Jurnal Sains Informatika Terapan (JSIT)

E-ISSN: 2828-1659, Volume: 04, Issue: 03, Month: Oktober, Year: 2025
Page: 589 - 595, Available online at: https://rcf-indonesia.org/home/

IMPLEMENTATION MACHINE LEARNING ALGORITHMS TO PREDICT THE FINANCIAL RESILIENCE OF COMPANIES BASED ON FINANCIAL STATEMENTS

Nadya Andhika Putri[1], Nico Gawa Lahara[2]

[1,2] Universitas Pembangunan Panca Budi

Corresponding Author: [1] nadyaandhika@dosen.pancabudi.ac.id

Article Info

Article history:

Received: Oct 20, 2025
Revised: Oct 22, 2025
Accepted: Oct 23, 2025
Published: Oct 30, 2025

Keywords:

Machine Learning
Financial Resilience
Predictive Modeling
Financial Statements
Data-Driven Decision Making

ABSTRACT

This study explores the application of machine learning algorithms to predict the financial resilience of companies based on their financial statements. In an era of data-driven decision-making, traditional financial analysis methods may fall short in providing timely and accurate insights. By leveraging advanced machine learning techniques, such as regression models, decision trees, and neural networks, this research aims to create predictive models that can effectively forecast a company's financial health. The study utilizes historical financial data, including balance sheets, income statements, and cash flow reports, to train and test various machine learning models. The findings highlight the potential of machine learning in identifying patterns and trends within financial data that may not be readily apparent through conventional methods. The results can provide valuable tools for financial analysts, investors, and company managers to assess and mitigate financial risks, enhancing decision-making processes and strategic planning.

1. INTRODUCTION

Machine learning (ML) has become an essential component in various fields across both theoretical and practical applications, enhancing the ability of systems to analyze data, learn patterns, and make informed decisions with minimal human intervention. This overview synthesizes recent literature regarding machine learning methodologies, their diverse algorithms, and the various domains they impact.

Machine learning encompasses several key categories, including supervised learning, unsupervised learning, and reinforcement learning, which are integral in areas such as consumer banking and image analysis. Kaminskyi et al. explored these classifications and confirmed their applicability in economic contexts, highlighting how ML algorithms adapt to changing economic environments and consumer behaviors [1]. Additionally, Khan et al. emphasized the effectiveness of deep learning techniques for tasks such as MRI image segmentation, noting that while traditional ML methods may struggle with large datasets, deep learning excels when provided with ample data [2]. This notion of leveraging various ML techniques for specific applications illustrates the versatility and adaptation of ML tools.

Specific algorithms play a pivotal role in harnessing the power of machine learning. For instance, Bellinger et al. presented a hybrid system incorporating diverse ML methods to forecast air pollution more accurately, showcasing the combination of traditional and advanced techniques [3]. Similarly, tools like Principal Component Analysis (PCA) and artificial neural networks (ANN) are noted for enhancing decision-making processes across different environmental assessments and diagnostics [4]. Furthermore, Zhao elucidates how the transition from manual feature extraction to more automated processes through end-to-end model data has broadened the scope of asset pricing methodologies, demonstrating shifts in application approaches within financial systems [5].

The convergence of machine learning with deep learning has also yielded significant advancements across various sectors, particularly in healthcare. Studies such as those by Deo and Johnson highlight both the potential and challenges that ML brings to medical diagnostics, stressing the need for integrative solutions between ML researchers and healthcare professionals to improve clinical outcomes [6][7]. Despite the promising capabilities of ML models in areas such as diabetic retinopathy screening and sepsis diagnosis, their actual impact on clinical practice has been less pronounced, as noted in extensive reviews [8][9]. This disconnect raises questions about how effectively ML can translate technical performance into real-world benefits, which remains an area ripe for exploration and research.

Within the classification realm, the recent literature reviews various algorithms, from decision trees to deep learning models. The synthesis of these approaches demonstrates that while traditional classification methods provide robust frameworks for general tasks, newer algorithms like convolutional neural networks (CNNs) have revolutionized fields such as image processing and sentiment analysis, proving superior in many contexts [10][11]. The integration of machine learning with domain-specific knowledge, such as in food safety Deng et al. [12] and flood prediction [13], reflects an increasing trend toward specialized applications, encouraging continued research in these intersections.

