Twitter Data Sentiment Analysis Using Naive Bayes Classifier and Generation of Heat Map for Analyzing Intensity Geographically

Chapter · April 2021

DOI: 10.1007/978-981-33-6919-1_10

6 authors, including:

Jyoti Gautam
JSS Academy of Technical Education
59 PUBLICATIONS 914 CITATIONS

Rabindra Nath Shaw
Bharath University
144 PUBLICATIONS 1,161 CITATIONS

Ankush Ghosh
The Neotia University
113 PUBLICATIONS 1,004 CITATIONS

Twitter Data Sentiment Analysis Using Naive Bayes Classifier and Generation of Heat Map for Analyzing Intensity Geographically

Jyoti Gautam, Mihir Atrey, Nitima Malsa, Abhishek Balyan, Rabindra Nath Shaw, and Ankush Ghosh

Abstract Analysis of tweets or microblogs posted by users of a popular microblogging service, Twitter, using natural language processing is a very powerful tool to understand public sentiment on any worldly issue. This work expounds a hybrid approach using both corpus-based and dictionary-based methods to determine the semantic orientation of the opinion words in tweets using Naïve Bayes classifier. The accuracy of the classification has been verified with the use of Weka Tool. The valuable knowledge gained should be very informative and easy to understand. The proposed system intends to do that and also formulate a way to plot this Twitter data on Google Maps. A heat map showcases the intensity of a particular sentiment from different locations in the world. The actual implementation of this project has been done on Python language using various libraries for specific tasks. We have utilized 'TextBlob' library for implementing Naïve Bayes classifier.

Keywords Twitter · Sentiment analysis · Machine learning · Naïve Bayes classifier · Python

J. Gautam · M. Atrey · N. Malsa · A. Balyan
Department of Computer Science and Engineering, JSS Academy of Technical Education, Noida, India
e-mail: jyotig@jssaten.ac.in

M. Atrey
e-mail: mihir.atrey@gmail.com

N. Malsa
e-mail: nitimamalsa@jssaten.ac.in

A. Balyan
e-mail: balyanabhishek@gmail.com

R. N. Shaw
Department of Electronics and Communication Engineering, Galgotias University, Greater Noida, India
e-mail: r.n.s@ieee.org

A. Ghosh (✉)
The Neotia University, Sarisha, West Bengal, India
e-mail: ankushghosh@gmail.com

J. C. Babsal et al. (eds.), *Advances in Applications of Data-Driven Computing*,
Advances in Intelligent Systems and Computing 1319,
https://doi.org/10.1007/978-981-33-6919-1_10

1 Introduction

Nowadays, Twitter has become a trendy way for sharing their views about every worldly issue to the point that is noteworthy for trading and marketing industry. Every person from a common man to a celebrity or a politician can share their opinion about anything within a limit of 280 characters. The people can give their opinions without necessarily belonging to a particular institution making their viewpoints more real and authentic sources for analysis. A lot of researchers and enthusiasts have worked upon this Twitter data to extract public sentiment or sentiment of masses [1]. Product reviewing has also been rapidly growing in recent years because more and more products are selling on the Web. The large number of reviews allows customers to make informed decisions on product purchases. However, it is difficult for product manufacturers or businesses to keep track of customer opinions and sentiments on their products and services. In order to enhance the customer shopping experiences, a system is needed to help people analyze the sentiment content of product reviews. Here, we intend to showcase the strength and usability of a system which handles live Twitter data in an efficient manner to display detailed analysis in the form of reports, pie charts, and heat maps [1–2].

This type of analysis can help the user to research upon public opinion on any topic that is related to any field. The user will be able to enter a hash tag and the number of tweets it wants the system to analyze [3–4]. The system will then perform sentiment analysis and generate an informative database on an excel sheet, further using that database to create a summarized report showcasing what percentage of people display what sentiment out of a wide range of sentiments [5, 6]. The report will also be displayed in the form of a pie chart to increase user friendliness of the report. Finally, the system will generate a heat map based on Google Maps API to showcase the intensity of a particular sentiment from a particular location in the world [7, 8].

