UE20CS390B – Capstone Project Phase – 2

Project Progress Review #3

Project Title : FeelSpeak: Generating Emotional Speech with Deep Learning
Project ID : PW23_VRB_07
Project Guide : Prof. V R Badri Prasad
Project Team : 235_320_345_362

PES UNIVERSITY

Outline

- Abstract and Scope of the Project
- Suggestions from Review – 2
- Implementation Details
- Project Demonstration
- Test Plan and Strategy
- Results and Discussion
- Project Report Draft
- Conclusion and Future Work
- References

PES UNIVERSITY

Abstract and Scope

Abstract:

- Goal: Develop a system for generating emotional speech from input text.
- Approach: Identify emotions in text, synthesize speech with appropriate prosodic features.
- Tasks: Natural Language Processing (NLP), text emotion detection, speech synthesis, emotion recognition.
- Integration: Fusion of NLP, speech synthesis, and emotion recognition for holistic interaction.

Scope:

- Objective: Create a system accurately identifying emotional content in input text.
- Implementation: Utilize NLP and text emotion detection for structural and emotional analysis.
- Output: Generate emotionally rich speech using prosodic features reflecting detected emotions.
- Components: Emotion detection from text, speech synthesis with recognized emotions .
- Significance: Extensive research at the convergence of NLP, speech synthesis, and emotion recognition.

PES UNIVERSITY

Suggestions from Review - 2

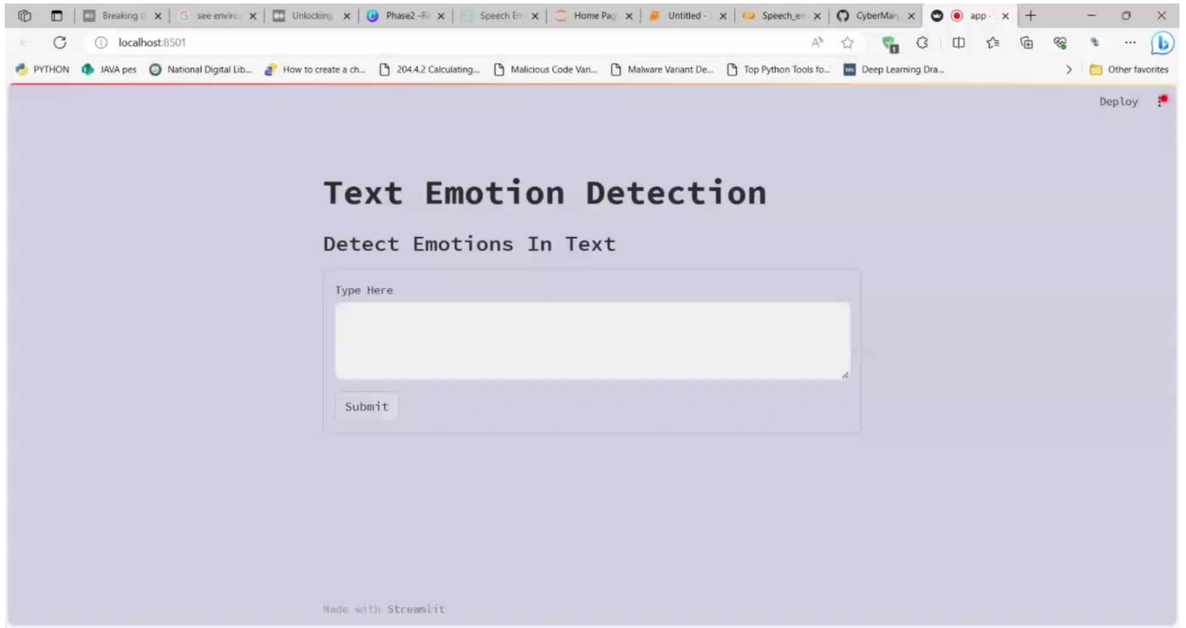
The suggestions given by the panel members were:

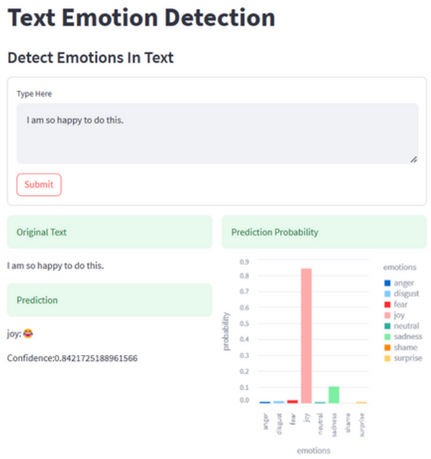
- To improve the selection of emotions for the emotion detection model.
- To work on attaching emotion to the particular input speech given by the user.
- Improve the quality of the output speech generated by the TTS model.
- Train the model further on chosen dataset to improve the speech clarity.

