Twitter Sentiment Analysis Documentation

Introduction

This document provides an overview of the Twitter Sentiment Analysis project, including the data preprocessing steps, data visualization, model building, evaluation, and model saving.

Project Overview

Twitter Sentiment Analysis is a machine learning project aimed at analyzing sentiments associated with tweets. The project involves various steps such as data preprocessing, feature extraction, model building, and evaluation.

Project Structure

- **Dataset**: The dataset used in the project is stored in a CSV file named `twitter_sentiment.csv`.
- **Notebook**: The project is implemented in a Jupyter Notebook file named `Twitter_Sentiment_Analysis.ipynb`.
- **Model**: The trained model is saved in a pickle file named `twitter_sentiment_Capstone_1.pkl`.

Requirements

To run the project, the following libraries need to be installed: - spacy - beautifulsoup4 - textblob - pandas - matplotlib - seaborn - wordcloud - scikit-learn

To install these dependencies, run the following commands: pip install spacy

python -m spacy download en_core_web_sm

pip install beautifulsoup4

pip install textblob

pip install pandas

pip install matplotlib

pip install seaborn

pip install wordcloud

pip install scikit-learn

Data Preprocessing

The data preprocessing steps include: - Basic feature extraction - Data cleaning (lowercasing, removing URLs, HTML tags, special characters, and retweets)

Data Visualization

- Histograms: Plots 2x4 grid histograms for each numerical feature, colored by sentiment.
- Pie chart: Shows the distribution of sentiments.
- Word clouds: Displays 2x2 grid word clouds for each sentiment.

Model Building

- Uses a TF-IDF Vectorizer for feature extraction.
- Implements a Random Forest Classifier for sentiment prediction.

Evaluation

- Evaluates the model using accuracy score on the test dataset.

Model Saving

- Saves the trained model using the pickle module.

Execution Time

The execution time for training the model and saving it is less than 1 minute.

Conclusion

The Twitter Sentiment Analysis project provides insights into the sentiment distribution of tweets and builds a predictive model for sentiment analysis. It can be further extended for real-time sentiment monitoring and analysis.

For more details, refer to the Jupyter Notebook `NLP Book /NLP_Book.ipynb`.