Sentiment Analysis with Machine Learning and R

**Total Marks: 10**

10 marks will be given, if detailed explanation of results is reported in this report by following steps from guided project documents

**Bonus: 5 Marks**

Additional 5 marks will be given to those who will develop a complete end to end project using **try out** tips in this document.

Note: For bonus, you don’t need to prepare another complete report, just bring out screen shot of experiment and results explaining the differences and your observation.

# Executive Summary

## Provide one paragraph summary of your findings: e.g.

## *This document presents use of Azure Machine Learning to build and operationalized sentiment analysis from tweet dataset.*

# Data Exploration

Explore data came along with this assignment. Explain what type of data is it. Upload data in Azure Machine Learning (AML) visualize and verify dataset contains the data you viewed in the source file. Provide screen shot of Data visualization in (AML)

## 

# Building Text Analytics Model

## Text Analytics model is built using following steps:

## Step 1: Clean and preprocess Data

The cleaning reduces the noise in the dataset, help you find the most important features, and improve the accuracy of the final model.

*We remove stopwords - common words such as ---- and numbers, special characters, duplicated characters, email addresses, and URLs.*

*We also convert the text to lowercase….*

*Stem the words, e. g.----*

## *Provide screen shot of experiment visualize result and provide screenshots before and after applying preprocessing*

**Try Out 1:** You can use **Preprocess Text Module** instead of using R code to preprocess data.Try it out and note differences between both options.

## Step2: Extract numeric feature vectors from pre-processed text

To build a model for text data, you typically have to convert free-form text into numeric feature vectors. In assignment you have used **Feature Hashing** to transform the text data to such format.

*Read more about “How feature Hashing works” from below link*

<https://msdn.microsoft.com/library/azure/dn906018.aspx>

*Run the experiment and provide screenshot showing how column look like after applying hashing.*

## **Try out 2:** You can use [**Extract N-Gram Features** from Text](https://msdn.microsoft.com/library/azure/mt762916.aspx) to transform the text data to numeric features. Try it out and note differences.

## Step3: Train classification or regression model

Now the text has been transformed to numeric feature columns. It is ready to train the classifier to transform into regular classification to predict sentiment scores.

*Explain which classification model is used and how training and testing data is split e.g.*

*e.g. The model was created using the Two-Class logistic regression algorithm and trained with 75% of the data and remining 25% for testing.*

**Try out 3:**You can try out other algorithm such as **Two Class SVM**, **Two Class Boosted Decision Trees** etc. to improve accuracy

You can also use other tools available in Azure Machine Learning to improve the model. For example, you have used **hyperparameter tuning** module to improve the accuracy.

**Try out 4:**Replace **hyperparameter tuning** module with **Train Model.** In the properties for the **Train Model**, use the column selector to select only the **sentiment** column.

## Step4: Score and validate the model

Model is evaluated by scoring it against the test dataset and accuracy is evaluated.

Testing the model with the remaining 25% of the data yielded the following results.

Provide screen shot of results and explain results based upon your observations.

## Step5: Deploy the model to production

You have deployed model as webservice which can take free text as input and return sentiment label and score as output.

Provide 10 rows of free text and use prediction to predict label. Copy and paste your results

# Bonus Points (5 Marks)

Explanation and screen shot of experiment and results using different module such as **Preprocess Text** for data cleaning, [**Extract N-Gram Features** from Text](https://msdn.microsoft.com/library/azure/mt762916.aspx) to transform the text data to numeric features, other classification algorithms such as **Two class SVM**, T**wo class Boosted Decision Trees** etc., and **Train Model.**

Note: it should be a complete experiment by applying above alternatives Module from AML along with deployment of webservice.