Predictive Analytics in Enterprise Risk Assessment: A Machine Learning Perspective

Narayana Maharana
Department of Management Studies, GVP College of Engineering (Autonomous), Visakhapatnam, Andhra Pradesh, India
Email Id: maharana.narayan@gmail.com

Suman Kalyan Chaudhury
Department Business Administration, Berhampur University, Berhampur, Odisha, India
Email Id: sumankchaudhury72@gmail.com

Chandra Sekhar Patnaik
Department of Business Administration, Omega Degree and PG college, Hyderabad, Telangana,India
Email Id: cpattnaik8@gmail.com

Sudesh Kumar Kuppili
Department of Management Studies, GVP College of Engineering, Autonomous, Visakhapatnam, Andhra Pradesh, India
Email Id: drsudesh@gvpce.ac.in

Abstract: *Scientific risk assessment serves as a crucial assurance for the sustainable growth of businesses. With the continual progress and maturation of machine learning technology, its significance in the realm of data prediction and risk assessment has become pivotal. This study investigates the utilization of machine learning in assessing enterprise risks, employing three distinct algorithms—namely, support vector machine (SVM), random forest (RF),and AdaBoost. The initial step involves establishing comprehensive risk assessment indexes for enterprises, capturing diverse risks through various parameters. Then, utilising previously collected secondary data, the three machine learning algorithms were trained to develop a reliable risk evaluation model. Lastly, the risk indices were produced by the applied risk assessment model using a collection of current risk indicators. The experimental phase involves the analysis and validation of the method using real data, demonstrating the efficacy of the proposed machine learning algorithms in accurately evaluating enterprise risks.*

Keywords: *Machine Learning, Artificial Intelligence, Financial Risk, Predictive analytics.*

Introduction

Given the rising popularity of artificial intelligence (AI) and machine learning (ML), and the onset of the age of big data analytics, numerous researchers have employed ML techniques to extensively investigate risk assessment [1-4]. In the recent times, enterprise risk management assumes a paramount significance in safeguarding the stable and uninterrupted functioning of financial institutions on a global scale. It serves as the linchpin for ensuring the resilience and sustainability of these institutions amidst the dynamic and often unpredictable landscape of the financial world. By meticulously identifying, assessing, and mitigating risks across various facets of their operations, financial institutions fortify their ability to weather adversities and capitalize on opportunities. Traditional methods for determining user defaultshave no longer suffice in meeting the demands of today's diverse data types, large user populations, and the need for high-risk prediction accuracy [5-7]. Many scholars have turned to machine learning methods, engaging in comprehensive discussions and presenting research findings to demonstrate the efficacy of these methods in prediction and generalization [8-10]. Originally, risk assessment techniques built around statistical learning techniques were the mainstay of research. Methods like regression analysis were extensively used for estimating credit risk of companies. Approaches to evaluating risk also used linear discriminant analysis, and some mathematical models developed in order to handle critical difficulties in analysing credit risk[11-14].Nonetheless, despite the utility of these methodologies, it is imperative to acknowledge their inherent limitations. One prominent issue lies in their tendency to operate within hypothetical frameworks, often placing stringent requirements on data distribution and relying excessively on sample classification rooted in variance rather than mean-based analysis. Consequently, the resulting classification outcomes may lack the desired robustness. In response to these challenges, efforts have been directed towards employing linear regression as a means to establish a comprehensive score rating system. This system operates by meticulously evaluating the credit status and actual circumstances of potential lenders (as referenced in[15-17], with the ultimate goal of forecasting the credit risk associated with bank customers. However, it's essential to recognize that linear regression is not without its own drawbacks, as highlighted in existing literature [18, 19]. One notable limitation is the expansive output range, extending from positive to negative infinity, which can present challenges in practical application. To mitigate this issue, logistic regression has emerged as a promising alternative. This approach, pioneered by Wiginton et al., introduces the sigmoid function, thereby transforming the linear regression output into a probability value. This transformation facilitates the establishment of an empirical threshold, typically ranging between 0 and 1, enabling the effective resolution of binary classification problems. By harnessing logistic regression, practitioners can navigate the complexities of credit evaluation with enhanced precision and reliability, thereby advancing the state-of-the-art in risk assessment methodologies[20, 21]. In recent years, the utilization of machine learning-based risk assessment models has gained prominence, surpassing conventional risk assessment approaches[7, 13, 22-25]. In the contemporary landscape of machine learning methodologies, a multitude of approaches have emerged as significant contenders in the domain of risk assessment. Among these, prominent methodologies include the Backpropagation (BP) neural network, renowned for its capability to adaptively learn and model complex relationships within data. Similarly, the K Nearest Neighbours (KNN) algorithm stands out for its simplicity and effectiveness in classifying data points based on their proximity to neighbouring instances. Further expanding the repertoire, the Support Vector Machine (SVM) algorithm has garnered considerable attention for its ability to delineate intricate decision boundaries, particularly useful in scenarios with non-linearly separable data. Notably, the spectrum extends beyond these foundational methodologies to encompass a diverse array of tree-based machine learning models.These tree-based models constitute a cornerstone in contemporary risk assessment practices, with basic decision tree models serving as fundamental building blocks. Moreover, integrated models such as Random Forest (RF), Gradient Boosting Decision Trees (GBDT), XGBoost, and LightGBM have emerged as formidable tools in the risk assessment toolkit. Their ability to aggregate insights from an ensemble of decision trees offers heightened accuracy and robustness in discerning patterns and mitigating risks within complex datasets. As such, the widespread adoption of these methodologies underscores their efficacy and versatility in addressing the multifaceted challenges inherent to enterprise risk assessment in today's dynamic business landscape[4, 26-30]. Many scholars have highlighted the pivotal role of machine learning methodologies in the analysis of historical risk data, primarily executed through supervised learning techniques [22, 31, 32]. This process entails a sequence of operations commencing with meticulous data processing and feature extraction. Following this preparatory phase, the derived model is deployed to prognosticate user behaviour and delineate various characteristics, thereby facilitating the discernment of enterprise risk. In alignment with the established corpus of literature, the present investigation leverages machine learning algorithms for the assessment of enterprise risk by using the ML algorithms such as;"Random Forest, Support Vector Machine (SVM), and AdaBoost". These methods have been meticulously selected and employed to scrutinize and evaluate the risk profile associated with individual companies. The endeavour to establish a comprehensive corporate risk indicator system unfolds through the systematic training of the aforementioned machine learning algorithms. This training process involves the utilization of extensive historical corporate data, meticulously curated to ensure the derivation of a robust and reliable evaluation model. Subsequent to this model's construction, the current state of the enterprise under examination undergoes thorough scrutiny, and its risk profile is meticulously evaluated. Employing actual data sets, the experiment rigorously assesses the performance of the three ML algorithms, thereby substantiating their efficacy and reliability in the domain of enterprise risk assessment.

