Departamento de Ciencias de la Computación
Universidad de Ingeniería y Tecnología

Inteligencia Artificial
Proyecto #2: Classification
Prof. Cristian López Del Alamo

The objective of this project is to classify patients as either having COVID or not, using only the sound of the patient's cough. For this, your group can use libraries to obtain the best feature vector to represent the sound of the cough.

The dataset contains the sound signals of Covid-19 positive and negative patients. The Dataset is created from the samples collected from COSWARA and Virufy which are highly reliable. There are 1207 coughs which are of people tested negative and 150 coughs are of COVID-19 positive people.

Activities

1. Use the *dataset* cough sounds and apply the following classification algorithms: Logistic Regression, SVM, Decision Trees, and KNN.

2. Implement (score out of 20) or use libraries (score out of 15) to classify the dataset using SVM, KNN, and Decision Trees.

3. Perform the training process using *K-fold cross-validation* and *Bootstrap* to estimate the error.

4. In a table, present the Precision, *Recall*, and *F1-Score* values for each hyperparameter test in each model.

5. Finally, conclude which models provide the best results.

NOTE: Each team must upload a single document with the following structure.

If your team implements all the algorithms, the grade will be evaluated out of 20; if you only use libraries, the grade is out of 15.

1. Introduction.

2. Brief explanation of the models.

3. Experiments. This is the most important part of the project and must be done thoroughly. Use graphs and tables to show your findings. It is important that your graphs and tables allow for a visual comparison between the models.

4. Conclusions: Write the conclusions based on the results you have obtained.

5. As part of the experiments, you will test your best model with the test data (Tab 2) and generate a confusion matrix with the percentages obtained for each class and the total **accuracy** of your model.

6. The document must be written in a *paper* style, considering the following format ieee, and **it must have a maximum of 9 pages, including references.**

7. The best work will be submitted to an event for publication.

Finally, include the GitHub or Colab link to verify the source code. Also, define a seed so that we can replicate your results and verify that your source code truly works.

Note: One learns by working honestly, and it goes without saying that any form of plagiarism would result in the application of UTEC's plagiarism policy.