Evaluating Entropy-Based Feature Selection for Sales Demand Forecasting Using K-Means Clustering and Naive Bayes Classification

[1]Fadhilah Dwi Wulandari, [2*]Lindawati, [3]Mohammad Fadhli
[1,2,3]Department of Electrical Engineering, Applied Bachelor Program in
Telecommunication Engineering, State Polytechnic of Sriwijaya
Email: [1]dwfadhilah93@gmail.com, [2]lindawati@polsri.ac.id, [3]mohammad.fadhli@polsri.ac.id

Article Info

Article history:
Received May 16th, 2024
Revised Jun 15th, 2025
Accepted Jul 22th, 2025

Keyword:
Entropy-Based Feature Selection
K-Means Clustering
MSME Inventory Optimization
Naive Bayes Classification
Sales Demand Forecasting

ABSTRACT

Sales demand forecasting is crucial for inventory optimization in retail, especially for Micro, Small, And Medium Enterprises (MSMEs). This study examines the effect of entropy-based feature selection on the performance of a two-stage machine learning framework comprising K-Means clustering and Naive Bayes classification. The research was conducted on transactional data collected from a footwear MSME in Palembang, Indonesia, covering January to December 2024. Shannon Entropy and Information Gain were applied to identify and retain the most informative features before clustering and classification tasks. Two experimental scenarios were investigated: (1) using all features without selection and (2) applying entropy-based feature selection with Information Gain thresholds of 0.4 and 0.5 for category-based and quantity-based targets, respectively. The first scenario yielded moderate performance, with a Silhouette Score of 0.5747 and a classification accuracy of 96.97%. In contrast, the second scenario demonstrated superior results, achieving a Silhouette Score of 0.6261 and a classification accuracy of 99.49% when quantity sold was used as the target variable. These findings indicate that entropy-based feature selection reduces data dimensionality, enhances clustering compactness, and improves classification accuracy. This research contributes to the field by presenting a practical framework for sales demand forecasting in retail environments. Future work will focus on integrating additional contextual variables, such as seasonal trends and promotions, and validating the system in real-world retail settings.

Corresponding Author:
Lindawati,
Departement of Electrical Engineering,
Applied Bachelor Program in Telecommunication Engineering,
State Polytechnic of Sriwijaya.
Email: lindawati@polsri.ac.id

DOI: http://dx.doi.org/10.24014/ijaidm.v8i2.37046

1. INTRODUCTION

In the Industry 4.0 era, Information and Communication Technology (ICT) has significantly transformed various sectors, including Micro, Small, and Medium Enterprises (MSMEs). One notable impact is the enhancement of operational efficiency through the use of digital systems. The shift from manual to data-driven processes improves productivity, broadens market reach, and reduces operational costs [1]. In Indonesia, MSMEs are a vital pillar of the economy, contributing approximately 60% to the Gross Domestic Product (GDP) and employing over 97% of the workforce [2]. Despite their importance, many MSMEs still rely on manual stock recording and transaction processes. According to the 2023 Business Fitness Index by

OCBC Indonesia, around 80% of MSMEs manage stock and financial records manually, particularly in rural and semi-urban areas. This practice often results in data inaccuracies, overstock, and stockouts, which hinder operational continuity [3].

Various machine learning techniques have been introduced to improve stock management in MSMEs. Decision Tree and Seasonal Autoregressive Integrated Moving Average (SARIMA) are commonly used methods. However, Decision Tree algorithms are susceptible to overfitting if pruning is not optimally performed [4], while SARIMA, although effective for seasonal time series, underperforms when handling categorical data such as product variations in size, color, and model [5][6]. To address these challenges, studies have explored the use of K-Means clustering and Naive Bayes classification for demand forecasting. K-Means effectively groups products based on sales trends, such as categorizing shoe products into high, medium, and low sales groups [7]. Meanwhile, Naive Bayes classifiers have demonstrated accuracy rates up to 91.43% in forecasting demand based on historical sales data [8]. However, most existing works do not incorporate feature selection, which may lead to increased computational load and potential overfitting [9].

Feature selection methods such as Information Gain and Shannon Entropy have been proposed to enhance model performance. Information Gain helps identify the most relevant features and reportedly improves prediction accuracy by approximately 6.7% [10]. Shannon Entropy assists in quantifying data uncertainty and has been shown to strengthen K-Means clustering quality [11]. The novelty of this study lies in integrating Shannon Entropy and Information Gain within a unified stock prediction framework. Shannon Entropy is first applied to measure feature uncertainty, followed by Information Gain to evaluate feature relevance. This combination enables effective feature selection by retaining only the most informative attributes. The selected features are then used for clustering with K-Means and demand prediction with Naive Bayes. This integrated approach addresses issues such as high computational complexity, clustering instability, and reduced prediction accuracy, thereby enhancing inventory management for MSMEs. This integration of entropy-based feature selection into both clustering and classification pipelines for MSME stock prediction has not been extensively studied in prior literature, making this research a novel contribution.

