Machine Learning Project Presentation

Title:Gender and Emotion Classification by Voice

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Group Members:

Abdullah Shujat

Anant Kaushal

Anikait Agrawal

Ansh Varshney

Motivation

Voice: used for human communication.

- Audio categorization

- Speech emotion recognition

- Mental Health and Well-being Monitoring

Emotion: used for expressing feelings.

- Enhancing Human-Computer Interaction

- Education and Training

- Automotive Industry

- Voice based Age, Accent and Gender Recognition [1]
  - Utilizing five hidden layers for gender (91% accuracy) and grid search pipeline for age (59% accuracy).

  - Mel-Frequency Cepstral Coefficients (MFCCs) and Spectrogram analysis are the major features extracted for model training.

  - Potential applications could be biometric identification and NLP.

  - Challenges in enhancing age detection accuracy and handling various accents.

Literature Review (2/2)

- Emotion Recognition using pitch , tone and rate.[2]
  - The paper discusses extracting features like pitch, tone, and rate for emotion detection.

  - It reviews several classification schemes, including hidden Markov models (HMMs) and neural networks.

  - Other classification methods mentioned include support vector machines (SVMs) and decision trees.

  - The study also discusses publicly available databases used for training and testing emotion recognition systems.

Dataset Description – Voice Gender Dataset

- Dataset consists of 3168 supervised learning samples where each voice recording is labeled with either male or female gender.

- Acoustic properties that help in distinguishing male and female voice are:
  - Mean frequency (Hz)
  - Standard Deviation of frequency
  - Median Frequency
 - Skewness of Frequency Distribution etc.

- [Kaggle Link]

Dataset Description – Emotion Gender Dataset

- Consists of 1440 supervised audio samples from 24 actors (12 male + 12 female), capturing a diverse range of emotions.

- Encompasses 8 distinct emotions:

  neutral, calm, happy, sad, angry, fearful, disgust, and surprised.
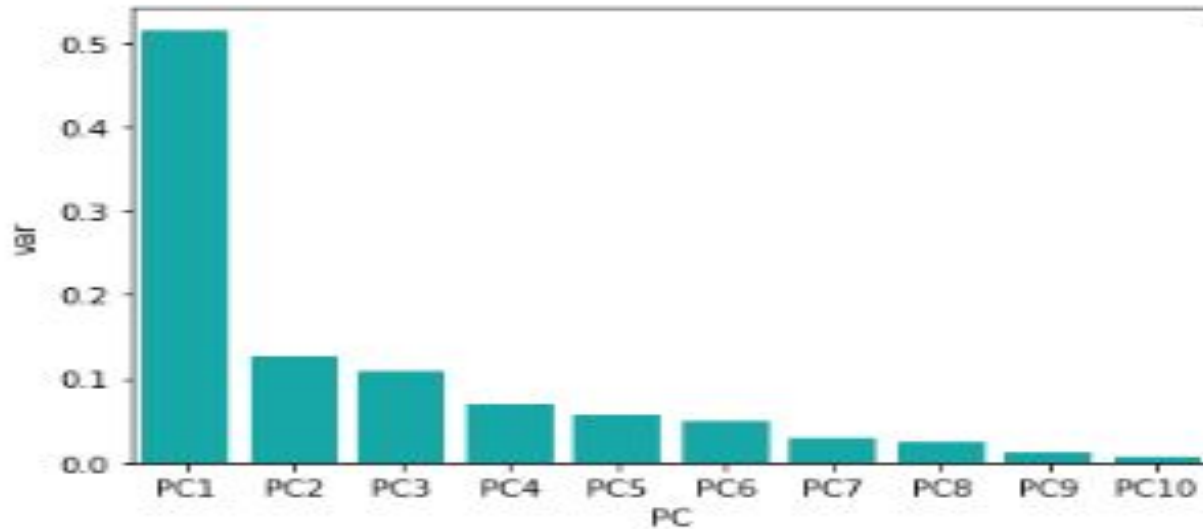
- [Kaggle Link]

