



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 5, May 2023



Impact Factor: 8.423

9940 572 462

6381 907 438

ijrset@gmail.com

www.ijrset.com

Sentiment Manifestation of Kannada Tweets

Sandeep B¹, Adithya S Nair², Anirudh D Shasthry³, Atishay SG⁴, Disha R⁵

Assistant Professor, Dept. of CS&E, JNNCE, Shimoga, India¹

Research Scholar, Dept. of CS &E, JNNCE, Shimoga, India²

Research Scholar, Dept. of CS &E, JNNCE, Shimoga, India³

Research Scholar, Dept. of CS &E, JNNCE, Shimoga, India⁴

Research Scholar, Dept. of CS &E, JNNCE, Shimoga, India⁵

ABSTRACT- In the current world scenario, Internet has become a major platform for online learning, exchanging ideas and sharing opinions. Social networking sites like Twitter, Facebook, Instagram are rapidly gaining popularity as they allow people to share and express their views about topics, have discussion with different communities, or post messages across the world. The main goal of sentiment analysis is to detect and analyse attitude, opinions or sentiments in the text. Sentiment analysis has reached its popularity by extracting knowledge from huge amount data present online. The Process of analysis includes selecting features and opinion which is a challenging task in languages other than English. In this project, we have targeted one of the most promising and widely used social media platform-Twitter where opinions are shared as “Tweets”. We have specifically chosen the tweets made in Kannada and analyse the sentiment of those tweets. Kannada is a Dravidian language spoken majorly in the south Indian state of Karnataka, with minorities in all neighboring states. Emotion analysis is the method of defining and evaluating the emotions conveyed in textual data. Emotion detection and classification are straightforward tasks that can be completed based on the emotions conveyed in the text, such as fear, rage, happiness, sorrow, affection, motivation, or neutral.

KEYWORDS: Sentiment Analysis, Tweets, Kannada, Opinions, emotion analysis

I. INTRODUCTION

Social Networking has evolved to become one of the major sources of information. Micro blogging websites are one among these social networking websites. Social Media platforms such as Facebook, Instagram, Twitter, and Google+ have evolved itself through technology to become a major part of people’s lives. Twitter is one of the very popular micro blogging website where the users have freedom to create and share short messages, also called as “tweets”. Sometimes, these tweets are used to express the opinions as well as comments of the users on different topics, that maybe political, social etc. Sentiment as defined in the oxford dictionary “is a feeling or an opinion, especially one based on emotions”. A manifestation is the public display of emotion or feeling, or something theoretical made real. Sentiment manifestation means being able to display sentiment and emotions into the real world. Kannada is a Dravidian language spoken majorly in the south Indian state of Karnataka, with minorities in all neighboring states. Kannada is known to be one of the oldest languages known to mankind. It has around 47 million native speakers, and was additionally a second or third language for around 13 million non-native speakers in Karnataka. The proposed work explains sentiment manifestation of Kannada tweets, which have been classified into positive and negative using different machine learning algorithms. One of the significant challenges in the sentiment manifestation of tweets is the unavailability of sentiment tagged corpus. Therefore, one of the challenges in presenting a sentiment tagged corpus is we will have to work with a limited quantity of data or create a dataset manually.

II. LITERATURE SURVEY

SarithaShetty et al., [1] have pointed out the importance of analysing product or movie reviews. Sentiment analysis of the tweets is much harder than the conventional documents we find.

The author proposed a system where we define three different models for English, Hindi and Kannada languages. The training inputs to these models are pre-processed dataset. The pre-processing involves removal of stop words, unnecessary characters, lemmatization in case of English language and finally tokenizing the sentence. A Bag of words model is created which contains a list of all words from the dataset. A unique integer mapping is created for each token and we use this to map our sentence to a new integer vector. This vector is given to each model for training along with

their corresponding positive or negative labels. After the training process, we save the model weights and the bag of words model.

KakuthotaRakshitha et al., [2] have stated that sentiment analysis is a vital current research area within the field of web page mining.

