

project

by Kingshuk Srivastava

Submission date: 03-May-2023 10:27AM (UTC+0530)

Submission ID: 2082790363

File name: final_minor_project_2_report111.docx (735.39K)

Word count: 1261

Character count: 7337

Minor Project-II

Mid Semester Report

for

Election outcome prediction using social media analysis

Date: 30/03/2023

Prepared by

Specialization	SAP ID	Name
B.TECH Big Data (NON-HONS.)	500084655	Aditya Aditya
B.TECH Big Data (NON-HONS.)	500083050	Ayush Upadhyay
B.TECH Big Data (HONS.)	500085037	Ali Amaan Zaidi
B.TECH Big Data (HONS.)	500084199	Raghav Prashar



4

Department of Informatics
School of Computer Science

UNIVERSITY OF PETROLEUM & ENERGY STUDIES,
DEHRADUN- 248007. Uttarakhand

Dr. Kingshuk Shrivastav
Project Guide

Dr. TP Singh
Cluster head



School of Computer Science
University of Petroleum & Energy Studies, Dehradun

Mid Semester Report

Project Title: Election outcome prediction using social media analysis

Abstract:

The Election Outcome Prediction using Social Media Analysis project aims to develop a predictive model that can forecast the results of an election using social media data analysis. With the increasing usage of social media platforms, individuals and organizations have access to vast amounts of data about people's opinions, beliefs, and preferences. This project seeks to leverage this data to create a machine learning model that can accurately predict election outcomes based on social media sentiment analysis and other relevant metrics.

The project involves collecting and analyzing social media data related to the election, cleaning and processing it using natural language processing techniques, and extracting meaningful insights and trends. These insights are then used to train a machine learning model that takes into account various factors such as sentiment analysis, candidate mentions, and social media activity related to the election.

Introduction:

The Election Outcome Prediction using Social Media Analysis project aims to leverage the power of social media data to predict the results of an election. With the rise of social media platforms, individuals and organizations have unprecedented access to vast amounts of data about people's opinions, beliefs, and preferences. This project seeks to harness this data to create a predictive model that can accurately forecast election results based on social media sentiment analysis and other relevant metrics. The project involves collecting and analyzing social media data related to the election, including posts, comments, and hashtags on various platforms such as Twitter, Facebook, and Instagram. This data is then cleaned and processed using natural language processing techniques to extract meaningful insights and identify trends. The insights gained from the data analysis are then used to train a machine learning model to predict the outcome of the election. The model takes into account various factors such as sentiment analysis, candidate mentions, and the overall social media activity related to the election.

The results of this project could be used by political parties, campaign strategists, and other stakeholders to make informed decisions about their election campaigns. By accurately predicting the election outcome, this project could also help to enhance the transparency and accountability of the electoral process..

Problem Statement:

The problem statement for the Election Outcome Prediction using Social Media Analysis project is the lack of accurate and reliable methods for predicting the results of an election. Traditional polling methods can be expensive, time-consuming, and may not always reflect the true sentiments of the electorate. Furthermore, the increasing use of social media platforms as a means of political discourse and communication has created an enormous amount of unstructured data that is difficult to analyze and interpret.

The aim of this project is to leverage the power of social media data analysis to develop a predictive model that can accurately forecast the results of an election. This would provide a cost-effective, efficient, and reliable alternative to traditional polling methods, allowing political parties, campaign strategists, and other stakeholders to make more informed decisions about their election campaigns. The project seeks to address the challenges of analyzing and interpreting vast amounts of unstructured social media data to gain meaningful insights and trends that can be used to train a machine learning model for election outcome prediction.

Literature Survey

- "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment" by Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2011): This study analyzed Twitter data from the 2010 German federal election to predict the election outcome. The authors found that Twitter data could predict the election outcome with an accuracy of 73.3%.
- "Can Social Media Predict Election Results? Evidence from New Zealand" by Borquez, J. C. (2018): This study analyzed Twitter data from the 2017 New Zealand general election to predict the election outcome. The author found that Twitter data could predict the election outcome with an accuracy of 62.5%.

