Semiconductor Optoelectronics, *Vol. 42 No. 1 (2023), 1324-1335*
https://bdtgd.cn/

Semiconductor Optoelectronics ISSN: 1001-5868

A SYSTEM BASED ON AI AND ML ENHANCED TO INVESTIGATE PHYSIOLOGICAL MARKERS FOR USER FORECASTING DECISION-MAKING

Sunil Chahal
Director- Technology Delivery, Concepts IT Inc.
messc0610@gmail.com

Abstract- This paper suggests a system for the automatic, low-cost, widespread, nonintrusive recognition of human needs that draws on a multi-layered psychological reference model and is built using various modules, such as those for data collecting, preprocessing, feature extraction, and contextualization. The reference model consists of various folding of SVM models to assess people's environments in relation to various aspects of their lives during any subjective event or toward emerging topics at anytime and anywhere using their publicly available social media content. It also measures people's satisfaction levels. Various textual, psychological, semantic, lexicon-based, and Twitter-specific elements are assessed for their predictive abilities. We test and compare the performance of different folding in order to offer benchmark results. Our findings attest to the usefulness of the created reference model. This work is used to identify citizen needs. Finally with the cross validation values found with the accuracy of 10 folds then the correct and more accurate model was predicted. Dataset used for this research is the question and answer discussed in the Council of States in each session which is available in kaggle for research purposes.
Keywords: Citizen needs, Social media analytics, Machine learning, Psychological based actions, New assessment.

I.INTRODUCTION

The upper house of the Indian Parliament is called the Council of States, or "Council of States." This "Question Hour" is of utmost priority in legislative proceedings since it allows members to learn information by asking questions about commonplace issues that the Ministries collectively answer to the legislature about. This essay primarily examines the Featured and Unfeatured categories. The questions with a star next to them are those to which verbal responses were provided on the House floor. Questions without stars will have written responses posted on the House table at the conclusion of the hour. In the political sphere, NLP is a formidable technology that can be utilised to unleash the potential of massive unstructured data. When people began to realise the value of a machine-based method for translating languages from one to another during the Second World War in the 1940s, the area of NLP was born [NLP - overview]. Recent developments in NLP techniques based on deep learning imitate the neuronal functions of the human brain to deliver improved performance and precise outcomes. AI is transforming politics more quickly than we can think, from political campaigns to social good [How AI will choose your next Prime Minister]. Numerous studies on sentiment analysis and aspect extraction have been conducted in recent years in the Indian political sphere, spanning from the study of parliamentary speeches to those of political tweets. Analysis of the Council of States Question Hour's elements and sentiments, however, has not received much attention. The dataset used in this study[18], covers more than 88000 questions and responses from the Council of States Question Hour from 2009 to 2017, organized withquery information, response period, government bureau, and more. Due to its fact that interpretation from database utilized for the current work which are uncategorized and there isn't a golden ordinary database by learning category which are exists in the area of the current work, this paper combines supervised and unsupervised learning techniques.

Political science learners still read a lot of old political writings every day, and text is likely the most significant artifact of political commentary [1]. This is crucial for understanding the script as information but as individual word individual texts given the growing digitization of political materials. Better analysis as well as outcomes will arise from this, with addition to a large reduction in the amount of unstructured data that must be stored. The key subjects of the question are explained by keyword extraction techniques [2][3][4], but they do not reveal the inquiry's larger concept. Let's take the query, "Does the Government plan to put doctors on board the Duronto Express?," as an example. keypoints retrieval approaches would find the keypoints such as"doctors; express," but further classification would require a rule-based technique. In the unsupervised phase of the study, opinion tags can be manually annotated; however, under the study researcher eliminates tagging emotions manually by using VADER (valence aware dictionary and sentiment analyzer) for generating sentiment labels.

II.RELATED WORK

By fostering the relationship between citizens and elected officials, a much more immediate and responsive democracy could de facto be brought back via the use of AI in politics [5]. Decision-making in democracy would be based more on facts and less on politics thanks to AI. Every democracy shares the fundamental ability of learning from the faults, educate, recover and as well as adjust new circumstances, which is precisely how the reinforcement training algorithm of supervised learning works.

