BiLSTM-ANN Based Employee Job Satisfaction Analysis From Glassdoor Data Using Web Scraping

Muhammed Yaseen Morshed Adib
Computer Science and Engineering
BRAC University
Dhaka Bangladesh
yaseen.morshed.adib@g.bracu.ac.bd

Sovon Chakraborty
Computer Science and Engineering
BRAC University
Dhaka Bangladesh
sovon.chakraborty@bracu.ac.bd

Munim Bin Muquith
Computer Science and Engineering
BRAC University
Dhaka Bangladesh
munim.bin.muquith@g.bracu.ac.bd

Annajiat Alim Rasel
Senior Lecturer
Computer Science and Engineering
BRAC University
Dhaka Bangladesh
annajiat@bracu.ac.bd

***Abstract*—The job sector of a country varies due to numerous factors. Employee job satisfaction is a vital element for the growth of an organization. IT and Software is a growing industry in Bangladesh. An extensive number of students pushes their career to be setted in renowned IT and Software companies after the study. During the period of Covid-19 the job sectors have shifted a lot. Companies have changed their strategies in order to survive. These strategies have a direct effect on employee job satisfaction. Analyzing employee job satisfaction will lead to understanding major pros and cons of a country's economy. Social media platform where people often express their opinions and views. Glassdoor is a USA based social networking platform where current and former employees express the pros and cons of organizations. Taking all of these into consideration, authors in this research attempted to analyze the job satisfaction of employees. A hybrid deep learning based architecture BiLSTM-ANN is proposed in order to understand employee job satisfaction in a shorter period of time. At first reviews of 12 renowned IT and Software companies were scraped from Glassdoor.com. After Preprocessing a specialized dataset is built to feed into the proposed architecture. Five polarities namely Super positive, Positive, Neutral, Negative and Super negative are assigned based on the rating given in Glassdoor.com Earlier result shows, the proposed BiLSTM-ANN model outperformed the state-of-the-art architectures in terms of different evaluations. The model exhibits 88.64trainable parameters than other architectures. This model is suitable for detecting polarities of employees in computing devices.**

***Index Terms*—sentiment analysis, deep learning, BiLSTM, BiLSTM-ANN, BIGRU (key words)**

I. Introduction

Because of the advancements in science and information technology, the world is now seen as a global village [1]. Social media is used by more than 50% of the world's population. The utilization of are spread in social media for information, marketing, other online activities, and entertainment in addition to entertainment. People all over the world can claim that today is a digital era in which information technology is thoroughly utilized [2]. Social media is also another medium where people their daily happy and frenetic activities [3]. Since there is now more competition than before due to COVID-19, finding a job has grown more challenging. The outbreak caused many workers to lose their jobs, some of them to change fields, and many more. In Japan, 21% of people are unhappy with their current jobs [4]. How satisfied a worker is at work is influenced by a variety of factors, including the task at hand, the quality of the work, their relationship with the boss, and a host of others. A company will remain in operation for a very long time if its employees are content. Enhancing employee satisfaction may increase morale and satisfaction with the organization, which enhances overall organizational effectiveness. A worker's ability to progress professionally, maintain their physical and mental health, and a variety of other things depend on how satisfied they are at work [4]- [6]. It's important for both the employer and the employee to be happy in their jobs. IT and Software is a developing sector in Bangladesh where a certain number of companies are larger in size. During the pandemic the job sector has been affected heavily. Glassdoor is a reputed social media site where people throughout the whole world review their working companies [7]. A number of reviews concerning Bangladeshi top IT and Software companies are also available in Glassdoor.com. Online reviews provide the employee a chance to be heard, and the attributes that the employee values the most should be investigated to ascertain

Identify applicable funding agency here. If none, delete this.

the degree of enjoyment it brings. However, there are some restrictions on these internet statistics. For instance, this reviews quality cannot be guaranteed because anyone can express their thoughts online without boundaries. For this research, authors have considered analyzing the reviews from Glassdoor.com related to the top 12 Software and IT companies of Bangladesh. Web scraping is a procedure through which authors have gathered necessary information into a Comma separated value(CSV). Sentiment analysis is a method for locating ambiguity in words, opinions etc. It reveals the sentiments of certain users regarding a specific topic [8]. Numerous Deep learning (DL) architectures and Machine learning(ML) algorithms are used to identify polarity of users [9]- [12].DL architectures are heavily used in the case of product reviews, movie reviews and other sentiment analysis related tasks. Observing these incidents, authors have proposed a hybrid BiLSTM-ANN architecture in order to analyze reviews with regard to some of the top software companies of Bangladesh. The major contributions of this research paper are given below:

- Proposing a Deep learning based BiLSTM-ANN based architecture that can detect sentiments of an employee more efficiently than the state-of-the-art architectures.
- the job market of Bangladesh IT and Software sector along with understanding the perspective of the attached employees.
- price increases are a recurring phenomenon and wages are insufficient to cover expenses, the analysis can also help studies about changes in human behavior patterns over time.
- A specialized dataset is built and preserved for further utilization in the concerned research field.

