

National Institute of Technology, Calicut

Interview Score Prediction and Analysis Using Audio Features and NLP

SUBMITTED BY:–
GROUP 38

Palash Bajpai
B180759CS

Ritik Gautam
B180630CS

Aditya Jha
B180648CS

GUIDED BY:–

Dr. Raju Hazari
(Department of CSE)
NIT Calicut

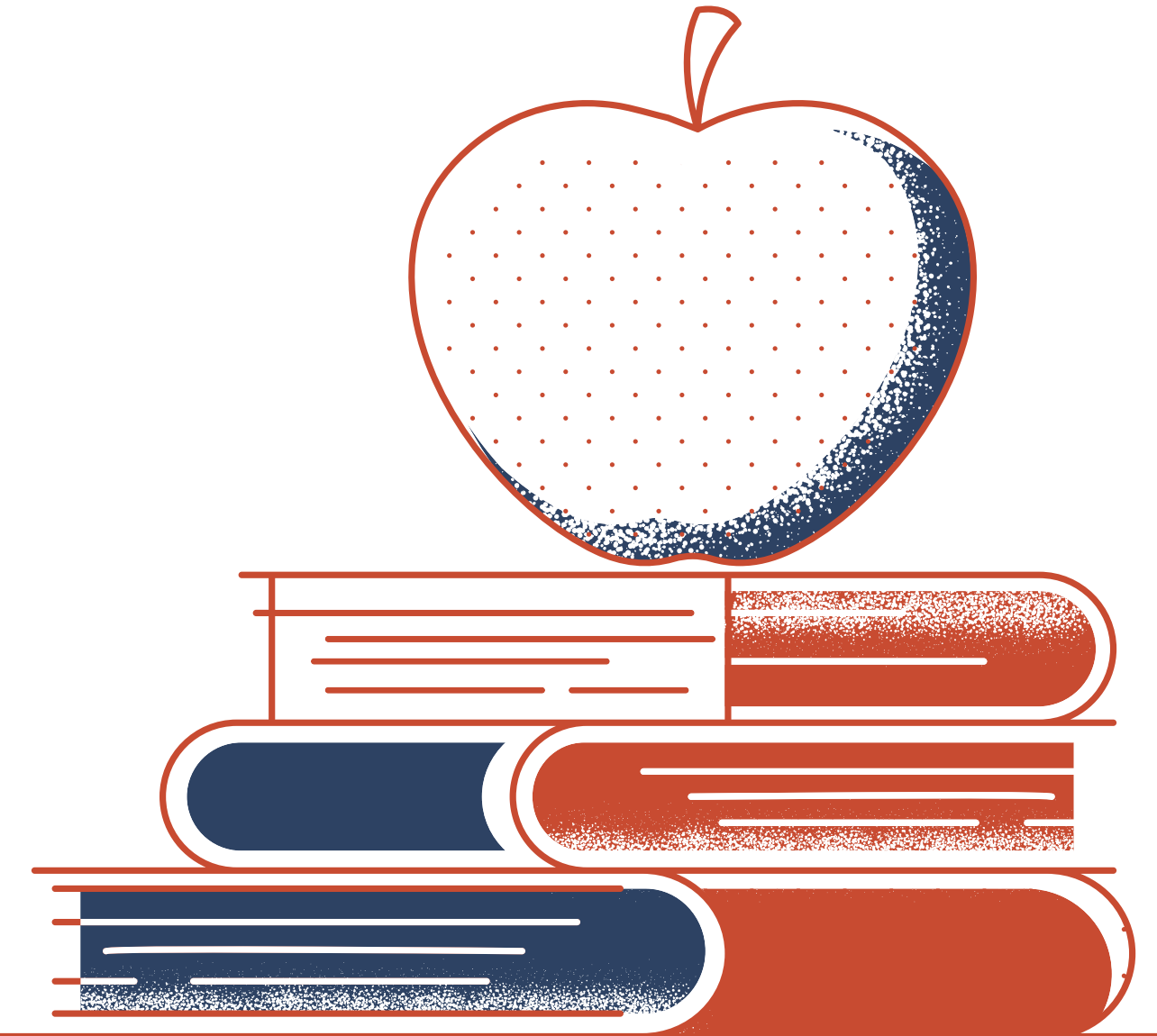


To build a suggestion-based interview analysis tool that can help companies to score candidates on the basis of lexical and prosodic features also gives feedback and suggestions to interviewees.