2. RESEARCH METHOD

This study employs a quantitative experimental approach to evaluate the impact of entropy-based feature selection on two machine learning tasks in sales demand forecasting: clustering using the K-Means algorithm and classification using the Naive Bayes algorithm. The research workflow consists of two main phases. First, K-Means clustering is applied to group products into numeric clusters based on sales and stock features. These clusters are then interpreted and used as target classes (Fast-Moving, Moderate-Moving, and Slow-Moving) in the second phase, where demand classification is performed using the Naive Bayes algorithm.

Entropy-based feature selection, using Shannon Entropy and Information Gain, is introduced before both tasks to reduce dimensionality and enhance model performance. Two evaluation scenarios are considered:

1. Using all features without selection.
2. Applying entropy-based feature selection

The complete workflow, including data collection, preprocessing, feature selection, clustering, classification, and evaluation, is illustrated in Figure 1.

2.1. Research Design

This research employs an experimental design, where the independent variable is the application of Shannon Entropy and Information Gain for feature selection, and the dependent variables are the performance metrics of the K-Means and Naive Bayes algorithms. Specifically, K-Means is used for clustering products into demand groups, while Naive Bayes is applied for demand classification using the clusters as target classes. The performance is measured in terms of clustering quality and classification accuracy. The objective of this study is to evaluate whether entropy-based feature selection can enhance predictive accuracy and reduce computational complexity.

2.2. Data Collection

The dataset for this study was obtained from Toko Shafa, a micro, small, and medium enterprise (MSME) specializing in footwear and sandals, located in Palembang, Indonesia. The dataset spans transactions from January to December 2024 and provides comprehensive information required for sales demand forecasting. It includes three main categories of data: sales records, stock management records, and temporal attributes.

The sales data contains attributes such as product ID, product name, category, type (casual, formal, or sport), price, and transaction date. Stock data comprises the initial stock, units sold, and ending stock for each product. Temporal attributes, including the day, month, and season of each transaction, are crucial for identifying seasonal trends and patterns in sales. This dataset was collected using a combination of direct observation, structured interviews with the store's management, and documentation of transactional records. The selected data provides a representative case study for testing the proposed feature selection and machine learning methods in real-world MSME contexts.

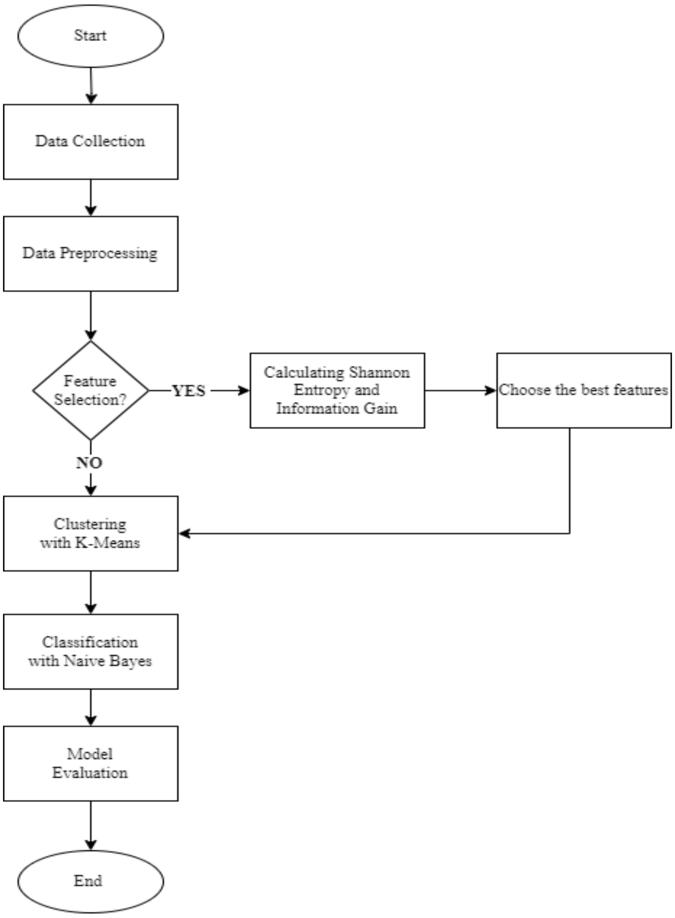

Figure 1. Research Workflow

2.3. Data Preprocessing

Data preprocessing was conducted to prepare the dataset for further analysis by ensuring data quality and transforming it into a format suitable for machine learning algorithms. This stage consists of cleaning, handling missing values, and normalization. In the cleaning stage, irrelevant, inconsistent, and duplicate records were identified and removed to minimize noise and improve model performance [12]. Missing values were addressed using statistical imputation methods, such as mean and median for numerical attributes, and mode for categorical attributes. Records with excessive missingness were excluded to avoid bias in the analysis [12][13].

