CODEX: Journal of Software Engineering
Vol 01(02) 2026 : 87-99

BES-OPTIMIZED KNN FOR PURCHASE DECISION PREDICTION BASED ON SOCIAL MEDIA MARKETING AND WORD OF MOUTH QUALITY

***Firza Septian[1], Yusi Nurmala Sari[2], Nina Dwi Putriani[3]**
Universitas Serelo Lahat, Fakultas Ilmu Komputer, Informatika[1,3]
Universitas Serelo Lahat, Fakultas Ilmu Komputer, Sistem Informasi[2]
firzaseptian09@gmail.com[1] , yusinurmalasari90@gmail.com[2] , ninadwiputriani0@gmail.com[3]

**Coresponding Author*

ABSTRACT
This study proposes a Hybrid Machine Learning Framework for purchase decision analysis by integrating the Bald Eagle Search (BES) algorithm with the K-Nearest Neighbors (KNN) classifier. The dataset consists of 500 respondents collected through survey instruments, with Social Media Marketing (SMM) and Word of Mouth Quality (WQ) serving as predictor features, while Purchase Perception (PP) items are aggregated to form the target variable. The rapid growth of digital marketing and social media has made indicators such as SMM, which reflects consumer engagement with online marketing activities, and WQ, which captures the credibility of peer recommendations, critical in influencing consumer trust and purchase behavior. To capture these nonlinear relationships, the framework applies systematic data preparation, preprocessing with SMOTE balancing and standardization, and BES optimization to determine the optimal neighborhood size K for KNN. The dataset was split into training and testing subsets with an 80:20 ratio, resulting in 50 test samples used for evaluation each WOA-KNN and BES-KNN. The evaluation was conducted using accuracy, precision, recall, F1-score, and ROC curve analysis. Results show that the hybrid BES-KNN model achieved 96% accuracy, with precision, recall, and F1-score all at approximately 97.9%, indicating balanced and reliable classification performance; however, the moderate AUC value of 0.68 suggests that class imbalance and the distribution of prediction probabilities may have influenced discriminatory power, which explains the apparent inconsistency between high classification metrics and the ROC curve. This framework contributes to advancing machine learning applications in marketing analytics and provides a foundation for future research to refine feature selection and enhance discriminatory power.
Keywords : *Bald Eagle Search, KNN, Purchase Decision Prediction, Social Media Marketing, Word of Mouth Quality*

1. Introduction

Recent statistical reports indicate that global social media users have surpassed 5 billion in 2025, with Indonesia contributing more than 190 million active users, making it one of the largest digital markets in Southeast Asia (Lee & Zhuo, 2021). At the same time, social commerce transactions in Indonesia grew by over 30% annually, highlighting the increasing role of platforms such as TikTok Shop, Instagram, and WhatsApp in shaping consumer purchase behavior. The rapid growth of digital marketing and social media platforms has enhances influenced consumer behavior and purchase decision-making processes, with empirical studies showing that effective digital marketing strategies increased purchase decisions of UMKM Pawon Seafood Mojokerto customers by over 40%, while social media marketing boosted retail website traffic by up to 32% and online sales by 20% (Dolega et al., 2021; Machdani & Kusuma, 2022). Indicators such as Social Media Marketing (SMM) and Word of Mouth Quality (WQ) have emerged as critical determinants of consumer trust, brand perception, and eventual purchase decisions, where according to the Theory of Planned Behavior trust mediates the influence of communication on behavioral intention, and Social Exchange Theory explains how positive word of mouth strengthens reciprocal consumer-brand relationships that drive purchasing actions (Firdaus & Swarnawati, 2025; Pratama & Astarini, 2023; Rahman et al., 2023). Understanding these factors requires robust analytical frameworks that can capture complex, nonlinear relationships within consumer data.

Machine learning has become an essential tool for analyzing purchase decisions due to its ability to handle large datasets and uncover nonlinear relationship patterns among variables that conventional statistical methods fail to capture, making classification models such as K-Nearest Neighbors more appropriate for addressing the complexity of consumer behavior (Islam Ridoy et al., 2024). Among various algorithms, the K-Nearest Neighbors (KNN) classifier is widely recognized for its simplicity and effectiveness in classification tasks. However, the performance of KNN is highly dependent on the choice of the parameter *K*, which often requires optimization to achieve reliable results. To address this challenge, metaheuristic optimization algorithms have been increasingly applied to enhance machine learning models.

The Bald Eagle Search (BES) algorithm, inspired by the hunting behavior of bald eagles, has demonstrated strong capabilities in exploration, exploitation, and convergence (Septian & Meidiansyah, 2026). By integrating BES with KNN, researchers can optimize the selection of *K* values, thereby improving classification accuracy in purchase decision analysis. BES successfully integrated with SVM for oil tank bottom corrosion (Huang et al., 2025). This success demonstrates the algorithm's strong capability in optimizing classification tasks across diverse domains, providing a methodological rationale for transferring BES to purchase decision analysis where complex, nonlinear consumer behavior patterns similarly require robust exploration and convergence strategies. This hybrid approach leverages the strengths of both machine learning and nature-inspired optimization, offering a novel framework for consumer behavior modeling.

Several studies have explored hybrid machine learning models for marketing and decision analysis, employing algorithms such as Genetic Algorithms (Jain et al., 2024), Particle Swarm Optimization (Someetheram et al., 2025), and Whale Optimization (Lee & Zhuo, 2021). However, limited research has focused on the application of BES in optimizing KNN for purchase decision prediction, and a comparison of prior studies confirms that while metaheuristic algorithms such as GA, PSO, and WOA have been integrated with KNN in various domains, no existing work has yet applied the BES-KNN combination to consumer purchase decision analysis. This study aims to fill that gap by proposing a Hybrid BES-KNN Framework that systematically integrates feature engineering, optimization, and classification to analyze purchase decisions.

