Sentiment Analysis of Twitter User Data on Punjab Legislative Assembly Election, 2017

Akhilesh Kumar Singh
Department of Computer Science & Engineering, Dr.BR Ambedkar National Institute of Technology, Jalandhar
Email: akhilesh7210@gmail.com

Deepak Kumar Gupta and Raj Mohan Singh
Department of Computer Science & Engineering, Dr.BR Ambedkar National Institute of Technology, Jalandhar
Email: guptadk,@nitj.ac.in, rm_singh14@yahoo.com

Abstract—Sentiment Analysis is the way of gathering and inspecting data based on the personal emotions, reviews, and contemplations. The sentimental analysis is also recognized as opinion mining since it mines the major feature from people opinions. There are various social networking platforms, out of which Twitter is praised by lawmakers, academicians, and journalists for its potential political values. In literature, numerous studies have been performed on the data ecstatic to elections on Twitter. The greater part of them has been on the U.S Presidential Elections where there are two main applicants who fight it out. Since individuals discuss so many political parties and candidates and their prospects too in rendered messages, the issues of distinguishing their political feeling become extensive and fascinating. Consideration of all these aspects along with a sheer volume of data propelled us to look into the data and find interesting inferences in it.

To select the 117 members of the Punjab Legislative Assembly, Legislative Assembly election was held in Punjab, the State of India on 4 February 2017. As per the Election Commission, a total of 1.9 crore voters is eligible for voting in August 2016 in Punjab. The data set that is collected with the help of Twython was analyzed to find out trivial things and interesting patterns in the data. The central idea of this research paper is to carry out the sentiment analysis on Legislative Assembly election 2017 that was held in the Punjab, a state of India for the Political Parties like BJP, INC, and AAP. We have analyzed and fetch significant implications from the tweets gathered over the whole period of elections.

Index Terms—Sentiment Analysis, Twython, Textblob, NLTK, Twitter, NLP.

I. INTRODUCTION

To select the 117 members of the Punjab Legislative Assembly, Legislative Assembly election was held in Punjab a territory of India on 4 February 2017. These kinds of elections are held every five years when the sitting government is nearing its completion. As per election commission, in Punjab total of 1.9 crores, voters are eligible for voting in August 2016. Drug peddling is the first and primary concern in this election. Other election issues include unemployment & absence of skills, agriculturist crises, repeatedly deteriorating economy, hysterical crime and the function of goons frequently are the matters of concerns of the voter, violence hostile to Dalits and Dalit land issues in Sangrur area, the 1984 anti-Sikh riots and the delivery of drugs & obsession to them. Punjabi Non-resident Indians (NRIs) play a key character in elections.

A. Election and online social network platform

There was a significant change in 2017 Legislative Assembly election from the 2012 Legislative Assembly election. This was the revolution in the function performed by the social media throughout the elections. Mobile Internet user's statistics in India is probably going twofold and move beyond 300 million spots by 2017 from current 159 million users at hand, a new tale by Internet and Mobile Association of India (IAMAI) and consultancy firm KPMG. There is a range of social media platforms available, e.g., Facebook, Twitter. The public is gradually more using these platforms to articulate their views on different topics. India has been at the front line of the development stories in a number of users of these online social networking sites. In 2018, a total number of Facebook users in India are anticipated to reach 254.9 million, up from 195.2 million in 2016. Twitter position second in the number of users with 23.2 million monthly active users in the area, increase from 11.5 million in 2013.This is like taking an inspiration of the General Elections 2014 or not, but the social networking sites were used by the parties for their campaigns. Approximately all the foremost parties made their existence felt on the social sites with the social accounts and authenticated pages of their influential leaders and the parties.

Punjab has over 13.46 million internet users out of which 4.71 million are rural and 8.75 million users are urban. With the introduction of 3G and 4G networks in

the state, the number of internet users has grown manifold as compared to 2012 Assembly elections.

Social media is free and gives a chance for the people to have their say. There are more than 30 lakh internet users who fall within the age group of 18-35 years. At one hand, the internet has given a chance for the parties to reach voters; it has also become an instrument for common people to voice their concern.

The Facebook page of Aam Aadmi Party (AAP Punjab) has been liked by nearly 9 lakh followers. Apart from the videos of Bhagwant Mann, this page also has a video which shows Arvind Kejriwal giving third-degree torture to Deputy Chief Minister Sukhbir Singh Badal. The page also shows Sukhbir Badal liking AAP's FB page and an AAP missile chasing Congress vice-president Rahul Gandhi.

Similarly, the Congress page (Punjab captain) with 9.31 lakh followers also contains cartoons targeting either Badals or Arvind Kejriwal. The well-liked hashtags being used by the Congress are #AAP-Hi-SADhai (AAP is Akali Dal), #AAPPunjabVichSaaf (AAP has been cleaned in Punjab), #SADisBAD (Akali Dal is bad). Masking the faces in a Bollywood song, the party has also unofficially released a video showing Captain Amarinder Singh beating up Badals and Kejriwal.

