

위치	오류유형	수정 전	수정 후																																				
46p 번호 : 11	문제-문항	① 제품과 서비스의 유형	① 거래의 형태																																				
61p	문제-본문	<table border="1"> <thead> <tr> <th>유형</th> <th>데이터 소스</th> </tr> </thead> <tbody> <tr> <td>정형 데이터</td> <td> <ul style="list-style-type: none"> 업무 처리, 매개거래, 로그 데이터, 시계열 데이터 DB to DB, Sqoop, EAI(Enterprise Application Integration) ETL(Extract, Transform, Load) 등을 이용한 수집 </td> </tr> <tr> <td>반정형 데이터</td> <td> <ul style="list-style-type: none"> 데이터 내부에 데이터 구조에 대한 메타 정보 포함 HTML, XML, JSON, RSS, 웹로그, 센서 데이터 등 </td> </tr> <tr> <td>비정형 데이터</td> <td> <table border="1"> <thead> <tr> <th>유형</th> <th>설명</th> </tr> </thead> <tbody> <tr> <td>일반</td> <td> <ul style="list-style-type: none"> 소셜 미디어(SNS), 고객 서비스, 품질보증 이벤트, 이메일, 자유형식의 텍스트 </td> </tr> <tr> <td>센서 데이터</td> <td>온도, QR 코드, RFID, GPS</td> </tr> <tr> <td>새로운 데이터 유형</td> <td>GPS 매칭, 오디오, 정치화상, 비디오</td> </tr> <tr> <td>데이터 수집</td> <td>Crawler, FTP, HTTP Protocol 수집, Parsing</td> </tr> </tbody> </table> </td> </tr> </tbody> </table>	유형	데이터 소스	정형 데이터	<ul style="list-style-type: none"> 업무 처리, 매개거래, 로그 데이터, 시계열 데이터 DB to DB, Sqoop, EAI(Enterprise Application Integration) ETL(Extract, Transform, Load) 등을 이용한 수집 	반정형 데이터	<ul style="list-style-type: none"> 데이터 내부에 데이터 구조에 대한 메타 정보 포함 HTML, XML, JSON, RSS, 웹로그, 센서 데이터 등 	비정형 데이터	<table border="1"> <thead> <tr> <th>유형</th> <th>설명</th> </tr> </thead> <tbody> <tr> <td>일반</td> <td> <ul style="list-style-type: none"> 소셜 미디어(SNS), 고객 서비스, 품질보증 이벤트, 이메일, 자유형식의 텍스트 </td> </tr> <tr> <td>센서 데이터</td> <td>온도, QR 코드, RFID, GPS</td> </tr> <tr> <td>새로운 데이터 유형</td> <td>GPS 매칭, 오디오, 정치화상, 비디오</td> </tr> <tr> <td>데이터 수집</td> <td>Crawler, FTP, HTTP Protocol 수집, Parsing</td> </tr> </tbody> </table>	유형	설명	일반	<ul style="list-style-type: none"> 소셜 미디어(SNS), 고객 서비스, 품질보증 이벤트, 이메일, 자유형식의 텍스트 	센서 데이터	온도, QR 코드, RFID, GPS	새로운 데이터 유형	GPS 매칭, 오디오, 정치화상, 비디오	데이터 수집	Crawler, FTP, HTTP Protocol 수집, Parsing	<table border="1"> <thead> <tr> <th>유형</th> <th>데이터 소스</th> </tr> </thead> <tbody> <tr> <td>정형 데이터</td> <td> <ul style="list-style-type: none"> 업무 처리, 매개거래, 로그 데이터, 시계열 데이터 DB to DB, Sqoop, EAI(Enterprise Application Integration) ETL(Extract, Transform, Load) 등을 이용한 수집 </td> </tr> <tr> <td>반정형 데이터</td> <td> <ul style="list-style-type: none"> 데이터 내부에 데이터 구조에 대한 메타 정보 포함 HTML, XML, JSON, RSS, 웹로그, 보안 및 특정 센서 데이터 등 </td> </tr> <tr> <td>비정형 데이터</td> <td> <table