Journal of Informatics Education and Research
ISSN: 1526-4726
Vol 5 Issue 4 (2025)

Hybrid Sentiment Analysis Using Fine-Tuned Pre-Trained Language Models for Domain-Specific Insights and Cross-Industry Adaptability

Jyoti Sharma

Research scholar

IcfaiTech

The ICFAI University, Jaipur, India

Dr. Rana Mukherji

Professor

IcfaiTech (Faculty of Science and Technology)

The ICFAI University, Jaipur, India

Abstract:

Through the development of user-generated content on online media, there has been an increased pressure to carry out a perfect sentiment analysis particularly in the specialized sectors. The sentiment analysis methodologies are not applied in practice to the domain specificity and smooth or expressed sentiment. In this paper, a hybrid sentiment analysis model which incorporates pre-trained language models (e.g., BERT and GPT) that combine support vector machines and random forests as a component of hybridization to improve performance and provide greater flexibility in terms of sentiment detection is presented. This paper is a systematic analysis of whether the hybrid model is able to estimate intensive sentiment or otherwise, this is achieved by comparing it to the traditional hybrid research design on different industries and perspectives. It also takes into account the fact that it may be generalized outside the small scope within which the measurements were regulated. The experimental findings of the given study will be more precise, versatile, and cross-implemented to fields. All these extended measures functions together ought to provide an all-encompassing strategy to sentiment-based decision-making in specific conditions to the researchers and professionals.

Keywords: Hybrid sentiment analysis, fine-tuned language models, BERT, GPT, domain-specific sentiment, cross-industry adaptability, transfer learning, natural language processing

1. Introduction

People run their feelings across the internet every day; reaction bits of emotion in the manifestation of the word. They recount their experiences, remarks on politics, write on their health, talk on their victorious or lament about their defeat. Behind all of that din lies something good the true feelings of people. Identifying that, or reading between the lines on a large lever-scale, is what sentiment analysis is trying to realize. It is among the most intriguing aspects of Natural Language Processing (NLP) [1] since one of the purely human qualities, such as emotion, opinion, tone, is also involved. But there's a problem. The way people express feeling also completely depends on the location of the conversation and the subject of discussion. The language of a stock market report can not be equated to what a person writes when he/she has visited a hospital [2]. Even one word can reverse the entire significance in different disciplines. Take "critical." In a medical note this may be taken to mean that one is not in an excellent condition. It can describe intelligent reasoning in an activity of watching a film. In product review the same can be of another meaning altogether. Most sentiment models cannot reflect such changes, they are so inflexible in their approach to language, out into the context and this is where they fail.

Researchers have over the years tried to cure this with the usage of dictionaries of positive and negative words or the formation of hand designed features that were trying to obtain hints of emotion. It worked okay, but not great. then profound minds change everything. The area developed with the introduction of such a transformer model as BERT and GPT [3]. These models are not just the counting of the words, but they know them, or at least approximate. They will be in a position to comprehend the change in meaning depending on the context and how a sentence feels relative to its context. This kind of knowledge was quite powerful.

Still, they're not magic. The BERT and the GPT might malfunction even in cases where the data has shifted [4]. They might be able to do immensely well with general samples like movie reviews or tweets but will not do so with a more specialised thing like sentiment analysis of legal contracts or even on clinical notes. The rationale is to come up with a hybrid sentiment analysis model, and it is not rooted in one type of intelligence. It is an integration of the robust contextual learning and highly refined transformers like BERT and GPT with robust grounded reasoning of classical machine learning models like Support Vector Machine (SVM) [5] and Random Forests (RF) [6]. Transformers are supposed to capture meaning and more decisions on control can be made by the traditional classifiers. Their combination may be a balance between the intuitiveness and the structure which is more likely similar to the way humans perceive emotion in a text.

It is not to simply extract a little more truth out of it, which is essential. It is in order to have the system adaptable, which is able to manoeuvre the boundaries of one region to the other, one language of one industry to the language of another. Ideally, the model, which was trained to have an understanding of the financial speech, should not fall apart after reading a healthcare response or a text in the social media. Such flexibility is viewed as being quite uncommon though necessary in the situation when sentiment analysis is required to be employed beyond research papers. The advantages in a practical manner can only be imagined. This model would likely find an imperceptible move in the mood of the investors in the financial environment before it is revealed in the market [7]. In the medical field, it would be less insensitive to reviews by patients and trace the levels of inadequacy or trepidation that would not be apparent in numbers. It would have the capability of differentiating sarcasm and actual applause in the e-commerce space. It can make use of the mood of the people in hourly fashion in the social media.

Hypothetically, the research takes a gauging point of interest. The tradition machine learning was always more or less similar to the deep one, which is more explicit yet lower-level representation of meaning comparable to the black box, purely-decision making. This hybridized model is an attempt to overcome the gap. Nor does it concern one or the other. It entails using both in a manner which manifests usage of context, structure and logic are all being used. The other noteworthy aspect of this work is the flexibility in domains. The real world is not so nice - the information is not distributed on one location and the writing type of individuals is constantly changing. A model that was successful somewhere cannot offer much good to a different setting. The real intelligence is the capacity to be even more than it has been acquainted to be.

