Objective:

Developing model which can classify voice samples in authentic vs synthetic

Key Instructions:

1. Data Exploration and Analysis:

- Conduct thorough exploratory data analysis (EDA) to understand the distribution of samples across authentic and synthetic categories.

To tackle class imbalance we can utilise methords like oversampling, under sampling, based on amount of data available

2. Feature Engineering:

- Extract relevant acoustic features from the audio recordings. Pitch, base, amplitude, pauses in speech, mean and std deviation of each voice.

- Normalize or standardize the features to ensure consistency and facilitate model convergence.

3. Data Preprocessing:

- Segment the audio recordings into fixed-length segments to ensure uniform input size for the model. Truncation and padding both can be utilised.

- Handle diverse acoustic conditions by applying appropriate preprocessing techniques such as noise reduction, equalization, or augmentation.

4. Speaker Variability Handling:

Each person has a unique way of speaking. To make sure a computer program can understand

everyone equally well, we can use techniques to recognize and adjust for these differences. This helps the program to work better for all kinds of speakers.

5. Model Validation and Evaluation:

- Split the dataset into training, validation, and test sets using an appropriate ratio.

- Employ robust validation strategies such as cross-validation or stratified sampling to ensure reliable model performance estimation.

- Utilize relevant evaluation metrics such as accuracy, precision, recall, and F1-score to assess model performance.

6. Documentation and Version Control:

- Maintain detailed documentation of all preprocessing steps, feature extraction methods, and model configurations.

- Git will be used to store the track changes and to store versions

Assumptions:

1. The database contains sufficient samples of both authentic and synthetic voices across various acoustic conditions.

2. The provided recordings are labelled accurately, with clear distinctions between authentic and synthetic voices.

3. The model deployment environment and target platform requirements have been considered in the preprocessing and model development stages.