International Journal of Applied and Scientific Research (IJASR)
Vol. 4 No. 6, 2026: 343-356

Sentiment Analysis of Generative AI Adoption: A Comparative Study of Naive Bayes and Support Vector Machine Algorithms

Yosep Moleng[1*], Felix Wijaya[2], Manda Reksi Saputri [3], Besus Maula Sulthon[4]
Bina Sarana Informatika University
Corresponding Author: Yosep Moleng josephino89@gmail.com

A R T I C L E I N F O
Keywords: Sentiment Analysis, Generative AI, Naive Bayes, Support Vector Machine, TF-IDF, SMOTE

Received : 27, April
Revised : 28, May
Accepted: 30, June

A R T I C L E I N F O
The rapid adoption of Generative Artificial Intelligence (AI) has led to increasing public discussions and diverse opinions across social media platforms. This study aims to analyze public sentiment toward the adoption of Generative AI and compare the performance of the Naive Bayes and Support Vector Machine (SVM) algorithms in multi-class sentiment classification. The dataset consists of 1,125 Twitter/X posts categorized into positive, neutral, and negative sentiments. The research methodology includes text preprocessing through case folding, tokenization, stopword removal, and stemming, followed by TF-IDF feature weighting and data balancing using the Synthetic Minority Over-sampling Technique (SMOTE). Model performance was evaluated using 10-fold cross-validation to ensure reliable and robust results. The findings show that SVM achieved an accuracy of 62.13%, outperforming Naive Bayes, which achieved 52.44%, representing an improvement of 9.69%. These results demonstrate that SVM is more effective in handling high-dimensional TF-IDF feature spaces for multi-class sentiment classification and can support the development of public sentiment monitoring systems for Generative AI technologies.

DOI: https://doi.org/10.59890/ijasr.v4i6.250
E-ISSN: 3025-7670
https://nvlmultitechpublisher.my.id/index.php/ijasr/index

INTRODUCTION

Generative Artificial Intelligence (AI) is an advanced technology capable of automatically generating various forms of content, including text, images, audio, and video. One of the most widely recognized applications of generative AI is ChatGPT, developed by OpenAI. This technology generates human-like natural language responses and has been widely adopted for diverse purposes, ranging from information retrieval to digital content creation. Since its launch in November 2022, ChatGPT has attracted global attention and experienced rapid user growth, accelerating the adoption of generative AI technologies across academic, professional, and industrial settings.

The emergence of ChatGPT has created numerous opportunities in education and information technology. AI-powered chatbots can enhance learning by providing instant access to information, enabling personalized learning experiences, and improving communication between users and digital systems. Furthermore, generative AI has been increasingly utilized to support information retrieval, document preparation, idea generation, programming assistance, and various academic and professional activities.

The rapid expansion of ChatGPT reflects the growing acceptance of generative AI among the public. Within the educational sector, this technology has the potential to transform teaching and learning by providing quick access to knowledge, assisting students in completing academic tasks, and enriching the overall learning experience. However, the widespread adoption of generative AI has also generated diverse opinions regarding its usefulness, ease of use, and potential impact on academic and professional activities. These differing perspectives indicate that the adoption of generative AI remains a subject of ongoing debate among researchers, educators, and practitioners.

The use of ChatGPT in education also presents new challenges for academic institutions in maintaining academic honesty and integrity. The increasing integration of generative AI into higher education has encouraged many universities to develop institutional guidelines and policies to ensure that the technology is used responsibly while preserving academic integrity and achieving intended learning outcomes. The coexistence of perceived benefits and potential risks has consequently generated diverse public opinions regarding the adoption of generative AI technologies.

Although research on generative AI continues to expand, studies specifically investigating public sentiment toward its adoption remain relatively limited. Existing research has primarily relied on social media data to identify public opinions and attitudes toward generative AI technologies, particularly ChatGPT. Moreover, most previous studies have focused solely on sentiment identification without comparing the performance of different classification algorithms to determine the most effective approach for sentiment analysis.
Research specifically examining public sentiment toward generative AI adoption within the field of Information Systems is also still scarce. Understanding public opinions and sentiments is essential for assessing technology acceptance, identifying user perceptions, and recognizing potential challenges associated with the implementation of generative AI. Therefore, further research is needed

to analyze public sentiment regarding generative AI adoption while comparing multiple classification algorithms to identify the most effective predictive model. This study addresses this research gap by applying a text mining approach based on a comparative analysis of classification algorithms using public opinion data related to generative AI adoption in the Information Systems domain. Accordingly, this study aims to analyze public sentiment toward the adoption of generative AI and compare the performance of different classification algorithms to identify the model with the highest classification accuracy.

