Ahliya Journal of Business Technology and MEAN Economies [Published by Palestine Ahliya University]
Vol. 01 Issue 02 (2024) ISSN: 3007-9691

Stock Pledge Defaults Prediction Using Machine Learning

Hala Shaheen[1], Sadi Irziqat[2], Elias Al-Dardour[3]
[1] Faculty of Engineering and Information Technology, Palestine Ahliya University (Palestine)
✉ Halashahin98@gmail.com
[2] Faculty of Administrative and Finance Sciences, Palestine Ahliya University (Palestine)
✉ sadi@paluniv.edu.ps
[3] Ez-Zitouna University (Tunisia)
✉ elyes.dardour@yahoo.com

Received:23/10/2024 Accepted:13/11/2024 Published:15/12/2024

Abstract:

This study addresses the prediction of stock pledge defaults using machine learning techniques, a crucial topic in the field of economics and finance. Specifically, the study aims to develop models capable of identifying the risks associated with stock pledge defaults, thereby enabling investors and lenders to make informed financial decisions. The dataset used consists of 1,442 entries and 62 columns, including financial indicators such as pledge ratios, stock volatility, and returns. The results showed that using machine learning algorithms like K-Nearest Neighbors (KNN), Support Vector Classifier (SVC), and Decision Tree (DT) resulted in high accuracy in predicting defaults. These models demonstrated excellent performance, achieving accuracy rates of up to 99% in some cases, reflecting the ability of these techniques to handle complex financial data more effectively than traditional methods. This study contributes to the literature on financial risk prediction by providing an advanced framework that utilizes machine learning techniques, thus enhancing the understanding of the factors influencing stock pledge defaults. The study also highlights the importance of addressing data imbalance, opening new avenues for future research in this field.

Keywords: *Default Prediction; Machine Learning; Stock Pledge; Classification Algorithms; Risk Management; Financial Data Analysis.*

Ahliya Journal of Business Technology and MEAN Economies [Published by Palestine Ahliya University]
Vol. 01 Issue 02 (2024) ISSN: 3007-9691

التنبؤ بالتعثر عن سداد تعهدات الأسهم باستخدام التعلم الآلي

هلا شاهين[1]، سعدي ارزيقات[2]، إلياس الدردور[3]

[1] كلية الهندسة وتكنولوجيا المعلومات، جامعة فلسطين الأهلية (فلسطين)
Halashahin98@gmail.com
[2] كلية العلوم الإدارية والمالية، جامعة فلسطين الأهلية، (فلسطين)
sadi@paluniv.edu.ps
[3] جامعة الزيتونة (تونس)
elyes.dardour@yahoo.com

تاريخ الاستلام:2024/10/23 تاريخ القبول:2024/11/13 تاريخ النشر:2024/12/15

ملخص:

تتناول هذه الدراسة التنبؤ بالتعثر عن سداد تعهدات الأسهم باستخدام تقنيات التعلم الآلي، وهو موضوع بالغ الأهمية في مجال الاقتصاد والتمويل. وعلى وجه التحديد، تهدف الدراسة إلى تطوير نماذج قادرة على تحديد المخاطر المرتبطة بالتعثر عن سداد تعهدات الأسهم، وتمكين المستثمرين والمقرضين من اتخاذ قرارات مالية مستنيرة. تتكون مجموعة البيانات المستخدمة من (1442) صفًا و(62) عمودًا، بما في ذلك المؤشرات المالية مثل نسب التعهدات وتقلب الأسهم والعوائد. أظهرت النتائج أن استخدام خوارزميات التعلم الآلي مثل أقرب جيران (KNN) ومصنف متجه الدعم (SVC) وشجرة القرار (DT) أدى إلى دقة عالية في التنبؤ بالتخلف عن السداد. أظهرت هذه النماذج أداءً ممتازًا، حيث حققت معدلات دقة تصل إلى (99٪) في بعض الحالات، مما يعكس قدرة هذه التقنيات على التعامل مع البيانات المالية المعقدة بشكل أكثر فاعلية من الطرق التقليدية. تسهم هذه الدراسة في الأدبيات المتعلقة بالتنبؤ بالمخاطر المالية من خلال توفير إطار متقدم يستخدم تقنيات التعلم الآلي، وبالتالي تعزيز فهم العوامل المؤثرة على التعثر عن سداد تعهدات الأسهم. وبهذا تسلط الدراسة الضوء أيضًا على أهمية معالجة اختلال التوازن في البيانات، مما يفتح آفاقًا جديدة لأبحاث مستقبلية في هذا المجال.