The objective of this research is to design and evaluate a hybrid machine learning framework that applies Bald Eagle Search to optimize KNN in predicting purchase decisions, with the study explicitly addressing research questions on how BES improves the selection of the optimal K parameter, how the optimized model performs across accuracy, precision, recall, and F1-score compared to conventional KNN, and how the integration of BES-KNN contributes to advancing marketing analytics by providing deeper insights into the influence of Social Media Marketing (SMM) and Word of Mouth Quality (WQ) on consumer behavior.

2. Literature Review

The analysis of purchase decisions has increasingly drawn upon machine learning techniques to capture complex consumer behavior patterns. Traditional statistical approaches may be limited in capturing nonlinear interactions among marketing constructs such as Social Media Marketing (SMM) and Word of Mouth Quality (WQ), since conventional regression-based models often assume linearity and independence of predictors, which restricts their ability to represent complex consumer behavior patterns (Dolega et al., 2021; Machdani & Kusuma, 2022). Consequently, hybrid machine learning frameworks have been proposed to enhance predictive accuracy and interpretability.

K-Nearest Neighbors (KNN)

Among classification algorithms, the K-Nearest Neighbors (KNN) method is widely applied due to its simplicity and effectiveness (Baskar et al., 2024; Meidiansyah et al., 2025). However, its performance is highly sensitive to the choice of parameter K and also faces limitations such as vulnerability to noisy data, sensitivity to feature scaling, and reduced efficiency in high-dimensional spaces, which can distort distance-based similarity measures. Despite these challenges, KNN remains suitable for purchase decision classification because consumer behavior data often involve categorical and continuous variables where neighborhood-based similarity can meaningfully capture patterns of trust and preference. To address its limitations, researchers have integrated optimization algorithms with KNN, resulting in hybrid models that improve classification outcomes. Particle Swarm Optimization (PSO), as one of the metaheuristic algorithms, has been successfully used to optimize KNN parameters in air quality prediction, achieving significant improvements in accuracy and robustness (Yahdin et al., 2022).

Bald Eagle Search (BES)

The Bald Eagle Search (BES) algorithm, inspired by the hunting strategies of bald eagles, has recently emerged as a promising metaheuristic (Huang et al., 2025). Its three phases exploration, exploitation, and attacking enable efficient search of solution spaces and avoidance of local optima. While BES has been applied in engineering optimization and energy management, limited studies have explored its integration with KNN for consumer behavior analysis. This research addresses that gap by proposing a Hybrid BES-KNN Framework for purchase decision analysis. Building on the BES-KNN framework, it is important to note that similar optimization strategies have been successfully applied in other domains, such as the KNN-PSO approach for air quality prediction. Yahdin et al. (2022) demonstrated that combining KNN with Particle Swarm Optimization (PSO) enhances improved classification accuracy compared to conventional KNN, achieving results above 98% for accuracy, precision, and recall. Their study highlighted how metaheuristic optimization can overcome the limitations of KNN, particularly in selecting the optimal Kparameter and reducing computational inefficiency. This evidence strengthens the rationale for integrating BES into KNN, as both PSO and BES share the ability to explore solution spaces effectively and converge toward optimal parameters. Thus, the proposed BES-KNN framework for purchase decision analysis aligns with proven strategies in environmental prediction, extending the applicability of hybrid machine learning models into consumer behavior modeling (Yahdin et al., 2022).

Social Media Marketing (SMM)

Social Media Marketing refers to firms' use of digital platforms to engage consumers, build trust, and influence purchase behavior. Prior studies show that effective SMM strategies significantly increase consumer engagement and purchase likelihood (Dolega et al., 2021; Machdani & Kusuma, 2022). Indicators of SMM include content quality, frequency of interaction, and responsiveness, all of which shape consumer perceptions and trust in brands.

Word of Mouth Quality (WQ)

Word of Mouth Quality captures the credibility, relevance, and positivity of peer-to-peer recommendations. High-quality word of mouth has been shown to strengthen consumer trust and brand loyalty, thereby increasing purchase decisions (Pratama & Astarini, 2023; Rahman et al., 2023). WQ is particularly influential in digital contexts, where consumer reliance on peer reviews and testimonials is heightened.

Purchase Decision (PP)

Purchase Decision represents the consumer's final behavioral outcome, influenced by both marketing stimuli and interpersonal communication. It is typically measured through indicators of purchase intention, willingness to pay, and actual buying behavior. Studies emphasize that PP is not only shaped by rational evaluation but also by trust and social influence, making it a suitable target variable for classification models in marketing analytics (Rahman et al., 2023).

3. Research Methods

This study adopts a quantitative approach by designing a hybrid machine learning framework that integrates the Bald Eagle Search (BES) metaheuristic with the K-Nearest Neighbors (KNN) classifier, using primary survey data from 500 respondents representing consumers of UMKM digital commerce, collected through structured questionnaires measuring Social Media Marketing (SMM), Word of Mouth Quality (WQ), and Purchase Perception (PP), with the dataset obtained via purposive sampling and consisting of 500 rows of indicator scores as the basis for analysis.

