Emoji Predictor From Twitter Data
(Deep Learning)

Introduction to Bidirectional LSTM

Bidirectional Long Short-Term Memory networks (Bidirectional LSTM) are a type of recurrent neural network (RNN) architecture that enhances the capability of standard LSTM networks by capturing information from both past and future contexts. This bidirectional processing makes them particularly effective for tasks requiring understanding of context in sequential data, such as natural language processing (NLP) tasks.

Overview

This document outlines the process and methodology used to predict emojis from tweets using deep learning techniques. It covers data preprocessing, model development, training, evaluation, and application for emoji prediction.

1. Data Preparation and Cleaning

- **Data Sources**: Raw tweet data was obtained from a dataset.
- **Cleaning**: Tweets were cleaned to remove mentions, links, numbers, and punctuation using Python libraries such as NLTK and spaCy.

2. Tokenization and Preprocessing

- **Tokenization**: Tweets were tokenized into sequences and padded to a fixed length for uniform input using the Keras `Tokenizer` and `pad_sequences` functions.
- **Text Preprocessing**: Further preprocessing involved lemmatization and removal of stop words using spaCy.

3. Model Development

- **Bidirectional LSTM Model**:

- ○ Enhanced the model with Bidirectional LSTM layers to capture bidirectional context.
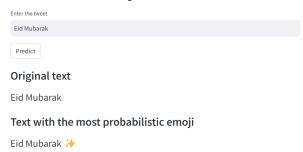- ○ Included GlobalMaxPooling and additional Dense layers for improved feature extraction and classification.

Metrics

	precision	recall	f1-score	support
0	0.56	0.31	0.40	3211
1	0.69	0.64	0.66	3249
2	0.84	0.73	0.78	3240
3	0.79	0.85	0.82	3175
4	0.93	0.93	0.93	3208
5	0.82	0.88	0.85	3281
6	0.89	0.93	0.91	3255
7	0.87	0.94	0.90	3172
8	0.93	0.93	0.93	3273
9	0.84	0.94	0.89	3226
10	0.96	0.89	0.92	3202
11	0.93	0.98	0.95	3216
12	0.95	0.97	0.96	3277
13	0.95	0.97	0.96	3202
14	0.95	0.97	0.96	3225
15	0.95	0.98	0.97	3177
16	0.94	0.96	0.95	3234
17	0.97	0.98	0.97	3224
18	0.87	0.99	0.92	3235
19	0.96	0.97	0.96	3272
accuracy			0.89	64554
macro avg	0.88	0.89	0.88	64554
weighted avg	0.88	0.89	0.88	64554

Additional info:

To open the streamlit app you will need to install streamlit via pip and use the command streamlit run app.py

Output :

Emoji Predictor

Enter the tweet

Eid Mubarak

Predict

Original text

Eid Mubarak

Text with the most probabilistic emoji

Eid Mubarak ✨

Done By
Adeib Arief