

2304516_Ran_Zhuo_AI_7026_as signment_1_report

by Ran Zhuo

Submission date: 19-Apr-2024 01:07PM (UTC+0100)

Submission ID: 229992471

File name: 50243_Ran_Zhuo_2304516_Ran_Zhuo_AI_7026_assignment_1_report_571450_1260827194.docx
(1.69M)

Word count: 3308

Character count: 18377

the application of meta ensemble deep learning techniques for sentiment analysis on the IMDB

name:Ran Zhuo id:2304516

Abstract

Nowadays, trillions of data have been added to the web as a result of its immensity, including blog entries, reviews, and comments. Analysis of these is crucial for a service provider, as is identifying the trends in which sentiment analysis is most useful. This is different from standard topic-based text categorization in that text is categorized according to the sentiment it expresses. Opinion and topic information for sentiment analysis are derived from user-generated text. We discuss the sentiment analysis of the IMDB movie review dataset in this report. We investigate the efficacy of various deep learning methods. Additionally, we employ an ensemble learning strategy that combines predictions from the chosen base models through stacking. Analyze the constructed model's performance using F1 score, accuracy, precision, recall, and other pertinent measures.

Introduction

Millions of comments are exchanged by users on social networks, financial websites, and e-commerce sites in the world of data. Although this kind of data is useful for suppliers and customers as reference material, processing such a large volume of data by hand is challenging for humans. Thus, in order to display the customer's emotional state, we must develop an automatic scoring system. The content of praise or criticism can be predicted by the system based on its emotion. It is important to take into account both the polarity and the customer's emotional perspective on the material.

 1 Sentiment analysis tools can be used to convert unstructured data on products, brands, films, or subjects that allow people to express their emotions into structured data. It helps companies to have a deeper comprehension of how consumers react to their offerings generally and where they fit into the market. They can also support their company's growth while better meeting the requirements and desires of their customers [4]. It enables the associations to comprehend the public's response to their products and their continued vigilance all the more quickly. Additionally, by using literature to support their problem statement, they can solve the buyer's demands and difficulties more quickly while also fostering a stronger working relationship. To contrast and compare a number of important learning models. We consider important learning models and their hybrids and we then suggest meta ensemble deep learning technique

for sentiment analysis on the IMDB.

problem statement with literature support

2 Sentiment analysis is used in analytics of online user material. In this regard, social media has numerous difficulties. This encompasses both the vast amounts of data that are currently available and their transient character. Collectively, these factors add to the challenge of sentiment analysis. In order to tackle this matter, several methods have been employed. Both the application of machine learning and the building of lexical change are frequently helpful strategies for this objective. Statistical techniques and domain-specific knowledge-driven analysis are further choices. These methods had a favorable impact on the sentiment analysis challenge. We focus on the many machine learning methods that are applied to sentiment analysis. Lijun Liu and associates used word2vec output vectors to pretrained a seven-layer CNN model. This made the distinction between each word easier to understand. They decided on a seven-layered model in order to achieve a compromise between high processing loads and the accuracy of the data acquired from multiple tests.[1] In a similar spirit, Jian Wang and associates created a model that analyzes sentiment in Chinese text using an LSTM. [2] However, Salma Akter Lima and associates created a system that examines IMDB movie reviews. Every one of the three models functions properly. [3] In order to gauge public sentiment toward US equities, Chien Cheng Lee and colleagues analyzed text texts relevant to the stock market using a pre-trained BERT Model. [4] In a similar vein, Manish Munikar and colleagues have conducted sentiment analysis on a SST dataset using a BERT model. [5] Junchao Dong and his colleagues have gone one step further, using a hybrid BERTCNN Model to do sentiment analysis on commodities reviews. The hybrid model was superior to the performance displayed by the individual models examined since it was specifically created to enhance the individuals. [6] Recent NER research studies have shown impressive potential for feature abstraction skills through deep learning. The GRU and LSTM are the models used in NLP the most frequently. Additionally, it has been demonstrated that Bidirectional RNNs (BRNNs) are advantageous in handling a variety of natural language processing (NLP) tasks because they process the input sequence from both directions, producing superior representations and identifying patterns that unidirectional RNNs could have overlooked.[8] A system that applies a classification algorithm to sentiment analysis of Hindi movie reviews was created by Charu Nanda and colleagues. Analogous endeavors had been made to evaluate movies

in Thai, Bengali, and English. [7] Even though sentiment analysis can be done in a variety of ways, little effort has been made in this area to evaluate the many approaches and determine which algorithm works best. Furthermore, as far as we are aware, not much research has been done on the topic of selecting the ideal sentiment analysis algorithm for IMDB movie reviews. The observed ones draw findings from a comparison of a limited number of models. Here's where our efforts are useful. Our research uses a variety of deep learning models to perform sentiment analysis on movie reviews from IMDB. After that, the results will be contrasted and compared to determine which method is the most effective and useful for further research.

