

ORACLE

코알못, 개알못도 클릭 한번으로 가능한 머신러닝

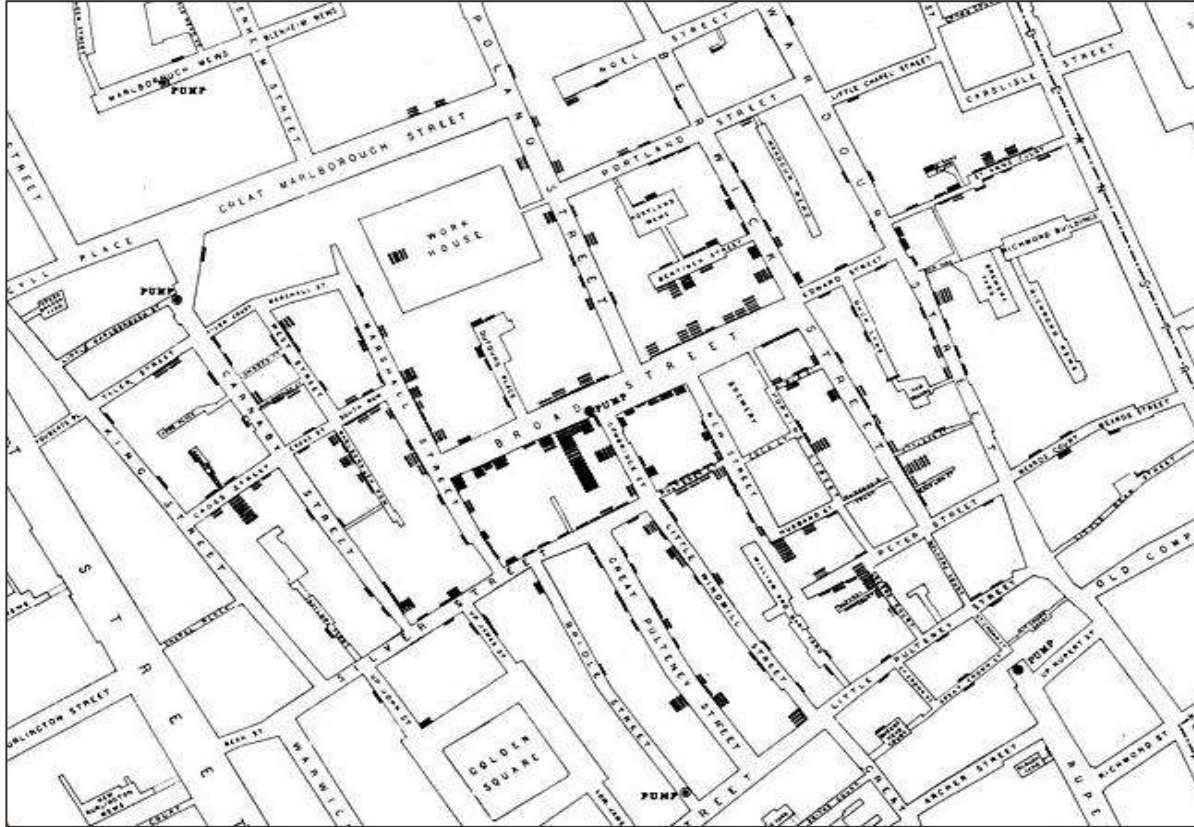
Sohee Park
Cloud Architect
Oracle, ODP Cloud

Agenda

1. 머신러닝의 이해
2. 오라클의 머신러닝
3. 비전문가를 위한 AutoML
4. 클릭기반 OAC ML



1854년 영국 런던 에서는...



1854년 영국 런던에서는...



