International Conference on Data Mining, Big Data, Database and Data Management (ICDMBDDDM 2021)
New York USA Jan 28-29, 2021, Part XI

# Emotion Classification using Recurrent Neural Network and Scalable Pattern Mining

Angelina A. Tzacheva, Jaishree Ranganathan, MuthuPriya Shanmugakani Velsamy, Shamika Kulkarni
Department of Computer Science
University of North Carolina at Charlotte
Charlotte, North Carolina, USA, 28223
Email: aatzache@uncc.edu
Email: jrangan1@uncc.edu
Email: mshanmu2@uncc.edu
Email: skulka26@uncc.edu

*Abstract*—Emotions play an important role in everyday life. Analyzing these emotions or feelings from the social media platforms like Twitter, Facebook, Blogs, and Forums based on user comments and reviews plays an important role in various factors. Some of them include brand monitoring, marketing strategies, reputation, and competitor analysis. The opinions or sentiments mined from such data helps understand the current state of the user. It does not directly provide intuitive insights on what actions to be taken to benefit the end user or business. Actionable Pattern Mining method provides suggestions or actionable recommendations on what changes or actions need to be taken in order to benefit the end user. In this paper we propose automatic classification of emotions in twitter data using Recurrent Neural Network - Gated Recurrent Unit. We achieve training accuracy of 87.58% and validation accuracy of 86.16%. Also, we extract action rules with respect to the user emotion, that help provide actionable suggestion on how to enhance user emotion from negative or neutral to more positive emotion ($Anticipation \rightarrow Trust$), and ($Sadness \rightarrow Joy$).

*Keywords*—Emotion Mining, Twitter, Recurrent Neural Network, Gated Recurrent Unit, Actionable Pattern Mining.

## I. Introduction

Emotion is an instinctive or intutive feeling. There are 6 different types of emotions such as happy, sad, anger, fear, disgust and surprise [1]. These emotions can be manual arousal or automatic (body's reaction) arousal. The automatic emotions are called as Fight-Flight-Freeze response from the brain. This kind of emotions are controlled by the 'amygdala' region of the brain and is hard for the humans to control. Manual emotions like happy, surprise, sad, etc. are reflected by the current mood and situation.

In this digital era, people tend to express their emotions/opinions, from politics to products review in social media because of its popularity and accessibility [2]. Social media is one of the most powerful platforms which is not only used for source of news but also as a place where users tend to provide feedbacks about their experience with the product or the company. The most popular social media platforms like Twitter, Instagram, Facebook, Yelp, LinkedIn etc. are used by variety of people to share a message, react to a review, post their opinion, etc. In addition to the texts provided by the user, there are also several reactions to convey the user's opinion. They are called emojis. Emojis are nothing but visual complement of the text [3].

Emotion mining techniques are used by many leading technology companies like Google, Microsoft, IBM, SAP, etc. [4] to build their own in-house industry activities like product improvement, and competitor analysis. For example, in Amazon, the customer reviews are scaled 1 to 5 with positive, neutral and negative comments. This is used to improve product analysis and overall customer satisfaction.

The process of analyzing the customer's opinion or feeling is done through determining the type of emotion in the words of their responses and this is an important task in Natural Language Processing (NLP). We focus on 8 different types of emotions such as Joy, Fear, Sadness, Anticipation, Surprise, Anger, Disgust and Trust.

Neural Networks (NN) which mimic the human brain, is utilized for better understanding of human emotions. Among other Neural Networks, Recurrent Neural Network (RNN) is one network which has internal memory, for processing the input. This feature of RNN is used for processing sequential information. The issue with the standard RNN is that, it is limited to look back only few steps in the internal memory, since it will have vanishing gradient or exploding gradient problem [5]. To overcome this, techniques like Long Short-term memory (LSTM), Gated Recurrent Unit (GRU) were introduced [6]. LSTM consists of input gate, output gate and forget gate. In addition to this, it contains a memory cell. LSTM is applied in Google Gmail Smart Compose [7], hand writing recognition, etc. GRU is also similar to LSTM [8] except that, it combines the input and the forget gate into a single gate called Update gate. In this paper we use Recurrent Neural Network - Gated Recurrent Unit for automatic classification of emotion from Twitter data.

