

Machine Learning Techniques in Financial Applications

Sami MESTIRI 

Faculty of Economic Sciences and Management of Mahdia
University of Monastir, Tunisia

Abstract

Over the past few years, the financial sector has witnessed an increase in the adoption of machine learning models within banking and insurance domains. Advanced analytic teams in the financial community are implementing these models regularly. This paper aims to explore the various machine learning approaches utilized in these sectors and offers recommendations for selecting suitable methods for financial applications. Additionally, the paper provides references to R packages that can be used to compute the machine learning methods. Our aim is to bring a valuable contribution to the field of financial research by providing a more comprehensive and advanced method of credit scoring, which in turn improves assessments of customers' debt repayment capabilities and improves risk management tactics.

Keywords: financial applications; machine learning (ML); credit scoring; R software.

JEL Classification: C45; G00.

Introduction

Artificial intelligence (AI) is used in machine learning (ML), which enables systems to learn from experience and get better without explicit programming. In effect, it is about developing predictive models that can access data and use it to learn on their own. There are several types of learning; we distinguish:

Supervised learning is done using a truth; that is, we have prior knowledge of what the output values for our samples should be. Therefore, the goal of this type of learning is to learn a function that, given a sample of data and the desired results, best approximates the relationship between observable inputs and outputs. There are two types of supervised learning. Classification algorithms, which seek to predict a class/ category, and regression algorithms, which seek to predict a continuous value.

Unsupervised learning aims at data structure inference. The two most common subcategories in unsupervised learning are clustering and dimensionality reduction. In clustering, observations are grouped in such a way as to produce high intra-group similarity and low inter-group similarity. The different types of clustering methods that have been proposed are entropy-based, density-based, and distribution-based methods. Reduction of dimensionality aims to increase the information density of the data by reducing its dimensionality while retaining most of the inherent information. There are different techniques based on principal component analysis (PCA) that derive linear combinations of the original variables to cover as much of the variance in the data as possible. Second, neural network-based methods reduce dimensionality with particular architectures.

In the dynamic and complex world of finance, accurately assessing market sentiment based on social media data is a crucial challenge with significant implications for investment decisions and risk management. Machine learning (ML) has become a powerful tool, offering new ways to extract information from vast volumes of text on social networks and uncover the collective emotional currents impacting markets. However, existing ML approaches often come up against the subjectivity and nuances inherent in human language, which can lead to misleading sentiment assessments.

This paper discusses the use of ML to solve problems in finance research. The contribution of this work is threefold. First, it provides an introduction to machine learning. Next, it pays particular attention to the different R packages implemented (Mestiri 2021). A taxonomy of current and future ML applications in finance is built. Finally, the prospects of ML applications in finance are studied. The research paper is organized as follows: Section 2 presents the different machine learning techniques used. Section 3, deals with a case study of credit risk. The fourth section is devoted to limitations and perspective. Finally, conclusion, limitation and perspectives are presented in Section 5.

1. Literature Review

A significant and growing part of economists is moving towards using the tools offered by ML to conduct innovative empirical analyses. The reason is twofold. ML has allowed economists to use new databases (multidimensional, images, texts) which until then remained unusable with traditional methods; it also opened the way for exploring new problems important to the discipline, notably problems where the prediction of an event is the main research question.

Thus, ML can be understood as a methodological but also conceptual advance in the discipline. It broadens the deductive approach in economics, it now also proposes to explore fields of research where we let the data speak in order to predict certain processes. In this sense, ML poses today as the best way to listen carefully to what the data has to tell us. ML therefore adds to the economist's toolbox not only to exploit new data and incorporate new methods, but also, ultimately, to address and solve new problems.

Algorithmic trading

Algorithmic trading refers to the use of algorithms to make better trading decisions. Usually, traders build mathematical models that monitor economic news and trading activities in real time to detect any factors that could force security prices up or down. The model comes with a set of predetermined instructions on various parameters such as timing, price, quantity and other factors to make trades without the active participation of the trader.

Unlike human traders, algorithmic trading can analyse large volumes of data simultaneously and make thousands of trades every day. Machine learning enables rapid trading decisions, giving human traders an advantage over the market average. Additionally, algorithmic trading does not make trading decisions based on emotions, which is a common limitation among human traders whose judgment may be affected by emotions or personal aspirations. The trading method is mainly used by hedge fund managers and financial institutions to automate trading activities.