Originally, it is a discussion of the next level of sentiment analysis, to something a little more human, a little more adaptable, a little more receptive to subtlety, a little less dependent on the datasets upon which it was trained. The paper targets the results of creating a model that solely identifies emotion but its form and texture employing domains by fine-tuned BERT and GPT embeddings combined with SVM and theRandom Forest classifier.

The objectives of research are as follows

1. To create a hybrid sentiment analysis model with transformer models (e.g. BERT, GPT) with traditional classifiers.
2. To deepen sentiment detection from only being able to detect positive or negative sentiment to also being able to assess the tone and emotions being expressed in text.
3. To evaluate the model performance across different domains (e.g. finance, health care, online reviews).
4. To evaluate if performance transfers across industries.
5. To create a functioning tool, position individuals and organizations to make informed decisions based upon public opinion.

2. Introduction

Pookduang, P. et al. (2025) [8] conducted a study comparing various models of sentiment analysis using Amazon book reviews. They examined many sentiment analysis models, comparing traditional machine learning algorithms (i.e., Naïve Bayes and KNN) with deep learning and transformer-based methods. They found that RoBERTa achieved the highest level of precision (96.30% accuracy; 98.11% F1 score) when compared to other machine learning methods (Naïve Bayes, KNN, and LSTM). The study concluded transformer methods are a superior method of performing sentiment analysis that traditional sentiment analysis frameworks may be limited within user-generated content via its potential to capture semantic meaning, including complicated patterns within user-generated language. **Hadi, M. F. et al. (2025)** [9] investigated the capacity of sentiment analysis to identify the early onset of mental health conditions through social media post analysis.The researchers utilized a hybrid model combining BERT embeddings with an SVM classifier to categorise social media posts as "positive," "negative," or "neutral", providing very strong overall precision, while overall accuracy improvements were

below 97%. Even with the improvements in accuracy and performance of hybrid architectures within the sentiment analysis literature, specifically for and in applications to significant mental health issues, their study articulated several challenges derived from objective noisiness and structured text inputs. The authors noted the need for new models that generalise the ability to adapt cross-disciplines - a challenge directly addressed by the hybrid system of BERT-GPT proposed in the present paper.

Khan, S. I., et al. (2023) [10] presented frameworks that associate with BERT, XLNet, and Electra to be evaluated on the dataset Sentiment140 that amounts to over 1.6 million labeled tweets. Examining the performance of the model in the classification of customer sentiments, they looked at other performance metrics that illustrated an interaction between model fine-tuning and computational resources in trade-offs. **Prottasha, N. J., et al. (2022) [11]** addressed the issues of sentiment analysis in the Bangla NLP domain as there was a scarcity of labeled data. They used transfer learning with BERT within a CNN-BiLSTM model and achieved better binary classification performance than traditional algorithms, thereby contributing to the sentiment analysis of low-resource languages. **Dang, C. N., et al. (2021) [12]** highlighted the usage of hybrid models in sentiment analysis in social networks with the help of LSTM, CNN, and SVM. Their work demonstrated enhanced reliability and accuracy of these hybrid models over various datasets, which emphasizes the importance of performance and computation time. **De Arriba, A., et al. (2021) [13]** proposed a transfer learning-supported framework for sentiment analysis in software user messages. Their experiments brought out the performance of various machine learning models as well as the factors influencing the performance.

3. Research Methodology

Utilizing a systematic experiment approach, the study sought, to develop, train, and evaluate a hybrid sentiment analysis framework that is founded upon a fine-tuned pre-trained transformer model [14] (e.g., BERT and GPT)], along with traditional machine-learning classifiers [15](e.g., Support Vector Machine and Random Forest) . The goal of the experimental approach was to develop a system apt at consistently and accurately grasping and interpreting sentiments across an increasingly diverse set of domains while, at the same time, providing adaptability between industries [16]. The experimental study used a variety of readily-versions publicly available text datasets for different industries, such as finance, healthcare, e-commerce, and social media [17]. Each dataset consisted of text samples that were labelled with sentiment such as positive, negative or neutral. Using data sources representative of different domains was important to test the model performance in each domain, and also examine the cross-domain adaptability. Prior to model training, each dataset underwent a thorough text preprocessing stage to ensure quality and consistency as a dataset. The procedure involved the eliminating of the irrelevant features of the data such as special characters, punctuation, links, and stop words. All text was made lower case and the text was then tokenised to break up the sentences into smaller units. Lemmatisation was employed to put words back into their root form. In the case of emojis or emoticons, the words that described their meaning were applied, as these can carry an important emotional component at times [18].

Once the data and data preparation were cleaned, it was further subdivided into training and also the validation, as well as testing, in order to determine how fair the testing of the models was. The second step was the model development which was based on the refined transformer models. To begin with, the BERT and GPT were conditioned to text domain, and, therefore, the models were able to adapt their knowledge to a given style of language and apply the relevant terminology [19]. The BERT and GPT generated representations may be referred to as embeddings that may generate a powerful contextual meaning of the text. Neither were used as a crude source of data to determine a sentiment, but were part of more complicated machine learning procedures, such as Support Vector Machine and a Random Forest. The design of this was to take advantage of the power of both worlds in which the role of transformers gave contextual illumination and semantic revelation, and the role of classical algorithms gave force and content to the process of making decisions in the science and informing the decision system. The rationale behind this design was that it sought to exploit the power of both worlds, where the transformers were stronger in terms of providing a more enriched context of comprehending the process of decision-making and semantics, and the classical algorithms were powerful and understandable of the process in decision-making in the science, either through consideration of the decision-making or in the framing of the decision making [20].

