Engineering and Technology Journal e-ISSN: 2456-3358

Volume 09 Issue 07 July-2024, Page No.- 4511-4519

DOI: 10.47191/etj/v9i07.21, I.F. – 8.227

Opinion-Mining Technique on Generative Artificial Intelligence Topic Using Data Classification Algorithms

Michael G. Albino

School of Graduate Studies, AMA Computer University, Quezon City, Philippines

ABSTRACT: The study employed an opinion-mining technique using data classification algorithms on the topic of Generative Artificial Intelligence (GenAI) to determine the sentiments of Twitter users. The researcher used a sentiment analysis framework to gather the datasets for dataset training and predict the results using Naïve Bayes, Random Forest, and SVM algorithms. The result shows that SVM and Random Forest algorithms had the same precision and recall of 1.000 indicating that the result has no false positive values. On the other hand, the Naïve Bayes algorithm garnered a .949 precision and .939 recall which means fewer false positive results on the trained models. The overall result shows that the trained datasets indicate a successful prediction with fewer false positive results. Moreover, the result of the sentiment analysis shows that more positive sentiments were drawn on the topic of generative artificial intelligence indicating the use and benefits of using AI. Furthermore, based on the result of the study, the research recommended the use of the sentiment analysis framework through an opinion-mining technique using data classification algorithms as it may help analyze different emotions of social media users.

KEYWORDS: Algorithms, data classification, generative artificial intelligence (GenAI), opinion-mining, **s**entiment analysis

I. INTRODUCTION

Mining is a technique used to assist individuals extract valuable information from massive amounts of data. Sentiment analysis is concerned with the analysis and comprehension of emotions expressed through text patterns. It identifies the opinion or attitude that a person has towards a topic or an object, as well as the viewpoint underlying a text span (Mouthami et al., 2013). Opinion mining or sentiment analysis is useful in social media monitoring to automatically characterize the overall feeling or mood of consumers as reflected in social media toward a specific brand or company and determine whether they are viewed positively or negatively on the web to improve products and services (Aggarwal, 2018).

On the other hand, opinion mining (OM) is the use of Natural Language Processing (NLP), computational linguistics, and text analytics to discover and extract subjective information from source materials, such as debates about certain products and services. (Patacsil et al., 2015). Opinion mining, also known as sentiment analysis, has become increasingly popular in recent years because social networking sites allow users to freely express their views, opinions, and perceptions on a given issue. Nowadays, any type of marketing business is engaged in the latest business trends. Aside from printed surveys, the organizations also extend their customer satisfaction study through the web, gathering a vast amount of data. (Troussas et al., 2019).

Related studies show the relatedness of using sentiment analysis on Facebook comments to track cyberbullying incidents using Naïve Bayes classifiers (Mathapati et al., 2017). Surroop et al., 2016, presented in their paper on the use of sentiment analysis. Among 1031 study participants, it was discovered that 97.8% of the very bad attitudes, 70.7% of the negative sentiments, and 77.0% of the good sentiments were successfully extracted. Sentiment analysis is a more effective approach for extracting specific sentiments than manually recognizing them. The researchers can already discern the respondents' favorable, negative, and neutral comments using sentiment analysis.

Thus, this research paper focused on opinion mining techniques using data classification algorithms from the extracted Kaggle datasets on Twitter posts for the topic “Generative Artificial Intelligence (GenAI)”. The data classification algorithms were used to analyze data from the text for sentiment analysis.

Opinion Mining Technique

Natural Language Processing (NLP) includes the discipline of opinion mining, sometimes known as sentiment analysis. It extracts people's ideas, including evaluations, attitudes, and feelings about people, topics, and events. The assignment is technically tough but extremely beneficial. With the rapid rise of digital platforms in cyberspace, such as blogs and social networks, individuals and organizations are increasingly relying on public opinion to make decisions. In recent years, extensive studies about mining people's sentiments based on text in cyberspace utilizing opinion mining have been investigated (Razali et al., 2021). Researchers have used a variety of opinion mining techniques, including machine learning and lexicon-based approaches, to assess and categorize people's attitudes based on text and debate the existing gap. As a result, it provides a chance for other researchers to examine and suggest superior methodologies and new domain applications to close the gap.

Moreover, opinion mining or sentiment analysis helps in achieving various goals such as observing public mood regarding political movements (Preotiuc-Pietro et al., 2017), customer satisfaction measurement (F. Xu et al., 2020), movie sales prediction (Rachiraju & Revanth, 2020), etc. However, the existing opinion mining method, which includes machine learning and a lexicon-based approach, is ineffective in analyzing and classifying people's sentiments and emotions in cyberspace in terms of national security because some opinion mining methods only focus on existing domains such as business and education.

According to Shah et al. (2023), opinion mining/sentiment analysis is the computational study of people's perceptions, evaluations, attitudes, and emotions about individuals, people, issues, events, subjects, and their characteristics. It is also the study of people's opinions based on the feelings, attitudes, or emotions expressed in a product (Isabelle et al., 2019).

