

Homework-03 Report

Introduction

In this homework, I was asked to do classification over a given dataset. The dataset was appropriate for analysis and preprocessing, so that we can try to implement the classification models.

Data Pre-Processing

I used OneHotEncoding since we have categorical features. Also, I used feature selection technique like Chi2 Square to select most important features based on feature scores.

Models

I have tried three models for classifications: **logistic regression**, **KNneighbors classifier**, and **GaussianNB**.

KNneighbor Classifier

The K value was optimized by trying many different K values in order to find best K that maximize the prediction score of the model. I used cross validation score method for hyperparameter optimization.

Logistic Regression

I used the GridSearchCV method to find the best parameter that fits the logistic regression model and maximizes the accuracy score.