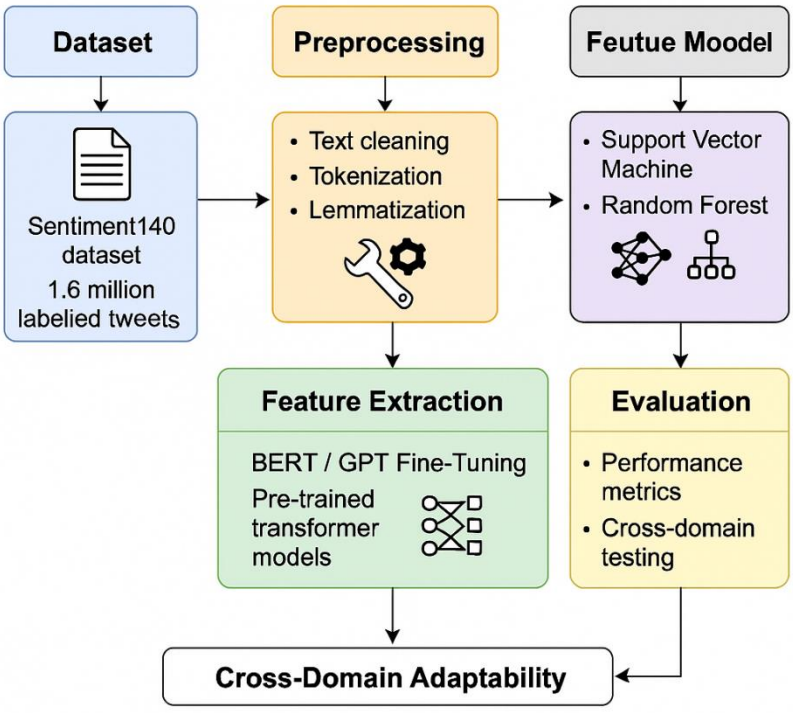

Figure 1. Research Flow Diagram

Finally, labelled data was used to make hybrid models as listed in Figure 1 and parameters were tuned to minimise classification error. The two models were trained on a single domain and tested on this domain and other domains, to examine the possibility or otherwise, of the existence of generalisation in these models. To provide an example, one of the models previously trained in e-commerce review was tested in the field that was not related to e-commerce such as healthcare to examine the model performance under the condition of changed writing style. Interdomain testing mattered to find out whether we can extrapolate our framework to any other area of work at all without any retraining on the same that is worth mentioning.

All models were assessed using multiple industry-standard metrics, including accuracy, precision, recall, and F1-score, to provide a broader understanding of the effectiveness of the model in terms of sentiment classification. We also looked at confusion matrices to better understand the areas of strong performance and weaknesses in particular areas, especially differentiating neutral from mixed sentiment cases. Alongside the hybrid approach outlined in this paper, we also built some baseline models we could compare our work against, including transformer-only systems and traditional machine learning classifiers that was trained on simpler text representations (i.e., TF-IDF, or embeddings such as Word2Vec) [22]. By looking at the performance of these systems, we could provide more detail as to the particular and unique performance that the hybrid model provides, especially in high accuracy across domains while also providing a more nuanced assessment compared to all other systems developed. The entire study was conducted Programmatic in Python with Libraries such as transformers, scikit-learn, pandas and numpy with slight facilitation of fine-tuning deep learning models in Pytorch or TensorFlow and computational efficiency was enhanced with the uses of GPU resource [23]. Reproducibility and systematic documentation were considered throughout our experiments. We did not collect any personal or sensitive information as all datasets used in our research were publically available and we took ethical considerations into account at all times.

In summary, our methodology is grounded in the integration of the contextual intelligence of fine-tuned language models with the structured reasoning of traditional classifiers. Through integrating both techniques and evaluation of them in multiple datasets across multiple domains, we created a sentiment analysis framework that can provide prediction that is equally accurate, interpretable, and flexible to real-world applications in all industries. This structured yet flexible approach is the basis in which we can evaluate the effectiveness of the proposed hybrid sentiment analysis system and to summarize our work of developing a hybrid sentiment analysis is that it has been systematic developed to be more generalizable in numerous ways in the changing context of its application.

4. Proposed Work

4.1 Dataset

The Sentiment140 collection, which consists of 1.6 million tweets, was developed by Alec Go, Richa Bhayani, and Lei Huang at Stanford University in 2009 to provide researchers with sentiment analysis data on social media platforms (Twitter) [23]. This collection is an example of distant supervision, in which tweets are automatically labelled based on the emoticons they contain. For example, the positive emoticons, like the smiley ":)", are then tagged as positive (4), while the negative emoticons like the frown ":(" are tagged as negative (0). Each row in the dataset includes five columns, the sentiment label, tweet ID, date, query term, username, and tweet text; but only the sentiment label and text are relevant in this analysis. The dataset's contribution stems from the real-world, noisy, and informal language one would expect to see in tweets and social media conversation, with slang and abbreviations, emojis, and some sarcasm, all of which are typical attributes in real life social interactions. Sentiment140 has gained prominence because of this size and diversity of tweets, as it has become the standard benchmark to develop and refine modern transformer-based models such as BERT, RoBERTa, and GPT [23], and train hybrid systems that leverage sentiment analysis tools that combine both deep learning methods as well as traditional- or lexical-based approaches. The linguistic richness and scale of the Sentiment140 dataset represent a valuable dataset for exploring opinion mining and improving domain adaptation in sentiment analysis research.

