STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–9.
Published online in International Academic Press (www.IAPress.org)

Implementation of Logistic Regression and Naïve Bayes Methods in Analyzing Employee Sentiment at Dr. Iskak Hospital Regarding Organizational Performance

Kethut Chandra[1,*], Indasah[2], Ratna Wardani[2], Teguh Herlambang[3,4]

[1]*Postgraduate Student of Public Health Program, Universitas Strada Indonesia, Indonesia*
[2]*Department of Public Health, Universitas Strada Indonesia, Indonesia*
[3]*Department of Information Systems, Faculty of Economics Business and Digital Technology, Universitas Nahdlatul Ulama Surabaya, Indonesia*
[4]*Center of Excellence in Science and Technology for Digital Health, Universitas Nahdlatul Ulama Surabaya, Indonesia*

Abstract This study combines the McKinsey 7S model and machine learning-based classification techniques to study employee sentiments towards organization performance at Dr. Iskak Hospital, Tulungagung, East Java. The data were obtained by interviewing 56 employees and managers which can be classified into positive, neutral, and negative class by referring to 7S aspects in McKinsey model as previously defined. Term Frequency-Inverse Document Frequency (TF-IDF) is used for feature extraction in which it is implemented for translating words into numeric vectors. There are two classification methods namely Logistic Regression (LR) and Naive Bayes (NB) applied and compared under two simulations with variations in training-test dataset split (70:30 and 80:20) by applying precision, recall, F1-score and accuracy as evaluation measures. It can be concluded that both methods produced the same evaluation scores in the two simulations. In the first simulation, both methods yielded an accuracy rate of 71 percent, which increased to 75 percent in the second simulation. The results obtained from the implementation of these two methods provide a broader perspective on organizational climate, which will serve as valuable insights for data-driven decision-making and corrective actions within a healthcare organization.

Keywords Hospital, Logistic Regression, Naïve Bayes, Sentiment Analysis, Sustainable Management

DOI: 10.19139/soic-2310-5070-3699

1. INTRODUCTION

In the analysis of organizational performance—and more specifically, hospital performance—it is common practice to use conceptual models such as the McKinsey 7S Model to gain a comprehensive understanding of the components of strategy, structure, systems, style, staff, and so on. In this case, Dr. Iskak Hospital in Tulungagung, East Java. The McKinsey 7S Model was developed by Waterman, Peters, and Phillips of McKinsey & Company in the early 1980s as a framework for holistically analyzing organizational effectiveness. This model emphasizes that an organization's success is determined by the degree of alignment among seven key elements: strategy, structure, systems, shared values, skills, staff, and style [1]. The McKinsey 7S Model can analyze the extent to which each component supports the others in improving organizational efficiency. This model provides an in-depth examination of organizational structure, authority, hierarchy (structure), the organization's strategic direction (strategy), operational processes (systems), leadership approaches (style), the capabilities of the organization's employees (staff), and the values upheld by the organization's members (shared values). When applied to a hospital, as in the case of Dr. Iskak Hospital, this method helps researchers gain a comprehensive view of organizational

*Correspondence to: Kethut Chandra (Email: kethut.28@gmail.com).

ISSN 2310-5070 (online) ISSN 2311-004X (print)

performance because it examines both tangible aspects—such as organizational structure and systems—and intangible aspects, including the influence of leadership and organizational values, thereby providing a holistic perspective on the organization's health and sustainability [2].

For the utilization of machine learning (ML) approaches, the classifier which is mainly used for classifying qualitative data such as sentiments, which in this case obtained from the interview are Logistic Regression (LR) and Naive Bayes (NB). LR is a statistical regression model used to analyze the probability of a binary outcome based on one or more predictor variables. The LR model calculates the probability that a text belongs to a specific sentiment class by applying the logistic function to a linear combination of input features (e.g., word frequency, term frequency vectors), thereby modeling the probability that a text will be classified into a specific sentiment class. This model is based on the assumption that there is a linear relationship between the predictor variables and the log-odds of the outcome variable. It is often chosen because it is less sensitive to irrelevant input data or even outliers, and it handles correlated predictor variables better than basic probabilistic models. This model is capable of delivering reliable text classification performance when applied to the appropriate datasets and input features. Specifically, this model has been applied to analyze sentiment related to health and organizational issues, and has demonstrated an adequate level of classification accuracy in text analysis [3].

