Mood Based Music Recommendation using NLP and TinyML

CACSC19 : AI HARDWARE AND TOOLS W/S

Aayush Bagga (2022UCA1806)
Saaz Gupta (2022UCA1860)
Ayush Arora (2022UCA1822)

Dataset and Preprocessing

Dataset :-
https://www.kaggle.com/datasets/abhi8923shriv/sentiment-analysis-dataset/data

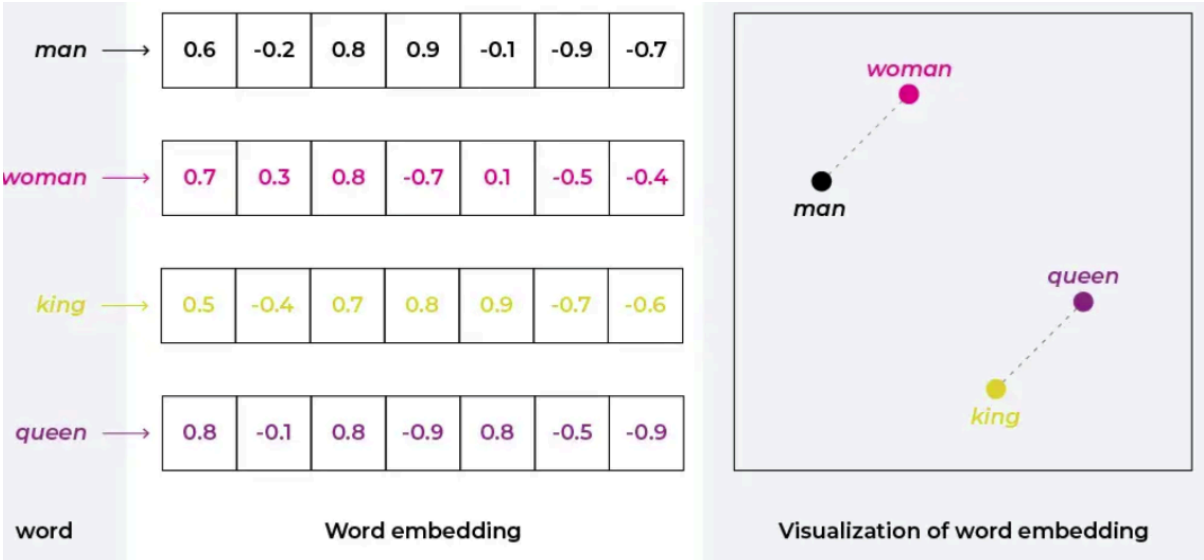
Features in dataset include:-

1. **textID**: A unique identifier for each text entry.
2. **text**: The original text of the tweet or message.
3. **selected_text**: A portion of the original text that strongly represents the sentiment.
4. **Time of Tweet**: The timestamp when the tweet was posted.
5. **Age of User**: The age of the person who posted the tweet.
6. **Country**: The country of the user who posted the tweet.
7. **Population** - 2020: The population of the country in 2020.
8. **Land Area (Km²) & Density (P/Km²)**: Country's land area and population density.