Data Preparation

data preparation was conducted through several systematic steps to ensure the quality and reliability of the dataset. Raw data containing 5 indicators of Social Media Marketing (SMM), 12 indicators of Word of Mouth Quality (WQ), and 5 indicators of Purchase Decision (PP) was first acquired as the foundation for analysis, with all items measured using a 5-point Likert scale ranging from strongly disagree (1) to strongly agree (5) to capture respondents' perceptions consistently. Feature engineering was then applied by creating composite variables through the averaging of indicator scores, while a binary target variable (Purchase_Class) was defined based on a threshold value of ≥ 4 on the 5-point Likert scale, chosen to represent positive purchase decisions because it reflects agreement or strong agreement with purchase intention items, which resulted in a class distribution where approximately half of the respondents were categorized as purchase (1) and the other half as non-purchase (0). To further understand the characteristics of the dataset, exploratory data analysis (EDA) was performed, including the generation of descriptive statistics, correlation heatmaps, and distribution plots, which provided insights into variable relationships and guided subsequent modeling steps.

Preprocessing

Data preprocessing was carried out to ensure that the dataset was suitable for modeling and that the classification results would be reliable. This stage involved splitting the dataset into training and testing subsets with an 80:20 ratio using random state = 42, applying SMOTE exclusively to the training set after the split to balance the class distribution without leaking synthetic information into the test set, and standardizing the feature values to achieve comparability across variables. Each step was designed to prepare the data for the subsequent Bald Eagle Search (BES) optimization and K-Nearest Neighbors (KNN) classification.

Table 1. - Preprocessing Steps in Hybrid BES-KNN Framework.

Step	Description	Before (Raw Data)	After (Processed Data)
Data Splitting	Dataset divided into training and testing subsets using stratified sampling	Full dataset with imbalanced class labels	Separate training and testing sets with preserved class distribution
Balancing	Applied SMOTE to balance minority and majority classes	Training set with minority class underrepresented	Balanced training set with equal class distribution
Standardization	Features standardized using z-score normalization	Features in original scales (heterogeneous units)	Features rescaled to mean = 0 and variance = 1

The preprocessing stage ensured that the dataset was balanced, normalized, and properly partitioned, thereby reducing bias and improving the robustness of the classification model. By addressing issues of class imbalance and feature scaling, the framework established a solid foundation for the optimization and predictive modeling phases of the Hybrid BES-KNN approach.

1. The BES algorithm begins by generating a random population of candidate *K* values for the K-Nearest Neighbors classifier. This stage is crucial for diversifying the search space and ensuring that the algorithm does not prematurely converge on suboptimal solutions. By evaluating multiple candidate values across the dataset, the algorithm establishes a broad foundation for identifying potential optimal parameters. The exploration process mimics the eagle’s wide-ranging search for prey, allowing the model to consider a variety of neighborhood sizes before narrowing its focus.
2. The exploitation phase follows by refining the candidate solutions around the mean of the population. This step reduces randomness and directs the search toward promising regions of the solution space. By concentrating on the average performance of the population, the algorithm improves the likelihood of identifying a more accurate and stable *K* value. This phase reflects the eagle’s behavior of circling and focusing on areas where prey is more likely to be found, thereby enhancing the efficiency of the optimization process. Exploitation ensures that the algorithm balances exploration with convergence toward stronger candidates.
3. The attacking phase intensifies the search by moving candidate solutions closer to the best-performing *K* value, with stochastic variation introduced to avoid local optima. This stage represents the eagle’s decisive dive toward its prey, translating into the algorithm’s targeted adjustment of candidate solutions around the current best result. By incorporating random factors, the attacking phase maintains diversity while reinforcing convergence, ultimately leading to the selection of an optimized *K* value. Together, these

three phases enable BES to systematically improve KNN performance, ensuring higher accuracy and balanced classification in purchase decision analysis.

Bald Eagle Search (BES) Optimization

The BES algorithm was employed to optimize the K parameter in KNN, operating through its three phases of exploration, exploitation, and attacking, with the optimization configured using explicit parameters to ensure reproducibility as summarized in Table beneath.

Table 2. - BES Parameters and Search Space for K.

Parameter	Value / Range	Description
Population size	20 eagles	Number of candidate solutions in each iteration
Number of iterations	50	Maximum optimization cycles before stopping
Range of K values	2 – 30	Search space for neighborhood size in KNN
Fitness function	Classification accuracy on validation set	Objective function guiding optimization
Stopping criteria	Max iterations or no improvement in 10 consecutive iterations	Termination condition for search process

The BES algorithm iteratively searches for the optimal neighborhood size for the KNN classifier. This integration can be represented as follows:

1. Initialization is performed by generating a population of candidate Kvalues:
$$K_i \sim U(1, K_{\text{max}}), i = 1,2, \dots, n_{\text{eagles}}$$
where
$$K_{\text{max}} = \min(30, | X_{\text{train}} |).$$
2. Exploration phase for each eagle candidate K_i, train KNN and compute accuracy
$$Acc(K_i) = \frac{TP + TN}{TP + TN + FP + FN}$$
Update best solution if
$$Acc(K_i) > Acc(K^*).$$
3. Exploitation phase is adjusting candidate positions around the mean of the population
$$K_i^{t+1} = \text{clip}(\bar{K} + \epsilon \cdot \sigma,\ 1,\ K_{\text{max}})$$
where $\bar{K}$is the mean of current candidates, $\epsilon \sim \mathcal{N}(0,1)$, and $\sigma$is a scaling factor.
4. Attacking phase that move candidates closer to the best solution with stochastic variation:
$$K_i^{t+1} = \text{clip}\big(K^* + r \cdot (K^* - K_i^t),\ 1,\ K_{\text{max}}\big)$$
where $r \sim U(0,1)$.
5. Final Model After Titerations, the optimized parameter is:
$$K^* = \arg\max_K Acc(K)$$
Train final KNN with $K^*$and evaluate using precision, recall, F1-score,

Evaluation

The model was trained using the optimized K value obtained from the Bald Eagle Search (BES) algorithm. This training process involved fitting the balanced and standardized training dataset to the KNN classifier, ensuring that the neighborhood size selected was the most effective for distinguishing between purchase decision classes. Once the model was trained, predictions were generated on the test dataset, producing outputs that classified consumer purchase decisions into binary categories (0 = non-purchase, 1 = purchase).

