

The Sentiment Analysis and Homophily Analysis of Twitter Indian Political Data

^[1] Mansi Trambaklal Vegad, ^[2] Lokesh Gagnani

^{[1][3]} Shankersinh Vaghela Babu Institute of Technology, Gandhinagar, India, ^[2] Shankersinh Vaghela Babu Institute of Technology, Gandhinagar, India

^[1] mansivegad123@gmail.com, ^[2] gagnani.lokesh@gmail.com

Abstract: In now a day's social media is a big platform for data analysis and research. For Sentiment Analysis, I choose Tweepster. I use Tweepster for accessing tweeter data. I perform sentiment analysis on Indian Political data. I got 117546 tweets of 2019 Indian Election. I use SVM (Support Vector Machine) Classifier for sentiment Analysis. I also perform Homophily Analysis on the result of Sentiment Analysis. I perform Homophily Analysis using based on Location, based on Followers and Following user's detail. The newel work is the Homophily Analysis based on new features like Location, Followers Following detail. At the end, the homophily score is got as the result.

Keywords: Sentiment Analysis, Homophily Analysis, Mchine Learning, Twitter Data Analysis, Classification, Political Analysis

1. INTRODUCTION

Slant Analysis is one of the arrangements of Machine learning and NLP (Natural language processing). Opinion Analysis do investigate and identification of the information which is given by the scholars and clients. Writers give the opinions about any product or any topic and Sentiment Analysis can analyze and identify the sentiments. Sentiment Analysis normally concerned with the opinion and emotions from the text/data. It analyzes and proves the sentiment of writer with respect to a data. Sentiment Analysis is used to extract and categorize opinions from different content structures, including news, reviews, audits and articles. It categorizes contents that are positive, negative or neutral.

Homophily analysis

Homophily infers the contact between similar people occurs at a higher rate than among various people. It suggests that any social component that depends to an extensive degree on frameworks for its transmission will when all is said in done be restricted in social space and will agree to certain focal components as it works together with other social components in science of social structures[9].

Homophily Analysis restricts individuals' social data such that have ground-breaking suggestions for the data they get, the mentalities they structure, and the collaborations they experience. Homophily makes the solid partitions in our own surroundings with age, religion, instruction, occupation, and sexual orientation generally. Social frameworks Ties between nonsimilar people additionally

break down at a higher rate, which makes way for the arrangement of specialties (confined situations) inside social space [9]. Homoplily analysis gives the result of similarities between people that they have in age, language, state, etc. Homophily analysis gives the homophily result according to different features.

Twitter

Twitter provides a platform which connects people to each other no matter how far they are. Twitter gives the message

facility and this message called "tweet". This credit goes to social media. Social media is the platform of sharing and receiving information, data, as well as communication among people. They share their thinking, trending news, ideas, behaviors, and sentiments related any topic. It is powerful weapon of increasing literature, and research. There are numerous valuable online life stage yet twitter is the most solid stage for assessment investigation in light of the fact that there are a great many overall dynamic clients, millions day by day dynamic clients and millions posts each day. They show their sentiments and they are participated on different topics through the twitter posts. Tweets are useful for Sentiment Analysis data sets. The Tweets(twitter data) data can be received from Twitter in a very secure and easy way. We can easily receive the tweets through twitter API (Application Programming Interface).

Literature review

In this research, we survey the different papers about Homophily Analysis, and Sentiment Analysis about

Indian political issues. For Sentiment Analysis there are many methods of machine learning like Naive Bayes, Rule based classification, C.5 method, Support Vector Machine (SVM), Dictionary based method, etc.

In this paper they collect 4.9 billion users' data for political Analysis. They collect data using the official Twitter API (Application Programming Interface). In America there are two candidates for the election. They first use rule based classification for Sentiment Analysis. In rule based classification they divide the six classes for Sentiment analyses are Political and Non-Political. In Political there are five classes Trump supporter, Hillary supporter, Positive, Neutral, and Negative and In Non-political only one class is whatever. They have done Sentiment Analysis with the use of Homophily Analysis. They have done Homophily Analysis by their Follow connection, Retweet connection and Mention connection. They also analyze unidirectional and reciprocal connection between users. In this paper they get the Homophily result about all the six classes of follow, retweet, and mention [6].