Knowledge Discovery is the process of extracting interesting patterns and applying such patterns to specific areas of interest. To find such interesting patterns from data there are wide range of techniques. Actionable Pattern Discovery is one of the knowledge discovery approach. Action Rule mining helps fins actionable suggestions for recommendations. It helps identify appropriate interesting measure from the data. Emotion mining from the data provide the current state or feeling of the user. In order for the applications to provide valuable insights on the mined data for applications such as public policy making - by analysing the social platform of a particular community or area, it is required to further process the data to discover interesting measures. In this work we propose the use of Action Rule mining for discovering valuable suggestions or recommendation on how to enhance user emotion.

## II. Related Works

Sentiment analysis is based on three approaches such as Document level, Sentence level and aspect-based sentiment analysis [9]. Document level analysis is based on classifying the opinionated document either positive or negative. Sentence level analysis is based on classifying each individual sentence in the document by either positive, negative or neutral. Aspect-based analysis focuses on aspect extraction, entity extraction and aspect sentiment classification.

Common approach to analyze emotions are through bag-of-words or bag-of-n-grams, where a document will be analyzed by certain combinations of terms and ignores grammar. The limitation is complexity in the elaborate vocabulary design, sparsity, discarding word/contextual meanings [10](‘this is fake’ vs ‘is this fake’).

In Machine Learning, there are several algorithms used for sentiment analysis like Support Vector Machine (SVM) [11], Random Forest, Naive Bayes, K-Nearest neighbor [12], Convolutional Neural Network (CNN), Recurrent Neural Network (RNN). Especially Neural Networks (NN) is used to analyze large datasets of emotions. CNN have been used for the sentiment analysis [9], where human emotions are discovered from image filtering and also used for short sentence sentiment classification [13].

RNN is a network which has internal memory, for processing the input. This feature of RNN is used for processing sequential information. The issue with the standard RNN is that, it is limited to look back only few steps in the internal memory, since it will have vanishing gradient or exploding gradient problem [5].

Hierarchical Bi-directional Recurrent Neural Network (HBRNN) is used for providing output at the end of the sentence, instead of output for each word [14]. To overcome the vanishing gradient problem, algorithms like Long Short-term memory (LSTM), Gated Recurrent Unit (GRU) were used. Both these recurrent units have gated structure and GRU is the simplified version of LSTM without output gate [15]. Like LSTM, GRU doesn’t make use of memory unit [16] for the information storage, instead, it uses Update and Reset gates. GRU overperformed LSTM [14] on banks and telecommunication domains.

GRU model outperformed other models like tf-idf, word2vec and LSTM in sentiment analysis on IMDB dataset [17] with 97 % AUC score. Also in [18], GRU model have been used to analyze airline sentiment in twitter data where they acquired 83% accuracy in 2-layers GRU and 88% accuracy in 3-layers GRU. In [19], GRU algorithm is used on Hepsiburada Turkish e-commercial dataset and have got 95% accuracy using sigmoid function where the output is classified as positive and negative reviews.

In [20] with size of 500 tweets, they achieved an accuracy of 94.19 % on twitter dataset with the application of supervised learning algorithms such as Naïve Bayes (NB) and Random Forest (RF). In another similar work on Sentiment analysis of Twitter data especially McDonalds and KFC tweets data [21] using various supervised learning algorithms such as NB, RF, SVM, Decision Tree, Bagging and Maximum entropy (ME) proved that ME was the best model for both KFC and McDonalds data with 78 % and 74 % accuracy after 4-fold cross validation respectively.