- The traditional interview is a test of personality traits for many hiring managers.
- Prosodic features can help to get insights on emotions, confidence, anxiety, etc which can help interviewers to judge the candidate.
- By extracting the transcript from available audio and applying NLP and other machine learning models we can score the answers given by interviewees.
- Our project uses machine learning in order to facilitate the faster assessment of candidates by using prosodic features and NLP, ultimately improving the interviewer's ability to make hiring decisions based on more reliable data that is less susceptible to human biases

Our project proposes a model to design and implement an automated prediction framework for quantifying the ratings of job interviews based on audio features (prosodic, lexical), given the audio recordings. The prediction framework automatically extracts a diverse set of multimodal features (lexical and prosodic) and quantifies the overall interview performance, the likelihood of getting hired, and other social traits relevant to the job interview process. We would be extending the idea of interview score prediction to more like a virtual coach or a tool that would help the candidates to improve them by going through the feedback and suggestions given by the tool.



AUDIO DATASET

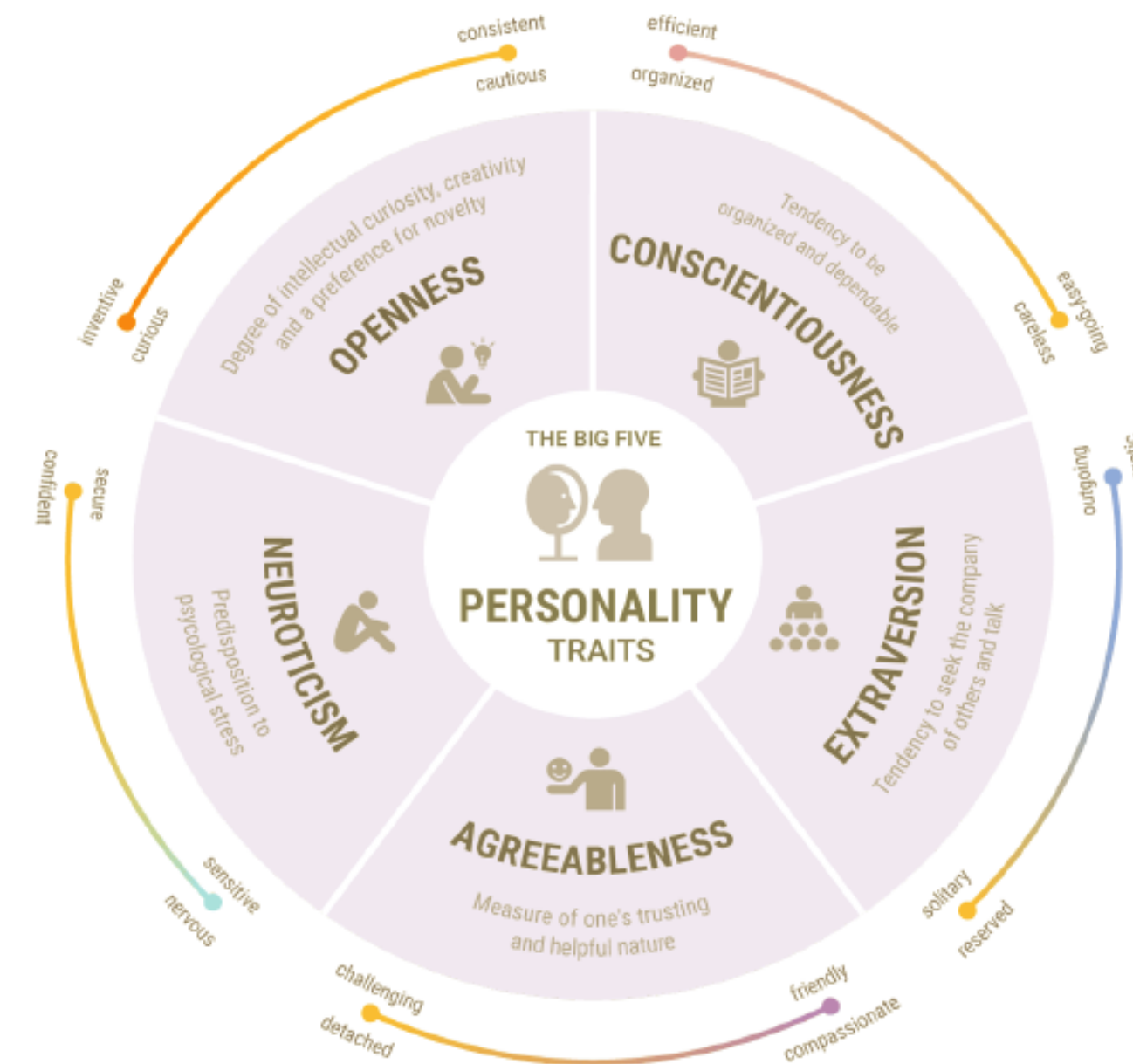
- For the audio dataset, we have used the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS).
- For the preparation of this dataset, researchers asked 24 professional actors (12 female, 12 male) to pronounce two identical statements in a North American accent. This produces 7356 files which end up being 24.8 GB of data.
- Each file is scored 10 times on emotional validity, intensity, and genuineness. Ratings were provided by 247 individuals who were characteristic of untrained adult research participants from North America.
- A further set of 72 participants provided test-retest data. Since this dataset includes both speech and songs, we had a speech that contains calm, happy, sad, angry, fearful, surprised, and disgusted expressions. Song includes emotions such as calm, happy, angry, sad, and fearful emotions