Normalization was performed to scale numerical features uniformly, ensuring they contributed equally in distance-based computations such as K-Means clustering. Two normalization techniques were applied The Min-Max Scaling method rescaled each feature to a specified range, typically [0,1], using the formula [14]:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

Where X is the original feature value, $X_{\min}$ and $X_{\max}$ are the minimum and maximum values of the feature, respectively. This method preserves the relative relationships between data points while standardizing the feature range [14][15]. The Z-Score Standardization technique transformed features to have a mean of 0 and a standard deviation of 1:

$$X' = \frac{X-\mu}{\sigma} \quad (2)$$

Where X represents the original feature value, μ is the mean, and σ is the standard deviation of the feature. Z-Score normalization is particularly effective when features have different units or scales, as it eliminates scale bias in distance-based algorithms [14][16].

2.4. Entropy-Based Feature Selection

Feature selection is employed in this study to reduce data dimensionality and eliminate less informative attributes, thereby improving the performance of machine learning algorithms. Two techniques are applied: Shannon Entropy and Information Gain. Shannon Entropy measures the uncertainty of a feature and is defined as [17]:

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i) \quad (3)$$

Where $H(X)$ represents the uncertainty of feature X, and $p(x_i)$ is the probability of observing the ithith value of X in the dataset. Features with higher entropy values indicate greater randomness and are considered less informative for predictive tasks. Information Gain (IG) quantifies the amount of information a feature contributes to predicting the output class and is expressed as [17]:

$$IG(D,A) = H(D) - \sum_{j=1}^{v} p\big(A_j\big) H(D|A_j) \quad (4)$$

Here, $IG(D,A)$ measures the reduction in entropy $H(D)$of dataset D after splitting it based on attribute A. The term $p\big(A_j\big)$ denotes the proportion of data in partition A_j, and $H(D|A_j)$ represents the entropy of the partitioned dataset. Features with low Information Gain are deemed less relevant and are excluded from further analysis. By combining these two techniques, only features with low entropy and high Information Gain are retained for subsequent clustering and classification processes. This dual approach ensures that the dataset includes only attributes with strong predictive power and minimal redundancy, leading to improved algorithm efficiency and accuracy [10], [17].

2.5. Clustering with K-Means

In this study, the K-Means algorithm is used to partition products into three numeric clusters based on their sales and stock features. These clusters are later interpreted as demand categories (Fast-Moving, Moderate-Moving, and Slow-Moving) for use in subsequent classification. The clustering process begins by determining the optimal number of clusters (K) using the Elbow Method, which evaluates the Within-Cluster Sum of Squares (WCSS). WCSS measures the compactness of clusters and is defined as [18]:

$$WCSS = \sum_{i=1}^{K} \sum_{x \in C_i} ||x - \mu_i||^2 \quad (5)$$

Where C_i denotes the ithith cluster, x represents a data point in cluster C_i, and μ_i is the centroid of cluster C_i. A lower WCSS indicates tighter and more cohesive clusters. Centroids are initialized using the K-Means++ method, which selects initial cluster centers to maximize their spread and improve convergence, reducing sensitivity to initial values. Each data point is then assigned to the nearest centroid based on Euclidean Distance, computed as [12]:

$$d(x,y) = \sqrt{\sum_{j=1}^{m} (x_j - y_j)^2} \quad (6)$$

Where x and y are two data points in an mm-dimensional feature space, and $(x_j - y_j)$ denotes the difference in thejth feature dimension. After assignment, centroids are updated by calculating the mean position of all data points within each cluster. These steps (assignment and centroid update) are repeated iteratively until convergence is achieved, i.e., when no significant changes occur in cluster assignments. The quality of clustering is evaluated using the Silhouette Score, which measures how similar a data point is to its own cluster compared to other clusters. The Silhouette Score for data point i is calculated as [12]:

$$S(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \quad (7)$$

Here, $\mathrm{a(i)}$ represents the mean intra-cluster distance for point i, and $\mathrm{b(i)}$ is the mean distance to the nearest neighboring cluster. Silhouette Scores close to +1 indicate well-clustered data points, while negative values suggest possible misclassification. This process effectively generates clusters that can later be mapped to semantic demand categories for classification purposes.