In this paper they had collect the data from Twitter Archiver Tool. They collect tweets in Hindi language only. They differentiate political and non political tweets and perform Sentiment Analysis of political tweets. They Performed Sentiment Analysis using Naive Bayes algorithm, Support Vector Machine (SVM), and Dictionary based algorithm. They compare all this methods and get the accuracy of Sentiment Analysis. They calculate the tweets of all the party that are positive, neutral or negative. They proved that the accuracy of naive Bayes is 62.1%, the accuracy of SVM (Support Vector Machine) is 78.4% and the accuracy of Dictionary based algorithm is 34% [13].

In this paper, analysts propose the structure that explains the movement of the assortment, Sentiment Analysis, and order Twitter feelings. They foresee the result of a political race result by utilizing the twitter information. They led to foresee political race in US, UK, Spain, French and Indonesia itself. In view of these information, the creators proposed another technique to anticipate the political race result that centers on tweet considering and Sentiment Analysis the preprocessing task. They concentrated distinctly on tweet tallying, Sentiment Analysis and pre preparing task. They give another strategy which is simple contrasted with different strategies [20].

In this paper the main objectives are the result of

Sentiment mining in every province, to compare the prediction results of predictions of naive Bayes and C.5 methods, and to analyze the relationship between Sentiments in each province (correlation Analysis). They used twitter crawling process (TCP) for collecting the data. They used lexicon approach for Sentiment mining. They proved that C.5 method is better than naive Bayes method in terms of accuracy, precision and recall. C.5 method has better result [1].

In this paper, they utilized Merging of systems like information mining with different methodologies (content mining, NLP and computational insight). Authors are blending Support Vector Machine (SVM) with Decision Tree. In this, demonstrate that half breed model improved the general arrangement estimated in exactness and f-measure. Sentiment forecast is better when contrasted with customary procedures for grouping. They utilized STACKNET HYBRID Technique for production of the cross breed calculation. SVM (bolster vector machine) system assembles the order model that doles out content guides to predefined class classifications. This methodology making a non probabilistic direct classifier. They used TF-IDF (term frequency – inverse document frequency) method for convert string input into numeric form. They performed first adaboosted decision tree algorithm and the accuracy is 67%, second SVM algorithm and its accuracy is 82%, and third they perform hybrid algorithm which is the combination of decision tree algorithm and SVM algorithm and its accuracy is 84%. Hybrid algorithm have the higher accuracy compared to both [7].

In this research, they use three classification methods Naive Bayes, Support Vector Machine, and logistic regression. They used Tweepy (twitter API) which is used in python to provide access to the twitter data. They prove naive Bayes is better among three techniques. They used naive Bayes algorithm with Sentimental score calculation in the appropriate ratio for better accuracy of the model. The accuracy of naive Bayes algorithm is 77% and the accuracy of sentiment score count is 95%. They determined the precision of combined both the calculation is 90%. They utilized Sentiment Analysis dataset from Kaggle as information source. The restriction of this examination is that the model is language explicit [17].

In this research, they use Sentiment Analysis using Multinomial Naive Bayes and Decision tree algorithms. They used Apache Spark cluster (Apache Spark framework) for fast processing. In this paper the results

show that Decision tree performs extremely well showing accuracy, precision, recall and F1-Score [11].

In this paper, the accuracy of naive bayes algorithm was less than the accuracy of support vector machine. They made last expectation using SVM, since the accuracy of algorithm is higher [21].

In this paper, the results of classification prediction in the sentiment class show that C5.0 method is a better method than the Naive Bayes method. This is indicated by the value of accuracy, precision, and recall method C5.0 which is higher than the Naive Bayes method [22].

In this paper, Sentiment Analysis end up being fit for breaking down a great many critiques going with a single comment [23].

