

An Experimental Study on Credit Score Prediction Using K-Nearest Neighbors

Taylor Stonelake

Marymount University, Arlington, Virginia, United States

Abstract—Credit score classification is a vital component of risk management in the financial sector. It conventionally relies on obsolete models that fail to capture dynamic patterns. This study utilizes K-Nearest Neighbor (KNN) on a dataset comprised of demographic and financial features. The methodology involved preprocessing steps, followed by model training and performance calculations. The KNN model produced an accuracy of 88%, which exceeded the desired threshold of 70%. However, performance varied across the classes, with Low credit scores demonstrating lower precision and F1-scores. This shows the model had difficulty in predicting minority classes. The correlation matrix showed feeble linear relationships amongst the features, which significantly limit predictive power. The results of this study demonstrate strong potential for credit scoring but should be explored with substitute models and feature engineering to provide precision and impartiality in credit score classification.

Keywords—Correlation matrix; dataset; k-nearest neighbors; machine learning

I. INTRODUCTION

As technological innovation of technology has rapidly increased, machine learning (ML) has been a common component in data-focused decision-making. It allows systems to learn patterns and make predictions without specific programming across all industry sectors and in personal use. K-Nearest Neighbors (KNN) is a form of artificial intelligence (AI) that learns from labelled datasets [1]. KNN is simplistic in nature and used for both classification and regression, yet can still handle complex data for real-world situations [1, 2].

KNN can be traced back to the United States Air Force in 1951 and has evolved as well as been enhanced throughout recent decades [3]. It operates on proximity in feature space, meaning new data points are categorized based on the majority class of its K closest neighbors. It predicts this by averaging their values in regression tasks [1]. This approach does not assume underlying data separation, but it is influenced by components such as the choice of k, distance metric, and feature scaling, which have the potential to cause issues such as over- and underfitting.

The dataset utilized for credit score classification was from Kaggle. This data is essential in the financial sector to perform risk management and loan decisions [4]. KNN improves the accuracy and flexibility of credit score classification by examining historical financial data. The dataset used the following features: age, gender, income, education, marital status, number of children, home ownership, and credit score.

All these factors can be examined to determine an individual's creditworthiness. This study calculates the number

of High, Average, and Low credit scores and uses a label encoder to change the features to numerical substitutes that the model can comprehend.

The study encompasses data preprocessing, feature scaling, and KNN model training. The performance of the model is measured by accuracy, precision, recall, and F1-score; it also includes a confusion matrix. To mitigate the risk of class imbalance in the data, the Synthetic Minority Over-sampling Technique (SMOTE) is utilized.

II. LITERATURE REVIEW

Financial institutions have used credit scoring models to determine a consumer's likelihood to repay a loan for many decades. Traditional models have relied on statistical models and rules to assess lending risks based upon the consumer's history of repaying loans. These types of models failed to adapt to changing environmental conditions and recognize trends found in large datasets. As a result, researchers have explored the utility of ML to better predict the likelihood of an individual repaying a loan and provide more flexible models for assessing credit risk.

ML is specifically designed to discover hidden relationships in data and to predict outcomes without making assumptions. Thus, ML holds much promise for its role in guiding financial decisions. One of the most well-known supervised classification algorithms is the KNN algorithm. KNN uses classification based upon the labels of the closest neighbors to the unknown sample in a predefined distance metric. Therefore, KNN is a direct and adaptable method for classification problems. Unlike parametric-based models, which require knowledge of the underlying distributional characteristics of the data, KNN relies on no such assumptions and can thereby easily develop complex decision boundaries without requiring extensive model design.

A number of studies have examined the use of ML in Credit Scoring. More specifically, these studies examine if alternative data sets (as opposed to traditional financial data sets) can be used to create better predictive ability of credit scoring models and extend access to credit. According to the World Bank, current credit scoring models use both alternative data sets and financial record data to increase the predictive ability of credit scoring models and to provide greater access to credit [12, 13, 14]. Similarly, Xu (2025) found that machine learning models using alternative data sets were able to evaluate a borrower's risk much more comprehensively than traditional credit scoring models [3]. It appears from this research that machine learning models are able to find relationships/ patterns in the data that

traditional statistical methods cannot find and thus can provide superior predictive results.

