


A New Technique for Sentiment Analysis System Based on Deep Learning Using Chi-Square Feature Selection Method

Mohammed Hussein and Fatih Özyurt


Abstract— The sentiment analysis system uses natural language processing techniques and a sentimental vocabulary network. Sentiment analysis means discovering and recognizing people's positive or negative feelings about an issue or product in the texts. Increasing the importance of sentiment analysis has coincided with social media's growth, such as opinion polls, weblogs, Twitter and other social networks. One of the applications of deep learning in NLP is sentiment analysis. The most common and successful type of RNN is the LSTM network. There is a lot of research that uses the LSTM ability to analyze sentiment. But large data volumes reduce the accuracy of LSTM network results in test data; in other words, the problem of over-fitting occurs. This problem occurs when there is a high correlation between independent variables. The model may not have high validity despite the high value of the correlation coefficient between the independent and dependent variables. In other words, although the model looks good, it does not have significant independent variables. Combining the LSTM network with feature selection methods can increase sentiment analysis accuracy to select effective features and solve this problem. In this study, we review state of the art to determine how previous research has addressed these tasks. We also proposed combining the feature selection method, Chi-Square with LSTM, Bi-LSTM and GRU models, the performance of each measured and compared in terms of accuracy, precision, recall, and F1 score for two benchmark datasets, YELP and US Airline. The results show that feature selection methods significantly increases classification accuracy in all cases. In the Yelp dataset, the maximum attained an accuracy of Bi-LSTM is 100% using chi-square when the number of features is 500 In the US Airline dataset, the maximum achieved an accuracy of GRU-LSTM is 97.9% using chi-square when the number of features is 20.

Index Terms— *deep learning, feature selection, chi-square, sentiment analysis, Gated Recurrent Unit Model, LSTM, Bi-Lstm.*

Mohammed Hussein, is with Department of Computer Science University of Raparin, Sulaimani, Iraq, (e-mail: muhamad_it@uor.edu.krd).

 <https://orcid.org/0000-0003-0973-8389>

Fatih Özyurt, is with Department of software Engineering Firat University, Elazig, Turkey, (e-mail: ozyurtfatih@gmail.com).

 <https://orcid.org/0000-0002-8154-6691>

Manuscript received February 28, 2021; accepted September 19, 2021.
DOI: [10.17694/bajece.887339](https://doi.org/10.17694/bajece.887339)

I. INTRODUCTION

INTERNET UTILIZATION has become a fundamental necessity for people to buy products or services online. Social networking sites are now a simple communication tool that allows users to communicate individually or via a public forum. They are the most useful resources to collect information about people's thoughts and feelings on various issues. Thus, if someone needs to buy an item, they need to ask their friends and family for opinions on the products. Presently multiple user surveys are available in public web forums. Gathering and analyzing people's opinions are crucial, mainly when they are extracted and investigated appropriately. Manual extraction and opinion analysis are impossible because the content is disorganized and written in natural language. Sentiment analysis can be used to analyze opinions automatically, that usually modelled as a text classification problem. Text classification is a crucial task for natural language processing that can be performed in many applications that understand the natural language to determine the purpose and meaning and apply it to resolve multiple issues [1], Fig. 1 shows the general sentiment analysis system.

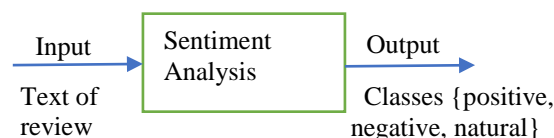


Figure.1. General sentiment analysis system

Sentiment analysis is the classification of sentiments within text data using text analysis techniques. It can automatically change people's public view from unstructured data into structured data about brands, items, services, and politics. This data can be considerably helpful for business applications to revise marketing strategy by understanding customer feelings on products [2]. Furthermore, a sentimental analysis of social media commentary will quickly show if customers are happy with the services. It will help businesses and companies gain input from target customers to identify their strengths and weaknesses and know precisely how to enhance the quality of their products and services [3]. To encourage the development of automated systems using deep learning, we compare the chi-square effectiveness used with deep learning models, i.e. LSTM, BiLSTM and, GRU for two benchmark datasets Yelp and US Airline.

In this paper, the feature set is first passed with

preprocessing to the embedding layer to transform features to word-vectors. Another set of analyses is carried out using Chi-square to show component level analysis and obtain the decreased feature set. The rest of the paper includes four parts organized in the following manner: Part II summarizes related work. Part III shows the Proposed Method of our work, Part V presents the results and discussion, and finally, part VI concludes the paper.

II. RELATED WORK

Recently, the use of neural network-based methods has become increasingly common for sentiment classification. These methods have been common for their ability to learn the distinguishing features of data [8] and determine general context information. With the development of distributed representation, neural networks have significantly improved sentiment classification. It has been shown that embedding good words as input can improve neural network models.

