Unveiling Tomorrow: Forecasting Attendance Trends and Patterns in Face Recognition-Based Attendance Systems through Deep Learning and Machine Learning Techniques

Joseph Teguh Santoso1, Danny Manongga2, Hendry3
[1] Dept. Computer Science, University of Science and Computer Technology, Semarang, Indonesia
[1] Dept. Informatic Engineering, Satya Wacana Christian University, Salatiga, Indonesia
[2] Dept. Informatic Engineering, Satya Wacana Christian University, Salatiga, Indonesia
[3] Dept. Informatic Engineering, Satya Wacana Christian University, Salatiga, Indonesia
Email: [1]Joseph_teguh@stekom.ac.id [2]danny.manongga@eksw.edu, [3]hendry@uksw.edu
Orchid Id number: [1] https://orcid.org/0000-0001-6227-1111, [2] https://orcid.org/0000-0002-7430-8740, [3] https://orcid.org/0000-0002-7387-2622
Corresponding Author*: Joseph Teguh Santoso.

ABSTRACT:

In a competitive business environment, creating an efficient workplace is crucial. Integrating digital technology into employee attendance systems is vital due to its significant impact on workforce efficiency and regulatory compliance. This study aims to formulate a predictive model for analyzing attendance patterns within a facial recognition-based attendance system, leveraging methodologies rooted in machine learning (ML) and deep learning (DL) paradigms. This research integrates regression and classification models derived from ML theory with DL techniques to enhance predictive precision. Utilized models include Random Forest, XGBoost, SVM, KNN, and Neural Network. Assessment of model effectiveness involves the evaluation of four metrics: accuracy, precision, recall, and the F1 score. Data collection relies on a facial recognition-based attendance system, trained, and tested within the Google Colab environment using Python. Findings reveal Random Forest and XGBoost as the most precise predictors of timeliness or tardiness among employees, considering age range and other pertinent factors, achieving an accuracy rate of 99%. Random Forest marginally outperforms XGBoost in both accuracy and F1-score by 0.01. This study is notable for its incorporation of attendance system data with ML and DL methodologies to predict attendance patterns based on age and diverse parameters, consequently enhancing decision-making processes and performance management.

Keywords: Machine Learning, Deep Learning, Pattern Prediction, Attendance Systems, Facial Recognition-Based Systems.

1) Introduction:

Data has emerged as one of the most significant assets for businesses in the ever-changing digital age. Data gathering and analysis give valuable insights for making better strategic decisions. Companies encounter significant challenges in monitoring and anticipating employee absence trends. Irregular attendance patterns can have a detrimental influence on productivity, operational efficiency, and overall business success. The incorporation of modern technologies such as machine learning and deep learning into corporate attendance systems has demonstrated considerable promise, particularly with the adoption of facial recognition-based systems. Face recognition-based attendance systems not only record employee attendance in an efficient and precise manner, but they also collect extensive and specific data on their attendance habits and trends. Machine learning and deep learning algorithms can use data from facial recognition technology to study attendance patterns more thoroughly. This technique allows for more precise projections of future absence by considering characteristics such as arrival timings, absenteeism frequency, and individual past tendencies. This implementation not only improves attendance monitoring accuracy but also gives information that can help management build more effective strategies for reducing absenteeism and increasing overall productivity.

In this environment, the presence of disciplined employees who regularly clock in on time is critical to a company's flow and performance. However, efficiently applying attendance data to forecast attendance trends inside the firm is still an area that requires further investigation. This is confirmed by bibliometric analysis data obtained by *Vosviewer*, which shows that the issue of employing machine learning technology, particularly in forecasting trends inside attendance systems, is an empty niche. Furthermore, while constructing predictive models, there is a research gap in using machine learning approaches to anticipate staff attendance patterns depending on the company's classifications and demands. Previous research has mostly concentrated on the application of face recognition technology in attendance systems, without thoroughly investigating the predictive power of this attendance data. Studies such as [1], [2], [3] have concentrated on the implementation and integration of usage techniques. Furthermore, research [2], [4], [5], and [6] used machine learning and deep learning to improve accuracy in facial recognition detection rather than as prediction models. Existing research has also focused on a few machine-learning models, such as [7] and [8], whereas many additional models have yet to be thoroughly investigated. Using suitable theories can help create more comprehensive and efficient prediction models.

To close this gap, this research attempts to create a prediction model for staff attendance patterns inside a face recognition-based attendance system, leveraging machine learning and deep learning models to improve resource management. Furthermore, forecasted patterns might give critical information for better decision-making, disciplinary action, and employee retention. By combining these technologies, the predictive model is projected to increase performance management and forecast accuracy, simplifying successful implementation and providing more holistic and practical insights into the organizational environment.

This study has the potential to change the landscape of employee performance management and workplace discipline. This study builds a solid basis for accurately forecasting employee attendance patterns by incorporating digital technologies into the face recognition-based attendance system and employing machine learning and deep learning methodologies. A greater knowledge of these attendance patterns helps businesses devise more effective ways to control employee performance and increase overall productivity.

Therefore, this research not only makes a significant contribution to academic literature but also has the potential to build a more predictable and efficient future workplace. This study contributes to the improvement of employee performance management and workplace discipline through the development of predictive models to analyze employee attendance trends using a facial recognition-based attendance system. It also offers useful insights on how to properly estimate staff attendance patterns utilizing machine learning and deep learning technology. Thus, not only does this study fill a vacuum in the scholarly literature, but it also has a direct influence on improving employee performance management and workplace discipline.

(a) Machine Learning

Machine learning (ML), a subset of artificial intelligence (AI), allows systems to learn from data, recognize patterns, and make choices with little or no human interaction. ML algorithms may be divided into three types: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning uses labeled data to train models on input-output pairings, such as linear regression and Support Vector Machine (SVM). Unsupervised learning uses unlabeled data to identify hidden structures or patterns; in this study, the clustering model is K-Nearest Neighbor (KNN). Machine learning has transformed many organizations by producing precise forecasts and doing in-depth data analysis. Multiple machine learning models are used to predict trends in a range of fields. For example, in finance, ML algorithms help with fraud detection, credit analysis, and stock market forecasts. Studies such as [9] and [10] have effectively predicted market prices using several machine learning classifiers. Furthermore, [11] and [12] increased profitability by merging deep learning, machine learning, and e-commerce. Research by [13], and [14] focused on forecasting online sales, and [15] effectively deployed machine learning models for shopping cart prediction. Furthermore, ML is used for analysis; for example, [16] used machine learning to assess sales offers, indicating that ML can efficiently analyze and decide whether to accept or reject sales pitches.