For the election techies sitting in the election war rooms not only organize animations and funny jingle but in addition the cartoons and videos which are unofficially circulated using different apps like WhatsApp. There are more than a dozen such animations being dispersed on social media. We found Facebook, Twitter and WhatsApp sections or social media sections in the war rooms of the Congress and the Shiromani Akali Dal (SAD). The techies working by these parties not only regularly post content but also comment on every post, giving an idea to the organic user that the page, tweet or post is popular among the people.

AAP Punjab's social media war room head, Akashnoor Gary said they simply overlook such comments or content which the rivals use to target their party. The rivals do post offensive comments and employ techies to post replies and tweets but we ignore such comments. We do not have money to employ techies. Rivals are spending crore of rupees on social media campaigns and on war rooms. Interestingly, all the parties spending huge money on social and digital media campaigns refused to give details about the expenses, following the Election Commission capping.

The Communication approach revolved around the philosophy of fighting in opposition to corruption, promoting transparency and pursuing grassroots democracy both within and outside the organization. The content for this long drawn campaign has been articulated well, keeping in mind the symbolic icons of Aam Aadmi Party, the objective at hand and the elements of language communication.

The Communication approach for Congress focuses on the shortcomings of the incumbent government and the general issues of agriculture, water, electricity and drug menace in Punjab. The idea is to position Congress as a representative of change with Captain Amarinder Singh spearheading this vibrant political movement hence the position statements like Chahta Hai Punjab Congress ki Sarkar and Punjab Da Captain.

Punjab Congress handle has over 12k followers on Twitter with decent engagement numbers ranging from 50-70 RT's for original tweets, the tweets from INC India handle and that of Captain Amarinder get over 200+ RT's on an average.

B. Sentiment Analysis

Textual information can be of two kinds: facts and opinions. While facts are basically concerned with objective data transmission while the opinions articulate the sentiment of their authors. The goals of sentiment analysis are to unearth the thoughts of the author on a specific topic from the written text.

Sentiment Analysis is the way of gathering and inspecting data based on the person emotions, reviews, and contemplations. Sentiment analysis is also expressed as opinion mining since it mines the major feature from people opinions. Sentimental Analysis utilizes a variety of machine learning methods, Natural Language Processing (NLP) and statistical models for the mining of the characteristics of a large volume of data.

Sentiment analysis can be used for various purposes. When we analyze the feelings or thoughts of the people that they express in form of text on social media we can make the opinion about the people. Since sentiment Analysis is field centered so, results of one field cannot be applied to another field. The utilization of sentimental analysis is also in the several real life events like, to get reviews about any product for the online shopping sites or movies, to obtain the monetary document of any corporation, for forecasting or marketing. Twitter is a microblogging platform where anyone can read or write the short form of the message which is referred to as tweets. The volume of data amassed on Twitter could be very huge. The data is amorphous in nature and written in natural language. Sentimental Analysis of Twitter data is the way of gaining access to tweets for a specific subject matter and anticipate the sentiment of these tweets as positive, negative or neutral with the help of different system studying set of rules i.e. machine learning algorithm.

C. Tool Used

Python is a great programming language with top notch capability for processing linguistic facts. We selected Python because its syntax and semantics are opaque, and it has a noble string-managing capability. Python facilitates interactive exploration as an interpreted language. When Python is used as an object-oriented language it allows facts and techniques to be encapsulated and reused easily.

Python is a dynamic language because in Python variable name is only bound to an object. Also, it permits variables to be typed dynamically, expediting rapid improvement. Python includes the broad standard library, modules for graphical programming, numerical

processing, and web connectivity. The syntax and semantics of the Python are fresh and length of the code is comparatively shorter than other programming languages. It's amusing to work in Python because it permits you to consider the problem in preference to focusing on the syntax.

Python is broadly utilized in industry, clinical studies, and training around the sector. Python facilitates productivity, quality, and maintainability of software. This is the main reason it is getting praised by the various field.

II. The Goal of Paper

As the number of individuals making their view available on social media has dramatically increased in the last decade, the chance to use these data for understanding what individuals think prevails. In the past decade growths of social networking, websites have grown like Twitter, Facebook, Tumbler, and so on. Out of which one of the website which is extensively used all over the world is Twitter. As per Twitter statistics, it has been recorded that around 6000 tweets per second which correspond to 350000 tweets sent per minute and around 500 million tweets per day and roughly 200 billion tweets per year. Twitter permits individuals to express their belief, view, emotions, opinions, reviews, etc. regarding any topic in natural language within 140 characters.

Python is a high-level programming language which is best suited for natural language processing (NLP). Natural Language Toolkit is a Python library that is used for the natural language processing. NLTK offers a lot of corpora. These corpora support the user in training classifiers. This also allows a user to perform all NLP methodology like tokenizing, part-of-speech tagging, stemming, lemmatizing, parsing and performing sentiment analysis for known data sets.