Methodology

- 1) CNN: A multi-layer supervised learning neural network is called a convolutional neural network. The essential components for implementing the convolutional neural network's feature extraction function are the pool sampling layer of the hidden layer and the convolutional layer. The network model uses the gradient descent approach to minimize the loss function and inversely modify the layer-by-layer weight parameters. Iterative training is used frequently to increase the accuracy of the network. The upper layer of the convolutional neural network consists of a hidden layer and a logistic regression classifier that corresponds to the conventional multi-layer perceptron. The lower hidden layer of the network is made up of a convolutional layer and a maximum pool sampling layer alternatively. The feature image acquired by feature extraction from the convolutional layer and the subsampling layer serves as the input for the first fully connected layer. The classifier in the final output layer can classify the input image using logistic regression, Softmax regression, or even support vector machines. As seen in Figure 1, the convolutional layer, down sampling layer, and fully-linked layer make up the convolutional neural network topology.

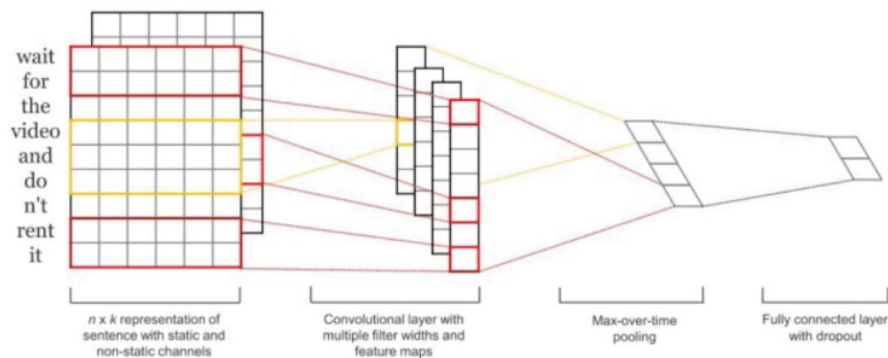


Fig. 1. CNN network

- 2) LSTM: A recurrent neural network is what it is. Its ability to handle successions by utilizing its internal state memory is unparalleled. It can be used to present tasks in a variety of arrangements. It is ready to handle storing a state for a considerable amount of time. addresses the shortcomings of recurrent neural networks, namely the gradient vector's gradual decline or extinction.[9] A forgot gate is an LSTM's middle layer. For effective long-term learning, the forget gate is utilized to identify which information should be retained and which should be deleted.[8] The middle layer extracts the data from the input layer, and the output layer shows the outcome.[10]

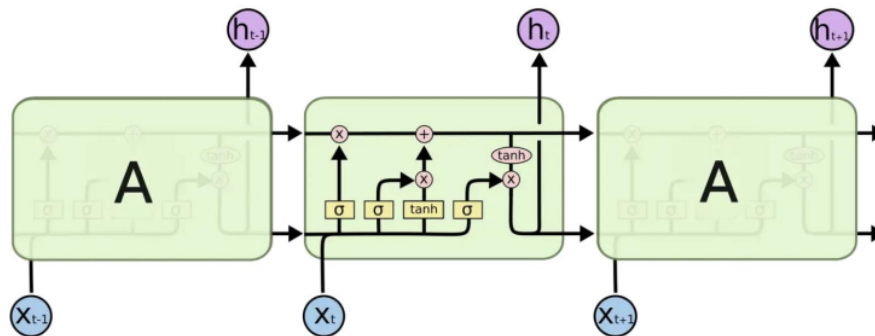


Fig. 2. LSTM network

- 3) BERT: Based on the Transformer framework, Figure 3 depicts the structure.

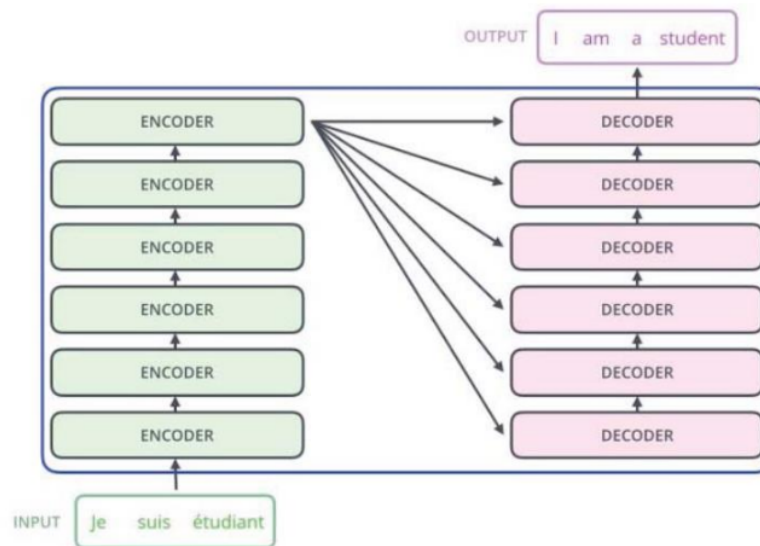


Fig. 3. Transformer framework

Transformer employs the same Encoder-Decoder architecture as the Attention model, except it has a far more intricate structure. Six encoders make up its encoder module, while six decoders make up its decoder module. Each encoder consists of a self-attention layer and a feed-forward neural network, the current node can focus on the current word while still capturing the semantics of the context. Attention layer, feedforward neural network, and a self-attention layer are present in every decoder node. The present node can receive the important stuff that requires attention right now with the assistance of the attention layer. Self-attention is depicted in Figure 4.

