CO544 - PROJECT PROPOSAL

Group 15

Group Members

- E/16/127 - Yoshith Sri Harashana
- E/16/290 - Kalani Proboda
- E/16/039 - Nimashi Uthpala

Background

- **Student Performance Classification System**

This is a machine learning project that can be used to predict the student performance of the common entrance examination(CEE), who qualified the medical entrance examination for admission to a medical college.

According to the project the student performance can be predicted using information such as their previous educational records and their family background.This research can be used in the Sri Lankan educational system as well, to predict the performance of University entrance students. So that it helps to improve the educational system of schools in the country and produce more efficient students. Also it can encourage the school system to obtain better results.

Source of data

This dataset contains data of the candidates who qualified the medical entrance examination for admission to medical colleges of Assam of a particular year and collected by Prof. Jiten Hazarika.

Provided by : Dr. Sadiq Hussain, Dibrugarh University, sadiq **'@'** dibru.ac.in

Repository:
https://archive.ics.uci.edu/ml/datasets/Student+Performance+on+an+entrance+examination

Modelling Problem

The purpose of the research is to Predict student's performance in Common Entrance Examination using a classification method.As the dataset is labeled, it is a supervised learning problem.In this problem we are going to predict a label for student's performance,such as "Excellent" , " Very Good " ," Good" or "Average" according to their independent variable values.Since the performance labeled as multivalued, the modeling problem is a **multiclass classification** problem.

- Output labels : Performance (Excellent, Very Good, Good, Average)
- Inputs: All the attributes except 'Performance' (attribute details are given below under "Summary of details of the dataset ")

Summary of details of the dataset

Number of instances : 666

Number of Attributes : 12

Attributes :

- Performance in Common Entrance Examination (CEE)
  - **Performance** : {'Excellent','Vg','Good','Average'}
- Gender of the Candidate
  - **Gender** : {'male','female'}
- Caste of the Candidate
  - **Caste** : {'General','OBC','SC','ST'}
- Time
  - **Time** : {'ONE','TWO','THREE','FOUR','FIVE','SEVEN'}
- Whether the candidate attended any coaching classes within Assam, outside Assam or not
  - **Coaching (coaching class attended)** : {'NO','WA','OA'}
- Name of the board where the candidate studied at Class X level
  - **Class_ten_education** : {'SEBA','OTHERS','CBSE'}
- Name of the board where the candidate studied at Class XII level
  - **Twelve_education** : {'AHSEC','CBSE','OTHERS'}
- Medium of instructions for the study at Class XII level
  - **Medium** : {'ENGLISH','OTHERS','ASSAMESE'}
- The percentage secured by the candidate at Class XII standard
  - **Class_ X_Percentage** : {'Excellent','Vg','Good','Average'}
- The occupation of the father of the candidate
  - **Class_XII_Percentage** : {'Excellent','Vg','Good','Average'}
- The occupation of the father of the candidate
  - **Father_occupation** {'DOCTOR','SCHOOL_TEACHER','BUSINESS','COLLEGE_TEACHER','OTHERS', 'BANK_OFFICAL','ENGINEER','CULTIVATOR'}
- The occupation of the mother of the candidate
  - **Mother_occupation** : {'OTHERS','HOUSE_WIFE','SCHOOL_TEACHER','DOCTOR','COLLEGE_TEACHER', 'BANK_OFFICIAL','BUSINESS','CULTIVATOR','ENGINEER'}

All the data in the dataset are categorical data.There does not exist any null values in the dataset .

Modelling plan

Modeling plan is to use a **multiclass classification mode**l such as **logistic regression** with 80% of the dataset as the training dataset and the remaining 20% as the test dataset.

Number of instances in training set : 533

Number of instances in test set : 133