Action Rules Mining is a method to discover Actionable Patterns from large datasets. Action Rules are rules that describe a possible transition of data from one state to another. In Data Mining literature, we see two pre-dominant frameworks for Action Rule generation: Rule based (loosely coupled) and Object based (tightly coupled) methods.

Author dardzinska [22], summarize the frameworks for generating Action Rules from [23] as follows: loosely coupled and tightly coupled. The loosely coupled framework is often called rule-based. It is based on pairing certain classification rules which have to be discovered first by using for instance algorithms such as LERS [24] or ERID [25]. The tightly coupled framework is often called object-based and it assumes that Action Rules are discovered directly from a database [26], [27]. Classical methods for discovering them follow algorithms either based on frequent sets (called action sets) and association rules mining [28] or they use algorithms such as LERS or ERID with atomic action sets used as their starting step. Action Rules are one way to mine Actionable knowledge from large dataset.

## III. Methodology

### A. Data

We use Twitter dataset which is obtained using Twitter API. The dataset contains 174,090 tweets and 9 columns namely Anger Score, Trust Score, Fear Score, Sadness Score, User Language, Is Possibly sensitive, Media Entities, Tweet Source and Tweet Emotion. We focus on eight different types of emotions such as Joy, Fear, Sadness, Anticipation, Surprise, Anger, Disgust and Trust as the output of the trained model. Figure 1 shows the 8 different emotions types in the twitter data along with its distribution across the data.

### B. Preprocessing

Preprocessing the data is one of the important tasks in data analysis. Earlier, the twitter tweet length is limited to 140 characters but suddenly twitter increased its tweet length to 280 characters [29]. In Figure 2 we can see that, most of the tweets fall between 130 to 140 characters length and only few are in the range of 140 to 160 characters length in our dataset.

During the data cleaning process, we removed redundant characters such as @mentions, hashtags and numbers in the tweets [30]–[32] to reduce the noise in the data. While dealing with text mining, it is important to split the text into words. Words are called as tokens and this process is called as

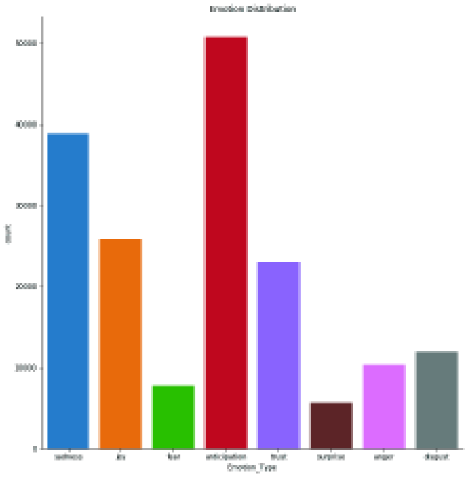


Fig. 1. Data Distribution Based on Emotion Class

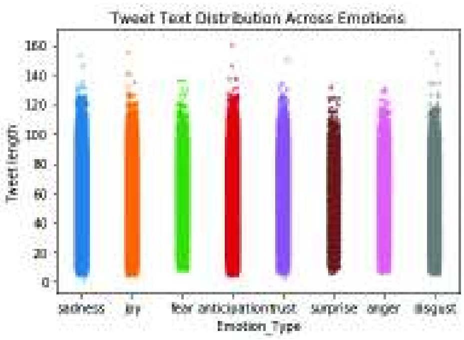



Fig. 2. Tweet Text Distribution accross Emotions

tokenization. This is achieved by using one of the keras library function [33] and the texts is case folded (lowercase).

We used 'Stop Words' for filtering and removing the most common English words like 'the', 'a', 'an', 'in', etc. from the dictionary. 'Lemmatization' analyzes a word morphologically and removes its inflectional ending, producing its base form or lemma as it is found in a dictionary [32]. For lemmatization we used 'Word Net Lemmatizer' from Natural Language Toolkit (NLTK). The above processes like tokenization, filtering, lemmatization is called as Feature Engineering. Feature engineering is the process of preprocessing the collected data into more meaningful data and features to gather more insights from the data.