Forecasting is an essential aspect of many fields, including healthcare, economics, and telecommunications. Its significance lies in the ability to predict future events based on historical data, thereby assisting in decision-making processes. One widely used method in time series forecasting is the Seasonal Autoregressive Integrated Moving Average (SARIMA) model, which captures temporal structures in data while accommodating seasonal variations. SARIMA extends the Autoregressive Integrated Moving Average (ARIMA) model by adding seasonal components, effectively modeling a series with seasonal patterns. The SARIMA model is commonly applied in various domains, including financial forecasting and demand planning in healthcare settings.

While the SARIMA model addresses predictive analytics, survival analysis methods, particularly the Kaplan-Meier estimator, are crucial for analyzing time-to-event data. The Kaplan-Meier estimator provides a non-parametric approach for estimating survival probabilities from incomplete data. This method, first proposed by Kaplan and Meier in 1958, accounts for censored observations—cases where the event of interest (e.g., death, failure) has not occurred or is unknown at the time of the analysis. Its applicability spans diverse fields, such as medicine and reliability engineering, where it is essential for estimating the time until an event occurs and for visualizing this data through survival curves [14][15][16].

The integration of the Kaplan-Meier estimator with other statistical models enhances its application in various research areas. For instance, it is often employed alongside the Cox proportional hazards model to investigate the effects of covariates on survival times, allowing researchers to evaluate prognostic factors while accounting for censoring [17][18]. In clinical trials, Kaplan-Meier curves assist in estimating the efficacy of treatments over time, providing valuable insights into patient outcomes under different treatment conditions [19][20]. However, the methodology has limitations, particularly in the presence of competing risks—events that preclude the occurrence of the primary event of interest. Such scenarios may lead to biased Kaplan-Meier estimates, prompting researchers to consider alternative estimators, such as cumulative incidence functions, to address competing risks more adequately [21][22][23][24].

In recent literature, various extensions and modifications of the Kaplan-Meier estimator have been proposed to mitigate its shortcomings. Techniques such as the inverse probability treatment weighting (IPTW) offer a way to handle bias due to non-random censoring, enhancing the validity of Kaplan-Meier estimates by accounting for confounders [25][26]. Moreover, studies have indicated that adjusted estimators can provide a more accurate representation of survival probability in cases with time-varying covariates [27][28]. These advancements illustrate a growing recognition of the need for robust methodologies in survival analysis, particularly as datasets become increasingly complex.

2. MATERIALS AND METHODS

The Materials and Methods section suitable for your study titled *"Implementation of Machine Learning Algorithms to Predict the Financial Resilience of Companies Based on Financial Statements"*

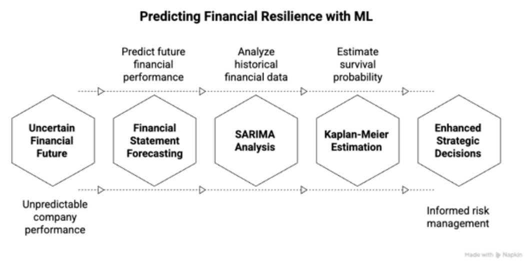

Figure 1. Research Methode

2.1 Data Collection

The dataset used in this study consists of financial statement data from publicly listed companies obtained from reliable financial databases such as Yahoo Finance, Bloomberg, or corporate annual reports. The dataset includes key financial indicators such as total assets, liabilities, equity, revenue, net income, and cash flow from the past five to ten years. The data were cleaned and standardized to ensure accuracy and consistency before analysis.

2.2 Data Preprocessing

Preprocessing steps included handling missing values, normalization, and feature selection. Outliers were detected using the interquartile range (IQR) method and treated appropriately. All numeric variables were normalized using Min-Max scaling to ensure uniformity across features. Feature selection was performed using correlation analysis and recursive feature elimination (RFE) to identify the most relevant variables contributing to financial resilience.

