FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

Declaration

Plagiarism is defined as “the unacknowledged use, as one’s own work, of work of another person, whether or not such work has been published” (Regulations Governing Conduct at Examinations, 1997, Regulation 1 (viii), University of Malta).

I / We*, the undersigned, declare that the [assignment / Assigned Practical Task report / Final Year Project report] submitted is my / our* work, except where acknowledged and referenced.

I / We* understand that the penalties for making a false declaration may include, but are not limited to, loss of marks; cancellation of examination results; enforced suspension of studies; or expulsion from the degree programme.

Work submitted without this signed declaration will not be corrected, and will be given zero marks.

* Delete as appropriate.

(N.B. If the assignment is meant to be submitted anonymously, please sign this form and submit it to the Departmental Officer separately from the assignment).

Edward Thomas Sciberras

Student Name Signature

Student Name Signature

Student Name Signature

Student Name Signature

ICS2203 & ARI2203 Speech Phoneme Analysis and Classification

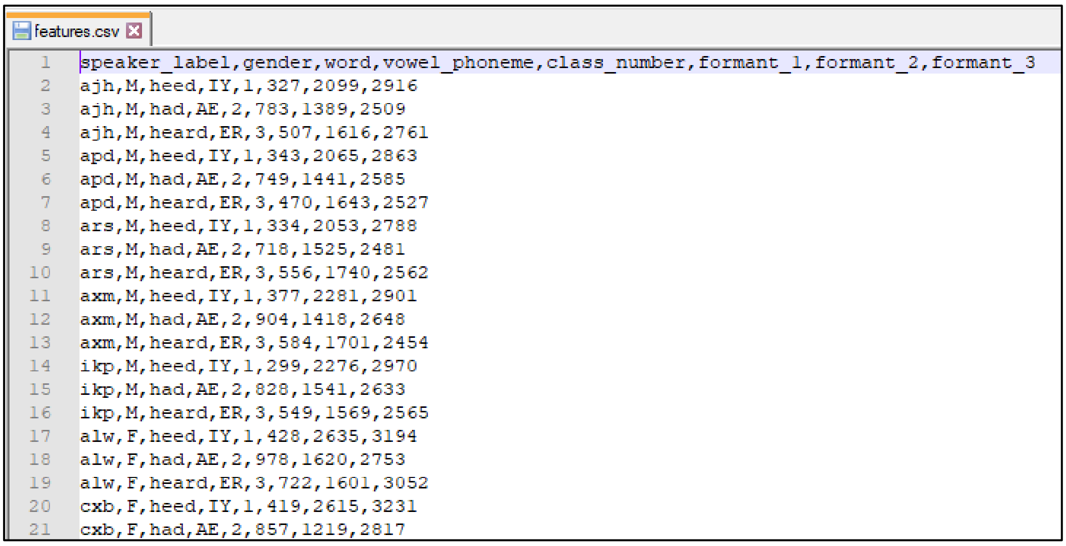
Course Code Title of work submitted

13/04/22

Date

Speech Phoneme Analysis and Classification

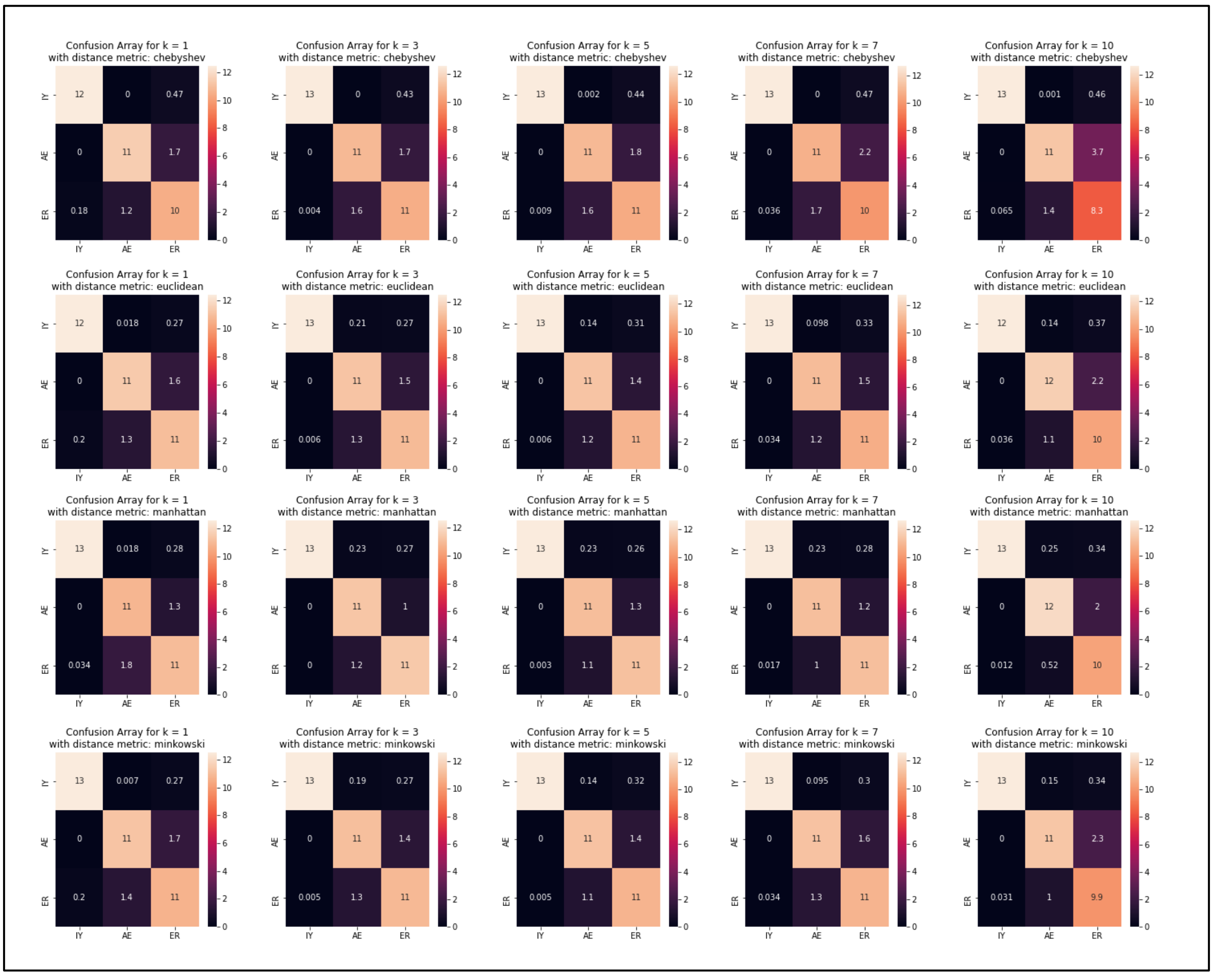
This report will feature the steps that have been taken to collect valid data correctly, how the model was trained (using different parameters) and testing the model to see the accuracy. Firstly, five accents were chosen with five speakers from each gender selected. The words along with their phoneme that were picked out are the following: ‘heed’ (IY), ‘had’ (AE) and ‘heard’ (ER). The three formants were analysed for each of the speakers and the frequencies taken down. The data was then exported to a CSV file for further processing.



```
speaker_label,gender,word,vowel_phoneme,class_number,formant_1,formant_2,formant_3
ajh,M,heed,IY,1,327,2099,2916
ajh,M,had,AE,2,783,1389,2509
ajh,M,heard,ER,3,507,1616,2761
apd,M,heed,IY,1,343,2065,2863
apd,M,had,AE,2,749,1441,2585
apd,M,heard,ER,3,470,1643,2527
ars,M,heed,IY,1,334,2053,2788
ars,M,had,AE,2,718,1525,2481
ars,M,heard,ER,3,556,1740,2562
axm,M,heed,IY,1,377,2281,2901
axm,M,had,AE,2,904,1418,2648
axm,M,heard,ER,3,584,1701,2454
ikp,M,heed,IY,1,299,2276,2970
ikp,M,had,AE,2,828,1541,2633
ikp,M,heard,ER,3,549,1569,2565
alw,F,heed,IY,1,428,2635,3194
alw,F,had,AE,2,978,1620,2753
alw,F,heard,ER,3,722,1601,3052
cxb,F,heed,IY,1,419,2615,3231
cxb,F,had,AE,2,857,1219,2817
```

Figure 1 - First records of feature CSV file

Once the algorithm was working for a singular random seed, k-value and distance metric, it was then decided to change these parameters and start seeing the differences it would cause in the outcome and observe them as well as possible. Moreover, the parameters were tested are as follows, using five different values of ‘k’ (1, 3, 5, 7, 10) and using four different distance metrics (Euclidean, Manhattan, Chebyshev, Minkowski). Not only were these variables testing individually, but over several loops with different training sets and testing sets. In essence, each set-up of the algorithm was run one thousand times each and an average was kept in the confusion matrix to hold the best and most accurate results possible.



