

**SWINBURNE UNIVERSITY OF TECHNOLOGY**

COS40007 – Artificial Intelligence for Engineering

Portfolio Assessment-2: “Systematic approach to develop ML model”

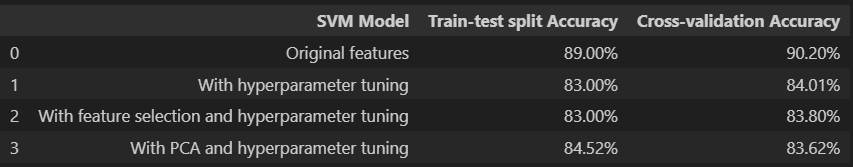
Student Name: Nguyen Cuong Nhat

Student ID: 104178590

Instructor: Dr. Pham Thi Kim Dung

Submission Date: 6/10/2024

# Summary Table of Studio 3: Activity 6 [1 mark]

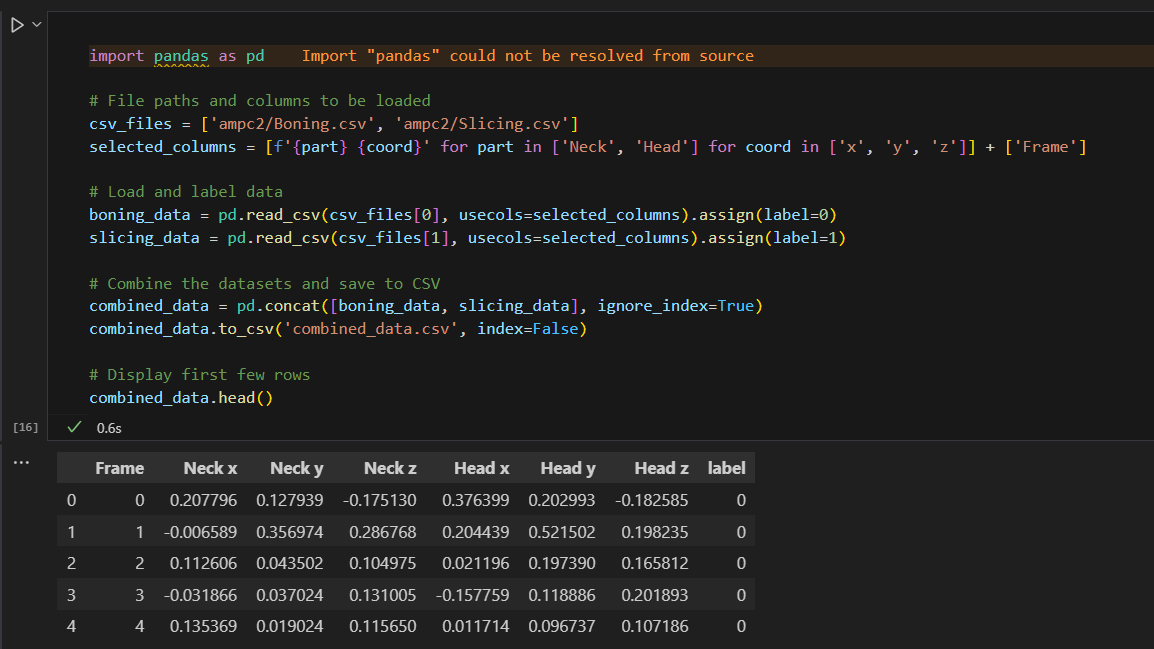


