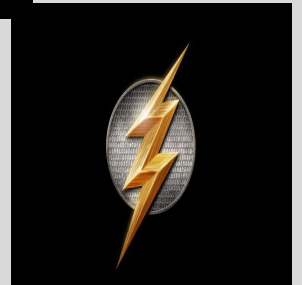
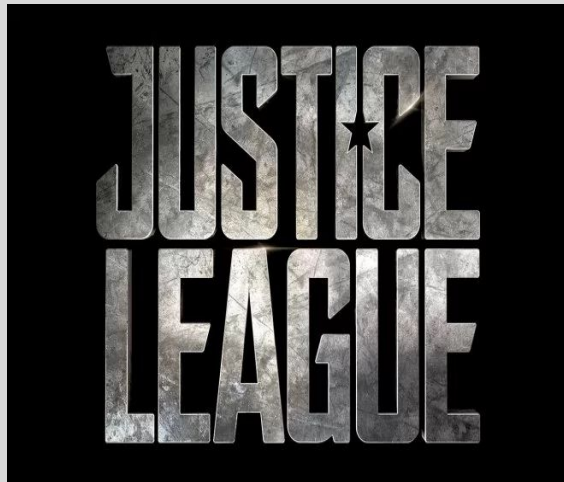
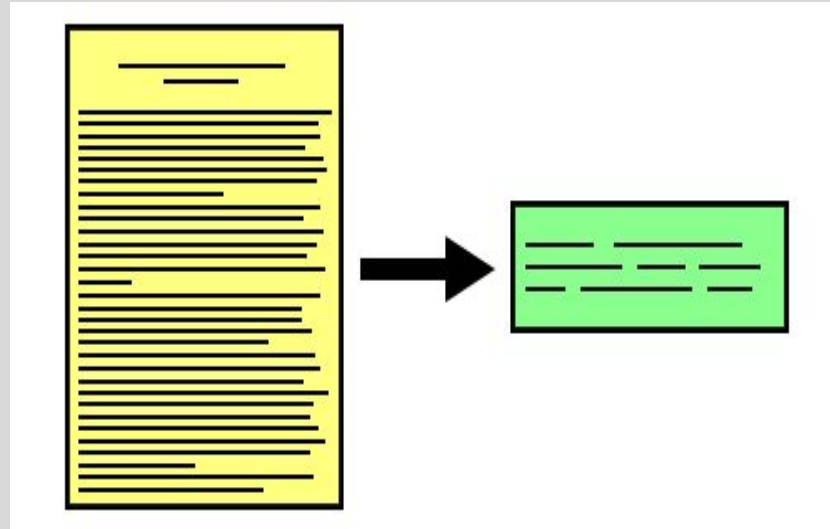


# Automated Document Summarization

Machine Learning - Term Project



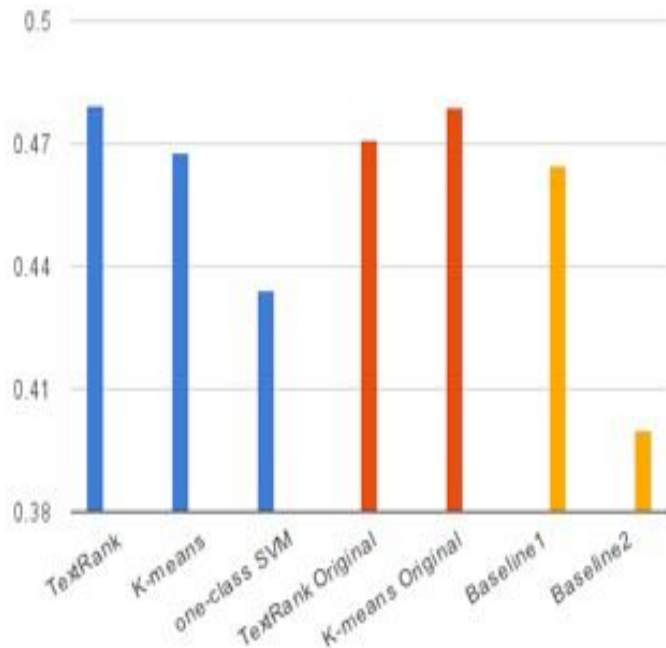
# Problem Statement/Objective

To build an **Automatic Text Document Summarization System** to help users to obtain essence of large documents with least effort using a **Machine Learning Approach**.

Wikipedia defines Automated Text Summarization as:

Process of reducing a text document with a computer program in order to create a summary that retains the most important points of the original document

# Some of the Existing Methods and their performances



Method	Rouge N-gram
TextRank	0.4797
K-means	0.4680
One-class SVM	0.4343
TextRank Original	0.4708
K-means Original	0.4791
Baseline-1	0.4649
Baseline-2	0.3998

# Method of approach

- Follows a statistical classification problem
- Summarizer is to be trained with a set of documents and their respective gold standard summaries using supervised learning algorithms
- Training involves selecting among potential features and choosing a weighted combination of these to produce a good scoring scheme
- To accomplish this, a classification function that estimates the probability of a sentence being selected in the summary is to be developed
- The estimated probability can be expressed using the Bayes' rule as follows:

$$P(s \in S | F_1, F_2, \dots, F_k) = \frac{\prod_{i=1}^k P(F_i | s \in S) P(s \in S)}{\prod_{i=1}^k P(F_i)}$$

