

# OPINION MINING ON PULWAMA ATTACK

Harshit Gupta<sup>1</sup>, Ritik Jain<sup>2</sup>, Sahiba<sup>3</sup>, Dilleshwar Pandey<sup>4</sup>

<sup>1</sup>Student, Computer Science & Engineering, ABES Institute of Technology Delhi, India

<sup>2</sup>Student, Computer Science & Engineering, ABES Institute of Technology Delhi, India

<sup>3</sup>Student, Computer Science & Engineering, ABES Institute of Technology Delhi, India

<sup>4</sup>Professor, Computer Science & Engineering, ABES Institute of Technology Ghaziabad, India

\*\*\*

**Abstract**— Social media has received greater attention these days. Public and private opinion is an extensive form of subjects that are expressed and unfold always through several social media. Twitter is the most trending social media platform that is gaining popularity. Twitter gives companies a quick and powerful way to research clients' perspectives toward the vital to success within the market place. Developing software for opinion evaluation is an approach to be used to computationally analyze degree of clients' perceptions. The key motive of this research paper is to design a sentiment evaluation, extracting a massive quantity of tweets. Results classify customers' perspective through tweets into good and bad, which is represented in a pie chart and flow chart.

**Keywords**— Hadoop Cluster, Tokenization, Opinion Mining, Twitter, Unstructured Data, Sentiment analysis.

## 1. INTRODUCTION

Micro scale blogging today has turned into an extremely famous specialized device among Internet users. Internet based social website gets a large number of tweets each day on assortment of imperative issues. Writers of those messages expound their opinion on collection of themes and examine current problems[4]. The analysis of these tweets can be utilized for Elections, Business, Product survey and so forth. Likewise sentiment analysis is one of the critical zone of analysis of tweets that can be exceptionally useful in basic leadership. Performing analysis of expansive surveys is less complex as compared to Sentiment Analysis on Twitter. This is on the grounds that the tweets are short and generally contain hash tags, slangs, emoticons etc. The challenge about which we need to perform sentiment analysis is submitted to the twitter API's which does further mining and gives the tweets identified with just that question. Twitter data is mostly unstructured i.e, utilization of shortened forms is high. Likewise it permits the utilization of emoticons which are immediate pointers of the creator's perspective regarding the matter. Tweet messages comprise of a timestamp and the client name. This time stamp is valuable for speculating the future pattern utilization of our venture. For doing twitter data analysis first data is gathered utilizing NIFI in HDFS. Tweets are preprocessed for expelling Null values and aimless images. After that HIVE can be used for twitter posts analysis [2].

## 2. PROBLEM STATEMENT

Tweet analysis can be valuable in understanding the sentiments behind a tweet which are in large numbers. The tweets produced can be helpful in numerous walks of life for analyzing the sentiments. In general, numerous tweets may be fake and thus having lesser sentimental importance behind them. With the utilization of Apache Hadoop we accelerated the procedure of breaking down the tweets into lexicons[3].

The conventional framework can't foresee and break the tweets using Hadoop and had some downsides such as:

- ❖ The framework did not made great utilization of Twitter API.
- ❖ It can't utilize machine learning algorithms.
- ❖ It can't utilize JSON results.
- ❖ It can't demonstrate the outcomes properly.
- ❖ It can't help in sentiment analysis of tweets.
- ❖ It can't make proficient utilization of big data.

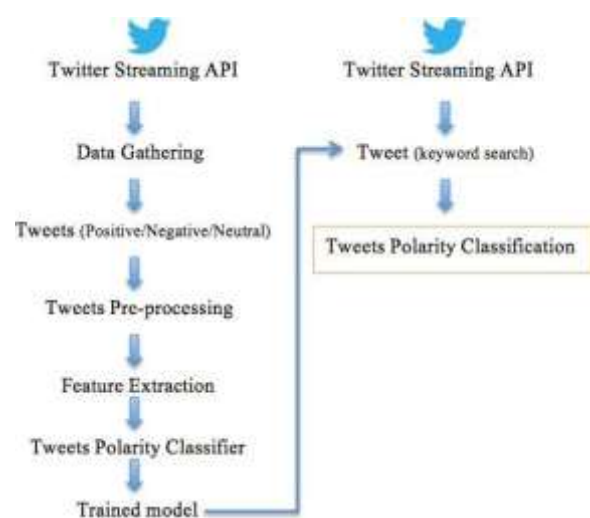


Figure 1: Twitter Analysis Steps[2]