Visualizations
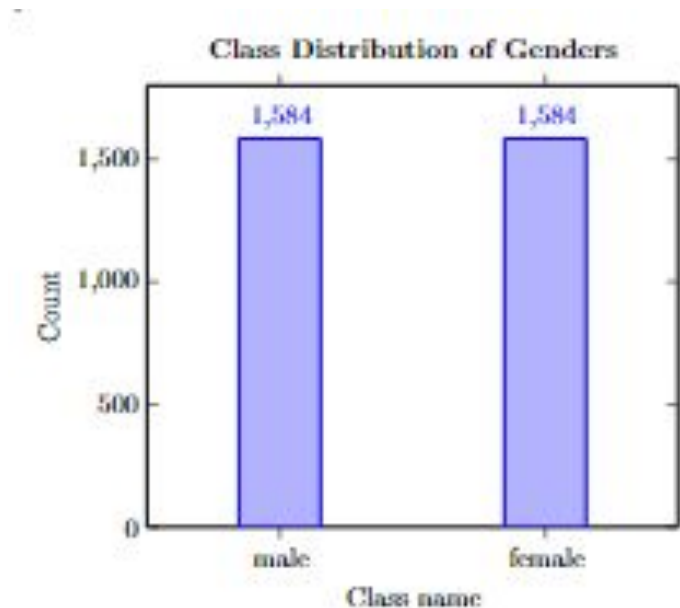
Class Distribution for output label
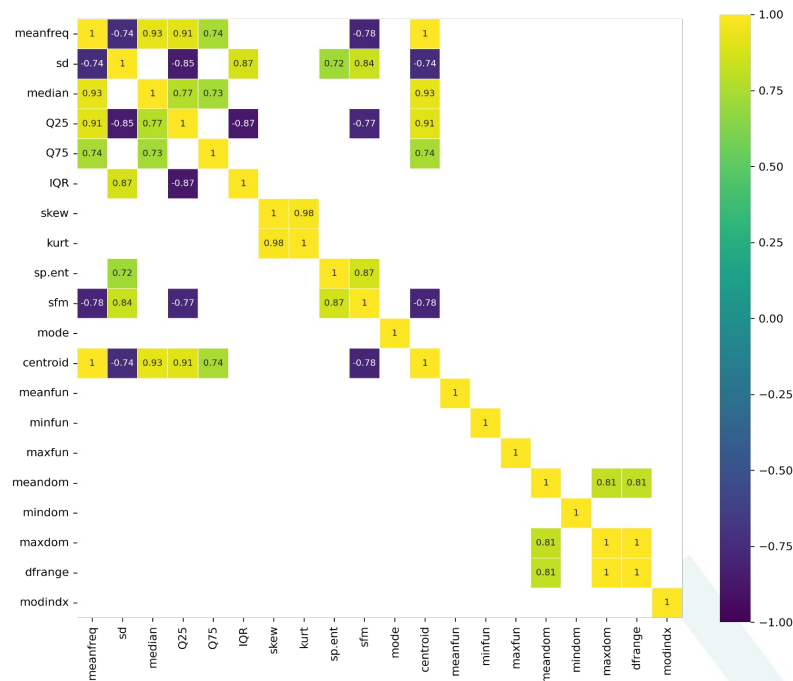
Heatmap for highly correlated columns

Visualizations

Feature Importance Graph using PCA

Methodology

- Preprocessed and cleaned data by normalizing, null checking, duplicate handling etc.

- Classified the speaker's gender using selected voice attributes.

- The speaker's voice was categorized as either 'male' or 'female.'

- Three different models/methodologies were used for classification:

  - Logistic Regression

  - Naive Bayes:

    - Gaussian

    - Bernoulli

  - Support Vector Machine

Results
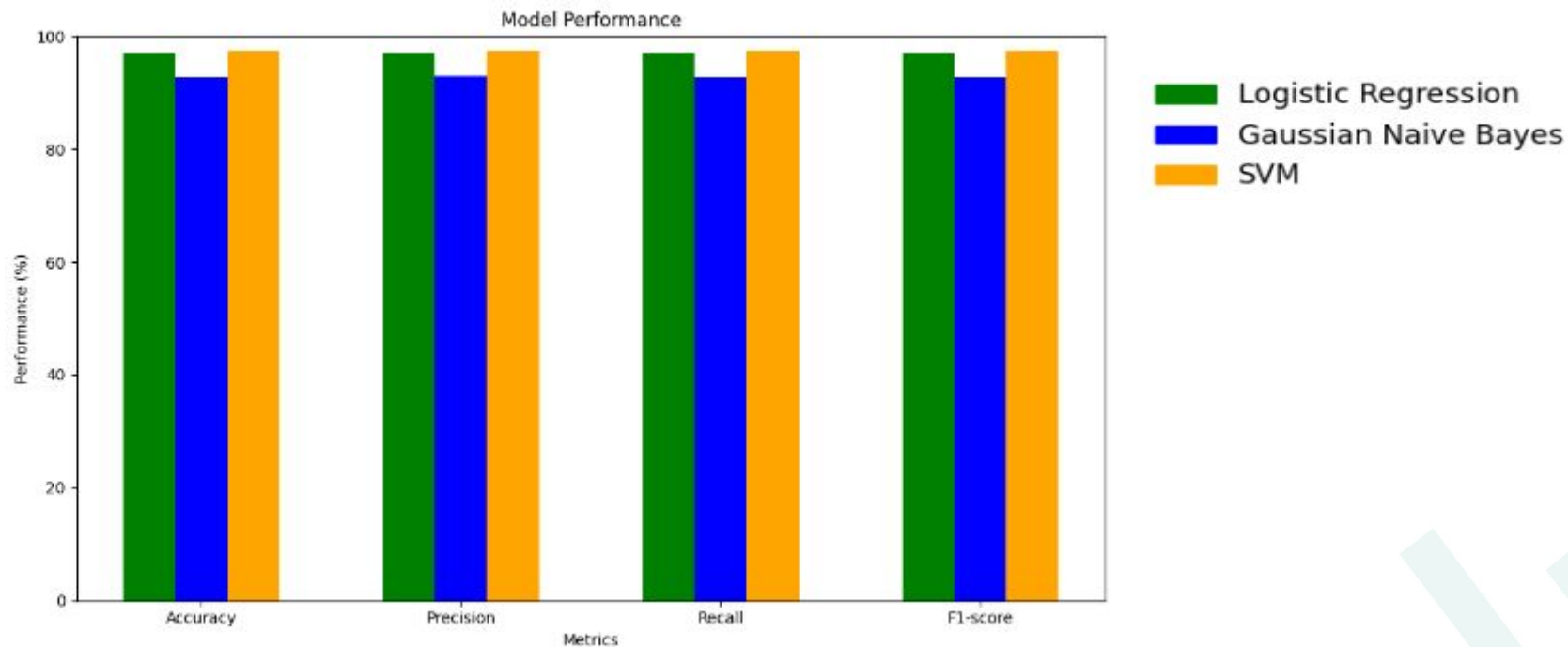
Conclusion

● Logistic Regression provided good results for each of the metrics chosen to judge the model.

● Gaussian naive bayes also reported good accuracy and other performance metrics, therefore, it implies that:

- The Gaussian Naive Bayes model fits perfectly to the data.

- The voice dataset is produced according to the Gaussian distribution for male and female categories.

● In SVM model, linear kernel was observed to be the best one.

Timeline

Timeline followed:

● Week 1-2: Data Collection , Pre-processing and Data Visualization for Gender Recognition.

● Week 3-4: Logistic Regression, Naïve Bayes, Support Vector Machine.

● Week 5: EDA, Feature Analysis, Heatmaps, PCA, TSNE (Analysis of Models).

● Week 6: Feature Selection and Observing the performance of each model .

● Week 7(after midsem):Data Collection for Emotion Recognition.

● Week 8:Report and PPT.

Contributions

Abdullah Shujat: Pre-processing and Data Visualization, Logistic Regression, Naive Bayes, Report

Anant Kaushal: Data Collection and Visualization, SVM , Feature Extraction , PPT

Anikait Agrawal: Data Collection, Logistic Regression, Naive Bayes , Report

Ansh Varshney: Pre-processing and Data Visualization , Feature Extraction , SVM , PPT

THANK YOU! 😎💪