Multi-lingual Twitter sentiment analysis using machine learning

K. Arun[1], A. Srinagesh[2]
[1]Department of Computer Science and Engineering, Acharya Nagarjuna University, India
[2]Department of Computer Science and Engineering, Rayapati Venkata Ranga Rao and Jagarlamudi Chandramouli (R.V.R. and J.C.) College of Engineering, Acharya Nagarjuna University, India

Article Info

Article history:

Received Aug 2, 2019
Revised May 2, 2020
Accepted May 17, 2020

Keywords:

Machine learning algorithms
Multilingual
NLP
Twitter sentiment analysis

ABSTRACT

Twitter Sentiment Analysis is one of the leading research fields nowadays. Most of the researchers have contributed to the research in twitter sentiment analysis in English tweets, but few researchers have focused on the multilingual twitter sentiment analysis. Still, some more challenges are present and not yet addressed in the domain of multilingual twitter sentiment analysis (MLTSA). Research is highly warranted in these unexplored areas. This study presents the implementation of sentiment analysis in multilingual twitter data and improves the data classification up to the adequate level of accuracy. Twitter is the sixth leading social networking site in the world. Active users for twitter in a month are 330 million. People can tweet or retweet in their languages and allow users to use emoji's, abbreviations, contraction words, misspellings, and shortcut words. The best platform for sentiment analysis is twitter. Multilingual tweets and data sparsity are the two main challenges. In this paper, the MLTSA algorithm gives the solution for these two challenges. MLTSA algorithm is divided into two parts. One for detecting and translating non-English tweets into English using natural language processing (NLP) and the second one is an appropriate pre-processing method with NLP support that can reduce the data sparsity. The result of the MLTSA with SVM achieves good accuracy by up to 95%.

Corresponding Author:

K. Arun,
Department of Computer Science Engineering,
Acharya Nagarjuna University,
NH16, Nagarjuna Nagar, Guntur, Andhra Pradesh, India.
Email: karun014@gmail.com

1. INTRODUCTION

Twitter is microblogging, which allows a maximum of 280 characters per tweet; twitter is one of the best social networks. Twitter has 330 million active users per month [1]. It provides a good platform to share user's opinions and views about the trending topics. Twitter has no local groups; any user can post tweets and has open access.

Twitter provides a good environment for sentiment analysis [2]. Sentiment classification can be done on twitter data like positive, negative and neutral. Sentiment analysis can be applied to movie reviews, product reviews, market review analysis, twitter trending news [3], restaurant review analysis [4], age, gender sentiment classification [5], future predictions like election prediction, forecast applications [6, 7], and forecasting stock market movements [8]. Multilingual tweets can be seen in trending news in the twitter data. Extending the sentiment analysis into multilingual [9] tweets is very essential to improve the accuracy in sentiment analysis without losing any user tweets. The efficiency of this analysis was improved by the best pre-processing techniques [10].

Increasing the accuracy in sentiment analysis is a research challenge in big data. By reducing the data sparsity and translating the multilingual tweets accuracy can be increased [7]. In the sentiment analysis, pre-processing plays an important role [10, 11]. Mostly sentiment analysis is applied only on

English tweets. In the pre-processing level non-English tweets are avoided, i.e. multilingual tweets are not considered in this analysis. On this occasion, the accuracy of the sentiment analysis gets to drop. Very valuable tweets and opinions of the public are not considered here. In twitter sentiment analysis, multilingual tweets must be translated into English tweets [12], and noise data like misspellings, emphasized words, code-mixed tweets need to be solved [13].

Multilingual twitter sentiment analysis is still prominent in reaserch area [14, 15]. In this context, analysis of multilingual tweets is not a direct process. The first step in this sequence is that it requires a translation from non-English tweets to the English language and the complementary steps are applied for the sentiment classification. Spelling mistakes, emphasized words, contraction words are called as data sparsity. Data sparsity can be reduced in pre-processing using NLP. In this way, the accuracy of the classification can be improved drastically through multilingual twitter sentiment analysis [16, 17].

In this paper, MLTSA (Algorithm-1) is a new proposed algorithm, which is used to define and solve the data sparsity by using efficient preprocessing methods and translation from non-English to English is by using Google translator. Finally, the sentiment analysis process has been applied in this algorithm. In this proposed system accuracy is increased up to 95%.