In modern era, Natural Language Processing combined using DL has been excelled in the numerous fields. An introduction of deep learning methods is given in [6]. Using networks with numerous layers, deep learning can take use of the artificial neural networks' increased learning capacity. Neural networks are made up of a sizable number of information processing units organised logically in numerous layers and imitate the organisation of the biological brain. To carry out learning tasks, neurons modify the weights of their connections. Deep learning extracts and transforms features using a flood of non-linear processing units from various layers. Deep networks assimilate more complicated information received from lower levels, whereas lower layers are typically nearer for the continuous input data as well as learning from basic attributes. Word embeddings are a common input for deep learning models. Word embedding involves moving with the complex dimensionally spaces which contains only sparse vector to a minor dimensionally values representation space in order to depict phrases as vector of actual numbers.

In order to determine the feature connected to a statement, aspect-based emotion analysis takes into account the aspect words (or target words) in the paragraph. Since it is usually tricky to describe the semantic relationship between an element and the words around it, aspect-level emotion detection is regarded as a difficult process [7]. The model of neural networks must therefore accurately represent the link among aspect-words and context-words. [8] Investigatethe twitter data on the basis of aspect level emotion detection using richer automated features acquired through unsupervised learning. The analysis demonstrated the ability of sentiment words, numerous clustering techniques, and various pooled methods to extract rich information and enhance performances.

.

The applications of aspect-based sentiment classification are vast and may be astounding, particularly in the political sphere. The analysis of speeches from the United States presidential election of 2016 that were recorded [9] demonstrates the accuracy of hand tagging of features in an unmarked schema. Contributions from either the 2016 US presidential discussions make up the corpus used in this research. On eight pre-defined classes, a marked annotation schema with 3 levels, Aspect, and Sentiment—is applied. The nouns and adjectives only are included in the annotation schema utilised in this research.

Numerous political opinion analyses have been conducted in several nations using a range of linguistic corpora [10] [11]. The classification of the polarization of political posts on twitter during the Delhi-NCR 2014 polling as well as the state wide polling during 2016 has been tested using unlabelled dictionary based and the condition based, NB, Support Vector Machine, with the lexicon based techniques [12] [13]. The study by [14] on debates in the house of council categorises remarks into 4 groups: issue, blame, appreciation, and call for action. The study bases its assessment on a physicallyannotate databasewith one of an aforementioned stance. In this paper, VADER is used for sentiment classification, and Author’s opinion is used to locate keywords in the corpus. Ansari et alrecent.' Study [15] taking place on political related emotiondetection process on twitter tweets based on common talk about election of 2019 makes use of tf-idf for extracting features and goes on to use Regression Techniques and Random forest classification to derive the most prevalent words that are being used.

A study of the literature on sentiment analysis and opinion mining of social issues was done by the researchers in [16]. The data for the chosen publications came from social media websites. According to authors, combining various categorization methods can produce superior outcomes.In [17], 3 data mining methods tree (C4.5), Multi - layer perceptron, and Naive Bayes—were used to predict students' academic achievement. Those methods were used using student data that was gathered over the course of two quarters from 2 undergraduate programmes. Results revealed that Naive Bayes performed better than MLP and Decision tree, with an accuracy rate of 86%.

III.Methodology

The work [19] makes use of structured information that contains more than 88000 questions and responses from the Council of States Question Hour from 2009 to 2017 organised by query information, response date(period), assembly, and so on. Among the areas of both the Indian

Parliamentary dataset that has received the least amount of attention is sentiment analysis and aspect extraction. The need for a goldenordinary training database containing featureas well asemotionlabel in the study region was also addressed in the sections before. The study is based on verbally given qualitative data about the Council of States Question Hour.

Fig. 1: Methodology

(i) Data Collection

This information makes it easier to comprehend what was being debated in India's Council of States. Council of States debated more than 88000 questions and answers between 2009 and September 2017.