The first section showcases the features of the system model. The second section showcases the system architecture and its implementation.

The third section showcases the case study of a test case. The fourth section is the accuracy test of the Naïve Bayes classifier using Weka Tool.

2 Literature Survey

Lot of research is available in the domain of sentiment analysis. Unluckily, informal tone of tweets is available in the Twitter data. There is a lot of biasing available in the tweets for sentiment analysis in the category like movies, products, popularity or anything of that sort. Sentiment analysis offers range of researches ranging from document-level classification [9] to sentence level [10] which leading to phrases [11]. Distant supervision was used for classifying tweets into positive and negative classes which was presented by Go et al. [12]. They presented an automatic approach for classifying the sentiment of Twitter messages with respect to a query term. Results of machine learning algorithms were presented (Naive Bayes, Maximum Entropy and Support Vector Machines) for classification. Analysis of linguistic features was done for detecting sentiment of Twitter messages and was investigated by Koulompis et al. [13]. Hash-tagged dataset (HASH) was used for training and development purposes. Supervised approach was applied by using the accessible hash tags in Twitter data for structuring training data [14]. Semantics was added as additional features into training set for analyzing the sentiments. Real-time sentiment analysis of tweets using Naive Bayes classifier was published by (Ankur Goel, Dr. Jyoti Gautam and Sitesh Kumar) along with SentiWordNet to improve classification accuracy.

3 System Model

The new system showcased here consists of features that convert Twitter data into knowledge with graphical outputs for much easier understanding of the user.

3.1 Database Creation Using Twitter Data

The system fetches live data from Twitter based on the user inputs and creates a detailed database on excel sheets which can be easily understood by the user [15]. Information such as tweeter's screen name, real name, time of tweeting, followers, following, location, the tweet, tagged users as well as sentiment of the tweet is displayed that can be used for further analysis.

3.2 Sentiments

With the use of 'TextBlob' library of Python, the preprocessed tweets are assigned a float number between -1 and 1 [1]. Seven types of sentiments are assigned to small ranges of values such as:

≥ -1.0 & < -0.6	Strongly Negative
≥ -0.6 & < -0.3	Negative
≥ -0.3 & <0	Weakly Negative
$=0$	Neutral
>0 & ≤ 0.3	Weakly Positive
>0.3 & ≤ 0.6	Positive
>0.6 & ≤ 1.0	Strongly Positive

3.3 Summarized Report on the Information

The system uses the information from the database created earlier to analyze what percentage of the total people whose tweets were analyzed feel what sentiment toward a particular topic. The report is showcased as a set of percentages corresponding to one out of seven emotions that the tweets were analyzed for.

3.4 Pie Chart

To make it easier for the user to compare the results, a pie chart based on the report is created.

3.5 Heat Map

The key feature of our system is its ability to create a heat map using the GPS location of the tweeter when making the tweet. The location is plotted as a point on the map, and this point gets bigger as more tweets pop up from the same location. This allows for an insight about the people's views about a particular topic location wise. It can also let the user understand the public opinion about an issue on a specific location.

4 System Architecture

The system being formulated will be implemented in six steps. The first step will involve inputting a topic in the form of a hash tag that the system will perform its analysis on, along with the number of tweets that the system will work with. The next step will be data extraction from Twitter servers using 'tweepy' library and performing data cleaning on that data to remove any unnecessary information and garbage values. The third step will be the sentiment analysis of individual tweets using 'TextBlob' library to find the polarity of each tweet as a float value between '-1' and '1'. The fourth step will be the construction of a database of the gathered tweets and information associated with each tweet such as the real name of the tweeter, screen name, time of tweeting, polarity, and GPS location on an excel sheet. Fifth step is the analysis of the database created in the fourth step to create a summarized report of the sentiments of tweeters and plotting that information on a pie chart using 'pyplot' library Fig. 1.

