Predictive Analytics for Financial Risk Management in Dynamic Markets

Serkan Yilmaz Kandir[1,*], Murat Ismet Haseki[2]

[1]Faculty of Economics and Administrative Sciences, Adana, Turkey

[2]Faculty of economics and administrative sciences, Cukurova University, Adana, Turkey

Emails: skandir@cu.edu.tr; mhaseki@cu.edu.tr

Abstract

effective risk management is an indispensable requirement for improving the flow of transactions in dynamic financial markets. To this end, this study presents an applied predictive analytics methodology, that integrate gradient boosting algorithm to model the risk behavior in dynamic markets. This study, based on predictive analytics in monetary and financial systems, faces an urgent need for robust models that can overcome the uncertainties inherent in dynamic markets. Holistic experimentations on public case study of U.S retail data demonstrate the predictive power of the proposed approach of the state-of-the-art techniques across different performance metrics. This in turn highlights the nuanced interaction between variables and delivering intuitions into crucial risk determining factor.

Keywords: Business Intelligence; Machine Learning; Risk Managements; Market Analysis; Predictive Modeling

1. Introduction

The modern economic landscape is constantly evolving, with market conditions changing in ways bringing both opportunities and challenges. In this realm effectively managing risk is crucial, for stability and growth. The emergence of analytics has revolutionized the way we understand and navigate the risks present in these ever-changing markets [1-3]. It is essential to explore the role of analytics in financial risk management and its ability to identify and mitigate risks in dynamic markets. By harnessing the power of data driven approaches predictive analytics allows us to anticipate market volatility, recognize trends and proactively manage risks [4]. This empowers decision makers, with foresight. As financial markets venture into territories, the importance and urgency of leveraging predictive analytics cannot be emphasized enough when it comes to reducing risks and optimizing strategies without going overboard [5-7].

Moreover, this research paper aims not to explore the foundations and principles that support predictive analytics but also to provide a practical understanding of its application in ever changing markets [8-11]. It intends to shed light on the relationship between analytics and financial risk management showcasing how these tools serve as valuable assets for organizations to adapt, thrive and maintain resilience in an environment characterized by constant change. Through the integration of insights, real world applications and case studies our study endeavors to elucidate the multifaceted aspects of using analytics for managing financial risks in dynamic markets [12-14]. Specifically, we propose an applied framework that combines gradient boosting trees to conduct risk analysis on data. in addition, our framework is armed with set of preprocessing techniques to ensure clarity of the data before fed into analysis phases. The framework also integrates actionable data exploration techniques to navigate the labyrinthine panorama of economic retail risk in recent day volatile markets. The remainder of this work is structured as summarized in Figure 1.

DOI: https://doi.org/10.54216/AJBOR.110106
Received: May 17, 2023 Revised: October 08, 2023 Accepted: December 07, 2023

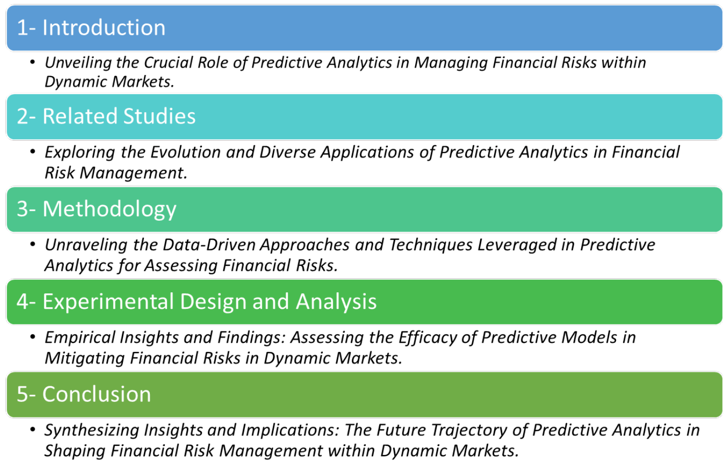

Figure 1: Roadmap to the main sections of our paper.

