

데이터 분석 개요

9회 기출

01. 데이터가 가지고 있는 특성을 파악하기 위해 해당 변수의 분포 등을 시각화하여 분석하는 분석 방식은 무엇인가?

- ① 전처리분석
- ② 탐색적자료분석(EDA)
- ③ 공간분석
- ④ 다변량분석

02. 데이터 마이닝의 모델링에 대한 설명이다. 설명이 가장 잘못된 것은?

- ① 데이터마이닝 모델링은 통계적 모델링이 아니므로 지나치게 통계적 가설이나 유의성에 집착하지 말아야 한다.
- ② 모델링 방법은 여러 가지가 있으므로 모델링 시 반드시 다양한 옵션을 줘서 모델링을 수행하여 최고의 성과를 도출하여야 한다.
- ③ 분석데이터를 학습 및 테스트 데이터로 6:4, 7:3, 8:2 비율로 상황에 맞게 실시한다.
- ④ 성능에 집착하면 분석 모델링의 주목적인 실무 적용에 반하여 시간을 낭비할 수 있으므로 훈련 및 테스트 성능에 큰 편차가 없고 예상 성능을 만족하면 중단한다.

10회 기출

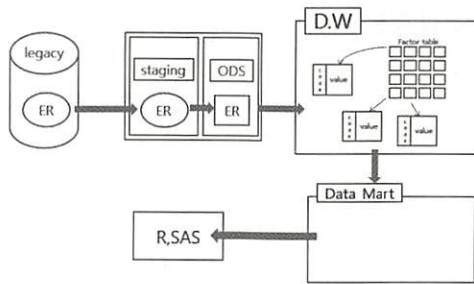
03. 모델링 성능을 평가함에 있어, 데이터마이닝에서 활용하는 평가 기준이 아닌 것은?

- ① 정확도(Accuracy)
- ② 리프트(Lift)
- ③ 디텍트 레이트(Detect Rate)
- ④ Throughput

04. 탐색적 데이터 분석의 목적은 데이터를 이해하는 것이다. 다음 중 이에 대한 설명으로 가장 부적절한 것은?

- ① 데이터에 대한 전반적인 이해를 통해 분석 가능한 데이터인지 확인하는 단계이다.
- ② 탐색적 데이터 분석 과정은 데이터에 포함된 변수의 유형이 어떻게 되는지를 찾아가는 과정이다.
- ③ 데이터를 시각화하는 것만으로는 이상점(outlier) 식별이 잘 되지 않는다.
- ④ 알고리즘이 학습을 얼마나 잘 하느냐 하는 것은 전적으로 데이터의 품질과 데이터에 담긴 정보량에 달려 있다.

05. 아래의 그림은 데이터 처리 구조를 나타내고 있다. 그림에 대한 설명으로 잘못 된 것은?



- ① 데이터를 분석에 활용하기 위해 데이터웨어하우스와 데이터마트에서 데이터를 가져 온다.
- ② 신규시스템이나 DW에 포함되지 않은 데이터는 기존 운영시스템(legacy)에서 직접 데이터를 DW와 전처리 없이 바로 결합하면 된다.
- ③ ODS는 운영데이터저장소로 기존 운영시스템의 데이터가 정제된 데이터이므로 DW나 DM과 결합하여 분석에 활용할 수 있다.
- ④ 스테이지 영역에서 가져온 데이터는 정제되어 있지 않기 때문에 데이터의 전처리를 해서 DW나 DM과 결합하여 사용한다.

06. 최근 시각화 기법의 활용이 높아지면서 데이터의 특성을 파악하는데 많은 기여를 하고 있다. 다음 중 최근의 시각화의 발전된 형태가 아닌 것은?

- ① 텍스트 마이닝에서의 워드 클라우드를 통한 그래프화
- ② SNA(social network analysis)에서 집단의 특성과 관계를 그래프화
- ③ 통계소프트웨어의 기초통계정보를 엑셀에서 그래프화
- ④ polygon, heatmap, mosaic graph 등의 그래프 작업