This will give a visual representation to the user showcasing what percentage of tweets out of the ones searched and show what sentiment, thus enabling a direct comparison for the user to access the public opinion. Finally, the last step will use

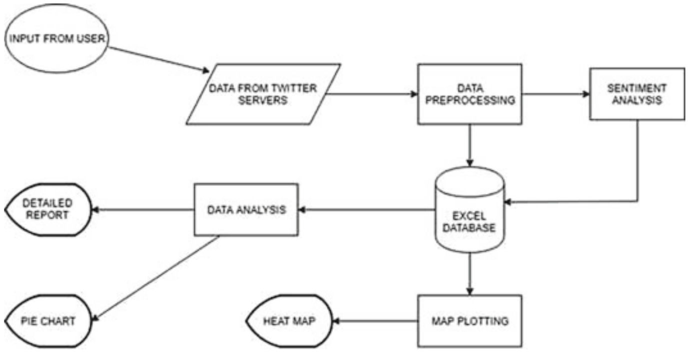


Fig. 1 Flowchart of the system architecture

the GPS location fetched in the database and plot that information as points on a heat map. The points will get bigger as more tweets pop up from the same location.

4.1 Algorithms for Analyzing Sentiments

Different algorithms are available which can be implemented for analyzing sentiments. Rule base is available for sentiment analysis which uses NLP techniques and lexicon.

Other approaches include automatic approaches and hybrid approaches.

Automatic approaches are generally based on machine learning techniques. One of them uses Naive Bayes classifier which is discussed in the coming section.

4.2 Naive Bayes Classifier

Naive Bayes classifiers belong to the probabilistic classifiers and are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem.

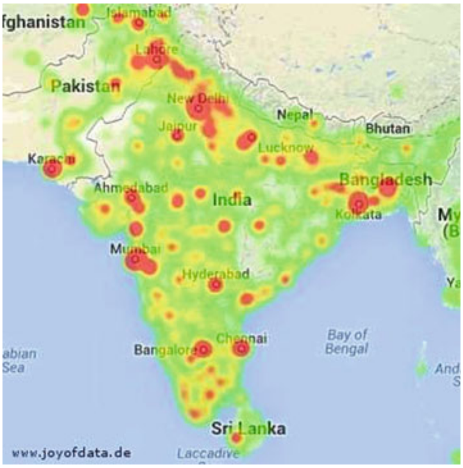
Maximum-likelihood training can be done by evaluating a closed-form expression, which takes linear time, rather than by expensive iterative approximation as used for many other types of classifiers.

This system uses a hybrid approach using both rule-based systems and automatic systems in order to achieve higher accuracy of sentiment calculation. 'TextBlob' library has been used for processing text data. It provides an API for natural language processing (NLP) tasks.

4.3 Google Maps Heat Maps Layer

Intensity of data is visualized at geographical points using a heat map. A colored overlay will appear on top of the map by enabling a heat map. Higher intensity areas will be colored red, and lower intensity areas will appear green. The heat map layer provides client side rendering of heat maps.

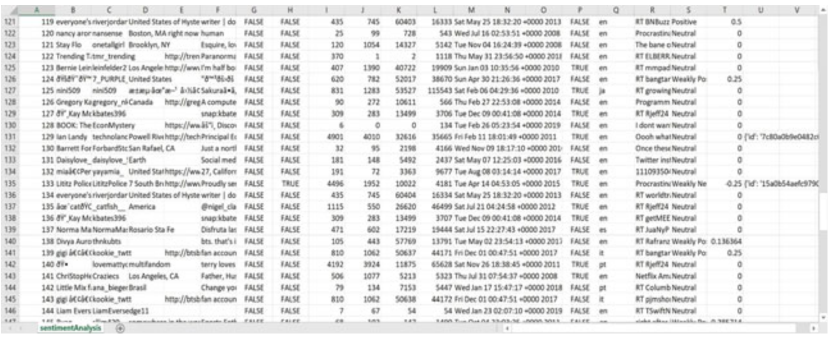
The size and intensity of the plotted points/dots increase with the increasing traffic of tweets of a particular sentiment in this system.



