Summarize

project in ML

Introduction to Data Science Final Project 2021 sem B- Data Visualization

1. In this project I use all the tools that we have learned during the course on machine learning by the book.

2. My vision was to try and get the best accuracy based on the little data we can get with dimensionality reduction.

3. I did many grid Search to improve the models, also I tried to use K-Means model and pipeline by the explanation in the book but it was not really helpful.

4. The hands data was confusing to import with all the manipulation but was very satisfy to get a good score.

5. The models & methods I used(I used all the models by the instruction of the book):

Models: K Neighbors Classifier, Logistic Regression, Random Forest Classifier, XGB Classifier ,Ada Boost Classifier , Gaussian Classifier , Voting Classifier (hard and soft) , Bagging Classifier(pasting as well) , Stacking Classifier.

Methods: Classification report, Data Scaling, Dimensionality reduction using PCA, Grid/Randomized Search ,confusion matrix.

Notebook number 1(classification):

Starting accuracy from last semester: 94% with knn

After utilize ensemble learning methods: 95.41% with 75% less data by pca algorithm.

Notebook number 2(fminst):   50 features,88% on the test set.

Notebook number 3(cats vs dogs):   125 features,68.5%.

Notebook number 4(hands sync):   25 features,89%.

The project improved my skills in ML- understanding better data and how to deal with it, adhering the book during the work was very helpful to me I recommend it for the next students in this course.