الكلمات المفتاحية: *التنبؤ الافتراضي؛ التعلم الآلي؛ تعهد الأسهم؛ خوارزميات التصنيف؛ إدارة المخاطر؛ تحليل البيانات المالية.*

1. Introduction

Financial predictions are a vital topic in economics, and recent years have seen significant advancements due to machine learning techniques. These techniques are increasingly being used to analyze complex financial data, including stock pledge default predictions. Investors and lenders face significant challenges in identifying risks associated with potential defaults in stock pledges, making the need for accurate predictive methods more crucial than ever (Liu et al., 2023).

Machine learning algorithms, such as Support Vector Machines (SVM) and Neural Networks, have improved predictions related to financial markets, including stock pledge defaults. By using classification algorithms, these tools can predict financial failure or default before it occurs, helping improve economic decision-making and reducing risks for financial institutions (Nazareth & Reddy, 2023). Predictive analysis using machine learning is an effective means of identifying factors that influence the likelihood of a stock pledge default, allowing these models to handle large volumes of financial and economic data with greater accuracy and efficiency than traditional methods. Stock pledge default prediction models require handling various dynamic factors such as historical stock prices, trading volumes, and macroeconomic factors like interest rates and economic growth. Recent studies have shown that integrating these factors into a single model can significantly improve predictive performance (Xia et al., 2024).

It is important to note that data imbalance poses a challenge in developing these models. Default cases, such as those related to stock pledges, tend to be rare, which can degrade model performance if not addressed. Therefore, techniques like data rebalancing and boosting are used to improve model performance in these situations (Song et al., 2023). Additionally, integrating economic data with stock data is an innovative approach that enhances the accuracy of predictive models, contributing to more precise predictions of the risks associated with stock pledge defaults (Behera et al., 2024).

In conclusion, this study provides an advanced framework for using machine learning techniques to predict stock pledge defaults, representing an important step toward improving predictive models in finance and risk management.

2. Literature Review:

The prediction of corporate default risks has become a critical topic in finance, particularly after the global financial crisis of 2008. The crisis, which began with the collapse of Lehman Brothers, highlighted the need for developing accurate models to predict companies at risk of financial distress. Recent developments in machine learning (ML) have significantly contributed to improving these models.

Hanstan et al. (2024) reviewed the use of machine learning models such as XGBoost, Support Vector Machines (SVM), and Random Forests to predict corporate defaults. By analyzing financial data from the Taiwanese Economic Index from 1999 to 2009, they identified key variables influencing default risk, such as after-tax interest rates, profitability over the past four quarters, and the debt-to-equity ratio. The results highlighted the importance of profitability and obligations in determining default risk, enabling proactive measures for management and investors.

Rahayo et al. (2022) explored the gap between economic theory and machine learning by comparing the Merton model with Case-Based Reasoning (CBR). The Merton model calculates default probability by comparing expected assets with projections, while CBR relies on recognizing patterns in data. The study revealed the strengths and weaknesses of each approach and suggested a hybrid model combining both. The integrated model achieved an accuracy rate of 84%, with a Type II error rate of 8%, supporting its effectiveness.

