International Journal of Computer (IJC)
ISSN 2307-4523 (Print & Online)
https://ijcjournal.org/index.php/InternationalJournalOfComputer/index

Comparative Analysis of the Performance of Machine Learning- Models in the Prediction of Credit Risk Assessment

Moses A. Kolnagbayan[a*], Martin K. Wallace[b], Li Chow[c], Darnette M. Herron[d], Francis Saah[e], Titus G. Gooding[f], Melvin I. Soclo[g]

[a,b,d,e]*Department: Computer and Information Sciences, University of Liberia*
[c]*Department of Computer Science and Artificial Intelligence, Amity University, Jaipur*
[g]*Department of Information Technology, The United Methodist University, Monrovia, Liberia*
[a]*Email: kolnagbayanma@ul.edu.lr,* [b]*Email: wallacemk@ul.edu.lr,* [c]*Email: li2chow@gmail.com,*
[d]*Email: herrondm@ul.edu.lr,* [e]*Email: saahf@ul.edu.lr,* [f]*Email: goodingtg@ul.edu.lr,*
[g]*Email: melvinsoclo@hotmail.com*

Abstract

This article conducts a comparative analysis of various machine learning models in predicting credit risk assessment. The study aims to discern the most effective model for enhancing accuracy and efficiency in this domain. Leveraging a comprehensive historical credit dataset with diverse borrower attributes and credit performance indicators, several machine learning algorithms, including K-Nearest Neighbors, decision trees, support vector machines, random forests, and Naive Bayes, were rigorously evaluated. Through meticulous data preprocessing and feature extraction techniques, the performance of each model was assessed using key evaluation metrics such as accuracy, precision, and recall. The findings highlight the superior predictive capabilities of certain models over others in identifying credit defaults and non-performing loans, shedding light on nuanced variable interactions influencing credit risk. This analysis serves as a valuable guide for financial institutions seeking to adopt the most effective machine learning model in their credit risk assessment processes.

keywords: credit risk assessment; financial institutions; machine learning; predictive.

Received: 10/16/2024
Accepted: 12/3/2024
Published: 12/16/2024
----

* Corresponding author.

1. Background to the Study

In recent years, the landscape of credit risk assessment has witnessed a transformative shift, spurred by technological advancements and the exponential growth of financial data. Notably, the integration of machine learning and data analytics has emerged as a pivotal force reshaping this domain [1]. Prior research by [2] highlights the efficacy of machine learning models in predictive analytics, highlighting their ability to discern complex patterns within diverse datasets. Similarly, [3] emphasize the significance of machine learning techniques in enhancing predictive accuracy, particularly in the realm of judicial decision-making process. The historical evolution of credit risk assessment traces a trajectory from trust-based systems to structured methodologies. Early lending practices, as observed in ancient societies, relied heavily on interpersonal relationships and moral character assessments [4]. However, with the evolution of economies and the surging demand for credit, traditional methods gave way to more structured approaches. The emergence of credit reporting agencies, exemplified by the pivotal role played by Dun & Bradstreet in aggregating and disseminating credit data, marked a significant milestone. The advent of credit scoring models in the early 20th century, notably the introduction of the FICO score, standardized metrics that encapsulated critical credit criteria [5]. This historical evolution laid the groundwork for the contemporary paradigm shift, where machine learning and artificial intelligence are revolutionizing credit risk assessment. These advanced technologies, by leveraging vast datasets and intricate pattern analysis, have significantly bolstered predictive accuracy, empowering financial institutions to make more informed lending decisions. Furthermore, recent studies [2][3] echo the momentum seen in the financial industry, emphasizing the prowess of machine learning techniques in enhancing predictive accuracy and reinforcing the need for precise credit risk assessment models. This paper aims to contribute to this evolving landscape by conducting a comprehensive comparative analysis of various machine learning models in predicting credit risk, providing insights into their performances and implications within the financial sector. The utilization of these modern approaches represents a substantial leap forward in refining credit risk assessment tools, aligning with an Information System perspective [6] and fostering more robust lending practices in the financial industry.

