








## Travel Vlog Reviews: Support Vector Machine Performance in Sentiment Classification

Yerik Afrianto Singgalen<sup>1\*</sup>, Sih Yuliana Wahyuningtyas<sup>2</sup>, Yohanes Eko Widodo<sup>3</sup>, Muhamad Nur Agus Dasra<sup>1</sup>,  
Ruben William Setiawan<sup>1</sup>

<sup>1</sup> Department of Tourism, Atma Jaya Catholic University of Indonesia, Jakarta 12930, Indonesia

<sup>2</sup> Department of Law, Atma Jaya Catholic University of Indonesia, Jakarta 12930, Indonesia

<sup>3</sup> Department of Business Administration, Atma Jaya Catholic University of Indonesia, Jakarta 12930, Indonesia

Corresponding Author Email: [yerik.afrianto@atmajaya.ac.id](mailto:yerik.afrianto@atmajaya.ac.id)

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.300109>

### ABSTRACT

**Received:** 20 September 2023

**Revised:** 10 December 2024

**Accepted:** 15 January 2025

**Available online:** 25 January 2025

#### Keywords:

*Support Vector Machine (SVM), Synthetic Minority Over-sampling Technique, sentiment analysis, imbalanced datasets, machine learning*

This research investigates the combination of the Support Vector Machine algorithm with the Synthetic Minority Over-sampling Technique to improve classification performance in sentiment analysis, especially in handling imbalanced datasets. Employing a dataset comprising 1,928 text entries, the research highlights SVM's challenges in managing imbalanced data, where a predisposition toward the majority class leads to less-than-optimal classification results. Through the application of SMOTE, synthetic samples were generated to balance the minority class, resulting in notable performance improvements, including an accuracy of 83.12%, a precision of 75.76%, a recall of 97.53%, and an Area Under the Curve (AUC) score of 0.978. These outcomes emphasize the effectiveness of integrating SVM and SMOTE to balance class distributions and enhance the model's capacity to distinguish between positive and negative sentiments. The findings underscore the importance of strategic model optimization to achieve balanced results and contribute to advancements in sentiment analysis methodologies.

## 1. INTRODUCTION

The foundation of this study lies in the dynamic progression of digital marketing and the growing intricacy of consumer behavior within online contexts. The emergence of sophisticated digital platforms and the widespread creation of user-generated content have compelled businesses to analyze extensive datasets to gain insights into consumer preferences and emerging trends. This need has catalysed a shift from traditional marketing methods to more data-driven, analytical approaches that leverage machine learning and artificial intelligence. The intersection of these technological advancements and marketing strategies presents an opportunity to explore how digital interactions shape consumer decision-making processes. This study aims to address the gap in understanding these dynamics by employing innovative methodologies that combine computational techniques with behavioral analysis. This approach is anticipated to generate both academically valuable and practically applicable insights, offering a thorough understanding of the digital consumer environment.

The urgency of analyzing travel vlog reviews using the SVM method stems from the rapid proliferation of user-generated content, significantly influencing consumer behavior and decision-making in the tourism industry [1-3]. As digital platforms become increasingly integral to travel planning, the sentiments and preferences conveyed in travel vlogs offer critical insights into tourist expectations and experiences [4]. These reviews reflect genuine user

perspectives and serve as a valuable resource for businesses and policymakers to optimize destination marketing strategies and enhance service delivery [5]. By systematically analyzing this content, tourism stakeholders can identify emerging trends, address service gaps, and refine promotional efforts to align with evolving consumer demands [6]. This analytical capability, supported by SVM's precision, underscores the indispensable role of sentiment analysis in sustaining competitive advantage in the rapidly transforming travel industry landscape.

This study seeks to systematically explore the incorporation of sentiment analysis in assessing travel vlog reviews, emphasizing both methodological and practical aspects. Specifically, the study addresses the following questions: (1) How can SVM techniques be optimized to improve sentiment classification accuracy in imbalanced datasets? (2) What are the critical factors influencing the effectiveness of sentiment analysis in capturing nuanced consumer preferences in travel-related content? (3) How does integrating synthetic data generation techniques like SMOTE impact model performance and generalizability? By addressing these questions, the research intends to achieve a dual objective: advancing methodological innovation in sentiment analysis and generating actionable insights that tourism businesses and policymakers can leverage to refine marketing strategies and enhance consumer engagement. The results are anticipated to enhance understanding of digital consumer behavior and establish a solid framework for applying machine learning in tourism analytics.

This research breaks new ground by developing a sophisticated methodology combining computational learning systems with human behavior research to examine customer engagement across digital platforms. While existing research has typically separated qualitative observations from statistical analysis, our framework introduces a unified system that harmonizes algorithmic exactitude with behavioral science perspectives. This dual approach transcends conventional research paradigms, offering more profound insights into consumer patterns [7-11]. This integration allows for a more nuanced understanding of complex consumer behaviors, particularly how they respond to digital marketing stimuli across various platforms [12, 13]. Such a methodological advancement enriches the theoretical framework surrounding consumer behavior analysis and offers practical applications for optimizing digital marketing strategies. This study advances the field by establishing innovative connections between data-driven analysis and human behavior research, offering valuable insights that expand current academic understanding. The methodological framework developed here opens new pathways for researchers seeking to combine these historically separate domains, potentially catalysing future developments in cross-disciplinary research approaches.

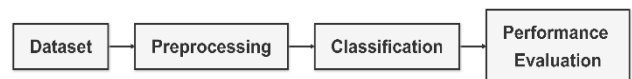
The significance of this investigation extends across multiple dimensions, generating substantial insights for theoretical discourse and industry implementation. Our findings advance the academic knowledge body while providing actionable frameworks that enhance operational effectiveness. Theoretically, this study offers a novel conceptual framework that integrates elements of consumer psychology with machine learning algorithms, thereby enriching the academic discourse on digital marketing strategies and consumer behavior. This interdisciplinary approach deepens the understanding of how digital stimuli affect consumer decision-making and provides a robust foundation for future studies to build upon. From a practical standpoint, the findings of this research are poised to inform the development of more effective marketing campaigns, particularly by enabling businesses to tailor their digital strategies more precisely to target specific consumer segments. The integrated outcomes of this investigation demonstrate its transformative influence across conceptual frameworks and operational applications. This comprehensive approach establishes new standards for research excellence while providing actionable insights that advance scholarly discourse and market-driven solutions.