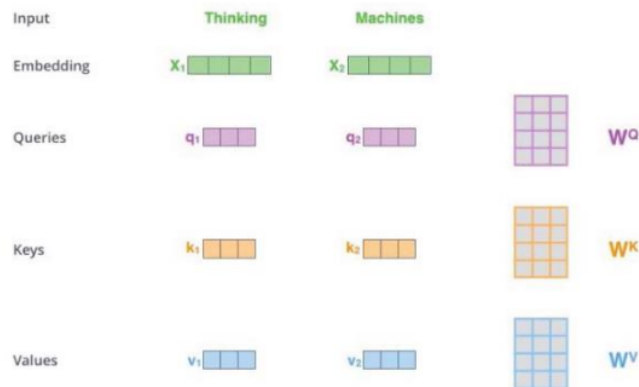


Fig. 4. The Self Attention

- 4) GRU: Another well-liked RNN variant that S. Hochreiter and J. Schmidhuber

proposed is called GRU. In order to resolve this issue, During testing, To identify sentence endings, GRU collects longitudinal data [13, 14]. This is accomplished by utilizing the concepts of input units and cells. Below is a summary of GRU's foundational engineering. 1. The input gate, bypass gate, pass gate, and memory cell are its constituent parts. Value recall is made possible by the information, yield, and bypass gates, which temporarily distribute values throughout the memory cell. These input circumstances demonstrate how the various gates affect the GRU network using weight network, historical data, and actuation (tanh and sigmoid). Figure 5 shows the overall layout of the GRU.

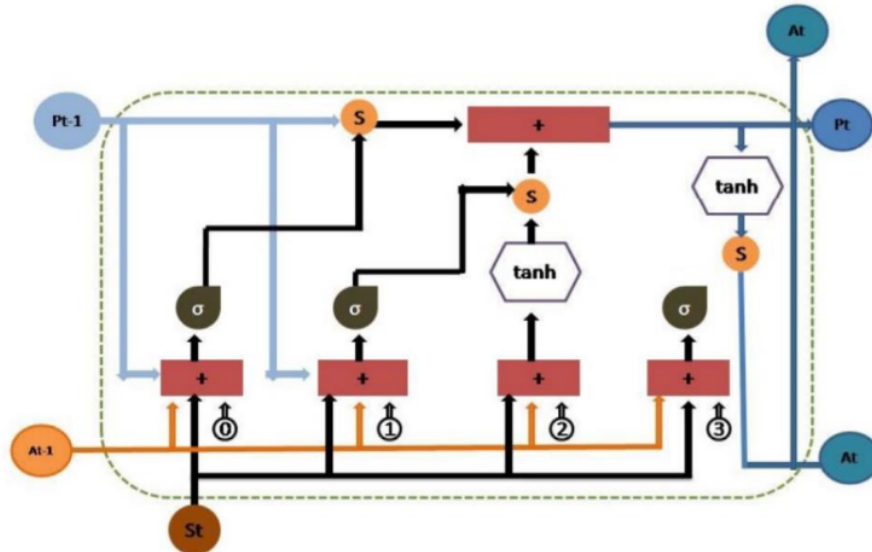


Fig. 5. Working of GRU

5) Stacking: The purpose of the so-called stacking is to use the model to simulate how much power the original data used. The original data is first learned by the basic learner, and then the original data is performed by these basic learners. The models' output is then listed in the following order. The process is used to create new data with (m, p) dimensions; m denotes the sample data and p the number of basic learners. The new sample data is then swapped out for the second layer model.