2.3 Definition of Financial Resilience

Financial resilience was defined as a company's ability to maintain stable financial performance and solvency under economic stress conditions. A binary classification label was created: *resilient* (1) and *non-resilient* (0), determined based on financial ratios such as the debt-to-equity ratio, current ratio, and profitability trends over time.

2.4 Machine Learning Algorithms

This study employs two advanced techniques to predict financial resilience:

1. SARIMA (Seasonal AutoRegressive Integrated Moving Average)**:** SARIMA was used for time-series forecasting of key financial indicators such as revenue, cash flow, and profitability trends over time. This model captures both the linear and seasonal components of the financial data, allowing for accurate forecasting of a company's financial health in the future. The SARIMA model was fitted on historical data, with the hyperparameters (p, d, q, P, D, Q) selected based on the AIC (Akaike Information Criterion) for model optimization.
2. Kaplan-Meier Estimator: The Kaplan-Meier estimator was used to analyze the survival function of companies, helping to understand the probability of a company's financial resilience over time. By treating the time to financial distress (e.g., bankruptcy, liquidity crises) as the event of interest, Kaplan-Meier survival curves were estimated. The data was grouped based on key financial indicators, and the survival probability for each group was calculated to predict the likelihood of financial resilience.

2.5 Model Training and Validation

The dataset was divided into training (80%) and testing (20%) sets. Cross-validation with k-fold (k=10) was used to ensure generalization and prevent overfitting. Model performance was evaluated using metrics including accuracy**,** precision**,** recall**,** F1-score**,** and ROC-AUC**.**

2.6 Forecasting Approach

A time-series forecasting approach was integrated to predict future financial indicators. The Long Short-Term Memory (LSTM) model was applied to sequential financial data to forecast future resilience trends. The forecasted results were compared with actual data to evaluate predictive reliability.

2.7 Tools and Environment

All experiments were conducted on a system with:

1. Processor: Intel Core i7 or equivalent
2. Memory: 16 GB RAM
3. Software Environment: Python 3.10, Jupyter Notebook, TensorFlow 2.x, Scikit-learn, and NumPy.

3. RESULTS AND DISCUSSION

In this section, the results of the research will be discussed in detail. This discussion aims to analyze and interpret the findings obtained, as well as relate them to theories and previous research. The research findings will be explained thoroughly, covering an analysis of the variables studied and the implications of these findings for the phenomenon under investigation. Each finding will be discussed in a relevant context, considering factors that may influence the results, and providing an explanation of the limitations of this research and suggestions for future studies.

3.1. *Forecasting Process*

The data used in this study is sourced from internal financial reports, including cash flow, project value and duration, as well as monthly cash balances periode January 2023 until Maret 2025. The data is organized in a time-series format on a weekly and monthly basis, with adjustments made to time formats, nominal units, and the cleaning of missing values and anomalies. The preprocessing process includes cash aggregation for SARIMA analysis and survival analysis, as well as the conversion of project data into a numerical format. All data has been validated and serves as the foundation for SARIMA modeling and Kaplan-Meier analysis.

3.2. Visualization and Initial Interpretation

The heatmap visualization of the correlation between variables shows that net income is strongly correlated with gross profit, operating expenses, and revenue. This indicates that controlling the cost of goods sold (COGS) and operating expenses is a crucial factor in maintaining the company's profitability.