F Rustam, I Ashraf have voting classifier (VC) to help sentiment analysis for such organizations on US Airline dataset, Positive, negative and neutral tweets were classified based on their sentiments. A range of machine learning classifications has also been evaluated used as performance metrics for accuracy, accuracy, recall and F1 score. When they trained a model using, they achieved an accuracy of %78.9 and 791 with TF and TF-IDF feature extraction with US Airline dataset, respectively [4].

Y Cheng, L Yao provides a multi-channel model that combines the CNN and the Bi-LSTM network with an attention mechanism (MC-AttCNN-AttBiGRU). The model will pay the model's capacity. The IMDB and Yelp 2015 data basket's experimental results indicate that the model proposed can extract more rich text features than other simple models and achieve 92.90% [5].

YM Wazery, HS Mohammed implemented two main sentiment analysis methods support vector machine, naive Bayes, decision tree, and K-nearest neighbor. The second method is the deep neural network, an RNN with LSTM. Two approaches were also used by three Twitter datasets: IMDB, Amazon, and Airline. We also highlight a comparison between different algorithms, and the experiment results show the recurrent neural network using LSTM with the highest accuracy of 88%, 87%, and 93% [6].

Rane, A., & Kumar, A, worked on a dataset from US Airlines and performed a multi-class analysis of sentiment. Seven classification strategy analyses were carried out: Gaussian Naïve Bayes, Random Forest, SVM, K-Nearest Neighbors, Decision Tree, Regression Logistics, and AdaBoost. Classifiers were trained on 80% of the data and tested on 20% of the data. The test result is the positive, negative and neutral feeling of the tweet. Based on the results achieved, the accuracy of each classification approach was calculated. The classification techniques used include Adaboost, which combine several other classifications into a strong classification with a precision of %84.5 [7].

A. Kumar, A. Jaiswal used a particle swarm-based feature selection was performed to increase the efficiency of

sentiment analysis on Twitter data. Tweets received from two standard Twitter datasets, SemEval 2016 & SemEval 2017, have been reviewed. The proposed feature selection using PSO is better than the basic group learning algorithm trained by conventional Tf-Idf. By implementing particle swarm optimization, an average of 8.5% improvement is achieved inaccuracy with a 33% reduction in the feature set [8].

Zainuddin, N., Selamat, A., & Ibrahim, R. according to the accuracy of sentiment analysis, a principal component analysis (PCA) based feature selection method is presented that can determine the most appropriate feature set for sentiment analysis. Feature selection helps to reduce redundant features and eliminate irrelevant features that affect classification accuracy. In this study, a feature selection method is presented to classify Twitter sentiment based on the principal component analysis (PCA). PCA is integrated with the Sentiwordnet dictionary-based method to carry out classification with the Support Vector Learning (SVM) framework. Experiments performed on the HCTS dataset and STS benchmark has 94.53% and 97.93%, respectively. Compared with other statistical feature selection methods, the proposed method of this study shows promising results in improving sentiment analysis performance [9].

S. Paudel, P.W.C. Prasad, has implemented a new approach for selecting the appropriate number of features using chi-square, and features are selected using the default score threshold. It has been found that there is a relationship in learning-based techniques among the logarithm of the number of features selected, the output of the sentiment classification, and this relationship is independent of the learning-based process. New results in this study show that researchers can often select the right number of features in learning-based methods to achieve the best performance in sentiment classification. This will help researchers select the right features to improve the efficiency of learning-based algorithms [10].

Sharma S. et al. analyzed Twitter data through real-time data extraction. Preprocessing and feature extraction was applied to the text data, Chi-square test and principal component analysis (PCA), and different machine learning classifications (SVM, Naïve Bayes, Random Forest) Logistic Regression) were conducted against performance indicators. The study concludes that feature selection led to enhancing the accuracy of the backup vector classifier [11].

Rana, S., Singh, A. have introduced a model for classifying movie reviews using the Naïve Bayes classifier and Linear SVM classifiers. They obtained that using the classifiers after omitting synthetic words gives a more accurate result. Their result reveals that SVM achieves better accuracy than the Naïve Bayes classifier. Both algorithms distinctly performed better for genre drama, reaching 87% with SVM and 80% with the Naïve Bayes algorithm [12].

