

## ✓ 1. 다음 중 빅데이터 시대의 위기요인과 사례가 아닌 것은?

1/1

- ☐ 사생활 침해-개인정보를 동의 없이 수집하여 맞춤형 광고 제작
- ☐ 책임 원칙 훼손-특정 집단에 소속되어 있다는 이유로 부당한 직원 해고
- ☐ 책임 원칙 훼손-특정성향의 직원에 대한 채용거부
- ☒ 데이터 오용-상업적목적으로 데이터 크롤링 하여 개인정보 수집 ✓

의견 보내기

1-16. 빅데이터 위기요인과 통제방안

4번의 개인정보 수집은 사생활 침해에 대한 것임

- 사생활 침해: 우리를 둘러싼 정보 수집 센서들의 수가 점점 늘어나고 있고, 특정 데이터가 본래 목적 외에 가공 처리돼 2차 3차적 목적으로 활용될 가능성이 증가
- 책임 원칙 훼손: 빅데이터 기반분석과 예측 기술이 발전하면서 정확도가 증가한 만큼, 분석 대상이 되는 사람들은 예측 알고리즘의 희생양이 될 가능성이 증가함
- 데이터 오용: 주어진 데이터에 잘못된 인사이트를 얻어 비즈니스에 직접 손실을 불러 올 수 있음

## ✓ 2. 다음 중 빅데이터 분석의 특성에 대한 설명으로 옳지 않은 것은?

1/1

- ☐ 데이터가 방대하다고 무조건 더 좋은 가치를 창출하는 것은 아니다
- ☒ 데이터 크기가 커질수록 더 많은 분석을 수행하는 것이 경쟁우위 확보의 원천이다 ✓
- ☐ 분석적 방법과 성과에 대한 이해부족은 빅데이터 과제에 대한 걸림돌이다
- ☐ 비즈니스의 핵심에 더욱 객관적이고 통찰력 있는 데이터를 추출하는 것이 중요하다

의견 보내기

데이터의 크기와 분석의 수는 비례하여 수행하는 관계는 아니다.

✓ 3. 다음 중 빅데이터의 가치측정이 어려운 이유가 아닌 것은?

1/1

- ☐ 데이터 재사용이 일반화되어 특정데이터를 언제 누가 사용하였는지 알기 힘들다
- ☐ 빅데이터는 기존에 존재하지 않던 가치를 창출한다
- ☒ 빅데이터 전문 인력의 증가로 다양한 곳에서 빅데이터가 활용되고 있다 ✓
- ☐ 데이터 분석기술의 발전으로 과거에는 분석할 수 없었던 데이터도 분석할 수 있게 되었다

의견 보내기

빅데이터의 가치 산정이 어려운 이유

1. 데이터의 활용 방식: 데이터 재사용이 일반화되어 특정데이터를 언제 누가 사용하였는지 알 수 없게 되었다
2. 새로운 가치 창출: 빅데이터는 기존에 존재하지 않던 가치를 창출해낸다
3. 분석기술의 발달: 데이터분석기술의 발전으로 과거에는 분석할 수 없었던 데이터도 분석할 수 있게 되었다

✓ 4. 다음 중 분석 과제 정의서에 필수적으로 포함되야 할 항목이 아닌 것은? 1/1

- ☐ 필요 소스 데이터/ 분석 정의
- ☐ 데이터 입수 난이도
- ☐ 분석 수행주기
- ☒ 상세 분석 알고리즘 ✓

의견 보내기

2-15. 분석 프로젝트의 특징

분석 과제 정의서- 다양한 분석 과제 도출 방법을 통해 도출된 분석 과제를 정리한 문서  
작성 항목- 필요한 소스 데이터, 분석 방법, 데이터 입수 난이도, 데이터 입수 사유, 분석 수행주기, 분석결과에 대한 검증, 분석 과정 상세 등을 작성함



✓ 5. 다음 중 데이터 사이언스에 대한 설명으로 옳바르지 않은 것은?

1/1

- ☐ 데이터에서 의미 있는 정보를 찾는 학문이다
- ☒ 주로 분석 정확성에 초점을 맞춰 수행한다
- ☐ 정형 데이터 뿐만 아니라 다양한 유형의 데이터를 활용한다
- ☐ 기존의 통계학과는 달리 총체적 접근법을 사용한다



의견 보내기

1-21. 데이터 사이언스

- 데이터로부터 의미 있는 정보를 추출해내는 학문으로 통찰력 있는 분석에 초점
- 정형, 반정형, 비정형의 다양한 유형의 데이터를 대상으로 함
- 분석 뿐 아니라 이를 효과적으로 구현하고 전달하는 과정까지 포함한 포괄적 개념
- 데이터 공학, 수학, 통계학, 컴퓨터 공학, 시각화, 해커의 사고방식, 해당 분야의 전문 지식을 종합한 학문 => 총체적(holistic) 접근법을 사용함

✓ 6. DBMS에 대한 설명이 옳지 않은 것은?

