Applied Information System and Management (AISM)
Volume 8, (1) 2025, p. 141–150
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v8i1.45600

A Comparative Study of Machine Learning Models for Fashion Product Demand Prediction: Exploring Algorithms, Data Splitting, and Feature Engineering

Reviana Siti Mardiah[1], Fitrianingsih[2*]

***Abstract*—The fashion industry faces challenges in accurately predicting demand due to inherent uncertainty, leading to suboptimal inventory and financial losses. Machine learning (ML) offers a robust solution by analyzing large and complex data, identifying non-linear patterns, and providing more accurate predictions than conventional methods that rely on limited factors. This research aims to compare and evaluate the performance of six different ML models—XGBoost, SVM, RF, GBM, KNN, and NN, considering the influence of feature engineering and various data split ratios on predicting fashion product demand. KNN and NN were included due to distinct modeling approaches and competitive capabilities in identifying local and non-linear patterns across numerical, categorical, and time series data. Techniques such as feature extraction and selection and various data split ratios (70:30, 80:20, 90:10) were used. Using Adidas sales data, the models were evaluated based on Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). The results indicate that the XGBoost-based model with feature engineering consistently outperforms the other models across all data split ratios. Particularly, XGBoost with feature engineering at a data split ratio of 90:10 achieved the best performance with an RMSE of 4.46 and an MAE of 1.51. Analyzing model performance shows that the predictive ability of ML models is influenced by the implementation of feature engineering and the selection of the data split ratio. These results demonstrate the potential of using feature-engineered XGBoost models and optimized data ratios to mitigate the risk of stockouts or overstocks, and reduce financial losses and environmental waste.**

***Index Terms*—Data splitting, demand prediction, fashion product, feature engineering, machine learning.**

Received: 26 March 2025; Revised: 18 May 2025; Accepted: 30 May 2025.
*Corresponding author

[1]Reviana Siti Mardiah, Universitas Gunadarma Jakarta, Indonesia (e-mail: revianasitimardiah@staff.gunadarma.ac.id).
[2*]Fitrianingsih, Universitas Gunadarma Jakarta, Indonesia (e-mail: fitrianingsih@staff.gunadarma.ac.id).

I. Introduction

The fashion industry is a dynamic sector characterized by short product life cycles and unpredictable demand, primarily due to constantly changing customer preferences, trends, and consumer behaviour. This unpredictability makes it challenging to predict demand accurately [1], [2], leading to potential stockouts or overstocks, which can negatively affect the company [3]. Stockouts result in lost sales and decreased customer satisfaction, while overstocks lead to deadstock [4], which causes financial losses and environmental waste [3], [4], [5]. Therefore, accurate demand prediction is crucial to optimize inventory and minimize negative environmental impacts [6].

Conventional demand prediction methods such as naíve, moving average, trend, multiple linear regression, Holt-Winters, exponential smoothing, and ARIMA have been employed in fashion product demand prediction. However, these methods often fail to capture market volatility and rapid trend changes because it usually only consider one or a few simple factors such as trends, seasonality, or cycles [7], [8]. Additionally, conventional methods demand expertise and involve time-consuming processes that are susceptible to human error [9]. Furthermore, fashion data generally consists of a combination of numerical, categorical, and time features, necessitating a more flexible and adaptive approach [8].

To address these shortcomings, the use of AI-based methods, particularly Machine Learning (ML), can be a robust solution. ML can process diverse datasets and uncover hidden patterns within the data [7], [8]. This research aims to evaluate and compare the performance of six current machine learning algorithms, namely XGBoost, SVM, GBM, RF, KNN, and NN to predict the demand for fashion products. In addition, this research also considers the influence of feature engineering as well as data split ratio variation on the performance of the prediction model.

Research shows that ML has been widely applied in demand prediction and consistently outperforms conventional prediction methods by analyzing more influencing factors. ML can identify key underlying demand factors and uncover new insights by processing a large number of predictor variables and determining which ones are most significant. ML can also handle different types of data (numerical, categorical, nominal), identify non-linear patterns and demand predictors, and provide more accurate and adaptive predictions [8].

Several research studies have proposed machine learning algorithms to address the demand prediction problem. Recent research from [10] introduced a demand prediction model for fashion products using diverse retail data. The results showed that RF outperformed KNN, with an average Root Mean Square Error (RMSE) value of 209 for RF and 457 for KNN. This advantage is due to RF's ability to handle non-linear data, outliers, and multiple predictors, while KNN struggles with large, variable datasets.

In other research, [11] proposed a prediction model for cell phone products. This research analyzed twelve recent machine learning algorithms, namely LightGBM, XGBoost, Ridge, Recurrent Neural Network (RNN), Lasso, CatBoost, Deep Neural Network (DNN), SVM, RF, Gradient Boosting Machine (GBM), AdaBoost, and LSTM, and found that the RF algorithm has the best performance with MAPE 42.6258, RMSE 8443.3328, and correlation coefficient 0.8629. This shows that RF consistently produces higher prediction accuracy than other models on complex data.

Similarly, [12] proposed a demand prediction model for medicinal products in the pharmaceutical industry. This research analyzed six machine learning models, namely Simple Tree (ST), Gradient-Boosted Trees (GBT), LR, RF, Polynomial Regression (PR), and Tree Ensemble (TE). This research shows that RF and ST algorithms have the best performance and can improve prediction accuracy by 10%–40% by effectively managing varied data and highlighting important features.

Reference [13] compared five methods for predicting retail sales at Walmart, including RF, XGBoost, gradient boosting, AdaBoost, ANN and RF-XGBoost-LR. The combined model outperforms the others based on data of weekly sales, achieving MAE of 0.0024 and MSE of 4.7932e-05. However, both standalone RF and XGBoost showed strong performance in datasets with multiple factors.

In the energy sector, [14] proposed a demand prediction model by analyzing three machine learning algorithms: Medium-size Neural Networks (MNN), SVM, and Whale Optimization Algorithms (WOA). It was found that MNN has the best performance, proving its strength in modeling non-linear and handling fluctuating time-series data.

Although various machine learning algorithms such as RF and XGBoost, as well as neural networks have been often reported to perform well in terms of prediction accuracy, most previous research has generally been limited to specific sectors or datasets. There is an obvious gap in applying these models to demand prediction of complex fashion products, with numerical features, categorical variables, and time-series data. In particular, there is a lack of research that evaluates model performance by considering a combination of factors during the model development stage. These factors include the selection of ML algorithms, the implementation of feature engineering, and the selection of the data split ratio used. This research fills this gap by evaluating and comparing six recent machine learning models by considering the influence of feature engineering and different splitting ratios of training and testing data.

The results of this research are expected to provide recommendations for a more accurate demand prediction model for the fashion industry, which would support better decision-making in the industry's supply chain management.

The paper is structured as follows: Section 2 reviews related work, Section 3 outlines the research methods, Section 4 presents the results, discusses the findings, and provides future research suggestions, and Section 5 concludes with the main contributions of this research.