A hedge fund wants to capitalize on short-term market movements. Their ML algorithm continuously analyses real-time news feeds, social media sentiment and market data to identify ephemeral opportunities and automatically execute trades, capturing profitable micro-trends that humans might miss. Renaissance Technologies, a quantitative investment company, uses complex ML algorithms to manage its high-frequency trading strategies, achieving impressive returns over decades.

In finance, our interest lies primarily in sentiment aggregated to markets such as the stock market, which is the most common target of ML-based sentiment measures. The majority of relevant studies use measures of sentiment towards stocks to study their effect on future stock returns. There are many studies that construct a measure of investor sentiment from social media e.g., Antweiler & Frank (2004) use naïve Bayes and SVM methods to classify user posts on the Yahoo Finance forum as positive or negative where they aggregate their classifications to construct a measure of stock market sentiment. In addition to text analyses, Obaid & Pukthuanthong (2022) apply machine learning to news photos to derive a sentiment measure for stocks and find that it can replace text-based measures. Other studies use analyst reports or annual reports to measure sentiment.

Sentiment analysis

Sentiment analysis be used to generate trading signals. Although correlations may exist, sentiment is only one factor among others in a complex market, and relying on it alone can be risky. It is essential to combine it with other fundamental and technical analyses. Sentiment data can be noisy and biased, reflecting individual opinions and agendas. Evaluate sources and potential biases carefully before drawing conclusions. Even if market sentiment could perfectly predict market movements, this would not guarantee profits due to market inefficiencies and competition. Market manipulation based on sentiment analysis raises ethical concerns. Ensure compliance with regulations and responsible use of information.

Measurements of the business leader's characteristics

The large quantity of image data available free of charge on the Internet allows numerous studies to exploit this information and extract several criteria such the appearance of business leaders, their personality traits, and these own convictions. Indeed, recent progress in ML also allows studies that construct measures of executive emotions. For example, Akansu et al. (2017) apply ML-based face reading to videos of CEOs during press interviews to extract facial emotions and quantify the CEO's mood, emotions such as anger, disgust, fear are measured, sadness, happiness or surprise and study their effect on company performance. Aslam et al. (2022) proposed a framework for fraud detection in the auto insurance industry, employing three predictive models (logistic regression, support vector machine, and naïve Bayes) to develop a fraud detection mechanism. Their study revealed that the support vector machine outperforms in terms of accuracy, while logistic regression achieves the highest f-measure score.

By analysing past data and executive characteristics, algorithms can predict potential future performance measures such as employee turnover, customer satisfaction or even financial results. Based on managers' individual characteristics and their impact on the organization, ML models can suggest personalized development plans, focusing on areas with the greatest potential for improvement. ML algorithms, when trained on diverse datasets and checked for fairness, can help identify and mitigate potential biases in measurement methods.

Machine learning models are only as good as the data they are trained on. Biased or incomplete data can lead to biased and inaccurate results, perpetuating existing inequalities. Ensure that diverse and representative datasets are used. Many ML models are complex and opaque, making it difficult to understand how they arrive at their conclusions. This can be problematic, especially when evaluating individuals. Choose interpretable models and explain their reasoning as much as possible. Data confidentiality and security are crucial, especially when dealing with sensitive information such as leadership assessments. Implement robust security measures and obtain informed consent from all participants. Human oversight and accountability: machine learning should not entirely replace human judgment. Use ML as a tool to inform decisions, not to automate them completely. Maintain human oversight and accountability throughout the process.

Studies construct measures of company characteristics with ML methods are mainly based on measures of financial characteristics and risk exposures of companies. Buehlmaier & Whited (2018) apply ML to annual reports to construct a measure of financial constraints. Cheraghali & Molnár (2023) review the methodologies used in the literature to predict failure in small and medium-sized enterprises. So, ML can also help study corporate culture. Li et al. (2021) extract aspects of corporate culture from conference call transcripts. Indeed, they study the effect on company performance measures such as operational efficiency and company value. Finally, the capabilities of ML enable the construction of new measures of business connectivity.

Measures of Credit Risk

Credit risk is a typical economic forecasting problem (Mestiri & Hamdi 2012). Its goal is to detect which potential borrowers will eventually default. Tantri (2021) predicts consumer credit default with strengthened regression trees based on borrower data. Mestiri (2024) use credit card transaction data to predict repayment patterns. Corporate credit risk is another area where ML can provide superior credit risk predictions. Tian et al. (2015) and Hamdi & Mestiri (2014) directly predict corporate bankruptcy from corporate financial statements and market data.