1/1

- ☒ 데이터베이스 관리로 모든 데이터 문제를 해결 할 수 있다
- ☐ 우리나라에서 일반적으로 많이 사용되는 데이터베이스 관리 시스템은 관계형 DBMS다
- ☐ 사용자들이 동시에 트랜잭션 하는 경우에도 즉각적으로 결과값을 확인할 수 있다
- ☐ 계정을 가진 모든 사용자가 접근이 가능하다



의견 보내기

데이터베이스는 장애가 발생했을 때 원인, 상태 파악이 어려우며, 전체 시스템의 업무처리가 중단되는 단점을 가지고 있다.



✓ 7. 다음 중 데이터웨어하우스 활용 관련 내용으로 틀린 것은?

1/1

- ☐ 데이터웨어하우스는 저장 수단이고 데이터 마트는 소비자에게 제공하는 수단이다
- ☒ 데이터 마트는 특정 조직이나 부서의 사람들 뿐만 아니라 모든 사람이 사용할 수 있다 ✓
- ☐ 데이터 마트는 데이터웨어하우스로부터 특정 주제, 부서 중심으로 구축된 소규모 단일 주제 데이터웨어하우스이다
- ☐ 데이터웨어하우스는 다양한 운영 시스템에서 추출, 변환, 통합되고 요약된 데이터베이스이다

의견 보내기

1-09. 기업 내부 데이터베이스 솔루션

- 데이터웨어하우스: 사용자의 의사 결정에 도움을 주기 위하여, 다양한 운영 시스템에서 추출, 변환, 통합되고 요약된 데이터베이스이다
- 데이터 마트: 전사적으로 구축된 데이터웨어하우스로부터 특정 주제, 부서 중심으로 구축된 소규모 단일 주제의 데이터웨어하우스로, 대개 특정 조직 혹은 팀 등 제한된 사용자 그룹에게 서비스가 제공된다.

✓ 8. 빅데이터 전략 도출을 위한 가치기반 분석에 대한 내용으로 옳지 않은 것은? 1/1

- ☒ 일차적 분석만으로 해당 부서 및 업무에 효과가 없다 ✓
- ☐ 빅데이터는 가치창출이 가능해야 하고, 그 시점이 빠를수록 좋다
- ☐ 빅데이터의 걸림돌은 분석적 방법과 성과에 대한 이해부족이다.
- ☐ 기업의 핵심 가치와 관련해 전략적 통찰력을 가져다 주는 데이터 분석을 내재화 하는 것이 어렵다

의견 보내기

1-19. 빅데이터 분석

- 일차적 분석을 통해서도 부서나 업무 영역에서 상당한 효과를 얻을 수 있음
- 일차적 분석 경험이 증가하고 분석의 활용 범위를 더 넓고 전략적으로 변화시켜야 함



✓ 9. 데이터의 가공 및 상관관계 간 이해를 통해 패턴을 인식하고 그 의미를 부여한 데이터를 무엇이라 하는가? 1/1

정보



의견 보내기

1-04. 데이터와 정보의 관계

- 정보(Information) : 데이터의 가공 및 상관/연관 관계 속에서 의미가 도출된 것
- 지식(Knowledge) : 상호 연결된 정보 패턴을 이해하여 이를 토대로 예측한 결과물
- 지혜(Wisdom) : 근본 원리에 대한 깊은 이해를 바탕으로 도출되는 아이디어

✓ 10. 인터넷으로 연결된 기계마다 통신 장치를 갖추고 있는 환경에서 사람 또는 기계끼리 자동으로 통신하는 기술로써 사물과 사람, 사물과 사물 간의 정보를 상호 소통하는 방식을 무엇이라 하는가? 1/1

IoT



의견 보내기

- 34회(주관식) 33회(객관식) 기출에서 언급됨

IoT란, 인터넷에 연결되어 IoT 애플리케이션이나 네트워크에 연결된 장치, 또는 산업 장비 등의 다른 사물들과 데이터를 공유할 수 있는 수많은 '사물'을 말합니다. 인터넷에 연결된 장치는 내장 센서를 사용하여 데이터를 수집하고, 경우에 따라 그에 맞게 반응합니다



✓ 1. 데이터 분석과제에서 분석 프로젝트 관리에 대한 설명 중 틀린 것은? 1/1

- ☒ 분석 과제는 분석 전문가의 상상력을 요구하므로 일정을 제한하는 일정계획은 적절하지 못하다 ✓
- ☐ 분석 과제는 적용되는 알고리즘에 따라 범위가 변할 수 있어 범위관리가 중요하다.
- ☐ 분석 과제에서 다양한 데이터를 확보하는 경우가 있어 조달관리 또한 중요하다.
- ☐ 분석 과제에는 많은 위험이 있어 사전에 위험을 식별하고 대응방안을 수립해야 한다.