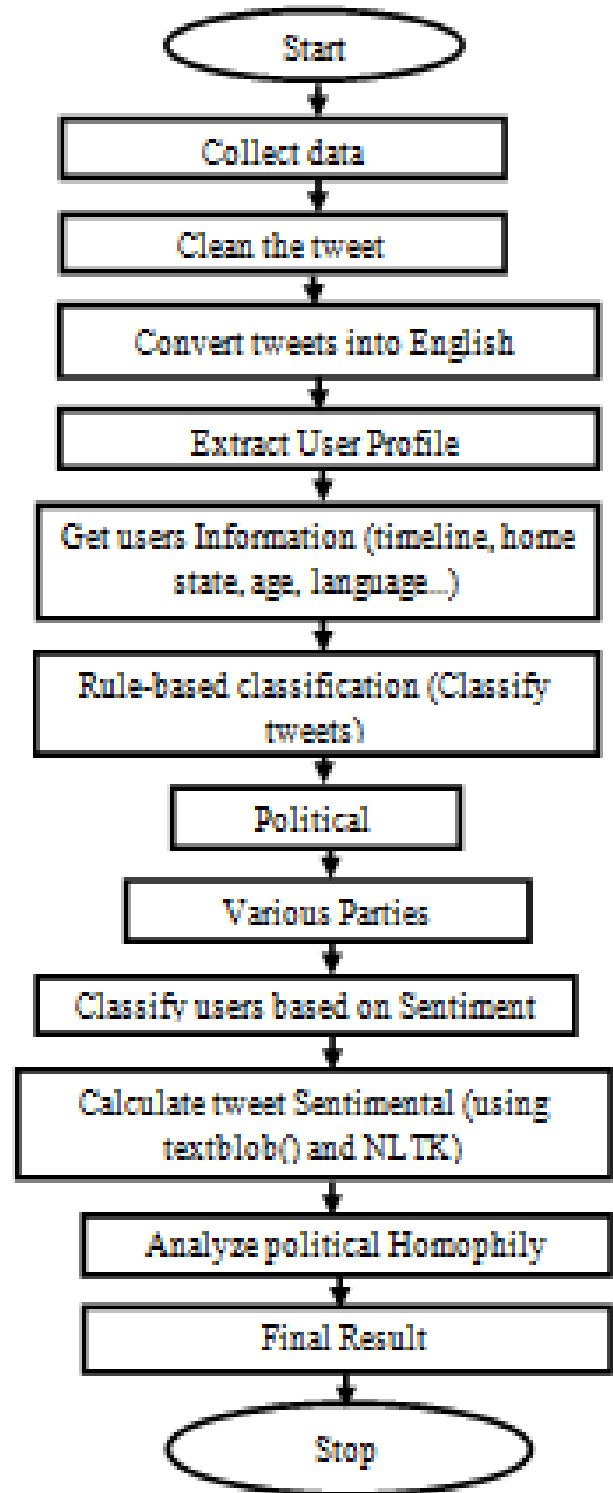
In this paper, the framework can break down shared web based life substance uninhibitedly utilized, all things considered, progressively to discover occasions identified with fiascos or mishaps, and examine and foresee wistful ways, the framework can be used for catastrophe notice administration for seismic tremors and waves or continuous car crash educating administration to quickly distinguish calamities and mishap in this manner decreasing harm [24].

proposed method

Sentiment Analysis is useful in every type of business, social media or in personal use. Sentiment Analysis for Indian political issues by using the twitter data is the difficult task because there are many parties and We also use Homophily Analysis for find the similarities of users by their age, home state, their timeline, language, etc.

We also classify parties by using rule based classification method. In India there is no only English tweets founded so we have to convert the tweets also. The research proposes to the Analysis of Indian political parties. In India there are many parties in election so we divide them in two part one in current ruling party and second is opposition parties. In India, there are many languages tweets are there. So we convert all tweets into English. In addition, we apply Homophily Analysis by age, home state, profile, language, timeline, etc.