Kim et al. (2021) investigated the use of "geometric delay variables" in predicting corporate defaults using machine learning techniques. These variables are based on geometrically weighted averages of time series financial data, which reduces computation time and improves predictive performance. The results showed that actual default rates increase with higher expected default probabilities, confirming the model's effectiveness in supporting investment decisions.

Incerti et al. (2022) presented a comparative study using Adaptive Bayesian Regression Trees with missing data (BART-MIA) to predict corporate defaults in the Netherlands and Italy. The study demonstrated the model's ability to handle structural differences in financial data between the two countries and achieve superior performance compared to traditional models, reinforcing the use of Bayesian models in international settings.

Aliyaj et al. (2020) conducted a study using machine learning techniques to predict corporate defaults, relying on a large-scale credit database from Italy's Central Credit Register. The results showed that ensemble learning techniques, particularly Random Forests, provided the best performance in predicting defaults. This study supports previous findings on the effectiveness of these models in identifying companies at risk of default.

Collectively, these studies indicate that the use of machine learning techniques combined with economic models can enhance the accuracy of default risk predictions, helping companies and investors make more informed and proactive decisions.

3. Methodology:

Equity pledge financing is one of the commonly used financing methods in publicly listed companies, particularly those with clear ownership structures of their shares, which contributes to good liquidity and fair pricing of these stocks. Due to these characteristics, equity pledging is more accepted by financial institutions, as it can effectively meet the financing needs of shareholders in listed companies. However, at the same time, this type of financing faces high credit risks, necessitating the urgent need for effective mechanisms to manage these risks. Therefore, this study focuses on examining how to manage the risks associated with equity pledge financing and evaluating the effectiveness of predictive models in mitigating these risks.

3.1 Data Description:

The dataset used in this research includes information regarding the default status of equity pledge financing for controlling shareholders in Chinese publicly listed companies from 2017 to 2022. The dataset contains 1,442 entries and 62 columns, including significant financial data such as pledge ratios, stock volatilities, returns, and other financial indicators that may affect a company’s ability to repay financing. The target column, "IsDefault," contains two values: 1 (default) and 0 (normal), reflecting the default status in equity pledge financing.

3.2 Preliminary Data Analysis:

The data was examined using the data.info() function, which showed that the dataset contains 1,442 entries and 62 columns. During this phase, data cleaning was performed, where certain columns, such as "P/E ratio," were adjusted by removing commas and converting values to the appropriate data type (float). Initial operations revealed no missing values in the dataset, facilitating further data processing.

3.3 Data Processing Procedures:

Regarding data processing, it was confirmed that there were no missing values using the data.isna().sum() function. To address the imbalanced classes between defaulted companies (class 1) and non-defaulted companies (class 0), "upsampling" techniques were applied to balance the classes. Subsequently, the data was split into two sets: a training set (80%) and a test set (20%) using the train_test_split function from the sklearn library, ensuring that the split was stratified across the target variable (stratify=y).

3.4 Feature Encoding and Scaling:

Appropriate encoding techniques such as target encoding were used for columns with non-numeric values, such as the "P/E ratio" column. Additionally, feature scaling techniques like standardization or normalization were applied to improve model performance, as these techniques effectively handle features with different ranges.

3.5 Machine Learning Algorithms

3.5.1 K-Nearest Neighbors

The KNN algorithm uses the concept of “feature similarity” to predict the values of the latest data instances, meaning that a new data point will be assigned a value based on how well it matches the data points in the training set. The KNN algorithm is classified as a supervised machine learning technique capable of tackling both classification and regression problems. The key parameter associated with KNN is called K, which denotes the number of neighbors used to determine the classification of an unlabeled data instance

Below is an illustration of the basic K-nearest neighbors’ model that aims to make it easier for us to understand the methodology used by the model to generate predictions. The user specifies a specific value for K. In this particular case, we choose K = 3 neighbors surrounding the green dot. The goal is to identify the K observations within the dataset that are in closest proximity to unknown sample features (Wang et al., 2023).