It's far tough assignment to cope with a huge dataset, however, with using NLTK and TextBlob we are able to simply classify our data and provide greater accurate results based on different classifiers. The main objective of this research paper is to carry out sentiment analysis of Punjab Legislative Assembly Election, 2017 for various political parties of Punjab, like BJP-Akali Dal Alliance, AAP (Aam Aadmi Party), INC (Indian National Congress). The tweets of these parties are collected from Twitter and after that labeled into sentiments; i.e. tweets are positive or negative by using polarity and subjectivity parameter which is available in TextBlob. These outcomes provide the reviews of individuals on these local and national political parties. To achieve this goal, a module is created which can accomplish sentimental analysis

Various agencies and organizations want to know the customer's perception about the services or products that they deliver. This information is immensely valuable. With this information that is gained from an evaluation company can apprehend issues with their merchandise, spot trends earlier than their competitor, build better relations with their target spectators. In this research

paper, we stint on distinct political parties since in our country politics performs the special role. From this perspective sentiment analysis being carried out for these applications, so that outcomes will consent to a user to gain insight into how each party is being perceived by the community. The perception of the individual is too much valuable because with this perception they express their belief, reviews etc. Therefore this simply revolutionizes the feedback procedure.

III. Related Work

McLuhan (1964) [29] in his work laid down a distinction in the midst of old and new media, concluding that there is no fixed border between the two. He coined the terminology "the medium is the message", where he strongly asserted that new media is fed by older media. In nutshell, it can be summarized that the new medium of words consumed the older medium of sounds, and in a more modern context the newer medium of film consumed the older medium of the novel. In the same context, it can be added that "new media" inculcates older media and can be separated when the time arrives, and the existing media can be termed as "new media". In the present scenario, the newer media is overpowering the older media at a rapid rate and there is nothing and there is no stoppage to it. Due to the digitization of information such as speech, text, image and film, and the motive of connecting each one to everyone on the Internet, has made McLuhan's theories increasingly significant (Lifvergren, 2011).

Mcquail (2010), in his work, stated that digital media can be used to depreciate the unseen approach of escaping from the unfair "top-down" politics of mass democracies in which rigid organized political parties develop the strategy that is one sided and propel support behind them with least negotiation and grassroots input. "The most widely noted potential effect for the media institution is the convergence between all existing media forms in terms of their organization, distribution, reception and regulation" (p.138).

Garcia (2010), in his work, stated that interactive social media platforms (Facebook, LinkedIn, Delicious, Twitter, YouTube, Foursquare, Digg in…) tremendously altered the communication standard. Garcia (2010) briefed Chadwick Martin Bailey and iModerate Research Technologies in February 2010,works and concluded that 67% of consumers are more likely to buy from the brands they follow on Twitter, and 51% are likely to buy from a brand they follow on Facebook, clearly, indicating the influence of social media on their users.

Numerous researches have already been conducted in the area of sentiment analysis. The present day research is done by performing sentiment analysis on data collected by the end user from different social networking platforms like Facebook, Twitter, and Amazon and so on. Machine learning algorithms are the basis of research for sentiment analysis. The main focuses of these algorithms are to find whether given text is in support or in opposed and also recognize polarity of text.

Turney[28] in their work stated, a straightforward unsupervised machine learning algorithm for the purpose of classification. This classification algorithm classifies whether reviews are justified or not (Turney 2002).On the basis of the average semantic orientation of phrases in the review that contain adjectives or adverbs classification of a review is done. (Pang, Lee, and Vaithyanathan 2002[8]) were the first to work on sentiment analysis. Their foremost goal was to classify text by inclusive sentiment, not just by topic like they classify movie review as either positive or negative. They applied machine learning algorithm on movie review database and concluded that these algorithms outperform human produced algorithms. They used Naïve-Bayes, maximum entropy, and support vector machines machine learning algorithm. By exploring various factors, they found that classification of sentiment is very challenging. They demonstrated that supervised machine learning algorithms are the basis for sentiment analysis. Most of the unsupervised sentiment classification methods produce a prevailing or domain dependent opinion lexicon for words or opinion phrases.

Riloff and Wiebe [27], in their work, stated about subjectivity details as a part of their research. These details were then used by (Wiebe, Wilson and Cardie 2005) to perceive linguistic orientation. In this paper, the author introduced a bootstrapping process in which high precision classifiers make use of famous subjective vocabulary. This subjective vocabulary isolates subjective and objective sentences from a non-annotated text collection. Sentiment analysis of Twitter data is significantly different from normal sentiment analysis techniques. Common opinion lexicons cannot be used as replacement or alternatives for the informal language used on social communication platforms. Special emphasis is laid upon pre-processing due to certain unique properties of the Twitter language. The nature of Twitter messages, called tweets, cannot be associated with the structure of the current areas of studies.