Prior digital commerce and consumer psychology investigations have concentrated on leveraging computational analysis of extensive datasets to forecast customer tendencies and purchasing patterns. The conventional approach has emphasized quantitative methodologies to extract meaningful patterns from consumer interactions [14-18]. Existing scholarly work has demonstrated the capabilities of advanced data interpretation methodologies to uncover meaningful consumer insights. By systematically examining digital interactions and user-generated content, these investigations have revealed underlying patterns in customer sentiment and behavior [19]. Evidence from these investigations confirms the measurable advantages of implementing sophisticated analytical approaches to enhance market comprehension and strategic

planning. This empirical validation strengthens the case for adopting advanced methodologies in modern marketing operations. However, while these methodologies provide valuable contributions, there remains a gap in fully integrating psychological theories of consumer behavior with advanced data-driven techniques. Addressing this gap by combining computational analytics with a deeper understanding of psychological factors could lead to more robust models that better capture the complexities of consumer decision-making in digital environments. This synthesis of existing research highlights the ongoing need for interdisciplinary approaches that leverage technological advancements and behavioral science to advance the field further.

## 2. METHOD

The stages of this research are systematically designed to ensure a comprehensive and rigorous examination of the proposed hypotheses. The research commences with a methodical compilation phase, integrating crucial elements from various sources to develop an extensive information base. Following this, data preprocessing is conducted to cleanse and prepare the data for analysis, removing any inconsistencies or irrelevant information that might affect the outcome. The analytical phase implements advanced computational frameworks to extract meaningful correlations from the collected information. Raw datasets are systematically evaluated using intelligent processing systems to reveal underlying trends. This critical process transforms unstructured information into actionable knowledge supporting research conclusions. Finally, the results are interpreted to contextualize the findings within the broader theoretical framework, providing a deeper understanding of the implications for academic discourse and practical applications. This sequential approach ensures that each stage builds upon the previous one, leading to a coherent and insightful exploration of the research questions.



**Figure 1.** Research methodology framework

Figure 1 shows that the research framework encompasses a substantial dataset of 1,928 textual entries designated for sequential refinement and analytical processing. The methodology prioritizes data refinement protocols to eliminate extraneous elements and inconsistencies, establishing a foundation for precise analytical outcomes. This meticulous preparation safeguards against potential analytical distortions from unprocessed information. The subsequent extraction methodology identifies and isolates fundamental elements aligned with research parameters, emphasizing critical indicators essential for core analytical procedures. The research ensures data integrity and relevance through the methodical implementation of these protocols, establishing robust analytical groundwork. This systematic framework enhances the validity of the research and strengthens its scholarly contribution through precise data handling and processing mechanisms.

## 2.1 Data preprocessing

The implemented data preparation framework follows a structured sequence that transforms raw information into an analysis-ready format through distinct processing phases. The initial refinement protocol addresses data quality through systematic identification and correction of anomalies, inconsistencies, and null values, establishing a foundation of analytical reliability. The framework then proceeds to information synthesis, where distinct data sources undergo consolidation to generate a unified analytical base, enabling comprehensive examination. The final preparation phase encompasses structural optimization, where information undergoes standardization and organizational protocols to enhance its analytical utility and modeling compatibility. This systematically engineered approach ensures the creation of a robust analytical foundation through methodical data refinement processes. The implemented protocols maintain strict data integrity standards while optimizing information structure for subsequent analytical procedures, supporting the research objectives through precise data preparation methodologies. Following transformation, data reduction is done to simplify the dataset by reducing its dimensionality, thus improving computational efficiency while retaining the most relevant information. The final step, data discretization, transforms continuous data into discrete intervals, facilitating the application of specific statistical or machine learning models that require categorical inputs. The methodological framework incorporates sequential processing protocols where each procedural element is a foundational component for subsequent phases, establishing a comprehensive analytical preparation system. This integrated approach ensures methodical refinement at each processing interval, reinforcing the analytical framework's capacity to generate meaningful insights. The systematic progression through interconnected preparation stages strengthens the research methodology's empirical foundation, ultimately enhancing the validity and reliability of investigative outcomes.

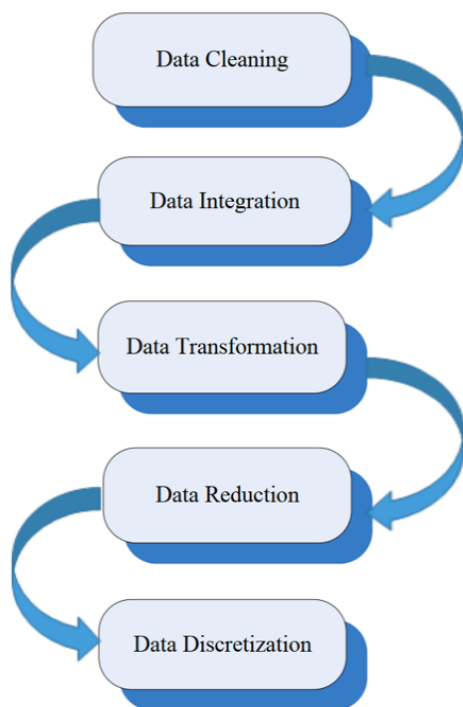


Figure 2. Data preprocessing (Source: [20])

Figure 2 shows the data preprocessing process, which includes several stages. Information refinement protocols encompass systematic detection and correction of data irregularities, gaps, and discrepancies to optimize quality standards. The primary objective is establishing dataset accuracy and analytical readiness for subsequent examination procedures. These refinement methodologies incorporate duplicate elimination, null value resolution, and structural standardization processes to ensure data integrity. Data Integration: This process combines data from various sources into a unified dataset. It is often required when data originates from different systems or databases. Data integration aims to produce consistent data free from conflicts between different data sources. Data Transformation: This stage entails converting or mapping data from one format or structure to another that is more suitable for analysis. The structural optimization process encompasses standardization procedures, consolidation mechanisms, and derivation of enhanced parameters from established variables within the dataset. Data Reduction: In this step, the volume of data is reduced without losing significant information. This is often achieved through feature selection, data compression, or eliminating redundant data to enhance analysis efficiency and reduce data complexity. Data Discretization: This stage involves converting continuous data into categorical data by grouping data values into intervals or bins. Discretization is commonly used in machine learning algorithms that perform better with categorical data or when the analysis objectives require data in categorical form.

The data preprocessing workflow is a structured sequence designed to transform raw data into a format suitable for analysis. Each step is outlined below with specific operation methods and parameter settings to ensure reproducibility. Data Cleaning involves identifying and correcting inaccurate, missing, or inconsistent data. Missing values were handled using median imputation for numerical features and mode imputation for categorical features. Outliers were detected and treated using the interquartile range (IQR) method, and duplicate entries were removed to maintain dataset integrity. Textual data underwent tokenization, lowercasing, and removal of stop words, punctuation, and special characters using the Natural Language Toolkit (NLTK). Data Integration: Data from multiple sources, including structured datasets and unstructured text files, were consolidated into a single dataset using unique identifiers. Consistency across data types and formats was ensured using Python's panda's library for merging and alignment.

