

UCI 데이터셋 기반 데이터 분석 프로젝트

프로젝트 수행 내용

- 프로젝트 주제
 - 구체적인 목표 설정
- 자료 출처 명시
 - 사용한 데이터에 대한 URL 또는 출처 표시
- 프로젝트 발표 내용
 - 주제 선정 배경 설명
 - 통계 분석 프로세스 설명
 - 개인별 발표
 - 결론 도출
 - 팀별 대표자는 팀원들의 자료를 정리해서 도출된 결과 발표
- 제출물
 - 발표자료 및 개인별 소스코드 및 작업산출물

당뇨병 발병 위험 예측

- 데이터셋: Pima Indians Diabetes Database
- URL: <https://archive.ics.uci.edu/ml/datasets/pima+indians+diabetes>
- 목표: 특정 인구 집단의 당뇨병 발병 위험을 예측하는 모델을 구축하는 것.
- 설명: 피마 인디언 여성의 의료 기록 데이터를 분석하여, 연령, BMI, 혈압 등 다양한 변수와 당뇨병 발병 간의 관계를 평가합니다. 로지스틱 회귀와 같은 분류 모델을 사용하여 발병 위험이 높은 그룹을 예측하고, 정규분포를 통해 건강 지표의 분포를 시각화합니다.
- 기대 결과: 당뇨병 예방을 위한 조기 진단 도구 개발, 고위험 환자 식별.

와인 품질 예측

- 데이터셋: Wine Quality Data Set
- URL: <https://archive.ics.uci.edu/ml/datasets/wine+quality>
- 목표: 화학적 특성(예: 산도, 알코올 함량 등)을 기반으로 와인의 품질을 예측하는 모델을 구축하는 것.
- 설명: 레드 와인과 화이트 와인의 화학적 특성을 분석하여 품질 점수를 예측합니다. 회귀 분석 및 분류 모델을 사용하여 와인 품질에 영향을 미치는 주요 요인을 식별하고, 와인 품질의 분포를 시각화합니다. 이 프로젝트는 와인 생산자들이 품질 관리를 최적화하는 데 도움을 줄 수 있습니다.
- 기대 결과: 와인 품질 예측 모델 구축, 품질 향상을 위한 생산 공정 최적화.

온라인 쇼핑 고객 세분화

- 데이터셋: Online Retail Data Set
- URL: <https://archive.ics.uci.edu/ml/datasets/Online+Retail>
- 목표: 온라인 쇼핑 데이터를 기반으로 고객을 세분화하고, 각 세그먼트에 맞춤형 마케팅 전략을 제안하는 것.
- 설명: 온라인 소매 데이터를 분석하여 고객의 구매 패턴, 지출 수준, 구매 빈도 등을 기준으로 고객을 세분화합니다. K-means 클러스터링과 같은 군집화 알고리즘을 사용하여 고객 세그먼트를 정의하고, 각 그룹의 특성을 정규분포로 시각화하여 이해를 돕습니다.
- 기대 결과: 맞춤형 마케팅 전략 수립, 고객 이탈 방지 및 충성도 향상.

스마트폰 사용자의 활동 인식

- 데이터셋: Human Activity Recognition Using Smartphones Data Set
- URL:
<https://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones>
- 목표: 스마트폰 센서 데이터를 기반으로 사용자의 활동(걷기, 달리기, 계단 오르기 등)을 인식하는 모델을 개발하는 것.
- 설명: 가속도계와 자이로스코프 데이터를 분석하여 사용자가 수행 중인 활동을 분류합니다. 정규분포를 사용하여 각 활동의 센서 데이터 분포를 모델링하고, 딥러닝 또는 기계학습 모델을 활용하여 높은 정확도의 활동 인식 시스템을 구축합니다.
- 기대 결과: 웨어러블 디바이스와 통합된 활동 추적 시스템, 건강 관리 응용 프로그램 개발.

교통사고 발생 분석

- 데이터셋: Motor Vehicle Crash Data
- URL: <https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars>
- 목표: 교통사고 발생의 주요 요인을 분석하고, 사고의 빈도 및 심각도 분포를 정규분포로 모델링하는 것.
- 설명: 사고 데이터의 시각화 및 분포 분석을 통해 특정 요일, 시간대, 기상 조건에서 사고 발생 확률을 평가합니다. 정규분포를 사용하여 사고 발생 빈도를 분석하고, 사고의 심각도를 평가하여 위험이 높은 조건을 식별합니다.
- 기대 결과: 사고 예방을 위한 정책 제안, 위험 시간대 및 조건 예측.