The authors have proposed a system of Natural Language Processing. Natural Language Processing (NLP) is a subpart of artificial intelligence and linguistics it mainly concentrates on the interaction between human language and computer focuses on how to handle or analyse a huge amount of regional language data. Therefore, the computer can be able to “understand” the concept present in the document. This technology then can particularly abstract the data included in the documents, and then categorize based on the polarity.

Sentiment analysis using Python is done using TextBlob which is a python package for implementing the sentiment analysis. Sentiments are determined based on the frequency of each word and semantic relations which are provided by the input data, which leads to accurate output.

The output of Text Blob’s polarity will be in the form of floating points that is in the range of[-1.0, 1.0].

In the research paper published by Satyam Dutta et al., [3] Machine Learning, Deep Learning, and Transformer models have been described. They have focused on a code mixed language, i.e. people who are multilingual (particularly non-native English speakers), have a tendency to code-mix in their primary language by employing English-based phonetic typing and inserting Anglicism. On the Kannada dataset, we first used several basic Machine Learning models like Random Forest, Support Vector Machine, and Logistic Regression with default values of hyper-parameters. The performances of these models were evaluated in terms of precision, recall and F1-score. The dataset consisted of a large number of invalid contents like emoji, URLs, and others. Hence, data cleaning and the preprocessing task is performed prior to developing the model, such as eliminating all the emojis, emoticons, and special symbols. We also removed all of the numbers and transformed the data to lowercase. For text encoding, we employed the n-gram Term Frequency-Inverse Document Frequency Vectorizer. We created a hybrid model combining Convolutional Neural Networks (CNN) [20, 21] and Bidirectional Long Short-Term Memory (Bi-LSTM) networks after the machine learning models failed to produce promising results for the hybrid model’s details, as well as the hyper parameters.

The authors Mr. Pavan Kumar P.H.V et al., [4] have discusses the methodology and results submitted to the shared task of sentiment analysis for Malayalam-English, Tamil-English, and Kannada-English languages. They have implemented three Deep Neural network architectures for classifying code-mixed text: Convolution Neural Network (CNN) combined with LSTM (CNN-LSTM), Bidirectional-Long Short Term Memory (Bi-LSTM) network, and Deep Neural Network (DNN) with one hidden layer.

The dataset used in the shared task contains bilingual and native texts of three different languages, Malayalam-English, Tamil-English and, Kannada-English.

Authors Clarissa Xavier et al., [5] have targeted the Brazil audience and said that Social media has a strong position in the lives of people in Brazil. To access tweets, the access to Twitter APIs is necessary. In order to access Twitter API, it is required to register an application. By default, applications can access only public information. The API search methods returns tweets corresponding to the search parameters. These parameters may include keywords, hash tags, phrases, regions, usernames, or ids.

III.PROPOSEDMETHODFOR SENTIMENT MANIFESTATION OF KANNADA TWEETS

SYSTEM ARCHITECTURE:

DESCRIPTION

Twitter: Twitter is an online social media and social networking service owned by American company Twitter, Inc., on which users post and interact with messages known as "tweets". Registered users can post, like, and retweet tweets, while unregistered users only have the ability to read public tweets. Users interact with Twitter