In the above example, bubbles having a redder tone and thicker size are places having more intensity of a particular sentiment than places having a lighter and greener tone.

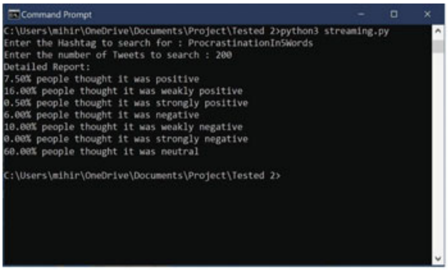
5 Case Study

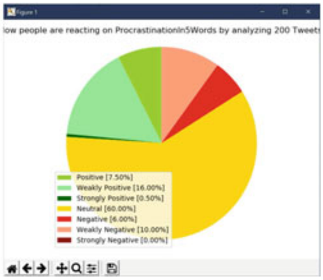
The following test cases were run:

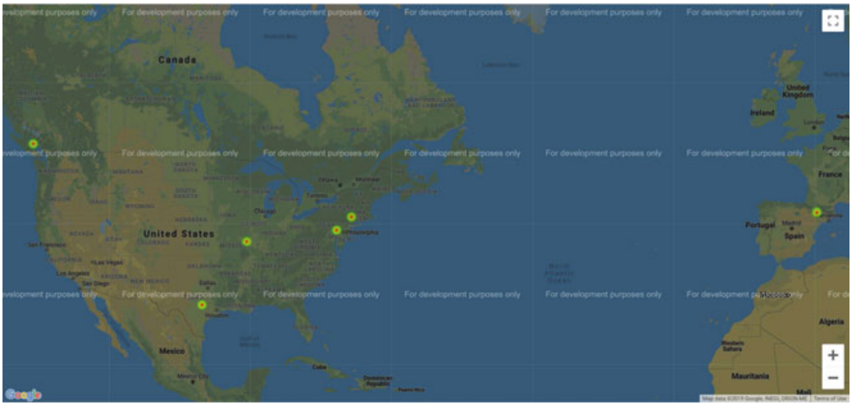


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
121	119	everyone's	riverjordar	United States of Hyste		writer | do	FALSE	FALSE	435	745	60403	16333	Sat May 25 18:32:20 +0000 2013			FALSE	en	RT BNBuzz	Positive	0.5		
122	120	nancy aror	nansense	Boston, MA right now		human	FALSE	FALSE	25	99	728	543	Wed Jul 16 02:53:51 +0000 2008			FALSE	en	Procrastin	Neutral	0		
123	121	Stay Flo	onetallgirl	Brooklyn, NY		Esquire, lo	FALSE	FALSE	120	1054	14327	5142	Tue Nov 04 16:24:39 +0000 2008			FALSE	en	The bane o	Neutral	0		
124	122	Trending T	tmr_trending		http://tren	Paranorma	FALSE	FALSE	370	1	2	1118	Thu May 31 23:56:50 +0000 201			FALSE	en	RT ELBERR	Neutral	0		
125	123	Bernie Lein	leinfelder2	Los Angele	http://ww	I'm half bo	FALSE	FALSE	407	1390	40722	19909	Sun Jan 03 10:35:56 +0000 2010			TRUE	en	RT mmpad	Neutral	0		
126	124	ðŸ'ðŸ"ðŸ™	7_PURPLE_	United States		"ð™'ðŸ-ðŸ	FALSE	FALSE	620	782	52017	38670	Sun Apr 30 21:26:36 +0000 2017			FALSE	en	RT bangtar	Weakly Po	0.25		
127	125	nini509	nini509	æ±æµ·ãœ"æ–'	ã›¾ãC	Sakuraã•ã,	FALSE	FALSE	831	1283	53527	115543	Sat Feb 06 04:29:36 +0000 2010			TRUE	ja	RT growing	Neutral	0		
128	126	Gregory Ka	gregory_nl	Canada	http://greg	A compute	FALSE	FALSE	90	272	10611	566	Thu Feb 27 22:53:08 +0000 2014			FALSE	en	Programm	Neutral	0		
129	127	ðŸ'_Kay Mc	kbates396			snap:kbate	FALSE	FALSE	309	283	13499	3706	Tue Dec 09 00:41:08 +0000 2014			TRUE	en	RT Rjeff24	Neutral	0		
130	128	BOOK: The	EconMystery		https://ww	ã‰"ï, Disco	FALSE	FALSE	6	0	0	134	Tue Feb 26 05:23:54 +0000 2019			FALSE	en	I dont wan	Neutral	0		
131	129	Ian Landy	technolanc	Powell Rive	http://tech	Principal E	FALSE	FALSE	4901	4010	32616	35665	Fri Feb 11 18:01:49 +0000 2011			TRUE	en	Oooh what	Neutral	0	{'id': '7c80a0b9e0482c	
132	130	Barrett For	ForbardStc	San Rafael, CA		Just a nort	FALSE	FALSE	32	95	2198	4166	Wed Nov 09 18:17:10 +0000 201			FALSE	en	Once these	Neutral	0		