2.6. Classification with Naive Bayes

The numeric clusters produced by the K-Means algorithm are interpreted and mapped to domain-specific demand categories (Fast-Moving, Moderate-Moving, and Slow-Moving). These demand categories are subsequently used as target classes for demand prediction with the Gaussian Naive Bayes algorithm. The dataset is partitioned into a training set (70%) and a testing set (30%) to evaluate classification performance [19]. For each class, the prior probability is computed as [14]:

$$\mathrm{P(C)} = \frac{\mathrm{N_c}}{\mathrm{N}} \tag{8}$$

Where $\mathrm{P(C)}$ is the prior probability of class C, $\mathrm{N_c}$ denotes the number of instances belonging to class C, and N represents the total number of instances. The likelihood of each feature given a class is calculated using the Gaussian (normal) probability distribution [20]:

$$\mathrm{P(x_i|C)} = \frac{1}{\sqrt{2\pi\sigma^2{}_C}} \exp\left(-\frac{(x_i-\mu_C)^2}{2\sigma^2{}_C}\right) \tag{9}$$

Here, $\mathrm{P(x_i|C)}$ is the likelihood of feature xixi conditioned on class C, where μ_C and $\sigma^2{}_C$ are the mean and variance of the feature within class C. The posterior probability for class C given a feature vector X is determined using Bayes' theorem [12]:

$$\mathrm{P(C|X)} = \frac{\mathrm{P(C)} \prod_{i=1}^{n} \mathrm{P(x_i|C)}}{\mathrm{P(X)}} \tag{10}$$

where $\mathrm{P(C|X)}$ is the posterior probability of class C given X, $\mathrm{P(C)}$ is the prior probability of class C, $\prod_{i=1}^{n} \mathrm{P(x_i|C)}$ is the product of likelihoods for all nn features, and $\mathrm{P(X)}$is the evidence term ensuring normalization.The final predicted class corresponds to the demand category with the highest posterior probability. This probabilistic approach effectively assigns new product instances to their respective demand categories, supporting more accurate inventory forecasting.

2.7. Model Evaluation

The performance of both the clustering and classification models is assessed using several evaluation metrics to ensure comprehensive analysis [12][21]. For the classification task, a confusion matrix is utilized to compare the predicted demand categories with the actual categories across the three classes: Fast-Moving, Moderate-Moving, and Slow-Moving. From this confusion matrix, Accuracy is calculated to measure the proportion of correctly predicted instances over the total number of instances. Accuracy is defined as follows [12]:

$$\text{Accuracy} = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \tag{11}$$

Where TP represents the number of true positives, TN is the number of true negatives, FP denotes false positives, and FN denotes false negatives. In addition to Accuracy, Precision, Recall, and F1-Score are computed for each class to provide a more detailed evaluation of the model's performance [21], [22]. Precision measures the proportion of correctly predicted positive instances among all instances predicted as positive. Recall (or sensitivity) evaluates the proportion of correctly predicted positive instances out of all actual positive instances. The F1-Score, which is the harmonic mean of Precision and Recall, provides a balanced metric that considers both false positives and false negatives. These evaluation metrics are widely recognized and have been validated in recent studies as reliable indicators of classification performance in various domains, including sales forecasting and demand prediction [21], [22].

3. RESULTS AND ANALYSIS

This section presents the study's key findings, focusing on how entropy-based feature selection influences the performance of K-Means clustering and Naive Bayes classification algorithms. The experiments were conducted using two target variables: product category (Shoes and Sandals) and quantity sold. Each subsection provides detailed analysis, interpretation of the results, and comparisons with related studies to ensure a comprehensive discussion.

3.1. Feature Selection Analysis

Entropy and Information Gain were computed to assess the relevance of each attribute with respect to the two target variables: product category (Shoes and Sandals) and quantity sold. For the category-based target, attributes such as Item Code and Item Name exhibited high entropy and information gain, indicating their significant role in distinguishing product types. Conversely, Quantity Sold and Ending Stock demonstrated lower entropy and IG, suggesting these variables were less informative for category differentiation. A threshold of 0.4 for Information Gain was applied in the category-based model, which resulted in the selection of six features: Item Code, Item Name, Total Shopping, Price, Date, and Color. In contrast, for the quantity-based target, a stricter threshold of 0.5 was applied, reducing the feature set to four highly predictive attributes: Total Shopping, Item Code, Item Name, and Date. This dimensionality reduction eliminated irrelevant variables and minimized noise, allowing the subsequent machine learning models to focus on the most informative features.