# Summary Table of Studio 3: Activity 7 [1 mark]

# Step 1: Data collection (link of your source code and data) [1 mark]

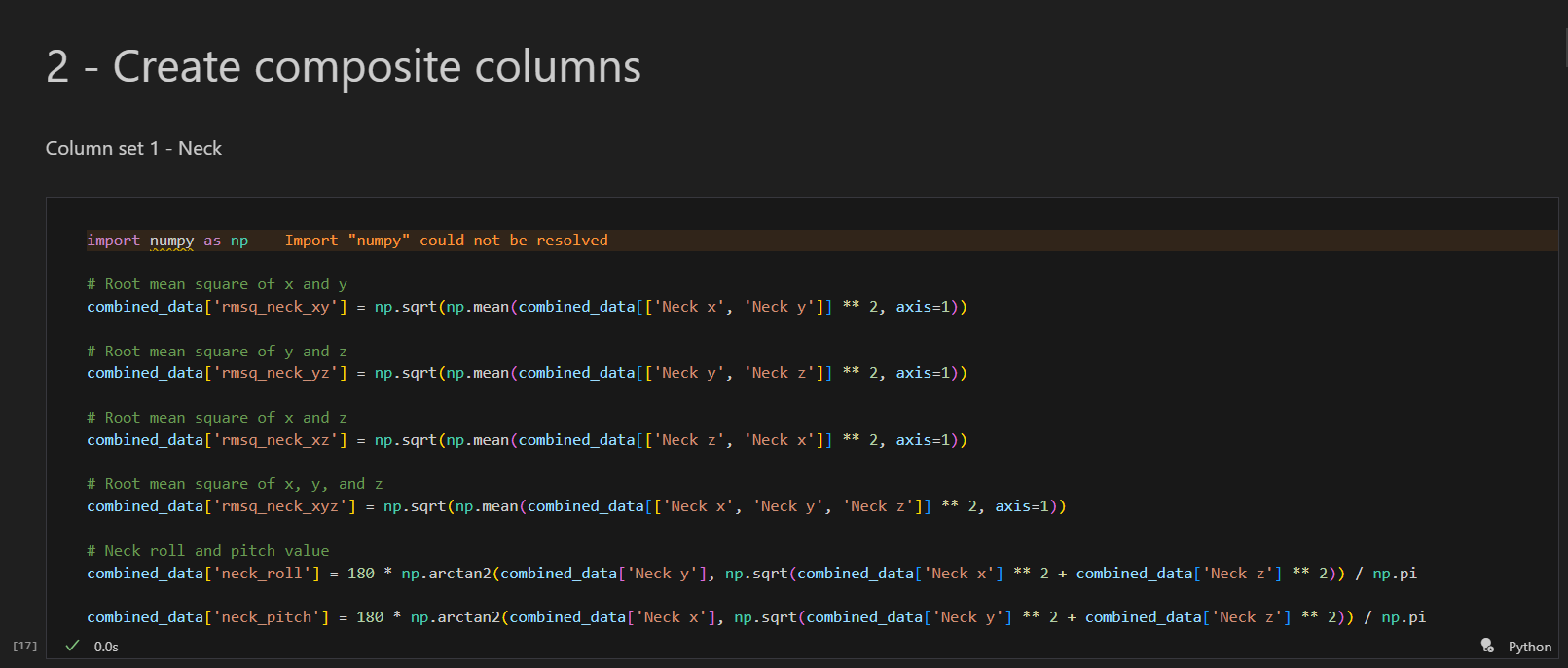
Link of source code and data: <https://drive.google.com/drive/folders/1sTtohPmIwONP5OappyijrrVhUiEIzfue?usp=sharing>

Extract the columns from the 2 files and combine them along with class value (0- boning, 1 – slicing)



Step 2: Create composite columns (link of your source code and data) [1 mark]

