Using Transfer Learning Approach to Implement Convolutional Neural Network model to Recommend Airline Tickets by Using Online Reviews

Maryam Heidari
George Mason University
mheidari@gmu.edu

Setareh Rafatirad
George Mason University
srafatir@gmu.edu

***Abstract*—Social Media provides an opportunity for people to share their idea about different aspects of life. Traveling is one of the essential aspects of life. In this paper, we use Bidirectional Encoder Representations from Transformers(BERT) for sentiment classification of online reviews to implement a Convolutional Neural Network model that can recommend airline tickets. Different sentiment classification models are compared based on people's online reviews from six online social platforms. The new model can classify airline tickets as an economical choice or not economical choice ticket based on various aspects of the trips, including airline customer satisfaction, travel destination, hotel information, restaurants, and tourist attractions. This work contribution is two-fold: first, it examines the importance of choosing a transfer learning model for sentiment classification of online social platforms in the recommender system's prediction accuracy. Second, the implementation of the Convolutional Neural Network model, which can classify airline tickets based on data generated by multiple online social platforms. Using the natural language processing approach to use the transfer learning model to improve the CNN model's prediction accuracy is a new approach to use online social platforms to recommend economical airline tickets to consumers.**

I. INTRODUCTION

Social Media are increasingly used by people to review brands and share their opinion about specific services. Consumers' opinions are an integral part of all business owners' economic success to achieve success in competitive markets. Customer feedback helps companies to choose the best marketing strategies to increase their profit. Social Media and online reviews data can determine the popularity of the company's services among people worldwide. In the Airline industry, customer's feedback, which social media provides, can be considered essential information to enhance the quality of services and identify the weaknesses to achieve success in a competitive market. Consumer satisfaction level about specific Airline companies can be provided by a sentiment analysis of Twitter data, one of the most powerful social media platforms. Sentiment analysis of Twitter data about brand-specific US airlines can help Airline companies to attract more customers. More importantly, the Airline industry is considered as a billion-dollar industry with millions of customers. To survive in these competitive markets building consumer loyalty is an integral part of success for airline companies. In the past, people filled the customer satisfaction surveys manually. Then Airline companies try to analyze thousands of surveys manually, which demanded much effort. This process was very time consuming, and many work forces were forced to infer consumer's ideas from texts. The problem of analyzing the consumer's questioners was not limited to the time. All surveys also ask a similar type of question that addresses airline company concerns rather than consumers' views about their expectations about the quality of services. These important reasons create motivation for this research to value customer's feedback by sentiment analysis of online reviews on Twitter and other online platforms to increase the quality of services. In this research, we use five different online platforms in addition to Twitter, to create a reliable convolutional neural network model for airline ticket recommendation.

II. RELATED WORKS

Consumer concerns are important in providing an efficient recommender systems algorithms [1] [2]. In this work, online reviews' Sentiment classification is an important component in implementing the CNN model. Turney [3] introduced a method called "bag of words. In his approach, the average semantic orientation of the phrase is used to classify the reviews. In this method, if the phrase contains positive adjectives or adverbs, it is classified as good, and if the phrase associates with bad adjectives, it is classified as a bad phrase. The problem with his approach is the individual words were not considered in sentiment analysis.Also, Online multimedia and digital marketing can have specific effects on consumer ideas, which can be detected by powerful machine learning models [4] [5]. Another study that used a symbolic approach was called Wordnet. It is the lexical database used by Kamps *et al.* [6] to calculate the semantic orientation of adjectives. The main focus of their research was the similarity and distance between words and concepts. Some previous works which were adopted machine learning techniques for sentiment classification. Sentiment analysis of Twitter data by using machine learning methods is conducted by Sanket Sahu *et al.* [7]. In this work, the researchers improved the preprocessing phase by using the spell checking algorithm. Also, they introduced a scoring system that ranks the tweets by a degree of positivity or negativity. They defined the polarity of tweets by different classifiers such as the Maximum Entropy and Support vector machine. They concluded that the preprocessing methods play a major role in enhancing the sentiment classification accuracy for short length phrases like people's ideas about specific

978-1-7281-5919-5/20/$31.00 © 2020 IEEE

subjects on Twitter.