LITERATURE REVIEW

Generative Artificial Intelligenc

Generative Artificial Intelligence (Generative AI) is a branch of artificial intelligence capable of creating new content, including text, images, audio, and video, by learning patterns from training data. Unlike traditional AI, which primarily focuses on classification and prediction tasks, generative AI is designed to produce original outputs that closely resemble human-created content.

ChatGPT

ChatGPT is a generative AI-powered chatbot developed by OpenAI. It is built on a Large Language Model (LLM) that enables the system to understand and generate natural language responses, allowing more interactive and human-like communication with users.

ChatGPT can be applied to various tasks, including information retrieval, text summarization, document drafting, and educational support. Its ability to generate fast, relevant, and context-aware responses has made it one of the most widely adopted generative AI applications in both academic and professional environments.

Sentiment Analysis

Sentiment analysis is a technique within the fields of text mining and Natural Language Processing (NLP) used to identify and classify opinions or attitudes toward a particular topic into positive, negative, or neutral categories. It is widely employed to understand public perceptions based on textual data collected from social media, discussion forums, and other digital platforms. The insights generated from sentiment analysis can support decision-making processes and evaluate public responses to specific issues or emerging technologies.

Text Mining

Text mining is the process of extracting meaningful information and knowledge from unstructured textual data using techniques from Natural Language Processing (NLP), statistics, and machine learning. This approach is employed to identify patterns, relationships, and valuable information contained within textual data, transforming it into structured knowledge that is easier to analyze and interpret. In sentiment analysis research, text mining plays a

fundamental role in converting public opinion data into a format suitable for machine learning-based sentiment classification.

Text Mining Preprocessing

Text preprocessing consists of a sequence of procedures performed to transform unstructured text into a clean and standardized dataset suitable for analysis. In this study, the preprocessing stage includes case folding, tokenization, stopword removal, and stemming.

Case Folding

Case folding is a text normalization process that converts all alphabetic characters into lowercase letters to eliminate inconsistencies caused by differences in letter capitalization.
Example:
Original text:
"*ChatGPT Sangat Membantu Mahasiswa*"

After case folding:
"*chatgpt sangat membantu mahasiswa*"
Tokenization

Tokenization is the process of splitting text into smaller units called tokens, typically individual words, enabling each token to be processed independently during text analysis.
Example:
Input:
"*chatgpt sangat membantu mahasiswa*"

Output:
chatgpt | sangat | membantu | mahasiswa
Stopword Removal

Stopword removal is a preprocessing technique that eliminates commonly occurring words with little semantic value for classification. Removing these words helps reduce noise and improves the quality of textual features used by machine learning algorithms.
Stemming

Stemming is the process of reducing inflected or derived words to their root forms, thereby minimizing vocabulary variation and simplifying feature representation within text documents.

Original Word	**Root Word**
Helping	Help
Learning	Learn
Using	Use

Through these preprocessing steps, unstructured textual data are transformed into a cleaner and more standardized format, enabling more effective TF-IDF weighting and sentiment classification.

TF–IDF (Term Frequency–Inverse Document Frequency)

Term Frequency–Inverse Document Frequency (TF-IDF) is a feature weighting technique used to measure the importance of a word within a document relative to an entire document collection. The method combines Term Frequency (TF), which measures how often a term appears in a document, with Inverse Document Frequency (IDF), which measures how unique or rare the term is across the corpus. Words with higher TF-IDF values contribute more significantly to representing the document's content.

The TF-IDF formula is expressed as:

[\mathbf{TF\text{-}IDF}(t,d)=TF(t,d)\times\log\left(\frac{N}{DF(t)}\right)]

where:

- t = the term being weighted;
- d = a specific tweet or document;
- TF(t,d) = the frequency of term *t* in document *d*;
- DF(t) = the number of documents containing term *t*;
- N = the total number of documents in the dataset (1,019 tweets).

The IDF component reduces the weight of frequently occurring words while increasing the importance of more distinctive terms, resulting in a feature representation that is more suitable for sentiment classification.

