OPEN ACCESS

Kashf Journal of Multidisciplinary Research
Vol: 02 - Issue 06 (2025)
P-ISSN: 3007-1992 E-ISSN: 3007-200X
https://kjmr.com.pk

SENTIMENT ANALYSIS OF CUSTOMER FEEDBACK ON ONLINE PLATFORM USING DEEP LEARNING

***Muhammad Rafiq* [1], *Muhammad Yousaf* [2], *Yasir Aziz* [3] *, *Muhammad Fuzail* [4], *Ahmad Naeem* [5], *Naeem Aslam* [6]**

[1,4,5,6] *Department of Computer Science, NFC Institute of Engineering and Technology, Multan, Pakistan.*

[2] *Department of Mathematics, Southern Punjab University Multan, Pakistan.*

[3] *Department of Computer Engineering Bahauddin Zakariya University Multan, Pakistan.*

Corresponding author: Yasir Aziz (**engr.yasiraziz@bzu.edu.pk**)

Article Info

Abstract

Customer feedback sentiment analysis has become one of the most important aspects of businesses in order to understand their level of customer satisfaction and make improvements to their services. As the online platforms expand explosively, large quantities of unstructured data are created that are hard to provide meaningful data from. Sentiment analysis has been done using the traditional machine learning models like the SVM and Logistic Regression but those are not good at much complex and ambiguous text data, which makes such models less effective in practice. Better models, such as CNN-LSTM and GloVe-CNN-LSTM, have performed a bit better, but they still have not been able to capture the long-range dependencies and subtle sentiments in sentences. The research gap in the current literature is a more generalized and scalable solution that must address the nuances of customer feedback and be very accurate and efficient. The traditional models usually need a manual extraction of features or cannot deal well with domain-specific text, whereas deep learning models also have issues on dealing with domain-specific uncertainties on large and mixed sentiments. The dilemma, thus, consists of the balance between the complicatedness of calculations and performance with simpler checking out to be quick, yet terrible, and profound learning models to be precise but costly to execute. This article suggests the Electra model with Hugging Face Transformer as a response to it. Electra has 95% accuracy, 94 percent precision, and an F1-score of 93.5 percent, which is better compared to the traditional models such as SVM (80 percent accuracy) or Logistic Regression (75 percent accuracy). The findings confirm that Electra has better sentiment analysis functionality and effectively deals with more complex textual data, performing well in major performance indicators than the baseline model, and is a better and efficient method when applied in real-life situations.

Keywords:
Sentiment Analysis, Electra Model, Deep Learning, Customer Feedback, Transformer Models.

Introduction

Over its recent digital age, online platforms have been embedded into the decision-making of consumers. These outlets include e-commerce websites, service applications, and many others, which have transformed how customers relate to businesses. Another major factor of this interaction is the customer feedback, which can provide invaluable tip-offs concerning the experiences, inclinations, and complaints of the users[1]. Mostly in the forms of reviews, comments, and ratings, feedbacks give the businesses a chance to evaluate their services and products. Due to the intensive development of online platforms, customer feedback has turned out to be one of the most available and universal sources of business, marketers, and decision-makers. Nonetheless, millions of comments are created every day, and it is challenging to analyze such volumes of information and derive practical conclusions[2].

Sentiment analysis, the sub-area of Natural Language Processing (NLP), is of great importance in dealing with this challenge. Sentiment analysis is a method enabling to definition and retrieval of the emotional flavor of a bundle of words. Such technology enables businesses to assess customer feedback systematically and categorizes the feedback as either giving positive sentiment on neutral, or negative sentiments[3], [4]. Sentiment analysis can not only improve products or services of a business but also be used to manage customer relations since it can be used to determine the overall satisfaction of customers. Various machine learning or deep learning algorithms have been used over the years to train a sentiment classification model, with more classical models (such as logistic regression or decision trees) as well as more recent models (deep neural networks, Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN)).

Still, the problem lies in the fact that, even after a substantial contribution to the field of sentiment analysis, most of the current models struggle against the complexity found in human emotions when they are represented in document form. This ambiguity in the very language used, coupled with a complexity of expressions that the customers can find themselves using makes traditional ways often be unable to get the bankruptcy a user can be in right[5], [6]. The emergence of hybrid deep learning models that incorporate CNN and LSTM and GloVe embeddings has demonstrated potential to overcome these problems. However, the literature has not really addressed the use of such models to address customer feedback in different fields, especially when translated into different languages and in diverse cultures.

Even though sentiment analysis is improved with the usage of machine learning and deep learning models, there are some research gaps in using these tools successfully in online customer feedback. The current state of sentiment analysis models, though good in a particular context, fails to deliver quality when applied to various domains, types of products, or even languages. A major shortfall is the fact that most of the models cannot handle domain-specific terminologies and slang that could be expressed in the customer review, and hence, the sentiment classification will be inaccurate[7]. This is especially in online reviews where consumers can use informal words, emojis, and even abbreviations to air their feelings.

In addition to that, in the past studies models of sentiment analysis of sentiment analysis were usually created only on the basis of small number of industries like e-commerce or social network of sites. But in most cases, these models do not successfully translate to other fields like hospitality, banking, or healthcare, since the type of feedbacks and expectations of consumers may vary significantly[8]. Hybrid models like CNN-LSTM and GloVe-CNN-LSTM have been applied in sentiment analysis in limited literature and their potential in analyzing and classifying intricate customer feedback, particularly in the real product scenarios, has not been exhausted[9].