2. Related Works

In this section, we begin to explore the major seminal studies and research efforts that have laid the foundation for predictive analytics in financial risk management. Qiao and Belling [9] examined decision analysis and machine learning in economics and monetary policy and emphasized their critical role in strategic decision making Bouchaud and Potters [10] provided a theoretical foundation a about financial risk and price derivatives presented, presented a unique approach to statistical physics risk management. Kim et al [11] presented a complex case using deep learning to predict risk in retail investors shedding light on the ability of advanced methods to predict financial risk behavior. Bahrami and Shokouhyar [12] advocated a dynamic capability approach, examined the importance of big data analytics the ability to strengthen supply chain efficiency and firm performance. Tran [13] Proposed a conceptual framework for the application of big data analytics, especially in the banking sector, outlining its potential uses and implications.

Furthermore, Zhu and Yang [14] examined the transformational role of big data analytics in economic performance and sustainable development, emphasizing its multidimensional impact. Rasmussen [15] presented an original work on risk management in dynamic societies, which addresses complex risk modeling in such contexts. Clintworth, Liridis, Boulogouris [16] introduced a machine learning-based approach to financial risk assessment in the shipping industry, aimed at improving risk management DiCuonzo et al. [17] by jokhim management 4.0, especially in the banking community, has been analyzed to have been analyzed, as well as the possibilities and probabilities and possibilities and possibilities and possibly improvements, Shih et al. [18] and the amount of Stock volume to enhance the pre-chetavankana. Jokhim Business and Policy Contributions to Hall [19] examined the use of text analysis in risk management in financial institutions and highlighted the potential of data analysis techniques in improving risk assessment processes.

3. Methodology

This part of our paper is a foundation for developing a predictive analytics framework for managing financial risk through outlining the systematic approach used to build, train, and validate the predictive models used in this study.

In the present research, the GBDT was used as a novel machine-learning technique in the discipline of urban growth and planning. Using this method, recent research has looked into the association between sales information, as well as the market dynamic surroundings and sales movement [4, 12]. GBDT is superior to the conventional regression model in several ways. First off, while retaining a comparatively high prediction accuracy, GBDT can effectively handle complex and non-linear correlations [3, 13]. One explanation for this is that the GBDT method classifies predictors using decision trees and estimates the result by reducing the loss rate. The study states that "models are fit by decreasing a measure of loss, which might be the squared-error or a likelihood-based loss function, aggregated over the learning data" [10]. Second, instead of picking lacking data at arbitrary, the GBDT can manage lacking data properly by marking it with constituting information. As a result, the missing value is handled as a new category

DOI: https://doi.org/10.54216/AJBOR.110106
Received: May 17, 2023 Revised: October 08, 2023 Accepted: December 07, 2023

throughout the tree-building process. Finally, while this is difficult to accomplish with typical statistical models, the GBDT can calculate and order the level of significance among forecasters pertaining to the response variable [44].

The following is a summary of the equation, which comes from the research [4]:

Start from scratch:

$$f_0(x) = \operatorname{argmin}_\gamma \sum_{i=1}^{N} L(y_i, \gamma) \quad (1)$$

The formula is generated in three phases. In order to minimize the loss function, the method first sets up the potential fixed model. The formula is then built in the second phase, which involves four sub-steps for each relationship m.

$$r_{\min} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f=f_{m-1}} \quad (2)$$

The estimation of a generalization or pseudo residue is obtained by calculating the loss function's downward gradient. The correlation tree is further fitted to the goal.

$$R_{jm},\ j = 1, 2, \ldots, J_m \quad (3)$$

It is significant to point out that the algorithm has two layers of cycles stacked within it: the number of samples i and the number of repetition cycles m. Using the computed formula to fit into a regression tree of CART yields the regression tree of cycles m. J denotes every component tree's size as a function of the sub-step.

$$\gamma_{\min} = \underset{\gamma}{\operatorname{argmin}} \sum_{x_i \in R_{jm}} L(y_i, f_{m-1}(x_i) + \gamma) \quad (4)$$

$$f_m(x) = f_{m-1}(x) + \sum_{j=1}^{Jm} \gamma_{jm} I\left(x \in R_{jm}\right) \quad (5)$$

$$f(x) = f_M(x) = f_0(x) + \sum_{m=1}^{M} \sum_{j=1}^{J} c_{jm} I\left(x \in R_{jm}\right) \quad (6)$$