Term Frequency-Inverse Document Frequency (TF-IDF) has been used to weigh the word in the content and assign score or value to the word based on the number of times it appears in the document.

### C. Recurrent Neural Network

Neural Networks are the resemblance of our human brain and the models are trained in the way how our brain works and behaves. In particular, RNN have internal memory which is used for processing sequential information. One of the gated Recurrent Neural Network approach called Gated Recurrent Unit is implemented to train the model based on the twitter data and get better accuracy results.

As stated in [34] there are 3 main aspects of neural network which plays important role in the network performance: the network architecture and the pattern of connections between units, the learning algorithm, and the activation functions used in the network.

The RNN-GRU model is implemented using Keras libraries in python [35], [36]. Activation layer decide which neurons will push forward the values into the next layer based on the input parameters weight and bias, then it will activate the neurons and help the network to use the important information and suppress the irrelevant data samples. There are many different types of activation functions such as SoftMax, Sigmoid, tanh, ReLU, etc. SoftMax layer is used as the activation layer in our model at the end of the dense layer since we have multi class output variables [17] i.e., it produces 8 different emotion outputs as stated in sections I and III.

We use Sequential API from Keras library to build the GRU model layer by layer. Layers like Embedding, Spatial Dropout 1D, Dropout, Dense and Activation are used when building the neural network. Embedding layer is used to encode each word into a unique integer.

Dropout layer is applied to prevent the model from overfitting. It is one of the regularization techniques for neural networks. The term "dropout" refers to dropping out the units (hidden or visible) in the neural network. A hyperparameter is introduced to specify the threshold or probability at which the outputs of the layer should drop out.

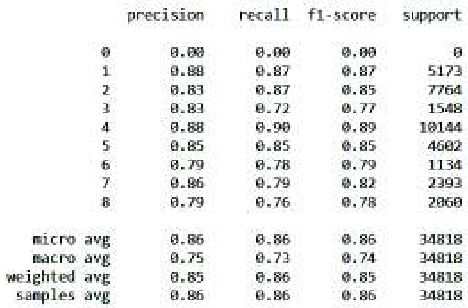
Adam optimizer algorithm is an adaptive learning algorithm which is used for efficient stochastic optimization [37]. This method computes individual adaptive learning rates for different parameters from estimates of first and second moments of the gradients. We use Adam optimizer because it works well with sparse gradients [37], doesn't require a stationary objective. Also Adam optimizer is a better choice for sparse data [38] and complex neural network.

### D. Action Rule Mining

In this paper, we use the approach of hydrid Action Rule [39], combining the rule based and object based approach of Action Rule mining to reduce the overhead of Action Rule iterative procedure. Association Action Rule extraction is an exhaustive Apriori based method which extracts complete set of Action rules by taking all possible combinations of the action terms. It is an iterative procedure and does not scale very well in case of dense and high dimensional dataset. This method create subtables by using the Action Rule Schemas in a highly dense data as explained above. Then performs Association Action Rule extraction algorithm on each of the subtables in parallel which allows the algorithm to complete and generate rules in a much faster time compared to the existing algorithms and systems.

## IV. Experiments and Results

The dataset contains 174,090 tweets. The data is split in the ratio of 80:20 for training and testing the model respectively. In Embedding layer, we used 100000 as the number of distinct words in the dataset and 100 as the size of the embedding



| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.00 | 0.00 | 0.00 | 0 |
| 1 | 0.88 | 0.87 | 0.87 | 5173 |
| 2 | 0.83 | 0.87 | 0.85 | 7764 |
| 3 | 0.83 | 0.72 | 0.77 | 1548 |
| 4 | 0.88 | 0.90 | 0.89 | 10144 |
| 5 | 0.85 | 0.85 | 0.85 | 4602 |
| 6 | 0.79 | 0.78 | 0.79 | 1134 |
| 7 | 0.86 | 0.79 | 0.82 | 2393 |
| 8 | 0.79 | 0.76 | 0.78 | 2060 |
| micro avg | 0.86 | 0.86 | 0.86 | 34818 |
| macro avg | 0.75 | 0.73 | 0.74 | 34818 |
| weighted avg | 0.85 | 0.86 | 0.85 | 34818 |
| samples avg | 0.86 | 0.86 | 0.86 | 34818 |