## 3. NEED FOR SENTIMENT ANALYSIS

Sentiment analysis is greatly valuable in social media observing as it enables us to gather a review of the more extensive general assessment behind specific themes. The

utilizations of sentiment analysis are wide and intense. Works in sentiment on social media have been appeared to connect with moves in money market. The Obama organization utilized sentiment analysis to measure general feeling to approach declarations and battle messages in front of 2012 presidential race [6].

#### 4. SENTIMENT ANALYSIS

Sentiment is characterized as an articulation or opinion by an author about any protest or any perspective. Breaking down, examining, removing user's opinion, sentiment and inclinations from the abstract content is known as sentiment analysis. The fundamental focal point of sentiment analysis is parsing the content. In straightforward terms, sentiment analysis can be characterized as recognizing the polarity of the content. Polarity can be positive, negative or neutral. It is likewise alluded to as opinion mining as it infers opinion of the user. Opinions differ from user to user and sentiment analysis incredibly helps in understanding user's point of view. Sentiment can be, coordinate opinion as the name recommends the opinion around a protest is given straightforwardly and the opinion might be either positive or negative. For instance, "The video lucidity of the cell phone is poor" communicates a negative opinion [1].

Correlation opinion, it is a similar articulation which comprises of examination between two indistinguishable articles. The announcement, "The image nature of camera-x is superior to anything that of camera-y" is one conceivable case for communicating a relative opinion [2].

Sentiment analysis is performed at three unique levels:

- ❖ Sentiment analysis at sentence level recognizes whether the given sentence is abstract or goal. Analysis at sentence level expect that the sentence contains just a single opinion.
- ❖ Sentiment analysis at archive level arranges the opinion about the specific element. Whole archive contains opinion about the single protest and from the single opinionholder.
- ❖ Sentiment analysis at highlight level concentrates the component of a specific protest from the surveys and decides if the expressed opinion is certain or negative. The extricated highlights are then assembled and their abridged report is delivered [1].

#### 5. TOOLS & TECHNOLOGIES USED

##### 5.1 HADOOP

For pre-processing the large data in parallel crosswise over cluster of computers, Apache thought of an open source structure known as Hadoop. The significant segments of Hadoop are Hadoop Distributed File System (HDFS) and the MapReduce programming model. Hadoop is open source, since it keeps running on distributed

computing administrations or product machine crosswise over cluster of computers. Any number of computers can be added to the Hadoop group with the end goal to manage immense data in parallel. To every single computer data is dispersed and thus activity is performed in parallel in Hadoop group [5].

##### 5.2 MAP-REDUCE

The MapReduce empowers the composition of uses in a viable way and furthermore handling of colossal arrangements of data is productive with Hadoop MapReduce. The MapReduce has two distinct tasks:

**5.2.1 The Map Task:** The Map catches the information and this information data are partitioned according to similarity of data. This data is additionally separated into tuples to frame a key value pair.

**5.2.2 The Reduce Task:** The contribution to the Reduce errand is the yield from the Map task. All the partitioned tuples in the Map task is joined to frame small arrangement of tuples. Map Task is trailed by Reduce Task [5].

##### 5.3 HDFS

Hadoop has its own filesystem known as Hadoop distributed file system dependent on Google File System (GFS). The engineering of HDFS portrays the master-slave design. Master Node deals with the record framework and the capacity of real data is taken consideration by the slave Node [1].

A record in a HDFS namespace is partitioned into a few sections and these portions are put away in Data Nodes. The plotting of these fragments to the Data Nodes is recognized by the Name Node. The Data Node performs read and compose tasks.