A thought, opinion, or concept based on a feeling about a situation is the definition of the term "sentiment" according to the Cambridge Dictionary (2021). Opinion mining is the process of gathering opinions and categorizing them based on their polarity, whether positive, negative, or other emotions. They can be used at several levels, including document-level sentiment analysis, sentence-level sentiment analysis, and feature/aspect-level sentiment analysis.

Relatively, opinion mining has been a research interest since the early twenty-first century. In 2003, Dave et al., (2003) they explored opinion mining and provided a methodology for document polarity classification (either recommended or not recommended) based on feedback analysis from certain entities. Following that discovery, additional researchers got interested in including opinion mining in their text-mining studies. It then became new extensive research in the following years. Hu & Liu (2004) investigated the mining approach for summarizing product reviews by recognizing opinion sentences in each review and determining whether they are good or negative. In 2008, Abbasi et al. conducted research on sentiment analysis techniques and their applications (Abbasi et al., 2008; Zhang et al., 2008). In 2009, Tang et al. (2009) they reviewed document sentiment classification and opinion extraction, as well as experimented with classifying web review opinions for consumer product analysis. In 2010, Chen & Zimbra (2010) assessed the attitudes of various business constituents about the organization using an analysis framework that employed automated topic and sentiment extraction approaches to various online forums. Based on a survey of chosen articles, this study discovered that between 2016 and today, opinion mining-related research is still an interesting subject field for researchers (Kaur et al., 2022).

Generative Artificial Intelligence

Generative AI (GenAI) refers to a class of machine learning algorithms that create new data samples that closely resemble current datasets. One of the core approaches of GenAI is the Variational Autoencoder (VAE), which is a form of neural network that learns to encode and decode input in a way that preserves its important qualities (H. Xu, 2018). Another prominent GenAI technology is Generative Adversarial Networks (GANs), which are made up of two neural networks competing to generate realistic data samples (Goodfellow et al., 2014). GenAI models employ powerful algorithms to understand patterns and generate new content, including text, images, audio, videos, and code. GenAI tools include ChatGPT, Bard, Stable Diffusion, and Dall-E. Its capacity to handle complicated cues and produce human-like output has sparked study and interest in incorporating GenAI into a variety of industries, including healthcare, medicine, education, media, and tourism.

In addition, ChatGPT, for example, has caused a surge of interest in the use of GenAI in higher education since its release in November 2022 (K. Hu, 2023). OpenAI created a conversational AI system using an autoregressive big language model (more than 175 billion parameters) that was pre-trained on a massive corpus of text data. It can produce human-like answers to a variety of text-based inputs. The model has been trained on a variety of texts, including books, papers, and webpages, enabling it to comprehend user input, generate responses, and maintain coherent discussions on a wide range of topics. There has been substantial discussion about its potential to revolutionize disciplinary practices like medical writing (Feng & Shen, 2023; Kitamura, 2023), surgical practice (Bhattacharya et al., 2023) and healthcare communications (Eggmann et al., 2023) as well as enhancing higher education teaching and learning (Adiguzel et al., 2023; Anu & Ansah, 2023).

Generative AI, particularly Large Language Models (LLMs), an advanced type of transfer learning model, offers the potential to further alter sentiment analysis (Krugmann & Hartmann, 2024). The large amount of data utilized for LLM training could greatly improve performance across sentiment analysis tasks, influencing the choice of approaches in this domain. This versatility not only improves the accessibility of LLMs but also makes them appropriate for diverse sentiment classification tasks, ranging from binary to multi-class (Kheiri & Karimi, 2023; Wang et al., 2023), and enables their application in zero-shot or few-shot scenarios (Simmering & Huoviala, 2023).

Data Classification Algorithms

Classification is a data mining (machine learning) approach that predicts group membership for data examples (Kesavaraj & Sukumaran, 2013). Although there are a range of accessible approaches for machine learning, classification is the most generally utilized method (Singh et al., 2013). Classification is a popular problem in machine learning, particularly for future and knowledge discovery. Researchers in the disciplines of machine learning and data mining consider classification to be one of the most extensively studied problems (Pisani et al., 2002). A general model of supervised learning (classification techniques) is shown in Figure 1.

Although classification is a well-known machine learning technique, it has some limitations, such as dealing with missing data. Missing values in the dataset can cause issues during both the training and classification stages. Some of the possible causes for missing data are presented in (Aized Amin Soofi & Arshad Awan, 2017) non-entry of a record due to a misunderstanding, data identified as irrelevant at the time of input, data removal due to departure from other documented data, and equipment malfunction. Missing data issues can be addressed through techniques (Bhukya & Ramachandram, 2010); Data miners can ignore omitted data, replace entire omitting values with a single global constant, replace an omitting value with its feature mean for the given class, manually observe samples with omitted values, and insert a feasible or probable value. In this study, we will simply look at a few classification approaches.