Naïve Bayes

Naïve Bayes is one of the most widely used machine learning algorithms for sentiment analysis. It classifies text into positive, negative, or neutral categories based on the probabilistic occurrence of words within each class. Owing to its simplicity, computational efficiency, and strong performance in text classification tasks, Naïve Bayes remains a popular baseline algorithm in sentiment analysis research.

Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised machine learning algorithm commonly applied to sentiment analysis because of its ability to effectively process high-dimensional textual data. SVM separates different classes by constructing an optimal hyperplane that maximizes the margin between data points from different categories. Its robustness and high classification accuracy have made SVM one of the most effective algorithms for sentiment classification on social media and textual datasets.

Confusion Matrix

A confusion matrix is an evaluation tool used to assess the performance of a classification model by comparing predicted labels with actual labels. Based on this comparison, several evaluation metrics—including accuracy, precision, recall, and F1-score—can be calculated, providing a comprehensive assessment of model performance.

Cross-Validation

Cross-validation is a machine learning evaluation technique used to estimate a model's ability to generalize to unseen data. The dataset is partitioned into several subsets (folds), which are alternately used as training and testing data so that every observation is evaluated exactly once.

This study employs 10-fold cross-validation, where the dataset is divided into ten equally sized subsets. During each iteration, nine folds are used for model training while the remaining fold is used for testing. The process is repeated ten times until every fold has served as the test set once. The final evaluation score is calculated as the average performance across all ten iterations.

The 10-fold cross-validation approach was selected because it provides a more stable estimate of model performance and reduces evaluation bias compared with a single train-test split. Furthermore, stratified cross-validation is recommended for imbalanced datasets because it preserves the class distribution within each fold, thereby producing more reliable and representative evaluation results.

METHODOLOGY

Research Framework

This study was systematically designed following a text mining workflow for sentiment analysis. The overall research process consists of several sequential stages, including data acquisition, text preprocessing, feature extraction, sentiment classification, and model evaluation. The complete end-to-end research framework is illustrated in Figure 1.

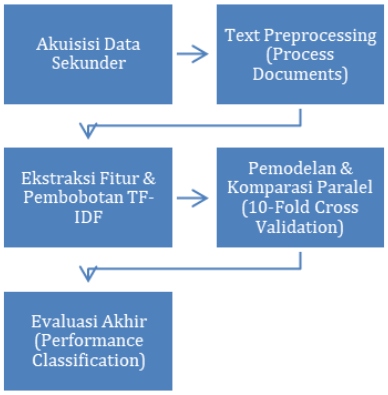

Figure 1. Research Framework

Data Acquisition

This study employed secondary data obtained from the publicly available ChatGPT Sentiment Analysis dataset hosted on the Kaggle repository. The dataset contains opinions and comments posted by Twitter (X) users regarding the use of ChatGPT.

A total of 1,000 labeled tweets were selected for this study. Each record had been manually annotated into one of three sentiment categories: positive, neutral, or negative. The dataset was subsequently imported into RapidMiner using the Read CSV operator for text mining and sentiment classification.

Text Preprocessing in RapidMiner

To prepare the textual data for machine learning, a series of preprocessing operations was performed using the Process Documents from Data operator in RapidMiner. This stage aims to clean, normalize, and standardize the text while reducing noise and improving feature quality for classification.

The preprocessing pipeline consisted of the following sequential steps:

a. Tokenization

The text was segmented into individual tokens (words) using non-letter characters as delimiters, allowing each word to be processed independently.

b. Case Transformation

All characters were converted to lowercase to ensure consistent word representation and eliminate differences caused by letter capitalization.

c. Stopword Removal

Common English stopwords, including conjunctions, prepositions, and auxiliary words that contribute little semantic information, were removed to improve classification performance.

d. Stemming

Words containing prefixes or suffixes were reduced to their root forms using the Porter Stemmer algorithm. This process minimizes vocabulary variations by representing words with similar meanings in a standardized form.

Feature Extraction

After preprocessing, the cleaned textual data were transformed into numerical feature vectors using the Term Frequency–Inverse Document Frequency (TF-IDF) weighting scheme.

Within the Process Documents from Data operator, the Text Term Weighting parameter was configured to use TF-IDF. This method calculates the importance of each term by combining its frequency within a document (Term Frequency) with its rarity across the entire document collection (Inverse Document Frequency). The resulting TF-IDF matrix represents the relative importance of each term and serves as the input features for the sentiment classification models.

Classification Modeling

Model development was performed using the Cross Validation operator in RapidMiner with a 10-fold cross-validation strategy to obtain robust and unbiased performance estimates.