Fig. 3. RNN Model - Precision, recall, f-Measure

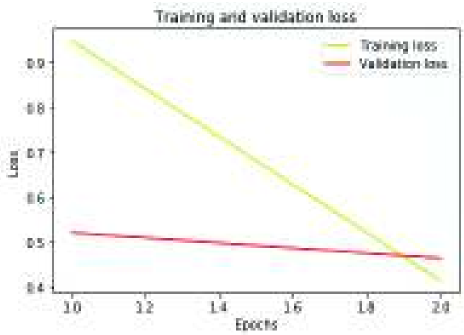



Fig. 4. RNN Model - Training and Validation - Loss Evaluation

vectors. Dropout layer which is applied to prevent overfitting of the model is given the hyperparameter value of 0.5 for dropping out the units from one layer to another. SoftMax layer have been used at the end of the dense layer since we have multi class output variables [17] i.e., it produces 8 different emotion outputs as stated in sections I and III.

We tried different epoch values such as 2,4,8 and batch size as 128, 256 and 2048 along with different layers of GRU such as single layer GRU and Bidirectional GRU. If epoch is set to 2 then the model will run through the entire dataset twice splitting it accordingly to the batch size. If the batch size is set to 128 then, first 128 samples are split and train the network, then take the next 128 samples and train the network again. This procedure continues until we propagated all samples through the neural network model.

Classification report is used to determine the quality of predictions of the trained model. How many predictions are True and how many are False? More specifically, True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN) are used to predict the metrics of a classification report using accuracy, precision, recall and f1 score as shown in Figure 3.

We have got the training and validation result as 87.58% and 86.16% respectively. From Figure 4 we can clearly see that the training and validation loss lowers as the epoch continues.

Confusion matrix is used for summarizing the performance of trained model. The number of correct and incorrect predictions are summarized with count values and broken down by each class. Figure 5 shows the count of the actual and predicted values where the model made errors in making predictions.

The GRU model is trained and tested with various changes

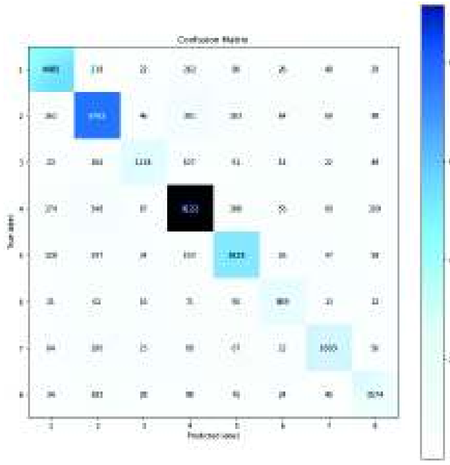



Fig. 5. RNN Model - Confusion Matrix

TABLE I
CLASSIFICATION ACCURACY

| Epoch | Batch Size | GRU Model | Accuracy |
|---|---|---|---|
| 2 | 128 | 1 Layer | 0.8616 |
| 2 | 128 | Bi-Directional | 0.8585 |
| 2 | 256 | 1 Layer | 0.8573 |
| 2 | 256 | Bi-Directional | 0.8547 |
| 4 | 256 | 1 Layer | 0.8546 |
| 4 | 256 | Bi-Directional | 0.8488 |
| 8 | 128 | 1 Layer | 0.8437 |
| 8 | 2048 | Bi-Directional | 0.8463 |

TABLE II
PARAMETERS USED FOR ACTION RULE EXPERIMENTS

| Property | Values |
|---|---|
| # of Attributes | 9 Attributes |
| Decision Attribute | Sadness → Joy<br>Anticipation → Trust |
| # of Instances | 174090 |

in the hyperparameter values. Table I shows the classification accuracy of the trained GRU model for different parameter values of Epoch & Batch size and different layers of GRU such as single layer GRU and Bidirectional GRU.