Implementation Details - Emotion detection from text

Emotion detection from text.

- Detecting emotion from text using models like SVM, Random forest, linear regression and EmoRoBERTa Model from Huggingface. Transformers.
- Saving the model using pickle library
- Visualizing the model with graph and confidence score using streamlit.





PES UNIVERSITY

Implementation Details - Emotion detection from text

- LeXmo: The first Python package for classifying emotions in English texts
- LeXmo converts text into a pandas data frame, calculating emotion weights by dividing emotional association by word count.
- Find the demo here.

- It uses Emo-Roberta model to detect text from emotions from hugging face transformer see emotions below.
- It calls the model use this link and predicts the emotion.
- The models gives dictionary with key as label(emotion) and score.
- Best result f1-score: 49.03%

Dataset labelled 58000 Reddit comments with 28 emotions

- admiration, amusement, anger, annoyance, approval, caring, confusion, curiosity, desire, disappointment, disapproval, disgust, embarrassment, excitement, fear, gratitude, grief, joy, love, nervousness, optimism, pride, realization, relief, remorse, sadness, surprise + neutral

PES UNIVERSITY

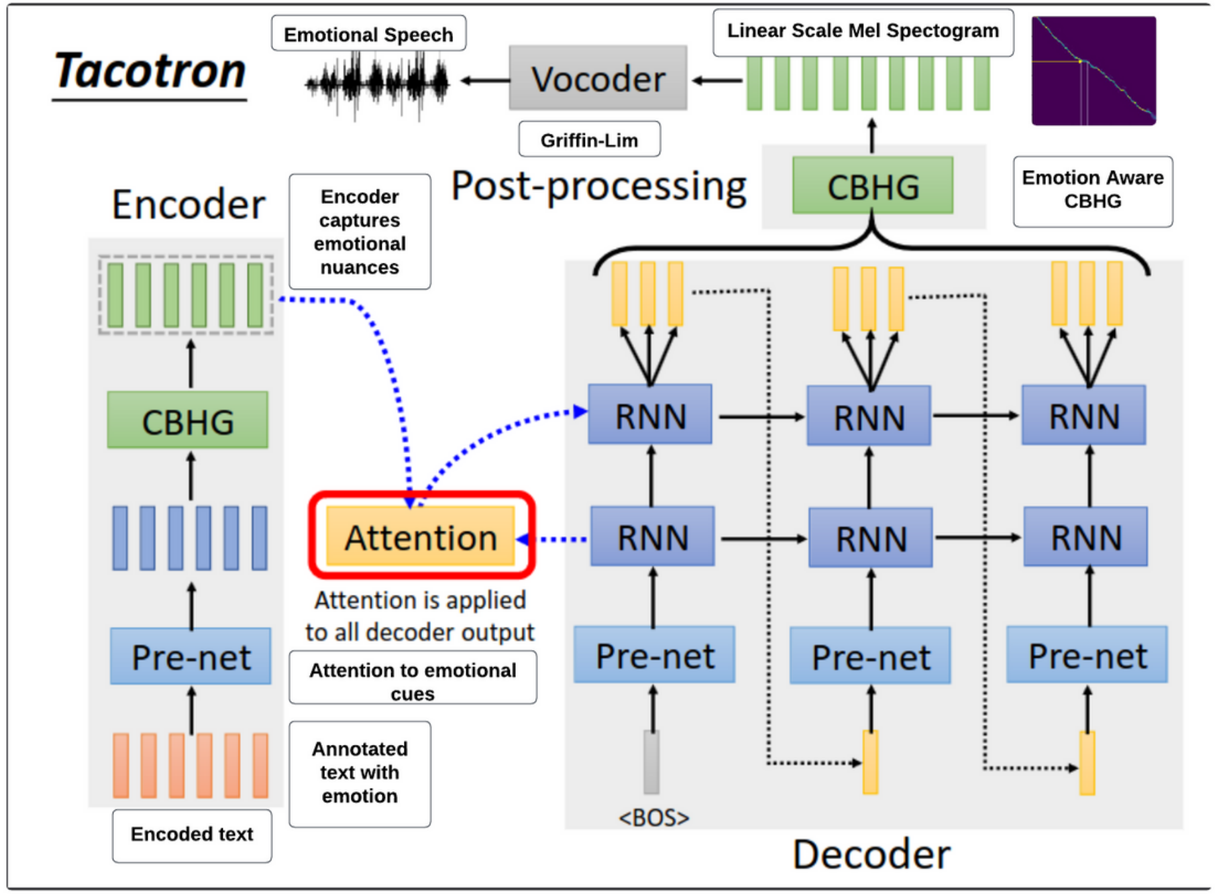
Implementation Details -Text to speech with emotion

1. Dataset:
Objective: Download the LJ Speech dataset for English speech samples.
Action Taken:
Downloaded the LJ Speech dataset.
Organized the dataset, including audio files and text transcripts, into a structured directory.