III. PROPOSED METHODOLOGY

Selecting the input that is right, intelligent, and fit to designing a learning model for prediction is of particular importance. The structure of the LSTM is not a predetermined

program, and its weights in the learning process are determined based on the input data. Therefore, if the input data be more prosperous, the network is trained better, and also its performance will be better in predicting new data. This approach seeks to find the relations between the input variables to achieve the desired goal without considering and solving the model's equations. After identifying the goal-related data, preprocessing methods are used to know more detailed information about this set. There are unknown and unusual data in the selected set of inputs. On the other hand, the raw data used to learn the model includes data related to each other due to overlapping inputs. This correlation throughout the learning process can confuse the network in achieving the desired goal and reduce its generalization ability. Data preprocessing avoids using trial-and-error methods and recognizes the most important variables affecting modelling using intelligent techniques [13].

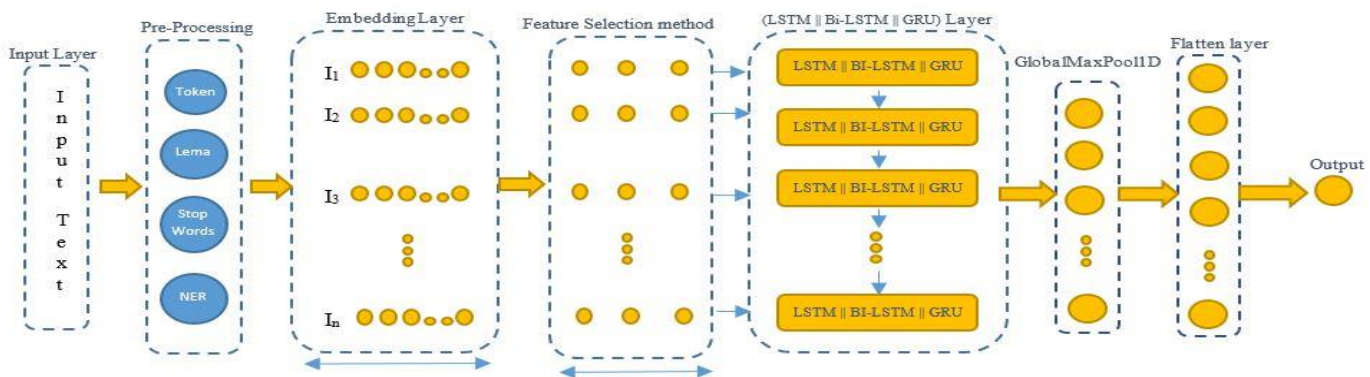


Fig. 2. The proposed method

Fig. 2 shows the structure of the proposed method in sentiment analysis, a combination of chi-square methods with LSTM, Bi-LSTM, and GRU models. In this study the proposed model includes six steps to perform sentiment analysis: Loading dataset from CSV file, pre-processing, word embedding, sentiment detection, feature selection methods and classification using LSTM, Bi-LSTM, and GRU. In this work, all dataset are the benchmark.

A. Dataset

Two benchmark twitter datasets, Yelp, US Airline, were used. Each dataset, including the negative and positive labels, consists of the tweet column and the label column. We also divided the dataset into an 80% training dataset used to train the model and a 20% testing dataset to evaluate the model [5]. In the details of each database.

1) US Airline

The US Airline dataset was extracted from multiple sources for the top 10 US airline carriers, such as Twitter tweets and online reviews of Skytrax. It analyzed a total of 14640 tweets from 7700 users. Twitter data was scrapped as of February 2015 and contributors were asked to identify positive, negative and neutral tweets first, followed by a categorization of negative reasons (such as "late flight" or "rude service") [19].

2) Yelp

Yelp dataset is collected from real Yelp businesses and aimed for academic purposes. It has 366K user profiles in a total of 2.9M edges; and 1.6M reviews and 500K tips for 61K businesses and total check-ins over time for each of the 61K businesses. The data of four countries (U.S., UK, Germany, and Canada) are collected in order to make the dataset diverse. All the data are separated into 5 JSON files, which are business, review, user, check-in, and tip. The content of them is easy to understanding. The business file presents explicitly the properties of each business, such as the internal unique ID of the business, name, geo-location, classes, and so on. The check-in file is a supplementary feature of each business. It offers the collected check-in time for each business in each hour of the week. The user file includes basic information about each user, like yelping time, name, fans, and social correlations with other users. Tip and review files are the

opinion of a specific user to a business.

The difference between tip and review is that the tip present always short, while the review is long enough for users to express their feedback. In this work I used 10000 tweets that have been split into 8324 positive and 1676 negative tweets. You can find YELP Dataset in reference. The dataset is obtained from Yelp Dataset Challenge in 2015. There are five levels of ratings from 1 to 5 [20].

B. Preprocessing of Text Data

Pre-processing is a significant step in converting the text in a human language into a machine-readable form for more processing. It affects the efficiency of other steps. The pre-processing step aims to make the data more machines readable to reduce ambiguity in feature extraction. In this work, some steps were used to normalize the text, converting the upper case to lower case, removing duplicate text, and stopping words, numbers, multiple spaces, special characters, a single character, and punctuation marks URLs, Html, mention, and hashtags. Also performing lemmatization for words, it is the process of substituting words with a stem or base words to decrease inflectional structure to a typical root structure and expand slangs and abbreviations [14].

