

위치	수정 전	수정 후																				
19페이지 이론 바로 적용하기 해설	④ MySQL: 비정형데이터	④ MySQL: 정형데이터																				
36페이지 이미지 Hard Skill	① 빅데이터에 대한 이론적 지시	① 빅데이터에 대한 이론적 지식																				
38페이지 (1) 주요 기술 표	<table border="1"> <thead> <tr> <th>단계</th> <th>주요 기술</th> </tr> </thead> <tbody> <tr> <td>수집</td> <td>• ETL • EAI • 크롤링</td> </tr> <tr> <td>저장</td> <td>• 분산파일시스템 • RDBMS • Nosql</td> </tr> <tr> <td>처리, 분석</td> <td>• SQL • 머신러닝 • 통계분석</td> </tr> <tr> <td>분석</td> <td>• R • Python • BI • Open API</td> </tr> </tbody> </table>	단계	주요 기술	수집	• ETL • EAI • 크롤링	저장	• 분산파일시스템 • RDBMS • Nosql	처리, 분석	• SQL • 머신러닝 • 통계분석	분석	• R • Python • BI • Open API	<table border="1"> <thead> <tr> <th>단계</th> <th>주요 기술</th> </tr> </thead> <tbody> <tr> <td>수집</td> <td>• ETL • EAI • 크롤링 • Open API</td> </tr> <tr> <td>저장</td> <td>• 분산파일시스템 • RDBMS • Nosql</td> </tr> <tr> <td>처리, 분석</td> <td>• SQL • 머신러닝 • 통계분석</td> </tr> <tr> <td>분석</td> <td>• R • Python • BI</td> </tr> </tbody> </table>	단계	주요 기술	수집	• ETL • EAI • 크롤링 • Open API	저장	• 분산파일시스템 • RDBMS • Nosql	처리, 분석	• SQL • 머신러닝 • 통계분석	분석	• R • Python • BI
단계	주요 기술																					
수집	• ETL • EAI • 크롤링																					
저장	• 분산파일시스템 • RDBMS • Nosql																					
처리, 분석	• SQL • 머신러닝 • 통계분석																					
분석	• R • Python • BI • Open API																					
단계	주요 기술																					
수집	• ETL • EAI • 크롤링 • Open API																					
저장	• 분산파일시스템 • RDBMS • Nosql																					
처리, 분석	• SQL • 머신러닝 • 통계분석																					
분석	• R • Python • BI																					
38페이지 ①	HTML(Hypertext Makup Language	HTML(Hypertext Markup Language)																				
40페이지 ② 둘째 줄	일 을 적당한 블록 크기	파일 을 적당한 블록 크기																				
56페이지 5번 정답	④	③																				
65페이지 표 '1차 목표'	Accuracy& Deply	Accuracy& Deploy																				
67페이지 표 '지원 인프라 영역'	EDW 최저고하	EDW 최적화																				
81페이지 마지막 줄	- 대상 시스템별 데이터 수집을 위한 크롤러를 준비하고 저장소를 기획한다.	삭제																				
113페이지 (3) 데이터 수집 기술	● 수집 대상 데이터는 구조적 관점에 따라 정형 데이터, 반정형 데이터, 반정형 데이터로 나뉜다.	● 수집 대상 데이터는 구조적 관점에 따라 정형 데이터, 반정형 데이터, 비정형 데이터로 나뉜다.																				
130페이지 (2) ① 표 일관성 설명	데이터의 할 구조, 값, 형태가 일관되게 정의되는 것으로 신뢰를 보장하는 척도이다.	데이터의 구조, 값, 형태가 일관되게 정의되는 것으로 신뢰를 보장하는 척도이다.																				
131페이지 ② 표 기능성 설명	해당 데이터가 특정 조건에서 사용될 때, 명시된 요과 내재된 요구를 만족하는 기능 제공 여부	해당 데이터가 특정 조건에서 사용될 때, 명시된 요구와 내재된 요구를 만족하는 기능 제공 여부																				
169페이지 3. 표 표현형식 변환 예시	섭씨 → 화씨 변환	성별 → 0과 1																				
173페이지 표	AMASYN	ADASYN																				
174페이지 3번	③ 미성년자는 음주량을 응답을 기피함 - 무작위 결측	③ 미성년자는 음주량을 응답을 기피함 - 비무작위 결측																				
218페이지 ②	표본의 수가 충분히 크다면 모집단의 분포와 상관없이 표본분포 는 정규분포에 근사한다.	표본의 수가 충분히 크다면 모집단의 분포와 상관없이 표본평균의 분포 는 정규분포에 근사한다.																				