2. Preprocessing:
Objective: Prepare audio and text data for model training.
Actions Taken:
Extracted features, such as mel spectrograms, from the audio files.
Tokenized and preprocessed text data to ensure suitability for training.

Implementation Details -Text to speech with emotion

3. Model Architecture:
   - Objective: Design the Tacotron model for sequence-to-sequence mapping.
   - Actions Taken:
     - Developed the Tacotron model architecture, including an encoder, attention mechanism, and decoder.
     - Utilized recurrent neural networks (RNNs) or LSTM networks for effective sequence modeling.



Implementation Details -Text to speech with emotion

4. Training:
   - Objective: Train the Tacotron model using preprocessed data.
   - Actions Taken:
     - Defined loss functions, incorporating spectrogram loss and alignment loss.
     - Utilized the Adam optimizer, experimenting with learning rates and other hyperparameters.

5. Hyperparameter Tuning:
   - Objective: Optimize hyperparameters based on model performance.
   - Actions Taken:
     - Fine-tuned hyperparameters, including learning rates, batch sizes, and training epochs.

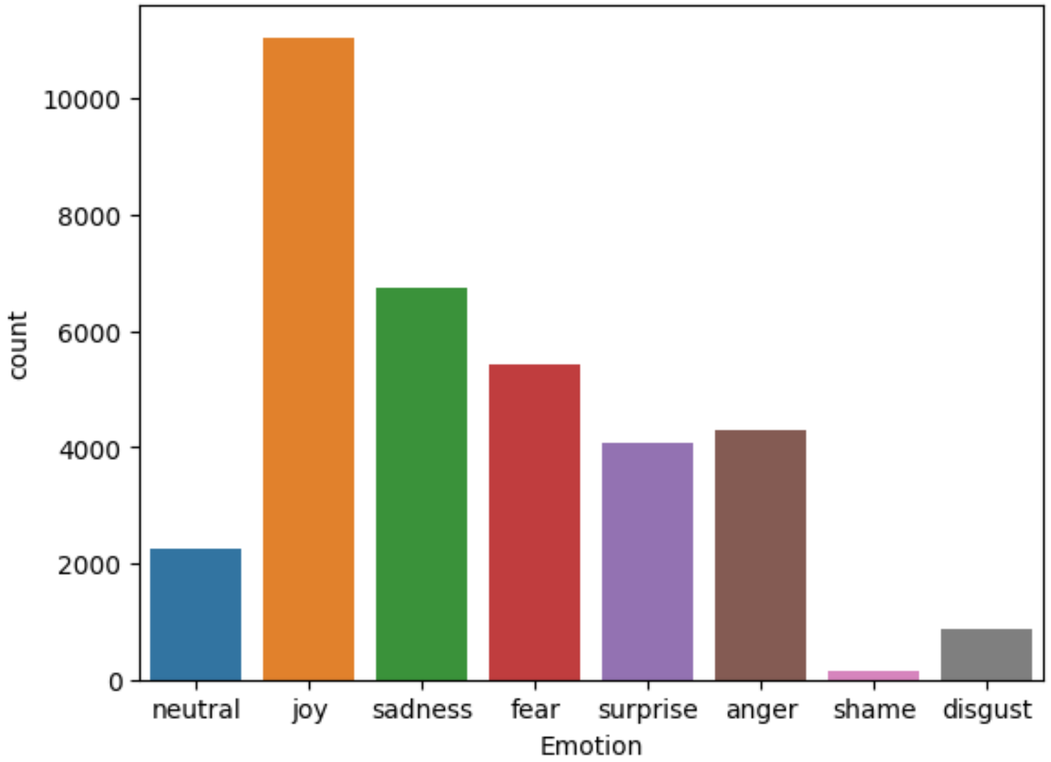
Implementation Details -Text to speech with emotion