C. Word Embedding

Word embedding is a leading deep learning technique used to the numeric representations of words that are useful to solve the problem in natural language processing. Neural networks in NLP, do not receive raw words as input since they can only

understand the numbers. Hence, words should transform into feature vectors, or word embedding's [10]. The word vectors can be learned by feeding a large group of raw text into a network and training it for a sufficient time. After training word embedding, it used to extract similarities between words or other relationships [14].

D. Chi-square Feature Selection Method

The large volume of features reduces the accuracy of classification results in the test data; in other words, over fitting occurs. Over fitting happens when the learning model has a high error in the test data and doesn't have generalizability on test data. This problem occurs when there is a high correlation between the independent variables. The model may not have a high validity despite the high value of the correlation coefficient between the independent and dependent variables. In other words, although the model looks good, it does not have meaningful independent variables [15]. The feature selection method has been used to select effective features and solve this problem. Selecting effective features for classification is an NP-Hard problem and can be solved using evolutionary algorithms. Therefore, appropriate feature selection methods were presented to reduce calculation time and increase prediction accuracy. Chi-square aims to improve the scalability of the text classification approaches to eliminate two dimensions used in the combination of deep learning models.

$$chi - 2(t, c) = \sum_{t \in (I)} \sum_{c \in (I)} \frac{(N_{t,C} - E_{t,C})^2}{E_{t,C}} \quad (1)$$

In this formula, we show the observed frequency with N, the expected frequency with E, period t and class C. The high value of chi-square indicates that the hypothesis of independence of two features is not correct. As the correlation between the independent and dependent variables increases, the chi-square index shows a weaker fitting. The chi-square value's sensitivity is to the sample size (this index is usually significant in the high number of samples). If there is a correlation between the independent category feature (predictor) and the dependent category feature (response), the probability of correct prediction of the class occurrence increases. As a result, the feature with the high value of chi-square is selected as the associated feature [15]. In Yelp dataset usually, there are 512 features, but I chose 500 features, and in US Airline dataset usually, there are 23 features, but I chose 20 features. As presented in Table 1.

TABLE I
NUMBER OF FEATURES SELECTED BY EACH FEATURE SELECTION METHODS ON DATASETS.

Datasets	YELP	US Airline
number of features	512	23
chi-square feature selection	500	20

E. Classification Techniques

In this part, we explore some various techniques used to sentiment analysis. They are as follows:

1) Long Short-Term Memory (LSTM)

A Recurrent Neural Network is an artificial deep neural network. This method is used in several NLP studies. They are designed to identify the characteristics of a sequence of data.

LSTM is a specific type of RNN that has more advanced functions and manages information flow. The standard RNN has an issue of gradient vanishing or exploding. To conquer these issues, an LSTM intended by Hochreiter and Schmidhuber [16]. LSTM involved a memory cell, input gate, output gate, and a forget gate. Data can be saved, read, or write from cell-like information in a computer's memory [16]. The cell makes decisions about what to read, write, or erase through opened and closed gates. These gates work on the signals they receive and pass or block data due to its strength or weakness. This division of responsibilities enables the network model to retain information for long periods.

2) Bidirectional Long Short-Term Memory (Bi-LSTM)

Bidirectional LSTMs are an extension of standard LSTMs that can enhance the model's efficiency in sequencing classification issues. Bi-LSTMs train on two LSTMs rather than one LSTM on the input sequence, where all time steps of the input sequence are available. The first one passed on the input sequence without modification, and the other one passed on a reversed copy of the input sequence. It connects them to the same output. Thus at each time step, the networks can have backward and forward information about the sequences. This extra setting adds to the network and enhances the accuracy of the network. Bi-LSTMs are mainly useful when an input context is needed. For example, in sentiment analysis, performance can be improved by knowing the words before and after existing words [17].

3) Gated Recurrent Unit (GRU)

GRU is a gating mechanism in RNN. It is similar to LSTM but has fewer parameters. It's also easy to teach. A GRU unit has two valves: an update gate and a reset gate. The update gateway determines how much of the previous memory should be kept around. The reset gate determines how the new input combines with the previous memory [18].