Paroubek and Pak [8] and Rajabi [9] introduced an approach that can extract sentiment oriented tweets from Twitter API. The classifier using bigram features provides the highest classification accuracy from their search result because creating a balance between coverage and balance. Since the extracted tweets had specific emoticons, their database was becoming biased, and it affected their results. Another weakness of their studies was the lack of neutral sentiment analysis. Neutral tweets are considered very important for Sentiment analysis, as same as both negative and positive tweets. Adeborna and Keng [10] and Jafari [11] applied Correlated Topic Model(CTM) with Variational Expectation-Maximization algorithm. In their research, the lexicon was developed based on Airline Quality Rating(QAR). The tweets were collected and categorized using the CTM with the VEM algorithm. The author only used unigrams as sentiment classification features in the Naive Bayesian classifier. This issue can cause problems because phrases and terms can change the unigrams' sentiment orientation in sentences. Their study's problem was the authors considered only unigrams for sentiment classification features in the Naive Bayesian classifier. As a result, phrases and negation lexicons were overlooked by their approach. However, in this research, we provide a comparative analysis of different sentiment classification methods to choose the best model for online reviews sentiment classification. In this work, We use Bidirectional Encoder Representations from Transformers(BERT) [12], which provide language understanding of each online review rather than just sentence representation of the review. Using BERT in this work plays a significant role in the prediction accuracy of the CNN model.

Neural network models implement successful applications in various aspects of life, such as advanced image processing in Health [13], [14], health products [15] [16],robust hardware security methods [17] [18]and advanced applications in cyber security [19] [20] [21]. Online information plays a significant role in creating different trends in society [22] and can affect consumer decisions. Social bots can create fake trends and affect consumer ideas [23] [24].Also,computing devices are critical part of deep learning applications which can be improved by different models [25] [26] [27]. So in this work, we examine the performance of the convolution neural network model in our new proposed model for airline system based on online user's reviews.

III. Data preprocessing and Data collection

All data collection phase is based on answering these tree critical questions: What is consumer idea about Airline quality of services with regards to both Airline company and Travel destination? It is crucial for airline companies to find their problem to improve the quality of services to meet their customer expectations. To achieve this goal, they need to know precisely which travel destination creates negative reviews about their companies.

Does the geographical location affect people's ideas about the quality of services of a specific Airline Brand? People may think differently about specific brand quality based on their geographical location. Various Airline brands reviews, in different places, can be affected by other factors. For example, a popular vacation destination may receive better reviews because all passengers are in a good mood since they are going on vacation. On the other hand, some cities are business and work destinations, and when people go to these cities or live in these types of cities, they are more on the life pressure. In this case, The airline company needs to improve their quality of services in more creative ways to address customer needs. For example, they can offer free and exciting services in the airport rather than just in the airplane to make their customer more relaxed and happy in big cities' hectic environments, which are more business destinations rather than popular vacation spots.

Data is collected based on four major US Airlines: American Air, Southwest Air, Delta, and United Airlines. Data extracted from six different online platforms such as Google Flight(googleflights.com), Kayak(www.kayak.com), SkyScanner(www.skyscanner.com), Twitter, Hotels.come(hotlels.com) and Trip advisor(www.tripadvisor.com). From Google flights, Kayak flights and Skyscanner API, all flights information about a ticket based on destination and departure location is collected. This information includes destination, departure time, arrival time, Ticket price, Flight class, departure date, return date, and Airline company. So the flights are saved in the data set based on their prices from most expensive to the list expensive. The non-stop flights with the lowest ticket price, the departure and arrival time(before midnight and after 4 AM) are labeled as good deal flights. This phase can be customized based on the user's needs and the user's definition of an acceptable ticket. This information is extracted from these websites based on the selected ticket, time, departure, and destination location. All customers' feedback about some major airlines is collected using twitter API. Twitter provides passenger reviews about four major US Airlines, which are: American Air, Southwest Air, Delta, United Airlines. The tweets are collected via twitter API for four months. Analyzing tweets' sentiment is more challenging than written text because the user's reviews do not follow a specific language model and have more slangs. In this phase, Since the tweet data about customer reviews are noisy, we apply different prep processing techniques to clean data. Since twitter data have inconsistency and missing values, it is crucial to adopt suitable data preprocessing steps before starting the sentiment analysis of online tweets.