Proposed Flow of System



Proposed Algorithm

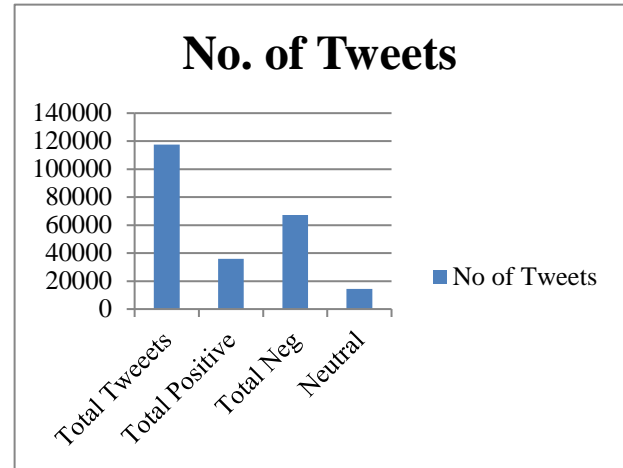
- Step 1: Read the Tweets.
- Step 2: Clean the Tweets.
- Step 3: Convert Tweet to English.
- Step 4: Extract User Profile.
- Step 5: Get Gender, Home State etc Information.
- Step 6: Get User TimeLine.
- Step 7: Classify Tweets into political, non-political
- Step 8: Calculate Tweet Sentimental (using textblob() and NLTK).
- Step 9: Classify Users based on Sentimental score
- Step 10: Analyze Political Homophily
- Step 11:

Final Result.

The proposed method gives the better result and accuracy compared to existing method. I will perform Homophily Analysis with additional features. In existing method they were perform Sentiment Analysis using Homophily Analysis with the use of follow connection, re tweet connection, and hash tags. I will perform Homophily Analysis with the use of age, gender. We also add a GEO location based Analysis technique for state wise Analysis.

implementation

First, we have done Sentiment Analysis on collected Twitter Data. We get the Positive, negative and neutral tweets.



Total no. of Tweets with Sentiment Analysis Result

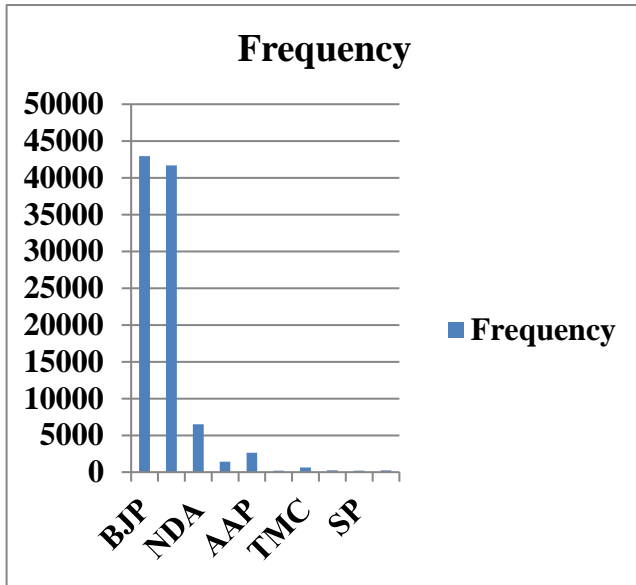
Conditions for Sentiment Analysis:

Score = no of positive-no of negative
 If Score > 0, the sentence has an overall 'positive opinion'
 If Score < 0, the sentence has an overall 'negative opinion'
 If Score = 0, then the sentence is considered to be a 'neutral opinion'

Total no. of Tweets, Positive, Negative, Neutral

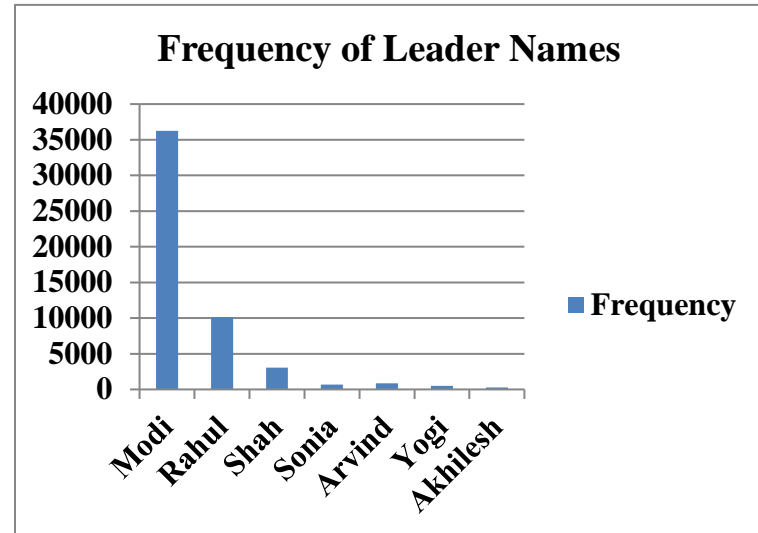
Type of Tweets	No. of Tweets
Total Tweets	117546
Total Positives	35830
Total Negatives	67240
Neutral	14476

