

SENTIMENTAL ANALYSIS USING TWEETS

Kingston W

200701126

ABSTRACT

The objective of this task is to detect hate speech in tweets. For the sake of simplicity, we say a tweet contains hate speech if it has a racist or sexist sentiment associated with it. So, the task is to classify racist or sexist tweets from other tweets. The preprocessing of the text data is an essential step as it makes the raw text ready for mining, i.e., it becomes easier to extract information from the text and apply machine learning algorithms to it. If we skip this step then there is a higher chance that you are working with noisy and inconsistent data. The objective of this step is to clean noise those are less relevant to find the sentiment of tweets such as punctuation, special characters, numbers, and terms which don't carry much weightage in context to the text . It is a Natural Language Processing Problem where Sentiment Analysis is done by Classifying the Positive tweets from negative tweets by machine learning models for classification, text mining, text analysis, data analysis and data visualization . Sentiment analysis deals with identifying and classifying opinions or sentiments expressed in source text. Social media is generating a vast amount of sentiment rich data in the form of tweets, status updates, blog posts etc. Sentiment analysis of this user generated data is very useful in knowing the opinion of the crowd. Twitter sentiment analysis is difficult compared to general sentiment analysis due to the presence of slang words and misspellings. The maximum limit of characters that are allowed in Twitter is 140. Knowledge base approach and Machine learning approach are the two strategies used for analyzing sentiments from the text. In this project, we try to analyze twitter posts about electronic products like mobiles, laptops etc using the Machine Learning approach. By doing sentiment analysis in a specific domain, it is possible to identify the effect of domain information in sentiment classification. We present a new feature vector for classifying the tweets as positive, negative and extract peoples' opinion about products.

NEED FOR SENTIMENT ANALYSIS

1.Applications that use Reviews From Websites

Today the Internet has a large collection of reviews and feedback on almost everything. This includes product reviews, feedback on political issues, comments about services, etc. Thus there is a need for a sentiment analysis system that can extract sentiments about a particular product or services. It will help us to automate the provision of feedback or rating for the given product, item, etc. This would serve the needs of both the users and the vendors.

2. Applications as a Sub-component Technology

A sentiment predictor system can be helpful in recommender systems as well. The recommender system will not recommend items that receive a lot of negative feedback or fewer ratings. In online communication, we come across abusive language and other negative elements. These can be detected simply by identifying a highly negative sentiment and correspondingly taking action against it.

3. Applications in Business Intelligence

It has been observed that people nowadays tend to look upon reviews of products which are available online before they buy them. And for many businesses, the online opinion decides the success or failure of their product. Thus, Sentiment Analysis plays an important role in businesses. Businesses also wish to extract sentiment from the online reviews in order to improve their products and in turn their reputation and help in customer satisfaction .