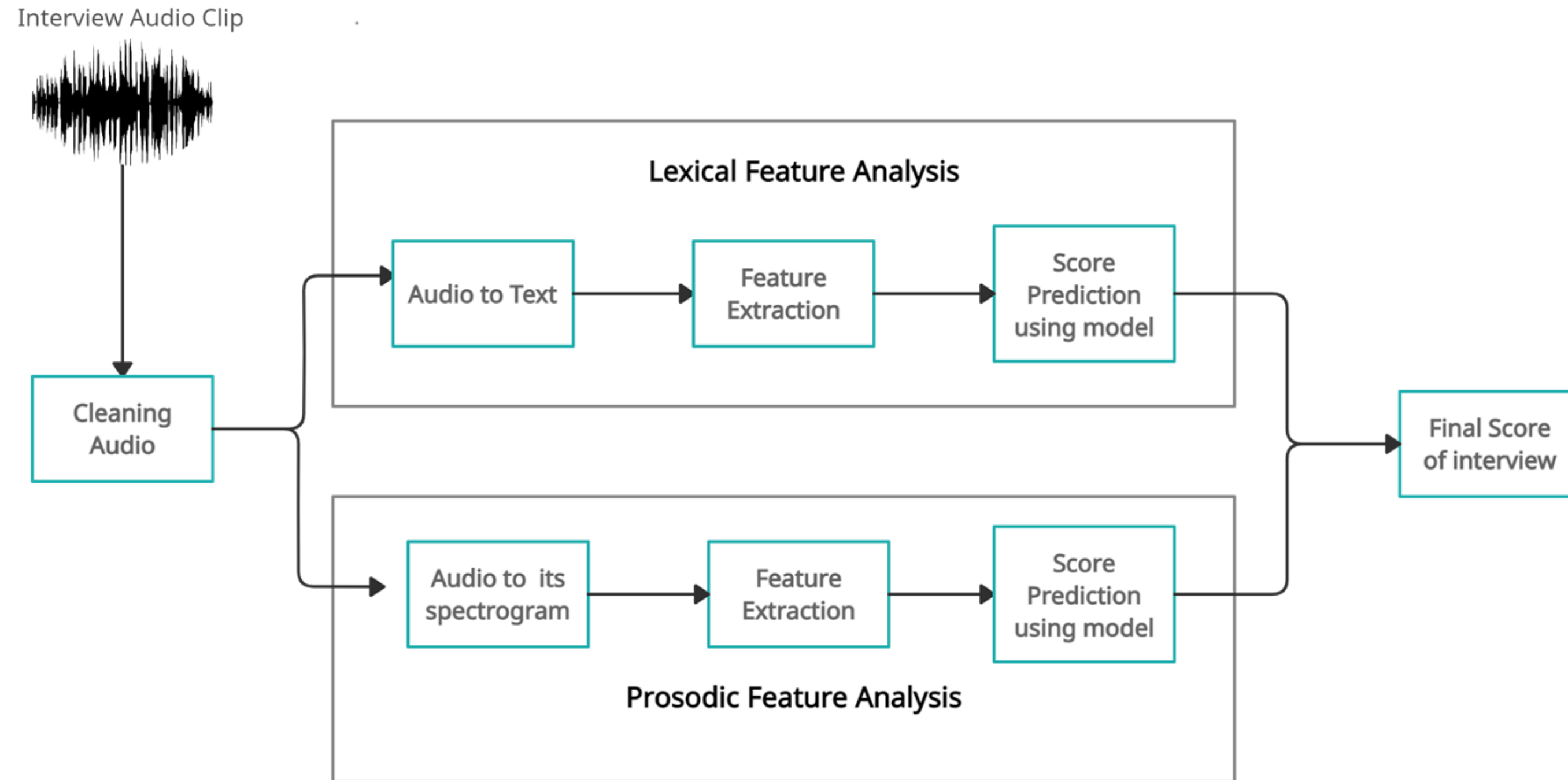
TEXT DATASET

- To train our lexical model we used a Stream-of consciousness data set created during the study of language styles by Pennebaker and King. Their research investigates the authenticity, character structure, and legitimacy of the written language using a computer-based word analysis system. This database used daily texts for 34 psychology students.
- In this case, 34 students include 29 females and 5 males, with ages ranging from 18 to 67 with an average of 26.4. All these writings form a dataset comprising 2468 records. Students were expected to write at least 20 minutes a day on a specific topic for each activity.
- The scope of personality traits in the sheet is expressed in terms of "y" and "n" representing yes and no to indicate high and low scores.
- To earn points they used the Big Five innovation techniques to select the best features to score human goals.

BIG FIVE PERSONALITY TRAITS

1. **Openness:** This trait reflects a person's imagination and insight. People with high values for this trait tend to be creative, adventurous, and very curious.
2. **Conscientiousness:** People with strong conscious tendencies are more organized and focused on their goals. Such people like a set schedule and have self-control.
3. **Extraversion:** People with this tendency are considered lively, talkative, assertive, and emotionally expressive. They like to hang out with other people, start conversations, and enjoy the fact that they are in the spotlight. Such people can make new friends easily.
4. **Agreeableness:** This dimension of personality includes attributes such as trust, affection, compassion, kindness, and other prosocial behaviors.
5. **Neuroticism:** People with this trait tend to be emotionally unstable. They are sad, irritable, and suffer from anxiety and mood swings.





Input: The input to our framework will be an audio file that contains the answer to the question asked.

Output: It would be a score given to each candidate out of 100.

