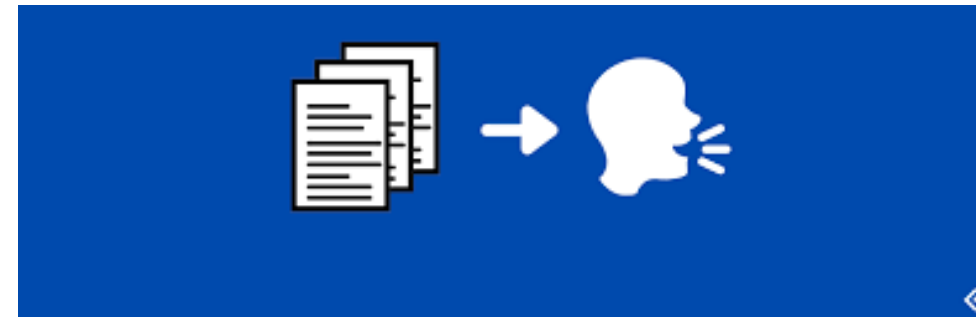


UE20CS390A - Capstone Project Phase - 1

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

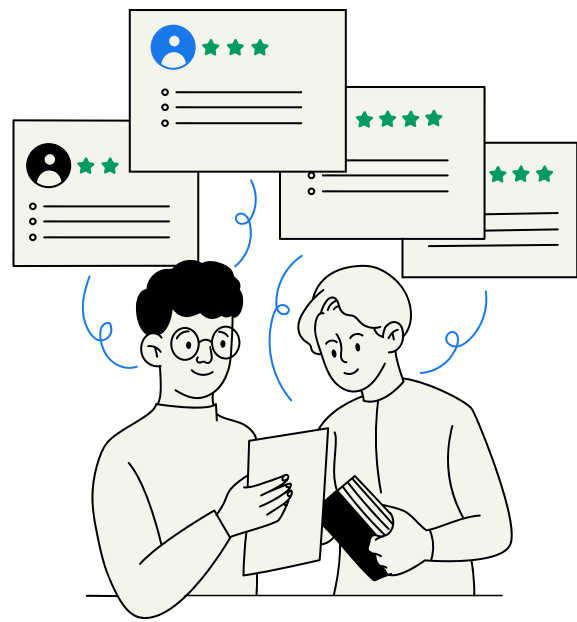
Abstract and Scope

The project "Generation of emotional speech from text" aims to develop a system that can generate speech with appropriate emotional content based on a given input text.

- The system will be able to identify the emotions expressed in the text and generate speech with appropriate prosodic features (such as pitch, duration, and intensity) that convey those emotions effectively.
- The scope of this project includes various tasks such as natural language processing (NLP), speech synthesis, and emotion recognition.

The NLP component will involve parsing the input text to identify its structure, meaning, and emotional content. The speech synthesis component will be responsible for generating speech that accurately reflects the emotions expressed in the input text. The emotion recognition component will involve identifying the emotional content of the input text and mapping it to appropriate prosodic features.

Suggestions from Review - 2

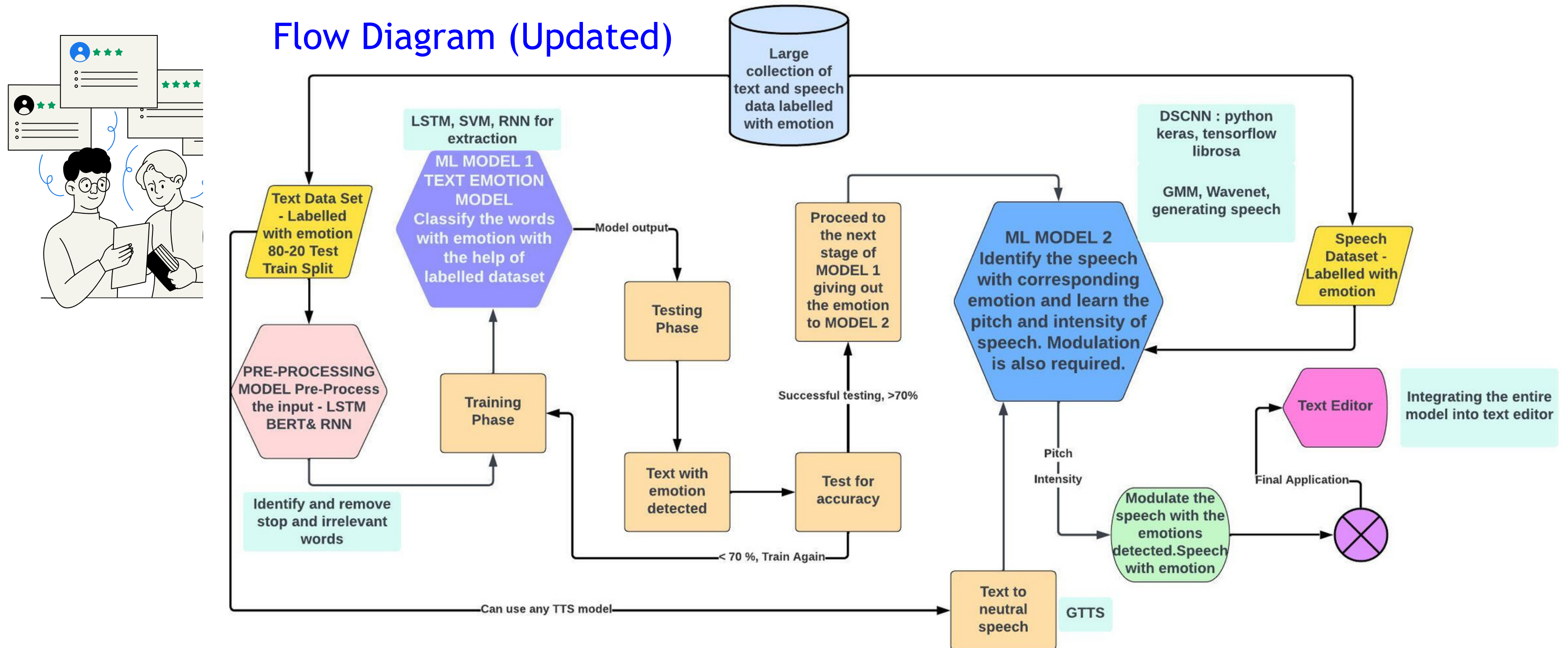


- Provide the suggestions and remarks given by the panel members.
 - To refactor the flow diagram.
 - To do extensive literature survey for some specific models.
 - Limitations to the models
 - Dataset collection
- Mention the feasibility on the same showing the progress.
- Extensive Literature Survey
- Modification of flowchart

All the progress are mentioned in the upcoming slides

Suggestions from Review - 2

Flow Diagram (Updated)



Design Approach

Iterative Design Approach



1. Gather requirements: Understand the user requirements and project goals.
2. Design: Create a design for the system architecture, including the emotion detection and speech conversion models.
3. Implement: Implement the models and integrate them into the system architecture.
4. Test: Test the system using sample inputs and evaluate the accuracy of emotion detection and speech conversion.
5. Evaluate: Analyze the results of testing and gather feedback from users to identify areas of improvement.

Benefits:


- Allows for continuous improvement and refinement of the system.
- Can help identify and address issues early in the development process.

Drawbacks:

- Can be time-consuming and costly.
- May require significant changes to the system architecture if major issues are identified.

