Airline Tweets Sentiment Analysis

Introduction

This project investigates the application of Machine Learning (ML) and Deep Learning (DL) techniques for sentiment analysis of airline-related tweets. The primary objective is to accurately classify tweets into one of three sentiment categories: **Positive, Neutral, or Negative**. By leveraging advanced natural language processing (NLP) methods, the project aims to extract meaningful insights from textual data, enabling airlines to better understand customer opinions, grievances, and satisfaction levels.

Team Members

- Shannon Kwok
- Javier Chin
- Tan Hon Jung
- Vanisha Muthu

Project Overview

We finetuned and evaluated several ML and DL models, integrating sentiment analysis while doing so. We also preprocessed the text data beforehand, due to informalities and inconsistencies in social media language. This was done using techniques such as lemmatization, converting emojis to text and removing special characters. Our project consists of 4 different models, which combine both ML and DL to achieve more accurate results.

Exploratory Data Analysis

We used visualisation plots such as barplots and word clouds to understand class distribution and conduct feature analysis in the data. More details on the methodology of our exploratory data analysis can be found in our Jupyter notebook.

Models Used

Machine Learning Models

1. Latent Dirichlet Allocation (LDA) with Convolutional Neural Network (CNN)
2. Bidirectional Encoder Representations from Transformers (BERT) Embeddings with K-means Clustering
3. GMM with BERT Embeddings

Deep Learning Models

1. DistilBert Classification with Bayesian Optimisation

Results

Our models yielded the following results:

Models	F1-score
Latent Dirichlet Allocation with Convolutional Neural Network	78%
BERT with K-Means Clustering	67.60%
GMM with BERT Embeddings	62%

DistilBert with Bayesian Optimization	83.59%

Detailed performance metrics including precision, recall, and accuracy are available in the results section of the project report in this repository.

Usage

To replicate our findings or use the models:
Close this pdf. Upload the Tweets.csv file manually to Google Colab. Run the Jupyter notebooks provided in the code/ directory to train the models on your data. Alternatively, you may run the notebooks on Google Colaboratory from the links in the respective notebooks.