



Program: BS (CS)
Semester: Spring 2022
Course: CS4051 - Information Retrieval

Assignment#02
Marks: 10
Instructor: Waqas Ali

1. You are given two files (tab separated values) named **film-genres-train.tsv** and **film-genres-test.tsv**. As the names suggest one file contains the training dataset and the other contains the test dataset. Both the test and train datasets contains some text describing each movie and the genre of the movie. You have to solve a classification problem where you are required to predict the genre of a movie (where the description of the movie is available). You are required to use the Naive Bayes classifier to solve this problem.

You are required to provide the *confusion matrix* and calculate the *Precision*, *Recall* and *Accuracy* for each class (genre of the movie).

	Actual: Yes	Actual: No
Predicted: Yes	tp	fp
Predicted: No	fn	tn

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fn + fp}$$