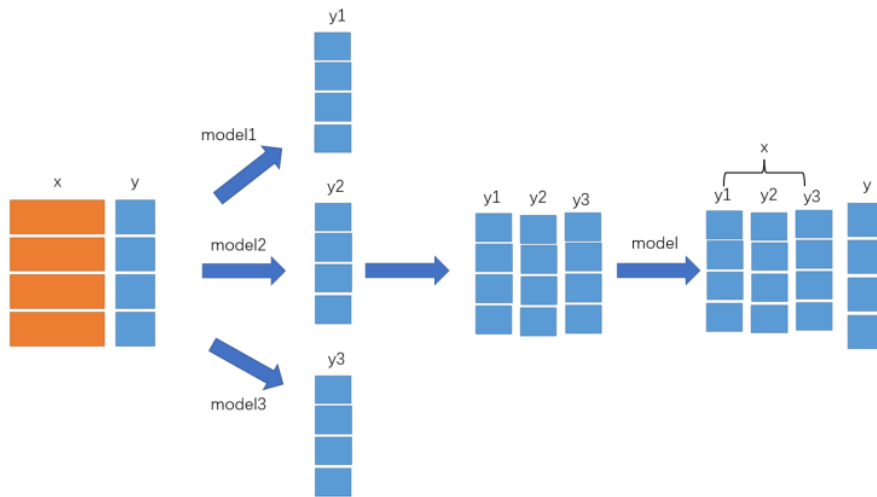


Fig. 6. Working of stacking

Experiments

The movie review dataset is gathered from IMDB, a well-known website that rates films and provides information on television shows, films, video games, and other media. In addition, IMDB offers details on directors, producers, actors, storyline synopses, unrelated information, movie statements, fan reviews, and audits. The majority of the data on IMDb is obtained from clients and content created by clients. Users can rate movies they've viewed on IMDB from 1 to 10 and post reviews about them. The rating that the user assigns indicates the polarity of the review, or whether the user expresses a positive or negative viewpoint [15, 16].

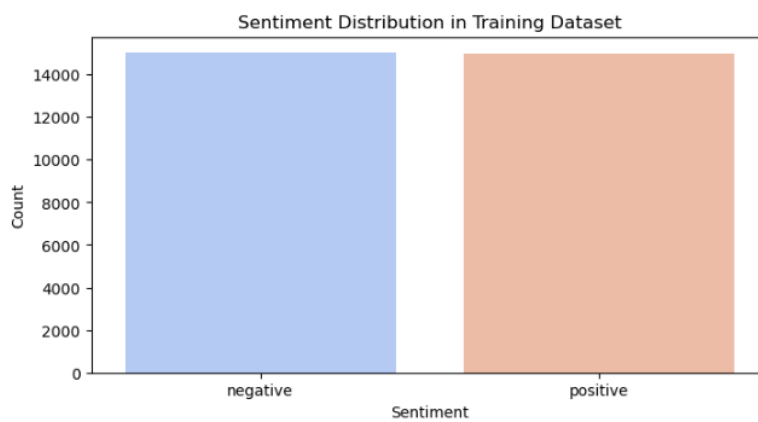


Fig.7. sentiment distribution in training dataset

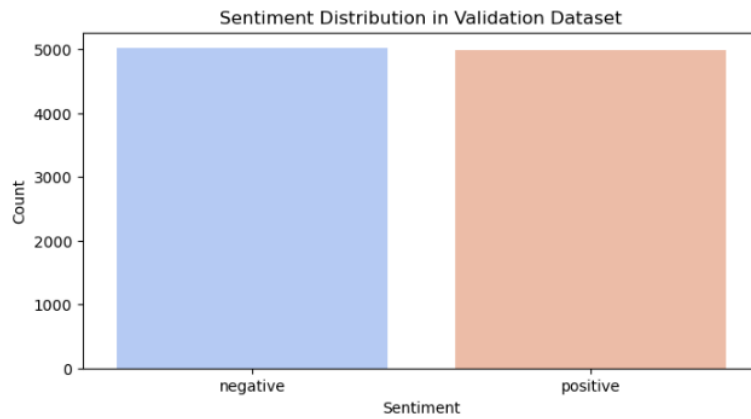


Fig.8. sentiment distribution in validation dataset

Results

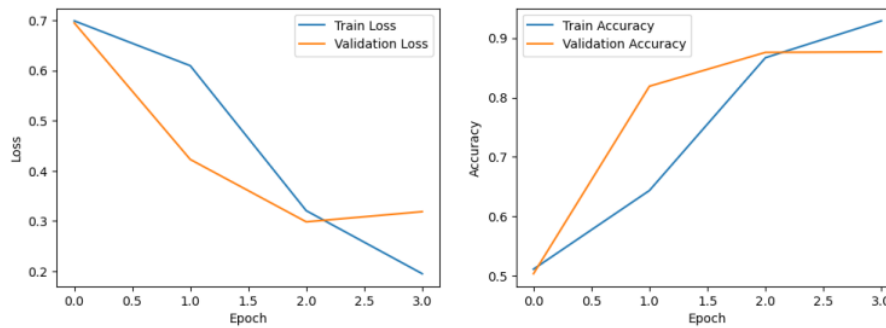


Fig 9. Loss and Accuracy Curve of GRU

The training loss decreases significantly as the epochs progress, indicative of learning. However, there's an uptick in the validation loss after initially decreasing, which suggests overfitting; the model is performing better on the training data but worse on unseen validation data. Both accuracies improve over time, although the validation accuracy seems to plateau, suggesting the model may be reaching its limit in generalizing from the training data to the validation data.