II. Related Work

Demand prediction plays a crucial role in optimizing business and manufacturing processes, significantly affecting a company's success [15]. This requires predicting future demand based on historical data, a task that becomes complicated in scenarios with incomplete or unpredictable information [16]. The nonstationarity and complexity of influencing factors further challenge accurate demand prediction.

Enhancing the accuracy of demand prediction can be effectively achieved through advanced methodologies, especially Machine Learning (ML). Various research has explored different ML algorithms designed to address demand prediction in various fields. For example, [17] proposed a Gated Recurrent Unit (GRU)-based approach that can manage large-scale, high-dimensional datasets and discern complex patterns better than traditional statistical methods. However, these models are prone to overfitting, thus requiring sophisticated regularization strategies.

Meanwhile [18] proposed four models specifically used to predict demand in catering services, aiming to mitigate the problem of over- and underproduction. This research includes various ML models, such as LightGBM, Long Short-Term Memory (LSTM), Random Forest (RF), and transformers, with LSTM achieving the most reliable predictions, with RMSE values between 60.9 and 173.36, reducing food wastage by 52%. However, this model lacks adaptability.

Other study [19] proposed a hybrid model that integrates ElasticNet, Gaussian Process Regression (GPR), and K-means clustering for demand prediction. This approach shows excellent prediction accuracy, reflected by the mean absolute error (MAE) of 5.57. However, its generalizability is still limited. In a similar research, [20] combined the GRU and Prophet to predict electricity demand, reducing the prediction error and revealing consumption patterns. However, the model had difficulty in dealing with data anomalies related to holidays.

Moreover [10] proposed a demand prediction model that utilizes RF and K-Nearest Neighbors (KNN). The results showed that the average RMSE of RF was 209, while KNN reached 457. Although the model successfully predicted the sales period, the findings faced challenges related to complex variable selection. In the fashion retail industry, [21] examined various analysis approaches, including Extreme Learning Machine (ELM), Support Vector Regression (SVR), and k-Means clustering. The results showed that the KM-ELM model was highly effective, exhibiting high accuracy even amidst demand uncertainty, an inherent characteristic of retail

P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v8i1.45600

dynamics. However, it ignored important variables such as fashion trends, economic conditions, and the impact of weather.

In other research, [11] proposed a prediction model for cell phone products. This research analyzed twelve recent machine learning algorithms, namely LightGBM, XGBoost, Ridge, Recurrent Neural Network (RNN), Lasso, CatBoost, Deep Neural Network (DNN), SVM, RF, Gradient Boosting Machine (GBM), AdaBoost, and LSTM, and found that the RF algorithm has the best performance with MAPE 42.6258, RMSE 8443.3328, and correlation coefficient 0.8629. Similarly, Also [12] proposed a demand prediction model for medicinal products with a case study in the pharmaceutical industry. This research analyzed six machine learning models, namely Simple Tree (ST), Gradient-Boosted Trees (GBT), LR, RF, Polynomial Regression (PR), and Tree Ensemble (TE). This research shows that RF and ST algorithms have the best performance and can improve prediction accuracy by 10%-40%.

Another study [13] compared five methods for retail sales, including RF, XGBoost, gradient boosting, AdaBoost, ANN, and RF-XGBoost-LR. The combined model outperforms the others based on weekly sales data from a retail company in the US. In the energy sector, [14] proposed a demand prediction model by analyzing three machine learning algorithms: Medium-size Neural Networks (MNN), SVM, and Whale Optimization Algorithms (WOA). It was found that MNN has the best performance.

Previous research on demand prediction models utilizing machine learning has revealed several significant shortcomings, including issues with overfitting that necessitate more sophisticated regularization techniques, as well as limitations in model adaptation and generalization. Furthermore, some research has struggled to effectively manage specific datasets or to select complex features, often overlooking critical variables such as fashion trends or economic conditions. These weaknesses are exacerbated by inadequate performance analysis and a lack of discussion surrounding data splitting strategies, both of which are crucial for establishing model validity. Addressing these shortcomings could enhance the accuracy of prediction models under real-world conditions across various industry scenarios [22].

There is considerable room for improvement, particularly in the fashion industry, which experiences significant demand uncertainty due to constantly changing customer preferences, trends, and consumer behavior. While a range of algorithms has been explored, the tendency to overlook essential aspects such as feature engineering and data splitting frequently results in suboptimal research outcomes when faced with realistic data conditions.

Given these deficiencies, it becomes clear that there is an urgent need for more comprehensive research that analyzes the influence of feature engineering and data splitting strategies on the performance of various machine learning algorithms.

Based on this urgency, this research not only explores various algorithms but also examines how feature engineering and data splitting strategies can significantly influence the results. By exploring these factors, it is anticipated that this research will yield deeper and more applicable insights into predictive models tailored for the fashion industry, thereby improving their effectiveness and reliability in practical applications.

III. Research Method

This research utilized a quantitative methodology using supervised machine learning for regression to predict the demand for Adidas products. This approach was chosen because this research focuses on analyzing numerical sales data from the Adidas Sales Dataset [23], which consists of numerical feature (retailer ID, price per unit, units sold, total sales, operating profit, operating margin), categorical feature (retailer, region, state, city, product, sales method, day classification, season, school vacation), and time feature (invoice date).

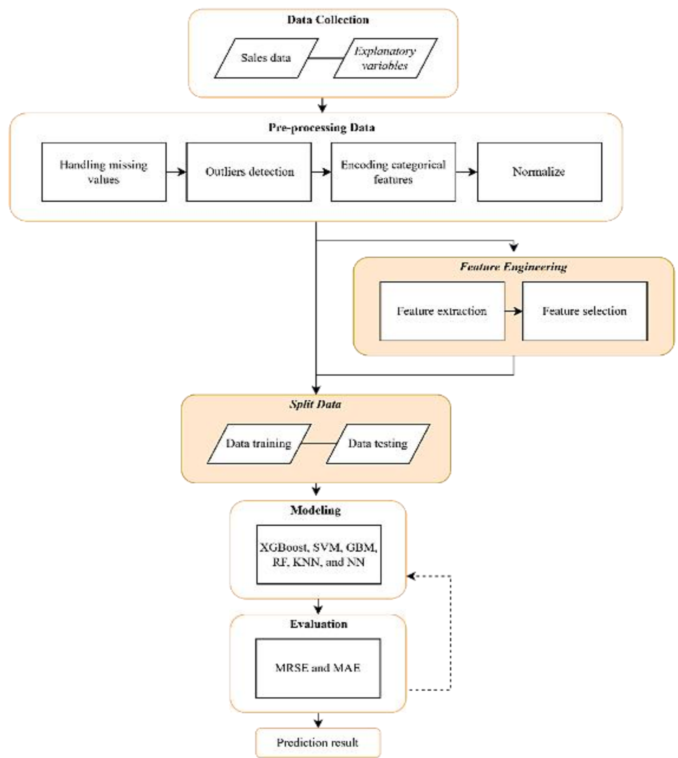

Fig. 1. The model development stage.