Pak and Paroubek 2010, in their work, stated that tweets comprising emoticons are used as training corpus to evade manual annotation. The data were sorted into two categories: happy emoticons from the positively marked set and sad emoticons from the negatively marked set.

E. Loper, S. Bird [10], has developed Natural Language Toolkit (NLTK) library that comprises several program modules, the huge set of organized files, many tutorials, problem sets, several statistics module, and built-in machine learning classifiers, computational linguistics program and so on. Natural language processing is the key task of NLTK. Corpora are used for training of classifier which is already included in NLTK. Developers devise new modules and exchange them with a current module. This is the main reason why NLTK provide more structured programs and more sophisticated results for given data set.

H. Wang, D. Can, F. Bar, S. Narayana [11] in his work, propose a system that performs real-time analysis of public reactions to 2012 presidential elections in U.S. They accumulated these reactions from Twitter, a microblogging platform. Social networking site, Twitter is a platform where individuals share their observations, feelings, and opinions on any trending topic. In the reference to U.S. election huge amount of data is generated on the Twitter. These data help in creating a sentiment for every individual candidate as well as it also gave a prediction about which candidate can win the election. They also analyzed the effects of sentiment analysis on these various public events that are organized during the election. The author also demonstrated that this live sentiment analysis was very fast in contrast to outdated content analysis. Since previous content analysis method takes many days or up to certain weeks to complete the analysis process. The system they revealed analyzed the sentiment of whole Twitter data about the election, candidates, promotions, etc. and conveyed outcomes at a rapid rate. This system offers media, politicians, and researchers a new way which is timely operative, completely based on public opinion.

O. Almatrafi, S. Park, B. Chavan [12]), in their work stated a system which is based on location. They explained how Sentiment Analysis is accomplished by Natural Language Processing (NLP) and machine learning algorithms. This system extracts sentiment from a text unit which is from a specific location. They studied several functions of location-based sentiment analysis by using a data source in which facts can be extracted from different locations easily. Twitter, provides space of tweet location which can easily be accessed by a script. The facts (tweets) collected from the precise location can be poised for describing patterns and trends.

Their work is basically concerned with Indian general elections 2014. They collected 600,000 tweets over a period of 7 days for two political parties. To construct a classifier they used supervised machine learning approach, like the Naïve-Bayes algorithm. These algorithms can classify the tweets as either positive or negative. They recognized the views and opinions of users of these two political parties in different locations. They also plotted their finding on India map by using a Python library. An example of their results on tweets of BJP in 2014 is shown in Figure 2.1, which shows different locations in India where BJP got the positive review.

L. Jiang, M. Yu, M. Zhou, X. Liu, T. Zhao [13], generated a system that spots on target dependent classification. The system theme is based on Twitter in which a query is generated first; followed by classification of tweets as positive, negative or neutral sentiments. This classification is based on the query that comprises sentiment as positive, negative or neutral. Here, query sentiment served as the target. Since sometimes target assigned immaterial sentiments, hence to resolve these issues they use target autonomous approach. Another advanced methodologies that are used for classification, merely take tweet into deliberation. The methods explained by the author ignore correlated tweet since they classify tweets based on the present tweet.
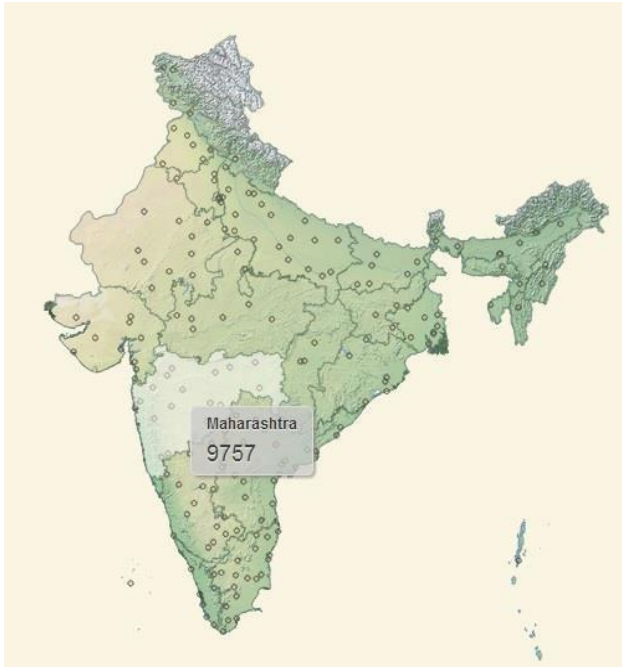
Fig.1. Positive tweets of BJP for different states in 2014

The tweets are small and generally vague. The present tweet solitary for sentiment analysis is not sufficient. The author suggested a system that expands target-dependent Twitter sentiment classification by:

1) Integrating target-dependent features, and
2) Taking related tweets into consideration.

As per their experimental results, this new improvement highly increases the performance and competence of target-dependent sentiment classification.

