

# Opinion Analysis of Digital India Mission

M. SreeVani

**ABSTRACT**---“DigitalIndia” project has launched thru Indian authorities. This venture end up anticipated with purpose to digitally empower the people parents of a. authorities of India can offer unfastened wireless in Trains & Railway Stations, excessive tempo internet offerings, Boast New Scheme-Digi-Locker.severa amount of information is accrued approximately DigitalIndia venture in social media.in this paper, opinion evaluation is completed to discover how human beings enjoy about a DigitalIndia. ten thousand tweets have been accumulated and statistics mining became achieved to recognize the real opinion of the citizens about DigitalIndia, the use of Random forest classifier set of guidelines. The effects of evaluation were very convincing and strong that the Indian residents had a nice opinion inside the course of the mission “DigitalIndia”.

**Keywords**—DigitalIndia, Twitter data, opinion analysis.

## 1. INTRODUCTION

In recent times Social net net websites become a supply of humans’s critiques. those internet internet web sites will offer an get entry to to public to post their opinions on an expansion of subjects which is probably presently occurring u . s .. In reality, many organizations have began to poll those social networks to get a experience of citizen’s opinion of any government recommendations or any public and political troubles. via social networks authorities or any organization can accumulate public opinion on any government policy or any authorities preference. One tough venture is to collect technology to turn out to be privy to and summarize an traditional opinion.The people during India also can posted their evaluations on DigitalIndia in Social media. NLTK 3.0 modules protected with Python code. project consists of StopWords removal, Tokenization and Stemming. The tweets gathered from twitter are a combination of urls, and particular non-opinional facts like hashtags “#”, annotation “@” and retweets “RT”. One advantage of this information is that the tweets are accrued in a streaming fashion and therefore represent a actual sample of actual tweets in phrases of language use and content material fabric material.

In this paper, Twitter information may be take into account and bring together fashions for classifying tweets into splendid, terrible and independent opinion. Then study the twitter information to discover the impact of DigitalIndia on the humans. The opinion of the Twitter tweets may be analyzed the use of Random woodland set of policies. This paintings may be prolonged to offer summarization of opinions clearly so authorities could have participation of citizens, via records their views/reviews, in approach and implementation of numerous governance insurance. segment 2 communicate approximately associated paintings and segment three describes opinion evaluation of twitter

records and phase 4 communicate about consequences and decided thru prevent.

## 2. RELATED WORK

PriyaAnanthram [1] Discussed about opinion analysis. It measures like, dislike opinions of the people and can be used to find the customer responses about products or events.

Xiaowen Ding, Bing Liu [3], In this authors are proved that word presence is more effective than word frequency for opinion analysis. Word location in sentence is important to recognize the priority of word in sentence.

Pak, A., Paroubek, P [4], authors are demonstrated the procedure to segregate negative and positive opinions from collected twitter data.

Go, A., Bhayani, [5][6] discussed tweets classification based on number of negative opinion and positive opinion.

Sunil B. Mane, Kruti Assar [7], In this authors are discussed about product rating using opinion mining.

But, The Twitter follows micro-on foot a weblog nature, small tweets and moreover helps first-rate herbal languages, so Twitter is the brilliant supply for the opinion assessment.three. Opinion assessment of Tweets the usage of Random wooded area

Random wooded location Classifier combines the predictions of some base estimators constructed with desire tree algorithm to enhance robustness over an individual estimator. it's miles used due to its easiness in every at some stage in education and classifying steps. Pre-processed data is given as input to teach input set the usage of Random forest classifier and that knowledgeable model is completed on check to generate each remarkable or terrible opinion. If we need to categorise a new statistics, every tree offers its category prediction as one vote. In famous, the more timber within the random wooded place the better accuracy outcomes given.

### 3.1 Information Pre-Processing

Twitter has created its private API for tweets retrieval. we've were given got used this Twitter API in our Python code for Twitter corpus Retrieval related to “#DigitalIndia”. We were able to effectively retrieve ten thousand tweets from Twitter using our Python code. The amassed records had been in addition processed. factsalong side numbers, symbols, non english characters and further areas have been removed in order that the facts is first-class to perform records mining. records preprocessing is performed the use of NLTK 3.0 modules protected with Python code. project includes StopWords removal, Tokenization and Stemming.

Revised Manuscript Received on May 15, 2019.

M. SreeVani, Dept of Computer Science and Engineering, MGIT, Gandipet, Hyderabad, Telangana, India. (osreevani@gmail.com)

The tweets collected from twitter are a combination of urls, and one-of-a-kind non-opinional statistics like hashtags “#”, annotation “@” and retweets “RT”. To gather n-gram features, as established in determine 1, we first ought to tokenize the text input. Tweets pose a problem for stylish tokenizers designed for formal and everyday text.

three.2 teach a Random woodland Classifier  
Use an item clf to create a random forest classifier and furthermore educate the classifier as proven beneath.

```
1. clf =
    RandomForestClassifier(n_estimators=10000,random
        _state=0, n_jobs=-1)
2. clf.wholesome(X_train, y_train)
three.3 characteristic Extraction
```

Term Frequency is assigning weights with the aid of assuming that each time duration has a contribution that is proportional to the big form of its occurrences in files. To understand the maximum crucial features use selector object that in flip will use the random wooded place classifier to select out capabilities which have an importance of more than zero.15.sfm = SelectFromModel(clf, threshold=0.15)four. results and dialoguenside the take a look at, Random wooded location will be examined using severa term weighting method which includes Binary TF, uncooked TF, Logarithmic TF, and TF.IDF. The results may be seen in figure 1. The phrase ratings of the capabilities are tested based totally on Chi-rectangular technique. Then build frequency distribution of all words after which frequency distribution of terms internal wonderful and terrible labels.