Sentiment analysis for Swiss politician tweets; in this study tweets are available in multiple national languages like Swiss. The language was translated from Swiss to English with the help of machine learning and a lexicon-based approach [18]. Norah Fahad Alshammari [19] Sentiment analysis classification for twitter data from Arabic to English with deep learning model. A complete review of Arabic sentiment analysis is given by Abdullatif Ghallab [20].

Md. Al-Amin improves the accuracy of the Bengali comments up to 75.5%, with the combination of word2vec and sentiment extraction of words. This approach performance is not compromising in increasing the size of the dataset [21]. Narr, Presents multilingual tweets sentiment analysis with the language independent classification approach. Here four languages were translated using the Amazon Mechanical Turk such as English, German, French, and Portuguese. Language tweets are independently translated and classified on human-annotated tweets with good accuracy [22]. Nankani H., Dutta H focuses on sentiment analysis of various low resources multi-languages [23] having limited sentiment analysis resources such as annotated datasets, word embeddings and sentiment lexicons, along with the English language.

2. RESEARCH METHOD

Multilingual twitter sentiment analysis (MLTSA) contains the following steps.

- Pre-processing.
- Language translation for each multi-language tweets.

These two methods are very clearly defined in the flowchart Figure 1 and algorithm as follows.

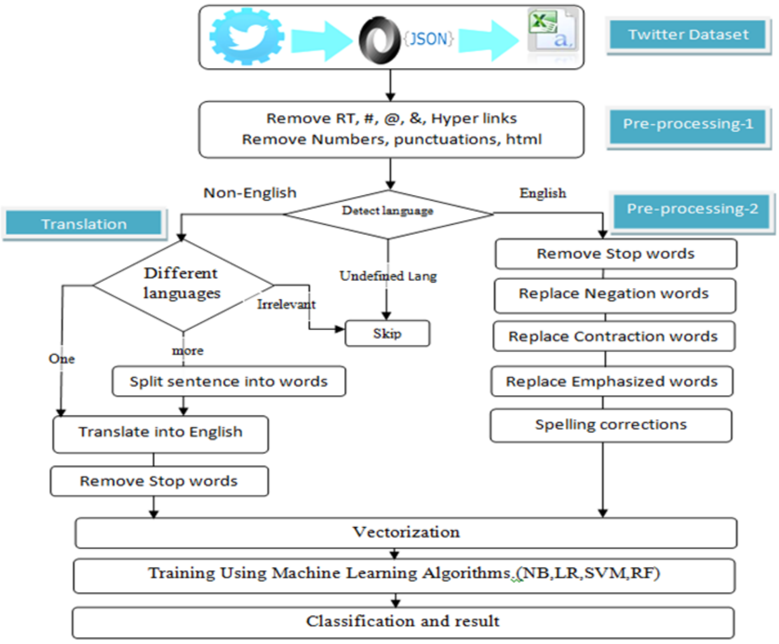


Figure 1. Flowchart for multilingual twitter sentiment analysis

Algorithm 1. Multilingual twitter sentiment analysis (MLTSA)

```
ALGORITHM: - "Multilingual Twitter Sentiment Analysis (MLTSA)"
Purpose: - "Prediction of sentiment Analysis from the Multilingual and sparsity of tweets".
Input: - Tweets Γ, Tweet_language T_L, Sentiment_Lexicon L
Output: - Tweet_Sentiment TS {P, Neg, NT}, Sentiment_Score SS
         Where P: Positive, Neg: Negative, and Nt: Neutral.
Initialize: - POS,NEG, and NT=0,E="English Language"
    While Tokenize Tweets Γ into Tweet($t_i$)
            Remove  RT, Numbers, URL, Hyperlinks from $t_i$
        if T_L($t_i$) ≠ English then
                Each $t_i \in Γ$ is Tokenize into word set $W_i$
                //translating into English using Google Translator(GNMT)//
                        $t_i=P(E_i|W_i)=\prod_{i=1}^{n}P(E_i|E_0,E_1,E_3…E_{i-1};W_1,W_2,W_3…W_i)$
            Tokenize  $t_i \in Γ$ into word set $W_i$
            Remove stop words, punctuation symbols and special symbols from $W_i$
            if  contraction words in $W_i$ then
                        Replace with the complete word.
            else if emphasized words in $W_i$ then
                  Replace with the proper word.
            //Lexicon values from the Lexicon Dictionary for each word//
            Search for $W_i$ in L
            if  $W_i \in$ L.POS then
                  POS←POS+L.val
            else if $W_i \in$ L.NEG then
                  NEG←NEG+L.val
            else
                  NT←NT+1
          // Tweets Sentiments and Sentiments scores calculation //
            if  POS>|NEG| then
                  TS=P,  SS=POS|(POS+NEG)
            else if POS<|NEG| then
                  TS=Neg,   SS=NEG|(POS+NEG)
            else
                  TS=Nt
    End
```

