# MS302

# Business Analytics Project

## (Business Intelligence in Education)

Group 11

# Project Objective

Predicting Student Performance using Educational Data Mining (EDM) Techniques in Secondary School.

## Problem Statement

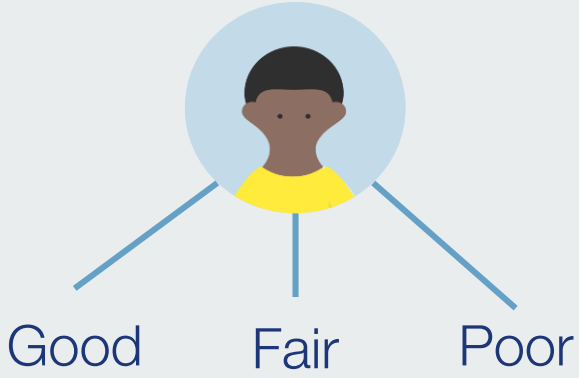
- How to accurately analyse the student’s performance in school and find out which factors affect their academic performance?

# Our Flow

1. Data Collection
2. Data Analysis
3. Selection of Classification Model
4. Prepare Dataset for Modelling
5. Predict the student’s performances
6. Results and Conclusions

# 1. Data Collection

## Three Level Classification




## Features Analysed

- Romantic Status
- Alcohol Consumption
- Parents Education Level
- Frequency Of Going Out
- Desire Of Higher Education
- Living Area

# 2. Data Analysis

**Dataset contains features like-**

Student's school, student's sex, student's age, student's home address type, family size, parent's cohabitation status, mother's education, father's education, mother's job, father's job, reason to choose this school, student's guardian, home to school travel time, weekly study time, failures, extra-curricular activities, attended nursery school, wants to take higher education, Internet access at home, with a romantic relationship, free time after school, going out with friends, workday alcohol consumption, weekend alcohol consumption, current health status, number of school absences

## Math/Portuguese

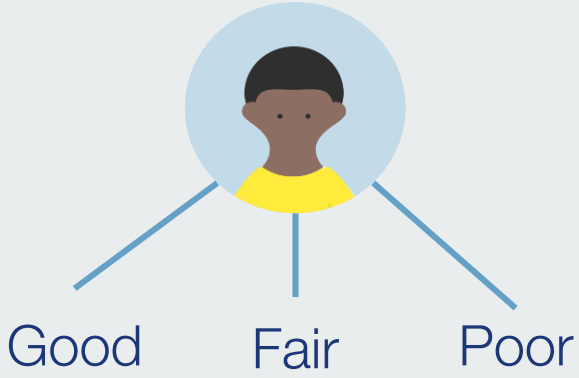
- G1 - first-period grade (numeric: from 0 to 20)
- G2 - second-period grade (numeric: from 0 to 20)
- **G3 - Final Grade (numeric: from 0 to 20, output target)**

these affect

*Exploratory Data Analysis in the report.

# 3. Classification

## Three Level Classification




## Models

- Decision Trees
- Random Forest
- SVM
- Logistic Regression
- ANN
- Naïve Bayes

# 4. Prepare Dataset for modelling

## Steps

- Slicing Dataset (Taking out Final Grade)
- Test Train Split **(7:3)**
- Total features - **58**

# 5. Prediction

| | Models | | | | | |
|---|---|---|---|---|---|---|
| | Decision Tree | Random Forest | SVM | Logistic Regression | ANN | Naive Bayes |
| Model Score | 0.90 | 0.95 | 0.94 | 0.90 | 0.90 | 0.67 |
| Cross-Validation Score | 0.88 | 0.86 | 0.86 | 0.86 | 0.85 | 0.65 |
| Accuracy Percentage | 91.93 | 90.66 | 90.87 | 90.87 | 90.65 | 76.85 |

*More detailed description in the report.