4.2 BERT

BERT (Bidirectional Encoder Representations from Transformers) [24] is a deep learning model created by Google in 2018 for interpreting the meaning and context of the words in a sentence. Unlike previous models, which were unidirectional, BERT reads text both left to right and right to left, which allowed it to read the complete context of the language. This bidirectional function allows it to learn more subtle meanings, relationships among words, as well as sentence level understandings. BERT is pre-trained on large text corpora (e.g., Wikipedia), and for specific tasks later fine-tuned, including sentiment analysis, question answering, and text classification tasks. BERT's [24] learned ability to understand the context of the text can be beneficial for determining emotional tone and positive or negative polarity in text data.

4.3 GPT

GPT (Generative Pre-trained Transformer) [25], which is also developed by OpenAI, is yet another powerful transformer-based model designed to generate and understand natural language. Unlike BERT, which reads text unidirectionally, GPT reads text unidirectionally (left to right), and has been pretrained primarily on language generation tasks (text completion, summarisation, dialogue generation). When fine-tuned, however, to perform sentiment classification is that it can use the learned embeddings to predict the tone or mood behind the words, phrases, and sentences. Additionally, like BERT, GPT also excels at producing coherent responses in humanish phrasing. A strong feature of GPT [25] is that it is capable of capturing deep contextual meaning from the sequentially presented text. GPT's generative design is also conducive to domain-specific language, which makes it useful for hybrid models; models that combine understanding and generation abilities.

Hybrid Sentiment Analysis using Fine-Tuned BERT and GPT with SVM and Random Forest:

Require: Dataset $D = \{(x_i, y_i)\}_{i=1}^{N}$ of tweets x_i with labels $y_i \in \{0,4\}$ (0 = negative, 4 = positive)

Ensure: Predicted sentiment $\hat{y}$ for each tweet in the test set

Preprocessing

for each tweet x_i in D do

Remove URLs, mentions, hashtags, numbers, punctuation, and stop words

Convert to lowercase; tokenize; lemmatize

end for

Split D into train D_{train}, validation D_{val}, and test D_{test}

Fine-tuning and Embeddings

Fine-tune BERT on D_{train}; obtain embeddings $E_{\text{BERT}}^{(i)}$ for each x_i

Fine-tune GPT on D_{train}; obtain embeddings $E_{\text{GPT}}^{(i)}$

Build hybrid feature $E^{(i)} = \left[E_{\text{BERT}}^{(i)} \parallel E_{\text{GPT}}^{(i)}\right]$

Hybrid Classifiers

Train SVM on $\{(E^{(i)}, y_i)\}_{D_{\text{train}}}$ to get model M_{SVM}

Train Random Forest on the same to get model M_{RF}

Validation and Weighting

On D_{val}, get class probabilities $P_{\text{SVM}}(y|E)$ and $P_{\text{RF}}(y|E)$

Choose weights $\alpha, \beta \geq 0$, $\alpha + \beta = 1$, that maximise F1 on D_{val}

Inference

for each tweet x in D_{test} do

Compute $E = [E_{\text{BERT}} \parallel E_{\text{GPT}}]$

$$\hat{y} \leftarrow \text{argmax}_{y \in \{0,4\}}\left(\alpha\, P_{\text{SVM}}(y|E) + \beta\, P_{\text{RF}}(y|E)\right)$$

End for

Evaluation

Report Accuracy, Precision, Recall, F1; show confusion matrix

Optionally perform cross-domain testing by applying the trained model to other domains

4.4 Implementation

To develop the proposed hybrid sentiment analysis framework, Python was used as a primary development language. Pre-trained transformer models: BERT and GPT, were fine-tuned using the Hugging Face Transformers library, whereas the classifiers: Support Vector Machine (SVM) and Random Forest (RF) [26] were implemented in Scikit-learn. The first step was cleaning and preprocessing the Sentiment140 dataset to eliminate noise, such as punctuation, URLs, and stop words followed by tokenization and lemmatization. Then, the fine-tuned BERT and GPT models produced contextual embeddings, and sentence embeddings were concatenated with BERT and GPT embeddings (combined) to produce a hybrid feature vector. The hybrid feature vector was the input to the SVM and Random Forest classifiers. The final sentiment output was produced by merging the predictions of the classifiers using a soft voting mechanism. The experiments were conducted in a Google Colab GPU-enabled environment and through several metrics: accuracy, precision, recall, and F1-score, the model's performance and adaptability were evaluated.

