

2022년 32회 기출복원 인쇄용 (50문제)

총점 47/49 ?

✓ 1. 다음 SQL의 명령어 중 DML이 아닌 것은 무엇인가?

1/1

가 - SELECT 나 - UPDATE 다 - INSERT 라 - DELETE 마 - CREATE

☐ 가, 나

☐ 다

☐ 라

☒ 마



의견 보내기

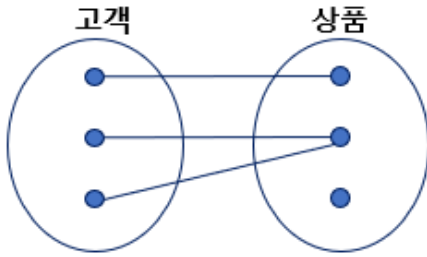
DML(Data Manipulation Language) – SELECT, UPDATE, INSERT, DELETE

DDL(Data Definition Language) – CREATE, DROP, ALTER

DCL(Data Control Language) – GRANT, REVOKE, COMMIT, ROLLBACK



- ✓ 2. 고객과 상품의 대응관계를 도식화 한 것이다. 대응비 관점에서 고객과 상품 간의 관계가 옳은 것은?



- ☐ 1:1
- ☒ N:1
- ☐ N:M
- ☐ 1:N



의견 보내기

1:1 => 하나의 개체가 하나의 개체에 대응
1:N => 하나의 개체가 여러 개체에 대응
N:1 => 여러 개체가 하나의 개체에 대응
N:M => 여러 개체가 여러 개체에 대응

제조업을 포함한 다양한 비즈니스 분야에서 생산, 구매, 재고, 주문, 공급자와의 거래, 고객 서비스 제공 등 주요 프로세스 관리를 돕는 여러 모듈로 구성된 통합 애플리케이션

- ☒ ERP
- ☐ CRM
- ☐ SCM
- ☐ KMS



의견 보내기

- Enterprise Resource Planning, 제조업을 포함한 다양한 비즈니스 분야에서 생산, 구매, 재고, 주문, 공급자와의 거래, 고객 서비스 제공 등 주요 프로세스 관리를 돕는 여러 모듈로 구성된 통합 애플리케이션 소프트웨어 패키지
- CRM : Customer Relationship Management, 고객별 구매 이력 데이터베이스를 분석하여 고객에 대한 이해를 돕고 이를 바탕으로 각종 마케팅 전략을 통해 보다 높은 이익을 창출할 수 있는 솔루션
- SCM : Supply Chain Management, 기업이 외부 공급업체 또는 제휴업체와 통합된 정보시스템으로 연계하여 시간과 비용을 최적화 시키기 위한 것으로, 자재구매 데이터, 생산, 재고 데이터, 유통/판매 데이터, 고객 데이터로 구성됨
- Knowledge Management System, 지식관리시스템의 약자, 조직 내의 지식을 체계적으로 관리하는 시스템을 의미

- ☐ 데이터, 인프라, 기술
- ☒ 데이터, 인력, 기술
- ☐ 인프라, 프로세스, 인력
- ☐ 프로세스, 인력, 기술



의견 보내기

빅데이터 활용 요소 3가지: 데이터, 인력, 기술



✓ 5. 다음 중 딥러닝의 종류가 아닌 것은?

1/1

- ☒ SVM
- ☐ DNN
- ☐ CNN
- ☐ RNN



의견 보내기

딥러닝의 종류: ANN, DNN, CNN, RNN

SVM: 지도학습 모델 (회귀, 분류 모델이 있음)

✓ 6. 지도 학습은 입력과 출력 모두 주어진 상태에서의 학습이며, 비지도 학습 1/1
은 입력만 주어진 상태에서의 학습이다. 다음의 보기에서 그 종류가 다른
것은 무엇인가?

가. 회귀 분석 나. 군집 분석 다. 감정 분석 라. 분류 분석

- ☐ 가
- ☒ 나
- ☐ 다
- ☐ 라



의견 보내기

회귀, 분류, 감정분석은 지도 학습
군집 분석은 비지도 학습



✓ 7. 빅데이터와 인문학 열풍과 관련된 내용으로 상관 없는 것은?

1/1

- ☐ 단순 세계에서 복잡한 세계로의 변화
- ☐ 비즈니스의 중심이 제품생산에서 서비스로 이동
- ☒ 빅데이터 분석 기법 및 방법론의 확대
- ☐ 경제와 산업의 논리가 생산에서 시장창조로 바뀜



의견 보내기

단순 세계에서 복잡한 세계로의 변화: 다양성과 각 사회의 정체성, 연결성, 창조성 키워드 대두
비즈니스의 중심이 제품생산에서 서비스로 이동: 고객에게 얼마나 뛰어난 서비스를 제공 여부가 관건
경제와 산업의 논리가 생산에서 시장창조로 바뀜: 무형자산이 중요

✓ 8. 데이터 사이언티스트의 역량에 대한 설명이 아닌 것은?