Design Approach

User-Centered Design Approach:

- 
1. Research: Conduct user research to understand user needs and preferences.
 2. Design: Create a system architecture that meets user needs and preferences, including the emotion detection and speech conversion models.
 3. Implement: Implement the models and integrate them into the system architecture.
 4. Test: Test the system using sample inputs and gather user feedback to evaluate the system's effectiveness and usability.
 5. Refine: Refine the system based on user feedback.

Benefits:


- Prioritizes the needs and preferences of users, which can lead to a more user-friendly and effective system.
- Helps ensure that the system meets user expectations and requirements.

Drawbacks:

- Can be time-consuming and costly.
- May require significant changes to the system architecture if user feedback identifies major issues.

Design Approach

Agile Design Approach

- 
1. Plan: Create a high-level plan for the system architecture, including the emotion detection and speech conversion models.
 2. Develop: Develop the models and integrate them into the system architecture in small, incremental stages.
 3. Test: Test the system using sample inputs and gather feedback to evaluate the system's effectiveness and usability.
 4. Evaluate: Analyze the results of testing and use them to guide further development.

Benefits:


- Allows for flexibility and adaptability in response to changing user needs or system requirements.
- Enables the development team to focus on high-priority features and functionality.

Drawbacks:

- Can lead to a lack of clarity around the overall system architecture and goals.
- May result in technical debt if short-term solutions are prioritized over long-term considerations.

Design Constraints, Approach, Assumptions & Dependencies

Design constraints and assumptions:

- 
1. **Availability of labeled speech and text dataset:** The project assumes that there is a sufficient amount of labeled speech and text data available to train and test the emotion recognition and conversion models. Without this data, it would not be possible to develop accurate models for recognizing emotions and converting neutral speech to emotional speech.
 2. **Processing power and resources:** The project assumes the availability of adequate processing power and resources to train and test the machine learning models. The deep learning models require significant computational resources, including GPUs and memory, to train effectively.
 3. **Accuracy of emotion recognition:** The accuracy of emotion recognition is an important assumption of the project. The emotion recognition model needs to accurately identify the emotion in the input speech or text data to produce the correct emotional speech output.
 4. Text entered will be in English Language with a UK/US accent.
 5. A total of five emotions are proposed to be detected, which include happy, sad, neutral, angry, sarcastic.
 6. Text is assumed to be an annotated form of text, i.e, conversational format of text, role based dialogue format, etc.

Design Constraints, Approach, Assumptions & Dependencies



Dependencies:

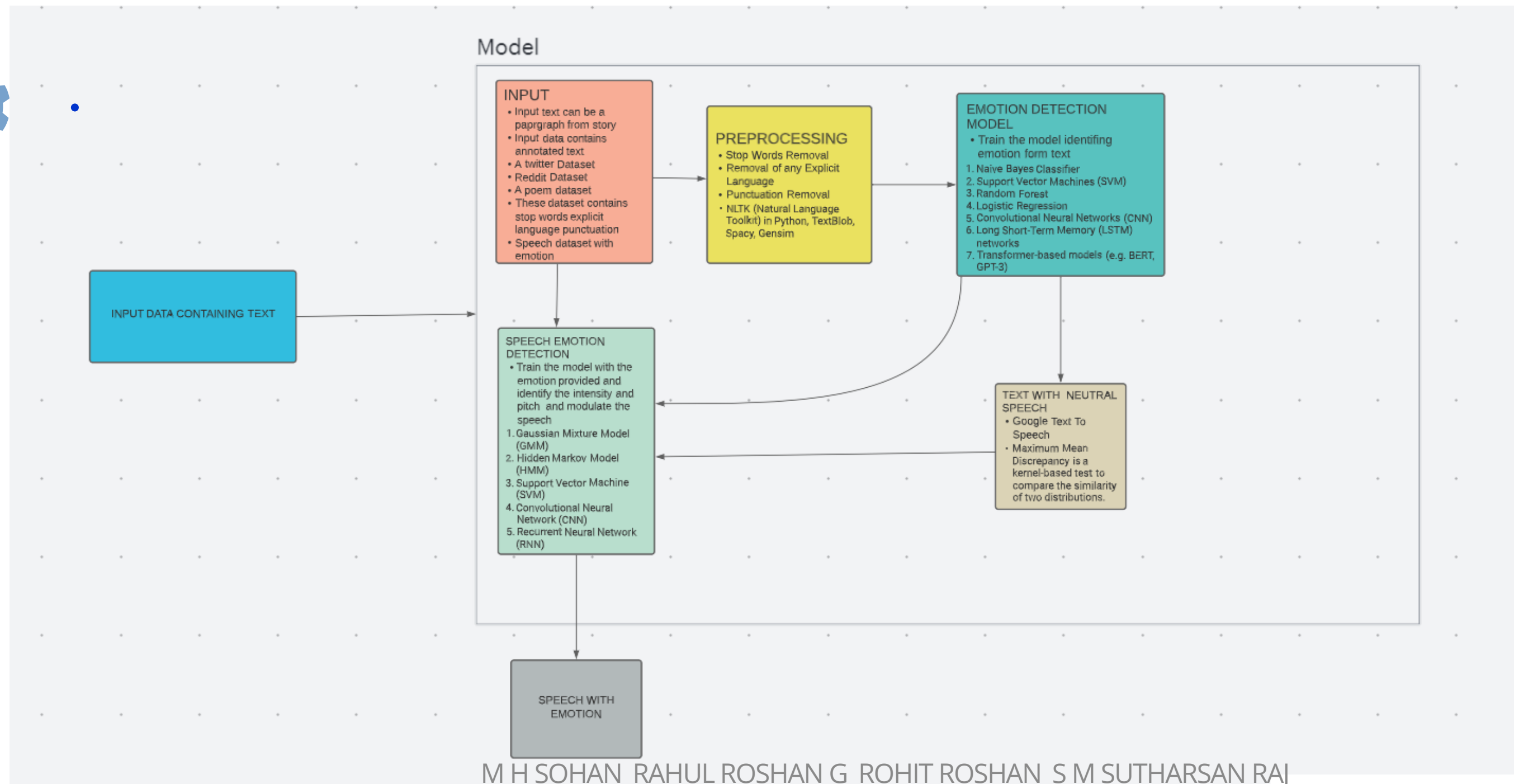
1. Availability of labeled speech and text dataset: The availability of labeled speech and text data is a critical dependency for the project. Without this data, it is impossible to train and test the emotion recognition and conversion models.
2. Accuracy of the emotion recognition model: The accuracy of the emotion recognition model is a key dependency for the project. If the emotion recognition model is not accurate, it will produce incorrect emotional speech outputs, leading to a poor user experience.
3. Availability of processing power and resources: The availability of processing power and resources is also a significant dependency for the project. Without these resources, it is not possible to train and test the machine learning models effectively.

Impact of dependencies:

The impact of these dependencies is that they affect the project's timeline and overall success. Without adequate resources, and accurate models, the project may experience delays, quality issues, and poor user adoption. It is essential to manage these dependencies carefully and plan for contingencies to ensure project success.

Architecture

Provide high-level design view of the system.