In healthcare, machine learning is used to identify illnesses, provide personalized patient care, and evaluate medical pictures. For example, [17] and [18] investigated diabetes prediction, [19] utilized ML to predict cancer, [20] to anticipate food safety risks, and [21] to predict eye disease. Beyond sickness prediction, [22] claimed that AI and machine learning can spot patterns, identify biomarkers, and anticipate disease progression with remarkable accuracy. In other disciplines, [23] utilized machine learning to predict network attack patterns, and [24], [25], and [26] employed ML models to forecast energy use. In the industrial business, machine learning is used to forecast employee performance [27], [28], and [29] used ML for this purpose. In this study, different machine learning models will be utilized to forecast attendance patterns in an attendance system to uncover trends and patterns depending on employee age. These projections can then be used to help guide future employment decisions based on age criteria.

(b) Deep Learning

Deep learning has shown dominance in a variety of applications, including image identification, natural language processing (NLP), and video analysis. Neural networks (NN) in deep learning are capable of learning more complicated feature representations than typical machine learning techniques. However, the ability of deep learning to extract features from raw data and handle unstructured data makes it extremely powerful in prediction and application across various domains. Deep learning continues to push the frontiers of machine learning capabilities as technology and optimization algorithms progress, making it an essential tool in current data analysis and artificial intelligence.

Deep learning techniques, like Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), are essential for forecasting attendance patterns in a facial recognition system. CNNs excel in image processing and can improve facial recognition accuracy, resulting in reliable attendance tracking. CNNs improve attendance systems by extracting and learning complex features from facial images. On the other hand, RNNs and Long Short-Term Memory (LSTM) networks excel at sequence prediction, making them ideal for forecasting attendance trends. These algorithms may find temporal relationships and trends in attendance data over time, predicting future events based on past behaviors. RNNs can predict temporal shifts and seasonal trends based on previous attendance data. This deep learning approach establishes a solid platform for studying and forecasting attendance trends, allowing institutions to better anticipate attendance variations and enhance operational efficiency. However, in this work, deep learning is not used for facial identification inside the face recognition-based attendance system, but rather to increase the accuracy of anticipating employee attendance trends. This technique aims to give a complete understanding of predicted outcomes with high accuracy, which may be applied in a real-time integrated system.

(c) Integration of Machine Learning and Deep Learning for Comprehensive Solutions

In terms of business attendance systems, combining machine learning and deep learning approaches provides a comprehensive solution for forecasting attendance patterns. Ensemble and supervised learning models lay the groundwork for early prediction and classification, but deep learning models can improve forecasts by incorporating complicated and high-dimensional data patterns. This combination enables the creation of a strong and adaptable attendance forecasting system that takes use of the capabilities of both methodologies.

In the context of forecasting attendance patterns, deep learning provides the capacity to handle complicated data and model more abstract and nonlinear patterns. According to [3], [30], [31], deep learning is used in time series forecasting, including the usage of RNN and LSTM to predict employee performance in enterprises. Meanwhile, [32] investigates the use of deep learning in anomaly identification, including anticipating productivity trends and employee retention. Although ensemble ML models, as well as classification models, often yield good results in predicting attendance patterns, the deep learning approach using neural networks also has the potential to enhance prediction performance in cases of highly complex attendance patterns or very large datasets. In the context of developing predictive models, machine learning models can be used to identify factors influencing employee performance, predict future performance, and provide recommendations for steps to improve performance. In the development of predictive models in this research, ML and DL models will learn from existing attendance data to make predictions on attendance patterns based on employee age to determine which employees of a certain age are consistently on time or consistently late for future decision-making by the company.

(d) The Implementation of Machine Learning and Deep Learning in Attendance Systems

The application of machine learning models for prediction has brought significant benefits in various contexts, including employee performance prediction, sales forecasting, risk analysis, and others. However, there are still some gaps that can be filled. Recent research, particularly in the industrial sector, has started to combine various technologies to innovate and enhance company performance in predictive model development for attendance systems. Studies by Online attendance systems based on facial recognition with face mask detection, [4], [6], [7], [33] and [34] used machine learning and deep learning to successfully find and distinguish faces in attendance photos. Similar studies by [1], [2], [4], [7], [35], [36], and [37], show that using deep learning and machine learning techniques in attendance systems may greatly enhance individual recognition accuracy when compared to traditional approaches. However, few studies have concentrated on constructing more complex prediction models for estimating individual attendance, especially when several internal and external variables are considered. The outcomes of this study suggest that more advanced machine learning algorithms have the potential to dramatically increase the efficacy and accuracy of attendance systems, as well as provide a full knowledge of how external variables may affect attendance.

2) Methodology:

(a) The Machine Learning and Deep Learning Models Used in the Research

In this study, several machine learning and deep learning models were selected for use as predictors trained and tested within the system. The models used include Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Neural Network (NN), in this case is RNN. The RF model was chosen for its ability to handle a wide range of intricate properties while reducing overfitting in the dataset. This method generates many decision trees independently before combining the results to increase prediction accuracy. On the other hand, XGBoost, as an ensemble model, is regarded to be useful for anticipating attendance patterns because of its ability to improve performance repeatedly, hence increasing prediction accuracy by resolving earlier model flaws and developing robust and accurate models. Furthermore, SVM can easily extract intricate classes from data, yielding complex and non-linear attendance patterns. KNN is used in attendance pattern prediction to identify patterns that are like previous attendance patterns and predict future attendance patterns based on those similarities. Finally, as a deep learning model, Neural Network allows for an improved understanding of complex feature representations in attendance data, such as repeated attendance patterns over time or patterns dependent on specific conditions.

(b) Approach and Technical

This work combines machine learning theory with a deep learning technique to improve the accuracy of forecasting attendance patterns and developing trends using a facial recognition-based attendance system. The machine learning models utilized in this study are divided into two categories: ensemble and supervised learning. The ensemble machine learning models chosen for this study are RF and XGBoost, while the supervised learning regression classification machine learning models used are SVM and KNN. In this case, the approach used is deep learning with a neural network model. These five models are used to predict and identify attendance patterns of various categories of employees in the attendance system, namely late, on-time, and trends emerging from attendance pattern predictions based on age.