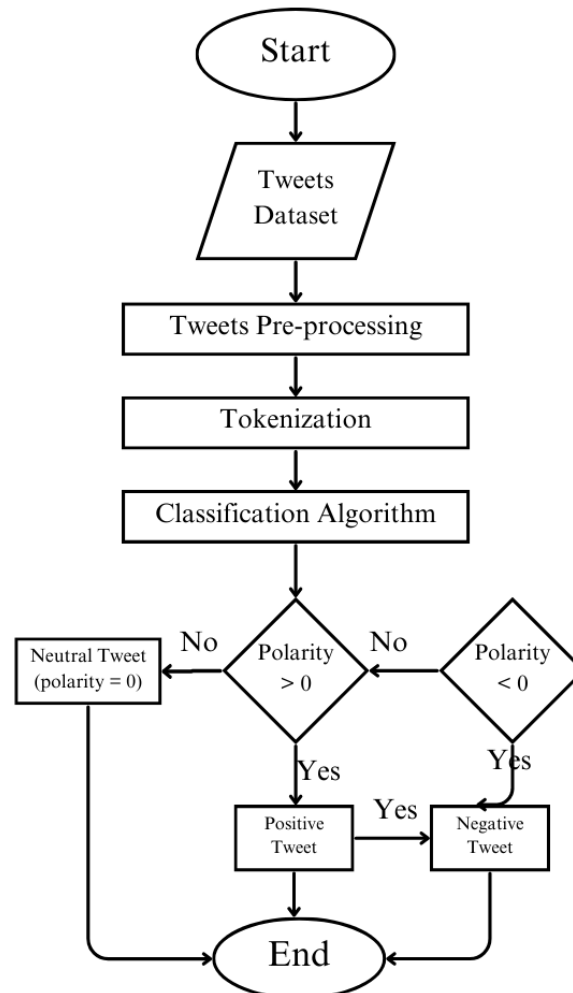


Fig 1 Flowchart

through browser or mobile frontend software, or programmatically via its APIs. Before April 2020, services were accessible via SMS. Tweets were originally restricted to 140 characters, but the limit was doubled to 280 for non-CJK languages in November 2017. Audio and video tweets remain limited to 140 seconds for most accounts.

Tweets Dataset:

Twitter dataset includes tweets, retweets, and replies that has ever been published containing a hashtag, keyword, account or user mention. Datasets are something that you can leverage to better understand the likes and preferences of your target audience and fine-tune your marketing strategy accordingly.

Data Pre-processing:

A tweet contains a lot of opinions about the data which are expressed in different ways by different users. The twitter dataset used in this survey work is already labelled into two classes' viz. negative and positive polarity and thus the sentiment analysis of the data becomes easy to observe the effect of various features. Pre-processing of tweet include following points,



Classifier:

In data science, a classifier is a type of machine learning algorithm used to assign a class label to a data input. Sentiment classification is the automated process of identifying opinions in text and labelling them as positive, negative, or neutral, based on the emotions customers express within them.

Positive and Negative:

Once the task of finding whether a sentence is opinionated is done, we have to find the polarity of the sentence i.e., whether it expresses a positive or negative opinion. Sentiment classification can be a binary classification (positive or negative), multi-class classification (extremely negative, negative, neutral, positive or extremely positive), regression or ranking.

RESULTS

The data for this study consisted of a twitter dataset which consisted of over 11020 tweets on corona virus and about the pandemic era. The posts were randomly sampled and represented a diverse range of topics and emotions. The posts were analyzed using natural language processing techniques to identify and classify emotional content. Sentiment analysis was done on this dataset which gave us the polarity and sentiment of the texts as shown in Fig 1. The analysis identified a total of 5 emotions in the social media posts: happy, angry, surprise, sad and fear as shown in Fig 2. Logistic regression technique for predicting the probability of a binary outcome based on one or more predictor variables was used. The precision, recall, and F1 score are commonly used metrics to evaluate the performance of a logistic regression model and were recorded as shown in Fig 3.

	text	polarity	sentiment
0	Same folks said daikon paste could treat a cyt...	0.000000	Neutral
1	While the world has been on the wrong side of ...	-0.500000	Negative
2	#coronavirus #SputnikV #AstraZeneca #PfizerBio...	0.000000	Neutral
3	Facts are immutable, Senator, even when you're...	-0.050000	Negative
4	Explain to me again why we need a vaccine @Bor...	0.000000	Neutral
5	Does anyone have any useful advice/guidance fo...	0.400000	Positive
6	it is a bit sad to claim the fame for success ...	-0.100000	Negative
7	There have not been many bright days in 2020 b...	0.675000	Positive
8	Covid vaccine; You getting it?\n\n #CovidVacci...	0.000000	Neutral
9	#CovidVaccine \n\nStates will start getting #C...	0.000000	Neutral

