Machine Learning for Understanding Emotions in Speech

Team Members (Group-8)
Sahithi Nara
Revanth Kumar Bondada
Catherine Raj
Pranathi Sai Maddigunta

Table of Contents

Q&A
Thank You.

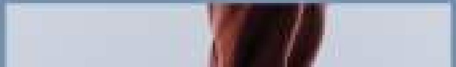
Conclusion

How Related to Sustainability

Table of Contents

Aim

Develop a highly accurate model capable of recognizing emotions from speech data, enhancing user interaction and responsiveness in applications.



Motivation

- Real-World Applications, Practical applications in areas like criminal justice (detecting truth in speech) and healthcare (assessing patient emotions remotely)

- Tackle existing biases in emotion recognition systems through balanced datasets and equitable model training approaches.

How Related to Sustainability

- Enhancing Communication

- Energy Efficiency

- Mental Health and Well-being

- Reducing Physical Resources

Methodology Overview

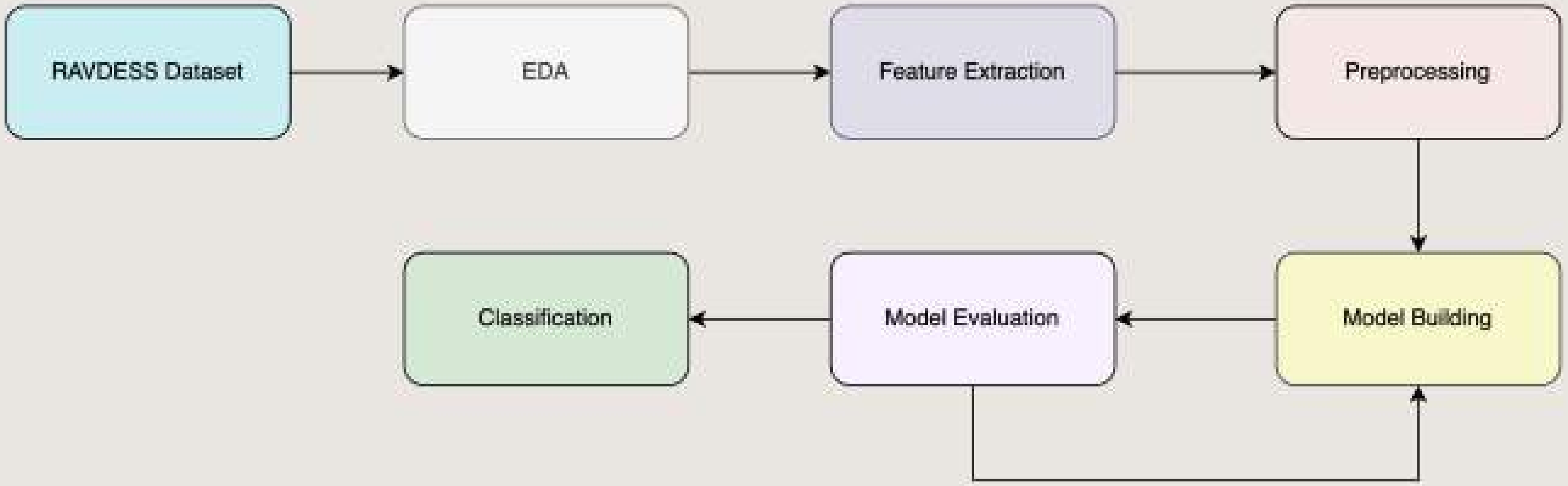
Start with data acquisition from RAVDESS dataset on Kaggle.

Conduct data labeling and balancing for fairness in model training.

Implement feature extraction techniques like MFCC for detailed audio analysis.

Employ multiple ML models for robust evaluation.

Architecture



Dataset Introduction

- Use RAVDESS Emotional Speech Audio dataset, containing various emotions.

- Initially unlabelled, emotions categorized into eight types: happy, calm, etc.

- Aimed to balance the dataset to ensure uniform data for all emotions.

- Applied normalization techniques for consistent data input to models.

Feature Extraction Techniques

- Utilized histogram plots to visualize data distribution.

- Employed spectrogram plots to observe decibel levels across audio files.

- Feature extraction focused on capturing nuances in speech.

- StandardScaler used for data normalization before model training.