133	131	Daisylove_	daisylove_'	Earth		Social med	FALSE	FALSE	181	148	5492	2437	Sat May 07 12:25:03 +0000 2016			FALSE	en	Twitter inst	Neutral	0		
134	132	miaã€‹Per	yayamia_	United Stat	https://ww	27, Califor	FALSE	FALSE	191	72	3363	9677	Tue Aug 08 03:14:14 +0000 2017			TRUE	en	11109350	Neutral	0		
135	133	Lititz Police	LititzPolice	7 South Bri	http://www	Proudly ser	FALSE	TRUE	4496	1952	10022	4181	Tue Apr 14 04:53:05 +0000 2015			TRUE	en	Procrastin	Weakly Ne	-0.25	{'id': '15a0b54aefc979C	
136	134	everyone's	riverjordar	United States of Hyste		writer | do	FALSE	FALSE	435	745	60404	16334	Sat May 25 18:32:20 +0000 2013			FALSE	en	RT worldtr	Neutral	0		
137	135	âœ¨catðŸC	_catfish__	America		@nigel_cla	FALSE	FALSE	1115	550	26620	46499	Sat Jul 21 04:24:58 +0000 2012			TRUE	en	RT Rjeff24	Neutral	0		
138	136	ðŸ'_Kay Mc	kbates396			snap:kbate	FALSE	FALSE	309	283	13499	3707	Tue Dec 09 00:41:08 +0000 2014			TRUE	en	RT getMEE	Neutral	0		
139	137	Norma Ma	NormaMa	Rosario Sta Fe		Disfruta la	FALSE	FALSE	471	602	17219	19444	Sat Jul 15 22:27:43 +0000 2017			FALSE	es	RT JuaNyP	Neutral	0		
140	138	Divya Auro	thnkubts			bts. that's i	FALSE	FALSE	105	443	57769	13791	Tue May 02 23:54:13 +0000 201			FALSE	en	RT Rafranz	Weakly Po	0.136364		
141	139	gigi ã€cã€	kookie_twt		http://btsb	fan accoun	FALSE	FALSE	810	1062	50637	44171	Fri Dec 01 00:47:51 +0000 2017			FALSE	it	RT bangtar	Weakly Po	0.25		
142	140	ðŸ•	lovemattyc	multifandom		terry loves	FALSE	FALSE	4192	3924	11875	65628	Sat Nov 26 18:38:45 +0000 2011			TRUE	pt	RT Rjeff24	Neutral	0		
143	141	ChriStopHe	Craziecs	Los Angeles, CA		Father, Hu	FALSE	FALSE	506	1077	5213	5323	Thu Jul 31 07:54:37 +0000 2008			TRUE	en	Netflix Am	Neutral	0		
144	142	Little Mix f.	ana_bieger	Brasil		Change yo	FALSE	FALSE	79	134	7153	5447	Wed Jan 17 15:47:17 +0000 2018			FALSE	pt	RT Columb	Neutral	0		
145	143	gigi ã€cã€	kookie_twt		http://btsb	fan accoun	FALSE	FALSE	810	1062	50638	44172	Fri Dec 01 00:47:51 +0000 2017			FALSE	it	RT pjmsho	Neutral	0		
146	144	Liam Evers	LiamEversedge11				FALSE	FALSE	7	67	54	54	Wed Jan 23 02:07:10 +0000 2019			FALSE	en	RT TSwiftN	Neutral	0		