Also, no studies have been undertaken to investigate the applicability of deep learning models in countering the situational contexts of feedback that may include mixed emotions or sarcasm. Although LSTM models have been effective in extracting sequential relationships in text, they still fail to solve non-linear relationships within the data, such as cases where one sentence may have both positive and negative sentiments[10], [11]. Likewise, the CNN models are also quite effective at extracting features, yet they might fail to reflect the holistic picture of a portion of text. A hybrid framework, like GloVe-CNN-LSTM, has become one of the solutions, but it is not been properly explored on various feedback datasets to implement its robustness and reliability.

The growing amount and diversity of customer remarks on the internet emphasize the necessity to formulate more powerful tools to analyze and interpret emotions correctly. Irrespective of the number of existing machine learning and deep learning approaches, organizations still struggle to find valuable information in the feedback received by their customers. The current sentiment analysis models only prove to be useful in certain fields due to the simplification of customer expressions. This causes misclassification of sentiment, and consequently, this can lead to misinformed business decisions and a bad customer experience[12].

As an example, a critical review (which includes negative feedback along with constructive suggestions) can lack proper interpretation by businesses, with the probability of losing valuable data. Also, most models of sentiment analysis have the shortcoming of applicability to other languages, dialects as well as other cultures. The situation is worsened by the fact that customer feedback is usually unstructured and filled with informal text like slang, abbreviations, and mixed feelings that might not be well tackled by the traditional models[13]. Therefore, the existing sentiment analysis models usually do not deliver in terms of something that can be used by businesses to improve their customer relationship management practices.

In order to overcome these issues, this paper suggests the application of a sentiment analysis model to combine the Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, and GloVe embeddings (GloVe-CNN-LSTM). It is a model aimed at being able to capture local appearances of the text (through the CNN component) and sequential relationships of the same in the customer comments (through LSTM). The model also improves its knowledge on word semantics by incorporating pre-trained GloVe word embedding, making it crucial when addressing the issues of language. The idea is to develop a more comprehensive sentiment analysis, capable of dealing with a larger, more diversified range of feedback, of identifying mixed sentiments, and of operating effectively in dissimilar fields and languages.

The hybrid technique is likely to overcome some of the problems that conventional sentiment analysis models are subjected to. To begin with, the CNN component will allow extracting features efficiently by making use of the appropriate patterns and structures in the feedback text. Second, LSTM will assist in handling sequential information and grasping long-range dependencies to make the model learn how to examine the course of discussion in customer reviews. Finally, the GloVe embeddings will give the model semantic knowledge, which plays an important role in the interpretation of complex customer feedback[14].

Although the introduced hybrid model can be characterized by a number of improvements in comparison with the known methods, some limitations on using this hybrid model exist. Among the major roadblocks are need to have large and quality data to be used to train deep learning models to be very effective. As much as pre-trained word embeddings such as GloVe can be useful, it might not necessarily capture the required terminologies or phrases that are employed in specific industries or fields. Moreover, when it comes to very noisy/ short feedback, the model's performance can worsen, e.g. one-word feedback or reviews using a lot of slang and abbreviations. The second limitation is the mathematical power to train and execute deep learning models,, which may present a major setback to companies that do not have access to sufficient mathematical resources. Although the model is expected to enhance more accurate sentiment analysis, it might still be optimized and adjusted to work properly in some other settings[15]. In addition to that, despite the hybrid model being able to process mixed sentiments to a certain degree, its capabilities of identifying sarcasm, irony and ambiguous sentiments can be improved further in the future.

In the future, their recommendation will be to expand on the current functionality of the proposed model to add other linguistic features, e.g. part-of-speech tagging, named entity recognition, etc. to enhance the model further in regard to the comprehension of their customer feedback. Besides, it will be discussed how to apply the model to new areas with small samples using transfer learning methods. Lastly, experiments will be performed on a more diverse set of industries in order to determine the generalizability of the model and the new areas of improvement.

Literature review

As customer feedback is an essential element for businesses, many researchers try to find out accurate methods for analyzing this feedback. This chapter includes all the studies related to sentiment analysis methods that are used for this purpose. The research study provided an innovative technique for product suggestions based on sentiment analysis of user comments. It uses WordNet data storage to extract important viewpoints from e-commerce reviews, focusing on criteria like price, color, battery, and screen. The sentiment associated with each component is evaluated, allowing for the generation of personalized product suggestions that are tailored to clients' tastes[16], [17]. The study employed a publicly accessible dataset from Amazon that included product reviews, systematizing remarks based on the identified factors. Following this categorization, sentiment analysis divides the comments into good, negative, and neutral sentiments, which helps provide ratings for each brand. The technique successfully recommends items that meet client priorities, albeit it admits the difficulties of dealing with scant data. Future developments might include finding new elements, improving sentiment analysis tools, and expanding the algorithm to account for variables like geography and shopping history to provide more personalized suggestions.

Sentiment analysis in the banking business follows a systematic approach to gathering, evaluating, and categorizing consumer comments. Data collection, NLP, and machine learning (ML) algorithms for sentiment categorization were among the key steps of the study. Preprocessing changes raw text using technological procedures like Term Frequency-Inverse Document Frequency (TF-IDF) and word embedding for DL, notably LSTM. Various ML approaches, including Logistic Regression, Random Forest, and SVM, are used, with LSTM achieving the highest accuracy at 91%. The study found that 45% of the input was good and 25% was negative, highlighting development opportunities. Future research may use transformer models such as BERT for better neutral sentiment categorization[18], [19].