(c) Development of Machine Learning Predictive Models

Models for predicting absenteeism patterns were trained and tested in the Google Colab environment, using the Python programming language. The dataset used was collected from an attendance system-based facial recognition, with a total of 998 data points. To maintain data accuracy, only relevant characteristics were preserved for analysis following selection and normalization. During data preparation, a normalization test was performed to determine the distribution spread. Deep data analysis and identification were carried out utilizing the Interquartile Range (IQR), a statistical approach for detecting outliers in data. Outliers were detected using Q1-1.5xIQR for the lower bound and

Q3+1.5xIQR for the upper bound. Outlier detection continued until no new data were recognized as outliers in the dataset. The data was then divided into training and testing sets using the *train_test_split* approach, which applied the Pareto principle [38] with an 80/20 split, with 80% of the data set aside for training and 20% for testing.

(d) Evaluation Metrics

To evaluate the performance of the developed machine learning models, various evaluation metrics were utilized. These metrics include accuracy, precision, recall, F1-score, and ROC-AUC. By employing these metrics, the performance of both ML and DL models in predicting absenteeism patterns and trends can be more thoroughly assessed. This comprehensive evaluation enables the selection and application of the best predictive model according to specific needs, thereby allowing decision-makers to have greater confidence in the model's predictions and make more informed decisions based on the information provided within a company. Given that the research employs five models across different categories, the evaluation metrics are also varied. This diversity in metrics can be seen in Figure 2, where each model is associated with distinct measurement criteria.

3) Result:

(a) Evaluation Metrics

The testing results of five models, comprising two ensemble machine learning models (RF and XGBoost), two supervised learning classification models (SVM and KNN), and one model from the Deep Learning approach (NN), demonstrate the superiority of ensemble machine learning models over the other three in predicting attendance patterns and trends in an attendance system-based face recognition scenario.

Specifically, both the Random Forest and XGBoost models have the greatest accuracy level, both scoring 0.99, as well as similarly good F-1 and recall scores. However, Random Forest has a slightly higher accuracy value of 0.01 points than XGBoost, making it the best model for future attendance pattern prediction tasks. Furthermore, the AUC-ROC values for the Random Forest and XGBoost models exceed 0.8, showing strong and consistent performance suited for most actual applications. Table 1 shows the metric measurement results for reference.

Among the remaining three models, KNN exhibits superior evaluation metrics compared to the other two models, albeit with a marginal difference from the top two ensemble models. Specifically, the accuracy rate of KNN is 0.95, which is 0.04 points lower than the preceding ensemble models. Similarly, precision, recall, and F1-score values for KNN, while surpassing those of NN and SVM, position KNN in third place in the prediction ranking of this study. Conversely, the other two models demonstrate nearly identical low evaluation scores, despite perfect recall values. However, they appear insufficient in predicting employee absenteeism patterns.

From the measurement results of the five metrics applied to the five models, it is evident that the Random Forest model outperforms others in predicting absenteeism patterns and trends in attendance systems. However, XGBoost performs as well as Random Forest. As a result, both models are appropriate for use in corporate attendance systems to forecast punctual and tardy employee attendance patterns and identify developing trends within the company's attendance system. Fig. 3, Fig. 4, Fig. 5, and Fig. 6, provide extensive comparisons between the five models utilized in this study.

(b) Normality Test

The dataset used in the study included over 1000 data points; however, after preprocessing, it was discovered that the data did not follow a normal distribution. As a result, data normalization was performed using two methods: min-max scaling and outlier elimination by detecting abnormalities in the dataset using the IQR approach. Outliers were eliminated methodically until the dataset included no outliers. The dataset was further checked with a Boolean array to confirm that there were no outliers in any of the values. A hypothesis test (p-value) was then performed to see if the results obtained from the sample data were robust enough to reject the null hypothesis, which normally states that there is no significant impact or difference. Table 2 also includes the results of the p-value test.

The p-values from the Shapiro-Wilk test indicate the normality of the error or residual distribution based on model predictions. When the p-value exceeds α ($\alpha = 0.05$), the data is considered to have a normal distribution. In this context, Random Forest and XGBoost have Shapiro-Wilk p-values of 0.826500 and 0.829337, respectively, which are considerably more than 0.05. Thus, the null hypothesis is not rejected, indicating that the error distributions for both models are normal. Similarly, the p-values for the KNN and Neural Network models (0.862744 and 0.787177) exceed 0.05, indicating normal distribution adherence. In comparison, the Shapiro-Wilk p-value for the SVM model is 0.184938, making it the weakest of the models but still more than 0.05. Although marginally deviating from the normal distribution, the difference is not large enough to rule out normalcy. As a result, all examined models had Shapiro-Wilk p-values greater than 0.05, suggesting no compelling evidence to reject the hypothesis that their error distributions are normal. These data indicate that the model is doing well. However, the Kolmogorov-Smirnov test p-values for all models are less than 0.05, indicating that the null hypothesis that the error distributions are normal is rejected. In other words, all models' error distributions are non-normal, as shown by the Kolmogorov-Smirnov test. Despite this, the assessment findings reveal great performance for all five models, since they are consistent with the normal distribution based on the Shapiro-Wilk test, with minor differences suggested by the Kolmogorov-Smirnov test.

4) Discussion:

With an accuracy rate of 99%, the study's findings demonstrate that the Random Forest and XGBoost models produced the most accurate forecasts of employee attendance patterns. As demonstrated by [4] and [7], which also discovered that the Random Forest method is quite successful at handling complicated datasets for predictive tasks, this is in line with earlier research. The efficacy of Random Forest in this investigation may be ascribed to its capacity to mitigate overfitting using the consolidation of several decision trees; this conclusion is corroborated by [9] and [10] in distinct scenarios.

Meanwhile, XGBoost, while somewhat less accurate than Random Forest, performed admirably, with nearly equal F1 scores. XGBoost is well-known for its greater capacity to handle different data and improve on the flaws of earlier models via the boosting strategy, as demonstrated by the findings of [12] and [13], both of which used XGBoost in market prediction scenarios.