GBDTs modify models by decreasing the anticipated value of a loss function, and they develop models step-by-step. The GBDT uses a shrinkage method to reduce over-fitting and boost prediction accuracy [5, 6]. But if the training data are matched too tightly, an excessive fitting issue may arise. Consequently, in order to mitigate the over-fitting issue and enhance prediction precision, three factors are incorporated: the quantity of trees (M), learning ratio (ξ), and tree intricacy (C). Keep in mind that the learning rate (decrease) is determined to scale the role of each basic tree model by adding a factor of ξ, as follows. The tree difficulty influences the model difficulty and the degree to which the model fits.

$$f_m(x) = f_{m-1}(x) + \xi \times \beta_m \sum \frac{J}{j=1} \gamma_{jm} I(x \in R_{jm}), where\ 0 < \xi \leq 1 \quad (7)$$

The complete procedure of gradient boosting is given in Algorithm 1.

Algorithm 1: Gradient Boost algorithm

1: **Inputs:**
2: input data $(x, y)_{i=1}^{N}$
3: number of iterations M
4: choose of the loss-function $\Psi(y, f)$
5: choosing of the base-learner model $h(x, \theta)$
6: **Procedure:**
7: initialize $\hat{f}_0$ with a constant
8: for $t = 1$ to M do
9: calculate the negative gradient $g_t(x)$
10: fit a new base-learner function $h(x, \theta_t)$
11: discover the best gradient descent step-size ρ_t :

DOI: https://doi.org/10.54216/AJBOR.110106
Received: May 17, 2023 Revised: October 08, 2023 Accepted: December 07, 2023

12: $\rho_t = \arg\min_{\rho} \sum_{i=1}^{N} \Psi[y_i, \hat{f}_{t-1}(x_i) + \rho h(x_i, \theta_t)]$
13: update the function estimate:
14: $\hat{f}_t \leftarrow \hat{f}_{t-1} + \rho_t h(x, \theta_t)$
15: End for.

In our experiments, we used common regression analysis metrics to evaluate the performance of the predictive models. The mathematical definitions of these metrics are expressed as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \quad (8)$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \quad (9)$$

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \cdot 100\% \quad (10)$$

$$R2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}, for\ \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i \quad (11)$$

$$\text{RMSLE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(log(\hat{y}_i + 1) - log(y_i + 1)\right)^2} \quad (12)$$

4. Experimental Design and Analysis

In this section, we discuss the details of conducted experimentations and related results to prove the validity of the presented predictive modeling approach.

The experimentation of this work is conducted on public case study containing sales details of different stores of a supermarket chain that has multiple stores in different parts of the US. The data is composed of the following set of features Ship Mode, Segment, Country, City, State, Postal code, Region, Category, Sub-category, Sales, Quantity, Discount, and Profit. Data preparation is required for the application of ML algorithms. Variable selection determines which variables have importance for forecasting green building performance. Pre-processing includes outlier removal, data noise reducing, standardization, and normalization. The data collected includes numerical values for all variables. After selecting the appropriate variables, pre-processing is required, which includes outlier elimination and dataset normalization. During the initial compilation of data process, outliers are eliminated from the collected data. Interquartile intervals were employed in the present investigation to detect extreme and outlier outcomes.

In the early phase of our experiments, we conduct some exploratory analysis to get some valuables insights about the dataset. we show the results of Statistical analysis in Table 1, in terms of central tendency measures for each feature. It is worth noting that average values of different features vary significantly, which might cause difficulty for model training. Also, we can observe that the count statistics are the same for all features, which implies that the dataset is free from null or missing values.