In contrast, NB is a probabilistic classifier that utilizes Bayes' theorem. NB classifies text based on the assumption of feature independence given the class label. NB calculates the posterior probability for each class based on the conditional probabilities of the features, which are typically word occurrences or frequencies in the text. Even with this simple assumption of independence, NB classification demonstrates surprising performance and speed in text classification; and for features with very high dimensions—such as customer review analysis and sentiment classification in social media posts and company reviews—NB classification performs optimally. The NB model is fast and does not require a large number of training samples. Several research results show that NB can achieve better accuracy, on par with LR, although it performs slightly worse in medical diagnosis and sentiment analysis when features exhibit high dependency [4].

In sentiment analysis, both LR and NB convert text data into numerical feature vectors by applying text vectorization approaches such as Bag of Words (BoW) or Term Frequency-Inverse Document Frequency (TF-IDF). The resulting features are then fed into the classifier so that it can learn the patterns associated with each sentiment class. Due to its ability to effectively optimize the decision boundary between classes, LR performs well on sentiment datasets that share the same vocabulary. NB is suitable for classifying text using probability estimates and can efficiently assign labels to each sentence, but it does not perform well when two words frequently appear together or have a specific relationship. Therefore, LR performs better than NB when dealing with complex sentiment words or correlated features; however, NB performs very well when the data has high dimensionality and simple characteristics, and it can reduce complexity while efficiently processing text data [5].

The combination of LR and NB offers complementary benefits in sentiment analysis, and when used alongside organizational models such as the McKinsey Organizational Model, the two can provide in-depth analysis by gathering sentiment data from various sources—such as interviews with employees and managers—thereby helping us gain a comprehensive understanding of the organization's climate and performance [6].

1.1. Focus of the Study

This study focuses on the implementation of the LR and NB methods to classify sentiment into three categories—positive, neutral, and negative—based on data from interviews with several staff members and managers at Dr. Iskak Hospital.

1.2. Contributions

Several contributions resulting from this study include:

1. Providing insights into the overall management and performance of Dr. Iskak Hospital based on the results of direct interviews with the staff and managers involved.
2. Implementing the LR and NB methods to obtain accurate classification results as a basis for data-driven evaluation and improvement.

1.3. Limitations

This study employs 56 data rows, based on the number of respondents involved. The respondents were staff members and managers from Dr. Iskak Hospital.

2. RELATED WORKS

In 2022, [7] conducted study regarding to classification of sentiment analysis through chat bot for counselling in mental health diagnoses. In this study, the authors classifying several statements into eight emotional states, ranging from neutral to unpleasant. The methods employed in this study are LR and NB. The data processed is obtained from 34,792 tweets. The result obtained from these methods states that the LR method achieves better accuracy that the NB method, which the LR method obtained 0.62, while the NB method obtained 0.56. Previously in 2021, [8] conducted study related to sentiment analysis of COVID-19 vaccine in Indonesia. This study utilized the data from Twitter/X platform in January 2021 and classified to positive and negative opinions. The method applied in this study is NB. By utilized 6000 tweets, the NB succeed to classify 39% of positive sentiment, 56% of negative sentiment, and 1% favorable sentiment. For more advanced, [9] conducted study about sentiment analysis and utilized the dataset from Twitter/X platform. This study combined both techniques, ML and DL. The dataset obtained 14,460 reviews and will be classified as positive, neutral, and negative. In this study also utilized TF-IDF, BoW, n-grams, and emoticon lexicon. Six classification methods were employed, such as SVM, NB, RF, LR, LSTM, and CNN. All ML techniques obtained 52%-78% in range of accuracy value, while all DL techniques reached 81%-89% in range of accuracy value. Based on several examples from previous studies, it was found that the LR and NB methods, as well as several basic NLP methods, have been applied in various case studies related to sentiment analysis. In this study, these two methods were reapplied to analyze employee sentiment within a healthcare organization, in this case Dr. Iskak Hospital, Tulungagung, East Java, which was categorized into several groups according to the McKinsey 7S framework. Based on the performance of these two methods, the accuracy levels achieved for the predefined classes will be analyzed next.