sentimentAnalysis







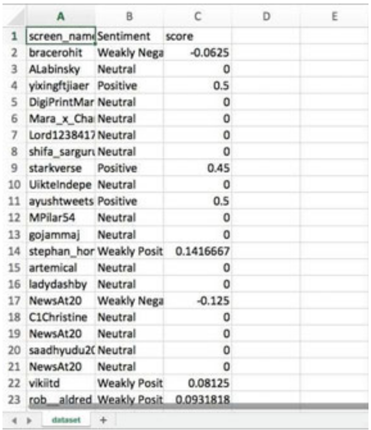
The number of locations plotted on the heat map depends on the number of tweeters who have provided GPS permission to their Twitter app or to the Twitter Web site.

The system can be modified to make points be plotted on the heat map only for tweets showing negative emotions or positive emotions or for any sentiments out of the seven sentiments included in the system.

6 Naive Bayes Accuracy Test Using Weka

Weka tool is used for machine learning algorithms. It has functions available for preparation of data, classification and regression techniques, clustering, etc.

For Naive Bayes classification, the very first step is to extract the relevant fields from the csv file, which in our case is screen-name, sentiment, and score.



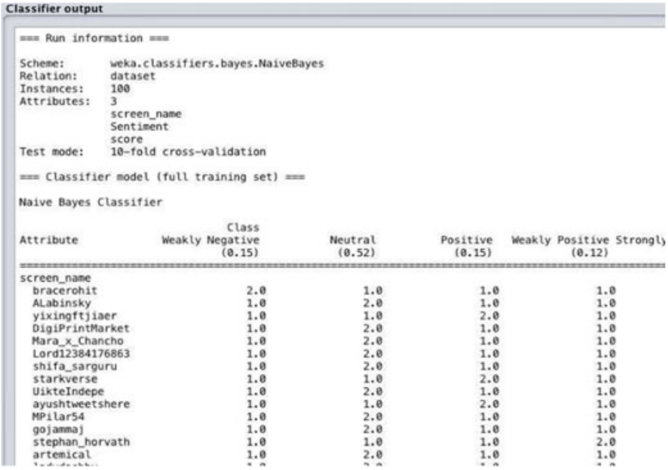
	A	B	C	D	E
1	screen_name	Sentiment	score		
2	bracerohit	Weakly Nega	-0.0625		
3	ALabinsky	Neutral	0		
4	yixingftjiaer	Positive	0.5		
5	DigiPrintMar	Neutral	0		
6	Mara_x_Cha	Neutral	0		
7	Lord1238417	Neutral	0		
8	shifa_sargur	Neutral	0		
9	starkverse	Positive	0.45		
10	UikteIndepe	Neutral	0		
11	ayushtweets	Positive	0.5		
12	MPilar54	Neutral	0		
13	gojammaj	Neutral	0		
14	stephan_hor	Weakly Posit	0.1416667		
15	artemical	Neutral	0		
16	ladydashby	Neutral	0		
17	NewsAt20	Weakly Nega	-0.125		
18	C1Christine	Neutral	0		
19	NewsAt20	Neutral	0		
20	saadhyudu2(	Neutral	0		
21	NewsAt20	Neutral	0		
22	vikiitd	Weakly Posit	0.08125		
23	rob__aldred	Weakly Posit	0.0931818		

dataset

From the tools section in Weka, the extracted dataset is viewed by using arff viewer, and it is then saved in the ".arff" extension file format, so that later Naive Bayes classification can be performed upon it.