Architecture

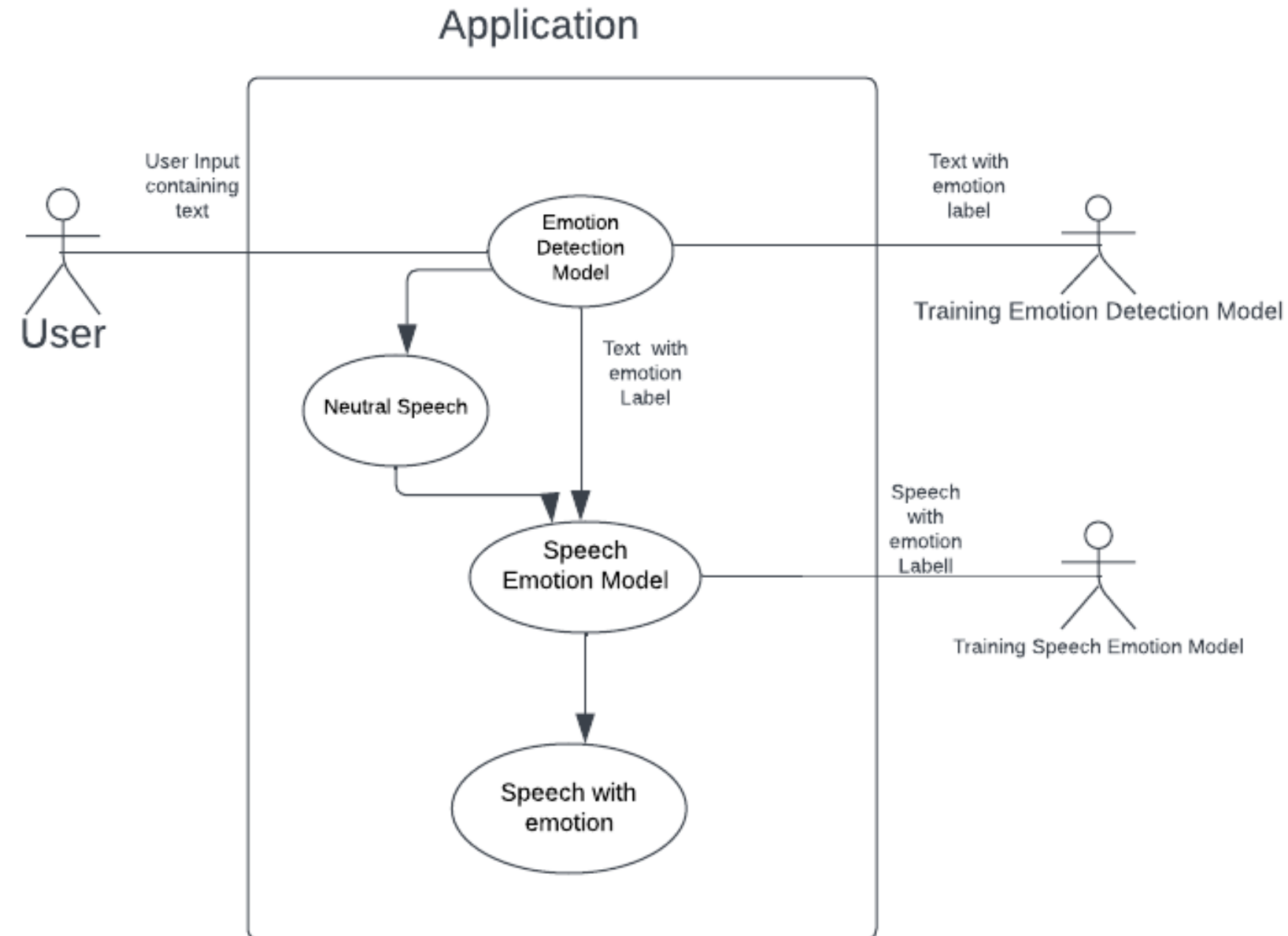


- The input data consist of dataset from various sources like social media, poetry, paragraph .Preprocessing is done on the text dataset like removing stop words, explicit words,punctuation.
- The labeled text data being provided as input to the emotion detection training model. The model analyzes the text and identifies the corresponding emotion. This output is then passed on to the speech emotion detection model.
- The speech emotion detection model receives speech data that has been labeled with an emotion. The model analyzes the speech and identifies the corresponding emotion. This output is also passed on to the speech emotion model.
- The labeled text data is then passed through the speech emotion model. The model adjusts the speech to reflect the appropriate emotion based on the input text and the corresponding emotion identified by the emotion detection training model.
- The resulting output is speech that accurately reflects the emotional content of the input text. This scenario can be repeated for multiple input texts, allowing the system to generate emotional speech for a variety of different inputs.