F. Performance Evaluation Metrics

In building any deep learning model, one of the primary tasks is to evaluate its performance. The performance of each technique used in this work is measured by computing different metrics. The ultimate purpose behind working with other metrics is to understand how well a deep learning model will perform on unseen data. In this work, the following metrics are used:

- Accuracy is the proportion of the accurately analyzed samples to the total number of samples.

$$Accuracy = (T_p + T_N) / (T_p + T_N + F_p + F_N) \quad (2)$$

- Precision is the number of accurate positive analyzed samples to the classifier's predicted positive results.

$$Precision = T_p / (T_p + F_p) \quad (3)$$

- The recall is the number of accurate positive analyzed samples to the number of all relevant samples.

$$Recall = T_p / (T_p + F_N) \quad (4)$$

- F1 score computes precision and recall of the test to calculate the score.

$$F_1 \text{ score} = 2 * (Precision * Recall) / (Precision + Recall) \quad (5)$$

In the above equations, TP is the true positive and predicted correctly, FP is the false positive and predicted incorrectly, TN is the true negative and predicted correctly, FN is the false negative and predicted incorrectly [23-26].

G. Experimental Setup

In this paper, to perform sentiment analysis, we used YELP and US Airline. Anaconda open-source tool is used for python language to perform machine learning task. This study utilized Jupiter notebook and Keras sequential model to implement all models. LSTM, Bi-LSTM, and GRU-LSTM models are utilized to evaluate the performance.

The result of each technique achieved in the following configurations: each of the models configured with a dropout layer to restrict the neural network from memorizing the training set, which is useful to prevent the over fitting. The models compiled with the Adam optimizer with the batch size of 128 for 10 epochs, the network for LSTM and Bi-LSTM models has 32 neurons, and 256 for GRU model, performed using Tensor flow and Keras. Also, we have created a network configuration, as explained in Table 2.

TABLE II
PARAMETERS SETTING IN LSTM, BI-LSTM AND, GRU MODELS.

Parameter	LSTM	Bi-LSTM	GRU
Training approach	CV	CV	CV
Optimizer	Adam	Adam	Adam
Loss function	Binary cross entropy	Binary cross entropy	Binary cross entropy
Learning rate	0.0001	0.0001	0.0001
Batch size	300	300	300
Max. length	6000	7000	7000
Epochs	10	10	10
Hidden layer size	32	32	256
Drop out	0.05	0.25	0.25
Activation function for hidden layer	tanh	tanh	tanh
Activation function for output layer	Sigmoid	Sigmoid	Sigmoid

IV. RESULT

To determine the efficiency of combining LSTM models with correlation-based feature selection method such as chi-square test), training and testing each of the hybrid models has done on YELP and US Airline datasets. To have a proper horizon for evaluating the results, four evaluation criteria have been used: accuracy, precision, recall, and F1 score. When we design a model for analyzing input data, Generalization is the essential feature of this model on the unseen data. When examining the LSTM, Bi-LSTM, and GRU models results without using feature selection methods, we found a large difference between the accuracy of these models in training and testing data. Therefore, the prediction accuracy of LSTM

models before and after combination with feature selection methods were compared with each other. The results of this evaluation are presented on the following YELP and US Airline datasets.

A. Evaluation of the Results in the YELP Dataset

The evaluation of the results of LSTM, Bi-LSTM and GRU models before and after combination with chi-square feature selection methods in YELP dataset are presented in Table 3.

TABLE III
PERFORMANCE OF THE PROPOSED MODEL FOR YELP DATASET.

Techniques	Class	Precision	Recall	F1 score	Accuracy
LSTM	positive	65.3%	68.2%	66.7%	78.7
	negative	52.3%	83.5%	84.3%	
Bi-LSTM	positive	67%	74.6%	76%	85
	negative	87.8%	83.2%	85.4%	
GRU-LSTM	positive	71.8 %	58.7%	64.6%	79.9
	negative	82.6%	89.5%	85.9%	
Chi- square with LSTM	positive	98.4%	99.8%	99.1%	99.4
	negative	99.9%	99.2%	99.6%	
Chi- square with Bi-LSTM	positive	100%	100%	100%	100
	negative	100%	100%	100%	
Chi- square with GRU-LSTM	positive	100%	99.8%	99.9%	99.9
	negative	99.9%	100%	99.9%	

Based on the results obtained in Table 3, the prediction accuracy of the LSTM model on the YELP dataset before combining with feature selection methods was 78.7%, and the accuracy of this model has been increased to 99.4% after combining with feature selection method based on chi-square correlation. The prediction accuracy of combining Bi-LSTM and GRU models with the chi-square feature selection method was 100% and 99.9%, respectively. Examining the results obtained from the hybrid models, we conclude that combining the Bi-LSTM model and the chi-square feature selection method has the best result or 100% in the YELP dataset.

Chen et al. [21] used the CNN multi-channel network as a feature extractor and BI-GRU as a learning model in the YELP dataset, but their proposed method's accuracy also reached 99.2% at the best situation. In contrast, the GRU model's proposed approach with the chi-square feature selection method in the YELP dataset was 99.95%. Fig. 3 shows that we have over fitting due to a large number of features in the YELP dataset. Fig. 4 shows this problem has solved with the feature selection methods, and the accuracy of the model has been increased.