During each iteration, the dataset was partitioned into ten equal subsets (folds). Nine folds were used for model training, while the remaining fold served as the testing set. This procedure was repeated ten times until every fold had been used once as the testing dataset. The final model performance was calculated as the average of all evaluation results.

Two machine learning algorithms were implemented and compared:

a. Naïve Bayes (NB)

Naïve Bayes is a probabilistic classification algorithm based on Bayes' Theorem, assuming conditional independence among features. Due to its computational efficiency and simplicity, it is widely used as a baseline algorithm for text classification.

b. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised learning algorithm that constructs an optimal hyperplane to separate data into different classes within a high-dimensional feature space. Owing to its robustness in handling sparse textual data, SVM is widely recognized as one of the most effective algorithms for sentiment classification.

Model Evaluation

During the testing phase, the trained models were evaluated using the Apply Model and Performance (Classification) operators available in RapidMiner.

Model performance was analyzed using a Confusion Matrix, which compares predicted labels with the actual sentiment labels for each class. Based on the confusion matrix, the following evaluation metrics were calculated:

- Accuracy, representing the proportion of correctly classified instances relative to the total number of observations.
- Precision, measuring the proportion of correctly predicted instances within each predicted sentiment class.
- Recall, indicating the model's ability to correctly identify all instances belonging to each actual sentiment class.

The accuracy, precision, and recall values obtained from the Naïve Bayes and Support Vector Machine models were subsequently compared to determine the algorithm that achieved the best performance for multi-class sentiment analysis of public opinions regarding ChatGPT.

RESULT AND DISCUSSION

Performance of the Naïve Bayes Algorithm

The Naïve Bayes algorithm was evaluated using 10-fold cross-validation on the sentiment dataset related to Generative Artificial Intelligence (AI) adoption after applying text preprocessing, TF-IDF feature weighting, and SMOTE-based data balancing. The objective of this experiment was to assess the algorithm's ability to classify public opinions into three sentiment categories: positive, neutral, and negative.

Based on the RapidMiner evaluation results, the Naïve Bayes classifier achieved an accuracy of 52.44%. The confusion matrix reveals an uneven classification pattern. The model correctly classified 424 out of 453 negative instances, corresponding to a recall of 93.60% for the negative class. However, its performance in distinguishing positive and neutral sentiments was considerably lower. The recall values for the positive and neutral classes were only 23.18% (52 out of 224 instances) and 27.85% (101 out of 362 instances), respectively.

The relatively poor performance on the positive and neutral classes indicates a strong classification bias toward the negative category. This limitation is mainly attributed to the probabilistic assumption underlying Naïve Bayes, which considers each feature to be conditionally independent of the others. In sentiment analysis, however, words often exhibit complex contextual relationships rather than acting independently. When high-dimensional TF-IDF features are used, this independence assumption becomes increasingly unrealistic, causing the classifier to favor the dominant class. Although SMOTE was employed to reduce class imbalance, its effectiveness remains limited for algorithms that are highly sensitive to class distributions during the learning process

Table 1. Naïve Bayes Evaluation Results

Evaluation Metric	Value
Accuracy	52.44%

Model performance was further analyzed using a confusion matrix, as presented in Table 2.

Table 2. Confusion Matrix of Naïve Bayes

Actual / Predicted	Positive	Neutral	Negative
Positive	105	27	16
Neutral	103	61	13
Negative	245	131	424

The confusion matrix demonstrates that Naïve Bayes performs relatively well in identifying negative sentiment, correctly classifying 424 negative instances. Nevertheless, the algorithm struggles to distinguish between positive and neutral sentiments, as reflected by the substantial number of misclassified samples in both categories.

The recall values obtained were 23.18% for the positive class, 27.85% for the neutral class, and 93.60% for the negative class. These findings indicate that the classifier tends to assign most observations to the negative category, resulting in relatively poor predictive performance for the remaining sentiment classes.

Performance of the Support Vector Machine (SVM) Algorithm

The Support Vector Machine (SVM) classifier was evaluated using the same dataset and experimental configuration as the Naïve Bayes model, ensuring a fair comparison under identical conditions.

The experimental results show that SVM achieved an accuracy of 62.13%, correctly classifying 699 out of 1,125 observations. The confusion matrix indicates that SVM produced a more balanced distribution of predictions across sentiment classes. The model correctly identified 440 negative instances (97.13% recall), 382 positive instances (50.77% recall), and 48 neutral instances (13.24% recall).