The further sectors of this research paper are segmented as follows: In Section 2, a concise survey of pertinent literature on sentiment analysis is provided. The suggested process to fulfill the required assignment of classification of job satisfaction reviews is described in Section 3. Section 4 displays the outcomes of the experiment using the proposed model. In the mentioned sections, a comparative analysis between the proposed model and baseline models are also presented.

II. LITERATURE REVIEW

Luo et al. has Offered a fresh perspective regarding employee job satisfaction with organizational performance [13]. Anonymous employee reviews are gathered from Glassdoor.com. The primary purpose was to perform textual analysis. In total 2,74,061 reviews were considered between 2008-2014. Inability of using proper regression analysis is the main limitation of this research. In [14] the study did find a correlation between employee satisfaction, operating company financial health, and top management's investment stance, but it did not fully establish the mechanism underlying this relationship. The proposed architecture had an accuracy rate of 81.7%. Authors in [15] considered an aspect-based sentiment analysis to let job seekers know which firm best meets their needs. Parts of Speech Tagger (POS Tagger) is used for data preprocessing purposes. Though, no new architecture has been proposed in this research. ML algorithms have a great impact in understanding sentiments of users. Supervised ML algorithms are used in [16] to identify the employee job satisfaction comparing the salary they are earning. The Multivariate Bernoulli Naive Bayes classification and Multinomial Naive Bayes classification are used by Singh et al. to identify whether a certain news story is positive or negative [17]. Only two polarities are considered here for sentiment analysis. ML algorithms often suffer in numerous cases where DL architectures perform constantly better. LSTM and BiLSTM are two Recurrent Neural Network (RNN) based architectures is a widely used architecture that is used to identify polarities in different languages [18]- [20]. Artificial Neural Network (ANN) is renowned for understanding complex patterns indata. Hybrid models are applied in order to improve the performance in identifying certain tasks [21]. In terms of analyzing Job satisfaction, the application of DL architectures is very limited. In the respected field, there is no utilization of hybrid models. Taking these research gaps into consideration, authors have proposed a DL based hybrid architecture combining BiLSTM and ANN. A specialized dataset is built through web scraping from Glassdoor.com. The proposed model outperforms the state-of-the-art models in terms of various performance metrics.

III. RESEARCH METHODOLOGY

In this section, a detailed description of the proposed methodology has been discussed. A brief analysis of the gathered dataset along with the programming environments are depicted in Figure 1. At first data was scraped from Glassdoor.com. After that preprocessing is performed upon the scraped data that includes stopwords removal, punctuation removal, hyperlink removal, URLs removal and unnecessary space removal. After that tokenization is performed upon that dataset. Before feeding the dataset to the proposed model, word embedding is applied. 70% data are utilized for training purposes where 30% data are preserved for validation purposes. After proper training and validation, a detailed comparative analysis is shown between the proposed model and baseline architectures.

A. Programming Environments

Python programming language is considered here for implementing the DL architecture related tasks. It offers a wide range of libraries that may be used to put different deep learning methods into practice. For implementation the latest version of Python 3.10.8 is utilized. A large-scale machine learning package for

numerical computing is called TensorFlow. A Python data visualization package built on matplotlib is called Seaborn.

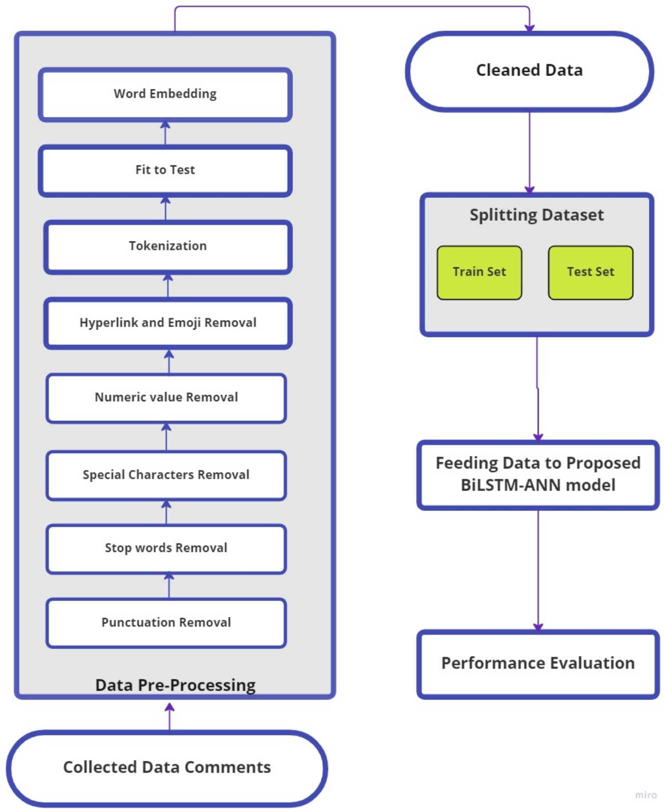

Fig. 1. Elaboration of the proposed methodology

The tool BeautifulSoup is used to extract data from HTML and XML after web scraping. Open-source neural network library Keras is listed on the Python library list. It can be used with R, Theano, Microsoft Cognitive Toolkit, or PlaidML. The final section provides an explanation of the experimental details of deep learning classifiers.