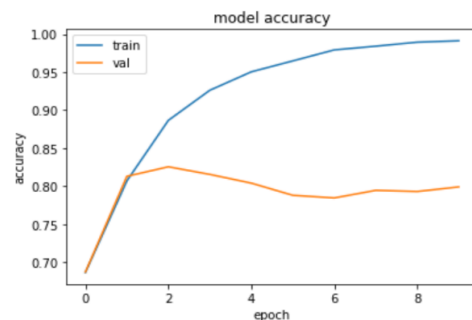


Fig.3.The accuracy of the model before of the feature selection on YELP

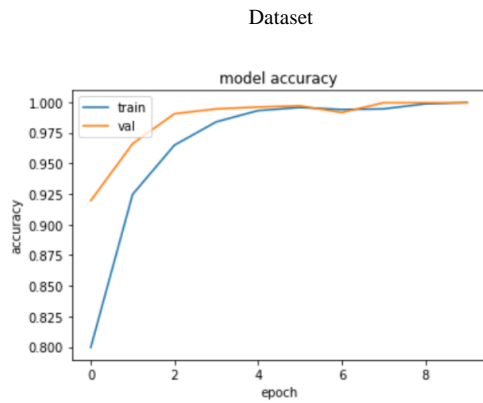


Fig.4.The.accuracy of the model after of the feature selection on YELP Dataset

B. Evaluation of the Results in the US Airline Dataset

The evaluation of the results of LSTM, Bi-LSTM and GRU models before and after combination with chi-square feature selection methods in YELP dataset are presented in Table 4.

TABLE V
PERFORMANCE OF THE PROPOSED MODEL FOR US Airline DATASET

Techniques	Class	Precision	Recall	F1 score	Accuracy
LSTM	positive	83.9%	83.9%	83.9%	73.3%
	negative	55.9%	56.9%	56.4%	
	neutral	70.2%	58.9%	64%	
Bi-LSTM	positive	85.5%	83.6%	84.5%	74.2%
	negative	57.9%	57.6%	57.7%	
	neutral	69%	63.1%	65.9%	
GRU-LSTM	positive	84%	84.6%	84.3%	73.1%
	negative	56.1%	52.9%	54.5%	
	neutral	68.5%	59.8%	63.8%	
Chi- square with LSTM	positive	97.6%	97.3%	97.4%	93.9%
	negative	90.6%	88.4%	89.5%	
	neutral	92.5%	89.2%	90%	
Chi- square with Bi-LSTM	positive	98.5%	97.9%	98.2%	96.1%
	negative	93.8%	93.1%	93.4%	
	neutral	95%	94.2%	94.6%	
Chi- square with GRU-LSTM	positive	99.6%	98.7%	99.2%	97.9%
	negative	96.9%	97.2%	97.1%	
	neutral	96.4%	97.3%	96.8%	

Based on the results obtained in Table 4, the prediction accuracy of the LSTM model on the US Airline dataset before combining with feature selection methods was 73.36%, and the accuracy of this model has increased to 93.95% after combining with feature selection based on chi-square correlation method. The prediction accuracy of combining Bi-LSTM and GRU models with the chi-square feature selection method is 96.1% and 97.9%, respectively. Examining the results obtained from the combined models, we conclude that the combination of the GRU model and the chi-square feature selection method has the best possible result of 97.9% in the US Airline dataset. Rustam, F., Ashraf, I., [22] proposes a voting classifier (VC) to help sentiment analysis for such organizations in the US Airline dataset, but their proposed method's accuracy also reached 79.1% at the best situation

Fig. 5 shows that we have over fitting due to a large number of features in the US Airline dataset. In Fig. 6 this problem has solved with feature selection methods, and the accuracy of the model has been increased.

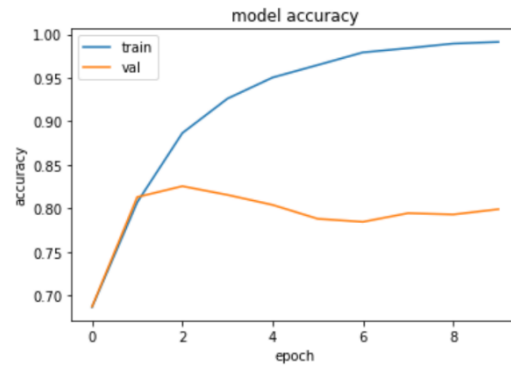


Fig.5.The accuracy of the model before of the feature selection on US Airline Dataset