Data Transformation: Textual data was vectorized using the Term Frequency-Inverse Document Frequency (TF-IDF) method with parameters set to  $\text{max\_features}=5000$  and  $\text{ngram\_range}=(1, 2)$  to capture both unigrams and bigrams. Using Min-Max scaling to eliminate scale disparities, numerical variables were normalized to a  $[0,1]$  range. Data Reduction: Feature selection was performed using recursive feature elimination (RFE) with an SVM as the estimator, retaining the top 20% of features based on model importance. Principal Component Analysis (PCA) was applied to reduce dimensionality further while preserving 95% of the variance. Data Discretization: Continuous numerical variables were discretized into quartile-based bins using equal-width binning. For text data, sentiment scores were discretized into three categories: negative, neutral, and positive.

The methodological alignment between Knowledge Discovery in Databases (KDD) and this investigation centres

on extracting meaningful insights from complex information structures. The KDD framework encompasses sequential analytical phases, including information selection, preprocessing protocols, transformation mechanisms, pattern discovery, and interpretative analysis, which directly parallel this study's methodological architecture. This procedural symmetry manifests through corresponding research stages, incorporating data refinement, consolidation, structural optimization, and analytical procedures. Implementing KDD methodologies enables systematic identification of underlying patterns and correlations within the dataset, facilitating a comprehensive understanding of the investigated phenomena. Integrating KDD principles enhances the research framework's capacity to process extensive datasets while maintaining analytical precision. This methodological synthesis establishes robust foundations for knowledge extraction, advancing theoretical frameworks and practical applications through systematically transforming raw information into actionable insights. Through this structured approach, the research leverages KDD's analytical power to ensure methodological rigor and enhance the validity of research outcomes, ultimately contributing to both theoretical advancement and practical implementation within the field.

## 2.2 SVM

The sentiment classification framework employs Support Vector Machine methodology, demonstrating notable effectiveness in processing multidimensional datasets and managing intricate classification boundaries. The SVM architecture establishes an optimal separating hyperplane that maximizes the margin between distinct sentiment categories. This mathematical framework facilitates precise differentiation among positive, negative, and neutral sentiment indicators within textual information structures [21]. This ability is particularly advantageous when dealing with textual datasets characterized by vast and sparse feature spaces, where traditional classification methods may struggle [22]. Furthermore, SVM's flexibility in employing different kernel functions allows it to model non-linear relationships in sentiment data, enhancing its classification accuracy [23]. The model's strength lies in its capability to generalize well from limited training data, making it a preferred choice for sentiment analysis tasks requiring high precision and reliability [24]. As a result, SVM's application in sentiment classification improves the efficiency of analyzing textual data. It contributes to more nuanced and actionable insights into consumer opinions, ultimately informing more effective decision-making strategies.

The implemented classification framework employs a supervised learning methodology, precisely the Support Vector Machine approach, renowned for its classification capabilities in multidimensional and sparse information structures. This algorithmic framework uses optimization principles that identify optimal separation boundaries between categorical elements, maximizing the discriminative margin to enhance predictive robustness. The methodology implements dimensional transformation procedures through kernel functionalities, facilitating non-linear relationship analysis within the feature space. The framework incorporates kernel implementations, including linear, polynomial, and Gaussian transformations. For this analytical context, the investigation employed the Gaussian Radial Basis Function transformation, selected for its capacity to address complex non-linear patterns

inherent in sentiment-based textual analysis. The decision enhances the framework's capability to capture intricate relationships within the analyzed sentiment parameters.

The SVM framework demonstrates multiple analytical advantages in sentiment-based classification scenarios. The methodology performs better in processing high-dimensional feature representations, particularly when analyzing textual data transformed through TF-IDF vectorization protocols. Furthermore, this analytical approach maintains computational stability against overfitting phenomena, especially in contexts where feature dimensions exceed sample quantities. The framework's adaptability through diverse kernel implementations enables optimal analysis across varying data structures and relationship patterns. Lastly, SVM demonstrates strong performance even with limited training data, making it suitable for applications like sentiment analysis, where labeled data might be sparse. This research applied SVM to classify travel vlog reviews into positive and negative sentiments. Its capability to distinguish subtle differences in textual expressions enables a more nuanced analysis of user-generated content, providing valuable insights into consumer opinions. By leveraging SVM, the study ensures accurate sentiment classification, contributing to a deeper understanding of digital consumer behavior in the tourism industry.

The relevance of SVM in this research is underscored by its ability to efficiently handle high-dimensional data and its robustness in classification tasks, particularly in contexts where clear margins of separation between classes are crucial. SVM's methodological strength lies in its capacity to construct optimal hyperplanes that maximize the margin between different classes. It is particularly pertinent for analyzing complex datasets where patterns are not easily discernible [25]. This characteristic makes SVM an ideal choice for tasks such as sentiment analysis or any domain requiring nuanced differentiation between closely related categories [26]. Additionally, SVM's flexibility in utilizing various kernel functions allows it to model non-linear relationships effectively, thereby enhancing the model's accuracy and generalization capability [27]. This is particularly advantageous when the research involves intricate and non-linear data distributions, as it facilitates the extraction of more profound insights. By integrating SVM into the analytical framework, the research gains a powerful tool that improves classification accuracy and contributes to a more refined understanding of the underlying patterns within the data, ultimately enhancing the validity and impact of the findings.

## 2.3 Optimizing classification models using SMOTE

Optimizing classification models using the SMOTE is a pivotal strategy for addressing the inherent challenges of imbalanced datasets in machine learning. Imbalanced data, where one class significantly outnumbers others, often leads to biased models that perform poorly on the minority class, skewing predictions and reducing overall model efficacy [28]. SMOTE mitigates this issue by synthetically generating new examples for the minority class, creating a more balanced dataset without merely replicating existing samples [29]. This process enhances the model's ability to learn from both classes more effectively and prevents overfitting, which can occur when models are trained on duplicated data. The synthetic samples generated by SMOTE introduce variability and robustness, which are crucial for the model's generalizability

and accuracy. By redistributing the data balance, SMOTE enables the development of more reliable and fair classification models, thus significantly contributing to the advancement of machine learning applications in fields where class imbalance is a common obstacle.