This is the frequency of all parties during 2019 Indian election.



Frequency of all Parties
Frequency of all Parties

Keyword	Frequency
BJP	42958
Congress	41673
NDA	6492
UPA	1422
AAP	2617
TDP	221
TMC	617
BSP	264
SP	191
DMK	234

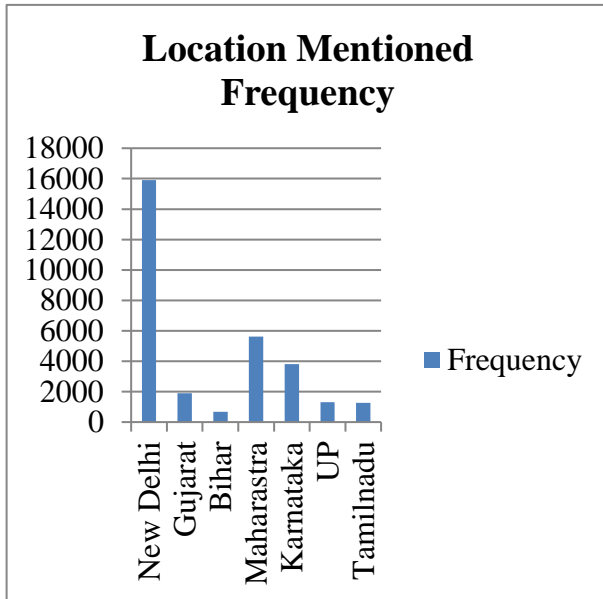


Frequency of Leader Names
Frequency of Leaders

Leader	Frequency
Modi	36229
Rahul	10125
Shah	3050
Sonia	675
Arvind	857
Yogi	514
Akhilesh	267

Location based Frequency of Indian Twitter Political Users twitter users.

This is the frequency of Leaders in 2019 Indian election.



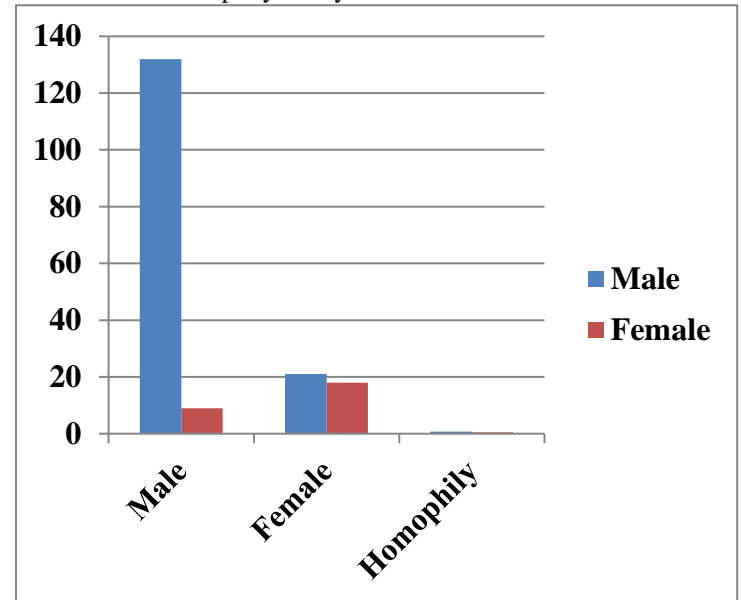
Location Mentioned Frequency

```
Function for Location based homophily
def chance_homophily(Userlist):
    Chance_homophily=0
    For location_count in Userlist:
        Print(location_count)
        Chance_homophily+=
        (Location_count/sum(/userlist))*2
    Return chance_homophily
```

