Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871

Vol. 6, No. 5, October 2025, Page. 3832-3843
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5292

MSMEs Recommendation System using Item-Based Collaborative Filtering and LightGBM Machine Learning

Mar'atuttahirah*[1], Khaera Tunnisa[2], Danang Fatkhur Razak Ra[3], Hafizah Najwa[4], Januar Fahrisal[5]

[1,2,3,4,5]Information System, Institut Teknologi Bacharuddin Jusuf Habibie, Indonesia

Email: [1]maratuttahirah.ir@ith.ac.id

Received : Aug 26, 2025; Revised : Oct 8, 2025; Accepted : Oct 16, 2025; Published : Oct 23, 2025

Abstract

Micro, Small, and Medium Enterprises (MSMEs) face challenges in recommendation systems for digital economy growth, particularly in participatory development and sustainable revenue optimization. This study aims to develop a recommendation system using Item-Based Collaborative Filtering and LightGBM for stock prediction and item recommendation at Kedai Pesisir MSME. Based on 1,229 transaction records from January to July 2025, we performed preprocessing, feature engineering, and LightGBM training to generate daily stock predictions and monthly priorities for August 2025 to January 2026. Evaluation yielded RMSE 0.069, MAE 0.034, and MAPE 1.14%, indicating high accuracy. This advances informatics by providing a scalable AI tool for MSME inventory management and revenue enhancement, supporting strategic decisions in dynamic markets

Keywords : *Collaborative Filtering, Item, Feature, Machine Learning, Revenue, Recomendation.*

1. INTRODUCTION

Information technology serves as a tool for sustaining business activities by facilitating transactions and managing innovation aspects to ensure more effective business processes[1] .High-quality economic development is characterized by innovation as the main driver, along with coordination, openness, sustainability, and the capacity to meet society's increasing needs for a better quality of life[2]. The digital economy, as a technology-driven system, accelerates innovation, efficiency, and productivity in supporting efforts toward achieving high-quality economic development[3]. The presence of technology fosters continuous advancements across various disciplines, particularly in the field of economics[4].

Micro, Small, and Medium Enterprises (MSMEs) are business actors engaged in diverse sectors that directly affect society. In Indonesia, MSMEs are currently considered an effective means of poverty alleviation. Based on statistics and research, MSMEs represent the largest share of business groups. They are legally regulated through Law Number 20 of 2008 concerning Micro, Small, and Medium Enterprises, serving as a safeguard for the national economy during crises and as a catalyst for economic recovery[5]. Data from the Ministry of Cooperatives and MSMEs show that MSMEs contribute significantly to Indonesia's economy, accounting for more than 61% of GDP, or approximately IDR 9,580 trillion[6].

Parepare City, as a rapidly developing municipality, has placed strong emphasis on MSME development due to its crucial role in economic growth. According to the Parepare Office of Cooperatives and MSMEs, more than 2,000 MSMEs have been established and continue to grow in the city. Among various business sectors, the culinary industry dominates[7]. Within this context, the fish-processing MSME group *Kedai Pesisir*, located in Ujung District, Parepare, operates in the food sector,

Vol. 6, No. 5, October 2025, Page. 3832-3843
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5292

serving as a source of income for coastal communities. The independence of *Kedai Pesisir* is closely tied to one of its main objectives—utilizing technology for two essential purposes: participatory growth processes and sustainable development to improve income levels.

The revenue of a business is influenced by rising demand and the availability of sufficient stock[8]. Collaborative Filtering has emerged as a paradigm for parameterizing users and items based on correlation patterns derived from interaction or purchase data[9]. The application of Collaborative Filtering models in business recommendation systems, which present items to users based on preferences and behavioral patterns, is particularly relevant for influencing strategic decision-making, revenue generation, and inventory management[10].

Machine learning has become a central component of recent advancements in Artificial Intelligence (AI)[11], particularly in recommendation systems, personalization trends, and internet accessibility. However, these systems face various limitations and challenges, including scalability and the cold-start problem[12]. Numerous studies have been conducted on recommendation systems using various Machine Learning methods. For instance, Jing Yu et al. developed a recommendation system using the MovieLens 1M dataset, which demonstrated superior performance in predicting customer preferences[13]. Recommendation systems operate by identifying items with similar characteristics, such as music tracks and movies[14]. Xu Yua et al. introduced a recommendation system that incorporated customer behavioral dynamics through Self-Supervised Imitation (SSI), demonstrating flexibility and superior performance [15]. Similarly, item-based recommendation systems utilizing customer clickstream data showed that SQN and SAC methods significantly improved item recommendation accuracy [16]. Among widely used Machine Learning models, the LightGBM algorithm has been applied in various domains. For example, Chyntia et al. employed LightGBM to predict research article methodological quality, achieving 57% accuracy in retrospective validation and 53% in prospective studies[17], Yue Wu et al. applied LightGBM for market volatility prediction, reporting an RMSPE score of 0.211 [18], Dhruva et al. utilized LightGBM to detect Parkinson's disease, demonstrating higher efficiency and accuracy compared to XGBoost [19], Xiaomin et al. implemented LightGBM in a personalized web content recommendation system, achieving MAE of 1.08% and RMSE of 2.41%, while highlighting challenges related to big data management and diverse user preferences [20].

