**요약**

데이터를 활용하여 성적 예측 모델을 개발하고자 하였다. 2000년대부터 “자녀의 성적에 영향을 미치는 요소”를 분석하고자 하는 움직임이 대두되며 가정환경과 학생의 학업성취도가 유의미한 영향 관계를 보이는 것으로 분석되었다.

해당 논문 및 기사들을 바탕으로 가정 환경을 기준으로 예측모델을 개발하여, 환경에 따른 교육 방식의 더 나은 발전을 위하여 사용할 수 있는 모델을 개발하고자 한다.

키워드

도입 : 분석 배경 및 목적

방법

논리 결과 논의

1. **데이터**
   1. **사용 데이터**

1. school – 학생의 학교 ("GP" - Gabriel Pereira / "MS" - Mousinho da Silveira)

2. sex – 학생의 성별 ("F" – 여자 / "M" - 남자)

3. age – 학생의 나이 (범위 : 15 ~ 22세)

4. address – 학생의 거주 지역 타입 ("U" – 도시 / "R" – 도시 외곽)

5. famsize – 가족 크기 ("LE3" - 3인가족 이하 / "GT3" – 3인 초과 가족)

6. Pstatus – 부모와의 동거 상태 ("T" - 동거 / "A" - 별거)

7. Medu – 어머니의 가정교육 (0 – 하지 않음, 1 - 초등교육, 2 – 중등교육, 3 – 고등교육, 4 – 고등교육 이상)

8. Fedu – 아버지의 가정교육 (0 – 하지 않음, 1 - 초등교육, 2 – 중등교육, 3 – 고등교육, 4 – 고등교육 이상)

9. Mjob – 어머니의 직업 ("teacher" : 선생님, "health": 건강 관련 직업, "services": 서비스직, "at\_home": 주부, "other": 그 외)

10. Fjob – 아버지의 직업 ("teacher" : 선생님, "health": 건강 관련 직업, "services": 서비스직, "at\_home": 주부, "other": 그 외)

11. reason – 해당학교를 선택한 이유 ("home": 자택과의 거리, "reputation": 학교 평판, "course": 학교 강의, "other": 그 외)

12. guardian – 학생의 보호자 ("mother": 어머니, "father": 아버지, "other": 그 외)

13. traveltime – 자택에서 학교까지 걸리는 시간 (numeric: 1 : 15분 이하, 2 : 15분에서 30분, 3 : 30분에서 1시간, 4 : 1시간 이상)

14. studytime – 주간 공부 시간( 1: 2시간 이하, 2 : 2시간에서 5시간, 3 : 5시간에서 10시간, 4 : 10시간 이상)

15. failures – 과거 수업 실패 횟수 (numeric: n if 1<=n<3, else 4)

16. schoolsup – 추가 교육 지원 (O/X)

17. famsup – 가족 교육 지원 (O/X)

18. paid – 과목 내 추가 유료 수업 (수학 혹은 포르투갈어) (O/X)

19. activities – 교과 외 활동 (O/X)

20. nursery – 보육원 여부 (O/X)

21. higher – 고등교육 원하는 여부 (O/X)

22. internet – 자택 내 인터넷 연결 여부 (O/X)

23. romantic – 연애 여부 (O/X)

24. famrel – 가족 곾계의 질 (1: 아주 나쁨 ~ 5: 아주 좋음)

25. freetime – 방과후 자유시간 (1: 아주 나쁨 ~ 5: 아주 좋음)

26. goout – 친구들과의 외출 (1: 아주 나쁨 ~ 5: 아주 좋음)

27. Dalc – 평일 음주량(1: 아주 나쁨 ~ 5: 아주 좋음)

28. Walc – 주말 음주량 (1: 아주 나쁨 ~ 5: 아주 좋음)

29. health – 현재 건강 상태 ( 1: 아주 나쁨 ~ 5: 아주 좋음)

