



## Research Article

Copyright © All rights are reserved by Qazi Waqas Khan

# Automating Loan Status Prediction Using Machine Learning: A Comparative Study of NODE, Tab Net, and ANN Models with Chi-Square Feature Selection

Qazi Waqas Khan\*

Department of Computer Engineering, Jeju National University, Jeju 63243, Jeju Special Self-Governing Province, Republic of Korea

\*Corresponding author: Qazi Waqas Khan, Department of Computer Engineering, Jeju National University, Jeju 63243, Jeju Special Self-Governing Province, Republic of Korea.

Received Date: July 12, 2024

Published Date: July 29, 2024

## Abstract

The bank makes loans to customers at an interest rate. Loan interest is the bank's principal income source. After checking the customer's loan details, the bank approves the loans because it does not want to invest money in an untrustworthy consumer. Our model aims to automate the prediction of loan statuses using machine learning methods. Loan status prediction is a classification problem, and this work employs classification approaches such as NODE, TabNet, and ANN models for loan status classification. The experimental results demonstrate that the ANN and TabNet model has 83 and 82% accuracy, respectively, for the classification of loan status. It concludes that using a chi-square feature selection improves the prediction results for loan status classification.

## Introduction

Many banks make their money primarily through loan distribution. Most of the bank's income is derived from loan interest [1]. Banks make these loans with interest and make a lot of money. They always want to invest their money in a reliable customer who will repay it quickly [2]. Customer information and certain documents are required with the loan application to determine whether the consumer can repay the amount [3]. The data provides information about siblings, home ownership, gender, property, and income. This detail determines whether the bank approves or rejects the loan application [4]. An automated system that approves or denies applications in less than one minute is necessary to minimize the processing time [5].

Machine learning involves learning from the data. Several researchers offered machine learning-based loan prediction systems. Srinivasa et al. [6] applied Random Forest to predict

loans. Sivagaminathan et al. [7] employed a multi-relational fuzzy classifier to predict loans. Kavita et al. [8] used Random Forest to forecast the loan approval status. However, these techniques imbalance class labels and rely on feature selections.

## Methodology For Loan Status Prediction

Data is initially selected for the analysis. After this, data is prepared using preprocessing methods. Classification algorithms are employed on the preprocessed data, and lastly, the model is evaluated using different metrics.

## Machine Learning Models

### Tab Net

Tab Net is a neural network framework designed specifically for handling tabular data. The sequential attention process combines

the strength of neural networks with decision trees. By focusing on the essential features at each decision point, TabNet enhances accuracy and interpretability. It outperforms the traditional method, making it a robust choice for different predictive tasks [9].

### Neural Decision Ensemble

NODE is a state-of-the-art neural network model that combines decision tree concepts with deep learning capabilities. Complex traditional trees are transformed into a distinct format, allowing smoother training. It is designed to handle tabular data, and this model efficiently utilizes neural networks' symbolic power and

hierarchical decision mechanism [10].

### Artificial Neural Network

Artificial neural networks (ANNs) are computing models influenced by the structure of human brains. They consist of layers organized of interconnected nodes that process and learn from input data. ANNs are good at finding patterns and relationships in large, complex datasets, making them essential choices for tasks like natural language processing, image and speech recognition, and predictive analytics [11].

