

PG Certificate in Software Engineering for Data Science

Explainability Assignment

In this assignment you will get familiar with Explainability methods used to understand ML models decisions. You can make use of the boiler plate code to begin with.

Programming Language: Python

Introduction

In this assignment, we will aim to model bug prediction using Machine Learning and explore the explainability of the model.

This data set consists of the parameters of the code — line count of code, cyclomatic complexity, essential complexity, design complexity, total operators + operands, volume, program length, difficulty, intelligence, effort, time estimator, line count, count of lines of comments, count of blank lines, count of Code And Comment, unique operators, unique operands, total operators, total operands, branchCount of flow graph, defects.

In this data we will be predicting whether a code can have a defect or not looking at the other listed parameters.

Dataset

You can explore the dataset here:

<http://promise.site.uottawa.ca/SERepository/datasets/cm1.arff>

Task:

explainable_AI_starter.ipynb

1. Understand how the h2o package helps in explainability with various plots on relation between attributes and the defect prediction.
2. Try modifying the code on Bike rental dataset and see how the package tries to explain the amount of bikes rented correspond to the various environmental conditions. You can find the dataset for this task here.

<http://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset>

visualize_cnn_starter.ipynb

3. The code given shows the image output after passing through each layer of a model. Implement a model with CNN layers (preferably similar to the model from NN assignment2) and show the output of a sample image after each CNN layer similar to the starter code.

explainable_DT_starter.ipynb

4. The starter code shows how an iris data is classified on a decision tree with one attribute at each node position. Change/Replace the iris dataset to the dataset of your choice and show what attribute is being considered at each node in the decision tree.

Submission

Make changes in the starter notebook and submit the same on the portal.