Overall, these studies suggest that social media analysis can be used to predict election outcomes with varying levels of accuracy. However, there are also concerns about the impact of social media on political polarization and the potential for biases in social media data.

Existing System Issue

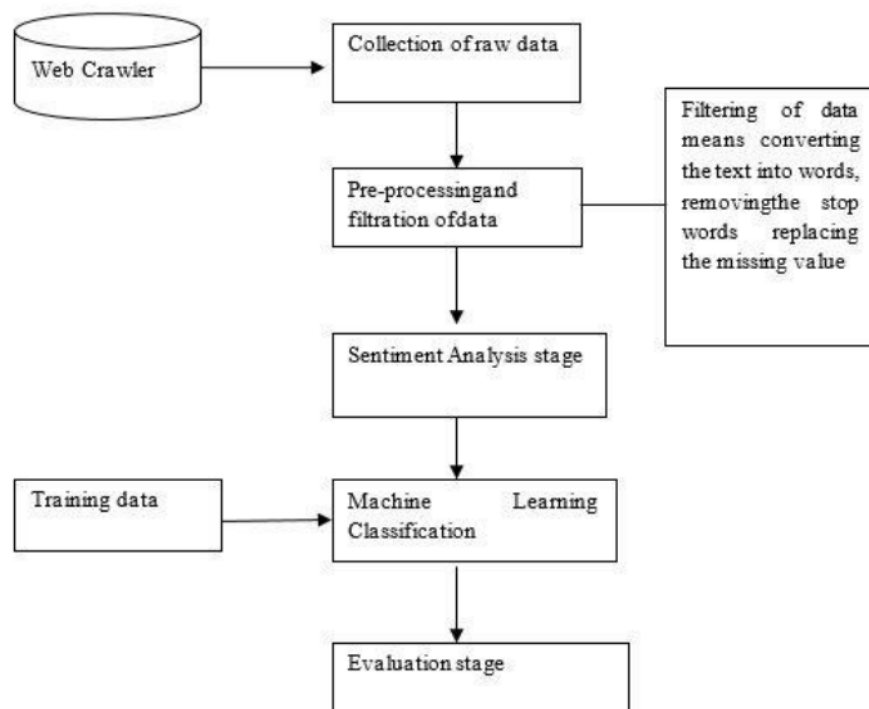
The existing system for predicting election outcomes typically relies on traditional polling methods, such as conducting surveys and interviews with voters. However, these methods have several limitations, including:

- Time-consuming: Traditional polling methods require significant time and resources to plan, execute, and analyze the data collected.
- Limited sample size: Polls typically sample only a small percentage of the population, which may not be representative of the wider electorate.

- Limited accuracy: Polls can suffer from bias, measurement error, and other sources of inaccuracy that may impact the accuracy of their predictions.
- Limited real-time feedback: Traditional polling methods may not provide real-time feedback on the sentiments of the electorate, which can limit the ability of campaigns to adjust their strategies and messaging.

The limitations of the existing system highlight the need for more efficient and accurate methods for predicting election outcomes. This is where the Election Outcome Prediction using Social Media Analysis project aims to address the shortcomings of traditional polling methods by leveraging social media data analysis to create a predictive model that can accurately forecast election outcomes in a cost-effective and efficient manner.

Proposed System Design



Algorithms Discussed

- Web Scraping and API Call using requests, BeautifulSoup and praw.
- NLP² models such as nltk , RoBERTa and huggingface hinglish model .NLP stands for Natural Language Processing, which is a branch of artificial intelligence that focuses on the interaction between computers and humans using natural language.
- RoBERTa (Robustly Optimized BERT Pre-training Approach)¹ is a large-scale language model developed by Facebook AI Research (FAIR). It is based on the BERT (Bidirectional Encoder Representations from Transformers) model architecture and was trained using an improved pre-training methodology.