220페이지 2.	점 추정 조건 통계는 표본평균, 표본분산, 위수 , 최빈수가 있다.	점 추정 조건 통계는 표본평균, 표본분산, 중위수 , 최빈수가 있다.
225페이지 4. (1) ① 표	Alternative hypothesis: H_a	Alternative hypothesis: H_1
264페이지	정확도	정밀도
267페이지 (2) ② 표 분산 분석에서 F-통계량 내용	p-값 이 가장 작은 예측 변수와 그 당시의 최적 분리를 통해서 자식마디를 형성한다.	f-값 이 가장 작은 예측 변수와 그 당시의 최적 분리를 통해서 자식마디를 형성한다.
344페이지 표	<div style="border: 1px solid black; padding: 5px;"> <div style="display: flex; justify-content: space-between; align-items: center;"> <div style="background-color: #f0e6ff; padding: 2px 5px;">L2 규제</div> <div style="text-align: center;"> 실제값과 예측값 오차들의 제곱의 합 $\sum_{i=0}^N (y_i - \sum_{j=0}^M x_{ij} W_j)^2 + \lambda \sum_{j=0}^M W_j$ </div> </div> </div>	<div style="border: 1px solid black; padding: 5px;"> <div style="display: flex; justify-content: space-between; align-items: center;"> <div style="background-color: #f0e6ff; padding: 2px 5px;">L2 규제</div> <div style="text-align: center;"> 실제값과 예측값 오차들의 제곱의 합 $\sum_{i=0}^N (y_i - \sum_{j=0}^M x_{ij} W_j)^2 + \lambda \sum_{j=0}^M W_j$ </div> </div> <p style="color: red; margin-top: 5px;">(w_j의 절대값 삭제)</p> </div>
387페이지 24번	③ 미성년자는 음주량을 응답을 기피함 - 무작위 결측	③ 미성년자는 음주량을 응답을 기피함 - 비무작위 결측
402페이지 40번 발문 아래 추가		x_1, x_2 가 서로 독립일 때, $f(x_1, x_2 \theta) = f(x_1 \theta) \cdot f(x_2 \theta)$ 임 $f(\theta) = f(tvert\theta) = \begin{cases} \theta_e^{-te} & (t \geq 0) \\ 0 & (t < 0) \end{cases}$
429페이지 49번 발문 아래 추가		TP/(TP+FN)
443페이지 56번	분석에 사용되는 변수를 랜덤하게 추출하는 것으로서, 훈련단계에서 훈련목적함수를 최대화 만드는 노드 분할 함수의 매개변수 0 의 최적값을 구하는 과정이다.	분석에 사용되는 변수를 랜덤하게 추출하는 것으로서, 훈련단계에서 훈련목적함수를 최대화 만드는 노드 분할 함수의 매개변수 θ 의 최적값을 구하는 과정이다.
446페이지 75번	① 계단식 그래프(Step Line Graph): 수치화된 데이터 값을 서로 다른 선의 굵기로 표시하는 그래프	① 계단식 그래프(Step Line Graph): 수치화된 데이터 값을 서로 다른 원형으로 표시하는 그래프
450페이지 21번 정답	③	②
451페이지 33번 정답 및 해설	33 ① ② 피어슨 상관계수는 +1과 -1 사이의 값을 가진다. ③, ④ 스피어만 상관계수에 대한 설명이다.	33 ①, ③ (중복 정답) ② 피어슨 상관계수는 +1과 -1 사이의 값을 가진다. ④ 스피어만 상관계수에 대한 설명이다.
463페이지 6번 정답	③	②
464페이지 12번 정답	①	③
468페이지 54번 정답	③	②
477페이지 68번 정답	④	③

481페이지 23번 정답 및 해설	<p>23 ① 대치단계에서 만든 m개의 완전한 가상 자료 각각을 표준 적 통계분석을 통하여 관심이 있는 추정량과 분산을 계산 한다.</p>	<p>23 ② 다중대치방법은 결측값을 2번 이상 대치 후 m개의 가상 적 자료 만들어 분석한다.</p>
482페이지 44번 정답	②	③