의견 보내기

2-17. 10 개 주제별 프로젝트 관리 체계

분석 프로젝트의 경우 관리 영역에서 일반 프로젝트와 다르게 유의해야 할 요소 존재

시간, 범위, 품질, 통합, 이해관계자, 자원, 원가, 리스크, 조달, 의사소통

시간: 데이터 분석 프로젝트는 초기에 의도했던 결과(모델)가 나오기 쉽지 않기 때문에 지속적으로 반복되어 많은 시간이 소요될 수 있다. 분석 결과에 대한 품질이 보장된다는 전제로 Time Boxing 기법으로 일정관리를 진행하는 것이 필요하다

✓ 2. 분석준비도의 분석업무파악 영역이 아닌 것은? 1/1

- ☒ 업무별 적합한 분석기법 사용 ✓
- ☐ 최적화 분석 업무
- ☐ 발생한 사실 분석 업무
- ☐ 예측 분석 업무

의견 보내기

2-25. 분석 준비도

- 분석준비도의 분석업무파악 영역: 발생한 사실 분석 업무, 예측 분석 업무, 시뮬레이션 분석 업무, 최적화 분석 업무, 분석 업무 정기적 개선

- 분석준비도의 분석기법 영역: 업무별 적합한 분석 기법 사용, 분석 업무 도입 방법론, 분석 기법 라이브러리, 분석 기법 효과성 평가, 분석 기법 정기적 개선

- 이외 분석 준비도 영역: 인력 및 조직, 분석 데이터, 분석 문화, IT 인프라(=분석 인프라)

### ✓ 3. 분석 마스터 플랜에 대한 내용으로 틀린 것은?

1/1

- ☒ 전체 과정을 순환적이고 반복적인 단계로 작성한다 ✓
- ☐ 분석 과제의 적용 범위 및 방식에 대해서도 종합적으로 고려하여 결정한다
- ☐ 일반적인 IT 프로젝트의 우선순위로는 전략적 중요도와 실행 용이성이 있다.
- ☐ 분석 마스터 플랜의 순서는 중장기 마스터 플랜 수립 - 단기적인 세부 이행계획 수립 - 과제별 우선순위 설정 순서이다.

의견 보내기

분석 마스터 플랜- 분석 마스터 플랜의 모든 단계를 반복하기보다

- 데이터수집 및 확보와 분석 데이터를 준비하는 단계를 순차적으로 진행하고,
- 모델링 단계는 반복적으로 수행하는 혼합형을 많이 적용하며,
- 이런 특성을 고려하여 세부적인 일정계획도 수립해야 한다.

분석 마스터 플랜 수립 시 고려 요소

- 우선순위 고려 요소: 전략적 중요도, ROI(투자자본수익률), 실행 용이성
- 적용 범위/방식 고려 요소: 업무 내재화 적용 수준, 분석 데이터 적용 수준, 기술 적용 수준

### ✓ 4. 하향식 접근방법에 대한 설명으로 옳지 않은 것은?

1/1

- ☐ 빠짐없이 문제를 도출하고 식별하는 것이 중요하다
- ☐ 문제를 해결함으로써 발생하는 가치에 중점을 두는 것이 중요하다
- ☐ 비즈니스 모델 캔버스를 문제 탐색 도구로 활용한다
- ☒ 분석 유스케이스 활용보다는 새로운 문제 탐색이 우선한다 ✓

의견 보내기

하향식 접근 방법

- 문제가 확실할 때 사용함, 문제가 주어지고 해법을 찾기 위해 사용함
- 문제 탐색에 있어 유스케이스 활용이 중요하다

가. 데이터 이해에서 데이터 준비 단계로 갈 수 있다  
 나. 모델링 단계에서 학습용/테스트용 데이터를 사용해  
 과소적합을 확인한다  
 다. 비즈니스 이해, 데이터 이해 간 피드백이 가능하다  
 라. 평가에서 적합하면 바로 프로젝트 투입이 가능하다

- ☐ 가, 라
- ☐ 가, 나, 라
- ☒ 가, 나, 다
- ☐ 나, 라



의견 보내기

평가단계에서는 모델이 프로젝트의 목적에 부합하는지 평가하는 것으로, 이 단계의 목적은 데이터 마이닝 결과를 수용할 것인지 판단하는 것이지 프로젝트 투입에 대한 것이 아님

✓ 6. 빅데이터 분석 방법론 시스템 구현 단계에 해당하는 것으로 짝지어진 것은? 1/1

- ☒ 설계 및 구현, 시스템 테스트 및 운영
- ☐ 모델링, 모델 평가 및 검증
- ☐ 설계 및 구현, 프로젝트 평가 보고
- ☐ 모델링, 데이터 스토어 설계



의견 보내기

2-10. 빅데이터 분석 방법론

- 분석 기획: 비즈니스 이해 및 범위 설정, 프로젝트 정의 및 계획수립, 프로젝트 위험 계획 수립
- 데이터 준비: 필요 데이터 정의, 데이터 스토어 설계, 데이터 수집 및 적합성 점검
- 데이터 분석: 분석용 데이터 준비, 텍스트 분석, 탐색적 분석, 모델링, 모델 평가 및 검증
- 시스템 구현: 설계 및 구현, 시스템 테스트 및 운영
- 평가 및 전개: 모델 발전 계획, 프로젝트 평가 보고, 평가 및 전개





✓ 7. 빅데이터 분석 방법론 분석기획시 고려사항으로 틀린 내용은?