UML Diagram

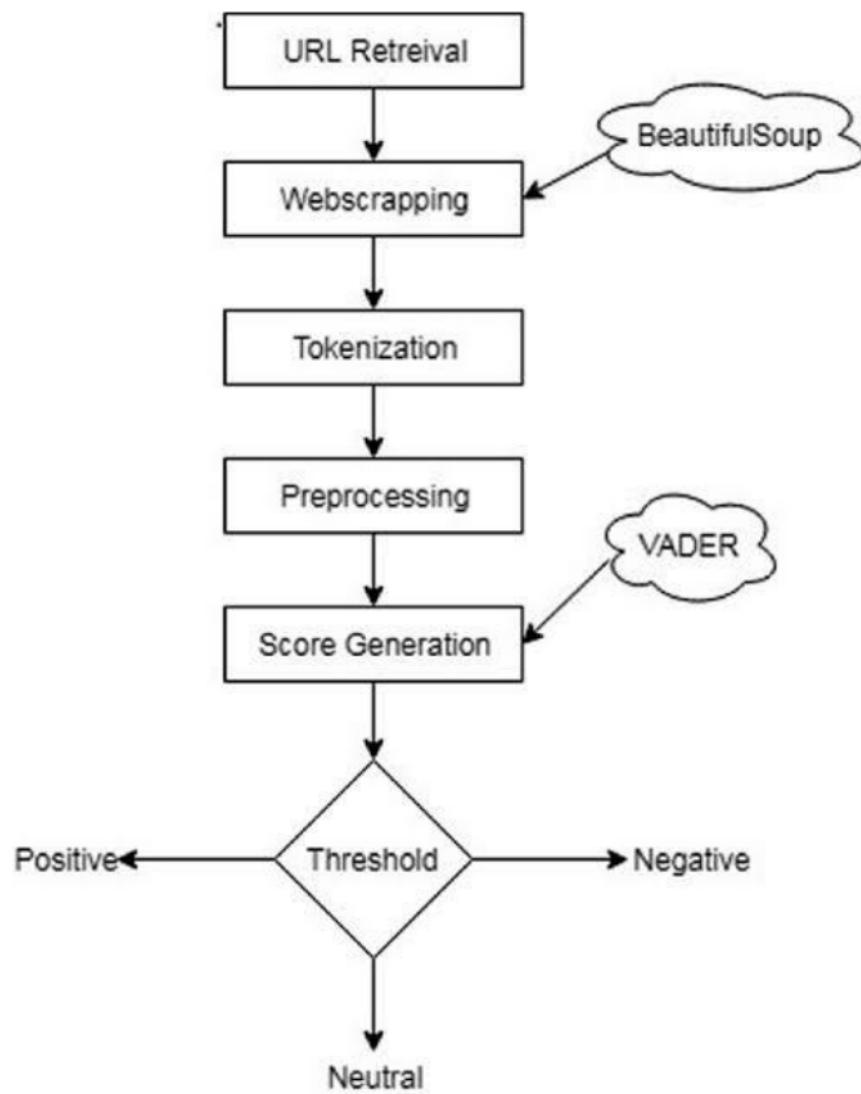


Fig 1.2 UML Diagram

Results and Discussion

- Dataset compiled from various sources like: reReddit and various news websites and further preprocessed by various task. Figure 1.1 and 1.2 show the datasets compiled f

```
,headline
0,amit shah on veer savarkar news18 rising india summit 2023 bjp viral videos shorts
1,"washing machine, bhajapa': mamata banerjee's laundry act on stage to protest against bjp watch"
2,piyush goyal at rising india 2023: 'pm modi stands out as the hero number 1' shorts viral videos
3,"rahul gandhi gets readyNBSFfor k'takeNBSFfight, to launch campaign from 2019 'modi surname' remark site"
4,4 bjp mlas marshalled out of delhi assembly; aap mlas hit out at l-g over indraprastha college incident
5,"karnataka congress news: dk shivakumar throwing money at a rally, alleges bjp english news"
6,"karnataka elections 2023: what's at stake for bjp, congress and jd(s)"
7,rahul gandhi's disqualification shows bjp's dictatorial mindset: chhattisgarh cm bhupesh baghel
8,bjp will win 200+ seats in madhya pradesh polls: jp nadda
9,"bjp hits back, says congress 'satyagrah' show of arrogance"
10,"congress did not seek legal remedy on purpose, says bjp"
11,lost lok sabha seat as pm modi was afraid of my next speech: rahul gandhi
12,"congress wanted rahul barred because of political calculations, says bjp"
13,opposition parties rally around rahul gandhi; bjp calls his disqualification lawful
14,"bjp using legal cover to silence inconvenient voices, says congress"
15,"bjp schemes emerge from bureaucracy, congress's from the people: rahul gandhi"
16,"bjp is just a tenant, not owner of democracy: congress"
17,"pm modi, mamata banerjee have deal to defame rahul gandhi, congress: adhir ranjan chowdhury"
18,samajwadi party to keep equal distance from both congress and bjp: akhilesh yadav
19,"no gujarat-style experiment in madhya pradesh, bjp to stick with cm shivraj singh chouhan"
20,rahul gandhi has become permanent part of 'toolkit' working against india: bjp chief jp nadda
21,congress leader randhawa calls to finish off pm modi to 'save' country
22,"pm modi tallest leader of country, any party entering northeast will have to support him: assam cm himanta"
23 the harp of helish dharkhan l how big's evensess ex-head shaped india
```