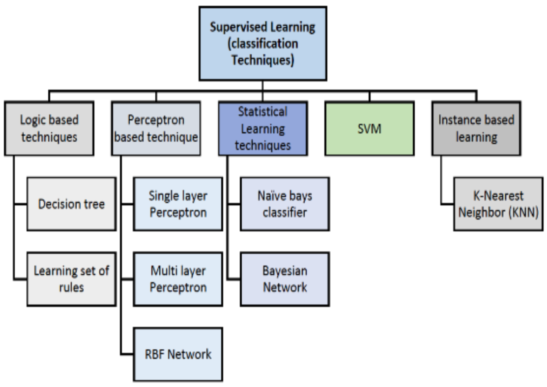


Figure 1. Supervised Learning Classification Technique

Text classifiers are capable of organizing, arranging, and categorizing nearly any sort of text, including documents, medical studies, files, and online material (Amanat et al., 2022). Unstructured data makes up more than 80% of all data, with text being one of the most common categories. Because analyzing, interpreting, organizing, and sifting through text data is complex and time-consuming due to its chaotic nature, most firms do not use it to its full potential. Text categorization is a technique that involves extracting usable information from text (Bashir et al., 2022). Here is where machine learning and text classification come into play. Text classifiers can help businesses swiftly and cost-effectively organize all important text kinds, such as emails, legal documents, social media, surveys, and more (Abbasi et al., 2021, 2022; Hina, Ali, Javed, Ghabban, et al., 2021).This technology enables businesses to save time examining text data, automate business processes, and make data-driven business decisions. Many businesses employ text analysis technologies to examine the text. Text analysis technologies enable businesses to arrange massive volumes of information, such as emails, chats, social media, support tickets, papers, and so on, in seconds rather than days. Therefore, we can dedicate more resources to critical tasks (Hina, Ali, Javed, Srivastava, et al., 2021; Rafat et al., 2022).

II. METHODS

This research paper aims to gather posts of 1000 Twitter users through Kaggle datasets and will apply machine learning techniques to extract the opinions or sentiments of users on the topic "Generative Artificial Intelligence (GenAI)". Qualitative text analysis will be used by the researchers to gather qualitative data, assign code labels, and iteratively develop findings (Guetterman et al., 2018; Sarfo et al., 2017) and apply Natural Language Processing Techniques (NLPT) to automate the rigorous process of analyzing qualitative statements through opinion mining or sentiment analysis (Kaufman et al., 2016).

The application of the NLPT to this research study was beneficial since it completed the automation of text cleaning, data visualization, and sentiment analysis. The trained model will be easily computed for sentiment analysis through this process and provide output to the researcher for compilation and analysis.

In this research paper, the researcher evaluated the accuracy of the trained model through several machine learning algorithms but not limited to K-Nearest Neighbors (K-NN), Tree (T), and Naïve Bayes (NB) to get the precision and recall. These algorithms learn quickly to adapt as the researcher feeds new, smaller but domain-relevant data into the pre-trained algorithm for model refinement (Lazrig & Humpherys, 2022; Zhuang et al., 2021). A preprocessing technique was used to clean the unnecessary data from the datasets such as uppercase words, unwanted spaces, URLs, HTML encoding, usernames, punctuations, numbers, emoji, and three or fewer characters which are all unnecessary to the processed text as shown in Table 1.

Table 1. Text Processing Features

Basic Text Preprocessing	Intermediate Text Preprocessing
Set text to lowercase	Remove stop words
Remove unwanted spaces	Lemmatizing
Remove URL	Remove duplicate

Remove HTML encoding	
Remove usernames	
Remove punctuations, numbers, and emoji	
Remove three or fewer character	

Afterward, the researcher employed an intermediate text processing technique to further enhance and remove some unnecessary words and characters in the pre-processed text such as removing stop words, lemmatizing, and removing duplicate words to the text. These steps were necessary as part of the Natural language Processing Technique (NLPT) before proceeding to the actual sentiment analysis process and acquiring sentiments to the gathered datasets.

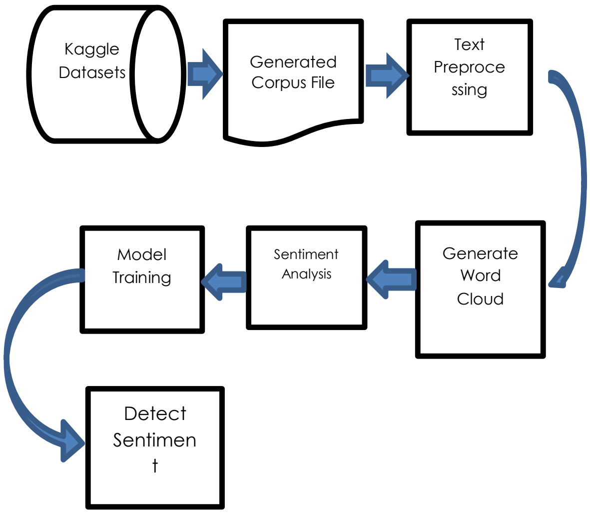


Figure 2. Data Classification in Opinion Mining Technique Framework

The data classification process framework that the researcher utilized in this research study is shown in Figure 2. First, the researchers will use the Kaggle datasets to gather 1000 tweeter posts of users related to Generative AI. Then, the researcher generated a corpus file for the machine learning application to read and use the data extracted easily and readily. To ensure that the data extracted are clean and free from unnecessary symbols and words, a preprocessing technique was administered by the researchers and generated a word cloud visualization of the clean text. After this, sentiment analysis was administered by the researcher using (Liu Hu and Vader) to get the positive, negative, and neutral sentiments of the users. The model was trained using a machine learning application and generated the results of the training through precision and recall or the accuracy of the data generated by the application.

