Capstone Project

Airline Tweets Sentiment Analysis Final Report

**The Problem**

Many Twitter users tweet about there travel experiences, particularly with regard to Airlines, and the service they received on particular airlines. This can have an influence on the future purchase decisions of potential passengers, so it is important for Airline management to know how its Airline is perceived on Twitter.

The objective is to build a model than can analyses the sentiment of a tweet. That way a Website can be built around the model which can get a quantative information on an airlines twitter perception over a time period and examine the trend of that perception.

**Model Outcomes**

This will be a classification model with the outcome being either a negative or positive sentiment. It will be a supervised learning algorithm that will be used.

**Data Acquisition**

I plan to use the Airline Tweets Sentiment Dataset found on Kaggle, originally from Crowdflower. IF I get a chance I will validate the model on data I download using the Twitter API.

1. **Data Preprocessing/Preparation:** For this deliverable, you are tasked with detailing how you cleaned the data for your notebook.
2. What techniques did you use to ensure your data was free of missing values, and inconsistencies?
3. How did you split the data into training and test sets?
4. Please include any necessary analysis and encoding steps you took as well.

**Modeling**

I evaluated 7 different models

1. Logistic Regression
2. Support Vector Machine (SVM)
3. Decision Tree
4. Random Forest
5. K-Nearest Neighbors (KNN)
6. Naive Bayes
7. Neural Networks

**Model Evaluation**

**:** Share your model evaluation here. What types of models did you consider for your problem (classification, regression, unsupervised)?  Articulate the evaluation metrics you used and how you determined which model was most optimal for your problem.