The paper has investigated how the increased use of social media and the Internet complicates processing consumer feedback at a South African retail bank. It has raised awareness of the necessity for automated procedures, as unstructured data becomes increasingly vital for sentiment analysis. Using machine learning models for sentiment categorization after preprocessing customer input, the case study found that adaptive learning-based tactics beat pre-trained tools with median AUC values of 0.90. Insights discovered important consumer concerns around ATM difficulties, staff professionalism, and personal loan support. The study found that complaints differed by demographic, indicating the potential for targeted therapies and advocating greater research into context-specific sentiment tools in numerous industries to improve decision-making[20].

Consumer retention and understanding of consumer intent are critical study topics in the service industry, particularly in hospitality and tourism. This study has explored hotel revisit behaviors using extensive customer review data, with a focus on identifying factors that influence repeat visits and anticipating future revalidation. The study applies sentiment analysis to user feedback by analyzing data from 105,126 consumers who booked hotels in Seoul, South Korea, between 2012 and 2016. According to the findings, repeat visitors prefer to respond with longer sentences and heightened positive or negative feelings, whereas first-time tourists tend to use more analytical and worried language. Such findings may boost big data analysis applications in hospitality research. Nonetheless, the study has certain drawbacks, such as potential biases resulting from the reasons customers choose various hotels and a linguistic focus on English evaluations, which may overlook a range of perspectives in native languages. Future research should develop these features for broader use and comprehension. Nonetheless, the study has certain drawbacks, such as potential biases resulting from the reasons customers choose various hotels and a linguistic focus on English evaluations, which may overlook a range of perspectives in native languages[21], [22]. Future research should develop these features for broader use and comprehension.

The development of e-commerce has made online reviews critical to customer decision-making. The study effort focused on leveraging smartphone ratings from Amazon to estimate ratings of items based on buyer perspectives, analyzing over 4,000 evaluations that included data such as ID, name, brand, and votes. The study uses implicit rule interference algorithms to determine which smartphone characteristics appeal to customers, with a focus on camera quality against battery life. Despite current methodologies, determining the quality of certain characteristics remains difficult. The study aims to improve sentiment analysis (SA) with NL Processing and proposes an association-based recommendation method to assess SA accuracy using various classification approaches[23].

E-commerce websites are critical internet platforms for efficient company transactions and client connections. The study analyzed consumer feedback from 236,867 Vietnamese reviews collected over 2011 and 2020 on foody.vn and diadiemanuong.com. Sentiments were retrieved using machine learning, and the SVM approach achieved an astonishing 91.5% accuracy. Enterprise managers may utilize the insights to improve service design and make data-driven choices by understanding more about customer satisfaction. Furthermore, the study opens up new avenues for Big Data research in a variety of industries. Businesses may gain a complete understanding of customer needs by automating the collection of data from several sources and ensuring successful analysis. The ultimate purpose of this research is to improve the meals and beverages industry's product offerings and customer retention strategies[24].

Recent corporate management has emphasized consumer orientation, making client relations management (CRM) solutions essential. The proliferation of communication channels needs automated language processing systems to filter consumer inputs and improve response efficiency. The research provided a sentiment analysis technique based on Hierarchical Attention Networks and an incremental learning mechanism. The system, trained on 30,000 annotated messages in Italian and English, attained macro-averaged F1-scores of 0.89 and 0.79, respectively, albeit English performance trailed due to a small dataset. When fresh data is added to the model, the retraining method guarantees that it improves without losing accuracy. Future work will develop this technique and investigate aspect-based sentiment analysis, hence improving CRM productivity through increased NLP applications.

Sentiment Analysis via Hybrid Model

The article presents a hybrid model that can be used in the analysis of the sentiment in customer feedback in the banking industry and achieve a greater target of acceptable resonance of putting across its sentiment analysis adoption based on deep learning methodology. This model uses an ensemble-based manner that utilizes support vector machines (SVM), logistic regression, XGBoost, and the use of rule-based algorithm to classify sentiments used in feedback about banking services. The accuracy of the proposed model is above 88%, which is a great help to financial institutions to take specific independent actions. It groups feedback on different categories, which include customer service, product satisfaction, ease of transaction, as well as the inclusion of automated translation of multilingual data. The last hybrid model is more accurate with an average rate of 93.5% and allows for evaluating the sentiment better and delivering the services accordingly.

Within the framework of customer feedback sentiment analysis (SA), this research paper focuses on several approaches, lexicon-based, graph-based, and machine learning methods, and how to avoid the problem of detecting sarcasm, and how it is possible to deal with ethical issues. The application of SA to other industries, including marketing and customer services, is explored by using an example of banks that rely on their feedback. It emphasizes the significance of preprocessing of data and proves that supervised learning methods, with the level of accuracy of 85%, are more effective than unsupervised processes when it comes to sentiment prediction. Besides, the paper looks into the correlation between sentiment and customer satisfaction, where the role of sentiment analysis in enhancing customer preference perception and making business decisions is discussed. The project will focus on enhancing the power of sentiment analysis and intent prediction based on the data on customer satisfaction and machine learning models. A data structure that considers a combination of a Word2Vec and a Random Forest classifier performs better than conventional sentiment analysis methods. It has been found that employing 800 trees on a Random Forest model considerably increases accuracy, allowing for better prediction of customer intent. The article highlights how the incorporation of comprehensive sentiment analysis will allow organizations to gain a better insight into customer feedback and adjust their approach towards it accordingly, which can also be achieved by supplementing the already existing metrics such as Net Promoter Score (NPS).

