Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
May 2024 5(5), 658–665

UTILIZING MACHINE LEARNING TECHNIQUES TO IDENTIFY EMOTIONAL CORRELATES IN PHRASE ARTICULATION

Devendra Singh Rathore[1], Dr. Pratima Gautam[2]

[1]Rabindranath Tagore University, Raisen, M.P. India
[2]Rabindranath Tagore University, Raisen, M.P. India

Corresponding Author
Devendra Singh Rathore, devendrarathore2007@yahoo.com

DOI
10.29121/shodhkosh.v5.i5.2024.2107

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

Emotions are the strands that crisscross human communication, influencing our views, responses, and interactions with the environment. Comprehending and interpreting the emotional connotations included in textual information have become essential tasks in the field of natural language processing. This study tries to explore the complex field of sentiment analysis by closely examining the feelings that are ingrained in language construction. It is impossible to overestimate the importance of emotional analysis in textual communication. Key words are added than just information carriers; they also have layers of emotional meaning that have a considerable influence on how they are understood and received. Regardless of Emotions, such as happiness, sadness, rage, or ambivalence, affect how we interpret and react to the values that are told via language. We breaks down the emotional connections found in phrases using a framework for systematic analysis using sophisticated NLP methods and sentiment analysis algorithms, we set out to interpret the complex emotional aspects included in written communication. By use of lexical feature extraction, syntactic structure extraction, and semantic context extraction, our goal is to reveal the many aspects of affective expression that are contained in sentences. As part of the study process, a variety of textual datasets covering a range of genres, styles, and situations are collected. Our empirical research is based on these datasets, which allow us to investigate the subtleties of emotional expression in many language areas. Through painstaking annotation and classification of phrase emotional content, we aim to build an all-encompassing knowledge of the emotional terrain present in textual communication.

Keywords: Speech recognition, emotion analysis, social media, NLP, sentiment analysis.

1. INTRODUCTION

Natural language processing (NLP) has two components, human language creation and human language interpretation [1]. However, genuine language holds ambiguity, the latter is more challenging because genuine language has ambiguity, the earlier is more difficult. NLP is used in speech recognition, structure conversion, examine answering, document summarization, speech synthesis, and other applications. The two most important aspects of natural language processing are emotion recognition and sentiment analysis. These two names are not equal while they exist occasionally used synonymously [2]. Data may be evaluated to figure out condition it is favorable, negative, or neutral using emotion analysis. On the other hand, emotion detection is the process of recognizing various kinds of human emotions, including anger, happiness, or depression. There are occasions when the terms "feeling finding," "affecting determining," "emotion analysis," and "emotion classification" are used simultaneously. The internet has become a fashionable way for people to express their emotions because Internet connectivity has gotten easier. Persons freely communicate their thoughts, disagreements, and ideas on a broad spectrum of subjects on social networks [3].

[4] Furthermore, many people provide reviews and comments on a range of goods and services on different e-commerce websites. Vendors and service providers are encouraged to improve their present systems, commodities, or services by the ratings and reviews left by users on various platforms. These days, practically every business or sector is going through a digital transformation, producing enormous volumes of increased organized and unstructured data. Businesses face a huge challenge in turning unstructured data into visions that are valuable for making decisions. For case, in the business world, vendors use social media sites like Facebook, Instagram, YouTube, Twitter, and LinkedIn to effectively spread information about their products and ask for customer feedback. Active user feedback is helpful for company marketers not just to gauge client happiness and check the rivalry as well as for customers who wish to research a good or service already making a purchase. The examination of sentiment helps companies better grasp the viewpoints of individuals who are buying their outcomes so they may make the required improvements to their goods or services [5, 6].

2. OBJECTIVE & PROBLEMS

The goal of the research paper is to systematically investigate the emotional associations embedded within the articulation of sentences. The primary goals include:

- Understanding Emotional Expression: To comprehensively recognize how reactions are communicated within textual content and the diverse linguistic mechanisms employed to convey different emotional states.
- Developing Analysis Frameworks: To develop systematic analysis frameworks and methodologies for dissecting and categorizing emotional associations within sentences, encompassing lexical, syntactic, and semantic dimensions.
- Exploring Sentiment Analysis Techniques: To explore and evaluate many natural speech processing methods and sentiment analysis algorithms for extracting and interpreting emotional content from textual data.
- Uncovering Emotional Nuances: To unravel the nuanced emotional nuances, present within sentences across different linguistic domains, genres, and contexts, including positive, negative, and neutral sentiments.