Figure -1.1

```
,headline
0,amit shah on rahul gandhi kolar rally in karnataka rising india summit 2023 short video
1,"does congress believe that courts have closed in this nation": home minister amit shah"
2,news18 rising india home minister amit shah on rahul gandhi rahul gandhi disqualified news18
3,jaishankar on rahul gandhi's disqualification at news18 rising india summit 2023 viral shorts
4,"rahul gandhi gets readyNBSFfor k'takeNBSFfight, to launch campaign from 2019 'modi surname' remark site"
5,s jaishankar on rahul gandhi's disqualification rising india 2023: jaishankar interview news18
6,congress is prepared to face elections: siddaramaiah on karnataka poll schedule
7,"karnataka congress news: dk shivakumar throwing money at a rally, alleges bjp english news"
8,karnataka karnataka electiosn 2023 bommai claims a meeting was held between congress & jds
9,congress news rahulgandhi and sonia gandhi meet sanjay raut
10,"karnataka elections 2023: what's at stake for bjp, congress and jd(s)"
11,"gandhi family is congress pivot, keeps party united: rajasthan cm gehlot"
12,"after lok sabha disqualification, rahul gandhi gets notice to vacate official bungalow"
13,rahul gandhi's disqualification shows bjp's dictatorial mindset: chhattisgarh cm bhupesh baghel
14,rahul has forgotten indira gandhi's letter on savarkar: anurag thakur
15,won't tolerate insult to savarkar: uddhav thackeray on rahul gandhi's remark
16,"bjp hits back, says congress 'satyagrah' show of arrogance"
17,"congress did not seek legal remedy on purpose, says bjp"
18,lost lok sabha seat as pm modi was afraid of my next speech: rahul gandhi
19,"congress wanted rahul barred because of political calculations, says bjp"
20,"rahul gandhi disqualified as mp, can't contest polls for 8 years if conviction not stayed"
21,opposition parties rally around rahul gandhi; bjp calls his disqualification lawful
22,black day in history of democracy in india: kcr on rahul gandhi's disqualification
23 this union legal cover to silence inconvenient voices save congress"
```