Review of Literature

Machine learning techniques are increasingly applied to assess the financial risk of enterprises, such as credit risk, market risk, operational risk, and compliance risk. Machine learning techniques can leverage large amounts of data, both structured and unstructured, to build models that improve decision making, tailor services, and enhance risk management. However, machine learning techniques also pose some challenges and limitations, such as model interpretability, data quality, regulatory compliance, and ethical issues. Enterprise risk is the term used to describe the various types of risks that an organization may face in its operations, such as financial, operational, strategic, reputational, legal, or compliance risks. Enterprise risk management (ERM) is the process of identifying, assessing, and managing these risks in a holistic and integrated way, aligned with the organization's objectives and strategy. ERM can help organizations to improve their decision-making, performance, and resilience in the face of uncertainties and challenges.

Credit risk: Credit risk is the risk of loss due to the default or deterioration of the creditworthiness of a borrower or a counterparty. Machine learning techniques can help to improve the accuracy and efficiency of credit scoring, credit rating, and default prediction, by using various types of data, such as financial statements, transaction records, social media, and alternative data. For example, Babel, Buehler [33] used machine learning techniques to assess the credit risk of retail customers in China, who have limited or no credit history. Similarly,Aziz and Dowling [34] proposed an enhanced hybrid integration algorithm based on random space and MultiBoosting to predict the credit risk of small and medium-sized enterprises.Huang, Wei [24] used RF, SVM, and AdaBoost to evaluate the credit risk of enterprises based on financial and non-financial indicators. They found that the machine learning models outperformed the traditional logistic regression model in terms of accuracy, precision, recall, and F1-score.Yang, undefined [35] constructed a variability model based on big data machine learning algorithms to provide early warning and control of financial risks. He used a combination of principal component analysis, k-means clustering, and support vector machine to analyse the financial data of listed companies, and then used a deep neural network to predict the financial risk level of the companies.

Market risk: Market risk is the risk of loss due to adverse movements in market prices or rates, such as interest rates, exchange rates, equity prices, or commodity prices. Machine learning techniques can help to model the dynamics and dependencies of market factors, to optimize the portfolio allocation and hedging strategies, and to detect and prevent market anomalies and frauds. For instance, Kim, Yang [36] introduced an innovative machine learning approach for financial risk detection and classification, which has demonstrated the potential to mitigate enterprise financial risks by up to 10%. Their methodology hinges upon the meticulous selection of features, accomplished through the application of a genetic algorithm. Subsequently, a diverse array of machine learning techniques, including decision trees, k-nearest neighbour, naive Bayes, and support vector machine, are deployed to categorize enterprises into distinct risk groups: low-risk, medium-risk, and high-risk.Similarly, Gulsoy and Kulluk [37] presented a methodical framework for the objective measurement of risk within the loan processes of small and medium-sized business entities. Their approach relies on sophisticated data mining techniques to extract valuable insights. By leveraging these techniques, they offer a systematic and unbiased assessment of risk factors inherent in the lending practices directed towards such enterprises.

Operational risk: Operational risk represents the potential for financial loss stemming from deficiencies or breakdowns in internal processes, human resources, technological systems, or unforeseen external occurrences.Machine learning techniques can help to identify, monitor, and mitigate the sources and impacts of operational risk, such as human errors, cyberattacks, natural disasters, or legal disputes. For example, Yang, Yu [38] constructed a variability model based on big data machine learning algorithms to provide early warning and control of financial risks. Kim, Jung and Kim [39] developed a deep learning-based anomaly detection system to prevent fraudulent transactions in online banking.

Compliance risk: This category of risk encompasses the potential consequences arising from non-compliance with regulatory mandates or legal obligations, spanning diverse sectors and jurisdictions. Organizations confront compliance risk in navigating a complex web of regulatory frameworks governing their operations, encompassing areas such as financial reporting, consumer protection, data privacy, environmental regulations, and industry-specific standards. Machine learning techniques can help to automate and streamline the compliance processes, such as anti-money laundering, know your customer, fraud detection, and regulatory reporting. For example, Tertychnyi, Slobozhan [40] reported that machine learning techniques can reduce the false positive rates of anti-money laundering alerts by up to 80%, and increase the accuracy of fraud detection by up to 50%. [34]reviewed the applications of machine learning techniques in the field of regulatory technology (RegTech), which aims to enhance the efficiency and effectiveness of regulatory compliance.

