Training Program on AI
Applied / Data Scientist

Twitter Sentiment Analysis

by

Suraj Kumar Yadav
(02106)

Organized by
Department of Computer Science,
Banaras Hindu University, Varanasi

Supported By:

MINISTRY OF ELECTRONICS & INFORMATION TECHNOLOGY GOVERNMENT OF INDIA

OUTLINE

- INTRODUCTION
- METHODOLOGY
- DATA DESCRIPTION
- RESULTS AND DISCUSSION
- REFERENCES

INTRODUCTION

In today's digital world, the internet holds a vast collection of opinions and feedback on products and services. The aviation sector, especially US airlines, is greatly influenced by online reviews and customer sentiment. Analyzing these reviews is vital for airlines to understand their performance and enhance customer satisfaction. Due to the exponential growth of the airline industry, automated sentiment analysis techniques are necessary. Our project focused on using machine learning algorithms to automate sentiment analysis for US airline reviews.

Problem statement

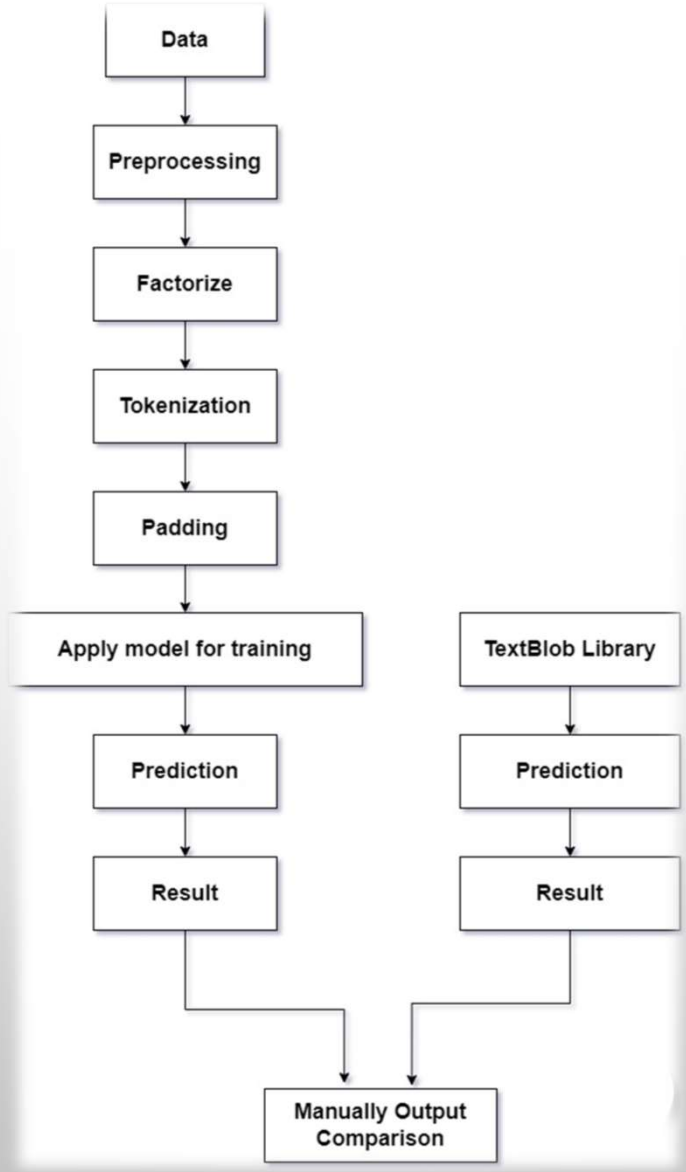
Classification of tweet sentiments into positive and negative categories to help the US airlines.

LIBRARIES USED

- Pandas
- Matplotlib
- Tensorflow
  - tensorflow.keras.preprocessing.text.Tokenizer
  - tensorflow.keras.preprocessing.sequence.pad_sequences
  - tensorflow.keras.models.Sequential
  - tensorflow.keras.layers.LSTM
  - tensorflow.keras.layers.Dense
  - tensorflow.keras.layers.Dropout
  - tensorflow.keras.layers.Embedding
- scikit-learn (`sklearn`)
  - WordCloud
- seaborn

METHODOLOGY

- This project uses an LSTM-based neural network with dropout layers for sentiment analysis of tweets. The model is designed with Keras and TensorFlow. It processes tweet texts, uses an embedding layer, LSTM layer with dropout, and Dense layer with sigmoid activation for sentiment probabilities. The model is trained with binary cross-entropy loss, Adam optimizer, and evaluated using accuracy. Dropout prevents overfitting. The goal is robust sentiment analysis by capturing tweet context and generalizing well to new data.