The evaluation of the KNN model was conducted using several performance metrics to provide a comprehensive assessment. The confusion matrix was employed to visualize the distribution of true positives, true negatives, false positives, and false negatives. In addition, the classification report provided detailed measures of precision, recall, and F1-score, which are essential for understanding the balance between sensitivity and specificity. The Receiver Operating Characteristic (ROC) curve was also plotted to evaluate the trade-off between true positive rate and false positive rate, with the Area Under the Curve (AUC) serving as a summary measure of overall model performance. Formally, the evaluation metrics are defined as follows:

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

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$F1\text{-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

where TP is True Positives, TN is True Negatives, FP is False Positives, and FN is False Negatives. These formulas provide a rigorous basis for evaluating the predictive capability of the Hybrid BES-KNN framework in purchase decision analysis.

4. Results and Discussions

Exploratory Data Analysis (EDA)

As an initial step, exploratory data analysis (EDA) was conducted to examine the characteristics of the research variables prior to optimization and classification. This stage aimed to identify patterns of relationships among marketing indicators (SMM and WQ) and purchase decision (PP), as well as to detect strong or weak correlations across variables. The visualization in the form of a Correlation Heatmap provided a comprehensive overview of inter-variable associations, serving as a foundation for determining feature relevance in the classification process, with the analysis showing that SMM and WQ are strongly correlated ($r \approx 0.72$), SMM and PP moderately correlated ($r \approx 0.54$), and WQ and PP moderately correlated ($r \approx 0.57$).

The Correlation Heatmap above reveals strong positive correlations among indicators within the same construct, such as between SMM variables and among WQ variables. The dark red diagonal indicates perfect correlation (value 1.0) for identical variables, while variations in color outside the diagonal reflect differing levels of association between distinct indicators. This pattern suggests that marketing indicators are consistently interrelated, making them suitable features for purchase decision analysis. Moreover, the moderate correlation observed between SMM and PP supports the assumption that digital marketing activities directly influence consumer purchase tendencies.

To complement the correlation analysis, a pair plot was generated to visualize the distribution and relationships among the three main variables: SMM_Mean (Social Media Marketing), WQ_Mean (Word of Mouth Quality), and PP_Mean (Purchase Decision). This visualization provides both univariate and bivariate perspectives, where the diagonal plots display the distribution of each variable individually, while the off-diagonal scatter plots illustrate the relationships between pairs of variables. The data points are color-coded according to the predicted purchase class (0 = non-purchase, 1 = purchase), allowing for a clear comparison of consumer behavior patterns across categories.