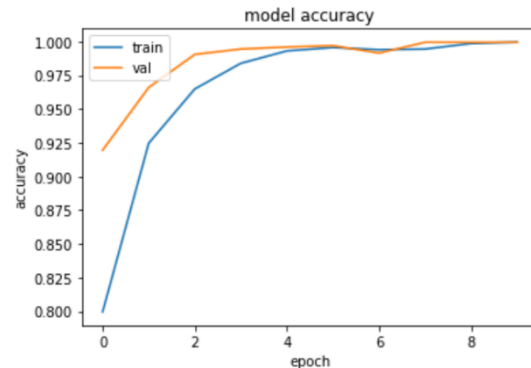


Fig.6.The accuracy of the model after of the feature selection on US Airline Dataset

C. The overall evaluation of the results

The overall results of the evaluation of LSTM, Bi-LSTM and GRU models before and after combination with chi-square feature selection in the YELP and US Airline datasets are presented in Fig. 7.

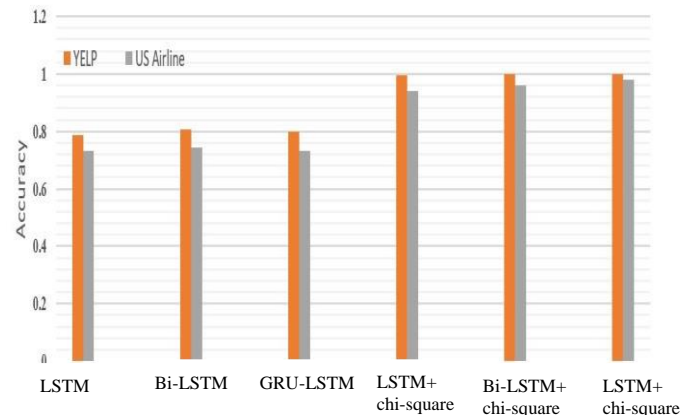


Fig.7.The accuracy of the models on YELP and US Airline datasets

V. CONCLUSION

In this study, chi-square is introduced as one method for feature selection in Sentiment analysis. Feature selection is an

essential step in machine learning. The ultimate goal of feature selection is to select a subset of the main feature set to enhance learning algorithms' performance. The proposed model includes LSTM, Bi-LSTM and GRU with chi-square, in which chi-square extract the quality features passed on to all models. First, with the neural network preprocessing, we give the unreduced feature set. The techniques are then used to reduce dimensionality, and the results are compared. For US Airline and Yelp datasets, chi-square increases classifier performance by removing the feature vector's irrelevant, noisy, and redundant features. By scoring up to %97.9 and % 100 accuracies for US Airline and Yelp datasets, respectively, which is significantly better than the previous studies, this process produces promising results. It is appropriate to say that selecting features can reduce the number of features while preserving classifiers' high performance.