Design Description



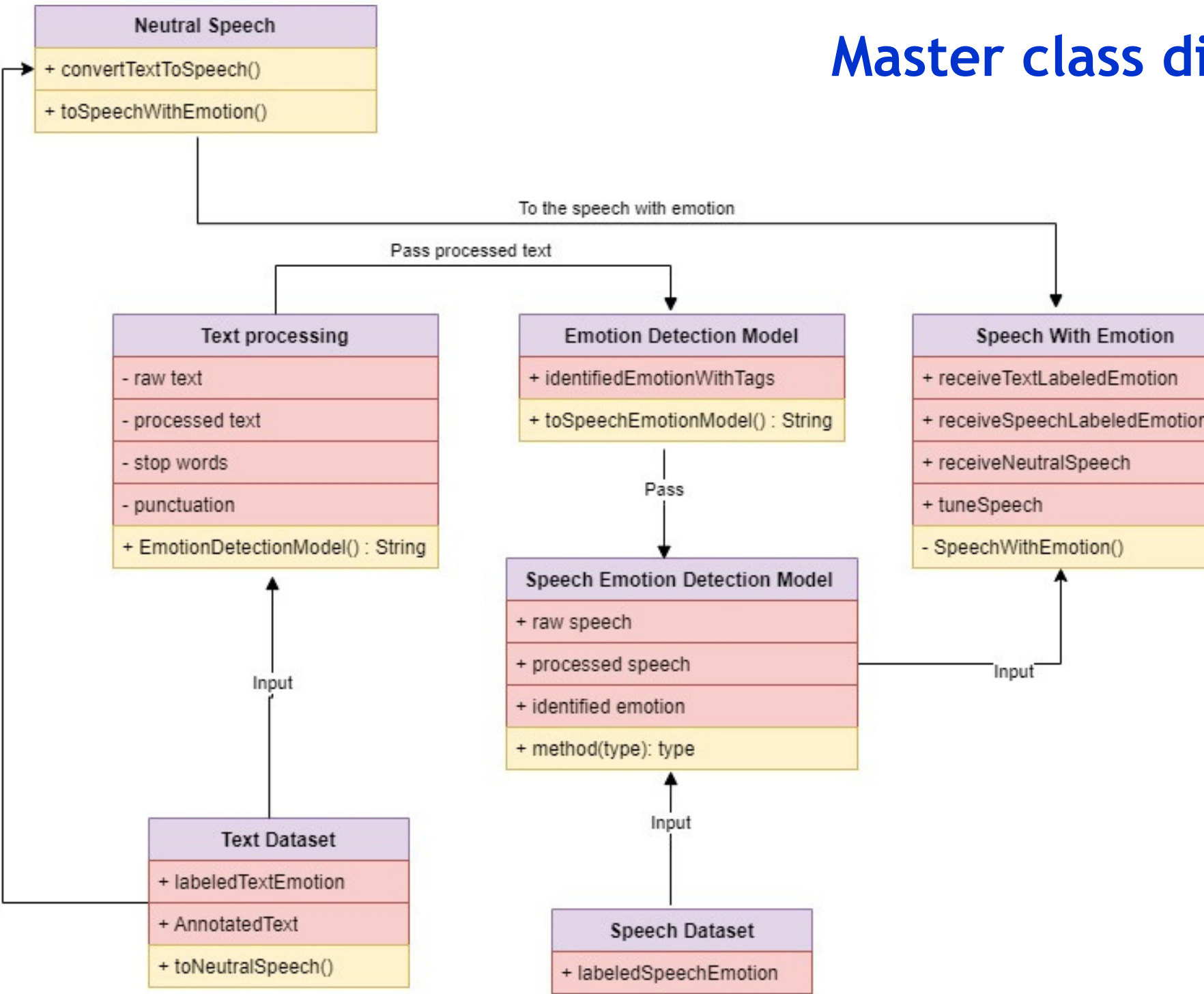
UML DIAGRAM



Design Description



Master class diagram



Design Description



"Twitter dataset for emotion detection from text"

This dataset has 40001 rows.

1. tweet_id: This column contains a unique identifier for each tweet in the dataset.
2. sentiment: This column contains the sentiment associated with each tweet. In this case, it looks like the sentiment can be one of several values, including "empty," which suggests a neutral sentiment or an absence of any strong emotion. It's possible that there are other sentiment categories in the dataset as well, but without more information it's difficult to say for sure.
3. content: This column contains the text content of each tweet. In this case, the tweets appear to be personal statements or opinions about various topics, and they may include references to other people or events.

The purpose of this dataset may be to train or test a machine learning model for emotion detection in text. By analyzing the content of each tweet and comparing it to the associated sentiment label, a model could learn to recognize patterns and predict the sentiment of new text inputs.

Design Description



"Twitter dataset for emotion detection from text"

A	B	C	D	E	F	G	H
tweet_id	sentiment	content					
1956967341	empty	@tiffanylue i know i was listenin to bad habit earlier and i started freakin at his part =[
1956967666	sadness	Layin n bed with a headache ughhhh...waitin on your call...					
1956967696	sadness	Funeral ceremony...gloomy friday...					
1956967789	enthusiasm	wants to hang out with friends SOON!					
1956968416	neutral	@dannycastillo We want to trade with someone who has Houston tickets, but no one will.					
1956968477	worry	Re-pinging @ghostridah14: why didn't you go to prom? BC my bf didn't like my friends					

Design Description



"SemEval 2018 - task E-c"

The dataset you provided is in text file consists of 12 columns:

1. ID: This column contains a unique identifier for each tweet in the dataset.
2. Tweet: This column contains the text content of each tweet, which includes a mention of two Twitter users, and a reference to a past incident involving them.
3. Anger, Anticipation, Disgust, Fear, Joy, Love, Optimism, Pessimism, Sadness: These columns contain binary values (0 or 1) that indicate whether the corresponding emotion is present in the tweet. For example, the "Anger" column has a value of 1 for this tweet, indicating that the text contains some degree of anger.
4. Trust: This column contains a binary value (0 or 1) that indicates whether the text expresses a sense of trust or confidence in something or someone.

The purpose of this dataset appears to be to train or test a machine learning model for emotion detection in text. By analyzing the content of each tweet and comparing it to the associated emotion labels, a model could learn to recognize patterns and predict the presence or absence of certain emotions in new text inputs. In this case, the tweet expresses some degree of anger and disgust, and does not express any of the other emotions in the dataset.

This dataset has 10,987 rows.

Design Description



"SemEval 2018 - task E-c"

Train dataset:

2018-E-c-En-train - Notepad

ID	Tweet	anger	anticipation	disgust	fear	joy	love	optimism	pessimism	sadness	surprise	trust
2017-En-21441	"Worry is a down payment on a problem you may never have". Joyce Meyer. #motivation #leadership #worry	0	1	0	0	0	0	0	0	0	0	0
2017-En-31535	Whatever you decide to do make sure it makes you #happy.	0	0	0	0	1	1	1	0	0	0	0
2017-En-21068	@Max_Kellerman it also helps that the majority of NFL coaching is inept. Some of Bill O'Brien's play calling was wow, ! #GOPATS	0	0	0	0	0	0	0	0	0	1	0
2017-En-31436	Accept the challenges so that you can literally even feel the exhilaration of victory.' -- George S. Patton	0	0	0	0	0	0	0	0	0	0	1
2017-En-22195	My roommate: it's okay that we can't spell because we have autocorrect. #terrible #firstworldprobs	1	0	1	0	0	0	0	0	0	0	0

Test dataset:

2018-E-c-En-test - Notepad

ID	Tweet	anger	anticipation	disgust	fear	joy	love	optimism	pessimism	sadness	surprise	trust
2018-En-01559	@Adnan__786__ @AsYouNotWish Dont worry Indian army is on its ways to dispatch all Terrorists to Hell	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE
2018-En-03739	Academy of Sciences, eschews the normally sober tone of scientific papers and calls the massive loss of wildlife a "biological annihilation	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE
2018-En-00385	I blew that opportunity -_- #mad	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE
2018-En-03001	This time in 2 weeks I will be 30... ☹	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE

dev dataset:

2018-E-c-En-dev - Notepad

ID	Tweet	anger	anticipation	disgust	fear	joy	love	optimism	pessimism	sadness	surprise	trust
2018-En-00866	@RanaAyyub @rajnathsingh Oh, hidden revenge and anger...I remember the time, she rebutted you.	1	0	1	0	0	0	0	0	0	0	0
2018-En-02590	I'm doing all this to make sure you smiling down on me bro	0	0	0	0	1	1	1	0	0	0	0
2018-En-03361	if not then #teamchristine bc all tana has done is provoke her by tweeting shady shit and trying to be a hard bitch begging for a fight	1	0	0	0	0	0	0	0	0	0	0
2018-En-03230	It is a #great start for #beginners to jump into auto #trading. PROFITABLE FX EA will give you full support, manuals & Team Viewer support.	0	0	0	0	0	0	0	0	0	0	0
2018-En-01143	My best friends driving for the first time with me in the car #terrifying	0	0	0	0	1	0	0	0	0	0	0
2018-En-04301	Hey @SuperValuIRL #Fields in #skibbereen give your online delivery service a horrible name. 1.5 hours late on the 1 hour delivery window.	1	0	0	0	0	0	0	0	0	0	0
2018-En-02651	Why have #Emmerdale had to rob #robron of having their first child together for that vile woman/cheating sl smh #bitter	1	0	1	0	0	0	0	0	0	0	0
2018-En-03058	@ThomasEWoods I would like to hear a podcast of you going off refuting her entire article. Extra indignation please.	1	0	1	0	0	0	0	0	0	0	0

Design Description



“GoEmotions: A Dataset of Fine-Grained Emotions”,

- The dataset consists of labeled text samples with corresponding ratings for various emotions and sentiments.
- Each row in the dataset represents a single text sample and contains information about the ID, author, subreddit, link ID, parent ID, and creation time.
- The dataset includes 34 different emotion/sentiment categories, including admiration, amusement, anger, annoyance, approval, caring, confusion, curiosity, desire, disappointment, disgust, embarrassment, excitement, fear, gratitude, grief, joy, love, nervousness, optimism, pride, realization, relief, remorse, sadness, surprise, and neutral.
- Each emotion/sentiment category has a corresponding column in the dataset, and the values in each column represent the rating for that emotion/sentiment on a scale of 0-1.
- A rating of 0 indicates the absence of the emotion/sentiment, while a rating of 1 indicates the presence of the emotion/sentiment.
- The dataset appears to have been annotated by multiple raters, as there is a separate column for each rater ID.
- Additionally, there is an "example_very_unclear" column, which may indicate that some samples were difficult to categorize.

