# Sentiment Analysis on New Currency in Kenya Using Twitter Dataset

## **Ibrahim Moge Noor**

Department of engineer, Faculty of Computer engineer, Istanbul Commerce University Turkey. Email: engineermoge@gmail.com

**Abstract.** Social media sites recently became popular, it is clear that it has major influence in society, and almost one third of the entire world are in social media. It became a platform where people express their feelings, share their ideas, wisdoms and give feedback of an event or a product, with help of new technology it gave us an opportunity to analyse these contents easily. Twitter being one of these sites, with full of people opinions, where one can truck sentiment express about different kind of topics, instead of wasting time and energy for long surveys, due to advance sentiment analysis we can now collect a huge data of opinions of people. Sentiment analysis was one of the major interesting research area nowadays. In this paper we focused Sentimental insight into the 2019 Kenya currency replacement. Kenya government has announced that the country currency is to be replace with new generation of bank notes, the government ordered the Kenyan citizen to return back the old 1000 shilling notes (\$10) to bank by 1st October 2019, in a bid to fight against corruption and money laundering. Kenyans citizen expressed their reaction over new banknotes. We perform sentiment analysis of the tweets using Multinomial Naïve Bayes algorithm by utilizing data from one of the social media platform–Twitter and I have collected during this period of demonetization, 1122 tweets from twitter using web scrapper with help of twitter advance search.

Keywords: Demonetization, sentiment analysis, multinomial naïve base, twitter data, machine learning

# **INTRODUCTION**

Demonetization is withdrawal of currency from circulation and replace the old currecy to new currency (Ghosh et al., 2017). The early June 2019 during celebration of Madaraka day, Kenya decided to withdraw 1000 Kenya shilling note which equivalent to 10 dollar from circulation by 1st October 2019. The change of old currency with new one is something started back 2010 when new constitution was promulgated, that mandated the change of currency. The 2010 constitution commanded the Central Bank of Kenya (CBK) to spearhead the creation new notes which should be fashioned to allow the visually impaired to use them. The constitution forbids the utilization of an individual's picture on monetary standards. The old notes has images of the successor of the first president Daniel Arap Moi and first president Jomo Kenyatta, the most interesting thing is that why the Central Bank of Kenya delay for almost a decade for fulfilment of constitution. The Central Bank of Kenya claim that its immediate decision is base on fighting corruption such illict financial flow and money laundering, but the CBK central bank decision has face strong headwind from private sector. although is their constitutional provision command this alter of banknotes.

The Central Bank of Kenya of after its decision of new currency faced two legal suits, of them was felt by kenyan activist he argued that the designs of the new generation currency notes were not subjected to public participation in line with the requirements of the Kenyan Constitution, aslo he argued that new note was include an image of former President Jomo Kenyatta which against constitution which clearly said that the new notes should not be include the portrait of any president but instead should have national symbols and other side central bank of kenya claimed that actuallay not an portrait it is astatue, which at Kenya International Convention, but still what is appering on Kenyan new notes is an image of 1st president. The second suit was filled by a legislature of the east africa legislative assemply because its against certaint articles of constitional, specifically article 231 that was excuted without involvement or partipation of public, just it come above directive it get implemented. Perceiving the untrustworthiness of the governing body, nations, for example United Kingdom and Australia have set up foundations that look for to upgrade its everyday accountability. These responsibility instruments are introduced on the conviction that administrators are not just political on-screen characters but on the other hand are "open authorities" who ought to work openly on the eyes of public (Akach, 2011). After demonetization in kenya people across country post their view on demonetization on social media. Specifically we use twitter as a source of our data set.

Twitter is one of the best online social network site, the microblogging service had over 48.6 million active users and 330 million active users per month, where users share either opinion or information about product, events, politics (Prentice and Huffman, 2008). In this paper, web scraper with help of twitter advance search was used to extract tweets data from twitter on related topic as discussed above.

# MATERIALS AND METHODS

## **Data preprocessing**

Data preprocessing is one way of preparing the data in a way that is suitable to analyze. The preproceessing data shown in Figure 1. Demonetiztion data was collected between june and October, 1128 tweets were collected though their was some shortage of tweets, the dataset was collected using web Scraper and twitter advance search, then the gathered data was applied an important techniques inorder to reduce the noise and dimensionality of sentence.the data was cleaned by removing symbols, non english words, extra spaces and numbers. Also the collected tweets was mixed of hashtags '#', url links, annotation '@', also we remove the stopword, these are common words which doesn't add a values for classification such as and, either, to, the so on. Stemming also was applied, we take out the root of word.

Removing Stopwords: these are common words which doesn't add value to the classification such as *but*, *and*, either so on.

Stemming: Stemming to take out the root of the word.



Figure 1. Data preprocessing.

#### Sentiment analysis of tweets

It's estimated that 80% of the world's information is unstructured and not sorted out in a precharacterized way. Sentiment analysis algorithm: There are various methods that can be implement sentiment analysis, see in Figure 2, which can be group as.

- 1. Automatic depend on machine learning techniques to learn the data.
- 2. Rule-based system which perform sentiment analysis by set of physically created principles.
- Hybrid system compine both. Machine learning and Rule-based system approaches to address Sentiment Analysis is called Hybrid.