Table 1: Summary of central tendency measure across different features of the training dataset.

	count	mean	std	min	0.25	0.5	0.75	max
Postal Code	9994	55190.38	32063.69	1040	23223	56430.5	90008	99301
Sales	9994	229.858	623.2451	0.444	17.28	54.49	209.94	22638.48
Quantity	9994	3.789574	2.22511	1	2	3	5	14
Discount	9994	0.156203	0.206452	0	0	0.2	0.2	0.8
Profit	9994	28.6569	234.2601	-6599.98	1.72875	8.6665	29.364	8399.976

In addition, as an essential part of exploratory analysis, we provide a visualizing the distribution of sales in different scenarios to elucidate the underlying characteristics and relationships within the data. As shown in Figure 2, we display

DOI: https://doi.org/10.54216/AJBOR.110106
Received: May 17, 2023 Revised: October 08, 2023 Accepted: December 07, 2023

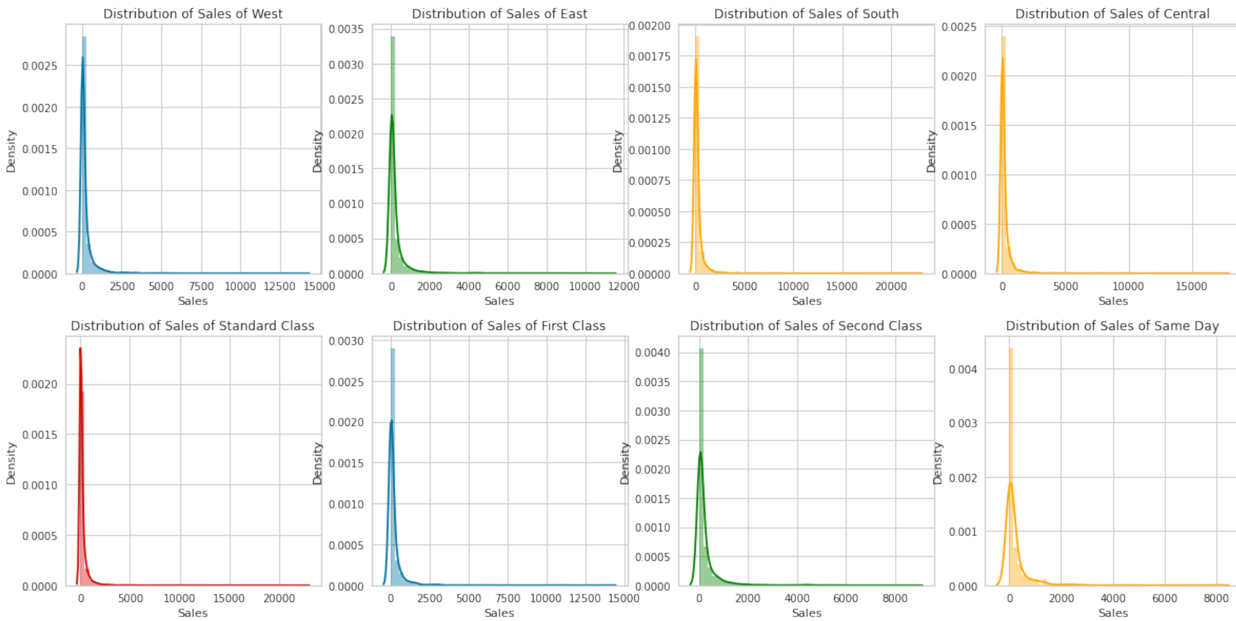

Figure 2: visualization of sales information under different settings.

the distribution of sales of standard class, distribution of sales of first class, distribution of sales of second class, distribution of sales of same day. Also, we provide visualization of distribution of sales of west, east, south, and central. Besides, in Table 2, we tabulated the results obtained from the Pearson correlation analysis, which reflect the interrelationships between variables in the market data with correlation coefficients ranging from -1 to 1.

Table 2: The results of Pearson correlation analysis between features.