존 스노 박사의 콜레라 지도

사용지역	가구수	사망자수	1만 가구당 사망자수
A 펌프	40,046	1,263	315
B 펌프	26,107	98	37

Broad Street에 있는 펌프 주변으로 많은 사망자수가 발생

머신러닝 기법 중 군집

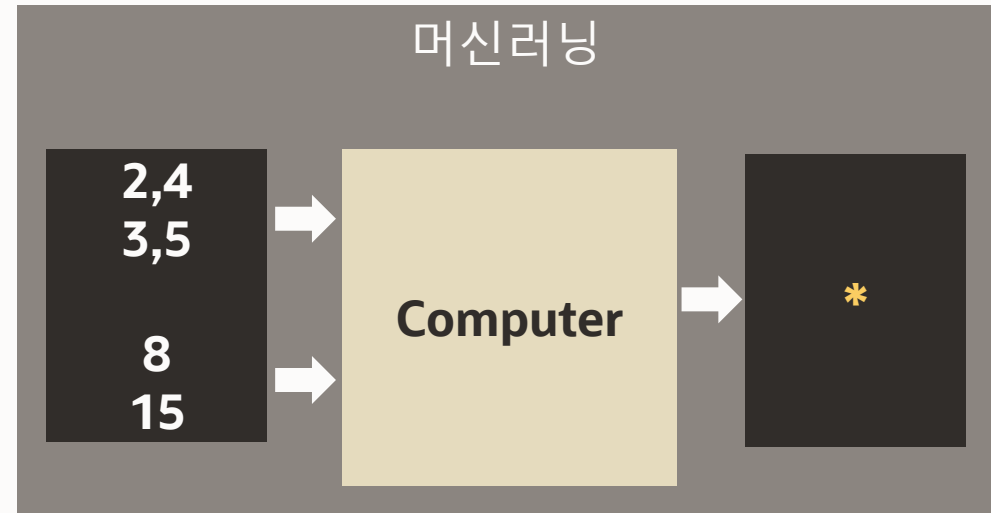
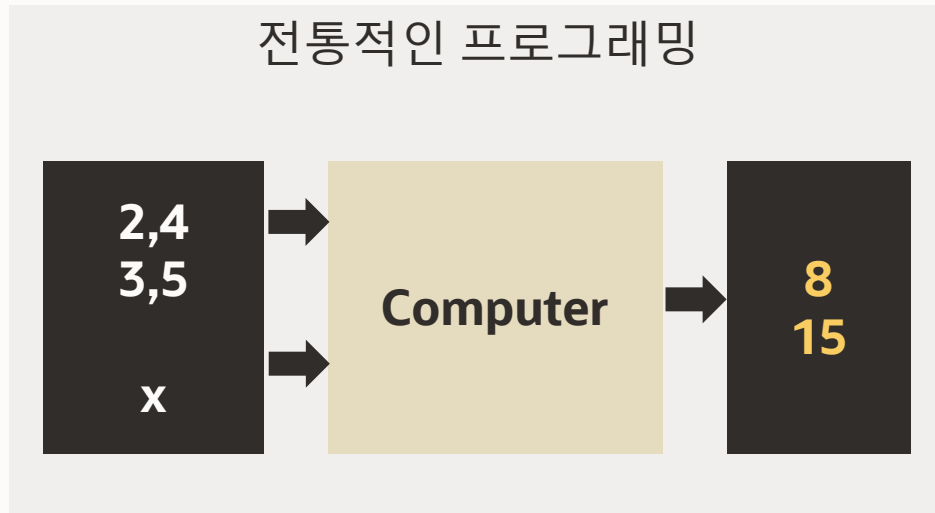
기록된 데이터를 합산하고 평균을 내고 공통점과 차이점을 찾는 것에서 시작