Although the recall for the neutral class remained relatively low, SVM demonstrated a more balanced classification performance than Naïve Bayes. This superiority can be attributed to its ability to construct an optimal hyperplane that effectively separates high-dimensional TF-IDF feature vectors. By maximizing the margin between classes, SVM can better handle overlapping sentiment categories, particularly between positive and neutral opinions. Furthermore, the kernel mechanism enables SVM to capture complex nonlinear relationships among textual features that cannot be effectively modeled by Naïve Bayes.

Table 3. SVM Evaluation Results

Evaluation Metric	Value
Accuracy	62.13%

The confusion matrix of the SVM classifier is presented in Table 4.

Table 4. Confusion Matrix of SVM

Actual / Predicted	Positive	Neutral	Negative
Positive	230	40	9
Neutral	31	29	4
Negative	192	150	440

The confusion matrix confirms that SVM outperformed Naïve Bayes in classifying sentiment data. In particular, the algorithm successfully recognized 440 negative instances while substantially improving the identification of positive sentiments compared with Naïve Bayes.

The recall values were 50.77% for the positive class, 13.24% for the neutral class, and 97.13% for the negative class. These results indicate that SVM is highly effective in recognizing negative sentiment, although distinguishing neutral opinions remains challenging due to lexical similarities with both positive and negative expressions.

Comparison Between Naïve Bayes and Support Vector Machine

The experimental results demonstrate that Support Vector Machine (SVM) significantly outperformed Naïve Bayes, achieving an accuracy improvement of 9.69 percentage points (62.13% versus 52.44%).

Several factors explain this performance difference.

1. Algorithm Characteristics and High-Dimensional Feature Space

Naïve Bayes assumes conditional independence among features, an assumption that rarely holds for textual data represented by thousands of TF-IDF dimensions. Consequently, its predictive capability deteriorates when contextual relationships among words become important. In contrast, SVM is specifically designed to operate effectively in high-dimensional spaces by constructing an optimal separating hyperplane, enabling it to learn more complex decision boundaries.

2. Class Separation Strategy

Naïve Bayes relies solely on posterior probabilities for each class without considering the geometric relationships between classes. As a result, it often produces suboptimal decision boundaries when sentiment classes overlap substantially. Conversely, SVM explicitly maximizes the margin between classes, generating more robust classification boundaries and improving generalization performance.

3. Impact of Class Imbalance

Although **SMOTE** was applied to balance the dataset, Naïve Bayes continued to exhibit a strong bias toward the majority class during testing. SVM proved more resilient to residual class imbalance because its margin-maximization objective and optimization framework are less affected by uneven class distributions.

4. Complexity of Social Media Sentiment

Twitter/X posts discussing generative AI often contain mixed or nuanced opinions. Statements such as "*ChatGPT is useful, but it still requires stronger regulation*" simultaneously express positive and neutral viewpoints. Naïve Bayes struggles to capture such contextual nuances because it evaluates words independently, whereas SVM considers combinations of features, enabling it to distinguish more complex sentiment patterns.

Practical Implications

Although an accuracy improvement of 9.69% may appear moderate, it is substantial in a three-class sentiment classification problem. Given the evaluation dataset of 1,125 observations, SVM correctly classified approximately 109 more instances than Naïve Bayes. For large-scale public sentiment monitoring systems, such an improvement can significantly enhance the reliability of generated insights and support better strategic decision-making.

However, this improvement comes with certain trade-offs. Compared with Naïve Bayes, SVM generally requires longer training and inference times and involves more complex hyperparameter tuning, including kernel selection, the regularization parameter (C), and gamma. Therefore, in real-time applications with strict latency constraints, Naïve Bayes may still be a practical alternative despite its lower predictive performance.

Table 5. Accuracy Comparison Between Classification Algorithms

Algorithm	Accuracy
Naïve Bayes + SMOTE	52.44%

SVM + SMOTE	62.13%

The findings of this study are consistent with previous sentiment analysis literature, which consistently identifies Support Vector Machine as one of the most effective algorithms for text classification, particularly in multi-class sentiment analysis (Pang & Lee, 2008; Liu, 2012). Similarly, studies by Tripathi et al. (2016) and Medhat et al. (2014) reported that SVM generally outperforms Naïve Bayes when dealing with high-dimensional social media datasets.