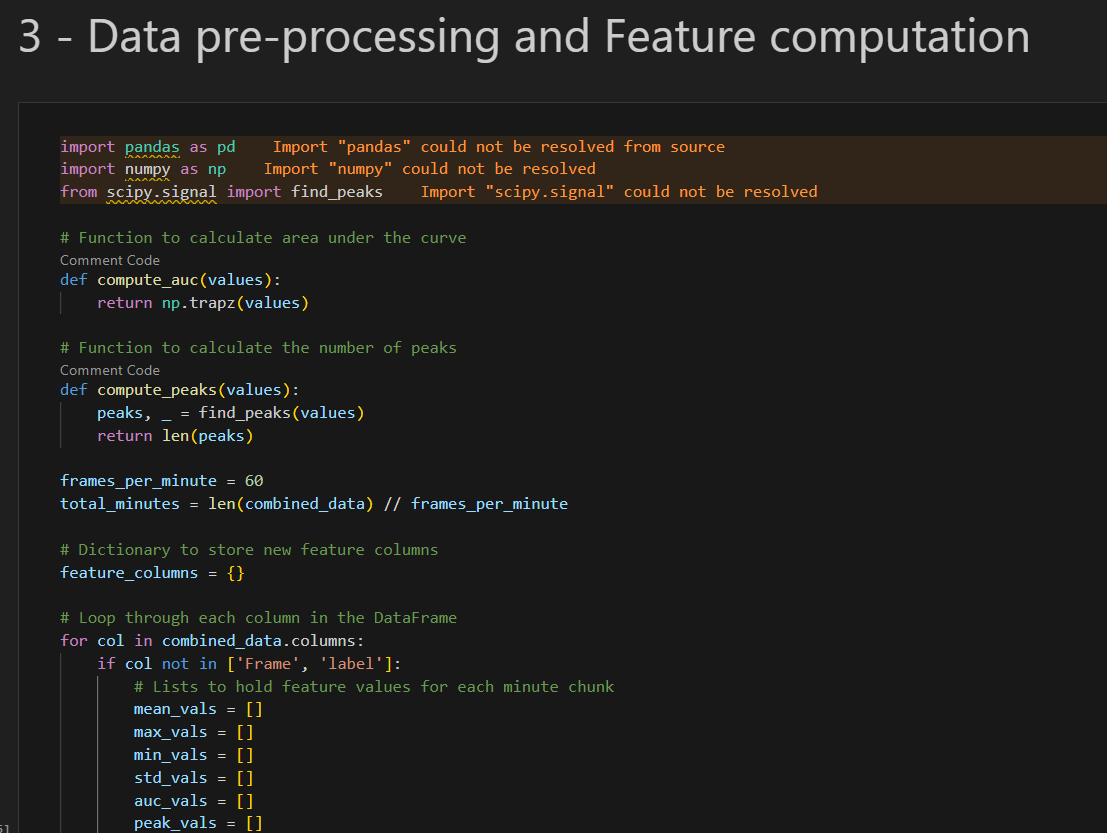
Link of source code and data: <https://drive.google.com/drive/folders/1sTtohPmIwONP5OappyijrrVhUiEIzfue?usp=sharing>



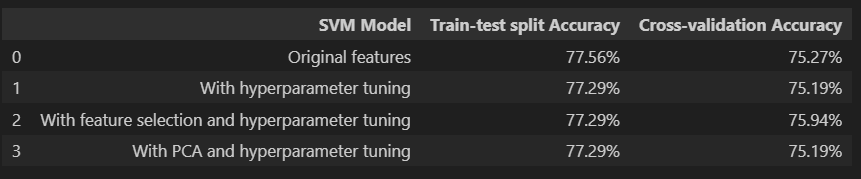
  
Step 3: Data pre-processing (link of your source code and data) [3 marks]

Link of source code and data:

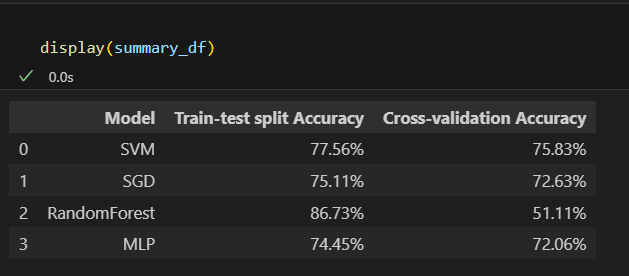
<https://drive.google.com/drive/folders/1sTtohPmIwONP5OappyijrrVhUiEIzfue?usp=sharing>

  
Step 4: Training (outcome summary tables) [2 marks]

SVM classifiers with different settings comparison table



SGD, RandomForest and MLP classifier with original dataset summary table



# Step 5: Model Selection (explain in 1-2 line the reason of your selection) [1 mark]

## Which SVM model will be the best for your problem?

The **SVM model with feature selection and hyperparameter tuning** seems to perform well in general. However, **cross-validation accuracy** is a more reliable metric for evaluating model generalization, and this SVM model has the highest cross-validation accuracy among the models compared. Therefore, the SVM model that applied feature selection and hyperparameter tuning is a best choice for the problem.

## Which ML model will be the best for your problem?

The RandomForest model achieves the highest train-test split accuracy (86.73%) but has a significantly lower cross-validation accuracy (51.11%), indicating potential overfitting. The model with the highest cross-validation accuracy is SVM (75.83%), followed by MLP (72.06%). Since cross-validation accuracy better reflects the performance on unseen data, **SVM** is likely the best overall model for this problem based on the summary table.