Analysing the determinants of firm-specific outcomes is an important topic of study in the field of corporate finance that can also be the target of ML-based forecasting. Two studies use ML to predict different financial results. Amini et al. (2021) study corporate capital structure as a typical problem in corporate finance and predict corporate leverage based on standard determinants of capital structure. Mestiri (2024) applied random forests to predict future profits of companies based on their accounting data. Corporate misconduct represents another forecasting problem. Bao et al. (2020) type of corporate misconduct we will study is accounting fraud. Rahman & Zhu (2024) utilize machine learning techniques to develop financial distress prediction (FDP) models for companies, then compare the classification performance of these models with conventional Z-Score models. Their results confirm that machine learning classifiers can effectively predict financial distress, highlighting the potential of these advanced techniques in enhancing predictive accuracy and reliability. Tron et al. (2023) conducted an analysis of the relationships between corporate governance characteristics and financial distress status comparing the predictive performance of corporate governance variables in anticipating corporate defaults, employing both the Logit and Random Forest models due the previous research which has recognized these models as among the most efficient machine learning techniques available. Xiang et al. (2012) apply ML-based textual analysis to predict start-up acquisitions based on company data. Non-parametric models (Mestiri & Farhat 2021) have been investigated in the literature.

An insurance company assesses the risk of insuring a new driver. Traditionally, they relied on factors such as age, location and driving history. Now, ML algorithms can analyse additional data such as social networking behaviour, driving telematics (collected via smartphone apps) and even weather conditions to create a more nuanced risk profile, leading to fairer, more accurate insurance premiums. Many banks use machine learning to assess creditworthiness beyond traditional credit ratings, making financial services accessible to underserved communities.

2. Research Methodology

Linear Discriminant Analysis (LDA)

Fisher (1933) pioneered work on discriminant analysis. In his work, he developed a statistical technique for defaults prediction, by developing a linear combination of quantitative predictor variables. This linear combination

of descriptors is called discriminant function. The output of ADL is a score that is consists of classify a data observation between the good and bad classes.

$$Score = \sum_{i=0}^p a_i X_i$$

where: a_i are the weights associated with the quantitative input variables X_i .

The *lda* function from the MASS library (Venables and Ripley, 2002) have been used to implement the discriminant analysis as follows:

```
lda_mod <- lda(Y~.,training_data)
```

Logistic Regression (LR)

Logistic regression is a statistical method used for binary classification tasks (e.g., 0 or 1, bad or good, health or default, etc.). Corresponding to Ohlson (1980), the outcome of LR model can be written as:

$$P(y = 1|X) = \text{sigmoid}(z) = \frac{1}{1+\exp(-z)}, \text{ where: } P(y = 1|X) \text{ is the probability of } y \text{ being 1, given the input variables } X, z \text{ is a linear combination of } X: z = a_0 + a_1X_1 + a_2X_2 + \dots + a_pX_p,$$

where: a_0 is the intercept term, a_1, a_2, \dots, a_p are the weights, and X_1, X_2, \dots, X_p are the input variables.

The *glm* functions from the stats library have been utilized for the estimation of the Logit:

```
logit_mod <- glm(Y ~ .,family=binomial, data = Training_data).
```

Decision Trees (DT)

Decision trees (DT) are typically not formulated in terms of mathematical equations, but rather as a sequence of logical rules that describe how the input variables are used to predict the output variable. However, the splitting criterion used to select the best split at each decision node can be expressed mathematically. The Gini impurity measures the probability of misclassifying an observation in S if we randomly assign it to a class based on the proportion of observations in each class (Gelfand et al., 1991). A small value of G(S) indicates that observations in S are well-separated by the input variables. The split with the smallest value of ΔG is chosen as the best split. The decision tree algorithm proceeds recursively, splitting the data at each decision node based on the best split, until a stopping criterion is met, such as reaching a maximum depth or minimum number of observations at a leaf node. The following R script runs the *rpart* function from the *rpart* package (Therneau & Atkinson, 1997), used for the Decision trees model:

```
DT_Mod <- rpart(formula = Y~ ., data = training_data, method = "class", parms = list(loss = , nrow = ))
```

Support Vector Machine (SVM)

Support vector machine (SVM), developed by Vapnik (1998), is a supervised learning algorithm used for classification, regression, and outlier detection. The basic idea of this technique is to find the best separating hyperplane between the two classes in a given dataset. The mathematical formulation of SVM can be divided into two parts: the optimization problem and the decision function.

