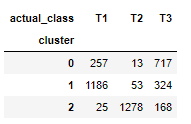
**Assignment 7 Analysis:**

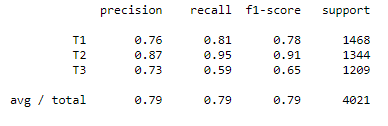
Task 1) K-Mean Clustering

Results:

Confusion Matrix:



Classification Report:



Based on cluster centroids/samples, give a meaningful name (instead of T1, T2, T3) to each cluster.

Top words for each cluster are as follows:

Cluster 0: said; crash; bus; rail; plane; train; police; passengers; car; cruise; airlines; speed; flight; road; says; driver; traffic; accident; airport; people

Cluster 1: said; oil; people; bp; japan; water; spill; gulf; disaster; nuclear; earthquake; pakistan; tsunami; quake; floods; coast; million; plant; government; killed

Cluster 2: percent; tax; said; year; economy; rate; obama; comment; government; economic; billion; budget; debt; bank; growth; new; market; spending; jobs; report

Cluster Names:

Cluster 0 (T3): Travel

Cluster 1 (T1): Bad News Natural and Manmade

Cluster 2 (T2): Finance

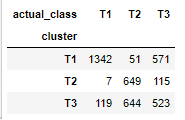
Task 2:

1. Print out: Feature Names and Shapes:

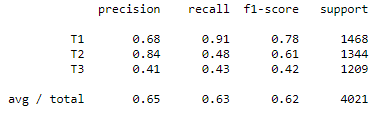
['abandoned', 'abc', 'ability', 'able', 'aboard', 'abroad', 'absolutely', 'abuse', 'accept', 'access']

(4021, 2749)

1. Confusion Matrix:



1. Classification Report:



1. Display Topics:

Topic 0:

said people comment oil news sign users water japan police city bp officials report killed says new told year crash hit time government rate coast plane disaster world nuclear area state just gulf passengers spill million video miles ap earthquake

Topic 1:

percent said year economy rate market economic government growth bank billion new debt report china financial prices month week global news unemployment jobs recession markets fed banks crisis data rates time quarter months expected recovery world european finance airlines central

Topic 2:

said tax new obama year com million state president rail high house money budget federal government billion business years time www people plan service security speed health information make public travel program work pay taxes cruise services help company spending

1. Cluster Names:

Topic 0: Accidents and Disaster

Topic 1: Global Finance

Topic 2: Government Topics

Analysis:

For me, I am getting a higher accuracy for K-Mean than LDA.

But if we increase clusters than the accuracy will increase for both the models, I assume that.