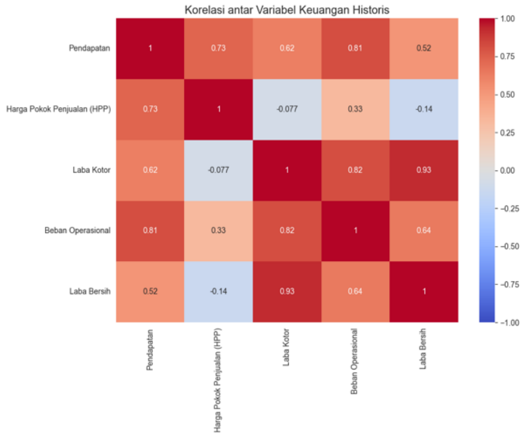

Figure 2. Financial Correlation Heatmap

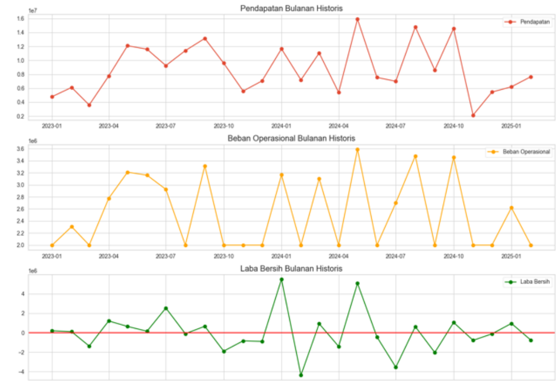

Figure 3. Historical Graph of Revenue, Expenses, and Net Income

In addition, the historical trends of revenue, operating expenses, and net income are also presented to assess the stability of the company's financial performance. This visualization helps management understand the trends that need to be anticipated in cash projections.

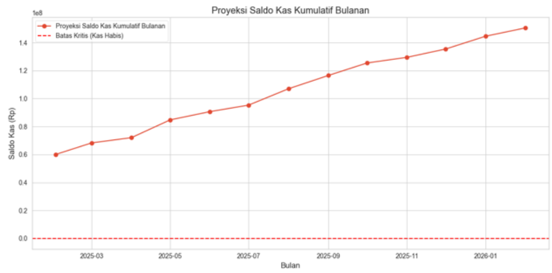

Figure 4. Cumulative Cash Balance Graph

The monthly cumulative cash balance forecast shows an upward trend, indicating an improvement in the financial position. However, fluctuations in revenue and expenses still need to be monitored to ensure that the cash balance does not enter a critical zone.

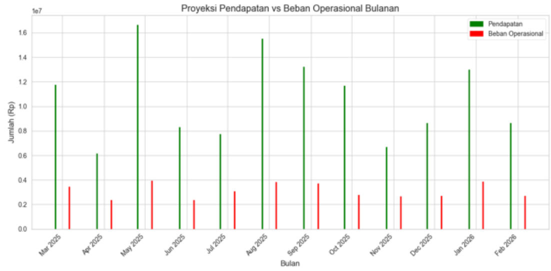

Figure 5. Revenue vs Expense Comparison Graph

The comparison graph between revenue forecasts and operating expenses reinforces insights into cost management efficiency.

3.3. Integration of SARIMA and Kaplan-Meier Models

The monthly cash balance forecast uses SARIMA. Historical cash balance data is modeled with SARIMA to generate predictions for the upcoming monthly cash balance. The output includes the cash balance values along with confidence intervals for the next 6–12 months. Calculate the average weekly fixed expenses. The average weekly expenses are calculated from historical operating expense data. This value is used to simulate weekly cash flows and serves as a reference for the cash balance's resilience against routine expenses.

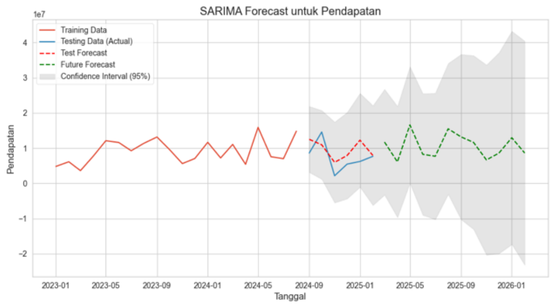

Figure 6. SARIMA Forecast Results for Revenue

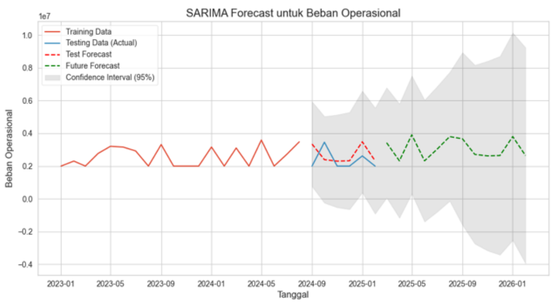

Figure 7. SARIMA Forecast Results for Operating Expenses

Simulate the week when the cash balance will no longer be sufficient. The predicted cash balance is divided by the weekly expenses to estimate the week when the cash funds will run out, assuming no additional income. Use the simulation data for the Kaplan-Meier Estimator. The weeks when the cash is insufficient are considered as "events" in the survival analysis. Kaplan-Meier is used to measure the survival probability each week until the cash balance reaches the critical point.