##### 5.4 HIVE

Hive is an information distribution center foundation apparatus to process organized information in Hadoop. It dwells over Hadoop to condense Big Data, and makes questioning and examining simple. At first Hive was created by Facebook, later the Apache Software Foundation took it up and created it further as an open source under the name Apache Hive. It is utilized by various organizations. For instance, Amazon utilizes it in Amazon Elastic MapReduce [8].

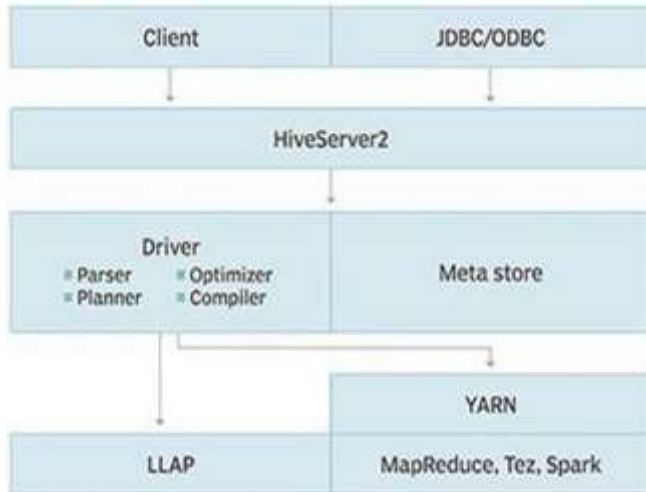


Figure 2: Architecture of Hive

### 5.5 NIFI

Apache NiFi is a real time information ingestion stage, which can exchange and oversee information exchange between various sources and goal frameworks. It bolsters a wide assortment of information groups like logs, geo location data, social feeds, etc. It likewise underpins numerous conventions like SFTP, HDFS, and KAFKA, and so forth. This support to wide assortment of information sources and conventions making this stage mainstream in numerous IT associations.

### 5.6 TABLEAU

Tableau is a powerful and quickest developing data representation tool utilized in the Business Intelligence Industry. It helps in improving raw data into the effortlessly reasonable configuration.

the data [3,7].

- ❖ Using this structured data we analyzed sentiments with the help of a BI tool i.e. Tableau.

## 7. CONCLUSION

As twitter posts are important source of opinion on various issues and subjects. It can give a sharp knowledge about a point and can be a good source of analysis. Analysis can help in basic leadership in different zones. Apache Hadoop is outstanding amongst other choices for twitter post analysis. When the framework is set up utilizing FLUME and HIVE, it helps in analysis of diversity of topics by simply changing the keywords in query.

Likewise, it does the analysis on real time data. The analysis what we did could be useful in breaking down individuals' sentiment. Opinion mining should likewise be possible on that data for finding polarity (Positive, Negative, Neutral) of tweets gathered.