Variable specifications:

- id - Unique identifier
- answer_date - Answer date
- ministry - Ministry name
- question_type - Type of question (Starred or Unstarred)*
- question_no - Question number. (This question no. is unique per session)
- question_by - Minister who has raised the question.
- question_title - Discussion title
- question_description - Detailed question.
- answer - Detailed answer to the above question.
-

(ii)Dataset Preprocessing

Dataset preprocessing completes the following procedures and aids in removing null data.

A.Data gathering

This section gathered the publicly State of council and-answer dataset from Kaggle for research purposes.

B. Tokenization

Here, the information was gathered, or the entire sentence broken down word by word. Ex: Do you want to profit from the new project? Each word in the statement is separated by the tokenization process. For example: Do, you, want, to profit, from, the, new, project?

C. Filter the stopwords

Stopwords like "is,""was,""that,""has," and "been" are all eliminated throughout this step from the acquired data.

D.Negation Handling

Automatically detecting the range of a negative and reversing the polarity of opinionated words that are really impacted by a negative is known as "negation handling."

E.Stemming

The stemming procedure takes the word's suffix out and gives the core response back.
Example: Processed became Process.

(iii)Feature Extraction

Here, python code is used to identify and display the dataset's features.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8388 entries, 0 to 8387
Data columns (total 9 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   id                    8388 non-null   int64
 1   answer_date           8388 non-null   object
 2   ministry              8388 non-null   object
 3   question_type         8388 non-null   object
 4   question_no           8388 non-null   int64
 5   question_by           8388 non-null   object
 6   question_title        8388 non-null   object
 7   question_description  8388 non-null   object
 8   answer                8388 non-null   object
dtypes: int64(2), object(7)
memory usage: 589.9+ KB
```

Figure 2: Feature Extraction

Feature extraction of 2017 dataset depicts in the figure 2.

(iv)Psychological aware

	2017	2016	2015	2014	2013	2012	2011	2010	2009
0	development	development	development	projects	projects	projects	power	projects	cases
1	land	railway	projects	coal	coal	coal	schemes	national	judges
2	pradesh	scheme	funds	development	schemes	funds	projects	power	currency
