**Data Science for Engineers**

**Lab Report 9**

**Hafiz Ahmad**

**19l-1316**

**Section-7A**

Feature Selection

**INTRODUCTION:**

Feature Selection is the method of reducing the input variable to your model by using only relevant data and getting rid of noise in data. It is the process of automatically choosing relevant features for your machine learning model based on the type of problem you are trying to solve. There are three categories of feature selection methods, depending on how they interact with the classifier, namely, filter, wrapper, and embedded methods. The main goal of feature selection is to improve the performance of a predictive model and reduce the computational cost of modeling. In embedded techniques, the feature selection algorithm is integrated as part of the learning algorithm. The most typical embedded technique is the decision tree algorithm. Decision tree algorithms select a feature in each recursive step of the tree growth process and divide the sample set into smaller subsets.

**OBJECTIVES:**

• To get familiarized with feature selection in data science

**Application:**

The Feature Selection is the method of reducing the input variable to your model by using only relevant data and getting rid of noise in data. It is the process of automatically choosing relevant features for your machine learning model based on the type of problem you are trying to solve. In the machine learning process, feature selection is used to make the process more accurate. It also increases the prediction power of the algorithms by selecting the most critical variables and eliminating the redundant and irrelevant ones. Three key benefits of performing feature selection on your data are: Reduces Overfitting: Less redundant data means less opportunity to make decisions based on noise. Improves Accuracy: Less misleading data means modeling accuracy improves. Reduces Training Time: Less data means that algorithms train faster.

**Issues:**

we never find any issue regarding this lab.

**Conclusion:**

In this lab we understand in short, feature selection helps solve two problems: having too much data that is of little value, or having too little data that is of high value. Your goal in feature selection should be to identify the minimum number of columns from the data source that are significant in building a model. feature selection is one of the most crucial tasks for machine learning problems, after performing data wrangling and cleaning. you can find the functions implementing the feature selection process using XGBOOST feature importance here.