III. RESULTS AND DISCUSSIONS

The Kaggle datasets were the primary source of the data used by the researchers. Through the collection of the data from the Kaggle website, the researcher was able to gather tweets from users regarding the topic "Generative Artificial Intelligence" which is considered a trending topic now in all social media websites.

Figure 2. Generated Word Cloud

Figure 2 shows the cleaned text viewed in a Word Cloud. The following words are generative ai, ChatGPT, generative art, artificial intelligence, stable diffusion, images, open ai, new, and many more. In this manner, the text presented is already cleaned and unwanted symbols are no longer presented. The data presented are ready for processing for sentiment analysis of the documents presented in the corpus file.

	Text	sentiment
1	Is creativity at risk? As Generative AI enters the creator economy, we must ensure that human creativity and origi...	0
2	Three Ways Media Leaders Can Leverage Generative AI https://t.co/1qPjNTbQTQ #documentai #generativeai #ve...	4.34783
3	Unbelievable AI Auto Designer.. I just signed up to try Autodesigner, a new generative AI by @uizard for UI/UX d...	0
4	Generative AI promises to streamline health care thanks to Microsoft and Epic	8.33333
5	REG NOW #Roundtable on #GenerativeAI on 5/16 hosted @MayfieldFund in Menlo Park. Discover the opportuni...	0
6	As a Startup Attorney, I've combined Generative AI tools with my legal expertise to create winning outcomes for ...	5.26316
7	I just signed up to try Autodesigner, a new generative AI by @uizard for UI/UX design. Looks incredible! https://t....	2.5641
8	Engaged in productive meetings with photography association and research faculty, gathering insights on conte...	3.44828
9	Preparing for the future of AI: discover the latest developments in generative AI. #GenerativeAI #FutureTech #AI ...	0
10	By taking these steps, businesses can build trust in generative AI and leverage its full potential for content creation.	9.52381
11	Regulators need to take action to prevent Big Tech from consolidating their power and blocking competition. Wi...	2.7027
12	Between the fear & hype, marketers are looking to drive business results with new AI technologies. Join me ...	0
13	There is definitely an underlying meaning in these abstract art created with generative AI.	0
14	We use generative AI very carefully, respecting privacy and keeping humans in the loop. -- Centific's Wei Zhang,...	-2.22222
15	Generative AI Is Here: Forrester Offers Tech Execs Tips on Next Steps...Gen AI has arrived, and it's time to look at ...	1.72414
16	Solving the mystery of how ChatGPT and generative AI surprisingly pick up foreign languages, says AI Ethics and...	-1.88679
17	Generative AI Is Here: Forrester Offers Tech Execs Tips on Next Steps...Gen AI has arrived, and it's time to look at ...	1.96078
18	Augmented Reality and Generative AI: A Comprehensive Guide from Basics to Advance #AR #generativeAI #Aug...	6.45161
19	Generative AI is revolutionizing the art world by empowering artists to create unique and mesmerizing pieces wi...	2.38095
20	Empower your business with generative AI! Delve into the game-changing potential of using AI to automate the ...	2.77778

Figure 3. Corpus File Sentiment Analysis using Liu & Hu

Figure 3 shows the result of the sentiment analysis using the Liu & Hu sentiment analysis algorithm. The Liu and Hu Lexicon-based sentiment analysis presented the value based on the assessment of the sentiments in each document, as can be seen in the figure. The formula: Count of Positive Words + Count of Negative Words + 1 was used to calculate the sentiment score. By adding the document's positive and negative sentiments plus one, the lexicon-based approach to sentiment analysis evaluated each word in the document.

	Analysis	Text	sentiment
1	neutral	Is creativity at risk? As Generative AI enters the creator economy, we must ensure that human creativity and origin...	0
2	positive	Three Ways Media Leaders Can Leverage Generative AI https://t.co/1qPjNTbQTQ #documentai #generativeai #ver...	4.34783
3	neutral	Unbelievable AI Auto Designer.. I just signed up to try Autodesigner, a new generative AI by @uizard for UI/UX de...	0
4	positive	Generative AI promises to streamline health care thanks to Microsoft and Epic	8.33333
5	neutral	REG NOW #Roundtable on #GenerativeAI on 5/16 hosted @MayfieldFund in Menlo Park. Discover the opportuniti...	0
6	positive	As a Startup Attorney, I've combined Generative AI tools with my legal expertise to create winning outcomes for ...	5.26316
7	positive	I just signed up to try Autodesigner, a new generative AI by @uizard for UI/UX design. Looks incredible! https://t.c...	2.5641
8	positive	Engaged in productive meetings with photography association and research faculty, gathering insights on conte...	3.44828
9	neutral	Preparing for the future of AI: discover the latest developments in generative AI. #GenerativeAI #FutureTech #AI h...	0
10	positive	By taking these steps, businesses can build trust in generative AI and leverage its full potential for content creation.	9.52381
11	positive	Regulators need to take action to prevent Big Tech from consolidating their power and blocking competition. Wit...	2.7027
12	neutral	Between the fear & hype, marketers are looking to drive business results with new AI technologies. Join me a...	0
13	neutral	There is definitely an underlying meaning in these abstract art created with generative AI.	0
14	negative	We use generative AI very carefully, respecting privacy and keeping humans in the loop. -- Centific's Wei Zhang, ...	-2.22222
15	positive	Generative AI Is Here: Forrester Offers Tech Execs Tips on Next Steps...Gen AI has arrived, and it's time to look at u...	1.72414
16	negative	Solving the mystery of how ChatGPT and generative AI surprisingly pick up foreign languages, says AI Ethics and ...	-1.88679
17	positive	Generative AI Is Here: Forrester Offers Tech Execs Tips on Next Steps...Gen AI has arrived, and it's time to look at u...	1.96078
18	positive	Augmented Reality and Generative AI: A Comprehensive Guide from Basics to Advance #AR #generativeAI #Aug...	6.45161
19	positive	Generative AI is revolutionizing the art world by empowering artists to create unique and mesmerizing pieces wit...	2.38095
20	positive	Empower your business with generative AI! Delve into the game-changing potential of using AI to automate the ...	2.77778