Now, for classification, click on the Explorer's tab, and load the dataset in Weka tool. After that, click on classify tab, and select the classifier as Naive Bayes classifier.

6.1 Classifier Output



```
Classifier output
=== Run information ===

Scheme:       weka.classifiers.bayes.NaiveBayes
Relation:     dataset
Instances:    100
Attributes:   3
              screen_name
              Sentiment
              score
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

Naive Bayes Classifier

                                 Class
Attribute            Weakly Negative      Neutral      Positive  Weakly Positive Strongly
                              (0.15)       (0.52)        (0.15)           (0.12)
==========================================================================================
screen_name
  bracerohit                     2.0          1.0           1.0              1.0
  ALabinsky                      1.0          2.0           1.0              1.0
  yixingftjiaer                  1.0          1.0           2.0              1.0
  DigiPrintMarket                1.0          2.0           1.0              1.0
  Mara_x_Chancho                 1.0          2.0           1.0              1.0
  Lord12384176863                1.0          2.0           1.0              1.0
  shifa_sarguru                  1.0          2.0           1.0              1.0
  starkverse                     1.0          1.0           2.0              1.0
  UikteIndepe                    1.0          2.0           1.0              1.0
  ayushtweetshere                1.0          1.0           2.0              1.0
  MPilar54                       1.0          2.0           1.0              1.0
  gojammaj                       1.0          2.0           1.0              1.0
  stephan_horvath                1.0          1.0           1.0              2.0
  artemical                      1.0          2.0           1.0              1.0
```

```
Classifier output

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          96               96      %
Incorrectly Classified Instances         4                4      %
Kappa statistic                          0.9377
Mean absolute error                      0.0228
Root mean squared error                  0.1147
Relative absolute error                 10.3622 %
Root relative squared error             34.8039 %
Total Number of Instances              100

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     Weakl
                 1.000    0.022    0.982      1.000    0.991      0.980    0.979     0.950     Neutr
                 0.933    0.024    0.875      0.933    0.903      0.886    0.976     0.745     Posit
                 0.917    0.011    0.917      0.917    0.917      0.905    0.947     0.931     Weakl
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     Stron
                 0.000    0.000    ?          0.000    ?          ?        0.179     0.018     Stron
Weighted Avg.    0.960    0.017    ?          0.960    ?          ?        0.963     0.907

=== Confusion Matrix ===

  a  b  c  d  e  f   <-- classified as
 15  0  0  0  0  0 |  a = Weakly Negative
  0 54  0  0  0  0 |  b = Neutral
  0  0 14  1  0  0 |  c = Positive
  0  1  0 11  0  0 |  d = Weakly Positive
  0  0  0  0  2  0 |  e = Strongly Negative
  0  0  2  0  0  0 |  f = Strongly Positive
```

6.2 Results

100 tweets were used for the testing purpose, giving 96% efficiency with relative absolute error of 10.36% and root relative squared error of 34.80%.

This proves that our classification of tweets using NAIVE BAYES classifier is fairly accurate and can be relied upon.

7 Conclusion

From above results, it can be concluded that this system is easy to use and understand and can allow anyone to easily research about people's views on any topic. Rather than having to go through Twitter data manually or using a Web site to download datasets and then using a separate tool to analyze that dataset with, a single simple tool can be used. The added functionality with this system is that heat maps are constructed letting the user of this system easily visualize trends location wise. Such a tool that performs analysis on micro blogging sites is highly useful for corporations that depend upon people's opinions and market trends to create newer products that can be sold easily. During election time, the opinion polls that are often done right before elections can also be automated with this system. The opinion of people in a location for a particular political party can let the parties mend their ways or make better decisions. Hence, this system has use everywhere.

8 Future Scope

This system has immense future scope of improvement.

(1) A graphical user interface or GUI can be implemented to make the system more user-friendly.
(2) A method to allow the user to import older datasets of Twitter data for analysis.
(3) All the set of inputs can be displayed in a single window to enable a more comfortable viewing for the user.