B. Dataset Description and Analysis

Authors have considered two datasets for evaluating the performance of the proposed DL architecture. The first dataset is scraped from Glassdoor.com and the second dataset is the Amazon product review dataset. Authors have observed the performance of the proposed model in case of the both datasets and draw a proper conclusion. The descriptive analysis of each dataset is stated below:

1) Scraped Dataset From Glassdoor: For the purpose of data collection, authors have used Glassdoor.com as a typical social networking website. The gathered information is taken directly from Glassdoor.com. The Python programming language is used for the entire scrapping process. A good number of Python libraries are available for scraping data. The "urllib" library is one that Python programmers must have because it has the modules they need to interact with the required URLs. The "urllib" request module specifically contains the "urlopen()" function. In most cases, urlopen gives back an HTTPResponse object. In order to scrape the HTML from the page, the HTTPResponse object's.read() method is first invoked, which results in a string of bytes. Use decode() to turn the data into a UTF-8 string after that. The HTML code of the contents is then correctly visible after that. After the completion of web scraping finally the dataset is built. In the collected dataset, there are in total 11,505 reviews gathered from 12 companies. These companies recruit a good number of employees each year. Three major columns are scraped namely "Rating", "Pros" and "Cons". When talking individually Rating consists of a value between 1 - 5. The rating can be considered into fractional numbers also. But in the pre-processing section, the numbers are converted into the nearest decimal point. The collected data is categorized into five major polarity values. The annotation is performed based on rating. A comment attached with a 5 star is annotated as a polarity of super positive where 1.00 is considered as super negative Table I describes the attributes of Values regarding the polarity.

Corresponding Polarity	Rating	Total Value Counts
Super Positive	5.00	3312
Positive	4.00	2662
Neutral	3.00	2764
Negative	2.00	1625
Super Negative	1.00	1142

TABLE I
POLARITY VALUE COUNTS OF THE COLLECTED DATASET FROM GLASSDOOR.COM

2) Amazon Product Review Dataset: For this research, this dataset is also considered to evaluate the performance of the baseline approaches. In this dataset, 1600 items are considered, where the attributes are ID, Rating, review sources, review comments, weights, prices and manufacturers name are provided. For this research, only the review comments and rating are being considered. The dataset is gathered from kaggle.com (https://www.kaggle.com/datasets/saurav9786/amazon-product-reviews).

C. Data Preprocessing

The primary purpose of data preprocessing is the removal of inconsistencies between data. Besides that, increasing reliability and accuracy is another reason for preprocessing the achieved dataset. The description of data preprocessing phases is described sequentially. For both dataset, the same preprocessing techniques have been applied.

1) Erase Punctuation: The percentage of punctuation inside a text is usually relatively higher than other characters. In the case of the sentiment analysis models, punctuations do not bear any specific meaning. For presenting data in a normalized form all the punctuations are erased from the dataset. Usually, text length gets reduced with the removal of punctuation marks. Consider the below sentence:

" Pipilika software!!!, They are great in terms of employee satisfaction ...."

After the removal of punctuations, the sentence will be:

"Pipilika software They are great in terms of employee satisfaction"

2) Converting into Lowercase: Usually, reviews in Glassdoor data contain reviews without following any grammar norms. The text can be in both lower-case and upper-case characters. The same word can be available in both options also. In this research, authors have considered many cases that are case-sensitive. Deep learning architectures might face difficulty determining the proper meaning of the provided text with characters in both lowercase and uppercase. For that reason, all the alphabets are converted into lowercase. In the case of the previously mentioned sentence, the sentence will turn into the following form after converting it into lower case.
"pipilika software they are great in terms of employee satisfaction"

3) Stopwords, Hyperlink, and Emoji removal: ambiguous when there exists a huge number of stopwords inside text. As a result, there is a chance of producing unexpected predictions. The sentence **"pipilika software they are great in terms of employee satisfaction"** converts into **"pipilika software great terms employee satisfaction"** after removing all the unnecessary stopwords. Hyperlinks and emojis are two criteria that are removed from the collected text also. Hyperlinks are helpful for functionality purposes. On the contrary, Emojis are also removed in account to learn deep learning classifiers properly.

4) Unnecessary Space Removal: After scraping the dataset, authors observed there exists unnecessary spaces in between words in a text. For mining the meaning of a text in a more precise way, these spaces have been removed from the acquired dataset.

5) Tokenization: For dividing the text streams into phrases or small slabs of textual materials tokenization is usually performed. Complex textual meanings can be understood in a precise way with the use of tokenization. Both semantics and lexical analysis are crucial while determining polarity from a certain text. Tokenization preserves the semantic meaning of a text properly. After tokenization the sentence turned into
[[pipilika], [software], [great], [terms], [employee], [satisfaction]]

The above operation is performed using the keras.tokenizer where all the necessary words are tokenized.

