**Speaker Accent recognition and detection**

**Introduction**

**Background information:** Speaker accent recognition and detection is the process of identifying and categorizing the accent of a speaker based on their speech patterns. Accents are an important aspect of language and culture, and the ability to accurately identify them can have applications in a variety of fields, including speech recognition, language learning, and forensic linguistics.

**Research problem:** Accurate accent recognition and detection can be challenging, as accents can vary widely depending on factors such as regional dialects, cultural background, and individual speaking styles. Additionally, there may be overlap between different accents, making it difficult to distinguish between them.

Despite these challenges, significant progress has been made in the field of accent recognition and detection in recent years, with many state-of-the-art systems achieving high levels of accuracy.

**Research questions:**

* How accurately can we classify different accents using the frequency coefficients?
* How does the accuracy of accent classification vary across different speakers and language backgrounds?

**Project Data set** :

* I obtained the data for this project from the following link: [http://archive.ics.uci.edu/ml/datasets/Speaker+Accent+Recognition#](http://archive.ics.uci.edu/ml/datasets/Speaker+Accent+Recognition).
* The attribute information for the Speaker Accent Recognition data set can be summarized as follows:
* Response variable:
* language: a categorical variable indicating the accent of the speaker in the recording. There are six possible categories: ES (Spanish), FR (French), GE (German), IT (Italian), UK (British English), and US (American English).
* Explanatory variables:
* X1, X2, ..., X12: a set of twelve variables obtained using MFCC (Mel Frequency Cepstral Coefficients) on the original time domain soundtrack of the maximum 1 second of reading of a word. These variables capture information about the frequency content of the speech signal and are commonly used in speech analysis and recognition tasks.
* In summary, this dataset contains recordings of speakers reading a word in one of six different accents, along with 12 variables derived from the speech signal. The goal of the analysis would be to predict the accent of a speaker given their speech signal features.

**Challenges :**

* creating dummy variables for the categorical response consists of 6 types of accents.

I have used as. factor() function to create dummy variables for the each of the accents. This function helps to create the categorical variables into levels. Here I have created 6 dummy variables for a language categorical variable based on the accent types.

* Finding the reduced model for each accent categorical models

As I have 12 predictor variables, selecting the variables based on the p-values less than 0.05, And building the model for the each of accent dummy variables

* Calculating accuracy of each of the model such as original model and reduced model of each accent category.
* selecting the best regression model.

**Models I have Developed:**

1. Since the task is to classify different accents, a logistic regression model has been selected as the classification algorithm. The first step in this approach is to develop separate logistic regression models for each accent category, including ES, FR, GE, IT, UK, and US.

|  |  |  |  |
| --- | --- | --- | --- |
| Model-ES | AIC | BIC | ACCURACY(%) |
| Original(12) | 69.37381 | 118.72256 | 97.872 |
| Reduced() | 64.00139 | 98.16591 | 97.26 |

Upon examining the accuracy of language ES, it was observed that the accuracy is approximately 97%, indicating that detecting this language is not very challenging. As a result, the developed model is able to accurately detect the language with an accuracy of 97%.

|  |  |  |  |
| --- | --- | --- | --- |
| Model FR | AIC | BIC | ACCURACY(%) |
| ORGINAL(12) | 142.4397 | 187.9924 | 92.401 |
| REDUCED(8) | 138.5204 | 172.6849 | 92.40122 |

It was observed that the accuracy of detecting language FR is approximately 92%, indicating that it is not very challenging to detect this language. Therefore, the developed model can accurately detect the language with an accuracy of 92%.

|  |  |  |  |
| --- | --- | --- | --- |
| MODEL-GE | AIC | BIC | ACCURACY(%) |
| ORGINAL(12) | 118.5769 | 167.9257 | 94.44073 |
| REDUCED(4) | 117.8806 | 136.8609 | 94.52888 |

Upon examining the accuracy of language GE, it was found to be around 94%, suggesting that detecting this language is not too difficult. The developed model is capable of detecting the language with a high degree of accuracy, achieving a 94% accuracy rate. It is worth noting that the accuracy of detecting language GE is 2% higher compared to language FR.

|  |  |  |  |
| --- | --- | --- | --- |
| MODEL IT | AIC | BIC | ACCURACY(%) |
| ORGINAL MODEL(12) | 127.8822 | 177.2309 | 92.40122 |
| REDUCED(10) | 125.4917 | 167.2483 | 92.70517 |