### A. *Action Rule Mining - Results*

The decision problem here is to suggest possible recommendations to the user's - in applications such as public policy making, Education, Customer Care Service, and other promising applications. In this work, we use Emotion as the decision attribute and collect Action Rules that help identify changes that are required for the emotion to be more positive. For instance, to change the emotions from "Sadness" to "Joy" and "Anticipation" to "Trust".Parameters used for the experiments is described in Table II

The experiments are performed in both a single node machine with 16GB Random Access Memory (RAM), 64-Bit operating system; and on Amazon Elastic Map Reduce (EMR) cluster with m4-large instance (1 - Master node and 2

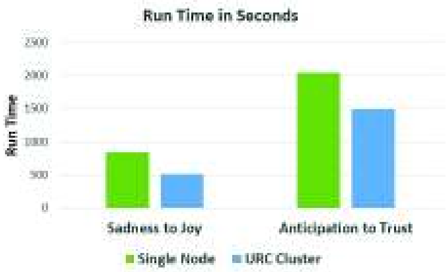



Fig. 6. Action Rule Mining - Algorithm - Execution Time

TABLE III
SAMPLE ACTION RULES

| **Emotion Change - Twitter Data** |
|---|
| 1) $AR_{C1}$ : $(TrustScore, VeryLow \rightarrow VeryHigh) \wedge (MediaEntities, 0) \wedge (TweetSource, 2 \rightarrow 4) \Longrightarrow (Emotion, Anticipation \rightarrow Trust)[Support : 141, Confidence : 54.7\%]$ |
| 2) $AR_{C2}$ : $(TrustScore, VeryLow \rightarrow VeryHigh) \wedge (FearScore = VeryLow) \wedge (TweetSource, 2 \rightarrow 4) \Longrightarrow (Emotion, Anticipation \rightarrow Trust)[Support : 141, Confidence : 58.59\%]$ |
| 3) $AR_{C3}$ : $(SadnessScore, Medium \rightarrow VeryLow) \wedge (MediaEntities, 0) \Longrightarrow (Emotion, Sadness \rightarrow Joy)[Support : 648, Confidence : 95\%]$ |

- Worker nodes).Figure 6. shows the run time in seconds for both Single node machine and university research cluster.

Table III shows sample Action Rules extracted for the Twitter dataset. Let us consider the action rule $AR_{C3}$ from “Table. III”, this rule suggest possible changes to achieve a desirable emotional state of ‘joy’. If user tends to reduce use of negative words as denoted by $(SadnessScore, Medium \rightarrow VeryLow)$ and if there is media (photo or video) associated with the tweet, then it is possible to change the emotion from ‘sadness’ to ‘joy’. This method could be useful in applications where emotions or feeling of people in a particular city or county could be analysed in order to understand the life satisfaction.

## V. CONCLUSION

Sentiment analysis is widely used in social discussions like politics, technical advantages like customer preferences on products/brands and marketing strategies. With no doubt, this is a valuable asset for many industry firms. To the best of our knowledge, as stated in the related works section, the previous Sentiment analysis work on Twitter data focuses on specific community or product. Our proposed GRU model shows better accuracy in predicting 8 different emotion type in the tweets with an accuracy of 86.16%. This generalized GRU model could also be used to visualize the results of particular product or brand or company too. The action rules obtained suggest in brief that decreasing the ‘SadnessScore’ and increasing the ‘TrustScore’ can help improve joy and trust emotions in tweets.