A. Pak, P. Paroubek [15] anticipated a way to decide positive, negative and neutral sentiments of the Twitter document by linguistic examination of collected corpus. They form a sentiment classifier. They have suggested a system for emoticons which is used in tweets. With the help of emoticons, they generate a corpus for emoticons. These emoticons are substituted with their relevant meaning so that they can pull out the characteristics from emoticons.

B. Sun, V. Ng, [16] has done many efforts to gather information from social networking platform. The purpose of the collection is to find out the sentiment of internet users. Their fundamental aspiration is to find sentiment analysis impacts from social media posts. The author also compares the result on several themes on diverse social networking sites. Every day huge volume of data in the form of messages, post feedback is generated; people are too interested in finding new similar individuals among them. Several researchers consider a number of likes and replies as the measures of impact for any post. However, they are not confident whether the effect is positive or negative in another post. Certain questions are elevated and novel approaches are developed in their research for the sentimental influence of post.

IV. PROPOSED WORK

The main goal of this paper is to accomplish the task of sentiment analysis on Legislative Assembly election 2017 that was held in the Punjab a state of India. We perform sentiment analysis for the Political Parties like BJP, INC, and AAP. We have collected tweets during an entire period of an election after that we analyze and fetch significant implications from the collection of tweets. Therefore we have created a classifier using supervised machine learning method and carry out live sentiment analysis on tweets gathered for different political parties.

A. Methodology

The proposed method for achieving our goal is as follows:

- First of all, in depth examination of current techniques in the domain of sentiment analysis is carried out.
- In second step we collected data from Twitter
- In third step pre-processing of data is performed, since gathered data is not in proper format.
- After that, a classifier is created with the help of Naive Bayes algorithm.
- Finally, we plot a graph which shows the polarity and subjectivity of the gathered data for different political parties.

V. SYSTEM ARCHITECTURE

Our main goal is to perform sentiment analysis for data gathered from Twitter. Hence we have created a classifier with the help of Naive Bayes algorithm. When we create the classifier it is trained and then it follows the steps shown in Figure 2
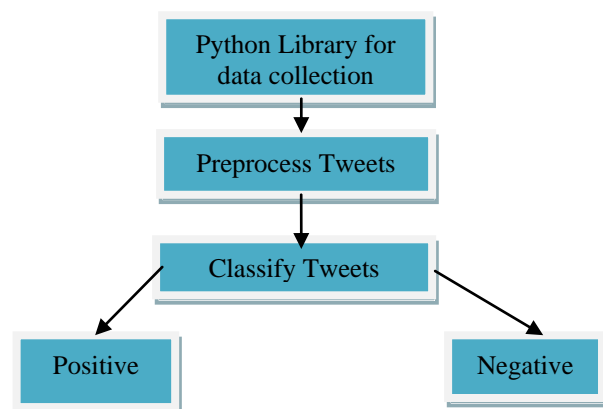

Fig.2. Classification of tweets

Step-1 Firstly we are going to collect tweets with the help of Twython library in python. We have downloaded the Python library on GitHub. Then we install Twython packages too. We have done installation very easily via pip or easy_ install.

Step-2 For the purpose of mining and extracting features from these tweets pre-processing is performed In a case of the social media, a language used by the users is

very casual. Users generate their particular words and spelling shortcuts and punctuation, misspellings, slang, new words, URLs, and genre vague terminology and abbreviations. Therefore such kind of text demands to be corrected.

Step-3 We classify the data as the positive and negative class by passing them into the classifier which is build using text blob. We have used Twitter as the main source of data for analysis. We have collected the tweets from Twitter in our database.

A. Twython

"Python is an interpreter, interactive, object-oriented programming language. It incorporates modules, exceptions, dynamic typing, very high-level dynamic data types, and classes. Python combines remarkable power with very clear syntax. It has interfaces to many system calls and libraries. . . "-Python website (https://docs.python.org/2/faq/general.html#what-is-python)

Python makes collecting Twitter data easy!

Twython is the leading Python library that provides access to Twitter data in a simple way. Twython provides support for Python 2.6+ and Python 3. Twython is widely used by companies, educational institutions and individuals and so on. Twython is a very vigorous, advanced library for Twitter. It has been preserved for over two years. With the support of Twython, we can get the user timeline. Twython also offers features of searching and fetching daily and weekly Twitter trends. Twython is a set of Python procedures built for the purpose of sending queries to the Twitter API and returning the results. For every REST API method, there is an equivalent Twython method. There are also other familiar packages that perform the similar task as Twython, such as Tweepy or Python-Twitter. In the present scenario, Twython is widely accepted by the user since it is dynamically enabled and contains all the features that are required by the user.
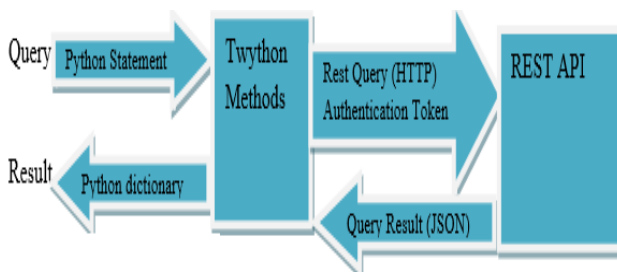