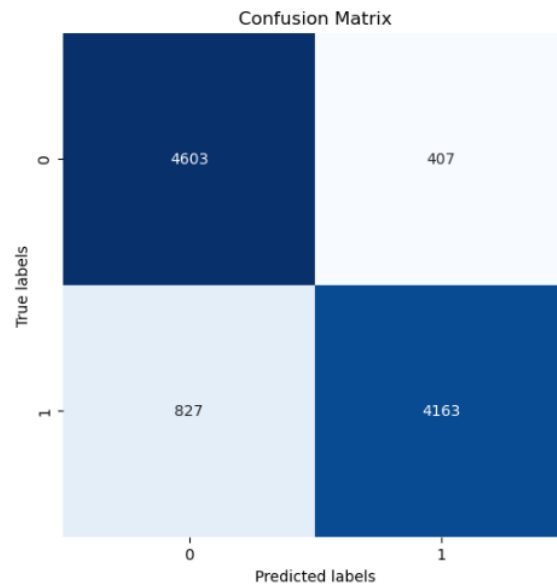


Fig 10. Confusion Matrix of GRU

This indicates the model is more accurate in predicting the negative class than the positive class.

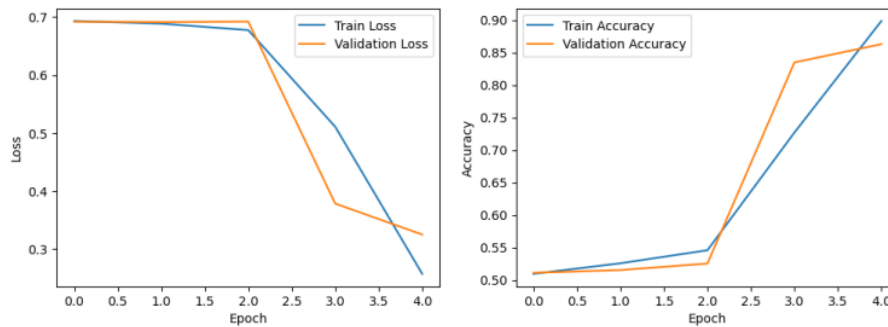


Fig 11. Loss and Accuracy Curve of LSTM

As the number of epochs rises, both losses reduce, suggesting that the model is learning. But the lack of training plateaus much earlier than the validation loss, which continues to decrease even at the last epoch, suggesting that further training could still be beneficial and that the model may not yet be overfitting. There's a notable increase in both accuracies as the training progresses, with a sharp uptick in validation accuracy in the last epoch. This is a good sign that the model is improving in its predictive capabilities on both the training and validation datasets.

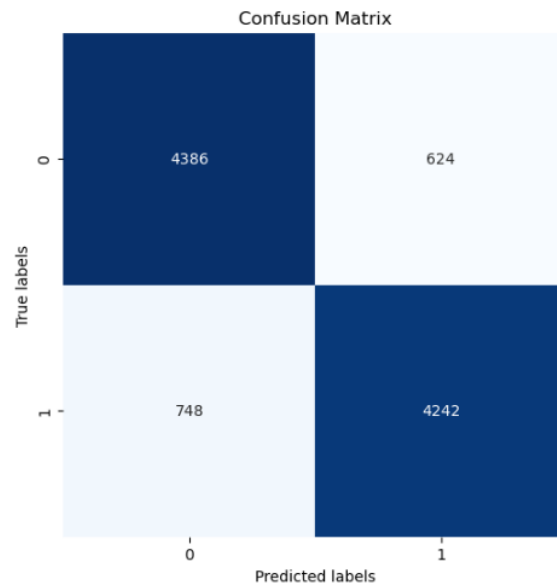


Fig 12. Confusion Matrix of LSTM

It misclassified 624 negative sentiments as positive and 748 positive sentiments as negative. This suggests that the model has a balanced performance in predicting both classes but with a slightly higher misclassification rate for the positive class.

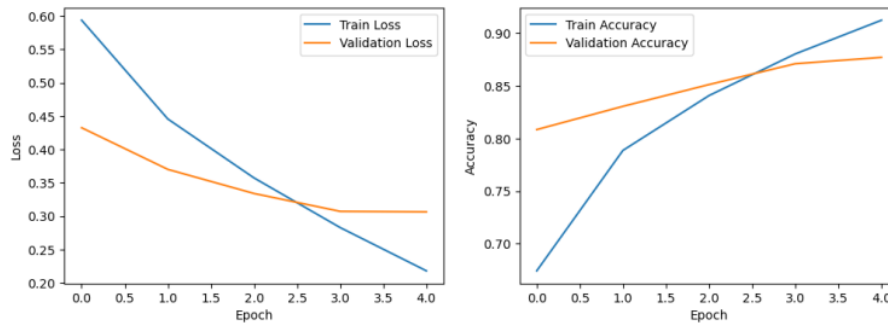


Fig 13. Loss and Accuracy Curve of CNN

Both losses decrease over time, with the training loss decreasing more rapidly than the validation loss. This pattern typically indicates that the model is learning effectively, and since the validation loss continues to decrease without increasing, it suggests that overfitting has not occurred by the final epoch. Both accuracies increase over time, with the training accuracy plateauing towards the end of the epochs. The validation accuracy continues to increase, indicating that the model might still benefit from further training.