The model development stage involves several steps, starting with data collection of historical sales and explanatory variables. Followed by preprocessing to prepare the dataset for ML analysis. Further, feature engineering was performed to improve the interpretability of the model by constructing features that are easier to understand [24]. The dataset was then split into training and testing sets. Furthermore, the modeling process was carried out with hyperparameter optimization to improve the performance of the model, thus ensuring the development of a reliable prediction model. Then the model was evaluated using RMSE and MAE values. The most effective model can then be used for prediction and decision-making. Fig. 1 illustrates the sequential stages of the research methodology used in this research.

A. Data Collection

This research analyzed Adidas product sales data from Kaggle, consisting of 9,648 rows across six categories: men's and women's apparel, street shoes, and athletic shoes. The dataset covers sales from 2020 to 2021 and includes variables such as city, invoice date, operating margin, price per unit, product, retailer, region, units sold, total sales, sales method, and school holidays. Table 1 explains each variable in this research.

Table 1.
Variable Description

Variable	Description	Data Type
Retailer	Name of the retailer that sells the product	Categorical
Retailer ID	Retail identification number	Numerical
Invoice date	Date of transaction	Date
Region	The retail geographical area is located	Categorical
State	Retail states located	Categorical
City	Retail city located	Categorical
Price per unit	Price per product unit	Numerical
Product	Product category	Categorical
Units sold	Number of products sold in one transaction	Numerical
Total sales	Total sales in one transaction	Numerical
Operating profit	Profit to be gained	Numerical
Operating margin	Percentage of revenue	Numerical
Sales method	Selling method	Categorical
Day classification	Day of the week classification	Categorical
Season	Season of the year	Categorical
School vacation	Indicators of whether or not the school is on vacation	Categorical

B. Preprocessing Data

The preprocessing in this research involves handling missing values, outlier detection, encoding categorical features, and normalization. Since the dataset had no missing values, no replacement of missing values was necessary. Outliers were addressed using the nearest neighbor approach, which can provide valuable insights [25]. This research used one-hot encoding to convert categorical variables into binary vectors for analysis [26]. Lastly, the Interquartile Range (IQR) method was used for data normalization to ensure uniform feature scaling [27].

C. Features Engineering

This research applied two conditions: one with feature engineering and one without. This approach was taken to validate the findings from the research of Swaminathan and Venkitasubramony [2], which emphasizes that feature engineering is essential for predictive modeling in the fashion industry to achieve more accurate and reliable predictions. Additionally, it aimed to explore the influence of various variables and feature engineering on algorithm performance. This research would only involve feature extraction and selection, as shown in Fig. 2. Feature extraction was performed on the date attribute, which was further divided into six sub-features: date, day of week, day of year, month, week of year, and year.

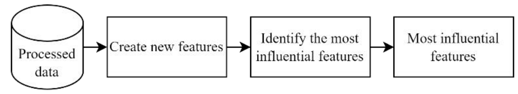

Fig. 2. Feature engineering stage.

D. Split Data

When developing a ML model, data is divided into training and testing sets. The model was trained with training data and tested with testing data to evaluate its performance in achieving the research objectives. To determine the optimal data division ratio, this research used three commonly adopted ratios, which were 70:30, 80:20, and 90:10, for training and testing data [28], [29].

E. Modeling with GridSearchCV

The modeling stage in this research was conducted for six machine learning algorithms: XGBoost, SVM, GBM, RF, KNN, and neural network, along with the implementation of hyperparameter tuning using GridSearchCV. GridSearchCV is a popular hyperparameter tuning method for machine learning [30], [31], [32]. Hyperparameters are key parameters that control model performance and can be tuned to optimize the results. GridSearchCV helps determine the best hyperparameter combination by implementing cross-validation and returning the combination with the best performance on the data [9]. Through the implementation of GridSearchCV, this research successfully identified configurations that significantly enhance the prediction accuracy of the model.

F. Model Evaluation

After training the model, the error rate was evaluated using RMSE and MAE. This was performed to measure each model's predictive accuracy, thus allowing an objective comparison of model performance. A lower RMSE indicates a more accurate prediction, while a lower MAE indicates a better match between demand results and actual data [33]. RMSE and MAE were determined using equations (1) and (2), where the number of data points (m), predicted value (x_i), and actual value (y_i).

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(x_i - y_i)^2} \quad (1)$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|x_i - y_i| \quad (2)$$

IV. RESULT

This section presents a model performance comparison, an overall discussion, and suggestions for future research.

A. Comparison of the Performance of Machine Learning Models

To identify the best-performing fashion product demand prediction model, this research conducted a comparative study involving various machine learning algorithms (XGBoost, SVM, GBM, RF, KNN, NN). In addition, to evaluate the

Applied Information System and Management (AISM)
Volume 8, (1) 2025, p. 141–150
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v8i1.45600

influence of feature engineering on model performance, two research scenarios were examined: with and without feature engineering. Furthermore, the influence of data splitting was also evaluated by experimenting with various ratios of training and testing data. The approach was applied to a sales dataset from six US retailers from 2020 to 2021.

Performance evaluation was performed using two key metrics, namely RMSE and MAE, as detailed in (1) and (2). To demonstrate its applicability, Table 2 presents an overview of data from a scenario without feature engineering using XGBoost, along with a 70:30 data split.

Table 2.
The First Three Actual and Predicted Data

No.	Unit Sold (Actual)	Predicted
1	75	77.83
2	200	200.06
…	…	…
2894	56	57.42

According to the data presented in Table 2, the RMSE (equation 1) and MAE (equation 2) can be calculated using the Unit sold as variable x and the predicted as variable y, as follows:

$$RMSE = \sqrt{\frac{(75-77.83)^2 + (200-200.06)^2 + \ldots + (56-57.42)^2}{2894}}$$
$$= 10.56$$
$$MAE = \frac{|75-77.83| + |200-200.06| + \ldots + |56-57.42|}{2894}$$
$$= 4.15$$

The calculation outlines the manual descent of the two matrices for the scenario without feature engineering, with a 70:30 data split using XGBoost. The manual and computational results exhibit similarities that align with the values presented in the first row of Table 3. In addition, Table 3 also presents the computational results for the scenario without feature engineering with variations in data split and other machine learning models, while Table 4 contains the results after the application of feature engineering. Furthermore, Fig. 3 offers a graphical representation comparing the results in the two conditions that enables a comprehensive understanding of the predictive ability of each model under various training and testing ratios.