The main challenge in Collaborative Filtering lies in handling increasingly complex data [21], as user–item interactions are often sparsely distributed, which limits the representational capacity of the paradigm [9]. In the business context, the ability to identify customer purchasing patterns significantly influences strategic decisions, such as inventory management [22]. These challenges underscore the necessity for innovation in recommendation systems for business applications. Therefore, this study aims to develop a system capable of recognizing patterns within data, improving prediction and recommendation accuracy based on customer–item interaction data, and providing MSMEs with predictive insights and stock recommendations for several upcoming months. This serves as an added value in the broader effort of national economic transformation toward an innovation-driven economy, utilizing an Item-Based Collaborative Filtering model with a Machine Learning approach through the LightGBM algorithm.

2. METHOD

This study employs the LightGBM algorithm with the objective of identifying patterns within the data, enhancing the accuracy of recommendations based on customer–item interaction data from MSME transactions, and providing predictive insights as well as stock recommendations for the upcoming months. The algorithm was selected due to its capability to efficiently discover patterns, its computational speed, and its low memory consumption [20]. As a result, the developed model is

Vol. 6, No. 5, October 2025, Page. 3832-3843
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5292

expected to accurately capture sales data patterns across multiple attributes obtained from MSMEs. It subsequently generates predictions and provides recommendations on products likely to experience higher demand in the following months, thereby enabling MSMEs to prioritize inventory for those products. The research workflow is illustrated in Figure 1.

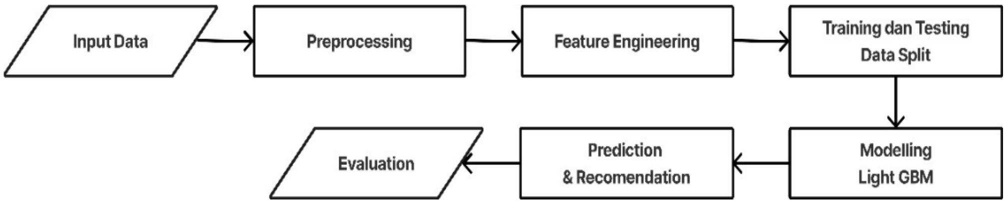

Figure 1. Methodological Steps used in this Study

2.1. Data Collection

This study utilizes historical sales data from the *Kedai Pesisir* MSME, comprising 1,229 transaction records collected over the period of January 2025 to July 2025.

2.2. Preprocessing

At this stage, data cleaning is a process used to detect, correct, and handle incomplete and inaccurate datasets, tables, or databases. This process refers to the treatment of *dirty data*, which may be replaced, modified, or removed after the identification of incomplete, invalid, inconsistent, or irrelevant data [23]. In this study, checking for missing values was the initial step in data preprocessing to identify potential data gaps, followed by either removing rows or columns with missing values or imputing them with appropriate substitutions.

Furthermore, the Date column was converted into the datetime format, while the Unit Price column was cleaned by removing non-numeric characters (e.g., “Rp” and “.”). Subsequently, a data transformation step was performed, converting the dataset from .xlsx format into .csv format for model input. The sales data covering the period of January 2025–July 2025 in both .xlsx and .csv formats are presented in Table 1 and Table 2, respectively.

Table 1. Data in .xlsx format

No.	Date	Product Name	Variant	Package Weight	Quantity	Unit Price	Total
1	13/01/2025	Fish Floss	Spicy	100 gr	19	Rp 22.000	Rp 218.000
2	13/01/2025	Fish Floss	Sweet	200 gr	3	Rp 38.000	Rp 114.000
3	13/01/2025	Chiken Floss	Spicy	200 gr	7	Rp 43000	Rp 301.000
4	13/01/2025	Chiken Floss	Spicy	100 gr	2	Rp 27000	Rp 54.000
5	13/01/2025	Fish Floss	Spicy	500 gr	2	Rp 80.000	Rp 160.000
6	13/01/2025	Fish Floss	Sweet	500 gr	3	Rp 88.000	Rp 176 .000
…	…	…	…	…	…	…	…
1226	31/07/2025	Chiken Floss	Sweet	200 gr	2	Rp 43.000	Rp 86.000
1227	31/07/2025	Beef Floss	Spicy	500 gr	1	Rp 88.000	Rp 88.000
1228	31/07/2025	Fish Floss	Sweet	200 gr	5	Rp 38000	Rp 190.000
1229	31/07/2025	Chiken Floss	Spicy	500 gr	1	Rp 85.000	Rp 85.000

Vol. 6, No. 5, October 2025, Page. 3832-3843
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5292