3. RESEARCH METHODS

3.1. Data Acquisition

As mentioned in the previous section, the data utilized in this study were obtained from an interview with several staff members and managers at Dr. Iskak Hospital in Tulungagung, East Java. A total of 56 responses were successfully collected and aligned with the McKinsey framework.The questions posed to respondents were limited to specific areas, namely coordination, procedures, communication, services, accountability, governance, budgeting, services, data, systems, risk, controls, leadership, selection processes, rewards, competencies, and innovation. In the subsequent stages, the data will be classified into Positive and Neutral classes to assess sentiment according to the LR and NB methods.

3.2. Data Preprocessing

Data processing involving responses from interviews was conducted using an Indonesian lexicon to represent sentiment weights, where sentiment weights are divided into three categories—positive, neutral, and negative. These categories are then applied as labels for each response in every row of the data. Furthermore, the sentences containing opinions from respondents were converted into a numerical matrix using the TF-IDF approach. Table 1 below presents the classification of word meanings (positive and negative) based on the lexicon.

Table 1. Classification of Word Meanings

Word	Category
facilitate, clear, harmonious, accountability, discipline, excellent, balance, empowerment, competency, innovation, efficiency,	Positive
risk, obstacle, infect, not, not yet, impede, inadequate, grey	Negative

3.3. Sentiment Analysis

Sentiment can be defined as a human expression regarding an issue that involves emotional elements. The emotions in this context can vary quite widely, ranging from joy, sadness, disappointed, and so on. In the context of information technology, sentiment plays a distinct and significant role and can provide unique insights. ML processes this sentiment or opinion using a technique known as Natural Language Processing (NLP). In several recent studies, sentiment data was obtained from social media platforms and, in part, from websites [10].

3.4. Classification Methods and Evaluation Metrics

Previous studies have described various classification methods for sentiment analysis, ranging from traditional machine learning methods to deep learning-based approaches. When selecting a model, it is important to consider supporting factors such as the amount of the data and computational time.In this study, also utilized by TF-IDF. The main of this function is become statistical indicator that calculated importance of a word in information retrieval [11]. Adapted from [12], below in Eq. (1) is the formula of TF-IDF method.

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right) \tag{1}$$

where:
$tf_{i,j}$: The frequency of appearance of i in j
df_i : The number of documents containing i
N : The number of documents

Furthermore, regarding the classification methods used—namely LR and NB—both can be explained as follows, the LR method can be defined as the statistical model and its baseline form employ logistic function to model a binary response variable, while the NB techniques incorporate the independence property among the feature identified, whereas Bayesian networks are based on the dependency property of those features [13]. Below in Eq. (2) is the formula of NB method.

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \tag{2}$$

Description:
$P(Y|X)$: The likelihood of a condition-dependent hypothesis
$P(X|Y)P(Y)$: The previous probabilities of a class based on a hypothetical condition
$P(X)$: The probability of Y

Meanwhile, several methodological approaches were employed to evaluate the classification results, as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{5}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + TN + FP} \tag{6}$$

4. RESULTS AND DISCUSSION

4.1. Result of The Labelling Process

This subsection discusses the results of the labelling process based on responses obtained from interviews with staff and managers at Dr. Iskak Hospital. The several detail of these results are summarized in Table 2 below. Furthermore, in Fig. 1 depicts a diagram representing the distribution of the sentiment labels that were successfully identified.