Based on the study of the five models utilized, it is clear that the K-Nearest Neighbor (KNN) model produced relatively excellent results, with an accuracy of 95%, but fell short of ensemble methods such as Random Forest and XGBoost. This suggests that KNN can detect attendance patterns similar to earlier ones, but it is less effective in dealing with the total complexity of the data, as reported in research [14] and [15]. This drawback might be attributed to KNN's reliance on closest neighbor data, which may not be successful in huge datasets or when there are non-linear trends.

In contrast, the Neural Network in this study performed worse than the other models. Although deep learning excels at processing unstructured data, the outcomes of this study imply that Neural Networks may require further tweaking to handle structured attendance data. This is consistent with research by [16] and (Kasula, 2021), which found that implementing a Neural Network is frequently more complicated and needs more resources to reach the same level of accuracy as other machine learning models.

From a practical standpoint, the findings of this study support the hypothesis that a mix of machine learning and deep learning may be utilized to improve face recognition-based attendance systems. For example, the Random Forest and XGBoost models predict attendance patterns while also providing additional insights on age features and other parameters, which are consistent with the findings of [18] and [19] on the use of machine learning algorithms in employee performance assessments. This enables management to make better-informed judgments about human resource management methods, such as defining discipline policies and employee retention.

5) Conclusion:

This study uses machine learning theory and a deep learning method to forecast absence patterns and explore developing trends across various employee age groups. Among the models tested, Random Forest and XGBoost produced the greatest prediction values with a noticeably high accuracy level (0.99), followed by KNN in third place with a little lower accuracy. The high accuracy scores achieved by the top two models (Random Forest and XGBoost) demonstrate their superior prediction ability in identifying and distinguishing absence patterns depending on employee age. Thus, for prediction tasks in this study, Random Forest and XGBoost are judged superior to KNN, SVM, and Neural Network. However, the Neural Network model in this context performs poorly on all criteria, highlighting the need for additional development. As a result, future model additions may include more tuning and improvements in architecture or training approaches.

Acknowledgment: The authors would like to express their sincere gratitude to the University of Science and Computer Technology (STEKOM University) and Satya Wacana Christian University (SWCU) for their support and collaboration in conducting this research. Special thanks go to the Faculty of Informatics Engineering for providing the necessary resources and technical assistance throughout the research process. The authors also extend their appreciation to the research team members for their valuable insights and contributions. Lastly, our deepest gratitude goes to all participants and stakeholders who supported this study with their time and efforts, making this research possible.

Funding Statement: There is no found received for this article

Data Availability: No new data were created or analyzed in this study. Data sharing does not apply to this article
Declare the availability of the data used in the manuscript

Conflict of interest: The authors declare that there is **no conflict of interest**.

References

[1] M. H. M. Kamil, N. Zaini, L. Mazalan, and A. H. Ahamad, "Online attendance system based on facial recognition with face mask detection," *Multimed Tools Appl*, vol. 82, no. 22, pp. 34437–34457, Sep. 2023, doi: 10.1007/S11042-023-14842-Y/TABLES/6.

[2] A. Kumar, S. Samal, M. S. Saluja, and A. Tiwari, "Automated Attendance System Based on Face Recognition Using Opencv," *2023 9th International Conference on Advanced Computing and Communication Systems, ICACCS 2023*, pp. 2256–2259, 2023, doi: 10.1109/ICACCS57279.2023.10112665.

[3] P. B. Weerakody, K. W. Wong, G. Wang, and W. Ela, "A review of irregular time series data handling with gated recurrent neural networks," *Neurocomputing*, vol. 441, pp. 161–178, Jun. 2021, doi: 10.1016/J.NEUCOM.2021.02.046.

[4] H. Burhan, U. Haq, W. Akram, M. Nauman Irshad, A. Kosar, and M. Abid, "Enhanced Real-Time Facial Expression Recognition Using Deep Learning Acadlore Transactions on AI and Machine Learning Enhanced Real-Time Facial Expression Recognition Using Deep Learning," *Trans. Mach. Learn*, vol. 3, no. 1, pp. 24–35, 2024, doi: 10.56578/ataiml030103.

[5] M. Rostami, A. Farajollahi, and H. Parvin, "Deep Learning-based Face Detection and Recognition on Drones," *J Ambient Intell Humaniz Comput*, vol. 15, no. 1, pp. 373–387, Jan. 2024, doi: 10.1007/S12652-022-03897-8/METRICS.

[6] S. Hangaragi, T. Singh, and N. Neelima, "Face Detection and Recognition Using Face Mesh and Deep Neural Network," *Procedia Comput Sci*, vol. 218, pp. 741–749, Jan. 2023, doi: 10.1016/J.PROCS.2023.01.054.

[7] C. J. Harshitha, R. K. Bharathi, and C. A. Kushalappa, "Machine learning models for real time applications in face recognition system for masked and unmasked images," *AIP Conf Proc*, vol. 2966, no. 1, Mar. 2024, doi: 10.1063/5.0189905/3279449.

[8] S. Sharma, M. Bhatt, and P. Sharma, "Face Recognition System Using Machine Learning Algorithm," pp. 1162–1168, Jul. 2020, doi: 10.1109/ICCES48766.2020.9137850.

[9] M. Nikou, G. Mansourfar, and J. Bagherzadeh, "Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms," *Intelligent Systems in Accounting, Finance and Management*, vol. 26, no. 4, pp. 164–174, Oct. 2019, doi: 10.1002/ISAF.1459.

[10] W. Khan, M. A. Ghazanfar, M. A. Azam, A. Karami, K. H. Alyoubi, and A. S. Alfakeeh, "Stock market prediction using machine learning classifiers and social media, news," *J Ambient Intell Humaniz Comput*, vol. 13, no. 7, pp. 3433–3456, Jul. 2022, doi: 10.1007/S12652-020-01839-W/METRICS.

[11] M. Sarkar *et al.*, "Optimizing E-Commerce Profits: A Comprehensive Machine Learning Framework for Dynamic Pricing and Predicting Online Purchases," *Journal of Computer Science and Technology Studies*, vol. 5, no. 4, pp. 186–193, Dec. 2023, doi: 10.32996/JCSTS.2023.5.4.19.

[12] X. Zhang, F. Guo, T. Chen, L. Pan, G. Beliakov, and J. Wu, "A Brief Survey of Machine Learning and Deep Learning Techniques for E-Commerce Research," *Journal of Theoretical and Applied Electronic Commerce Research 2023, Vol. 18, Pages 2188-2216*, vol. 18, no. 4, pp. 2188–2216, Dec. 2023, doi: 10.3390/JTAER18040110.