All information is collected from Twitter. At first, different hashtags such as #airline, #aviation, #airfare, #plane, #flight are used, the result was large, but there were few relevant data. For example, from every 1000 tweets, just one tweet was related to user's reviews. After that, we used #United,#Delta,#Southwest, and other airlines' hashtags, but the total collected tweet reviews for each airline company was very small. Each Airline company has a twitter account, which starts with @ symbol, followed by its name. As a result, we decide to use @ the company's name to retrieve the user's review tweets. This filter for using APIs enabled us to collect an acceptable number of tweets about user reviews posted by customers about all airline companies. We collected 3.5 million tweets related to the customer's review. We extracted the tweet Id, Tweet's text, Date and time of the tweet, User screen name, user Id, location of the tweet, and hashtags for each tweet. Tweet Id is a unique number that is assigned to each tweet. User Id is a unique number for each user. The location of the tweets shows the user location in the time of the tweets. Each tweet's text is

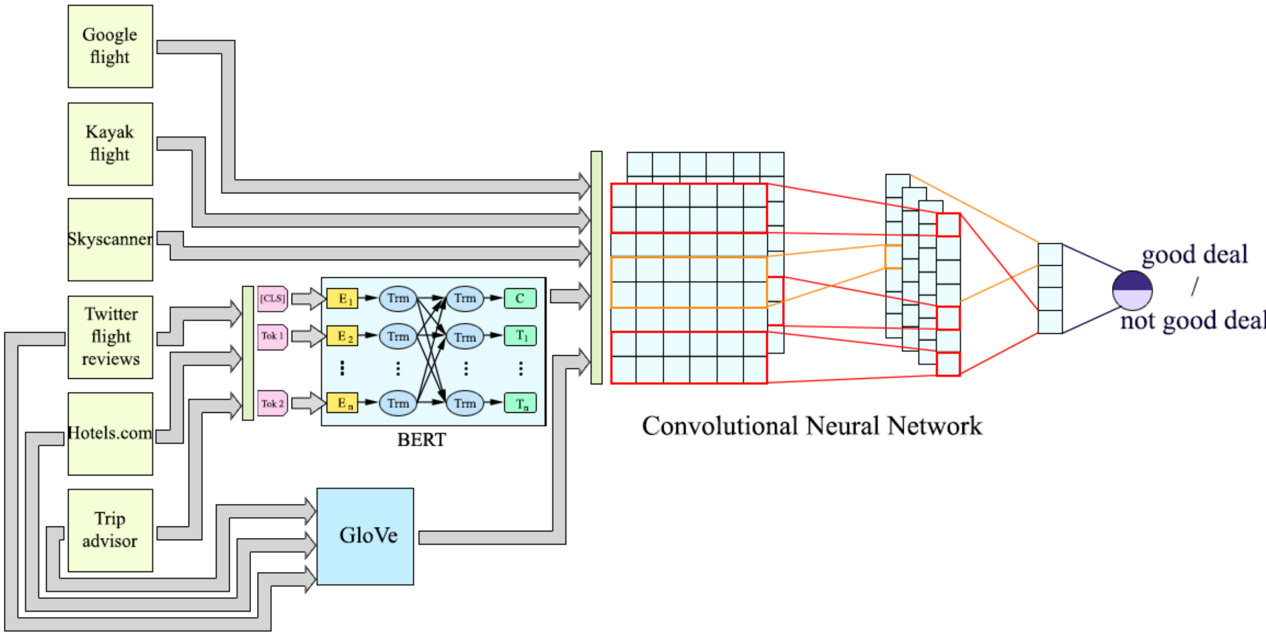

Fig. 1. Our new model to recommend airline tickets

about 280 characters. We choose four major airline companies that have more than 3000 users review tweets. Some example of these airlines is American Air, Delta, Southwest Air, and United.

In this step, we remove all links, pictures, and numbers for a customer's tweets. Tweets that are not in English are eliminated from the original data for this research. Since the focus in this work is sentiment analysis of English vocabulary. After removing stop words, all tweets are lemmatized. In the Lemmatiztaion process, each word, are written just based on its root base of English vocabulary. For example, walking is reduced to walk, and both Stemming and lemmatizing help have clearer text for the classification phase. Different features regarding single tweets are collected and stored in the database such as User ID, Date and time of each tweet, Tweet's text, and tweet ID. We removed all personally identifiable information from the data set before we start sentiment analysis to comply with and respect the user's privacy for online activities. We collect restaurants and hotel information from both Trip Advisor and hotels.com websites by using their website APIs. For each ticket, The reviews related to the top 10 cheapest restaurants and hotels and the average price of hotels and restaurants will be added to each ticket based on the flight destination.