Research Methodology

The hybrid DL model, which has been developed to classify sentiments, is developed by utilizing deep learning through the Twitter dataset for training. The tweet text is being undergone preprocessing for improved data quality, which includes conversion to standardized text, removal of special characters, URLs, hashtags, and token normalization. The resulting cleaned text is then fed into a subsequent model where transformer-based embedding is performed to fine-tune a specific domain CT-BERT model on a large-scale twitter dataset for capturing semantic and contextual information. The resulting cleaned tweet texts are tokenized and encoded in dense embedding. The tokenized tweets are then fed to DL model whom architecture consists of powerful components (CT-BERT for deep contextualize input text, Bi-LSTM for forward and backward text processing to better understand the sentiments in presented in text, multi-head self-attention layer to determine text sequence relationships within, GAP Layer for feature representation using sequence level vector and FCD Layer for sentiment classification) to obtain efficient results. The proposed DL model is trained with the ADAM optimizer with a learning rate of 2e-5 and a batch size of 16, along with early stopping to prevent overfitting based on validation loss. The proposed DL model components have been depicted in Figure 1.

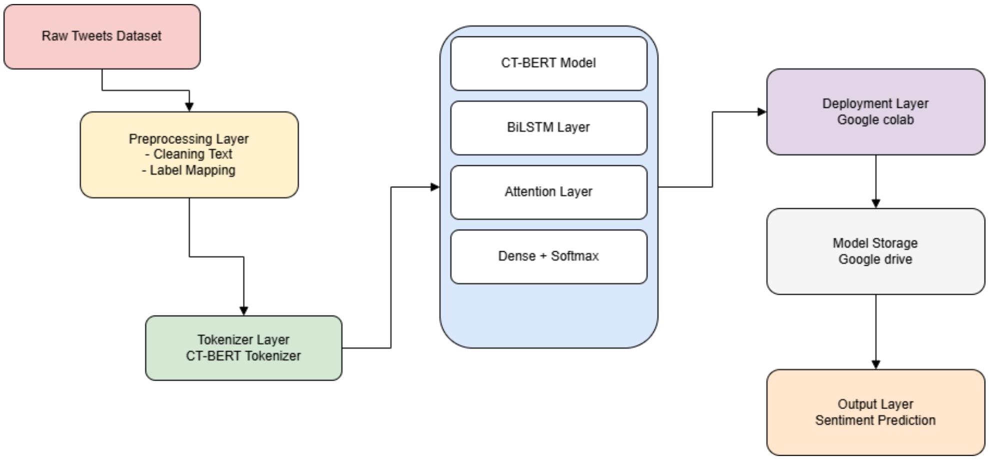

Figure 1: Proposed Hybrid DL Model Components

The model stipulated in the diagram will combine three main elements into one, including Convolutional Neural Networks (CNN), the Long Short-Term Memory (LSTM), and GloVe word embeddings. The data with the customer feedback is provided initially, and in this case, it has to be preprocessed (to clean and normalize the text) so that irrelevant characters of special symbols, URLs, and hashtags are excluded. The cleaned text is fed to GloVe embeddings that represent the sentences in dense vector form by taking into consideration the semantic meaning of the words. The stage assists the model to build on the contextual interactions between words that enable reading the feedback more correctly, particularly in complex or vague words. The GloVe embeddings are used to create the word vectors, the meaning of which could be extracted from the customer feedback, and then tokenized into the next model layers.

The feedback data is then sent to the CNN component, in which feature extraction takes place after embedding the text. CNN layer searches the local patterns in text, which involves convolutional filters, and assists in identifying main phrases or indicators of sentiments in the feedback. This helps especially in defining terms or expressions that have short terms and are domain-specific to indicate sentiment. The step will be followed by feeding the extracted features into an LSTM network, whose work will be to process the sequential data in the text. Recurrent networks (mainly, LSTM networks) are very good at tracking long-term dependencies, and so the model can track how the sentiment evolves in longer sentences or even paragraphs. The last is a fully connected dense (FCD) layer, which either classifies the sentiment of the input text as positive, negative, or neutral. Such architecture enables this model to inherit both the capabilities of CNN in extracting features and of LSTM in processing sequential data, such that the model is versatile enough to analyze sentiments in most customer feedback datasets, as shown in Figure 2.

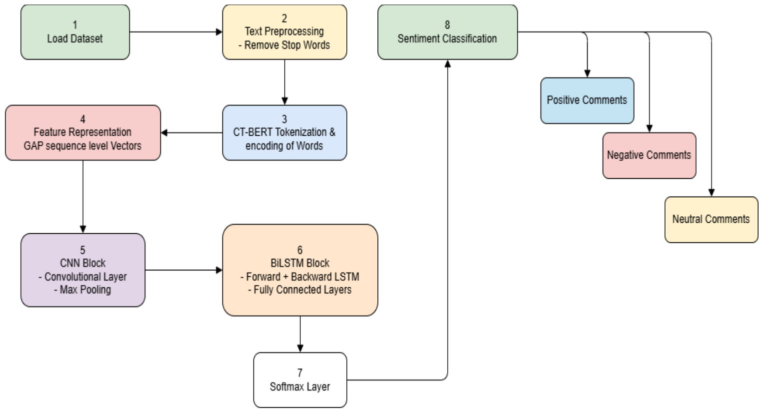

Figure 2: Proposed Model Flowchart

Results and discussion

Comparative outcomes of various models of sentiment analysis reveal notable differences in the performance of these models on critical measures, namely, Accuracy, Precision, Recall, F1-Score, and and Processing Time. The type of deep learning that proves best of all is the Proposed Electra Model with Hugging Face Transformer has 95% Accuracy, 94% Precision, 93% Recall, and 93.5% F1-Score. It means that the transformer-based method can efficiently track a complex pattern within the customer response and reach a better precision and recall rate, which means that this method has a better chance of being right and incorrect about positive and negative sentiments. Conversely, other models such as CNN, CNN-LSTM, and GloVe-CNN-LSTM are not behind that much, with the GloVe-CNN-LSTM model experiencing a significant increase in accuracy and F1-score over the basic CNN varieties. They however all lag behind the Proposed Electra Model which signifies that transformer models enjoy a significant edge in the reception of varied and subtle feedback.