Overall, the goal is to deepen our comprehension of the intricate relationship between language and emotion, paving the way for advancements in sentiment analysis research and applications across diverse domains.

3. DATA COLLECTION AND FEATURE SELECTION

Web scraping, social media, news channels, e-commerce websites, forums, weblogs, and other websites can all be used to gather data from the internet. The first step in the sentiment analysis process is data collection. Video, audio, location, and other forms of data can be merged with text data, depending on the task sentiment analysis results. A few crucial places to get data are [7, 8].

- Social media: Communication obtained through social broadcasting platforms is referred to as social data. It illustrates how users' access, post, and exchange information about the product. Social media is a dynamic data source used in academic research on individuals, groups, and behavioral behavior. It describes web-based or mobile applications on the Internet that let users exchange, produce, and consume user-generated content.
- Forums**:** Message boards allow users to text each other for help, debate a range of subjects, and share thoughts and opinions. Because user-generated content on forums is dynamic, it's a fascinating source for sentiment research. Additionally, by using forums as a source, researchers may do sentiment analysis on a particular subject.
- Weblog: A brief weblog is made up of paragraphs that provide information, links, personal journal entries, and points of view. Collectively referred to as postings, they are organized chronologically in the manner of a research paper, with the most recent item appearing first. A useful tool for conducting sentiment analysis on a range of things is a blog.
- Electronic commerce websites: These are online platforms that allow people to rate and comment on specific companies or organizations. In this case, e-commerce sites with product reviews1 or expert review sites like

conducted a descriptive analysis of the several airline service classes [9] are examples of websites that do not explicitly evaluate products and hold millions of reviews.

4. FEATURE SELECTION APPROACH

The characteristics of the data are found by evaluating the choice of features approach. A trait may be redundant, important, or inconsequential. Quaternion overall groupings are exploited to group statistical strategies for feature selection: filter, embedding, wrapper, and hybrid. The method for choosing characteristics that is most often employed is the filter approach. Based on the overall characteristics of the training data, it chooses features devoid of the use of any machine learning techniques. The feature is rated using a few statistical parameters, and the features that receive the highest rankings are after chosen. They work effectively with datasets that have an extensive number of attributes and are inexpensive to compute. The terms "basic filter methods" are referred to by the terms "data gain," "chi-square," "document time," and "shared knowledge." [10,11]

- Wrapper approach - Because this method depends on the output of the machine learning algorithm, it is proven on machine learning algorithms. Because of this reliance, approaches are often iterative and computationally costly, but they can find the ideal feature set for that modeling procedure. Feature subset creation (forward or backward selection) and different learning algorithms (NB or SVM) are examples of wrapper approaches.
- Embedded approach- This approach integrates the process of picking features with the modeling algorithm's running. It uses classification techniques with an integrated feature selection mechanism. It is therefore theoretically more efficient than the wrapping method. This method is technique precise, though.
- Hybrid approach- This tactic blends filters & wrappers techniques; mixed strategies typically combine many techniques to provide the best feature group. Hybrid strategies usually use many ways to reach outstanding efficiency and precision. [12,13]

5. RESULT ANALYSIS

We evaluated the GRU neural network model, BiLSTM, and LSTM. The tables that follow display the results obtained for these types of models.

Table 1.1: LSTM Emotion Evaluation Performance Metrics

Epoch	Emotion Evaluation Performance									
	Train dataset (40K)					Examine dataset (10K)				
	loss	accuracy	precision	recall	f1 score	loss	accuracy	precision	recall	f1 score
1	0.379	0.837	0.830	0.854	0.819	0.286	0.884	0.901	0.858	**0.863**
2	0.271	0.892	0.886	0.893	0.874	0.322	0.868	0.929	0.785	**0.832**
3	0.268	0.894	0.893	0.892	0.877	0.271	0.890	0.894	0.883	**0.874**
4	0.221	0.913	0.908	0.911	0.897	0.327	0.871	0.923	0.799	**0.840**
5	0.202	0.921	0.918	0.915	0.905	0.294	0.884	0.911	0.844	**0.860**
6	0.253	0.897	0.889	0.906	0.882	0.313	0.884	0.901	0.863	**0.865**
7	0.180	0.930	0.929	0.927	0.918	0.335	0.866	0.920	0.787	**0.831**
8	0.182	0.928	0.922	0.928	0.914	0.331	0.883	0.867	0.906	**0.872**
9	0.153	0.942	0.938	0.937	0.929	0.332	0.876	0.897	0.841	**0.851**
10	**0.138**	**0.949**	**0.947**	**0.946**	**0.939**	**0.401**	**0.846**	**0.796**	**0.931**	**0.842**