Figure 4. Trained Datasets based on Sentiment Results

The trained dataset is presented in Figure 4. The document is in the corpus file which was viewed using the machine learning application. The sentiment per document was analyzed and the weight of each sentiment using color coding. The color coding for positive sentiments is green color, blue for negative, and red for neutral sentiments. Using the sentiment analysis scoring guide, the researchers were able to analyze properly and accurately the sentiments in the corpus file. The trained dataset is now ready for analysis of the sentiments using Data Classification algorithms such as SVM, Random Forest, and Naïve Bayes respectively.

Model	AUC	CA	F1	Precision	Recall
SVM	1.000	1.000	1.000	1.000	1.000
Random Forest	1.000	1.000	1.000	1.000	1.000
Naive Bayes	0.975	0.939	0.924	0.949	0.939

Figure 5. Test and Score Result

Figure 5 shows the test and score results on the model SVM, Random Forest, and Naïve Bayes algorithms. The result shows that the **SVM** and **Random Forest** have the same precision and recall **(1.000)** indicating that the result of the test has no false positives while the **Naïve Bayes** algorithm garnered a **.949** precision and **.939** recall which means fewer false positive results on the trained models.

	SVM	Random Forest	Naive Bayes
SVM		0.500	0.833
Random Forest	0.500		0.833
Naive Bayes	0.167	0.167	

Figure 6. Comparison Result of the three Algorithms

Figure 6 shows the probability that the score for the model row is higher than that of the model in the column. Small numbers show the probability that the difference is negligible. The cross-validation splits the data into a given number of folds (3 folds). The algorithms were tested by holding out examples from one fold at a time; the model is induced from other folds and examples from the held-out fold were classified. This was repeated in all the folds in the data.

The researcher used a cross-validation technique on the trained model to test its accuracy. Also, the models were tested using the recall classification to test the proposition of the true positives among all the positive instances in the data. The result shows that the SVM evaluation on the Random Forest algorithm was .500 and Naïve Bayes with .833. The Random Forest evaluation on SVM was .500 and Naïve Bayes with .833. Lastly, the Naïve Bayes evaluation on SVM and Random Forest was .167. This indicates that the model trained was successful in evaluating the probability of having a false positive on the result of the test and score.

		Predicted			
		negative	neutral	positive	Σ
Actual	negative	2	6	0	8
	neutral	0	32	0	32
	positive	0	0	59	59
	Σ	2	38	59	99

Figure 7. Confusion Matrix for Naïve Bayes Algorithm

The result of a prediction using the confusion matrix and the Naïve Bayes algorithm is shown in Figure 7. It shows how many or what percentage of instances belong to the actual class as opposed to the predicted class. The estimated value of each sentiment was displayed in the figure. Negative emotions totaled 8 but were predicted to be 2, with a wrong value of 6. The neutral sentiments prediction was 32 with no incorrect predictions. Finally, the positive sentiments were 59 with no incorrect predictions. Because more models would decrease the accuracy of the final result, the researchers only used 99 trained models to generate an accurate prediction.

		Predicted			
		negative	neutral	positive	Σ
Actual	negative	8	0	0	8
	neutral	0	32	0	32
	positive	0	0	59	59
	Σ	8	32	59	99

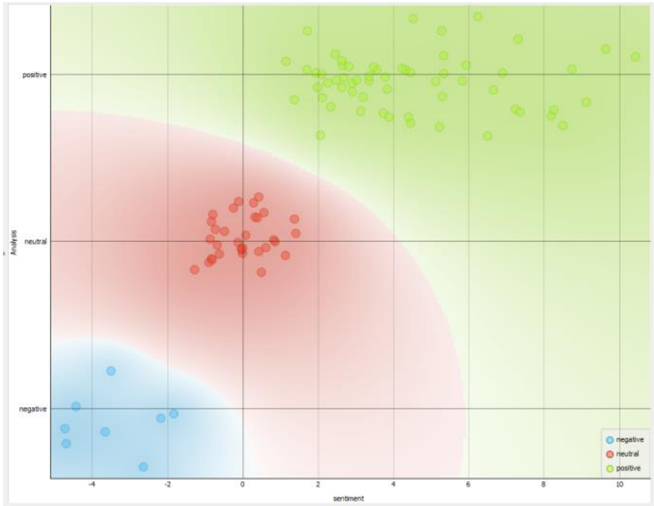
Figure 8. Confusion Matrix for Random Forest Algorithm