NEED FOR SENTIMENT ANALYSIS

4. Applications across Domains

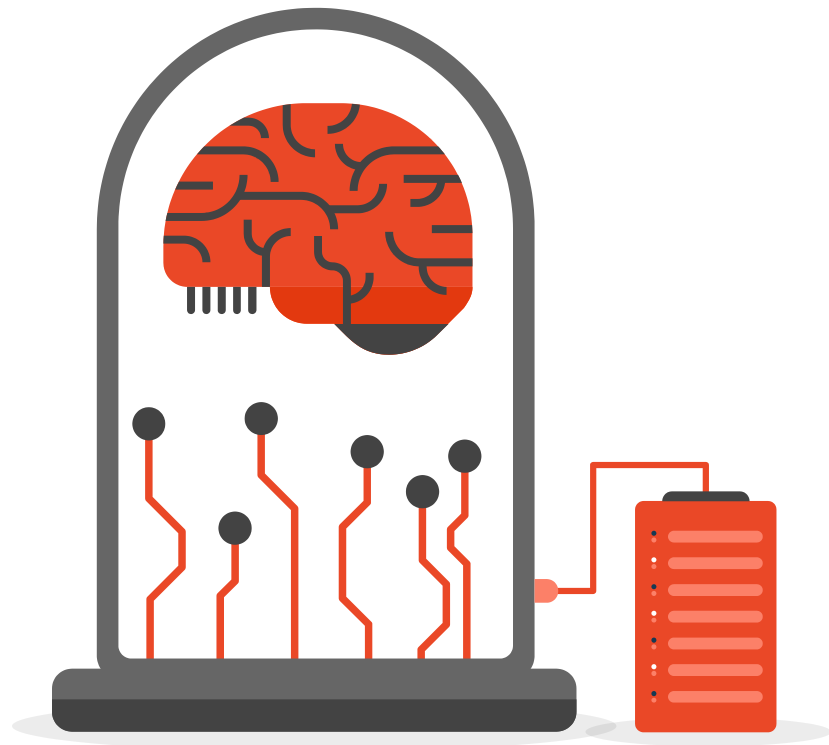
Recent Researches in sociology and other fields like medical, sports have also been benefited by Sentiment Analysis that show trends in human emotions especially on social media.

5. Applications In Smart Homes

Smart homes are supposed to be the technology of the future. In future entire homes would be networked and people would be able to control any part of the home using a tablet device. Recently there has been a lot of research going on the Internet of Things(IoT). Sentiment Analysis would also find its way in IoT. For example, based on the current sentiment or emotion of the user, the home could alter its ambiance to create a soothing and peaceful environment. Sentiment Analysis can also be used in trend prediction. By tracking public views, important data regarding sales trends and customer satisfaction can be extracted.

ADVANTAGES OF THE SENTIMENT ANALYSIS

1. Discover new market trends
2. Design targeted marketing campaigns
3. Track brand awareness
4. Develop a personalized brand experience
5. Conceive new product ideas
6. Discover and use influencer marketing
7. Develop more engaging online content
8. Monitor competitors