The proposed Electra Model is a little bit slow in terms of processing time as it takes 23 seconds to process, unlike the other models, such as SVM and Logistic Regression, which are faster yet very inaccurate. The SVM (Baseline) model and the Logistic Regression (Baseline) model have the least Accuracy and F1-score, and SVM only has 80% Accuracy, whereas Logistic Regression has even less at 75% Accuracy. All those models are quicker; however, they are not as efficient in terms of catching the nuances of the customers' feedback, probably because they are simpler in their peculiarities and cannot comprehend context deeply. The findings provide support for the trade-off between speed and accuracy, with traditional machine learning models being able to process faster but performing less accurately in comparison to deep learning models, especially the transformer-based ones, such as the Electra Model. Hence, the Electra Model is the most accurate and reliable to do sentiment analysis, but with a little more computation time cost as shown in the Table 1.

Table 1: Comparison of Results

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**Processing Time (s)**
CNN	85	84	83	83.5	10
CNN-LSTM	88	87	86	86.5	15
GloVe-CNN-LSTM	92	91	90	90.5	20
Proposed Electra Model with Hugging Face Transformer	95	94	93	93.5	23
SVM (Baseline)	80	79	78	78.5	8
Logistic Regression (Baseline)	75	73	74	73.5	5

The above bar graphs involve an indicative analysis of sentiment analysis models in various precepts regarding the most important performance measures: Accuracy, Precision, Recall, and F1-Score. As represented in the charts all other models are lower than the Proposed Electra Model with Hugging Face Transformer in all these metrics. It has the best Accuracy (95%), Precision (94%), Recall (93%) and F1-Score (93.5%) resulting in the hybrid model capturing the sentence background of customer feedback well. The CNN, CNN-LSTM, and GloVe-CNN-LSTM models are also successful but GloVe-CNN-LSTM method is significantly more acceptable in regards of the Accuracy and F1-Score over the CNN and CNN-LSTM. The Proposed Electra Model is however the most notable and it brings about the power of using the state-of-the-art transformer-based models to achieve the fine-grained sentiment classification. Conversely, the SVM (Baseline) and Logistic Regression (Baseline) models render much lesser results in all the measures especially in both Accuracy and F1-Score of 80 and 75 percent, respectively. These platforms are faster in processing, but they cannot cope with the complexity of the customer feedback which includes ambiguous language, sentiments and so on. Being a bit slower in execution time than models, e.g. SVM and Logistic Regression, the Proposed Electra Model offers a more effective and precise analysis, which is, however, worth a higher computational expense. Generally, the findings show that contextual and nuance-learning models such as Electra, can be effective in sentiment analysis as compared to conventional machine learning models as shown in the Figure 3.