Table 2. The Result of Labelling Process

Response	Label
The organizational structure has been established in accordance with a clear legal basis.	Positive
The structure is designed to ensure that service, management, and support functions work in harmony.	Positive
Undocumented flexibility can pose a risks.	Negative
Digitization as a necessity	Neutral

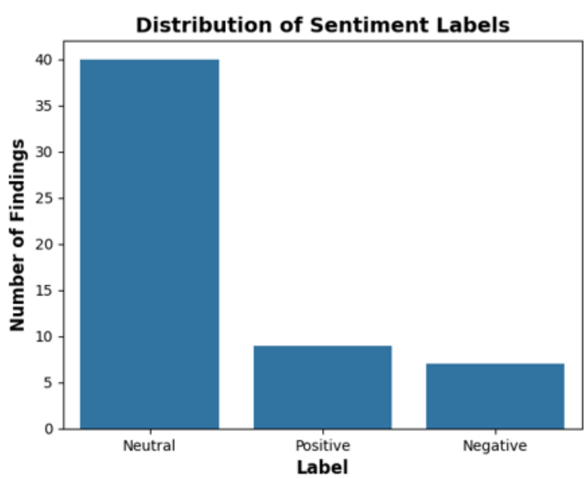

Figure 1. The Distribution of the Sentiment Labels

4.2. Result of Simulation

When processing data to be analysed using ML-based classification techniques, several adjustments must be performed on the hyper parameters to be used. The purpose is to achieve optimal result and also can be utilized as decision making parameters. Table 3 and 4 below provide details of the hyper parameter values, while Table 4 is adjustment of TF-IDF.

Table 3. Hyperparameter Tuning of Logistic Regression

Name of Hyperparameter	**Value**
max_iter	1000
class_weight	balanced

Table 4. Hyperparameter Tuning of Naive Bayes

Name of Hyperparameter	**Value**
Type	Multinomial

Table 5. Details of TF–IDF Hyperparameter Tuning

Name of Hyperparameter	**Value**
max_features	1000
ngram_range	(1,2)
stop_words	none

This section presents the results of a classification comparison between the two methods based on the previously defined hyper parameter values. Furthermore, the data was divided into 70% training data and 30% test data for the first simulation, then 80% training data and 20% test data for the second simulation. The visualizations are illustrated in Fig. 2 until 5 below.

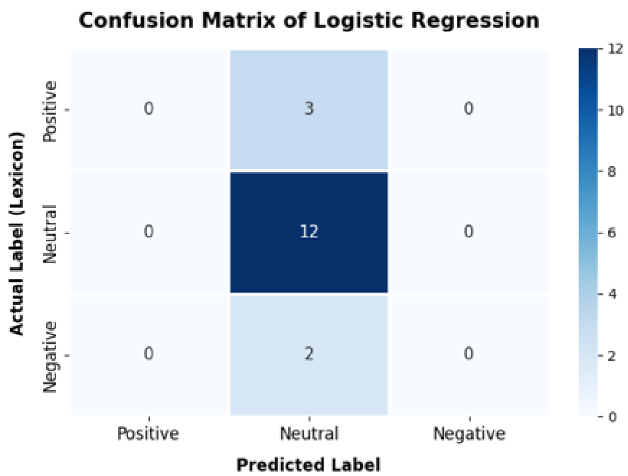

Figure 2. The Confusion Matrix of Logistic Regression (First Simulation)

Based on the Figures 2 and 3, it can be observed that the classification results of the two methods, represented as heat maps. Both methods produced the same output, with an accuracy of 71 percent. The number of True Positive (TP) and True Negative (TN) depicted in the diagrams for both methods is also the identical. Furthermore, Figures 4 and 5 present the classification results from both methods, which are not significantly different from those of the first simulation. In both figures, the two methods still produce identical outputs, with their accuracy increasing to 75 percent. Table 6 below provides details on the metric evaluations produced by two methods in the two simulations that were conducted.