The SMOTE addresses the challenges associated with imbalanced datasets by generating synthetic samples for the minority class, thus balancing the dataset and improving model performance. SMOTE operates by interpolating between existing samples of the minority class, creating new data points that are not exact duplicates but are generated within the feature space. Pseudocode:

```

Input: Minority class samples X_minority, number of
synthetic samples N, number of neighbors k
Output: Augmented dataset with synthetic samples
1. Initialize an empty list to store synthetic samples: S = []
2. For each sample x in X_minority:
    a. Find k-nearest neighbors of x in X_minority
    b. For each required synthetic sample (N/k times):
        i. Randomly select a neighbor x_neighbor from k-
nearest neighbors
        ii. Compute the difference vector: diff = x_neighbor
- x
        iii. Generate a random value: rand_val ∈ [0, 1]
        iv. Create a synthetic sample: synthetic_sample = x
+ rand_val * diff
        v. Append synthetic_sample to S
3. Return the augmented dataset: X_augmented =
X_minority ∪ S

```

This pseudocode demonstrates how SMOTE generates synthetic samples systematically and integrates them into the original dataset. The SMOTE implementation in this research used the Python imbalanced-learn library with the following parameters: `sampling_strategy=0.5`, `k_neighbors=5`, and `random_state=42`. This ensured a balanced representation of the minority class, resulting in a more equitable training dataset.

The SMOTE offers several distinct advantages, making it an effective solution for dealing with class imbalance in machine learning. One of its primary strengths lies in its ability to generate synthetic samples for the minority class. It enhances the model's learning process by providing a more balanced training set without duplicating existing data points [30-32]. This approach mitigates the risk of overfitting and ensures that the minority class is well represented, allowing the model to develop a more nuanced understanding of all classes. Additionally, SMOTE introduces variability into the minority class by creating new instances that are interpolations of existing ones, thereby enriching the dataset with diverse scenarios that the model must learn to handle [33-36]. This diversity is critical in preventing the model from becoming overly simplistic and biased toward the majority class, ultimately leading to more robust and generalizable predictions. By addressing the challenges posed by imbalanced data with such precision and efficiency, SMOTE proves to be an invaluable tool in enhancing the performance and fairness of machine learning models across various applications [37-40].

The experimental design systematically evaluates the performance of the SVM model under varying hyperparameter settings to identify the optimal configuration for sentiment classification. (1) The kernel type was tested using linear,

polynomial, and Radial Basis Function (RBF) kernels, with the RBF kernel outperforming the others by achieving an accuracy of 83.12% and an AUC of 0.978 due to its ability to capture non-linear relationships in the feature space. (2) The regularization parameter (C), which controls the trade-off between minimizing training error and model complexity, was varied from 0.01 to 100. A moderate value of  $C=1.0$  provided the best balance between underfitting and overfitting, improving generalization and stable validation performance. (3) The gamma parameter for the RBF kernel, which determines the influence of individual training examples, was adjusted between 0.001 and 1.0. An optimal gamma value of 0.1 was identified, enabling the model to capture data patterns effectively without overfitting, resulting in a recall of 97.53% and precision of 75.76%. (4) The SMOTE sampling strategy, defined as the percentage of synthetic samples generated for the minority class, was evaluated across a range of 0.1 to 1.0. A sampling strategy of 0.5 was found to achieve the best balance, significantly reducing false negatives and improving minority class recognition. (5) The optimal combination of parameters was determined to be an RBF kernel with  $C=1.0$ ,  $\gamma=0.1$ , and a SMOTE sampling strategy of 0.5. This configuration achieved the highest overall performance, including an F-measure of 85.26% and a micro-averaged AUC of 0.974. The systematic exploration of these parameters demonstrates the critical role of hyperparameter tuning in optimizing model performance, improving robustness, and providing insights into the model's behavior across different configurations.

### 3. CLASSIFICATION

#### 3.1 SVM without SMOTE

The performance metrics of the SVM model without applying the SMOTE indicate high accuracy yet reveal significant limitations in handling imbalanced datasets. The model achieved an accuracy of 95.00% with a micro-average and a precision of 96.20%, suggesting its effectiveness in correctly predicting the majority class. However, the confusion matrix shows a substantial misclassification of the minority class, with only four true negatives and 57 false negatives, indicating the model's difficulty in identifying the minority class correctly. Furthermore, while the recall rate for the positive class is notably high at 98.70%, the AUC metrics present a concerning disparity. The optimistic AUC of 0.886 contrasts sharply with the pessimistic AUC of 0.239, reflecting a significant variance in the model's ability to distinguish between the classes under different scenarios. This disparity underscores the model's inherent bias toward the majority class, highlighting a need for balancing techniques like SMOTE to enhance its robustness and generalization capabilities across all classes. Therefore, while the SVM model demonstrates strong performance metrics on the surface, the underlying imbalance significantly affects its predictive reliability, emphasizing the importance of addressing data imbalance for more equitable and accurate model outcomes.

The reported AUC of 0.563, with a micro-average of 0.553, suggests a performance marginally better than random guessing, indicating significant room for improvement in the model's discriminatory power. The corresponding Receiver Operating Characteristic (ROC) curve demonstrates a lack of consistency, with sharp drops that reflect the model's struggles

with classification thresholds, particularly when distinguishing between the positive and negative classes. The considerable variance represented by the shaded areas further highlights instability in the model's predictive capabilities across different thresholds. Such results underscore the impact of imbalanced data on model performance, where the minority class is inadequately represented, leading to biased predictions. The observed AUC performance suggests the necessity for advanced balancing techniques like SMOTE, which could enhance the model's ability to learn from minority class examples, ultimately leading to a more robust and accurate classification performance.

Evaluating the SVM model without applying the SMOTE reveals significant limitations, primarily due to the challenges posed by data imbalance. Although the model achieves a high overall accuracy of 95.00%, a closer examination of the confusion matrix and class-specific metrics indicates substantial misclassification of the minority class. Specifically, the model produces only four true negatives while misclassifying 57 instances as false negatives, demonstrating a pronounced bias toward the majority class. The poor performance of the SVM model underscores the critical need for balancing techniques like SMOTE. By generating synthetic samples for the minority class, SMOTE addresses the skew in class distribution, enabling the model to learn the decision boundaries of both classes more effectively. This approach mitigates the bias introduced by data imbalance and enhances the model's ability to generalize across all classes, leading to more reliable and equitable classification outcomes. In summary, the observed limitations in the SVM model without SMOTE highlight the pivotal role of data balancing in ensuring robust sentiment classification, particularly in contexts where accurate representation of minority classes is essential.