The decision function takes an input vector x and returns its predicted class label based on whether the output of the hyperplane is positive or negative. The details of the optimization process are discussed in Acosta et al. (2022), Chang & Lin (2004), Cristianini & Shawe-Taylor (2000) and Gunn (1998).

Thereafter, SVM finds the best separating hyperplane by solving an optimization problem that maximizes the margin between the two classes, subject to constraints that ensure all data points are correctly classified with a margin of at least $1 - \xi_i$. The decision function then predicts the class label of new data points based on the output of the hyperplane. The *svm function* from the *e1071* library available on CRAN has been employed (Karatzoglou et al., 2004).

```
svm_mod <- svm(as.factor(Y) ~ ., data=training_data, , cost = 10, gamma = 1/length(data), probability = TRUE)
```

Random Forests (RF)

Random Forest is an ensemble of learning algorithm developed by Breiman (2001). It is a type of ensemble learning method that combines multiple decision trees for making predictions. The algorithm is called "random" because it uses random subsets of the features and random samples of the data to build the individual decision trees. The data is split into training and testing sets. The training set is used to build the model, and the testing set is used to evaluate its performance. At each node of a decision tree, the algorithm selects a random subset of the features to consider when making a split. This helps to reduce overfitting and increase the diversity of the individual decision trees. The following R script runs the *Random Forest* function from the *Random Forest package* (Liaw & Wiener 2002).

```
RF_mod <- randomForest (as.factor (Y) ~., data = Training_data, mtry=ncol(data)-1, ntree=1000)
```

3. Case Study

This section deals with a case study to demonstrate the effectiveness of machine learning techniques in addressing specific financial problems. In this empirical study, it is used a personal loan data set provided by a commercial bank in Tunisia. This data set contains both categorical and continuous data. For this analysis, a total of 12 variables are used; the first 12 variables are used to characterize each instance, and the final attribute is used to classify a transaction as good or bad. The various attributes, which are either categorical or numerical in nature, are shown in Table 2. The data consists of 688 personal loans with 577 good loans and 111 bad loans. The proportion of bad loans (default) compared to good loans (non-default) is 19.23.

Predictive performance measures

There are several criteria that can be utilized to compare and evaluate the predictive ability of the employed techniques including accuracy rate, *F1* score and AUC: The accuracy rate is the most famous performance metric, deduced from the matrix confusion The *F1* score is also computed from the confusion matrix. The value of *F1* score varies between 0 and 1, since 1 is the best possible score. A high *F1*-score indicates that the model shows both high precision and high recall, meaning it can correctly identify positive and negative cases. Area under Curve (AUC) is a synthetic indicator derived from the ROC curve. This curve is a graphical indicator used to assess the forecasting accuracy of the model The ROC curve is based on two relevant indicators that are specificity and sensitivity (Mestiri & Hamdi, 2012). This curve is characterized by the 1- specificity rate on the x axis and by sensitivity on the y axis.

Results and discussion

Table 1. Prediction results and models accuracy

Models	Accuracy rate	F1- score	AUC	Rank
Linear Discriminant Analysis (LDA)	70.9%	0.790	0.464	4
Logistic Regression (LR)	75.8%	0.822	0.533	2
Decision Trees (DT)	64.3%	0.738	0.575	5
Random Forest (RF)	78.2%	0.833	0.715	1
Support Vector Machine (SVM)	74.8%	0.810	0.563	3

According to Table 1, Random Forest outperforms the other techniques in terms of all prediction performance metrics. RF shows the highest accuracy rate with 78.2% for RF whereas 75.8% for LR. The lowest rate of prediction accuracy was found by the use of LDA (70.9%). For the same objective to assess the predictive ability of the proposed algorithms, *F1*-score equal to 0.833 proves *RF*'s ability to identify with a great precision good from bad customers. Since 1 is the best desired *F1* score, *RF* reaches the highest score while *F1* score value was equal to 0.822, 0.810, 0.790 and 0.738 for *LR*, *SVM*, *LDA* and *DT* respectively.