Despite the superior overall performance of SVM, neutral sentiment remains the most difficult class to identify accurately. The recall values of 13.24% for SVM and 27.85% for Naïve Bayes indicate that both models are more sensitive to extreme sentiments (positive and negative) than to balanced opinions. Within the context of generative AI, neutral sentiments are particularly important because they often represent nuanced perspectives that acknowledge both the benefits and the limitations of the technology. Future studies should therefore investigate cost-sensitive learning, class weighting, or other imbalance-aware approaches to improve the recognition of neutral sentiment.

The results also highlight the critical role of feature engineering and text preprocessing. While the TF-IDF representation proved effective for both classifiers, future research could explore alternative feature extraction techniques such as BM25, Word2Vec, FastText, or transformer-based embeddings (e.g., BERT) to further improve classification performance.

From a practical perspective, these findings are valuable for organizations seeking to develop public sentiment monitoring systems for emerging technologies such as generative AI. With an accuracy of 62.13%, SVM provides a reliable baseline for initial deployment. Nevertheless, further improvements are still needed, particularly in identifying neutral opinions. For applications requiring higher predictive accuracy, future work should investigate ensemble learning methods and deep learning-based sentiment classification models.

CONCLUSIONS AND RECOMMENDATIONS

Conclusion

Based on the findings of this study on sentiment analysis of Generative Artificial Intelligence (AI) adoption using the Naïve Bayes and Support Vector Machine (SVM) algorithms, the following conclusions can be drawn:

1. The text preprocessing pipeline—including case folding, tokenization, stopword removal, stemming, and TF-IDF feature weighting—successfully transformed unstructured textual data into numerical representations suitable for sentiment classification.
2. The implementation of the Synthetic Minority Over-sampling Technique (SMOTE) effectively mitigated class imbalance among the sentiment categories, enabling a more balanced and reliable classification process.
3. The Naïve Bayes classifier achieved an accuracy of 52.44%. Although the algorithm demonstrated strong performance in identifying negative sentiment, it showed limited capability in distinguishing positive and neutral sentiments.

4. The Support Vector Machine (SVM) classifier achieved an accuracy of 62.13%, outperforming Naïve Bayes in classifying sentiment across the dataset. The superior performance of SVM highlights its effectiveness in handling high-dimensional textual features generated through TF-IDF.
5. Comparative analysis revealed that Support Vector Machine (SVM) was the most effective classification algorithm in this study, surpassing Naïve Bayes by 9.69 percentage points in overall accuracy.

Overall, the findings indicate that Support Vector Machine (SVM) is more suitable than Naïve Bayes for performing sentiment analysis on public opinions regarding the adoption of Generative Artificial Intelligence (AI) within the dataset used in this research.

Recommendations

Based on the findings and limitations of this study, several recommendations are proposed for future research:

1. Expand the dataset size to enable the classification models to learn more comprehensive sentiment patterns, thereby improving predictive accuracy and model generalization.
2. Utilize more diverse data sources, including other social media platforms, online discussion forums, and digital news websites, to obtain a broader representation of public opinion regarding Generative AI adoption.
3. Incorporate feature selection methods and hyperparameter optimization techniques to enhance the performance of sentiment classification algorithms and improve classification accuracy.

Compare additional machine learning and deep learning algorithms, such as Random Forest, Decision Tree, K-Nearest Neighbors (KNN), Logistic Regression, and transformer-based or other deep learning models, to identify more effective approaches for multi-class sentiment classification.

FURTHER STUDY

This study still has limitations; therefore, further research on the topic is required.

REFERENCES

A. Brasoveanu, M. Moodie, and R. Agrawal, "Textual evidence for the perfunctoriness of independent medical reviews," in *CEUR Workshop Proceedings*, CEUR-WS, 2020, pp. 1–9. doi: 10.1145

A. Dabis and C. Csáki, "AI and ethics: Investigating the first policy responses of higher education institutions to the challenge of generative AI," *Humanit. Soc. Sci. Commun.*, vol. 11, no. 1, Dec. 2024, doi: 10.1057/s41599-024-03526-z.

A. Haleem, M. Javaid, and R. P. Singh, "An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges," *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, vol. 2, no. 4, Oct. 2022, doi: 10.1016/j.tbench.2023.100089.

A. Palanivinayagam, C. Z. El-Bayeh, and R. Damaševičius, "Twenty Years of Machine-Learning-Based Text Classification: A Systematic Review," *Algorithms*, vol. 16, no. 5, May 2023, doi: 10.3390/a16050236.

