**SYNOPSIS**

**Report on**

**Sentiment Analysis**

**by**

Vanshita Gupta (2000290140130)

Nitin Goyal (2000190140079)

**Session:2021-2022 (4th Semester)**

Under the supervision of

**Prof. Vidushi Mishra**

### KIET Group of Institutions, Delhi-NCR, Ghaziabad



### Department Of Computer Applications

**KIET GROUP OF INSTITUTIONS, DELHI-NCR, GHAZIABAD-201206**

(march- 2022)

**ABSTRACT**

In this Project there are different machine learning techniques which are used for sentiment analysis. Mostly sentiment analysis done by using machine learning classifier like SVM (support vector machine), Random forest, Naïve Bayes. In this we are seeing some paper which are help new researcher to found a proper path for their new research. In this there is a proposed method of new research program. Social media is biggest medium to share people’s opinion on different topics. Sentiment analysis uses machine learning technique and without any human interruption machine will give and accurate sentiment of the people. Sentiment analysis turn text into positive, negative or neutral. So, any company or foundation or movie reviewer can take the opinion of the people and take further steps according that.

**TABLE OF CONTENTS**

1. Introduction 4
2. Literature Review 5-6
3. Project / Research Objective 7
4. Research Methodology 7
5. Project / Research Outcome 7
6. Proposed Time Duration 7

References 8

### 

### Introduction

Sentiment analysis is a machine learning tool which is used for analyze the texts for polarity from positive to negative. Machine automatic learn how to analyze the sentiment of the human without the human input or interruption. Nowadays social media is a part of the people’s life; people uses social media for give their review over some political field, movie review or marketing area. There are many social media sites like Twitter, Facebook, Instagram etc. They use this social media sites as the medium to express their view on many topics. So, sentiment analysis analyzes the text which inputted by any person from the different country by using the training data set it will analyze the sentiment of that particular text by knowing the emotion of that people. The application of the sentiment analysis very broad and powerful like Expedia Canada; Canadian take the advantage of sentiment analysis when they notice that people are giving negative comments on the music used by their television channel. Rather than chalking by negative comment, Expedia manages to take advantage of that negative comment and air all new soulful music in their channel.

### Literature Review

In this paper 1) Tweets are classify into the positive or negative comments using machine learning algorithm such as Naïve Bayes, Random forest (RF), Support vector machine (SVM), Unigram with Sentiwordnet and unigram with Sentiwordnet including negations are using as the input in this paper. Author derived three thousand one hundred eighty-four (3184) tweets using the tweeter API. Nine hundred fifty-four (954) positive, one thousand eighteen (1318) negative, 145 stop words have been identifying from 3184 tweets. Using. Author used feature of sentiment analysis like Bag of words (BOW), Term frequency vs Inverse document frequency (TF-IDF), Unigram with Sentiwordnet, Unigram with Sentiwordnet including negation words as an input. Author gets a conclusion that all the classifier with Unigram with Sentiwordnet and Unigram with Sentiwordnet including negation word shows higher accuracy the Bags of words (BOW) and term frequency vs Inverse document frequency (TF-IDF). Random forest algorithm with Unigram with Sentiwordnet including negation words get highest accuracy of 95.6%.

In this paper 2) authors try to use machine learning algorithm for Arabic customer’s feedback. They study two different type of methods which are voting and metaclassifier combination. They collecting data using Tweepy Application Programing Interface (API)17. There are many sarcastic and neutral tweets with the positive and negative tweets. Total 438,931 tweets were collected from that 75,774 are positive and 75,774 negative. Removing the al noisy data from the tweets like pictures, hashtags, retweets, emotions; second tokenization removing non Arabic letters, normalizing Arabic analogue letters. 10 classifiers NB, ME, LR, RR, PA, MNB, SVM, SGD and Ada boost BNB were used to extract and discover the polarity of given tweets. The highest accuracy achieved by PA and RR is 99.96%. Lowest accuracy achieved by Ada boost, LR and BNB which is less than 60%.

This paper 3) uses Amazon customer review data to find out the positivity, negativity and neutrality on customer’s review. In this they compare two machine learning algorithms Naïve Bayes algorithm and Support vector machine (SVM). The input is the customer review of the Amazon products. The review maybe negative, positive or neutral. Apriori algorithm is used to extract the frequently used aspects from the input dataset. Sentiwordnet is used to calculate positivity, negativity and neutrality score and after that the classifier will apply. The comparison of the algorithm based on the performance can be calculated by using the Accuracy, Precision, Recall and F-1 Measure of each classification. By the experimental result Naïve Bayes classification is batter accuracy then Support vector machine (SVM). Calculation were done by True positive sample (TP), False positive samples (FP), True negative samples (TN) and False negative samples (FN).