3	farmers	funds	national	national	water	water	india	pradesh	court
4	railway	country	plan	water	country	power	national	development	banks
5	funds	new	power	new	scheme	development	country	coal	loan
6	water	water	scheme	power	pradesh	pradesh	act	country	judiciary
7	policy	production	foreign	implementation	power	country	coal	cwg	essential
8	india	pradesh	country	country	development	national	pradesh	sector	courts
9	country	services	new	pradesh	national	production	water	indian	credit
10	sector	projects	sector	sector	railway	sector	new	implementation	act
11	projects	national	production	railway	production	assam	funds	india	sezs
12	scheme	india	farmers	delhi	schools	india	sector	railway	central
13	new	sector	india	construction	sector	railway	development	water	price
14	rural	states	railway	funds	act	cases	scheme	production	tax
15	women	women	implementation	setting	new	rural	delhi	scheme	disclosure
16	construction	project	areas	india	women	setting	status	project	loans
17	power	increase	women	production	delhi	assistance	states	tamil	disinvestment
18	demonetisation	farmers	schemes	shortage	construction	scheme	cases	rural	insurance
19	increase	setting	states	cases	rural	states	shortage	land	laws

Figure 3: Displaying the shifting trends in State of Council questions

From 2009 to 2017, psychological (emotional) awareness of pupils was used to demonstrate how well the questionnaire changed over time.

(v)Contextualization

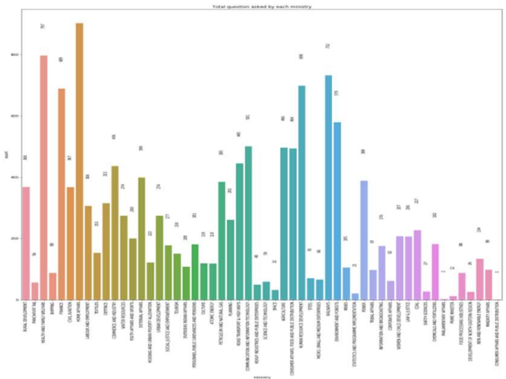


Chart 1: Contextualization

In the chart 1: contextualization process displays all questions from each ministry.

(vi)Analysis and visualization

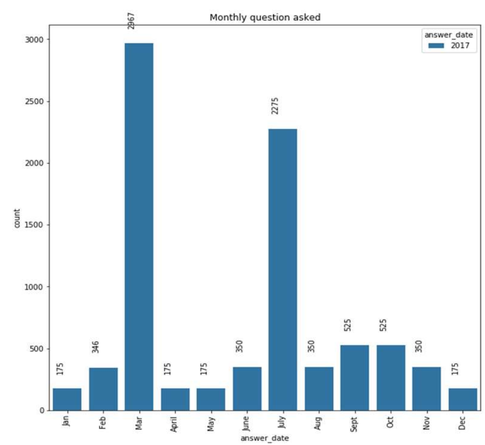
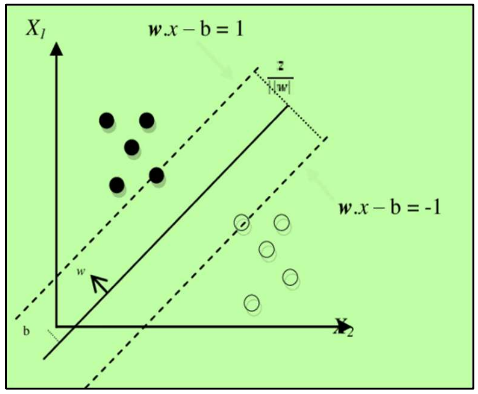


Chart 2: Examining and visualising monthly queries from the year 2017.

The Council of States question and answer dataset for 2017 is shown in chart 2, along with the response dates and the quantity broken down by month.

Algorithm used- SVM (Support Vector Machine)

The supervised (feed-me) machine learning algorithm known as prediction or categorization issues can be solved with SVM. While classification forecasts a labeling or group, regression predicts a continuous value. SVM achieves classification by determining the hyper-plane that separates the categories we displayed in n-dimensional dimensions.



SVM produces a hyper - plane by transforming the inputs using what are known as "Kernels," or statistical processes. Several varieties of kernels include linear, sigmoid, RBF, non-linear, polynomial, etc.