Fig 1 Sentiment Analysis



	text	polarity	sentiment	emotion
0	Same folks said daikon paste could treat a cyt...	0.000	Neutral	Surprise and Sad
1	While the world has been on the wrong side of ...	-0.500	Negative	Fear
2	#coronavirus #SputnikV #AstraZeneca #PfizerBio...	0.000	Neutral	
3	Facts are immutable, Senator, even when you're...	-0.050	Negative	Sad
4	Explain to me again why we need a vaccine @Bor...	0.000	Neutral	
5	Does anyone have any useful advice/guidance fo...	0.400	Positive	
6	it is a bit sad to claim the fame for success ...	-0.100	Negative	Surprise and Sad
7	There have not been many bright days in 2020 b...	0.675	Positive	Happy
8	Covid vaccine; You getting it?\n\n #CovidVacci...	0.000	Neutral	
9	#CovidVaccine \n\nStates will start getting #C...	0.000	Neutral	Sad

Fig 2 Emotion analysis

	precision	recall	f1-score	support
Negative	0.00	0.00	0.00	245
Neutral	0.47	1.00	0.64	1046
Positive	0.00	0.00	0.00	912
accuracy			0.47	2203
macro avg	0.16	0.33	0.21	2203
weighted avg	0.23	0.47	0.31	2203

Fig 3 Logistic Regression

IV.CONCLUSION

Social media like Twitter is the biggest platform where people are sharing their opinions and reviews about the people, place, products, etc. Out of all those reviews, differentiating sentiments behind each review is a difficult task. Because nowadays people are expressing their views in their regional languages. This can be achieved using TextBlob where text blob helps to classify the customer’s product reviews as positive, negative, or neutral along with the polarity and also calculated the sentiment score behind each customer’s review. So, this proposed method helps the people analyze the sentiment of kannada tweets. Also helps people know the emotions behind that specific tweet. The same proposed methodology can also be helpful to know the sentiment analysis in several sectors.

V. ACKNOWLEDGEMENTS

The authors would like to thank principal and staff of JNNCE for their support in this work.

REFERENCES

- [1] SarithaShetty et al., “Sentiment Analysis of Twitter Posts in English, Kannada and Hindi languages”, Publication at ResearchGate (2022 January).
- [2] KakuthotaRakshitha, et al, “Sentimental analysis of Indian regional languages on social media”, publishing services by Elsevier B.V (5 July 2021).



- [3] Satyam Dutta et alii., “Sentiment Analysis on Multilingual Code-Mixed Kannada Language”, FIRE 2021, Forum for Information Retrieval Evaluation, December 2021.
- [4] Pavan Kumar P.H.V et al., “Deep Learning Based Sentiment Analysis for Malayalam, Tamil and Kannada Languages”. FIRE 2021: Forum for Information Retrieval Evaluation, December 2020.
- [5] C. C. Xavier et al., “A Basic Approach for Extracting and Analysing Data from Twitter”, Springer Nature Switzerland AG 2020.
- [6] Soumya S., et al., “Sentiment analysis of Malayalam tweets using machine learning techniques”, publishing services by Elsevier B.V. 14 April 2020.
- [7] Vishal A. Kharde et al., “Sentiment Analysis of Twitter Data: A Survey of Techniques”, International Journal of Computer Applications (0975 – 8887) Volume 139 – No.11, April 2016.
- [8] Rohini V et al., “Domain Based Sentiment Analysis in Regional Language-Kannada using Machine Learning Algorithm”, IEEE International Conference on Recent Trends In Electronics Information Communication Technology, May 2016.
- [9] Maryam Hasan et alii., “Automatic emotion detection in text streams by analysing Twitter data”, Springer International Publishing AG, 16 January 2018.
- [10] Kashfia Sailunaz et al., “Emotion and sentiment analysis from Twitter text”, All rights reserved by Elsevier B.V. 2019.



Impact Factor: 8.423



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details