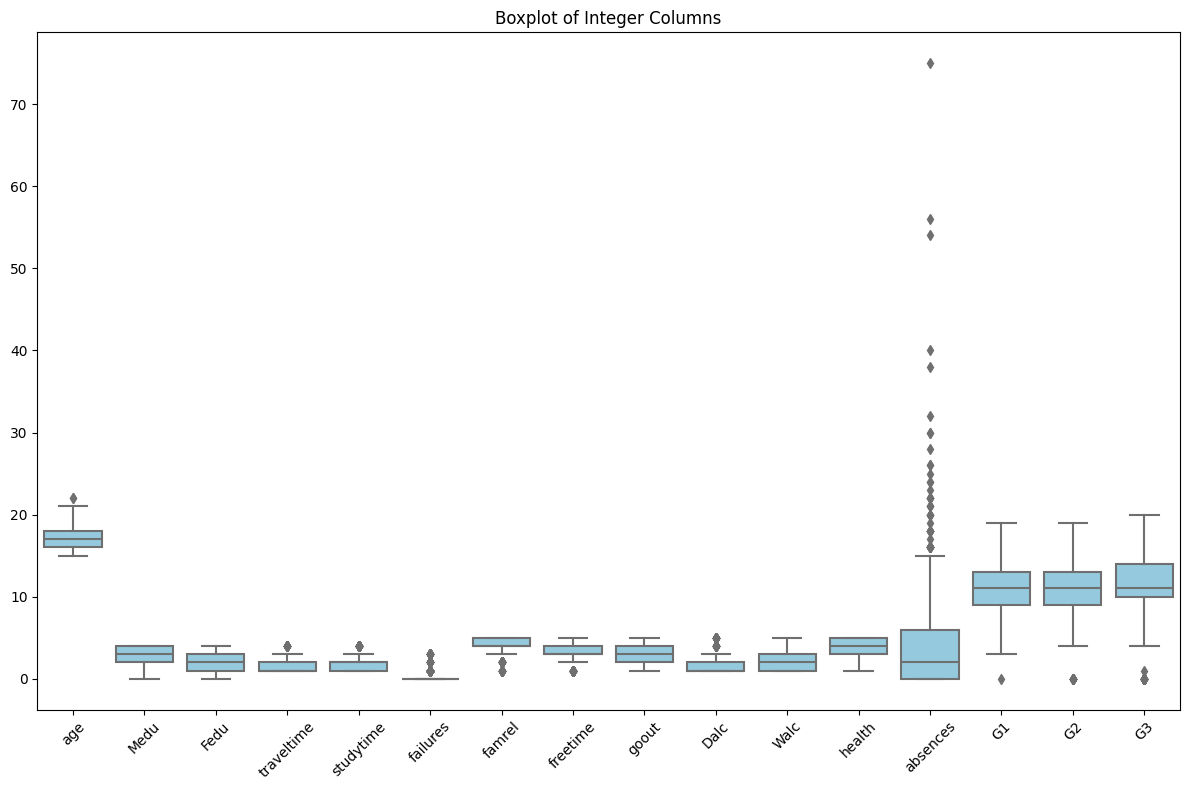
30. absences – 결석 횟수 (범위 : 0~93번)

31. G1 - 1학년 성적 (범위 : 0~20)

32. G2 – 2학년 성적 (범위 : 0~20)

* 1. **이상치 확인**

데이터 전처리 과정을 통해 이상치 처리를 수행하였다. 박스플롯을 확인하여 int64의 타입의 값들의 이상치를 판단하면 데이터를 확인 후 이상치 제거를 수행하고자 하였지만, 해당 데이터는 모든 데이터가 정상범위에 위치했으므로 이상치 없음으로 제거하지 않고 진행했다.



1. 모델링
2. 부록
3. 코드 첨부
4. 참고문헌

김현짂. (2007). 가정배경과 학교교육 그리고 사교육이 학업성취에 미치는 영향 분석. 교육행정학연구, 25(4), 485-508.

- 추상엽, 임성문. (2007). 가족구조와 고등학생의 학업성취 간의 곾계. 교육심리연구, 21(3), 497-512.

- 김현주. (2007). 학업성적에 영향을 미치는 가족배경과 사회ㆍ문화적 자원의 다중상호작용방식. 핚국사회학, 41(6), 171-203.

- 정예화, 정제영. (2017). 초·중·고등학생의 학업성취도에 영향을 미치는 요인 분석. 교육행정학연구, 35(2), 1-20.

- 방하남, 김기헌. (2002). 기회와 불평등. 핚국사회학, 36(4), 193-222.

- 장상수, 손병선. (2005). 가족배경이 학업성적에 미치는 영향. 한국사회학, 39(4), 198-230.

- 박상은, 서봉언. (2018). 대구 중학생의 고등학교 유형 선택 결정 요인. 지역사회연구, 26(3), 285-311, 10.31324/JRS.2018.09.26.3.285

별첨