

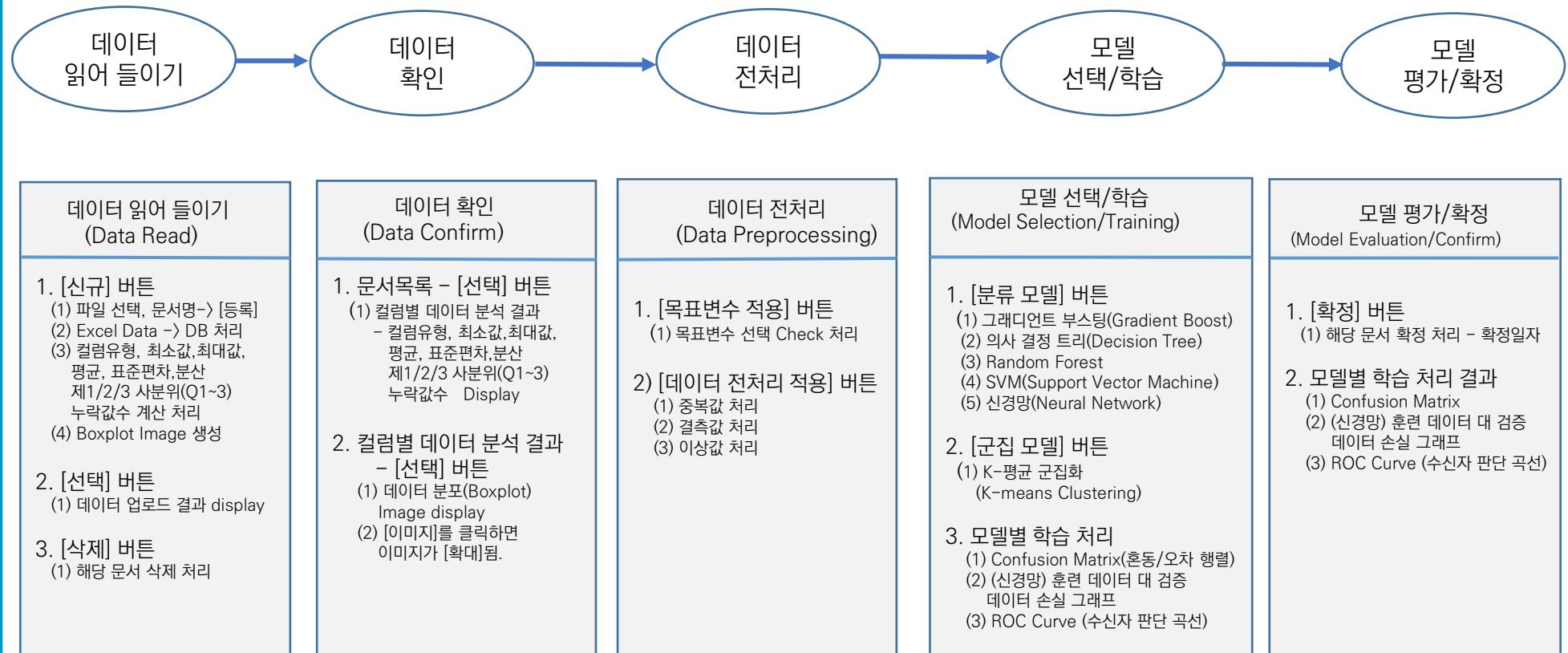
올바로시스템 노후 장비 교체 및 클라우드 인프라 증설

프로세스 설계서

[머신러닝 소프트웨어]

문서 번호	B2022_02
작성 일자	2022.10.

1. 프로세스 개요



2. 기능 프로세스

2.1 데이터 읽어 들이기 (Data Read)

1

A	B	C	D	E	F	G	H	I	J	K	
1	Channel	Region	Fresh	Milk	Grocery	Frozen	Detergent	Delicassen	Sugar	Apple	y
2	2	3	705	980	9566	7561	214	2674	1338	4656	3174 no
3	2	3	705	980	9566	7561	1762	3285	1776	6008	2379 yes
4	2	3	663	800	7684	6404	3516	788	1776	6008	2379 yes
5	1	3	13265	1196	4221	6404	507	1788	2196	5404 no	
6	2	3	22615	5410	719	3915	1777	5185	4470	2915 no	
7	2	3	9413	8259	5126	466	1789	1451	7259	766 yes	
8	2	3	12126	3199	6975	480	3140	545	4199	580 no	
9	2	3	7579	9423	1042	196	332	2568	399	1659 no	
10	1	3	3463	3648	6192	425	1716	750	3148	475 yes	
11	2	3	6006	11093	18881	1159	7425	2098	11023	1139 yes	

1. Excel 파일(CSV 포맷)을 업로드 한다.

데이터 읽어 들이기

신규 파일 업로드

파일
파일 선택 선택된 파일 없음
문서 명



문서 명	처리상태	등록일	선택	삭제
테스트 - 2022.11.21(01)	업로드	2022-11-21	<input checked="" type="button"/> 선택	<input type="button"/> 삭제

등록 닫기

2

automl_dataread.py

2. 업로드된 파일을 적용하여 Auto ML 관련 테이블을 생성한다.