Whenever the input is unknown beforehand, the overall kernel for non-linear circumstances, the kernel variable "RBF" is used as a hyper parameter. For problems with linear specificity, use the kernel option "linear". Because our issue is linear, humans shall make use of "linear SVM" (consisting only of positive and negative values).

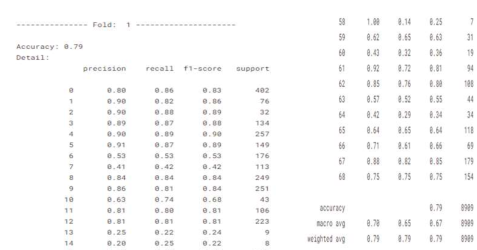
Steps for creating a model

- Optimal data collection for testing and training
- Data vectorization
- Constructing a linear SVM model that is both predictive and trainable

IV.Results and Analysis

It produced remarkably positive outcomes. Additionally, it is planned to comment extra documents to see whether the accuracy improves any further. The number of annotations made was the restriction. Support vector machine (SVM problem was finally used to approach the classification challenge. Initially it classified the documents individually. Later, it would like to conduct sentence-level analysis. This study will motivate scholars to pursue additional study on textual appreciation and blame extraction in accordance with current approaches to argument mining, hate speech, sarcasm production, etc.

(1)Fold 1



--------------- Fold: 1 --------------------

Accuracy: 0.79
Detail:

	precision	recall	f1-score	support
0	0.80	0.86	0.83	402
1	0.90	0.82	0.86	76
2	0.90	0.88	0.89	32
3	0.89	0.87	0.88	134
4	0.90	0.89	0.90	257
5	0.91	0.87	0.89	149
6	0.53	0.53	0.53	176
7	0.41	0.42	0.42	113
8	0.84	0.84	0.84	249
9	0.86	0.81	0.84	251
10	0.63	0.74	0.68	43
11	0.81	0.80	0.81	106
12	0.81	0.81	0.81	223
13	0.25	0.22	0.24	9
14	0.20	0.25	0.22	8
15	0.82	0.89	0.85	102
58	1.00	0.14	0.25	7
59	0.62	0.65	0.63	31
60	0.43	0.32	0.36	19
61	0.92	0.72	0.81	94
62	0.85	0.76	0.80	108
63	0.57	0.52	0.55	44
64	0.42	0.29	0.34	34
65	0.64	0.65	0.64	118
66	0.71	0.61	0.66	69
67	0.88	0.82	0.85	179
68	0.75	0.75	0.75	154
accuracy			0.79	8909
macro avg	0.70	0.65	0.67	8909
weighted avg	0.79	0.79	0.79	8909

Fig. 4: Fold 1 – Accuracy 79%

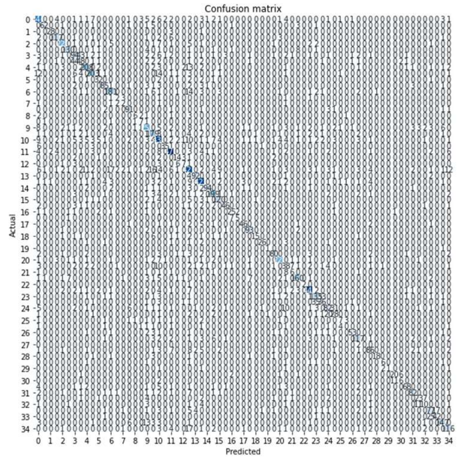


Fig. 5: Fold 1 – Confusion Matrix

(2)Fold 2

--------------- Fold: 2 ---------------------

Accuracy: 0.79

Detail:

	precision	recall	f1-score	support
0	0.79	0.84	0.81	402
1	0.92	0.88	0.90	76
2	0.85	0.91	0.88	32
3	0.89	0.84	0.86	134
4	0.94	0.90	0.92	257
5	0.89	0.85	0.87	149
6	0.53	0.56	0.54	176
7	0.38	0.42	0.40	113
8	0.81	0.84	0.83	249
9	0.84	0.82	0.83	251
10	0.59	0.77	0.67	43
11	0.87	0.84	0.85	107
12	0.78	0.80	0.79	223
13	0.50	0.10	0.17	10
14	0.17	0.38	0.23	8
56	1.00	0.62	0.76	26
57	0.78	0.50	0.61	14
58	0.80	0.57	0.67	7
59	0.77	0.55	0.64	31
60	0.55	0.63	0.59	19