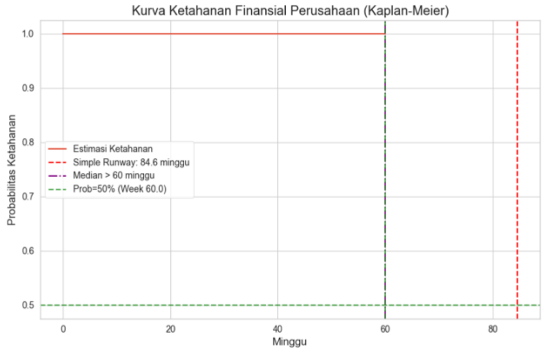

Figure 8. Kaplan-Meier Estimator Curve for Financial Resilience

With this approach, an estimate of resilience is obtained that is not only numerical (e.g., the company can survive for 24 weeks) but also accompanied by the survival probability for each specific week. This allows management to create contingency plans based on data.

3.4. Evaluation of Financial Resilience Based on Forecast

3.4.1 Analysis of Financial Runway

In the past two years, PT Moukoe Indonesia Kreatif's financial condition has shown fluctuations due to its project-based business model, with inconsistent contract values and durations each month. Revenue is highly dependent on the success of project acquisitions and smooth payments from clients. Forecasting results using SARIMA predict that the cash balance as of March 2025 will remain positive but is vulnerable to revenue variability and expenditure efficiency. Meanwhile, Kaplan-Meier analysis indicates that the cash resilience drastically decreases after week 6, which suggests that the company can only survive for about six weeks without additional income, known as the financial runway.
The runway is calculated using the formula:

Runway = Current Cash Balance / Average Weekly Expenses

Where:

1. Cash Balance as of March 31, 2025: Rp X (based on actual data)
2. Average weekly expenses: Rp Y

Thus, runway = X / Y weeks
Example illustration:
Cash Balance = Rp 60,000,000
Weekly Expenses = Rp 10,000,000
Runway = 6 weeks

3.4.2. Scenario Simulation

To provide a more strategic overview, two scenarios are conducted:
Worst-Case Scenario

1. No new projects enter in the next 2 months.
2. Fixed expenses continue as planned (Rp 10 million/week).
3. No additional funding or loans are obtained.

Outcome: The company will run out of cash by week 6 and will not be able to continue operations without intervention.
Best-Case Scenario

1. A new project worth Rp 100 million starts in the 2nd week of April.
2. The first payment (50%) is received in the 3rd week.
3. Expenses remain constant.

Outcome: The cash balance increases significantly, and the runway extends to over 12 weeks. Financial resilience returns to a stable position.

3.4.3. Evaluation of Financial Resilience Indicators

Several key indicators evaluated in this analysis include:

Table 1. Financial Resilience Indicators

Indicator	Value	Interpretation
Burn Rate	Rp X million/week	Indicates the rate of expenditure. The smaller, the more efficient.
Runway	6 weeks	Maximum time to survive with the current cash balance.
Solvency Ratio	> 1.5x	Still healthy (assets are greater than liabilities).
Liquidity Ratio	2.1x	Very good, indicating smooth short-term cash flow.

From the evaluation above, financial condition is generally stable in the short term, but there is a moderate risk if no new project income is received in the next 1–2 months.

4. CONCLUSION

Based on the forecasting results, it is recommended that the company implement operational cost efficiencies through routine audits and renegotiation of vendor contracts. Accounts receivable management should also be tightened with a more disciplined collection system, as well as incentives or penalties to regulate client payment timelines. To reduce the risk of income fluctuations, the company is advised to diversify projects and clients by undertaking recurring short-term projects. Additionally, the pricing strategy should be evaluated by considering project complexity and payment risks, while implementing a value-added approach.