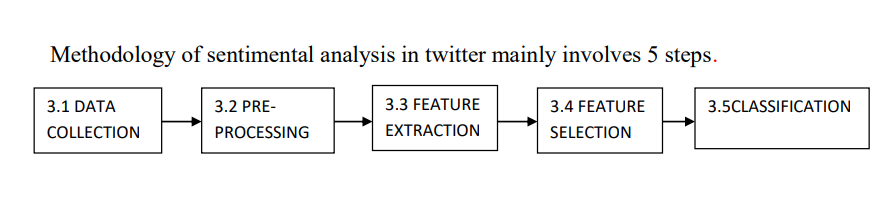
In this paper 4); There are many unsolicited email campaigns are one of the biggest threats affecting the users. Author combine both Sentimental analysis and personality recognition for analyze the email content. They use two different datasets to validate the proposed method. The first dataset is original dataset (CSDMC 2010 dataset) and second dataset validation dataset (TREC 2007). CSDMC 2010 spam corpus: - This composed 2949 emails messages to carry out original experiments. TREC 2007 public corpus: - In this there are 75419 emails in which 25220are legitimate 50199 spam emails. This method validated in two different datasets improving the best accuracy in the both the cases (from 99.15% to 99.24% and 98.98% to 99.18%). Further this method is also using for different validation like SMS and social media validation. This paper 5) shows; During the pandemic of the COVID19 whole word is suffering. Social media is the vast platform to share your thoughts any situations. Author uses the social media to analyze the people’s reaction on this situation. Author portray the fact that how irrationally people are behaving in this situation. It would be easier for victim to gather some structured information from social media. Two sets of datasets have been used in this paper. #corona, #covid19, #coronavirus mostly used for this survey. In dataset-1 there were 2,26,668 tweets used as the preliminary for dataset-2 they use the tweets which were retweeted most. To fit in the model data have been categorized in train, validation and test sets. To show the accuracy unigram, bigram and trigram performed. The accuracy of dataset 1 is 81% and accuracy of dataset 2 is 75% using different classifiers. By the conclusion author came to know that social media is not useful enough to help people.

In this paper 5) author examine the Alzheimer disease stigma on twitter using machine learning technique Machine learning technique modeled stigmatization expressed in 31150 Alzheimer disease-related tweets collected via tweeter API. In this 1% of the dataset used to train a classifier the tweet and rest 99% of the dataset. In this paper author discuss that how social media outlet affect attitude bearing in other development outcomes. Retweet were removed, other tweets which are not related to Alzheimer were removed, the keywords “alz”, “Alzheimer”, “dementia”, “memory loss”, “senility” which defined the sample of analysis. Lastly they removed the username which contain the topic name they removed. Two researcher manual coding and result are as follow: 43.41% informative, 23.79% joke, 21.22% metaphorical, 19.29% organization, 24.50% ridicule.

### Research Objective

The main objective of sentiment analysis is **to review different algorithm and techniques to extract feature wise summary of a twitter data and analyze it to form an authentic review**. Sentiment analysis for a particular data can help companies know about these expectations before product launch.

### Research Methodology



### Research Outcome

Research Paper

### Proposed Time Duration

This project is estimated to be completed in 2 months.

**REFERENCES**

1) Soumya, S. and K. J. I. E. Pramod (2020). "Sentiment analysis of malayalam tweets using machine learning techniques."

2) Gamal, D., M. Alfonse, E.-S. M. El-Horbaty and A.-B. M. J. P. C. S. Salem (2019). "Implementation of Machine Learning Algorithms in Arabic Sentiment Analysis Using N-Gram Features." 154: 332-340.

3) Vanaja, S. and M. Belwal (2018). Aspect-level sentiment analysis on e-commerce data. 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), IEEE.

4) Ezpeleta, E., I. Velez de Mendizabal, J. M. G. Hidalgo and U. J. L. J. o. t. I. Zurutuza (2020). "Novel email spam detection method using sentiment analysis and personality recognition." 28(1): 83-94.

5) Chakraborty, K., S. Bhatia, S. Bhattacharyya, J. Platos, R. Bag and A. E. J. A. S. C. Hassanien (2020). "Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media." 97: 106754.