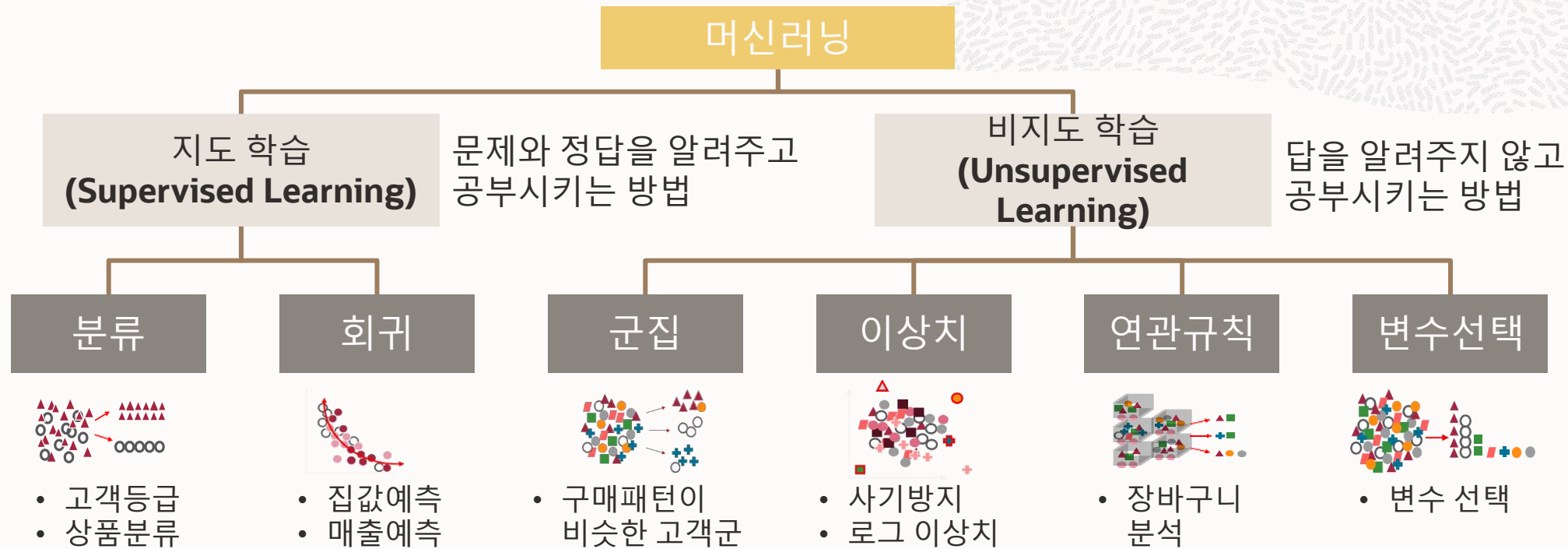
머신러닝이란

1959년 Arthur Samuel은 기계학습에 대해서 다음과 같이 정의를 했다.

머신러닝은 컴퓨터 공학의 한 분야로 통계적 기법을 사용하여 명백하게 프로그램 된 것이 아닌 컴퓨터 시스템에 데이터를 이용하여 “학습” 능력을 주는 것을 말한다.



머신러닝의 분류 및 알고리즘



- 머신러닝 알고리즘**
- K-최근접이웃(K-Nearest Neighbors)
 - 선형회귀(linear regression)
 - 로지스틱 회귀(Logistic regression)
 - 서포트벡터머신(SVM)
 - 결정트리(Decision Trees)
 - 랜덤포레스트(Random Forests)
 - 신경망분석(Neural Networks)
 - K-평균
 - 계층형 군집
 - 기대값 최대화(EM) 군집
 - One-class SVM
 - 장바구니분석(Apriori)
 - 주성분 분석(PCA)
 - 최소설명길이(MDL)



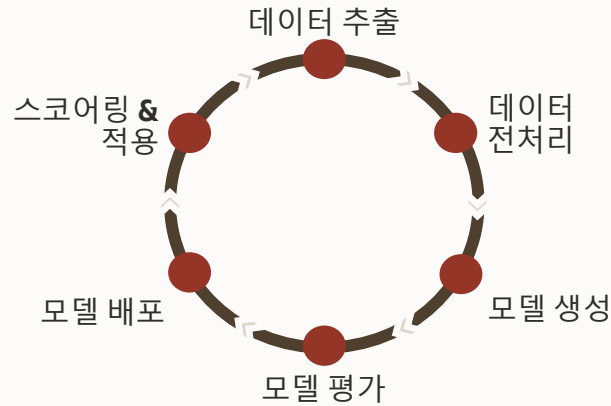
머신러닝을 하기 위해서는....

데이터 추출



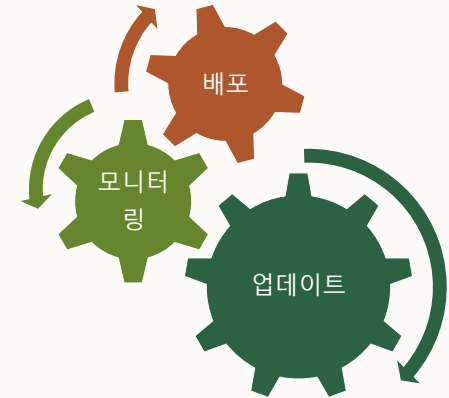
- 자사/클라우드/하이브리드 데이터 관리
- 데이터 검색
- 데이터 결합

모델 생성의 복잡성



- 다양한 역할의 사용자 필요
- 지난 작업의 재사용
- 툴 및 인프라 접근

머신러닝 운영



- 사용을 위한 모델 배포
- 모델 모니터링
- 모델 업데이트



머신러닝을 하기 위해서는...