Table 3 shows that in the scenario without feature engineering and using a 70:30 data split, the XGBoost model outperforms the other models with an RMSE of 10.56 and MAE of 4.15 ± 9.71. In contrast, the Support Vector Machine (SVM) performs poorly, with high RMSE and MAE values of 88.22 and 72.5 ± 50.25, respectively. Similar to XGBoost, gradient boosting also performs well, with an RMSE of 11.18 and MAE of 4.76 ± 10.12. In contrast, RF performs very poorly, with an RMSE of 150.42 and an MAE of 123.04 ± 86.53. K-nearest Neighbors (KNN) and neural networks have moderately good performance, with RMSE of 79.88 and 73.53, along with MAE of 62.04 ± 50.32 and 59.28 ± 43.51.

Table 3.
Performance Results Without Feature Engineering

Data Splitting	Model	RMSE	MAE
70:30	**XGBoost**	**10.56**	**4.15 ± 9.71**
	SVM	88.22	72.5 ± 50.25
	GBM	11.18	4.76 ± 10.12
	RF	150.42	123.04 ± 86.53
	KNN	79.88	62.04 ± 50.32
	NN	73.53	59.28 ± 43.51
80:20	**XGBoost**	**8.8**	**2.74 ± 8.37**
	SVM	70.78	56.14 ± 43.1
	GBM	10.23	3.49 ± 9.62
	RF	143.1	118.74 ± 79.87
	KNN	63.68	49.62 ± 39.91
	NN	41.34	33.75 ± 23.88
90:10	**XGBoost**	**5.65**	**2.28 ± 5.17**
	SVM	64.57	50.01 ± 40.85
	GBM	5.61	2.85 ± 4.83
	RF	142.69	113.51 ± 86.47
	KNN	58.96	44.82 ± 38.31
	NN	38.86	31.02 ± 23.41

In the 80:20 data split scenario, XGBoost continues to outperform, recording an RMSE of 8.8 and an MAE of 2.74 ± 8.37. SVM and RF still show the worst performance, with RMSE values of 70.78 and 143.1, and MAE values of 56.14 ± 43.1 and 118.74 ± 79.87. Gradient boosting remains competitive, with an RMSE value of 10.23 and an MAE value of 3.49 ± 9.62. KNN and neural networks also perform moderately well, with an RMSE of 63.68 and 41.34, along with an MAE of 49.62 ± 39.91 and 33.75 ± 23.88.

In the 90:10 data split scenario, XGBoost once again shows the best performance, with an RMSE value of 5.65 and MAE 2.28 ± 5.17. In contrast, SVM and RF continue to show the worst performance, with RMSE values of 64.57 and 142.69, and MAE values of 50.01 ± 40.85 and 113.51 ± 86.47. Gradient boosting and neural networks show significant improvement, with RMSE values of 5.61 and 38.86, and MAE values of 2.85 ± 4.83 and 31.02 ± 23.41. KNN records an RMSE of 58.96 and an MAE of 44.82 ± 38.31.

In Table 4, the results show that with feature engineering and a data split of 70:30, XGBoost markedly outperforms the other models, with an RMSE of 8.86 and an MAE of 2.87 ± 8.38. SVM shows improved performance, resulting in an RMSE of 78.02 and MAE of 59.6 ± 50.34. However, it still performs poorly compared to other models. Both gradient boosting and RF show significant improvement, achieving RMSE values of 10.06 and 39.72, along with MAE values of 3.41 ± 9.47 and 29.7 ± 26.37. KNN and neural networks also show improvement, with RMSE values of 47.83 and 46.54, along with MAE values of 34.87 ± 32.75 and 34.97 ± 30.71.

In the 80:20 data split scenario, XGBoost continues its excellent performance, with a significant decrease in RMSE to 7.51 and MAE to 1.65 ± 7.33. SVM improves but lags with an RMSE of 66.4 and an MAE of 53.63 ± 39.16. Gradient boosting and RF show more improvement, with RMSE values of 9.5 and 32.08, along with MAE values of 2.47 ± 9.18 and 25.09 ± 19.99. KNN also shows an improvement with an RMSE of 34.36 and an MAE of 24.03 ± 24.55, while the neural network shows a decrease in performance with an RMSE of 51.97 and an MSE value of 39.51 ± 33.76.

Table 4.
Performance Results With Feature Engineering

Data Splitting	Model	RMSE	MAE
70:30	**XGBoost**	**8.86**	**2.87 ± 8.38**
	SVM	78.02	59.6 ± 50.34
	GBM	10.06	3.41 ± 9.47
	RF	39.72	29.7 ± 26.37
	KNN	47.83	34.87 ± 32.75
	NN	46.54	34.97 ± 30.71
80:20	**XGBoost**	**7.51**	**1.65 ± 7.33**
	SVM	66.4	53.63 ± 39.16
	GBM	9.5	2.47 ± 9.18
	RF	32.08	25.09 ± 19.99
	KNN	34.36	24.03 ± 24.55
	NN	51.97	39.51 ± 33.76
90:10	**XGBoost**	**4.46**	**1.51 ± 4.2**
	SVM	61.61	49.13 ± 37.17
	GBM	**3.75**	**1.99 ± 3.18**
	RF	28.93	23.16 ± 17.33
	KNN	28.36	20.4 ± 19.7
	NN	41.16	30.25 ± 27.91

In the 90:10 data split, XGBoost and gradient boosting excelled with an RMSE of 4.46 and 3.75, along with an MAE of 1.51 ± 4,2 and 1.99 ± 3.18. SVM remains the most underperforming model with an RMSE of 61.61 and MAE of 49.13 ± 37.17. RF has an RMSE of 28.93 and an MAE of 23.16 ± 17.33. KNN and neural networks show RMSE of 28.36 and 41.16, with MAE of 20.4 ± 19.7 and 30.25 ± 27.91.

Overall, the XGBoost model consistently outperforms the other models in all scenarios, especially with feature-engineering scenarios. This implies that non-linear relationships and interactions between features play an important role in fashion product prediction [34], which XGBoost can effectively capture. On the other hand, SVM has a higher error rate, which indicates that the SVM model cannot capture demand patterns well or is less suitable for this type of time series prediction. GBM performs equally well with XGBoost, as both ensemble methods use decision trees [35]. However, XGBoost has a slight advantage over GBM, as it has more sophisticated regularisation techniques that prevent overfitting [36], [37].

On the other hand, RF shows high variability and generally higher errors, which can be caused by overfitting or not having the right ensemble size [38]. Meanwhile, KNN does not perform as well as tree-based models, which suggests that nearness-based methods are less able to capture patterns in demand forecasting effectively, and NN also shows moderate performance. However, the results could potentially be improved with a more complex architecture, more data, or further hyperparameter tuning.

