**Yarmouk University**



**Collage of Information Technology**

**and**

**Computer Science**

**Department of Information Systems**

**Deep Learning – DA450**

**[EchoGen]**

**Students name:**

Ahmed AL-Jarrah , Sewar Qudah , Shaima AL-Hussain , Sura Keelani ​

**Supervisor: Dr. Zeyad Hailat**

**FIRST TERM 2024**

**Table of content**

|  |  |  |
| --- | --- | --- |
|  | Introduction | 3 |
|  | Dataset Description | 3 |
|  | Methods | 4 |
|  | Preprocessing Steps | 4 |
|  | Results | 4 |
|  | Discussion | 7 |
|  | Conclusion and Lessons Learned | 8 |
|  | References | 8 |

**List of Tables**

|  |  |  |
| --- | --- | --- |
| Table 1 | Method Name | 5 |

**List of Figures**

|  |  |  |
| --- | --- | --- |
| Figure 1 | Model Loss | 6 |
| Figure 2 | Model Accuracy | 6 |

1. **Introduction:**

Have you ever found it difficult to distinguish between male and female voices? Certainly not, but what about computers? It is known that computers can't recognize gender through voice directly, but with the advent of artificial intelligence, the impossible has become possible. Once computers gain this ability, it will enhance several fields, including marketing, commerce, and security.

**An example from the marketing side:** A telecommunications company employs an AI-powered voice recognition system in its customer service call center. When a customer calls, the system analyzes the voice to predict the speaker's gender. This information is used to:

* **Tailor Greetings:** The AI selects personalized scripts or greetings that resonate better with the predicted gender, creating a more welcoming customer experience. For example, using tone, language, or offers that appeal more to the caller's preferences.
* **Customized Recommendations:**
  + For male customers: Suggest plans emphasizing sports-related content, business tools, or tech-related offers that relate to men.
  + For female customers: Highlight plans that focus on family-oriented packages, lifestyle benefits, makeups, and so on.
* **Market Segmentation:** Analyze the gender distribution of callers to adjust marketing campaigns, such as targeting ads for specific genders in regions where the data suggests a higher presence of a particular demographic.

**So, where is this technique, and how does it work?** We’re eager to use it in our application. All you need we will discuss in our project **"Gender Detection Using Voice."**

**2.Dataset Description**

* **Source of the Dataset**

**Kaggle :** <https://www.kaggle.com/datasets/primaryobjects/voicegender/data>

* **Description of the Dataset**

1. **Features**:

The dataset consists of 21 columns. Key features include acoustic properties such as:

* + meanfreq, sd, median: Statistical measures of frequency.
  + skew, kurt: Measures of distribution shape.
  + sp.ent, sfm: Spectral entropy and spectral flatness.
  + meandom, mindom, maxdom, dfrange: Measures related to the dominant frequency.
* The last column, label, indicates the gender (e.g., "male" or "female").

1. **Size:**

* **Rows**: 3,168
* **Columns**: 21

**3. Preprocessing Requirements:**

* No missing values were detected in the dataset.

Features may require normalization or standardization depending on the modeling approach

**3. Methods:**

The model is a sequential neural network designed for binary classification, consisting of dense layers with 128, 64, and 32 neurons using ReLU activation, and a final dense layer with Sigmoid activation for output. Dropout layers with a rate of 0.2 are applied after each dense layer to prevent overfitting by deactivating 20% of the neurons during training while retaining 80% for effective learning. The initial layer with 128 neurons handles the complexity of high-dimensional audio features, capturing intricate patterns, while the gradual reduction to 64 and 32 neurons simplifies extracted features and prevents overfitting. ReLU activation ensures efficient learning of complex patterns by introducing non-linearity and avoiding the vanishing gradient problem, while the Sigmoid activation in the final layer outputs probabilities suitable for binary classification. The Binary Crossentropy loss function is used to measure the difference between predicted probabilities and actual labels, ensuring effective training. This architecture provides a balanced and efficient approach for tasks like gender identification.

**4. Preprocessing:**

In this project, we did not perform complex pre-processing on the audio data; instead, we focused solely on extracting key audio features. The audio was converted into usable data by extracting important features.These features were then transformed into structured data, making it easier to process and analyze using DL techniques.

**5. Results:**

**Explanation of Results:**

-**Precision**:

* For **Female (Class 0)**, the model achieved a precision of **0.95**, meaning that 95% of the instances predicted as female were actually female.
* For **Male (Class 1)**, the precision is **0.96**, meaning that 96% of the instances predicted as male were actually male.

- **Recall**:

* For **Female (Class 0)**, the recall is **0.96**, meaning that the model correctly identified 96% of the actual female voices.
* For **Male (Class 1)**, the recall is also **0.96**, meaning that the model correctly identified 96% of the actual male voices.

- **F1-Score**:

* The **F1-Score** is the harmonic mean of precision and recall, and it provides a balanced measure. The model achieved an **F1-score of 0.96** for both classes, indicating that the model performs well in balancing both precision and recall.

- **Support**:

* **Support** refers to the number of instances in each class. There are **310 female instances** (class 0) and **324 male instances** (class 1).