Steps involved in preprocessing include:-

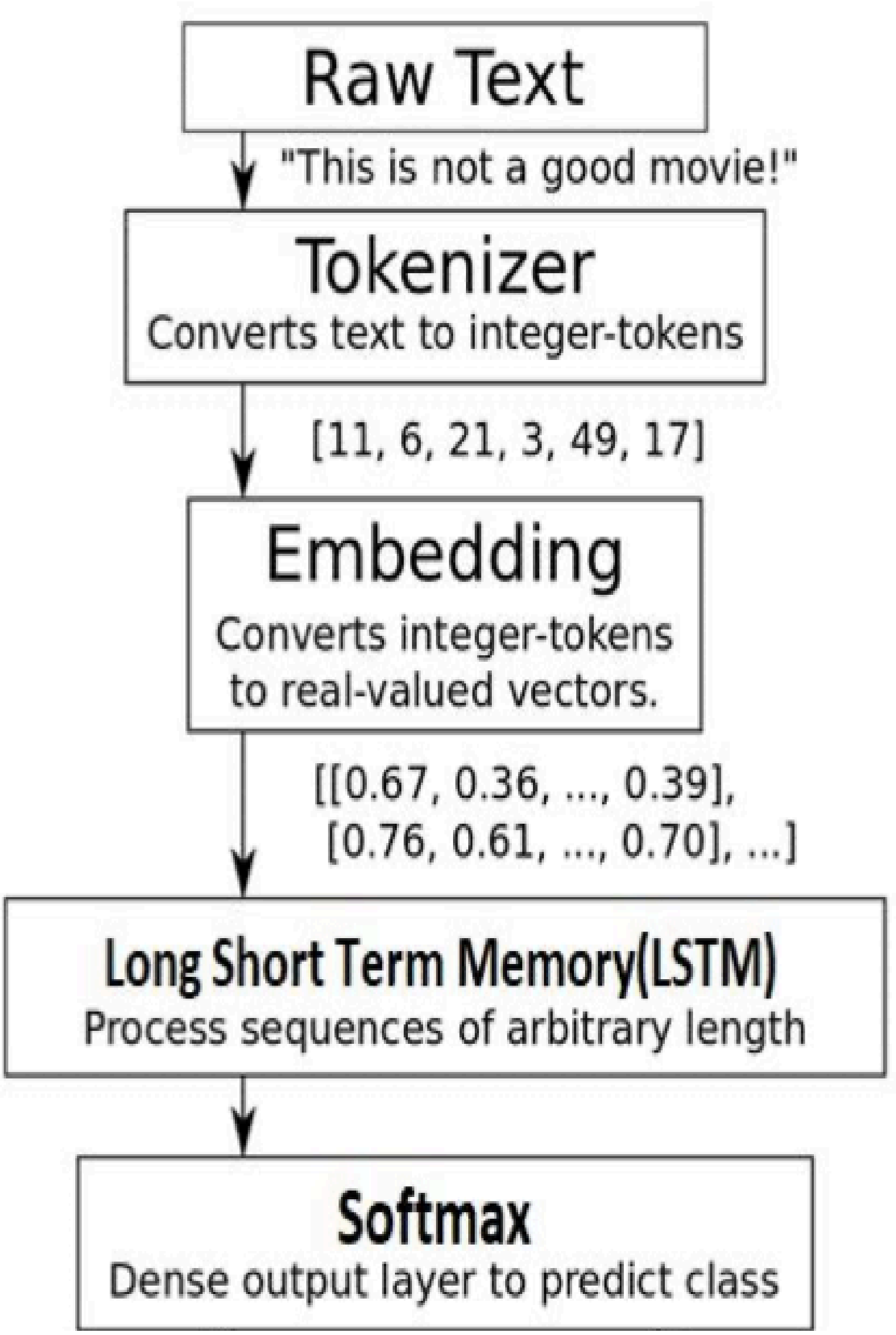
1. Selected features include " text" with target variable as "sentiment". Common stopwords were removed to retain only the most relevant and meaningful words in the selected text
2. Tokenisation = An index mapping dictionary is created by collecting frequently occurring 5000 words. Other rare words are ignored due to minimal impact. Tokenization is the process of tokenizing or splitting a string, text into an inventory of tokens. Each string token is assigned index according to vocab-integer dictionary. Most frequent words are assigned smaller indices.Eg:- `"I love programming"` -> `[0, 1, 2]`
3. Each token vector is padded for uniform size to create X_padded
4. Target labels "neutral","positive","negative" are encoded as 0 1 2 and are one hot encoded to create y_onehot
5. Dataset is split into 60-20-20 where 60% is training set , 20% is validation set and 20% is test set.