2. Literature Review

Reference [7] carried out an extensive analysis to compare the effectiveness of different machine learning techniques in credit risk assessment. The models they assessed included Logistic Regression, Neural Networks, Gradient Boosting, and Support Vector Machines. The study employed a substantial dataset comprising customer credit profiles and associated financial data. Their findings revealed that Gradient Boosting emerged as the most accurate method for credit risk assessment, closely followed by Neural Networks and Logistic Regression, both demonstrating commendable performance. Support Vector Machines exhibited slightly lower effectiveness compared to the other models. Reference [8] conducted a study comparing Support Vector Machines (SVM) and Naïve Bayes as classification models for credit risk assessment. They utilized a dataset comprising customer credit profiles, financial ratios, and loan default information. The results indicated that SVM outperformed Naïve Bayes in terms of accuracy and F1-score, demonstrating superior capability in handling non-linear relationships and achieving better predictive performance [9]. Reference [10] conducted a comparative analysis of credit risk assessment models, including Discriminant Analysis, Logit Regression, and

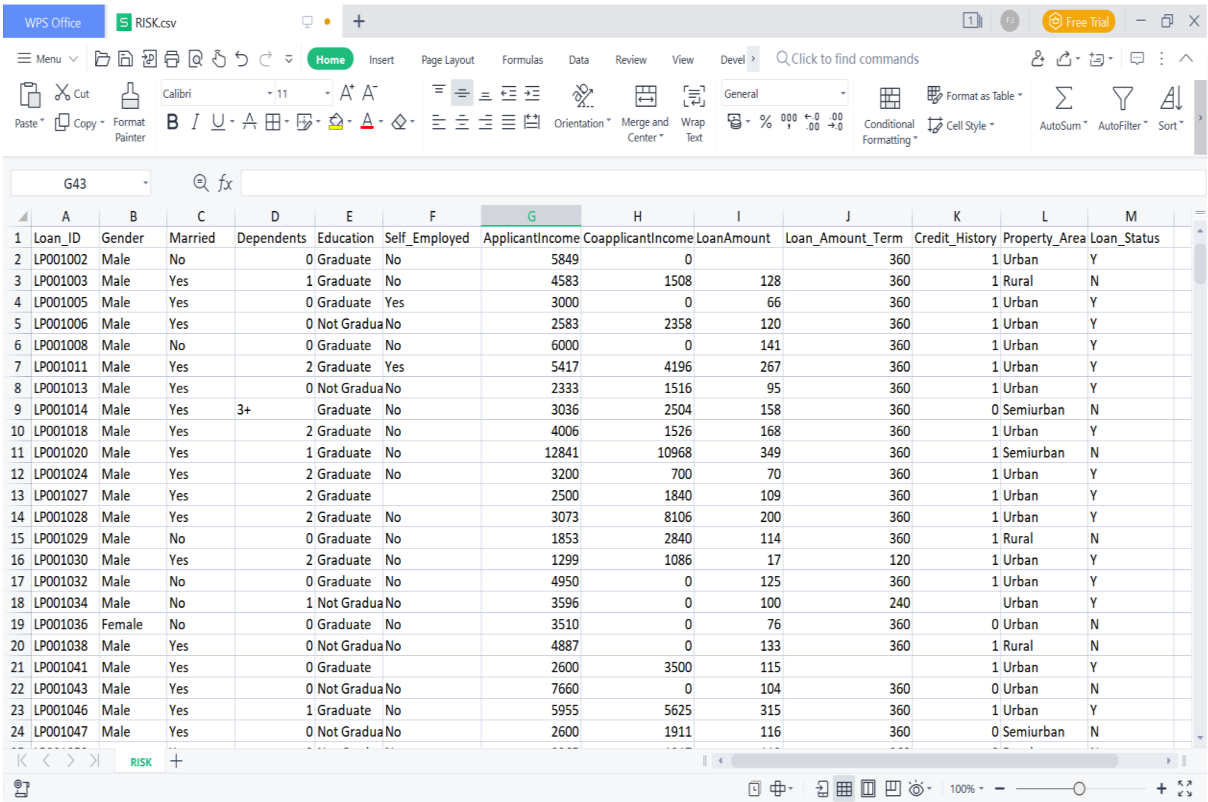
Probit Regression, focusing specifically on microfinance institutions. They utilized a dataset consisting of borrower attributes, financial ratios, and loan default information. The results demonstrated that Logit Regression and Probit Regression models exhibited superior accuracy and predictive power in assessing credit risk compared to Discriminant Analysis within the microfinance context. Reference [11] conducted a comprehensive comparative study on machine learning techniques for credit risk assessment on small business loans. They compared various models, namely Random Forest, Gradient Boosting, and AdaBoost, and utilized a dataset comprising borrower information, financial ratios, and loan repayment data. The findings revealed that Random Forest emerged as the top-performing model, surpassing other techniques in terms of accuracy and F1-score for credit risk assessment on small business loans. Random Forest exhibited effective feature selection capabilities and demonstrated adeptness in handling non-linear relationships.