Data
Preprocessing
Factorize
Tokenization
Padding
Apply model for training
Prediction
Result
TextBlob Library
Prediction
Result
Manually Output Comparison

DATA DESCRIPTION

- We obtained our dataset from Kaggle, a popular online platform for data science projects. The dataset consists of a large collection of US airline reviews along with their corresponding sentiment labels. Each review in the dataset is labeled as positive, negative or neutral, but here we use only positive and negative, indicating the sentiment expressed by the reviewer. This binary sentiment classification setup allowed us to train a model that can accurately predict whether a given review reflects a positive or negative sentiment.

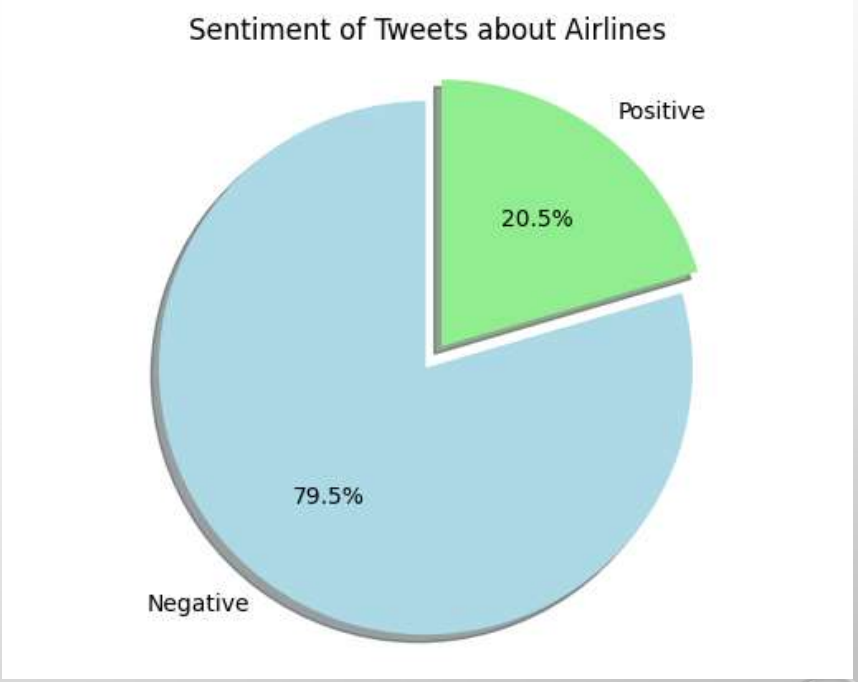
- It has 14640 rows and 15 columns, columns are: 'tweet_id', 'airline_sentiment', 'airline_sentiment_confidence', 'negativereason', 'negativereason_confidence', 'airline', 'airline_sentiment_gold', 'name', 'negativereason_gold', 'retweet_count', 'text', 'tweet_coord', 'tweet_created', 'tweet_location', 'user_timezone'.

- But we work on only two columns: 'text','airline_sentiment'

