**CAPSTONE PROJECT PROPOSAL**

Prepared for:

CKME 136 - Data Analytics: Capstone Course

Prepared by:

A. B. M. Naimul Huq

Ryerson ID: 500672338

May 10, 2015

**Capstone Project Proposal**

### Brief context about the problem

A number of airlines from different operators are operating in different routes and assisting people to travel their preferred destination. Due to a great demand of air travel and strong competition between airlines, travellers are getting an increased choice of airline, airport, price and service. However, the standard of services varies. On the contrary, it is important for airline operators to remain competitive in terms of price and being preferred by travellers based on the service quality. Therefore, it is important for airline operators to understand traveller’s emotions and identify factors related to their services which might affect their brand preference.

**Statement of the Problem**

Air travellers have stated issues related to their travel experience through tweets which needs to be analyzed so that air line operators can identify, understand and fix those issues to improve their overall service quality.

**Data Sources**

We will be using the **“Airline Twitter Sentiment”** dataset from “**CrowdFlower”** website. Twitter data was captured from February, 2015.

#### <http://www.crowdflower.com/data-for-everyone> Added: February 12, 2015 by CrowdFlower

#### Data Analysis Techniques

The purpose of this report is to perform a **“Sentiment Analysis”** job about problems of each major U.S. airlines regarding their services. We will classify tweets based on emotion (e.g. anger, joy etc.) and polarity. In contrast to the classification of emotions, the classification of polarity function will allow us to classify some text as positive or negative. We will separate the text by emotions and visualize the words with a comparison cloud after removing stop words and applying stemming. This word cloud model will assist airline operators to understand traveller’s emotion / experience regarding their services.

1. Perform sentiment analysis
2. Perform exploratory data analysis
3. Perform ranking of words related with emotion;

#### Data Analysis Tools

1. Hadoop Ecosystem Tools - Data processing, Twitter Sentiment Analysis
2. R – Data processing, Twitter Sentiment Analysis, Exploratory Data Analysis
3. R gg plot – Data visualization
4. Tableau – Data visualization
5. GitHub – Reporting