border="1"> <thead> <tr> <th>유형</th> <th>설명</th> </tr> </thead> <tbody> <tr> <td>일반</td> <td> <ul style="list-style-type: none"> 소셜 미디어(SNS), 고객 서비스, 품질보증 이벤트, 이메일, 자유형식의 텍스트 </td> </tr> <tr> <td>센서 데이터</td> <td>온도, QR 코드, RFID, GPS</td> </tr> <tr> <td>새로운 데이터 유형</td> <td>GPS 매칭, 오디오, 정치화상, 비디오</td> </tr> <tr> <td>데이터 수집</td> <td>Crawler, FTP, HTTP Protocol 수집, Parsing</td> </tr> </tbody> </table> </td> </tr> </tbody> </table>	유형	데이터 소스	정형 데이터	<ul style="list-style-type: none"> 업무 처리, 매개거래, 로그 데이터, 시계열 데이터 DB to DB, Sqoop, EAI(Enterprise Application Integration) ETL(Extract, Transform, Load) 등을 이용한 수집 	반정형 데이터	<ul style="list-style-type: none"> 데이터 내부에 데이터 구조에 대한 메타 정보 포함 HTML, XML, JSON, RSS, 웹로그, 보안 및 특정 센서 데이터 등 	비정형 데이터	<table border="1"> <thead> <tr> <th>유형</th> <th>설명</th> </tr> </thead> <tbody> <tr> <td>일반</td> <td> <ul style="list-style-type: none"> 소셜 미디어(SNS), 고객 서비스, 품질보증 이벤트, 이메일, 자유형식의 텍스트 </td> </tr> <tr> <td>센서 데이터</td> <td>온도, QR 코드, RFID, GPS</td> </tr> <tr> <td>새로운 데이터 유형</td> <td>GPS 매칭, 오디오, 정치화상, 비디오</td> </tr> <tr> <td>데이터 수집</td> <td>Crawler, FTP, HTTP Protocol 수집, Parsing</td> </tr> </tbody> </table>	유형	설명	일반	<ul style="list-style-type: none"> 소셜 미디어(SNS), 고객 서비스, 품질보증 이벤트, 이메일, 자유형식의 텍스트 	센서 데이터	온도, QR 코드, RFID, GPS	새로운 데이터 유형	GPS 매칭, 오디오, 정치화상, 비디오	데이터 수집	Crawler, FTP, HTTP Protocol 수집, Parsing
유형	데이터 소스																																						
정형 데이터	<ul style="list-style-type: none"> 업무 처리, 매개거래, 로그 데이터, 시계열 데이터 DB to DB, Sqoop, EAI(Enterprise Application Integration) ETL(Extract, Transform, Load) 등을 이용한 수집 																																						
반정형 데이터	<ul style="list-style-type: none"> 데이터 내부에 데이터 구조에 대한 메타 정보 포함 HTML, XML, JSON, RSS, 웹로그, 센서 데이터 등 																																						
비정형 데이터	<table border="1"> <thead> <tr> <th>유형</th> <th>설명</th> </tr> </thead> <tbody> <tr> <td>일반</td> <td> <ul style="list-style-type: none"> 소셜 미디어(SNS), 고객 서비스, 품질보증 이벤트, 이메일, 자유형식의 텍스트 </td> </tr> <tr> <td>센서 데이터</td> <td>온도, QR 코드, RFID, GPS</td> </tr> <tr> <td>새로운 데이터 유형</td> <td>GPS 매칭, 오디오, 정치화상, 비디오</td> </tr> <tr> <td>데이터 수집</td> <td>Crawler, FTP, HTTP Protocol 수집, Parsing</td> </tr> </tbody> </table>	유형	설명	일반	<ul style="list-style-type: none"> 소셜 미디어(SNS), 고객 서비스, 품질보증 이벤트, 이메일, 자유형식의 텍스트 	센서 데이터	온도, QR 코드, RFID, GPS	새로운 데이터 유형	GPS 매칭, 오디오, 정치화상, 비디오	데이터 수집	Crawler, FTP, HTTP Protocol 수집, Parsing																												