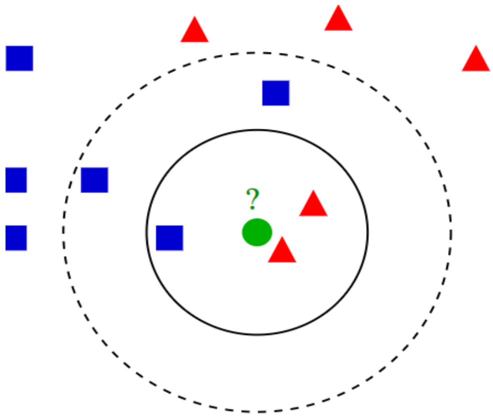

Figure (1): KNN Classifier

3.5.2 Logistic Regression (LR)

Logistic regression constitutes a method of supervised learning. Despite the inclusion of the term regression in its nomenclature, logistic regression models are primarily employed to address classification challenges.

Logistic regression employs a sigmoid function that constrains the output within the interval of 0 to 1. In scenarios where the outcome of interest is binary, this mechanism facilitates the output of the logistic regression to represent the probabilities associated with a specific class. These probabilities are amenable to conversion into class predictions. Through the application of this function, one is able to transform a linear regression model into a classification model, exemplified by logistic regression.

Consequently, when implementing a logistic regression model to forecast defaults on stock covenants, the probability of default is computed based on the provided input characteristics. Should the predicted value (probability) meet or exceed the threshold of 0.5, the subject is categorized as defaulting (class 1). Conversely, if the predicted value falls below 0.5, the subject is categorized as not defaulting (class 0).

3.5.3 Decision Tree

The Decision Tree methodology is capable of addressing both classification and regression challenges; however, its predominant application lies within the realm of classification tasks. This model is characterized by a branching structure, wherein the internal nodes signify the attributes of a dataset, the branches denote the decision-making criteria, and each terminal node indicates the resultant decision. The nomenclature of this algorithm as a Decision Tree is derived from its structural resemblance to a tree, commencing with a root node that proliferates into additional branches, culminating in a tree-like configuration of decisions (Alam, 2022).

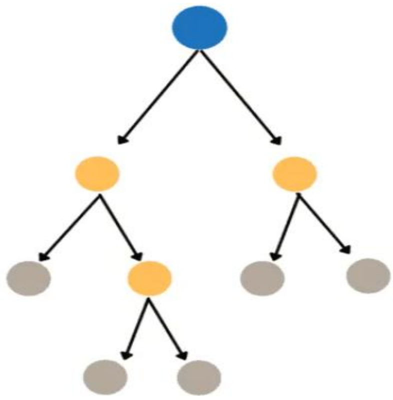

Figure (2): Decision tree structure

3.5.4 Support Vector Machine

It is a supervised learning algorithm used in the fields of classification and regression, and aims to determine the optimal interval that separates data into homogeneous classes by maximizing the distance between different classes. SVM is based on the principle of dividing data into regions based on the similarity between records, allowing new records to be classified into the region whose features are similar to those of previously classified records. SVM is a flexible technique that has been developed and explored extensively in many machine learning applications. The following figure shows a visual representation of how the SVM algorithm works (Hasni et al., 2024).

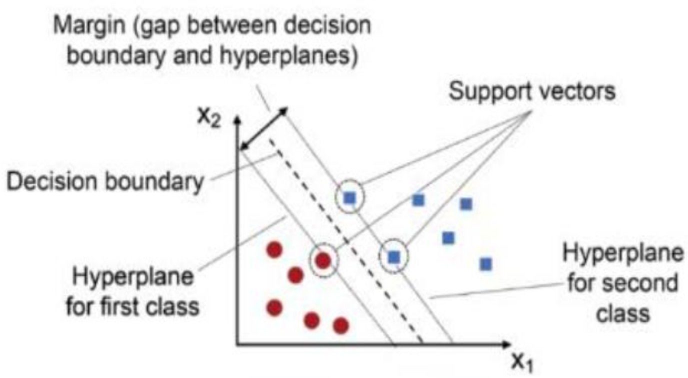

Figure (3): SVM Algorithm, (Kumar, 2023).