We perform our twitter sentiment analysis using Multinomial Naive Bayes algorithm which is a type Naive Bayes, Naive Bayes classifier to get higher accurate and we come up a lexicicon analysis which contains a words list which is negative and positive.



Figure 2. Various methods for implementing sentiment analysis.

In order to classify the sentiment behind the tweets, count the negative and positve words allocate a score for each tweet. In view of the score, the tweet will be classiffied into negative, positive and neutral. Extremity scores are additionally relegated to each tweet based emotónal of tweets such joy,sad,happiness,anger likewise, and base on polarity such negative, positive and neutral.

## **Multinomial Naive Bayes method**

Multinomial Naïve Bayes method used to represent the recurrence in the text of the document. We found first the prior probability of our document, by just dividing number of document of that class (in our case class can be positive, negative or neutral) by total number of documents.

- 1.  $P(c) = \frac{Nc}{N}$
- 2. We calculated for word given a class

 $P(w|c) = \frac{count(w,c) + 1}{count(c) + |V|}$  the addition of 1 and V is for smoothing purpose in case some word got zero count. V is an array of all the words in the vocabulary.  $Pr(c) \propto \prod_{w=1}^{|v|} Pr(w|c)^{fw}$ .

We have used the sum of logs to avoid underflow. Pr (c)  $\alpha \log (\pi c \Pi \Pr(w|c) f w | V | w = 1)$ 

The Pseudo Code for Twitter Sentiment analysis:

```
for (each extracted tweet)
     do
     remove(Symbols)
     remove(non-English words)
     remove (url links)
     remove (annotation)
     remove(stop word)
 TrainedWords[n]
                  =
                      Split
                             Tweet
                                           into
                                     text
words
 for (each word in TrainedWords)
do
   for each word in TrainedWords
do
   if TrainedWords matches negative
   negative count++
   if TrainedWords matches positive
   positive count++
   if TrainedWords matches neutal
   neutral count++
  done
 done
```

```
if (negatives - positives) > 0:
return = Negative sentiment
elif (positives - negatives) > 0:
return = Positive sentiment
else
return Neutral
reset (negative count, positive count, neutal
count)
done
```

# Predicted



#### **RESULTS AND DISCUSSION**

I have 1122 tweets from twitter, between june to October. The collected data was split into a test and train set. 997 tweets for training and 125 for test. The keyword which carrying polarity was collected. The dataset collected was split as 89% for train data and 11% for test data as shown in Figure 3 and Figure 4 down below. The polirity of 88% of train data was as follow.



Figure 3. Count of negative.

### 12 % of data collected from twitter was splits for test.





Train words:

"Precious ","1","0","0" "Debates","0","1","0" "shame","0","1","0" "abuse","0","1","0" "treason","0","1","0" "Could this","0","0","0"

As illustrated above if word is positive the first columns was assign one, but if word is negative the second columns was assign one, finally we assign all column zero for neutral words.



We use a confusion matrix to summarize the performance and prediction results of classification algorithm. Confusion matrix give us a better idea of our classification model. To validate our results the accuracy, precision, and recall are also calculated and the results shows in Figure 6.

		precision	recall	fl-score	support
	1	0.81	0.68	0.74	44
	2	0.72	0.76	0.74	45
	3	0.59	0.67	0.62	36
micro	avg	0.70	0.70	0.70	125
macro	avg	0.71	0.70	0.70	125
weighted	avg	0.71	0.70	0.71	125

Figure 6. Validation of accuracy, precision, and recall.

# CONCLUSIONS

We perform twitter sentiment analysis to understand people's view on demonetization. Dataset collected from twitter was about 1121 tweets, we use web scraper to extract data from twitter and splited into training set and test set, 89% and 11% respectively. We perform sentiment analysis using Multinomail Naive Bayes, the test set which was tested was we were able to get 0.704 for accuracy score. The calculation shown in Figure 7.

Figure 7. Accuracy score.

For future work we are planning to analyze words and document frequency in order to get more accuracy.

#### REFERENCES

Agarwal A, Bhattacharyya P. 2005. Sentiment Analysis: A New Approach for Effective Use of Linguistic Knowledge and Exploiting Similarities in a Set of Documents to be classified.

- Akach M. 2011. Abuse of Power and Corruption in Kenya: Will the New Constitution Enhance Government Accountability 18(1): 242-294.
- Ghosh J, Chandrasekhar CP, Patnaik P. 2017. Demonetisation Decoded. New York.
- Hernandez-Suarez A, Sanchez-Perez G, Toscano-Medina K, Martinez-Hernandez V, Sanchez V. 2018. A Web Scraping Methodology for Bypassing Twitter API Restrictions. 27(3) [27 March 2018]
- Pak A, Paroubek P. 2010. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. 7(10): 1320-1326.
- Barbosa L, Feng J. 2010. Robust Sentiment Detection on Twitter from Biased and Noisy Data: 36-44.
- Susanti AR, Djatna T, Kusuma WA. 2013. Twitter's Sentiment Analysis on GSM Services using Multinomial Naïve Bayes. 15(3): 1354-1361.
- Twitter. 2019. about Twitter. https://about.twitter.com/company [29 July 2019]
- Zhao B. 2017. Web Scraping 1-3.