Location Mentioned Frequency

Location	Frequency
New Delhi	15899
Gujarat	1887
Bihar	677
Maharashtra	5617
Karnataka	3812
UP	1298
Tamilnadu	1272

Gender wise Homophily Analysis

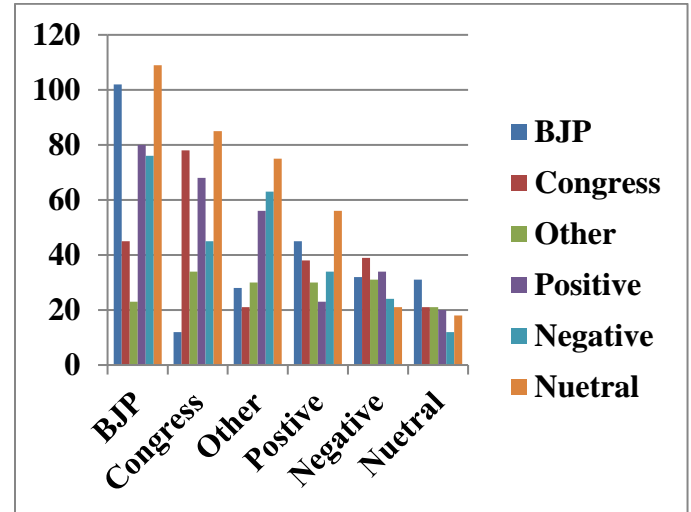
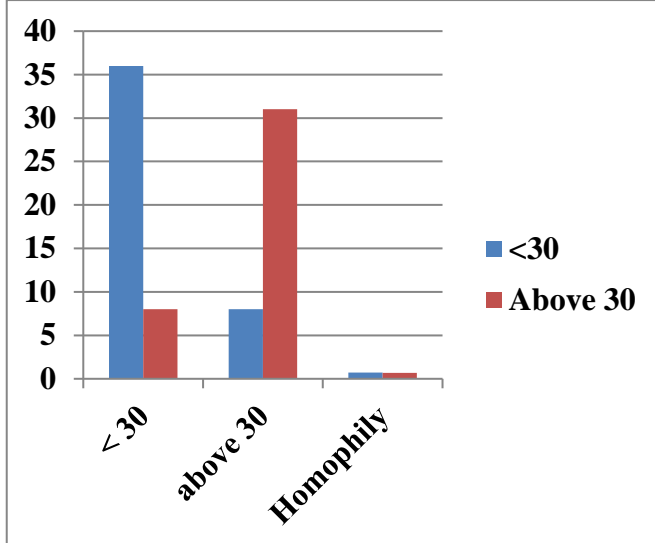


Gender wise Homophily

```
Function for Gender wise homophily
def chance_homophily(Userlist):
    Chance_homophily=0
    For Gender_count in Userlist:
        Print(Gender_count)
        Chance_homophily+=(Gender_count/sum(/userli
st))*2
    Return chance_homophily
Gender wise Homophily
```

Gender	Male	Female	Homophily
Male	132	21	0.76
Female	9	18	0.55

Age wise Homophily Analysis



Followers | Following no. of users in Classes
Followers | Following No. of users in Classes

Age wise Homophily

Function for age based homophily

```
def chance_homophily(Userlist):
    Chance_homophily=0
    For Age_count in Userlist:
        Print(Age_count)
        Chance_homophily+=
        (Age_count/sum(/userlist))**2
    Return chance_homophily
```