### 3.2 SVM with SMOTE

Implementing the SVM with SMOTE demonstrates a significant shift in performance metrics, particularly in the model's ability to handle imbalanced datasets. While the accuracy slightly decreases to 83.12%, this is compensated by a dramatic improvement in the model's discriminatory capabilities, as reflected by the AUC values optimistic, standard, and pessimistic, all of which approach near-perfect levels, with micro averages of 0.978, 0.974, and 0.969, respectively. This suggests that the SVM model, enhanced with SMOTE, is far more effective in distinguishing between positive and negative classes across various decision thresholds. The precision, at 75.76%, while lower than traditional accuracy measures, indicates a more balanced approach to classification, where both classes receive fair consideration. Moreover, the recall rate of 97.53% highlights the model's improved sensitivity to the positive class, significantly reducing the occurrence of false negatives. The F-measure, at 85.26%, further corroborates the model's enhanced overall performance by balancing precision and recall. This comprehensive improvement illustrates the crucial role of SMOTE in addressing class imbalance, ultimately leading to a more equitable and accurate model, especially in applications where the cost of misclassification is high.

The AUC values, with an optimistic average of 0.978, a standard average of 0.974, and a pessimistic average of 0.969, all approaching 1.0, reflect the model's superior performance in correctly distinguishing between the positive and negative

classes across different thresholds. This high AUC demonstrates that when augmented with SMOTE, the SVM effectively addresses the class imbalance by enhancing the model's sensitivity and specificity, leading to more accurate predictions. The corresponding ROC curve displays a steep rise and a long plateau, signifying highly accurate positive rates with minimal false positives. Such consistency in the ROC profile across varying thresholds underscores the model's robustness and reliability in real-world applications where balanced classification is crucial. The improved AUC values, therefore, highlight the efficacy of SMOTE in mitigating bias towards the majority class, facilitating a more equitable evaluation of both classes and ultimately enhancing the model's overall predictive performance.

Implementing the SMOTE significantly enhanced the performance of the SVM model, particularly in recognizing the minority class. A comparison of the confusion matrices before and after applying SMOTE demonstrates its quantitative impact. Without SMOTE, the confusion matrix shows 1,424 true positives (TP), four true negatives (TN), two false positives (FP), and 57 false negatives (FN). After applying SMOTE, the confusion matrix reflects notable improvements with 1,424 true positives (TP), 1,003 true negatives (TN), 57 false positives (FP), and 36 false negatives (FN). These changes illustrate a more balanced classification, as evidenced by improved recall and precision for the minority class. Specifically, the recall rate increased from 96.10% to 97.53%, while the F-measure rose to 85.26%, demonstrating the model's enhanced generalization ability. These improvements underscore the effectiveness of SMOTE in addressing class imbalance and mitigating bias toward the majority class. By enabling equitable learning across all classes, SMOTE significantly enhances the reliability and applicability of machine learning models in real-world scenarios, especially in sentiment analysis tasks involving imbalanced datasets.

## 4. RESULT

### 4.1 SVM performance evaluation in sentiment classification

The performance evaluation of SVM in sentiment classification reveals its strengths and limitations when dealing with varying class distributions. SVM's robust algorithmic structure allows it to construct optimal hyperplanes for separating sentiment classes, which is particularly beneficial in high-dimensional feature spaces typical of text data. However, its effectiveness is highly contingent on the balance of the dataset, as imbalanced data can lead to a bias towards the majority class, diminishing its ability to classify minority class sentiments accurately. This limitation becomes evident when examining precision, recall, and F-measure metrics, which fluctuate significantly depending on class distribution. Introducing data balancing techniques such as the SMOTE substantially enhances SVM's classification performance by generating synthetic samples for the minority class, mitigating the skew, and improving the model's ability to generalize across different sentiment classes. Consequently, while SVM exhibits strong baseline performance in sentiment classification, its efficacy is significantly bolstered when complemented by strategies that address data imbalance, leading to more reliable and equitable



outcomes in sentiment analysis tasks.

SVM achieves optimal performance when paired with the SMOTE, as evidenced by various performance metrics highlighting classification accuracy and balance improvements. With an overall accuracy of 83.12%, the SVM model shows a notable enhancement in its ability to correctly classify positive and negative sentiments, reflected in the confusion matrix where 1,003 true negatives and 1,424 true positives were accurately identified. This improvement is further substantiated by the AUC scores, with an optimistic value of 0.978, a standard AUC of 0.974, and a pessimistic AUC of 0.969, all demonstrating the model's robust discriminative power across different scenarios. Additionally, the recall rate of 97.53% indicates a high sensitivity in identifying positive class instances, while the precision of 75.76% suggests a balanced approach to reducing false positives. The F-measure, at 85.26%, effectively balances precision and recall, comprehensively evaluating the model's performance. These metrics collectively illustrate that integrating SMOTE into the SVM framework significantly enhances the model's capability to handle imbalanced data, resulting in more accurate and equitable sentiment classification outcomes.

The combined use of SVM and SMOTE in this research offers distinct advantages, particularly in enhancing the model's effectiveness in handling imbalanced datasets. SVM is renowned for its ability to construct optimal hyperplanes that separate classes in high-dimensional spaces, making it a powerful tool for classification tasks where clear boundaries are required [41-46]. However, imbalanced data often challenge its performance, where the model may become biased towards the majority class. SMOTE addresses this limitation by generating synthetic examples for the minority class, thereby balancing the class distribution and allowing SVM to learn the characteristics of both classes better. This synergy between SVM's precise classification capabilities and SMOTE's data balancing approach enhances the overall model performance, reflected in improved accuracy, precision, recall, and AUC scores. Such a methodological combination mitigates the risk of misclassification and ensures more reliable and generalizable outcomes, making it particularly valuable in domains where data imbalance is prevalent. Therefore, the integration of SVM and SMOTE within this research framework underscores a strategic approach to leveraging the strengths of both techniques, resulting in a more robust and equitable analytical model.

SVM demonstrates several limitations and opportunities for optimization when applied to sentiment classification tasks. While SVM effectively identifies an optimal hyperplane for separating classes, its performance can decline when encountering datasets with complex, non-linear relationships or high variability [47-50]. This decline arises from its reliance on fixed kernel functions, which can limit its adaptability to diverse data distributions. For instance, in sentiment classification, the intricate nuances and contextual dependencies in textual data pose significant challenges to the model's capacity to generalize beyond the training set, particularly for minority class instances. Additionally, SVM is sensitive to noise and outliers in datasets, which can disproportionately affect the decision boundary. This issue is especially pronounced in sentiment analysis, where user-generated content often features informal language, spelling errors, and ambiguous expressions, potentially compromising the model's robustness and predictive reliability without

adequate noise-handling mechanisms. Furthermore, as dataset sizes increase, the computational cost of training SVM rises considerably, particularly when employing non-linear kernels. This scalability limitation hinders its applicability to large sentiment datasets, necessitating optimization strategies or alternative methods.