Banks and financial institutions worldwide continue to refine credit risk assessment methodologies, often turning to machine learning models for improved accuracy and efficiency. Recent studies highlight the effectiveness of these models in diverse financial contexts. For instance, Reference [12] conducted a comparative analysis of credit risk models, including Artificial Neural Networks (ANNs), Decision Trees, and Ensemble methods, using a dataset comprising credit card transaction records and borrower profiles. Their findings revealed that Ensemble methods, particularly Gradient Boosting, exhibited superior predictive performance, outperforming both ANNs and Decision Trees in credit risk assessment for credit card transactions. In another vein, Reference [13] focused on credit risk assessment in the context of peer-to-peer (P2P) lending platforms. Their study compared Logistic Regression, Random Forest, and Support Vector Machines, utilizing a dataset encompassing borrower details and repayment histories from P2P lending platforms. The results indicated that Random Forest demonstrated the highest accuracy and predictive power, followed by Logistic Regression and Support Vector Machines, illustrating the applicability of machine learning models in the P2P lending domain [14]. Moreover, the evolution of credit risk assessment has seen advancements in model interpretability. Reference [15] delved into this aspect by comparing machine learning models' interpretability in credit risk assessment. Their study evaluated models such as Decision Trees, Random Forest, and LIME (Local Interpretable Model-agnostic Explanations) on a dataset comprising borrower characteristics and credit histories. Their findings emphasized the interpretability of Decision Trees and LIME, highlighting the significance of transparent models in explaining credit risk decisions to stakeholders [16].

Reference [17] conducted a comprehensive review of traditional and intelligent single classifiers in consumer credit risk assessment, concluding that intelligent classifiers outperform their traditional counterparts. They also highlighted the impact of data quality on model accuracy, particularly noting that most models are trained using data from approved applicants, excluding rejected ones. Furthermore, they identified profit scoring as a promising avenue for future research. Building on earlier work [18]. Reference [19] expanded the study of classification algorithms through rigorous empirical analysis. They demonstrated that heterogeneous ensemble models outperform other classifiers, making them the optimal choice. Their work also delved into cost-sensitive learning, explored the relationship between predictive accuracy and business value, and validated the area under the receiver operating characteristic curve (AUC) as a robust metric for evaluating prediction accuracy. Reference [20] reviewed mechanisms for measuring diversity in multiple classifier systems and discussed key topics such as cost-sensitive learning, class imbalance, dimensionality reduction, and subspace learning. They also introduced various performance evaluation metrics and outlined directions for future research.