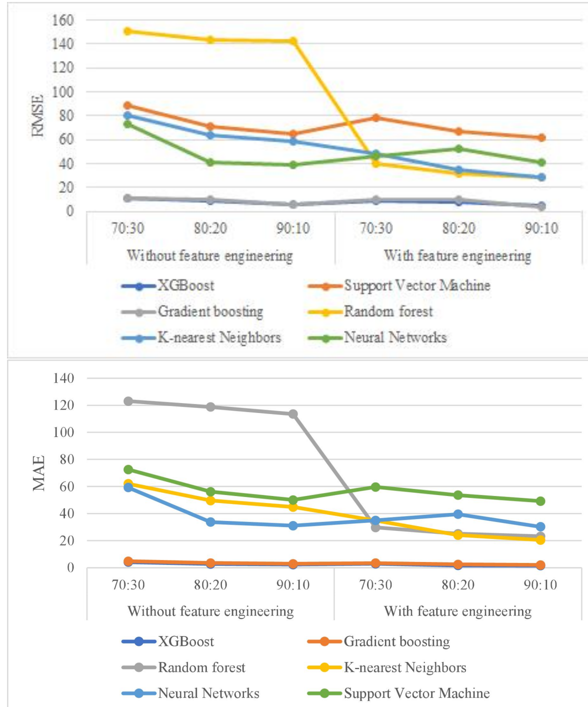

Fig. 3. Comparison of RMSE and MAE for six ML models.

The results also highlight that increasing the training data size can enhance most of the model's performance, as more data can help the model to learn better. This suggests that there is an influence of the data split ratio on model performance. Feature engineering, which involves feature extraction and selection, also significantly enhanced performance across all models. This suggests that feature engineering is influential and plays a crucial role in developing prediction models, especially in the context of fashion production demand prediction, where seasonality, trends, and other factors can affect demand [39].

This result indicates that engineered features in feature-engineered datasets are more predictive for fashion product demand prediction compared to non-engineered features. As a result, the research found that XGBoost with feature engineering is the most suitable choice for fashion product demand prediction, followed by GBM. Moreover, the analysis shows that the strategy to split the dataset and the implementation of feature engineering are important factors that influence the performance of the model. Therefore, practitioners must consider these strategies during the prediction modeling process to achieve optimal results.

B. Discussion

This section presents a detailed discussion of several aspects that influence the model's performance in predicting fashion product demand, ranging from the characteristics of the data explored through Exploratory Data Analysis (EDA), the influence of preprocessing and feature engineering techniques, the role of data split ratios, and the insights gained from the modeling results.

EDA was conducted to understand the data distribution patterns and relationships between variables. The pairplot of numerical features in Fig. 4 shows a strong non-linear relationship between units sold, total sales, and operating profit, highlighted in navy. This result is reinforced by the correlation

Applied Information System and Management (AISM)
Volume 8, (1) 2025, p. 141–150
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v8i1.45600

heatmap in Fig. 5, which shows high correlation coefficients between these features (highlighted in black), suggesting that the modeling process needs to be aware of potential data leakage, especially when total sales is used as a predictor of units sold.

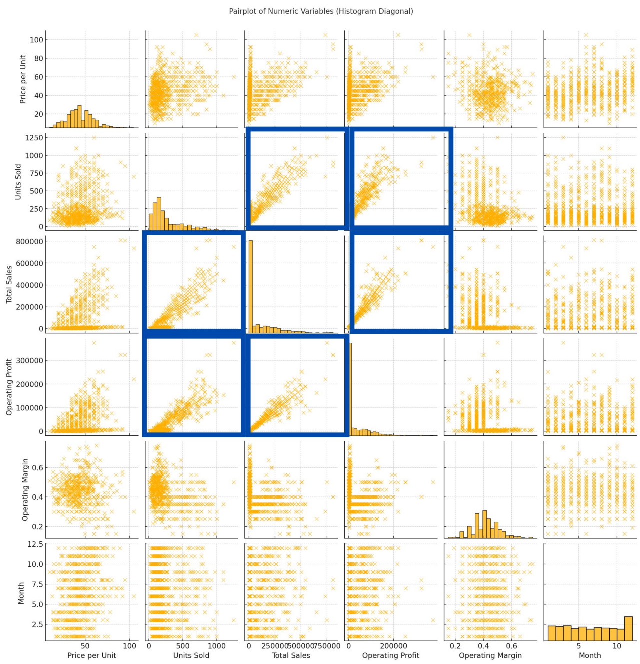

Fig. 4. Pairplot of numeric variables.

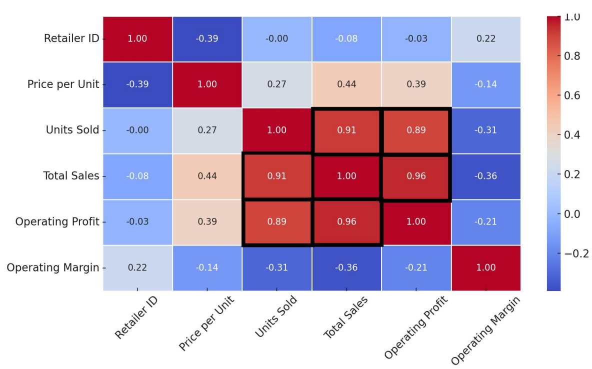

Fig. 5. Heatmap.

The histogram presented in Fig. 6 shows that the numerical features exhibit a right-skewed distribution, and outliers are present in almost all features, with total sales and operating profit being the most affected.

At the same time, the barplot illustrating the categorical features in Figure 7 shows the dominance of certain categories, such as Foot Locker among retailers and Portland among cities, which may cause model bias if not addressed with proper preprocessing.

Upon entering the preprocessing stage, it was discovered that there were no missing values, thus no replacement of missing values was necessary.

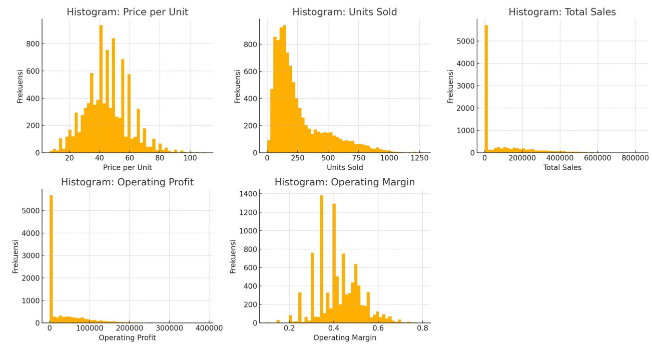

Fig. 6. Histogram of numeric features.

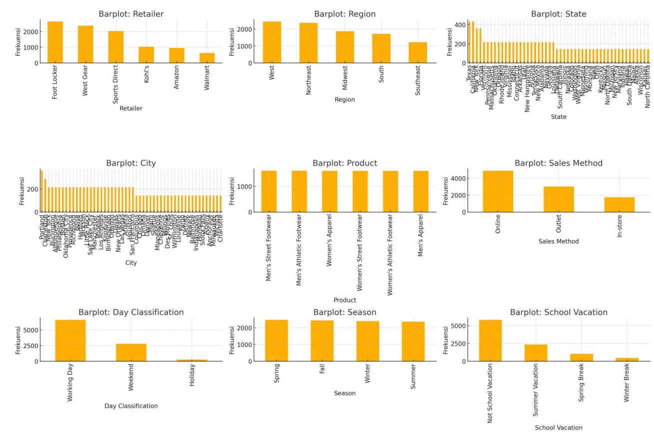

Fig. 7. Barplot of categorical features.