Addressing these challenges involves exploring optimization spaces that enhance SVM's effectiveness. Selecting alternative kernel functions or hybrid methods can improve its adaptability to complex data structures. Automated hyperparameter optimization techniques like grid search or Bayesian optimization may further refine kernel and parameter configurations. Advanced text preprocessing methods, including embedding representations like Word2Vec or BERT, can mitigate the impact of noisy data, while feature engineering approaches, such as sentiment lexicons or domain-specific features, provide a richer context for the model. Combining SVM with other machine learning models in ensemble frameworks, such as boosting or bagging, can bolster generalization ability and robustness by mitigating sensitivity to noise and outliers. To address scalability concerns, distributed computing frameworks or parallelized implementations of SVM can facilitate its application to larger datasets without excessive computational overhead. By addressing these limitations and exploring these optimization avenues, the utility of SVM in sentiment classification can be significantly enhanced, improving its generalization and robustness while broadening its applicability across diverse real-world scenarios.

## 5. CONCLUSION

The conclusion of this research underscores the effectiveness of integrating SVM with SMOTE in enhancing the accuracy and reliability of classification models, particularly in the context of imbalanced datasets. Based on a dataset comprising 1,928 text entries, the findings reveal that SVM is highly effective in constructing optimal hyperplanes for distinguishing between classes. However, its performance is significantly compromised in scenarios with unequal class distribution. By incorporating SMOTE, which generated additional synthetic samples to balance the dataset, the model's sensitivity to minority classes was substantially improved, leading to a more balanced and nuanced understanding of the data. This combination not only improves key performance metrics such as accuracy (83.12%), precision (75.76%), recall (97.53%), and AUC scores (up to 0.978) but also contributes to the robustness and generalizability of the model across different applications. The research demonstrates that addressing data imbalance through advanced techniques like SMOTE is crucial for achieving equitable and accurate outcomes in machine learning tasks. Thus, the study provides a valuable framework for future research and practical applications in fields that frequently encounter challenges with imbalanced data, advocating for a more strategic and integrated approach to model optimization.

The research demonstrates the effectiveness of integrating an SVM with the SMOTE to address class imbalance challenges in sentiment classification. Significant improvements in model performance, including enhanced accuracy, precision, recall, and AUC metrics, underline the importance of data balancing techniques in achieving equitable and reliable classification outcomes. Building on

these findings, several future directions are proposed to enhance the research's continuity and extensibility. Expanding the data scale by incorporating more extensive and diverse datasets, including multi-lingual and culturally varied user reviews, could improve the generalizability and robustness of the results. Additionally, exploring alternative data balancing methods, such as Adaptive Synthetic Sampling (ADASYN), Borderline-SMOTE, or hybrid approaches, could provide comparative insights into their effectiveness. Extending sentiment analysis to other domains, such as healthcare, education, and finance, would demonstrate the methodology's versatility and practical utility. Integrating advanced machine learning models, such as deep learning techniques like BERT or LSTM, could further enhance the ability to capture contextual and sequential dependencies in text data, improving performance in complex tasks. Moreover, developing real-time sentiment analysis systems for monitoring customer feedback or analyzing social media trends would bridge the gap between theoretical research and practical implementation. These proposed directions address current limitations while paving the way for broader applications and innovative solutions in data-driven decision-making.

## ACKNOWLEDGMENTS

Gratitude is extended to Indonesia's Directorate of Research, Technology, and Community Service (DRTPM) for the grant awarded under decree number 0375.20/III/LPPM-PM.10.03.01/06/2024. We also express our appreciation to LLDIKTI 3, the Faculty of Business Administration and Communication, the Faculty of Law, LPPM, and Atma Jaya Catholic University of Indonesia for their invaluable contributions.

## REFERENCES

- [1] Zhao, C., Yan, Z., Sun, X., Wu, M. (2024). Enhancing aspect category detection in imbalanced online reviews: An integrated approach using Select-SMOTE and LightGBM. *International Journal of Intelligent Networks*, 5: 364-372. <https://doi.org/10.1016/j.ijin.2024.10.002>
- [2] Zhang, R., Qi, Y., Kong, S., Wang, X., Li, M. (2024). A hybrid artificial intelligence algorithm for fault diagnosis of hot rolled strip crown imbalance. *Engineering Applications of Artificial Intelligence*, 130: 107763. <https://doi.org/10.1016/j.engappai.2023.107763>
- [3] Yilmaz Eroglu, D., Pir, M.S. (2024). Hybrid oversampling and undersampling method (houm) via safe-level smote and support vector machine. *Applied Sciences*, 14(22): 10438. <https://doi.org/10.3390/app142210438>
- [4] Sabri, N.M., Muhamad Subki, S.N.A., Bahrin, U.F.M., Puteh, M. (2024). Post Pandemic Tourism: Sentiment Analysis using Support Vector Machine Based on TikTok Data. *International Journal of Advanced Computer Science & Applications*, 15(2): 323-330. <https://doi.org/10.14569/IJACSA.2024.0150234>
- [5] Qin, J. (2023). Analysis of factors influencing the image perception of tourism scenic area planning and development based on big data. *Applied Mathematics and Nonlinear Sciences*, 9(1): 1-16. <https://doi.org/10.2478/amns.2023.1.00486>
- [6] Liu, X. (2024). Tourism destination recommendation based on bag of visual word combined with SVM classification. *Informatika*, 48(17): 6431. <https://doi.org/10.31449/inf.v48i17.6431>
- [7] Hariyono, H., Wibawa, A.P., Noviani, E.F., Laurretta, G.C., Citra, H.R., Utama, A.B.P., Dwiyanto, F.A. (2024). Exploring visitor sentiments: A study of Nusantara temple reviews on tripadvisor using machine learning. *Journal of Applied Data Sciences*, 5(2): 600-612. <https://doi.org/10.47738/jads.v5i2.208>
- [8] Gu, Y. (2024). Decision-making support platform and security design for rural leisure tour industry based on SEA method and SVR model. *Scalable Computing: Practice and Experience*, 25(4): 3086-3099. <https://doi.org/10.12694/scpe.v25i4.2888>
- [9] Gu, H. (2023). The use of red cultural and creative design in tourism cultural and creative design based on the background of big data. *Applied Mathematics and Nonlinear Sciences*, 9(1): 00179. <https://doi.org/10.2478/amns.2023.2.00179>
- [10] Chang, V., Islam, M.R., Ahad, A., Ahmed, M.J., Xu, Q.A. (2024). Machine learning for predicting tourist spots' preference and analysing future tourism trends in Bangladesh. *Enterprise Information Systems*, 18(12): 2415568. <https://doi.org/10.1080/17517575.2024.2415568>
- [11] Bhardwaj, M., Mishra, P., Badhani, S., Muttoo, S.K. (2024). Sentiment analysis and topic modeling of COVID-19 tweets of India. *International Journal of System Assurance Engineering and Management*, 15(5): 1756-1776. <https://doi.org/10.1007/s13198-023-02082-0>
- [12] Alemerien, K., Al-Ghareeb, A., Alksasbeh, M.Z. (2024). Sentiment analysis of online reviews: a machine learning based approach with TF-IDF vectorization. *Journal of Mobile Multimedia*, 20(5): 1089-1116. <https://doi.org/10.13052/jmm1550-4646.2055>
- [13] Waqas, S., Harun, N.Y., Arshad, U., Laziz, A.M., et al. (2024). Optimization of operational parameters using RSM, ANN, and SVM in membrane integrated with rotating biological contactor. *Chemosphere*, 349, 140830. <https://doi.org/10.1016/j.chemosphere.2023.140830>
- [14] Prabhavathy, T., Elumalai, V.K., Balaji, E. (2024). Hand gesture classification framework leveraging the entropy features from sEMG signals and VMD augmented multi-class SVM. *Expert Systems with Applications*, 238: 121972. <https://doi.org/10.1016/j.eswa.2023.121972>
- [15] Feng, F., Ghorbani, H., Radwan, A.E. (2024). Predicting groundwater level using traditional and deep machine learning algorithms. *Frontiers in Environmental Science*, 12: 1291327. <https://doi.org/10.3389/fenvs.2024.1291327>
- [16] Dong, J., Wang, Z., Wu, J., Cui, X., Pei, R. (2024). A novel runoff prediction model based on support vector machine and gate recurrent unit with secondary mode decomposition. *Water Resources Management*, 38(5): 1655-1674. <https://doi.org/10.1007/s11269-024-03748-5>
- [17] Demilie, W.B. (2024). Plant disease detection and classification techniques: A comparative study of the performances. *Journal of Big Data*, 11(1): 5. <https://doi.org/10.1186/s40537-023-00863-9>
- [18] Daniel, C., Khatti, J., Grover, K.S. (2024). Assessment of compressive strength of high-performance concrete using soft computing approaches. *Computers and Concrete*, 33(1): 55-75.