Number of scholars have made significant contributions to the field of systematic risk analysis within the financial sector through the application of machine learning methodologies. Notably, Kou, Chao [41] proposed a machine learning technique aimed at systematic risk analysis, shedding light on the crucial role of financial systems in managing financial systemic risk. The utilization of machine learning techniques has become increasingly prevalent among researchers striving to discern and mitigate systemic risk by harnessing vast datasets extracted from financial markets. Building upon this foundation, A Financialization Risk Assessment method for controlling Risk based on artificial intelligence and Machine Learning was presented by Song and Wu [25], highlighting the use of big data as well as machine learning to improve trade financing capabilities and lessen the risks related to over securitization. Employing genetic algorithms, neural networks, and principal component analysis, Clintworth, Lyridis and Boulougouris [42] proposed a comprehensive machine learning approach for financial risk evaluation in the shipping industry. This method addresses the imperative need for accurate default risk assessments, a concern shared by stakeholders such as regulators and banks. In a related context, Hansen and Borch [43]investigated Uncertainty Absorption and Amplification in machine learning-driven finance, elucidating the impact of market changes and associated uncertainties on financial products. Collectively, these research endeavours contribute to a more nuanced understanding of risk management within the financial domain, showcasing the evolution and effectiveness of machine learning applications in addressing complex financial challenges.

There is no definitive answer to what type of machine learning techniques can be best used in identifying and addressing enterprise risks, as different techniques may have different advantages and disadvantages depending on the data, the problem, and the context[44-47]. However, some general guidelines for use of different techniques are given in Table 1.

Table 1: Machine Learning tools and their usage

Type	Description	Tools
Supervised learning problems	Credit scoring, default prediction, fraud detection where the data is Imbalanced, Nonlinear relationships, feature interactions.	Random forest, Gradient Boosting, neural networks can show better result than logistic regression, linear discriminant analysis.
Unsupervised learning problems	Anomaly detection, clustering, dimensionality reduction where the aim is to analyse more complex data.	Deep autoencoders, generative adversarial networks, or self-organizing maps are better techniques than k-means, principal component analysis, or isolation forest.
Reinforcement learning problems	portfolio optimization, hedging, or stress testing, techniques that can learn from feedback and adapt to changing environments and scenarios.	Q-learning, policy gradient, or actor-critic methods, are better techniques over mean-variance optimization, delta hedging, or Monte Carlo simulation.

Source: Algorithmic and machine learning risk management | Deloitte US

Of course, these are not exhaustive or exclusive lists, and other factors, such as data quality, model interpretability, computational efficiency, and regulatory compliance, should also be considered when choosing the appropriate machine learning techniques for enterprise risk management. Moreover, hybrid or ensemble techniques, such as stacking, bagging, or boosting, may offer better performance and robustness than single techniques, by combining the strengths and mitigating the weaknesses of different techniques.

Machine learning (ML) has become increasingly prominent as a means to transform unknown variables into manageable threats, offering a logical and economically sound approach to reducing uncertainty. In the research conducted by Jomthanachai, Wong and Lim [48], developed a Risk Management model utilizing Data Envelopment Analysis (DEA) which is a non-parametric method used in operations research and management science to evaluate the relative efficiency of decision-making units (DMUs), such as organizations, departments, or processes and they incorporated Machine Learning algorithm to it for better result. This integration of DEA with machine learning methods resulted in enhanced risk control and improved the treatment procedures. ML techniques are applied to forecast risk levels at real time based on simulated data aligned with the assumptions of hazard analysis. Another notable contribution by Shimin, Ke [49] involves the creation of a platform based on XGboost for detecting financial fraud, providing an innovative solution for safe operation of the e-commerce by identifying possiblefrauds through a combination of manual and automatic classifications. Zhang [50] addressed the critical issues relating to fraud happening in digital transactions using a content and graphical transaction identity verification model. He used the XGboost ML algorithm in his proposal for detecting customer transaction fraud and claimed better reliability in predicting fraud. Orlova [51]recognised the growing level of economic crisis in banks as a result of higher workloads in credit operations, which make a major contribution to the major chunk of income of such organisations, and therefore concentrated on techniques of credit risk management decision-making using ML and AI. These research endeavours collectively contribute valuable insights that motivate the development and implementation of online transaction systems by researchers and engineers in the field.

The literature review elucidates that a central concern in finance pertains to the speed of innovation facilitated by information technology. The significant role of financial risk management in influencing organisational performance is paramount. Researchers are increasingly incorporating machine learning techniques to discern and address hazards utilizing the burgeoning volume of data amassed in financial markets and systems. Encompassing the development of mobile banking services, crowdfunding, stock trading using various online platforms, online money transactions,and cryptocurrenciesthese activities generate substantial gigabytes of data. The critical consideration in this context revolves around the secure handling of such voluminous data, given its inclusion of sensitive personal and financial information. The mishandling of such data could lead to severe consequences for numerous businesses and industries. Standard models are inadequate in providing precise credit risk forecasts since they primarily rely on identity and demographic data. The fact that these models are restricted to demographic factors like “ID, name, age, marital status, and education level, etc.” makes it difficult for financial institutions to attract new customers. It is difficult to integrate diversified and fragmented information gathered from multiple sources into the models and statistical techniques that are currently in use in financial risk management.