Upon analyzing the accuracy of language IT, it was found to be around 92%, suggesting that detecting this language is not particularly challenging. The model that was developed is able to accurately detect the language with a 92% accuracy rate.

|  |  |  |  |
| --- | --- | --- | --- |
| Model-UK | AIC | BIC | ACCURACY(%) |
| ORGINAL MODEL(12) | 117.3777 | 166.7625 | 92.70517 |
| REDUCED(8) | 113.7499 | 147.9144 | 93.31307 |

Upon analyzing the accuracy of language UK, it was found to be around 92%, indicating that detecting this language is not overly difficult. The developed model is capable of accurately detecting the language with an accuracy rate of 92%.

|  |  |  |  |
| --- | --- | --- | --- |
| Model-US | AIC | BIC | ACCURACY(%) |
| Original(12) | 305.8747 | 355.2234 | 77.81155 |
| Reduced(7) | 300.0943 | 330.4628 | 77.81155 |

Upon analyzing the accuracy of language US, it was found to be around 77%, suggesting that detecting this language US is more challenging compared to the other language models.

**Summary:** The results indicate that detecting language ES, FR, GE, IT, and UK is not very challenging, with accuracy rates ranging from 92% to 97%. However, detecting language US is more difficult, with an accuracy rate of around 77%. The developed model accurately detects all languages with high accuracy rates. Additionally, the AIC and BIC values improved for the reduced model in all language models, indicating better model performance.

2. I have Also developed the naïve bayes, QDA and LDA models of original and reduced models

I have developed a logistic regression model for the language variable ,After considering the p-values less than 0.05, the coefficients X4-X10 and X12 were selected from logistic regression for the reduced models of QDA, LDA, and Naive Bayes.

|  |  |  |  |
| --- | --- | --- | --- |
| Models | Naive Bayes-Accuracy(%) | LDA Accuracy | QDA Accuracy |
| Original(12) | 62.61% | 76.8997 | 92.09726 |
| Reduced model(7) | 55.92705 | 57.14286 | 81.45897 |

When compared to LDA, QDA, and Naive Bayes, logistic regression exhibits higher accuracy and is better at accurately detecting accents. Therefore, logistic regression is considered to be the best model for accent recognition and detection.

**Result:** The above results indicate that Logistic regression is the most suitable model for the given accent recognition and detection dataset with the highest accuracy for the accent rcognition .

**Conclusion :** In conclusion, the developed model accurately detects all accents with high accuracy rates, ranging from 77% to 97%. Logistic regression is the best model for accent recognition and detection compared to LDA, QDA, and Naive Bayes due to its higher accuracy. Additionally, the AIC and BIC values improved for the reduced model in all language models, indicating better model performance.

**Applications:**

1. **Speech recognition systems**: Automatic speech recognition (ASR) systems can benefit from accent recognition to improve their accuracy in recognizing spoken words. By detecting the accent of the speaker, the ASR system can adapt its language models and acoustic models to better match the speaker's dialect.
2. **Language learning:** The dataset can be used to train models that can help language learners identify different accents and improve their ability to understand and communicate with speakers from different regions.
3. **Forensic investigations:** The dataset can be used in forensic investigations to identify the accent of a suspect in a recorded conversation. This information can be used to narrow down the pool of potential suspects and provide valuable evidence in criminal investigations.
4. **Sociolinguistics research:** The dataset can be used to study the relationship between accent and social factors, such as geography, age, and gender. This can provide insights into how accents are acquired and how they evolve over time.

**References:**

1. Garimella, K. R., et al. "Speaker accent recognition using deep neural networks." 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2017.
2. Sridharan, S., et al. "Automatic speaker recognition: A study on the TIMIT database." Proceedings of the Australian Speech Science and Technology Conference. 1998.
3. Ko, T., et al. "A study on convolutional neural network for speech recognition." IEEE Signal Processing Letters 23.5 (2016): 641-645.
4. Lee, K. A., et al. "Multilingual accent recognition: Exploring the impact of training and testing with non-native speakers." 2019 IEEE Spoken Language Technology Workshop (SLT). IEEE, 2019.
5. Sainath, T. N., et al. "Deep convolutional neural networks for LVCSR of low-resource languages." 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2015.
6. Choi, K., et al. "Speaker accent recognition using transfer learning from speech recognition." 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU). IEEE, 2019.
7. Alam, M. J. E., et al. "An overview of speaker recognition: Technology, databases and evaluation." Pattern Recognition Letters 33.7 (2012): 859-872.