The short-term financial strategy (1–3 months) focuses on real-time cash flow tracking, allocation of reserve funds from ongoing project profits, and simplifying expenses to prioritize projects with quick ROI. Meanwhile, the mid-term strategy (4–12 months) includes the development of an integrated digital financial monitoring system linked with operations, routine evaluations of predictive models based on the latest data, and the preparation of expansion or service pivot plans based on client demand trends.

REFERENCES

[1] S. Xu *et al.*, "Extension of Kaplan-Meier Methods in Observational Studies With Time-Varying Treatment," *Value in Health*, vol. 15, no. 1, pp. 167–174, 2012, doi: 10.1016/j.jval.2011.07.010.

[2] S. Chen, "A Novel Estimator of Survival Without Assumption of Censored Survival Time," 2021, doi: 10.21203/rs.3.rs-141531/v1.

[3] S. Willems, A. Schat, M. S. v. Noorden, and M. Fiocco, "Correcting for Dependent Censoring in Routine Outcome Monitoring Data by Applying the Inverse Probability Censoring Weighted Estimator," *Stat Methods Med Res*, vol. 27, no. 2, pp. 323–335, 2016, doi: 10.1177/0962280216628900.

[4] J. Xie and C. Liu, "Adjusted Kaplan–Meier Estimator and Log-rank Test With Inverse Probability of Treatment Weighting for Survival Data," *Stat Med*, vol. 24, no. 20, pp. 3089–3110, 2005, doi: 10.1002/sim.2174.

[5] C. D. Serio, "Competing Risk Problems With No Independence Assumed: Does It Make a Difference?," *Stat Methods Appt*, vol. 9, no. 1–3, pp. 39–51, 2000, doi: 10.1007/bf03178957.

[6] M. S. Pepe and M. Mori, "Kaplan—meier, Marginal or Conditional Probability Curves in Summarizing Competing Risks Failure Time Data?," *Stat Med*, vol. 12, no. 8, pp. 737–751, 1993, doi: 10.1002/sim.4780120803.

[7] D. Biau, A. Latouche, and R. Porcher, "Competing Events Influence Estimated Survival Probability," *Clin Orthop Relat Res*, vol. 462, pp. 229–233, 2007, doi: 10.1097/blo.0b013e3180986753.

[8] M. Gillam, P. Ryan, S. E. Graves, L. Miller, R. d. Steiger, and A. Salter, "Competing Risks Survival Analysis Applied to Data From the Australian Orthopaedic Association National Joint Replacement Registry," *Acta Orthop*, vol. 81, no. 5, pp. 548–555, 2010, doi: 10.3109/17453674.2010.524594.

[9] T. Martinussen and T. Scheike, "Estimating the Population Survival Function Using Additional Information Recorded Over Time: A Filter Based Approach," *Scandinavian Journal of Statistics*, vol. 25, no. 4, pp. 621–635, 1998, doi: 10.1111/1467-9469.00125.

[10] J. Péron, P. Roy, B. Ozenne, L. Roche, and M. Buyse, "The Net Chance of a Longer Survival as a Patient-Oriented Measure of Treatment Benefit in Randomized Clinical Trials," *JAMA Oncol*, vol. 2, no. 7, p. 901, 2016, doi: 10.1001/jamaoncol.2015.6359.

[11] R. Sharma, "Customer Churn Analysis in Telecom Industry Using Kaplan–Meier and Cox Proportional Hazards Model," *Interantional Journal of Scientific Research in Engineering and Management*, vol. 08, no. 06, pp. 1–5, 2024, doi: 10.55041/ijsrem36121.

[12] M. Resche-Rigon, É. Azoulay, and S. Chevret, "Evaluating Mortality in Intensive Care Units: Contribution of Competing Risks Analyses," *Crit Care*, vol. 10, no. 1, 2005, doi: 10.1186/cc3921.