Table 2. Data in .csv format

No.	Date, Product Name, Variant, Package Weight, Quantity, Unit Price, Total
1	13/01/2025,Fish Floss,Spicy,100 gram,19,"Rp22.000,00","Rp218.000,00"
2	13/01/2025,Fish Floss,Sweet,200 gram,3,"Rp38.000,00","Rp114.000,00"
3	13/01/2025,Chiken Floss,Spicy,200 gram,7,"Rp43.000,00","Rp301.000,00"
4	13/01/2025,Chiken Floss,Spicy,100 gram,2,"Rp27.000,00","Rp54.000,00"
5	13/01/2025,Fish Floss,Spicy,500 gram,2,"Rp80.000,00","Rp160.000,00"
6	13/01/2025,Beef Floss,Sweet,500 gram,2,"Rp88.000,00","Rp176.000,00"
…	…
1226	31/07/2025,Chiken Floss,Sweet,200 gram,2,"Rp43.000,00","Rp86.000,00"
1227	31/07/2025,Beef Floss,Spicy,500 gram,1,"Rp88.000,00","Rp88.000,00"
1228	31/07/2025,Fish Floss,Sweet,200 gram,5,"Rp38.000,00","Rp190.000,00"
1229	31/07/2025,Chiken Floss,Spicy,500 gram,1,"Rp85.000,00","Rp85.000,00"

2.3. Feature Engineering

Feature engineering involves the design of several well-constructed features [24], such as Temporal Features, Contextual Features to mark periods of high demand, Lag Features to capture short-term trends, Rolling Features to represent long-term trends, and Annual Cyclical Features to assist the model in recognizing 12-month seasonal patterns. Several of the features utilized in this study are presented in Table 3.

Table 3. Features Table

No.	Features	Feature Description
1	Time Features	Month, Year, Week_of_Year, Day_of_Week, and Week_of_Month
2	Contextual Features	Season
3	Lag Features	Sales_Last_Month, Sales_2Months_Ago, and Sales_3Months_Ago
4	Rolling Features	Avg_Sales_3Months, Avg_Sales_6Months, and Std_Sales_3Months
5	Annual Cyclical Features	Month_Sin and Month_Cos

2.4. Training and Testing Data Split

The dataset was divided into training and testing subsets to evaluate model performance and ensure its ability to generalize effectively. In this study, 80% of the data was allocated for training and 20% for testing.

2.5. Item Based Collaborative Filtering

Item-based collaborative filtering, or product-based collaboration, provides recommendations to users by assessing the similarity values between items or products. These similarity values are calculated based on the ratings given by users. Adjusted Cosine Similarity is one of the key stages in the item-based collaborative filtering method, used to compute similarity scores between items and to identify the most similar ones.

Vol. 6, No. 5, October 2025, Page. 3832-3843
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5292

In this approach, for any pair of items i and j, a similarity measure (S_{ij}) is determined by applying the following formula:

$$sim\ (i,j) = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}u)(R_{u,j} - \bar{R}u)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}u)^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}u)^2}} \qquad (1)$$

Where,

sim(i,j)	= Similarity value between items *i* and *j*
u € U	= Set of users who have rated both items
$R_{u,i}$	= Rating given by user *u* to item *i*
$R_{u,i}$	= Rating given by user *u* to item *j*
$\bar{R}$u	= Avarage rating value of user *u*

2.6. LightGBM

This study employed the LightGBM algorithm. The fundamental principles and processes of LightGBM are described as follows:

First, initialize *F0(x)* learner based on a priori information, as in Equation (2)

$$f_0(\mathrm{x}) = lg\,\frac{P(y=1|x)}{1-P(y=1|x)} \qquad (2)$$

where *P(y = 1|x)* represents the proportion of sample *y = 1* in training samples.

Then, *M* regression trees are established, and the residual value of the *m*th tree is calculated according to Equation (3):

$$r_{m,i} = yi - \frac{1}{1+e-F(xi)} \qquad (3)$$

where *F(xi)* and *yi* are fitting data and *rm,i* stands for leaf node region.

Finally, strong learner *FM(x)* is calculated according to Equation (4):

$$f_M(\mathrm{x}) = f_0(\mathrm{x}) + \sum_{m=1}^{M} \sum_{j=1}^{Jm} \left|\frac{yi-\hat{y}i}{yi}\right| c,jI, \quad x \in r_{m,j} \qquad (4)$$

where *j* represents the *j*th tree; *cm,jI* is the residual value corresponding to the *j*th tree; *I* is a random number[25].

In this study, we adopted optimization techniques for training and prediction in LightGBM, namely Gradient-Based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB). With GOSS, LightGBM discards the majority of data instances with small gradients and retains the remaining ones to construct a representative data subset. Meanwhile, EFB reduces the feature space by combining mutually exclusive features into the same vector, with offset indices applied to separate the feature values during decoding.

