THE OXFORD COLLEGE OF SCIENCE

Department of Computer Science and Applications

(AFFILIATED TO BANGALORE UNIVERSITY)

19TH Main, 17th B Cross, Sector-IV, HSR Layout, Bangalore-560102

PROJECT REPORT

ON

STUDENT ALCOHOL CONSUMPTION

SUBMITTED BY

Nitheesh Kumar M

(U03MS21S0)

Under the guidance of
DR. SUSHIL KUMAR SAHOO

STUDENT ALCOHOL CONSUMPTION Project Report submitted in partial fulfillment of the requirements of Sixth Semester BCA, Bangalore University

MAY -*2024*

Table of Content:

Introduction

The student performance dataset is derived from a survey conducted on secondary school students enrolled in Math and Portuguese language courses. It encompasses a wide array of social, gender, and study-related information, making it suitable for exploratory data analysis (EDA) and predictive modeling to forecast students' final grades.

Objectives

- To analyze and understand the factors influencing student performance.
- To predict students' final grades based on various attributes.
- To provide insights into the social and educational factors affecting academic outcomes.
- To support educational institutions in identifying areas for improvement.

Key Features of the Dataset

- **Attributes**: Detailed information about students, including demographics, social factors, and academic records.
- **Courses**: Separate datasets for Math (`student-mat.csv`) and Portuguese (`student-por.csv`) courses.
- **Grade Information**: Grades for the first period (G1), second period (G2), and final grade (G3) for each student.

Content Description

The dataset includes the following attributes for both `student-mat.csv` and `student-por.csv`:

- **Demographics and Personal Information**:
  - `school`: Student's school (binary: 'GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira)
  - `sex`: Student's sex (binary: 'F' - female or 'M' - male)
  - `age`: Student's age (numeric: 15 to 22)
  - `address`: Student's home address type (binary: 'U' - urban or 'R' - rural)
  - `famsize`: Family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3)
  - `Pstatus`: Parent's cohabitation status (binary: 'T' - living together or 'A' - apart)
- **Parental Information**:
  - `Medu`: Mother's education (numeric: 0 - none to 4 - higher education)
  - `Fedu`: Father's education (numeric: 0 - none to 4 - higher education)
  - `Mjob`: Mother's job (nominal: 'teacher', 'health', 'services', 'at_home', 'other')
  - `Fjob`: Father's job (nominal: 'teacher', 'health', 'services', 'at_home', 'other')
- **School-Related Information**:
  - `reason`: Reason to choose the school (nominal: 'home', 'reputation', 'course', 'other')

- o `guardian`: Student's guardian (nominal: 'mother', 'father', 'other')
- o `traveltime`: Home to school travel time (numeric: 1 - <15 min to 4 - >1 hour)
- o `studytime`: Weekly study time (numeric: 1 - <2 hours to 4 - >10 hours)
- **Support and Activities**:
  - o `failures`: Number of past class failures (numeric: 0 to 4)
  - o `schoolsup`: Extra educational support (binary: yes or no)
  - o `famsup`: Family educational support (binary: yes or no)
  - o `paid`: Extra paid classes within the course subject (binary: yes or no)
  - o `activities`: Extra-curricular activities (binary: yes or no)
  - o `nursery`: Attended nursery school (binary: yes or no)
  - o `higher`: Wants to take higher education (binary: yes or no)
  - o `internet`: Internet access at home (binary: yes or no)
  - o `romantic`: With a romantic relationship (binary: yes or no)
- **Student Lifestyle and Health**:
  - o `famrel`: Quality of family relationships (numeric: 1 - very bad to 5 - excellent)
  - o `freetime`: Free time after school (numeric: 1 - very low to 5 - very high)
  - o `goout`: Going out with friends (numeric: 1 - very low to 5 - very high)
  - o `Dalc`: Workday alcohol consumption (numeric: 1 - very low to 5 - very high)
  - o `Walc`: Weekend alcohol consumption (numeric: 1 - very low to 5 - very high)
  - o `health`: Current health status (numeric: 1 - very bad to 5 - very good)
  - o `absences`: Number of school absences (numeric: 0 to 93)
- **Academic Performance**:
  - o `G1`: First period grade (numeric: 0 to 20)
  - o `G2`: Second period grade (numeric: 0 to 20)
  - o `G3`: Final grade (numeric: 0 to 20, output target)

