# Using twitter to do analysis on Oscars

**.**

**Team members:**

**1.K.Arya Prakash**

**2.Srinivas Attili**

**3.Vennel Mudireddy**

**Introduction:**

In our day to day life we see machine learning and analysis of open data set useful to the developers across the world to build a tool to explore some useful information . We would like to use data from our twitter accounts and to use the sentiment analysis on the data present on twitter regarding the current Oscars. With latest accusations in Oscars on it being white has sparked a great deal of debate. Somewhere positive about it and somewhere negative. Some hash tags like #OscarsSoWhite is causing a lot of backlash. Some has termed it as causing a genuine conversation.

**Research Question:**

We're going to explore a test set of tweets about Oscars, specifically tweets which included the hash tag #OscarsSoWhite,and the Tweepy API was used for obtaining the data from Twitter. Exploratory data analysis of Twitter data.  
Development of API to classify text; endpoints non-functional until classifiers complete. Development of classifiers; creation of test and training sets; tuning of classifiers.  
Development of build and deployment script. Development of API test harness.

**Data Extraction and Modelling**

The main file is result-oscars-with-header.csv this is the filw which contains all tweets on oscars collected over a period of time.

the training data is in result-oscars-train.csv

this data is normalized to get equal counts of features. The normalized trainig data is result-oscars-normalized-train.csv

**Cleaning Data:**

The test data is in the file result-oscars-test.csv

**Extracting Sentiment:**

The final classification results are in the file result-oscars-with-labeled.csv and the classification is present in the column 'classifier'.

we are using nltk NaiveBayesClassifier as classifying algorithm. We are first extracting features from training set creating a

word list and using these features and word list create training set which is passed to classifier.

Each of the tweet we are classifying as either info/funny/pos/neg