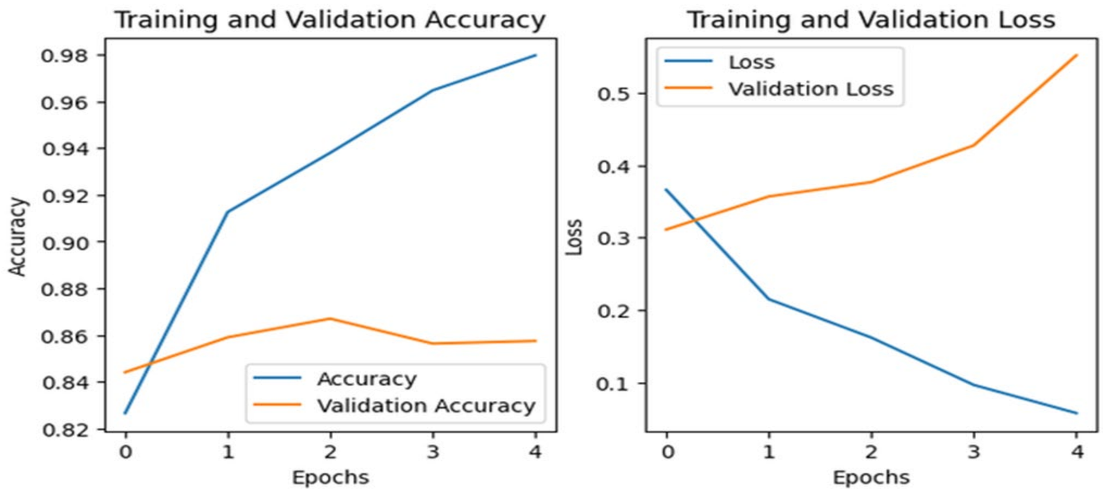

Figure 1.1: The plot of training and validation accuracy and loss

The efficacy characteristics of an LSTM model evaluated on a different dataset of 10,000 samples across ten epochs and trained on a dataset of 40,000 samples are exposed in Table-1.1. The guidance and review datasets F1 scores, recall, accuracy, precision, and loss are among the parameters that are assessed. Reduced loss values and rising standards for truth, accuracy, memory, and F1 score sign that the model's performance has improved gradually across the epochs. Remarkably, by the last epoch, the model performs well on the test dataset, with accuracy reaching 94.9% and F1 score reaching 93.9%. This implies that the model learns to generalize to previously unknown data well, proving its effectiveness in finding the basic trends in the dataset.

The model's ability to learn is shown by the training graph (shown in blue), and its ability to generalize effectively to new data is shown by the validation graph (shown in yellow). Accuracy should rise with time, and one would prefer to see a steady decrease in error loss. This implies that the model has not yet overfitted and is still learning from the training set. The model may have learned everything it could from the training set of data when the loss plateaus.