The boxplot of numerical features in Fig. 8 confirms the existence of outliers in nearly all features. However, all data was still used because outliers can provide valuable information, and distance-based detection methods are commonly used to handle them. As a result, normalization was applied to the numerical features to lessen the impact of outliers, except the targeted sold units, to ensure the model remains stable and accurate.

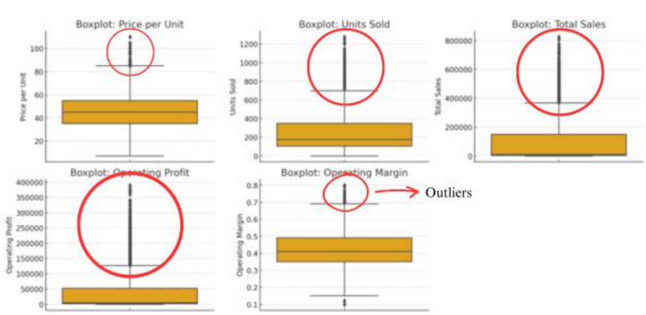

Fig. 8. Boxplots of numerical features show outliers in almost all features.

Feature selection produced a list of key features presented in Fig. 9, highlighting total sales, operating profit, and price per unit as the most significant, along with categories such as City_New York and Product_Men's Street Footwear. The prominence of these features suggests that the prediction model will rely heavily on transaction data associated with substantial sales in a particular category.

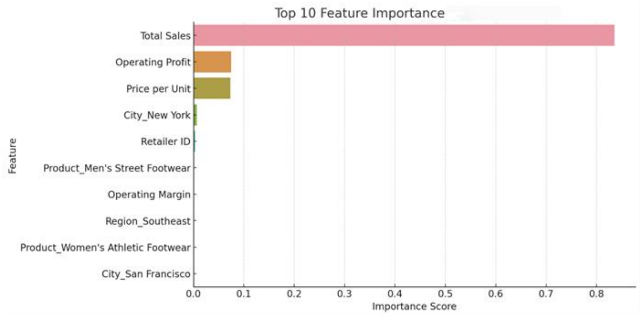

Fig. 9. The top ten features identified as important.

This analysis demonstrated the significance of feature selection and validation when creating an effective demand prediction model. Using features that are systematically connected to the target (units sold), such as total sales, should be avoided. Therefore, this feature was excluded as an input to ensure predictions rely solely on other variables, preventing data leakage.

Based on the research results presented in Table 4, two models demonstrate the best performance, namely the XGBoost and gradient boosting-based models. Both models show high prediction accuracy, with the main difference in the MAE and RMSE values. XGBoost has a lower MAE of 1.51 compared to gradient boosting's MAE of 1.99. In contrast, gradient boosting has a lower RMSE of 3.75 compared to XGBoost's RMSE of 4.46.

According to research by [40], MAE is easier to interpret because it directly measures the average error in the same units as the original data and in many cases. MAE is considered more reliable because it does not over-penalize large errors like RMSE. Thus, although the RMSE of XGBoost is higher, this is not necessarily an indicator of model weakness. The RMSE is more sensitive to outliers and can magnify the influence of some extreme errors, which may not be very relevant in certain contexts [41]. Therefore, MAE can be considered a more representative metric in assessing overall model performance, leading to the conclusion that the XGBoost-based model is more advantageous.

Furthermore, the results of this research demonstrate that XGBoost, when combined with feature engineering, has the potential to enhance the accuracy of fashion product demand prediction significantly. This enhancement is attributed to XGBoost's ability to handle complex data and mitigate overfitting through its built-in regularisation, such as L1 (Lasso) and L2 (Ridge) regularization, which helps the model adapt to large and dynamic datasets [36], [37], [42]. Given its effectiveness in demand prediction, particularly within the retail sector, integrating this model into enterprise analytics tools could lead to more precise demand predictions and improved inventory management. However, these results are not in line with previous research by [11], which indicated that RF was more advantageous than XGBoost. This difference may stem from product type, data, or methodology differences.

In this research, the XGBoost model used parameters such as n_estimators, max_depth, and learning_rate, which were adopted from the research of [43]. The range of parameters considered in this research can be seen in Table 5, and the result of hyperparameter optimization used GridSearchCV obtained the optimal combination of n_estimators set to 250, learning_rate at 0.1, and max_depth of 6 and 8.

Table 5.
Selected Parameters and The Values [43]

Hyperparameter	Value Range
Learning_rate	0,1-1
n_estimator	100-250
max_depth	1- 14

In addition, this research also highlights the importance of feature engineering in enhancing the performance of prediction models. Proper feature extraction and selection processes are crucial for allowing the model to capture more relevant patterns in the data, thereby improving prediction accuracy. This aligns with the research of Swaminathan and Venkitasubramony [2], which states that feature engineering is essential for fashion industry prediction models to produce more accurate and reliable predictions. Therefore, it is recommended that fashion retailers conduct a comprehensive analysis of the importance of features to identify and integrate the most predictive variables into their demand prediction models.

Differences in data-split ratios also influence the performance of the algorithms. For example, in some algorithms, the performance improvement was more pronounced at a 90:10 split ratio with feature engineering. XGBoost also showed greater robustness to changes in the data-split ratio and the utilization of feature engineering.

In general, this research confirms that XGBoost can be a robust model for the demand prediction of fashion products if supported with proper feature engineering and a data split ratio. While gradient boosting shows an advantage in reducing the impact of extreme errors, XGBoost is more flexible in handling complex data sets, making it an excellent choice for applications in the retail and inventory management sectors. This research illustrates the significant influence that feature engineering and data splitting have on model performance.

C. Future Research Area

This research comprehensively evaluates the effectiveness of tree-based machine learning (XGBoost, GBM, RF) and non-tree-based (SVM, KNN, and NN) across various dimensions. This includes not only investigating the influence of different data split ratios (70:30, 80:20, 90:10), but also the influence of feature engineering. By evaluating the model across various data splitting scenarios, this research seeks to determine model stability in response to changes in data proportions and ensure that the evaluation results are not biased towards any particular configuration. This approach is crucial to confirming that the model works effectively, not only under specific conditions but also maintains consistency across various data distributions.

Future research will delve into advanced feature engineering techniques, additional data variables, and more data split ratios to further optimize model performance. The effectiveness of the XGBoost algorithm in demand prediction presents further research opportunities to integrate it into other advanced technologies. For this reason, future research also plans to incorporate the demand prediction model with cloud-based IoT technology and Linear Programming optimization methods. IoT will be integrated for real-time inventory data collection, then feed data into the system that sends it to a database via the internet. Separately, historical sales and explanatory variable data will enter the demand prediction model. Finally, the data from the database and the prediction results will be integrated into the inventory optimization model to be analyzed to obtain the best decision regarding inventory purchases [6].