61	0.92	0.80	0.85	95
62	0.96	0.83	0.89	109
63	0.57	0.48	0.52	44
64	0.46	0.47	0.46	34
65	0.66	0.67	0.67	118
66	0.66	0.57	0.61	69
67	0.81	0.83	0.82	179
68	0.80	0.78	0.79	154
accuracy			0.79	8909
macro avg	0.69	0.66	0.67	8909
weighted avg	0.79	0.79	0.79	8909

Fig.6: Fold2 – Accuracy 79%

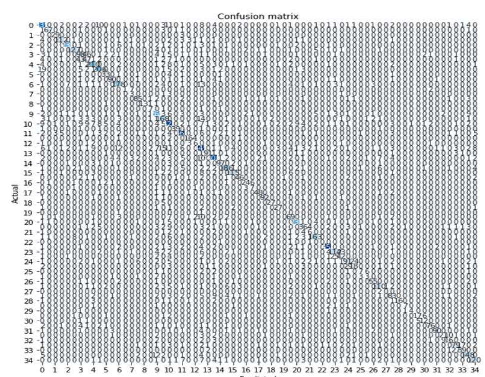


Fig. 7: Fold 2 – Confusion Matrix

(3)Fold 3

accuracy			0.79	8909
macro avg	0.72	0.67	0.69	8909
weighted avg	0.79	0.79	0.79	8909

Fig.8: Fold 3 – Accuracy 79%

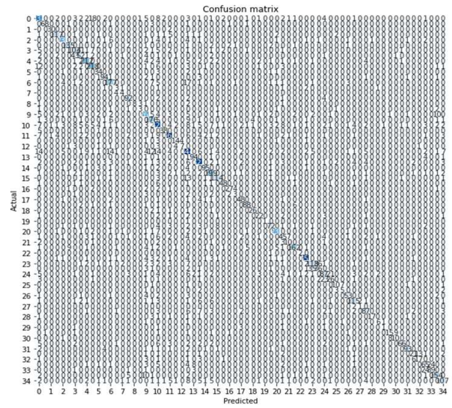


Fig. 9: Fold 3 – Confusion Matrix

(4)Fold 4

accuracy			0.79	8909
macro avg	0.69	0.65	0.67	8909
weighted avg	0.80	0.79	0.79	8909

Fig.10: Fold 4 – Accuracy 79%

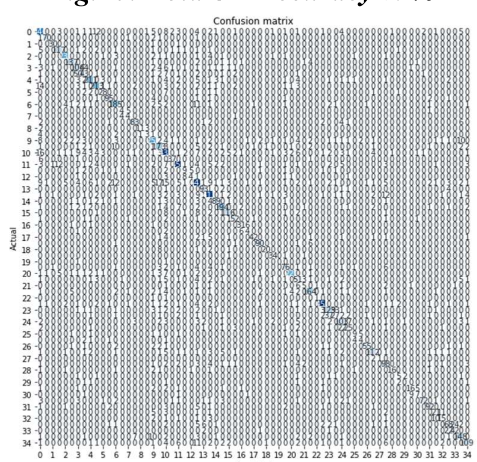


Fig. 11: Fold 4 – Confusion Matrix

(5)Fold 5

accuracy			0.79	8909
macro avg	0.69	0.67	0.67	8909
weighted avg	0.79	0.79	0.79	8909

Fig.12: Fold 5 – Accuracy 79%

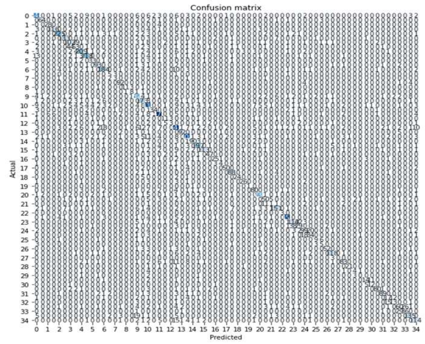


Fig. 13: Fold 5 – Confusion Matrix

(6)Fold 6

accuracy			0.79	8908
macro avg	0.67	0.65	0.66	8908
weighted avg	0.79	0.79	0.79	8908

Fig.14: Fold 6 – Accuracy 79%

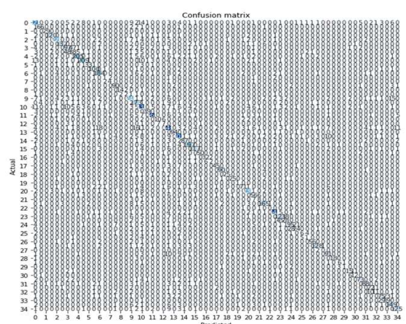


Fig. 14: Fold 6 – Confusion Matrix

CV accuracy: 78.80 +/- 0.42 max: 79.37

CV precision: 78.93 +/- 0.45 max: 79.56

CV recall: 78.80 +/- 0.42 max: 79.37

CV f1: 78.72 +/- 0.43 max: 79.32

Fig.6: Overall values for cross validation methods

Cross validation done with the 10batches that is 10 folds**.**

V.Conclusion:

In this paper, the author presented a dataset of State of Council question and answer. We analyzed the purpose of the speeches of the Member of Parliament and categorized them into 4 major categories and provided statistics of the categories. Also tried to identify and classify them automatically using SVM algorithm and provided the results. The analysis is done for understanding the purpose of the speeches in the parliament from 2009 to 2017. Finally 10 folds were used and finalised with the cross validation methods to identify the speeches whether the member is in favour of the debate topic or not.

REFERENCES

[1] Monroe BL, Schrodt PA. 2008. Introduction to the Special Issue: The Statistical Analysis of Political Text. Polit. Anal. 16: 351–355.

[2] Rose S, Engel D, Cramer N, Cowley W. 2010. Automatic Keyword Extraction from Individual Documents. In: Text Mining: Applications and Theory., p 1–20.

[3] Timonen M, Toivanen T, Teng Y, Chen C, He L. 2012. Informativeness-based Keyword Extraction from Short Documents. In: KDIR 2012 - Proceedings of the International Conference on Knowledge Discovery and Information Retrieval., p 411–421.