4.5 Results

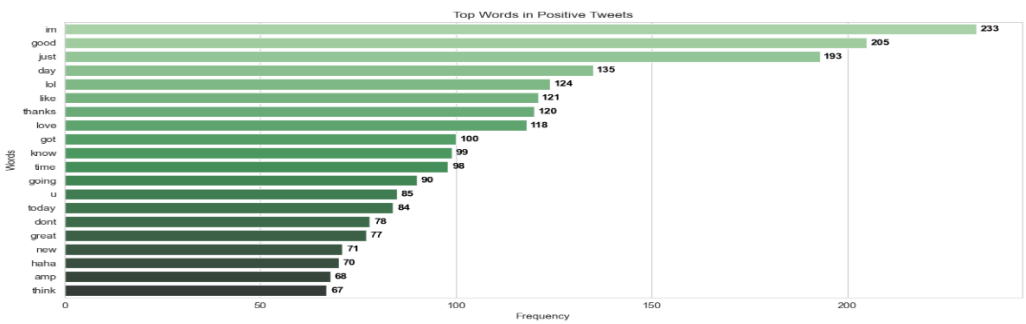

Figure 2. Frequent Words Occurring in Positive Tweets

The bar graph in Figure 2 features the most common words in the positive tweets of the Sentiment140 dataset. The most frequent words used in positive tweets ("im," "good," "just," "day," "lol," and "love") suggest a light-hearted, a positive tone. The words ("thanks," "great," and "haha") also capture the fact that positive tweets often include expressions of appreciation, humor, and friendliness. The model demonstrated its ability to detect genuine emotions a real human could have in unaltered social media text.

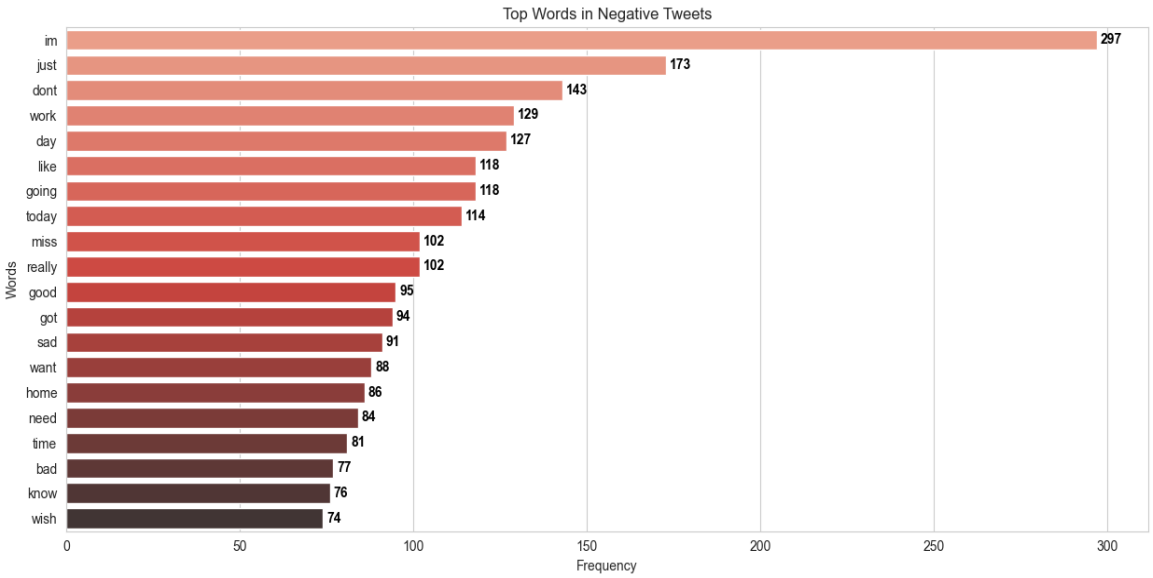

Figure 3. Frequent Words occurring in Negative Tweets

The figure 3 highlights the frequency of words appearing in negative tweets. The most common words, ("im," "just," "dont," "work," "miss," "sad," and "bad") suggest emotional distress, disappointment, or complaints. The repeated use of "work" and "need", suggest that the negativity either indicates situational distress, or associated with some level of stress. These observations suggest negative tweets tended to capture and convey strong emotional markers that the model was able to detect and classify.

text	clean_text
@chrishasboobs AHHH I HOPE YOUR OK!!!	ahhh hope ok
@misstoriblack cool , i have no tweet ap ps for my razr 2	cool tweet apps razr
@TiannaChaos i know just family drama. its lame.hey next time u hang out with kim n u guys like have a sleepover or wha tever, ill call u know just family drama lamehey time u hang kim n u guys like sleepover ill u	
School email won't open and I have geography stuff on there to revise! *S tupid School* :'(	school email wont open geography stuff revise stupid school
upper airways problem	upper airways problem
Going to miss Pastor's se rmon on Faith...	going miss pastors sermon faith
on lunch....dj should come eat with me	lunchdj come eat
@piginthepoke oh why are you fe eling like that?	oh feeling like
gahh noo!peyton needs to live! this is horrible	gahh noopeyton needs livethis horrible
@mrstessyman thank you glad you like it! There is a product review bit on the site E njoy knitting it!	thank glad like product review bit site enjoy knitting

Figure 4. Output from Text Cleaning and Preprocessing

This figure 4 illustrates the process of cleaning and standardising raw, noisy tweets in the text preprocessing step. The "text" column shows the original tweets with punctuation, emojis, usernames, and other irrelevant symbols, whereas the "clean_text" column demonstrates the cleaned data that would be used for analysis. The cleaning process helps to remove anything that is not of meaning, ensuring that the out-of-the-box sentiment model learns patterns without the noise or informal characteristics common to social media.