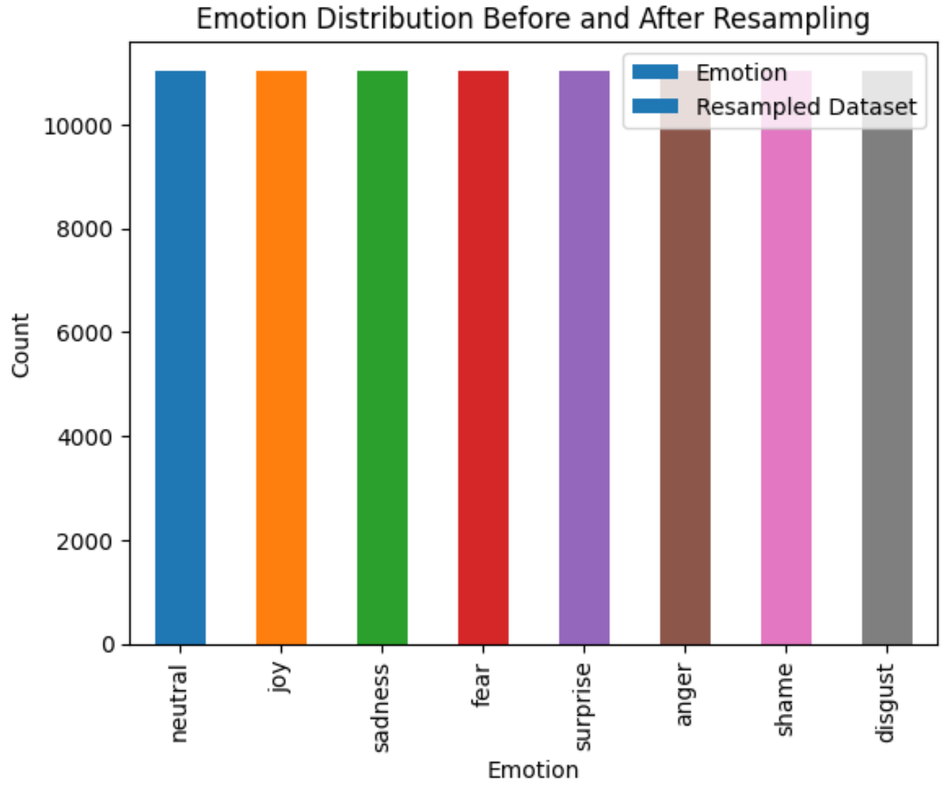
6. Evaluation:
   - Objective: Assess the performance of the trained model.
   - Actions Taken:
     - Evaluated the model on a validation set to ensure proper learning.
     - Leveraged metrics like Mean Opinion Score (MOS) in subjective listening tests for voice quality assessment.

7. Inference:
   - Objective: Implement an inference pipeline for synthesizing speech from text.
   - Actions Taken:
     - Developed an inference pipeline to synthesize speech using the trained Tacotron model.
     - Combined Tacotron output with a vocoder (e.g., Griffin-Lim) to generate the final waveform.

Project Demonstration - Emotion detection from text

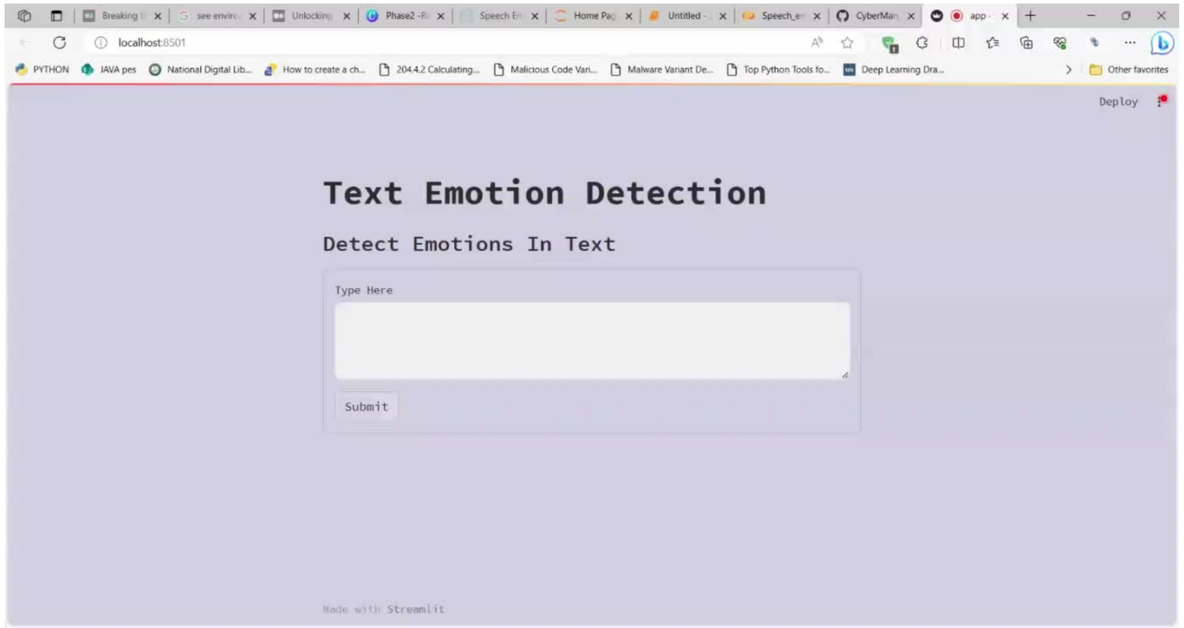
EMOTION DETECTION FROM TEXT - BALANCING DATASET