Data Collection

- **Source**: The data were obtained through a survey of secondary school students.
- **Format**: The datasets are available in CSV format (`student-mat.csv` for Math and `student-por.csv` for Portuguese).
- **Period**: The survey was conducted over a specific period, capturing a diverse set of student attributes.

Existing System & Its Drawbacks

Existing System

Currently, many educational institutions rely on manual record-keeping and basic statistical methods to monitor and predict student performance. This traditional approach often involves teachers and administrators using spreadsheets and simple databases to track grades and other student information.

Drawbacks

1. **Limited Insights**: The existing system provides limited insights into the underlying factors influencing student performance.
2. **Time-Consuming**: Manual data entry and analysis are time-consuming and prone to human error.
3. **Lack of Predictive Power**: Traditional methods lack advanced predictive capabilities to forecast student outcomes.
4. **Inconsistent Data**: Inconsistencies and missing data due to manual processes can lead to inaccurate analyses.
5. **Inefficient Resource Allocation**: Schools may struggle to allocate resources effectively without advanced data analysis tools.

Proposed System

The proposed system leverages advanced data analytics and machine learning techniques to analyze the student performance dataset. This system aims to automate the process of data analysis, providing deeper insights and more accurate predictions of student outcomes.

Advantages of the Proposed System

1. **Enhanced Predictive Accuracy**: Utilizes machine learning models to accurately predict students' final grades.
2. **Data-Driven Insights**: Provides actionable insights into the factors affecting student performance.
3. **Efficiency**: Automates data processing and analysis, saving time and reducing the risk of errors.
4. **Resource Optimization**: Helps educational institutions allocate resources more effectively based on data-driven insights.
5. **Improved Decision-Making**: Supports administrators and educators in making informed decisions to enhance student performance.

Drawbacks

1. **Implementation Cost**: Initial setup and implementation of the advanced system may be costly.
2. **Data Privacy**: Ensuring the privacy and security of student data is critical and can be challenging.
3. **Technical Expertise**: Requires technical expertise to set up and maintain machine learning models and data analysis tools.
4. **Data Quality**: The effectiveness of the system is dependent on the quality and completeness of the data collected.

Hardware Requirements

- **Processor**: Intel Core i5 or higher
- **RAM**: 8 GB or more
- **Storage**: 500 GB HDD or SSD
- **Graphics**: Dedicated GPU (optional for deep learning models)
- **Internet**: Stable internet connection for data access and model updates

Software Requirements

- **Operating System**: Windows 10 or higher / macOS / Linux

- **Programming Language**: Python, R
- **Data Analysis Tools**: Jupyter Notebook, RStudio
- **Database Management**: MySQL, PostgreSQL
- **Machine Learning Libraries**: Scikit-learn, TensorFlow, Keras
- **Visualization Tools**: Matplotlib, Seaborn, Tableau

Applications

1. **Educational Research**: Analyzing factors influencing student performance for academic studies.
2. **Institutional Improvement**: Providing insights for schools to enhance teaching methods and support services.
3. **Personalized Learning**: Developing personalized learning plans based on individual student data.
4. **Policy Making**: Assisting policymakers in making data-driven decisions to improve education systems.
5. **Student Counseling**: Supporting counselors in identifying students in need of academic or social support.

Conclusion

The student performance dataset offers a rich source of information for analyzing and predicting student outcomes. By leveraging the detailed attributes provided, researchers and educators can gain valuable insights into the factors influencing academic performance. Continuous improvements in data collection methods and analytical techniques are essential to maximize the utility of this dataset and support student success.