Table 4. Features Table

No.	Features	Feature Description
1	n_estimators	200
2	learning_rate	0,05
3	num_leaves	31
4	max_depth	10
5	reg_alpha	0,1
6	reg_lambda	0,5

Furthermore, grid search was employed to tune several key hyperparameters, including *n_estimators* (the number of boosting rounds), *learning_rate* (the step size for tree combination), and *reg_alpha* (L1 regularization to prevent overfitting by encouraging sparsity). In this study, the *num_leaves* parameter for a single decision tree was set to 20 and 30. The complete parameter configuration used in this research is presented in Table 4.

2.7. Model Evaluation

The performance of the model was evaluated using Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Error (MAE). RMSE measures the square root of the average squared difference between the predicted values ($\hat{y}_i$) and the actual values (y_i). MAPE evaluates relative error in percentage terms, which facilitates comparison across datasets with different scales. The MAPE value is obtained by calculating the mean of the absolute percentage errors, where M denotes the number of data points, y_t the actual values, and $\hat{y}_t$ the predicted values [26]. MAE measures the average magnitude of absolute differences between actual and predicted values [27]. The formulations of RMSE, MAPE, and MAE used in this study are presented in Equations (4), (5), and (6).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (yi - \hat{y}i)^2} \quad (5)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{yi - \hat{y}i}{yi} \right| \quad (6)$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |yi - \hat{y}i| \quad (7)$$

3. RESULTS

The dataset used in this study covers the period of January-July 2025 from data of UMKM *Kedai Pesisir*, which were collected through observation, interviews, data gathering, and Focus Group Discussions (FGD). The dataset was subsequently preprocessed and subjected to feature engineering, including Month, Year, Week_of_Year, Day_of_Week, Week_of_Month, Season, Sales_Lag1, Sales_Lag2, Sales_Lag3, Rolling_Mean3, Rolling_Mean6, Rolling_Std3, Month_Sin, and Month_Cos. The data were then split into training and testing subsets with an 80:20 ratio. The LightGBM algorithm was employed with optimized hyperparameters (*n_estimators* = 200, *learning_rate* = 0.05, *num_leaves* = 31, *max_depth* = 10, *reg_alpha* = 0.1, *reg_lambda* = 0.5), yielding significant performance results.

The predictive analysis and stock recommendations were generated using validation data that were not included during training, thereby ensuring an unbiased evaluation of the model's forecasting capability. The prediction results and stock recommendations, aimed at anticipating demand volatility for the next six months (August 2025 – January 2026), are presented in Table 5.

Table 5 presents the prediction and recommendation results generated by the LightGBM-based recommendation system in this study. The evaluation period covers August 2025 to January 2026, focusing on demand forecasting for UMKM abon products. The 'Forecast' column represents the estimated demand, while 'Safety Stock' indicates the buffer stock recommended to mitigate uncertainties. The 'Suggested Stock' refers to the total inventory required to anticipate sudden increases in demand, and the 'Stock Priority' column classifies products into different priority levels based on predicted demand and risk of shortage.

Vol. 6, No. 5, October 2025, Page. 3832-3843
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5292

Table 5. Forecasted Demand and Stock Recommendations

No.	Evaluation Period	Product/Item	Forecast	Safety Stock	Suggested Stock	Stock Priority
1	2025-08	Chiken Floss (Spicy, 200 gram)	3	3	6	Medium Priority
2	2025-08	Beef Floss (Spicy, 100 gram)	6	7	13	High Priority
3	2025-08	Fish Floss (Spicy, 500 gram)	2	2	4	Low Priority
4	2025-08	Beef Floss (Spicy, 200 gram)	2	2	4	Low Priority
5	2025-08	Fish Floss (Sweet, 100 gram)	9	13	22	High Priority
6	2025-08	Fish Floss (Spicy, 100 gram)	3	4	7	Medium Priority
...	...	...	...	...	...	...
176	2026-01	Chiken Floss (Spicy, 200 gram)	3	3	6	Medium Priority
177	2026-01	Fish Floss (Spicy, 100 gram)	6	9	15	High Priority
178	2026-01	Beef Floss (Spicy, 500 gram)	1	2	3	Low Priority
179	2026-01	Chiken Floss (Sweet, 200 gram)	3	4	7	Low Priority
180	2026-01	Chiken Floss (Sweet, 100 gram)	8	10	18	High Priority
181	2026-01	Chiken Floss (Sweet, 200 gram)	3	4	7	Medium Priority

Furthermore, the recommended priority stock items for the next six months, as produced by the system, are summarized in Table 6 and illustrated in Figure 2.

