



# IMPLEMENTATION OF SPEECH BASED EMOTION RECOGNITION ON RAVDESS EMOTIONAL SPEECH DATASET

## **PRESENTED BY**

Devansh Tiwari (21070126025)

Dhrubo Bhattacharjee (21070126026)

Harshit Jain (21070126034)

# Motivation :

- Improving communication
- Mental health applications
- Marketing and advertising
- Entertainment
- Research

- **Improving communication**

Emotion recognition technology can help improve communication between people, particularly those with communication difficulties. For example, individuals on the autism spectrum may have difficulty interpreting and expressing emotions through verbal and nonverbal cues. Emotion recognition technology can help bridge this communication gap.

- **Mental health applications**

Emotion recognition technology can be useful in mental health applications, such as detecting and monitoring mood disorders like depression and anxiety. It can also be used to provide emotional support to people in distress, such as through virtual assistants or chatbots.

- **Marketing and advertising**

Emotion recognition technology can be used to understand customer reactions to marketing and advertising campaigns, helping businesses optimize their messaging and improve customer engagement.

- **Entertainment**

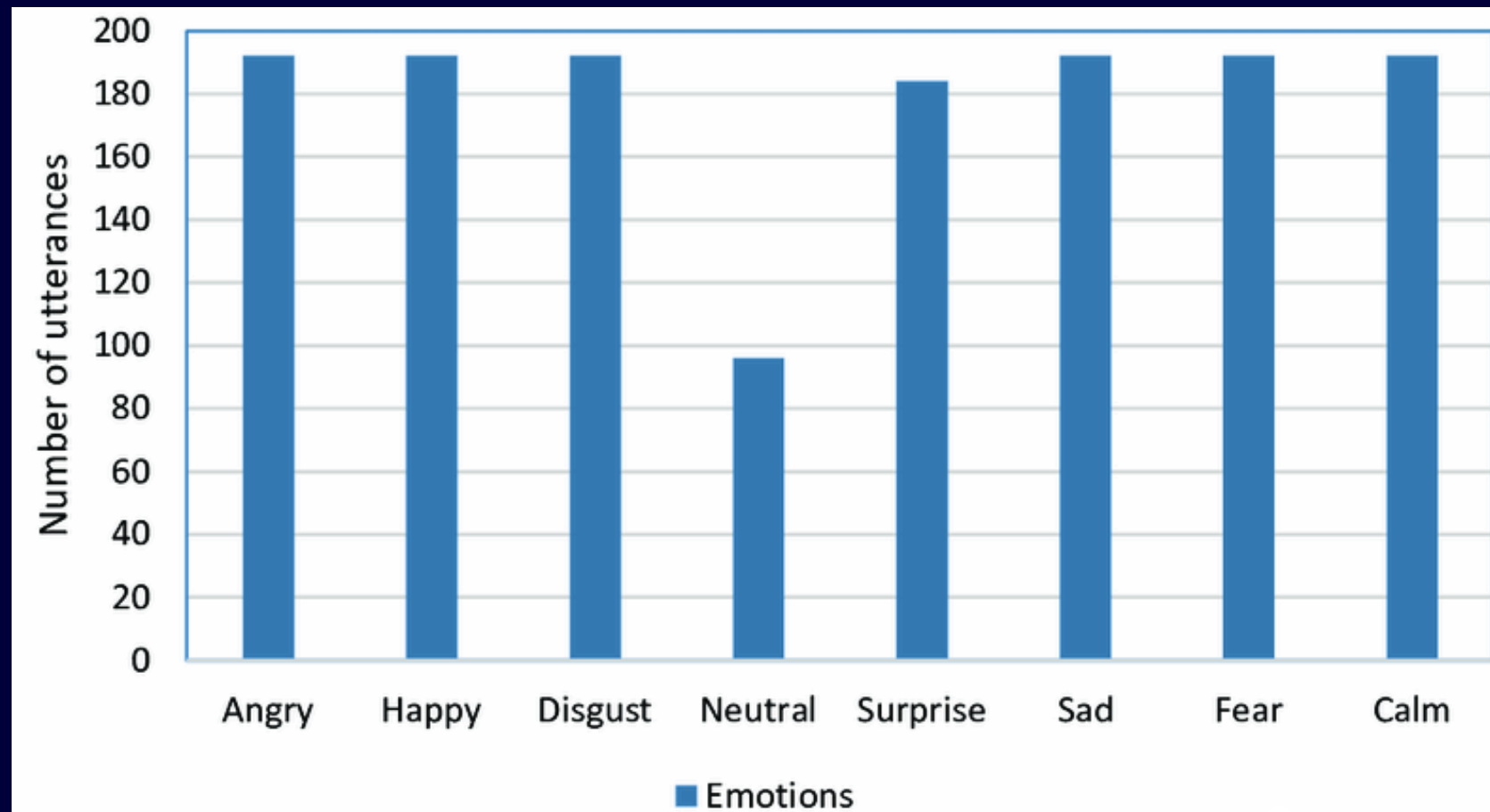
Emotion recognition technology can be used to create more immersive and interactive gaming and virtual reality experiences, as well as personalized music and media recommendations based on the user's emotional state.