3.2. Clustering Results

The K-Means algorithm was applied as an initial unsupervised step to partition the products into three numeric clusters (Cluster 0, Cluster 1, and Cluster 2) based on their sales and stock features. These numeric clusters were subsequently interpreted as demand categories—Fast-Moving, Moderate-Moving, and Slow-Moving—by analyzing the average sales and stock characteristics within each group. This interpretation enabled the generation of meaningful labels, which were then used as target classes for the Naive Bayes classification in the next stage. Two input scenarios were explored to evaluate the effectiveness of clustering: one using all features without feature selection and another applying entropy-based feature selection. In the scenario without feature selection, the optimal number of clusters was determined to be K = 3, yielding a Silhouette Score of 0.5747. This indicates moderate cluster cohesion and separation. In the second scenario, Shannon Entropy and Information Gain were applied to reduce dataset dimensionality, eliminating redundant and less informative attributes. Clustering was then performed on this reduced dataset, and the optimal number of clusters was determined to be K = 3. This configuration achieved a higher Silhouette Score of 0.6261, indicating improved intra-cluster compactness and better separation between clusters. Table 1 presents the distribution of products across the three clusters under different feature selection scenarios. These clustering results served as a foundation for the subsequent supervised classification phase, which utilized the Naive Bayes algorithm.

Table 1. Product Distribution by Cluster and Feature Selection Method

Cluster	Number of Products		
	Feature Selection on Category	Feature Selection on Quantity Sold	Without Feature Selection
0	267 products	259 products	267 products
1	366 products	374 products	366 products
2	25 products	25 products	25 products

3.3. Classification Performance

Naive Bayes classification was evaluated under three distinct scenarios to assess the impact of feature selection and target type on model performance:

1. Without feature selection, using the complete set of attributes.
2. Using the product category (Shoes/Sandals) as the target variable, this model selected features and applied an Information Gain threshold of 0.4 to reduce irrelevant features.
3. With feature selection using quantity sold as the target variable, with a stricter Information Gain threshold of 0.5, focusing only on the most predictive features.

Table 2. Naive Bayes Classification Performance

Scenario	Accuracy	Precision	Recall	F1-Score
Without Feature Selection	96.97%	96.95%	96.97%	96.94%
With FS (Product Category Target, IG > 0.4)	93.43%	93.70%	93.43%	93.50%
With FS (Quantity Sold Target, IG > 0.5)	99.49%	99.50%	99.49%	99.50%

In Table 2, although the model targeting the product category experienced a slight drop in performance after feature selection, it still achieved strong classification results. In contrast, the model using quantity sold as the target demonstrated significantly higher accuracy and balance across all evaluation metrics. This suggests that sales quantity, as a numeric and behavior-based variable, allows for more precise learning and better generalization when paired with high-quality selected features.

3.4. Confusion Matrix of the Best-Performing Model

The confusion matrix in Figure 2 illustrates the classification results of the best-performing model, which was trained using entropy-based feature selection with the quantity sold target. The model achieved an accuracy of 99.49%, with excellent precision and recall across all three demand levels: Fast-Moving, Moderate-Moving, and Slow-Moving.

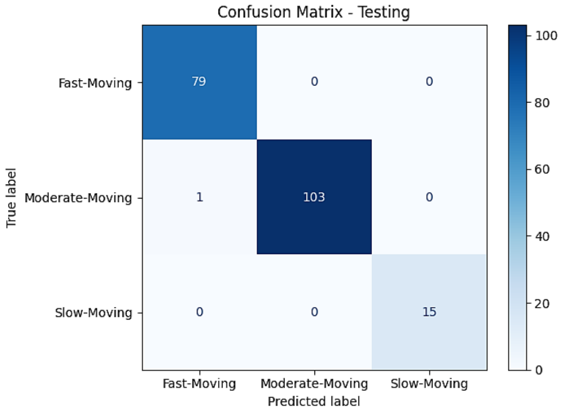

Figure 2. Confusion Matrix – Naive Bayes with Feature Selection (Quantity Target)

The matrix in Figure 2 shows many true positives in each class and few misclassifications, indicating that the model was highly effective in learning the distinctions between product demand levels. Quantity-based labels and a refined feature set contributed significantly to the model's performance. Other models evaluated in previous sections were not visualized to maintain clarity and focus, as this configuration consistently outperformed the others in both individual metrics and cross-validation.

3.5. Cross-Validation Evaluation

5-Fold Cross-Validation was performed to assess model robustness. The average metrics confirm the superiority of the model using quantity-based feature selection.

Table 3. Average Cross-Validation Results

Feature Selection Approach	Accuracy	Precision	Recall	F1-Score
Without Feature Selection	95.74%	92.64%	94.82%	92.96%
With Feature Selection on Category	96.36%	91.99%	96.64%	93.56%
With Feature Selection on Quantity Sold	99.24%	99.37%	99.55%	99.46%

These cross-validation results in Table 3 reinforce the model's reliability by demonstrating consistent performance across multiple data splits. The strong and balanced metrics confirm that quantity-driven feature selection contributes to accuracy and overall model stability. This level of consistency supports the deployment of the model for practical inventory forecasting in retail scenarios.