1/1

- ☐ 통찰력 있는 분석, Data Technical Skill
- ☐ 설득력 있는 전달
- ☐ 다분야간 협력
- ☒ 데이터 분석 알고리즘으로 인해 피해를 입은 사람을 구제함



의견 보내기

1-23. 데이터 사이언티스트의 역량
데이터 사이언티스트들은 하드 스킬과 소프트 스킬 능력을 동시에 갖추고 있어야 한다
하드스킬: Machine Learning, Modeling, Data Technical Skill
소프트 스킬: 통찰력 있는 분석, 설득력 있는 전달, 다분야간 협력
데이터 분석 알고리즘으로 인해 피해를 입은 사람을 구제하는 것은 "알고리즘미스트"의 역할이다

✓ 9. 다음은 무엇에 대한 설명인가?

구글의 인터넷에 연결된 네스트는 날씨와 기온 정보 그리고 집주인의 평소 온도 설정 DATA를 기반으로 사용자의 Context를 인식해 자동으로 온도를 설정해 주며, 아마존은 Dash라는 작은 장치는 WiFi가 내장된 바코드 인식기로 상품에 인쇄된 바코드를 Dash로 비추게 되면 그 상품을 아마존 장바구니에 저장할 수 있도록 해준다. 나이키의 경우 애플과 제휴하여 스마트한 운동 관리를 할 수 있도록 해주는 서비스로 자리 매김했다. 굳이 우리가 기계를 조작하지 않아도 모든 것이 사람을 위해 알아서 자동으로 돌아가는 세상이 이것이 보여줄 미래이다

IoT



의견 보내기

IoT(Internet of Things)

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.

✓ 10. 다음은 어떤 기업내부데이터 솔루션에 대한 설명인가?

1/1

제조, 물류, 유통업체 등 유통공급망에 참여하는 모든 업체들이 협력을 바탕으로 정보 기술 (Information Technology)을 활용, 재고를 최적화하기 위한 솔루션이다

SCM



의견 보내기

SCM

기업이 외부 공급업체 또는 제휴업체와 통합된 정보시스템으로 연계하여 시간과 비용을 최적화 시키기 위한 것으로, 자재구매 데이터, 생산, 재고 데이터, 유통/판매 데이터, 고객 데이터로 구성됨



✓ 1. 분석 기회 발굴의 범위 확장에서 '거시적 관점'이 아닌 것은?

1/1

- ☐ 사회, 기술
- ☐ 경제
- ☐ 환경, 정치
- ☒ 고객



의견 보내기

분석 기회 발굴의 범위 확장

거시적 관점의 요인: STEEP - 사회, 기술, 경제, 환경, 정치 영역
경쟁자 확대 관점: 대체재 영역, 경쟁자 영역, 신규진입자 영역
시장의 니즈 탐색: 고객(소비자) 영역, 채널 영역, 영향자들 영역
역량의 재해석 관점: 내부역량 영역, 파트너 네트워크 영역

✓ 2. ROI의 4V 중 Return에 해당하는 것은?

1/1

- ☐ Volume
- ☐ Variety
- ☐ Velocity
- ☒ Value



의견 보내기

Value (Return): 비즈니스 효과 요소

Volume, Variety, Velocity: 투자비용 요소

- ☐ 메타데이터와 데이터 사전의 관리 원칙을 수립한다
- ☒ 데이터 표준 용어 설정, 명명 규칙 수립, 메타데이터 구축, 데이터 사전 구축 등의 업무로 구성된다 ✓
- ☐ 메타데이터 및 표준 데이터를 관리하기 위한 전사 차원의 저장소를 구성한다
- ☐ 데이터 거버넌스 체계를 구축한 후 표준 준수 여부를 주기적으로 점검하고 모니터링을 실시한다

의견 보내기

2-28. 데이터 거버넌스 체계 수립

데이터 표준화: 데이터 표준용어 설정, 명명규칙 수립, 메타 데이터 구축, 데이터 사전 구축

데이터 관리체계: 메타데이터와 데이터 사전(Data Dictionary)의 관리 원칙 수립

데이터 저장소관리: 메타데이터 및 표준 데이터를 관리하기 위한 전사 차원의 저장소를 구성

표준화 활동: 데이터 거버넌스 체계 구축 후, 표준 준수 여부를 주기적으로 점검, 모니터링

✓ 4. 빅데이터 분석 방법론의 분석기획 단계의 Task 중에서 '비즈니스 이해 및 1/1 범위설정' 단계에서 프로젝트에 참여하는 관계자들의 이해를 일치시키기 위해 작성하는 것은 ?