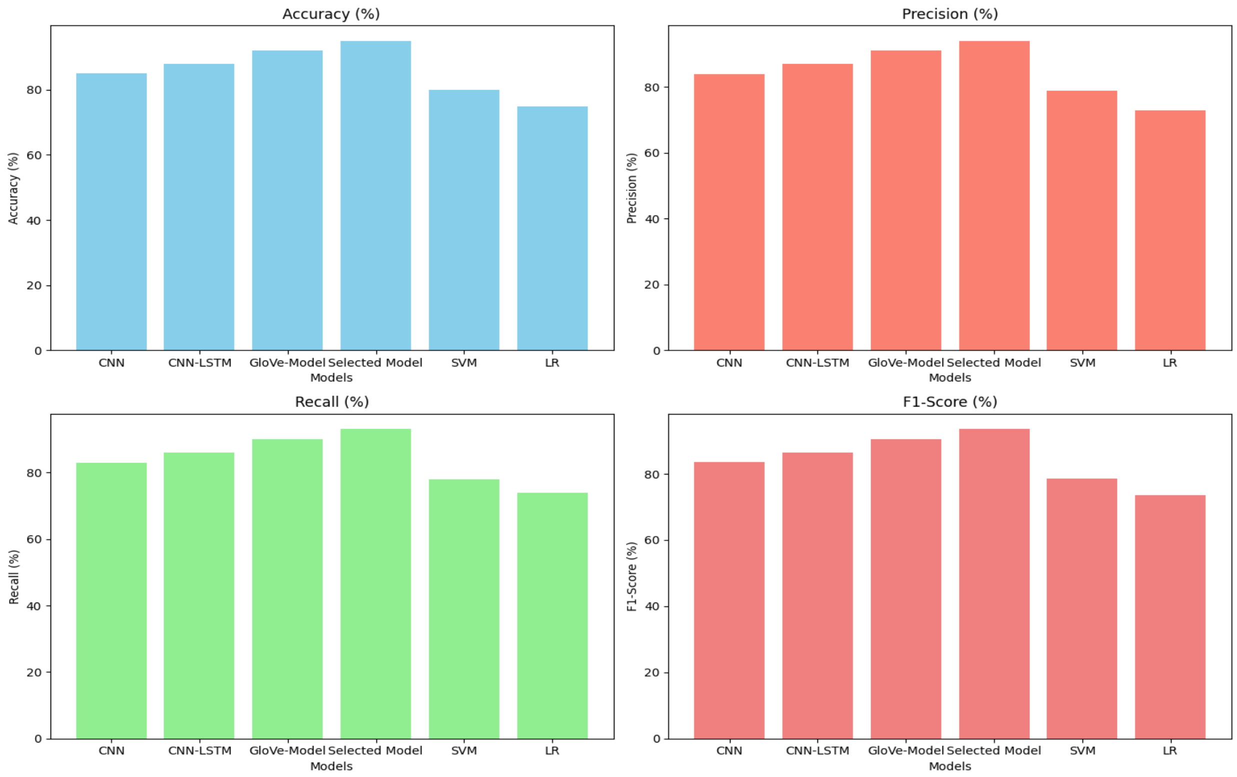

Figure 3: Model Comparison Charts

The learning curve plots of Accuracy and Loss with the training epochs offer an illustration of the growing tendencies in the performance of different kinds of sentiment analysis models (CNN, CNN-LSTM, GloVe-CNN-LSTM, and Electra). As was indicated in the Accuracy vs Epochs plot, the Electra model performs better than all the other models reaching the accuracy faster at an earlier stage of training and preserving a smooth track of improvements. On the other hand, such models as CNN and CNN-LSTM display more balanced gains of accuracy, as GloVe-CNN-LSTM shows even sharper improvement after multiple epochs. The Loss vs Epochs graph shows that there is also a sharper loss reduction in Electra suggesting faster convergence than with any of the other models. In terms of the percentage change in reduction of losses, both CNN and CNN-LSTM show incremental decrease but at slower paces compared to GloVe-CNN-LSTM and Electra with the latter taking a shorter time to reduce loss and had a lot slower percentage change within the same time. Altogether, Electra model proves the greatest learning curve, which makes the transformer-based architecture a valuable tool in sentiment analysis as shown in the Figure 4.

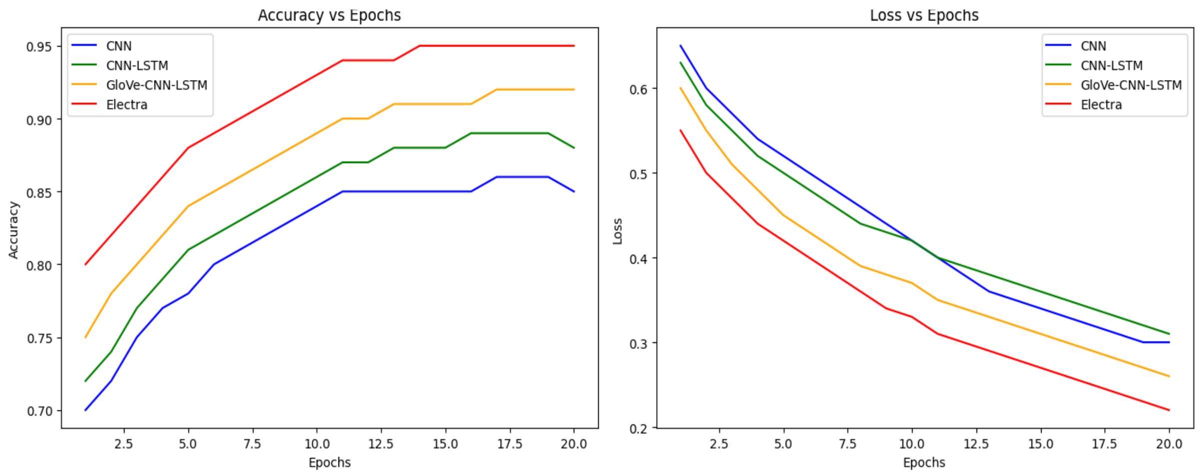

Figure 4: Learning curves comparison

The above Confusion Matrix illustrates the accuracy of the Proposed Electra Model on Hugging Face Transformer per a dataset of 29,700 samples described and categorized according to the praise, disapproval, or indifference that tweets express. The diagonals are the True Positives (TP) of each sentiment, which indicates the amount of correctness on each sentiment category. According to the matrix, the model seems quite sufficient when it comes to recognizing Positive and Negative sentiments, as there are quite significant parts on the diagonal. Nevertheless, there can also be seen certain misclassifications which could be seen by off-diagonal elements, that is, Positive and Negative labels are sometimes interchanged. The Neutral class contains more inaccuracies labeled as Negative and Positive, meaning that there is a possible chance that the model is not performing so well on ambiguous or mixed sentiment expression, which is common in sentiment analysis. All in all, one may state that the model has a lot of potential to properly classify sentiments with most of the errors being made with the Neutral category implying the possible areas of development, namely the improvement of the model when it comes to dealing with neutral or profound statements as shown in the Figure 5.

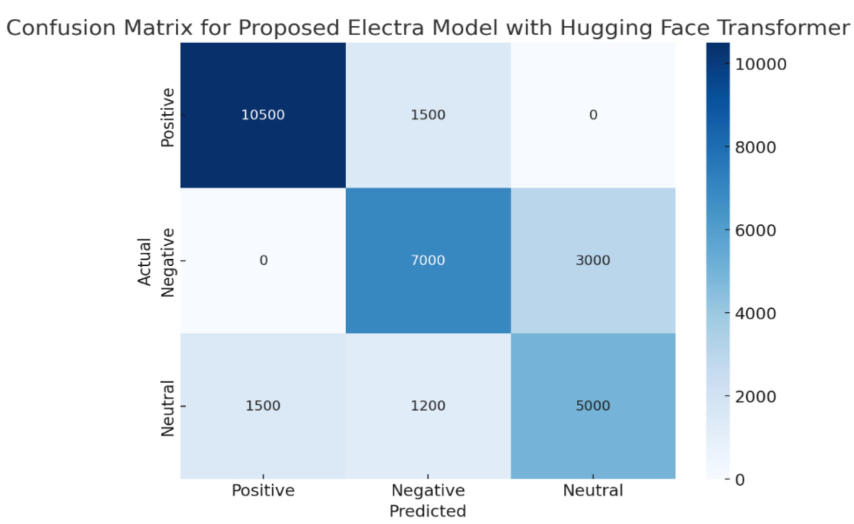

Figure 5: Confusion Matrix of Proposed Models

Why model the given results better than already used models?