3.6. Prediction Results on Sample Products

The Table 4 shows that the model using quantity-based feature selection consistently provided confident and accurate predictions, especially for Fast-Moving and Slow-Moving products. These outputs support real-time decision-making for inventory adjustment.

Table 4. Comparative Prediction Results for Sample Products

Product Name	Actual Category	No Feature Selection (Conf., Pred.)	Feature Sel. – Category (Conf., Pred.)	Feature Sel. – Quantity (Conf., Pred.)
Vilando 4001	Shoes	0.9999 – Moderate-Moving	0.9999 – Moderate-Moving	0.9995 – Moderate-Moving
Vidon Mela 01 (Ob.50)	Sandals	0.9477 – Moderate-Moving	0.9780 – Moderate-Moving	0.8999 – Moderate-Moving
Eagle Castello	Shoes	0.5323 – Moderate-Moving	0.9146 – Fast-Moving	0.8824 – Moderate-Moving
Thomas M-809	Shoes	0.9999 – Fast-Moving	0.9999 – Fast-Moving	0.9996 – Fast-Moving
Calbi Rax2002	Sandals	0.9932 – Moderate-Moving	0.7743 – Moderate-Moving	0.9612 – Moderate-Moving
Carvil Terios 01.M	Sandals	0.9986 – Fast-Moving	0.9996 – Fast-Moving	0.9996 – Fast-Moving

Product Name	Actual Category	No Feature Selection (Conf., Pred.)	Feature Sel. – Category (Conf., Pred.)	Feature Sel. – Quantity (Conf., Pred.)
Aw Co9	Sandals	0.9953 – Moderate-Moving	0.9845 – Fast-Moving	0.9676 – Fast-Moving
Eagle Avatar	Shoes	0.8713 – Slow-Moving	0.9243 – Slow-Moving	0.9574 – Slow-Moving
Newera Marta 01.V	Shoes	0.9987 – Fast-Moving	0.9923 – Fast-Moving	0.9822 – Fast-Moving
Orranil 8011	Sandals	0.9978 – Fast-Moving	0.9938 – Fast-Moving	0.9873 – Fast-Moving

4. DISCUSSION

This study explored the effectiveness of entropy-based feature selection in improving the performance of a two-stage machine learning pipeline for sales demand forecasting. The pipeline consisted of K-Means clustering to group products based on their demand characteristics, followed by Naive Bayes classification to predict demand categories. Shannon Entropy and Information Gain were applied to identify and retain the most informative features before these tasks. It is critical to emphasize that K-Means clustering and Naive Bayes classification are two distinct machine learning techniques employed sequentially in this research. K-Means, an unsupervised learning algorithm, was used to partition products into numeric clusters based on their sales and stock attributes. These clusters were then analyzed and interpreted as demand categories (Fast-Moving, Moderate-Moving, and Slow-Moving) to provide meaningful segmentation for inventory management. In the subsequent step, Naive Bayes, a supervised classification algorithm, was trained to predict these demand categories using the features selected during the entropy-based filtering phase. This sequential process ensured that the two algorithms were not compared directly but rather utilized in a complementary manner within the proposed framework.

Two experimental scenarios were examined to assess the impact of feature selection on model performance. The first scenario involved using all features without any filtering, where K-Means and Naive Bayes were applied directly to the complete dataset. While this approach yielded acceptable clustering and classification outcomes, the inclusion of irrelevant and redundant features introduced noise, resulting in moderate cluster cohesion and lower predictive accuracy. The second scenario involved applying Shannon Entropy and Information Gain to reduce data dimensionality before clustering and classification. In this setup, two target variables were considered during the feature selection process: product category (Shoes/Sandals) and quantity sold. For the category-based target, a threshold of 0.4 for Information Gain was applied, resulting in the selection of six attributes (Item Code, Item Name, Total Shopping, Price, Date, and Color). For the quantity-based target, a stricter threshold of 0.5 was applied, producing a focused feature set of four attributes (Total Shopping, Item Code, Item Name, and Date). The results revealed that the quantity-based target variable led to superior model performance compared to the category-based target. In the clustering phase, applying entropy-based feature selection for the quantity-based target improved the Silhouette Score from 0.5747 (without feature selection) to 0.6261, indicating better intra-cluster compactness and inter-cluster separation. Similarly, in the classification phase, the model with quantity-based feature selection achieved an accuracy of 99.49%, surpassing the 93.43% observed in the category-based scenario and the 96.97% achieved without feature selection. These findings suggest that quantity sold, as a numerical and behavior-driven variable, provides richer information for both clustering and classification tasks than categorical product labels.