6) Fixed Padding Length: One of the major challenges faced by authors in this research is not all the texts are of the same length. Usually, people place their comments distinctly from one to another. As a result, not all comments are not the same length. That's why padding is performed to make all the reviews of the same length. Padding merely entails the final addition of zero layers to our input review to guarantee that each employee review is the same length. Keras Pad-Sequence is utilized in order to perform the padding operation.

7) Word Embedding: Word embedding uses numerical and vector representations of each word. Word embedding describes texts that contain precise replicas of words that have the same meaning. Word embedding in particular is unsupervised word representation learning that is comparable to semantic similarity. In a coordinated scheme, comparable phrases are clustered closer together based on a set of relationships. After performing the word embedding the sentence is converted into:

[[234], [142], [1021], [533], [647], [1436]]
Along with all the polarity values are converted into numerical numbers. At first, the rating is converted into its closest decimal number. So, 4.2 becomes 4 where 4.7 is converted into 5. A rating of 4.5 is also considered as 5. After that the polarity values are converted into numerical numbers. So, 5 is assigned to Super positive, 4 is assigned to Positive. On the other hand, 3 is considered as Neutral, 2 as Negative and 1 as Super Negative.

IV. The Proposed Model Architecture

The detailed architecture of the proposed model is described in Figure 3. The authors primarily focused on the data preposessing phase. After the processing, the dataset is passed to the deep learning classifiers. The authors focused on the state-of-the-art architectures at first that showed excellent accuracy in previous research. After that authors proposed an ensemble model BiLSTM-ANN. The hybrid model is initiated using the Sequential() function from Keras. In the next phase, the words are embedded using the word embedding method. Dropping the feature map instead of individual elements helps the deep-learning classifiers to understand the meaning properly. To accomplish this task SpatialDropout() is applied. Then the data is fed into the Bi-LSTM model. The reasons behind selecting Bi-LSTM model are stated below:

- Bi-LSTM has the ability to flow information from both sides.
- Sequential dependencies between words and phrases are understood better by BI-LSTM architecture.
- The primary explanation is that each element of an input sequence contains data from the past and the present. For this reason, by integrating LSTM layers from both directions, Bi-LSTM can generate an output that is more meaningful.

Figure 2 shows a detailed view of the BiLSTM architecture. Bi-Directional LSTMs (BiLSTMs) bypass this restriction by analyzing the input sequence in both the forward and backward directions, combining the information from both ends to create a single representation, as opposed to LSTMs which can only use past contexts. As a result, they are able to catch significantly richer sequential patterns in both directions and learn a far better feature representation for the input query.

The weight matrix for Bi-LSTM is randomly initialized where the Dropout() function is applied in order to ignore the model to be overfitted. Later, the model is sent to a number of ANN layers for understanding the complex interactions. The hyperparameter details for this model are given below in Table II. The detailed architecture of the dense layer and fully connected layer are shown in Figure.. For introducing non-linearity ReLU activation function has been used by the authors. All the hyperparameters are fine- tuned.

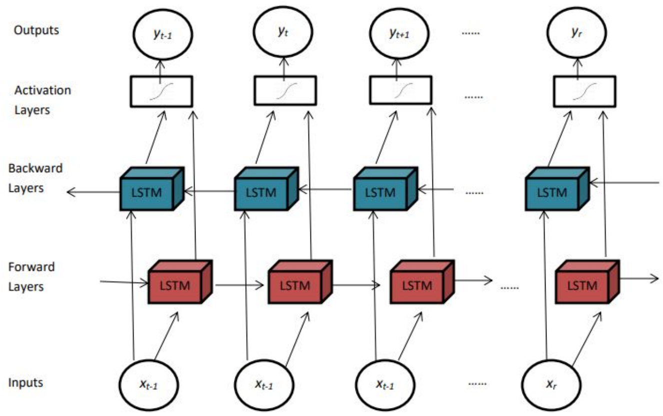

Fig. 2. BiLSTM Architecture for this research

Name of The parameters	Quantity
Maximum Length	24
Batch Size	256
Word Vector Dimensionality	124
Non Linearity	ReLU
Number of BiLSTM cells	120
Optimizer	Adam Stochastic Gradient Descent
Learning rate	0.0001
Epsilon	1e-06
Activation function	Softmax and ReLU
Loss function	Binary Cross$_{e}ntropy$
Recurrent Dropout	0.002
Number of epochs	20
Number of trainable parameters	4,24,435
Percentage of Test set	30%

TABLE II
HYPERPARAMETER TUNING FOR THE PROPOSED BILSTM-ANN ARCHITECTURE

Table 2 represents the detailed hyperparameter tuning of the proposed BiLSTM-ANN model. For introducing non linearity ReLU activation function is used. In the case of ANN dense layers, the learning rate is set to 0.0001. At the fully connected layers Soft max activation function is applied for achieving better performance. To ignore over-fitting, Dropout is applied where the loss function binary-cross entropy is used for multi class sentiment analysis. Batch size is set to 256 with a maximum length of 24. To understand from error efficiently Adam stochastic gradient descent is performed upon that layer. To understand the detailed working procedure, authors have provided Figure 3. First, prepossessing is done by mentioning all the necessary steps above. After performing the word embedding, the datasets are fed into the proposed model. The model consists of 120 BiLSTM blocks followed by a dropout layer. After that, flatten is applied to the dataset and finally it is passed to ANN layers. In the first fully connected layer ReLU activation function is used where Soft-max activation function is used in the second fully connected layer. Finally the sentiment prediction and detection is justified by the proposed model. Dropout is applied in between all layers to make sure no overfitting takes place in between the layers.