2.1. Pre-processing

In multi-lingual twitter sentiment analysis, pre-processing plays an important key role. The pre-processing step is used to prepare the raw data into suitable data for analysis. The basic method of this process is cleaning and replacing the required data. The pre-processing method directly affects the accuracy and efficiency of sentiment analysis. Here some of the pre-processing steps are used for the best analysis. In the Figure 1 and Algorithm 1 defines many steps in pre-processing.

- Remove "RT": In the twitter data set "RT" re-tweets are common, but it is not required for the processing, so it is removed from the data.
- Remove numeric values: Numbers are not required in this analysis part, so numbers are also removed from the tweets in the data set.
- Remove Isolated words: These words have no significant impact on sentiment analysis. Such as {i, a, y} which are single characters.
- Remove undefined language tweets from the data set: In the twitter data set, every tweet is written in a specific language. But some of the tweets are posted in the undefined language, this type of tweets will not support for the language translations and sentiment analysis. So it is better to avoid undefined language tweets.
- Remove punctuations and special symbols: Punctuation symbols do not affect the analysis; punctuations are removed from each and every tweet.The symbols are [!"#$%&\'()*+,-./:;<=>?@[\\]^_`{|}~], not only these symbols some more special symbols are available regularly that are used in the tweets such as [:" ' '…], these are also removed from the tweets.
- Remove hyperlinks: Hyperlinks do not affect the analysis; from each tweet, hyperlinks are removed.
- Remove Tags: there are different types of tags used in the tweets such as @, and &. This will not show any effect on the analysis.
- Remove Stop words: There is One Hundred and Seventy-Nine (179) list of English language stop words that were supported in the NLTK. All these are used in regular English sentences, but there is no effect in the sentiment analysis so, it is better to remove the stop words. In the stop word list, some of the negative words such as { 'nor', 'not', }has an effect in the sentiment classification, so these words must be excluded from the stop word list, otherwise negative tweets will be classified into positive.

- Replace the contraction words: Contraction words are commonly used in the tweets. The contraction words are {I'll, don't, I'm, won't, this'll, etc}. But this type of contraction words are not directly processed, it needs to be replaced with the proper words instead of contraction words.
  Like {I'll→I will, don't→do not, I'm→ I am, won't→ would not, this'll→this will, ..etc}
- Spelling corrections: More than 10% of tweets contain misspellings, which decreases the efficiency of classification [24]. Here every tweet and every word in the tweet is verified. If any spelling mistakes occur, that misspellings are corrected by using dictionary-based which is proposed by Peter Norvig [9].
- Replace the emphasized words: Emphasized words are used for expressive words and it is very common to express the feelings in the text form. These words are not supported from any dictionary, with the help of NLTK, regular expressions and spelling corrections can be replaced with proper words, {'cooooooool'→'cool','perrrrrrrrrrrffffect'→'perfect','gooooood'→'good','hooooot'→'hot','toooooo' →'too'}.

2.2. Language translation

Language translation is used to translate non-English language tweets into English tweets. The language of the tweets is specified in the "lang" attribute along with the "text". Three types of labels are defining the language of tweets. The labels are 1) English, 2) Non-English, 3) Undefined. Example "en", "hi", "und" in the Figure 1 and Algorithm 1 explains the translation procedure. If the 'lang' type is 'und', These types of tweets provide hyperlinks, videos, stickers, or any graphics but not the textual information and there is no use in the sentiment analysis process.

If the language is in the 'Non-English' type, it requires English translation. According to the given algorithm (MLTSA), Algorithm 1 uses one most Google translator that is GNMT(google neural machine translator). This GNMT [25-28] uses the LSTM RNN (long-short term memory recurrent neural networks) [29]. This GNMT architecture divides into four parts 1) Decoder LSTMs, 2) Encoder LSTMs, 3) Attention, 4) Softmax. Encoder LSTMs encodes the given input set of words from the source sentence into fixed-sized vectors in its hidden layers, Decoder LSTMs and Softmax can decode into target language words symbols, and attention provides the generated target word set and communication between encoder and decoder hidden layers.

