COMP6930 - Assignment #4
(Due Date: 24th April 2018 @ 11:59 PM)

Instructions: You are to submit a **.zip** folder named using your UWI ID number to inzamam.rahaman@outlook.com.

Part A - Diabetes Dataset Analysis (10 marks)

The file *diabetes.csv* contains data on several patients and an indication on whether they are suffering from Type II diabetes or not. Use an appropriate Naive Bayes classifier and Linear classifier to predict from the presented patient data whether they are afflicted or not. Compare the results of these two classifiers. What aspects of the data points do you believe contributed to the relative accuracy of these two classifiers? Which of the two classifiers do you believe to be better and under what circumstances? If you used a Naive Bayes method provided by *sklearn*, what extra data preprocessing steps or modifications to the method itself could have yielded better performance? You **do not** need to code these alternative steps/models.

Part B - Food Sentiment (10 marks)

The file *reviews.csv* - stored at https://goo.gl/ZnZFk6 - contains several gourmet foods from Amazon. A review is said to express positive sentiment about a product if it scores the product with 3 or more stars and negative otherwise. Use an appropriate Naive Bayes method to train a model to predict the sentiment of food reviews using this data. Evaluate this model using F1 scores and a plotted confusion matrix. Do you believe that this model performs well? Justify you conclusion using your F1 score and confusion matrix. Regardless of performance, train and evaluate a non-Naive Bayes model and compare it to the Naive Bayes model you developed. Time the training of both the Naive Bayes method and non-Naive Bayes method. Which of the two models would you choose and why?

Part C - Manual Calculation (10 marks)

Consider the following table that contains data on the measurements of several flowers of two different species. Assuming that Petal Length and Petal Width are normally distributed with respect to the species, manually classify a flower of Petal Length = 5.2 cm and Petal Width = 2.3 cm. Show your working a *.pdf* file called *PartC.pdf*

Petal Length (Cm)	Petal Width (Cm)	Species
1.4	0.2	0
1.4	0.2	0
1.3	0.2	0
1.5	0.2	0
1.4	0.2	0
1.7	0.4	0
5.7	2.3	1
4.9	2.0	1
6.7	2.0	1
4.9	1.8	1
5.7	2.1	1
6.0	1.8	1