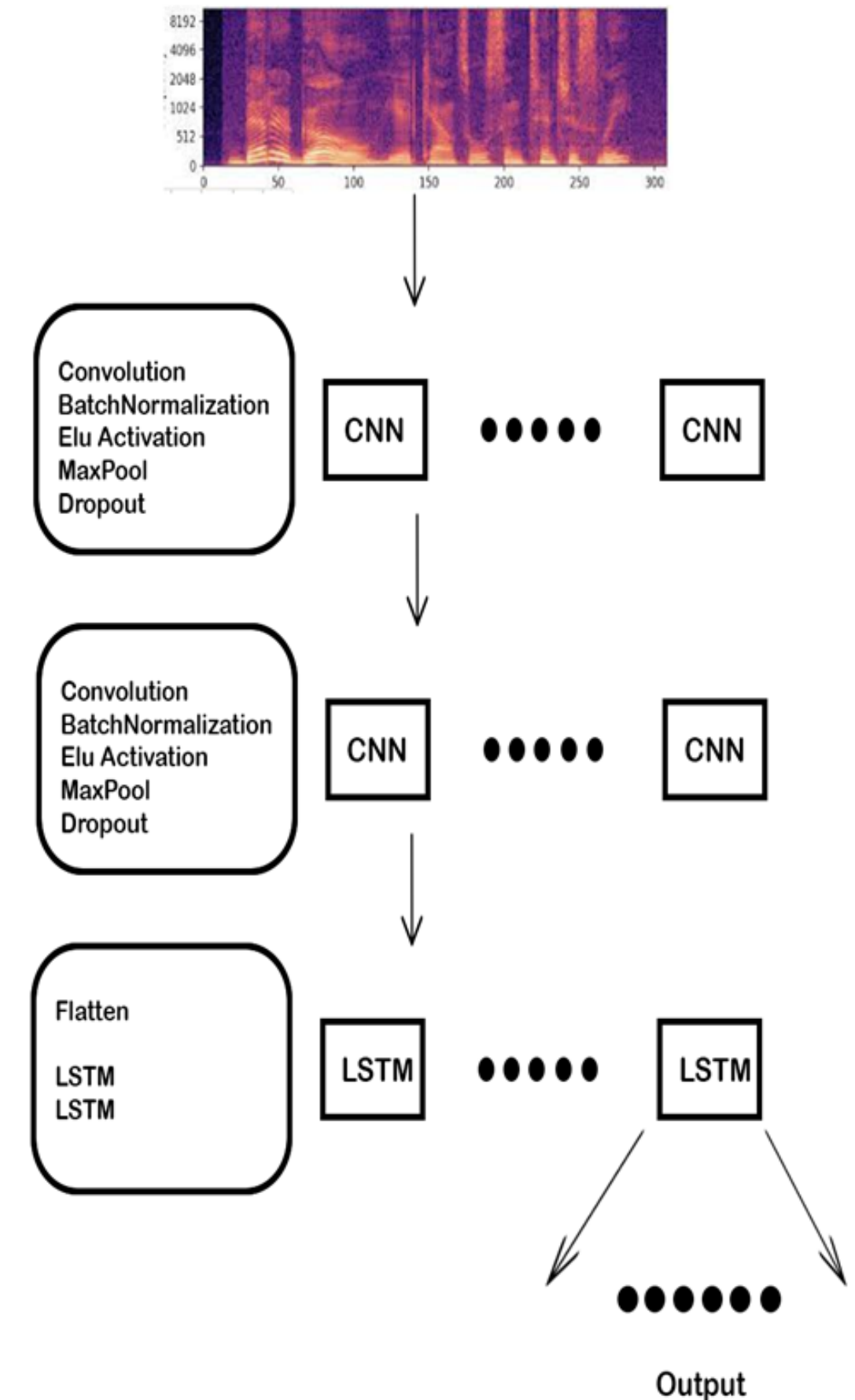
AUDIO ANALYSIS MODEL

PIPELINE:

1. Audio signal input
2. Signal discretization
3. Convert to spectrogram
4. Convert spectrogram to mel scale
5. Split spectrogram in windows
6. Predict using a pre-trained model

OUTPUT: A total of 6 emotions are recognized by our audio analysis model namely calm, happy, sad, angry, fearful, surprised, and disgusted. This model will tell which emotion was predominant for your speech and also shows emotion distribution. This will also return a comparison of your emotional distribution with that of others. This emotion distribution will be converted to scores to grade your interview.