In this process, Twitter dataset r is tokenized into single tweets ti, and each tweet has its language, the Language of ti is 'eng' or 'und' no need of translation, otherwise, translation is needed for any source language into English only. In this process ti tokenized into words Wi.

$$V1, V2, V3, ... Vn = \text{EncoderLSTM}(W1, W2, W3, ... Wn) \quad (1)$$

where Vn is the word symbol vector is the output of the encoder LSTM hidden layers.

Then decoderLMST and Softmax modules can make it into English word vector by using the conditional probability model.

$$P(E|V) = P(E| V1, V2, V3, ... Vn) \quad (2)$$

$$= \prod_{i=1}^{n} P(Ei|E0,E1,E3\ldots Ei\text{-}1;\ V1, V2, V3, ... Vn) \quad (3)$$

where E0 is the special symbol as the beginning of the sentence in any target language. Ei translated word vector.

This system can do the translation task for the valid inputs. If Wi is a non-English word and which is defined in a dictionary of a language, that word(s) will go for translation. But in other cases where Wi is non-English but lexically English, translation is not possible.

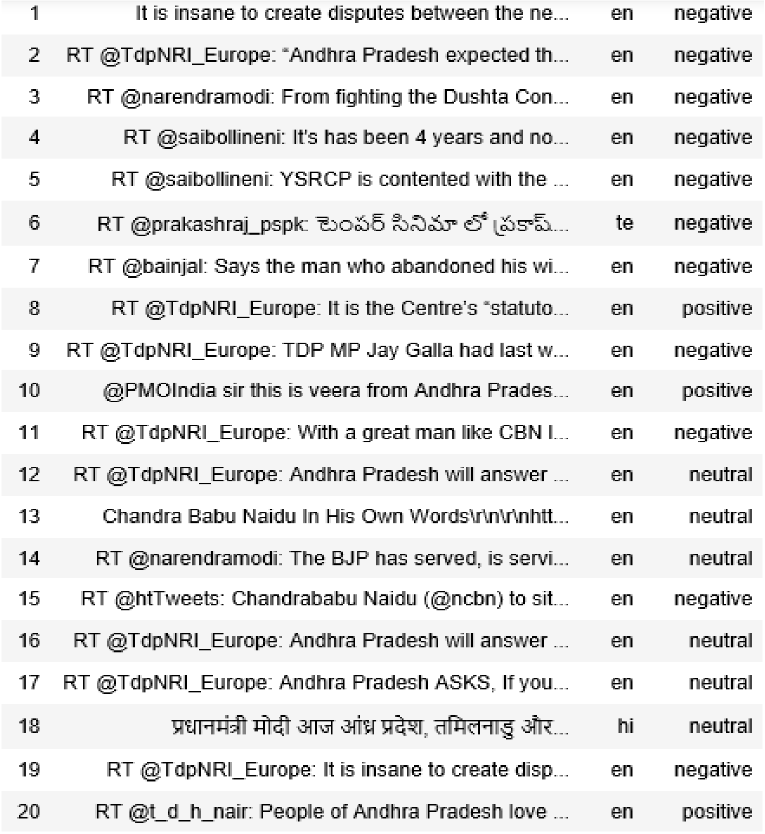
3. RESULTS

3.1. Data set

The data set is collected from the Twitter API. The topic that has been accessed from Twitter live stream is the honourable CM, he is the Andhra Pradesh honourable CM, and 200 tweets were collected. Tweets are in local languages, and mostly in English. The rest of the tweets are in Indian local languages such as Hindi, and Telugu. which is in the Figure 2.

Initially, each tweet contains 69 attributes, such as used_id,user_str_id, text, language, country, out of this process it requires only two attributes {text,lang}. Here text contains tweets, re-tweets and lang is referred to as the language of each tweet. The dimensionality of the data set was reduced up to two features.

One of the new feature added to the data set is "sentiment" as a third feature. This sentiment attribute contains three labels i.e. "positive, negative, neutral". An auto label is assigned with the support of a lexicon-based approach. Translation of different languages into English is done using a python programming language with NLP support. Finally data set contains three attributes. In Figure 3, 65-80% of tweets are in the English language (en), and remaining are in the local language [te: Telugu, hi: Hindi, be: Bengali, ar: Arabic, ur: Urdu] along with the undefined(und) language. Undefined language tweets are not considered in this process.