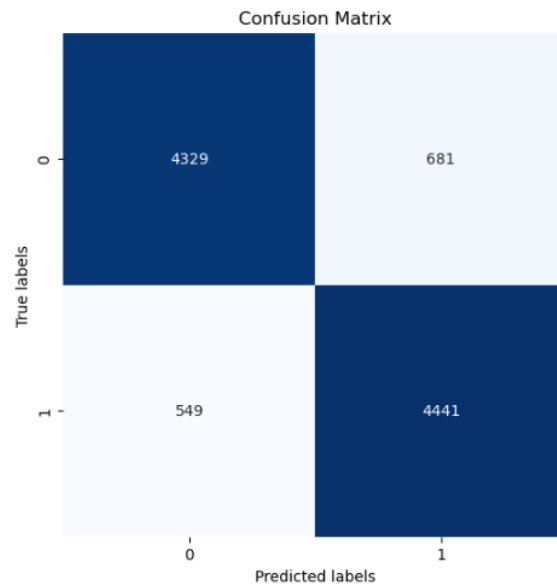


Fig 14. Confusion Matrix of CNN

It incorrectly predicted 681 negative reviews as positive and 549 positive reviews as negative. The numbers suggest that the model is fairly good at distinguishing between positive and negative sentiments but is slightly more prone to misclassifying negative sentiments as positive.

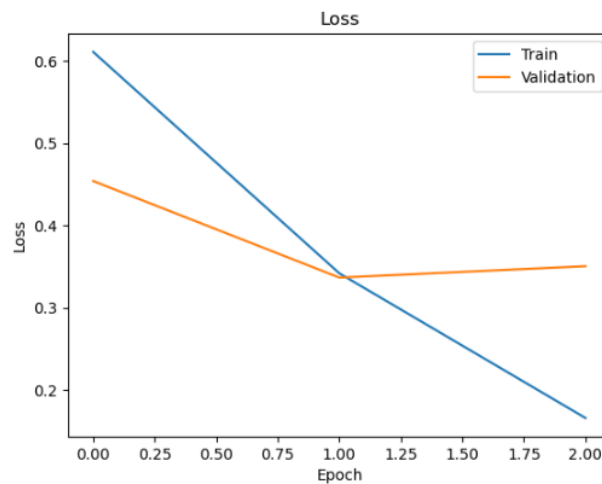


Fig 15. Loss Curve of BERT

The Figure 15. shows the training loss (blue) decreasing sharply, while the validation loss (orange) also decreases but starts to plateau towards the end of the training epochs. This is a common and desirable pattern during the training of

machine learning models, where the loss diminishes as the model learns from the data. The fact that the validation loss levels off suggests the model is beginning to fit the validation data well without further overfitting.

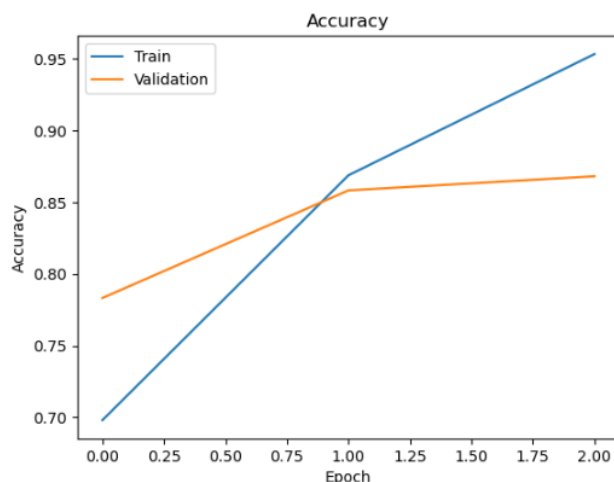


Fig 16. Accuracy Curve of BERT

The Figure 16. illustrates the accuracy of the model on both the training (blue) and validation (orange) data. The training accuracy climbs steadily and surpasses the validation accuracy after the first epoch. The validation accuracy improves initially but begins to flatten, which could indicate the model's generalization performance is reaching its limit with the given architecture and data.

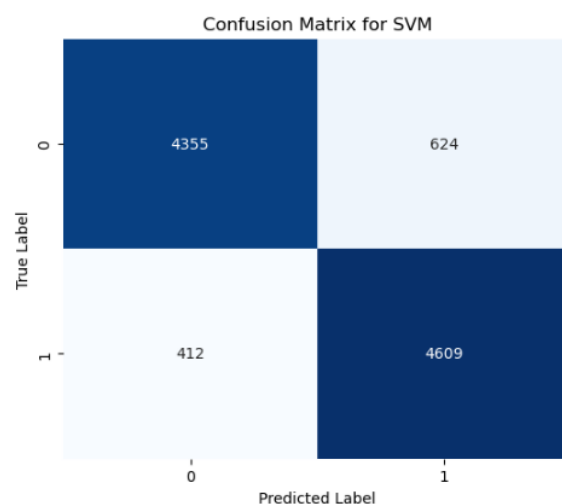


Fig 15. Confusion Matrix of SVM

The SVM model has demonstrated a strong ability to classify both positive and

negative sentiments, with a higher number of true positives and true negatives. The model may be more adept at identifying positive sentiment in this particular dataset than negative sentiment, as evidenced by the less amount of false negatives compared to false positives. Overall, the confusion matrix indicates that the SVM classifier is performing well on this sentiment analysis task. The relatively low number of misclassifications compared to correct predictions suggests a high accuracy.