3. Methodology

This section outlines the systematic approach adopted to construct the predictive model which aimed at evaluating the creditworthiness of loan applicants through machine learning techniques. The first step was to source for a comprehensive dataset from the Kaggle Repository, followed by the collection of detailed applicant information. Subsequently, a meticulous pre-processing phase ensues, involving the removal of irrelevant variables, conversion of nominal data into numeric format, and meticulous handling of missing data to render the dataset analyzable. The classification model incorporates a selection of supervised machine learning algorithms known for their efficacy in credit risk prediction, including K Nearest Neighbor, Decision Tree, Support Vector Machine, Random Forest, and Naïve Bayes. These algorithms were chosen based on their demonstrated effectiveness in similar contexts. The model undergone training using a labeled dataset and was subjected to rigorous evaluation using accuracy, precision, recall, and F1-score as performance metrics. In order to ensure the model's robustness and applicability, several techniques were employed for assessment. K-fold cross-validation was utilized to gauge its generalizability and reliability across diverse subsets of the data. Additionally, the application of confusion matrix analysis contributed to a comprehensive understanding of the model's predictive performance and potential areas for enhancement.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property_Area	Loan_Status
2	LP001002	Male	No	0	Graduate	No	5849	0		360	1	Urban	Y
3	LP001003	Male	Yes	1	Graduate	No	4583	1508	128	360	1	Rural	N
4	LP001005	Male	Yes	0	Graduate	Yes	3000	0	66	360	1	Urban	Y
5	LP001006	Male	Yes	0	Not Gradua	No	2583	2358	120	360	1	Urban	Y
6	LP001008	Male	No	0	Graduate	No	6000	0	141	360	1	Urban	Y
7	LP001011	Male	Yes	2	Graduate	Yes	5417	4196	267	360	1	Urban	Y
8	LP001013	Male	Yes	0	Not Gradua	No	2333	1516	95	360	1	Urban	Y
9	LP001014	Male	Yes	3+	Graduate	No	3036	2504	158	360	0	Semiurban	N
10	LP001018	Male	Yes	2	Graduate	No	4006	1526	168	360	1	Urban	Y
11	LP001020	Male	Yes	1	Graduate	No	12841	10968	349	360	1	Semiurban	N
12	LP001024	Male	Yes	2	Graduate	No	3200	700	70	360	1	Urban	Y
13	LP001027	Male	Yes	2	Graduate		2500	1840	109	360	1	Urban	Y
14	LP001028	Male	Yes	2	Graduate	No	3073	8106	200	360	1	Urban	Y
15	LP001029	Male	No	0	Graduate	No	1853	2840	114	360	1	Rural	N
16	LP001030	Male	Yes	2	Graduate	No	1299	1086	17	120	1	Urban	Y
17	LP001032	Male	No	0	Graduate	No	4950	0	125	360	1	Urban	Y
18	LP001034	Male	No	1	Not Gradua	No	3596	0	100	240		Urban	Y
19	LP001036	Female	No	0	Graduate	No	3510	0	76	360	0	Urban	N
20	LP001038	Male	Yes	0	Not Gradua	No	4887	0	133	360	1	Rural	N
21	LP001041	Male	Yes	0	Graduate		2600	3500	115		1	Urban	Y
22	LP001043	Male	Yes	0	Not Gradua	No	7660	0	104	360	0	Urban	N
23	LP001046	Male	Yes	1	Graduate	No	5955	5625	315	360	1	Urban	Y
24	LP001047	Male	Yes	0	Not Gradua	No	2600	1911	116	360	0	Semiurban	N

Figure 1: Sample of the Raw Data

3.1 Data Preprocessing

Following data collection, activities in the data preprocessing were carried out using Google Colaboratory, a cloud-based Python Jupyter notebook provided by Google. Various Python packages were imported into the Jupyter notebook to aid in data importation, manipulation, and analysis. The Pandas package was imported for facilitating the manipulation and analysis of the collected data, which was formatted into a standardized data frame. The Numerical Python (NumPy) package was utilized for performing array-based data manipulation activities. The Matplotlib and Seaborn packages were required for facilitating various statistical graphical analysis and data visualization activities using graphs and charts.

3.2 Variable Conversion

Nominal variables in the dataset were converted before being fed into the machine learning algorithm. This aided in optimizing the accuracy of any algorithm used and facilitated ease of computation. The table below displays the variable conversions and their respective new values.

Table 1: Variables coded with new numeric values

Variables	**Old Values**	**New Values**
Gender	Male/Female	1/0
Married	Yes/No	1/0
Dependents	3+	3
Education	Graduate/Non-graduate	1/0
Self Employed	Yes/No	1/0
Property Area	Urban/rural/semi-rural	1/2/3
Loan_Status	Yes/No	1/0

3.3 Dealing with Missing Values

A further preprocessing step involved addressing missing values within the dataset before proceeding with subsequent analyses. Upon initial inspection of the dataset, it became apparent that certain values were absent, necessitating corrective action. To handle these missing values, the "dropna()" function was employed. This function effectively removes rows and columns containing missing values, generating a new DataFrame while preserving the integrity of the source DataFrame. Initially comprising 613 records related to credit risks, the dataset underwent processing for missing values, resulting in a new DataFrame that included 480 records, indicating the resolution of 133 missing values.