**Table 1: Tweets-Opinion Analysis**

| Tweets Collected   | Dictionary                                      | Polarity Classification |
|--|---|-------------------------|
| Providing high speed internet is the ambitious plan of Reliance Group. Good going # DigitalIndia                                 | Positive Words<br>High<br>Ambitious<br>Good     | Positive                |
| @UIDAI plz fix d Aadhar android app SMS verification issue otherwise this will be alet down issue 4 @ DigitalIndia @NarendraModi | Negative Words<br>Letdown                       | Negative                |
| @Airtel_Presence 48 hrs landline dead no Internet no action.Is this the #DigitalIndia  | Negative Words<br>Landline Dead<br>no internert | Negative                |

Finally, the wide variety of tremendous and negative words in addition to the whole quantity of terms and the dictionary of phrase score based totally mostly on term Weighting method. As we are able to see in desk 1, tweets gathered are matched in opposition to first-class & poor

phrases used from Dictionary & then tweets are classified as notable & horrible. closing tweets are labeled as neutral.

Accuracy is the performance evaluation parameter and it's far calculated via wide variety of efficaciously decided on tremendous and terrible phrases divide with the resource of overall amount of phrases gift within the corpus. The approach is given as below.

$$Accuracy = \frac{\sum True Positive + \sum True Negative}{\sum Total number of words} \quad \text{Equation 1}$$

Where True positive is number of tweets recognized as positive and true negative is number of tweets recognized as negative respectively. Figure 1 describes opinion analysis based on Term Weighting method. Table 2 shows experimental results.

**Table 2: Experimental results**

| Opinion  | No.of samples | TP   | TN   |
|----------|---------------|------|------|
| Positive | 3000          | 2500 | 500  |
| Negative | 2500          | 1500 | 1000 |
| Neutral  | 1500          | 1000 | 500  |



**Figure 1: opinion analysis using Random Forest**

### 3. CONCLUSION

Opinion mining is completed to analyses the human beings’s opinion on DigitalIndia. Random wooded area classifier is used to research the opinion of the tweets. the kind consequences show that the netizens of India are having high superb opinion in the direction of the DigitalIndia. ultimately, it's miles essential to go through in thoughts that opinion are multidimensional object. Tweets can also have combined excessive top notch and horrible critiques. As in future art work, Opinion evaluation want to consist of negation managing and emphasis dealing with with a view to improve its class accuracy.

### REFERENCES

- Priya Ananthram, 2 month ago, Opinion Analysis of tweets. Retrieved from: <https://www.kaggle.com/priyaananthram/d/arathee2/demonetization-in-india-twitter-data/opinion-analysis-of-tweets>
- Xiaowen Ding, Bing Liu- “A Holistic Lexicon-Based approach to Opinion Mining,” Proceedings of the 2008 International Conference on Web Search and Data Mining, ACM, 2008
- Bo Pang, L.L.: Opinion mining and opinion analysis. Foundations and Trends in Information Retrieval January Volume 2 Issue 1-2, 1–94 (2008).
- Pak, A., Paroubek, P.: Twitter as a corpus for opinion analysis and opinion mining. In: Chair, .C.C., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M., Tapias, D. (eds.) Proceedings of the Seventh International Conference on



- Language Resources and Evaluation (LREC'10).
5. Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, "A Opinion strength detection in short Text ",Journal of the American Society for Information Science and Technology, 61(12), 2544–2558,2010
  6. Go, A., Bhayani, R., Huang, L.: Twitter opinion classification using distant supervision. *Processing* 150(12), 1–6 (2009).
  7. Sunil B. Mane, Kruti Assar, PriyankaSawant, & Monika Shinde.(2017). Product Rating using Opinion Mining.International Journal of Computer Engineering in Research Trends,4(5), 161-168.
  8. MostafaKaramibekr,Alia.Ghorbani, "A structure for opinions in Social Domains," Social Computing (SocialCom), International Conference, 2013
  9. Brendan O' Connor, RamnathBalasubramanyan, "From Tweets to polls:Linking Text Opinion to Public opinion time Series," Proceedings of the fourth International AAAI Conference on Weblogs and Social Media, 2010.
  10. M.LovelinPonnFelciah,R.Anbuselvi, "A Study on Opinion Analysis of Social Media Reviews," IEEESecond Conference on Innovations in Information Embedded & Communication Systems, 2015
  11. PrernaAwasthi, RanjanaRajnish, "Systematic Study on approach & tools used for opinion mining,". <https://www.scribd.com/doc/315237279/Opinion-Mining>
  12. AnantArora, ChinmayPatil, StevinaCorreia, "Opinion Mining: An Overview," International Journal of Advance Research in Computer and Communication Engineering, 2015