IV. METHOD

Figure1 shows our new model. We collect restaurants and hotel information from both Trip Advisor and hotels.com websites by using their website APIs. For each ticket, The reviews related to the top 10 cheapest restaurants and hotels and the average price of hotels and restaurants will be added to each ticket based on the flight destination. In this work, we also compare the result of sentiments analysis of airline customer's reviews by the Airways magazine and Airquality [28] websites to determine whether the user's reviews in social media is consistent with the real world news about specific Airline or it has inconsistency with other online data.

At first, for sentiment classification of online reviews, we examine two methods. We used the TextBlob library in python to label the tweets into three categories: Positive, negative, and neutral. TextBlob assigns a specific score and polarity to each tweet, which defines the category of the tweets. The TextBlob is based on applying a supervised learning method, Naive Bayes classifier.TextBlob assigns a score of 1 for positive tweets and -1 for negative tweets, and 0 for neutral tweets. The class label is assigned to tweets based on the sentiment label that has the biggest sentiment product.

A. TF-IDF

Also, we built a customized TF-IDF approach for tweet's sentiment classification. We built a customized TF-IDF method for the words in each tweet.TF-IDF method in this research is used for both relevancy of tweet topics and tweet text classification. For finding more relevant documents, we use TF-IDF, which is a term-frequency, inverse document-frequency technique, as a statistical approach to rank words in different documents. A word is considered an important word if it appears several times in one document(TF), and it appears less frequently in all relevant documents(Higher IDF). For the Tf-IDF method, we need to build a language corpus that is relevant to the topic. In this research, to create a query for the TF-IDF method, we consider two different queries. The first query is called a positive query, which checks the similarity of tweets with a positive compliment about airline companies and then classifies the tweet as a positive tweet if the similarity is relatively high. A second Query is called a negative query and checks the tweet's similarity with the list of all possible words and adjectives, which can be seen in complaint reviews.

We need to identify and create a positive word list for positive query and negative word list for the negative query. Positive and negative word lists can be extracted from online customer comments. We create a corpus based on information extracted from the Airline reviews website [28], which is available online. We calculate the multiplication of term frequency in Inverse-Document-Frequency. The following formula is used to find each tweet's similarity to a positive and negative customer's review list.

$$sim(d_j, q) = \frac{\sum_{i=1} w_{i,j} \times w_{i,q}}{\sqrt{\sum_{i=1} w_{i,j}^2} \times \sqrt{\sum_{j=1} w_{i,q}^2}}$$

We compare the textblob(built-in library in python)and the TF-IDF method for online reviews sentiment classification. As shown in this picture, the TextBlob Polarity is compared with a TF-IDF polarity for each tweet. Figure 2 and 3, the comparison between the TF-IDF and Textblob, indicates that the TextBlob can classify the text better than our customized TF-IDF method in this research. For each Airline company, we show that all calculated polarity values with respect to customer reviews. The number of neutral tweets in TF-IDF is more compared with Textblob; it indicates that Textblob is better for sentiment classification of online reviews and can classify tweets more precisely because it has a more comprehensive dictionary of words.