	Postal Code	Sales	Quantity	Discount	Profit
Postal Code	1	-0.02385	0.012761	0.058443	-0.02996
Sales	-0.02385	1	0.200795	-0.02819	0.479064
Quantity	0.012761	0.200795	1	0.008623	0.066253
Discount	0.058443	-0.02819	0.008623	1	-0.21949
Profit	-0.02996	0.479064	0.066253	-0.21949	1

In addition, in Table 3, we provide a comparative analysis among diverse predictive modeling algorithms (e.g., CatBoost Regressor, Light Gradient Boosting Machine, Random Forest Regressor, Extreme Gradient Boosting, Extra Trees Regressor, Decision Tree Regressor, K Neighbors Regressor, Huber Regressor, Passive Aggressive Regressor, Lasso Regression, Bayesian Ridge, Ridge Regression, Linear Regression, Least Angle Regression, Elastic Net, Lasso Least Angle Regression, Orthogonal Matching Pursuit, Dummy Regressor, AdaBoost Regressor) employed in assessing financial risks within dynamic markets. It could be noted that the Gradient Boosting Regressor seems to have the lowest prediction error (MAE: 91.73, MSE: 156071.3, RMSE: 344.3419) among different regressors involved in the comparison.

Table 3: Comparative analysis between different predictive modeling algorithms.

Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
Gradient Boosting Regressor	91.7307	156071.3	344.3419	0.6315	0.7335	1.0031
CatBoost Regressor	100.1096	171326.1	378.226	0.6246	0.8166	1.294
Light Gradient Boosting Machine	102.1397	159542.3	359.1345	0.6756	0.7599	1.189
Random Forest Regressor	106.8707	148552.6	347.7607	0.6696	0.8744	1.5912

DOI: https://doi.org/10.54216/AJBOR.110106
Received: May 17, 2023 Revised: October 08, 2023 Accepted: December 07, 2023

Extreme Gradient Boosting	107.1741	139814.6	334.5426	0.6939	0.9426	1.7777
Extra Trees Regressor	117.0605	144970.8	341.2783	0.7003	1.042	2.319
Decision Tree Regressor	124.0199	213363.2	422.5852	0.4405	0.9107	1.6922
K Neighbors Regressor	130.2343	161736.2	367.2034	0.6535	1.009	1.5814
Huber Regressor	146.3698	358164.8	551.5956	0.0877	1.0444	1.752
Passive Aggressive Regressor	147.4644	375190.9	565.0838	0.0375	0.998	1.4854
Lasso Regression	192.1984	282649.4	503.5164	0.3151	1.436	4.3952
Bayesian Ridge	192.6441	282654.1	503.5142	0.315	1.4421	4.4415
Ridge Regression	192.736	282662.8	503.52	0.3147	1.4423	4.4487
Linear Regression	192.7407	282664.9	503.5219	0.3147	1.4423	4.4491
Least Angle Regression	192.7407	282664.9	503.5219	0.3147	1.4423	4.4491
Elastic Net	194.221	291619.8	513.1678	0.3114	1.4622	4.3974
Lasso Least Angle Regression	199.2832	301547.2	522.4468	0.2859	1.5909	5.4857
Orthogonal Matching Pursuit	225.4904	400853.2	607.5682	0.011	1.6645	6.5842
Dummy Regressor	273.1895	411129.1	619.7359	-0.0026	2.0541	11.0882
AdaBoost Regressor	578.5163	703115	751.6361	-1.7038	2.474	26.0604

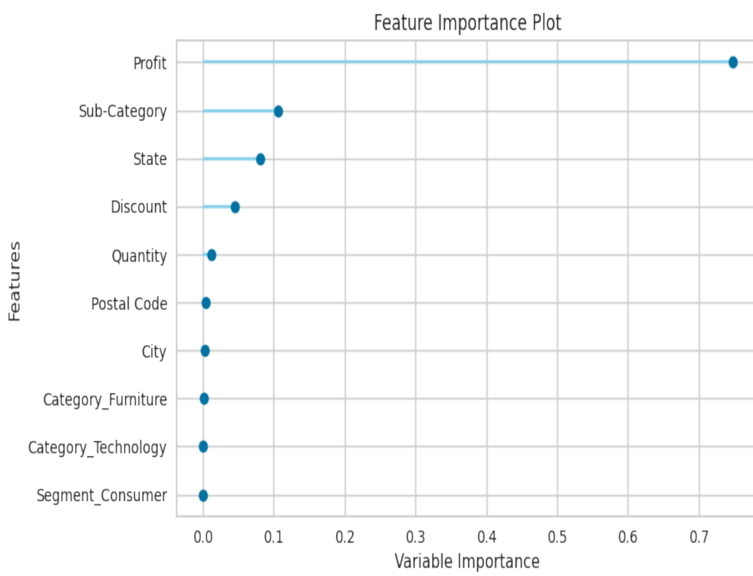

Figure 3: Visualization of feature importance for the applied gradient boosting.