Although ML models are superior to traditional statistical models in terms of determining a borrower's worthiness, several obstacles must be overcome in order to fully realize their potential. One major obstacle that researchers believe must be addressed is bias/fairness in machine learning predictive models. Martinez (2022) indicated that demographic information and/or an individual's borrowing behavior could lead to biased lending practices unless appropriately mitigated. As well, researchers continue to advocate for the development of fair/transparent predictive models that can effectively predict a borrower's likelihood of repaying a loan [4]. Wang (2024) further emphasized that machine learning models must follow ethical guidelines so that equal treatment can be afforded to all borrowers and still provide sufficient predictive capability to assess the financial risk associated with a loan [5].

Another important issue to consider when trying to use KNN in credit score classification is Feature Scaling. The reason for this is that KNN uses the distances among values in the database to calculate how close an unknown value is to the known values in the database. If features are scaled differently and/or normalized before classification takes place, then features measured on larger scales will dominate the classification process. Pinheiro et al. (2024) showed that feature scaling greatly impacts classification accuracy and can improve model reliability dramatically [6]. Another problem researchers often encounter when dealing with credit scoring databases is Class Imbalance. Most credit scoring databases are composed of large numbers of observations in specific categories of borrowers, which makes it difficult for machine learning models to accurately represent the under-represented categories. Researchers typically resolve this by employing Resampling techniques such as SMOTE. Wang (2024) indicated that balanced class distributions promote model fairness and enable detection of underrepresented credit categories. Due to the fact that most credit scoring applications involve classification into one of three categories (High Risk/Average Risk/Low Risk), this problem is particularly acute.

Although KNN offers advantages including simplicity, interpretability, and robustness for certain types of classification problems, KNN possesses some inherent disadvantages. The Trantor Team (2025) pointed out that KNN may suffer performance loss resulting from noisy attributes or high dimensionality within a database. Additionally, as a database grows larger, KNN can grow exponentially more computationally expensive as each prediction requires comparison with multiple neighbor observations [6]. These limitations indicate that, although KNN would make an acceptable baseline classifier, to achieve maximum results, optimization and comparisons with other Machine Learning classifiers will likely be required.

In summary, prior research shows that KNN can be utilized for the classification of credit scores, as well as some of the issues related to selecting features, class imbalance, fairness, and scalability. Further, prior research showed that through careful preprocessing of the data and tuning the parameters associated with the models, substantial improvements in classification

accuracy can be realized. The purpose of this study is to build on this body of literature and investigate if KNN can effectively classify credit scores based on demographics and financial characteristics as input features and to assess performance via accuracy metrics.

III. PROBLEM STATEMENT

Credit score classification is essential to the financial sector. It is a component in identifying creditworthiness for individuals in need of credit. Historically, financial models have been used based off statistical and rule-based systems [7]. Unfortunately, they fail to capture the overall picture. These models lead to faulty results and misclassifications.

Traditional methods focus on the dependency on historical data, assuming past behavior forecasts future performance. These methods fail to capture real-time updates such as financial changes, leading to inaccurate and unfair assessments [8]. Some adults lack sufficient credit history, preventing them from demonstrating financial responsibility [8]. There is limited data and predictive accuracy as mobile payments and subscription services are on the rise [9]. Lastly, biases are commonly based on demographics, underrepresented populations, and age [10].

To improve prediction accuracy and to adapt to evolving data, machine learning is needed. It offers a solution for establishing patterns, risk mitigation, inclusion of secondary data, real-time focus, and fairness/transparently [8, 9, 10]. KNN can increase prediction precision and the reliability of credit score classifications.

This research evaluates the effectiveness of KNN in classifying credit scores from historical data. The goal is to achieve reasonable levels of accuracy, precision, recall, and F1-scores while facing challenges like feature scaling and class imbalance; this can introduce prediction bias toward the majority class [11].

By identifying and addressing these roadblocks, this research seeks to evaluate how machine learning can improve traditional credit score methodology while increasing risk management and providing better credit score allocation,

IV. METHODOLOGY

This research project employed a supervised machine learning methodology to categorize credit scores into one of three categories: High, Average, or Low, utilizing the KNN classifier. KNN is a type of classifier that utilizes the class assignments of the closest neighboring data points to predict the class assignment of a new unclassified data point. A primary assumption underlying the KNN algorithm is that data points with like characteristics have like class assignments. To convert the categorical nature of the target variable into a format suitable for use with a machine learning model, the researcher encoded the credit score categories into numerical representations using label encoding. Specifically, the researcher assigned the integer value "0" to Low, "1" to Average, and "2" to High.