Research Methodology

The enterprise’s risk status directly influences the borrower’s capacity and desire to pay off the loan obligation. As a result, it is required to develop a scientific and instinctive signalling system to aid the creditors like financial institutions and bank while taking lending decisions, allowing for more scientific and unbiased assessment. For instance, the examination of variables impacting financing decision-making by considering all the parameter or factors that determines the credit risk position of a firm. Following the footsteps of many prior studies, we also used various ratios which are popularly used as the standard indicators of financial risk such as current ratio, quick ratio or acid test ratio, inventory turnover ratio, Debt to equity ratio, debt to tangible assets ratio, Return on assets ratio,interest coverage ratio etc. Table 2 provides the estimates used for enterprise financial risk assessment:

Table 2: Variables and their Estimation Methods

Estimator	Formula	Definition
Current Ratio (CR)	CR= Current assets/ Current liabilities	"This index reflects the company's ability to repay short-term debt. The more the current assets and the fewer the short-term debts, the greater the current ratio and the stronger the company's short-term debt repayment ability".
Quick Ratio (QR)	QR=Total Liquid Asset/ Total Current Liability	"This index can reflect the company's ability to repay short-term debt. Because current assets still include inventories that have a slower realization rate and may have depreciated, the current assets are deducted from inventories and then compared with current liabilities to measure the company's short-term debt solvency".
Inventory turnover ratio (ITR)	ITR= Total Sales/ Average Inventory	"This index is the main indicator of inventory turnover speed. Carrying high inventory turnover rate and shortening the business cycle can improve the company's liquidity".
Debt to Equity Ratio (DER)	DER= Total Debt/ Total Equity	"This index reflects the ratio of capital provided by creditors to total capital. This index is also called the debt-to-business ratio".
Debt to Tangible Assets Ratio (DTAR)	Total Debt/ (Total Asset- Total intangible Assets)	"The extension of the property rights ratio index more cautiously and conservatively reflects the degree to which the capital invested by creditors is protected by shareholders' rights during the liquidation of the enterprise. Regardless of the value of intangible assets, including goodwill, trademarks, patent rights, and nonpatent technologies, they may not be used to repay debts. For the sake of caution, they will all be regarded as insolvent".
Return on asset ratio (ROA)	Net Profit/ Average Assets	"This index compares the net profit of the company for a certain period with the company's assets, showing the comprehensive utilization effect of the company's assets. The higher the index, the higher the efficiency of asset utilization, indicating that the company has achieved good results in increasing income and saving funds. Otherwise, the opposite conclusion is true".
Interest coverage ratio (ICR)	Earnings Before Interest and Tax/Interest payable	"The ratio of business income to interest expense is used to measure the company's ability to repay the interest on borrowings. It is also called interest protection multiple. As long as the multiple of the interest earned is large enough, the enterprise has sufficient ability to repay the interest".
Credibility	Categorical Variable 1= performers, -1= defaulters	"These companies were categorised into two groups as "performing companies (p=1)" and "default companies (p=-1)" based on their financial position, credit records and operational viability".

Source: author's own compilation

Enterprise Risk Assessment Model Development

This paper rigorously selects three distinct machine learning algorithms, namely Random Forest (RF), Support Vector Machine (SVM), and AdaBoost, for the purpose of constructing enterprise risk assessment models. The fundamental principles underlying each algorithm are explicated as follows: [18–24].

A. Random Forestis a popular ensemble machine learning algorithm that combines multiple decision trees to improve predictive accuracy and control overfitting. However, explaining the entire process manually using statistical equations in just five steps is challenging, as it involves a series of computations and concepts. Nonetheless, I'll provide a simplified overview to give you a basic understanding. Keep in mind that in practice, these steps are typically implemented using programming languages like Python or R.

Step 1: Create Random Subsets of Data (Bootstrapping)

Equation**:** $D_i = [(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)]$

Randomly select subsets of the original dataset with replacement (bootstrapping) to create multiple training datasets D_i. This helps introduce diversity among the trees.

Step 2: Build Decision Trees

Equation: $\hat{y}_{i,t} = Decision\ Tree\ (D_i)$

For each subset D_i, build a decision tree using a random subset of features at each split. The decision tree predicts the target variable $\hat{y}_{i,t}$based on the input features.

Step 3: Aggregate Predictions

Equation**:**

$$\hat{Y} = \frac{1}{T}\sum_{t=1}^{T} \hat{y}_{i,t}$$

Combine the predictions of all individual trees $(\hat{y}_{i,t})$ by averaging (for regression problems) or using voting (for classification problems) to obtain the final prediction $\hat{Y}$.

Combine the predictions of all individual trees $(\hat{y}_{i,t})$ by averaging (for regression problems) or using voting (for classification problems) to obtain the final prediction $\hat{Y}$.

Step 4: Assess Model Performance

Equation**:** Depends on the problem (e.g., Mean Squared Error for regression, Accuracy for classification).Evaluate the model's performance on a separate validation dataset using an appropriate metric. This helps ensure that the model generalizes well to unseen data.

Step 5: Tune Parameters and Repeat:Adjust hyperparameters (e.g., tree depth, number of trees) based on the performance evaluation. Repeat the process until satisfactory results are achieved.While the above steps provide a high-level overview, each step involves more detailed computations and considerations, such as calculating impurity for tree splits, feature selection, and handling categorical variables.