Figure:1.2

- Further sentiment analysis is performed using vader, Roberta and hinglish model. Output of one of the models showing positive and negative scores.

```

1 Unnamed: 0,headline,positive_score,negative_score
2 0,0,amit shah on veer savarkar news18 rising india summit 2023 bjp viral videos shorts,0.9993688464164734,0.0005995629471726716
3 1,1,"washing machine, bhajapa": mamata banerjee's laundry act on stage to protest against bjp watch",9.343728015664965e-05,0.9987886548042297
4 2,2,piyush goyal at rising india 2023: 'pm modi stands out as the hero number 1' shorts viral videos,0.9996986389160156,0.00026994539075531065
5 3,3,"rahul gandhi gets readyNBSPror K'takaNBSPright, to launch campaign from 2019 'modi surname' remark site",0.006444282829761505,0.991431355476
6 4,4,4 bjp mlas marshalled out of delhi assembly; aap mlas hit out at l-g over indraprastha college incident,0.9998202919960022,0.0001303747703786
7 5,5,"karnataka congress news: dk shivakumar throwing money at a rally, alleges bjp english news",8.818224887363613e-05,0.9987786412239075
8 6,6,"karnataka elections 2023: what's at stake for bjp, congress and jd(s)",0.9995290040969849,4.882626308244653e-05
9 7,7,rahul gandhi's disqualification shows bjp's dictatorial mindset: chhattisgarh cm bhupesh baghel,0.00020192074589431286,5.092552237329073e-05
10 8,8,bjp will win 200+ seats in madhya pradesh polls: jp nadda,0.9998101592063904,6.626573303947225e-05
11 9,9,"bjp hits back, says congress 'satyagrah' show of arrogance",8.668555528856814e-05,0.9987719655036926
12 10,10,"congress did not seek legal remedy on purpose, says bjp",0.0008920224499888718,0.9978880286216736
13 11,11,lost lok sabha seat as pm modi was afraid of my next speech: rahul gandhi,8.837060886435211e-05,0.9987799525260925
14 12,12,"congress wanted rahul barred because of political calculations, says bjp",8.608039934188128e-05,0.9987722039222717
15 13,13,opposition parties rally around rahul gandhi; bjp calls his disqualification lawtul,8.701315528014675e-05,0.9987741112709045
16 14,14,"bjp using legal cover to silence inconvenient voices, says congress",9.223099186783656e-05,0.9987910389900208
17 15,15,"bjp schemes emerge from bureaucracy, congress's from the people: rahul gandhi",8.516432717442513e-05,0.998767614364624
18 16,16,"bjp is just a tenant, not owner of democracy: congress",8.647063077660277e-05,0.9987738728523254
19 17,17,"pm modi, mamata banerjee have deal to defame rahul gandhi, congress: adhir ranjan chowdhury",8.544828597223386e-05,0.9987688660621643
20 18,18,samajwadi party to keep equal distance from both congress and bjp: akhilesh yadav,0.9998288154602051,0.00011714826541719958
21 19,19,"no gujarat-style experiment in madhya pradesh, bjp to stick with cm shivraj singh chouhan",0.00013405113713815808,0.9988377690315247
22 20,20,rahul gandhi has become permanent part of 'toolkit' working against india: bjp chief jp nadda,0.9703811407089233,0.029184672981500626
  
```

- Final output showing the outcome of election by analyzing sentiments via different model.

```

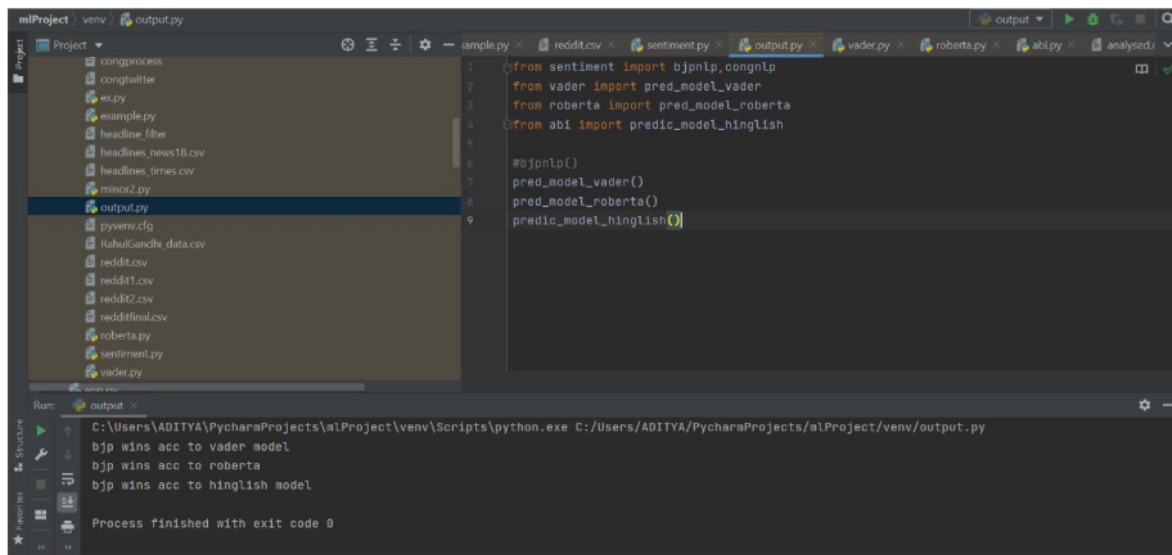
1 from sentiment import bjpnlp,congntp
2 from vader import pred_model_vader
3 from roberta import pred_model_roberta
4 from abi import predic_model_hinglish
5
6 #bjpnlp()
7 pred_model_vader()
8 pred_model_roberta()
9 predic_model_hinglish()
  