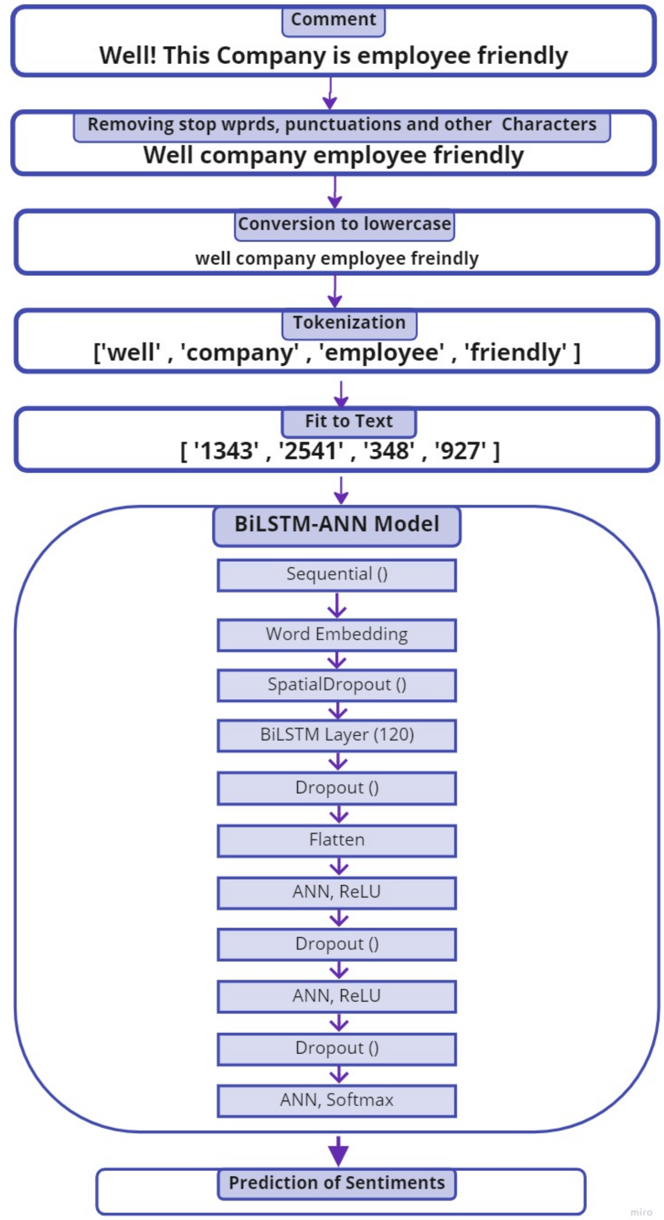

Fig. 3. BiLSTM Architecture for this research

V. RESULT ANALYSIS AND DISCUSSION

A. Evaluation Metrics

To evaluate how the model is performing, authors have compared the performance with some of the baseline approaches. For comparison purposes, some evaluation metrics have been taken into account.

1) Precision: The precision shows how accurate the classification is. Lower accuracy and fewer false positives are the results of both low accuracy and high precision [22]. Reduced sensitivity is a result of improved precision, which has an inverse relationship to sensitivity. Precision can be calculated

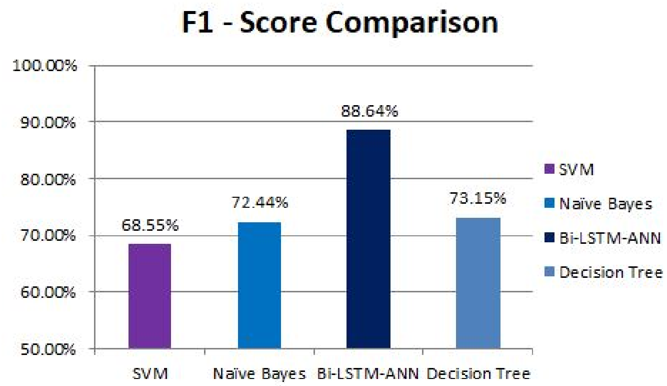

Fig. 4. Comparison of the proposed model with the previously studied ML algorithms