1	It is insane to create disputes between the ne...	en	negative
2	RT @TdpNRI_Europe: "Andhra Pradesh expected th...	en	negative
3	RT @narendramodi: From fighting the Dushta Con...	en	negative
4	RT @saibollineni: It's has been 4 years and no...	en	negative
5	RT @saibollineni: YSRCP is contented with the ...	en	negative
6	RT @prakashraj_pspk: టెంపర్ సినిమా లో ప్రకాష్...	te	negative
7	RT @bainjal: Says the man who abandoned his wi...	en	negative
8	RT @TdpNRI_Europe: It is the Centre's "statuto...	en	positive
9	RT @TdpNRI_Europe: TDP MP Jay Galla had last w...	en	negative
10	@PMOIndia sir this is veera from Andhra Prades...	en	positive
11	RT @TdpNRI_Europe: With a great man like CBN I...	en	negative
12	RT @TdpNRI_Europe: Andhra Pradesh will answer ...	en	neutral
13	Chandra Babu Naidu In His Own Words\r\n\r\nhtt...	en	neutral
14	RT @narendramodi: The BJP has served, is servi...	en	neutral
15	RT @htTweets: Chandrababu Naidu (@ncbn) to sit...	en	negative
16	RT @TdpNRI_Europe: Andhra Pradesh will answer ...	en	neutral
17	RT @TdpNRI_Europe: Andhra Pradesh ASKS, If you...	en	neutral
18	प्रधानमंत्री मोदी आज आंध्र प्रदेश, तमिलनाडु और...	hi	neutral
19	RT @TdpNRI_Europe: It is insane to create disp...	en	negative
20	RT @t_d_h_nair: People of Andhra Pradesh love ...	en	positive

Figure 2. Sample data set from twitter live

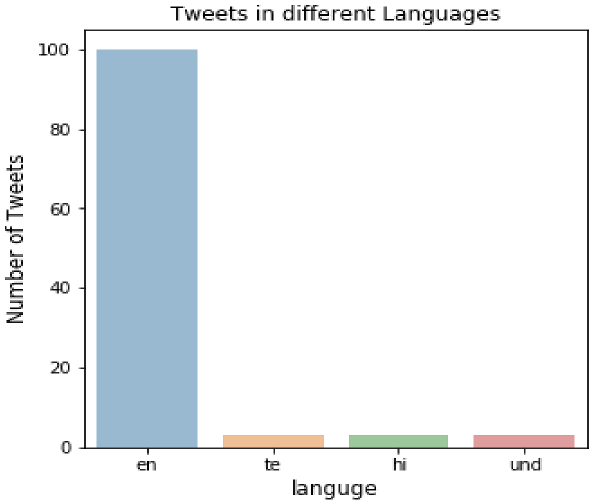


Figure 3. Multilingual tweets [English, Telugu, Hindi, und]

3.2. Pre-processing

The preprocessing method contains 11 steps as mentioned above. The data set contains a huge amount of noise data, even though the attributes were reduced up to three. The noise data must be cleared before analyzing it. Python language completely supports the pre-processing because of its rich set of packages. URL's, numbers, HTML tags, stop words, punctuation symbols, special symbols, and emoticons were removed from each and every tweet. Additionally, the pre-processing module replaces contraction words with actual words, Negation words with complete word and misspelling words were rectified. Finally, the two data sets were pre-processed very effectively.

Sample pre-processing steps:

Input-1: RT@ravi: Apple phone it is verrrry costlllllllllllllllllllllly

Output-1: apple phone costly.

Emphasized words and spelling corrections in the tweets are verrrrry, costlllllllllllllly.

These words can reduce the length like very, costly by using some of the regular expressions. The remaining corrections depend on the spelling corrections.

3.3. Language translation

According to the specified algorithm MLTSA, different language (non-English) tweets are translated. This translation is supported by one of the python modules, and that module depends on the Google translator. In this algorithm, it contains 3 cases.

3.3.1. Translating non-English into English

Tweet in Hindi: पुलवामा हमले पर बोले भारतीय बैडमिंटन टीम के राष्ट्रीय कोच पुलेला गोपीचंद पाक के साथ किसी तरह के

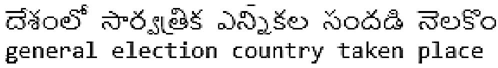
Translated to English: Pullela Gopichand, national coach of the Indian badminton team, spoke on the Pulwama attack.