Fig.3. Twython Working

B. Data Collection

Twitter Data

In this section of the paper, we explain how to create a Twitter Application, how to authenticate a user, and how to make basic API calls. First, we go over to https://dev.twitter.com/apps and register an application! After the registration process, we get applications Consumer Key and Consumer Secret from the application

details tab. On the API Keys tab following information is to be shown:

- API key – the First key provided to the user.
- API Secret – The secret key is in the form of code.
- Access level – It is used to "Read, Write and Access direct messages".
- Owner – It gives information about the Twitter username.
- Owner ID – It is Twitter Id in number form. Not important to you. To generate the access token we made API request on Twitter with the help of Twitter account. After that following information is shown:

- Access token
- Access token secret
- Access level
- Owner
- Owner ID

C. Data Storage

When we start collecting our data from Twitter API our subsequent phase is to store that data so that we can use it for sentiment analysis. We ran our scripts for a couple of months and collect the tweets for different political parties. Every time we ran the script described in figure a JSON file is created which comprises tweets that are extracted from Twitter API. We have used .ODT format for our collected data files because data consists of many fields. .ODT separates each field with a comma, thus making it very easier to access the particular field which consists of text.

When we start collecting our data from Twitter API our subsequent phase is to store that data so that we can use it for sentiment analysis. We ran our scripts for a couple of months and collect the tweets for different political parties. Every time we ran the script described in figure a JSON file is created which comprises tweets that are extracted from Twitter API. We have used .ODT format for our collected data files because data consists of many fields. ODT separate each field with a comma, thus making it very easier to access the particular field which consists of text.

We create separate directories to collect tweets of different political parties for respective month. We store them in our hard drive from where these can be simply imported to our snippet and additional analysis can be performed. When we stored our tweet we have to pre-process the data in advance before applying it to classifier since the data we collect from Twython library is not fit for analysis. Hence pre-processing the data is our subsequent step.

D. Data Pre-Processing

Tweets collected using Python library is not appropriate for extracting features. Generally, messages in conjunction with usernames, bare spaces, special characters, stop words, emoticons, abbreviations, hashtags, time stamps, URL's, etc. are included in the

tweets. Therefore, NLTK is using for the pre-processing of the data. When we perform the preprocessing of data we first fetch the key message from the tweet, after that elimination of all blank spaces, stop words (like is, a, an, when, them, etc.), hashtags, repeating words, URL's, and so on. While we are finished with it, we're ready with processed tweet that is given to classifier for required result. A model processed tweet is shown in Table 1

Table 1. Primitive Tweet and Processed Tweet

Tweet Type	Result
Primitive tweet	@ArvindKejriwal: Goa and Punjab will create history
Processed tweet	Arvind, Kejriwal, Goa, Punjab, create, history

Several syntactic features that may not be used for analysis are also included in the tweets. Hence, cleaning of Twitter data is necessary. The pre-processing of data is performed such that we characterized the tweets simply in terms of words so that simply classify the class. A code which is developed in Python will be used to get processed tweet.

E. Classification

For classification of tweets in the different class (positive and negative) we have created a classifier using Naive Bayes algorithm. It is a machine learning algorithm. For the development of classifier, we used a library of Python called, Textblob. The text blob.sentiments module comprises two sentiment analysis implementations, Pattern Analyzer (based on the pattern library) and NaiveBayesAnalyzer.

F. Extracted Features

When we ran the script shown in Figure 4.4, it will extract the features from the training data and also apply a Boolean value to each attribute. Classification problem is considered as sentiment analysis task. Fetching and selecting the text features are the first step in the sentiment classification problem. Some selected characteristics of text are

- **Terms presence and frequency:** Terms presence are the unique features of words or word n-grams whose frequency is counted. It helps in finding the words binary weighting. It can also use term frequency weights to indicate the significance of features.
- **Parts of speech (POS)** In order to complete an opinion and give it a proper meaning, appropriate adjective need to be found.
- **Opinion words and phrases:** The words used to express viewpoints or judgments like good or bad hate or like come under the category of opinion words. Phrases, on the other hand, express opinions or judgments without making use of opinion words.

- **Polarity:** By studying the existing frequency of the word, in a largely explained corpus of texts, the polarity of a word can be recognized. The polarity of the word is considered to be positive if the word occurs more frequently among positive texts and is negative if it occurs in the negative text. If the frequencies are equal, then it is a neutral word.