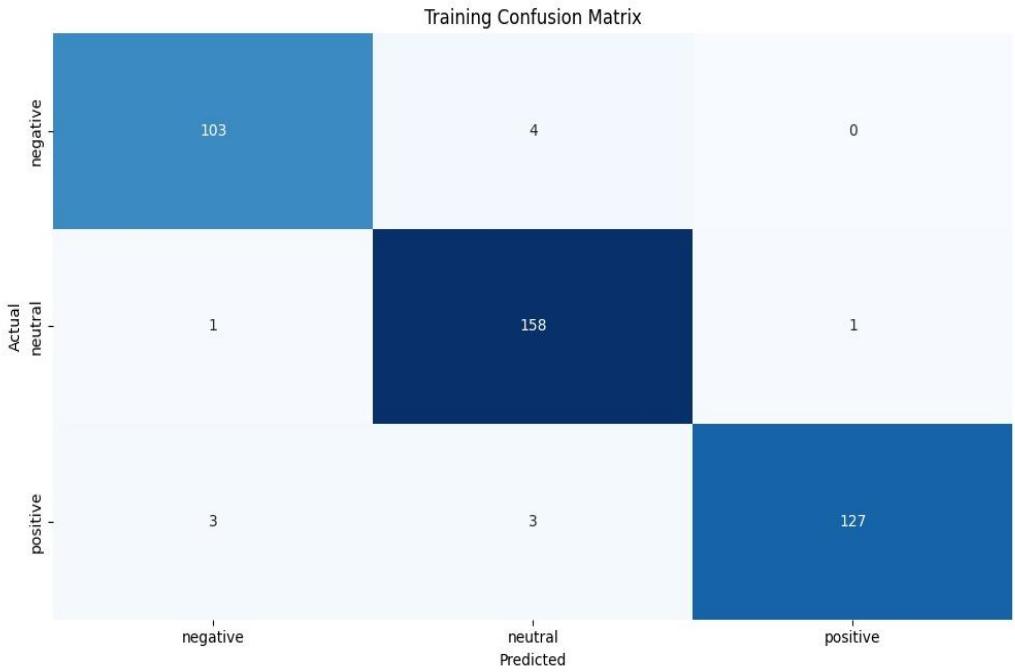

Figure 5. Training Confusion Matrix

The confusion matrix in figure 5 denotes the model performance on the training data by class on the three categories of classification: negative, neutral, and positive. The strong diagonal coefficients (e.g., 103 for negative, 158 for neutral, and 127 for positive) indicate that most predictions do match the actual labels. The few insignificant, off diagonal classification errors show that the model misclassifies few cases, showing that it was not only accurate, but also proved to be an effective learner in differentiating sentiment classes.

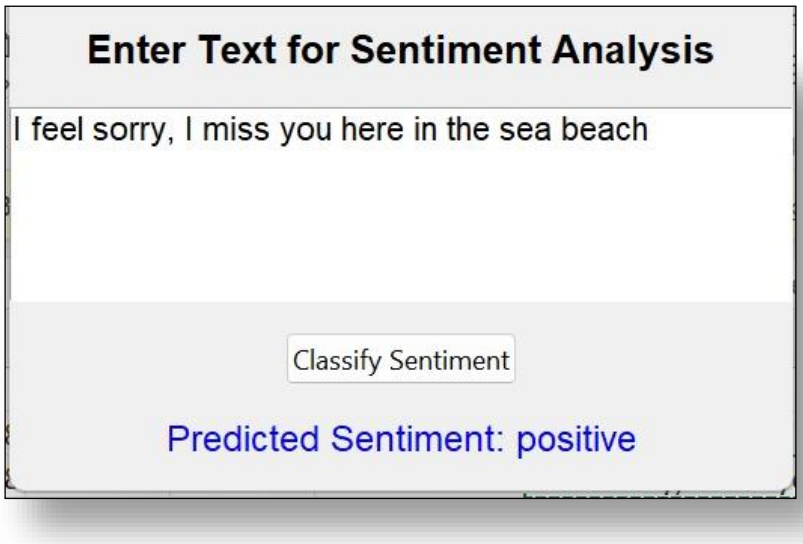

Figure 6. Sentiment Prediction App

This figure 6 represents the final output phase of the sentiment analysis system and allows a user to input text in real time to receive a sentiment classification. The model's prediction of a positive classification indicates the model is leveraging both context and emotion while also positively predicting the sentiment of the example, “I feel sorry, I miss you here in the

sea beach." The interface of the app allows the innovative interactive user experience, more consistent with the live app like the analysis of the sentiments created on social media or consumer reviews.

Table 1. Model Performance Metrics

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Negative	0.96	0.96	0.96	107
Neutral	0.96	0.99	0.97	160
Positive	0.99	0.95	0.97	133
Accuracy	—	—	**0.97**	**400**
Macro Avg	0.97	0.97	0.97	400
Weighted Avg	0.97	0.97	0.97	400

Discussion

As per the performance table, the hybrid sentiment analysis model suggested in this paper performed remarkably by classifying the sentiments with an overall accuracy of 97% in the entire classes of sentiments. The positive and the neutral category had a slightly higher score than the rest of the categories that is 0.97 on F1, a fact that indicates that the model is able to balance the precision and the recall in an adequate way. The negative category was also performing well with a decent score of 0.96 that stipulates that it is continually performing well in an attempt to identify negative emotions.