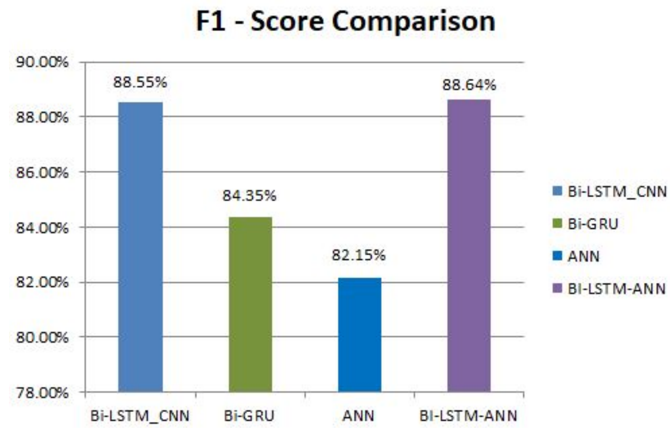

Fig. 5. Comparison of the proposed model with the previously studied DL architectures

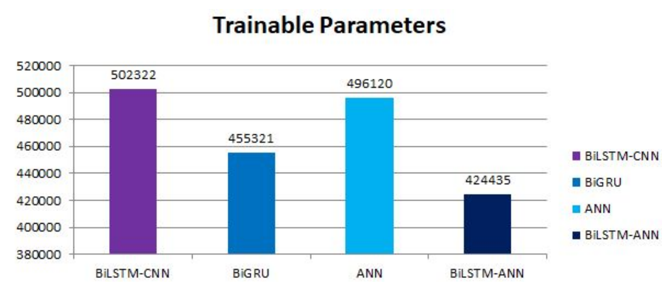

Fig. 6. Comparison of trainable parameters of the mentioned models

by the below mentioned equation,

$$Precision = TP/(TP + FP) \tag{1}$$

Where TP = True positive and FP = False positive.

2) Recall / Sensitivity: By using a few straightforward techniques, percentage can be rapidly determined [23]. Higher sensitivity leads to fewer false negatives, while lower sensitivity leads to more false negatives. Sometimes, accuracy degrades as sensitivity rises. False Negative (FN) plays a vital role in the case of determining recall.

$$Recall = TP/(TP + FN) \tag{2}$$

3) F1- Score: F1-Measure combines sensitivity and precision. The weighted harmonic method is a sensitive, accurate, and precise method [24]. The F1 measurement has been shown to be both precise and effective. All the four observation including True Negative (TN) and False positive (FP) is also considered here.

$$F1 - Score = TP + FP/(TP + FP + TN + FN) \tag{3}$$

B. Result Analysis

For this research, authors have compared the proposed model with the baseline approaches. Two datasets have been used to observe the performance of the proposed result. In both cases, the proposed BiLSTM-ANN has shown promising results.

1) Comparison With the State-of-the-art ML Algorithms: Sentiment analysis based on ML architectures are elaborately discussed in [25]- [27]. The baseline approaches worked on almost similar tracks. Though there are few major ML algorithms that are used for job satisfaction. Authors have not found any previous research on employee Job satisfaction in Bangladeshi IT and Software companies. A detailed comparison is shown in Figure 4 in between the proposed architecture and the state-of-the-art ML algorithms.

Figure 4, describes that, the proposed model shows a maximum F1-score of 88.64% where previously Decision Tree has the second highest F1-score. Understanding the complex text pattern and feature extraction is the main boundary for the ML algorithms.

2) Comparison With the State-of-the-art DL Architectures: In addition to F1-score, trainable parameters should also be taken into account while comparing various DL architectures. Figure 5 demonstrates that the suggested BiLSTM-ANN model reveals a higher F1- score when compared to state-of-the-art DL architectures, even if less trainable parameters are chosen.

The proposed architecture almost shows a similar F1-score with BiLSTM-CNN approach [28] where other architectures [29]- [30] fall short in terms of accuracy. In both cases, BiLSTM is common. The major reason behind the success of the proposed model is stated as follows:

- Though CNN has a more accurate way of solving complex problems yet fine tuning of proper parameters allows the proposed model to show more accuracy.
- CNN is extremely accurate in the case of Computer vision related problems where ANN allows better where complex patterns exist inside text data.
- For limited data, ANN performs better than CNN as there is less data available for training.

Figure 6 depicts that, less trainable parameters are used in the BiLSTM-ANN model. As all the necessary hyper parameters are fine-tuned that is why the proposed model detects the

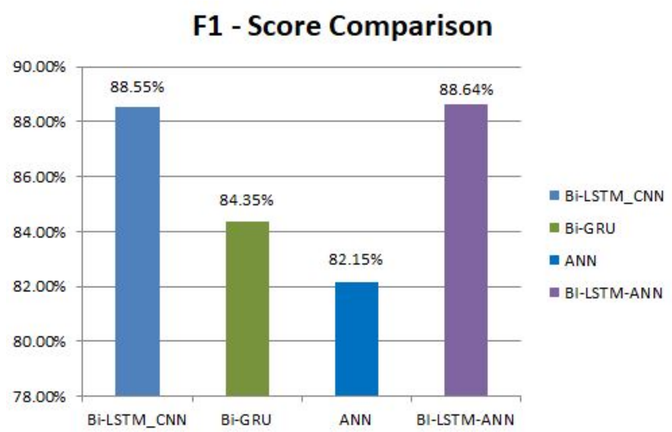

Fig. 7. Results Analysis on Amazon Product Review Dataset

sentiment of texts in a lesser amount of time. The training time is also fewer than the other baseline approaches.