In Figure 3, we visualize feature importance plot for the introduced gradient boosting regressor to elucidate the extent by with each feature contribute to the final model prediction. As shown, the bars distill complex insights by ranking the features according to their influence on the model's predictive capability. It is worth noting that the profit feature attains the highest importance with a 0.8 score. Then, the sub-category and state variables are the next important features with an average score near 0.1.

Moreover, in Figure 4, we present an insightful illustration of the residuals stanching from our model's predictions, which reflect the variance between the projected values and the actual viewed results within the data. The visualized residuals show the differences or flaws in our model's prediction ability for financial risk. A well-fitted model is expressed by a patternless distribution of residuals around the zero line, indicating that predictions closely match actual observations. In opposition, detectable patterns, or outliers in this plot highlight areas where the model may falter or exhibit biases, which promote the developer to improve and tune the model's architecture.

DOI: https://doi.org/10.54216/AJBOR.110106
Received: May 17, 2023 Revised: October 08, 2023 Accepted: December 07, 2023

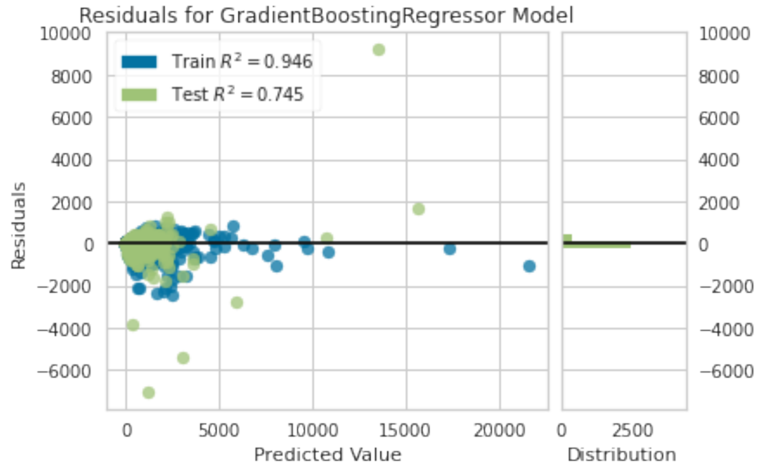

Figure 4: Visualization of residual plot for the applied gradient boosting

5. Conclusion and Future work

This study explores a new mechanism that integrates gradient boosting methodology for predictive modeling within financial risk management. The use of this approach allowed for the development of robust predictive models capable of navigating the complexities of ever-evolving economic events The findings demonstrated the importance of predictive analytics in the identification and management of risks and highlighted the critical role of gradient enhancement to increase predictive accuracy and provide information on strategic decision-making. Although feature importance analysis revealed important variables affecting prediction, which enabled stakeholders to prioritize risks and modify risk mitigation strategies, analysis of residuals revealed areas for model refinement, and emphasized the importance of continuous model calibration and improvement. Going forward, future research efforts in this area may explore more ways to enhance the use of gradient enhancement in financial risk management. First, diving in ensemble techniques and hybrid models that incorporate gradient enhancement can increase predictive power, especially in capturing nonlinear relationships and complex interactions between variables is possible. In addition, it can provide a comprehensive framework for overall risk management by expanding the analysis to include a wider range of risks, including operational and systemic risk. Finally, continuous validation and optimization of models.

References

[1] Broby, D. (2022). The use of predictive analytics in finance. *The Journal of Finance and Data Science*, *8*, 145-161.