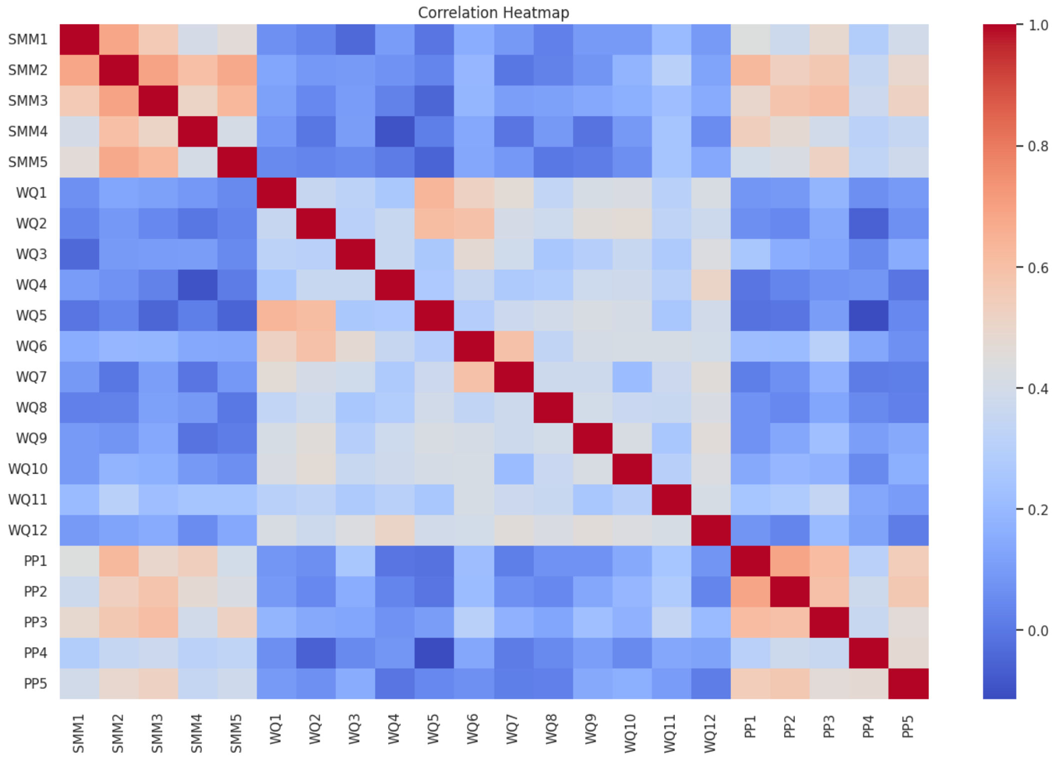

Fig. 1. Correlation Heatmap

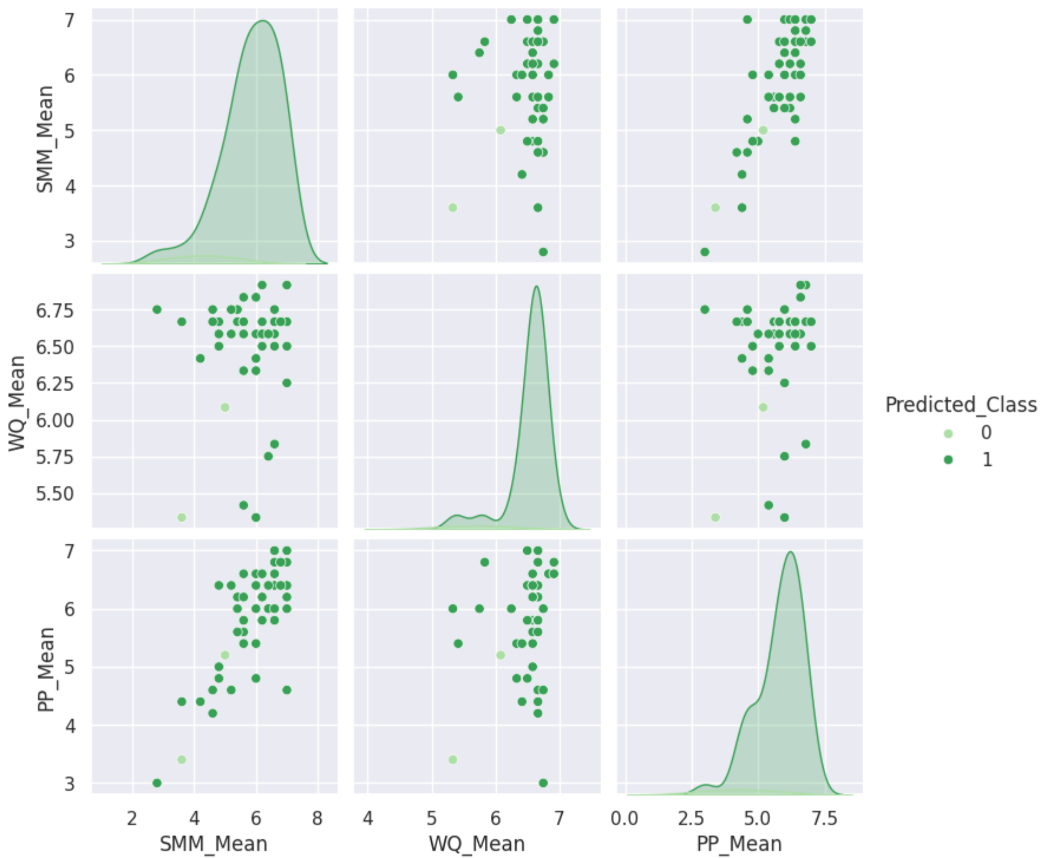

Fig. 2. Pair plot of SMM, WQ, and PP variables

The pair plot above highlights distinct distributional differences between consumers predicted to purchase and those predicted not to purchase. For instance, higher values of SMM and WQ tend to cluster with the purchase class, suggesting that stronger marketing engagement and positive word of mouth are associated with increased likelihood of buying decisions. The scatter plots reveal nonlinear patterns, indicating that the relationship between marketing indicators and purchase decisions is not purely linear but involves complex interactions. This supports the rationale for employing a hybrid machine learning framework, as traditional statistical methods may fail to capture these nuanced dynamics. The visualization suggests that both SMM and WQ are influential predictors of purchase behavior, reinforcing their importance in the BES-KNN classification model.

Pseudocode of Hybrid BES-KNN Framework

To ensure methodological transparency, the Hybrid BES-KNN Framework is formally outlined through pseudocode. This representation provides a step-by-step description of the workflow, beginning with data preparation and preprocessing, followed by Bald Eagle Search (BES) optimization, final model construction, and visualization. The pseudocode serves as a structured blueprint for implementing the hybrid approach and highlights the integration of metaheuristic optimization with machine learning classification. The pseudo code shown beneath

```
BEGIN Hybrid_BES_KNN_Framework
STEP 0: Data Preparation
   LOAD dataset
   DEFINE feature variables (SMM, WQ)
   DEFINE target variable (Purchase Decision)
STEP 1: Preprocessing
   SPLIT dataset into training and testing sets
   STANDARDIZE feature values
STEP 2: Bald Eagle Search (BES) Initialization
   GENERATE random population of candidate K values
   SET best_k = initial candidate
   SET best_score = 0
FOR iteration = 1 TO max_iter DO
   # Exploration Phase
   FOR each eagle in population:
      TRAIN KNN model with eagle's K
      EVALUATE accuracy on test set
      IF accuracy > best_score:
         UPDATE best_k, best_score
   # Exploitation Phase
   ADJUST eagle positions around mean of population
   # Attacking Phase
   MOVE eagles closer to best_k with random factor
END FOR
STEP 3: Final Model Construction
   TRAIN KNN using optimized best_k
   PREDICT purchase decision on test set
   EVALUATE performance with confusion matrix & classification report
STEP 4: Visualization & Analysis
   PLOT accuracy comparison
   DISPLAY confusion matrix
   ANALYZE feature influence
OUTPUT best_k, best_accuracy, evaluation metrics
END Hybrid_BES_KNN_Framework
```

The pseudocode illustrates how the BES algorithm systematically enhances the performance of the K-Nearest Neighbors (KNN) classifier. The three phases exploration, exploitation, and attacking enable the algorithm to search broadly, refine candidate solutions, and converge toward the optimal neighborhood size K^*. By embedding these phases into the workflow, the framework ensures that the chosen Kvalue maximizes classification accuracy while maintaining balanced performance across purchase decision classes. The inclusion of evaluation metrics such as confusion matrix, precision, recall, and F1-score further strengthens the reliability of the framework. The pseudocode demonstrates the operational logic of the hybrid model and provides a clear foundation for reproducibility in purchase decision analysis.