Table 6. Forecasted Demand and Recommendations Priority Month Stock

No.	Evaluation Period	Product/Item	Variant	Package Weight	Recommended Stock Quantity
1	2025-08	Fish Floss	Sweet	100 gram	22 Units
2	2025-08	Chiken Floss	Sweet	100 gram	14 Units
3	2025-09	Fish Floss	Sweet	100 gram	15 Units
4	2025-09	Chiken Floss	Sweet	100 gram	11 Units
5	2025-10	Fish Floss	Sweet	100 gram	16 Units
6	2025-10	Chiken Floss	Sweet	100 gram	11 Units
7	2025-11	Fish Floss	Sweet	100 gram	20 Units
8	2025-11	Chiken Floss	Sweet	100 gram	14 Units
9	2025-12	Fish Floss	Sweet	100 gram	15 Units
10	2025-12	Chiken Floss	Sweet	100 gram	11 Units
11	2026-01	Fish Floss	Sweet	100 gram	16 Units
12	2026-01	Chiken Floss	Sweet	100 gram	13 Units

Subsequently, the performance of the prediction results generated by the recommendation system in this study is presented in Figure 3.

Vol. 6, No. 5, October 2025, Page. 3832-3843
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5292

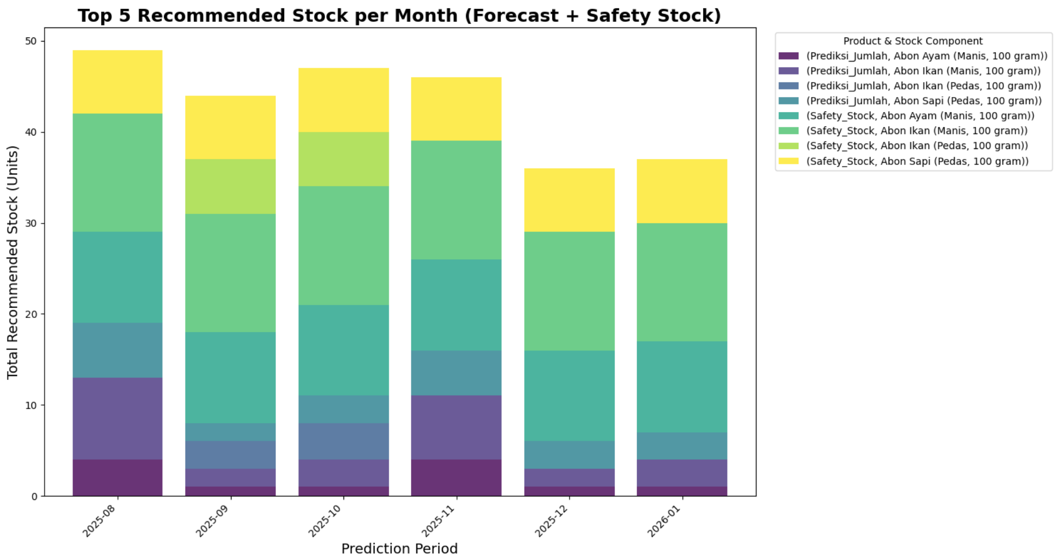

Figure 2. Priority Stock Item Recommendations for the Next Six Months

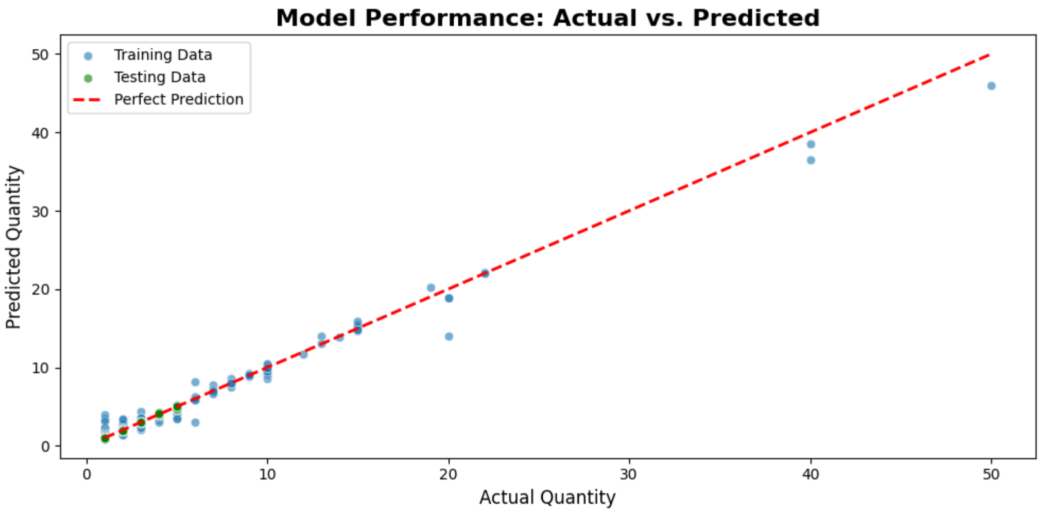

Figure 3. Model Prediction Results

Overall, as shown in Figure 4, the distribution of training and testing data indicates that the model does not suffer from overfitting, thereby ensuring accurate prediction performance. The alignment between predicted values, recommendations, and actual values further demonstrates that the model possesses strong predictive and recommendation capabilities, highlighting its applicability in forecasting sales and providing stock item recommendations to anticipate demand volatility. The model's performance metrics are presented in Table 7.