Word Embedding and Embedding layer



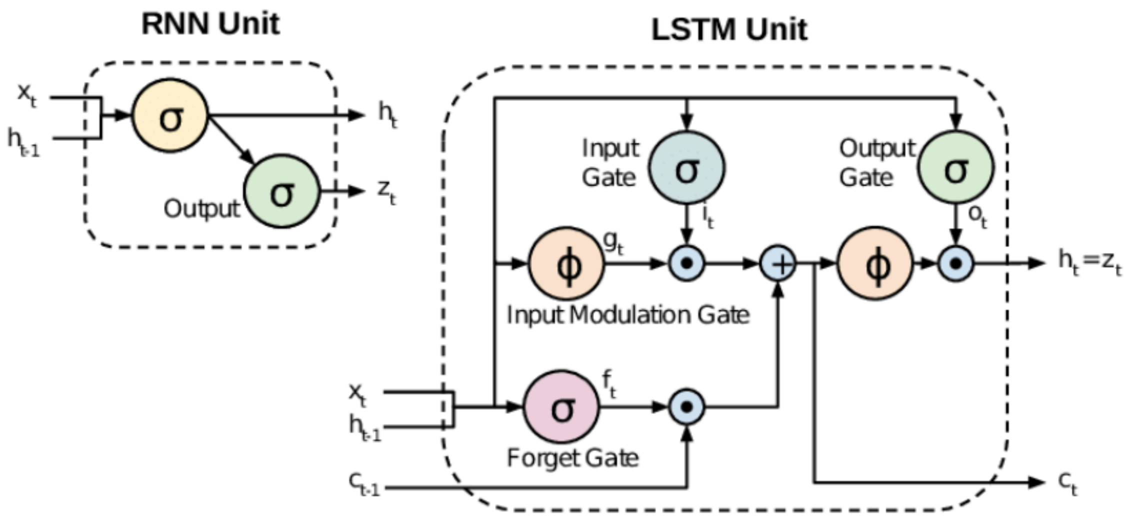
- Embeddings are dense vector representations of a token
- The semantic relationships between words are reflected within the distance and direction of the vectors. It's a representation of text where words that have the identical meaning have an analogous representation.
- In other words, it represents words in an exceedingly system where related words, supported a corpus of relationships, are placed closer together.
- Embedding layer stores a lookup table to map the words represented by numeric indexes to their dense vector representations.
- The embedding layer is used on the front end of a neural network and is fit in a supervised way using the Backpropagation algorithm.
- A random embedding is employed to seed the model, but the embedding is updated jointly during the training of the model

Generalised flow :-



Long short-term memory neural networks

- Recently, LSTM is most popular to deal with sentiment classificationThey work tremendously well on large different types of problems and are now widely used.
- LSTMs are explicitly designed to ignore the long-term dependency problem
- Long short-term memory (LSTM) is a synthetic recurrent neural network (RNN) architecture employed in the sphere of deep learning. Unlike standard feedforward neural networks,LSTM has feedback connections.
- LSTM networks are well-suited to classifying, processing, and making predictions supported statistic data since there may be lags of unknown duration between important events in a very statistic.
- LSTMs were developed to accommodate the exploding and vanishing gradient problems that may be encountered when training traditional RNNs.



Model Architecture

Model: "sequential"

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 100, 128)	640,000
lstm (LSTM)	(None, 100, 128)	131,584
dropout (Dropout)	(None, 100, 128)	0
lstm_1 (LSTM)	(None, 64)	49,408
dropout_1 (Dropout)	(None, 64)	0
dense (Dense)	(None, 3)	195

Total params: 2,463,563 (9.40 MB)
Trainable params: 821,187 (3.13 MB)
Non-trainable params: 0 (0.00 B)
Optimizer params: 1,642,376 (6.27 MB)

Each token is represented with an embedding vector of size 100 X 128. The top 5000 words are used in the vocabulary and rare words are removed from the dictionary to avoid unnecessary computations.

