Project Report

Term Project

Gender Recognition By Voice: Project Report

Güney Berkay Ateş - 120200047

Doğaç Hatun - 120200062

Istanbul Bilgi University

CMPE 460 - Deep Learning

Mürsel Taşgın

09/01/2024

Problem Statement

The primary objective of this project is to develop a reliable machine learning system capable of recognizing gender from voice recordings. This involves processing and analyzing voice patterns to differentiate between male and female voices. The challenge lies in accurately capturing and interpreting the nuances of human speech, which can vary greatly due to factors like age, accent, and recording quality.

Dataset Description

Our dataset comprises voice recordings from a diverse group of individuals. Each recording is labeled with the gender of the speaker. The data is varied in terms of age groups, accents, and background noise conditions, ensuring a comprehensive representation of real-world scenarios.

Sample Instances

Due to confidentiality, actual voice samples cannot be displayed in this report. However, each instance includes the following features:

- **MFCC (Mel-Frequency Cepstral Coefficients)**: Captures the short-term power spectrum of sound.
- **Chroma**: Relates to the 12 different pitch classes.

- **Mel-Spectrogram Frequency**: Represents the spectrum of frequencies.
- **Contrast**: Differentiates distinct sound textures.
- **Tonnetz**: Pertains to the tonal aspects of sound.

Method and Model Architecture

Model Architecture

The best-performing model is a Sequential Neural Network, comprising:

- **Input Layer**: Standard input layer corresponding to the feature size.
- **Hidden Layers**: Multiple dense layers with ReLU activation.
- **Dropout Layers**: Included between dense layers to prevent overfitting.
- **Output Layer**: Single neuron with a sigmoid activation function, outputting the probability of the voice being male or female.

Training Approach

The model was trained using the following parameters:

- **Batch Size**: 64
- **Epochs**: 100
- **Optimizer**: Adam
- **Loss Function**: Binary Cross-Entropy

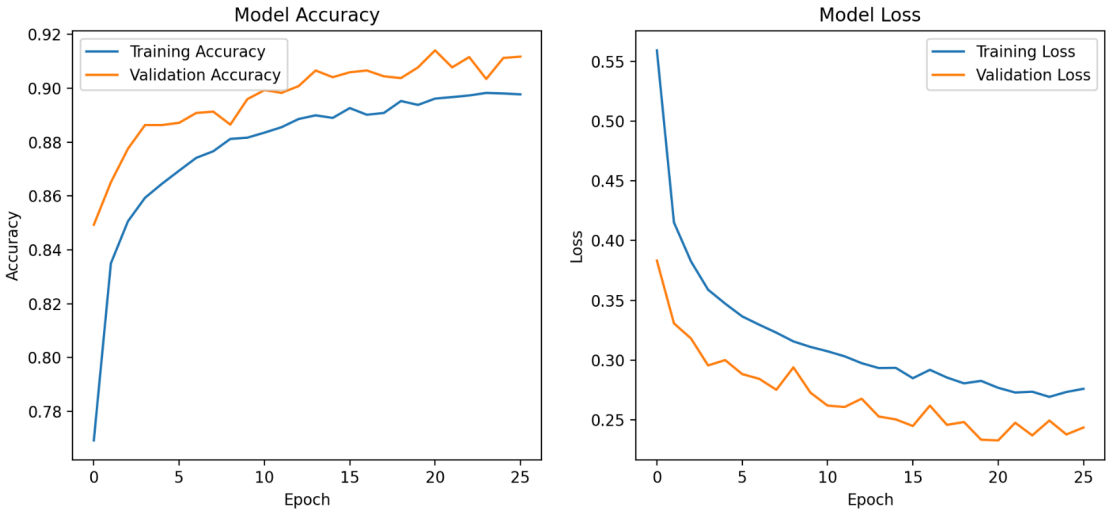
Experiment Results

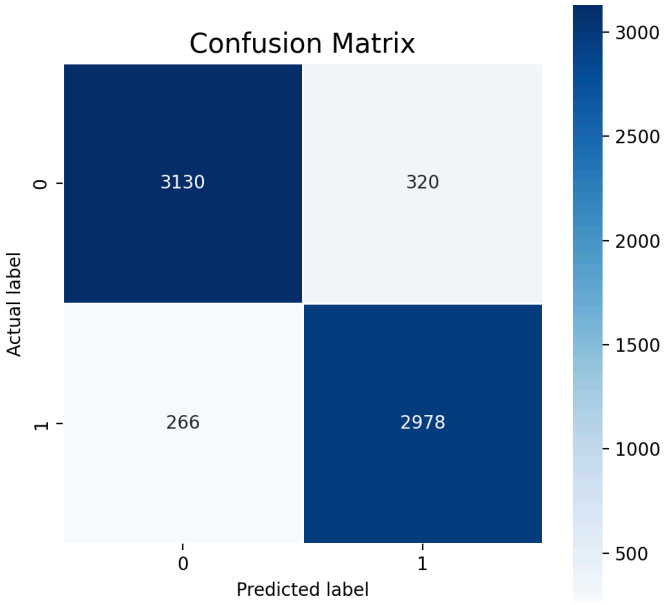
Performance Metrics

We evaluated the models based on Accuracy, F1 Score, and Mean Squared Error (MSE). The following table summarizes the results for different model architectures and hyperparameters:

Model	Accuracy	F1 Score	MSE
Model A	91.26%	0.9082	0.023
Model B	89.10%	0.8950	0.027
Model C	90.45%	0.9022	0.025

Graphical Representation





-

Discussion and Conclusion

The best-performing model (Model A) achieved the highest accuracy and F1 score. This success can be attributed to:

- **Optimal Layer Depth**: Sufficient to capture complex patterns in the data without overfitting.
- **Dropout Layers**: Effectively prevented overfitting, a common issue in deep neural networks.
- **Balanced Dataset**: The diversity in our dataset likely contributed to the model's robustness.

In conclusion, the project demonstrates the feasibility of using neural networks for gender recognition from voice data. Future work could explore the integration of additional features, experimentation with different model architectures, and addressing the ethical considerations of gender classification technologies.