The result of a prediction made using the confusion matrix and the Random Forest algorithm is shown in Figure 8. It shows how many or what percentage of instances belong to the actual class as opposed to the predicted class. The estimated value of each sentiment was displayed in the figure. Negative emotions totaled 8 but were predicted to be 8, with a wrong value of 0. The neutral sentiments prediction was 32 with no incorrect predictions. For the positive sentiments was 59 with no incorrect predictions. Because more models would decrease the accuracy of the final result, the researchers only used 99 trained models to generate an accurate prediction.

Actual \ Predicted	negative	neutral	positive	Σ
negative	8	0	0	8
neutral	0	32	0	32
positive	0	0	59	59
Σ	8	32	59	99

Figure 8. Confusion Matrix for SVM Algorithm

Figure 8 displays the outcome of a prediction made using the confusion matrix and the SVM algorithm. It displays the quantity/proportion of instances in the actual class compared to the predicted class. The figure shows the estimated value of each sentiment. Positive sentiments were predicted as 8 but with a total of 8 and a wrong value of 0. With no incorrect predictions, the neutral sentiments prediction was 32. Finally, the positive value was 59 and there were no incorrect predictions, giving a final score of 99 The researchers only used 99 trained models to produce an accurate prediction because using more models would reduce the accuracy of the final result.



Scatter plot visualization with enhancements for intelligent data visualization and exploratory analysis. A 2-dimensional scatter plot visualization is offered by the Scatter Plot widget. Each point in the data display has a value for the x-axis attribute that determines its position on the horizontal axis and a value for the y-axis attribute that determines its position on the vertical axis.

The scatter plot shows that positive sentiments are dominant in the documents. Neutral sentiments are also present together with the negative sentiments but are less compared to the positive sentiments. It is clearly stated in the graph that the topic "Generative Artificial Intelligence" has a more positive sentiment in tweeter posts.

CONCLUSIONS

Based on the results analyzed by the researchers, the following conclusions were drawn: (1) The Kaggle datasets used by the researcher were utilized to get the Twitter posts of the users on the topic "Generative Artificial Intelligence" which is free of charge; (2) Corpus file was successfully used for transforming Kaggle datasets into a document that machine learning application can read and write; (3) Sentiment Analysis was successfully utilized to the trained models for the application of the Naïve Bayes algorithm analysis using machine learning techniques; (4) SVM, Random Forest, and Naïve Bayes algorithms gained the same predictions using machine learning techniques; and (5) Positive emotions or sentiments were analyzed on the datasets on the topic "Generative Artificial Intelligence".

Moreover, some recommendations were drawn based on the derived conclusions that Kaggle datasets can be used for obtaining Twitter datasets for machine learning evaluation. The designed sentiment analysis framework yielded a successful result showing the test and result of the testing on the trained datasets. Also, the three (3) algorithms used in the prediction process show positive results. Lastly, there were more positive sentiments on the topic of "Generative Artificial Intelligence" extracted from tweeter posts and analyzed by a machine learning application.

ACKNOWLEDGMENT

The researcher would like to extend his heartfelt thanks to the AMA Computer University, Quezon City administration, and teaching force for allowing him to conduct this kind of study as part of his academic requirements. Special thanks to his professors and mentors who taught him in the different areas of Machine Learning.

REFERENCES

1. Abbasi, A., Chen, H., & Salem, A. (2008). Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums. *ACM Trans. Inf. Syst.*, *26*(3). https://doi.org/10.1145/1361684.1361685
2. Abbasi, A., Javed, A. R., Chakraborty, C., Nebhen, J., Zehra, W., & Jalil, Z. (2021). ElStream: An Ensemble Learning Approach for Concept Drift Detection in Dynamic Social Big Data Stream Learning. *IEEE Access*, *9*, 66408–66419.