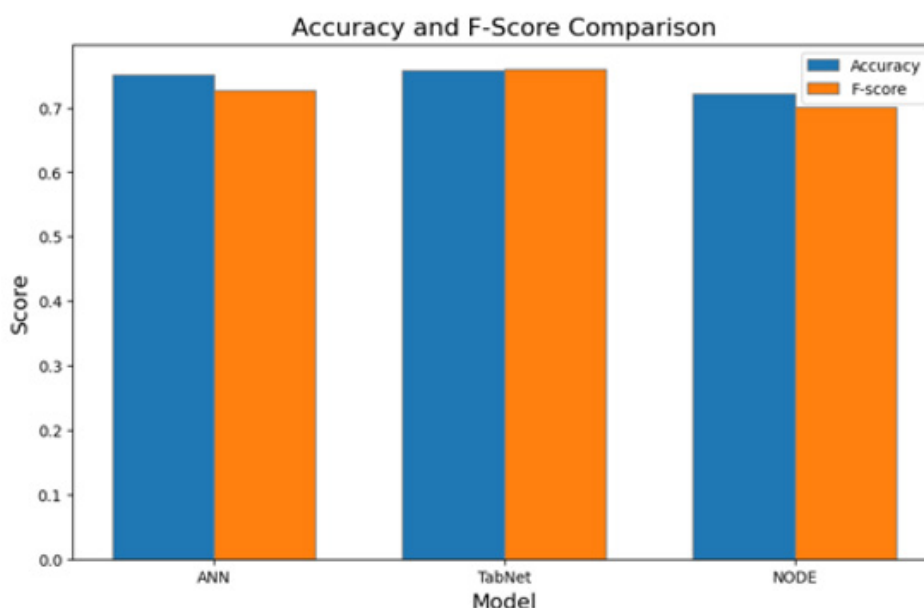
## Results

**Table 1:** Results of a deep learning model for loan status prediction without feature selection.

Method	Accuracy	Precision	Recall	F-score
ANN	0.7515	0.7089	0.7467	0.7273
Tab Net	0.7574	0.677	0.8667	0.7602
NODE	0.7219	0.6707	0.7333	0.7006

Table 1 demonstrates the results of ANN, TabNet, and NODE models for loan status classification. The results show that the TabNet surpasses other models as it has the highest accuracy

of 75.74%, recall (86.67%), and F-score (76.02%). While on the other hand, ANN has the highest precision rate (70.89%), Tab Net's overall performance makes it a more reliable model.



**Figure 1:** Without feature selection.

In a Figure 1 the results of a models are presented for loan status classification. From the figure we see that the TabNet model

has higher results for loan status classification. While the NODE has lower results than the ANN and TabNet model.

**Table 2:** Results of a deep learning model for loan status prediction with chi-square feature selection.

Method	Accuracy	Precision	Recall	F-score
ANN	0.8343	0.7701	0.8933	0.8272
Tab Net	0.8284	0.7674	0.88	0.8199
NODE	0.787	0.701	0.9067	0.7907

Table 2 presents the experimental findings of ANN, TabNet, and NODE models for the loan status prediction with chi-square feature selection. The result shows that all three models achieve the highest prediction results with chi-square feature selection. ANN achieves the highest accuracy, precision, and f-score, while TabNet

closely follows with a slightly lower metrics score. NODE improves recall but has lower accuracy, precision, and f-score than ANN and TabNet. Overall, results are enhanced with chi-square feature selection, with ANN becoming a more reliable model across all the metrics.

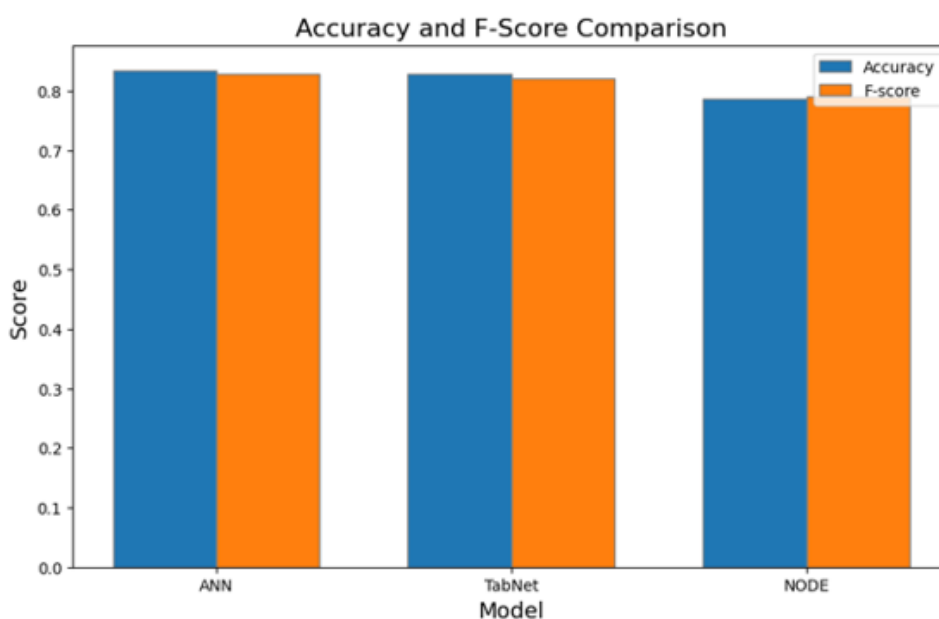
**Figure 2:** With chi-square feature selection.