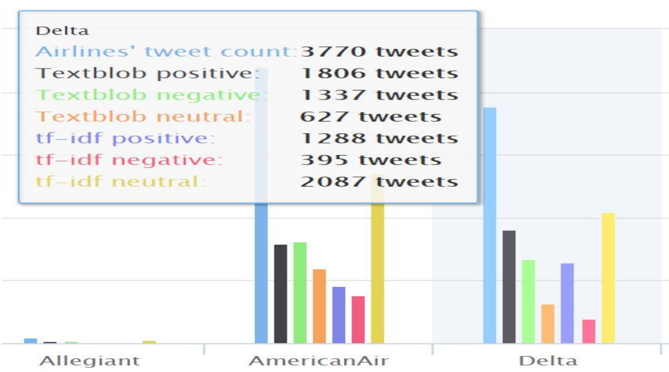

Fig. 2. delta Airline

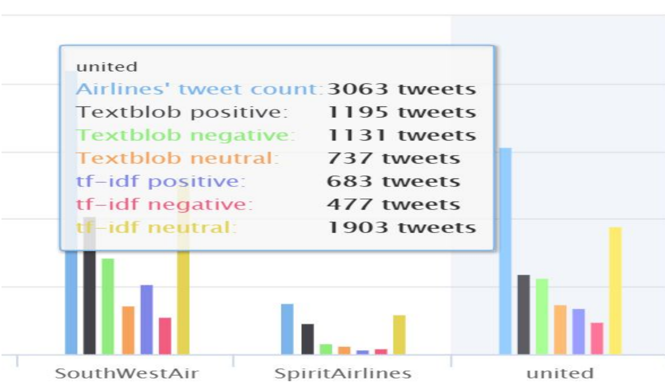

Fig. 3. United Airline

B. BERT

In this work, we use Bidirectional Encoder Representations from Transformers, which is a strong transfer learning model for sentiment classification. The reason that we do not use Textblob or TF-IDF for this research is that they are not computationally efficient for a huge online stream of big with respect to both time and memory space. The first component of the model is the Tweet sentiment classification. In this work, at the first phase of a model, BERT classifies tweets into positive and negative. At first, we trained BERT based on the 50,000 movie reviews [29]. In this paper, the BERT model uses the Stanford Sentiment Treebank $SST2$ [12]. SST2 is a single-sentence classification task based on sentences extracted from movie reviews labeled by human annotators. Using movie reviews in the first phase of the new model in this work is a new approach to classify online airline customer's reviews into positive and negative reviews.

Using BERT, we assign a polarity score, which classifies each online review into three different categories: positive, negative, and neutral. The polarity assigns X score for each tweet's text such that $-0.01 < X < 0.01$, the tweet is a neutral comment. If $0.01 < X$ means a positive comment, and $X < -0.01$ means a negative comment. After sentiment classification, sentiment scores are extracted from the online user's review based on BERT's polarity score. All sentiment scores are passed to the CNN model for the airline classification model. Figure 4 shows that even for the sentiment classification of nearly 14000 tweets, Bert outperformed the TF-IDF method. Figure 5 and Figure6 show an example Temporal and geospatial analysis of the BERT sentiment classification. Figure 5 shows United Airlines review results in two different months and Figure6 shows the sentiment of the online review based on different airline companies in California state.

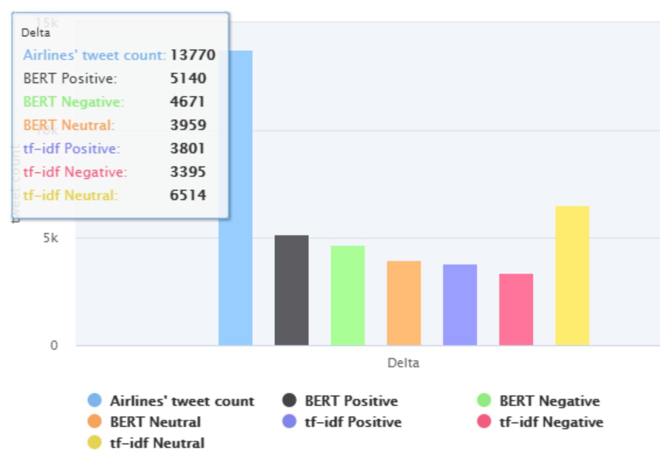

Fig. 4. BERT vs tf-idf comparison

V. THE NEW CNN MODEL FOR AIRLINE TICKET RECOMMENDATION

Based on Figure1, We use GloVe [30] for word embedding of all online user's reviews which extracted from hotels, restaurants,and Twitter. The word embedding phase results create an input space for the Convolutional neural network