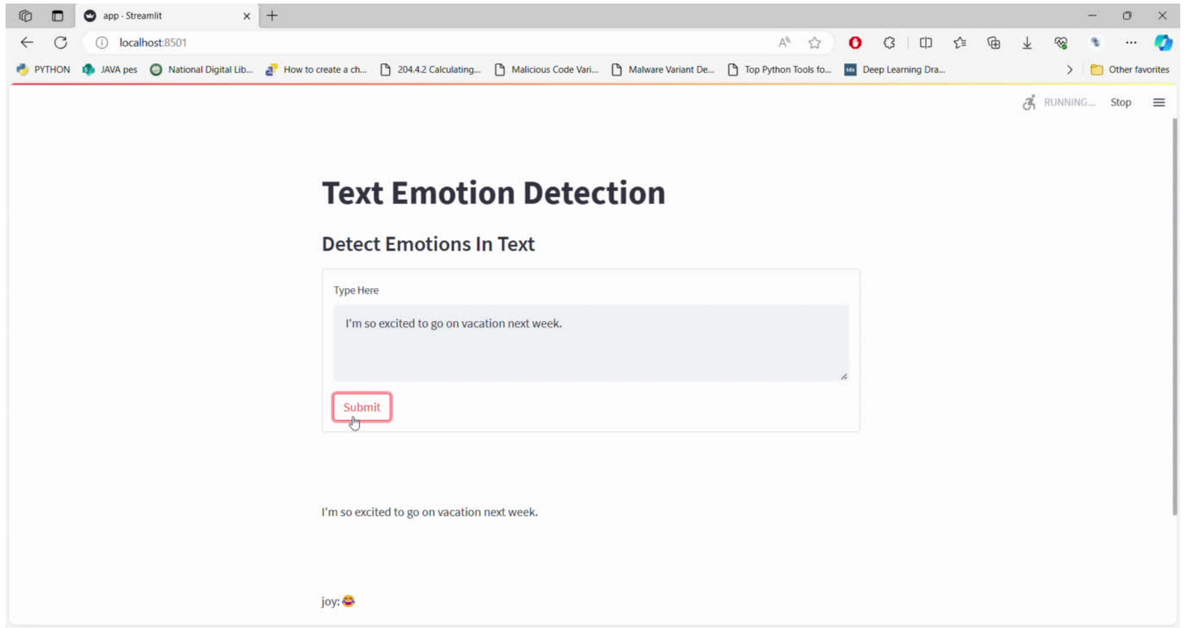


Project Demonstration - Emotion detection from text

With unbalanced dataset

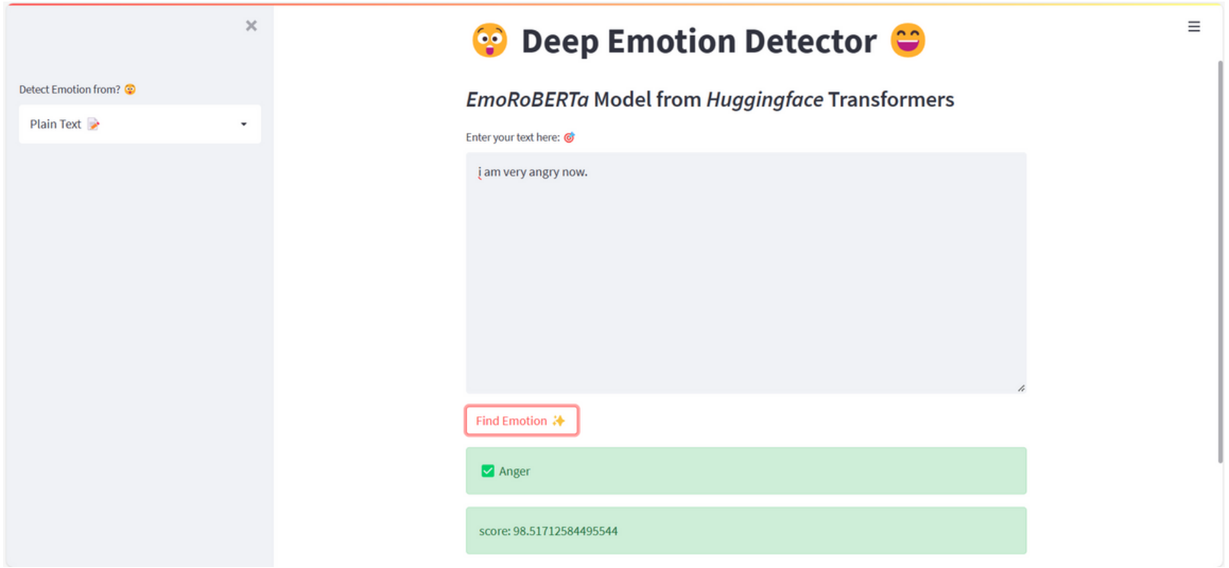


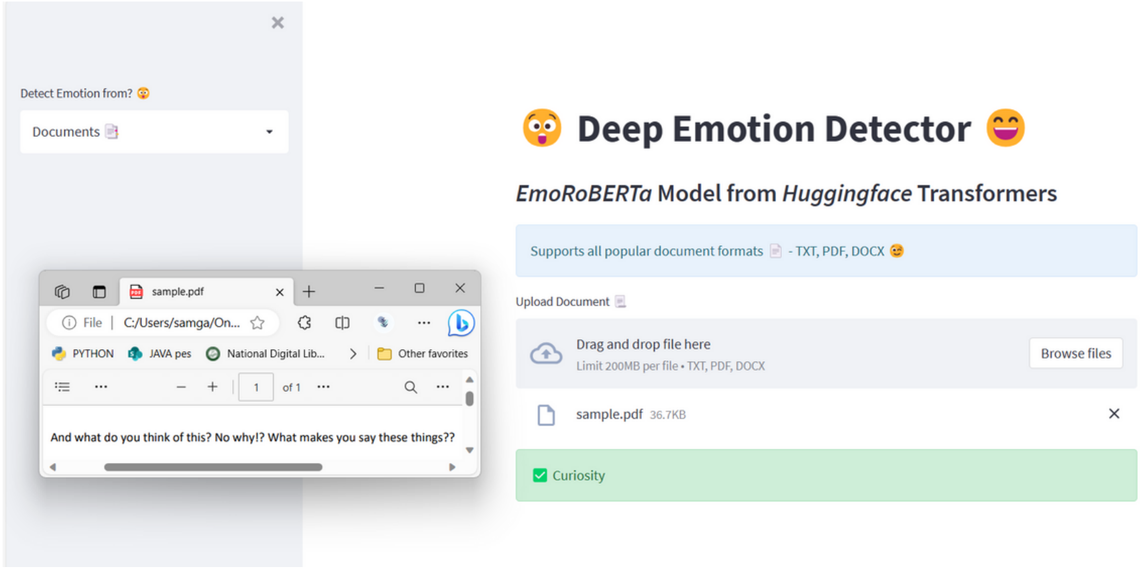
With balanced dataset



Project Demonstration - Emotion detection from text

EmoRoberta model from Hugging face Transformers



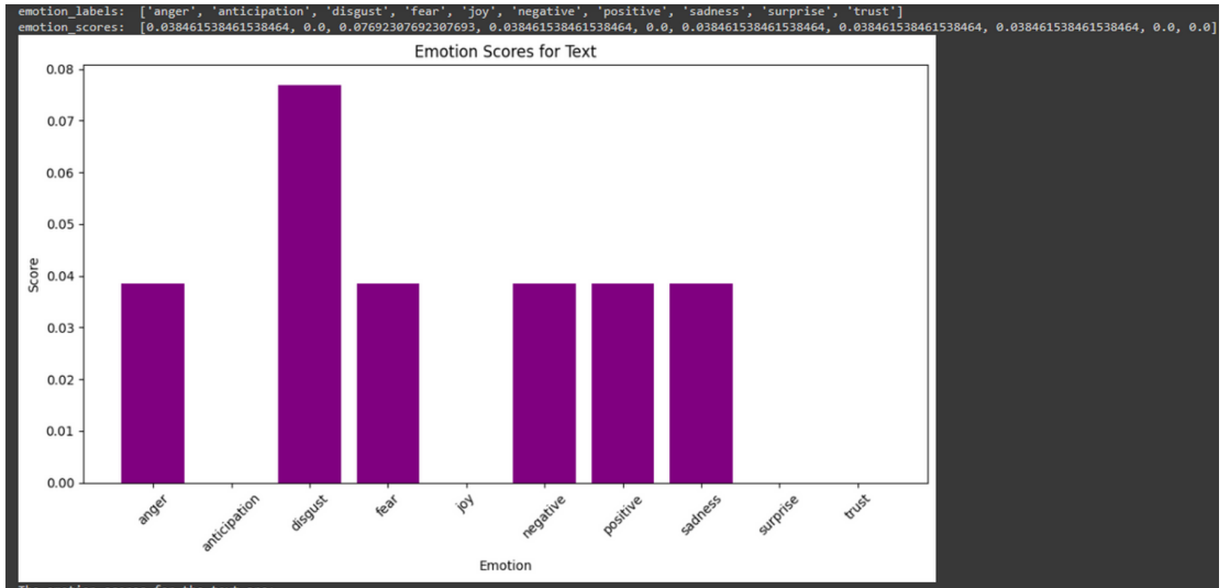


Project Demonstration - Emotion detection from text

LexMo python package

```
[ ] t= """From the beginning, she had sat looking at him fixedly.
    As he now leaned back in his chair, and bent his deep-set eyes upon her in his turn,
    perhaps he might have seen one wavering moment in her,
    when she was impelled to throw herself upon his breast,
    and give him the pent-up confidences of her heart.
    But, to see it, he must have overleaped at a bound the artificial barriers he had for many years been erecting,
    between himself and all those subtle essences of humanity which will elude the utmost cunning of algebra
    until the last trumpet ever to be sounded shall blow even algebra to wreck.
    The barriers were too many and too high for such a leap. With his unbending,
    utilitarian, matter-of-fact face, he hardened her again;
    and the moment shot away into the plumbless depths of the past,
    to mingle with all the lost opportunities that are drowned there."""