Tweet in Bengali: RT @IamSourav_b: যখন আজকালকার জীবনমুখী বাংলা চলচিত্রি বা গানে গালগিালাজ

ব্যবহার করা হয় তখন তাে @MyAnandaBazar এর গায়ে লাগনো দখেিা কন্তিু…

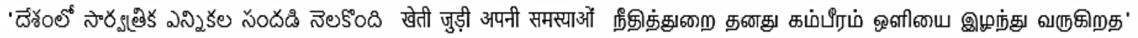
Translated to English: When we are now using the lyrics of the life-changing Bangla film or song, we do not see it

Tweet in Telugu and Translation:



దేశంలో సార్వత్రిక ఎన్నికల సందడి నెలకొంది|
general election country taken place

3.3.2. Translating mixed language words into English



'దేశంలో సార్వత్రిక ఎన్నికల సందడి నెలకొంది खेती जुड़ी अपनी समस्याओं நீதித்துறை தனது கம்பீரம் ஒளியை இழந்து வருகிறத'

Translating into English: Country universal election noise there farming attached mine the problems judiciary his majestically light losing Varukirata.

3.3.3. Code-mixed words

A tweet is typed in one language but the meaning is in other languages, this type of tweet is not getting translated(Code mixed). Tweet: “Antha Rajanna mahima Appatlo annagarini devudu ni ela chesaro ipudu ysr ni devudu cheyadam chusthunte ardamaipoth”. This tweet is written in English, but each and every meaning of the word is related to Telugu language. So this type of tweet is not converted.

3.4. Training and testing

Bag-of-words vector is created after the pre-processing and translations by using the counter vectorization. Data set splits into 7:3 ratio out of that 70% is for the train data, and 30% is for the test data. By using machine learning algorithms [30], the training and testing process is implemented. The ML algorithms are

- Multinomial Naïve Bayes (MNB),
- Logistic Regression (LR),
- Support Vector Machines (SVM),
- Decision Tree (DT),
- k-Nearest Neighbour (kNN),
- Random Forest (RF).

Performance measures are computed for all the algorithms. Performance measures such as precision, recall, f1-scores. Figure 4 to Figure 6 are used to represent the sentiment classification with pre-processes and MLTSA algorithm.