- ☒ SOW(Statement of Work) ✓
- ☐ WBS(Work Breakdown Structure)
- ☐ EDA(Exploratory Data Analysis)
- ☐ ETL(Extract Transformation Loading)

의견 보내기

SOW(Statement of Work): 비즈니스 이해 및 범위 설정 중 프로젝트 범위 설정에 있어 프로젝트에 참여하는 관계자들의 이해를 일치시키기 위하여 작성하는 구조화된 프로젝트 범위 정의서

WBS(Work Breakdown Structure): 프로젝트 수행 계획 수립 단계에서 전체 업무를 분류하여 구성 요소로 만든 후 각 요소를 평가하고 일정별로 계획하며 그것을 완수할 수 있는 사람에게 할당해주는 역할을 함

ETL(Extract Transformation Loading): 데이터 수집 및 정합성 점검 단계에서 다양한 데이터를 취합해 데이터를 추출하고 하나의 공통된 포맷으로 변환해 데이터웨어 하우스나 데이터 마트 등에 적재하는 과정을 지원하는 도구

EDA(Exploratory Data Analysis): 탐색적 데이터 분석, 이를 통해 자료에 대한 충분한 이해를 한 후에 모형 적합 등의 좀 더 정교한 모형 개발을 할 수 있음



✓ 5. 기업에서 사용하는 데이터의 가용성, 유용성, 통합성, 보안성을 관리하기 위한 정책과 프로세스를 다루며 프라이버시, 보안성, 데이터품질, 관리규정 준수를 강조하는 것을 무엇이라고 하는가? 1/1

- ☐ 데이터 표준화 활동
- ☐ 분석 마스터 플랜
- ☒ 데이터 거버넌스
- ☐ 메타 데이터



의견 보내기

분석 마스터 플랜: 일반적인 ISP 방법론을 활용하되 데이터 분석 기획의 특성을 고려하여 수행하고 기업에서 필요한 데이터 분석 과제를 빠짐없이 도출한 후 과제의 우선순위를 결정하고 단기 및 중·장기로 나누어 계획을 수립하는 것

데이터 표준화 단계: 데이터 표준용어 설정, 명명규칙 수립, 메타 데이터 구축, 데이터 사전 구축

메타 데이터: 데이터에 관한 구조화된 데이터로, 다른 데이터를 설명해 주는 데이터이며, 대량의 정보 가운데에서 찾고 있는 정보를 효율적으로 찾아내서 이용하기 위해 일정한 규칙에 따라 콘텐츠에 대하여 부여되는 데이터

✓ 6. 빅데이터 분석 방법론의 순서로 올바른 것은?

1/1

- ☒ 분석 기획 - 데이터 준비 - 데이터 분석 - 시스템 구현 - 평가 및 전개
- ☐ 데이터 준비 - 분석 기획 - 데이터 분석 - 시스템 구현 - 평가 및 전개
- ☐ 데이터 준비 - 분석 기획 - 데이터 분석 - 평가 및 전개 - 시스템 구현
- ☐ 분석 기획 - 데이터 준비 - 데이터 분석 - 평가 및 전개 - 시스템 구현



의견 보내기

2-10. 빅데이터 분석 방법론

분석기획- 데이터 준비- 데이터 분석- 시스템 구현- 평가 및 전개

✓ 7. 하향식 접근방식에 대한 설명으로 틀린 것은?

1/1

- ☐ 하향식 접근 방식의 데이터 분석 기획의 단계는 문제 탐색, 문제 정의, 해결 방안 탐색, 타당성 검토 순서이다
- ☐ 분석 유즈 케이스는 향후 데이터 분석 문제로의 전환 및 적합성 평가에 활용하도록 한다
- ☒ 하향식 접근 방식은 유사/동종 사례 벤치마킹을 통한 분석 기회 발굴을 하는 내부 참조 모델 기반 문제 탐색을 한다 ✓
- ☐ 분석 유즈 케이스는 풀어야 할 문제에 대한 상세 설명 및 해당 문제를 해결했을 때 발생하는 효과를 명시한다

의견 보내기

하향식 접근방식의 문제 탐색 단계

- 외부참조 모델 기반 문제 탐색

유사/동종 사례 벤치마킹을 통한 분석 기회 발굴

제공되는 산업별, 업무 서비스별 분석 테마 후보 그룹을 통해 Quick & Easy 방식으로 필요한 분석 기회가 무엇인지에 대한 아이디어를 얻고 기업에 적용할 분석 테마 후보 목록을 빠르게 도출

- 분석 유즈 케이스

풀어야 할 문제에 대한 상세 설명 및 해당 문제를 해결했을 때 발생하는 효과를 명시
향후 데이터 분석 문제로의 전환 및 적합성 평가에 활용하도록 함

✓ 8. 분석 준비도 프레임워크에 대한 내용이 아닌 것은?

