**HW3 Report**

**Files Description:**

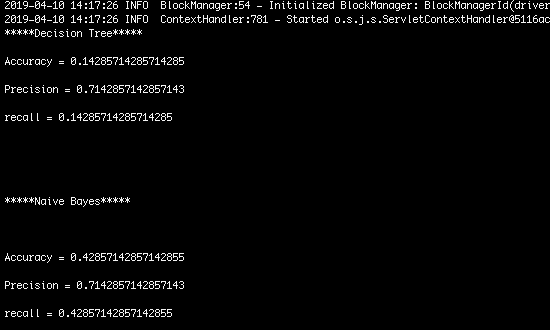
stream\_producer.py : modified to generate offline data for training

training.scala : Loading of data and then doing the pre-processing tasks - tokenizer, stop words removal, tf-idf. Training the dataset on Decision Tree Classifier and Naive Bayes Classifier

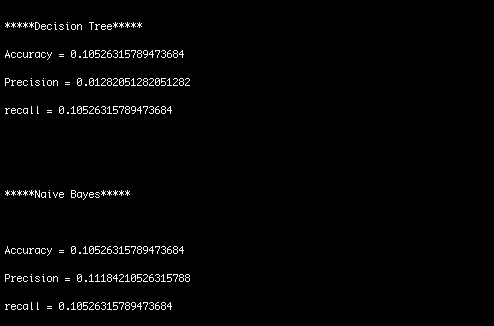
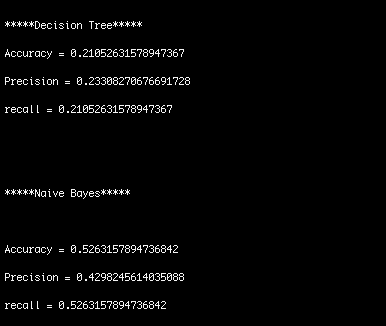
streaming.scala : Fetching the news from kafka server in intervals of 20 seconds and using trained model does the classification process and displays the performance results(accuracy,precision,recall)

**Results:**

Results of the HW3 are displayed on the terminal when running the streaming.scala file. It shows accuracy, precision,recall of the classification. Below are the screenshots for the same.







Also, results are visualized using Kibana and Elastic Search.

Over here, I have displayed results of trending news category in a window time frame by using a Tag cloud. Screenshots are below.