The model performance in Table 5 indicates that the Root Mean Squared Error (RMSE) of 0.069 reflects the average prediction error relative to the target output, suggesting that the model's predictions are consistently very close to the actual values. The Mean Absolute Error (MAE) of 0.034 demonstrates that, on average, each prediction deviates by only 0.034 units from the actual value, confirming the model's high accuracy. Furthermore, the Mean Absolute Percentage Error (MAPE) of 1.14% shows that the model's predictions deviate by just 1.14% on average from the actual values, which indicates an exceptionally high predictive accuracy, as it remains well below the 2% threshold.

Vol. 6, No. 5, October 2025, Page. 3832-3843
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5292

Table 7. Model Performance

No.	Metric	Value
1	Root Mean Squared Error (RMSE)	0.069
2	Mean Absolute Error (MAE)	0.034
3	Mean Absolute Percentage Error (MAPE)	1.14%

4. DISCUSSIONS

This study aimed to develop a recommendation system capable of identifying patterns in sales data and enhancing the accuracy of both predictions and recommendations based on customer–item interactions using the LightGBM algorithm in the context of UMKM Kedai Pesisir. The model was trained with optimized parameter configurations and an 80:20 train–test split, demonstrating that the model did not suffer from overfitting, thereby indicating reliable predictive performance. Although the predicted and actual values generally aligned well, minor discrepancies were observed at certain data points, particularly when deviations exceeded a value of 20. Nevertheless, the results confirm that the model performs effectively in predicting and recommending items for UMKM Kedai Pesisir. The system successfully generated daily stock predictions and recommendations for the next six months (August 2025–January 2026), including estimated demand, safety stock levels as a buffer, suggested stock levels for demand anticipation, and prioritized stock recommendations.

This study demonstrates that the LightGBM model, optimized through grid search and feature engineering, effectively captures sales patterns and provides accurate predictions and item recommendations to anticipate demand volatility in *UMKM Kedai Pesisir*. With evaluation metrics of RMSE at 0.069, MAE at 0.034, and MAPE at 1.14%, the model proved its ability to consistently learn complex patterns from the sales data within the period of January 2025–July 2025. These findings are consistent with previous research by Zhao et al. [28], who highlighted the strong performance of LightGBM in addressing data sparsity issues, achieving higher accuracy with reduced errors, a higher AUC, and a lower Logloss. Similarly, Sugiura et al. [29] noted that the limitation of insufficient external data could be mitigated through the integration of material price APIs. They also suggested future research directions, such as comparative analyses with algorithms like XGBoost. In their study, LightGBM, enhanced with optimized parameters and additional features, successfully tackled the challenge of predicting the most frequently selected news articles by users, achieving superior performance with an AUC of 0.8169.

Nevertheless, this study also acknowledges certain limitations. The analysis was restricted to internal sales data of UMKM Kedai Pesisir, specifically focusing on abon sales. External factors such as promotional activities or price fluctuations driven by raw material costs were not incorporated into the current dataset, limiting the model's ability to account for such variables. Future research could integrate these external factors into the recommendation framework, either by extending the current LightGBM-based approach or by exploring alternative models, to further enhance predictive performance and practical applicability.

5. CONCLUSION

This study presents the application of the LightGBM model in a recommendation system, demonstrating its applicability in sales prediction and item recommendation to anticipate demand volatility at UMKM Kedai Pesisir. The system is capable of generating daily stock predictions and recommendations for the next six months, including estimated demand, safety stock as a buffer,

Vol. 6, No. 5, October 2025, Page. 3832-3843
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5292

suggested stock levels to anticipate fluctuations, and prioritized product stock recommendations for the period of August 2025 to January 2026.

The evaluation results indicate a highly accurate performance. The Root Mean Squared Error (RMSE) of 0.069 suggests that the average deviation of the model's predictions from the actual values is minimal, reflecting consistent accuracy. The Mean Absolute Error (MAE) of 0.034 further confirms that, on average, each prediction deviates by only 0.034 units from the actual values, highlighting the model's strong predictive capability. Moreover, the Mean Absolute Percentage Error (MAPE) of 1.14% shows that the predictions deviate by only 1.14% from the actual values on average, underscoring the model's high accuracy, as it remains well below the 2% threshold. These findings confirm that the combination of data partitioning, parameter tuning, and feature selection effectively enables the model to capture sales patterns, addressing challenges highlighted in previous studies. However, this study also acknowledges certain limitations, particularly the reliance solely on internal sales data without incorporating external factors such as discounts, promotional activities, or raw material price fluctuations. The contribution to the field of informatics lies in the application of the LightGBM model for MSME recommendation systems, which enhances digital economic efficiency and business sustainability. Future research could integrate external data sources using a hybrid LightGBM–XGBoost approach to achieve higher predictive accuracy. Moreover, integrating these factors into the recommendation system—either by extending the current LightGBM framework or by adopting alternative models—could further improve predictive accuracy and strengthen the model's practical relevance in real-world business applications.