Other graphical indicator was also used to evaluate the quality of classification of the models under study, we talk about the *ROC* curve. From this curve, we deduce the *AUC* measure. More the *AUC* value is near to unity more the model shows high quality of classification between good and bad customers. Based on Table 2, the *AUC* of *RF* yields 0.715. In the second rank, we found the *DT* with *AUC* equals to 0.575. The *RL* and *LDA* models present the worst classification results as the *AUC* is 0.533 and 0.464, respectively, in the testing sample.

Table 2. List variables used in credit scoring modelling

ID variables	Description	Type
x_1	Age in years plus twelfths of a year	Numerical
x_2	Yearly income	Numerical
x_3	Credit length (in months)	Numerical
x_4	Amount of loans	Numerical
x_5	Length of stay (in years)	Numerical
x_6	Purpose	Categorical
x_7	Employment	Categorical
x_8	Type of house	Categorical
x_9	Gender	Categorical
x_{10}	Marital Status	Categorical
x_{11}	Education	Categorical
x_{12}	Number of dependents	Categorical
y	Default: Good Bad Indicator	Categorical

As a final conclusion, the ability of *RF* outperforms the statistical and conventional machine learning models in credit scoring. In the second rank we found that *LR* has a significantly higher prediction accuracy compared to other employed techniques. Based on our empirical investigation, the Random Forest can be considered as the best technique to detect customer's loan default and therefore can help to make managerial decisions.

4. Limitation and Perspective

Machine learning (ML) can be used to process numerical data that is high-dimensional or that has a high number of variables in comparison to the number of observations. These high-dimensional data arise if there are several economically relevant variables or if non-linearities and interaction effects play an important role. ML methods exploit the information content of this data to make predictions with small errors. Also, ML allows the exploitation of unconventional data, which is used to extract economically relevant information, which can then be a starting point for other economic analyses.

There are also limitations and drawbacks to using machine learning. First, ML methods tend to have low interpretability. It is often not directly observable how the algorithm generates its results, so ML is not generally suited to problems that require in-depth understanding. Second, ML requires large data sets. But unfortunately, large-scale data is not always available for many research questions in finance. Finally, the use of ML often incurs high computational costs, e.g., neural networks with complex architectures. However, many researchers are still unclear about how and where to apply ML in finance.

Machine learning relies heavily on data and models and lacks universality and autonomy. In terms of accuracy of forecasting, the performance of machine learning models is not always better than traditional methods, and the time and computing resources required for machine learning are much higher than traditional methods. This therefore requires high investments but has great potential.

Overall, machine learning holds immense potential in finance, but it's crucial to be aware of its limitations and to address them responsibly. As the technology evolves and ethical considerations become a priority, we can expect to see more robust and reliable applications emerge. Remember that ML is a tool, and, like any tool, its effectiveness depends on how it is used. Careful consideration of these limitations is essential to ensuring responsible and ethical implementation of ML in the financial sector.

Conclusion

In the finance industry, machine learning techniques are getting a lot of attention. The success of these methods stems from their ability to provide flexible, regularized approximations to the theoretically optimal decision rules in data-rich environments. The empirical application of machine learning in finance involves several methodological challenges. In this paper, I cover some of the interesting recent developments that address these challenges. This is an exciting and rapidly growing area of research that will need many interesting methodological developments and applications in the future.

Forecasting loan default has always been of great concern to credit financial institutions to make appropriate lending decisions. The purpose of this paper is to develop an effective model that can be used for the classification of credit applicants to make a good forecast of the financial difficulties of customers. In this paper, different machine learning techniques were compared to evaluate the credit risk in the Tunisian credit dataset. The empirical findings showed that RF is a highly suitable tool for studying financial defaults in Tunisian credit institutions. In future research, other technical aspects could be investigated, and the results of this study still need to be improved and discussed, especially to make a new model with higher performance than other previous models. Finding new or counterintuitive insights, finding a significant and meaningful new variable, etc.

The present work, as well as previous research, supports the idea that artificial intelligence models perform better than traditional methods. However, it will be interesting for further research to diversify the data sources and not use only classical numerical data (e.g., financial ratio data) but add textual data (e.g., news, public reports of companies, notes and comments from experts, auditors' reports, and management's statements) that can improve the accuracy of financial distress prediction.