VI. RESULTS

The polarity of tweets can be expressed at different levels whether the expressed opinions in a document or sentence is either positive or negative. The subjectivity of tweets is basically finding of subjective words and text that show the presence of opinions. In the result shown in Table 2 for various parties what we have seen AAP party tweets shows more subjectivity followed by INC and BJP-Akali Dal Alliance. Similarly, AAP has shown more polarity as compared to other two parties. Here polarity of AAP is tending towards negativity i.e. AAP posted more negative tweets as compared to other two parties. From our classifier result, we also conclude that INC has posted less number of negative tweets. This is examined after the calculation of positive and negative sentiment which is done with the help of TextBlob.

Table 2. Sentiment Analysis for BJP-Akali Alliance, APP, and INC for April 2016

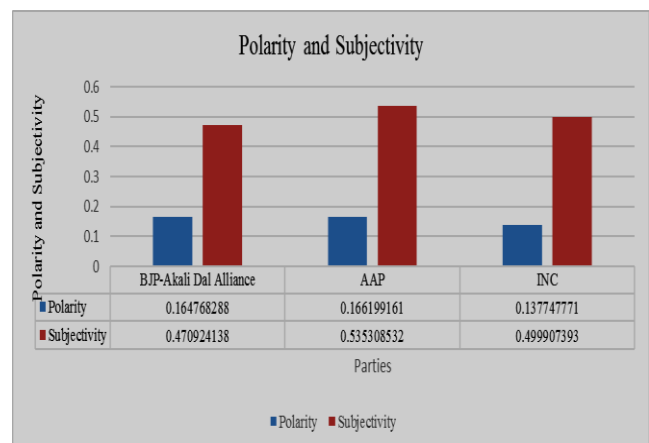| Party	Polarity	Subjectivity
BJP-Akali Dal Alliance	0.16476828792795484	0.470924138372633
AAP	0.16619916123762276	0.5353085319239169
INC	0.13774777079252399	0.4999073930855064

Fig.4. Polarities and Subjectivity

VII. CONCLUSIONS

We performed the sentiment analysis of various political parties, and analyzed the result, and found that

AAP has shown more polarity i.e. tweets expressed more emotions (either positive or negative). The positive reviews express greater polarity than the negative reviews. From the result, we also conclude that AAP has more negative tweets as compared to other two parties (i.e. INC and BJP-Akali). Our classifier which is built in Python is not only limited for use by political parties but it can be utilized for any rationale centered on tweets. Our classifier can also be used for news, finance, marketing, reviewing and much more. Even the actual results have also shown this. From the LIWC result what we found is that all the parties show more positive emotions as compared to negative words in the tweets. The work can be extended in the following areas:

- We can perform deep sentiment analysis of text, in different areas of application. It is not adequate to say that a text is an inclusive positive or inclusive negative. Users would like to know which separate topics are talked about in the text, which of them are positive and which are negative. So, there will be an inclination towards greater use of NLP techniques (such as syntactic parsing), in addition to machine learning methods.

- A more elaborate web-based application can be made for our work in future.

- We can improve our system that can deal with multiple local languages.

- By using various classification strategies we further improve the results.

- By the use of sentiment analysis, we forecast the future consequences or at least anticipate them better, when people tweet about present scenario.

REFERENCES

[1] H. Zang, "The optimality of Naïve-Bayes", Proc. FLAIRS, 2004

[2] C.D. Manning, P. Raghavan, and H. Schütze, "Introduction to Information Retrieval", Cambridge University Press, pp. 234-265, 2008

[3] A. McCallum and K. Nigam, "A comparison of event models for Naive Bayes text classification", Proc. AAAI/ICML-98 Workshop on Learning for Text Categorization, pp. 41-48, 1998

[4] M. Schmidt, N. L. Roux, and F. Bach, "Minimizing Finite Sums with the Stochastic Average Gradient", 2002

[5] Y. LeCun, L. Bottou, G. Orr and K. Muller, "Efficient BackProp", Proc. In Neural Networks: Tricks of the trade 1998.

[6] T. Wu, C. Lin and R. Weng, "Probability estimates for multiclass classification by pairwise coupling", Proc. JMLR-5, pp. 975-1005, 2004

[7] "Support Vector Machines" [Online], http://scikitlearn.org/stable/modules/svm.html#svm-classification, Accessed Jan 2016

[8] P. Pang, L. Lee and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques", Proc. ACL-02 conference on Empirical methods in natural language processing, vol.10, pp. 79-86, 2002

[9] P. Pang and L. Lee, "Opinion Mining and Sentiment Analysis. Foundation and Trends in Information Retrieval", vol. 2(1-2), pp.1-135, 2008

[10] E. Loper and S. Bird, "NLTK: the Natural Language Toolkit", Proc. ACL-02 Workshop on Effective Tools and Methodologies for teaching natural language processing and computational linguistics, vol. 1,pp. 63-70, 2002

[11] H. Wang, D. Can, F. Bar and S. Narayana, "A system for real-time Twitter sentiment analysis of 2012 U.S.presidential election cycle", Proc. ACL 2012 System Demonstration, pp. 115-120, 2012