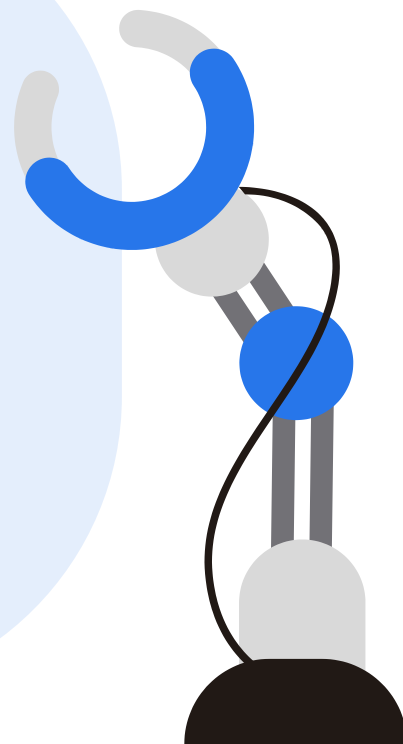


LITERATURE SURVEY

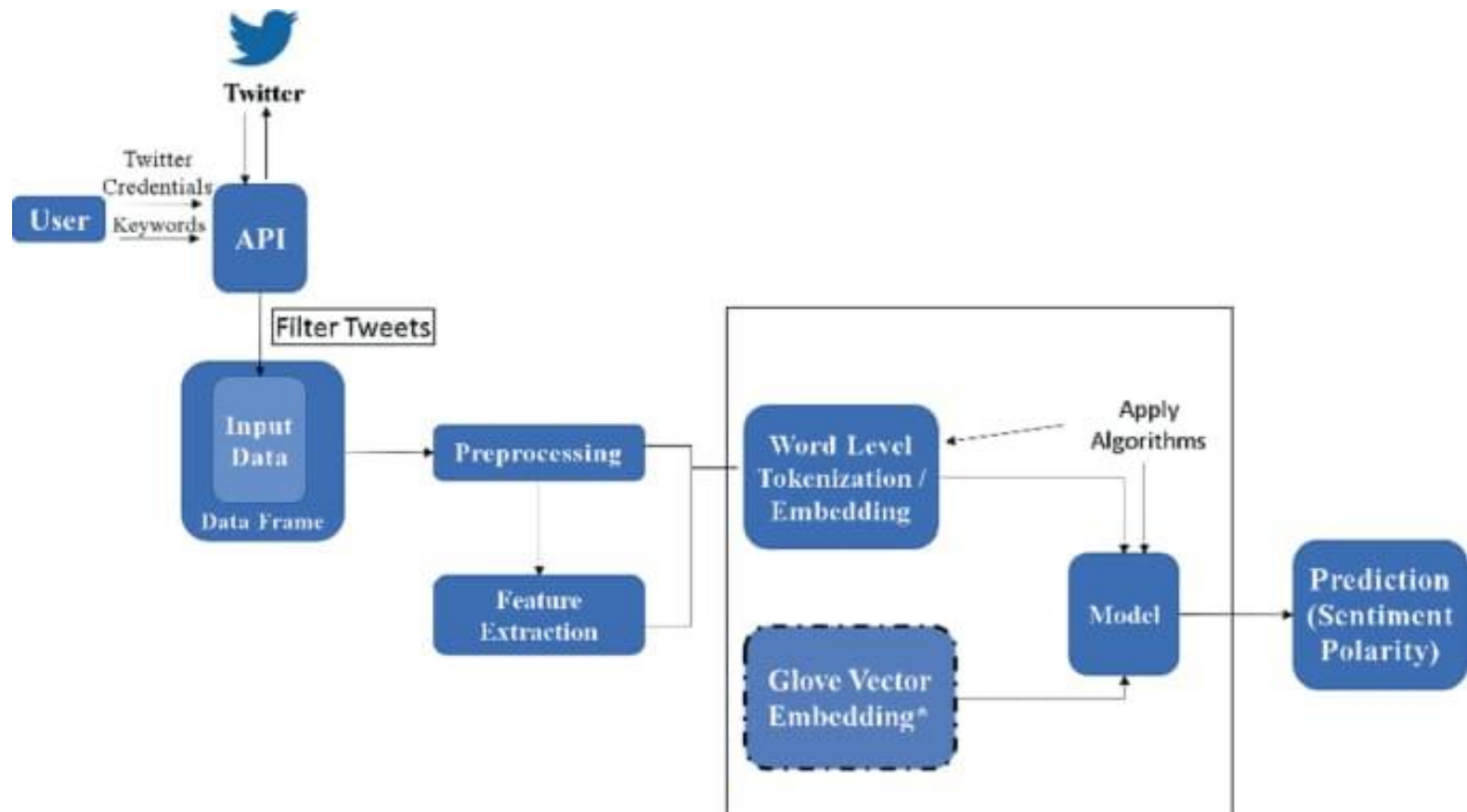
- **(Kaur, 2015)** describe regarding geographic area flood information set collected from twitter and realize the opinion of individuals. They used Naive Bayes formula for the classification of information and result they got 67% accuracy. They need collected several resolution from the individuals that are useful for each government and non-government organization to handle such scenario in an exceedingly higher manner. These strategies simpler than lexicon-based formula.
- **(Paul, 2017)** describe regarding the ultimate match of Indian premier league sport event 2015. Objective of this project to research standardity {the recognition} of IPL match and that player are popular and that team is dominate. They need used Hadoop and Map cut back artificial language. They got result like MS Dhoni is most talked regarding player and metropolis Indians team fairly dominated. This technique gave higher result.
- **(Mittal, 2016)** describe the requirement and impact of the sentiment analysis on on-line platform. They need additionally bestowed a listing of sentiments of emotions, interjections and comments that are extracted from posts and standing updates. They need got result to knowing whether or not {the on-line the web the net} reviews and posts are being useful to client or not and that on-line websites being most popular by the purchasers.