[ ] emo=LeXmo.LeXmo(t)

[ ] print(emo)

    {'text': 'From the beginning, she had sat looking at him fixedly.\n  As he now leaned back in his chair, and bent h

[ ] emo.pop('text', None)

    'From the beginning, she had sat looking at him fixedly.\n  As he now leaned back in his chair, and bent his deep-s
    \n  when she was impelled to throw herself upon his breast,\n  and give him the pent-up confidences of her heart.\n
    any years been erecting, \n  between himself and all those subtle essences of humanity which will elude the utmost
    o wreck.\n  The barriers were too many and too high for such a leap. With his unbending,\n  utilitarian, matter-of-
    f the past,\n  to mingle with all the lost opportunities that are drowned there.'

[ ] print(emo)

    {'anger': 0.023255813953488372, 'anticipation': 0.0, 'disgust': 0.005813953488372093, 'fear': 0.023255813953488372,
```

```
emotion_labels:  ['anger', 'anticipation', 'disgust', 'fear', 'joy', 'negative', 'positive', 'sadness', 'surprise', 'trust']
emotion_scores:  [0.038461538461538464, 0.0, 0.07692307692307693, 0.038461538461538464, 0.0, 0.038461538461538464, 0.038461538461538464, 0.038461538461538464, 0.0, 0.0]
```




```
The emotion scores for the text are:
anger: 3.8461538461538463
anticipation: 0.0
disgust: 7.6923076923076925
fear: 3.8461538461538463
joy: 0.0
negative: 3.8461538461538463