유형	설명																																						
일반	<ul style="list-style-type: none"> 소셜 미디어(SNS), 고객 서비스, 품질보증 이벤트, 이메일, 자유형식의 텍스트 																																						
센서 데이터	온도, QR 코드, RFID, GPS																																						
새로운 데이터 유형	GPS 매칭, 오디오, 정치화상, 비디오																																						
데이터 수집	Crawler, FTP, HTTP Protocol 수집, Parsing																																						
유형	데이터 소스																																						
정형 데이터	<ul style="list-style-type: none"> 업무 처리, 매개거래, 로그 데이터, 시계열 데이터 DB to DB, Sqoop, EAI(Enterprise Application Integration) ETL(Extract, Transform, Load) 등을 이용한 수집 																																						
반정형 데이터	<ul style="list-style-type: none"> 데이터 내부에 데이터 구조에 대한 메타 정보 포함 HTML, XML, JSON, RSS, 웹로그, 보안 및 특정 센서 데이터 등 																																						
비정형 데이터	<table border="1"> <thead> <tr> <th>유형</th> <th>설명</th> </tr> </thead> <tbody> <tr> <td>일반</td> <td> <ul style="list-style-type: none"> 소셜 미디어(SNS), 고객 서비스, 품질보증 이벤트, 이메일, 자유형식의 텍스트 </td> </tr> <tr> <td>센서 데이터</td> <td>온도, QR 코드, RFID, GPS</td> </tr> <tr> <td>새로운 데이터 유형</td> <td>GPS 매칭, 오디오, 정치화상, 비디오</td> </tr> <tr> <td>데이터 수집</td> <td>Crawler, FTP, HTTP Protocol 수집, Parsing</td> </tr> </tbody> </table>	유형	설명	일반	<ul style="list-style-type: none"> 소셜 미디어(SNS), 고객 서비스, 품질보증 이벤트, 이메일, 자유형식의 텍스트 	센서 데이터	온도, QR 코드, RFID, GPS	새로운 데이터 유형	GPS 매칭, 오디오, 정치화상, 비디오	데이터 수집	Crawler, FTP, HTTP Protocol 수집, Parsing																												
유형	설명																																						
일반	<ul style="list-style-type: none"> 소셜 미디어(SNS), 고객 서비스, 품질보증 이벤트, 이메일, 자유형식의 텍스트 																																						
센서 데이터	온도, QR 코드, RFID, GPS																																						
새로운 데이터 유형	GPS 매칭, 오디오, 정치화상, 비디오																																						
데이터 수집	Crawler, FTP, HTTP Protocol 수집, Parsing																																						
66p	문제-본문	<table border="1"> <thead> <tr> <th>유형</th> <th>데이터 종류</th> <th>저장 시스템</th> </tr> </thead> <tbody> <tr> <td>정형</td> <td>RDB, 스프레드시트</td> <td>RDB</td> </tr> <tr> <td>반정형</td> <td>HTML, XML, JSON, 웹문서, 웹로그, 센서 데이터</td> <td>RDB, NoSQL</td> </tr> <tr> <td>비정형</td> <td>소셜 데이터, 문서(워드, 한글), 이미지, 오디오, 비디오, IoT</td> <td>NoSQL, 분산파일시스템 (HDFS)</td> </tr> </tbody> </table>	유형	데이터 종류	저장 시스템	정형	RDB, 스프레드시트	RDB	반정형	HTML, XML, JSON, 웹문서, 웹로그, 센서 데이터	RDB, NoSQL	비정형	소셜 데이터, 문서(워드, 한글), 이미지, 오디오, 비디오, IoT	NoSQL, 분산파일시스템 (HDFS)	<table border="1"> <thead> <tr> <th>유형</th> <th>데이터 