

Personal Report

Introduction to AI

Taha Jomehpour Arashlou

Student ID: 2100330056

Our group decided to go with the Smoking detector through bio signal and I helped to develop our initial draft for the intermediate submission.

I removed ID as we already had numbers of the columns

As Gender, Oral and Tartar were objects I had to map them to make them a binary values so it can be used in my confusion metrics and my model.

I also have made a heatmap to show the correlation of each values so we can compare them with each other, For example the more red and higher the box is, it means the values have more impact on each other.

I used libraries such as sklearn and rbf kernel.

With the help of these libraries I got values for Accuracy, F1 Score, Precsion, Recall and err or rate.

Figures:

I have used 4 graphs to showcase my model and to help me compare to other models.

Models used: Confusion Matrix, ROC Curve, Precision Recall Curve

Confusion matrix is a helpful tool in svm and it breaks down the prediction into 4 sections tp, tn, fp and fn.

By observing these set of data it helped me understand specificity my model.

Difficulties and problems faced:

I started by using the same data cleaning and data splitting that we had for all our models but after developing my code further I had a result of 0 for everything except accuracy. with data cleaning, so I decided to load my data separately and do data cleaning and data training/splitting from scratch,

When I used the shared data training parts as my group, I would get a result of 0 for everything except accuracy.

While trying to get a feature importance to compare with other models, I realised that SVM does not allow feature importance.

Conclusion:

by making these model I have learned how to work better with github, python, sklearn libraries and mostly with working in a group and communication.