# Objective :

- **To build a model to recognize emotion from speech using the librosa and sklearn libraries and the RAVDESS dataset.**
- **To present a classification model of emotion elicited by speeches based on deep neural networks ML Classification based on acoustic features such as Mel Frequency Cepstral Coefficient (MFCC). The model has been trained to classify eight different emotions (calm, happy, fearful, disgust, angry, neutral, surprised, sad).**

# Dataset

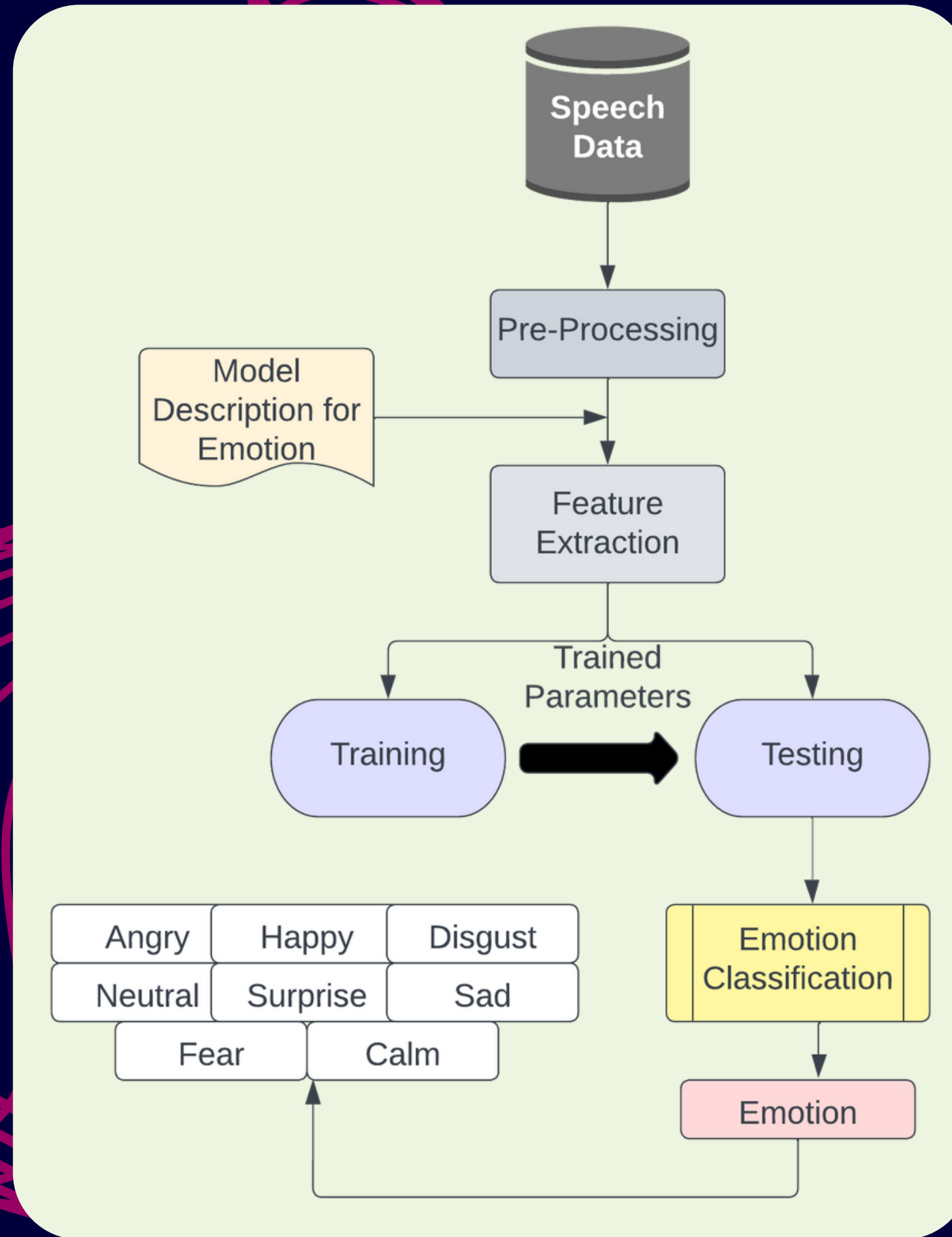


RAVDESS Data Count

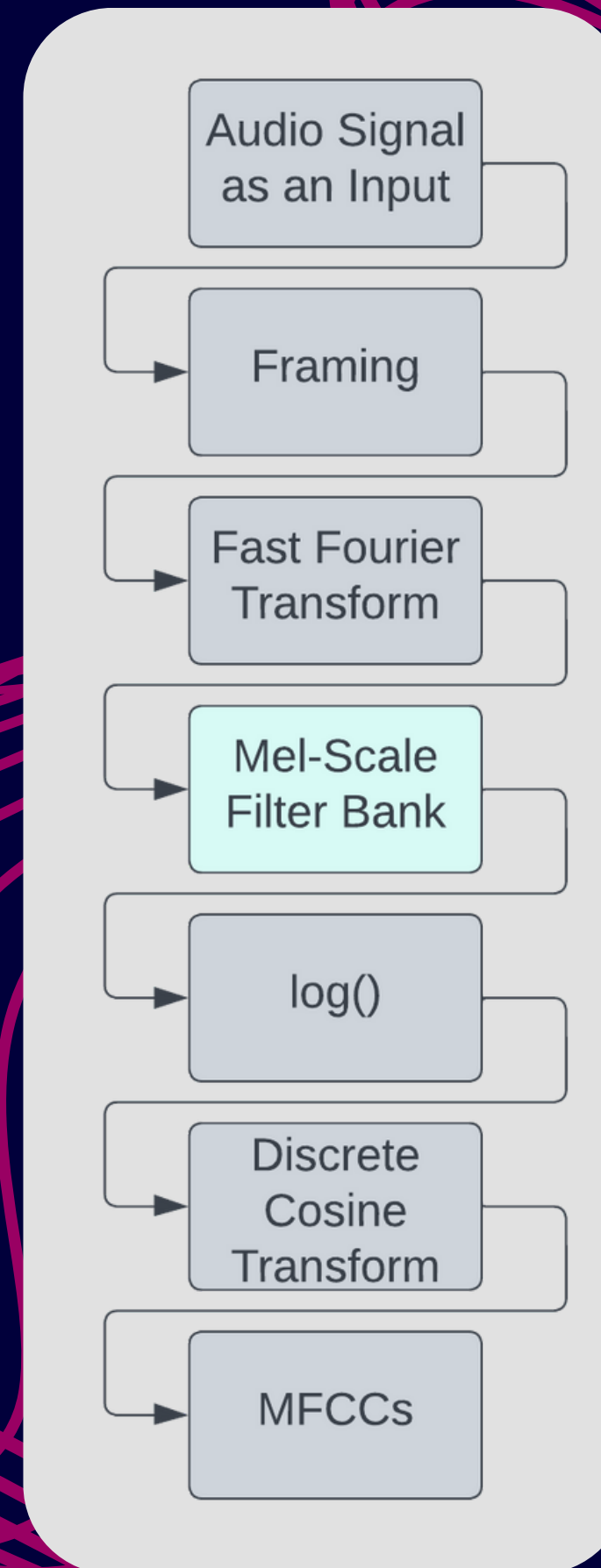
Gender	Count	Trials per Actors	Total Count
Male	12	60	720
Female	12	60	720
Total			1440

RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song) is a database of emotional speech and song. The RAVDESS dataset contains a total of 1440 files (1.15 GB) of audio, spoken and music. It is a multimodal and dynamic set of North American English facial and vocal expressions. The collection features 24 performers who deliver two lexically matched lines in a neutral North American accent. Neutral, calm, happy, sad, angry, terrified, disgusted, and astonished are the eight speech emotions. Each expression has two strength levels: strong and normal, with a neutral expression as an extra. All data is provided in three modes: audio-video, audio solo, and video only. In this investigation, only the audio files, comprising 1440 files (24 actors \* 60 trials per actor), are used.

# Work Flow :



# Feature Extraction :



**Other Useful Feature Extraction Methods are -**

- **ZCR (Zero-Crossing Rate)**
- **Mel Spectrogram**
- **Time-domain features**
- **Frequency-domain features**
- **Chroma features**
- **Wavelet-based features**
- **Spectro-temporal features**

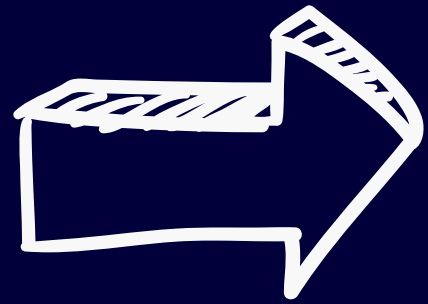


# Comparative Analysis of the Algorithms :

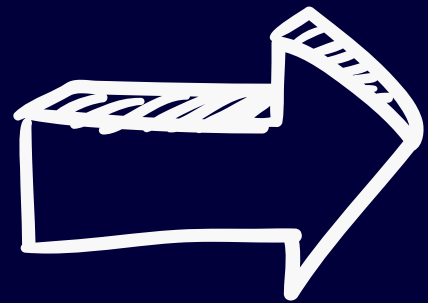


Algorithm	Accuracy
<b>Decision Tree</b>	49%
<b>SVM</b>	89%
<b>Naive Bayes</b>	46%
<b>KNN</b>	77%
<b>MLP</b>	84%
<b>CNN</b>	93%
<b>LSTM &amp; CNN</b>	82%