(1) 데이터 상세 테이블(TB_AUTOML_TEXT) Insert 처리

- Header 데이터
- 실제 데이터

(2) 데이터 컬럼 상세 테이블(TB_AUTOML_COLUMN_INFO) Insert 처리

- 컬럼유형, 최소값, 최대값, 평균, 표준편차, 분산, 제1/2/3 사분위(Q1~3)
- 누락값수

(3) 데이터 마스터 테이블(TB_AUTOML_MASTER) Update 처리

- 처리상태 코드 : 업로드
- 업로드 파일 행 개수, 업로드 파일 행 개수

3

automl_dataconfirm.py

3. 데이터 상세 테이블(TB_AUTOML_TEXT)을 select하여 데이터 컬럼 이미지 테이블을 생성한다.

(1) Boxplot(상자 수염 그림) Image 생성

- Boxplot(상자 수염 그림) - 서버에 저장하기
- IMAGE_GUBUN(이미지 구분) : '02'

2.2 데이터 확인 (Data Confirm)

1

문서 명 처리상태

Total : 13

문서 명	처리상태	등록일	선택
테스트 - 2022.11.01(03)	업로드	2022-11-18	<input type="button" value="선택"/>
테스트 - 2022.11.01(02)	업로드	2022-11-18	<input type="button" value="선택"/>
테스트 - 2022.11.01(01)	업로드	2022-11-18	<input type="button" value="선택"/>

<< < 1 2 > >>

1. 문서목록에서 [선택] 버튼을 클릭한다.

2

칼럼 개수 : 5

[컬럼별 데이터 분석 결과] ■ 제1사분위(Q1) : 25%의 위치, 제2사분위(Q2) : 50%의 위치, 제3사분위(Q3) : 75%의 위치

칼럼 명	칼럼 유형	사용여부	최소값	최대값	평균	표준편차	분산	25%	50%	75%	누적값 수	선택
공원명	문자형	Y									0	<input type="button" value="선택"/>
육지면적	실수형	Y	65.34	483.02	216.38	118.50	14042.81	149.14	178.58	236.57	0	<input type="button" value="선택"/>
답방객수	정수형	Y	851815.00	6439653.00	2012578.42	1706610.44	2912519178922.81	869313.25	1068332.50	2976096.50	0	<input type="button" value="선택"/>
도시면적객수	정수형	Y	851815.00	6439653.00	2012578.42	1706610.44	2912519178922.81	869313.25	1068332.50	2976096.50	0	<input type="button" value="선택"/>
y	문자형	Y									0	<input type="button" value="선택"/>

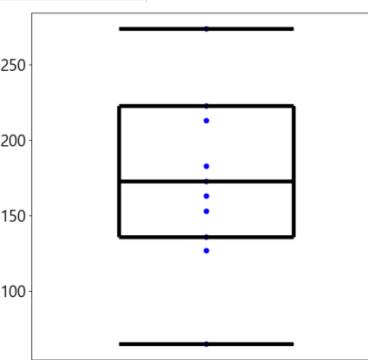
<< < 1 > >>

2. ‘컬럼별 데이터 분석 결과’ 목록에서 [선택] 버튼을 클릭한다.

(1) 3.번 처리

(2) [데이터 확인] 버튼을 클릭하면 처리상태가 ‘데이터 확인’으로 변경된다.

3



3. 데이터 분포(Boxplot) Image display
- [이미지]를 클릭하면 이미지가 [확대]된다.

2.3 데이터 전처리 (Data Preprocessing)

1.

2.

3.

1. 문서목록에서 [선택] 버튼을 클릭한다.