Hyperparameter Tuning

Tuning hyperparameter dropout and number of epochs:-

At dropout 0.1 and 5 epochs for normal model test accuracy = 70.7% , train accuracy = 83.68% , validation accuracy = 69.97%

At epoch=10,dropout = 0.2 for normal model:-

Results for epoch 1-5 are presented in Results section

Epoch	Train accuracy	Validation accuracy
6	87.35%	69.02%
7	89.19%	67.29%
8	90.76%	66.95%

9	91.98%	66.15%
10	93.29%	66.24%

Results and Applications

During training, the hyperparameters that resulted in the best performance are:

- Dropout is applied with a rate of 0.2 and 5 epochs.
- Adam optimizer is used to optimize the model and categorical_crossentropy is used as the loss function.
- A batch size of 64 is adopted.

Epochs	Train Accuracy	Train Loss	Validation Accuracy	Validation Loss
1	52.65%	0.9492	69.97%	0.7124
2	75.10%	0.6221	71.46%	0.6918
3	79.35%	0.5354	71.17%	0.7193
4	81.66%	0.4880	70.42%	0.7438
5	84.12%	0.4230	69.77%	0.8405

Results on test data by normal model :- 71.34% accuracy,

Size of normal model:- 9.4 mb

After creation of normal model it is optimised into a lite model by tensorflow lite library

Results on validation data by Lite model:- 69% accuracy

Results on test data by Lite model:- 71% accuracy

Size of lite model :- 3.14 mb

Sentiment analysis processes people's opinions and attitudes toward products, services, politics, social events, and company strategies. Reviews (from sources such as TripAdvisor, Amazon, and IMDB) and social network posts (mostly from Twitter and Facebook) are categories of textual documents that are the most interesting for sentiment analysis. DL methods such as LSTM show better performance of sentiment classification with 85% accuracy when there are more amounts of training data.

For recommendation industry text based adsense is more popular than audio,video due to privacy concerns

Applications include:-

- Lyrics-Based Recommendations: Recommends music by analyzing song lyrics for themes or emotional tones (e.g., love, heartbreak).
- Contextual Recommendations: Analyzes music articles and blogs to recommend trending or contextually relevant music.
- Social Media Insights: Leverages user-generated content on social media to understand musical preferences and recommend music accordingly.
- Enhanced Collaborative Filtering: Combines user behavior with text analysis to refine music recommendations beyond simple listening patterns.
- Trend Analysis: Uses NLP to identify emerging trends in music discussions and predict popular songs or artists.
- Voice Search Integration: Analyzes voice commands (e.g., “play relaxing music”) to generate personalized playlists based on user intent.

Simulation on Wokwi

For the simulation of our music recommendation model, we have utilized the ESP32 microcontroller in conjunction with a 20x4 LCD display. The system is designed to take user input, process it through our trained machine learning model, and display the music recommendations result on the LCD screen.

System Workflow:

1. User Input Acquisition: The user provides textual input via a designated interface.
2. Processing with ML Model: The input text is processed and passed through the deployed sentiment analysis model, which is optimized for embedded systems.
3. Prediction and Output Display: The model predicts the sentiment (e.g., positive, negative, or neutral), and the corresponding output of music recommendations are displayed on the 20x4 LCD screen connected to the ESP32 microcontroller.

This implementation demonstrates the integration of machine learning inference on embedded systems, ensuring an efficient and real-time sentiment classification process and recommendations on a resource-constrained device.

References and Demonstration

1. Text based sentiment analysis using LSTM by Dr. G. S. N. Murthy, Shanmukha Rao Allu, Bhargavi Andhavarapu,Mounika Bagadi, Mounika Belusonti,Department of Computer Science and Engineering,Aditya Institute of Technology and Management,Srikakulam, Andhra Pradesh.
2. Sentiment Analysis using Bidirectional LSTM Network,Author links open overlay panel,U.B. Mahadevaswamy, P. Swathi
3. Music Predictions Using Deep Learning.Could LSTM Networks be the New Standard for Collaborative Filtering? by EMIL KESKI-SEPPÄLÄ AND MICHAEL SNELLMAN
4. Deep neural networks for context aware personalized music recommendation by Oktay Bahceci

Demonstration video:-

[Mood based Music recommendation using NLP]