REFERENCES

- [1] S. Xu, H. Liang and T. Baldwin, "Unimelb at semeval-2016 tasks 4a and 4b: An ensemble of neural networks and a word2vec based model for sentiment classification," in Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016), 2016.
- [2] Hameed, Z., & Garcia-Zapirain, B. (2020). Sentiment classification using a single-layered BiLSTM model. *Ieee Access*, 8, 73992-74001.
- [3] F. Chollet, *Deep Learning with Python*, Manning Publications Co., 2018.
- [4] Rustam, F., Ashraf, I., Mehmood, A., Ullah, S., & Choi, G. S. (2019). Tweets classification on the base of sentiments for US airline companies. *Entropy*, 21(11), 1078.
- [5] Cheng, Y., Yao, L., Xiang, G., Zhang, G., Tang, T., & Zhong, L. (2020). Text sentiment orientation analysis based on multi-channel CNN and bidirectional GRU with attention mechanism. *IEEE Access*, 8, 134964-134975.
- [6] Wazery, Y. M., Mohammed, H. S., & Houssein, E. H. (2018, December). Twitter sentiment analysis using deep neural network. In *2018 14th International Computer Engineering Conference (ICENCO)* (pp. 177-182). IEEE.
- [7] Rane, A., & Kumar, A. (2018, July). Sentiment classification system of Twitter data for US airline service analysis. In *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)* (Vol. 1, pp. 769-773). IEEE.
- [8] A. Kumar, A. Jaiswal, Particle Swarm Optimized Ensemble Learning for Enhanced Predictive Sentiment Accuracy of Tweets. In: Singh P., Panigrahi B., Suryadevara N., Sharma S., Singh A. (eds) Proceedings of ICETIT 2019. Lecture Notes in Electrical Engineering, vol 605. (2020), Springer, Cham. https://doi.org/11007/978-3-030-30577-2_56.
- [9] Zainuddin, N., Selamat, A., & Ibrahim, R. (2016, August). Twitter feature selection and classification using support vector machine for aspect-based sentiment analysis. In *International conference on industrial, engineering and other applications of applied intelligent systems* (pp. 269-279). Springer, Cham.
- [10] Paudel, S., Prasad, P. W. C., Alsadoon, A., Islam, M. R., & Elchouemi, A. (2018, July). Feature selection approach for Twitter sentiment analysis and text classification based on Chi-Square and Naïve Bayes. In *International Conference on Applications and Techniques in Cyber Security and Intelligence* (pp. 281-298). Springer, Cham.
- [11] Sharma, S., & Jain, A. (2020). An Empirical Evaluation of Correlation Based Feature Selection for Tweet Sentiment Classification. In *Advances in Cybernetics, Cognition, and Machine Learning for Communication Technologies* (pp. 199-208). Springer, Singapore.
- [12] Rana, S., & Singh, A. (2016, October). Comparative analysis of sentiment orientation using SVM and Naive Bayes techniques. In *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)* (pp. 106-111). IEEE.
- [13] Hameed, Z., & Garcia-Zapirain, B. (2020). Sentiment classification using a single-layered BiLSTM model. *Ieee Access*, 8, 73992-74001.
- [14] Abdalla, G., & Özyurt, F. (2020). Sentiment Analysis of Fast Food Companies With Deep Learning Models. *The Computer Journal*.
- [15] Abdulkhaliq, S. S., & Darwesh, A. M. (2020). Sentiment Analysis Using Hybrid Feature Selection Techniques. *UHD Journal of Science and Technology*, 4(1), 29-40.
- [16] Nguyen, H. T., & Le Nguyen, M. (2019). An ensemble method with sentiment features and clustering support. *Neurocomputing*, 370, 155-165.
- [17] Dong, Y., Fu, Y., Wang, L., Chen, Y., Dong, Y., & Li, J. (2020). A sentiment analysis method of capsule network based on BiLSTM. *IEEE Access*, 8, 37014-37020.
- [18] Zhang, L., Wang, S., & Liu, B. (2018). *Deep learning for sentiment analysis: A survey*. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1253.
- [19] Eight, Figure. "Twitter US Airline Sentiment." Kaggle, 16 Oct. 2019, www.kaggle.com/crowdflower/twitter-airline-sentiment.
- [20] Mohit. "Yelp-dataset." Kaggle, 22 March. 2018, <https://www.kaggle.com/mohit473/yelp-data-set>.
- [21] Cheng, Y., Yao, L., Xiang, G., Zhang, G., Tang, T., & Zhong, L. (2020). Text sentiment orientation analysis based on multi-channel CNN and bidirectional GRU with attention mechanism. *IEEE Access*, 8, 134964-134975.
- [22] Rustam, F., Ashraf, I., Mehmood, A., Ullah, S., & Choi, G. S. (2019). Tweets classification on the base of sentiments for US airline companies. *Entropy*, 21(11), 1078.
- [23] Aydemir, E., Tuncer, T. Doğan, Ş. Unsal, M. (2021). A novel biometric recognition method based on multi kernelled bijection octal pattern using gait sound , *Applied Acoustics* , 173 (107701), 1-9.
- [24] Toslak, F, Sari, M, Aydemir, E, Altun, Y. (2020). Recommendation of a New Device for Calculation of Non-Planning Areas . *Journal of Soft Computing and Artificial Intelligence*, 1 (1), 42-49.
- [25] Tuncer, T, Aydemir, E. (2020). An Automated Local Binary Pattern Ship Identification Method by Using Sound , *Acta Infologica*, 4 (1), 57-63.
- [26] Abdalla, M.H. and Karabatak, M., 2020, June. To Review and Compare Evolutionary Algorithms in Optimization of Distributed Database Query, In *2020 8th International Symposium on Digital Forensics and Security (ISDFS)*, 1-5.

BIOGRAPHIES



than 3 articles in international journals and conferences. His research interests include Artificial Intelligent.



Mohammed Hussein received the bachelor's degree in computer science from University of Sulaimani, Iraq, in 2012, the master's degree in software engineering from Firat University, Elazig, in 2021. He is currently an Assistant Lecture with the Department of computer science, University of Raparin. He has more than 3 articles in international journals and conferences. His research interests include Artificial Intelligent.

Fatih Özyurt received the bachelor's degree in computer engineering from Eastern Mediterranean University, Cyprus, in 2011, the master's degree in computer engineering from Fatih University, Istanbul, in 2014, and the Ph.D. degree in software engineering from Firat University, Elazig, in 2019. He is currently an Associate Professor with the Department of Software Engineering, Firat University. He has more than 40 articles in national, international journals and conferences. His research interests include pattern recognition, artificial neural networks, image processing and deep learning.