1/1

- ☒ 타당성 단계에서 데이터 및 기술적 타당성은 복잡한 문제이기 때문에 다양한 사람들의 의견을 들어보는 것이 중요하다. ✓
- ☐ 장애요소 사전계획 수립이 필요하다
- ☐ 분석을 위한 데이터 확보 및 데이터 유형에 대한 분석이 선행되어야 한다
- ☐ 유사 분석 시나리오 및 솔루션이 있다면 이를 최대한 활용해야 한다

의견 보내기

2-04. 분석 기획 시 고려 사항

- 가용한 데이터: 데이터의 유형 분석이 선행적으로 이루어져야 함
- 적절한 유즈케이스 탐색: 유사분석 시나리오 솔루션이 있다면 이것을 최대한 활용
- 장애요소들에 대한 사전 계획 수립 필요

✓ 8. 분석 프로젝트의 관리 방안에 대한 설명 중 적절하지 않은 것은?

1/1

- ☐ 분석의 활용적인 측면에서는 정확도가 중요하며, 안정적인 측면에서는 정밀도가 중요하다.
- ☒ 모델을 지속적으로 반복했을 때 편차의 수준을 정확도라고 한다. ✓
- ☐ 난이도에 우선 순위 기준을 놓으면 시급성 높고 난이도 쉬운 과제가 가장 먼저 수행되어야 한다
- ☐ 시급성이 높고 난이도가 높은 분석과제는 경영진에 의해 조정 가능하다

의견 보내기

2-16. 분석 프로젝트의 특성 관리 영역

- Accuracy: 분석의 활용적인 측면 (모델과 실제 값의 차이)
- Precision: 분석의 안정성 측면 (모델을 반복했을 때의 편차)
- Accuracy, Precision은 트레이드 오프인 경우가 많음
- 모델의 해석 및 적용 시 사전에 고려해야 함

✓ 9. 식별된 비즈니스 문제를 데이터의 문제로 변환하여 정의하는 단계는 무엇인가? 1/1

문제 정의



의견 보내기

2-12. 하향식 접근 방식

문제 탐색 - 문제 정의 - 해결 방안 탐색 - 타당성 검토

✓ 10. 전사 차원의 모든 데이터에 대하여 정책 및 지침, 표준화, 운영조직 및 책임 등의 표준화된 관리체계를 수립하고 운영을 위한 프레임워크 (Framework) 및 저장소(Repository)를 구축하는 것을 말한다. 마스터 데이터(Master Data), 메타 데이터(Meta Data), 데이터 사전(Data Dictionary)은 ( )의 중요한 관리 대상이다. 1/1

데이터 거버넌스



✓ 1. boxplot에 대한 설명으로 적절하지 않은 것은? 1/1

- ☐ 박스플롯의 가운데 중심선은 중앙값을 의미한다.
- ☐ 다섯개 숫자(최소값, 최대값, 중앙값, 제1사분위수, 제3사분위수)를 확인할 수 있다.
- ☐ 데이터의 전체적인 분포를 확인할 수 있다.
- ☒ 이상치를 확인하는 데는 부적절하다.



의견 보내기

Boxplot은 이상치 확인에 매우 특화된 plot입니다.



✕ 2. 다음 중 아래 거래 전표에서 지지도 50%, 신뢰도 50% 이상인 규칙으로 가장 적절한 것은? .../1

물품	거래건수
A	10
B	5
C	25
A, B, C	5
B, C	20
A, B	20
A, C	15

- ☐ A -> C
- ☒ C -> A
- ☐ B -> C
- ☐ C -> B

✕

정답 없음

✓ 3. 카탈로그 배열, 교차 판매 등의 마케팅을 계획할 때 적절한 데이터마이닝 기법은? 1/1

- ☒ 연관규칙학습
- ☐ 군집분석
- ☐ 회귀분석
- ☐ 분류분석

✓

의견 보내기

연관규칙학습: 카탈로그 배열, 교차판매 등의 마케팅 계획에 사용되는 데이터마이닝 기법

✓ 4. 앙상블 설명으로 옳은 것은?

1/1

- ☒ 대표적인 기법으로 배깅과 부스팅이 있다
- ☐ Voting은 서로 같은 여러 개의 알고리즘을 사용한다
- ☐ Hard Voting 은 확률의 평균을 구해 높은 것을 선택한다
- ☐ RandomForest는 대표적인 부스팅 방식이다



의견 보내기

앙상블(Ensemble)

- 대표적인 기법으로 배깅과 부스팅이 있다
- Voting은 서로 다른 여러 개의 알고리즘을 사용한다
- Hard Voting과 Soft Voting이 있으며 Hard Voting은 다수결의 원칙에 따라 가장 많은 빈도의 것을 최종 결과로 하고, Soft Voting은 확률의 평균을 구해 높은 것을 선택한다
- RandomForest는 대표적인 배깅(bagging) 방식이다

✓ 5. apriori 알고리즘의 분석절차로 맞는 것은?

