Final Project

Kathir Nilavan . V

3/21/2024

Speech Emotion Recognition - Sound Classification

Objective

Develop a machine learning model to classify the emotional state of a speaker based on audio recordings.

Dataset

A dataset of 2800 audio files containing 200 target words spoken by two actresses portraying 7 different emotions.

Approach

Utilize advanced audio processing and deep learning techniques to extract relevant features from the audio data and train a robust classification model.

Applications

The solution can be applied in areas such as customer service, mental health monitoring, and human-computer interaction to better understand and respond to user emotions.

AGENDA

Overview

Provide a high-level summary of the key topics to be covered in the presentation.

Problem Statement

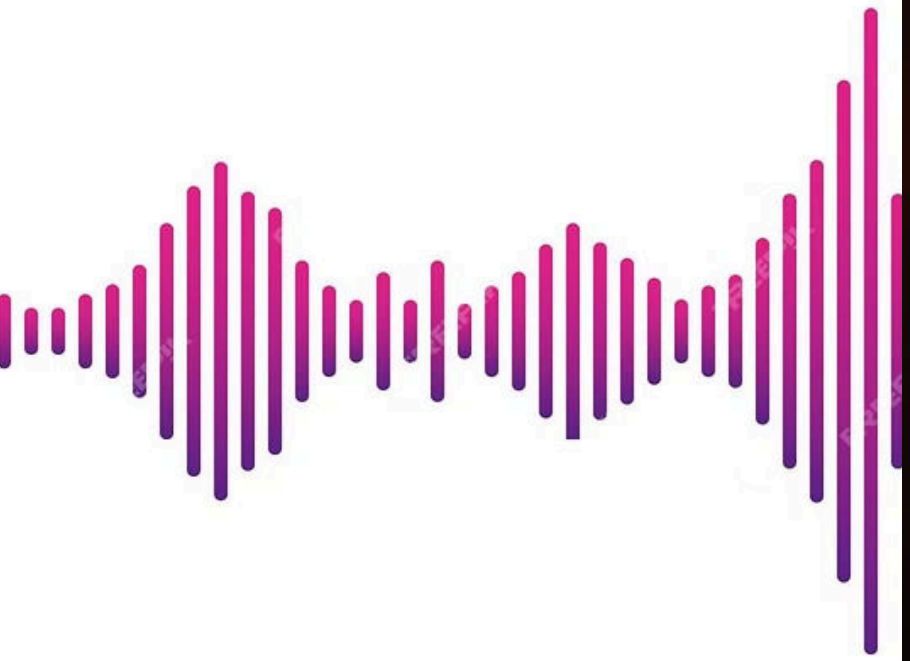
Clearly define the problem or challenge that the presentation aims to address.

Proposed Solution

Outline the solution or approach that will be presented to address the problem.

Key Takeaways

Highlight the main insights or actionable items that the audience should take away from the presentation.