Design Description



“GoEmotions: A Dataset of Fine-Grained Emotions”,

text	admiration	amusement	anger	annoyance	approval	confusion	curiosity	fear	grief	joy	love	nervousness	optimism	pride	realization	relief	remorse	sadness	surprise	neutral
We can hope	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Shhh don't give them the idea!	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Thank you so much, kind stranger. I really need that	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Design Description



“Voice Datasets” : https://github.com/jim-schwoebel/voice_datasets

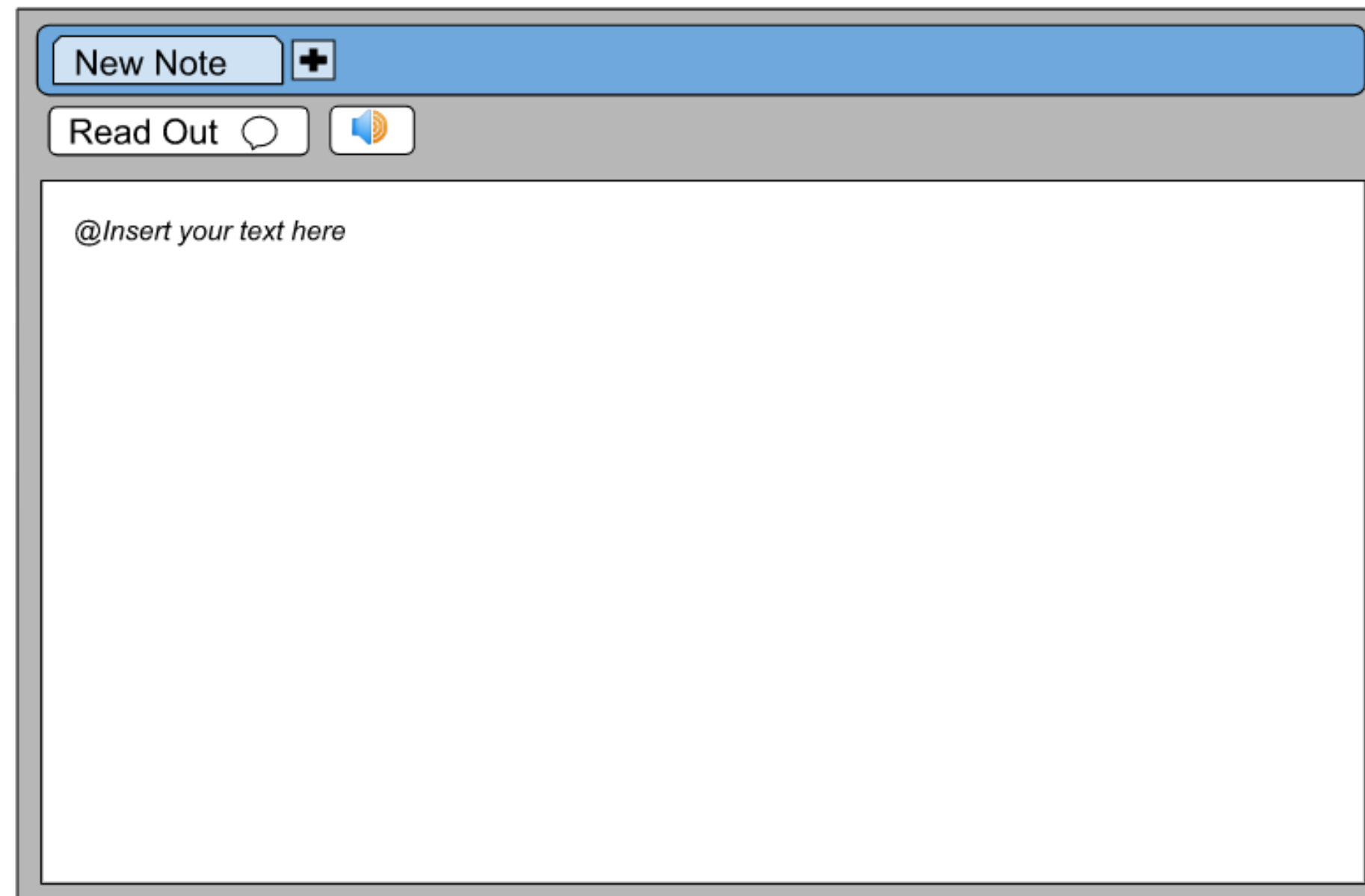
The voice datasets repository on GitHub contains a collection of various audio datasets that can be used for research and development of speech and audio processing applications. The datasets include recordings of speech, music, and ambient sounds, as well as metadata and annotations for some of the datasets.

- AESDD - around 500 utterances by a diverse group of actors (over 5 actors) simulating various emotions.
- ANAD - 1384 recording by multiple speakers; 3 emotions: angry, happy, surprised.
- BAVED - 1935 recording by 61 speakers (45 male and 16 female).
- Common Voice - Common Voice is Mozilla's initiative to help teach machines how real people speak. 12GB in size; spoken text based on text from a number of public domain sources like user-submitted blog posts, old books, movies, and other public speech corpora.
- EmotionTTS - Recordings and their associated transcriptions by a diverse group of speakers - 4 emotions: general, joy, anger, and sadness.
- Emov-DB - Recordings for 4 speakers- 2 males and 2 females; The emotional styles are neutral, sleepiness, anger, disgust and amused.
- EMOVO - 6 actors who played 14 sentences; 6 emotions: disgust, fear, anger, joy, surprise, sadness.
- SAVEE Dataset - 4 male actors in 7 different emotions, 480 British English utterances in total.