동영상

파이썬 텐서플로우 & 머신러닝 기초 1강 - 개발환경 구축하기 ...

파이썬 텐서플로우 & 머신러닝 기초 4강 - 선형 회귀 인공지능 ...

파이썬 텐서플로우 & 머신러닝 기초 2강 - 선형 회귀(Linear ...

→ 모두 보기

- 머신러닝 예제 - 데이터 분류:붓꽃
1. 필요 라이브러리 임포트
 2. 데이터 세트 로딩
 3. 데이터 탐색 및 변환
 4. 학습데이터와 테스트 데이터 세트로 분리
 5. 학습 데이터 세트로 학습 수행 : 알고리즘 설정
 6. 테스트 데이터로 예측 수행
 7. 예측 정확도 평가

사이킷런을 이용하여 붓꽃(Iris) 데이터 품종 예측하기

```

In [2]: # 사이킷런 버전 확인
import sklearn
print(sklearn.__version__)

0.23.2

**붓꽃 예측을 위한 사이킷런 필요 모듈 로딩**

In [4]: from sklearn.datasets import load_iris # Iris 데이터 로드
        # Iris 데이터 로드 시 사이킷런 인체 있는 데이터
        from sklearn.tree import DecisionTreeClassifier
        from sklearn.model_selection import train_test_split
    
```

데이터 세트를 로딩

```

In [5]: import pandas as pd

# 붓꽃 데이터 세트를 로딩함
iris = load_iris() # Iris 데이터 로드
iris # Iris 데이터 세트 정보
    
```

```

Out[5]: {'data': array([[5.1, 3.5, 1.4, 0.2],
                        [4.9, 3., 1.4, 0.2],
                        [4.7, 3.2, 1.3, 0.2],
                        [4.6, 3.1, 1.5, 0.2],
                        [5., 3.6, 1.4, 0.2],
                        [5.4, 3.9, 1.7, 0.4],
                        [4.6, 3.4, 1.4, 0.3],
                        [5., 3.4, 1.5, 0.2],
                        [4.4, 2.9, 1.4, 0.2],
                        [4.9, 3.1, 1.5, 0.1],
                        [5.4, 3.7, 1.5, 0.2],
                        [4.8, 3.4, 1.6, 0.2],
                        [4.8, 3., 1.4, 0.1],
                        [4.3, 3., 1.1, 0.1],
                        [5.8, 4., 1.2, 0.2],
                        [5.7, 4.4, 1.5, 0.4],
                        [5.4, 3.9, 1.3, 0.4],
                        [5.1, 3.5, 1.4, 0.3],
                        [5.7, 3.8, 1.7, 0.3],
                        [5.1, 3.8, 1.5, 0.3]])
    
```

"테스트 데이터" 세트로 예측(Predict) 수행

```

In [97]: # 학습이 완료된 DecisionTreeClassifier 객체에서 테스트 데이터 세트로 예측 수행.
pred = dt_clf.predict(X_test)

In [98]: print(len(pred))
pred

30

Out[98]: array([2, 1, 2, 2, 1, 0, 0, 0, 1, 0, 0, 1, 1, 1, 0, 1, 0, 1, 2, 0, 0, 0,
                2, 0, 2, 1, 0, 2, 0, 2])

In [99]: iris.target_names

Out[99]: array(['setosa', 'versicolor', 'virginica'], dtype='<U10')
    
```

예측 정확도 평가

```

In [100]: from sklearn.metrics import accuracy_score
print('예측 정확도: {:.4f}'.format(accuracy_score(y_test, pred)))

예측 정확도: 1.0000

-> 93%의 정확도로 Iris 데이터의 품종을 예측함.
-> 학습값이 높다고 무조건 정확도가 좋은 건 마님 (이 결과는 0.2 > 0.3 > 0.1 순으로 좋음)
-> test_size, random_state 값을 조절함에 따라 예측 정확도가 달라진다
    
```

