

Group 13 Proposal: Alcohol Effects On Academic Performance

Target

1. Analyze the impact of students' drinking habits and other factors on students' academic performance;
2. Use statistical and ML models to support conclusions and try to find some potential interactions among given variables.

Dataset

Kaggle: Alcohol Effects on Study

(<https://www.kaggle.com/datasets/whenamancodes/alcohol-effects-on-study>)

This dataset contains 1044 student exam scores and their personal information collected by two Portuguese secondary schools. Each piece of data includes 33 different students' personal information, including alcohol usage condition, student grades, family status (parental education level and job), and living habits (study time, drinking habits, etc.)

Proposed Solution

In order to analyze how alcohol and other factors affect academic performance, Exploratory Data Analysis (EDA) will play a critical role in discovering these potential relationships. We will use the histogram, heatmap, and scatterplot to help intuitively present the distribution of different factors including alcohol usage, sex, age, etc. Besides, some charts and quantitative indicators including the Pearson correlation coefficient or mutual information will be leveraged to help measure the correlation between the factors listed above and the final academic performance.

And we also want to use tree-based machine learning techniques including XGBoost, GBDT, etc, as auxiliary tools to help predict the grade based on other features provided. And this will help us find out the importance of overall features.

Future Applications

Improving students' studying achievement is a concern of every school and parents. In addition to learning efficiency in school, drinking habits and family factors also have a great impact on students' learning. This project aims to evaluate the impact of these factors on studying and provide guidance to students.

Timeline

Week	Task	Group Member
1	Data Cleaning and Exploration	Yuzhao Chen, Liyang Ru
2~3	Model & Feature Importance	Amber Szulc, Shril Mody
2~3	Data Analysis, Variable Correlation & Plot Trends	Linfeng Wen, Liyang Ru, Yuzhao Chen
4	Presentation	Linfeng Wen