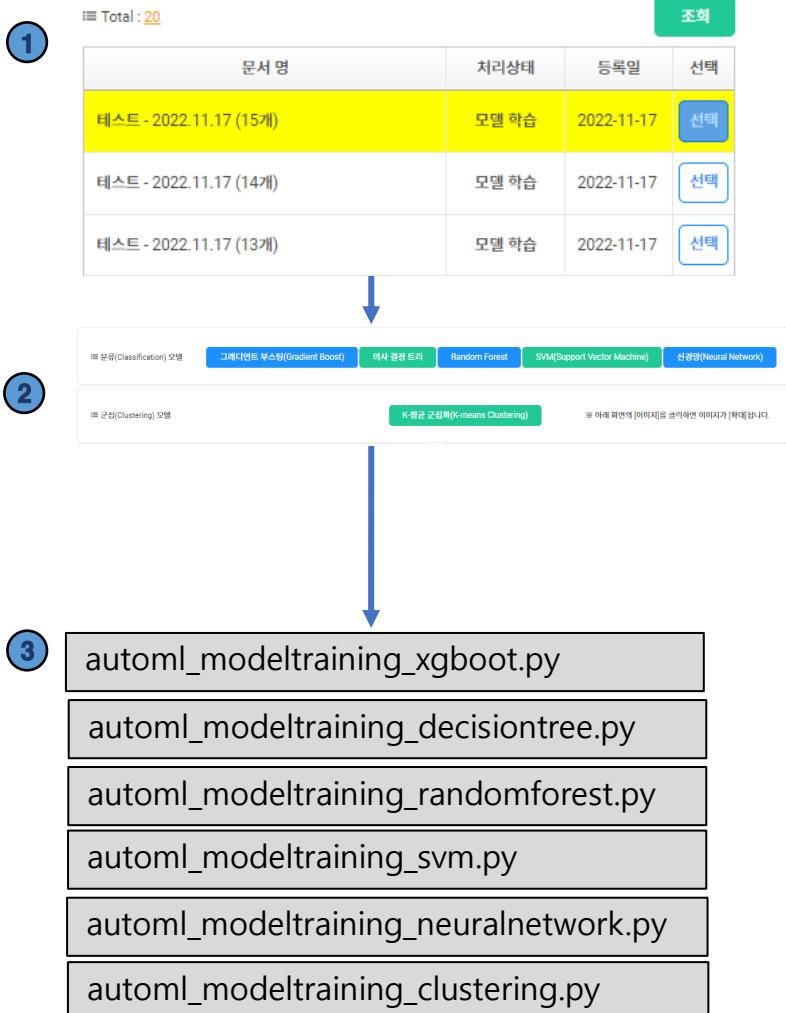
2. ‘컬럼별 데이터 분석 결과’ 목록에서 [목표변수 선택] Check 처리한다.

- (1) [목표변수 적용] 버튼을 클릭한다.
– 처리상태가 ‘데이터 전처리’로 변경된다.

3. [데이터 전처리] 구분별로 제거 여부, 사용 여부를 선택한다.

- (1) 중복값 처리
- (2) 결측값 처리
- (3) 이상값 처리

2.4 모델 선택/학습 (Model Selection/Training)



1. 문서목록에서 [선택] 버튼을 클릭한다.

2. 학습할 [분류 모델] / [군집 모델]을 클릭한다.

(1) 분류 모델

- 그레이디언트 부스팅(Gradient Boost)
- 의사 결정 트리(Decision Tree)
- Random Forest
- SVM(Support Vector Machine)
- 신경망(Neural Network)

(2) 군집 모델 : K-평균 군집화(K-means Clustering)

3. [분류 모델] / [군집 모델]을 학습하시겠습니까? → [예]를 클릭한 경우

- (1) 데이터 컬럼 상세 테이블(TB_AUTOML_COLUMN_INFO) select
- (2) 데이터 전처리(Data Preprocessing) – 목표변수 활용
 - 변수 정의
 - 범주형 데이터를 숫자형 데이터로 전환
 - 범주형 데이터와 숫자형 데이터 결합
 - 모든 특징의 이름 리스트 처리
- (3) 데이터 분할 – train_test_split
- (4) 알고리즘 선택
- (5) 학습, 예측
- (6) 학습 결과 테이블(TB_AUTOML_TRAINING_RESULT) 처리
- (7) ROC Curve(수신자 판단 곡선) 처리
- (8) Confusion Matrix(혼동/오차 행렬) 처리
- (9) (신경망) 훈련 데이터 대 검증 데이터 손실 그래프 처리
- (10) 처리상태 ‘모델 학습’으로 변경 처리