[4] Khurana D, Koli A, Khatter K, Singh S. 2017. Natural Language Processing: State of The Art, Current Trends and Challenges.

[5] A.I. Will Mark A Turning Point in the History of Politics | by SukhaylNiyazov | Towards Data Science.

[6] Zhang L, Wang S, Liu B. 2018. Deep Learning for Sentiment Analysis : A Survey. Wiley Interdiscip. Rev. Data Min. Knowl. Discov.

[7] Shen C, Sun C, Wang J, Kang Y, Li S, Liu X, Si L, Zhang M, Zhou G. 2020. Sentiment classification towards question-answering with hierarchical matching network.Proc.2018 Conf. Empir.Methods Nat. Lang. Process. EMNLP 2018: 3654–3663.

[8] Vo DT, Zhang Y. 2015. Target-dependent twitter sentiment classification with rich automatic features.IJCAI Int. Jt. Conf. Artif.Intell. 2015-Janua: 1347–1353.

[9] Gold D, Bexte M, Zesch T. 2018. Corpus of aspect-based sentiment in political debates. KONVENS 2018 - Conf. Nat. Lang. Process. / Die Konf. zurVerarbeitung Nat. Spr.: 89–99.

[10] Rauh C. 2018. Validating a sentiment dictionary for German political language—a workbench note. J. Inf. Technol. Polit. 15: 319–343.

[11] Sidarenka U. 2019.Sentiment analysis of German Twitter.

[12] Singhal K, Agrawal B, Mittal N. 2015. Modeling Indian General Elections: Sentiment Analysis of Political Twitter Data. In: Mandal JK, Satapathy SC, Kumar Sanyal M, Sarkar PP, Mukhopadhyay A, editors. Information Systems Design and Intelligent Applications. New Delhi: Springer India, p 469–477.

[13] Sharma P, Moh TS. 2016. Prediction of Indian election using sentiment analysis on Hindi Twitter. Proc. - 2016 IEEE Int. Conf. Big Data, Big Data 2016: 1966–1971.

[14] Rohit SVK, Singh N. 2018. Analysis of Speeches in Indian Parliamentary Debates.

[15] Ansari MZ, Aziz MB, Siddiqui MO, Mehra H, Singh KP. 2020. Analysis of Political Sentiment Orientations on Twitter. ProcediaComput. Sci. 167: 1821–1828.

[16] V. Singh and S. K. Dubey, "Opinion Mining and Analysis: A Literature Review," 5th International Conference on Confluence of Next-Generation Information Technology Summit, 2014, pp. 232-239.

[17]Modelling and Forecasting Students' Academic Performance Using Data Mining Techniques, no. November 2016, pp. 36–42, A. M

[18] Q & A Discussed in Parliament of India | Kaggle.

[19] Shreyas R Hegde, Yogesh R Gaikwad, A Hybrid Deep Neural Network For Aspect Based Sentiment Analysis On State of Council Questions, Indian Journal of Computer Science and Engineering (IJCSE), Vol. 12 No. 1 Jan-Feb 2021, pp. 112 - 128.