1/1

- ☐ 분석 인력 및 조직, 분석 인프라
- ☐ 분석 업무 파악, 분석 데이터
- ☒ 분석 성과, 분석 문화 ✓
- ☐ 분석 인력 및 조직, 분석 문화

의견 보내기

분석 준비도 프레임워크

- 분석 인력 및 조직, 분석 인프라(=IT 인프라), 분석 업무 파악, 분석 데이터, 분석 문화

✓ 9. 다음 분석 성숙도 단계 중 괄호에 들어갈 단어는 무엇인가?

1/1

도입 단계 - () 단계 - 확산 단계 - 최적화 단계

활용



의견 보내기

분석 성숙도 단계

도입 단계 - 활용 단계 - 확산 단계 - 최적화 단계

✓ 10. 합리적인 의사결정을 방해하는 요소로써 문제의 표현 방식에 따라 동일 1/1
한 사건이나 상황임에도 불구하고 사람들의 선택이나 판단이 달라지는 현상
을 무엇이라 하는가?

프레이밍 효과



의견 보내기

2-06. 분석 방법론 개요

기업의 합리적 의사결정 장애요소: 고정관념, 편향된 생각, 프레이밍 효과(Framing Effect)

프레이밍 효과(Framing Effect)

동일한 사건이나 상황임에도 불구하고 사람들의 선택이나 판단이 달라지는 현상으로, 특
정 사안을 어떤 시각으로 바라 보느냐에 따라 해석이 달라진다는 이론



✖ 1. 다음 중 R에서 서로 다른 데이터 타입을 담을 수 있는 구조는 무엇인가? 0/1

- ☐ matrix
- ☐ list
- ☐ array
- ☒ vector

✖

정답

- ☒ list

의견 보내기

단일 데이터 타입 구조 : vector, matrix, array, factor
복합 데이터 타입 구조 : list, data.frame

✔ 2. 박스플롯(boxplot)에서 중앙에 있는 값 50%를 무엇이라고 하는가? 1/1

- ☐ 범위
- ☐ 1 사분위수 범위
- ☒ IQR(사분위수 범위)
- ☐ 3 사분위수 범위

✔

의견 보내기

기술통계학에서 사분위수 범위는 간단히 IQR(InterQuartile Range)로 표시하며 확률분포, 또는 자료값의 산포도를 나타내는 측도의 하나이다. 확률분포 또는 자료의 가운데 50%가 포함되는 구간의 길이로 정의한다

✓ 3. 박스플롯(boxplot)에서 중간에 있는 선은 무엇인가?

1/1

- ☒ median
- ☐ mean
- ☐ IQR
- ☐ outlier



의견 보내기

박스플롯의 중간에 있는 선은 median으로 중앙값, 50%의 위치를 의미한다.

✓ 4. Boxplot에서 상한(최댓값)과 하한(최솟값)은 얼마인가?

Q1(1사분위 1/1 수) = 4, Q3(3사분위수) = 12

- ☒ 하한 = -8, 상한 = 24
- ☐ 하한 = -6, 상한 = 22
- ☐ 하한 = -4, 상한 = 20
- ☐ 하한 = -2, 상한 = 18



의견 보내기

$$IQR = 3\text{사분위수} - 1\text{사분위수} = 12 - 4 = 8$$

$$\text{상한: } Q3 + 1.5 * IQR = 12 + 1.5 * 8 = 24$$

$$\text{하한: } Q1 - 1.5 * IQR = 4 - 1.5 * 8 = -8$$



✓ 5. 1종 오류의 최대 허용 한계에 관한 확률을 나타내는 것은?

1/1

- ☒ 유의수준
- ☐ 기각역
- ☐ 검정통계량
- ☐ 2종 오류



의견 보내기

유의수준(α): 제 1종 오류의 최대 허용 한계

기각역: 귀무가설을 기각하고 대립(=연구)가설을 채택하게 되는 영역

검정통계량: 귀무가설을 채택할지 기각할지를 결정하는 판단 기준이 되는 통계량

2종 오류: 귀무가설이 거짓인데 채택한 오류

✓ 6. 표본조사에 대한 설명이 부적절한 것은?