"학습 데이터" 세트로 학습(Train) 수행

```

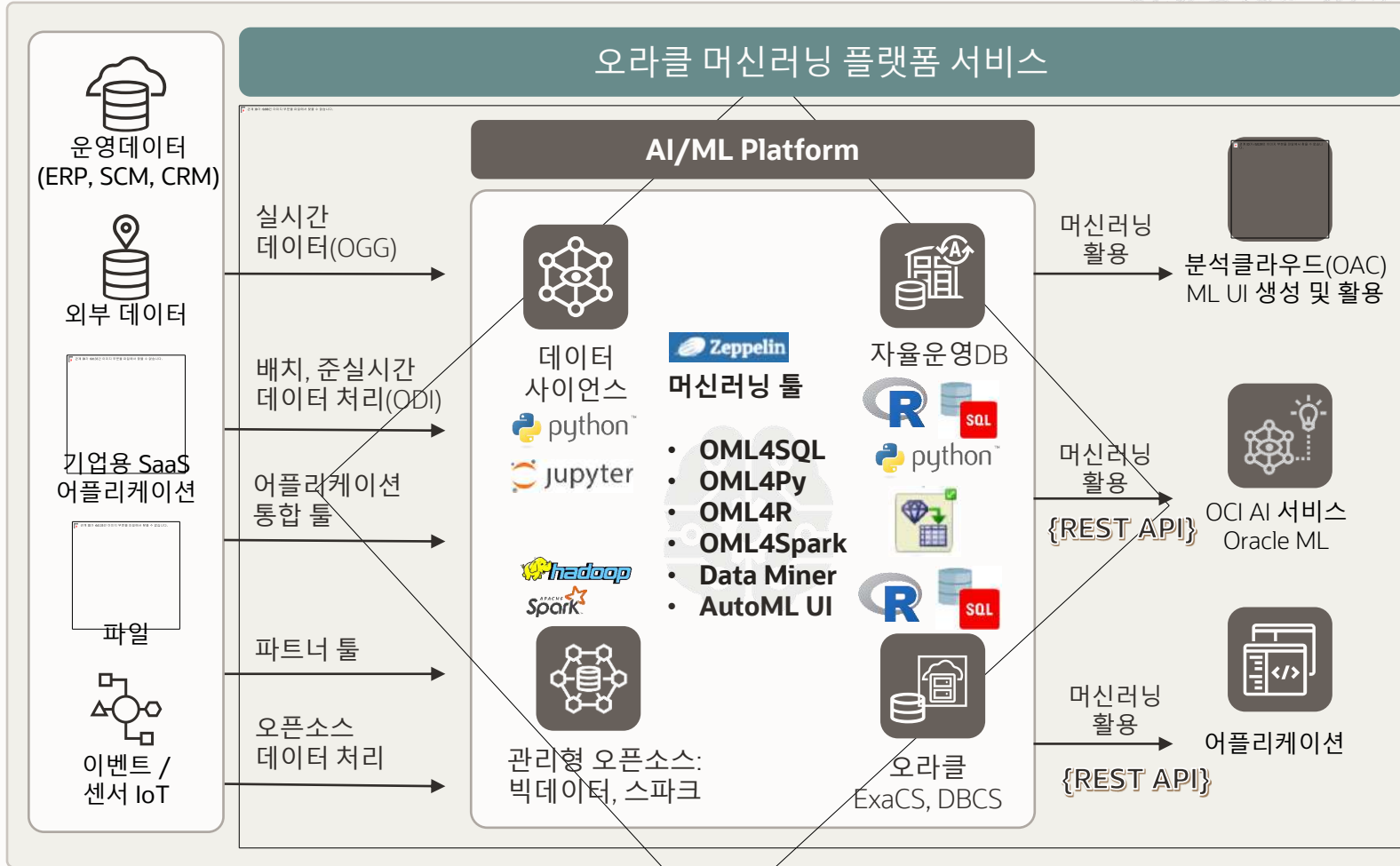
In [96]: # DecisionTreeClassifier 객체 생성
dt_clf = DecisionTreeClassifier(random_state=11)

# 학습 수행
dt_clf.fit(X_train, y_train)

Out[96]: DecisionTreeClassifier(random_state=11)
    
```



오라클 머신러닝(OML)