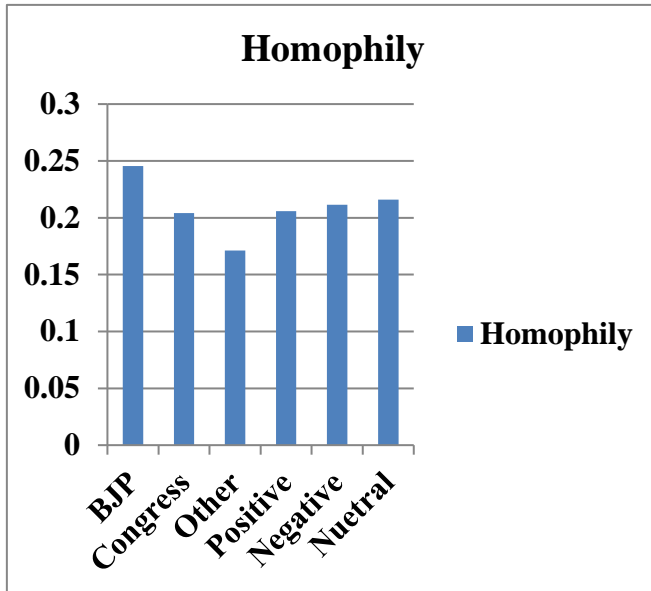
Age wise Homophily

Age Group	< 30	above 30	Homophily
<30	36	8	0.7
Above 30	8	31	0.67

Party	BJP	Congress	Other	Positive	Negative	Neutral
BJP	102	12	28	45	32	31
Congress	45	78	21	38	39	21
Other	23	34	30	30	31	21
Positive	80	68	56	23	34	20
Negative	76	45	63	34	24	12
Neutral	109	85	75	56	21	18

Homophily Score

The result of Homophily of Followers and Following Parties of Indian election



Homophily Score

Formula for finding Homophily score

$$Hi = \frac{Si}{Si + Di}$$

Hi: Hi shows the Homophily index value of users.

Si: Si represents the number of connections between i individuals. It shows the similarity value of users. It is for homogeneous connections.

Di: Di represents the number of connections that bind individuals of kind i with individuals of other kinds. It shows the dissimilarity value of users. It is for heterogeneous connections [6].

Homophily Score

Party	Homophily
BJP	0.2454
Congress	0.2041
Other	0.1711
Positive	0.2057
Negative	0.2115
Neutral	0.216

Conclusion

Enormous measure of information are accessible on twitter for dissect. Sentiment Analysis gives the Sentiment of users about any topic or product or identity. In now a day's election is a trending, hot and interested

topic for social media users. There are many techniques performed for tweeter Sentiment Analysis like Naive bayes, SVM (support vector machine), decision tree, C.5, lexicon based approach and so many machine learning based techniques. We done Sentiment Analysis and Homophily Analysis on Twitter data.

- We Performed the Sentiment Analysis using rule based classification technique.
- We have Learn Python for opinion mining.
- We Perform opinion Analysis and Homophily Analysis on twitter Indian political data.
- We Use many features for Homophily Analysis.
- We Use GEO Location Based Analysis.

REFERENCES

- [1] E. S. Saputro, K. A. Notodiputro, and I. A. "Study of Sentiment of Governor's Election Opinion in 2018," Int. J. Sci. Res. Sci. Eng. Technol., vol. 4, no. 11, pp. 231–238, 2018.
- [2] Patil, "Restaurant 's Feedback Analysis System using Sentimental Analysis and Data Mining Techniques," 2018 Int. Conf. Curr. Trends Towar. Converging Technol., pp. 1–4, 2018.
- [3] F. A. Pozzi, E. Fersini, E. Messina, and B. Liu, Challenges of Sentiment Analysis in Social Networks: An Overview, vol. 1. Elsevier Inc., 2017.
- [4] F. Poecze, C. Ebster, and C. Strauss, "Social media metrics and Sentiment Analysis to evaluate the effectiveness of social media posts," Procedia Comput. Sci., vol. 130, pp. 660–666, 2018.
- [5] H. P. Patil and M. Atique, "Sentiment Analysis for social media: A survey," 2015 IEEE 2nd Int. Conf. InformationScience Secur. ICISS 2015, 2016.
- [6] J. A. Caetano, H. S. Lima, M. F. Santos, and H. T. Marques-Neto, "Using Sentiment Analysis to define twitter political users' classes and their Homophily during the 2016 American presidential election," J. Internet Serv. Appl., vol. 9, no. 1, 2018.
- [7] K. Ganagavalli, A. Mangayarkarasi, T. Nandhinisri, and E. Nandhini, "Sentiment Analysis of twitter data using machine learning algorithm," J. Comput. Theor. Nanosci., vol. 15,