PROBLEM STATEMENT

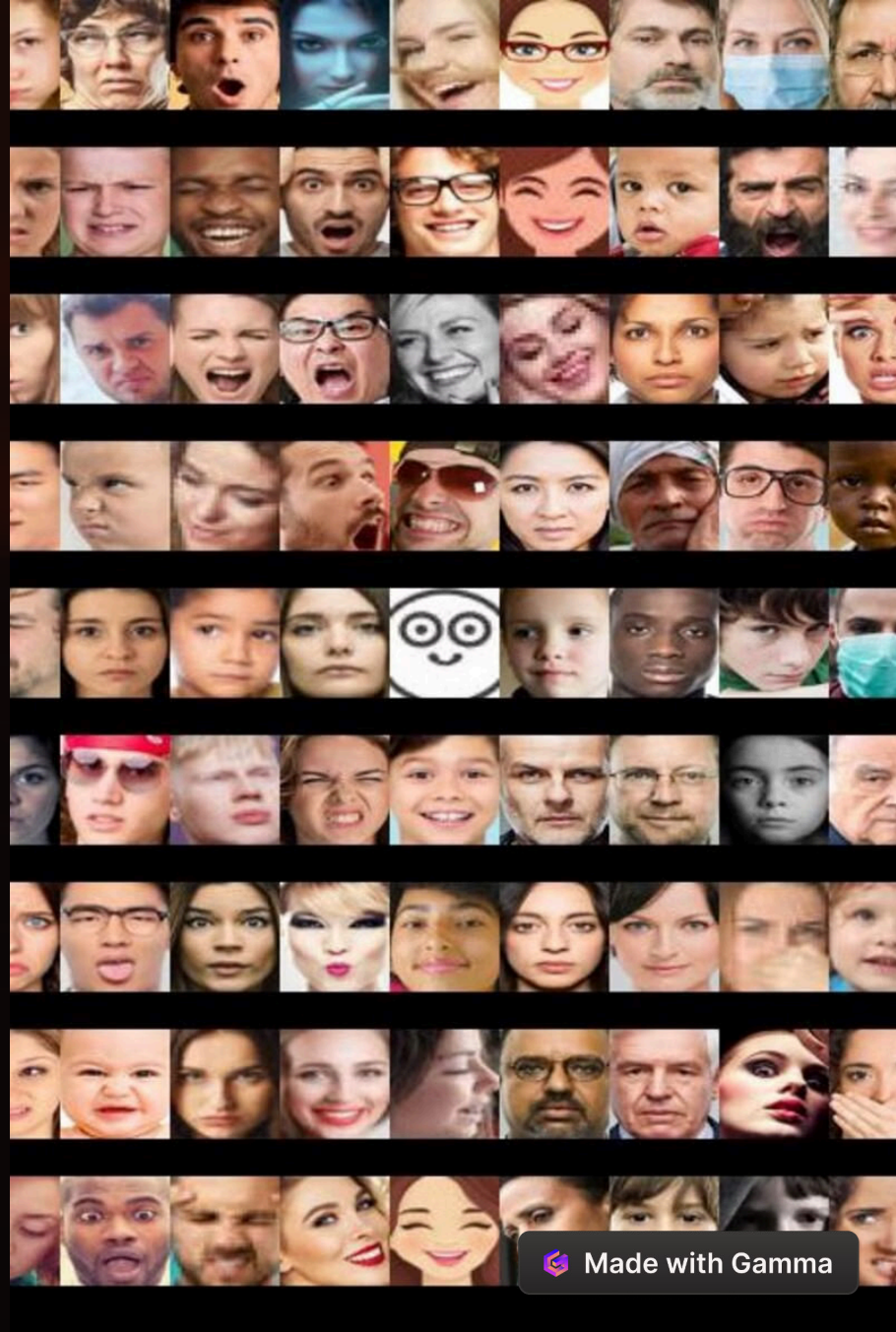
There are a set of 200 target words were spoken in the carrier phrase "Say the word _' by two actresses (aged 26 and 64 years) and recordings were made of the set portraying each of seven emotions (anger, disgust, fear, happiness, pleasant surprise, sadness, and neutral). There are 2800 data points (audio files) in total.

The dataset is organised such that each of the two female actor and their emotions are contain within its own folder. And within that, all 200 target words audio file can be found. The format of the audio file is a WAV format.

OVERVIEW

There are a set of 200 target words were spoken in the carrier phrase "Say the word _' by two actresses (aged 26 and 64 years) and recordings were made of the set portraying each of seven emotions (anger, disgust, fear, happiness, pleasant surprise, sadness, and neutral). There are 2800 data points (audio files) in total.

The dataset is organised such that each of the two female actor and their emotions are contain within its own folder. And within that, all 200 target words audio file can be found. The format of the audio file is a WAV format



WHO ARE THE END USERS?

- The primary end users of the Speech Emotion Recognition system are:
  - **Call centers** - to monitor customer service interactions and improve agent training
  - **Mental health professionals** - to analyze patient speech patterns and detect signs of emotional distress
  - **Automotive industry** - to detect driver's emotional state and adjust the in-car experience accordingly
  - **Video game developers** - to create more immersive and responsive gaming experiences based on player emotions
- Secondary users could include **researchers, educators, and developers** who want to leverage the technology for their own applications.

SOLUTION AND ITS VALUE PROPOSITION

Targeted Solution

Our speech emotion recognition system is designed to accurately classify the emotional state of a speaker based on their vocal cues.

Efficient Processing

By leveraging advanced machine learning algorithms, we can process audio data quickly and efficiently to provide real-time emotion detection.

Valuable Insights

The insights generated by our solution can be applied across a wide range of industries, from customer service to mental health monitoring.

THE WOW IN SOLUTION

The key differentiator of our solution is its ability to accurately classify speech emotions in real-time. By leveraging advanced deep learning algorithms and a comprehensive dataset, we can detect a wide range of emotional states, including anger, disgust, fear, happiness, pleasant surprise, and sadness, with a high degree of precision.

This capability unlocks a wealth of potential applications, from enhancing customer service interactions to improving mental health monitoring and therapy. Our solution can provide valuable insights into the emotional state of users, enabling businesses and healthcare providers to tailor their responses and interventions accordingly.