CONFLICT OF INTEREST

The authors declares that there is no conflict of interest between the authors or with research object in this paper.

ACKNOWLEDGEMENT

Acknowledgement is only addressed to funders or donors and object of research. Acknowledgement can also be expressed to those who helped carry out the research.

The authors would like to thank the 2025 DPPM (Grant No: 013/IT13.D/KH-DPPM/2024) from the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia for the financial support. The authors would like to thank Institut TeknologiBacharuddin Jusuf Habibie (ITH) and LPPM-PM ITH (Institute for Research, Community Service, and Quality Assurance) for their uidance and support throughout the research process.

REFERENCES

[1] M. Purwaningsih, B. Purwandari, M. Mishbah, and P. O. H. Putra, "Collaborative Commerce Model for SMEs A Knowledge and Resources Sharing Mechanism for Co Creation," *DESIDOC J. Libr. Inf. Technol.*, vol. 42, no. 2, pp. 80–87, Feb. 2022, doi: 10.14429/djlit.42.2.17294.

[2] C. Xiaochen, "Research on the Measurement of China's High-quality Economic Development Level," in *2020 2nd International Conference on Economic Management and Model Engineering (ICEMME)*, Chongqing, China: IEEE, Nov. 2020, pp. 779–782. doi: 10.1109/ICEMME51517.2020.00159.

[3] X. Xiao, "Research on the integration of digital economy and real economy to promote high-quality economic development," in *2020 Management Science Informatization and Economic Innovation Development Conference (MSIEID)*, Guangzhou, China: IEEE, Dec. 2020, pp. 8–11. doi: 10.1109/MSIEID52046.2020.00009.

[4] D. Rifai, S. Fitri, and I. N. Ramadhan, "Perkembangan Ekonomi Digital Mengenai Perilaku Pengguna Media Sosial Dalam Melakukan Transaksi," *ADI Bisnis Digit. Interdisiplin J.*, vol. 3, no. 1, pp. 49–52, Jun. 2022, doi: 10.34306/abdi.v3i1.752.

[5] I. Indrabayu, A. A. Ilham, I. S. Areni, A. Bustamin, Muh. A. F. I. Oemar, and S. W. Sidehabi, "Sosialisasi Rekondisi UMKM di Kelurahan Borongloe Kabupaten Gowa dengan Teknologi Informasi dan BMC di Era Pandemi Covid 19," *SIPISSANGNGI J. Pengabdi. Kpd. Masy.*, vol. 1, no. 3, p. 10, Sep. 2021, doi: 10.35329/sipissangngi.v1i3.2252.
[6] E. Haryo, "Dorong UMKM Naik Kelas dan Go Export, Pemerintah Siapkan Ekosistem Pembiayaan yang Terintegrasi," p. 2, Agustus 2023.
[7] R. A. Rifani, "Implementasi Layanan Food Delivery bagi Penjualan UMKM Makanan dan Minuman di Kota Parepare," *J. Manag.*, vol. 5, no. 2, Sep. 2022.
[8] S. Maratuttahirah and A. K. Jaya, "Klasifikasi Omset ATK menggunakan Algoritma Naïve Bayes," vol. 5, no. 1, 2024.
[9] L. Chao, "Hypergraph Contrastive Collaborative Filtering," *SIGIR 22 Proc. 45th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr.*, vol. 64, pp. 70–79, Jul. 2022, doi: https://doi.org/10.1145/3477495.3532058.
[10] Y. Afoudi, M. Lazaar, and M. Al Achhab, "Hybrid recommendation system combined content-based filtering and collaborative prediction using artificial neural network," *Simul. Model. Pract. Theory*, vol. 113, p. 102375, Dec. 2021, doi: 10.1016/j.simpat.2021.102375.
[11] G. Varoquaux and V. Cheplygina, "Machine learning for medical imaging: methodological failures and recommendations for the future," *Npj Digit. Med.*, vol. 5, no. 1, p. 48, Apr. 2022, doi: 10.1038/s41746-022-00592-y.
[12] D. Roy and M. Dutta, "A systematic review and research perspective on recommender systems," *J. Big Data*, vol. 9, no. 1, p. 59, Dec. 2022, doi: 10.1186/s40537-022-00592-5.
[13] J. Yu *et al.*, "Collaborative Filtering Recommendation with Fluctuations of User' Preference," in *2021 IEEE International Conference on Information Communication and Software Engineering (ICICSE)*, Chengdu, China: IEEE, Mar. 2021, pp. 222–226. doi: 10.1109/ICICSE52190.2021.9404120.
[14] E. Zangerle and C. Bauer, "Evaluating Recommender Systems: Survey and Framework," *ACM Comput. Surv.*, vol. 55, no. 8, pp. 1–38, Aug. 2023, doi: 10.1145/3556536.
[15] X. Yuan, H. Chen, Y. Song, X. Zhao, and Z. Ding, "Improving Sequential Recommendation Consistency with Self-Supervised Imitation," in *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, Montreal, Canada: International Joint Conferences on Artificial Intelligence Organization, Aug. 2021, pp. 3321–3327. doi: 10.24963/ijcai.2021/457.
[16] X. Xin, A. Karatzoglou, I. Arapakis, and J. M. Jose, "Self-Supervised Reinforcement Learning for Recommender Systems," Jun. 11, 2020, *arXiv*: arXiv:2006.05779. doi: 10.48550/arXiv.2006.05779.
[17] C. Lokker *et al.*, "Boosting efficiency in a clinical literature surveillance system with LightGBM," *PLOS Digit. Health*, vol. 3, no. 9, p. e0000299, Sep. 2024, doi: 10.1371/journal.pdig.0000299.
[18] Y. Wu and Q. Wang, "LightGBM Based Optiver Realized Volatility Prediction," in *2021 IEEE International Conference on Computer Science, Artificial Intelligence and Electronic Engineering (CSAIEE)*, SC, USA: IEEE, Aug. 2021, pp. 227–230. doi: 10.1109/CSAIEE54046.2021.9543438.
[19] G. V. D. Kumar, V. Deepa, N. Vineela, and G. Emmanuel, "Detection of Parkinson's disease using LightGBM Classifier," in *2022 6th International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India: IEEE, Mar. 2022, pp. 1292–1297. doi: 10.1109/ICCMC53470.2022.9753909.
[20] X. Li, "Design of a Web Content Personalized Recommendation System Based on Collaborative Filtering Improved by Combining k-means and LightGBM," *J. Web Eng.*, Apr. 2025, doi: 10.13052/jwe1540-9589.2425.
[21] L. Chen, L. Wu, R. Hong, K. Zhang, and M. Wang, "Revisiting Graph Based Collaborative Filtering: A Linear Residual Graph Convolutional Network Approach," *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 01, pp. 27–34, Apr. 2020, doi: 10.1609/aaai.v34i01.5330.
[22] A. Latif and S. Khotimatul Wildah, "ANALISIS KINERJA ALGORITMA ENSEMBLE DALAM PREDIKSI PERILAKU PEMBELIAN PELANGGAN," *JATI J. Mhs. Tek. Inform.*, vol. 9, no. 1, pp. 557–563, Dec. 2024, doi: 10.36040/jati.v9i1.12413.