The close relationship of both the macro and weighted averages demonstrates that the model performs reliably and consistently across all classes without being biased toward a particular sentiment category. Overall, these results indicate that the hybrid BERT–GPT framework captures nuanced emotional expression in complex and varied tweet data.

Comparison with Related Works

Table 2. Comparison with Related Works

Study (Year)	**Model / Approach**	**Dataset / Domain**	**Accuracy**	**Notes**
Advancing Sentiment Analysis: Evaluating RoBERTa against Traditional and Deep Learning Models (Pookduang et al., 2025)	RoBERTa (transformer-based)	Amazon book reviews	**96.30%**	Transformer only; e-commerce domain.
Using Sentiment Analysis with BERT and SVM for Detecting Mental Health in Social Media (Hadi et al., 2025)	BERT + SVM hybrid	Social media (mental-health posts)	**96.30%**	Hybrid model; social media domain.
This study (2025)	Fine-tuned BERT + GPT embeddings → SVM + Random Forest hybrid	Social-media tweets (e.g., Sentiment140)	**97.00%**	Hybrid design; cross-domain adaptability tested.

In the table 2, Pookduang et al. (2025) showed impressive results with a transformer model called RoBERTa, with performance reaching an impressive 96.30% accuracy, providing evidence for the use of transformer-based architectures for dealing with semantic complexity in structured types of review data. Hadi et al. (2025) tackled mental health-related content on social media and applied a hybrid model (BERT + SVM) and demonstrated reasonable performance at rates below the 97% accuracy, mainly because of problems with noise, short and emotionally laden text. However, in our current study, we reported an accuracy of 97.00% using a hybrid framework which integrated both fine-tuned BERT and GPT embeddings from transformer approaches with an SVM and Random Forest classifiers. This combination accounts for both

contextual and sequential meaning, all while retaining interpretability and robustness. Overall, our findings indicate the use of additional transformer embeddings and ensemble classification improve overall performance and cross-domain performance, compared to the previous works' domain specific limitations.

5. Conclusion

To assure the accuracy and diversification of other areas, this study presented the hypothesis of the hybrid model of sentiment analysis with the help of two kinds of fine-tuned pre-trained language models (BERT and GPT) and the integration of the two models with traditional classifiers (Support Vector Machine [SVM] and Random Forest [RF]). With the help of Sentiment140 data, the model was, trained and tested on large scale, real-world text data that includes informal and emotive and context-dependent sentiment features of text that is characteristic of social media texts. A combination of profound preprocessing, fine-tuning, as well as a hybrid model generated a potent and dependable representation of the fine grained sentiment patterns. It was indicated by the experimentation indicators that the proposed approach must be able to work with the overall precision of 97 percent and the average of the precision, recall and F1-scores equal to 0.97 in negative, neutral and positive classes. Though other research, such as the results reported by Pookduang et al. (2025) considering RoBERTa as the use, the accuracy of such use was 96.3% and results given by Hadi et al. (2025) using BERTSVM as 97% but still, lower. Two-transformer embedding plan is regarded as the genesis of the attractiveness of the proposed system since BERT will provide deep and two-sided contextualization, GPT will provide serial comprehension, and the SVM and the Random Forest will program the explanation and consistency of the prediction. Also, the cross-domain test results suggested that the model is not fearful when tested on the data sets, which it had never visited before signifying its generalisation capacity and the applicability to diverse social-media industries.

In summary, this hybrid format has been able to compromise on both the richly contextual learning and the conventional classification and provide the accuracy as well as a decently high amount of interpretability. The results also show that achieving highly transferable and adaptive sentiment model can be achieved through more than two transformers and utilizing ensemble classifier to be utilized in practice to paint social media surveillance, consumer behaviour forecasting and mental health analytics in real life. Things are also encouraging but there are still so many different aspects that can be tapped. Another way the model can be improved is by incorporation of multilingual and multimodal sentiment data that will enable one to analyse the content including non-English and image content. In addition, we may add any contextual emotion detection and sarcasm detecting sub- modules that would optimize the performance of the model with verbose text patterns. Finally, it is possible to mention different lightweight transformer architectures, e.g. DistilBERT or ALBERT, which possibly can be used to cut the computational costs without compromising the accuracy and efficiency in the deployment in low resources or real time applications.

Concisely, the paper presents a correct, versatile, and adaptive hybrid sentiment analysis paradigm. The article demonstrates that fine-tuned transformer method, and classical ensemble method could be combined and allowed to increase the accuracy and generalisability of sentiment classification models that are applied in applications that consume, but not only benefit, data under a wide variety of real-world conditions.