## CONCLUSION AND FUTURE SCOPE:

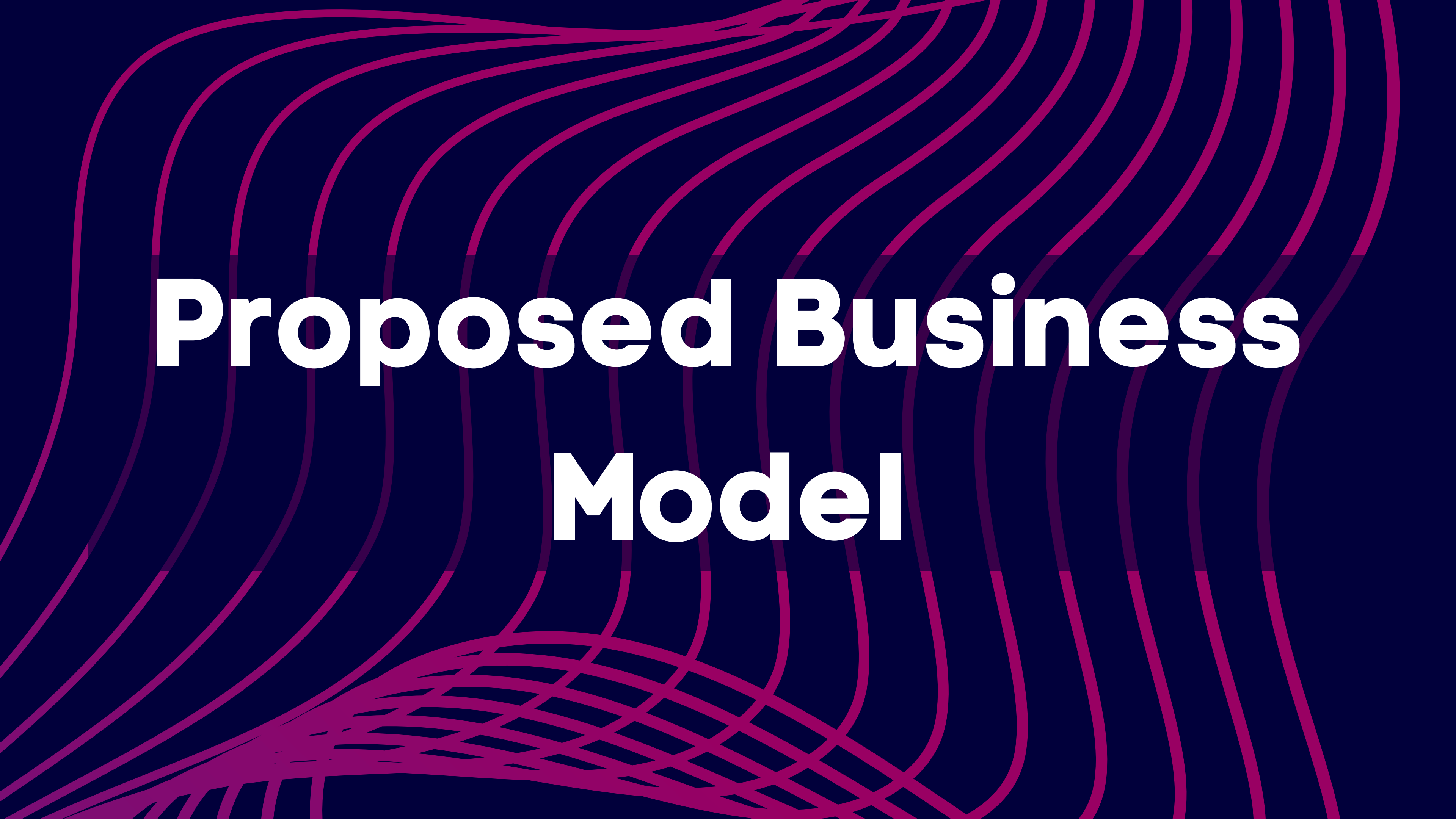


This is the first time, to the best of our knowledge, that we have used the RAVDESS dataset to apply Machine Learning approaches to our model. In the future, we plan to look more into deep learning algorithm, in such a way that our model will give us the highest possible accuracy.



The next challenge that we need to solve is Language Independency. Right now, we are expecting that our model perform good atleast on all forms of English accent.

- Human machine interaction is widely used nowadays in many applications. One of the medium of interaction is speech. The main challenges in human machine interaction is detection of emotion from speech.
- Emotion can play an important role in decision making. Emotion can be detected from different physiological signal also. If emotion can be recognized properly from speech then a system can act accordingly. Identification of emotion can be done by extracting the features or different characteristics from the speech and training needed for a large number of speech database to make the system accurate.
- An emotional speech RAVDESS dataset is selected then emotion specific features are extracted from those speeches and finally a ML classification model is used to recognize the emotions.



# **Proposed Business Model**



# SpeechURvey



Sentimental Analysis Using SER for  
Voice Based Survey

# Target Audience



Our main focus hole is on Startups, as each and every feedback of their product is very essential for them to summarize there Business in a ideal way.

At what area does this model will help them out ?

- ☒ **Sales**
- ☒ **Networking**



Select period:

from Jan 1, 2023

to Feb 28, 2023



Search...