positive: 3.8461538461538463
sadness: 3.8461538461538463
surprise: 0.0
trust: 0.0
```

PES UNIVERSITY

Test Plan and Strategy

Emotion detection from text using LR,RF, SVM

SENTENCE	EMOTION DETECTED	EXPECTED EMOTION
I am happy today.	JOY 76.5%	JOY
Alas, I lost all my project data due to a technical glitch with tacotron	SADNESS 77.9%	SADNESS
It's frustating how unreliable the results are . It's making me so angry!	ANGER 90.9%	ANGER

Test Plan and Strategy

Emotion detection from text using LR,RF, SVM

SENTENCE	EMOTION DETECTED	EXPECTED EMOTION
I'm scared of what the future holds.	FEAR 99.5%	FEAR
I didn't expect you to remember my birthday.	SURPRISE 73.64%	SURPRISE
I feel so embarrassed about what I did.	SHAME 99.5%	SHAME

Test Plan and Strategy

Emotion detection from text using EmoRoBERTa Model from Huggingface Transformers

SENTENCE	EMOTION DETECTED	EXPECTED EMOTION
Life’s good, you should get one.	NEUTRAL	NEUTRAL
The bear was ravenous, he was fierce and furious	ANGER 97.51%	ANGER
In sooth I know not why I am so melancholic.	SADNESS 73.61%	SADNESS

PES UNIVERSITY

Test Plan and Strategy

Emotion detection from text using EmoRoBERTa Model from Huggingface Transformers

SENTENCE	EMOTION DETECTED	EXPECTED EMOTION
Waaaaw!, this car is amazing!	EXCITMENT 77.86%	HAPPY
I'm so ashamed of my behavior.	EMBARRASSMENT 98.62%	SHAME
I'm afraid of public speaking.	FEAR 99.03%	FEAR