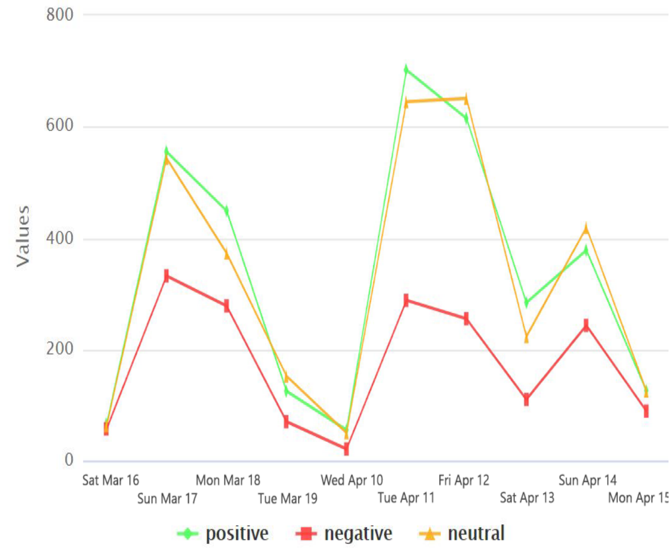

Fig. 5. BERT temporal nalysis for United Airline

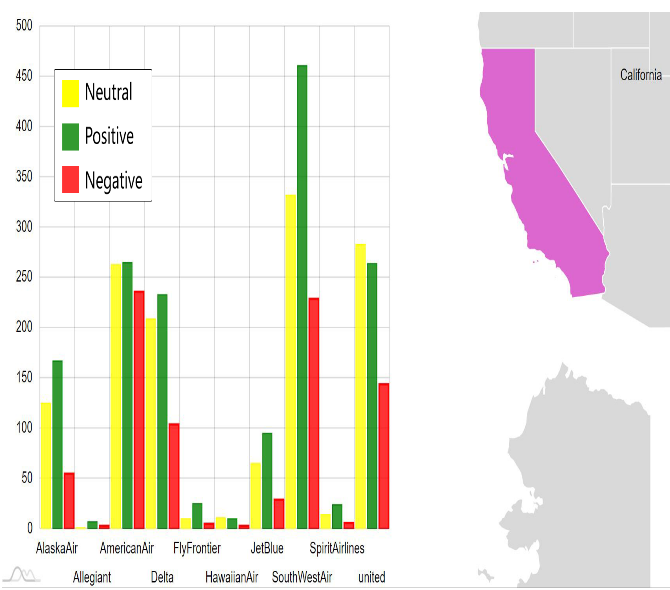

Fig. 6. BERT Sentiment classification for Airline companies in California State

model in the next phase. This work uses GloVe instead of the word2vec model for word embedding since Glove is more scalable and fast for converting huge numbers of reviews in this research to the vector. It also includes a huge global word corpus, which is essential when dealing with various people reviews about different aspects of airline services. This equation shows how the Glove creates the word vector based on a large corpus:

$$J = \sum_{i,j=1}^{V} f(X_{ij})({w_i}^T \widetilde{w}_j + b_i + \widetilde{b}_j - logX_{ij})^2 \qquad (1)$$

TABLE I. COMPARISON OF OUR NEW MODEL BY CONSIDERING BERT SENTIMENT SCORES AND BEST FLIGHT INFORMATION AS INPUT SPACE FOR CNN MODEL

algorithm	f1_score	mcc
CNN without BERT sentiment scores	0.81	0.76
CNN without Best Flight information	0.84	0.79
Our new model	0.92	0.88

V is a Vocabulary size, x is the co-occurrence, matrix, X_{ij} is occurrences of ith word in jth word context, f is weighting function to reduce the effect of the long tail, w_i is word vector (length d) for ith word, w_j is context word vector (length d) for jth word, and b_i and b_j are bias values for i and j, respectively. In our model's final phase, We classified Tickets as a "good deal," which means economic deals and "not a good deal" means not an economical deal. The tickets are labeled as a "good deal" that has the lowest prices(we labeled data as a good price which is explained in the data preprocessing phase), positive reviews for airline services(The data labeled based on hotels and restaurants with sentiment classification and its sentiment scores in the previous phase) in the training set. We split the data set into 70% for the Training set and 30% for the test set. In this work, the input space for convolutional neural network model includes Bert Classification Results, which assign a sentiment scores to each online customer's review based on Twitter passenger reviews, hotels and restaurants reviews, The GloVe word embedding of all online reviews, and the statistics that we extract from google flight, kayak, and Skyscanner about ticket information. Our Convolutional neural network model is a sequential model that contains two-dimensional convolution layer with a flattening layer and two dense layers. We define the hyperparameters and pooling size for the CNN model based on the Microsoft CNN design [31] [32]for semantic data extraction for sentences. We use the CLSM model for CNN in this work. It is one of the best models to extract text semantics. We used Relu as an activation function of the first dense layer and Sigmoid for the final layer. Table I shows the results based on F1 score and MCC. It shows a comparison of our new CNN model, with the CNN model without sentiment information of online reviews or Google flight and kayak and Skyscanner information. The results show that adding information from six different online platforms can increase the accuracy of the CNN model by more than 7%, compared with when we do not have this information.