Design Description

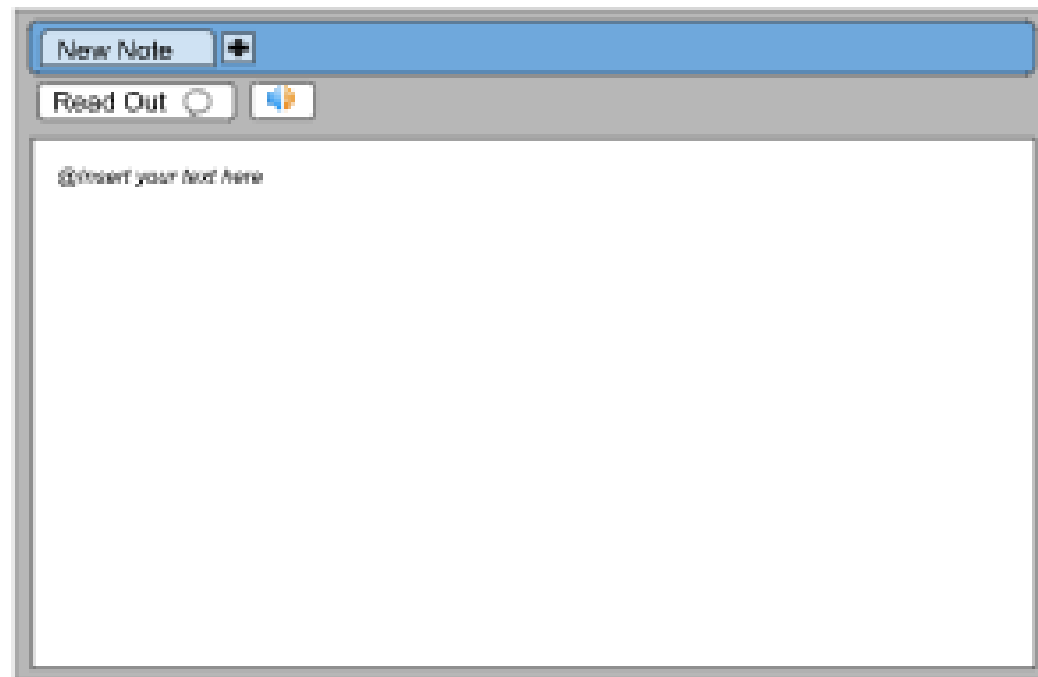


User Interface Diagram

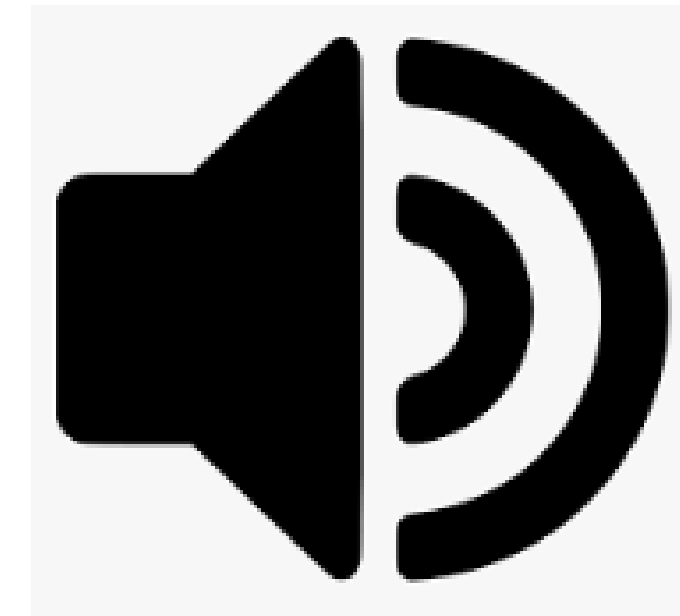


Design Description

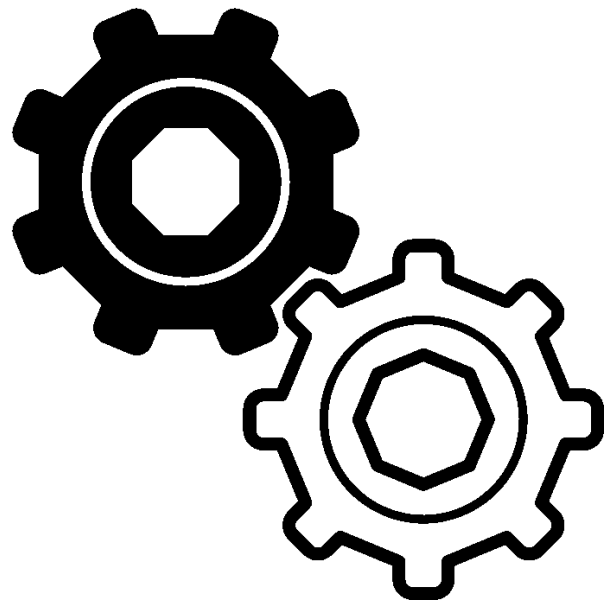
External Interfaces



Text Input From the user



Speech is generated with emotion



Technologies Used - M H Sohan

1. **Python programming language** - used to develop the software for building the sentiment lexicon and performing sentiment analysis.
2. **Natural Language Toolkit (NLTK)** - a Python library used for natural language processing tasks such as tokenization, stemming, and POS tagging.
3. **Deep Learning Models:** compares the performance of four different deep learning models, including Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and Bidirectional Encoder Representations from Transformers (BERT). They used pre-trained word embeddings such as GloVe and FastText for all models except BERT.
4. **Sentiment Analysis Tools:** VADER (Valence Aware Dictionary and sEntiment Reasoner) tool to evaluate the performance of their models.
5. **WordNet** - a lexical database of English words used to retrieve synonyms, antonyms, and related words.
6. **Google Translate API** - a cloud-based service used to translate text into different languages for cross-lingual sentiment analysis.



Project Demo - M H Sohan

```
import re
import nltk
import string
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from nltk.corpus import stopwords
from nltk.stem import SnowballStemmer, WordNetLemmatizer
from sklearn.feature_extraction.text import TfidfVectorizer

from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split

from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
import string
import re
from nltk.stem import WordNetLemmatizer
import tensorflow as tf
from tensorflow import keras

from keras.preprocessing.text import Tokenizer
from keras.preprocessing.sequence import pad_sequences
from keras.models import Sequential
from keras.layers import Dense, Embedding, LSTM, GRU, Bidirectional

from gensim.models import Word2Vec

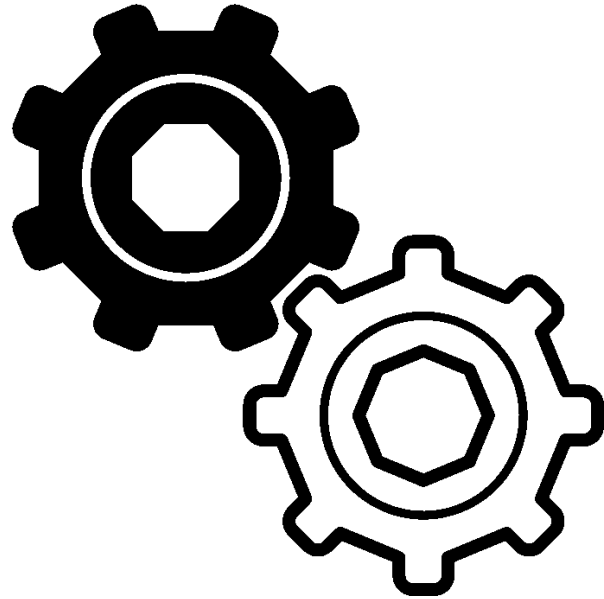
from keras.callbacks import EarlyStopping
from keras.models import load_model
```

```
df_train = pd.read_csv('train.txt', names=['Text', 'Emotion'], sep=';')
df_val = pd.read_csv('val.txt', names=['Text', 'Emotion'], sep=';')
df_test = pd.read_csv('test.txt', names=['Text', 'Emotion'], sep=';')
```

```
sentences = train_data.tolist()
```

```
# Tokenize the sentences into words
tokenizer = Tokenizer()
tokenizer.fit_on_texts(sentences)
sequences = tokenizer.texts_to_sequences(sentences)
```

```
## Basically we have extracted all the sentences and then assigned each word a number
```