https://doi.org/10.1109/ACCESS.2021.3076264

3. Abbasi, A., Javed, A. R., Iqbal, F., Jalil, Z., Gadekallu, T. R., & Kryvinska, N. (2022). Authorship identification using ensemble learning. *Scientific Reports*, *12*(1), 9537. https://doi.org/10.1038/s41598-022-13690-4
4. Adiguzel, T., Kaya, M. H., & Cansu, F. K. (2023). Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. *Contemporary Educational Technology*, *15*(3). https://doi.org/10.30935/cedtech/13152
5. Aggarwal, D. G. (2018). Sentiment Analysis An insight into Techniques, Application and Challenges. *International Journal of Computer Sciences and Engineering*, *6*(5), 697–703. https://doi.org/10.26438/ijcse/v6i5.697703
6. Aized Amin Soofi, & Arshad Awan. (2017). Classification Techniques in Machine Learning: Applications and Issues. *Journal of Basic & Applied Sciences*, *13*(August), 459–465. https://doi.org/10.6000/1927-5129.2017.13.76
7. Amanat, A., Rizwan, M., Javed, A. R., Abdelhaq, M., Alsaqour, R., Pandya, S., & Uddin, M. (2022). Deep Learning for Depression Detection from Textual Data. *Electronics (Switzerland)*, *11*(5), 1–13. https://doi.org/10.3390/electronics11050676
8. BAİDOO-ANU, D., & OWUSU ANSAH, L. (2023). Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. *Journal of AI*, *7*(1), 52–62. https://doi.org/10.61969/jai.1337500
9. Bashir, M. F., Javed, A. R., Arshad, M. U., Gadekallu, T. R., Shahzad, W., & Beg, M. O. (2022). Context Aware Emotion Detection from Low Resource Urdu Language using Deep Neural Network. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *April*. https://doi.org/10.1145/3528576
10. Bhattacharya, K., Bhattacharya, A. S., Bhattacharya, N., Yagnik, V. D., Garg, P., & Kumar, S. (2023). ChatGPT in Surgical Practice—a New Kid on the Block. *Indian Journal of Surgery*, *85*(6), 1346–1349. https://doi.org/10.1007/s12262-023-03727-x
11. Bhukya, D. P., & Ramachandram, S. (2010). Decision Tree Induction: An Approach for Data Classification Using AVL-Tree. *International Journal of Computer and Electrical Engineering*, *2*(4), 660–665. https://doi.org/10.7763/ijcee.2010.v2.208
12. Chen, H., & Zimbra, D. (2010). AI and Opinion Mining. *IEEE Intelligent Systems*, *25*(3), 74–80. https://doi.org/10.1109/MIS.2010.75
13. Dave, K., Lawrence, S., & Pennock, D. M. (2003). Mining the peanut gallery: opinion extraction and semantic classification of product reviews. *Proceedings of the 12th International Conference on World Wide Web*, 519–528. https://doi.org/10.1145/775152.775226
14. Dictionary, C. (2021). *Sentiment definition*. https://dictionary.cambridge.org/dictionary/english/sentiment
15. Eggmann, F., Weiger, R., Zitzmann, N. U., & Blatz, M. B. (2023). Implications of large language models such as ChatGPT for dental medicine. *Journal of Esthetic and Restorative Dentistry*, *35*(7), 1098–1102. https://doi.org/https://doi.org/10.1111/jerd.13046
16. Feng, S., & Shen, Y. (2023). ChatGPT and the Future of Medical Education. *Academic Medicine*, *98*(8), 867–868. https://doi.org/10.1097/ACM.0000000000005242
17. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Networks. *Advances in Neural Information Processing Systems*, *3*. https://doi.org/10.1145/3422622
18. Hina, M., Ali, M., Javed, A. R., Ghabban, F., Khan, L. A., & Jalil, Z. (2021). SeFACED: Semantic-Based Forensic Analysis and Classification of E-Mail Data Using Deep Learning. *IEEE Access*, *9*, 98398–98411. https://doi.org/10.1109/ACCESS.2021.3095730
19. Hina, M., Ali, M., Javed, A. R., Srivastava, G., Gadekallu, T. R., & Jalil, Z. (2021). Email Classification and Forensics Analysis using Machine Learning. *Proceedings - 2021 IEEE SmartWorld, Ubiquitous Intelligence and Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Internet of People, and Smart City Innovations, SmartWorld/ScalCom/UIC/ATC/IoP/SCI 2021*, 630–635. https://doi.org/10.1109/SWC50871.2021.00093
20. Hu, K. (2023). *ChatGPT sets record for fastest-growing user base—analyst note*. Reuters. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/
21. Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 168–177. https://doi.org/10.1145/1014052.1014073
22. Isabelle, G., Maharani, W., & Asror, I. (2019). *Analysis on Opinion Mining Using Combining Lexicon-Based Method and Multinomial Naïve*