- **Accuracy**:

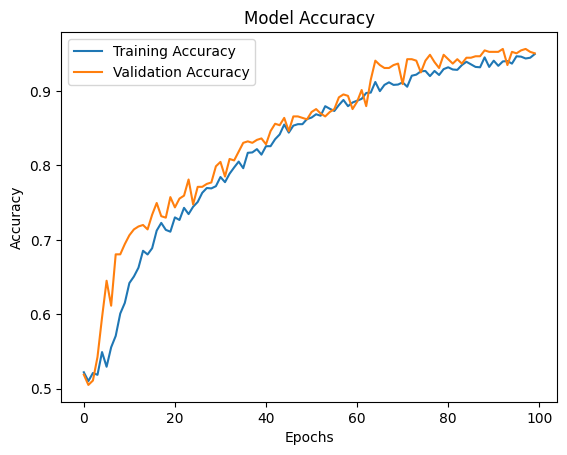
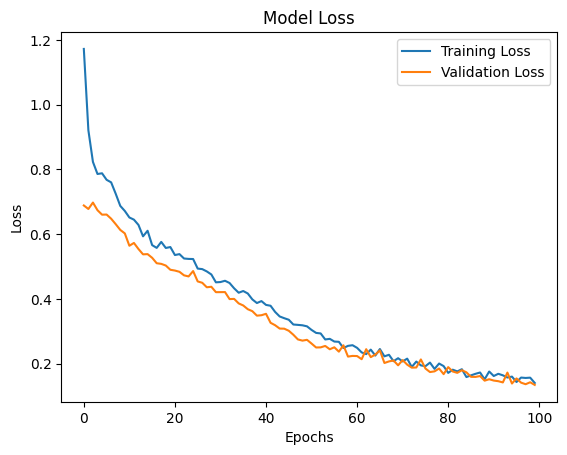
* The overall **accuracy** of the model is **0.96**, meaning that 96% of all instances were correctly classified.

- **loss**:

* The **loss value of 0.1180** shows that the model is performing well, with its predictions being quite close to the actual labels. It reflects a strong learning process, with minimal error in distinguishing between male and female voices.

|  |  |  |
| --- | --- | --- |
| **Method Name** | **Training accuracy** | **Testing accuracy** |
| A Baseline Algorithm | 50% | 50% |
| The frequency-based baseline model | 72% | 71% |
| The logistic regression model | 72% | 71% |
| The CART model | 81% | 78% |
| Random forest model | 100% | 87% |
| The boosted tree model | 91% | 84% |
| SVM | 96% | 85% |
| XGBoost algorithm | 100% | 87% |
| Random Forest or XGBoost model | - | 87% |
| The stacked model | - | 89% |
| DNN | 95.13% | 95.74% |

Glory for Deep Learning



**Discussion:**

1. **Model Performance Analysis:**
   * **Precision and Recall:** The model demonstrated high precision and recall for both classes (male and female) at 96%. This indicates strong performance in identifying the correct class and minimizing false positives and negatives.
   * **F1-Score:** An F1-Score of 0.96 for both classes reflects a balanced and effective classification process.
   * **Accuracy:** Achieving 96% overall accuracy highlights the model’s ability to correctly classify most instances.
   * **Comparison with Other Methods:**
     + The deep neural network (DNN) significantly outperformed traditional machine learning models like logistic regression and SVM.
     + Baseline algorithms performed much worse, confirming the value added by the DNN.
2. **Challenges Encountered and Solutions:**
   * **Imbalanced Features:** Managing 21 complex acoustic features could lead to overfitting. The dropout layers in the DNN architecture effectively addressed this, preventing overfitting and ensuring robust learning.
   * **Baseline Comparisons:** Traditional models like random forest and SVM struggled with generalization. The use of a DNN allowed the learning of intricate patterns in the data, overcoming the limitations of simpler models.
   * **Feature Extraction:** No extensive preprocessing was done, which might limit the model's performance. However, extracting well-defined audio features provided enough structure for effective modeling.
   * **Overfitting Risks:** Achieving similar training and testing accuracies (95.13% and 95.74%, respectively) suggests that the model avoided overfitting, likely due to effective dropout implementation.

**6.Conclusion**

Our deep neural network (DNN) achieved over 95% accuracy, recall, and precision on the testing dataset, outperforming all previously implemented algorithms. This establishes our model as a new benchmark for gender detection using voice data. Additionally, achieving over 95% accuracy on the training dataset indicates that our model is free from overfitting, demonstrating its reliability and suitability for deployment in real-world applications.

**Lessons Learned:**

Through this project, we gained valuable skills in sourcing and validating data, fostering effective team collaboration, managing time efficiently, and presenting detailed and structured reports. These experiences have not only contributed to the success of this project but also enhanced our overall technical and professional abilities.

**7. References**

<https://www.kaggle.com/datasets/primaryobjects/voicegender/data>

<https://www.primaryobjects.com/2016/06/22/identifying-the-gender-of-a-voice-using-machine-learning/>

ChatGPT

Introduction to Deep Learning