METHODOLOGY

- Sentiment analysis is perhaps one of the most popular applications of NLP, with a vast number of tutorials, courses, and applications that focus on analyzing sentiments of diverse datasets ranging from corporate surveys to movie reviews.
- The key aspect of sentiment analysis is to analyze a body of text for understanding the opinion expressed by it. Typically, we quantify this sentiment with a positive or negative value, called **polarity**. The **overall sentiment** is often inferred as *positive*, *neutral* or *negative* from the sign of the polarity score.



PROJECT ARCHITECTURE DIAGRAM



MAIN OBJECTIVE

Social media is generating a large volume of sentiment rich data in the form of tweets, status updates, blog posts, comments, reviews, etc. Moreover, social media provides an opportunity for businesses by giving a platform to connect with their customers for advertising. People mostly depend upon user generated content over online to a great extent for decision making. For e.g. if someone wants to buy a product or wants to use any service, then they firstly look up its reviews online, discuss it on social media before making a decision. The amount of content generated by users is too vast for a normal user to analyse. So there is a need to automate this, various sentiment analysis techniques are widely used. Sentiment analysis (SA) tells users whether the information about the product is satisfactory or not before they buy it. Marketers and firms use this analysis data to understand about their products or services in such a way that it can be offered as per the user's requirements. Textual Information retrieval techniques mainly focus on processing, searching or analysing the factual data present. Facts have an objective component but there are some other textual contents which express subjective characteristics. These contents are mainly opinions, sentiments, appraisals, attitudes, and emotions, which form the core of Sentiment Analysis (SA)

SYSTEM REQUIREMENTS

COMPONENTS	SPECIFICATION
PROCESSOR	Intel Core i5
RAM	8 GB RAM
GPU	NVIDIA GeForce GTX 1650
MONITOR	15" COLOR
HARD DISK	512 GB
PROCESSOR SPEED MINIMUM	1.1 GHz

FUNCTIONAL DESCRIPTION

A Twitter sentiment analysis identifies negative, positive, or neutral emotions within the text of a tweet. It is a text analysis using natural language processing (NLP) and machine learning. It identifies and extracts subjective information from original data, providing a company with a better understanding of the social sentiment of its brand, product, or service.

Positive words	love, great, good, thanks
Negative words	hate, shit, hell, tired
Emoticons	<i>P</i> (positive emoticon), <i>N</i> (negative emoticon)
Other	for, <i>U</i> (URL)

OUTPUT

⌨ How people are reacting on Russia war by analyzing 100 tweets.