References

1. A. Kumar, T.M. Sebastian, Machine Learning assisted sentiment analysis, in *Proceedings of International Conference on Computer Science and Engineering (*ICCSE'2012) (2012)
2. C.M. Bishop, Pattern Recognition and Machine Learning
3. Mark Lutz, Programming Python, 4th edn
4. S. Mandal, S. Biswas, V.E. Balas, R.N. Shaw, A. Ghosh, Motion prediction for autonomous vehicles from Lyft dataset using deep learning, in *2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA)* (Greater Noida, India, 2020), pp. 768–773. https://doi.org/10.1109/ICCCA49541.2020.9250790
5. Y. Belkhier, A. Achour, R. N. Shaw, Fuzzy passivity-based voltage controller strategy of grid-connected PMSG-based wind renewable energy system, in *2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA)* (Greater Noida, India, 2020), pp. 210–214. https://doi.org/10.1109/ICCCA49541.2020.9250838
6. S. Mandal, V. E. Balas, R. N. Shaw, A. Ghosh, Prediction analysis of idiopathic Pulmonary fibrosis progression from OSIC dataset, in *2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON)* (Greater Noida, India, 2020), pp. 861–865. https://doi.org/10.1109/GUCON48875.2020.9231239
7. R.N. Shaw, P. Walde, A. Ghosh, IOT based MPPT for performance improvement of solar PV arrays operating under partial shade dispersion, in *2020 IEEE 9th Power India International Conference (PIICON), SONEPAT*, India, 2020, pp. 1–4. https://doi.org/10.1109/PIICON49524.2020.9112952
8. M. Kumar, V.M. Shenbagaraman, R.N. Shaw, A. Ghosh, Predictive data analysis for energy management of a smart factory leading to sustainability, in ed. by M. Favorskaya , S. Mekhilef , R. Pandey, N. Singh, *Innovations in Electrical and Electronic Engineering. Lecture Notes in Electrical Engineering*, vol. 661 (Springer, Singapore). https://doi.org/10.1007/978-981-15-4692-1_58
9. B. Pang, L. Lee, A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts, in *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL'04)* (2004), pp. 271–278
10. M. Hu, B. Liu, Mining and summarizing customer reviews. in *Proceedings of ACM SIGKDD international conference on knowledge discovery and data mining (KDD-2004)*, Seattle (2004)
11. A. Esuli, F. Sebastiani, Determining term subjectivity and term orientation for opinion mining 2006, in *11th Conference of the European chapter of the association for computational linguistics* (2006)
12. A. Go, R. Bhayani, L. Huaug, Stanford University, Twitter sentiment classification using distant supervision, in *The Third International Conference on Data Analytics* (2009)
13. E. Koulompis, T. Wilson, J. Moore, Twitter sentiment analysis: the good the bad and the OMG! in *The fifth International AAAI Conference on Weblogs and Social Media* (2011)

14. H. Saif, Y. He, H. Alani, Semantic sentiment analysis of twitter, in *The 11th International Semantic Web Conference* (ISWC 2012) (11–15 November 2012, Boston, MA, USA)
15. G.A. Miller, R. Beckwith, C.D. Fellbaum, D. Gross, K. Miller, WordNet: An online lexical database. Int. J. Lexicograph. **3**(4), 235–244 (1990)
16. A. Bifet, E. Frank, Sentiment knowledge discovery in Twitter streaming data, in *Proceedings of 13th International Conference of Discovery Science*, Berlin (Springer, Germany, 2010)
17. B. Pang. L. Lee, Opinion Mining and Sentiment Analysis
18. T.M. Mitchell, Generative and Discriminative Classifiers: Naive Bayes and Logistic Regression
19. A. Goel, J. Gautam, S. Kumar, Real time sentiment analysis of tweets using Naive Bayes, in *Proceedings 2nd International Conference on Next Generation Computing Technologies*, 14–16 Oct 2016