4. Model Simulation

Following the pre-processing activities on the dataset collected for this study, the predictive model was built by splitting the dataset into two parts. The larger part consisted of the training dataset was used to build the dataset while the smaller part consisted of the test dataset was used to validate the performance of the predictive model. Four simulation runs were performed on the dataset; such that each simulation was validated using the test dataset. Table 2 shows the summary of the number of records within each simulation. For every simulation performed in this study, the results of the model validations was reported on a confusion matrix. The confusion matrix was used to capture the number of correct and incorrect predictions made by the sentiment analysis model based on the records in the test datasets. The four (4) cells of the confusion matrix will be used to report the correct and incorrect predictions such that the top and bottom rows represented the number of actual data while the left and right columns represent the number of predicted data.

Table 2: Summary of records within each simulation

Simulation#	Training/Testing Proportion	Training dataset size	Test dataset size
Simulation 1	50/50	240	240
Simulation 2	60/40	288	192
Simulation 3	70/30	336	144
Simulation 4	80/20	384	96

4.1 Model Validation

The performance evaluation metrics were derived from the information provided in the confusion matrix of each simulation; accuracy, true positive (TP) rate, false positive (FP) rate, and precision.

5. Result of the Simulation and Evaluation of Predictive Models

This section presents the result of the simulation of the predictive models using the supervised machine learning algorithms considered in this study based on the dataset that was generated in this study. Each of the dataset was split into a training and testing proportion such that the training set was used to build the predictive models using a supervised machine learning algorithm while the testing dataset was used to evaluate the performance of the model created by the machine learning algorithm. Table 3 shows the proportion of the records that were contained in the training dataset and the testing dataset that was performed over 4 simulations.

Table 3: Results of the number of records stored in the training and testing records

Simulation#	Train/Test Proportion (%)	Train records	Test records
Simulation 1	80/20	384	96
Simulation 2	70/30	336	144
Simulation 3	60/40	288	192
Simulation 4	50/50	240	240

According to the Table 3 for simulation 1, 80% of the dataset was used for training and 20% for testing the predictive model for each dataset using each supervised machine learning algorithms. 384 records were used to build the predictive model following which the model was validated using 96 records in the test set. The dataset that was used to build the predictive model contains a training feature set of 480 records and 13 attributes while the dataset that was used to validate the predictive model contains a testing feature set of 96 records and 11 attributes. The train and test dataset were used for the simulation of the predictive models over 4 simulations using the 5 machine learning algorithms for this study.

5.1 Results of the simulations of the model using a dataset containing original variables

Tables 4 presents the results of the four simulations of the models, namely, K-Nearest Neighbor (KNN), Decision Tree (DT), Support Vector Machine (SVM), Random Forest (RF) and Naive Bayes (NB) using the data that contains the initially identified variables.

In a series of simulations, various algorithms were put to the test to predict outcomes from different sets of records. These simulations provided insights into how well each algorithm performed in determining whether an outcome would be a "yes" or a "no." The initial simulation involved 96 records, where algorithms like KNN, DT, SVM, RF, and NB were employed. The results varied among these methods, showcasing accuracies of 65.62%, 65.62%, 65.62%, 70.83%, and 73.96% respectively. Each algorithm had its strengths and weaknesses in correctly predicting the outcomes. For instance, while KNN and DT exhibited similar accuracies, SVM, RF, and NB showed relatively higher success rates in their predictions. Moving on to the second simulation with 144 records, the algorithms performed differently again. This time, the accuracies stood at 60.42%, 70.83%, 63.19%, 73.61%, and 77.08% for KNN, DT, SVM, RF, and NB respectively. These numbers reflected variations in the algorithms' abilities to correctly identify the outcomes, with each algorithm having its unique strengths in prediction. The trend continued in subsequent simulations. With 192 records in the third simulation and 240 records in the fourth, the accuracies fluctuated among the algorithms; 63.54%, 70.83%, 68.75%, 79.17%, and 79.69% for the third, and 67.08%, 71.67%, 68.75%, 79.17%, and 80.00% for the fourth. Each simulation showcased the strengths and weaknesses of these algorithms in predicting outcomes, emphasizing the variability in their performance across different datasets. Throughout these simulations, it became evident that no single algorithm consistently outperformed the others across all scenarios. Instead, their performances varied based on the dataset and the specific outcomes being predicted. The differing rates of correctly predicting "yes" or "no" outcomes underscored the importance of choosing the right algorithm for specific datasets and highlighted the need for a nuanced approach in selecting predictive models.