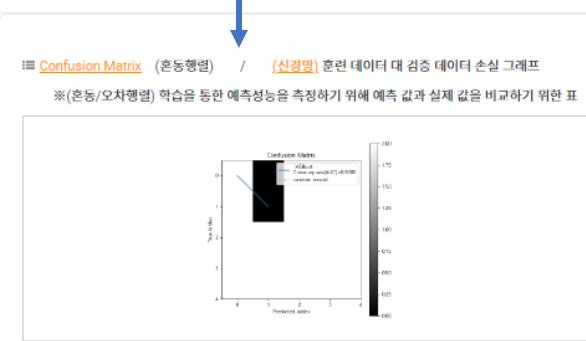
4

■ Total : 5 모델 학습 목록

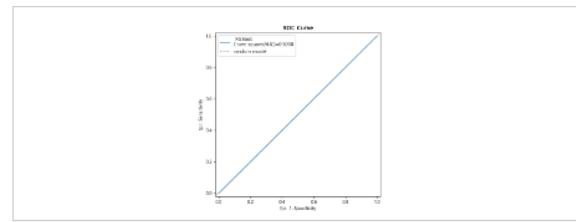
※ 정밀도(Precision) : 모델이 True라고 분류한 것 중에서 실제 True인 것의 비율
 ※ 민감도(Sensitivity) : 실제 True인 것 속에서 모델이 True라고 예측한 것의 비율
 ※ F1-score : 정밀도와 민감도의 조화평균 $> 2 * (\text{정밀도} * \text{민감도}) / (\text{정밀도} + \text{민감도})$

학습 구분	정밀도	민감도	F1-score	학습일자	학습횟수	선택
Gradient Boost	0.5	0.5	0.5	2022-11-17	1	<button>선택</button>
의사결정트리	0.4	0.4	0.4	2022-11-17	1	<button>선택</button>
Random Forest	0.4	0.4	0.4	2022-11-17	1	<button>선택</button>
SVM	0.4	0.4	0.4	2022-11-17	1	<button>선택</button>
신경망	0.25	0.25	0.25	2022-11-17	1	<button>선택</button>

5



■ ROC Curve ※ 특이도(Specificity) : 실제 진단결과가 음성 중에 음성이라고 맞춘 비율
 ※ (수신자 판단 곡선) 모델의 효율성을 민감도와 특이도를 이용하여 그래프로 나타낸 것



5. [모델 학습] 처리 후, '모델 학습 목록'에 학습한 데이터가 나타난다.

- (1) 학습구분, 정밀도, 민감도, F1-score, 학습일자, 학습횟수
- (2) [선택] 버튼을 클릭하면 6. 학습 결과 Image가 나타난다.

6. 학습 결과 만들어 진 Image

(1) Confusion Matrix (혼동 / 오차 행렬)

- 학습을 통한 예측성능을 측정하기 위해 예측 값과 실제 값을 비교하기 위한 표

(2) ROC Curve (수신자 판단 곡선)

- 모델의 효율성을 민감도와 특이도를 이용하여 그래프로 나타낸 것

2.5 모델 평가/확정 (Model Evaluation/Confirm)

1

Total : 22				조회
문서 명	처리상태	등록일	선택	
테스트 - 2022.11.17 (15개)	모델 학습	2022-11-17	<button>선택</button>	
테스트 - 2022.11.17 (14개)	모델 학습	2022-11-17	<button>선택</button>	
테스트 - 2022.11.17 (13개)	모델 학습	2022-11-17	<button>선택</button>	

1. 문서목록에서 [선택] 버튼을 클릭한다.



2

Total : 5 모델 학습 목록								
학습 구분	정밀도	민감도	F1-score	학습일자	학습횟수	확정일자	선택	확정
Gradient Boost	0.5	0.5	0.5	2022-11-17	1		<button>선택</button>	<button>확정</button>
의사결정트리	0.4	0.4	0.4	2022-11-17	1		<button>선택</button>	<button>확정</button>
Random Forest	0.4	0.4	0.4	2022-11-17	1		<button>선택</button>	<button>확정</button>
SVM	0.4	0.4	0.4	2022-11-17	1		<button>선택</button>	<button>확정</button>
신경망	0.25	0.25	0.25	2022-11-17	1		<button>선택</button>	<button>확정</button>

*정밀도(Precision) : 모델이 True라고 분류한 것 중에서 실제 True인 것의 비율
* 민감도(Sensitivity) : 실제 True인 것 중에서 모델이 True라고 예측한 것의 비율
※ F1-score : 정밀도와 민감도의 조화평균 > $2 * (\text{정밀도} * \text{민감도}) / (\text{정밀도} + \text{민감도})$

2. [확정] 버튼을 클릭한다..

- (1) 확정일자
- (2) 데이터 마스터 테이블(TB_AUTOML_MASTER) Update 처리
– 처리상태 코드 : 모델 확정