3.5.5 Naïve Bayes

The Naïve Bayes algorithm is a supervised, probabilistic learning algorithm based on Bayes' theorem. It is called "Naïve" because it assumes that all features are independent of each other. The algorithm is based on calculating the probability of each class based on the input features, and then selecting the class with the highest probability. This algorithm works efficiently on classification problems, especially when the relationship between features is relatively simple (Putri et al., 2022).

3.6 Model Building:

A variety of machine learning algorithms were used to develop the model for predicting equity pledge defaults. These algorithms included: K-Nearest Neighbors (KNN), Support Vector Classifier (SVC), Logistic Regression (LR), Decision Tree (DT), and Naïve Bayes (NB). This diversity reflects a mix of linear and nonlinear models, as well as probabilistic models, allowing for a comprehensive evaluation of model performance based on the nature and characteristics of the data. To ensure optimal performance, the hyperparameters of each algorithm were fine-tuned using grid search, a popular method for model optimization. For the SVC algorithm, the RBF kernel was chosen to provide a nonlinear representation of the relationships in the data, with the best hyperparameters being C = 1 (to balance classification error and model generalization) and gamma = 1 (to control the influence of data points on decision boundaries). In the KNN algorithm, experiments revealed that the optimal settings included n_neighbors = 70 for the number of neighbors, Euclidean distance metric = uclidean, and distance-based weights = distance to improve model accuracy. For the Decision Tree algorithm, the best settings involved using the entropy criterion for measuring the quality of splits, with max_depth = 10, min_samples_leaf = 1, and min_samples_split = 2. Careful tuning of the hyperparameters enhanced the performance of the algorithms, allowing for accurate and effective prediction of equity pledge defaults.

3.6 Model Evaluation:

The models were evaluated using multiple metrics to ensure prediction accuracy. Among these metrics, accuracy (the percentage of correct classifications) was measured. Additionally, a confusion matrix was used to understand the distribution between true and predicted classes and assess the model's ability to predict default cases accurately. Furthermore, the F1 score was used to balance precision and recall, particularly in cases where the class distribution was imbalanced.

The performance of the different classification algorithms was evaluated using a set of common metrics that express the quality of the model's performance and accuracy. These metrics include:

- Accuracy: The percentage of accurate predictions produced by the model out of all predictions.
- Precision: Focuses on how reliable the model is in predicting the positive class.
- Recall: The percentage of all samples belonging to the positive class (true positive cases) that the model correctly classified out of the total positive cases (true positive cases + false negative cases).
- F1-Score: This score indicates how effectively the model achieves a balance between recall and accuracy in the confusion matrix. The figure shows the formulas used to calculate precision, recall, and accuracy from the confusion matrix.

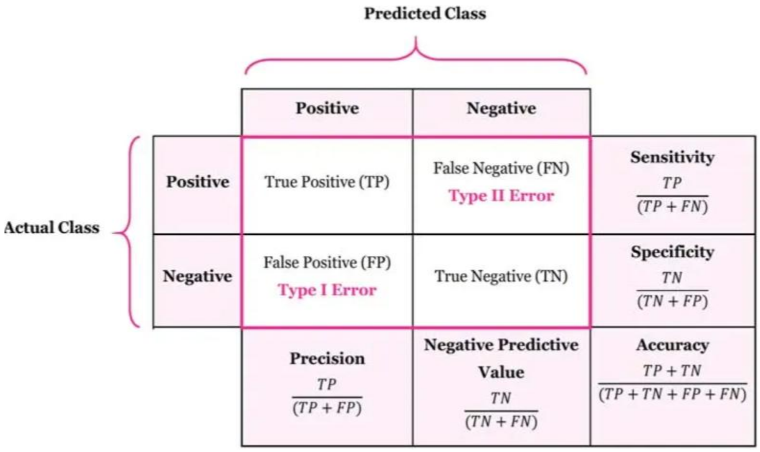

Figure (4): Mathematical formulas for performance analysis using confusion matrix.