Confusion Array for k = 1 with distance metric: chebyshev

Confusion Array for k = 3 with distance metric: chebyshev

Confusion Array for k = 5 with distance metric: chebyshev

Confusion Array for k = 7 with distance metric: chebyshev

Confusion Array for k = 10 with distance metric: chebyshev

Confusion Array for k = 1 with distance metric: euclidean

Confusion Array for k = 3 with distance metric: euclidean

Confusion Array for k = 5 with distance metric: euclidean

Confusion Array for k = 7 with distance metric: euclidean

Confusion Array for k = 10 with distance metric: euclidean

Confusion Array for k = 1 with distance metric: manhattan

Confusion Array for k = 3 with distance metric: manhattan

Confusion Array for k = 5 with distance metric: manhattan

Confusion Array for k = 7 with distance metric: manhattan

Confusion Array for k = 10 with distance metric: manhattan

Confusion Array for k = 1 with distance metric: minkowski

Confusion Array for k = 3 with distance metric: minkowski

Confusion Array for k = 5 with distance metric: minkowski

Confusion Array for k = 7 with distance metric: minkowski

Confusion Array for k = 10 with distance metric: minkowski

Figure 2 - Heatmaps of every variation with regards the k-value and distance metric.

A singular heat map from the photo above represents a single variation of the possible set ups with regards to k-value and distance metric. The heat maps have what phoneme sound the model has predicted on the y-axis against what the vowel sound actually was on the x-axis. Thus, the leading diagonal in each graph represents a correct prediction from the model and any other square was a mistake.

Once each variation was analysed and compared to the others. It can be noticed that the most accurate k-value was three and five as they had the most correct predictions with respect to their distance metric that was being used. This means that usually, the closest three to five nodes to the test subject are the correct ones. Using only one might cause the test node to get caught by an outlier and using ten might cause noise from far away nodes.

Regarding the distance metric, the metric that had the highest number of accurate predictions on average was the Manhattan metric. This being said, Minkowski was extremely close behind and the other two were not far off either.

In addition, when looking at the graphs individually, a pattern can be seen that the most wrong predictions came from the model thinking that the 'ER' sound is an 'AE' sound and vice-versa. This is not to say that there were not other errors but those two sounds were mixed up for each other more than any other mistakes.