Vol. 6, No. 5, October 2025, Page. 3832-3843
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.5292

[23] S. A. S. Purwandhani, A. A. N. Sajiatmoko, and C. S. K. Aditya, "Optimizing Indonesian Banking Stock Predictions with DBSCAN and LSTM," *J. Tek. Inform. Jutif*, vol. 6, no. 3, pp. 1173–1188, Jun. 2025, doi: 10.52436/1.jutif.2025.6.3.4439.
[24] X. Zhang, "Ion channel prediction Using Lightgbm Model," in *2020 International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, Fuzhou, China: IEEE, Jun. 2020, pp. 349–352. doi: 10.1109/ICBAIE49996.2020.00079.
[25] X. Li, "Design of a Web Content Personalized Recommendation System Based on Collaborative Filtering Improved by Combining k-means and LightGBM," *J. Web Eng.*, Apr. 2025, doi: 10.13052/jwe1540-9589.2425.
[26] R. Wahyusari, S. Sunardi, and A. Fadlil, "Comparative Analysis Of Ant Lion Optimization And Jaya Algorithm For Feature Selection In K-Nearest Neighbor (Knn) Based Electricity Consumption Prediction," *J. Tek. Inform. Jutif*, vol. 6, no. 3, pp. 1373–1388, Jun. 2025, doi: 10.52436/1.jutif.2025.6.3.4692.
[27] Y. D. Rosita and L. S. Moonlight, "Perbandingan Metode Prediksi untuk Nilai Jual USD: Holt-Winters, Holt's, dan Single Exponential Smoothing," *JTIM J. Teknol. Inf. Dan Multimed.*, vol. 5, no. 4, pp. 322–333, Jan. 2024, doi: 10.35746/jtim.v5i4.473.
[28] B. Zhao, B. Li, J. Zhang, W. Cao, and Y. Gao, "DCLGM: Fusion Recommendation Model Based on LightGBM and Deep Learning," *Neural Process. Lett.*, vol. 56, no. 1, p. 17, Feb. 2024, doi: 10.1007/s11063-024-11504-4.
[29] T. Sugiura, Y. Yamagishi, and Y. Kishimoto, "Leveraging LightGBM Ranker for Efficient Large-Scale News Recommendation Systems," in *Proceedings of the Recommender Systems Challenge 2024*, Bari Italy: ACM, Oct. 2024, pp. 27–31. doi: 10.1145/3687151.3687156.