4. Results:

The performance of five machine learning models—KNN, SVC, Logistic Regression (LR), Decision Tree (DT), and Naive Bayes (NB)—was analyzed for predicting equity pledge defaults. The models were evaluated using a comprehensive set of metrics, including accuracy, precision, recall, and the F1 score. The detailed results of these evaluations are as follows:

Table (1): Comparison of Machine Learning Classification Algorithms on the Dataset

		KNN	SVC	LR	DT	NB
Accuracy	Train	1.0	1.0	0.869	1.0	0.77
	Test	0.94	0.98	0.868	0.99	0.78
Precision	Train	1.0	1.0	0.85	1.0	0.74
	Test	0.90	1.0	0.88	0.99	0.78
Recall	Train	1.0	1.0	0.89	1.0	0.82
	Test	1.0	0.97	0.85	0.99	0.79
F1 Score	Train	1.0	1.0	0.87	1.0	0.78
	Test	0.94	0.98	0.86	0.99	0.79

The KNN, SVC, and Decision Tree models achieved perfect accuracy on the training data, with scores of (1.0), reflecting high performance for these models. However, the accuracy of the Logistic Regression model (0.869) was relatively lower, while Naive Bayes demonstrated the weakest performance with an accuracy of (0.770).

On the test data, the Decision Tree model continued to excel with an accuracy of (0.99), followed by SVC at (0.98). The other models performed at lower levels, with Logistic Regression at (0.868) and Naive Bayes at (0.78).

The KNN, SVC, and Decision Tree models demonstrated exceptional positive predictive values on the training data, with an ideal score of (1.0), indicating their ability to avoid false positive predictions. In contrast, the Logistic Regression and Naive Bayes models had lower rates of (0.85) and (0.74), respectively.

On the test data, the SVC model excelled with a perfect positive predictive value of (1.0), followed by the Decision Tree at (0.99), while Logistic Regression and Naive Bayes recorded lower rates of (0.88) and (0.78), respectively.

In terms of recall on the training data, the KNN, SVC, and Decision Tree models performed ideally with a score of (1.0), reflecting their ability to retrieve all positive cases. Logistic Regression and Naive Bayes had lower recall rates of (0.89) and (0.82), respectively.

On the test data, the KNN model continued to demonstrate ideal recall (1.0), while the Decision Tree achieved (0.99) and SVC (0.97). Logistic Regression and Naive Bayes showed lower recall values at (0.85) and (0.79), respectively.

The F1 score reflects the balance between precision and recall. The KNN, SVC, and Decision Tree models demonstrated perfect F1 scores of (1.0) on the training data, while Logistic Regression and Naive Bayes achieved F1 scores of (0.87) and (0.78), respectively.

On the test data, the Decision Tree model led with an F1 score of (0.99), followed by SVC at (0.98). Logistic Regression and Naive Bayes achieved lower F1 scores of (0.86) and (0.79), respectively.

The results highlight the superior performance of the Decision Tree and SVC models in terms of overall performance on the test data, making them strong candidates for predicting mortgage default. In contrast, the perfect performance of the other models on the training data, such as KNN, suggests a potential overfitting issue. The results recommend further analysis to avoid overfitting and achieve stable performance on real-world data.