## REFERENCES

[1] P. Ekman, “An argument for basic emotions,” *Cognition & emotion*, vol. 6, no. 3-4, pp. 169–200, 1992.

[2] C. Puschmann and A. Powell, “Turning words into consumer preferences: How sentiment analysis is framed in research and the news media,” *Social Media+ Society*, vol. 4, no. 3, p. 2056305118797724, 2018.

[3] F. Barbieri, M. Ballesteros, and H. Saggion, “Are emojis predictable?” *arXiv preprint arXiv:1702.07285*, 2017.

[4] F. A. Pozzi, E. Fersini, E. Messina, and B. Liu, *Sentiment analysis in social networks*. Morgan Kaufmann, 2016.

[5] S. Minaee, E. Azimi, and A. Abdolrashidi, “Deep-sentiment: Sentiment analysis using ensemble of cnn and bi-lstm models,” *arXiv preprint arXiv:1904.04206*, 2019.

[6] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[7] M. X. Chen, B. N. Lee, G. Bansal, Y. Cao, S. Zhang, J. Lu, J. Tsay, Y. Wang, A. M. Dai, Z. Chen *et al.*, “Gmail smart compose: Real-time assisted writing,” in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 2287–2295.

[8] L. Wu, C. Kong, X. Hao, and W. Chen, “A short-term load forecasting method based on gru-cnn hybrid neural network model,” *Mathematical Problems in Engineering*, vol. 2020, 2020.

[9] F. Siraj, N. Yusoff, and L. C. Kee, “Emotion classification using neural network,” in *2006 International Conference on Computing & Informatics*. IEEE, 2006, pp. 1–7.

[10] L. Guo, Y. Han, H. Jiang, X. Yang, X. Wang, and X. Liu, “Learning to make document context-aware recommendation with joint convolutional matrix factorization,” *Complexity*, vol. 2020, 2020.

[11] R. P. Mehta, M. A. Sanghvi, D. K. Shah, and A. Singh, “Sentiment analysis of tweets using supervised learning algorithms,” in *First International Conference on Sustainable Technologies for Computational Intelligence*. Springer, 2020, pp. 323–338.

[12] J. Ranganathan and A. Tzacheva, “Emotion mining in social media data,” *Procedia Computer Science*, vol. 159, pp. 58–66, 2019.

[13] F. Krebs, B. Lubascher, T. Moers, P. Schaap, and G. Spanakis, “Social emotion mining techniques for facebook posts reaction prediction,” *arXiv preprint arXiv:1712.03249*, 2017.

[14] J. Trofimovich, “Comparison of neural network architectures for sentiment analysis of russian tweets,” in *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference Dialogue*, 2016, pp. 50–59.

[15] J. Nowak, A. Taspinar, and R. Scherer, “Lstm recurrent neural networks for short text and sentiment classification,” in *International Conference on Artificial Intelligence and Soft Computing*. Springer, 2017, pp. 553–562.

[16] S. Sachin, A. Tripathi, N. Mahajan, S. Aggarwal, and P. Nagrath, “Sentiment analysis using gated recurrent neural networks,” *SN Computer Science*, vol. 1, no. 2, pp. 1–13, 2020.

[17] S. Biswas, E. Chadda, and F. Ahmad, “Sentiment analysis with gated recurrent units,” *Department of Computer Engineering. Annual Report Jamia Millia Islamia New Delhi, India*, 2015.

[18] Y. Tang and J. Liu, “Gated recurrent units for airline sentiment analysis of twitter data.”

[19] Y. Santur, “Sentiment analysis based on gated recurrent unit,” in *2019 International Artificial Intelligence and Data Processing Symposium (IDAP)*. IEEE, 2019, pp. 1–5.