```

```

C:\Users\ADITYA\PycharmProjects\mlProject\venv\Scripts\python.exe C:/Users/ADITYA/PycharmProjects/mlProject/venv/output.py
bjp wins acc to vader model
bjp wins acc to roberta
bjp wins acc to hinglish model
Process finished with exit code 0
  
```

Conclusion

- BJP will be the winning side according to the mode



The screenshot shows a PyCharm IDE with a project named 'mlProject'. The left sidebar displays a file explorer with various files including 'congress', 'conghwiter', 'ex.py', 'example.py', 'headline_filter', 'headlines_news18.csv', 'headlines_times.csv', 'minor2.py', 'output.py', 'pyvenv.cfg', 'RahulGandhi_data.csv', 'reddit.csv', 'reddit1.csv', 'reddit2.csv', 'reddittinal.csv', 'roberta.py', 'sentiment.py', and 'vader.py'. The main editor window shows the 'output.py' file with the following code:

```
1 from sentiment import bjpnlp, congnlp
2 from vader import pred_model_vader
3 from roberta import pred_model_roberta
4 from abi import predic_model_hinglish
5
6 #bjpnlp()
7 pred_model_vader()
8 pred_model_roberta()
9 predic_model_hinglish()
```

The bottom panel shows the output of the script:

```
C:\Users\ADITYA\PycharmProjects\mlProject\venv\Scripts\python.exe C:/Users/ADITYA/PycharmProjects/mlProject/venv/output.py
bjp wins acc to vader model
bjp wins acc to roberta
bjp wins acc to hinglish model
Process finished with exit code 0
```

References

- [1] Gayo-Avello, D., Metaxas, P. T., & Mustafaraj, E. (2011). Limits of electoral predictions using Twitter. ICWSM, 178-185.
- [2] Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2011). Predicting elections with Twitter: What 140 characters reveal about political sentiment. ICWSM, 178-185.
- [3] Jungherr, A., Jurgens, P., & Schoen, H. (2016). Why the pirate party won the German election of 2009 or the trouble with predictions: A response to Tumasjan, A., Sprenger, T.O., Sandner, P.G. & Welpe, I.M. "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment". Social Science Computer Review, 34(2), 229-234.
- [4] Borquez, J. C. (2018). Can social media predict election results? Evidence from New Zealand. Policy & Internet, 10(3), 326-342.





project

ORIGINALITY REPORT

10%

SIMILARITY INDEX

6%

INTERNET SOURCES

5%

PUBLICATIONS

8%

STUDENT PAPERS

PRIMARY SOURCES

1

Submitted to Kingston University

Student Paper

2%

2

Submitted to East Carolina University

Student Paper

2%

3

www.tandfonline.com

Internet Source

2%

4

Neeraj Kumar Singh, Amardeep Kumar, Rajesh Singh, Anita .. "Design and development of air quality management devices with sensors and web of things", International Journal of Engineering & Technology, 2018

Publication

1%

5

Submitted to The Manchester College

Student Paper

1%

6

morgenweb.ciando-shop.com

Internet Source

1%

7

Submitted to Liverpool John Moores University

Student Paper

1%

Exclude quotes On

Exclude matches

< 10 words

Exclude bibliography On