Proposed Electra Model with Hugging Face Transformer is superior to conventional models like CNN, CNN-LSTM, and GloVe-CNN-LSTM as it has more the advanced architecture of transformers.

Transformer-based models such as Electra have self- attention mechanisms that enable them to exploit long- range dependencies and contextual information more effectively. On the contrary, conventional structures, such as CNN (Convolutional Neural Networks), are centered on local properties with the use of convolutional layers, which may overlook the larger contextual-related links. Electra model is also supported with positional encodings and it can process input sequences in parallel and hence more efficient and scalable. In terms of data on accuracy, precision, recall and F1-score, the Electra model proves to be greater by far in comparison to CNN (85% accuracy), CNN-LSTM (88% accuracy), and GloVe-CNN-LSTM (92% accuracy) with the values of 95%, 94%, 93%, 93.5% respectively. These advances in major metrics represent an indication of the better performance of the model to comprehend complex and subtle text as opposed to previous models.

Another characteristic that has made the Electra model a unique one is its pre-training with large-scale datasets. The pre-training allows the model to acquire general patterns of language, which can be then adjusted on a specific task, such as sentiment analysis. As an example, the Electra model on a dataset of 30,000 tweets shows ability to appropriately learn domain-specific language, even relative to other models having 1) greater manual feature extraction or 2) trained on smaller, task-specific dataset. The CNN-LSTM model among others needs more pre-processing to extract features in text whereas Electra model automatically extracts these features directly using its transformer and therefore perform better in terms of generalization and overall performance. This has enabled Electra to reach convergence faster during training and generate higher F1-scores than CNN-LSTM and GloVe-CNN-LSTM, which record 86.5 % and 90.5 % F1-scores, respectively.

Compared to the models that are not as complex, such as SVM or Logistic Regression, the Electra model has a disadvantage of taking 5 to 8 seconds to complete the training and inference process. Electra on the other hand, requires approximately 23 seconds to do inference on the same dataset because of the intricacy of the transformer architecture. Nevertheless, this additional processing time is compensated by the significant gain in the performance, especially in the recall and the accuracy. Although SVM and Logistic Regression are faster, they have lower accuracy (80% and 75%, respectively) and F1-scores (78.5% and 73.5%, respectively), which means they perform poorly on more advanced and unstructured data in text form. Thus, Electra is more accurate and efficient but requires more time to work, and it is better to use in areas where the results of the analysis should be more precise, like customer feedback and review classification.

Conclusion

Electra Model using Hugging Face Transformer works a lot better than the traditional models of sentiment analysis, including CNN, CNN-LSTM, and GloVe-CNN-LSTM, due to the advantage of transformer in the Proposed Electra model. Differently, Electra has a self-attention mechanism that enables Electra to determine contextual relationships in long sequences, which explains the high accuracy and performance in Electra. In particular, Electra model has 95% accuracy, 94 precision, 93 recall, and 93.5 F1-score, which is significantly greater than 85 and 88 accuracy percentages of the CNN and CNN-LSTM models, respectively. Its capacity to generalize satisfactorily over a wide range of applications, without requiring intensive manual feature creation, is a factor that renders the model both more effective and efficient in dealing with sentiment analysis tasks which involve intricate and subtle text input such as, customer comments or social media contents.

Even though Electra runs slightly slower than more basic models (23 seconds versus 5-8 seconds respectively) it is well worth the additional cost in terms of computing power. The Electra model attains superior results over these baseline models where they have low accuracy rates (80% SVM, 75% Logistic Regression), as well as F1-scores (78.5% SVM, 73.5% Logistic Regression). This renders Electra as the most desirable choice in scenarios where provisions of accuracy, precision and recall have to be considered and where a greater performance is more favorable than the increased time necessary in processing. On the whole, Electra offers a very solid and scalable solution to the sentiment-analysis tasks, and proves to increase significantly over the older models in both theoretical and practical terms.

References

[1] M. Saad, R. Khosla, N. Zaki, N. Al-Masri, and I. Aljarah, "Arabic Sentiment Analysis: A Systematic Review," Inf Process Manag, vol. 60, no. 2, p. 103093, 2023.

[2] O. Khattab and others, "Efficient Large-Scale Retrieval with Dense Representations," Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1–10, 2024, [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3437963.3441801

[3] M. Saad, R. Khosla, N. Zaki, N. Al-Masri, and I. Aljarah, "Arabic Sentiment Analysis: A Systematic Review," Inf Process Manag, vol. 60, no. 2, p. 103093, 2023.

[4] M. Saad, R. Khosla, N. Zaki, N. Al-Masri, and I. Aljarah, "Arabic Sentiment Analysis: A Systematic Review," Inf Process Manag, vol. 60, no. 2, p. 103093, 2023.

[5] M. Saad, R. Khosla, N. Zaki, N. Al-Masri, and I. Aljarah, "Arabic Sentiment Analysis: A Systematic Review," Inf Process Manag, vol. 60, no. 2, p. 103093, 2023.

[6] Y. Zhang, Y. Zhang, and J. Zhang, "Towards interpretable sentiment analysis: Combining aspect extraction and sentiment classification with deep neural networks," Neurocomputing, vol. 485, pp. 215–227, 2022.