3) Classification report on the Glassdoor dataset: Table 3, shows a detailed classification report shown by the Proposed BiLSTM-ANN model. Authors have founds precision and recall is lesser in the case of Negative and Neutral data as sometimes they are hard to distinguish. For example, "This is okay", the internal meaning of this sentence depends on the users perspective. Apart from that other classifications are satisfactory.

Polarity Metrics	Precision	Recall	F1-Score
Super Positive	0.92	0.93	0.92
Positive	0.87	0.90	0.89
Neutral	0.74	0.81	0.79
Negative	0.77	0.84	0.80
Super Negative	0.89	0.91	0.90

TABLE III
CLASSIFICATION REPORT OF THE PROPOSED MODEL

4) Results Analysis on Amazon Product Review Dataset : Authors have also observed the result shown by the models in the case of Amazon product review dataset. Figure 7 describes the result shown by the proposed model.

VI. CONCLUSION

In this research, the authors have proposed a deep learning based BiLSTM-ANN model for analyzing employee job satisfaction scraped from Glassdoor.com where all the hyper-parameters are fine-tuned. 11,505 reviews from top IT companies in Bangladesh are only considered here. The proposed model outperforms all the state-of-the-art models in terms of numerous evaluations in sentiment classification and analysis. A specialized scraped dataset is collected from Glassdoor.com that can be further utilized for further research in this field. In the near future, authors will further investigate this model with other benchmark architecture to find best results in this field.

REFERENCES

[1] Ajmal, A.; Aldabbas, H.; Amin, R.; Ibrar, S.; Alouffi, B.; Gheisari, M.J.C.I. Stress-Relieving Video Game and Its Effects: A POMS Case Study. Comput. Intell. Neurosci. 2022,, 4239536.

[2] Khan, S.; Muhammad, K.; Mumtaz, S.; Baik, S.W.; de Albuquerque, V.H.C. Energy-efficient deep CNN for smoke detection in foggy IoT environment. IEEE Internet Things J. 2019, 6, 9237–9245.

[3] Shamrat, M.F.M.J.; Chakraborty, S.; Imran, M.M.; Muna, N.J.; Billah, M.M.; Das, P.; Rahman, O.M. Sentiment analysis on twitter tweets about COVID-19 vaccines using NLP and supervised KNN classification algorithm. Indones. J. Electr. Eng. Comput. Sci. 2021, 23, 463–470.

[4] Das Swain, V., Saha, K., Reddy, M. D., Rajvanshy, H., Abowd, G. D., amp; De Choudhury, M. (2020, April). Modeling organizational culture with workplace experiences shared on glassdoor. In Proceedings of the 2020 CHI conference on human factors in computing systems (pp. 1-15).

[5] Lee, Y., Ng, S., Shevlin, T., amp; Venkat, A. (2021). The effects of tax avoidance news on employee perceptions of managers and firms: Evidence from glassdoor. com ratings. The Accounting Review, 96(3), 343-372.

[6] Lam, J., Mulvey, M. S., amp; Robson, K. (2022). Looking through the Glassdoor: The stories that B2B salespeople tell. Industrial Marketing Management, 105, 478-488.

[7] 7.Gottipati, S., Shim, K. J., amp; Sahoo, S. (2021, April). Glassdoor Job Description Analytics–Analyzing Data Science Professional Roles and Skills. In 2021 IEEE Global Engineering Education Conference (EDUCON) (pp. 1329-1336). IEEE.

[8] Li, W., Shao, W., Ji, S., amp; Cambria, E. (2022). BiERU: Bidirectional emotional recurrent unit for conversational sentiment analysis. Neurocomputing, 467, 73-82.

[9] Chakravarthi, B. R., Priyadharshini, R., Muralidaran, V., Jose, N., Suryawanshi, S., Sherly, E., amp; McCrae, J. P. (2022). Dravidian-codemix: Sentiment analysis and offensive language identification dataset for Dravidian languages in code-mixed text. Language Resources and Evaluation, 1-42.

[10] Wang, W., Guo, L., amp; Wu, Y. J. (2022). The merits of a sentiment analysis of antecedent comments for the prediction of online fundraising outcomes. Technological Forecasting and Social Change, 174, 121070.

[11] Iqbal, A., Amin, R., Iqbal, J., Alroobaea, R., Binmahfoudh, A., amp; Hussain, M. (2022). Sentiment Analysis of Consumer Reviews Using Deep Learning. Sustainability, 14(17), 10844.

[12] Yi, S., amp; Liu, X. (2020). Machine learning based customer sentiment analysis for recommending shoppers, shops based on customers' review. Complex amp; Intelligent Systems, 6(3), 621-634.