1/1

- ☐ 표본 오차(Sampling Error)는 모집단의 일부인 표본에서 얻은 자료를 통해 모집단 전체의 특성을 추론함으로써 생기는 오차를 의미한다.
- ☐ 표본 편의(Sampling Bias)는 표본추출방법에서 기인하는 오차를 의미한다.
- ☐ 표본 편의는 확률화(Randomization)에 의해 최소화하거나 없앨 수 있다.
- ☒ 비표본 오차(non-sampling error)는 표본크기가 증가함에 따라 감소한다.



의견 보내기

표본 추출 시 발생하는 오차

표본 오차: 모집단의 일부인 표본에서 얻은 자료를 통해 모집단 전체의 특성을 추론함으로써 생기는 오차

비표본 오차: 표본 오차를 제외한 조사의 전체과정에서 발생할 수 있는 모든 오차로 표본의 크기에 비례하여 커지므로 표본의 크기가 크다고 반드시 좋은 것만은 아니다.

표본 편의: 비표본 오차의 한가지 종류로 모수를 크게 또는 작게 추정하는 것과 같이 표본추출방법에서 기인하는 오차를 의미하며, 확률화에 의해 최소화하거나 없앨 수 있다.

✓ 7. 다음 중 비모수 검정의 종류가 아닌 것은?

1/1

- ☐ Runs Test
- ☒ 카이제곱검정
- ☐ Wilcoxon Signed Rank Test
- ☐ Sign Test



의견 보내기

카이제곱 검정은 모수 검정, 비모수 검정이 모두 존재함

모수 검정: 모분산에 대한 검정

비모수 검정: 적합도 검정, 독립성, 동질성 검정

✓ 8. 파생 변수에 대한 설명으로 옳지 않은 것은?

1/1

- ☐ 요약 통계량을 이용한 변수 생성을 통해 데이터 분석이 가능하다
- ☐ 데이터에 포함된 나이 연속형 변수를 이용해 연령대 범주형 파생변수를 추가했다
- ☒ 파생 변수를 해당 데이터셋을 대표하는 변수로 사용할 수 있다
- ☐ 파생 변수는 매우 주관적일 수 있으므로 논리적 타당성을 갖추어 개발해야 한다



의견 보내기

파생 변수: 기존 변수를 조합하거나, 특정 조건 혹은 함수 등을 적용하여 새롭게 만든 변수
상관관계가 있는 변수들끼리 결합하여 분산을 극대화하는 변수로 만들어 변수를 축약하여
희생되는 정보를 최소화 함

데이터에 포함된 나이 연속형 변수를 이용해 연령대 범주형 변수 생성(단위 변환)

요약 통계량을 이용한 변수 생성(요약 통계량)

파생 변수는 매우 주관적일 수 있으므로 논리적 타당성을 갖추어 개발해야 함



공분산행렬 또는 상관계수 행렬을 사용해 모든 변수들을 가장 잘 설명하는 변수를 찾는 방법으로, 상관계수가 있는 변수들을 선형 결합에 의해 상관계수가 없는 새로운 변수를 만들고 분산을 극대화 하는 변수로 축약하는 방법으로 새로운 변수들은 변수들의 선형결합으로 이루어져 있다

- ☐ 요인 분석
- ☐ 회귀 분석
- ☒ 주성분 분석(PCA) ✓
- ☐ 다차원척도법

의견 보내기

주성분분석(PCA, Principal Component Analysis)

공분산행렬 또는 상관계수 행렬을 사용해 모든 변수들을 가장 잘 설명하는 주성분을 찾는 방법

상관계수가 있는 변수들을 선형 결합에 의해 상관계수가 없는 새로운 변수(주성분)를 만들고 분산을 극대화하는 변수로 축약함

주성분은 변수들의 선형결합으로 이루어져 있음

독립변수들과 주성분과의 거리인 '정보손실량'을 최소화하거나 분산을 최대화 함

- ☒ 가장 분산이 적은 것을 제1주성분으로 설정한다. ✓
- ☐ 주성분 분석은 상관계수가 있는 변수들을 결합해 상관계수가 없는 변수로 분산을 극대화하는 변수로 선형결합을 해 변수를 축약하는데 사용하는 방법이다
- ☐ 공분산 행렬은 변수의 특정단위를 그대로 반영한 것이고, 상관행렬은 모든 변수의 측정단위를 표준화한 것이다
- ☐ 공분산행렬을 이용한 분석의 경우 변수들의 측정 단위에 민감하다

의견 보내기

주성분 분석은 가장 분산이 큰 것을 제 1주성분으로 설정한다



✓ 11. 시간의 흐름에 따라 관측된 데이터를 무엇이라 하는가?