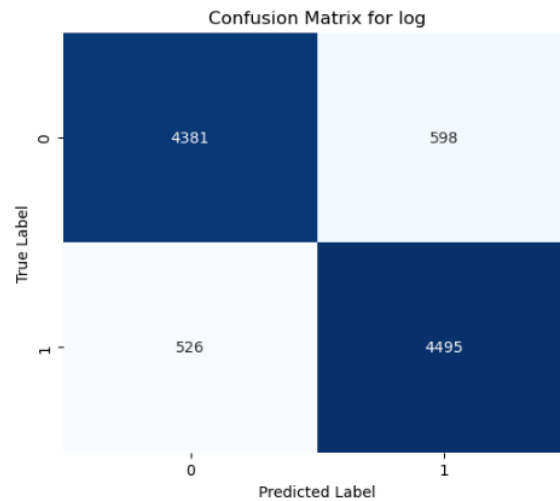


Fig 15. Confusion Matrix of Log

The logistic regression model has performed well in classifying sentiments as it has correctly classified a large number of both positive and negative sentiments.

The number of false negatives is slightly less than the number of false positives, suggesting the model may have a slightly higher precision in identifying negative sentiments than positive ones, but the difference is not substantial.

Model	Accuracy	Recall	F1 Score
SVM	0.8964	0.9179	0.8990
Log	0.8876	0.8952	0.8889

Tab 1. Model Performance on SVM and Log

SVM outperforms logistic regression across all three metrics for this particular task, which could indicate that SVM may be better suited for the nuances of sentiment analysis in the IMDB dataset. The differences in performance are not drastic but are significant enough to potentially influence the choice of model depending on the application's requirements. If the primary concern is to capture as many positive cases as possible (high recall), SVM seems to be the better choice.

discussion

The SVM model generally shows the best overall performance across accuracy, recall, and F1 score, indicating it is quite effective for this sentiment analysis task. This could be due to SVM's effectiveness in handling high-dimensional data and finding the optimal hyperplane for classification. For the models where learning curves (loss and accuracy over epochs) are provided, such as LSTM, GRU, CNN, and BERT, we can see that these models learn over time and improve their performance on both training and validation datasets. This is evidenced by decreasing loss and increasing accuracy. In the case of GRU and LSTM, we observed signs of overfitting where the validation loss started increasing after initially decreasing, while the training loss continued to decrease. For BERT and CNN, the models did not exhibit obvious overfitting within the given epochs. BERT, being a pre-trained model with capabilities to understand context through transformers, shows a good generalization ability as seen by the increasing validation accuracy. However, without further epochs or data, it's hard to conclude its full potential. The confusion matrices for SVM and logistic regression indicate that both models have a relatively balanced performance in terms of precision and recall. However, SVM has a slight edge with a higher recall rate, which could be crucial in applications where catching as many positive instances as possible is important.

Conclusion

The conclusion from evaluating different models on the IMDB sentiment analysis task is that SVM generally provides the best balance of accuracy, precision, and recall, with indications of robustness and good generalization. While deep learning models like LSTM, GRU, CNN, and BERT show potential, particularly in capturing sequential patterns and context, they require careful tuning to avoid overfitting. In addition to these performance measures, the computational cost and the particular requirements of the application should be taken into account when selecting the model for deployment.

references

- [1] X. Ouyang, P. Zhou, C. H. Li and L. Liu, "Sentiment Analysis Using Convolutional Neural Network," 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, 2015, pp. 2359-2364, doi: 10.1109/CIT/IUCC/DASC/PICOM.2015.349.
- [2] J. Wang and Z. Cao, "Chinese text sentiment analysis using LSTM network based on L2 and Nadam," 2017 IEEE 17th International Conference on Communication Technology (ICCT), 2017, pp. 18911895, doi: 10.1109/ICCT.2017.8359958.