[12] O. Almatrafi, S. Parack, and B. Chavan, "Application of location-based sentiment analysis using Twitter for identifying trends towards Indian general elections 2014". Proc. The 9th International Conference on Ubiquitous Information Management and Communication, 2015

[13] L. Jiang, M. Yu, M. Zhou, X. Liu and T. Zhao, "Target-dependent Twitter sentiment classification", Proc. The 49th Annual Meeting of the Association 39 for Computational Linguistics: Human Language Technologies, vol. 1, pp. 151-160, 2011

[14] C. Tan, L. Lee, J. Tang, L. Jiang, M. Zhou and P. Li, "User-level sentiment analysis incorporating social networks", Proc. The 17th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 1397- 1405, 2011

[15] A. Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining", vol. 10, pp. 1320-1326, 2010

[16] B. Sun and TY. V. Ng, "Analyzing the Sentimental influence of Posts on Social Networks", Proc. The 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design, 2014

[17] A. Go, R. Bhayani and L. Huang, "Twitter sentiment classification using distant supervision", CS224N Project Report, Stanford, vol.1-12, 2009

[18] A. Barhan and A. Shakhomirov, "Methods for Sentiment Analysis of Twitter Messages", Proc.12th Conference of FRUCT Association, 2012

[19] T. Mitchell, "Machine Learning", McGraw-Hill, 1997 [20] F. Jensen, "An Introduction to Bayesian Networks", Springer, 1996

[20] T. C. Peng and C. C. Shih, "An Unsupervised Snippet-based Sentiment Classification Method for Chinese Unknown Phrases without using Reference Word Pairs". IEEE/WIC/ACM Int. Conf. on Web Intelligence and Intelligent Agent Technology, vol. 3, pp. 243-248, 2010

[21] R. Feldman, "Techniques and applications for sentiment analysis", Proc.ACM, pp. 56-82, 2009

[22] N. Cristianini and J. Shawe-Taylor, "An Introduction to Support Vector Machines and Other Kernel-based Learning Methods", Cambridge University Press, March 2000

[23] "An Introduction to Python", v3.4.1, 2015 [Online], Available:https://docs.python.org

[24] http://indiatoday.intoday.in/story/punjab-assembly-political-parties-socialmedia/1/869440.html

[25] http://www.livemint.com/Industry/VThUq5I4BivpTDZdQb5sNN/Mobile-Internet-users-in-India-to-double-by-2017-says-study.html

[26] Conference on Design Education (Connected 2010), July 2010, UNSW, Sydney, Australia.

[27] E. Riloff, J. Wiebe, "Learning Extraction Patterns for Subjective Expressions" Proc. EMNLP-03 Conference on Empirical Methods in Natural Language Processing

[28] P. Turney "Thumbs up or thumbs down?" Semantic orientation applied to unsupervised classification of reviews. In Proceedings of the 40th annual meeting of the Association for Computational Linguistics, pp. 417424, 2002.

[29] M. McLuhan, "The Medium is the Message" Understanding Media: The Extensions of Man,1964

[30] Akshi Kumar, Teeja Mary Sebastian,"Sentiment Analysis: A Perspective on its Past, Present, and Future", IJISA, vol.4, no.10, pp.1-14, 2012. DOI: 10.5815/ijisa.2012.10.01

[31] Lopamudra Dey, Sanjay Chakraborty, Anuraag Biswas, Beepa Bose, Sweta Tiwari,"Sentiment Analysis of Review Datasets Using Naïve Bayes' and K-NN Classifier", International Journal of Information Engineering and Electronic Business(IJIEEB), Vol.8, No.4, pp.54-62, 2016. DOI: 10.5815/ijieeb.2016.04.07

Deepak Kumar Gupta is Associate Professor at the Department of Computer Science & Engineering, Dr.BR Ambedkar National Institute of Technology, and Jalandhar. He has total 29 years of including teaching and industries. In research, his current interests include Social Media, Data Analytics, and Operating System

Raj Mohan Singh received his M Tech degree in Computer Science and Engineering from Dr. B R Ambedkar National Institute of Technology, Jalandhar. His research interests include Cloud Computing,Operating Systems Distributed Computing, and Data Analytics.

Authors' Profiles

Akhilesh Kumar Singh is doing M.Tech in Computer Science and Engineering from, Dr.BR Ambedkar National Institute of Technology, and Jalandhar. He is currently working as a scholar in Social Network Analysis.

How to cite this paper: Akhilesh Kumar Singh, Deepak Kumar Gupta, Raj Mohan Singh," Sentiment Analysis of Twitter User Data on Punjab Legislative Assembly Election, 2017", International Journal of Modern Education and Computer Science(IJMECS), Vol.9, No.9, pp. 60-68, 2017.DOI: 10.5815/ijmecs.2017.09.07