Filter



Total Revenue Generated

₹ 60,589



Inquiry Success Rate

36.2%



No. of New Customers

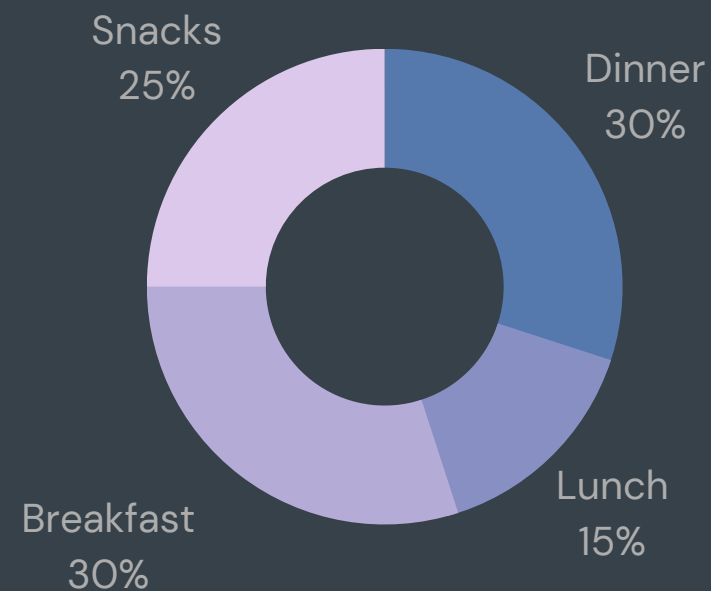
36



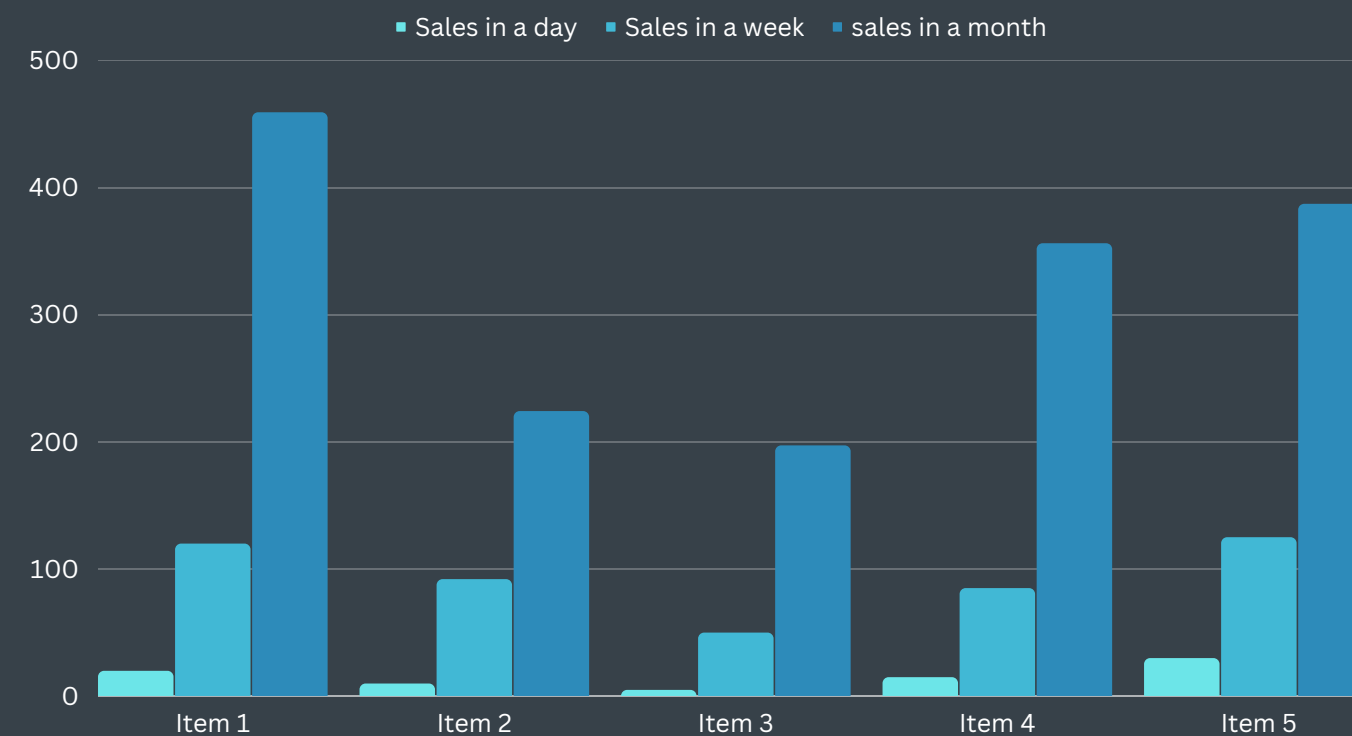
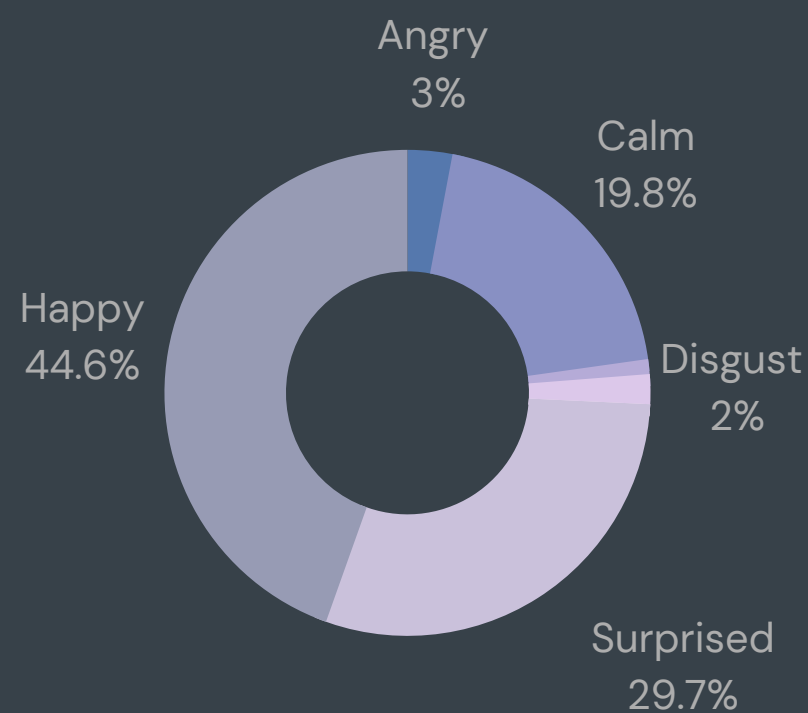
No of Customers Took Membership