1/1

- ☐ (가)-(마)-(다)-(나)-(라)
- ☒ (가)-(나)-(다)-(라)-(마)
- ☐ (가)-(마)-(라)-(나)-(다)
- ☐ (나)-(가)-(다)-(마)-(라)



✓ 6. 다음 오분류표를 사용하여 재현율을 구한 결과는 무엇인가?

1/1

오분류표		예측값		합계
		TRUE	FALSE	
실제값	TRUE	200	300	500
	FALSE	300	200	500
합계		500	500	1000

- ☐ 0.15
- ☐ 0.3
- ☒ 0.4
- ☐ 0.55



의견 보내기

재현율(Sensitivity, Recall)

= 실제 True 인 것 중 예측도 True로 한 것의 비율

=  $TP / (TP + FN) = 200 / 500 = 0.4$

✓ 7. 다음 연관 규칙에 대한 설명 중 잘못된 것은?

1/1

- ☐ 비목적성 분석 기법으로 계산이 간편하다
- ☐ 대표적인 알고리즘으로 Aprior가 있다
- ☐ 조건 반응으로 표현되는 연관분석의 결과를 이해하기 쉽다
- ☒ 분석 품목 수가 증가해도 계산은 복잡해지지 않는다



의견 보내기

연관 규칙의 단점

- 분석 품목 수가 증가하면 분석 계산이 기하급수적으로 증가함
- 너무 세분화된 품목을 가지고 연관규칙을 찾으려면 의미 없는 분석 결과가 도출됨
- 상대적 거래량이 적으면 규칙 발견 시 제외되기 쉬움

연관 규칙의 장점

- 조건반응(if-then)으로 표현되는 연관 분석의 결과를 이해하기 쉬움
- 강력한 비목적성 분석 기법이며, 분석 계산이 간편함



✕ 8. 잔차의 정규성 검토에 대한 설명 중 옳지 않은 것은?

.../1

- ☐ Q-Q plot으로 대략적인 확인이 가능하다
- ☐ 잔차의 히스토그램이나 산점도를 그려서 정규성 문제를 검토하기도 한다
- ☐ 정규성을 검정하는 방법으로 Shapiro-Wilk test, Anderson-darling test 등을 이용할 수 있다
- ☒ 정규성을 만족하지 못할 때 상관계수가 가장 큰 값을 제거한다 ✕

정답 없음

✓ 9. 777개 대학을 대상으로 졸업률이 비용과 관련 있는가를 회귀분석한 결과 1/1 이다.

잘못 해석한 것은 무엇인가? 등록금(Outstate), 기숙사비용(Room.board), Personal(개인비용지출)

```
> summary(lm(Grad.Rate~Outstate+Room.Board + Personal, data=College))
```

Call:

```
lm(formula = Grad.Rate ~ Outstate + Room.Board + Personal, data = College)
```

Residuals:

Min	1Q	Median	3Q	Max
-47.794	-8.769	-0.124	8.312	51.737

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	41.6291341	2.5177761	16.534	< 2e-16 ***
Outstate	0.0020537	0.0001692	12.136	< 2e-16 ***
Room.Board	0.0013903	0.0006045	2.300	0.02172 *
Personal	-0.0027346	0.0007760	-3.524	0.00045 ***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.97 on 773 degrees of freedom

Multiple R-squared: 0.3415, Adjusted R-squared: 0.339

F-statistic: 133.6 on 3 and 773 DF, p-value: < 2.2e-16

- ☐ 등록금이 높아지면 졸업률이 높아진다
- ☐ 개인비용지출이 높아지면 졸업률이 낮아진다
- ☒ 비용과 졸업률 사이의 인과관계를 확인 할 수 있다. ✓
- ☐ 모든 변수의 회귀계수는 유의미하다

의견 보내기

회귀분석의 결과를 사용하여 인과관계 확인을 할 수는 없다

✓ 10. 의사결정 나무에 대한 설명 중 적절하지 않은 것은?

1/1

- ☒ 비지도 학습으로 상향식 접근 방법을 이용한다. ✓
- ☐ 구조가 단순하여 해석이 용이하다.
- ☐ 목표변수가 이산형인 경우 분류나무, 목표변수가 연속형인 경우 회귀나무가 있다.
- ☐ 부모마디보다 자식마디의 순수도가 증가하도록 분류나무를 형성해 나간다

의견 보내기

의사결정 나무

- 지도학습, 비모수적 모형
- 구조가 단순하여 해석이 용이
- 목표변수가 이산형인 분류나무, 연속형인 회귀나무가 있음
- 부모마디보다 자식마디의 순수도가 증가하도록 분류나무를 형성해 나감

✓ 11. 아래 인공지능망 활성화함수는 무엇인가?