V. Conclusion

The findings of this research indicate that the XGBoost model demonstrates the best prediction performance for fashion product sales data from Adidas across all conditions and data split ratios. This model outperforms other models, namely KNN, GBM, RF, SVM, and NN. The optimal results were achieved under a scenario that employed feature engineering and a data split of 90:10. This scenario yielded an RMSE of 4.46 and MAE of 1.51, compared to the next best model (GBM), which had an RMSE of 3.75 and MAE of 1.99. This shows that XGBoost can produce more accurate and consistent predictions for this dataset.

This research also highlights the significant influence of feature engineering and data split ratio on model performance. The superior performance in the 90:10 scenario shows that a larger proportion of training data contributes to the model's improved capability in learning data patterns. Despite variations between tests, the average performance of the model remains consistent.

This research emphasizes the potential to integrate XGBoost into enterprise analytics for improved demand prediction, thereby improving inventory planning, stock reordering, and promotion scheduling based on demand patterns related to time, location, and sales channels. Therefore, this demand model can serve as a decision-making tool within the fashion industry.

While the application of the XGBoost algorithm has shown strong predictive capabilities, this research acknowledges limitations concerning the exploration of feature engineering techniques and the relatively conventional approach to data splitting. Consequently, future research is encouraged to explore more advanced feature engineering techniques and examine data split strategies. Additionally, future studies should consider the integration of multiple models to further refine prediction accuracy. This research also recommends the incorporation of the XGBoost algorithm with cloud-based IoT technology and optimization methods to streamline inventory management, effectively balancing demand with stock levels and mitigating the risks associated with excess or insufficient inventory, which can lead to revenue losses, competitive disadvantages, and adverse environmental effects.

References

[1] M. Koren and M. Shnaiderman, “Forecasting in the fashion industry: a model for minimising supply-chain costs,” *International Journal of Fashion Design, Technology and Education,* vol. 16, no. 3, pp. 30–318, 2023. doi: 10.1080/17543266.2023.2201508

[2] K. Swaminathan and R. Venkitasubramony, “Demand forecasting for fashion products: A systematic review,” *International Journal of Forecasting,* vol. 40, no. 1, pp. 247–267, 2024, doi: 10.1016/j.ijforecast.2023.02.005

[3] R. Szabó-Geletóczki, E. Szabó, and I. Rudnák, “Stockpile management through the everyday operation of a pharmaceutical company,” *Management/Vadyba,* vol. 38, no. 1, pp. 51–62, 2022, doi: 10.38104/vadyba.2022.1.06.

[4] X. Long and L. Gui, “Waste not want not? The environmental implications of quick response and upcycling,” *SSRN Electronic Journal*, Jan. 2022, doi: 10.2139/ssrn.4013877.

[5] A. R. Chowdhury, A. M. Mithu, S. Ahmad, and A. A. Malek, “A data-driven approach to inventory control under uncertain demand for pharmaceutical products using continuous review policy,” *GPH-International Journal of Business Management,* vol. 7, no. 01, pp. 32–51, 2024. doi: 10.5281/zenodo.10700678.

[6] A. Shah, R. M. Ellahi, U. Nazir, and M. A. Soomro, “Forecasting practices in textile and apparel export industry: A systematic review,” *International Journal of Circular Economy Waste Management,* vol. 2, no. 1, pp. 1–17, 2022, doi: 10.4018/IJCEWM.288501.

[7] C. Giri and Y. Chen, “Deep learning for demand forecasting in the fashion and apparel retail industry,” *Forecasting,* vol. 4, no. 2, pp. 565–581, 2022, doi: 10.3390/forecast4020031.

[8] M. Kharfan, V. W. K. Chan, and T. Firdolas Efendigil, “A data-driven forecasting approach for newly launched seasonal products by leveraging machine-learning approaches,” *Annals of Operations Research,* vol. 303, no. 1–2, pp. 159–174, 2021. doi: 10.1007/s10479-020-03666-w.

[9] Y. Ledmaoui, A. El Maghraoui, M. El Aroussi, R. Saadane, A. Chebak, and A. Chehri, “Forecasting solar energy production: A comparative study of machine learning algorithms,” *Energy Reports,* vol. 10, pp. 1004–1012, 2023, doi: 10.1016/j.egyr.2023.07.042.

[10] İ. Güven, Ö. Uygun, and F. Şİmşİr, “Machine learning algorithms with intermittent demand forecasting: An application in retail apparel with plenty of predictors,” *Textile Apparel,* vol. 31, no. 2, pp. 99–110, 2021, doi: 10.32710/tekstilvekonfeksiyon.809867.

[11] S. Hwang, G. Yoon, E. Baek, and B.-K. Jeon, “A sales forecasting model for new-released and short-term product: A case study of mobile phones,” *Electronics,* vol. 12, no. 15, Art. no. 3256, 2023, doi: 10.3390/electronics12153256.

[12] L. P. E. Yani and A. Aamer, “Demand forecasting accuracy in the pharmaceutical supply chain: a machine learning approach,” *International Journal of Pharmaceutical Healthcare Marketing,* vol. 17, no. 1, pp. 1–23, 2023. doi: 10.1108/IJPHM-05-2021-0056

[13] A. Mitra, A. Jain, A. Kishore, and P. Kumar, “A comparative study of demand forecasting models for a multi-channel retail company: a novel hybrid machine learning approach,” *Operations Research Forum*, 2022, vol. 3, no. 4, Sep. 2022, doi: 10.1007/s43069-022-00166-4.

[14] M. Saglam, C. Spataru, and O. A. Karaman, “Forecasting Electricity Demand in Turkey Using Optimization and Machine Learning Algorithms,” *Energies,* vol. 16, no. 11, Art. no. 4499, 2023. doi: 10.3390/en16114499.

[15] A. Brüggen, I. Grabner, and K. L. Sedatole, “The folly of forecasting: The effects of a disaggregated demand forecasting system on forecast error, forecast positive bias, and inventory levels,” *The Accounting Review,* vol. 96, no. 2, pp. 127–152, 2021, doi: 10.2308/tar-2018-0559.

[16] F. Yiğit, Ş. ESNAF, and B. Y. KAVUŞ, “A poisson-regression, support vector machine and grey prediction based combined forecasting model proposal: a case study in distribution business,” *Turkish Journal of Forecasting,* vol. 5, no. 2, pp. 23–35, 2021, doi: 10.34110/forecasting.957494.

