

LUDAtics

- Version: Studio 1.1

Common

1. 데이터 업로드

- Palette -> Data -> +Add
 - path : brightics-studio\brightics-server\data\brightics@samsung.com\upload
 - .snappy.parquet.crc 같은 확장명으로 저장되어 있음

3. 리포트 생성

- Main -> Reports -> + -> Model load -> Chart add -> Publish

4. Script 모델 사용

- (function) Load > String Summary / Query Executer / Python Script

Pre-processing

1. 데이터 결합

- (function) Load > Join

2. 파생변수 생성과 데이터 저장

- (function) Load > Add Function Columns > Unload

3. 데이터 변경

- (function) Load > Add Column / Binarizer

4. 데이터 샘플링

- Load > Split Data / Split Data / Random Sampling

• Parameter(Split Data)

- Train/Test Ratio(%)
- Seed

• Parameter(Random Sampling)

- Method: Number/Fraction(%)
- Replacement: True/False 복원/비복원 추출
- Seed
- Group By

5. 불균형 데이터 샘플링

- Load > String Summary / Random Sampling - Unload / Filter > Random Sampling > Bind Row Column / Filter > Bind Row Column

- Insurance 'yes'와 'no' 개수를 맞추기 위해 무작위 샘플링

6. 결측값 처리(1)
7. 결측값 처리(2)
8. 이상값 탐지 및 제거
9. 데이터 통합 처리
16. 문자 데이터의 더미 변수 변환
17. 날짜 관련 전처리 및 파생변수 생성

Exploratory Data Analysis

1. 요약통계량 생성
2. 그룹별 평균, 배열 생성
3. 파레토 분석
6. One Sample T-test
7. Two Sample T-test
8. Paired T-test
9. ANOVA와 사후검정
10. 카이제곱검정
11. 상관관계 분석(1)
12. 상관관계 분석(2)
13. 상관관계 분석(3)
14. 결측치 제거 및 선행변수 생성과 상관분석
15. 피어슨 상관분석 및 선형회귀분석
16. 설문 통계 분석
18. 데이터 시각화

Dimension Reduction

1. 잠재변수 탐지(1)
2. 잠재변수 탐지(2)
3. 설문지 데이터 요인분석

Clustering

1. K-means 군집분석
3. 계층적 군집분석

Regression

3. 단순 선형회귀분석
4. 다중 선형회귀분석(1)
5. 다중 선형회귀분석(2)

- 6. 다중 선형회귀분석(3)
- 10. 선형 회귀 계수 및 빈도 분석

Classification

- 2. 로지스틱 회귀분석
- 3. 나이브 베이즈 분류
- 4. GLM 이항 분류 모델
- 5. Tree 기반 분류 모델 비교
- 9. 이항분류 모델 비교
- 10. 다항분류 모델 비교

Recommendation

- 1. 연관성 분석(Association Rule)

Time Series

- 4. 시계열분해(Time Series Decomposition)와 EWMA
- 5. ARIMA 예측 모델
- 8. MA, EWMA 모델 비교
- 9. ARIMA, Holt-Winters 모델 비교
- 10. 로지스틱 분포 적합

Text Analysis