A subset of a large public dataset downloaded from Kaggle entitled Credit Score Classification was used to develop the predictive credit scoring model. The dataset included many demographics and financial characteristics commonly

associated with an individual's creditworthiness such as age, gender, annual income, educational level attained, marital status, number of dependents, whether or not the respondent owns their own home, and credit score classification. The dataset had tens of thousands of cases and included a wide variety of customer profiles, which made it well-suited for creating a predictive credit scoring model. A key step in evaluating how well the model performs is splitting the entire dataset into two smaller datasets; the first being used for training purposes and the second being used for testing purposes. For this research, the dataset was split so that approximately 70 percent of all cases would be part of the training dataset and approximately 30 percent would be part of the testing dataset.

To prepare the data prior to developing the models, several pre-processing steps were taken. Prior to developing any models, the researcher coded the credit score labels into numerical values since KNN does not recognize labeled data. Next, feature scaling was applied to normalize the range of values of the continuous features in order to avoid a potential loss in model accuracy due to differences in scale among some continuous features. KNN uses distances to determine similarities between cases. Therefore, if some features have much larger ranges than others, those features will dominate the calculation of similarity and may negatively impact model performance. Finally, in order to address issues related to class imbalances present in the dataset, the researcher applied SMOTE to generate additional synthetic cases for each minority class. As such, there are now equal numbers of cases in each class, which should minimize any negative impacts on model performance due to over-representation of certain classes. In addition, in order to maintain a consistent representation of each class across both the training and test datasets, the researcher also applied stratified sampling to create these two subsets of the full dataset. There were no missing values addressed during this experiment, and all analysis was performed using the original version of this dataset.

Once pre-processing was complete, the researcher developed the KNN classifier using only the training dataset and tested its performance against only the test dataset. The performance of the model was examined through a series of common classification measures including accuracy, precision, recall, and F1-score. Accuracy represents the proportion of total cases that were predicted correctly. Precision and recall represent the proportions of true positives that were identified out of all positive classifications made by the model. An ideal balance between precision and recall can be calculated through the use of the F1-score. In addition to assessing performance using various classification measures, the researcher created a confusion matrix to illustrate the performance of the model based upon accurate versus inaccurate classification decisions regarding each possible credit score category.

All experimentation occurred in a Jupyter Notebook environment utilizing Python as a programming language. Throughout this study, the researcher used several packages including Scikit-learn to support model development and evaluation, Pandas to support data manipulation and pre-processing, NumPy to perform numeric-based computations, and Seaborn to provide data visualizations. Each package provided a powerful combination for supporting data preparation tasks, building machine learning models, performing evaluations regarding model performance, and interpreting results.

V. RESULTS AND DISCUSSION

The results of this experiment were both positive and negative in performance. Accuracy achieved a score of .88 or 88%, beating the desired threshold of 70%. However, this value alone is often misleading, and examining metrics like precision, recall, and F1-score is essential (Fig. 1).

	Age	Gender	Income	Education	Marital Status	Number of Children	Home Ownership	Credit Score
Age	1.000000	0.235343	0.699464	0.170254	-0.517723	0.055390	-0.713803	0.205362
Gender	0.235343	1.000000	0.495738	0.248671	0.278362	-0.442139	-0.031519	-0.247729
Income	0.699464	0.495738	1.000000	0.369449	-0.471004	0.084547	-0.704928	0.083698
Education	0.170254	0.248671	0.369449	1.000000	-0.067797	0.047311	-0.397043	0.334424
Marital Status	-0.517723	0.278362	-0.471004	-0.067797	1.000000	-0.696984	0.708374	-0.205756
Number of Children	0.055390	-0.442139	0.084547	0.047311	-0.696984	1.000000	-0.497129	0.136517
Home Ownership	-0.713803	-0.031519	-0.704928	-0.397043	0.708374	-0.497129	1.000000	-0.293384
Credit Score	0.205362	-0.247729	0.083698	0.334424	-0.205756	0.136517	-0.293384	1.000000