A. Support Vector Machines (SVM) is a popular machine learning algorithm used forclassification and regression tasks. Conducting SVM manually involves optimizing the hyperplane that best separates the classes in the feature space. Here are five steps to manually implement SVM using statistics equations:

Step 1: Data Preparation:Start with a labelled dataset consisting of input features (X) and corresponding class labels (y).Normalize or standardize the features to ensure they are on a similar scale.

Step 2: Linear SVM:For simplicity, let's start with a linear SVM. The decision function for a linear SVM can be represented as: $f(x)=wx+b$ where,'w' is the weight vector,'x' is the input feature vector, and 'b' is the bias term.

Step 3: Objective Function:The goal is to find the optimal hyperplane that maximizes the margin between the classes. This can be formulated as an optimization problem:

Maximize: $\frac{2}{\|w\|}$

subject to the constraints:

$y_i\ (wx_i + b) \geq 1$, for all *i*, where y_i is the class label of the i^{th} sample.

Step 4: *Optimization:*Use optimization techniques (e.g., Lagrange multipliers) to solve the constrained optimization problem. The Lagrangian for this problem is:

$$\mathcal{L}(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \alpha_i \left[y_i (wx_i + b) - 1\right]$$

where α_i are the Lagrange multipliers.

Step 5: Decision Boundary:Once you've found the optimal values for *w* and *b*, the decision boundary is given by $f(x) = wx + b = 0$.Support vectors are the data points that lie on the margins or violate the margin constraints.

This is a simplified explanation, and in practice, machine learning libraries are commonly used to implement SVM due to their efficiency. Popular libraries like scikit-learn in Python can handle the optimization process and provide tools to train SVM models easily.

B. AdaBoost (Adaptive Boosting)AdaBoost, or Adaptive Boosting, is a machine learning algorithm that constructs a strong classifier by iteratively combining multiple weak classifiers. Initially, each training example is given equal weight, and a weak learner is trained on the data. In subsequent iterations, the algorithm increases the weight of misclassified examples, forcing the weak learners to focus more on difficult instances. Each weak learner is assigned a weight based on its accuracy, and the final prediction is made by a weighted sum of their outputs. AdaBoost is known for its ability to handle complex decision boundaries and its adaptability to various learning tasks, although it can be sensitive to noisy data. Here's a manual step-by-step explanation of AdaBoost using statistical equations in five steps:

Step 1: Initialize Weights:Initialize the weights for each training instance. Let w_i represent the weight of the i^{th} instance. For a dataset with '*N'* instances, initially set $w_i = \frac{1}{N}$.

Step 2: Train Weak Learner:Train a weak learner (e.g., a decision stump) on the dataset with the current weights. Let $h_t(x)$ be the hypothesis of the t^{th} weak learner.

Step 3: Compute Error:Calculate the weighted error of the weak learner using the formula:

$$\epsilon_t = \frac{\sum_{i=1}^{N} w_i \cdot I(y_i \neq h_i(x_i))}{\sum_{i=1}^{N} w_i}$$

wherey_i is the true label of the i^{th} instance, $h_t(x_i)$ is the prediction of the weak learner, and I(Å") is the indicator function.

Step 4: Compute Classifier Weight:Calculate the weight of the weak learner in the final combination:

$$\alpha_t = \frac{1}{2} ln\left(\frac{1 - \epsilon_t}{\epsilon_t}\right)$$

Step 5: Update Weights:Update the weights of the training instances based on the weak learner's performance:

$$w_{i,new} = w_i \cdot exp(-\alpha_t \cdot y_t \cdot h_t(x_t))$$

Normalize the weights so that they sum to 1:

$w_{i,new} = \frac{w_{i,new}}{\sum_{i=1}^{N} w_{i,new}}$ Repeat Steps 2-5 for a predefined number of iterations or until a certain criterion is met. The final strong learner is a weighted combination of the weak learners:

$$H(x) = sign\left(\sum_{i=1}^{N} \alpha_t \cdot y_t \cdot h_t(x_t)\right)$$

This process gives more emphasis to the misclassified instances in each iteration, improving the overall model's performance.

Result and Analysis

The study used the financial data collected from 600 BSE Listed manufacturing companies which obtained loan from banks in their capital structure. These companies were again sub-categorised into two groups as "performing companies (p=1)" and "default companies(p=-1)" based on their financial position, credit records and operational viability. According to the safety assessment tools specified above each company can considered a seven-dimensionalvector each dimensionrepresentingone safety measure. Initially, the sample data undergoes rigorous processing to ensure robustness and efficiency. Given the substantial volume of sample data and its overall smoothness, we used double triple standard deviation test to identify and eliminate aberrant data, resulting in a final count of 480 valid samples. Within this dataset, 242 companies fall in the performing category where they possess the capability to meet bank credit obligations, while the remaining 238 are default category that exhibit challenges in repaying interest and loans promptly. This inquiry used accuracy and the Receiver Operating Characteristic (ROC) curve as quantitative assessment criteria in order to comprehensively examine the effectiveness of the suggested model. The accuracy measure was used to determine the percentage of appropriately categorised samples relative to the entire sample size. This method is well appreciated for being easy to use and successful in evaluating classification and prediction abilities.The assessment of machine learning algorithms encompasses the comprehensive utilization of the Area under the Curve (AUC), which serves as a metric for evaluating posterior probability, classification proficiency, and ranking efficacy. This metric, extensively employed within the realm of machine learning algorithms, is derived by plotting false positive class rate (FPR) against true positive class rate (TPR) with FPR in the x-axis and TPR in the y-axis. The adjustment of classifier thresholds yields a series of distinct (FPR, TPR) points on the coordinate axis, subsequently connected to form the Receiver Operating Characteristic (ROC) curve. While the ROC curve itself is not directly employed as an assessment criterion for classifiers, the AUC value is commonly adopted as a quantitative benchmark for evaluating their performance.