종류</th> <th>저장 시스템</th> </tr> </thead> <tbody> <tr> <td>정형</td> <td>RDB, 스프레드시트</td> <td>RDB</td> </tr> <tr> <td>반정형</td> <td>HTML, XML, JSON, 웹문서, 웹로그, 보안 및 특정 센서 데이터</td> <td>RDB, NoSQL</td> </tr> <tr> <td>비정형</td> <td>소셜 데이터, 문서(워드, 한글), 이미지, 오디오, 비디오, IoT</td> <td>NoSQL, 분산파일시스템 (HDFS)</td> </tr> </tbody> </table>	유형	데이터 종류	저장 시스템	정형	RDB, 스프레드시트	RDB	반정형	HTML, XML, JSON, 웹문서, 웹로그, 보안 및 특정 센서 데이터	RDB, NoSQL	비정형	소셜 데이터, 문서(워드, 한글), 이미지, 오디오, 비디오, IoT	NoSQL, 분산파일시스템 (HDFS)												
유형	데이터 종류	저장 시스템																																					
정형	RDB, 스프레드시트	RDB																																					
반정형	HTML, XML, JSON, 웹문서, 웹로그, 센서 데이터	RDB, NoSQL																																					
비정형	소셜 데이터, 문서(워드, 한글), 이미지, 오디오, 비디오, IoT	NoSQL, 분산파일시스템 (HDFS)																																					
유형	데이터 종류	저장 시스템																																					
정형	RDB, 스프레드시트	RDB																																					
반정형	HTML, XML, JSON, 웹문서, 웹로그, 보안 및 특정 센서 데이터	RDB, NoSQL																																					
비정형	소셜 데이터, 문서(워드, 한글), 이미지, 오디오, 비디오, IoT	NoSQL, 분산파일시스템 (HDFS)																																					
67p	문제-본문	Google GFS(Google File System)	Google의 GFS(Google File System)																																				
152p 번호 : 53	해설	<p>정답 ④</p> <p>해설 키, 온도는 등간 척도, 순위는 서열 척도이다.</p>	<p>정답 ②, ④</p> <p>해설 온도는 등간 척도, 순위는 서열 척도이다.</p>																																				
204p 번호 : 38	문제-문항	④ MapReduce	④ MapR																																				
226p	문제-본문	<p>상자그림 그리는 방법</p> <ul style="list-style-type: none"> 위의 [그림]은 11명의 중간고사 성적(89점, 88점, 67점, 78점, 90점, 88점, 78점, 90점, 86점, 55점)에 대해서 상자그림을 그린 것이다. 상자그림을 그리는 방법은 다음과 같다. 	<p>상자그림 그리는 방법</p> <ul style="list-style-type: none"> 위의 [그림]은 11명의 중간고사 성적(89점, 88점, 67점, 78점, 90점, 88점, 78점, 90점, 86점, 55점, 78점)에 대해서 상자그림을 그린 것이다. 상자그림을 그리는 방법은 다음과 같다. 																																				
230p	문제-본문	$s = \sqrt{\frac{1}{(n-1)} (\sum x_i^2 - n\bar{x}^2)}$	$s = \sqrt{\frac{1}{(n-1)} (\sum x_i^2 - n\bar{x}^2)}$																																				
266p	문제-본문	㉡ 공차한계가 0.1 이하일 때 다중공선성 문제가 없다고 판정한다(분산팽창지수의 역수 값).	㉡ 공차한계가 0.1 이하일 때 다중공선성 문제가 있다고 판정한다(분산팽창지수의 역수 값).																																				

위치	오류유형	수정 전	수정 후
311p	문제-본문	① $\bar{x} = 77.5$ (표본평균)	① $\bar{x} = 76.7$ (표본평균)
313p	문제-본문	<p>⑥ 신뢰구간 예시 상기 통학시간 자료의 95% 신뢰구간을 구하시오. (60, 65, 50, 45, 35, 40, 45, 60, 50, 50)</p>	<p>⑥ 신뢰구간 예시 다음 통학시간(분) 자료의 95% 신뢰구간을 구하시오. (60, 65, 50, 45, 35, 40, 45, 60, 50, 50)</p>