29

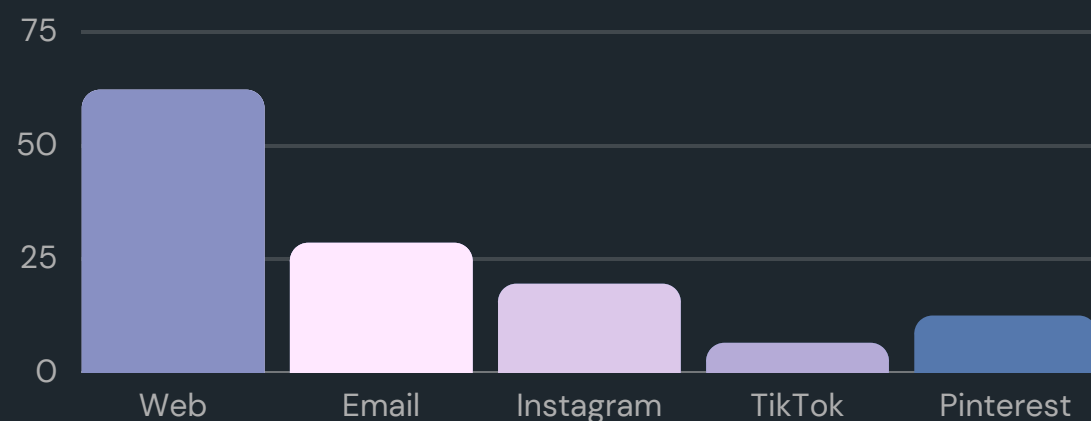
What type of meal does most of the customers prefer



