T.C.

TOROS UNIVERSITY

FACULTY OF ENGINEERING/ SOFTWARE ENGINEERING PROGRAM

STUDENTS PERFORMANCE IN EXAMS

İBRAHİM HALİL ALPA

205060007

MIDTERM PROJECT

MAY/ 2023

EXPLANATION

I will use the dataset named "Students Performance in Exams". This dataset is a dataset that includes students' performance in math, reading and writing exams. This dataset consists of 1000 students and exam results are recorded according to different characteristics of students such as gender, ethnicity, education level of their parents, type of school and type of food. This dataset can be used for classification and regression analysis. Classification analysis aims to classify students into a particular class or category. Regression analysis, on the other hand, can be used to predict students' test scores.

DATASET: https://www.kaggle.com/datasets/spscientist/students-performance-in-exams

DATASET PROPERTIES

Characteristics of the dataset include gender, ethnicity, parents' education level, school type, type of food, and scores on math, reading and writing tests. These features can be used to analyze students' exam performance. For example, characteristics such as students' gender and parents' education level may have an impact on test scores, and the dataset can be used to determine these effects.

ALGORITHMS THAT CAN BE USED

Random Forest: Random Forest is a machine learning algorithm created by combining many decision trees. I will use this algorithm to classify students' exam results using features from the dataset. Random Forest is also a good option to prevent overfitting.

K-Nearest Neighbors: K-Nearest Neighbors is an algorithm that groups samples in the dataset according to their similarity to each other. I intend to use this algorithm to predict students' exam results. For example, if a student's exam result is to be predicted, it can be estimated by looking at the exam results of similar students.

Decision Tree: Decision Tree is an algorithm that creates a decision tree based on the features in the dataset. These trees can be divided into certain classes or categories based on the features in the dataset. I intend to use this algorithm to classify students' exam results using features in the dataset.

CONCLUSION

In summary, the "Students Performance in Exams" dataset includes characteristics such as gender, ethnicity, parents' education level, school type, type of food, and scores on math, reading and writing tests, which can be used to analyze students' exam performance. Random Forest, K-Nearest Neighbors, and Decision Tree are the algorithms that can be used to classify and predict students' exam results. By utilizing these algorithms, we can gain insights into the factors affecting students' exam performance and make informed decisions to improve education systems.