- <https://doi.org/10.12989/cac.2024.33.1.055>
- [19] Binson, V.A., Thomas, S., Subramoniam, M., Arun, J., Naveen, S., Madhu, S. (2024). A review of machine learning algorithms for biomedical applications. *Annals of Biomedical Engineering*, 52(5): 1159-1183. <https://doi.org/10.1007/s10439-024-03459-3>
- [20] Alhaj, Y.A., Xiang, J., Zhao, D., Al-Qaness, M.A., Abd Elaziz, M., Dahou, A. (2019). A study of the effects of stemming strategies on Arabic document classification. *IEEE Access*, 7: 32664-32671. <https://doi.org/10.1109/ACCESS.2019.2903331>
- [21] Zuhairi, A.H., Yakub, F., Omar, M., Sharifuddin, M., Razak, K.A., Faruq, A. (2024). Imbalanced flood forecast dataset resampling using SMOTE-tomek link. *International Exchange and Innovation Conference on Engineering & Sciences*, 10: 845-850. <https://doi.org/10.5109/7323359>
- [22] Puspitasari, D., Aprian, A.J., Sikumbang, E.D., Ramanda, K., Sukmana, S.H., Azizah, Q.N. (2024). Heart disease: Application of the k-nearest neighbor (KNN) method. *Ingenierie des Systemes d'Information*, 29(4): 1275. <https://doi.org/10.18280/isi.290403>
- [23] Zeng, X., Shahzeb, M., Cheng, X., Shen, Q., et al. (2024). An enhanced gas sensor data classification method using principal component analysis and synthetic minority over-sampling technique algorithms. *Micromachines*, 15(12): 1501. <https://doi.org/10.3390/mi15121501>
- [24] Tran, T.O., Le, N.Q.K. (2024). SA-TTCA: An SVM-based approach for tumor t-cell antigen classification using features extracted from biological sequencing and natural language processing. *Computers in Biology and Medicine*, 174: 108408. <https://doi.org/10.1016/j.compbimed.2024.108408>
- [25] Swarnalatha, K., Narisetty, N., Kancherla, G.R., Bobba, B. (2024). Analyzing resampling techniques for addressing the class imbalance in NIDS using SVM with random forest feature selection. *International Journal of Experimental Research and Review*, 43: 42-55. <https://doi.org/10.52756/ijerr.2024.v43spl.004>
- [26] Singh, N., Tripathi, P. (2023). Efficient model for prediction of parkinson's disease using machine learning algorithms with hybrid feature selection methods. In *Biomedical Engineering Science and Technology: Second International Conference, Raipur, India*, pp. 186-203. [https://doi.org/10.1007/978-3-031-54547-4\\_15](https://doi.org/10.1007/978-3-031-54547-4_15)
- [27] Santiago-Gonzalez, F., Martinez-Rodriguez, J.L., García-Perez, C., Juárez-Saldivar, A., Camacho-Cruz, H.E. (2024). Hybrid class balancing approach for chemical compound toxicity prediction. *Current Computer-Aided Drug Design*. <https://doi.org/10.2174/0115734099315538240909101737>
- [28] Ndama, O., Bensassi, I., En-Naimi, E.M. (2024). Integrating artificial neural networks and support vector machines machine learning algorithms for advanced credit card fraud detection. In *Modern Artificial Intelligence and Data Science 2024: Tools, Techniques and Systems*, pp. 453-461. [https://doi.org/10.1007/978-3-031-65038-3\\_36](https://doi.org/10.1007/978-3-031-65038-3_36)
- [29] Mujahid, M., Kina, E.R.O.L., Rustam, F., Villar, M.G., Alvarado, E.S., De La Torre Diez, I., Ashraf, I. (2024). Data oversampling and imbalanced datasets: an investigation of performance for machine learning and feature engineering. *Journal of Big Data*, 11(1): 87. <https://doi.org/10.1186/s40537-024-00943-4>
- [30] Khotimah, B.K., Setiawan, E., Anamisa, D.R., Puspitarini, O.R., Rachmad, A. (2024). Least squares support vector machine ensemble based on sampling for classification of quality local cattle. *Communications in Mathematical Biology and Neuroscience*, 2024: 113. <https://doi.org/10.28919/cmbn/8838>
- [31] Herianto, H., Kurniawan, B., Hartomi, Z.H., Irawan, Y., Anam, M.K. (2024). Machine learning algorithm optimization using stacking technique for graduation prediction. *Journal of Applied Data Sciences*, 5(3): 1272-1285. <https://doi.org/10.47738/jads.v5i3.316>
- [32] Zhang, S., Chen, C., Chen, C., Chen, F., Li, M., Yang, B., Yan, Z., Lv, X. (2021). Research on application of classification model based on stack generalization in staging of cervical tissue pathological images. *IEEE Access*, 9: 48980-48991. <https://doi.org/10.1109/ACCESS.2021.3064040>
- [33] Chou, E.P., Yang, S.P. (2024). A virtual multi-label approach to imbalanced data classification. *Communications in Statistics-Simulation and Computation*, 53(3): 1461-1471. <https://doi.org/10.1080/03610918.2022.2049820>
- [34] Al-Azani, S., Alkhnbashi, O.S., Ramadan, E., Alfarraj, M. (2024). Gene expression-based cancer classification for handling the class imbalance problem and curse of dimensionality. *International Journal of Molecular Sciences*, 25(4): 2102. <https://doi.org/10.3390/ijms25042102>
- [35] Barid, A.J., Hadiyanto, H., Wibowo, A. (2024). Optimization of the algorithms use ensemble and synthetic minority oversampling technique for air quality classification. *Indonesian Journal of Electrical Engineering and Computer Science*, 33(3): 1632-1640. <https://doi.org/10.11591/ijeecs.v33.i3.pp1632-1640>
- [36] Fitriyani, N.L., Syafrudin, M., Alfian, G., Rhee, J. (2020). HDPM: An effective heart disease prediction model for a clinical decision support system. *IEEE Access*, 8: 133034-133050. <https://doi.org/10.1109/ACCESS.2020.3010511>
- [37] Atta Mills, E.F.E., Deng, Z., Zhong, Z., Li, J. (2024). Data-driven prediction of soccer outcomes using enhanced machine and deep learning techniques. *Journal of Big Data*, 11(1): 170. <https://doi.org/10.1186/s40537-024-01008-2>
- [38] Vives, L., Cabezas, I., Vives, J.C., Reyes, N.G., Aquino, J., Córdor, J.B., Altamirano, S.F.S. (2024). Prediction of Students' academic performance in the programming fundamentals course using long short-Term memory neural networks. *IEEE Access*, 12: 5882-5898. <https://doi.org/10.1109/ACCESS.2024.3350169>
- [39] Agustina, C., Purwanto, P., Farikhin, F. (2024). Enhancing sentiment analysis accuracy in borobudur temple visitor reviews through semi-supervised learning and SMOTE upsampling. *Journal of Advances in Information Technology*, 15(4): 492-499. <https://doi.org/10.12720/jait.15.4.492-499>
- [40] Gufroni, A.I., Hoeronis, I., Fajar, N., Rachman, A.N., Ramdani, C.M.S., Sulastri, H. (2024). Implementation of ensemble machine learning classifier and synthetic minority oversampling technique for sentiment analysis of sustainable development goals in Indonesia. *JOIV: International Journal on Informatics Visualization*, 8(2): 678-685. <https://doi.org/10.62527/joiv.8.2.1949>