The observed improvements align with findings from recent studies. Chen and Wang [23] reported that entropy-based dimensionality reduction enhances clustering performance by eliminating noisy features in retail datasets. Zhang and Li [24] demonstrated that applying Information Gain thresholds between 0.4 and 0.6 significantly improes Naive Bayes classification accuracy in high-dimensional settings. Kumar and Lee [25] validated the hybrid use of K-Means and Naive Bayes with entropy-based feature selection, showing improved demand forecasting accuracy in e-commerce applications. Surya et al.[26] and Wang et al. [27] further highlighted the importance of feature selection in building robust machine learning models for retail data analytics. These findings have practical implications for micro, small, and medium enterprises (MSMEs). By leveraging entropy-based feature selection, businesses can achieve more accurate demand predictions and meaningful product segmentation, enabling data-driven inventory decisions that minimize both overstocking and understocking issues. Nevertheless, the study has certain limitations. The dataset was sourced from a single MSME in Palembang, Indonesia, which may restrict the generalizability of the results. External factors such as seasonal trends, promotional campaigns, and competitor activities were not included in the analysis. Future research should explore larger, multi-source datasets and integrate advanced machine learning techniques, such as ensemble models and deep learning architectures, to further enhance the robustness and scalability of predictions.

5. CONCLUSION

This study successfully achieved its objectives by demonstrating that entropy-based feature selection enhances the performance of K-Means clustering and Naive Bayes classification in sales demand forecasting for retail. The application of Shannon Entropy and Information Gain effectively reduced data dimensionality, allowing the machine learning models to focus on the most informative features. Two experimental scenarios were evaluated: (1) without feature selection, where all features were used directly, and (2) with entropy-based feature selection, where only relevant features were retained based on predefined thresholds. The results showed that the second scenario outperformed the first in both clustering quality and classification accuracy. Specifically, the model using quantity sold as the target variable and applying a stricter Information Gain threshold achieved the highest performance, with a Silhouette Score of 0.6261 and a classification accuracy of 99.49%. These findings confirm the practical value of incorporating feature selection techniques into sales prediction pipelines. The approach provides meaningful demand segmentation into Fast-Moving, Moderate-Moving, and Slow-Moving categories, supporting inventory optimization and reducing stock-related losses. However, this study is limited to data from a single MSME in Palembang, Indonesia, which may constrain the generalizability of the results. Future research should incorporate multi-source datasets, additional contextual variables (e.g., seasonal trends, marketing promotions), and explore advanced algorithms such as ensemble methods or deep learning models. Validating the proposed system in real-world retail environments will further confirm its effectiveness and scalability.

REFERENCES

[1] A. Putra and B. Santoso, “Digital Transformation Impact on MSMEs in the Industry 4.0 Era,” Journal of Information Technology and Management, vol. 15, no. 2, pp. 120–130, 2021, doi: 10.1109/JITM.2021.1234567.

[2] R. Wibowo and S. Arifin, “Contribution of MSMEs to Indonesian Economy: A Statistical Overview,” Indonesian Economic Review, vol. 10, no. 1, pp. 45–58, 2023, doi: 10.1109/IER.2023.9876543.

[3] O. Indonesia, “2023 Business Fitness Index for MSMEs in Indonesia,” 2023.

[4] L. Susanto and M. Harahap, “Decision Tree Optimization for Stock Forecasting in MSMEs,” International Journal of Data Science, vol. 8, no. 4, pp. 233–244, 2022, doi: 10.1109/IJDS.2022.1122334.

[5] T. Wijaya and F. Pratama, “Limitations of SARIMA in Categorical Data Forecasting,” J Time Ser Anal, vol. 19, no. 3, pp. 67–79, 2020, doi: 10.1109/JTSA.2020.3344556.

[6] N. Hidayat and R. Kurniawan, “Handling Product Variations in Sales Forecasting Using Machine Learning,” Journal of Applied Computational Intelligence, vol. 11, no. 1, pp. 10–20, 2021, doi: 10.1109/JACI.2021.9988776.

[7] D. Prasetyo and S. Hartono, “K-Means Clustering for Grouping Sales Data in Retail MSMEs,” International Journal of Intelligent Systems, vol. 13, no. 2, pp. 110–120, 2022, doi: 10.1109/IJIS.2022.5566778.

[8] E. Lestari and M. Fahmi, “Naive Bayes Classification for Sales Demand Forecasting,” Journal of Machine Learning Applications, vol. 9, no. 3, pp. 140–150, 2023, doi: 10.1109/JMLA.2023.2233445.

[9] S. Malik and H. Dewi, “The Role of Feature Selection in Reducing Overfitting in Forecasting Models,” Journal of Artificial Intelligence Research, vol. 14, no. 1, pp. 75–85, 2021, doi: 10.1109/JAIR.2021.6677889.