C. Leiter *et al.*, "ChatGPT: A Meta-Analysis after 2.5 Months," Feb. 2023, [Online]. Available: http://arxiv.org/abs/2302.13795

D. Heaton, J. Clos, E. Nichele, and J. E. Fischer, "'The ChatGPT bot is causing panic now – but it'll soon be as mundane a tool as Excel': analysing topics, sentiment and emotions relating to ChatGPT on Twitter," *Pers. Ubiquitous Comput.*, vol. 28, no. 6, pp. 875–894, Dec. 2024, doi: 10.1007/s00779-024-01811-x.

D. R. E. Cotton, P. A. Cotton, and J. R. Shipway, "Chatting and cheating: Ensuring academic integrity in the era of ChatGPT," *Innovations in Education and Teaching International*, vol. 61, no. 2, pp. 228–239, 2024, doi: 10.1080/14703297.2023.2190148.

D. Wilimitis and C. G. Walsh, "Practical Considerations and Applied Examples of Cross-Validation for Model Development and Evaluation in Health Care: Tutorial," *JMIR AI*, vol. 2, no. 1, 2023, doi: 10.2196/49023.

H. Setyawan, L. Ma'rifatul Azizah, and A. Y. Pradani, "Sentiment Analysis of Public Responses on Indonesia Government Using Naïve Bayes and Support Vector Machine," 2023. [Online]. Available: https://t.co/FKZjLtyWaN

M. Siino, I. Tinnirello, and M. La Cascia, "Is text preprocessing still worth the time? A comparative survey on the influence of popular preprocessing methods on Transformers and traditional classifiers," *Inf. Syst.*, vol. 121, Mar. 2024,

N. Arifah Wahyuni, D. Rizky Amelia, and tishom Al Khoiry, "Exploring Public Sentiment on ChatGPT-An Analysis of Twitter Data Article Info," *JEE*, vol. 13, no. 1, pp. 51–60, 2024, [Online]. Available:

OpenAI *et al.*, "GPT-4 Technical Report," Mar. 2024, [Online]. Available: http://arxiv.org/abs/2303.08774

P. P. Ray, "ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope," Jan. 01, 2023, *KeAi Communications Co.* doi: 10.1016/j.iotcps.2023.04.003.

R. Bommasani *et al.*, "On the Opportunities and Risks of Foundation Models," Jul. 2022, [Online]. Available: http://arxiv.org/abs/2108.07258

R. Peres, M. Schreier, D. Schweidel, and A. Sorescu, "Editorial: On ChatGPT and beyond: How generative artificial intelligence may affect research, teaching, and practice," *International Journal of Research in Marketing*, vol. 40, no. 2, pp. 269–275, Jun. 2023, doi: 10.1016/j.ijresmar.2023.03.001.

S. Bubeck *et al.*, "Sparks of Artificial General Intelligence: Early experiments with GPT-4," Apr. 2023, [Online]. Available: http://arxiv.org/abs/2303.12712

S. Farhadpour, T. A. Warner, and A. E. Maxwell, "Selecting and Interpreting Multiclass Loss and Accuracy Assessment Metrics for Classifications with Class Imbalance: Guidance and Best Practices," *Remote Sens. (Basel).*, vol. 16, no. 3, Feb. 2024,

S. Szeghalmy and A. Fazekas, "A Comparative Study of the Use of Stratified Cross-Validation and Distribution-Balanced Stratified Cross-Validation in Imbalanced Learning," *Sensors*, vol. 23, no. 4, Feb. 2023, doi: 10.3390/s23042333.

V. M. Ngo, "Does ChatGPT change artificial intelligence-enabled marketing capability? Social media investigation of public sentiment and usage," *Global Media and China*, vol. 9, no. 1, pp. 101–125, Mar. 2024, doi: 10.1177/20594364241228880.

W. M. Lim, A. Gunasekara, J. L. Pallant, J. I. Pallant, and E. Pechenkina, "Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators," *International Journal of Management Education*, vol. 21, no. 2, Jul. 2023, doi: 10.1016/j.ijme.2023.100790.

Y. K. Dwivedi *et al.*, "'So what if ChatGPT wrote it?' Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy," *Int. J. Inf. Manage.*, vol. 71, Aug. 2023, doi: 10.1016/j.ijinfomgt.2023.102642.