Figure 2 compares the Accuracy and F-score of proposed deep learning models to predict loan status using the chi-square feature selection method. The figure explains that the proposed models achieve a significant boost with the chi-square feature selection method. ANN surpasses other models in accuracy and F-score, while TabNet has lower accuracy and F-score. NODE is the least effective among all the proposed classifiers.

## Conclusion

Loan status prediction is crucial in the banking and finance sector to determine loan repayment probability depending on the applicant's attributes. This study outlines the systematic approach used for predicting loan status. Three proposed machine learning classifiers are evaluated, and the results highlight the importance of feature selection. TabNet performed well without feature selection, while ANN outperformed the proposed classifiers with chi-square feature selection. The findings suggest that feature selection

enhanced the performance of the models, and ANN emerged as the most effective and reliable model for predicting loan statuses, achieving a higher accuracy of 83.43%.

## Acknowledgement

None.

## Conflict of Interest

None.

## References

- Maryem N, Lahrichi Y (2022) The determinants of banks credit risk: Review of the literature and future research agenda. *International Journal of Finance & Economics* 27(1): 334-360.
- Wardhani Mazli M, Iskandar M, Mircea Nedelea A (2022) Indicators of Giving Interest Rates to Customers and Debtors at PT. Bank X in Medan, Indonesia. *Ecoforum Journal* 11(1).

3. Munyendo CW, Acar Y, Aviv AJ (2022) Desperate times call for desperate measures: User concerns with mobile loan apps in Kenya. IEEE Symposium on Security and Privacy.
4. Erasmo P, Lorenzo F, Fallucchi F, De Luca EW (2023) The use of responsible artificial intelligence techniques in the context of loan approval processes. International Journal of Human-Computer Interaction 39(7): 1543-1562.
5. Mourtas SD, Katsikis VN, Stanimirovic PS, Kazakovtsev LA (2024) Credit and Loan Approval Classification Using a Bio-Inspired Neural Network. Biomimetics 9(2): 120.
6. Srinivasa MPL (2022) Loan Approval Prediction System Using Machine Learning.
7. Sivagaminathan PG, Vijayalakshmi CR, Thangaraj M (2018) A New Framework for Loan Prediction Using Multi relational Fuzzy Classifier.
8. Afrah K, Bhadola E, Kumar A, Singh N (2021) Loan Approval Prediction Model a Comparative Analysis. Quot.
9. Termeh SVR, Niaraki AS, Sorooshian A, Abuhmed T, Choi SM, et al. (2024) Spatial mapping of land susceptibility to dust emissions using optimization of attentive Interpretable Tabular Learning (TabNet) model. Journal of Environmental Management 358: 120682.
10. Sergei P, Morozov S, Babenko A (2019) Neural oblivious decision ensembles for deep learning on tabular data. arXiv preprint arXiv.
11. Agatonovic Kustrin S, Beresford R (2000) Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. Journal of pharmaceutical and biomedical analysis 22(5): 717-727.