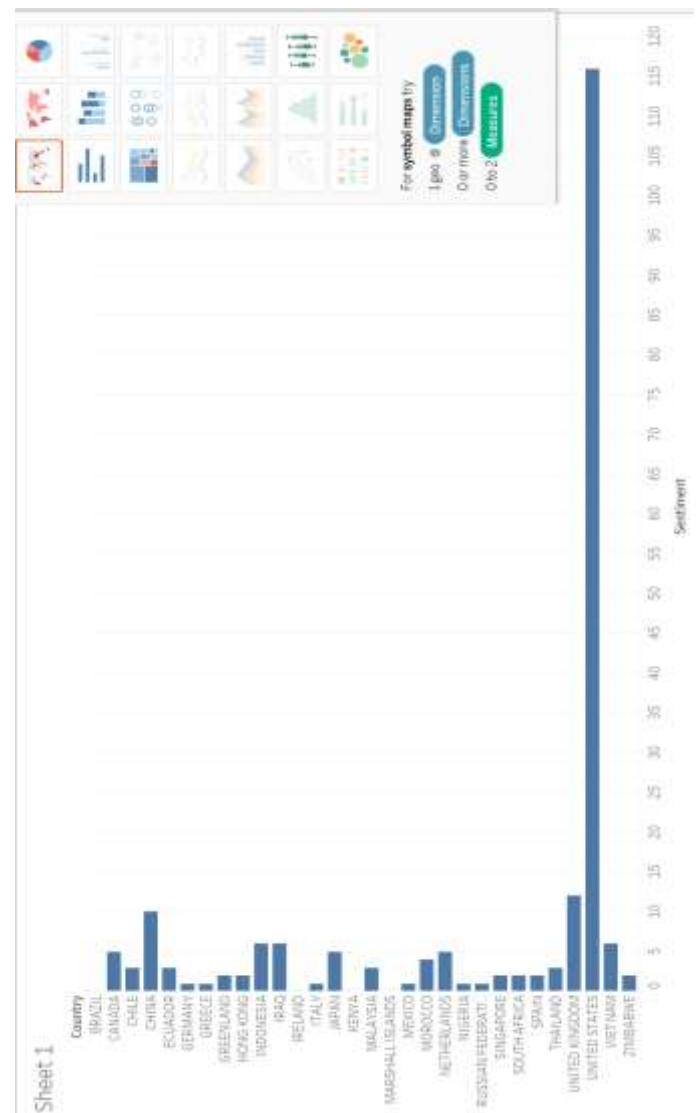
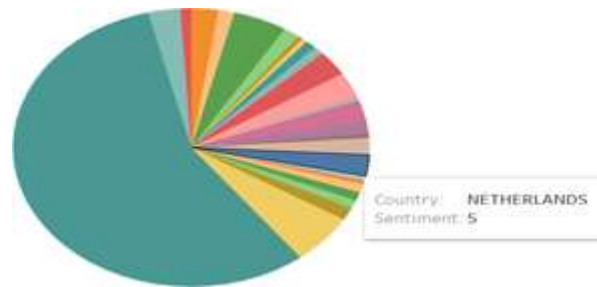


Figure 3: Observed output on Pulwama attack 2019



*Figure 4: Pie chart of observed outputs*

## 8. REFERENCES

1. Md. Nowraj Farhan, Md. Ahsan Habib and Md. Arshad Ali "A study and Performance Comparison of MapReduce and Apache Spark on Twitter Data on Hadoop Cluster". Published online on 08 |July-2018.
2. Harshal Kapase, Kalyani Galande, Tanmay Sonna, Deepali Pawar, Dipmala Salunke "A review on: Sentiment polarity analysis on Twitter data from different Events" Volume: 05 Issue:03 |Mar-2018.
3. Riya Bhatia, Prachi Garg, Rahul Joharic "Corpus based Twitter Sentiment Analysis" 3rdInternational Conference on Internet of Things and Connected Technologies (ICIoTCT), Jaipur(India) on 26 | March,2018.
4. Brinda Hegde, NagashreeH ,Madhura Prakash "Sentiment analysis of Twitter data: Amachine learning approach to analyse demonetization tweets" International Research Journal of Engineering and Technology (IRJET) ,Volume: 05 Issue: 06 |June-2018
5. Priyanshu Jadon, Miss Rupali Dave "A big data approach for Sentiment Analysis of Twitter Data using Naïve Bayes" International Journal of Innovations & Advancement in Computer Science. IJIACS ISSN 2347 - 8616 Volume 7, Issue: 04 |April 2018.
6. Mika V. Mäntylä, Daniel Graziotin, Miikka Kuuttila "The evolution of sentiment analysis—A review of research topics, venues, and top cited papers" M3S, ITEE, University of Oulu, Finland, Institute of Software Technology, University of Stuttgart, Germany, Issue : 21 | Nov-2017.
7. Amit Palve, Rohini D. Sonawane, Amol D. Potgantwar "Sentiment Analysis of Twitter Streaming Data for Recommendation using Apache Spark" Volume-5, Issue-3 | June 2017
8. Priya Gupta "Analysis of User Behavior for Twitter Posts on Hadoop" International Research Journal of Engineering and Technology (IRJET) Volume: 04 Issue: 05|May-2017
9. Jaba Sheela "A Review of Sentiment Analysis in Twitter Data Using Hadoop" International Journal of Database Theory and Application Vol.9, No.1 (2016).