[13] Chinazzo, G. (2021). Investigating the indoor environmental quality of different workplaces through web-scraping and text-mining of Glassdoor reviews. Building Research amp; Information, 49(6), 695-713.

[14] Uchida, E., amp; Kino, Y. (2021). Study on the Relationship Between Employee Satisfaction and Corporate Performance in Japan via Text Mining. Procedia Computer Science, 192, 1730-1739.

[15] Dina, N. Z., amp; Juniarta, N. (2020). Aspect-based Sentiment Analysis of Employee's Review Experience. Journal of Information Systems Engineering and Business Intelligence, 6(1), 79-88.

[16] Das Swain, V., Saha, K., Reddy, M. D., Rajvanshy, H., Abowd, G. D., amp; De Choudhury, M. (2020, April). Modeling organizational culture with workplace experiences shared on glassdoor. In Proceedings of the 2020 CHI conference on human factors in computing systems (pp. 1-15).

[17] Singh, G., Kumar, B., Gaur, L., amp; Tyagi, A. (2019, April). Comparison between multinomial and Bernoulli naïve Bayes for text classification. In 2019 International Conference on Automation, Computational and Technology Management (ICACTM) (pp. 593-596). IEEE.

[18] Kowsher, M., Sami, A. A., Prottasha, N. J., Arefin, M. S., Dhar, P. K., amp; Koshiba, T. (2022). Bangla- BERT: transformer-based efficient model for transfer learning and language understanding. IEEE Access, 10, 91855-91870.

[19] Setyanto, A., Laksito, A., Alarfaj, F., Alreshoodi, M., Oyong, I., Hayaty, M., ... amp; Kurniasari, L. (2022). Arabic Language Opinion Mining Based on Long Short-Term Memory (LSTM). Applied Sciences, 12(9), 4140.

[20] Tan, K. L., Lee, C. P., Lim, K. M., amp; Anbananthen, K. S. M. (2022). Sentiment analysis with ensemble hybrid deep learning model. IEEE Access, 10, 103694-103704.

[21] Thakkar, A., Mungra, D., Agrawal, A., amp; Chaudhari, K. (2022). Improving the performance of sentiment analysis using enhanced preprocessing technique and Artificial Neural Network. IEEE transactions on affective computing.

[22] Rahat, F. R., Mim, M., Mahmud, A., Islam, I., Mahbub, M. J. A., Mehedi, M. H. K., amp; Rasel, A. A.(2022). Data Analysis Using NLP to Sense Human Emotions Through Chatbot. In International Conference on Advances in Computing and Data Sciences (pp. 64-75). Springer, Cham.

[23] Tanvir, R., Shawon, M. T. R., Mehedi, M. H. K., Mahtab, M. M., amp; Rasel, A. A. (2022). A GAN-BERT Based Approach for Bengali Text Classification with a Few Labeled Examples (No. 8409). EasyChair.

[24] Biswas, A., Chakraborty, S., Rifat, A. N. M. Y., Chowdhury, N. F., amp; Uddin, J. (2020, August). Comparative Analysis of Dimension Reduction Techniques Over Classification Algorithms for Speech Emotion Recognition. In International Conference for Emerging Technologies in Computing (pp. 170- 184). Springer, Cham.

[25] AlBadani, B., Shi, R., amp; Dong, J. (2022). A novel machine learning approach for sentiment analysis onTwitter incorporating the universal language model fine-tuning and SVM. Applied System Innovation, 5(1),13.

[26] Kewsuwun, N., amp; Kajornkasirat, S. (2022). A sentiment analysis model of agritech startup on Facebook comments using naive Bayes classifier. International Journal of Electrical amp; Computer Engineering (2088-8708), 12(3). (EDUCON) (pp. 1329-1336). IEEE.

[27] Anwar, M. K. M. K., Yusoff, M., amp; Kassim, M. (2022, May). Decision Tree and Naïve Bayes for Sentiment Analysis in Smoking Perception. In 2022 IEEE 12th Symposium on Computer Applications amp; Industrial Electronics (ISCAIE) (pp. 294-299). IEEE.

[28] Rhanoui, M., Mikram, M., Yousfi, S., amp; Barzali, S. (2019). A CNN-BiLSTM model for document-level sentiment analysis. Machine Learning and Knowledge Extraction, 1(3), 832-847 29. Han, Y., Liu, M., amp; Jing, W. (2020). Aspect-level drug reviews sentiment analysis based on double BiGRU and knowledge transfer. IEEE Access, 8, 21314-21325.

[29] Han, Y., Liu, M., amp; Jing, W. (2020). Aspect-level drug reviews sentiment analysis based on double BiGRU and knowledge transfer. IEEE Access, 8, 21314-21325.

[30] Thakkar, A., Mungra, D., Agrawal, A., amp; Chaudhari, K. (2022). Improving the performance of sentiment analysis using enhanced preprocessing technique and Artificial Neural Network. IEEE transactions on affective computing.