1/1

$$\frac{1}{1 + e^{-x}}$$

- ☐ 부호함수
- ☐ 계단함수
- ☒ 시그모이드 ✓
- ☐ 소프트맥스

의견 보내기

Sigmoid 함수는  $x$  값에 따라 0 ~ 1의 값을 출력하는 비선형 함수이다.

✓ 12. 모수의 새로운 가설을 옳다고 판단할 수 있는지 표본 통계량 평가 판단 1/1  
과정을 설명하는 용어 중 적절한 것은?

- ☐ 점추정
- ☐ 구간추정
- ☒ 가설검정
- ☐ 연구가설



의견 보내기

가설검정: 귀무가설, 대립가설을 세우고, 유의수준을 설정 한 뒤, 검정통계량을 구하는 과정을 통해, 모수의 새로운 가설을 옳다고 판단할 수 있는지 표본 통계량 평가 판단하는 과정  
연구가설 = 대립가설

✓ 13. ROC Curve의 이상적 완벽 분류모형의 X축과 Y축 가장 적절한 것은? 1/1  
단(X값, Y값)

- ☐ (0, 0)
- ☒ (0, 1)
- ☐ (1, 0)
- ☐ (1, 1)



의견 보내기

X축:  $FP\ Rate (= 1 - Specificity)$  : 0 이 가장 좋음  
Y축:  $TP\ Rate (= Recall, Sensitivity)$  : 1 이 가장 좋음



✓ 14. 회귀분석 설명 중 가장 적절하지 못한 것은?

1/1

- ☒ 독립변수와 종속변수의 인과관계가 중요하다 ✓
- ☐ 성별과 같이 두 집단으로 분류된 명목형 자료여도 회귀분석에서 독립변수로 사용할 수 있다
- ☐ 잔차와 독립변수 간의 상관관계가 없다
- ☐ 결정계수는 독립변수가 종속변수의 분산을 설명할 수 있는 정도를 나타낸다.

의견 보내기

독립변수와 종속변수의 '상관관계'가 중요하다

✓ 15. 주성분 분석의 내용 중 가장 적절하지 않은 것은?

1/1

- ☐ 회귀분석의 다중공선성 문제 해결을 위해 사용한다
- ☐ 서로 상관성이 높은 변수를 선형 결합하여 변수를 축소, 해석상 구조적 문제해결을 위해 사용한다
- ☐ 다변량 데이터의 저차원 그래프 표시 및 이상치 탐색에 사용한다
- ☒ 원래 변수를 선형결합 할 때 목표변수를 고려할 필요는 없다 ✓

의견 보내기

주성분 분석

- 주성분분석은 비지도학습으로 회귀분석의 다중공선성 문제 해결을 위해 사용한다
- 차원 축소, 이상치 탐지, 자료의 그룹화에 사용할 수 있다
- 주성분분석은 목표변수를 고려해 목표변수를 잘 예측/분류할 수 있는 선형결합으로 이루어진 몇 개의 주성분을 찾아내기 위한 것이다
- 주성분 분석은 다른 통계적 분석을 하기 위한 사전 기초분석에 사용된다

✓ 16. 시계열 모형에서 ARIMA에 대한 설명으로 옳은 것은?

1/1

- ☐ ARIMA는 Auto-Regressive Improved Moving Average의 약자이다
- ☒ ARIMA(p, d, q)에서 p가 0이면 IMA(d, q)모형이다
- ☐ AR은 이동평균, MA는 자기회귀 모형이다
- ☐ ARIMA는 정상시계열 모형이다



의견 보내기

- ARIMA는 Auto-Regressive Integrated Moving Average의 약자이다
- ARIMA 모형은 비정상시계열 모형이며 차분/변환을 통해 AR, MA, ARMA 모형으로 정상화할 수 있음
- ARIMA(p, d, q)에서 p=0이면 IMA, d=0이면 ARMA, q=0이면 ARI 모형이 된다.

✓ 17. 다음 설명에 해당되는 앙상블 기법은?

1/1

"여러 개의 붓스트랩 자료를 생성하고 각 붓스트랩 자료에 예측 모형을 만든 후 결합하여 최종 예측 모형을 만드는 방법이다."

- ☒ 배깅(bagging)
- ☐ voting
- ☐ 부스팅(boosting)
- ☐ stacking



의견 보내기

3-83. 앙상블(Ensemble) 모형

- voting: 서로 다른 여러 개의 모형을 생성하고 결과를 집계하여 많은 표를 받은 것을 답으로 하는 방식
- 부스팅(boosting): 순차적인 학습, 붓스트랩 표본을 구성하는 재표본 과정에서 분류가 잘 못된 데이터에 더 큰 가중치를 주어 표본을 추출하는 기법
- 스택킹(stack): 두 단계의 학습을 사용하는 방식으로 서로 다른 여러 모형의 예측 결과를 다시 학습데이터로 하는 모형을 사용함



✓ 18. 시계열 데이터의 정상성 관련 내용으로 틀린 것은?