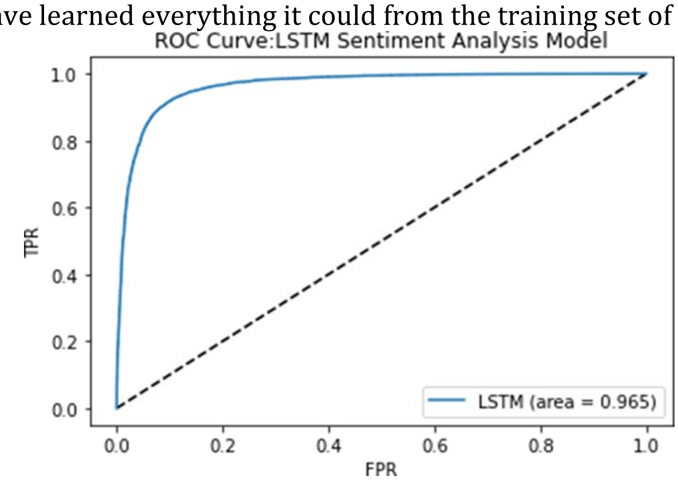

Figure 1.2: LSTM Model ROC Curve

The ROC Bend is exposed in this graph to illustrate the LSTM Sentiment Analysis Model. True positive rate (TPR) and false positive rate (FPR) are the two standard ROC curve axes. So, the "ideal" point, with a false positive rate of zero and a genuine positive rate of one, is in the upper left corner of the plot. This suggests that a greater zone under the arc (AUC) is generally preferable, although it is not particularly realistic. Since it is preferable to maximize the true positive rate while lowering the false positive rate, the "steepness" of ROC curves is also significant. When studying a classifier's output in binary classification, ROC curves are commonly employed. Expanding the ROC curve and ROC Due to multi-label classification, the output must be binarized. A ROC bend can be created for each label, or alternatively, its container be created by micro-averaging—that is, by treating each member of the label indicator matrix as a binary prediction.

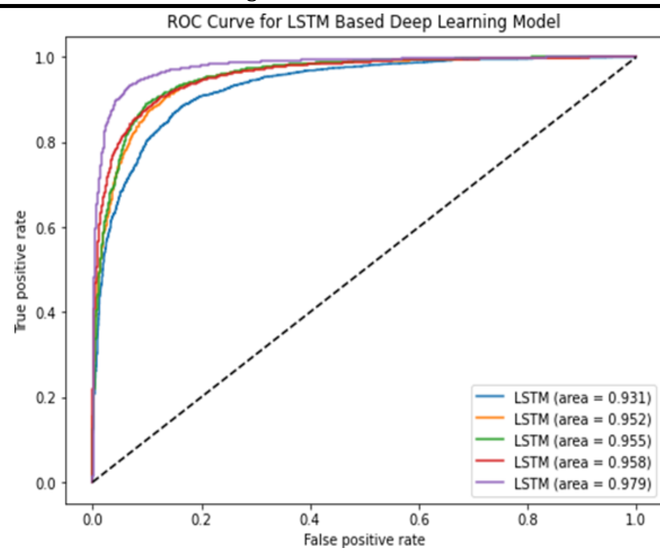

Figure 1.3: ROC Curve for LSTM Stratified Fold Cross Validation Model

A basic description of the ROC Curve for the LSTM Stratified 5-Fold Cross Confirmation Shape is provided by this graph. True positive rate (TPR) and false positive rate (FPR) are the two usual axes on ROC curves. So, the "ideal" point, with an FPR of zero and a TPR of one, is in the upper left corner of the plot. Although this is not particularly practical, it does write down that a region under the bend (AUC) that is greater is often preferable. Since it is best to maximize the TPR while reducing the FPR, the "steepness" of ROC curves is also significant. This chart, which was supplied applying 5-fold cross-validation, displays the ROC response of several datasets. Using all these curves, the possibility exists to figure out the variance of the curve and compute the mean AUC when the exercise set is separated into distinct subgroups. This illustrates, in general, how modifications to the working out data impact the classifier output and how distinct the splits produced by 5-fold cross-validation are fromoneanother

Table 1.2: BiLSTM Emotion Evaluation Performance Metrics

Epoch	Emotion Evaluation Performance									
	Dataset(40K)					Experiment dataset(10K)				
	loss	accuracy	precision	recall	f1 score	loss	accuracy	precision	recall	f1 score
1	0.452	0.784	0.761	0.789	0.704	0.411	0.800	0.901	0.767	0.741
2	0.330	0.865	0.869	0.845	0.844	0.344	0.816	0.833	0.564	0.832
3	0.271	0.894	0.892	0.889	0.833	0.233	0.876	0.844	0.654	0.864
4	0.244	0.904	0.904	0.888	0.826	0.222	0.800	0.802	0.654	0.876
5	0.223	0.912	0.911	0.901	0.867	0.232	0.898	0.987	0.345	0.875
6	0.206	0.920	0.916	0.912	0.990	0.250	0.838	0.897	0.675	0.867
7	0.184	0.930	0.926	0.922	0.911	0.211	0.887	0.090	0.564	0.853
8	0.175	0.934	0.931	0.922	0.910	0.327	0.823	0.987	0.987	0.874
9	0.163	0.940	0.937	0.933	0.921	0.334	0.887	0.125	0.456	0.875
10	0.148	0.944	0.942	0.931	0.930	0.401	0.844	0.786	0.988	0.861

The performance characteristics of a Bidirectional LSTM (BiLSTM) model that was trained on a 40,000-sample dataset and assessed on a 10,000-sample test dataset over ten epochs are presented in Table-1.2 For the training and test datasets, the metrics evaluated include F1 score, recall, accuracy, precision, and loss. Throughout the epochs, the BiLSTM model consistently performs better, showing declines in loss values and gains in accuracy, precision, recall, and F1 score. The model notably reaches a high accuracy of 94.4% on the work out dataset and 88.1% on the experiment dataset by the final epoch. Excellent values are also shown by precision, recall, and F1 score, which prove the model's ability to successfully name and generalize patterns found in the data. Although there was a little decline in the test dataset in the last epoch, the BiLSTM model continuously proves powerful performance in a few assessment measures.

