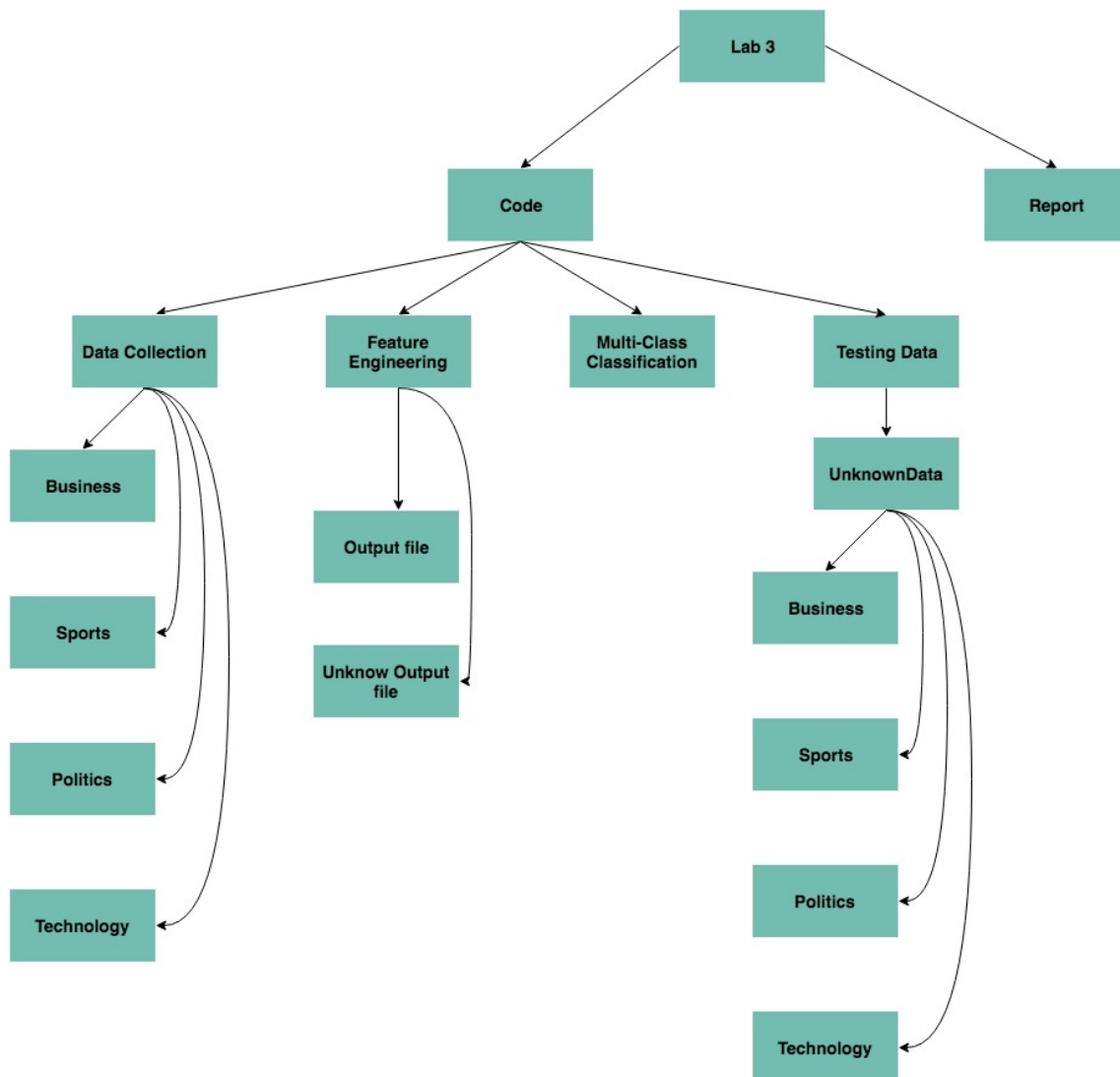


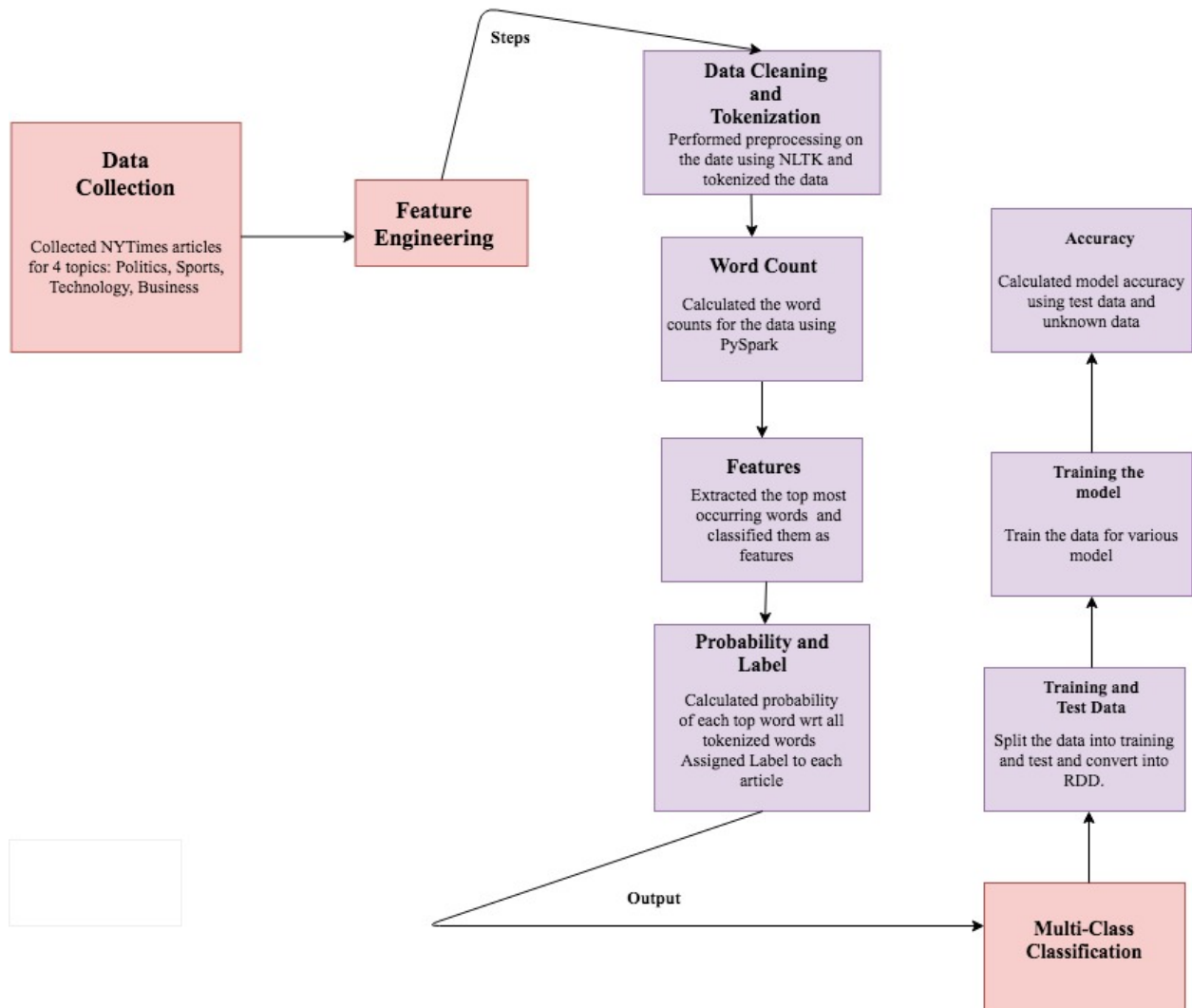
**CSE4/587 Data-intensive Computing Spring 2017**  
**LAB 3: DATA ANALYTICS PIPELINE USING APACHE SPARK: B. RAMAMURTHY**