A. Conclusion

In this work, customer's online reviews in Twitter, hotels, and restaurant reviews of Tripadvisor and extracting ticket information from multiple web sites, including Google flights, play a significant role in the Ticket recommendation's prediction accuracy model. We choose the BERT model over the traditional sentiment classification model(Textblob, TF-IDF) as a robust transfer learning model to classify more than 3 million online reviews based on pre-trained movie reviews. In this research, using movie reviews to create an input space for the Convolutional Neural Network model to recommend airline tickets was a new approach for the airline classification model based on online reviews. Using three components of GloVe, BERT, and Google Flight information to feed the CNN model provides a model with high prediction accuracy to offer the most economical choice for airline tickets to the customers.

References

[1] F. Ebrahimi and S. A. H. Golpayegani, "Personalized recommender system based on social relations," in *Iranian Conference on Electrical Engineering*, pp. 218–223, IEEE, 2016.

[2] G. Williams, M. Tushev, F. Ebrahimi, and A. Mahmoud, "Modeling user concerns in sharing economy: the case of food delivery apps," *Automated Software Engineering*, pp. 1–35, 2020.

[3] P. D. Turney, "Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews," *CoRR*, vol. cs.LG/0212032, 2002.

[4] M. Hosseini, S. Katragadda, J. Wojtkiewicz, R. Gottumukkala, A. Maida, and T. L. Chambers, "Direct normal irradiance forecasting using multivariate gated recurrent units," *Energies*, vol. 13, no. 15, p. 3914, 2020.

[5] M. Hosseini, A. S. Maida, M. Hosseini, and G. Raju, "Inception lstm for next-frame video prediction (student abstract)," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 13809–13810, 2020.

[6] J. Kamps, M. Marx, R. J. Mokken, and M. de Rijke, "Using wordnet to measure semantic orientations of adjectives," in *Proceedings of the Fourth International Conference on Language Resources and Evaluation, LREC 2004, May 26-28, 2004, Lisbon, Portugal*, European Language Resources Association, 2004.

[7] S. Sahu, S. K. Rout, and D. P. Mohanty, "Twitter sentiment analysis - A more enhanced way of classification and scoring," in *IEEE International Symposium on Nanoelectronic and Information Systems, iNIS 2015, Indore, India, December 21-23, 2015*, pp. 67–72, IEEE, 2015.

[8] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," in *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May 2010, Valletta, Malta* (N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, and D. Tapias, eds.), European Language Resources Association, 2010.

[9] Z. Rajabi, A. Shehu, and O. Uzuner, "A multi-channel bilstm-cnn model for multilabel emotion classification of informal text," in *2020 IEEE 14th International Conference on Semantic Computing (ICSC)*, pp. 303–306, 2020.

[10] E. Adeborna and K. Siau, "An approach to sentiment analysis -the case of airline quality rating," in *18th Pacific Asia Conference on Information Systems, PACIS 2014, Chengdu, China, June 24-28, 2014* (K. Siau, Q. Li, and X. Guo, eds.), p. 363, 2014.

[11] F. Jafariakinabad and K. A. Hua, "Maximal sequence mining approach for topic detection from microblog streams," in *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1–8, IEEE, 2016.

[12] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)* (J. Burstein, C. Doran, and T. Solorio, eds.), pp. 4171–4186, Association for Computational Linguistics, 2019.

[13] M. Saadati, J. Nelson, and H. Ayaz, "Mental workload classification from spatial representation of fnirs recordings using convolutional neural networks," in *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6, IEEE, 2019.

[14] M. Saadati, J. Nelson, and H. Ayaz, "Multimodal fnirs-eeg classification using deep learning algorithms for brain-computer interfaces purposes," in *International Conference on Applied Human Factors and Ergonomics*, pp. 209–220, Springer, 2019.