1/1

- ☐ 주성분분석
- ☐ 회귀분석
- ☒ 시계열 자료
- ☐ 군집 분석



의견 보내기

시간의 흐름에 따라 관측된 데이터는 '시계열 자료' 이다

✓ 12. 다음 시계열 데이터에 대한 설명 중 가장 옳지 않은 것은?

1/1

- ☐ 시계열 데이터의 모델링은 다른 분석 모형과 같이 탐색 목적과 예측 목적으로 나눌 수 있다.
- ☐ 짧은 기간 동안의 주기적인 패턴을 계절변동이라 한다.
- ☒ 잡음은 무작위적인 변동이지만 일반적으로 원인은 알려져 있다.
- ☐ 시계열분석의 주목적은 외부인자와 관련해 계절적인 패턴 추세와 같은 요소를 설명할 수 있는 모델을 결정하는 것이다.



의견 보내기

잡음은 무작위적인 변동이며 일반적으로 원인은 알려져 있지 않다.

✓ 13. 다음 시계열 자료의 정상성(stationary)에 대한 설명 중 가장 부적절한 것은? 1/1

- ☐ 모든 시점에 대해 일정한 평균을 가진다.
- ☐ 모든 시점에 대해 일정한 분산을 가진다.
- ☐ 공분산은 단지 시차에만 의존하고 시점 자체에는 의존하지 않는다.
- ☒ 데이터에 포함된 잡음은 독립성을 갖는다. ✓

의견 보내기

평균은 모든 시점(시간 t)에 대해 일정하다 $E(x_t) = \mu$

분산은 모든 시점(시간 t)에 대해 일정하다 $Var(x_t) = \sigma^2$

공분산은 시점(시간 t)에 의존하지 않고, 단지 시차에만 의존한다 $Cov(x_{t+h}, x_t) = \gamma_h$

✓ 14. 다음이 설명하는 데이터 마이닝 기법은 무엇인가?

1/1

미리 정의된 기준이나 예시에 의해서가 아닌 레코드 자체가 가진 다른 레코드와의 유사성에 의해 그룹화되고 이질성에 의해 세분화 되는 것으로 데이터 마이닝이나 모델링의 준비단계로서 사용되는 기법이다.

- ☒ 군집 분석 ✓
- ☐ 분류 분석
- ☐ 기술(description)
- ☐ 연관 분석

의견 보내기

분류 분석: 새롭게 나타난 현상을 검토하여 기존의 분류, 정의된 집합에 배정하는 것

기술(description): 데이터가 가진 특징 및 의미를 단순하게 설명하는 것

군집 분석: 미리 정의된 기준이나 예시에 의해서가 아닌 레코드 자체가 가진 다른 레코드와의 유사성에 의해 그룹화되고 이질성에 의해 세분화 됨

연관 분석: '같이 팔리는 물건' 같이 아이템의 연관성을 파악하는 분석

✓ 15. 다음 중 오분류표의 평가지표 중 True로 예측한 관측치 중 실제 True인 지표를 무엇이라 하는가? 1/1

- ☒ Precision
- ☐ Accuracy
- ☐ F1
- ☐ Sensitivity



의견 보내기

정밀도(Precision): 예측값이 True인 것에 대해 실제값이 True인 지표

재현율(Recall, Sensitivity): 실제값이 True인 것에 대해 예측값이 True인 지표

Accuracy: 전체 예측에서 옳은 예측의 비율

F1 Score: 오분류표 중 정밀도와 재현율의 조화평균을 나타내며 정밀도와 재현율에 같은 가중치를 부여하여 평균한 지표, $2 * (Precision * Recall) / (Precision + Recall)$

✓ 16. 아래 오분류표를 이용하여 구한 민감도(sensitivity) 값은 얼마인가? 1/1

Confusion matrix		예측값	
		TRUE	FALSE
실제값	TRUE	40	60
	FALSE	60	40

- ☒ 0.4
- ☐ 0.3
- ☐ 0.15
- ☐ 0.5



의견 보내기

재현율/민감도(Recall, Sensitivity): 실제값이 True인 것에 대해 예측값이 True인 지표

재현율/민감도 = $TP / (TP + FN) = 40 / (40 + 60) = 0.4$

✓ 17. 아래의 오분류표에서 특이도 값을 구하는 식을 구하시오.