Technologies Used - RAHUL ROSHAN G

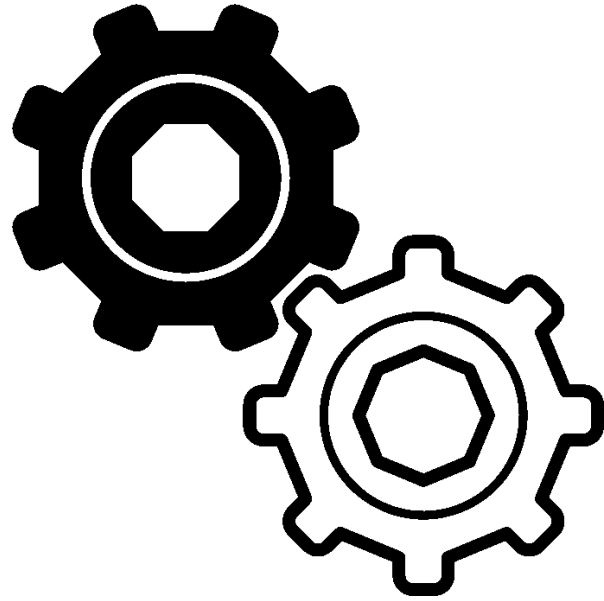
1. **Natural Language Processing (NLP) techniques:** This includes methods such as sentiment analysis, text classification, and topic modeling, which are used to extract emotions and other related information from text.
2. **Machine learning algorithms:** These are used to train models that can accurately detect emotions in text. Commonly used algorithms include decision trees, support vector machines (SVMs), and deep learning techniques such as Recurrent Neural Networks (RNNs).
3. **Lexicons and Dictionaries:** These are databases of words and phrases that are pre-labeled with their corresponding emotions. They are used to match the text with the most probable emotion.
4. **Emotion Recognition APIs:** There are several emotion recognition APIs available, such as Microsoft Azure's Text Analytics API, IBM Watson's Natural Language Understanding, and Google Cloud Natural Language API. These APIs provide pre-built models for detecting emotions from text, which can be integrated into applications.
5. **Hybrid models:** These are models that combine multiple techniques, such as NLP, machine learning, and lexicons, to improve the accuracy of emotion detection.

Designed Pseudocode/Algorithm for emotion from text



```
1 # Import the required libraries
2 import nltk
3 from nltk.sentiment.vader import SentimentIntensityAnalyzer
4
5 # Create an instance of the SentimentIntensityAnalyzer class
6 sia = SentimentIntensityAnalyzer()
7
8 # Define a function to detect emotions from text
9 def detect_emotions(text):
10     # Tokenize the text into individual words
11     tokens = nltk.word_tokenize(text)
12
13     # Calculate the sentiment scores for each word
14     scores = [sia.polarity_scores(token) for token in tokens]
15
16     # Calculate the average score for each emotion across all words
17     emotions = {}
18     for score in scores:
19         for emotion in score:
20             if emotion not in emotions:
21                 emotions[emotion] = score[emotion]
22             else:
23                 emotions[emotion] += score[emotion]
24     for emotion in emotions:
25         emotions[emotion] /= len(scores)
26
27     # Return the emotions and their corresponding scores
28     return emotions
29
```

Technologies Used - S M SUTHARSAN RAJ



There are several Python libraries that can be used for speech emotion recognition, including:

1. librosa: a library for audio and music analysis
2. PyAudio: a library for audio input and output
3. SpeechRecognition: a library for speech recognition
4. scikit-learn: a library for machine learning algorithms
5. TensorFlow: a popular library for deep learning
6. Keras: a high-level neural networks API
7. PyTorch: another popular library for deep learning

1.Convolutional Neural Networks (CNN):

CNNs are well-suited for speech recognition tasks, as they can effectively learn features from raw data and capture spatial and temporal dependencies.

2. Deep Stridal CNN (DSCNN)

DSCNN can automatically learn high-level features from the input data, allowing it to capture complex relationships between speech features and emotions. This means that it can achieve higher accuracy in emotion recognition compared to traditional feature-based methods.

Depending on the specific approach and model used for speech emotion recognition, some of these libraries may be more useful than others. For example, librosa may be useful for extracting features from audio data, while scikit-learn or TensorFlow may be useful for training machine learning or deep learning models.



Project Demo - S M SUTHARSAN RAJ

- Designed Pseudocode/Algorithm

```
from keras.models import Sequential
from keras.layers import Conv1D, MaxPooling1D, Dense, Dropout, Flatten
from keras.layers.normalization import BatchNormalization
from keras.layers.advanced_activations import ReLU

Define the DSCNN model architecture
del = Sequential()
del.add(Conv1D(64, kernel_size=3, padding='same', input_shape=input_shape))
del.add(BatchNormalization())
del.add(ReLU())
del.add(MaxPooling1D(pool_size=2))
del.add(Conv1D(64, kernel_size=3, strides=2, padding='same'))
del.add(BatchNormalization())
del.add(ReLU())
del.add(MaxPooling1D(pool_size=2))
del.add(Conv1D(128, kernel_size=3, padding='same'))
del.add(BatchNormalization())
del.add(ReLU())
del.add(MaxPooling1D(pool_size=2))
del.add(Flatten())
del.add(Dense(128, activation='relu'))
del.add(Dropout(0.5))
del.add(Dense(num_classes, activation='softmax'))
```

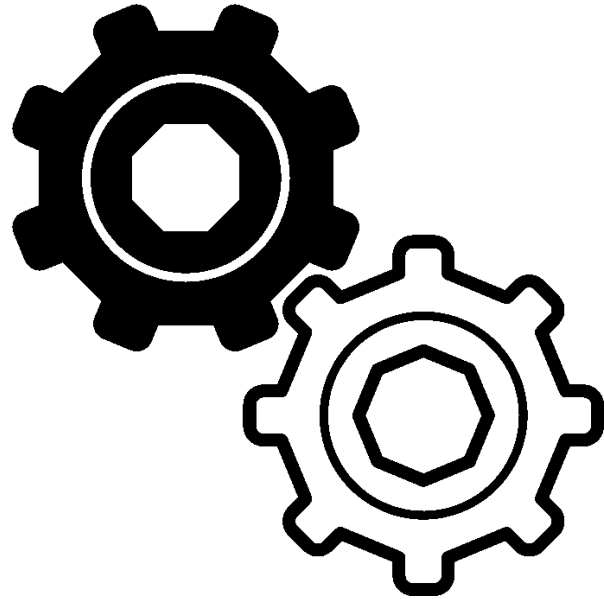
```
# Compile model with appropriate loss, optimizer, & metrics
model.compile(loss='categorical_crossentropy',
              optimizer='adam', metrics=['accuracy'])

# Train the model
history = model.fit(x_train, y_train, batch_size=batch_size,
                  epochs=epochs, validation_data=(x_val, y_val))

# Evaluate the model on the test set
loss, accuracy = model.evaluate(x_test, y_test,
                              batch_size=batch_size)

# Save the model
model.save('dscnn_model.h5')
```


