
Multinomial Naïve Bayes Classifier

Sentiment Analysis is the process of determining whether a piece of writing is positive, negative or neutral. In our approach, we proposed a machine learning approach to analyze emotions in text and social media content in EMMA. The core component of this model is a classifier model. In this model emotions are classified into 6 basic emotions identified by Ekman (Anger, Disgust, Fear, Happiness, sadness, Surprise). The data set is created from the posts shared on social medias and combining existing datasets. Altogether, the dataset is contained with the model accomplishes the feature selection, classification and result conferment in the system. The will consist of two phases; training and prediction. The training phase is used for the purpose of training and build the model.

The test phase is a view for the testing the accuracy of the prebuilt model

Raw data is arranged in a proper way by filtering some words in data preprocessing. Feature extraction helps to identify the most relevant features in building feature vector. By using a vectorizer object of scikit-learn package, it translates the textual collection of reviews into normalization to a sparse matrix of occurrence counts. After extracting features from given dataset, the most relevant features for classification identified through TF-IDF Score. TF-IDF is used for stop-words filtering in various subject fields including text summarization and classification. Several experimental models have been followed up to come up with the best-suited model. We have experimented our data set with Naïve Bayes and linear SVM since the efficiency of these models are higher in supervised learning setting.

Multinomial Naïve Bayes classifier is used to train the model since relatively its performance is higher and less complicated. Naïve Bayes is a probabilistic algorithm that is very simple and so efficient in performing. In addition, this classifier is very easy to build and particularly very effective in larger datasets. We experimented our model by training with Bernoulli and Multinomial Naïve Bayes classifiers which are variations of Naïve Bayes. Naive Bayes algorithm is derived from the Bayes theorem that states; Multinomial Naïve Bayes considers the counts of multiple features occur while Bernoulli Naïve Bayes considers counts for a single feature that occur and counts for the same feature that do not occur. When analyzing emotions in text, it is not enough to focus on occurrences of a single keyword. Therefore, Multinomial Naïve Bayes is more appropriate to use as our text classification model rather than using Bernoulli NB. Bernoulli Naïve Bayes classifier is more appropriate in classifying Spam or Adult Content Detection with accurate results.