1/1

- ☐ 모든 시점에 대해 일정한 평균을 가진다.
- ☐ 공분산은 단지 시차에만 의존하고 시점 자체에는 의존하지 않는다.
- ☐ 모든 시점에 대해 일정한 분산을 가진다
- ☒ 지수평활법은 최근 데이터를 이용하여 미래를 예측하는 방법이다 ✓

의견 보내기

지수평활법

- 전체 시계열 자료를 이용하여 평균을 구하고, 최근 시계열에 더 큰 가중치를 적용하는 방법
- 지수 평활을 사용하여 얻은 예측값은 과거 관측값의 가중평균(weighted average)
- 여기에서 과거 관측값은 오래될 수록 지수적으로 감소하는 가중치를 갖음

✓ 19. 다음 64개 변수에 대한 주성분분석의 Scree plot 결과에 대한 해석으로 1/1  
틀린 것은?

"점 64개 1번째 점이 0.12에 근접, 2번째 점이 0.07 근접, 7, 8번째에서 기울기 감소, 나머지 60까지 점은 밀집되게 나열"

- ☐ 첫 번째 주성분으로 10%이상 설명이 가능하다
- ☐ 주성분 64개로 100% 분산을 설명할 수 있다
- ☒ 두 개의 주성분으로 7% 분산을 설명한다 ✓
- ☐ 7, 8번째 성분에서부터 완만한 기울기를 가진다

의견 보내기

Scree Plot은 x 축을 주성분개수, y 축을 고윳값(설명가능한 분산값)으로 하는 그래프이며, 두 개의 주성분을 사용할 경우 두 주성분의 분산을 더해서 설명력을 구할 수 있다

- ☐ 인스턴스 러닝기법이다.
- ☒ K값이 클수록 과대적합(overfitting) 문제가 발생한다.
- ☐ 가까운 것으로 군집하는 것이다
- ☐ K는 가까운 이웃의 개수를 의미한다



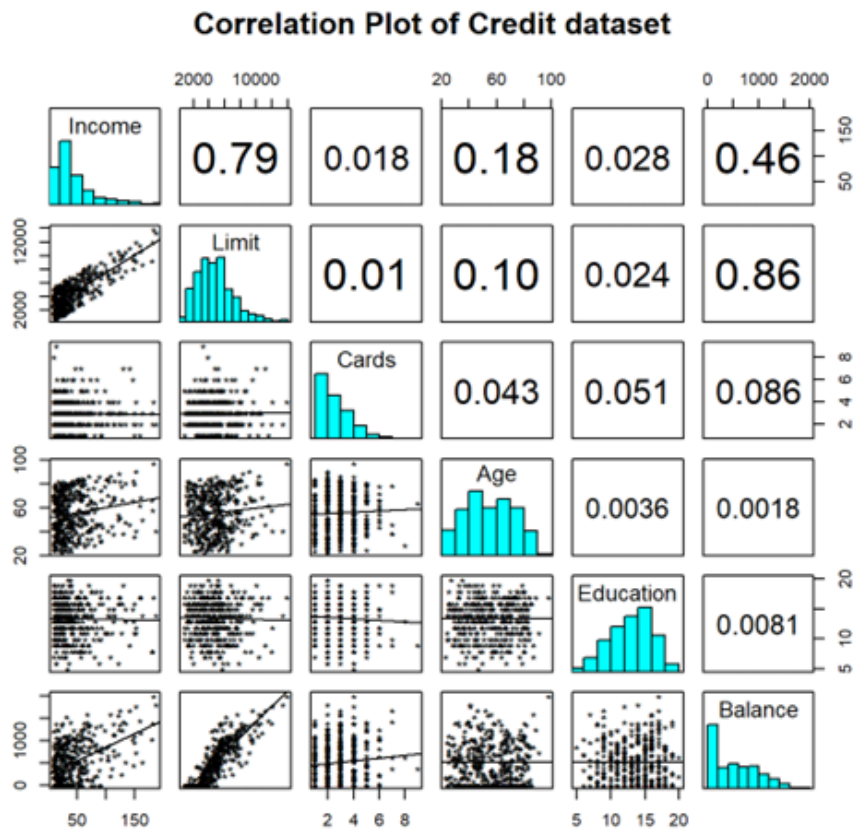
의견 보내기

*KNN 알고리즘*

*K가 커지면 커질 수록 과소적합, 작을수록 세부적으로 분류할 수 있는 과대적합을 만들 수 있다.*



- ✓ 21. Credit 데이터는 400명의 신용카드 고객에 대한 신용카드와 관련된 변수들이 포함되어 있다. 아래 변수 간의 산점도와 피어슨 상관계수를 나타내고 있다. 그림에 대한 설명 중 가장 부적절한 것은?