- [41] Lubis, A., Irawan, Y., Junadhi, J., Defit, S. (2024). Leveraging k-nearest neighbors with smote and boosting techniques for data imbalance and accuracy improvement. *Journal of Applied Data Sciences*, 5(4): 1625-1638. <https://doi.org/10.47738/jads.v5i4.343>
- [42] Villar, A., de Andrade, C.R.V. (2024). Supervised machine learning algorithms for predicting student dropout and academic success: a comparative study. *Discover Artificial Intelligence*, 4(1): 2. <https://doi.org/10.1007/s44163-023-00079-z>
- [43] Mahesh, T.R., Bhardwaj, R., Khan, S.B., Alkhalidi, N.A., Victor, N., Verma, A. (2024). An artificial intelligence-based decision support system for early and accurate diagnosis of Parkinson's disease. *Decision Analytics Journal*, 10: 100381. <https://doi.org/10.1016/j.dajour.2023.100381>
- [44] Sun, P., Wang, Z., Jia, L., Xu, Z. (2024). SMOTE-kTLNN: A hybrid re-sampling method based on SMOTE and a two-layer nearest neighbor classifier. *Expert Systems with Applications*, 238: 121848. <https://doi.org/10.1016/j.eswa.2023.121848>
- [45] Shaikh, R., Modak, M., Ravale, U., Punjabi, S. (2024). Quasi opposition-based learning in beluga whale optimization feature selection approach for diabetes prediction in IoT system. *International Journal of Intelligent Engineering & Systems*, 17(6): 1347-1357. <https://doi.org/10.22266/ijies2024.1231.98>
- [46] Mohamed, R., Azizan, N.H., Perumal, T., Abd Manaf, S., Marlisah, E., Hardhienata, M.K.D. (2023). Discovering and recognizing of imbalance human activity in healthcare monitoring using data resampling technique and decision tree model. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, 33(2): 340-350. <https://doi.org/10.37934/araset.33.2.340350>
- [47] Morasco, B.J., Pal, N., Ono, S.S., McPherson, S.M., et al. (2024). Tele-collaborative outreach to rural patients with chronic pain: pragmatic effectiveness trial protocol for the CORPs study. *Pain Medicine*, 25(Supplement\_1): S91-S98. <https://doi.org/10.1093/pm/pnae075>
- [48] Sugiharti, E., Arifudin, R., Prasetyo, B., Muslim, M.A. (2024). Optimizing support vector machine performance for Parkinson's disease diagnosis using GridSearchCV and PCA-based feature extraction. *Journal of Information Systems Engineering & Business Intelligence*, 10(1): 38-50. <https://doi.org/10.20473/jisebi.10.1.38-50>
- [49] Ahsan, M.M., Ali, M.S., Siddique, Z. (2024). Enhancing and improving the performance of imbalanced class data using novel GBO and SSG: A comparative analysis. *Neural Networks*, 173: 106157. <https://doi.org/10.1016/j.neunet.2024.106157>
- [50] Al-Fuhaidi, B., Farae, Z., Al-Fahaidy, F., Nagi, G., Ghallab, A., Alameri, A. (2024). Anomaly-based intrusion detection system in wireless sensor networks using machine learning algorithms. *Applied Computational Intelligence and Soft Computing*, 2024(1): 2625922. <https://doi.org/10.1155/2024/2625922>