Machine Learning Models

K-NN

68%

Machine Learning Models

SVM

53%

Machine Learning Models

Logistic Regression

55%

Machine Learning Models

Random Forest

71%

Model Training and Selection

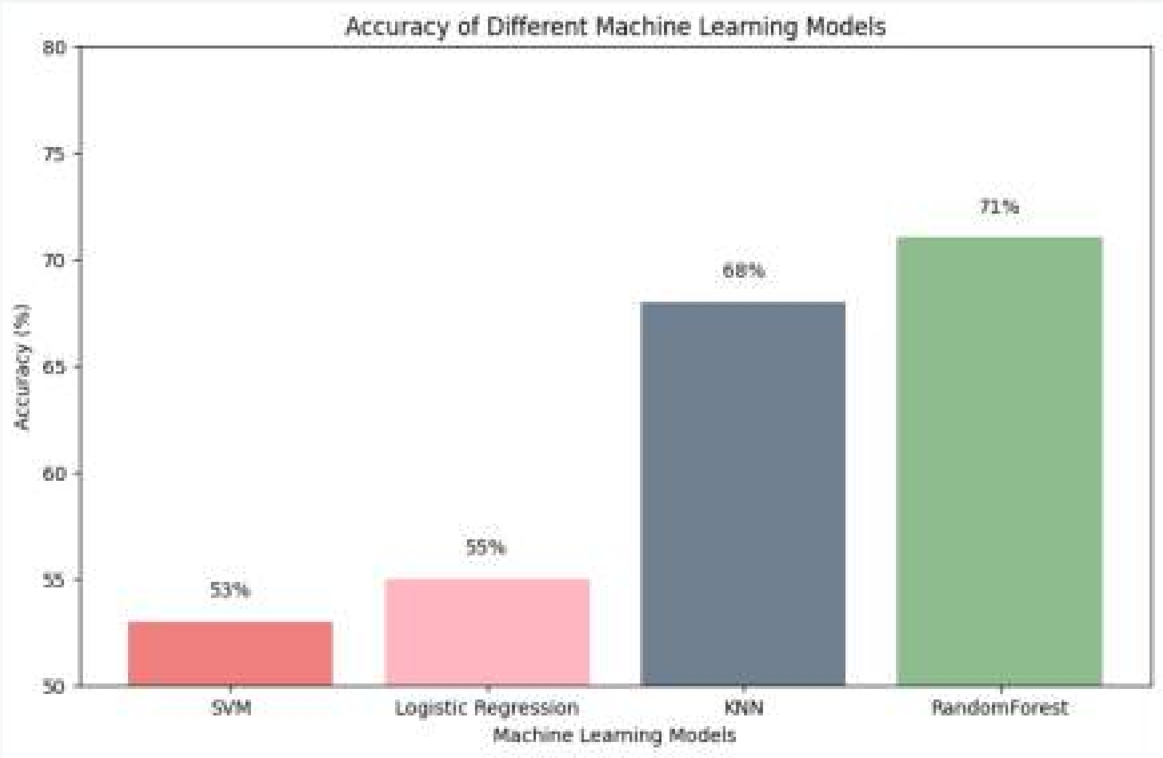
- Models trained on balanced and normalized data.

- Accuracy assessed to choose the best performing model.

- Emphasis on models that could handle diverse emotional expressions.

- Random Forest selected for final prediction model.

Results Overview



Discussion on Model Performance

- Analyzed each model's strengths and weaknesses in emotion recognition.
- Random Forest provided the best balance of accuracy and reliability.
- Discussed the impact of feature extraction on model success.
- Evaluated the role of data balancing in achieving fair model outputs.

Innovation

- Develop capabilities for real-time emotion recognition to improve interactive systems by directly utilizing the audio files, enabling dynamic responses in customer service and therapeutic settings.

- Implement sophisticated feature extraction methods, such as MFCC, SMOTE, CHROMA and advanced normalization to capture detailed emotional cues in speech and ensure data equity.

Comparing Solutions

- Most of the works were classified the emotion based on the RAVDESS CSV Data. We worked on Audio Data

- Models were trained with Imbalanced data. After Feature Extraction we balanced the data.

Technical Difficulties Encountered

- Challenges in data labeling and ensuring dataset balance.

- Difficulties in parameter tuning for optimal model performance.

- Overfitting issues encountered during model training phases.

- Addressed compatibility issues with different machine learning frameworks.

Lessons Learned

- Importance of comprehensive data preparation for ML projects.

- Gained insights into the effectiveness of various ML models for speech analysis.

- Recognized the need for continuous model updates based on new data.

- How to choose the best parameters for hyper-tune.

Best Practices Adopted

- Applied rigorous data cleaning and normalization procedures.

- Adopted systematic approaches to model selection and training.

- Emphasized the importance of diverse data to avoid bias in predictions.

Conclusion

- Successfully developed a speech emotion recognition model with 71% accuracy.

- Demonstrated the feasibility of using ML to enhance emotional recognition in speech.

- Proposed system holds potential for applications in mental health, HCI and Customer Support Service.

Q&A

Thank You