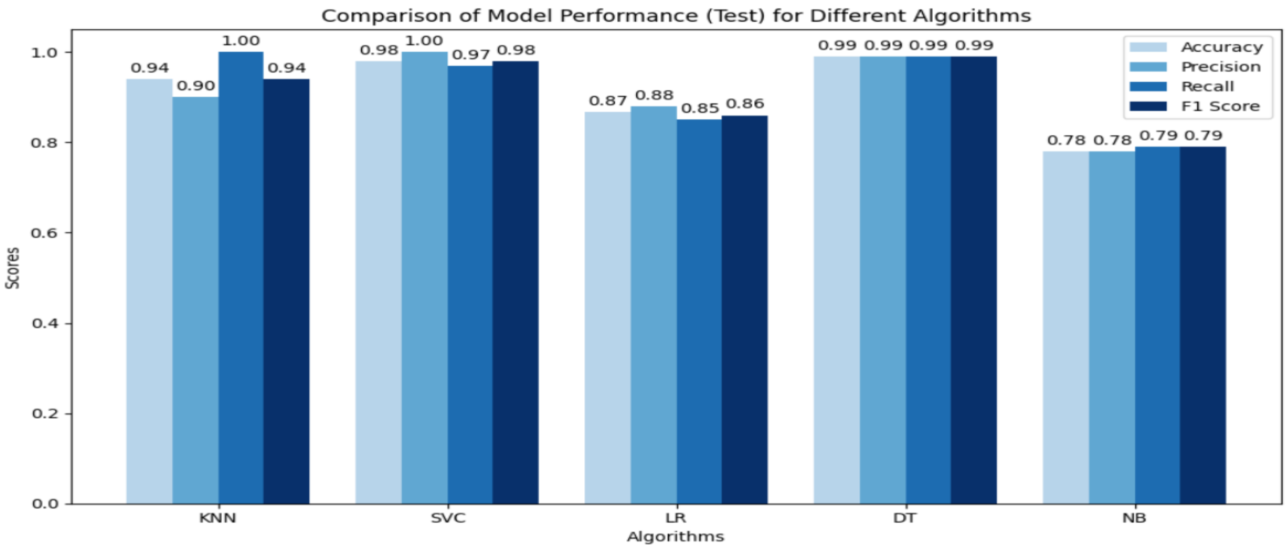

Figure (5): Comparison of Model performance (Test) for different algorithms

5. Discussion:

The KNN, SVC, and Decision Tree models exhibited perfect performance on the training data with accuracy, precision, recall, and F1 scores of (1.0). This indicates that these models adapted excellently to the training data. However, such results may signal overfitting, as this perfect performance was not equally replicated on the test data.

On the test data, the Decision Tree model outperformed others, achieving the highest accuracy (0.99) and near-perfect performance metrics, followed by the SVC model, which delivered advanced results across all metrics. This indicates that these models generalize well when faced with new data.

In contrast, the Logistic Regression and Naive Bayes models showed inferior performance across all metrics, which may suggest limitations in their ability to handle data complexities or the presence of nonlinear data distributions that they cannot efficiently model.

Despite the perfect performance of the KNN model on the training data, its results on the test data were less impressive (0.94 for accuracy and 0.90 for precision). This suggests that the model heavily relies on nearby data points in the training set, making it sensitive to variations or noise in new data. This sensitivity reflects an overfitting issue.

The Logistic Regression model showed relatively consistent performance between the training and test data, demonstrating moderate generalization capability. However, its lower performance compared to the other models may be attributed to its linear assumptions, which might be insufficient for capturing the complex relationships between variables in mortgage default prediction data.

Naive Bayes was the lowest-performing model across most metrics. This can be attributed to the model's simplified assumptions, such as feature independence, which may not hold true for complex financial data like mortgage default prediction. These assumptions hinder the model's ability to capture intricate relationships between features.

Both the Decision Tree and SVC models achieved outstanding results on the test data, suggesting their effectiveness in handling complex, nonlinear data. The Decision Tree model shows high flexibility in creating decision boundaries between classes, while the SVC excels at finding optimal decision boundaries using kernel methods.