[20] A. Srivastava, V. Singh, and G. S. Drall, “Sentiment analysis of twitter data: A hybrid approach,” *International Journal of Healthcare Information Systems and Informatics (IJHISI)*, vol. 14, no. 2, pp. 1–16, 2019.

[21] S. A. El Rahman, F. A. AlOtaibi, and W. A. AlShehri, “Sentiment analysis of twitter data,” in *2019 International Conference on Computer and Information Sciences (ICCIS)*. IEEE, 2019, pp. 1–4.

[22] A. Dardzinska, *Action rules mining*. Springer, 2012, vol. 468.

[23] H. Kaur, “Actionable rules: Issues and new directions.” in *in Transactions on Engineering, Computing and Technology, World Informatica Society*. Citeseer, 2005, pp. 61–64.

[24] J. W. Grzymala-Busse, “A new version of the rule induction system lers,” *Fundamenta Informaticae*, vol. 31, no. 1, pp. 27–39, 1997.

[25] A. Dardzinska and Z. W. Ras, “Extracting rules from incomplete decision systems: System erid,” in *Foundations and Novel Approaches in Data Mining, (Eds. T.Y. Lin, S. Ohsuga, C.J. Liau, X. Hu), Studies in Computational Intelligence*. Springer, 2005, vol. 9, pp. 143–154.

[26] Z. He, X. Xu, S. Deng, and R. Ma, "Mining action rules from scratch," *Expert Systems with Applications*, vol. 29, no. 3, pp. 691–699, 2005.

[27] S. Im and Z. W. Ras, "Action rule extraction from a decision table: Ared," in *Foundations of Intelligent Systems, Proceedings of ISMIS'08, A. An et al. (Eds.), Toronto, Canada, LNAI*, vol. 4994. Springer, 2008, pp. 160–168.

[28] R. Agrawal, R. Srikant *et al.*, "Fast algorithms for mining association rules," in *Proc. 20th int. conf. very large data bases, VLDB*, vol. 1215, 1994, pp. 487–499.

[29] K. Gligorić, A. Anderson, and R. West, "How constraints affect content: The case of twitter's switch from 140 to 280 characters," in *Twelfth International AAAI Conference on Web and Social Media*, 2018.

[30] A. S. M. Alharbi and E. de Doncker, "Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information," *Cognitive Systems Research*, vol. 54, pp. 50–61, 2019.

[31] E. UZUN, T. YERLİKAYA, and O. KIRAT, "Comparison of python libraries used for web data extraction," *FUNDAMENTAL SCIENCES AND APPLICATIONS*, p. 87, 2018.

[32] S. Symeonidis, D. Effrosynidis, and A. Arampatzis, "A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis," *Expert Systems with Applications*, vol. 110, pp. 298–310, 2018.

[33] J. Brownlee, "How to prepare text data for machine learning with scikit-learn," 2017.

[34] A. Farzad, H. Mashayekhi, and H. Hassanpour, "A comparative performance analysis of different activation functions in lstm networks for classification," *Neural Computing and Applications*, vol. 31, no. 7, pp. 2507–2521, 2019.

[35] B. Reddy, "Twitter sentiment analysis," 2019. [Online]. Available: "https://github.com/bhargavflash/Twitter_Sentiment_Analysis"

[36] N. Rahman, "Keras-models:lstm,cnn,gru,bidirectional,glove," 2018. [Online]. Available: "https://www.kaggle.com/nafisur/keras-models-lstm-cnn-gru-bidirectional-glove"

[37] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[38] Y. Zhang, B. Zhu, Q. Ma, and H. Wang, "Effects of gradient optimizer on model pruning," in *IOP Conference Series: Materials Science and Engineering*, vol. 711, no. 1. IOP Publishing, 2020, p. 012095.

[39] J. Ranganathan, S. Sharma, and A. A. Tzacheva, "Hybrid scalable action rule: Rule based and object based," in *Proceedings of the 2020 the 4th International Conference on Compute and Data Analysis*, 2020, pp. 107–112.