Observations

We have used K-fold cross validation were we first divided the total sample in to a multiple of 10 where each fold consists of 48 samples. These sample groups are then identified as T_i where i=1 to 10. T_i is used as test and training data sets on rotation basis for machine learning like for instance T1 to T9 for training and T10 for testing and so on. In this way we will have i^{th} group of combinations of training and testing data. Then the average value of mean absolute error, accuracy and AUC for all the three models were estimated and presented in Table 3.

Table 3: Comparison of performance of the three machine learning algorithms

Estimates	RF	SVM	AdaBoost
Mean Abs Error	0.1250	0.1666	0.3333
Accuracy (%)	0.9375	0.8333	0.9166
AUC	0.9598	0.9642	0.9267

Source: Primary

The observations derived from Table 3 indicate distinct performance trends among the evaluated models. Specifically, the SVM model demonstrates effectiveness, while both the RF and AdaBoost models exhibit outstanding performance. In terms of accuracy, the RF model surpasses the SVM and AdaBoost models. Conversely, regarding the AUC value, the RF model aligns closely with the SVM model, both surpassing the AdaBoost model. When considering both evaluation indicators, the RF model outperforms the AdaBoost model by 0.0209 in accuracy and 0.0331 in AUC value. The mean absolute errors estimated for three models suggests that RF model is better than the other two models with lowest mean absolute value of 0.1250. Furthermore, the relationship between enterprise risk levels, as inferred from the SVM and RF models, shows slight improvement compared to the AdaBoost model. To account for potential noise impacts, the study introduced varying degrees of noise levels to the sample containing 480 data sets, measuring the noise level using the Signal-to-Noise ratio (SNR). Figure 1 depicts the accuracy-performance curves of the machine learning methods across diverse SNR levels. The comparison reveals that the RF and SVM methods exhibit superior noise robustness compared to the AdaBoost method, highlighting their stronger resilience.

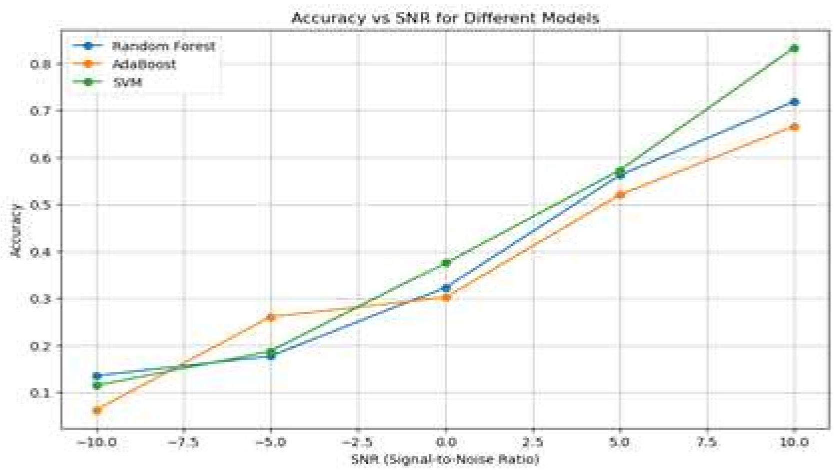

Figure 1: Accuracy of the Machine Learning Algorithms

Source: primary

Implications

The studyemployed K-fold cross-validation to evaluate the performance of the ML models developed using “RF, SVM, and AdaBoost” in predicting enterprise risk. The findings reveal that the AdaBoost model outperformed SVM and RF in terms of accuracy, while both RF and AdaBoost demonstrated superior noise robustness compared to SVM. The observed improvement in predicting enterprise risk levels with AdaBoost and RF models underscores their practical relevance for risk assessment applications. Additionally, the study’s rotation-based cross-validation approach ensures a thorough evaluation of model generalization. These insights contribute valuable information for practitioners seeking effective predictive analytics tools in the ground of enterprise risk assessment, emphasizing the significance of choosing models with high accuracy and resilience to noise for achieving robust and reliable predictions.

Conclusion

Machine learning algorithms are commonly utilised in risk evaluation because of their effective interpretability and simple mechanism. But the widely held belief that variables have a linear connection frequently results in forecasts that are not accurate enough to adequately convey the complex risk status in different circumstances. This study advocates for the application of modern machine learning methods in constructing risk assessment models, emphasizing their higher level of accuracy has been achieved through comprehensive data training and their expansive potential in enterprise risk evaluation. For corporate risk assessment, three well-known machine learning algorithms such as; "RF, SVM, and AdaBoost" are implemented and put through a rigorous testing process using actual data-driven scenarios. When compared, the study shows that RF and AdaBoost estimate risk more accurately than SVM method. Understanding the unique benefits of various machine learning techniques, the research proposes that merging or integrating these approaches for processing of data features can improve the suggested approach's overall reliability even more, leading to more precise and accurate risk assessments by handling complex real-world data.