General Report:
Weakly Negative

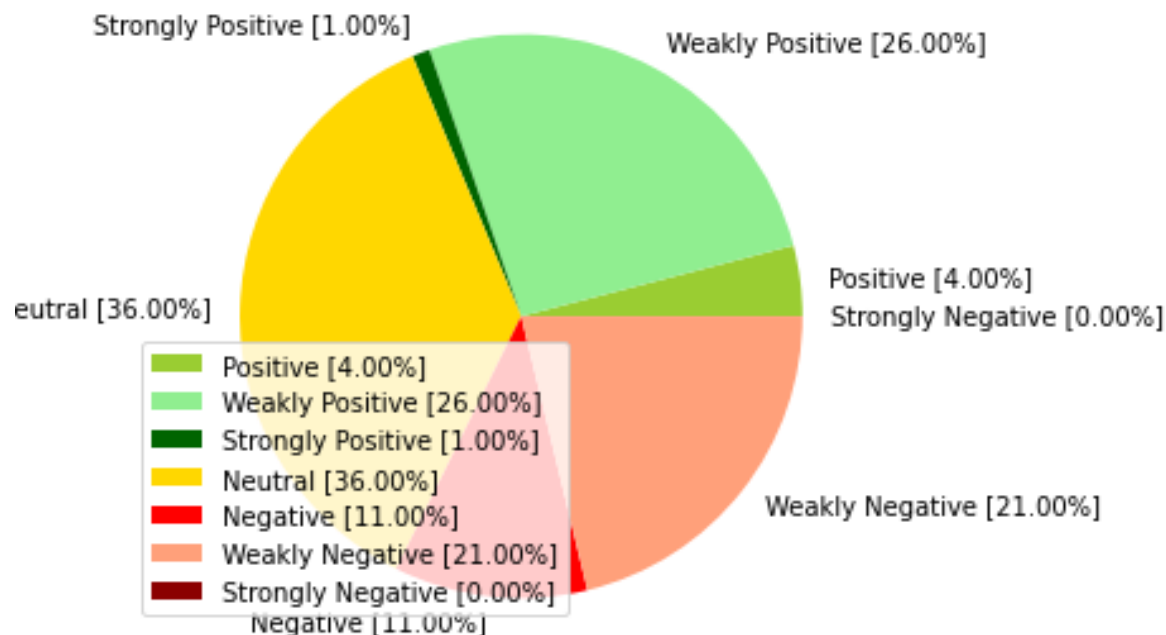
Detailed Report:
4.00% people thought it was positive
26.00% people thought it was weakly positive
1.00% people thought it was strongly positive
11.00% people thought it was negative
21.00% people thought it was weakly negative
0.00% people thought it was strongly negative
36.00% people thought it was neutral

