

머신러닝(나) 프로젝트 1차 보고서

20190501 수학과 김태균

1. 문제의 정의

- **목표** : 고등학생의 학업 성취도(GPA)에 영향을 주는 다양한 요인을 분석하고, 학업 성취도 향상에 기여할 수 있는 지침을 제안하고자 한다.
- **연구 가설** : 부모의 교육 수준이 높을수록 자녀의 학업 성취도가 높을 것이다.
- **Task** : 고등학생의 GPA 기반 성적 클래스를 예측하는 분류 문제
- **Subtask** : 다양한 요인(공부 습관, 과외 활동 등)과 학업 성취도의 상관 분석

2. 데이터 정의

- **목적 및 용도** : 고등학생의 학업 성취도에 영향을 미칠 수 있는 여러 요인을 기반으로 GPA 성적 분포를 예측하고 분석하는 데 사용된다.
- **데이터 형식** : csv 형식으로 제공되며, 각 행은 학생 한 명의 정보를 포함한다.

| | StudentID | Age | Gender | Ethnicity | ParentalEducation | StudyTimeWeekly | Absences | Tutoring | ParentalSupport | Extracurricular | Sports | Music | Volunteering | GPA | GradeClass |
|---|-----------|-----|--------|-----------|-------------------|-----------------|----------|----------|-----------------|-----------------|--------|-------|--------------|----------|------------|
| 0 | 1001 | 17 | 1 | 0 | 2 | 19.833723 | 7 | 1 | 2 | 0 | 0 | 1 | 0 | 2.929196 | 2.0 |
| 1 | 1002 | 18 | 0 | 0 | 1 | 15.408756 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 3.042915 | 1.0 |
| 2 | 1003 | 15 | 0 | 2 | 3 | 4.210570 | 26 | 0 | 2 | 0 | 0 | 0 | 0 | 0.112602 | 4.0 |
| 3 | 1004 | 17 | 1 | 0 | 3 | 10.028829 | 14 | 0 | 3 | 1 | 0 | 0 | 0 | 2.054218 | 3.0 |
| 4 | 1005 | 17 | 1 | 0 | 2 | 4.672495 | 17 | 1 | 3 | 0 | 0 | 0 | 0 | 1.288061 | 4.0 |

- 필드 설명

- 학생 ID : 학생을 구별하는 고유 번호 (총 2,392명)
- 나이 : 15 ~ 18세
- 성별 : {0:남자, 1:여자}
- 인종 : {0:백인, 1:아프리카계 미국인, 2:동양인, 3:기타}
- 부모의 교육 수준 : 부모의 최종 학력
{0:학력 없음, 1:고등학교 졸업, 2:대학 경험 있음, 3:학사 학위, 4:석사 학위 이상}
- 주당 공부 시간 : 0 ~ 20시간
- 연간 결석 횟수 : 0 ~ 30회
- 튜터링 : {0:No, 1:Yes}
- 부모의 지원 정도 : {0:없음, 1:낮음, 2:보통, 3:높음, 4:매우 높음}
- 과외 활동 : {0:No, 1:Yes}
- 스포츠 : {0:No, 1:Yes}
- 음악 : {0:No, 1:Yes}
- 봉사활동 : {0:No, 1:Yes}
- GPA(Grade Point Average) : 2.0 ~ 4.0
- GPA 기반 성적 클래스 : {0.0:'A'(GPA \geq 3.5), 1.0:'B'(3.0 \leq GPA<3.5), 2.0:'C'(2.5 \leq GPA<3.0), 3.0:'D'(2.0 \leq GPA<2.5), 4.0:'F'(GPA<2.0)}

- 메타데이터

- 소유자 : Rabie El Kharoua
- DOI(Digital Object Identifier) : 10.34740/kaggle/ds/5195702
- 생성 연도 : 2024
- 크기 : 166.9KB
- URL : <https://www.kaggle.com/ds/5195702>