higher than provided by Logistic Regression (78%).

7. Major Findings:

Machine learning programs have also introduced better precision when calculating startup credit risk.
Income and credit history are critical determinants of loan approval.
Decision Trees, while less precise, are simpler to comprehend.

8. Challenges:

- **Data Availability**: Restricted startup financial history influences model performance.

- **Model Transparency:** More advanced ML models need interpretability enhancements**.**
- **Bias and Fairness:** Algorithms can perpetuate inherited bias from past lending history.

9. Conclusion & Future Work :

Machine learning complements startup credit risk assessment, reducing loan choice uncertainty. Greater model transparency and the use of alternative information sources (i.e., social media, marketplace trends) will further improve decisions. Future research on hybrid ML-human choice platforms for financial risk analysis should follow.

10. REFERENCES :

1. Almeida, J. M. G., et al. (2020). Credit Scoring Using Machine Learning Techniques. *Journal of Applied Computer Science and Intelligent Technologies.*
2. Zhang, L. H., et al. (2019). Predicting Credit Risk with Machine Learning: A Case Study. *Journal of Financial Risk Management.* Sharma, S. K., & Gupta, R. K. (2021). Machine Learning Approaches for Credit Risk Assessment: A Review. *International Journal of Management and Technology.*
3. Lee, A. T., et al. (2022). Enhancing Credit Risk Prediction Using Ensemble Learning Techniques. *International Journal of Artificial Intelligence.*
4. Patel, M. R., &Singh, N. K. (2023). Deep Learning for Credit Risk Assessment: A Review and Framework. *Journal of Applied Computer Science and Intelligent Technologies.*