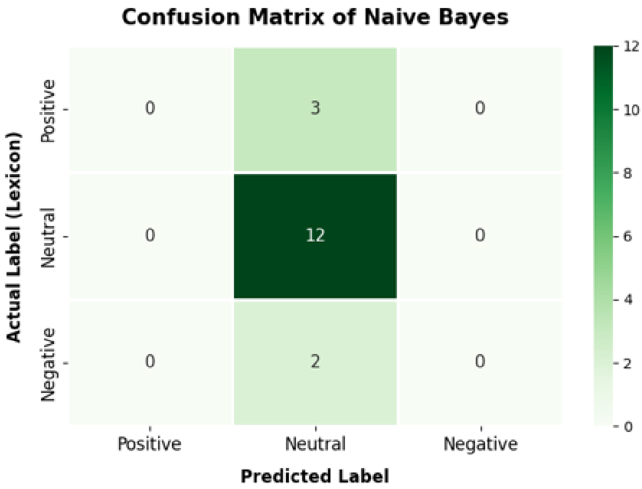

Figure 3. The Confusion Matrix of Naive Bayes (First Simulation)

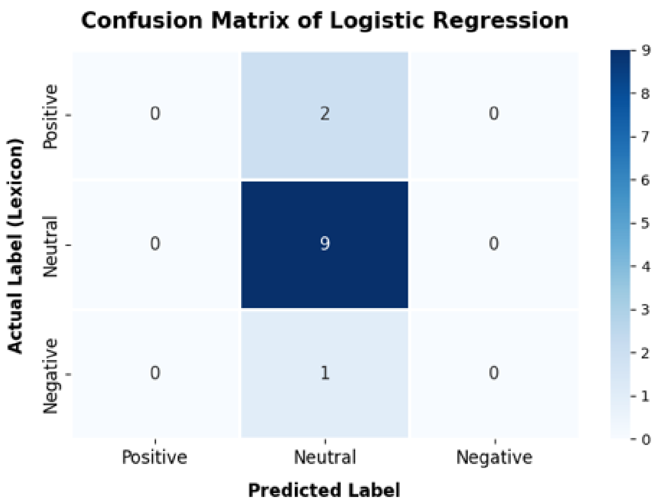

Figure 4. The Confusion Matrix of Logistic Regresion (Second Simulation)

Table 6 above provides details of the evaluation metrics from the classification simulations performed by both methods. As indicated in the table, the values of the other metrics (Precision, Recall, F1-score) obtained by both methods are identical. Given these results, a further test was conducted to assess the level of significance using McNemar test. Furthermore, Figures 6 and 7 above present the results of the McNemar test in the form of a matrix heatmap. This test was conducted to compare the TP, TN, FP, and FN values obtained by both methods. The test reveals that both methods again produced identical result, with a p-value of 1.00. Therefore, it can be concluded that there is no statistical significance or neither method is superior (a balanced condition).

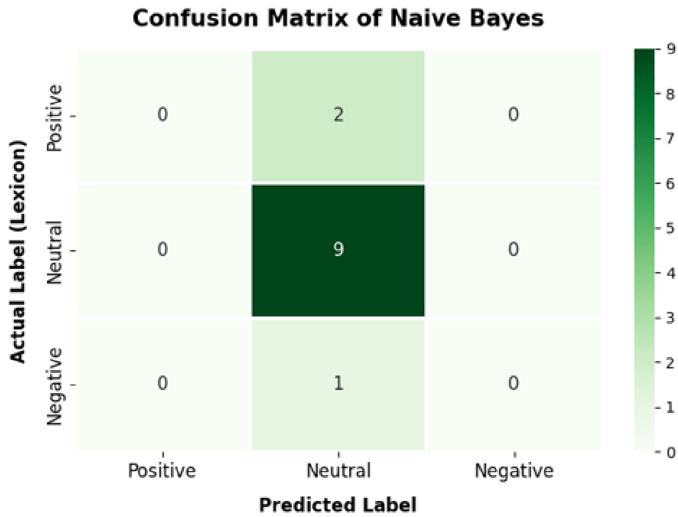

Figure 5. The Confusion Matrix of Naive Bayes (Second Simulation)