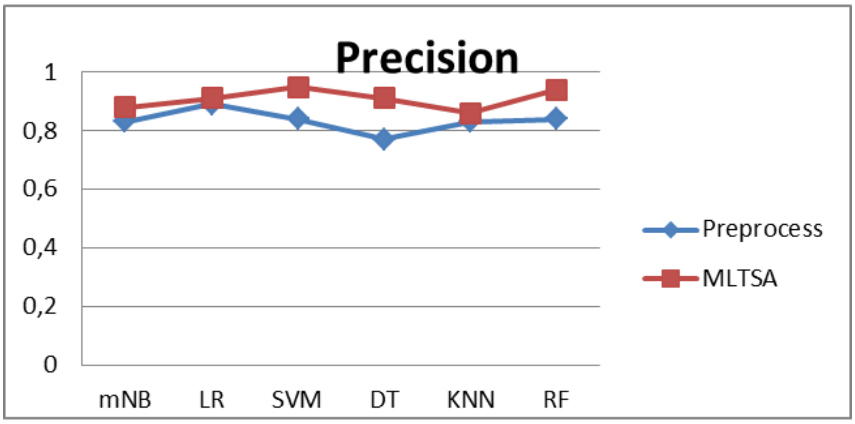


Figure 4. Precision in pre-process vs. MLTSA classification

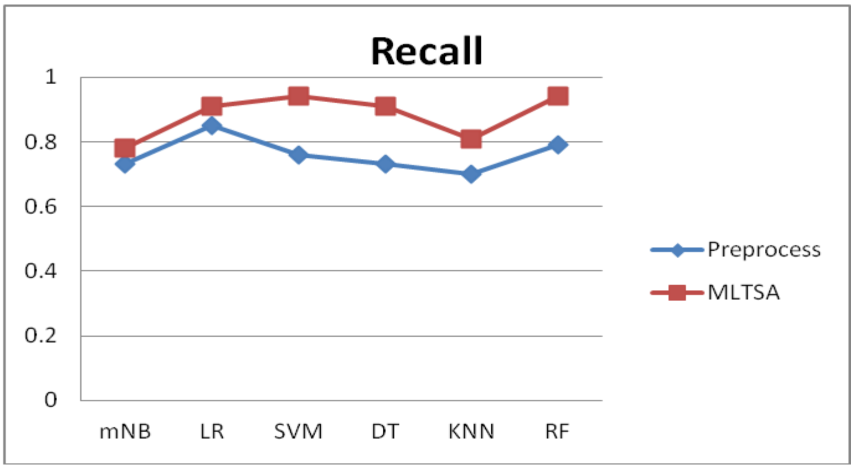


Figure 5. Recall in pre-process vs. MLTSA classification

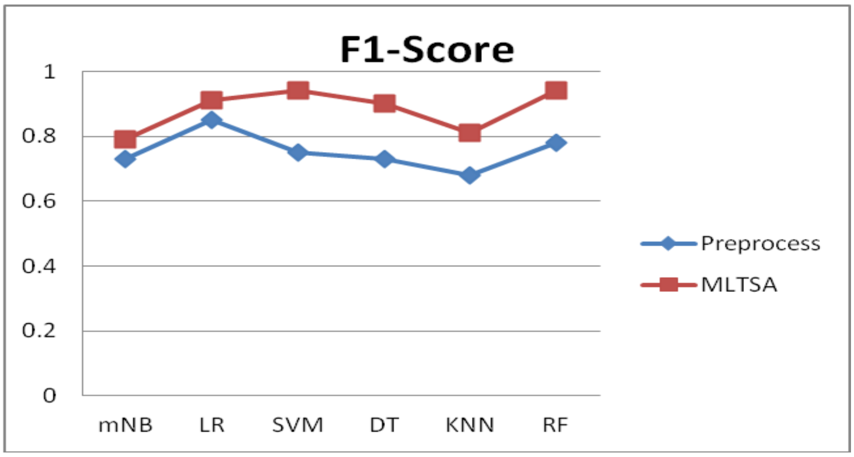


Figure 6. F1-Score in pre-process vs. MLTSA classification

In the Figures 4-6 the results show the comparison between pre-processed tweets and MLTSA algorithm. In all the cases, MLTSA leads to good and accurate results. Precision, Recall and F1-scores are the averages for the positive, negative, neutrals. MLTSA line is increased above the line of pre-processing. Training and testing score accuracy in machine learning algorithms for SVM (support vector machine) is 95% and for RF (random forest), it is 93%.

4. CONCLUSION AND FUTURE WORK

In this paper, sentiments are extracted from the multilingual tweets. This algorithm provides sufficient pre-processing techniques that were applied to set the dataset. Here totally 11 pre-processing techniques were implemented to improve the accuracy in the sentiment analysis. Multi-lingual tweets deal with multiple numbers of international or local languages. Python language translator module, which is supported by the Google translator, is used to translate the no English tweets into English, and then sentiments were extracted from the English data. MLTSA algorithm is a better pre-processing and language translation technique. Machine learning algorithms are applied in the trained and test data. Sentiment classifications are improved on the translated data. Machine learning algorithms are Multinomial naive bayes (MNB), logistic regression (LR), support vector machine (SVM), decision tree (DT), K-nearest neighbor (KNN), and random forest (RF) applied. The performance measures in machine learning are precision, recall, and f1-score. Acceptable improvements were recorded after the translation and SVM is the best classifier in multilingual twitter sentiment analysis. Accuracy is improved by up to 95%. In this work, some challenges need to be solved in situations like for example, accuracy drops when the tweets contain code-mixed and code-switched words and sentences.

REFERENCES

[1] J. Clement, “Twitter: number of active users 2010-2019,” *Statista*, Aug. 2019. [Online], Available: https://www.statista.com/statistics/282087/number-of-monthly-activetwitter-users.

[2] E. Kouloumpis, et al., “Twitter Sentiment Analysis: The Good the Bad and the OMG!” in *Proceedings of Fifth International AAAI Conference on Weblogs and Social Media*, pp. 538-541, 2011.

[3] K. Arun, et al., “Twitter Sentiment Analysis on Demonetization tweets in India Using R language,” *International Journal of Computer Engineering in Research Trends*, vol. 4, no. 6, pp. 252-258, 2017.

[4] S. A. El Rahman, et al., “Sentiment Analysis of Twitter Data,” *International Conference on Computer and Information Sciences (ICCIS),* Saudi Arabia, pp. 1-4, 2019.

[5] S. Kumar, et al., “Exploring Impact of Age and Gender on Sentiment Analysis Using Machine Learning,” *Electronics,* vol. 9, no. 2, pp. 1-14, 2020.

[6] H. Saif, et al., “Semantic sentiment analysis of twitter,” *International Semantic Web Conference*, pp. 508-524, 2012.

[7] S. Shayaa, et al., “Sentiment Analysis of Big Data: Methods, Applications, and Open Challenges,” in *IEEE Access*, vol. 6, pp. 37807-37827, 2018.