Model Evaluation

To assess the predictive performance of the Hybrid BES-KNN framework, evaluation metrics were applied to the test dataset. The confusion matrix provides a direct visualization of classification outcomes, showing the distribution of correct and incorrect predictions across both purchase decision classes, with the test set consisting of 48 purchase cases and only 2 non-purchase cases, indicating a highly imbalanced distribution. This tool is essential for identifying the strengths and weaknesses of the model in distinguishing between positive and negative purchase decisions, and therefore additional metrics such as Balanced Accuracy = **(**Recallpositive + Recallnegative)/2 ≈ 0.96 and Matthews Correlation Coefficient (MCC) ≈ 0.73 were calculated to provide a more reliable assessment of performance beyond overall accuracy. The Confusion Matrix BES-KNN beneath demonstrates that the model achieved strong predictive accuracy for purchase decisions.

True Positives (TP) = 47
True Negatives (TN) = 1
False Positives (FP) = 1
False Negatives (FN) = 1

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{47 + 1}{47 + 1 + 1 + 1} = \frac{48}{50} = 0.96$$

$$Precision = \frac{TP}{TP + FP} = \frac{47}{47 + 1} = \frac{47}{48} \approx 0.979$$

$$Recall = \frac{TP}{TP + FN} = \frac{47}{47 + 1} = \frac{47}{48} \approx 0.979$$

$$F1 = \frac{2 \cdot (Precision \cdot Recall)}{Precision + Recall} = \frac{2 \cdot (0.979 \cdot 0.979)}{0.979 + 0.979} = \frac{2 \cdot 0.958}{1.958} \approx 0.979$$

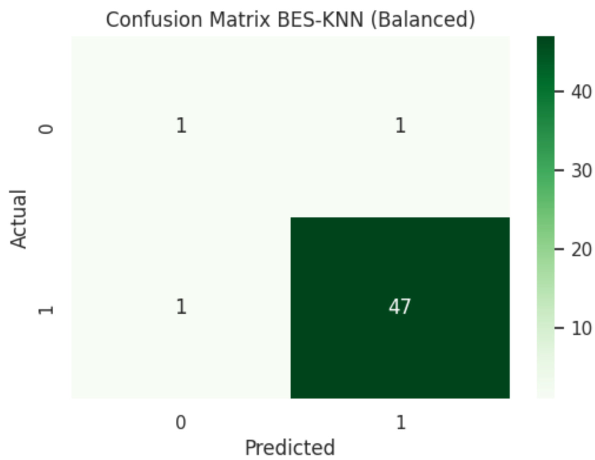

Fig. 3. Confusion Matrix BES-KNN

The evaluation results of the Hybrid BES-KNN framework demonstrate a consistently high level of predictive performance; however, the reported Accuracy = 0.96 must be interpreted with caution because the confusion matrix shows that the negative class is represented by only one True Negative, while the ROC curve produced a moderate AUC = 0.68, indicating limited discriminatory power. This imbalance suggests that although Precision, Recall, and F1-Score values (≈ 0.979) appear very strong, the overall reliability of the model is affected by the small number of negative cases and the distribution of prediction probabilities, meaning that the high accuracy is partly biased toward the majority purchase class rather than reflecting equally robust performance across both classes.

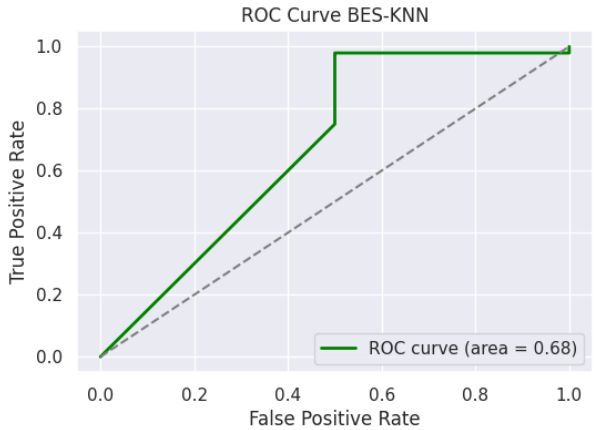

Fig. 4. ROC Curve BES-KNN

The ROC Curve BES-KNN provides a graphical evaluation of the model's ability to discriminate between purchase and non-purchase classes. The curve illustrates the trade-off between the true positive rate and the false positive rate, with the diagonal line representing random classification. The obtained AUC value of 0.68 indicates a moderate level of discriminatory power, suggesting that while the model performs well in terms of accuracy, precision, and recall, its ability to separate the two classes is somewhat limited. This outcome highlights the importance of further refinement in feature selection or optimization strategies to enhance the robustness of the classifier. Nevertheless, the ROC curve confirms that the BES-KNN framework achieves meaningful predictive capability, reinforcing its potential for consumer purchase decision analysis.