[13] C. Sun, P. Adamopoulos, A. Ghose, and X. Luo, "Predicting Stages in Omnichannel Path to Purchase: A Deep Learning Model," *Information Systems Research*, vol. 33, no. 2, pp. 429–445, Nov. 2021, doi: 10.1287/ISRE.2021.1071.

[14] D. Rohaan, E. Topan, and C. G. M. Groothuis-Oudshoorn, "Using Supervised Machine Learning for B2B Sales Forecasting: A Case Study of Spare Parts Sales Forecasting at an After-Sales Service Provider," *Expert Syst Appl*, vol. 188, p. 115925, Feb. 2022, doi: 10.1016/J.ESWA.2021.115925.

[15] T. M. Rausch, N. D. Derra, and L. Wolf, "Predicting online shopping cart abandonment with machine learning approaches," *International Journal of Market Research*, vol. 64, no. 1, pp. 89–112, Jan. 2022, doi: 10.1177/1470785320972526/

[16] P. Nguyen, S. B. Friend, K. S. Chase, and J. S. Johnson, "Analyzing sales proposal rejections via machine learning," *Journal of Personal Selling & Sales Management*, vol. 43, no. 1, pp. 24–45, 2023, doi: 10.1080/08853134.2022.2067554.

[17] Y. B. Kasula, "Machine Learning Applications in Diabetic Healthcare: A Comprehensive Analysis and Predictive Modeling," *International Numeric Journal of Machine Learning and Robots*, vol. 5, no. 5, 2021, Accessed: May 18, 2024. [Online]. Available: https://injmr.com/index.php/fewfewf/article/view/19/1

[18] C. Krishna Suryadevara, "Issue 4 Diabetes Risk Assessment Using Machine Learning: A Comparative Study Of Classification Algorithms," *International Engineering Journal For Research & Development*, vol. 8, 2023, Accessed: May 18, 2024. [Online]. Available: www.iejrd.com

[19] W. Zhu, L. Xie, J. Han, and X. Guo, "The Application of Deep Learning in Cancer Prognosis Prediction," *Cancers 2020, Vol. 12, Page 603*, vol. 12, no. 3, p. 603, Mar. 2020, doi: 10.3390/CANCERS12030603.

[20] B. Zhao *et al.*, "Prediction Heavy Metals Accumulation Risk in Rice Using Machine Learning and Mapping Pollution Risk," *J Hazard Mater*, vol. 448, p. 130879, Apr. 2023, doi: 10.1016/J.JHAZMAT.2023.130879.

[21] A. Bali and V. Mansotra, "Analysis of Deep Learning Techniques for Prediction of Eye Diseases: A Systematic Review," *Archives of Computational Methods in Engineering*, vol. 31, no. 1, pp. 487–520, Jan. 2024, doi: 10.1007/S11831-023-09989-8/METRICS.

[22] A. Husnain, K. H. Hussain, M. H. Shahroz, M. Ali, and Y. Hayat, "View of Advancements in Health through Artificial Intelligence and Machine Learning: A Focus on Brain Health," *Revista Espanola De Documentacion Cientifica*, vol. 18, no. 01, pp. 100–123, 2024, Accessed: May 18, 2024. [Online]. Available: https://redc.revistas-csic.com/index.php/Jorunal/article/view/195/162

[23] S. Nanda, F. Zafari, C. Decusatis, E. Wedaa, and B. Yang, "Predicting network attack patterns in SDN using machine learning approach," *2016 IEEE Conference on Network Function Virtualization and Software Defined Networks, NFV-SDN 2016*, pp. 167–172, May 2017, doi: 10.1109/NFV-SDN.2016.7919493.

[24] N. Somu, G. Raman M R, and K. Ramamritham, "A Deep Learning Framework for Building Energy Consumption Forecast," *Renewable and Sustainable Energy Reviews*, vol. 137, Mar. 2021, doi: 10.1016/J.RSER.2020.110591.

[25] L. Lei, W. Chen, B. Wu, C. Chen, and W. Liu, "A building energy consumption prediction model based on rough set theory and deep learning algorithms," *Energy Build*, vol. 240, p. 110886, Jun. 2021, doi: 10.1016/J.ENBUILD.2021.110886.

[26] R. Olu-Ajayi, H. Alaka, I. Sulaimon, F. Sunmola, and S. Ajayi, "Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques," *Journal of Building Engineering*, vol. 45, p. 103406, Jan. 2022, doi: 10.1016/J.JOBE.2021.103406.

[27] A. Pathak, C. K. Dixit, P. Somani, and S. K. Gupta, "Prediction of Employees' Performance Using Machine Learning (ML) Techniques," *Designing Workforce Management Systems for Industry 4.0: Data-Centric and AI-Enabled Approaches*, pp. 177–195, Jan. 2023, doi: 10.1201/9781003357070-11.

[28] A. S. Lather, R. Malhotra, P. Saloni, P. Singh, and S. Mittal, "Prediction of employee performance using machine learning techniques," *ACM International Conference Proceeding Series*, Nov. 2019, doi: 10.1145/3373477.3373696.

[29] Q. Cheng, L. Yang, J. Zheng, M. Tian, and D. Xin, "Optimizing Portfolio Management and Risk Assessment in Digital Assets Using Deep Learning for Predictive Analysis," Feb. 2024, doi: https://doi.org/10.48550/arXiv.2402.15994.

[30] K. Bandara, C. Bergmeir, and S. Smyl, "Forecasting across time series databases

using recurrent neural networks on groups of similar series: A clustering approach," *Expert Syst Appl*, vol. 140, p. 112896, Feb. 2020, doi: 10.1016/J.ESWA.2019.112896.

[31] P. Nath, P. Saha, A. I. Middya, and S. Roy, "Long-term time-series pollution forecast using statistical and deep learning methods," *Neural Comput Appl*, vol. 33, no. 19, pp. 12551–12570, Oct. 2021, doi: 10.1007/S00521-021-05901-2/FIGURES/13.

[32] M. M. Inuwa and R. Das, "A comparative analysis of various machine learning methods for anomaly detection in cyber attacks on IoT networks," *Internet of Things*, vol. 26, p. 101162, Jul. 2024, doi: 10.1016/J.IOT.2024.101162.