1/1

confusion matrix		예측값	
		T	F
실제값	T	TP	FN
	F	FP	TN

- ☐ $TP / (TP + FN)$
- ☐ $(TP + TN) / (TP + FP + FN + TN)$
- ☒ $TN / (TN + FP)$
- ☐ $TP / (TP + FP)$

✓

의견 보내기

특이도(Specificity) : 실제로 False 인 것들 중 예측이 False로 된 경우의 비율

특이도 : $TN / (TN + FP)$

Accuracy : $(TP + TN) / (TP + FP + FN + TN)$

Precision : $TP / (TP + FP)$

Recall : $TP / (TP + FN)$

✓ 18. 거리를 활용한 측도에 대한 설명으로 틀린 것은?

1/1

- ☐ 유클리드는 두 점 사이의 거리로, 가장 직관적이고 일반적인 거리의 개념이다
- ☐ 맨해튼 거리는 두 점의 좌표 간의 절대값 차이를 구하는 것이다
- ☒ 마할라노비스는 변수의 표준화를 고려하고, 변수 간의 상관성을 고려하지 않는다 ✓
- ☐ 표준화, 마할라노비스 거리는 통계적 거리의 개념이다

의견 보내기

마할라노비스 : 변수의 표준화와 함께 변수 간의 상관성을 동시에 고려한 통계적 거리

✓ 19. K-fold 교차검증에 대한 설명으로 옳바르지 않은 것은?

1/1

- ☒ k=2인 경우, LOOCV라고 한다. ✓
- ☐ 주어진 데이터를 가지고 K번 반복적으로 성과를 측정해 그 결과를 평균한다.
- ☐ K-fold 는 데이터를 K개로 분할하는 것을 의미한다.
- ☐ 데이터가 충분하지 않은 경우 주로 사용한다.

의견 보내기

LOOCV(Leave-One-Out Cross-Validation)

1개 관측값 만을 validation set으로 사용하고, 나머지 $n-1$ 개를 train set으로 사용하여 n 번 학습 진행

$k=n$ 인 경우의 교차검증

학습 후 n 개의 MSE를 평균하여 최종 MSE를 계산함

✓ 20. 비계층적 군집분석인 K-means 군집분석의 수행순서는?

1/1

가) 초기 군집의 중심으로 k 개의 객체를 임의로 선택한다
나) 각 자료를 가장 가까운 군집 중심에 할당한다
다) 각 군집 내의 자료들의 평균을 계산하여 군집의 중심을 갱신한다
라) 군집 중심의 변화가 거의 없을 때까지 나와 다를 반복한다

- ☒ 가-나-다-라 ✓
- ☐ 나-가-다-라
- ☐ 다-나-가-라
- ☐ 라-가-나-다

✓ 21. 여러 분포가 확률적으로 선형 결합된 형태로, 데이터가 k 개의 모수적 모 1/1
형의 가중합으로 표현되는 모집단 모형에서 나왔다는 가정하에, 추정된 k 개
의 모형 중 어느 모형으로부터 나왔을 확률이 높은지에 따라 군집 분류를
수행하는 것은 무엇인가?

- ☐ 평균연결법
- ☒ 혼합분포군집
- ☐ SOM
- ☐ DBSCAN



의견 보내기

평균연결법: 계층적 군집 중 응집형 군집 방법으로 모든 항목에 대한 거리 평균을 구하면
서 군집화, 계산량이 많아질 수 있음
SOM: 인공신경망의 한 종류로, 차원축소와 군집화를 동시에 수행하는 기법
DBSCAN: 어느 점을 기준으로 반경 x 내에 점이 n 개 이상 있으면 하나의 군집으로 인식하는
방식을 의미하며, 임의적 모양의 군집분석

✓ 22. 교차판매/물건배치 등에 이용되는 분석 기법은?

1/1

- ☒ 연관분석
- ☐ 주성분분석
- ☐ 회귀분석
- ☐ SOM



의견 보내기

연관규칙(Association rule): 항목들 간의 '조건-결과' 식으로 표현되는 유용한 패턴
이러한 패턴 규칙을 발견해내는 것을 연관분석이라 함
장바구니 분석이라고 함(미국 마트에서 기저귀를 사는 고객은 맥주를 동시에 구매한다는
연관규칙을 알아낸 것에 기인함)

장바구니	item
1	A, B, C
2	A, C
3	B, C
4	A, B
5	A, C, D

- ☐ 0.75
☐ 0.65
☒ 0.5
☐ 0.4



의견 보내기

신뢰도: 상품 A를 구매했을 때 상품 B를 구매할 확률이 어느 정도 되는지를 확인
 $\text{신뢰도} = P(B|A) = P(A \cap B) / P(A)$: A와 B가 동시에 포함된 거래 수 / A가 포함된 거래 수
 $= 2 / 4 = 0.5$
 지지도: 전체 거래항목 중 상품 A와 상품 B를 동시에 포함하여 거래하는 비율
 $\text{지지도} = P(A \cap B)$: A와 B가 동시에 포함된 거래 수 / 전체 거래 수