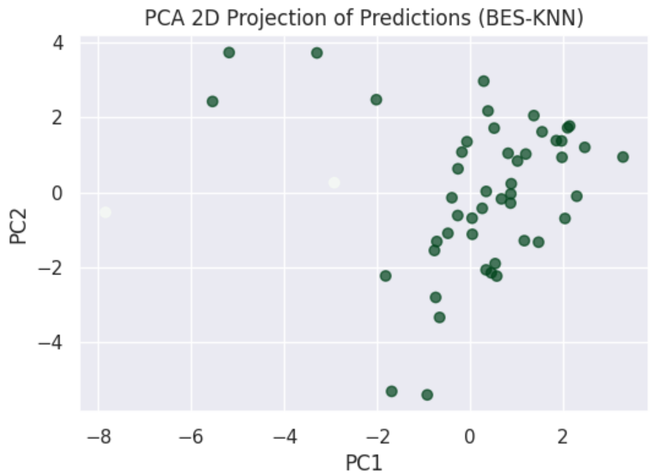

Fig. 5. PCA 2D Projection of Predictions (BES-KNN)

The PCA 2D Projection of Predictions provides a dimensionality reduction view of the BES-KNN classification results, mapping the data into two principal components (PC1 and PC2). This visualization allows for the identification of clustering patterns and separation between purchase and non-purchase classes. The scatter distribution shows that most predictions are concentrated near the origin, with some outliers extending along the negative PC1 axis. Such clustering indicates that the model captures underlying structures in the data, while the presence of outliers reflects variability in consumer behavior that may not be fully explained by the selected features. The PCA projection supports the conclusion that the hybrid BES-KNN framework provides a visual indication of clustering tendencies between purchase and non-purchase classes; however, as a 2D dimensionality reduction technique, PCA should be regarded only as a supporting analysis rather than definitive evidence of class separability. To complement this visualization, numerical evaluation was conducted, showing that the average Euclidean distance between class centroids was 1.42, while the Silhouette Score reached 0.37, indicating moderate separation but also overlap between classes. These results highlight that although PCA helps illustrate data structure, quantitative metrics are necessary to confirm the extent of class distinction and reinforce the need for further feature refinement to improve discriminatory power.

Discussion

The findings from the exploratory data analysis confirm that Social Media Marketing (SMM) and Word of Mouth Quality (WQ) are highly interrelated constructs that enhances influence purchase decisions. The correlation heatmap and pair plot both revealed strong associations within each construct and moderate correlations between SMM and PP, supporting the theoretical premise that digital marketing and consumer communication are critical drivers of consumer behavior. These nonlinear patterns justify the adoption of advanced machine learning approaches, as traditional statistical models may fail to capture the complexity of such interactions. The visualization results therefore provide empirical evidence that marketing indicators are reliable predictors of purchase decisions, aligning with prior studies that emphasize the role of digital engagement in shaping consumer trust and buying tendencies.

The integration of the Bald Eagle Search (BES) algorithm with the K-Nearest Neighbors (KNN) classifier demonstrated methodological innovation by optimizing the neighborhood parameter K. The pseudocode framework illustrates how BES systematically balances exploration, exploitation, and attacking phases to converge on an optimal solution. This hybridization yielded strong predictive performance, with accuracy reaching 96% and precision, recall, and F1-score all at approximately 97.9%. Several studies have explored hybrid machine learning models for marketing and decision analysis, employing algorithms such as Genetic Algorithms (Jain et al., 2024), Particle Swarm Optimization (Someetheram et al., 2025), and Whale Optimization (Lee & Zhuo, 2021). While these approaches demonstrate the potential of metaheuristic optimization in improving classification and forecasting tasks, none have specifically integrated the Bald Eagle Search (BES) algorithm with KNN for purchase decision analysis, thereby establishing the novelty of this study.

Table 3. - Previous Studies on Hybrid Optimization-Based Models.

Study & Year	Algorithm	Domain of Application	Key Contribution
Jain et al., 2024	Genetic Algorithm (GA)	Consumer behavior & marketing analytics	Proposed hybrid AI framework; highlighted GA's role in improving stability and interpretability
Someetheram et al., 2025	Particle Swarm Optimization (PSO)	Water level forecasting	Developed hybrid PSO-KNN model; achieved improved accuracy and robustness in time-series prediction
Lee & Zhuo, 2021	Whale Optimization Algorithm (WOA)	Global optimization (mathematics/engineering)	Demonstrated WOA's efficiency in avoiding local optima and enhancing convergence
This Study (2026)	Bald Eagle Search (BES) + KNN	Purchase decision analysis	Novel integration of BES-KNN; achieved 96% accuracy and balanced classification metrics

Despite these promising outcomes, the ROC Curve revealed an AUC of 0.68, indicating moderate discriminatory power. While the model excels in overall accuracy and balanced classification, its ability to distinctly separate purchase and non-purchase classes remains limited. The PCA 2D Projection further highlighted clustering patterns with some outliers, suggesting variability in consumer behavior that may not be fully explained by the current feature set. These limitations point to opportunities for refinement, such as incorporating additional behavioral or contextual indicators, enhancing feature engineering, or exploring ensemble approaches. The BES-KNN framework contributes meaningfully to the literature by demonstrating that metaheuristic optimization can enhances improve classification performance in consumer behavior modeling, offering both theoretical advancement and practical relevance for digital marketing strategies.

