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, positive, or neutral sentiment. It is be a supervised learning algorithm that is used.

**Data Acquisition**

Used the Airline Tweets Sentiment Dataset found on Kaggle, originally from Crowdflower.

A group of pie charts

AI-generated content may be incorrect.

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 Evaluation**

The problem required that I use classification model, and it was supervised. I evaluated the below 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 Train Time Train Accuracy Test Accuracy AUC Score

0 LogRegress 1.068357 0.923264 0.867224 0.943287

1 KNN 0.007240 0.378570 0.269266 0.621423

2 Decision Tree 3.433178 0.999303 0.761374 0.737086

3 SVC 135.185673 0.944973 0.778552 0.908386

4 Naive Bayes 0.000000 0.758649 0.732126 0.886241

5 Random Forest 29.060554 0.999303 0.803621 0.907840

6 Neural Net 241.408214 0.999303 0.836583 0.896468

A green bar graph with black text

AI-generated content may be incorrect.

**Model**

Articulate the evaluation metrics you used and how you determined which model was most optimal for your problem.