Hacettepe University

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT

BBM467 DATA INTENSIVE APPLICATIONS

SDSP - MACHINE LEARNING WEB APPLICATION

HÜSEYİN BERK YILMAZ – NEZİR TURHALLI

21627873 - 21591132

SPRING 2021

Problem and Purpose

There are diseases that are difficult to diagnose. It makes a disease with different symptoms better for doctors. Different tests can be applied to predict the patient's disease. These tests create time and cost. In this project, we are expected to create a machine learning model for predicting diseases. It is expected that a web application will be designed to facilitate this model.

Data Understanding and Data prepation

I used the mod value of the column it is in to complete the missing values in the data.

Features Selection

Until now, there has been no change in our data over any known situation.**Any action taken can be valid for any data set. However, we have used the 'Diseases' column while creating a train and test set below.** We learned the importance of features by creating a model by using all features for feature selection. On top of that, we separated the important features using **SelectFromModel** and created a new train and test set.

Model And Evaluation

I used the RandomForest classification algorithm. The algorithm produces results that are both fast and highly accurate. I created a new model using the train and test sets I created with important features and trained this model. The accuracy rate of our model is **97.5%**.

Create ML Model And Joblib

I used joblib to save my machine learning model.

Django

I tried to dynamically fit each value. Django works without being bound to any static variable. I rotate the view on the front. I used Html and Css files for the View. Additionally, I used bootstrap for design.

Setup Project

README.txt

Building a new model, If you want to create a new model using a different dataset, you need to run the file named 'SDSPModel.ipynb' in the Model folder again. You can change the file named 'sdsp_patients.xlsx' for rerun. Changing the file or features will not affect the execution of the code. However, it must have a 'Diseases' feature. When this file is recompiled using a different dataset, the

->python manage.py collectstatic

method must be run in the back folder.