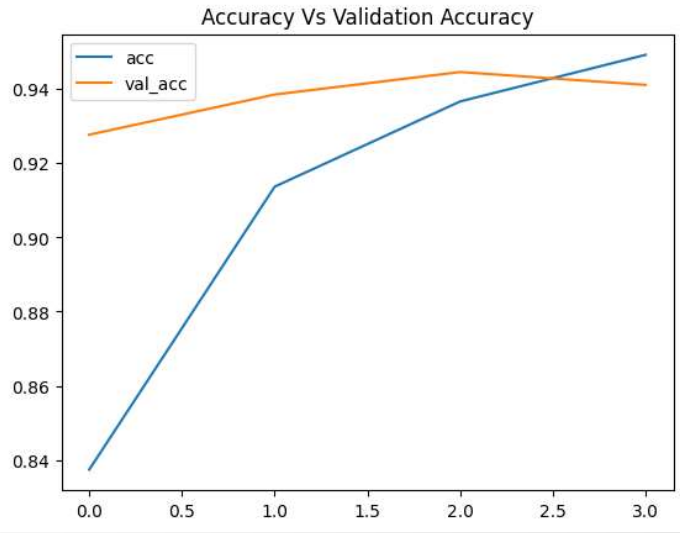
RESULTS AND DISCUSSION

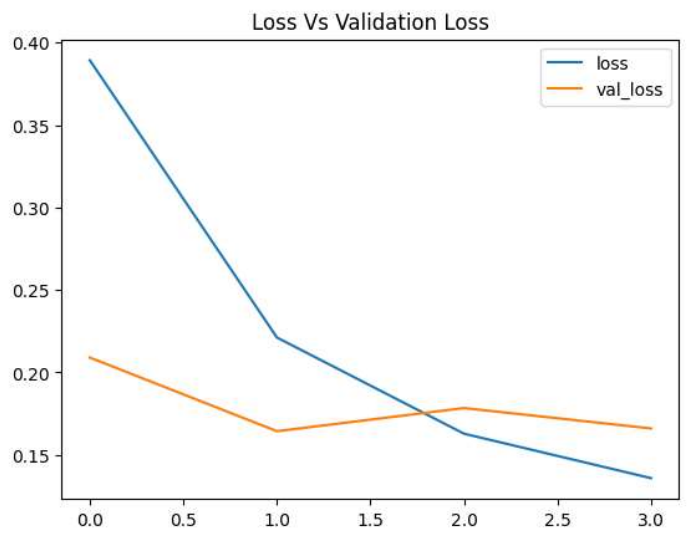
- Experimental Results

- Initially we have 79.5% negative and 20.5% positive tweets
- After performing the prediction from the trained model, we get:
  - Loss: 0.1021726205945015
  - Accuracy: 0.964214563369751

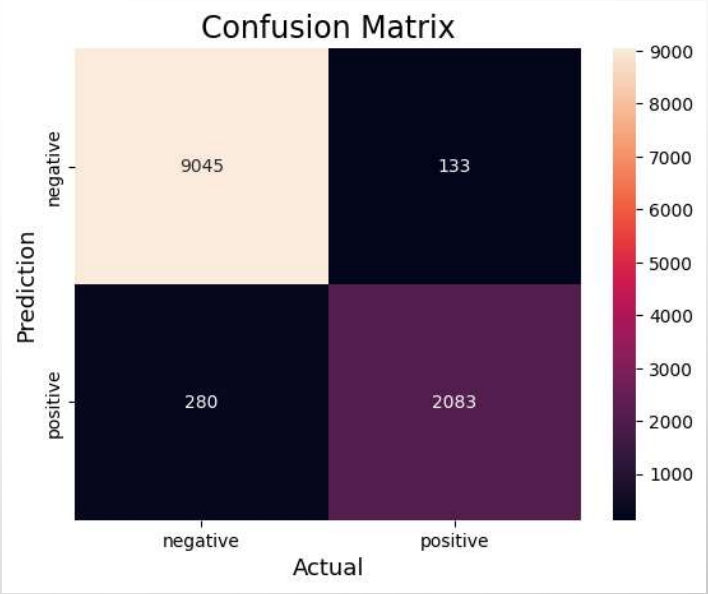


- OBSERVATIONS
- Here we plot the graph between Accuracy Vs Validation Accuracy and Loss Vs Validation Loss.
- In both plots, we see how both training and validation, accuracy and loss changes with epochs.





- Model evaluation
- We printed the confusion matrix to see how much data is predicted right or wrong.
- The final accuracy score is: 96.42%



REFERENCES

- https://towardsdatascience.com/step-by-step-twitter-sentiment-analysis-in-python-d6f650ade58d
- https://medium.com/@nikitasilaparasetty/twitter-sentiment-analysis-for-data-science-using-python-in-2022-6d5e43f6fa6e
- https://python.plainenglish.io/nlp-twitter-sentiment-analysis-using-python-ml-4b4a8fc1e2b
- YouTube
- ChatGPT
- Slide Share
- Kaggle
- GeeksForGeeks
- draw.io

THANK YOU