**PRIYA MURTHY:50248887**  
**SHEFALI SHARMA: 50247677**

**Directory Structure:**



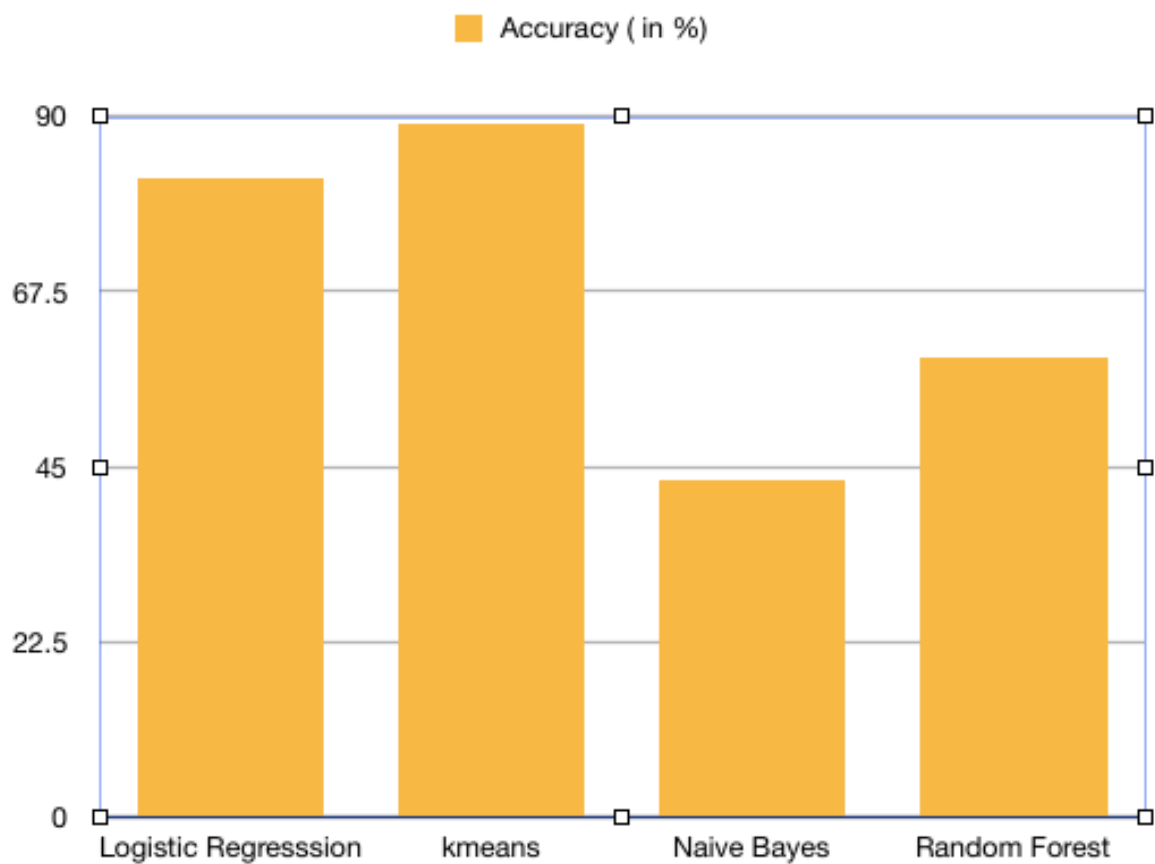
## DATA PIPELINE:



Model Accuracy for Test data:

Accuracy for Test Data

Model	Accuracy ( in %)
Logistic Regresssion	82
kmeans	89
Naive Bayes	43.21
Random Forest	59



### Model Accuracy for Unknown data:

Accuracy for Unknown Data

Model	Accuracy ( in %)
Logistic Regresssion	94.65
kmeans	89.62
Naive Bayes	40.9
Random Forest	31.1