Bayes. *2*(IcoIESE 2018), 214–219. https://doi.org/10.2991/icoiese-18.2019.38

23. Kaur, G., . P., Kaur, A., & Khurana, M. (2022). A Review of Opinion Mining Techniques. *ECS Transactions*, *107*(1), 10125. https://doi.org/10.1149/10701.10125ecst
24. Kesavaraj, G., & Sukumaran, S. (2013). A study on classification techniques in data mining. *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, 1–7.
25. Kheiri, K., & Karimi, H. (2023). *SentimentGPT: Exploiting GPT for Advanced Sentiment Analysis and its Departure from Current Machine Learning*. http://arxiv.org/abs/2307.10234
26. Kitamura, F. C. (2023). ChatGPT Is Shaping the Future of Medical Writing But Still Requires Human Judgment. *Radiology*, *307*(2), e230171. https://doi.org/10.1148/radiol.230171
27. Krugmann, J. O., & Hartmann, J. (2024). Sentiment Analysis in the Age of Generative AI. *Customer Needs and Solutions*, *11*(1). https://doi.org/10.1007/s40547-024-00143-4
28. Mathapati, P. M., Shahapurkar, A. S., & Hanabaratti, K. D. (2017). Sentiment Analysis using Naïve bayes Algorithm. *International Journal of Computer Sciences and Engineering*, *5*(7), 75–77. https://doi.org/10.26438/ijcse/v5i7.7577
29. Mouthami, K., Devi, K. N., & Bhaskaran, V. M. (2013). Sentiment analysis and classification based on textual reviews. *2013 International Conference on Information Communication and Embedded Systems (ICICES)*, 271–276. https://doi.org/10.1109/ICICES.2013.6508366
30. Patacsil, F. F., Malicdem, A. R., & Fernandez, P. L. (2015). Estimating Filipino ISPs Customer Satisfaction Using Sentiment Analysis. *Computer Science and Information Technology*, *3*(1), 8–13. https://doi.org/10.13189/csit.2015.030102
31. Pisani, S., Fioriti, D., Conte, M. P., Chiarini, F., Seganti, L., & Degener, A. M. (2002). Involvement of herpes simplex virus type 2 in modulation of gene expression of human papillomavirus type 18. *International Journal of Immunopathology and Pharmacology*, *15*(1), 59–63. https://doi.org/10.1177/039463200201500108
32. Preotiuc-Pietro, D., Hopkins, D. J., Liu, Y., & Ungar, L. (2017). Beyond binary labels: Political ideology prediction of twitter users. *ACL 2017 - 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, *1*, 729–740. https://doi.org/10.18653/v1/P17-1068
33. Rachiraju, S. C., & Revanth, M. (2020). Feature Extraction and Classification of Movie Reviews using Advanced Machine Learning Models. *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 814–817. https://doi.org/10.1109/ICICCS48265.2020.9120919
34. Rafat, K. F., Xin, Q., Javed, A. R., Jalil, Z., & Ahmad, R. Z. (2022). Evading obscure communication from spam emails. *Mathematical Biosciences and Engineering*, *19*(2), 1926–1943. https://doi.org/10.3934/mbe.2022091
35. Razali, N. A. M., Malizan, N. A., Hasbullah, N. A., Wook, M., Zainuddin, N. M., Ishak, K. K., Ramli, S., & Sukardi, S. (2021). Opinion mining for national security: techniques, domain applications, challenges and research opportunities. *Journal of Big Data*, *8*(1), 150. https://doi.org/10.1186/s40537-021-00536-5
36. Shah, P., Swaminarayan, P., & Patel, M. (2023). Sentiment Classification for Film Reviews in Gujarati Text Using Machine Learning and Sentiment Lexicons. *Journal of ICT Research and Applications*, *17*(1), 1–16. https://doi.org/10.5614/itbj.ict.res.appl.2023.17.1.1
37. Simmering, P. F., & Huoviala, P. (2023). *Large language models for aspect-based sentiment analysis*. *2022*, 1–12. http://arxiv.org/abs/2310.18025
38. Singh, M., Sharma, S., & Kaur, A. (2013). Performance Analysis of Decision Trees. *International Journal of Computer Applications*, *71*, 10–14.
39. Surroop, K., Canoo, K., & Pudaruth, S. (2016). A Novel Position-based Sentiment Classification Algorithm for Facebook Comments. *International Journal of Advanced Computer Science and Applications*, *7*(10), 261–268. https://doi.org/10.14569/ijacsa.2016.071035
40. Tang, H., Tan, S., & Cheng, X. (2009). A survey on sentiment detection of reviews. *Expert Systems with Applications*, *36*(7), 10760–10773. https://doi.org/https://doi.org/10.1016/j.eswa.2009.02.063
41. Troussas, C., Virvou, M., Espinosa, K. J., Llaguno, K., & Caro, J. (2019). Sentiment analysis of Facebook statuses using Naive Bayes Classifier for language learning. *IISA 2013 - 4th International Conference on Information, Intelligence, Systems and Applications*, 198–205. https://doi.org/10.1109/IISA.2013.6623713
42. Wang, Z., Xie, Q., Feng, Y., Ding, Z., Yang, Z., & Xia, R. (2023). *Is ChatGPT a Good Sentiment Analyzer? A Preliminary Study*.

http://arxiv.org/abs/2304.04339

43. Xu, F., Pan, Z., & Xia, R. (2020). E-commerce product review sentiment classification based on a naïve Bayes continuous learning framework. *Information Processing & Management*, *57*(5), 102221. https://doi.org/https://doi.org/10.1016/j.ipm.2020.102221
44. Xu, H. (2018). Generate Faces Using Ladder Variational Autoencoder with Maximum Mean Discrepancy (MMD). *Intelligent Information Management*, *10*(04), 108–113. https://doi.org/10.4236/iim.2018.104009
45. Zhang, C., Zuo, W., Peng, T., & He, F. (2008). Sentiment Classification for Chinese Reviews Using Machine Learning Methods Based on String Kernel. *2008 Third International Conference on Convergence and Hybrid Information Technology*, *2*, 909–914. https://doi.org/10.1109/ICCIT.2008.51.