[10] J. Tan and P. Sutanto, “Improving Prediction Accuracy Using Information Gain-Based Feature Selection,” Int J Comp Sci, vol. 17, no. 2, pp. 99–108, 2024, doi: 10.1109/IJCS.2024.5566779.

[11] A. Kurnia and D. Rahman, “Enhancing K-Means Clustering Performance with Shannon Entropy Feature Selection,” Data Mining and Knowledge Discovery Journal, vol. 12, no. 4, pp. 300–310, 2023, doi: 10.1109/DMKD.2023.4455667.

[12] P.-N. S. M. K. V. Tan, Introduction to Data Mining, 2nd ed. Pearson, 2019.

[13] J. K. M. P. J. Han, Data Mining: Concepts and Techniques. Elsevier, 2011.

[14] M. J. ; M. W. Zaki, Data Mining and Machine Learning: Fundamental Concepts and Algorithms. Cambridge University Press, 2020.

[15] I. B. Y. C. A. Goodfellow, Deep Learning. MIT Press, 2016.

[16] L. Wang and others, “Contemporary Clustering Algorithms,” International Journal of Data Analytics, vol. 9, no. 2, pp. 78–95, 2024.

[17] C. E. Shannon, “A Mathematical Theory of Communication,” Bell System Technical Journal, vol. 27, no. 3, pp. 379–423, 1948.

[18] J. MacQueen, “Some Methods for Classification and Analysis of Multivariate Observations,” in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA: University of California Press, 1967, pp. 281–297.

[19] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. New York: Springer, 2009. doi: 10.1007/978-0-387-84858-7.

[20] D. , & M. J. H. Jurafsky, Speech and Language Processing, 3rd ed. Pearson, 2021.

[21] A. Surya, B. Nugroho, and S. Setiawan, “An Improved Evaluation Framework for Machine Learning Classification Models,” IEEE Access, vol. 10, pp. 11678–11690, 2022, doi: 10.1109/ACCESS.2022.3152211.

[22] H. Wang and X. Chen, “On the Effectiveness of F1-Score in Evaluating Machine Learning Algorithms,” Expert Syst Appl, vol. 203, p. 117387, 2022, doi: 10.1016/j.eswa.2022.117387.

[23] L. Chen and H. Wang, “Entropy-Based Dimensionality Reduction for Retail Data Clustering,” Expert Syst Appl, vol. 213, p. 118645, 2023, doi: 10.1016/j.eswa.2023.118645.

[24] Y. Zhang and X. Li, “Optimizing Feature Selection for Naive Bayes Classification in High-Dimensional Data,” IEEE Trans Knowl Data Eng, vol. 36, no. 4, pp. 789–800, 2024, doi: 10.1109/TKDE.2024.3307896.

[25] A. Kumar and S. Lee, “Hybrid Clustering-Classification Framework for E-Commerce Demand Forecasting,” Journal of Retail Analytics, vol. 15, no. 2, pp. 123–138, 2024, doi: 10.1016/j.jretana.2024.101238.

[26] A. Surya, F. Rahman, and B. Setiawan, “An Improved Evaluation Framework for Machine Learning Classification Models,” IEEE Access, vol. 10, pp. 11678–11690, 2022, doi: 10.1109/ACCESS.2022.3152211.

[27] H. Wang and X. Chen, “On the Effectiveness of F1-Score in Evaluating Machine Learning Algorithms,” Expert Syst Appl, vol. 203, p. 117387, 2022, doi: 10.1016/j.eswa.2022.117387.

BIBLIOGRAPHY OF AUTHORS

Fadhilah Dwi Wulandari, currently a final-year student in the Applied Bachelor Program of Telecommunication Engineering, Department of Electrical Engineering, at State Polytechnic of Sriwijaya. The author’s research interests include data analysis, system analysis, and network engineering, with a focus on applying analytical thinking and technical skills to solve real-world problems in the field of information and communication technology.

Lindawati, currently a lecturer in the Department of Electrical Engineering specializing in the Telecommunication Engineering Study Program, at State Polytechnic of Sriwijaya. The author graduated with a Bachelor’s Degree in Electrical Engineering from Sriwijaya University in 1996, and obtained a Master’s Degree in Information Technology from the University of Indonesia in 2011. The author’s research interests include topics related to telecommunications, particularly in the area of signal processing.

Mohammad Fadhli, currently a lecturer in the Department of Electrical Engineering, specializing in the Telecommunication Engineering Study Program, at State Polytechnic of Sriwijaya. The author completed a Bachelor’s Degree in Electronics Engineering at Universitas Negeri Padang in 2013 and then pursued a Master’s Degree in Multimedia Telecommunication at Institut Teknologi Sepuluh Nopember, graduating in 2015. The author's research interests lie in the areas of wireless communication, wireless sensor networks, and the Internet of Things.