	Age	Gender	Income	Education	Marital Status	Number of Children	Home Ownership	Credit Score
Age	1.000000	0.235343	0.699464	0.170254	-0.517723	0.055390	-0.713803	0.205362
Gender	0.235343	1.000000	0.495738	0.248671	0.278362	-0.442139	-0.031519	-0.247729
Income	0.699464	0.495738	1.000000	0.369449	-0.471004	0.084547	-0.704928	0.083698
Education	0.170254	0.248671	0.369449	1.000000	-0.067797	0.047311	-0.397043	0.334424
Marital Status	-0.517723	0.278362	-0.471004	-0.067797	1.000000	-0.696984	0.708374	-0.205756
Number of Children	0.055390	-0.442139	0.084547	0.047311	-0.696984	1.000000	-0.497129	0.136517
Home Ownership	-0.713803	-0.031519	-0.704928	-0.397043	0.708374	-0.497129	1.000000	-0.293384
Credit Score	0.205362	-0.247729	0.083698	0.334424	-0.205756	0.136517	-0.293384	1.000000

Fig. 1. Classification report.

The Precision, recall, and F1-scores varied across the data. Unfortunately, Low credit scores produced an average score that could have resulted from possible errors. Both the Average and High credit scores performed well regarding precision, with the Average achieving a perfect score. All three features produced high recall scores, meaning that the model performed well in detecting positive instances while minimizing the number of false positives.

Lastly, all credit scores produced a sufficient F1-score; however, the Low credit score produced average results, which is often common in real-life situations. The F1-score balances the numerical mean of precision and recall scores; it can be assumed that because the Low credit score had a precision score of .67, this led to a lower F1-score; this interpretation may suggest that the Low credit score had an issue with miscalculating true positives for false positives.

	precision	recall	f1-score	support
0	0.67	0.91	0.77	11
1	1.00	0.85	0.92	34
2	0.83	1.00	0.91	5
accuracy			0.88	50
macro avg	0.83	0.92	0.87	50
weighted avg	0.91	0.88	0.89	50

Fig. 2. Correlation matrix.

The correlation matrix measures the coefficient of each pair of features or variables in a range from -1 to 1, where -1 indicates a perfect negative linear relationship, 0 indicates no linear relationship, and 1 indicates a perfect positive linear relationship (Fig. 2). The majority of the features indicated an extremely low linear relationship, whereas a sufficient number indicated a

perfect negative relationship. Due to the low perfect linear relationship, it can be assumed that the variables are not necessarily related to predicting credit scoring. As one feature increases, others increase but at a weak and not effectively predictable level; this leads to a lot of data being scattered in the heat map as shown below (Fig. 3).

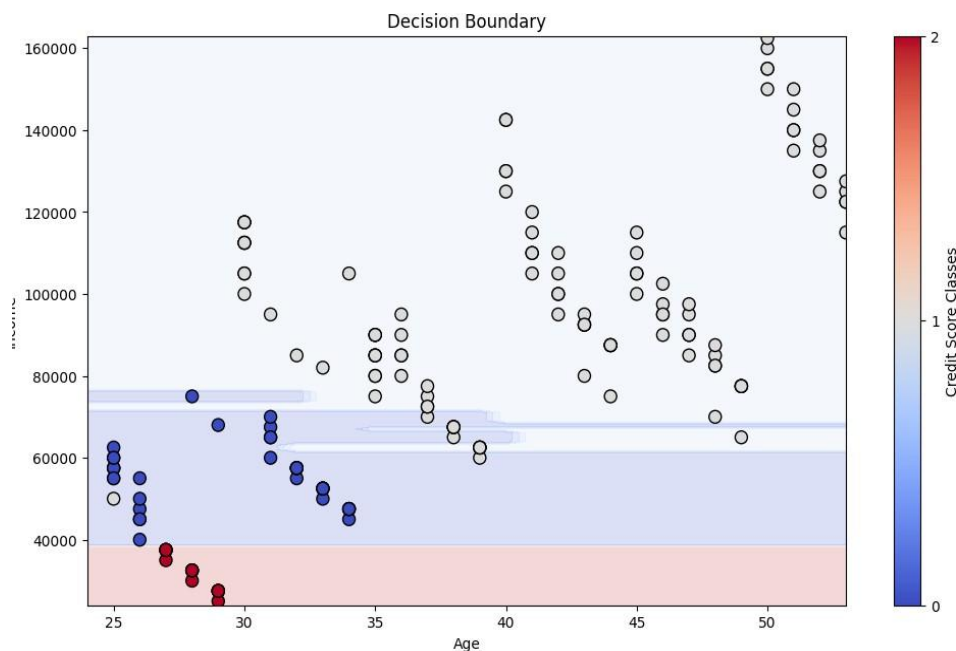


Fig. 3. Decision boundary (heat map).