[15] N. Nazari, S. A. Mirsalari, S. Sinaei, M. E. Salehi, and M. Daneshtalab, "Multi-level binarized lstm in eeg classification for wearable devices," in *2020 28th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP)*, pp. 175–181, IEEE, 2020.

[16] T. Razzaghi, I. Safro, J. Ewing, E. Sadrfaridpour, and J. D. Scott, "Predictive models for bariatric surgery risks with imbalanced medical datasets," *Annals of Operations Research*, vol. 280, no. 1-2, pp. 1–18, 2019.

[17] F. Behnia, A. Mirzaeian, M. Sabokrou, S. Manoj, T. Mohsenin, K. N. Khasawneh, L. Zhao, H. Homayoun, and A. Sasan, "Code-bridged classifier (cbc): A low or negative overhead defense for making a cnn classifier robust against adversarial attacks," *arXiv preprint arXiv:2001.06099*, 2020.

[18] A. Mirzaeian, H. Homayoun, and A. Sasan, "Tcd-npe: A re-configurable and efficient neural processing engine, powered by novel temporal-carry-deferring macs," in *2019 International Conference on ReConFigurable Computing and FPGAs (ReConFig)*, pp. 1–8, 2019.

[19] A. Mosallanezhad, G. Beigi, and H. Liu, "Deep reinforcement learning-based text anonymization against private-attribute inference," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, (Hong Kong, China), pp. 2360–2369, Association for Computational Linguistics, Nov. 2019.

[20] F. Jafariakinabad and K. A. Hua, "Style-aware neural model with application in authorship attribution," in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pp. 325–328, IEEE, 2019.

[21] N. Yousefi, M. Alaghband, and I. Garibay, "A comprehensive survey on machine learning techniques and user authentication approaches for credit card fraud detection," *arXiv preprint arXiv:1912.02629*, 2019.

[22] T. A. Oghaz, E. c. Mutlu, J. Jasser, N. Yousefi, and I. Garibay, "Probabilistic model of narratives over topical trends in social media: A discrete time model," in *Proceedings of the 31st ACM Conference on Hypertext and Social Media*, HT '20, (New York, NY, USA), p. 281–290, Association for Computing Machinery, 2020.

[23] M. Heidari, J. H. J. Jones, and O. Uzuner, "Deep contextualized word embedding for text-based online user profiling to detect social bots on twitter," in *IEEE 2020 International Conference on Data Mining Workshops (ICDMW), ICDMW 2020*, 2020.

[24] M. Heidari and J. H. J. Jones, "Using bert to extract topic-independent sentiment features for social media bot detection," in *IEEE 2020 11th Annual Ubiquitous Computing, Electronics Mobile Communication Conference, UEMCON 2020*, 2020.

[25] H. M. Makrani and H. Homayoun, "Mena: A memory navigator for modern hardware in a scale-out environment," in *2017 IEEE International Symposium on Workload Characterization (IISWC)*, pp. 2–11, IEEE, 2017.

[26] H. M. Makrani, H. Sayadi, S. M. P. Dinakarra, S. Rafatirad, and H. Homayoun, "A comprehensive memory analysis of data intensive workloads on server class architecture," in *Proceedings of the International Symposium on Memory Systems*, pp. 19–30, 2018.

[27] H. M. Makrani, H. Sayadi, T. Mohsenin, S. Rafatirad, A. Sasan, and H. Homayoun, "Xppe: cross-platform performance estimation of hardware accelerators using machine learning," in *Proceedings of the 24th Asia and South Pacific Design Automation Conference*, pp. 727–732, 2019.

[28] "Airline quality." https://www.airlinequality.com/.

[29] "Imdb dataset of 50k movie reviews." https://www.kaggle.com/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews.

[30] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL* (A. Moschitti, B. Pang, and W. Daelemans, eds.), pp. 1532–1543, ACL, 2014.

[31] J. Gao, P. Pantel, M. Gamon, X. He, and L. Deng, "Modeling interestingness with deep neural networks," Tech. Rep. MSR-TR-2014-56, October 2014.

[32] Y. Shen, X. He, J. Gao, L. Deng, and G. Mesnil, "A latent semantic model with convolutional-pooling structure for information retrieval," in *CIKM*, November 2014.