Limitations and Scope

Despite the valuable insights gained from the current study on applying different machine learning techniquesto verifying the enterprise risk levels, certain limitations warrant consideration and present avenues for further research. Firstly, the study relies on the assumption that risk relationships are linear, which may not always reflect the intricate nature of real-world scenarios. Future investigations should explore the incorporation of non-linear relationships and more complex interactions among variables to enhance the predictive accuracy of risk assessment models. Additionally, the study focuses on three specific machine learning algorithms, and while they demonstrate high accuracy, the field is rapidly evolving with the introduction of new algorithms and techniques. Exploring the effectiveness of emerging machine learning models and ensuring the scalability and adaptability of these models to diverse industry contexts is essential for comprehensive risk assessment strategies. Furthermore, the study primarily evaluates model performance using accuracy, mean absolute error, and AUC, providing a broad overview of predictive capabilities. However, for a more comprehensive understanding, future research could delve into the interpretability of the machine learning models employed. An exploration of model interpretability and explainability is crucial for building trust and acceptance in practical risk assessment models, especially in industries with stringent regulatory requirements. Additionally, investigating the robustness of these models in dynamic environments and under varying conditions, such as economic fluctuations or unforeseen events, would contribute to the practical applicability of the findings. Moreover, further research should focus on refining model assumptions, exploring other emerging machine learning techniques in order to enhance interpretability, robustness in diverse and dynamic risk assessment scenarios and provide more reliable and actionable insights for decision-makers.

References

1. Al-Haschimi, A., A. Apostolou, A. Azqueta-Gavaldon, and M. Ricci, Using Machine Learning to Measure Financial Risk in China. ECB Working Paper, No. 2767. 2023: European Central Bank (ECB), Frankfurt a. M.

2. Arora, N. and P.D. Kaur, A Bolasso based consistent feature selection enabled random forest classification algorithm: An application to credit risk assessment. Applied Soft Computing, 2020. 86: p. 105936.

3. Aziz, S. and M.M. Dowling, AI and Machine Learning for Risk Management. 2018.

4. Babel, B., et al., Derisking machine learning and artificial intelligence, in McKinsey

Quarterly, Business Technology Office. 2019, McKinsey and Company. p. 1-4.

5. Behera, J., A.K. Pasayat, H. Behera, and P. Kumar, Prediction based mean-value-at-risk portfolio optimization using machine learning regression algorithms for multi-national stock markets. Engineering Applications of Artificial Intelligence, 2023. 120(1): p. 105843.

6. Bhatore, S., L. Mohan, and Y.R. Reddy, Machine learning techniques for credit risk evaluation: a systematic literature review. Journal of Banking and Financial Technology, 2020. 4(1): p. 111–138.

7. Breiman, L. and R. Cutler, Random forests machine learning [J]. journal of clinical microbiology, 2001. 2: p. 199-228.

8. Burstein, G. and I. Zuckerman, Deconstructing Risk Factors for Predicting Risk Assessment in Supply Chains Using Machine Learning. Journal of Risk and Financial Management, 2023. 16(2): p. 97.

9. Chen, T., Q. Yang, Y. Wang, and S. Wang, Double-layer network model of bank-enterprise counterparty credit risk contagion. Complexity, 2020. 2020: p. 1-25.

10. Clintworth, M., D. Lyridis, and E. Boulougouris, Financial risk assessment in shipping: a holistic machine learning based methodology. Maritime Economics & Logistics, 2023. 25(1): p. 90-121.

11. Dong, S. and C. Liu, Sentiment classification for financial texts based on deep learning. Computational Intelligence and Neuroscience, 2021. 2021(9524705): p. 1-9.

12. Fan, C.-L., Defect Risk Assessment Using a Hybrid Machine Learning Method. Journal of Construction Engineering and Management, 2020. 146(9): p. 04020102.

13. Faridi, S., et al., Portfolio rebalancing based on a combined method of ensemble machine learning and genetic algorithm. Journal of Financial Reporting and Accounting, 2023. 21(1): p. 105–125.

14. Fonseca, D.P., P.F. Wanke, and H.L. Correa, A two-stage fuzzy neural approach for credit risk assessment in a Brazilian credit card company. Applied Soft Computing, 2020. 92: p. 106329.

15. Huang, B., et al., Enterprise Risk Assessment Based on Machine Learning. Computational Intelligence and Neuroscience, 2021. 2021: p. 1–6.

16. Gulsoy, N. and S. Kulluk, A data mining application in credit scoring processes of small and medium enterprises commercial corporate customers. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2019. 9(3): p. e1299.

17. Hansen, K.B. and C. Borch, The absorption and multiplication of uncertainty in machine learning driven finance. The British Journal of Sociology, 2021. 72(4): p. 1015-1029.

18. Huang, J. and C.X. Ling, Using AUC and accuracy in evaluating learning algorithms. IEEE Transactions on knowledge and Data Engineering, 2005. 17(3): p. 299-310.

19. Jiang, W., et al., Risk assessment and validation of flood disaster based on fuzzy mathematics. Progress in Natural Science, 2009. 19(10): p. 1419-1425.

20. Jianxing, Y., C. Haicheng, W. Shibo, and F. Haizhao, A novel risk matrix approach based on cloud model for risk assessment under uncertainty. IEEE Access, 2021. 9: p. 27884-27896.