Figure 1.4: GRU Model Loss Graph

This graph depicts the Training loss curve which blue in color versus Test which is orange in color for explaining the GRU model. Training halts due to early stopping after 10 epochs, writing down that the validation loss has reached its smallest evaluate at 10 epochs.

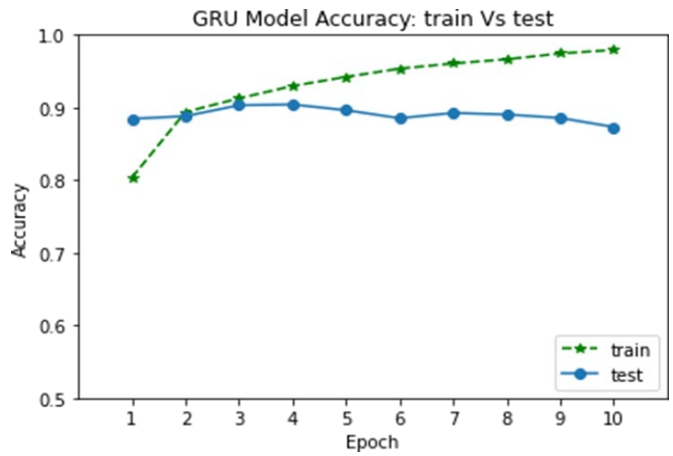

Figure 1.5: GRU Model Accuracy Graph

This graph depicts the Training loss curve which green in color versus Test which is blue in color for explaining the GRU model Accuracy . Training halts due to early stopping after 10 epochs, writing down that the Accuracy has reached its minimum value at 10 epochs.

Table 1.3: Comparative Machine Learning Model Accuracy

	Accuracy			
	Feature Size	**LSTM**	**BiLSTM**	**GRU**
Train accuracy	5000	94.92%	94.88%	96.17%
Test accuracy		85.33%	85.27%	86.47%
Train accuracy	10000	95.47%	94.65%	95.90%
Test accuracy		84.72%	83.94%	84.19%

The precision of three distinct regular neural system (RNN) architectures—LSTM, BiLSTM, and GRU—across a bounds of feature sizes is contrasted in table 1.3. There are two sizes of features that are assessed for the training and test datasets: 5000 and 10000. The GRU model shows the best performance for features with a size of 5000 on the training and test datasets, with 96.17% and 86.47%, respectively. The LSTM and BiLSTM models show a modest decrease in accuracy, trailing closely behind. The LSTM model obtains the best accuracy on the training dataset at 95.47% when the feature size is raised to 10,000. The GRU model comes in second place with 95.90% accuracy. All models' test accuracy does, however, decline when contrasted to the reduced feature size, with the GRU model continuing to have a little advantage over the others. In general, we spoke about a vector space model that picks up word representations that include emotion and semantic data. Examined is the impact of initializing embedding layer weights using pre-trained embedding models. There were two distinct word embeddings from movies that were employed.

The first movie pre-trained embedding model under examination is a publicly proposed CBOW 300 dimensions word representations, which were produced after training a sizable Modern Standard Movie dataset with over 5.8 billion words. Furthermore, another pre-trained Movie distributed word representation is used to investigate the effect of using a second pre-trained word embedding. The keyword implanting of 300 dimensions taught CBOW is employed. They trained a model on a sizable corpus of about 190 million words to produce their word embedding. The gathered dataset has been enhanced with dialectal language from several spoken motion pictures. The weights of these pre-trained speech paths are used to initialize the embedding layer weights.