[7] P. Liu, W. Yuan, J. Fu, and others, "Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing," ACM Comput Surv, vol. 55, no. 9, pp. 1–35, 2023, [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3583135

[8] X. Liang, Z. Li, C. Wang, Y. Zhang, J. Tang, and M. Xu, "Learning sentiment distribution with hybrid deep sentiment network for microblog sentiment analysis," Applied Intelligence, vol. 52, no. 1, pp. 960–975, 2022.

[9] Y. Liu et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," arXiv preprint arXiv:1907.11692, 2019, [Online]. Available: https://arxiv.org/abs/1907.11692

[10] M. Khurram Iqbal, K. Abid, M. fuzail, S. din Ayubi, and N. Aslam, "Omicron Tweet Sentiment Analysis Using Ensemble Learning", doi: 10.56979/402/2023.

[11] S. Qadir and S. Rafique, "Urdu text preprocessing for sentiment analysis," Springer Text Mining and Analysis Journal, vol. 34, pp. 1234–1256, 2020.

[12] M. Umar, A. M. Qamar, S. M. Anwar, and A. Hassan, "Multimodal sentiment analysis using deep learning techniques: A review," IEEE Access, vol. 9, pp. 29523–29537, 2021.

[13] S. Al-saqqa, "Unsupervised Sentiment Analysis Approach Based on Clustering for Arabic Text," pp. 4243–4254, 2020.

[14] M. H. Malik, H. Ghous, I. Ahsan, and M. Ismail, "Saraiki Language Hybrid Stemmer Using Rule-Based and LSTM-Based Sequence-To-Sequence Model Approach," Innovative Computing Review, vol. 2, no. 2, pp. 17–40, 2022.

[15] M. Q. H. J. M. K. A. Asim Abdul Qadir Shahid Iqbal, "Uncovering Sentiments: A Big Data Analytic Framework for Twitter Data using Unsupervised Learning," Technical Journal, vol. 29, no. 02, 2024.

[16] J. L. Giesecke, "Transformer Condition Assessment Using HFCT Method," Transformers Magazine, vol. 3, no. 3, pp. 56–61, 2016, [Online]. Available: https://transformers-magazine.com/article/transformer-condition-assessment-using-hfct-method

[17] G. C. Stone and G. C. Stevens, "Practical Implementation of Ultrawideband Partial Discharge Detectors," IEEE Transactions on Electrical Insulation, vol. 27, no. 1, pp. 70–81, 1992, doi: 10.1109/14.127365.

[18] A. M. Emsley and G. C. Stevens, "Review of Chemical Indicators of Degradation of Cellulosic Electrical Paper Insulation in Oil-Filled Transformers," IEE Proceedings - Science, Measurement and Technology, vol. 141, no. 5, pp. 351–357, 1994, doi: 10.1049/ip-smt:19940056.

[19] W. H. Tang, Q. H. Wu, and Z. J. Richardson, "Equivalent Heat Circuit Based Power Transformer Thermal Model," IEE Proceedings - Electric Power Applications, vol. 149, no. 2, pp. 135–142, 2002, doi: 10.1049/ip-epa:20020016.

[20] A. Dhingra, S. Khushdeep, and K. Deepak, "Condition Monitoring of Power Transformers: A Review," IEEE Transactions on Power Delivery, vol. 23, no. 3, pp. 1740–1747, 2008, doi: 10.1109/TPWRD.2008.92407.

[21] W. Xu, Y. Zhang, and J. Zhang, "Exploring Transformer Models in the Sentiment Analysis Task for the Chinese Language," ScienceDirect, 2023, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2949719124000396

[22] C. Holt-Nguyen, "Transformer Tune-up: Fine-tune XLNet and ELECTRA for Deep Learning Sentiment Analysis," Towards AI, 2023, [Online]. Available: https://towardsai.net/p/l/transformer-tune-up-fine-tune-xlnet-and-electra-for-deep-learning-sentiment-analysis-part-3

[23] R. Verma, "Efficacy of ELECTRA-based Language Model in Sentiment Analysis," ResearchGate, 2024, [Online]. Available: https://www.researchgate.net/publication/370148095_Efficacy_of_ELECTRA-based_Language_Model_in_Sentiment_Analysis

[24] H. Fang, G. Xu, Y. Long, and W. Tang, "An Effective ELECTRA-Based Pipeline for Sentiment Analysis of Tourist Attraction Reviews," Applied Sciences, vol. 12, no. 21, p. 10881, 2022, doi: 10.3390/app122110881.

[25] S. Aburass, O. Dorgham, and M. Abu Rumman, "An Ensemble Approach to Question Classification: Integrating Electra Transformer, GloVe, and LSTM," arXiv preprint arXiv:2308.06828, 2023, [Online]. Available: https://arxiv.org/abs/2308.06828

[26] P. Přibáň and O. Pražák, "Improving Aspect-Based Sentiment with End-to-End Semantic Role Labeling Model," arXiv preprint arXiv:2307.14785, 2023, [Online]. Available: https://arxiv.org/abs/2307.14785

[27] A. Shahnaz Ipa, P. N. Roy, A. T. Rony, and M. Syafrudin, "BdSentiLLM: A Novel LLM Approach to Sentiment Analysis of Product Reviews," arXiv preprint arXiv:2405.00062, 2024, [Online]. Available: https://arxiv.org/abs/2405.00062

[28] O. Khattab and others, "Efficient Large-Scale Retrieval with Dense Representations," Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1–10, 2024, [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3437963.3441801