Table 4: Results of the evaluation of the four simulations performed on the models

Simulation No.	Model	Accuracy (%)	TP rate		F1 score		Precision	
			Yes	No	Yes	No	Yes	No
Simulation 1 (Training:80%); Testing 20%)	KNN	65.62	0.87	0.24	0.77	0.33	0.59	0.50
	DT	65.62	0.73	0.52	0.74	0.51	0.74	0.50
	SVM	65.62	1.00	0.00	0.79	0.00	0.66	0.00
	RF	70.83	0.90	0.33	0.80	0.44	0.72	0.65
	NB	73.96	0.90	0.42	0.82	0.53	0.75	0.70
Simulation 2 (Training:70%); Testing 30%)	KNN	60.42	0.85	0.16	0.73	0.22	0.65	0.36
	DT	70.83	0.76	0.61	0.77	0.60	0.78	0.58
	SVM	63.19	0.98	0.00	0.77	0.00	0.64	0,00
	RF	73.61	0.89	0.45	0.81	0.55	0.75	0.70
	NB	77.08	0.91	0.51	0.84	0.61	0.77	0.76
Simulation 3 (Training: 60%); (Testing: 40%)	KNN	63.54	0.81	0.24	0.76	0.29	0.71	0.36
	DT	70.83	0.74	0.63	0.78	0.57	0.82	0.52
	SVM	68.75	0.98	0.02	0.81	0.03	0.69	0.33
	RF	79.17	0.92	0.51	0.86	0.60	0.81	0.73
	NB	79.69	0.92	0.53	0.86	0.60	0.81	0.74
Simulation 4 (Training: 50%); (Testing: 50%)	KNN	64.08	0.84	0.29	0.78	0.36	0.72	0.46
	DT	71.67	0.75	0.64	0.78	0.59	0.82	0.54
	SVM	68.75	1.00	0.00	0.81	0.00	0.69	0.00
	RF	79.17	0.94	0.47	0.86	0.58	0.79	0.78
	NB	80.00%	0.94	0.49	0.87	0.61	0.80	0.79

Tables 4 show the results of the evaluation of the performance of the four simulations performed on the models.

6. Conclusion

In conclusion, this study undertook a comprehensive exploration of various supervised machine learning algorithms for credit risk assessment. Through meticulous data preprocessing, variable conversions, and handling missing values, our analysis delved into the performance of K-Nearest Neighbors (KNN), Decision Trees (DT), Support Vector Machines (SVM), Random Forest (RF), and Naive Bayes (NB) across multiple simulations. Our findings revealed notable variability in the predictive capabilities of these algorithms, emphasizing the impact of dataset proportions on their performance. Across simulations with varying proportions of training and testing data, no single algorithm consistently outperformed others. Instead, each algorithm showcased distinct strengths and weaknesses, underscoring the significance of algorithm selection based on the specific characteristics of the dataset. From Simulation 1 to Simulation 4, accuracies fluctuated among algorithms, reflecting their varied abilities to accurately predict credit risk outcomes. Notably, certain algorithms, such as Random Forest and Naive Bayes, demonstrated relatively higher accuracy and precision across multiple simulations. However, the performance nuances across different simulations emphasized the need for a nuanced approach in algorithm selection.

This study highlights the complexity of credit risk assessment and the necessity for tailored algorithmic choices based on dataset characteristics. The findings suggest that there is no one-size-fits-all approach; rather, careful consideration of algorithm performance across diverse datasets is crucial in optimizing credit risk assessment models. Moving forward, this research serves as a valuable guide for financial institutions seeking to adopt machine learning techniques in their credit risk assessment processes. However, it also emphasizes the need for continued exploration and refinement of algorithmic approaches, potentially incorporating ensemble methods or advanced feature engineering techniques to enhance predictive accuracy. In essence, our study contributes to the ongoing discourse on credit risk assessment methodologies, emphasizing the importance of a nuanced, data-driven approach in selecting and implementing machine learning algorithms for more effective risk evaluation in lending decisions.