The probabilistic basis of the system provides a theoretically justifiable method for information path induction in place of the many, widely used matrix factorization-based methods. Our approach is connected to Bayesian latent subject models and is defined as a log-bilinear model, building on recent success in employing similar techniques for language models. Rather than capturing latent themes, we try to capture word representations when we parametrize the topical part of our model. These tasks need relatively simple sentiment information, but the representation is quite versatile in this sense; it can be secondhand extensively in the expanding fields of emotion analysis and retrieval, as its canister be secondhand to define an extensive choice of annotations. Another method for standing for texts numerically is to convert each textual word to a vector. Word semantics privation is to be preserved during this transformation; if two languages have parallel implications, then their vectors ought to be similar as well (in the view of L2-distance). The word2vec job is notable for its independence from the primary goal, which is sentiment analysis, and for not requiring a labeled dataset. To train word vectors, we used all 75,000 reviews—25,000 labeled and 50,000 unlabeled training sets—as the corpus in this instance. Apart from the standard pre-processing processes on the raw reviews, we must separate paragraphs into sentences to train word vectors. This is because the word2vec method requires sentences as input (since words that are not in the same phrase as the current word are not a part of its pertinent background)

6. CONCLUSION

For our investigation, we used the IMDB movie review dataset. There are 50,000 reviews in the dataset overall, with 25,000 reviews in the instructing set and 25,000 reviews in the test set. 50% positive and 50% negative reviews make up the balanced dataset used in both the test and the train. We constructed a machine learning model for SA using word embedding and an LSTM (long, short term memory) network. An RNN neural network variation called LSTM was created especially for sequence input, such as word sequences seen in phrases. On the training set of data (22,563 correct and 2,437 incorrect), the model achieves 90.25% accuracy; on the test set, it achieves 82.06% accuracy. Following training, the model is applied to categorize a brand-new, never-before-seen short film review of, "The film was a huge waste of my time. The algorithm writes down the review will be negative since the likelihood of a positive value, which is 0.1368, is less than 0.5.

CONFLICT OF INTERESTS

None

ACKNOWLEDGMENTS

None

REFERENCES

Abid F, Alam M, Yasir M, Li C (2019), "Sentiment analysis through recurrent variants latterly on convolutional neural network of twitter", Future Gener Compute Syst.

Acheampong FA, Wenyu C, Nunoo-Mensah H (2020), "Text based emotion detection: advances, challenges, and opportunities", Eng Rep 2(7).

Acheampong FA, Nunoo-Mensah H, Chen W (2021), "Transformer models for text-based emotion detection: a review of BERT-based approaches", Arif Intel Rev.

Ahmad S, Asghar MZ, Alotaibi FM, Awan I (2019), "Detection and classification of social media-based extremist affiliations using sentiment analysis techniques", Hum Centric Compute Inf Sci.

Akhtar MS, Ekbal A, Cambria E (2020), "How intense are you? predicting intensities of emotions and sentiments using stacked ensemble", IEEE Compute Intel Mag.

Algumwood SO, Al Wakrah AA (2021), "Arabic sentiment analysis using recurrent neural networks", a review. Arif Intell Rev 55:707–748

Ali SM, Noori an Z, Bagheri E, Ding C, Al-Obeidat F (2020), "Topic and sentiment aware microblog summarization for twitter", J Intel Inf Syst.

Arora A, Chakraborty P, Bhatia M, Mittal P (2021), "Role of emotion in excessive use of twitter during COVID-19 imposed lockdown in India", J Techno Behave Sci.

Baashar Y, Alhussian H, Patel A, Alkawsi G, Alzahrani AI, Alfarraj O, Hayder G (2020), "Customer relationship management systems (CRMS) in the healthcare environment", Compute Stand Interfaces.

Bandara K, Bergmeir C, Smyl S (2020), "Forecasting across time series databases using recurrent neural networks on groups of similar series: a clustering approach", Expert System Appl.

Bartusiak R, Augustyniak L, Janowicz T, Kaseko P (2015), Sentiment analysis for polish using transfer learning approach. In: 2015 second european network intelligence conference. IEEE.

Basiri ME, Nemati S, Abdar M, Cambria E, Acharya UR (2021), "An attention-based bidirectional CNN-RNN deep model for sentiment analysis", Future Gener Compute Syst.

Behdenna S, Barigou F, Belalem G (2018) Document level sentiment analysis: a survey. EAI Endorsed Trans Context-aware Syst Appl 4(13).

Bergsma S, McNamee P, Bagdouri M, Fink C, Wilson T (2012) Language identification for creating language-specific twitter collections. In: Proceedings of the second workshop on language in social media.

Bhaskar J, Sruthi K, Nedungadi P (2015), "Hybrid approach for emotion classification of audio conversation based on text and speech mining", Procedia Compute Sci 46:635–643..