Table 6. The Result of Evaluation

Stage	Method	Class	Accuracy	Precision	Recall	F1-Score
I	LR	Negative	0.71	0.00	0.00	0.00
		Neutral		0.71	1.00	0.83
		Positive		0.00	0.00	0.00
	NB	Negative	0.71	0.00	0.00	0.00
		Neutral		0.71	1.00	0.83
		Positive		0.00	0.00	0.00
II	LR	Negative	0.75	0.00	0.00	0.00
		Neutral		0.75	1.00	0.86
		Positive		0.00	0.00	0.00
	NB	Negative	0.75	0.00	0.00	0.00
		Neutral		0.75	1.00	0.86
		Positive		0.00	0.00	0.00

5. CONCLUSIONS

Based on the results of the analyses and simulations that have been conducted, this study has achieved its overall objectives. Other important findings include the following:

1. The TF-IDF approach has a positive impact on classification results and is compatible with integration using the LR and NB methods.
2. The LR and NB methods yield identical evaluation metrics in both the first and second simulations.

Based on these results, further study will be conducted to expand the scope in terms of the number of respondents and the methods implemented.

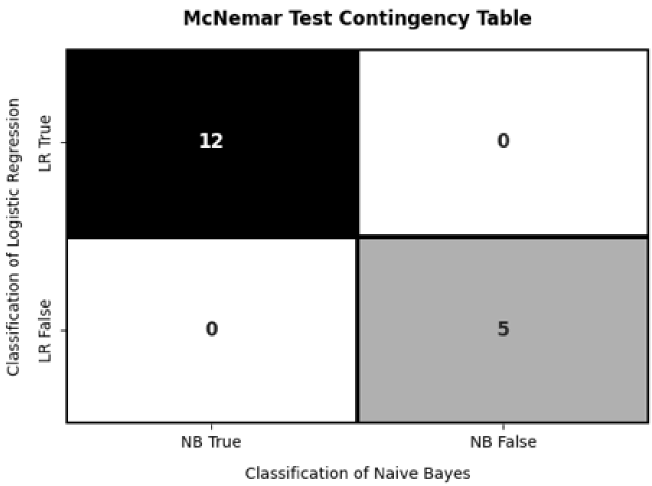

Figure 6. The McNemar Test Based on First Simulation

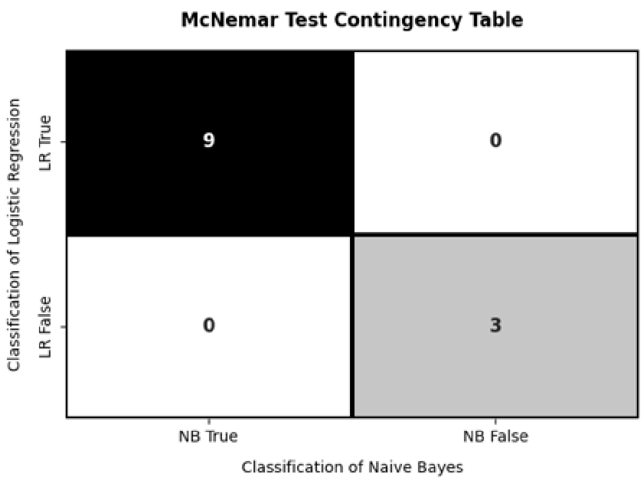

Figure 7. The McNemar Test Based on Second Simulation

COMPETING INTERESTS

The authors declare that they have no known competing financial interests or personal relationship that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY

The data utilized in this research is confidential.

DECLARATION OF GENERATIVE AI USE

In preparing and executing this research, the authors utilized Generative AI to assist with developing ideas and coding. The result obtained from this tool were subsequently processed and corrected to ensure accountability.