Results and Discussion

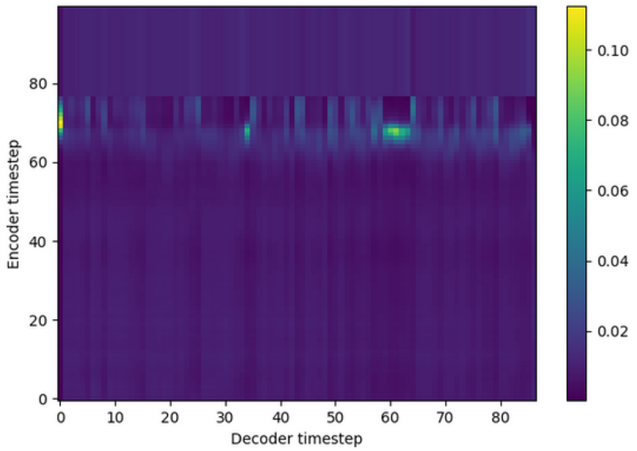
Emotion detection from text

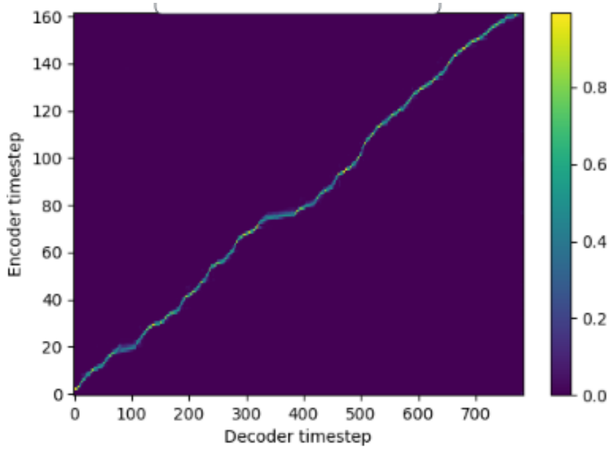
- After training and testing the model this are the accuracy score:
  - Linear Regression: 86.13%
  - Random Forest model: 89.38%
  - Support Vector Machine: 87.93%

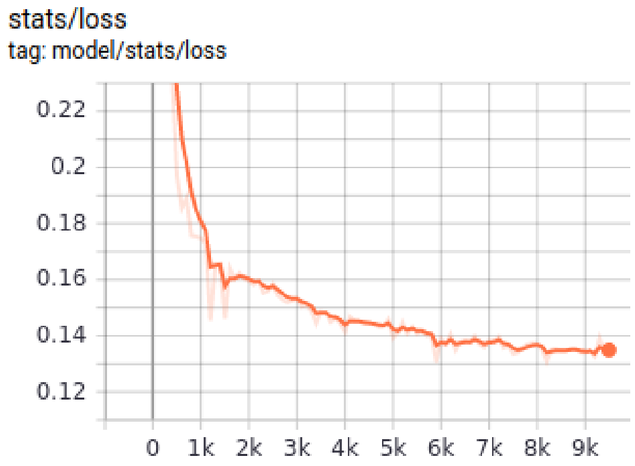
- EmoRoberta model of hugging face transformers is a pretrained model with the f1-score of 49.30%

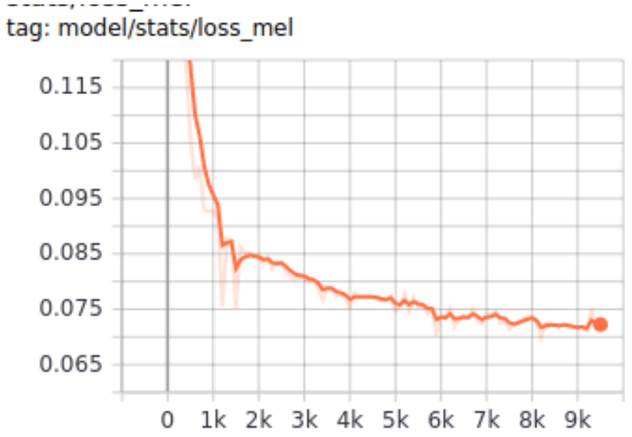
Results and Discussion

Emotional Speech









PES UNIVERSITY

Conclusion and Future work

Integration:

- Integrate the Tacotron-based TTS system into text editor read-aloud button.
- Use in any story books helps kids in classroom education.
- Help people with disabilty to listen to their favourite story book in a human way.
- Can be the voice to the text AI Assitant
- Can be extended to other accents.

PES UNIVERSITY

References

[1] Liu, Rui, et al. "Modeling prosodic phrasing with multi-task learning in tacotron-based TTS." IEEE Signal Processing Letters 27 (2020): 1470-1474.

[2] Huan, Jeow & Sekh, Arif Ahmed & Quek, Chai & Prasad, Dilip. (2022). Emotionally charged text classification with deep learning and sentiment semantic. Neural Computing and Applications. 34. 10.1007/s00521-021-06542-1

[3] P. Chandra et al., "Contextual Emotion Detection in Text using Deep Learning and Big Data," 2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA), Gunupur, India, 2022, pp. 1-5, doi: 10.1109/ICCSEA54677.2022.9936154.

[4] Speech Emotion Recognition using Convolution Neural Networks and Deep Stride Convolutional Neural Networks - 2020 Published in: 2020 6th International Conference on Wireless and Telematics (ICWT) ISBN Information:INSPEC Accession Number: 20133021
DOI: 10.1109/ICWT50448.2020.9243622

Thank You

M H SOHAN_RAHUL ROSHAN G_ROHIT ROSHAN_S M SUTHARSAN RAJ