Although machine learning offers interesting possibilities in finance, it is not without its limitations, as financial markets are inherently dynamic and unpredictable. Models trained on past data may have difficulty adapting to sudden changes or unforeseen events. Creating and maintaining complex ML models requires significant resources and expertise, which may not be feasible for all institutions. Regulatory frameworks around algorithmic trading and AI-based decision-making continue to evolve, posing implementation and compliance challenges.

Credit Authorship Contribution Statement

The author contributed to formulating research goals, collected, and compiled data from secondary sources to investigate it in a meaningful way. The methodology and formal analysis of the research was developed later by the author, followed by data investigation and validation through ML visualization techniques.

Acknowledgments

The author would like to thank eminent researchers across the globe who has worked in the domain of finance. The work of these professionals has helped author in framing the current research idea.

Conflict of Interest Statement

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- [1] Acosta, S. M., Amoroso, A. L., Sant'Anna, Â. M. O. et al. (2022). Predictive modelling in a steelmaking process using optimized relevance vector regression and support vector regression. *Annals of Operations Research*, 316, 905–926. <https://doi.org/10.1007/s10479-021-04053-9>
- [2] Antweiler, W., & Frank, M. Z. (2004). Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. *The Journal of Finance*, 59, 969-1442. <http://dx.doi.org/10.2139/ssrn.282320>
- [3] Akansu, A., Cicon, J., Ferris, S. P., & Sun, Y. (2017). Firm Performance in the Face of Fear: How CEO Moods Affect Firm Performance. *Journal of Behavioural Finance*, 18, 373-389. <https://doi.org/10.1080/15427560.2017.1338704>
- [4] Amini, S., Elmore, R., Öztekin, Ö., & Strauss, J. (2021) Can machines learn capital structure dynamics? *Journal of Corporate Finance*, 70. <https://doi.org/10.1016/j.jcorpfin.2021.102073>
- [5] Aslam,nF., Hunjra, A.I., Ftiti, Z., Louhichi, W., Shams, T. (2022). Insurance fraud detection: Evidence from artificial intelligence and machine learning, *Research in International Business and Finance*, 62, 101744. <https://doi.org/10.1016/j.ribaf.2022.101744>
- [6] Bao, Y., Ke, B., Li, B., Yu, Y.J. & Zhang, J. (2020). Detecting accounting fraud in publicly traded US firms using a machine learning approach. *The Journal of Accounting Research*, 58, 199 - 235 <https://doi.org/10.1111/1475-679X.12292>
- [7] Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- [8] Buehlmaier, M. M. M., & Whited, T.M. (2018). Are Financial Constraints Priced? Evidence from Textual Analysis. *Review of Financial Studies*, 31(7), 2693-2728. <https://www.jstor.org/stable/48615517>