- [3] M. R. Haque, S. Akter Lima and S. Z. Mishu, "Performance Analysis of Different Neural Networks for Sentiment Analysis on IMDb Movie Reviews," 2019 3rd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), 2019, pp. 161-164, doi: 10.1109/ICECTE48615.2019.9303573.
- [4] C. -C. Lee, Z. Gao and C. -L. Tsai, "BERT-Based Stock Market Sentiment Analysis," 2020 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan), 2020, pp. 1-2, doi: 10.1109/ICCETaiwan49838.2020.9258102.
- [5] M. Munikar, S. Shakya and A. Shrestha, "Fine-grained Sentiment Classification using BERT," 2019 Artificial Intelligence for Transforming Business and Society (AITB), 2019, pp. 1-5, doi: 10.1109/AITB48515.2019.8947435.
- [6] J. Dong, F. He, Y. Guo and H. Zhang, "A Commodity Review Sentiment Analysis Based on BERT-CNN Model," 2020 5th International Conference on Computer and Communication Systems (ICCCS), 2020, pp. 143-147, doi: 10.1109/ICCCS49078.2020.9118434.
- [7] C. Nanda, M. Dua and G. Nanda, "Sentiment Analysis of Movie Reviews in Hindi Language Using Machine Learning," 2018 International Conference on Communication and Signal Processing (ICCSP), 2018, pp. 1069-1072, doi: 10.1109/ICCSP.2018.8524223.
- [8] J. Li, S. Zhao, J. Yang, Z. Huang, B. Liu, S. Chen, et al., "WCPRNN: A novel RNN-based approach for bio-NER in Chinese EMRs", J. Supercomput., vol. 76, no. 3, pp. 1450-1467, Mar. 2020.
- [9] D. Li and J. Qian, "Text sentiment analysis based on long shortterm memory," 2016 First IEEE International Conference on Computer Communication and the Internet (ICCCI), 2016, pp. 471-475, doi: 10.1109/CCI.2016.7778967.
- [10] G. Xu, Y. Meng, X. Qiu, Z. Yu and X. Wu, "Sentiment Analysis of Comment Texts Based on BiLSTM," in IEEE Access, vol. 7, pp. 51522-51532, 2019, doi: 10.1109/ACCESS.2019.2909919.
- [11] H. Jangid, S. Singhal, R. R. Shah, and R. Zimmermann, "Aspect based financial sentiment analysis using deep learning", in Companion Proceedings of the Web Conference International World Wide Web Conferences Steering Committee, Lyon, France, April 2018.
- [12] V. Malik and A. Kumar, "Sentiment analysis of twitter data using Naive Bayes algorithm", International Journal on Recent and Innovation Trends in Computing and Communication, Vol. 6, No. 4, pp. 120-125, 2018.
- [13] P. Vateekul and T. Koomsubha, "A study of sentiment analysis using deep learning techniques on %ai T witter data", in 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), IEEE, 2016.
- [14] A. C. Pandey, D. S. Rajpoot, and M. Saraswat, "T witter sentiment analysis using hybrid cuckoo search method", Information Processing & Management, Vol. 53, No. 4, pp. 764-779, 2017.
- [15] N. Jnoub, F. Al Machot, and W. Klas, "A domain-independent classification model for sentiment analysis using neural models", Applied Sciences, Vol. 10, No. 18, pp. 6221, 2020.
- [16] J. F. Sanchez-Rada and C. A. Iglesias, "CRANK: a hybrid ´ model for user and

content sentiment classification using social context and community detection”,
Applied Sciences, Vol. 10, No. 5, pp. 1662, 2020.

Appendix

<https://youtu.be/psNGvZkOKiw>

2304516_Ran_Zhuo_AI_7026_assignment_1_report

GRADEMARK REPORT

FINAL GRADE

GENERAL COMMENTS

61 /100

Demonstrated the ability to build and evaluate AI models.
However writing skill and the presentation of the work need to be improved.

PAGE 1



Comment 1

Need to support with literature

PAGE 2



Comment 2

This should not be as a separate heading. This should be part of of the introduction that leads to the research aim/purpose



Comment 3

Wrong. The full stop should be after the in-text ciation.

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

GRADING FORM: ACCA 7026 GRADING

RAN ZHUO

61

ENGAGEMENT WITH LIT. SKILLS

(including current research, referencing, academic conventions and academic honesty)



Engagement with a limited range of literature.
Introduction is not well written, concepts are not
explained well, lack of literature support.
In-text citation syntax is not correct

6.5

KNOWLEDGE&UNDERSTANDING SKILLS

(A systematic, conceptual understanding of knowledge, and a critical awareness of current problems and/or new insights, much of which is at, or informed by, the forefront of the discipline; a comprehensive understanding of techniques applicable to their own research)



The knowledge about the ensemble technique and other AI models are explained in detail and illustrated by the coding provided.

7

COGNITIVE&INTELLECTUAL SKILLS

(Critical evaluation of current research and methodologies and develop critiques of them and, where appropriate, to propose new hypotheses; make sound judgements in the absence of complete data.)



Evaluation of the results is not presented well. Demonstrated results are based on the individual model, no discussion provided based on the ensemble technique

12.5

PRACTICAL APPLICATION SKILLS

(Originality / creativity in the application of knowledge, tools and techniques and in tackling and solving problems in complex and unpredictable professional situations; practical use of established techniques of research and enquiry to create and interpret knowledge in the discipline.)



Provided code demonstrate the application of all models.

29

TRANSF. SKILL & PROF.PRACTICE

(Exercise of self-direction, autonomy and personal responsibility; plan and implement tasks at a professional level; independent learning; use appropriate media to communicate effectively and professionally to a variety of audiences; fluency of expression; systematic approach; clarity and effectiveness in presentation and organisation.)



Need improvement in presentation of the abstract, methodology , results and discussion. Headings provided for some sections are not relevant.

6