- ☐ 가장 상관관계가 높은 두 변수는 Limit과 Balance이다
- ☐ 모든 상관관계가 양의 상관관계를 보인다
- ☒ Balance와 가장 상관관계가 높은 변수는 Income이다
- ☐ Age와 Balance는 매우 낮은 상관관계를 보인다



의견 보내기

Balance와 가장 상관관계가 높은 변수는 Limit 이다

✓ 22. 군집분석에 대한 설명으로 잘못된 것은?

1/1

- ☒ 형성된 군집에 대해 논리성보다 안정성이 더 중요하다 ✓
- ☐ 비지도학습으로 군집간 분산 최대화, 군집내 분산을 최소화 한다
- ☐ 집단별 특성이 유사할 경우 안정성이 높다
- ☐ 유사성을 이용하여 몇 개의 집단으로 그룹화하는 분석이다

의견 보내기

군집분석

- 유사성을 이용하여 몇 개의 집단으로 그룹화하는 분석이다
- 안정성은 일부 입력값이 변경되었을 때 군집의 변화가 유의하게 변하는지에 대한 개념이다
- 집단별 특성이 유사할 경우 안정성이 높다
- 군집분석에 있어 군집 타당성 검증을 위해 논리성과 안정성 모두가 중요한 부분이다.

✓ 23. 회귀 모형을 평가하는 방법으로 옳지 않은 것은?

1/1

- ☐ 모형이 통계적으로 유의미한가?
- ☐ 모형이 데이터를 잘 적합하는가?
- ☐ 회귀계수가 유의한가?
- ☒ 선형성, 정상성, 독립성을 만족하는가? ✓

의견 보내기

- 선형성, 정상성, 독립성은 데이터 탐색 단계에서 설명변수와 종속변수간 선형성이 있는지, 잔차가 정상성, 독립성을 갖는지 평가
- 회귀 모형 평가: 모형의 통계적 유의미, 모형의 데이터 적합, 회귀계수의 유의미성

✓ 24. 상관분석에 대한 설명으로 옳은 것은?

1/1

- ☐ 스피어만 상관계수는 선형관계 파악만 가능하다
- ☐ 피어슨 상관계수는 서열척도에서 사용한다
- ☒ 피어슨 상관계수가 0이면 상관관계가 없다
- ☐ 공분산은 측정단위에 영향을 받지 않는다.



의견 보내기

- 스피어만 상관계수는 비선형 관계의 파악도 할 수 있다
- 피어슨 상관계수는 비율척도에서 사용하며, 상관계수가 0이면 서로 선형관계가 없음을 의미한다
- 공분산은 측정단위에 영향을 받으며, 상관계수는 측정단위에 영향을 받지 않는다

✓ 25. 데이터 집합으로부터 크기가 같은 표본을 여러 번 단순 임의 복원 추출 하여 각 표본에 대해 분류기를 생성한 후 그 결과를 앙상블 하는 방법으로 노드 내 데이터를 자식 노드로 나누는 기준을 정할 때 설명변수의 일부분만을 고려함으로 성능을 높이는 방법을 사용하는 것을 무엇이라 하는가?

1/1

랜덤 포레스트



의견 보내기

랜덤 포레스트는 배깅(Bagging)에 랜덤 과정을 추가한 방법으로 노드 내 데이터를 자식 노드로 나누는 기준을 정할 때 설명변수의 일부분만을 고려함으로 성능을 높이는 방법을 사용한다.

✓ 26. 여러 개의 앙상블모형 중 붓스트랩 표본을 구성하는 재표본 과정에서 분류가 잘못된 데이터에 더 큰 가중치를 주어 표본을 추출하는 기법은?

1/1

부스팅



- ✓ 27. 다음이 설명하는 것은 무엇인가? 1/1  
"클러스터 안의 데이터들이 다른 클러스터와 비교해 얼마나 비슷한가를 나타내는 군집분석 타당성 평가지표로 1에 가까울수록 군집화가 잘 되었다고 판단한다."

실루엣 계수



- ✓ 28. 인공신경망 관련 알고리즘 중 다음 설명은 어떤 것을 말하는가? 1/1  
"지도학습에서 신경망을 학습시키는 방법으로, 출력층에서 제시한 값에 대해, 실제 원하는 값으로 학습하는 방법으로 사용되고, 동일 입력층에 대해 원하는 값이 출력되도록 개개의 weight를 조정하는 방법으로 사용되는 알고리즘이다. "

역전파 알고리즘



- ✓ 29. 모집단에서 표본을 추출하여 모집단의 모수를 추측하는 과정에서 범위가 아니라 가장 참값이라고 여겨지는 하나의 모수를 택하는 것을 무엇이라고 하는가? 1/1

점추정



- ✓ 30. CART에서 쓰이는 불순도 측정 지표로 데이터의 통계적 분산정도를 정량화해서 표현한 값은 무엇인가? 1/1

지니 지수



이 콘텐츠는 Google이 만들거나 승인하지 않았습니다. - [서비스 약관](#) - [개인정보처리방침](#)

Google 설문지