REFERENCES

1. M. Chmielewska, J. Stokwiszewski, J. Markowska, and T. Hermanowski, *Evaluating Organizational Performance of Public Hospitals using the McKinsey 7-S Framework*, BMC Health Services Research, vol. 22, no. 1, p. 7, Jan. 2022, doi: 10.1186/s12913-021-07402-3.
2. Silva, I. Maldonado, M. Da Silva, and C. Cepeda, *Sustainability Balanced Scorecard: Systematic Literature Review*, JRFM, vol. 18, no. 6, p. 324, June 2025, doi: 10.3390/jrfm18060324.
3. V. Mehta et al., *Machine Learning based Exploratory Data Analysis (EDA) and Diagnosis of Chronic Kidney Disease (CKD)*, EAI Endorsed Trans. Perv. Health Tech., vol. 10, Mar. 2024, doi: 10.4108/eetpht.10.5512.
4. B. Phatcharathada and P. Srisuradetchai, *Randomized Feature and Bootstrapped Naive Bayes Classification*, ASI, vol. 8, no. 4, p. 94, July 2025, doi: 10.3390/asi8040094.
5. P. Pookduang, R. Klangbunrueang, W. Chansanam, and T. Lunrasri, *Advancing Sentiment Analysis: Evaluating RoBERTa against Traditional and Deep Learning Models*, Eng. Technol. Appl. Sci. Res., vol. 15, no. 1, pp. 20167–20174, Feb. 2025, doi: 10.48084/etasr.9703.
6. X. Chen, Z. Shen, T. Guan, Y. Tao, Y. Kang, and Y. Zhang, *Analyzing Patient Experience on Weibo: Machine Learning Approach to Topic Modeling and Sentiment Analysis*, JMIR Med. Inform., vol. 12, no. 3, p. e59249, Nov. 2024, doi: 10.2196/59249.
7. S. Moulya and T. R. Pragathi, *Mental Health Assist and Diagnosis Conversational Interface using Logistic Regression Model for Emotion and Sentiment Analysis*, Journal of Physics: Conference Series, vol. 2161, no. 1, p. 012039, Jan. 2022, doi: 10.1088/1742-6596/2161/1/012039.
8. Pristiyono, M. Ritonga, M. A. A. Ihsan, A. Anjar, and F. H. Rambe, *Sentiment Analysis of COVID-19 Vaccine in Indonesia Using Naïve Bayes Algorithm*, IOP Conference Series: Materials Science and Engineering, vol. 1088, no. 1, p. 012045, Feb. 2021, doi: 10.1088/1757-899x/1088/1/012045.
9. T. Shaik, X. Tao, C. Dann, H. Xie, Y. Li, and L. Galligan, *Sentiment Analysis and Opinion Mining on Educational Data: A Survey*, Natural Language Processing Journal, vol. 2, p. 100003, Dec. 2022, doi: 10.1016/j.nlp.2022.100003.
10. X. Wang and H. Ning, *TF-IDF Keyword Extraction Method Combining Context and Semantic Classification*, Proceedings of the 3rd International Conference on Data Science and Information Technology, pp. 123–128, Jul. 2020, doi: 10.1145/3414274.3414492.
11. M. Demircan, A. Seller, F. Abut, and M. F. Akay, *Developing Turkish Sentiment Analysis Models Using Machine Learning and E-Commerce Data*, International Journal of Cognitive Computing in Engineering, vol. 2, pp. 202–207, Jun. 2021, doi: 10.1016/j.ijcce.2021.11.003.
12. A. Dadhich and B. Thankachan, *Opinion Classification of Product Reviews Using Naïve Bayes, Logistic Regression and SentiWordNet: Challenges and Survey*, IOP Conference Series: Materials Science and Engineering, vol. 1099, no. 1, p. 012071, Mar. 2021, doi: 10.1088/1757-899x/1099/1/012071.
13. N. R. Novita, T. Herlambang, F. Yduianto, and D. B. Magfra, *Implementation of Naïve Bayes Method for Sentiment Analysis Case Study of MBKM*, AIP Conference Proceedings, vol. 3372, p. 040004, Jan. 2025, doi: 10.1063/5.0299482.