21. Jomthanachai, S., W.-P. Wong, and C.-P. Lim, An Application of Data Envelopment Analysis and Machine Learning Approach to Risk Management. IEEE Access, 2021. 9: p. 85978–85994.

22. Kim, A., et al., Can deep learning predict risky retail investors? A case study in financial risk behavior forecasting. European Journal of Operational Research, 2020. 283(1): p. 217-234.

23. Kim, J., H. Jung, and W. Kim, Sequential pattern mining approach for personalized fraudulent transaction detection in online banking. Sustainability, 2022. 14(15): p. 9791.

24. Kou, G., et al., Machine learning methods for systemic risk analysis in financial sectors. Technological and Economic Development of Economy, 2019. 25(5): p. 716–742.

25. Luo, J., X. Yan, and Y. Tian, Unsupervised quadratic surface support vector machine with application to credit risk assessment. European Journal of Operational Research, 2020. 280(3): p. 1008-1017.

26. Moscatelli, M., F. Parlapiano, S. Narizzano, and G. Viggiano, Corporate default forecasting with machine learning. Expert Systems with Applications, 2020. 161: p. 113567.

27. Mhlanga, D., Financial Inclusion in Emerging Economies: The Application of Machine Learning and Artificial Intelligence in Credit Risk Assessment. International Journal of Financial Studies, 2021. 9(3): p. 39.

28. Nowak, J.W., S. Sarkani, and T.A. Mazzuchi, Risk assessment for a national renewable energy target part I: Developing the model. IEEE systems journal, 2014. 9(3): p. 1045-1056.

29. Ohlsson, C., Exploring the potential of machine learning : How machine learning can support financial risk management. 2017.

30. Orlova, E., Decision-Making techniques for credit resource management using machine learning and optimization. Information. 2020. 11(3): p. 144.1-144.17.

31. Sigrist, F. and N. Leuenberger, Machine learning for corporate default risk: Multi-period prediction, frailty correlation, loan portfolios, and tail probabilities. 0377-2217, 2023. 305(3): p. 1390–1406.

32. Qu, Q., C. Liu, and X. Bao, E-commerce enterprise supply chain financing risk assessment based on linked data mining and edge computing. Mobile Information Systems, 2021. 2021: p. 1-9.

33. Qu, Z. Application of improved PCA in risks assessment technology of enterprise information security. in 2009 2nd International Conference on Power Electronics and Intelligent Transportation System (PEITS). 2009. IEEE.

34. Ru, G., et al., Machine Learning Techniques applied in risk assessment related to food safety. EFSA Supporting Publications, 2017. 14(7): p. 1254E.

35. Song, Y. and R. Wu, The Impact of Financial Enterprises' Excessive Financialization Risk Assessment for Risk Control based on Data Mining and Machine Learning. Computational Economics, 2022. 60(4): p. 1245–1267.

36. Tertychnyi, P., I. Slobozhan, M. Ollikainen, and M. Dumas. Scalable and imbalance-resistant machine learning models for anti-money laundering: A two-layered approach. in Enterprise Applications, Markets and Services in the Finance Industry: 10th International Workshop, FinanceCom 2020, Helsinki, Finland, August 18, 2020, Revised Selected Papers 10. 2020. Springer.

37. Wan, N., L. Li, C. Ye, and B. Wang, Risk assessment in intelligent manufacturing process: a case study of an optical cable automatic arranging robot. Ieee Access, 2019. 7: p. 105892-105901.

38. Wang, Y., Y. Zhang, Y. Lu, and X. Yu, A Comparative Assessment of Credit Risk Model Based on Machine Learning——a case study of bank loan data. Procedia Computer Science, 2020. 174: p. 141-149.

39. Wellsandt, S., et al., Hybrid-augmented intelligence in predictive maintenance with digital intelligent assistants. Annual Reviews in Control, 2022. 53(2022): p. 382-390.

40. Williams, B., et al., Real-time fall risk assessment using functional reach test.

International journal of telemedicine and applications, 2017. 2017(2042974): p. 1-8.

41. Wu, D.D., D.L. Olson, and C. Luo, A decision support approach for accounts receivable risk management. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2014. 44(12): p. 1624-1632.

42. Wu, J., C. Li, and Y. Huo, Safety assessment of dangerous goods transport enterprise based on the relative entropy aggregation in group decision making model. Computational intelligence and neuroscience, 2014. 2014: p. 28-28.

43. Wu, S. and H. Nagahashi, Parameterized AdaBoost: introducing a parameter to speed up the training of real AdaBoost. IEEE Signal Processing Letters, 2014. 21(6): p. 687-691.

44. Yang, R., et al., Big data analytics for financial Market volatility forecast based on support vector machine. International Journal of Information Management, 2020. 50: p. 452-462.

45. Yang, Y., et al., Unlocking the Power of Voice for Financial Risk Prediction: A Theory-Driven Deep Learning Design Approach. MIS Quarterly, 2023. 47(1): p. 63–96.

46. Zhandong, X. and G. Chi. Bank-enterprise project risk assessment model based on the information entropy method. in 2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC). 2011. IEEE.

47. Zhang, P.-B. and Z.-X. Yang, A novel AdaBoost framework with robust threshold and structural optimization. IEEE transactions on cybernetics, 2016. 48(1): p. 64-76.

48. Zhang, N., RETRACTED: Big data simulation for financial risk assessment of real estate bubble based on embedded system and artificial intelligence algorithm. Microprocessors and Microsystems, 2021. 82: p. 103941.