(where  $s$  is a particular sentence,  $F_i$  is a feature and  $S$  is the set of sentences in the summary)

# Dataset / Implementation Details

**Coding platform** : Python

**Dataset** : The **CNN News Corpus** (500 docs. with summaries)

**Description** : Contains 2 archives training set, test set in .tgz format

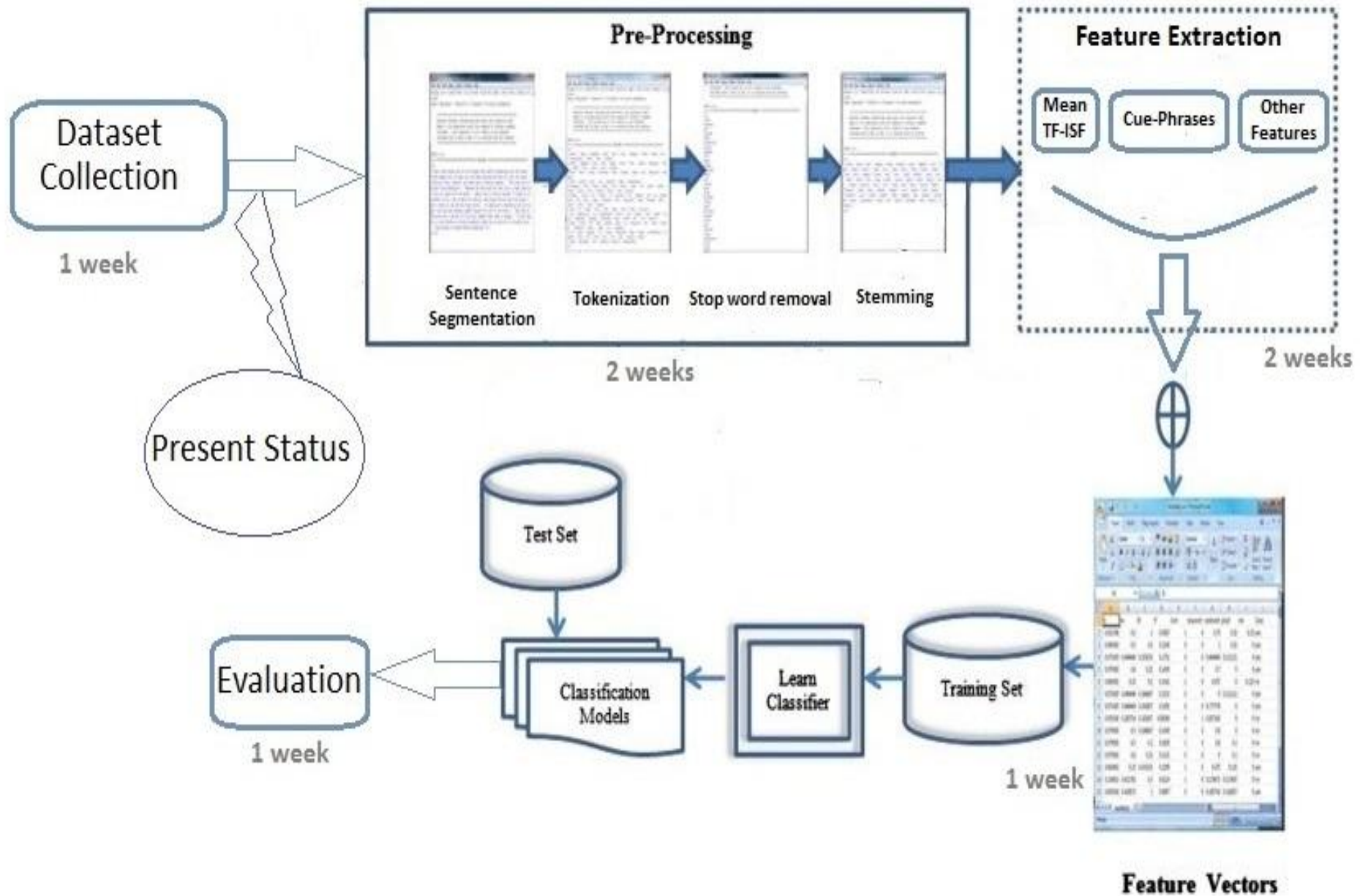
## Document before Summarization

` Enchanted ' star Amy Adams engaged  
Now Amy Adams ' love life is ` Enchanted , ' too .  
The Oscar-nominated Adams is engaged to fellow actor  
Darren Legallo , her publicist , Cari Ross , confirmed  
Thursday .  
Ross did not immediately provide further details .  
The couple reportedly met in acting class in 2001 , and  
plan to tie the knot next year .  
Adams , 33 , received an Oscar nomination for her  
supporting role in the 2005 family drama ` Junebug . '  
She had a breakout role as a Disney-style princess in the  
2007 romantic comedy ` Enchanted , ' and has appeared in  
such films as ` Charlie Wilson 's War , ' ` Talladega Nights :  
The Ballad of Ricky Bobby ' and ` Miss Pettigrew Lives for a  
Day

## Summary

Amy Adams engaged to actor Darren Legallo, publicist says. Couple met in acting class in 2001, plan to tie the knot next year. Adams, 33, nominated for Oscar for supporting role in "Junebug". She also played Disney-style princess in romantic comedy "Enchanted"

# Road Map



Thank You