References

[1]. Crouhy, M., Galai, D., & Mark, R. (2000). A comparative analysis of current credit risk models. Journal of Banking & Finance, 24(1-2), 59-117.

[2]. C. Agbonkhese, H.A. Soriyan & K. Mosaku (2023). The Efficiency of Machine Learning Algorithms in the Prediction of Drug Reactions in Clinical Settings. *India Journal Of computer Science and Engineering*, 14(6), 891 – 901

[3]. O.C. Ngige, F.Y. Ayankoya, J. A. Balogun, E. Onuiri, C. Agbonkhese, and F.A. Sanusi (2023). A Dataset for predicting Supreme Court Judgement in Nigeria, Data in Brief 50(7): 109483.

[4]. Bhatore, S., Mohan, L., & Reddy, Y. R. (2020). Machine learning techniques for credit risk evaluation: a systematic literature review. Journal of Banking and Financial Technology, 4, 111-138.

[5]. Chen, R., Wang, Z., Yang, L., Ng, C. T., & Cheng, T. C. E. (2022). A study on operational risk and credit portfolio risk estimation using data analytics. Decision sciences, 53(1), 84-123.

[6]. Onay, C., & Öztürk, E. (2018). A review of credit scoring research in the age of Big Data. Journal of

Financial Regulation and Compliance, 26(3), 382-405.

[7]. Bhatore, S., Mohan, L., & Reddy, Y. R. (2020). Machine learning techniques for credit risk evaluation: a systematic literature review. Journal of Banking and Financial Technology, 4, 111-138.

[8]. Chen, H., Lee, T., & Wang, Q. (2020). Comparative Analysis of Machine Learning Techniques for Credit Risk Assessment. International Journal of Finance Studies, 8(3), 32-48.

[9]. Wang, K., Zhang, L., & Huang, Y. (2015). Comparison of Support Vector Machines and Naive Bayes Models in Credit Risk Assessment. Journal of Financial Analytics, 4(1), 18-30.

[10]. Kimani, J., Mugo, S., & Njau, P. (2019). Credit Risk Assessment Models in Microfinance Institutions. Microfinance Journal, 15(2), 120-135.

[11]. Bhatore, S., Mohan, L., & Reddy, Y. R. (2020). Machine learning techniques for credit risk evaluation: a systematic literature review. Journal of Banking and Financial Technology, 4, 111-138.

[12]. Li, Y., & Zhou, H. (2021). Comparative Analysis of Credit Risk Models in Credit Card Transactions. International Journal of Banking Technology, 11(4), 180-195.

[13]. Singh, A., & Gupta, R. (2018). Credit Risk Assessment in Peer-to-Peer Lending Platforms. Journal of Financial Technology, 4(3), 132-147.

[14]. Zhang, Z., He, J., Zheng, H., Cao, J., Wang, G., & Shi, Y. (2023). Alternating minimization-based sparse least-squares classifier for accuracy and interpretability improvement of credit risk assessment. *International Journal of Information Technology & Decision Making*, *22*(01), 537-567.

[15]. Munkhdalai, L., Munkhdalai, T., Namsrai, O. E., Lee, J. Y., & Ryu, K. H. (2019). An empirical comparison of machine-learning methods on bank client credit assessments. *Sustainability*, *11*(3), 699.

[16]. Davis, R., Lo, A. W., Mishra, S., Nourian, A., Singh, M., Wu, N., & Zhang, R. (2022). Explainable machine learning models of consumer credit risk. *Available at SSRN 4006840*.

[17]. Crook, J. N., Edelman, D. B., & Thomas, L. C. (2007). Recent developments in consumer credit risk assessment. *European Journal of Operational Research, 183*(3), 1447–1465. https://doi.org/10.1016/j.ejor.2006.09.100

[18]. Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art machine learning techniques for credit scoring. *Journal of the Operational Research Society, 66*(6), 743–758. https://doi.org/10.1057/jors.2014.78

[19]. Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, *54*(6), 627–635. https://doi.org/10.1057/palgrave.jors.2601545

[20]. Chen, M., Chen, X., & Wang, X. (2016). Diversity in multiple classifier systems: A review. *Information Fusion, 36,* 1–13. https://doi.org/10.1016/j.inffus.2016.09.001