VI. CONCLUSION

The study evaluated the performance of KNN on a Kaggle credit classification dataset. The 88% accuracy exceeded the target of 70%, indicating good overall predictions. However, the metrics should not be limited to this value only. Precision, recall, and F1-scores varied yet produced decent results. The Low credit score class produced lower precision and an average F1-score, indicating it had challenges in identifying minority classes. Unfortunately, the correlation matrix provided insufficient results; it revealed weak linear relationships among the features, which can lead the researcher to assume demographic and financial characteristics do not strongly predict credit scores. These findings, combined with the sensitivity of KNN to feature scaling and noise, should require further optimization. Overall, KNN demonstrated promising results, with weak correlations suggesting a different set of features or more advanced models should be used for optimal credit score classification.

REFERENCES

- [1] Geeks for Geeks. K-Nearest Neighbor (KNN) Algorithm [Internet]. Geeks for Geeks; 2025 [Accessed 2025 Dec 1]. Available from: <https://www.geeksforgeeks.org/machine-learning/k-nearest-neighbours/>.
- [2] Suyal M, Goyal P. A Review on Analysis of K-Nearest Neighbor Classification Machine Learning Algorithms based on Supervised Learning. *Int J Eng Trends Technol*. 2022. 70(7):43-48.
- [3] Xu Z. Credit scoring using alternative data sources: A machine learning approach. *Acad J Comput Inf Sci*. 2025; Available from: <https://francispress.com/uploads/papers/ATiNq1cE9tMnqqoeLcYrkS8hXVJodbqLLL1zFdN.pdf>.
- [4] Caloca Martinez JJ. Machine learning in credit scoring: A fairness approach [Master's thesis]. Warsaw School of Economics; 2022. Available from: https://www.researchgate.net/profile/Jose-Javier-Caloca-Martinez/publication/379301243_Machine-learning-in-credit-scoring-A-fairness-approach/links/66032a76907e3f27ee6b8b8d
- [5] Wang ZQ. Artificial Intelligence and Machine Learning in Credit Risk Assessment: Enhancing Accuracy and Ensuring Fairness. *Open J Soc Sci*. 2024;12:19–34. Available from: https://www.scirp.org/pdf/jss20241211_221_76
- [6] Pinheiro JM, Oliveira SVB, Silva THS, et al. The Impact of Feature Scaling In Machine Learning: Effects on Regression and Classification Tasks [Internet]. arXiv.
- [7] Fix E, Hodges JL. Discriminatory analysis, nonparametric discrimination, consistency properties. *USAF School of Aviation Medicine Report No.4*; 1951.
- [8] Mladenova T, Valova I. Classification with K-Nearest Neighbors Algorithm: Comparative Analysis between Manual and Automatic Methods for K-Selection. *Int J Adv Comput Sci Appl*. 2023;14(4).
- [9] Syriopoulos PK, Kalampalikis NG, Kotsiantis SB, Vrahatis MN. kNN Classification: a review. *Ann Math Artif Intell*. 2023;93:43–75.
- [10] Trantor Team. K Nearest Neighbor Algorithm in Machine Learning: Advantages and Limitations [Internet]. Trantor Inc.; 2025 [cited 2025 Dec 1]. Available from: <https://www.trantorinc.com/blog/k-nearest-neighbor-algorithm-in-machine-learning>.
- [11] Hand DJ, Henley WE. Statistical classification methods in consumer credit scoring: a review. *J R Stat Soc Ser A Stat Soc*. 1997;160(3):523–541.
- [12] World Bank Group. Credit scoring approaches guidelines [Internet]. Washington, DC: World Bank; 2019 [cited 2025 Dec 3]. Available from: <https://thedocs.worldbank.org/en/doc/935890130022020/original/CREDITSCORINGAPPROACHESGUIDELINESFINALWEB.pdf>.
- [13] Laborda J, Ryoo S. Feature selection in a credit scoring model. *Mathematics*. 2021;9(7):746. Available from: <https://www.mdpi.com/2227-7390/9/7/746>.
- [14] Syriopoulos PK, Kalampalikis NG, Kotsiantis SB, Vrahatis MN. kNN Classification: a review. *Ann Math Artif Intell*. 2023;93:43–75.