- [9] Chang, C., & Lin, C. J. (2011). LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(27), 1-27. <https://doi.org/10.1145/1961189.1961199>
- [10] Cristianini, N., Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press, UK, 1-189. <https://doi.org/10.1017/CBO9780511801389>
- [11] Cheraghali, H. & Molnár, P. (2023). SME default prediction: A systematic methodology-focused review, *Journal of Small Business Management*. <https://doi.org/10.1080/00472778.2023.2277426>
- [12] Hamdi, M., & Mestiri, S. (2014). Bankruptcy prediction for Tunisian firms: An application of semi-parametric logistic regression and neural networks approach *Economics Bulletin*. 34(1), 133-143 <http://www.accessecon.com/Pubs/EB/2014/Volume34/EB-14-V34-11-P15.pdf>
- [13] Karatzoglou, A., Smola, A., Hornik, K. & Zeileis, A. (2004). Kernlab – An S4 Package for Kernel Methods in R, *Journal of Statistical Software*, 11(9), 1-20. [10.18637/jss.v011.i09](https://doi.org/10.18637/jss.v011.i09)
- [14] Li, K., Liu, X., Mai, F., & Zhang, T. (2021). The role of corporate culture in bad times: Evidence from the COVID-19 pandemic. *Journal of Financial and Quantitative Analysis*, 56, 2545 - 2583. <https://doi.org/10.1017/S0022109021000326>
- [15] Liaw, A., & Wiener, M. (2002). Classification and Regression by random Forest, *R news*, 2, 18–22.
- [16] Mestiri, S. (2021). Simulation de prêt personnel en utilisant R shiny. HAL Working Papers. <https://hal.science/hal-03448651>
- [17] Mestiri, S. (2024). Financial Applications of Machine Learning Using R Software. Available at SSRN Electronic. <http://dx.doi.org/10.2139/ssrn.4716425>
- [18] Mestiri, S., & Farhat, A. (2021). Using Non-parametric Count Model for Credit Scoring. *Journal of Quantitative Economics*, 19, 39-49. <https://doi.org/10.1007/s40953-020-00208-w>
- [19] Mestiri, S., & Hamdi, M. (2012). Credit Risk Prediction: A Comparative Study between Logistic Regression and Logistic Regression with Random Effects. *International Journal of Management Science and Engineering Management*, 7(3), 200-204. <https://doi.org/10.1080/17509653.2012.10671224>
- [20] Ohlson, J. A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy, *Journal of Accounting Research*, 18(1), 109–131. <https://doi.org/10.2307/2490395>
- [21] Obaid, K., & Pukthuanthong, K. (2022). A picture is worth a thousand words: Measuring investor sentiment by combining machine learning and photos from news, *Journal of Financial Economics*, 144(1), 273-297. <https://doi.org/10.1016/j.jfineco.2021.06.002>
- [22] Roy, T., Tshildzi, M., & Chakraverty, S. (2021). Speech emotion recognition using deep learning. *New Paradigms in Computational Modelling and Its Applications*, 177-187. <https://doi.org/10.1016/B978-0-12-822133-4.00009-8>
- [23] Rahman, J. & Zhu, H. (2024). Predicting financial distress using machine learning approaches: Evidence China, *Journal of Contemporary Accounting & Economics*, Volume 20, Issue 1, 100403. <https://doi.org/10.1016/j.jcae.2024.100403>
- [24] Shetty, S., Musa, M., & Brédart, X. (2022). Bankruptcy Prediction Using Machine Learning Techniques. *Journal of Risk and Financial Management*, 15(35), 1-10. <https://doi.org/10.3390/jrfm15010035>

- [25] Shin, K. S. & Lee, Y. J. (2002). A genetic algorithm application in bankruptcy prediction modelling. *Expert Systems with Applications*, 23, 321–328. [https://doi.org/10.1016/S0957-4174\(02\)00051-9](https://doi.org/10.1016/S0957-4174(02)00051-9)
- [26] Tantri, P. (2021). Fintech for the poor: Financial intermediation without discrimination, *Review of Finance*, 25, 561 - 593. <https://doi.org/10.1093/rof/rfaa039>
- [27] Tian, S., Yu, Y., & Guo, H. (2015). Variable selection and corporate bankruptcy forecasts, *Journal of Banking and Finance*, 52, 89 – 100 <https://doi.org/10.1016/j.jbankfin.2014.12.003>
- [28] Therneau, T. M., & Atkinson, E. J. (1997). An introduction to recursive partitioning using the RPART routines. Technical Report Mayo Foundation. <https://www.mayo.edu/research/documents/biostat-61pdf/doc-10026699>
- [29] Tron, A., Dallochio, M., Ferri, S. et al. (2023). Corporate governance and financial distress: Lessons learned from an unconventional approach. *Journal of Management and Governance*, 27, 425–456. <https://doi.org/10.1007/s10997-022-09643-8>
- [30] Vapnik, V. (1999). *The Nature of Statistical Learning Theory*. 2nd Edition New York: Springer Science & Business Media, 314 pp. ISBN-10: 0387987800, ISBN-13: 978-0387987804
- [31] Venables, W. N. & Ripley, B. D. (2002). *Modern Applied Statistics with S*, Springer, New York. <https://doi.org/10.1007/978-0-387-21706-2>

Cite this article

Mestiri, S. (2024). Machine Learning Techniques in Financial Applications. *Journal of Research, Innovation and Technologies*, Volume III, 1(5), 30-40. [https://doi.org/10.57017/jorit.v3.1\(5\).02](https://doi.org/10.57017/jorit.v3.1(5).02)

Article's history:

Received 5th of February, 2024; Revised 19th of February, 2024;

Accepted for publication 29th of February, 2024; Available online: 21st of March, 2024

Published 30th of June, 2024 as article in Volume III, Issue 1(5)

© The Author(s) 2024. Published by RITHA Publishing. This article is distributed under the terms of the license [CC-BY 4.0.](https://creativecommons.org/licenses/by/4.0/), which permits any further distribution in any medium, provided the original work is properly cited maintaining attribution to the author(s) and the title of the work, journal citation and URL DOI.