[33] M. Ali, A. Diwan, and D. Kumar, "Attendance System Optimization through Deep Learning Face Recognition," *International Journal of Computing and Digital Systems*, vol. 15, no. 1, pp. 1527–1540, Apr. 2024, doi: 10.12785/ijcds/1501108.

[34] Prof. M. Barhate *et al.*, "Innovative Attendance Tracking: Facial Recognition," *International Conference on Emerging Smart Computing and Informatics (ESCI)*, pp. 1–6, Mar. 2024, doi: 10.1109/ESCI59607.2024.10497364.

[35] C. D. Sawarkar and S. S. Alane, "Attendance System by Face Recognition Using Deep Learning," *International Research Journal of Modernization in Engineering Technology and Science*, Feb. 2024, doi: 10.56726/irjmets48994.

[36] Nagagopiraju, M. D. Deepak, V. M. Vinay, A. E. Kumar, and K. Supriya, "Machine Learning and Blockchain-Based Real-Time Facial Recognition Attendance System," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 15, no. 1, pp. 133–137, Mar. 2024, doi: 10.61841/TURCOMAT.V15I1.14554.

[37] A. Bhatta, V. Albiero, K. W. Bowyer, and M. C. King, "The Gender Gap in Face Recognition Accuracy Is a Hairy Problem," 2023.

[38] R. Sanders, "The Pareto Principle: Its Use and Abuse," *Journal of Services Marketing*, vol. 1, no. 2, pp. 37–40, 1987, doi: 10.1108/EB024706/FULL/XML.

Table and Figure:

TABLE I. Measurement Metric Results for the Five Models Used

Model	Measurement Metric				
	AUC-ROC	Accuracy	Precision	Recall	F1-Score
RF	0.88	0.99	0.99	1.00	0.99
XGBoost	0.80	0.99	0.98	1.00	0.99
SVM	-	0.56	0.56	1.00	0.72
KNN	-	0.95	0.93	0.99	0.96
NN	0.55	0.53	0.59	0.54	0.56

TABLE II. Result of p-Value

No	Model	Shapiro-Wilk	Kolmogorov-Smirnov
1	RF	0.826500	0.001337
2	XGBoost	0.829337	0.001410
3	SVM	0.184938	0.014016
4	KNN	0.862744	0.001896
5	NN	0.787177	0.026451