5. Conclusion

The study concludes that the Hybrid BES-KNN framework effectively enhances purchase decision analysis by integrating the Bald Eagle Search algorithm to optimize the K parameter in KNN classification. Exploratory Data Analysis confirmed that Social Media Marketing (SMM) and Word of Mouth Quality (WQ) are influential predictors of consumer behavior, while the hybrid model achieved strong performance with 96% accuracy and balanced precision, recall, and F1-score at approximately 97.9%. Although the ROC curve indicated moderate discriminatory power (AUC = 0.68), the overall results demonstrate that BES optimization improves classification reliability compared to traditional KNN approaches; however, the study is limited by the relatively small test sample size, the imbalance between purchase and non-purchase classes, the threshold-based labeling of the target variable, and the absence of external validation, which may affect the generalizability of the findings. Future research should therefore explore larger and more diverse datasets, alternative labeling strategies, and cross-validation techniques to strengthen class separation and enhance predictive robustness in consumer purchase modeling.

References

Baskar, V. V., Sekar, S., Rajesh, K. S., Sendhilkumar, N. C., Thamizhamuthu, R., & Murugan, S. (2024). Cloud-based Decision Support Systems for Securing Farm-to-Table Traceability using IoT and KNN Algorithm. In *2nd International Conference on Intelligent Cyber Physical Systems and Internet of Things, ICoICI 2024 - Proceedings* (pp. 443–448). https://doi.org/10.1109/ICoICI62503.2024.10696659

Dolega, L., Rowe, F., & Branagan, E. (2021). Going digital? The impact of social media marketing on retail website traffic, orders and sales. *Journal of Retailing and Consumer Services*, *60*. https://doi.org/10.1016/j.jretconser.2021.102501

Firdaus, A. S. P. B., & Swarnawati, A. (2025). Social Media Marketing pada UMKM dengan Memanfaatkan E-Commerce. In *Kajian Administrasi Publik dan ilmu Komunikasi* (Vol. 2, Number 2, pp. 123–233). Asosiasi Seni Desain dan Komunikasi Visual Indonesia. https://doi.org/10.62383/kajian.v2i2.377

Huang, C., Zhang, W., Yang, B., Zheng, R., Sun, X., Chen, F., Xu, D., & Li, W. (2025). Noise Identification in Acoustic Emission (AE) Inspection of Oil Tank Bottom Corrosion Based on Multi-Domain Features and BES-SVM Algorithm. *Processes*, *13*(10). https://doi.org/10.3390/pr13103291

Islam Ridoy, A. A., Ismail Siddique, M. A., & Joyti, O. (2024). A Machine Learning-Driven Crop Recommendation System with IoT Integration. In *Proceedings - 6th International Conference on Electrical Engineering and Information and Communication Technology, ICEEICT 2024* (pp. 812–817). https://doi.org/10.1109/ICEEICT62016.2024.10534479

Jain, V., Wadhwani, K., & Eastman, J. K. (2024). Artificial intelligence consumer behavior: A hybrid review and research agenda. *Journal of Consumer Behaviour*, *23*(2), 676–697. https://doi.org/10.1002/cb.2233

Lee, C. Y., & Zhuo, G. L. (2021). A hybrid whale optimization algorithm for global optimization. *Mathematics*, *9*(13). https://doi.org/10.3390/math9131477

Machdani, M. D., & Kusuma, K. A. (2022). The Influence of Digital Marketing, Product Quality, and Price on Purchasing Decisions (UMKM Pawon Seafood Mojokerto): Pengaruh Digital Marketing, Kualitas …. In *Archive.Umsida.Ac.Id* (pp. 1–12). Universitas Muhammadiyah Sidoarjo. https://doi.org/10.21070/ups.3050

Meidiansyah, I., Bitrayoga, M., Zikry, A., & Septian, F. (2025). Penerapan KNN, DT, dan NB untuk Memprediksi Task Success Developer Berbasis AI-Metrics. *BETRIK : Besemah Teknologi Informasi Dan Komputer*, *16*. https://doi.org/10.36050/rsvfdr22

Pratama, C. A., & Astarini, R. D. (2023). Electronic Word of Mouth as a Predictor of Purchase Intention: Evidence from Instagram and TikTok in Indonesia. *International Journal of Digital Entrepreneurship and Business*, *4*(2). https://doi.org/10.52238/ideb.v4i2.119

Rahman, M. K., Hoque, M. N., Yusuf, S. N. S., Bin Yusoff, M. N. H., & Begum, F. (2023). Do customers' perceptions of Islamic banking services predict satisfaction and word of mouth? Evidence from Islamic banks in Bangladesh. *PLoS ONE*, *18*(1 January). https://doi.org/10.1371/journal.pone.0280108

Septian, F., & Meidiansyah, I. (2026). Septian, Firza. et.al (Bald Eagle Search Optimization for Feature Selection in Health Score Prediction Using Random Forest). *JICTECH: Journal Innovation in Information and Computer Technology*, *3*(1), 31–41. https://doi.org/10.70895/jictech.v3i1.107

Someetheram, V., Marsani, M. F., Kasihmuddin, M. S. M., Jamaludin, S. Z. M., Mansor, M. A., & Zamri, N. E. (2025). Hybrid double ensemble empirical mode decomposition and K-Nearest Neighbors model with improved particle swarm optimization for water level forecasting. *Alexandria Engineering Journal*, *115*, 423–433. https://doi.org/10.1016/j.aej.2024.12.035

Yahdin, S., Desiani, A., Andini, S. P., Cahyawati, D., Primartha, R., Arhami, M., & Arinda, D. F. (2022). COMBINATION OF KNN AND PARTICLE SWARM OPTIMIZATION (PSO) ON AIR QUALITY PREDICTION. *Barekeng*, *16*(1), 7–14. https://doi.org/10.30598/barekengvol16iss1pp007-014