Школьники

Набор данных взят из репозитория машинного обучения UCL. Ссылка на набор данных

Источник:

Пауло Кортес, университет Минью, Гимарайнш, Португалия,

П. Кортес и А. Сильва. Использование интеллектуального анализа данных для прогнозирования успеваемости учащихся средней школы. В ред. А. Брито и Ж. Тейшейры. Труды 5-го будущего бизнес-технологии конференция (FUBUTEC 2008), стр. 5-12, Порту, Португалия, апрель, 2008, EUROSIS, номер ISBN 978-9077381-39-7.

веб-ссылка http://www3.dsi.uminho.pt/pcortez

Информация О Наборе Данных:

Эти данные отображают достижения учащихся в среднем образовании двух португальских школ. Атрибуты данных включают оценки учащихся, демографические, социальные и школьные характеристики), и они были собраны с помощью школьных отчетов и анкет. Представлены два набора данных, касающихся успеваемости по двум различным предметам: математике (mat) и португальскому языку (por). В работе [Cortez and Silva, 2008] эти два набора данных были смоделированы в рамках задач двоичной/пятиуровневой классификации и регрессии. Важное примечание: целевой атрибут G3 имеет сильную корреляцию с атрибутами G2 и G1. Это происходит потому, что G3-это итоговая оценка года (выданная в 3-м периоде), в то время как G1 и G2 соответствуют 1-му и 2-му классам периода. Предсказать G3 без G2 и G1 сложнее, но такое предсказание гораздо полезнее (Подробнее см. Источник статьи).

**Атрибутивная информация:**

школа - школьник (двоичный: 'ГП' - Габриель Перейра или 'MS' - Mousinho да Силвейра)

секс студент секс (двоичное: 'Ф' - женщина или " М " - мужчина)

возраст-возраст студента (числовой: от 15 до 22 лет)

адрес-тип домашнего адреса студента (двоичный: 'U' - городской или 'R' - сельский)

famsize-размер семейства (двоичный: 'LE3' - меньше или равно 3 или 'GT3' - больше 3)

Pstatus - родительского сожительства состояния (двоичное: 'Т' - жить вместе или " А " - отдельно)

Medu-образование матери (числовое: 0-Нет, 1-начальное образование (4-й класс), 2-5-9-й класс, 3-среднее образование или 4-высшее образование)

Феду-образование отца (числовое: 0-Нет, 1-начальное образование (4 класс), 2-5-9 класс, 3-среднее образование или 4-высшее образование)

Mjob - матери работа (номинальное: "учитель", "здоровье" медицинских услуг, связанных, гражданской "услуги" (например, административные или полицейские), 'at\_home' или 'другое')

Fjob-работа отца (номинально: "учитель", "здравоохранение", связанные с ним гражданские "услуги" (например, административные или полицейские), "at\_home" или "другое")

причина-причина выбрать эту школу (номинально: близко к "дому", школе "репутация", "курс" предпочтения или "другое")

опекун-опекун студента (номинально: "мать", "отец" или "другой")

время в пути - время в пути от дома до школы (числовое: 1-1 час)

studytime-еженедельное время обучения (числовое: 1-10 часов)

отказы-количество прошлых отказов класса (числовое: n, если 1<=n<3, иначе 4)

schoolsup-дополнительная образовательная поддержка (двоичный код: да или нет)

famsup-семейная образовательная поддержка (двоичный код: да или нет)

платные-дополнительные платные занятия в рамках предмета курса (математика или португальский язык) (двоичный код: да или нет)

мероприятия-внеклассные мероприятия (бинарные: да или нет)

детский сад-посещаемый детский сад (двоичный код: да или нет)

высшее - хочет получить высшее образование (бинарное: да или нет)

интернет - доступ в интернет дома (двоичный код: да или нет)

романтика - с романтическими отношениями (бинарно: да или нет)

фамрел-качество семейных отношений (числовое: от 1-очень плохо до 5-отлично)

freetime-свободное время после школы (числовое значение: от 1-очень низкий до 5-очень высокий)

goout-выход с друзьями (числовое значение: от 1-очень низкий до 5-очень высокий)

Dalc-потребление алкоголя в течение рабочего дня (числовое значение: от 1-очень низкое до 5-очень высокое)

Walc-потребление алкоголя в выходные дни (числовое значение: от 1-очень низкое до 5-очень высокое)

здоровье-текущее состояние здоровья (числовое значение: от 1-очень плохо до 5-очень хорошо)

прогулы - количество школьных прогулов (числовое значение: от 0 до 93)

эти оценки связаны с предметом курса, математикой или португальским языком:

G1-класс первого периода (числовой: от 0 до 20)

G2-класс второго периода (числовой: от 0 до 20)

G3-итоговая оценка (числовая: от 0 до 20, выходная цель)

соответствующие документы:

П. Кортес и А. Сильва. Использование интеллектуального анализа данных для прогнозирования успеваемости учащихся средней школы. В ред. А. Брито и Ж. Тейшейры. Труды 5-го будущего бизнес-технологии конференция (FUBUTEC 2008), стр. 5-12, Порту, Португалия, апрель, 2008, EUROSIS, номер ISBN 978-9077381-39-7.

Доступно: веб-ссылка

Attribute Information:

1. school - student's school (binary: 'GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira)
2. sex - student's sex (binary: 'F' - female or 'M' - male)
3. age - student's age (numeric: from 15 to 22)
4. address - student's home address type (binary: 'U' - urban or 'R' - rural)
5. famsize - family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3)
6. Pstatus - parent's cohabitation status (binary: 'T' - living together or 'A' - apart)
7. Medu - mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 â€“ 5th to 9th grade, 3 â€“ secondary education or 4 â€“ higher education)
8. Fedu - father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 â€“ 5th to 9th grade, 3 â€“ secondary education or 4 â€“ higher education)
9. Mjob - mother's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at\_home' or 'other')
10. Fjob - father's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at\_home' or 'other')
11. reason - reason to choose this school (nominal: close to 'home', school 'reputation', 'course' preference or 'other')
12. guardian - student's guardian (nominal: 'mother', 'father' or 'other')
13. traveltime - home to school travel time (numeric: 1 - 1 hour)
14. studytime - weekly study time (numeric: 1 - 10 hours)
15. failures - number of past class failures (numeric: n if 1<=n<3, else 4)
16. schoolsup - extra educational support (binary: yes or no)
17. famsup - family educational support (binary: yes or no)
18. paid - extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)
19. activities - extra-curricular activities (binary: yes or no)
20. nursery - attended nursery school (binary: yes or no)
21. higher - wants to take higher education (binary: yes or no)
22. internet - Internet access at home (binary: yes or no)
23. romantic - with a romantic relationship (binary: yes or no)
24. famrel - quality of family relationships (numeric: from 1 - very bad to 5 - excellent)
25. freetime - free time after school (numeric: from 1 - very low to 5 - very high)
26. goout - going out with friends (numeric: from 1 - very low to 5 - very high)
27. Dalc - workday alcohol consumption (numeric: from 1 - very low to 5 - very high)
28. Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)
29. health - current health status (numeric: from 1 - very bad to 5 - very good)
30. absences - number of school absences (numeric: from 0 to 93)

these grades are related with the course subject, Math or Portuguese:

1. G1 - first period grade (numeric: from 0 to 20)
2. G2 - second period grade (numeric: from 0 to 20)
3. G3 - final grade (numeric: from 0 to 20, output target)