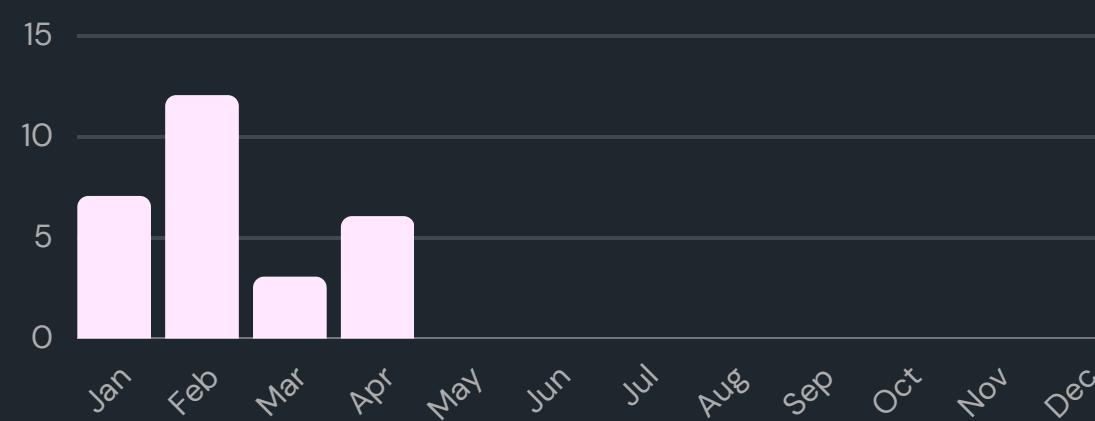
Customer Sentimental Review



Inquiry Source Breakdown



Inquiries per Month

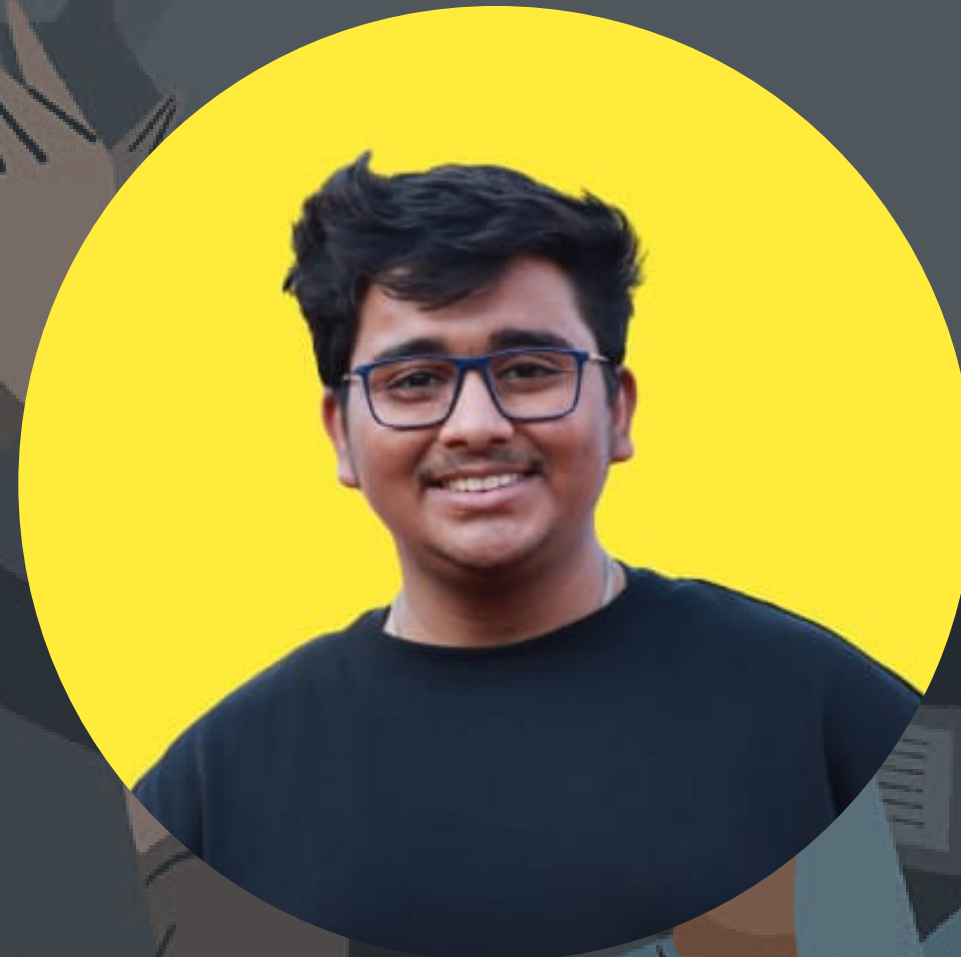




# Know Our Team



Harshit Jain



Dhrubo  
Bhattacharjee



Devansh Tiwari