[13] A. S. Hess and J. R. Hess, "Kaplan–Meier Survival Curves," *Transfusion (Paris)*, vol. 60, no. 4, pp. 670–672, 2020, doi: 10.1111/trf.15725.

[14] T. A. Gerds, "The Kaplan–Meier Theatre," *Teach Stat*, vol. 38, no. 2, pp. 45–49, 2016, doi: 10.1111/test.12095.

[15] Í. Etikán, "The Kaplan Meier Estimate in Survival Analysis," *Biom Biostat Int J*, vol. 5, no. 2, 2017, doi: 10.15406/bbij.2017.05.00128.

[16] F. Karim, M. A. Armin, D. Ahmedt-Aristizabal, L. Tychsen-Smith, and L. Petersson, "A Review of Hydrodynamic and Machine Learning Approaches for Flood Inundation Modeling," *Water (Basel)*, vol. 15, no. 3, p. 566, 2023, doi: 10.3390/w15030566.

[17] X. Deng, S. Cao, and A. L. Horn, "Emerging Applications of Machine Learning in Food Safety," *Annu Rev Food Sci Technol*, vol. 12, no. 1, pp. 513–538, 2021, doi: 10.1146/annurev-food-071720-024112.

[18] P. Poomka, N. Kerdprasop, and K. Kerdprasop, "Machine Learning Versus Deep Learning Performances on the Sentiment Analysis of Product Reviews," *Int J Mach Learn Comput*, vol. 11, no. 2, pp. 103–109, 2021, doi: 10.18178/ijmlc.2021.11.2.1021.

[19] P. Dhungana *et al.*, "Machine Learning in Data Abstraction: A Computable Phenotype for Sepsis and Septic Shock Diagnosis in the Intensive Care Unit," *World J Crit Care Med*, vol. 8, no. 7, pp. 120–126, 2019, doi: 10.5492/wjccm.v8.i7.120.

[20] "Review of Machine Learning Classification Algorithms," *International Journal of Progressive Research in Engineering Management and Science*, 2024, doi: 10.58257/ijprems36877.

[21] Z. Ren *et al.*, "The Value of Machine Learning in Preoperative Identification of Lymph Node Metastasis Status in Endometrial Cancer: A Systematic Review and Meta-Analysis," *Front Oncol*, vol. 13, 2023, doi: 10.3389/fonc.2023.1289050.

[22] A. E. W. Johnson, M. M. Ghassemi, S. Nemati, K. E. Niehaus, D. A. Clifton, and G. D. Clifford, "Machine Learning and Decision Support in Critical Care," *Proceedings of the Ieee*, vol. 104, no. 2, pp. 444–466, 2016, doi: 10.1109/jproc.2015.2501978.

[23] R. C. Deo, "Machine Learning in Medicine," *Circulation*, vol. 132, no. 20, pp. 1920–1930, 2015, doi: 10.1161/circulationaha.115.001593.

[24] F. Zhao, "A Literature Study of Asset Pricing Based on Machine Learning Method," *Advances in Economics Management and Political Sciences*, vol. 34, no. 1, pp. 27–36, 2023, doi: 10.54254/2754-1169/34/20231669.

[25] H. Anwar, T. Anwar, and M. S. Murtaza, "Applications of Electronic Nose and Machine Learning Models in Vegetables Quality Assessment: A Review," pp. 1–5, 2023, doi: 10.1109/icest56843.2023.10138839.

[26] C. Bellinger, M. S. M. Jabbar, O. R. Zäıane, and Á. Osornio-Vargas, "A Systematic Review of Data Mining and Machine Learning for Air Pollution Epidemiology," *BMC Public Health*, vol. 17, no. 1, 2017, doi: 10.1186/s12889-017-4914-3.

[27] M. K. H. Khan *et al.*, "Machine Learning and Deep Learning for Brain Tumor MRI Image Segmentation," *Exp Biol Med*, 2023, doi: 10.1177/15353702231214259.

[28] A. Kaminskyi, M. Nehrey, and L. Zomchak, "Machine Learning Methods Application for Consumer Banking," *SHS Web of Conferences*, vol. 107, p. 12001, 2021, doi: 10.1051/shsconf/202110712001.