ABSTRACT

Social media and Microblogging Sites have Transformed greatly due to their nature. That is, it has become a source of a disparate variety of information as people post their views on several topics, discuss current problems, issues, in addition to express positive thoughts that one faces in their daily life through Twitter, Facebook, etc. Any attitude, judgements, or thought that's caused by feelings is termed as the Sentiment. The user sentiments for specific entities are analyzed through tweets they post in Twitter with a procedure called Sentiment Analysis (SA). Social networking sites, blog and social media allows user to post their own content. However, as of a researcher's comprehension, many Social Media websites prompt researchers in addition to developers to collect and analyze data by releasing their Application Programming Interfaces (API). Nevertheless, those types of social media data encompass numerous defects that possibly hamper the Sentimental Analysis procedure. Some noise, such as Freestyle composing, typographical errors, incorrect grammar, along with abbreviations, can exist inside the text. Due to the limited dimensions, abbreviations, misspells, unstructured nature, and also slangs, it creates the investigators tiring while discovering the sentiments as of the concerned Twitter data. The significant purpose behind the thesis is to design a method to efficiently classify the user's opinion from unstructured tweets. This thesis provides 'Two' significant contributions.

The First contribution is that the data is accepted as of their Twitter database, and it's preprocessed wherein the stop word elimination, stemming, tokenization along with number removal are done. These pre-processed data's are inputted to the Hadoop Distributed File System (HDFS) which eliminates the repeated words using its Map-Reduce technique. Following that, both the

emoticons and non-emoticons are extracted as features. Grounded in their significance, the result features are rated. Then, the Deep Learning Modified Neural Network (DLMNN) does the classification and the Particle Swarm Optimization (PSO) perform weight optimization. Along with the attained results from the above process is validated using the K-fold Cross-Validation (KCV) methodology.

The Next contribution is that with the support of the GBDT (Gradient Boosted Decision Tree) classification algorithm a proficient Sentiment Analysis (SA) and Sentiment Classification (SC) of Twitter data, is proposed suggested. Originally, Twitter information is inputted for pre-processing. Afterward, these data are rendered to HDFS-Map Reduce. The features as of the processed data are extracted and from them, the I-EHO (Improved Elephant Herd Optimization) algorithm selects the efficient features. The rating values are gauged For each selected feature, and it's inputted into the classifier. The Gradient Boosted Decision Tree classifier classifies the data into negative, positive, or neutral.