6. Conclusion:

This study concludes that machine learning techniques are powerful and effective tools in the prediction of mortgage default, contributing to improved financial risk management. Through financial data analysis and the use of advanced models, financial institutions and investors can mitigate risks associated with mortgages.

The findings serve as a call to strengthen the use of machine learning techniques in financial analysis. However, relying on these models requires awareness of challenges related to training data, such as overfitting. Researchers and practitioners should work on enhancing these models by integrating economic data with financial data, using smart data balancing techniques, which will improve model accuracy in predicting defaults. Additionally, the results indicate the need for further research into hybrid models that combine traditional methods with machine learning techniques, which could lead to further improvements in prediction accuracy, enhancing financial system stability and reducing default risks.

References:

Aliaj, T., Anagnostopoulos, A., & Piersanti, S. (2020). Firms default prediction with machine learning. In *Mining Data for Financial Applications: 4th ECML PKDD Workshop, MIDAS 2019, Würzburg, Germany, September 16, 2019, Revised Selected Papers 4* (pp. 47-59). Springer International Publishing.

Behera, I., Nanda, P., Mitra, S., & Kumari, S. (2024). Machine Learning Approaches for Forecasting Financial Market Volatility. *Machine Learning Approaches in Financial Analytics*, 431-451.

Hanestan, F., Nathaniel, F., Kusuma, R. D., Gunawan, A. A. S., & Setiawan, K. E. (2024, July). Predicting Company Default Risks: A Machine Learning Perspective on Importance Features. In *2024 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)* (pp. 304-309). IEEE.

Incerti, F., Bargagli-Stoffi, F. J., & Riccaboni, M. (2022, September). A two-country study of default risk prediction using bayesian machine-learning. In *International Conference on Machine Learning, Optimization, and Data Science* (pp. 188-192). Cham: Springer Nature Switzerland.

Kim, H., Cho, H., & Ryu, D. (2021). Predicting corporate defaults using machine learning with geometric-lag variables. *Investment Analysts Journal*, *50*(3), 161-175.

Liu, J., Liu, Y., Ren, L., Li, X., & Wang, S. Trends and Trajectories: A Bibliometric Analysis of Financial Risk in Corporate Finance and Finance (2020-2024). *Available at SSRN 4960436*.

Nazareth, N., & Reddy, Y. V. R. (2023). Financial applications of machine learning: A literature review. *Expert Systems with Applications*, *219*, 119640.

Rahayu, D. S., Suhartanto, H., & Husodo, Z. A. (2022). Machine Learning and Finance Model in Predicting Default: Merton-based Reasoning. *Journal of Hunan University Natural Sciences*, *49*(2).

Song, Y., Wang, Y., Ye, X., Zaretzki, R., & Liu, C. (2023). Loan default prediction using a credit rating-specific and multi-objective ensemble learning scheme. *Information Sciences*, *629*, 599-617.

Xia, S., Zhu, Y., Zheng, S., Lu, T., & Ke, X. (2024). A deep learning-based model for P2P microloan default risk prediction. *International Journal of Innovative Research in Engineering and Management*, *11*(5), 110-120.

Kumar, A. (2023, March 27). Support vector machine (SVM) Python example. *Vitalflux.* https://vitalflux.com/

Hasni, M., Aguir, M. S., Babai, M. Z., & Jemai, Z. (2024). A Machine Learning Model for Accurate Credit Risk Forecasting in Banking Systems: An Empirical Investigation. *International Journal of Supply & Operations Management*, *11*(2).

Alam, B. (2022, January 12). Decision tree Python – Easy tutorial. *Hands-On.Cloud.* https://hands-on.cloud/

Putri, D. D., Nama, G. F., & Sulistiono, W. E. (2022). Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier. *Jurnal Informatika dan Teknik Elektro Terapan*, *10*(1).

Wang, A. X., Chukova, S. S., & Nguyen, B. P. (2023). Ensemble k-nearest neighbors based on centroid displacement. *Information Sciences*, *629*, 313-323.