Technologies Used - ROHIT ROSHAN



- Mel-Frequency Cepstral Coefficients (MFCC): A technique used to extract features from audio signals, which involves mapping the short-term power spectrum of a sound onto the mel scale and then applying the discrete cosine transform to obtain a set of cepstral coefficients.
- Gaussian Mixture Model (GMM): A statistical model used for clustering and density estimation, often used in speaker recognition and emotion recognition tasks.
- Support Vector Machine (SVM): A supervised learning algorithm used for classification and regression tasks, often used in speech and emotion recognition tasks.
- Deep Belief Network (DBN): A deep learning algorithm composed of multiple layers of restricted Boltzmann machines, often used in speech and image recognition tasks.
- Convolutional Neural Network (CNN): A deep learning algorithm that uses convolutional layers to extract features from input data, often used in image and speech recognition tasks.
- Recurrent Neural Network (RNN): A deep learning algorithm that uses recurrent connections to allow information to persist through time, often used in sequence-to-sequence tasks such as speech recognition and language translation.
- Convolutional Recurrent Neural Network (CRNN): A deep learning algorithm that combines the convolutional and recurrent layers, often used in speech and image recognition tasks.
- Variational Autoencoder (VAE): A type of generative model that learns a low-dimensional representation of input data, often used in speech and image synthesis tasks.
- Deep Bidirectional Long Short-Term Memory (DBLSTM): A variant of the RNN architecture that uses bidirectional connections and long short-term memory units, often used in speech and language modeling tasks.
- WaveNet: A deep generative model for audio waveforms that uses dilated convolutions to generate high-fidelity speech and music.
- MelGAN: A generative adversarial network (GAN) architecture for speech synthesis that uses the mel-spectrogram as input to generate high-quality speech.

Psuedo Code - ROHIT ROSHAN



Deep learning-based synthesis with PyTorch and the WaveNet architecture:

```
import torch
import torchaudio
import numpy as np
from models import WaveNet

# Load pre-trained model
model = WaveNet()
model.load_state_dict(torch.load("wavenet.pt"))
model.eval()

# Load audio file to synthesize
audio, sr = torchaudio.load("input_audio.wav")

# Normalize audio
max_amplitude = torch.max(torch.abs(audio))
audio = audio / max_amplitude

# Convert audio to mu-law encoding
mu_law_audio = torchaudio.functional.mu_law_encoding(audio)

# Synthesize audio using model
generated_audio = torch.zeros((1, 1))
with torch.no_grad():
    for i in range(len(mu_law_audio)):
        output = model(generated_audio)
        output = torch.softmax(output, dim=1)
        mu = torch.multinomial(output.squeeze(), 1)
        generated_audio = torch.cat((generated_audio, mu.float().unsqueeze(1)), dim=1)

# Convert mu-law encoding to audio
generated_audio = generated_audio.squeeze().cpu().numpy()
generated_audio = torchaudio.functional.mu_law_decoding(generated_audio, 256)
generated_audio = generated_audio * max_amplitude

# Save synthesized audio
torchaudio.save("synthesized_audio.wav", generated_audio, sr)
```

Open-source software ArtiSynth

```
# Import required packages
from maspack.matrix import VectorNd
from artisynth.core.mechmodels.Point import Point
from artisynth.core.modelbase import Model

# Define model parameters
length = 0.1
width = 0.02
height = 0.02
density = 1000.0

# Create model
model = Model("Articulatory Synthesis")
root = model.getRoot()

# Create tongue geometry
tongue = Point()
tongue.setSize(length, width, height)
tongue.setMass(density*length*width*height)
tongue.setLocalPosition(VectorNd([0.0, 0.0, 0.0]))
root.addRigidBody(tongue)

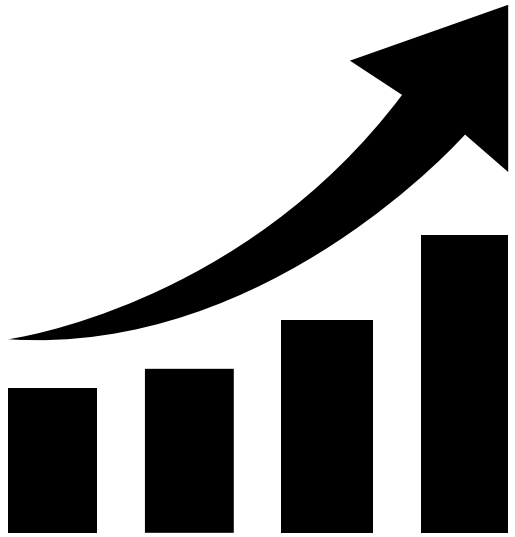
# Create jaw geometry
jaw = Point()
jaw.setSize(length, width, height)
jaw.setMass(density*length*width*height)
jaw.setLocalPosition(VectorNd([0.0, -width/2.0, 0.0]))
root.addRigidBody(jaw)

# Define movement parameters
tongueMovement = VectorNd([-0.1, 0.0, 0.0])
jawMovement = VectorNd([0.05, 0.0, 0.0])

# Move tongue and jaw
tongue.setPosition(tongue.getPosition().plus(tongueMovement))
jaw.setPosition(jaw.getPosition().plus(jawMovement))

# Run simulation
model.simulate(0.1)
```

Project Progress



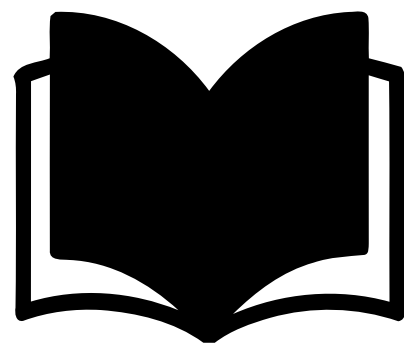
What is the project progress so far?

- In this project, we are now ready with all the architecture, high-level design, algorithms and tools required, methodologies from literature survey, etc...
- We are aware of the limitations and expected deliverables for the project.
- We have also looked for the datasets, both text and speech labelled with emotion which are annotated. We are ready with tools and other requirements and particularly the flow of the project.
- Now, we are to just filter our datasets used for training and then start with the implementation.
- We also have a text editor ready-in-hand for the model to be integrated with.

What is the percentage completion of the project?

- Therefore referring to the above progress we feel that we have completed around 30 % of the project

References



- [1] X. Cai, D. Dai, Z. Wu, X. Li, J. Li and H. Meng, "Emotion Controllable Speech Synthesis Using Emotion-Unlabeled Dataset with the Assistance of Cross-Domain Speech Emotion Recognition," ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 2021, pp. 5734-5738, doi: 10.1109/ICASSP39728.2021.9413907
- [2] *Effective Text Data Preprocessing Technique for Sentiment Analysis in Social Media Data* Saurav Pradha School of Computing and Mathematics Charles Sturt University Melbourne, Victoria, Senior Australia saurav.pradha54@gmail.com Malka N. Halgamuge Member, IEEE Dep. of Electrical and Electronic Engineering The University of Melbourne Victoria 3010, Australia malka.nisha@unimelb.edu.au Nguyen Tran Quoc Vinh Faculty of Information Technology The University of Da Nang - University of Science and Education, Vietnam ntquocvinh@ued.udn.vn.
- [3] Wang, Yuxuan, R. J. Skerry-Ryan, Daisy Stanton, Yonghui Wu, Ron J. Weiss, Navdeep Jaitly, Zongheng Yang, Ying Xiao, Z. Chen, Samy Bengio, Quoc V. Le, Yannis Agiomyrgiannakis, Robert A. J. Clark and Rif A. Saurous. "Tacotron: Towards End-to- End Speech Synthesis." Interspeech (2017).
- [4] 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.

Thank You