[17] I. Amellal, A. Amellal, H. Seghiouer, and M. Ech-Charrat, “An integrated approach for modern supply chain management: Utilizing advanced

machine learning models for sentiment analysis, demand forecasting, and probabilistic price prediction," *Decision Science Letters,* vol. 13, no. 1, pp. 237–248, 2024, doi: 10.5267/j.dsl.2023.9.003.
[18] M. Rodrigues, V. Miguéis, S. Freitas, and T. Machado, "Machine learning models for short-term demand forecasting in food catering services: A solution to reduce food waste," *Journal of Cleaner Production,* vol. 435, Art. no. 140265, 2024, doi: 10.1016/j.jclepro.2023.140265.
[19] D. Chung, C. G. Lee, and S. Yang, "A hybrid machine learning model for demand forecasting: combination of k-means, elastic-net, and gaussian process regression," *International Journal of Intelligent Systems Applications in Engineering,* vol. 11, no. 6s, pp. 325–336, 2023.
[20] N. Son and Y. Shin, "Short-and medium-term electricity consumption forecasting using prophet and GRU," *Sustainability,* vol. 15, no. 22, Art. no. 15860, 2023, doi: 10.3390/su152215860.
[21] I.-F. Chen and C.-J. Lu, "Demand forecasting for multichannel fashion retailers by integrating clustering and machine learning algorithms," *Processes,* vol. 9, no. 9, Art. no. 1578, 2021, doi: 10.3390/pr9091578.
[22] Q. H. Nguyen *et al.*, "Influence of data splitting on performance of machine learning models in prediction of shear strength of soil," *Mathematical Problems in Engineering,* vol. 2021, pp. 1–15, 2021, doi: 10.1155/2021/4832864.
[23] J. Kamiri and G. Mariga, "Research methods in machine learning: A content analysis," *International Journal of Computer Information Technology,* vol. 10, no. 2, pp. 78–91, Mar. 2021, doi: 10.24203/ijcit.v10i2.79.
[24] S. Ozdemir, *Feature Engineering Bookcamp*. Simon and Schuster, 2022.
[25] J. Yang, X. Tan, and S. Rahardja, "Outlier detection: How to select k for k-nearest-neighbors-based outlier detectors," *Pattern Recognition Letters,* vol. 174, pp. 112–117, 2023, doi: 10.1016/j.patrec.2023.08.020.
[26] M. K. Dahouda and I. Joe, "A deep-learned embedding technique for categorical features encoding," *IEEE Access,* vol. 9, pp. 114381–114391, 2021, doi: 10.1109/ACCESS.2021.3104357.
[27] A. Alabrah, "An improved CCF detector to handle the problem of class imbalance with outlier normalization using IQR method," *Sensors,* vol. 23, no. 9, Art. no. 4406, 2023, doi: 10.3390/s23094406.
[28] Q. H. Nguyen *et al.*, "Influence of data splitting on performance of machine learning models in prediction of shear strength of soil," *Mathematical Problems in Engineering,* vol. 2021, pp. 1–15, 2021, doi: 10.1155/2021/4832864.
[29] A. Grigorev, *Machine Learning Bookcamp: Build a Portfolio of Real-life Projects*. Shelter Island: Manning Publications co, 2021.
[30] N. Shirzadi, A. Nizami, M. Khazen, and M. Nik-Bakht, "Medium-term regional electricity load forecasting through machine learning and deep learning," *Designs,* vol. 2021, no. 5, Art. no. 27, 2021, doi: 10.3390/designs5020027.
[31] H. Darmawan, M. Yuliana, M. Hadi, and Z. Samsono, "GRU and XGBoost performance with hyperparameter tuning using gridsearchcv and bayesian optimization on an iot-based weather prediction system," *International Journal on Advanced Science, Engineering Information Technology,* vol. 13, no. 3, p[. 851–862, 2023, doi: 10.18517/ijaseit.13.3.18377.
[32] D. D. M. C. Maceda and J. C. D. Cruz, "Rainfall Classification Model for the Philippines using Optimized K-nearest Neighbor Algorithm with GridSearchCV Hyperparameter Tuning," in *2023 IEEE 13th International Conference on Control System, Computing and Engineering (ICCSCE)*, 2023, pp. 51–55, doi: 10.1109/ICCSCE58721.2023.10237156.
[33] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Computer Science,* vol. 7, Art. no. e623, 2021, doi: 10.7717/peerj-cs.623.
[34] P. Dhal and C. Azad, "A comprehensive survey on feature selection in the various fields of machine learning," *Applied Intelligence,* vol. 52, no. 4, pp. 4543–4581, 2022, doi: 10.1007/s10489-021-02550-9.
[35] S. Demir and E. K. Sahin, "An investigation of feature selection methods for soil liquefaction prediction based on tree-based ensemble algorithms using AdaBoost, gradient boosting, and XGBoost," *Neural Computing Applications,* vol. 35, no. 4, pp. 3173–3190, 2023, doi: 10.1007/s00521-022-07856-4.
[36] C. Qin, Y. Zhang, F. Bao, C. Zhang, P. Liu, and P. Liu, "XGBoost optimized by adaptive particle swarm optimization for credit scoring," *Mathematical Problems in Engineering,* vol. 2021, pp. 1–18, 2021, doi: 10.1155/2021/6655510.
[37] J. Wang and S. Zhou, "Particle swarm optimization-XGBoost-based modeling of radio-frequency power amplifier under different temperatures," *International Journal of Numerical Modelling: Electronic Networks, Devices Fields,* vol. 37, no. 2, Art. no. e3168, 2024, doi: 10.1002/jnm.3168.
[38] L. Barreñada, P. Dhiman, D. Timmerman, A.-L. Boulesteix, and B. Van Calster, "Understanding overfitting in random forest for probability estimation: a visualization and simulation study," *Diagnostic Prognostic Research,* vol. 8, no. 1, Sep. 2024, doi: 10.1186/s41512-024-00177-1.
[39] J. Oh, K.-J. Ha, and Y.-H. Jo, "A predictive model of seasonal clothing demand with weather factors," *Asia-Pacific Journal of Atmospheric Sciences,* vol. 58, no. 5, pp. 667–678, 2022, doi: 10.1007/s13143-022-00284-3.
[40] S. M. Robeson and C. J. Willmott, "Decomposition of the mean absolute error (MAE) into systematic and unsystematic components," *PloS one,* vol. 18, no. 2, Art. no. e0279774, 2023, doi: 10.1371/journal.pone.0279774.
[41] T. O. Hodson, "Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not," *Geoscientific Model Development Discussions,* vol. 2022, pp. 1–10, 2022. doi: 10.5194/gmd-15-5481-2022.
[42] J. Wu, Y. Li, and Y. Ma, "Comparison of XGBoost and the neural network model on the class-balanced datasets," in *2021 IEEE 3rd international conference on frontiers technology of information and computer (ICFTIC)*, 2021, pp. 457–461, doi: 10.1109/ICFTIC54370.2021.9647373.
[43] O.-E. Ørebæk and M. Geitle, "Exploring the Hyperparameters of XGBoost Through 3D Visualizations," in *AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering*, 2021.