OUTPUT

```
array[[ believe russia want war west mean u backed government could negotiated peace boris johnson stopped ask ,
'ntft nftart nftartist opensea poseidon doomsday weapon russia war nuclear wambo kun',
'version peace relinquishing part ukraine russia already engaged heinous war crime leftist bending backwards make excuse putin',
'yes case missed little miss genocide apologist russia war west',
'three russian tank explode flaming inferno ukrainian artillery strike war news russia ukraine ukrainenews',
'ukraine war map ukraine successfully attack russian fleet russia launch new donbas attack war news russia ukraine ukrainenews britain',
'biden asks covid amnesty afghanistan pullout amnesty gas price amnesty inflation amnesty student loan amnesty war russia amnesty nuclear armageddon amnesty weaponizing fbi political enemy amnesty via',
'good afternoon mr musk activist russia currently preparing project dedicated preventing nuclear war happening reducing military sentiment world make video interview opinio project',
'economy stronk exchange rate gud russia invasion ukraine central bank expected economy grow much mean war ha cost country gdp lost output year alone',
'russia export arm africa first africa population wa lost civil war turn interfere wheat export ukraine depleting africa population starvation russia seems get africa easily',
'yanno many people would homeless hungry biden put half money gave ukraine america instead remember street protesting war russia',
'use word build invasion wa going month tried talk month russia one attacked russia beat drum war russia started war russia fault',
'mad attempt talk child top natsec military leader saying publicly month goal instigate proxy war hope crippling russia military economy',
'cooperative even supportive russia putin illegal war',
'russian soldier flee ukrainian tank canada germany unitedstates ukrainewar ukraine',
'russia committing illegal war crime',
'voshchevska usa spends military budget nd th country biggest military budget world country spends much money war st usa g nd china g rd russia g',
'russia seek drone attack probe guarantee resume grain deal aljazeeraenglish via putin war management like businessman brutal militarily attacked want investigation like cerebral epileptic seizure stake',
'think fiona hill quite naive rd world war soon engulf asia middle east europe china iran silent',
'putin strategy clear oct ukraine daily update war news russia ukraine ukrainenews',
'russian missile strike continue vital infrastructure ukraine war news russia ukraine ukrainenews britain',
'russia hoax wa deliberate effort military industrial complex undermine trump platform lowering tension russia right brink nuclear war russia ukraine anyone belief otherwise sucker',
'nato already stated goal wipe russian race map russia forced take mean make sure achieved invading neighboring state sharing border nato expanding war fought russia',
'v rusiji se bojijo da bo veliko njihovih portnikov kon alo kariere russia russland sport urheilul war krieg vojna russiaukrainewar',
'k cook yes judged different standard russia war crime israel also war crime say continue analysing hypocrisy west',
'based opinion mass graf civilian russia ha leaving assorted war crime russia ha committing',
'turkey doe border russia country involved complicates thing amp brings another war putting gasoline across black sea happen sabotage',
'obvious correct answer wa also brief period russia dirty word post cold war pre crimea',
'escalate maybe weapon manufacture want war russia doe tho go back creation nato',
'ukraine tug war russia west toronto montreal vancouver calgary edmonton ottawa gatineau quebec winnepeg hamilton kitchener london victoria stcatharines niagara halifax oshawa windsor saskatoon barrie stjohns',
'russian soldier mobilized russia western country pushing world global war complete final victory russia guarantee world conflict giving territory ukraine threat state slavaukra ni slavaukraine ukraine',
'russia doe certain thing ukraine war russia china us military absorb taiwan back one china war china iran doe something arabia war iran add n korea please stop war',
'word know u soldier fighting russia ukrainian soil ukrainian fight send help see success ukraine success classic proxy war',
'nicest xbot going answer question dear sir war ukraine belong russia',
'trying pas sort victory lol many month war embarrassing imbe ile russia far superior force way supply missile men plane look crumbled multiple front',
'week ahead russia ukraine war u non farm payroll data opec meeting vienna trade balance germany read capitalzonefx money trade fx z forex eforfx h zl fx west global wforex forex rkiye doge btc',
'russia always say western world never understand u right guy either underestimate overestimate wa born soviet union always knew russian military terrible saw chechen war',
'onset russia ha every right intervene stop monstrous barbarity ukrainian nationalistic genocide racist ideology blame arse handed stop lying making thing justify war',
'max right advocating negotiated peace settlement u russia leave ukraine fewer ukrainian die war end reduce risk nuclear conflict exact opposite think war sport',
'ex russian president suggests way avoid nuclear war rt russia amp former soviet union',
'amen biden uk europe fomenting wwiii russia used ukraine proxy launch war russia congress allow biden go war russia russia enemy joe biden democrat enemy',
'russia ukraine war marathinews news newsinmarathi',
'oh god stop god damn world war ii comparison nazi conquered europe actual threat could taken even territory russia even take ukraine acting like big threat nazi',
'omg ya u know wan na hear ur opinion russian ukraine war u even understand russia imperial history bet ur opinion fully formed nuanced',
'last war start harden heart prepare shield god u great russia',
'oleksandr usyk ukraine v andrei kniazev russia knockout boxing fight hd war news russia ukraine ukrainenews',
```

OUTPUT

How people are reacting on Russia war by analyzing 100 Tweets.



CONCLUSION AND FUTURE WORKS

- This project's first contribution to the sentiment analysis research field is to detail the pre processing steps which have to be applied to extract bags of words from Twitter data. The second contribution is to propose a topic based sentiment analysis approach. This approach relies on the fact that the complexity of an analysis can be reduced when the algorithm focus on a smaller range of topic. The algorithms have to deal with less vocabulary and its estimation can be more accurate. Hence, the experiments have shown that, applying this method performs better than the classical sentiment analysis model. The proposed approach can be extended to other sentiment analysis problems.
- The method can be improved. The project has focused on exploiting the results of the default parameter for the topic modelling method. More research on the topic extraction may allow to have more distinct topics, hence the estimation should be more accurate . Moreover, the method proposed here cannot handle neutral data. In order to apply sentiment analysis on real application cases, neutral data need to be handled. Therefore, future works should focus on these neutral data. Two methods may allow to detect neutrality