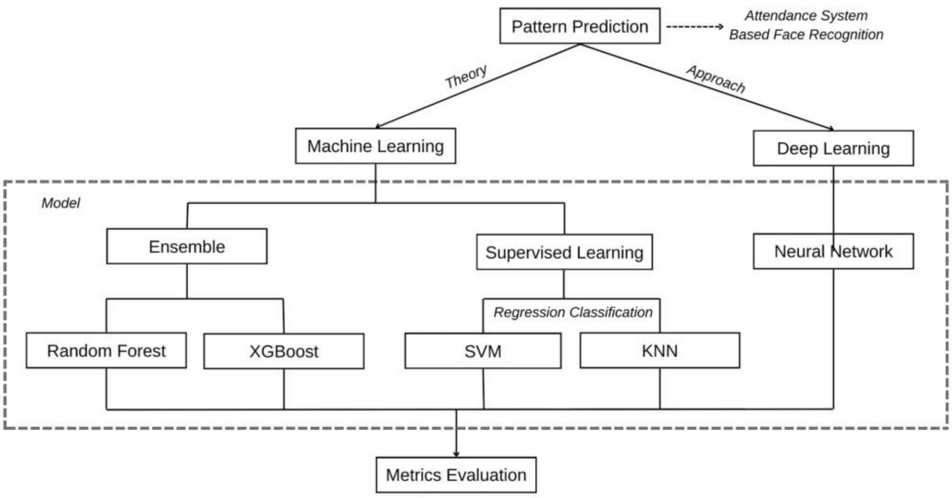

Fig. 1. Implementation of machine learning theory and research approach

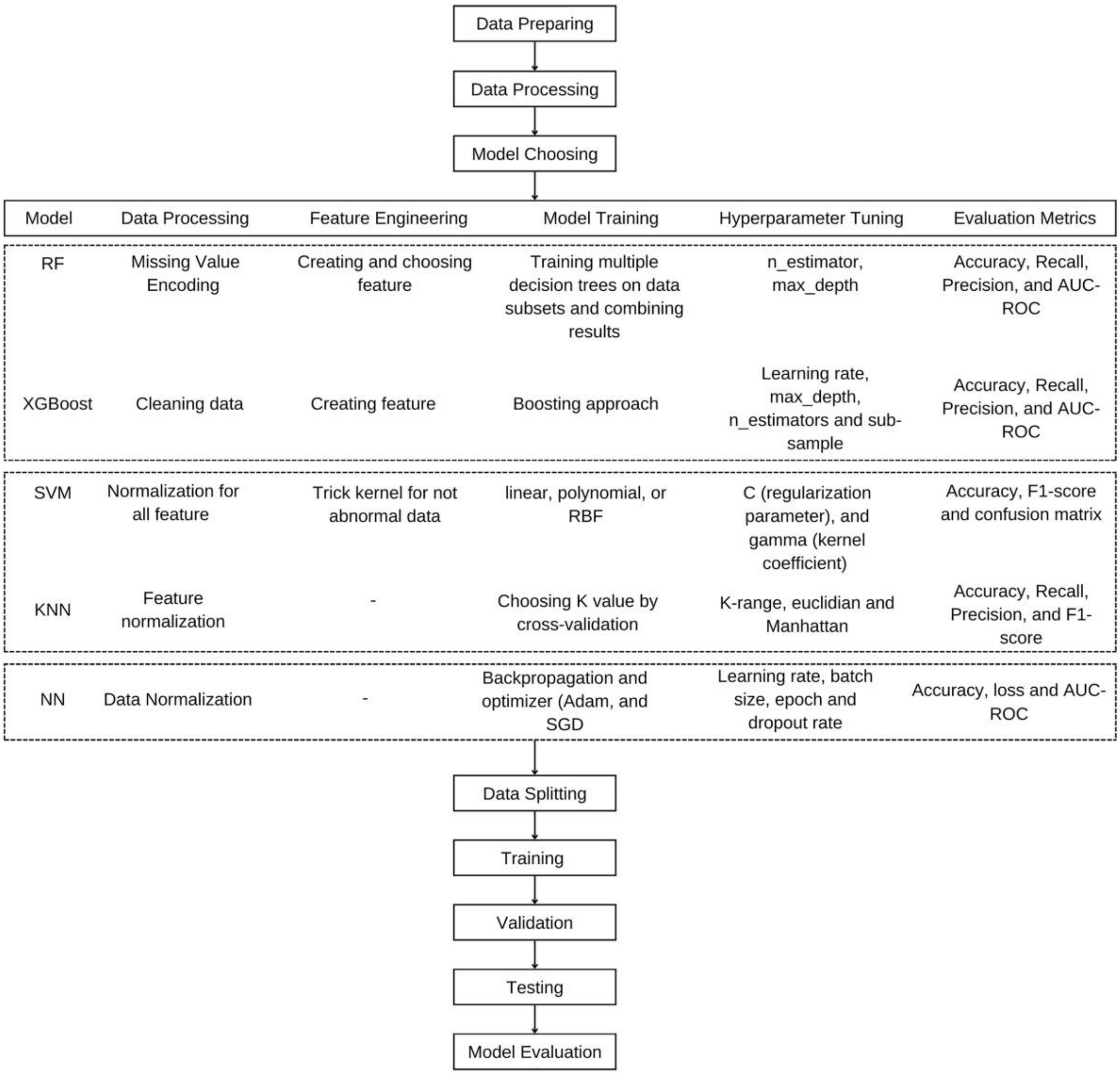

Figure 2. Training, testing, and prediction process in the developed ML and deep learning models in the study.

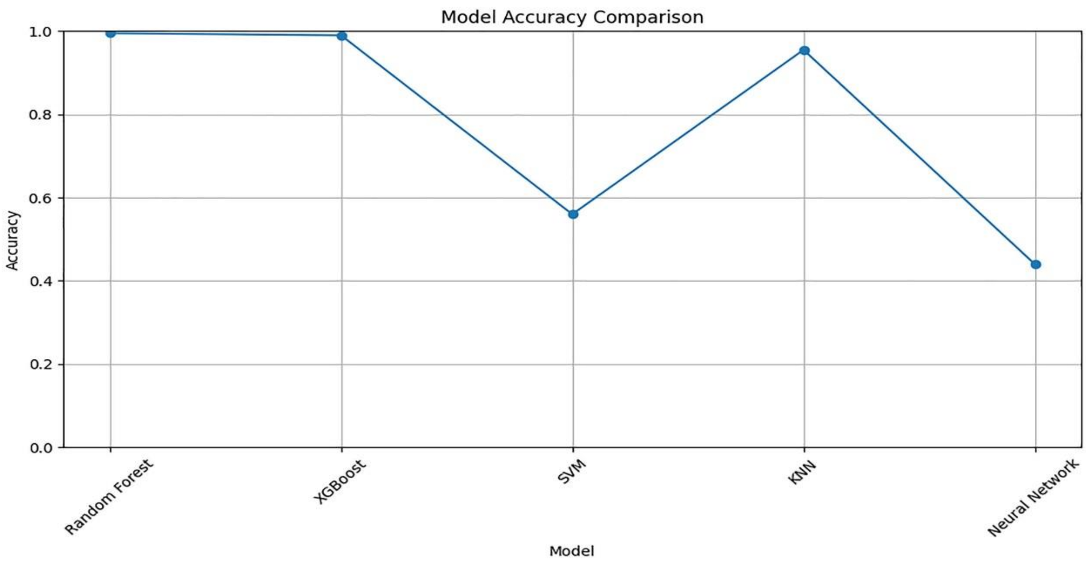

Fig. 3. Visualization of Accuracy Comparison Among the Used Models.

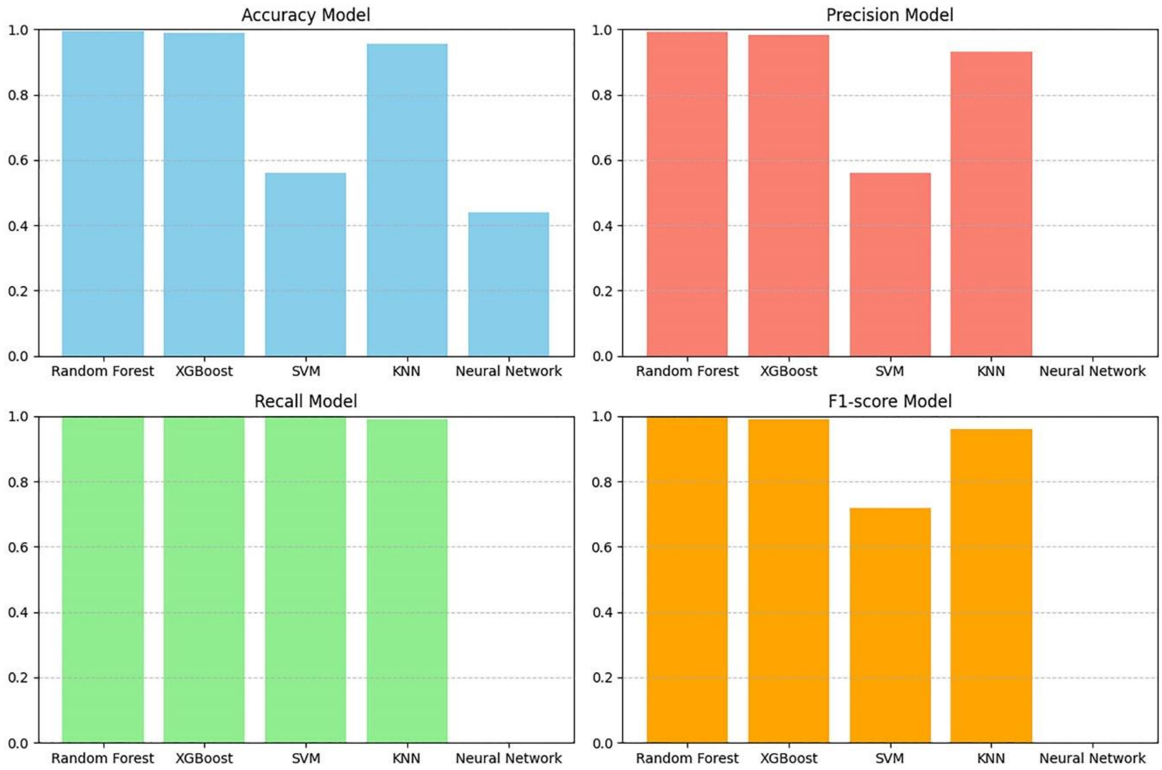

Fig. 4. Bar Chart Visualization of the Performance of Four Metrics in the Used Models.