- ✓ 24. 다음은 어느 마트의 판매품목에 대한 결과이다. 커피 -> 우유에 대한 향
상도는? 1/1

장바구니	item
1	커피, 빵
2	커피, 계란, 우유
3	계란, 커피, 빵, 우유
4	계란, 우유
5	커피, 우유
6	커피, 빵, 계란
7	우유, 빵

- ☐ 0.6
- ☐ 1.25
- ☐ 0.43
- ☒ 0.84



의견 보내기

향상도 : A가 주어지지 않았을 때 B의 확률 대비 A가 주어졌을 때 B의 확률 증가 비율
품목B를 구매한 고객 대비 품목A를 구매한 후 품목B를 구매하는 고객에 대한 확률
$$\text{향상도} = P(B|A)/P(B) = P(A \cap B) / (P(A) * P(B))$$
$$= (3/7) / ((5/7) * (5/7)) = 0.84$$

- ✓ 25. 시계열에 영향을 주는 일반적인 요인을 시계열에서 분리해 분석하는 방
법을 무엇이라 하는가? 1/1

분해시계열



의견 보내기

분해시계열 : 시계열에 영향을 주는 일반적인 요인을 시계열에서 분리해 분석하는 방법
분해시계열 분해 요인 : 추세요인, 계절요인, 순환요인, 불규칙요인이 있음

가) 시계열 모델 중 자기 자신의 과거 값을 사용하여 설명하는 모형임
나) 백색 잡음의 현재 값과 자기 자신의 과거 값의 선형 가중합으로 이루어진 정상 확률 모형
다) 모형에 사용하는 시계열 자료의 시점에 따라 1차, 2차, ..., p차 등을 사용하나
정상시계열 모형에서는 주로 1, 2차를 사용함

AR모형

✖

정답

AR 모델

ar model

AR model

AR MODEL

의견 보내기

AR 모델

- 시계열 모델 중 자기 자신의 과거 값을 사용하여 설명하는 모형임
- 백색 잡음의 현재 값과 자기 자신의 과거 값의 선형 가중합으로 이루어진 정상 확률 모형
- 모형에 사용하는 시계열 자료의 시점에 따라 1차, 2차, ..., p차 등을 사용하나 정상시계열 모형에서는 주로 1, 2차를 사용함

✓ 27. 학습 데이터에 너무 잘 맞게 학습되어 학습 데이터에 대한 성능은 매우 높지만 평가 데이터에 대한 성능은 낮은 상태를 무엇이라고 하는가? 1/1

과대적합

✓

의견 보내기

과대적합(Overfitting)

학습 데이터에 너무 잘 맞게 학습되어 학습 데이터에 대한 성능은 매우 높지만 평가 데이터에 대한 성능은 낮음

규제 모델, 앙상블 등의 방법으로 과대적합을 해결하거나 피할 수 있음

평가 데이터(test data)의 작은 변화에도 민감하게 반응 함

✓ 28. 다음이 설명하는 기업 내부 데이터베이스 솔루션은 무엇인가?

1/1

전사적으로 구축된 데이터웨어 하우스로부터 특정 주제, 부서 중심으로 구축된 소규모 단일 주제의 데이터웨어 하우스로 재무, 생산, 운영과 같이 특정 조직의 특정 업무 분야에 초점을 두고 있다.

데이터 마트



✓ 29. 사람 뇌의 뉴런을 모방하여 분류 및 예측에 사용할 수 있는 모형으로 입 1/1
력층, 은닉층, 출력층 3개의 층으로 구성되어 있고, 각 층에 뉴런이 여러 개
포함되어 있는 것은 무엇인가?

인공신경망



의견 보내기

인공신경망은 사람 뇌의 뉴런을 모방하여 분류 및 예측에 사용할 수 있는 모형으로 입력층, 은닉층, 출력층 3개의 층으로 구성되어 있고, 각 층에 뉴런이 여러 개 포함되어 있다.

✓ 30. 모형의 성능을 평가하기 위한 척도로 얼마나 예측이 잘 이루어졌는지를 1/1
나타내기 위해 임의로 나눈 각 등급별로 반응검출율, 반응률, 향상도 등의
정보를 산출하여 나타내는 것은 무엇인가?

이익 도표



의견 보내기

이익 도표(Lift Chart) : 모형의 성능을 평가하기 위한 척도로 얼마나 예측이 잘 이루어졌는지를 나타내기 위해 임의로 나눈 각 등급별로 반응검출율, 반응률, 향상도 등의 정보를 산출하여 나타내는 것을 말한다

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

Google 설문지