337p	문제-본문	표본 1	표본 1
410p	문제-본문	$= r \frac{s_y}{s_x}$	$\hat{\beta} = r \frac{s_y}{s_x}$
458p	문제-본문	$\beta_0 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$	$\beta_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$
474~475p	문제-본문	<p>장점</p> <p>⑦ 첫째, 다양한 형태의 데이터에 적용이 가능하고 분석방법의 적용이 용이하다.</p> <p>⑧ 둘째, 이해하고 구현하기가 쉬우며 결과 해석이 용이하다는 것이다.</p> <p>단점</p> <p>⑨ 첫째, 초기 군집수를 결정하는 것이 어렵고 결과 해석이 어렵다는 것이다.</p> <p>⑩ 둘째, 데이터의 평균값 정의가 가능한 데이터에 대해서만 사용이 가능하다는 것이다.</p> <p>⑪ 셋째, 아웃라이어에 매우 민감하다는 것이다.</p>	<p>장점</p> <p>⑦ 알고리즘이 쉬워 구현이 비교적 쉽다.</p> <p>⑧ 주어진 알고리즘으로 대규모 적용이 가능하다.</p> <p>⑨ 다양한 형태의 데이터에 적용이 가능하다.</p> <p>단점</p> <p>⑩ 초기 k를 찾기 어렵다.</p> <p>⑪ 모든 점에서 계산하기 때문에 속도가 느린다.</p> <p>⑫ 평균을 사용하기 때문에 아웃라이어에 영향을 많이 받는다.</p> <p>⑬ 가중치와 거리정의가 어렵다.</p> <p>⑭ 사전에 주어진 목적이 없으므로 결과를 해석하는 데 있어 어려움이 존재한다.</p>

위치	오류유형	수정 전	수정 후
478p	문제-본문	$S_{xy} = \sum(x_i - \bar{x})(y_i - y)$	$S_{xy} = \sum(x_i - \bar{x})(y_i - \bar{y})$
478p	문제-본문	$S_{yy} = \sum(y_i - y)^2$	$S_{yy} = \sum(y_i - \bar{y})^2$
590p	문제-본문	중앙값 = $\begin{cases} x_{(n+1)/2}, & n\text{이 홀수} \\ x_{n/2}, & n\text{이 짝수} \end{cases}$	중앙값 = $\begin{cases} x_{(n+1)/2}, & n\text{이 홀수} \\ \frac{1}{2}(x_{n/2} + x_{(n/2+1)}), & n\text{이 짝수} \end{cases}$
608p	문제-본문	<ul style="list-style-type: none"> 지지도(Support) : 전체 데이터세트에서 해당 아이템 집합이 포함된 비율 $S(X) = \frac{\text{Count}(X)}{N} = S(X, Y) = \frac{\text{Count}(X, Y)}{N} = P(X \cap Y)$ 	<p>지지도(Support) : 전체 데이터세트에서 해당 아이템 집합이 포함된 비율 $S(X) = \frac{\text{Count}(X)}{N} = P(X), S(X, Y) = \frac{\text{Count}(X, Y)}{N} = P(X \cap Y)$</p> <p>즉, 지지도 $S(X)$는 N명의 모집단에서 사건 X가 일어날 확률의 개념과 유사하게 정의됨</p>
608p	문제-본문	$\begin{aligned} Lift(X \Rightarrow Y) &= \frac{S(X, Y)}{S(Y)} = \frac{\frac{S(X, Y)}{S(X)}}{\frac{S(X, Y)}{S(Y)}} = \frac{S(X, Y)}{S(X) S(Y)} = \frac{\frac{\text{Count}(X, Y)}{N}}{\frac{\text{Count}(X)}{N} \frac{\text{Count}(Y)}{N}} \\ &= \frac{\frac{P(Y X)}{P(Y)}}{\frac{P(Y)}{P(Y)}} = \frac{P(X \cap Y)}{P(X)P(Y)} \end{aligned}$	$\begin{aligned} Lift(X \Rightarrow Y) &= \frac{S(X, Y)}{S(Y)} = \frac{\frac{S(X, Y)}{S(X)}}{\frac{S(X, Y)}{S(Y)}} = \frac{S(X, Y)}{S(X) S(Y)} = \frac{\frac{\text{Count}(X, Y)}{N}}{\frac{\text{Count}(X)}{N} \frac{\text{Count}(Y)}{N}} \\ &= \frac{\frac{P(X \cap Y)}{P(Y)}}{\frac{P(X)}{P(Y)}} = \frac{P(X \cap Y)}{P(X)P(Y)} \end{aligned}$
710p 번호 : 22	문제-본문	① 가장 약한 상관관계는 상관계수가 0이다.	① 상관계수가 0이면 서로 상관관계가 없다.

도서의 오류로 학습에 불편드린 점 진심으로 사과드립니다.
 더 나은 도서를 만들기 위해 노력하는 시대교육그룹이 되겠습니다.