Third Assigment (Data Science)

Due to: 25 Nov 2023

Feature Engineering

The objective of this assignment is to provide students with a comprehensive understanding of feature engineering techniques in the context of machine learning. Students will learn to manipulate and transform raw data into meaningful and informative features, as well as explore methods for selecting the most relevant features to improve model performance and interpretability. **There is no emphasis on your model accuracy in this assignment but you should discuss the effect of your process on the results.**

Data Preprocessing

1. Describe the steps involved in data preprocessing, including data cleaning, handling missing values, and dealing with outliers.
2. Explain the importance of scaling, normalization, and standardization of features.
3. Demonstrate how to handle categorical variables through techniques such as one-hot encoding, label encoding,... .

There is no limitation on using different methods and techniques in this process!!!

Feature Creation

1. Discuss techniques for creating new features from existing ones, such as polynomial features, interaction features, and domain-specific feature engineering.
2. Illustrate feature extraction methods, including text feature extraction or any other methods that you may need.
3. Provide examples and code to demonstrate the implementation of feature-creation techniques using popular machine-learning libraries.

Model

You are allowed to use a specified model for training and prediction. Be aware that the score of this assignment does not rely on your model accuracy. as we said above your preprocessing pipeline will be graded. **ALERT: IF YOU USE ANY OTHER MODELS, YOUR REPORT AND RESULTS WILL NOT GRADED.**

Data

Click to download it.

NOTE: Since there is no common criterion to evaluate your work properly, you should explain the methods and techniques you used in processing your features; and explain your techniques, or any other logic, behind your pipeline. Use figures, pictures, and tables, and DO NOT PUT ANY CODE IN THE REPORT.