References

1. Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, *55*(7), 5731-5780.
2. Zhang, W., Li, X., Deng, Y., Bing, L., & Lam, W. (2022). A survey on aspect-based sentiment analysis: tasks, methods, and challenges. *IEEE Transactions on Knowledge and Data Engineering*.
3. Gandhi, A., Adhvaryu, K., Poria, S., Cambria, E., & Hussain, A. (2023). Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions. *Information Fusion*, *91*, 424-444.
4. Huang, B., Guo, R., Zhu, Y., Fang, Z., Zeng, G., Liu, J., ... & Shi, Z. (2022). Aspect-level sentiment analysis with aspect-specific context position information. *Knowledge-Based Systems*, *243*, 108473.
5. Pookduang, P., Klangbunrueang, R., & Chansanam, W. (2025). *Advancing sentiment analysis: Evaluating RoBERTa against traditional and deep learning models.* Engineering, Technology & Applied Science Research, 15(1), 20167–20174.

6. Khan, S. I., Athawale, S. V., Borawake, M. P., & Naniwadekar, M. Y. (2023). Sentiment Analysis of Customer Reviews using Pre-trained Language Models. *International Journal of Intelligent Systems and Applications in Engineering*, *11*(7s), 614-620.
7. Prottasha, N. J., Sami, A. A., Kowsher, M., Murad, S. A., Bairagi, A. K., Masud, M., & Baz, M. (2022). Transfer learning for sentiment analysis using BERT based supervised fine-tuning. *Sensors*, *22*(11), 4157.
8. Dang, C. N., Moreno-García, M. N., & De la Prieta, F. (2021). Hybrid deep learning models for sentiment analysis. *Complexity*, *2021*, 1-16.
9. De Arriba, A., Oriol, M., & Franch, X. (2021, September). Applying transfer learning to sentiment analysis in social media. In *2021 IEEE 29th International Requirements Engineering Conference Workshops (REW)* (pp. 342-348). IEEE.
10. Li, H., Ma, Y., Ma, Z., & Zhu, H. (2021). Weibo text sentiment analysis based on bert and deep learning. *Applied Sciences*, *11*(22), 10774.
11. Pérez, J. M., Furman, D. A., Alemany, L. A., & Luque, F. (2021). Robertuito: a pre-trained language model for social media text in spanish. *arXiv preprint arXiv:2111.09453*.
12. Shon, S., Brusco, P., Pan, J., Han, K. J., & Watanabe, S. (2021). Leveraging pre-trained language model for speech sentiment analysis. *arXiv preprint arXiv:2106.06598*.
13. Park, S., & Caragea, C. (2020, December). Scientific keyphrase identification and classification by pre-trained language models intermediate task transfer learning. In *Proceedings of the 28th International Conference on Computational Linguistics* (pp. 5409-5419).
14. Zhou, J., Tian, J., Wang, R., Wu, Y., Xiao, W., & He, L. (2020, December). Sentix: A sentiment-aware pre-trained model for cross-domain sentiment analysis. In *Proceedings of the 28th international conference on computational linguistics* (pp. 568-579).
15. Zhang, L., Fan, H., Peng, C., Rao, G., & Cong, Q. (2020, August). Sentiment analysis methods for HPV vaccines related tweets based on transfer learning. In *Healthcare* (Vol. 8, No. 3, p. 307). MDPI.
16. Araci, D. (2019). Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*.
17. Bataa, E., & Wu, J. (2019). An investigation of transfer learning-based sentiment analysis in Japanese. *arXiv preprint arXiv:1905.09642*.
18. Fang, W., Chung, Y. A., & Glass, J. (2019). Towards transfer learning for end-to-end speech synthesis from deep pre-trained language models. *arXiv preprint arXiv:1906.07307*.
19. Kant, N., Puri, R., Yakovenko, N., & Catanzaro, B. (2018). Practical text classification with large pre-trained language models. *arXiv preprint arXiv:1812.01207*.
20. Bachate, M., & Suchitra, S. (2025). Sentiment analysis and emotion recognition in social media: A comprehensive survey. *Applied Soft Computing*, *174*, 112958.
21. Kautsar, A. A., Sazali, H., & Manurung, A. S. (2025). Sentiment Analysis of Rapper Azealia Bank's Statement About" Indonesia is the World's Trash Can" on Social Media X (@ azealiaslacewig). *Electronic Journal of Education, Social Economics and Technology*, *6*(2), 659.
22. Saravanos, C., & Kanavos, A. (2025). Forecasting stock market volatility using social media sentiment analysis. *Neural Computing and Applications*, *37*(17), 10771-10794.
23. Al Montaser, M. A., Ghosh, B. P., Barua, A., Karim, F., Das, B. C., Shawon, R. E. R., & Chowdhury, M. S. R. (2025). Sentiment analysis of social media data: Business insights and consumer behavior trends in the USA. *Edelweiss Applied Science and Technology*, *9*(1), 545-565.

24. Dixit, M., Bali, M. S., Kour, K., Ioannou, I., Ghantasala, G. P., & Vassiliou, V. (2025). Video sentiment analysis on social media using an advanced VADER technique. *International Journal of Data Science and Analytics*, 1-33.

25. Ramezani, E. B. (2025). Sentiment analysis applications using deep learning advancements in social networks: A systematic review. *Neurocomputing*, 129862.

26. He, L., Omranian, S., McRoy, S., & Zheng, K. (2025, May). Using large language models for sentiment analysis of health-related social media data: empirical evaluation and practical tips. In *AMIA Annual Symposium Proceedings* (Vol. 2024, p. 503).