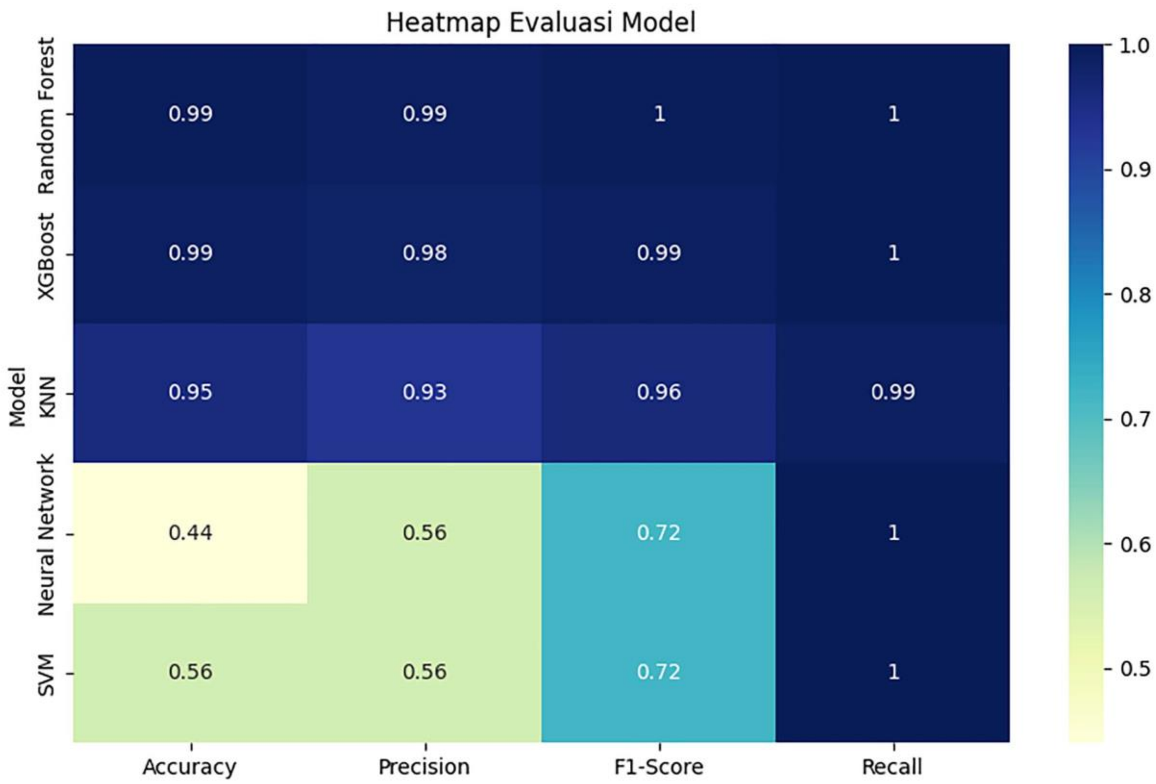

Fig.5. Heatmap representation depicting the performance of four evaluation metrics across the five models utilized.

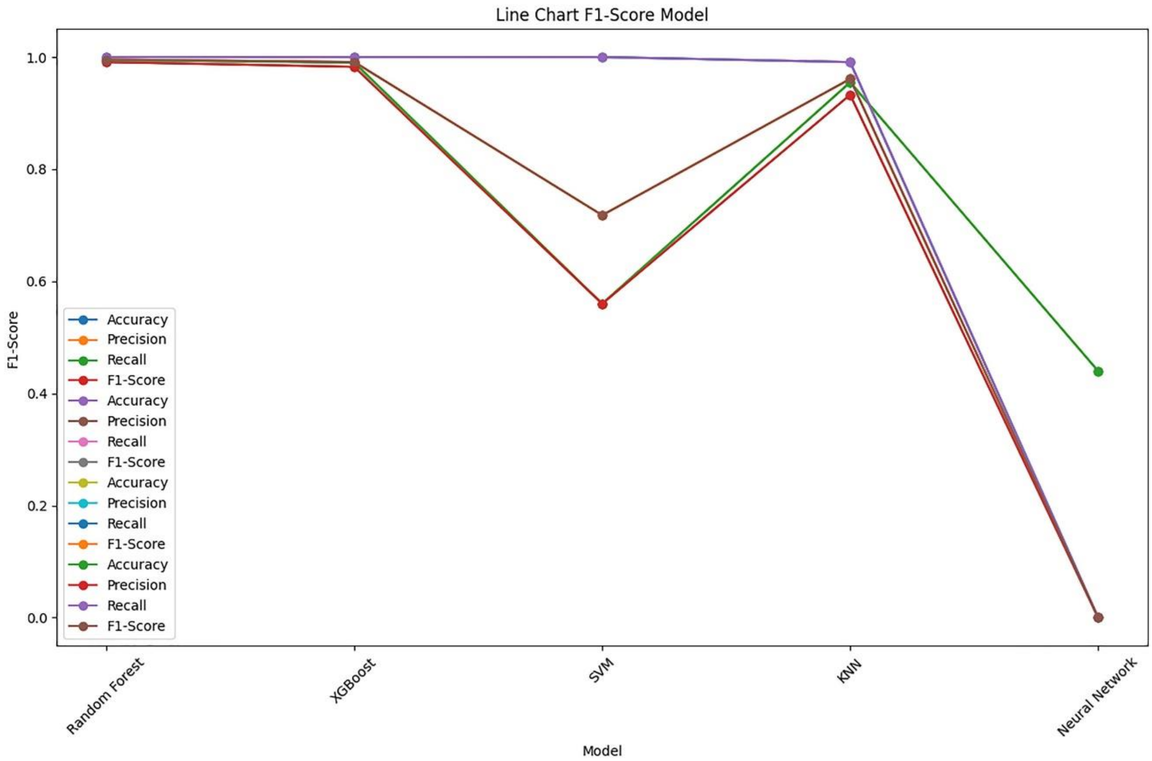

Fig. 6. Line graph illustrating the F1-score values across the five models utilized.