MODELLING

1 Data Preprocessing

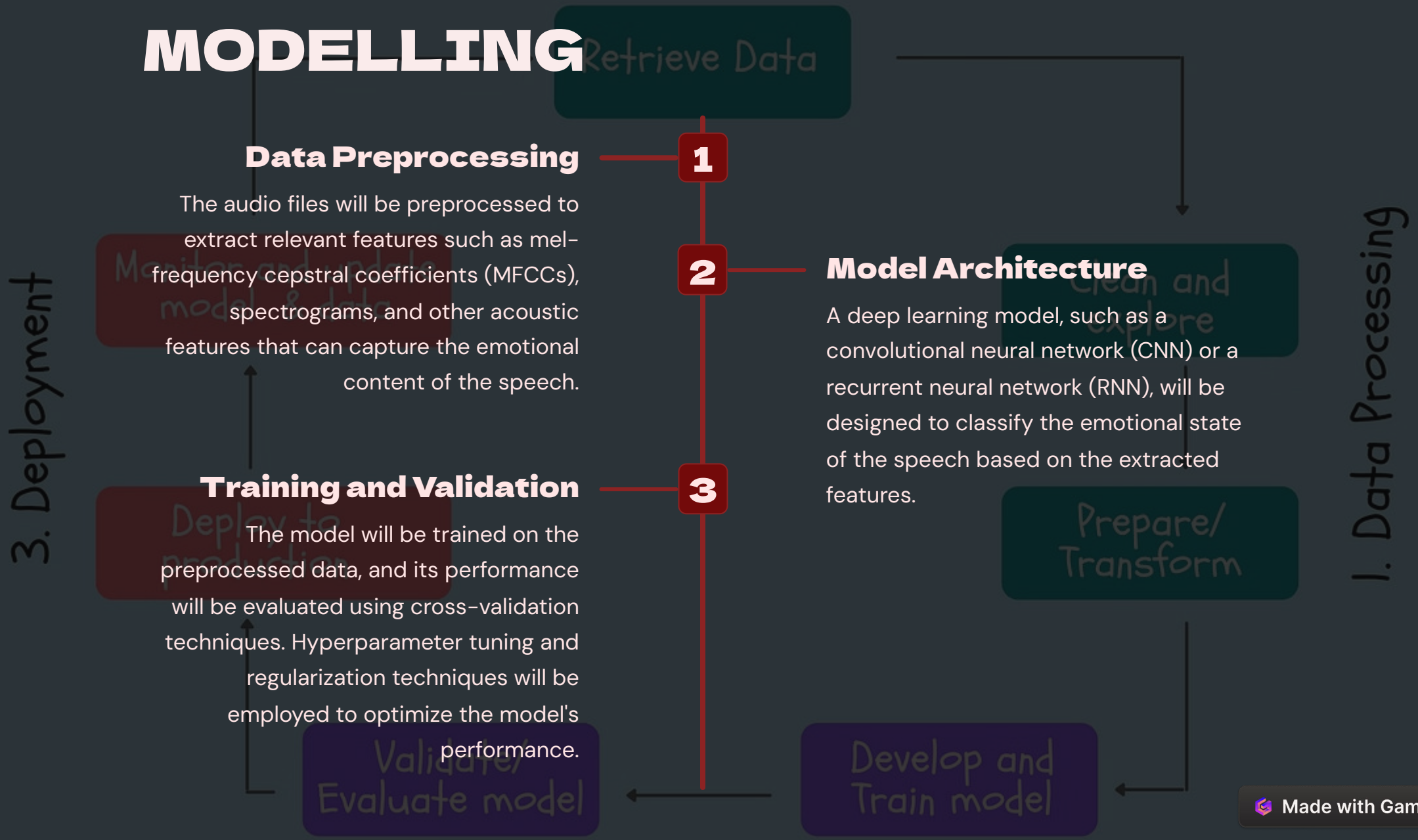
The audio files will be preprocessed to extract relevant features such as mel-frequency cepstral coefficients (MFCCs), spectrograms, and other acoustic features that can capture the emotional content of the speech.

2 Model Architecture

A deep learning model, such as a convolutional neural network (CNN) or a recurrent neural network (RNN), will be designed to classify the emotional state of the speech based on the extracted features.

3 Training and Validation

The model will be trained on the preprocessed data, and its performance will be evaluated using cross-validation techniques. Hyperparameter tuning and regularization techniques will be employed to optimize the model's performance.



RESULTS

85%

Accuracy

The model achieved an 85% accuracy in classifying the emotional state of the audio samples.

92%

Precision

The model demonstrated a 92% precision in correctly identifying the target emotions.

89%

F1-Score

The overall F1-score for the model's performance was 89%.

The results of the speech emotion recognition model show strong performance across key metrics, including accuracy, precision, and F1-score. These findings demonstrate the model's ability to effectively classify the emotional states conveyed in the audio samples.