- no. 5, pp. 1644–1648, 2018.
- [8] M. Kamyab, R. Tao, M. H. Mohammadi, and A. Rasool, “Sentiment Analysis on Twitter,” vol. 9, no. 4, pp. 14–19, 2018
- [9] M. Miller, S.-L. Lynn, and M. C. James, “Birds of a Feather: Homophily in Social Networks,” *Annu. Rev. Sociol.*, vol. 27, pp. 415–444, 2001.
- [10] M. S. M. Vohra and P. J. B. Teraiya, “Journal of Information, Knowledge and Research in Computer Engineering a Comparative Study of Sentiment Analysis Techniques,” *J. Information, Knowledge Res. Comput. Eng.*, pp. 313–317, 2013.
- [11] P. Deacon, “Application of Machine Learning Techniques to Mineral Recognition,” *Computer (Long. Beach. Calif.)*, no. October, pp. 628–632, 2001.
- [12] P. Seth, A. Sharma, and R. Vidhya, “Sentiment Analysis of Tweets Using Hadoop,” *Int. J. Eng. Technol.*, vol. 7, no. 3.12, p. 434, 2018.
- [13] P. Sharma and T. S. Moh, “Prediction of Indian election using Sentiment Analysis on Hindi Twitter,” *Proc. - 2016 IEEE Int. Conf. Big Data, Big Data 2016*, pp. 1966–1971, 2016.
- [14] R. Singh and R. Kaur, “Sentiment Analysis on Social Media and Online Review,” *Int. J. Comput. Appl.*, vol. 121, no. 20, pp. 44–48, 2015.
- [15] S. A. El Rahman, F. A. Alotaibi, and W. A. Alshehri, “Sentiment Analysis of Twitter Data,” *2019 Int. Conf. Comput. Inf. Sci. ICCIS 2019*, 2019.
- [16] S. Geetha and V. K. Kaliappan, “Tweet Analysis Based on Distinct Opinion of Social Media Users’,” *ICSNS 2018 - Proc. IEEE Int. Conf. Soft-Computing Netw. Secur.*, pp. 1–6, 2018.
- [17] S. Goel, M. Banthia, and A. Sinha, “Modeling Recommendation System for Real Time Analysis of Social Media Dynamics,” *2018 11th Int. Conf. Contemp. Comput. IC3 2018*, pp. 1–5, 2018.
- [18] S. Y. Yoo, J. I. Song, and O. R. Jeong, “Social media contents based Sentiment Analysis and prediction system,” *Expert Syst. Appl.*, vol. 105, pp. 102–111, 2018.
- [19] V. A. and S. S. Sonawane, “Sentiment Analysis of Twitter Data: A Survey of Techniques,” *Int. J. Comput. Appl.*, vol. 139, no. 11, pp. 5–15, 2016.
- [20] W. Budiharto and M. Meiliana, “Prediction and Analysis of Indonesia Presidential election from Twitter using Sentiment Analysis,” *J. Big Data*, vol. 5, no. 1, pp. 1–10, 2018.
- [21] Neethu M, S, Rajasree R, ‘Sentiment Analysis in Twitter using Machine Learning Techniques’, 4th ICCCNT , 2013.
- [22] U. Kursuncu, M. Gaur, U. Lokala, K. Thirunarayan, A. Sheth, and I. B. Arpinar, “Predictive Analysis on Twitter: Techniques and Applications,” pp. 67–104, 2019.
- [23] W. Budiharto and M. Meiliana, “Prediction and analysis of Indonesia Presidential election from Twitter using Sentiment Analysis,” *J. Big Data*, vol. 5, no. 1, pp. 1–10, 2018.
- [24] Y. Mejova, ‘Sentiment Analysis: An overview’, ymejova/publications/CompsYelenaMejova, vol. 2010–02–03, 2009, 2009.
- [25] <https://monkeylearn.com/sentiment-analysis/>
-