[2] Gu, X., Mamon, R., Duprey, T., & Xiong, H. (2021). Online estimation for a predictive analytics platform with a financial-stability-analysis application. *European Journal of Control*, *57*, 205-221.

[3] Yang, R., Yu, L., Zhao, Y., Yu, H., Xu, G., Wu, Y., & Liu, Z. (2020). Big data analytics for financial Market volatility forecast based on support vector machine. *International Journal of Information Management*, *50*, 452-462.

[4] Lin, E. M., Sun, E. W., & Yu, M. T. (2020). Behavioral data-driven analysis with Bayesian method for risk management of financial services. *International Journal of Production Economics*, *228*, 107737.

[5] Feng, R., & Qu, X. (2022). Analyzing the Internet financial market risk management using data mining and deep learning methods. *Journal of Enterprise Information Management*, *35*(4/5), 1129-1147.

[6] Dash Wu, D. (2020). Data intelligence and risk analytics. *Industrial Management & Data Systems*, *120*(2), 249-252.

[7] Ghosh, I., Jana, R. K., & Sanyal, M. K. (2019). Analysis of temporal pattern, causal interaction and predictive modeling of financial markets using nonlinear dynamics, econometric models and machine learning algorithms. *Applied Soft Computing*, *82*, 105553.

[8] Han, L. (2019). Correlation predictive modeling of financial markets. *Procedia Computer Science*, *154*, 738-743.

DOI: https://doi.org/10.54216/AJBOR.110106
Received: May 17, 2023 Revised: October 08, 2023 Accepted: December 07, 2023

[9] Qiao, Q., & Beling, P. A. (2016). Decision analytics and machine learning in economic and financial systems. *Environment Systems and Decisions*, *36*, 109-113.

[10] Bouchaud, J. P., & Potters, M. (2003). *Theory of financial risk and derivative pricing: from statistical physics to risk management*. Cambridge university press.

[11] Kim, A., Yang, Y., Lessmann, S., Ma, T., Sung, M. C., & Johnson, J. E. (2020). Can deep learning predict risky retail investors? A case study in financial risk behavior forecasting. *European Journal of Operational Research*, *283*(1), 217-234.

[12] Bahrami, M., & Shokouhyar, S. (2022). The role of big data analytics capabilities in bolstering supply chain resilience and firm performance: a dynamic capability view. *Information Technology & People*, *35*(5), 1621-1651.

[13] Phan, D. T., & Trang Tran, L. Q. (2022). Building a Conceptual Framework for Using Big Data Analytics in the Banking Sector.

[14] Zhu, X., & Yang, Y. (2021). Big data analytics for improving financial performance and sustainability. *Journal of Systems Science and Information*, *9*(2), 175-191.

[15] Rasmussen, J. (1997). Risk management in a dynamic society: a modelling problem. *Safety science*, *27*(2-3), 183-213.

[16] Clintworth, M., Lyridis, D., & Boulougouris, E. (2021). Financial risk assessment in shipping: a holistic machine learning based methodology. *Maritime Economics & Logistics*, 1-32.

[17] Dicuonzo, G., Galeone, G., Zappimbulso, E., & Dell'Atti, V. (2019). Risk management 4.0: The role of big data analytics in the bank sector. *International Journal of Economics and Financial Issues*, *9*(6), 40-47.

[18] Shih, D. H., Hsu, H. L., & Shih, P. Y. (2019, April). A study of early warning system in volume burst risk assessment of stock with Big Data platform. In *2019 IEEE 4th international conference on cloud computing and big data analysis (ICCCBDA)* (pp. 244-248). IEEE.

[19] Haile, I. M. (2020). *Data Analytics in Financial Institutions: How Text Analytics Can Help in Risk Management* (Doctoral dissertation, Colorado Technical University).

[20] Kothandapani, H. P. (2023). Applications of Robotic Process Automation in Quantitative Risk Assessment in Financial Institutions. *International Journal of Business Intelligence and Big Data Analytics*, *6*(1), 40-52.

DOI: https://doi.org/10.54216/AJBOR.110106

Received: May 17, 2023 Revised: October 08, 2023 Accepted: December 07, 2023