Tweet Classifier

Introduction:

Text classification is one of the problems that appeared in the natural language processing and big data analysis. It is mainly about how to assign each text to a specific class or category. Some problems are binary classification where the number of class are two such as spam or non-spam mails problem. On the other hand, there are multiple classes' classification such as movie genre classification. By using some NLP and machine learning algorithms, this problem can be solved.

Problem Statement:

Twitter has become as much of a news media as a social network, and much research has turned to analyzing its content for tracking real-world events, from politics to sports and natural disasters. Twitter users tweet their views in the form of short text messages.

Twitter generates 340+ million tweets per day, twitter is becoming a major source of information. This makes tweet classification a challenging problem for the Machine Learning researchers Twitter topic classification is classifying the tweets in to a set of predefined classes. We have a dataset of tweets and need to assign each tweet to the appropriate tags such as

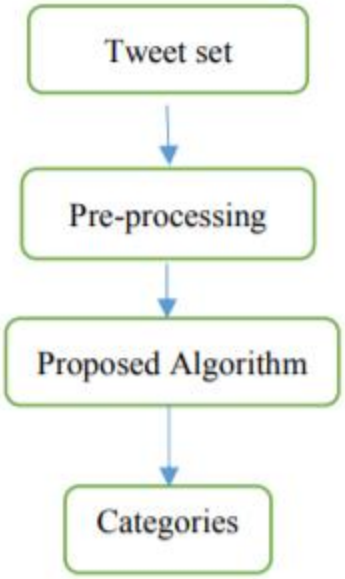
1. Technology
2. Business
3. Politics
4. Entertainment
5. Sports
6. Health

Dataset:

The dataset is obtained from Twitter API.

Problem Solution:

1. Feature extraction using natural language processing algorithms.
2. Use the extracted features to train the Multinomial Naive Bayes Classifier.
3. Use the extracted features to train SVM classifier.
4. Compare between the performance of each classifier



Team Members:

Karim Mohamed Ibrahim - 201301883

Omar Tarek Merghany - 201304322