ACCURACY: 75%



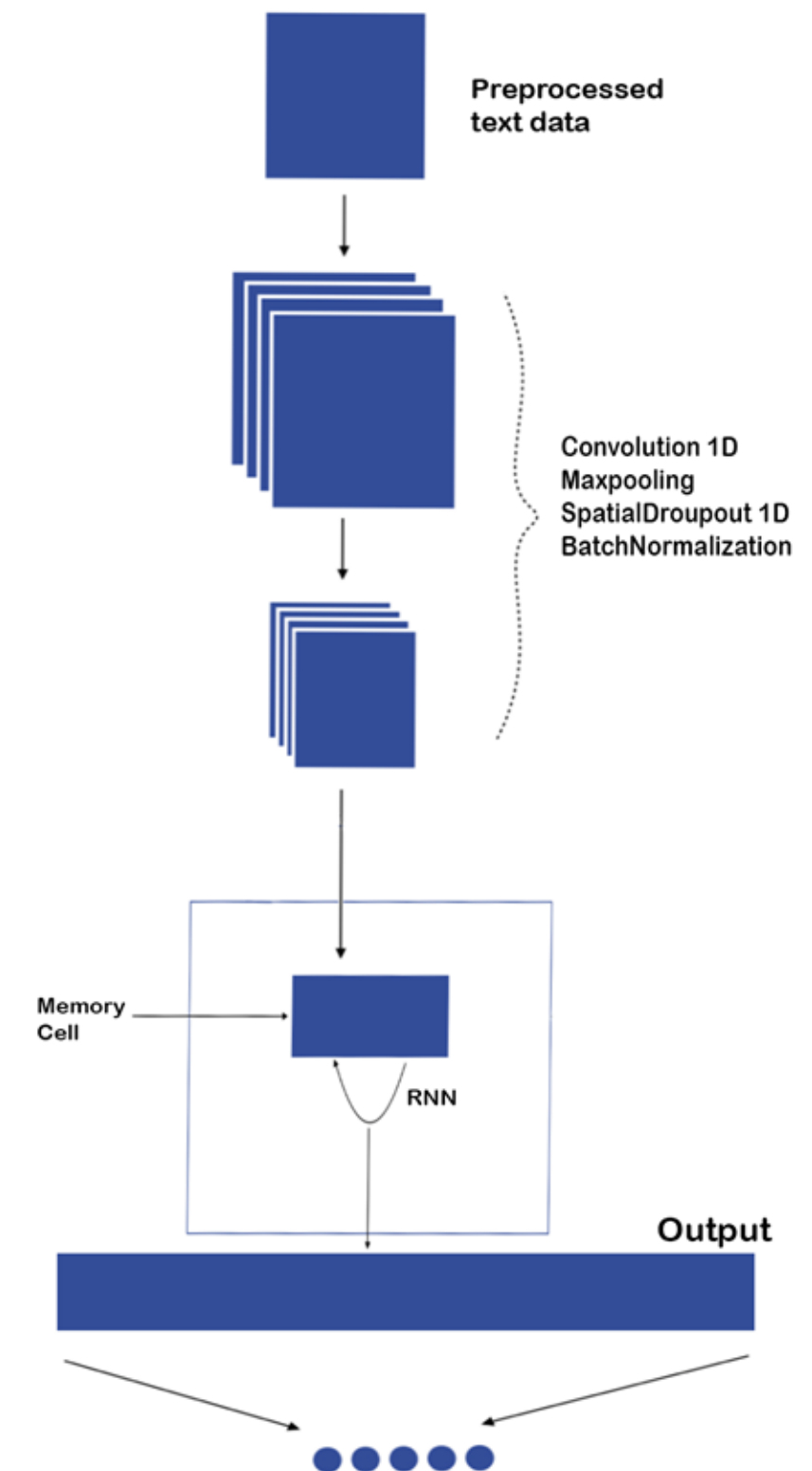
LEXICAL MODEL

PIPELINE:

1. Tokenization
2. Using regular expressions for standardization and deletion of punctuations
3. Converting tokens to lowercase
4. Removal of stopwords (like 'a', 'an')
5. Padding the sequence of tokens
6. Sending this data to a pre-trained model
7. Predict using a pre-trained model

OUTPUT: Our text analysis model will evaluate a total of five personality traits that can affect your interview namely openness, conscientiousness, extraversion, agreeableness, and neuroticism. The model shows which traits dominate the response and also shows the distribution of personality traits. It also returns the result of comparing your distribution to others. This distribution of personality traits translates into scores for interview scores.

ACCURACY: 53%



- We shifted from MIT dataset to 2 separate datasets for both audio and lexical models.
- We have understood the problem and the requirements to solve the problem efficiently and more accurately.
- We have fragmented our audio files into subparts based on the duration of questions asked, each part contains one question and its answer, and this helps to generate more data, which will help our model to work more accurately.
- We have extracted features from both audio files and stored their values in excel sheets.
- We have made a model for audio analysis using Time Distributed Convolution Neural Network.
- We made a model for lexical analysis using a one-dimensional convolution neural network and recurrent neural network. LSTM is also used to leverage on sequential nature of natural language and to get better accuracy.
- We have modularized our approach.
- Tested each component separately on different test cases.

- Explore more datasets like TESS and SAVEE for the audio analysis model.
- Modify dataset to make it more unbiased to factors like gender distribution.
- Work on the accuracy of all models and test them on different test cases.
- Build a framework to score interviews based on outputs from both lexical and audio analysis models.
- Test each component separately. The components should be loosely coupled and split the features for each component eg : (One component for prosodic features, one for lexical, etc).
- Try to build a suggestion-based tool that takes our model and gives feedback and suggestions to the candidate.
- Make a website and a mobile or pc based application for our model
- Present the final report and give a demo of our work to the panel.



The scope of this project has not been explored much, and there are many new ideas that can be implemented. Interview analysis is used by HR and companies to determine the suitability of a person for the job. This project can even be used by students to train and prepare themselves for interviews. With both audio and lexical features, we use two important techniques of human behavior analysis. This project also includes the use of quite different areas of machine learning and thus we will be learning quite new things. I hope this project gets built according to company standards and we can use this project in real life.



The End

Thank you