[8] T. M. Nisar and M. Yeung, “Twitter as a Tool for Forecasting Stock Market Movements: A Short-window Event Study,” *The Journal of Finance and Data Science*, vol. 4, no. 2, pp. 101-119, 2018.

[9] A. K. Soni, “Multi-Lingual Sentiment Analysis of Twitter data by using classification algorithms,” *2007 Second International Conference on Electrical, Computer and Communication Technologies,* pp. 1-5, 2017.

[10] Z. Jianqiang and G. Xiaolin, “Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis,” in *IEEE Access*, vol. 5, pp. 2870-2879, 2017.

[11] A. Deshwal and S. K. Sharma, “Twitter Sentiment Analysis using Various Classification Algorithms,” *2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)* (*ICRITO),* Noida, pp. 251-257, 2016.

[12] A. Balahur and M. Turchi, “Improving sentiment analysis in twitter using multilingual machine translated data,” *Proceedings of the International Conference Recent Advances in Natural Language Processing*, pp. 49-55, 2013.

[13] S. K. Mahata, et al., “Analyzing Code-Switching Rules for English–Hindi Code-Mixed Text,” *Emerging Technology in Modelling and Graphics*, pp. 137-145, 2020.

[14] K. Dashtipour, et al., “Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques,” *Cognitive Computation*, vol. 8, pp. 757-771, 2016.

[15] I. Mozetič, et al., “Multilingual Twitter Sentiment Classification: The Role of Human Annotators,” *PLOSONE*, vol. 11, no. 5, 2016.

[16] M. E. Basiri and A. Kabiri, “Uninorm operators for sentence-level score aggregation in sentiment analysis,” *4th International Conference on Web Research (ICWR*), Tehran, pp. 97-102, 2018.

[17] S. Sahu, S. K. Rout and D. Mohanty, “Twitter Sentiment Analysis - A More Enhanced Way of Classification and Scoring,” *2015 IEEE International Symposium on Nanoelectronic and Information Systems,* pp. 67-72, 2015.

[18] L. Brönnimann, “Multilanguage sentiment-analysis of Twitter data on the example of Swiss politicians,” M.Sc. Thesis, University of Applied Sciences Northwestern Switzerland, 2014.

[19] N. F. Alshammari and A. A. Al Mansour, "State-of-the-art review on Twitter Sentiment Analysis," *2nd International Conference on Computer Applications & Information Security (ICCAIS)*, Riyadh, Saudi Arabia, pp. 1-8, 2019.

[20] A. Ghallab, et al., "Arabic Sentiment Analysis: A Systematic Literature Review," *Applied Computational Intelligence and Soft Computing*, pp. 1-21, 2020.

[21] M. Al-Amin, et al., "Sentiment Analysis of Bengali Comments with words2VCE and Sentiment Information of Words," *2017 International Conference on Electrical, computer and Communication Engineering (ECCE)*, pp. 186-190, 2017.

[22] S. Narr, et al., "Language-Independent Twitter Sentiment Analysis," *Workshop on Knowledge Discovery, Data Mining and Machine Learning (KDML 2012)*, 2012.

[23] Nankani H., et al., "Multilingual Sentiment Analysis," in Agarwal B., et al. (eds), "Deep Learning-Based Approaches for Sentiment Analysis," *Algorithms for Intelligent Systems*, pp. 193-236, 2020.

[24] H. Saif, et al., "Alleviating data sparsity for Twitter sentiment analysis," *2nd Workshop on Making Sense of Microposts*, pp. 2-9, 2012.

[25] Y. Wu, et al., "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation," *arXiv: 1609.08144*, 2016.

[26] Q. V. Le and M. Schuster, "A Neural Network for Machine Translation, at Production Scale," Google AI Blog, Sep 2016. [Online], Available: https://ai.googleblog.com/2016/09/a-neural-network-for-machine.html.

[27] S. Strassel, et al., "Handbook of Natural Language Processing and Machine Translation: DARPA Global Autonomous Language Exploitation," *Springer-Verlag*, 2011.

[28] L. Specia, et al., "Quality Estimation for Machine Translation," *Morgan & Claypool Publishers*, 2018.

[29] Christopher Olah, "Understanding LSTM Networks," 2015. [Online], Available: https://colah.github.io/posts/2015-08-Understanding-LSTMs/.

[30] Jesus Rogel-Salazar, "Data Science and Analytics with Python," *1st. ed., Chapman & Hall/CRC*, 2017