Ready-to-build 머신러닝 플랫폼

- ML 툴, 데이터 관리
- 분석 클라우드
- 3rd 파티

Ready-to-work 플랫폼 서비스

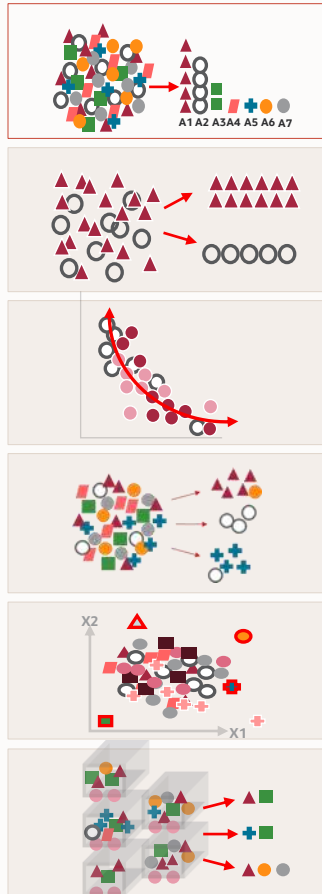
- 자율운영 데이터베이스
- 분석 클라우드(OAC)
- 빅데이터 클라우드(Hadoop)
- 데이터 플로우(Spark)
- 데이터 사이언스(Python)
- 클라우드 인프라(GPU)

Ready-to-go 어플리케이션, AI서비스

- 지능형 앱에 적용
- 지능형 UX
- 대화형 에이전트
- 언어
- 변형탐지
- 디지털 어시스턴트(챗봇)



오라클 머신러닝 알고리즘



- 가장 중요한 요소 식별(특성 중요도)
- 고객 행동 예측(분류)
- 대상 사용자 또는 항목의 프로필 찾기(분류)
- 값 예측 또는 추정(회귀)
- 인구 집단 세분화(클러스터링)
- 사기 또는 희귀 이벤트 찾기 (변칙 감지)
- "바스켓"에서 동시 발생 항목 결정 (연관)

지도학습, 비지도학습에 필요한 다양한 알고리즘

분류

- Naive Bayes
- Logistic Regression (GLM)
- Decision Tree
- Random Forest
- Neural Network
- Support Vector Machine (SVM)
- Explicit Semantic Analysis
- XGBoost

이상탐지

- One-Class SVM
- MSET-SPRT

군집

- Hierarchical K-Means
- Hierarchical O-Cluster
- Expectation Maximization (EM)

시계열

- Forecasting - Exponential Smoothing
- Includes popular models e.g. Holt-Winters with trends, seasonality, irregular time series

회귀

- Generalized Linear Model (GLM)
- Support Vector Machine (SVM)
- Stepwise Linear regression
- Neural Network
- XGBoost

변수중요도

- Minimum Description Length
- Principal Component Analysis (PCA)
- Unsupervised Pair-wise KL Divergence
- CUR decomposition for row & AI

연관어 규칙

- A priori

예측질의

- Predict, cluster, detect, features

SQL 분석

- SQL Windows
- SQL Patterns
- SQL Aggregates

변수 추출

- Principal Comp Analysis (PCA)
- Non-negative Matrix Factorization
- Singular Value Decomposition (SVD)
- Explicit Semantic Analysis (ESA)

행중요도

- CUR Decomposition

순위

- XGBoost

텍스트마이닝

- Algorithms support text columns
- Tokenization and theme extraction
- Explicit Semantic Analysis (ESA)

통계함수

- min, max, median, stdev, t-test, F-test, Pearson's, Chi-Sq, ANOVA, etc.



오라클 머신러닝의 활용

분석툴을 활용하는 OAC ML



“Invisible ML”
현업, 데이터 분석가

1 Click Insight
No Code



자연어



ML & 설명 AI



1-click 고급분석

증강분석, ML 데이터 준비

셀프보안, 패치

비전문가를 위한 Auto ML



신속한 프로토타입, 혁신
차세대분석가 / 파워사용자, DBA &
시티즌 데이터사이언스

다양한 알고리즘 내장
AutoML이 추천하는
최고의 알고리즘



데이터마이닝



데이터플로우
ML Models



장바구니분석



감성분석



공간/그래프

전문가를 위한 ML



생산성
데이터 사이언스, IT, 개발자

협업, 운영



신경망



OML 노트북



OCI Data
Science

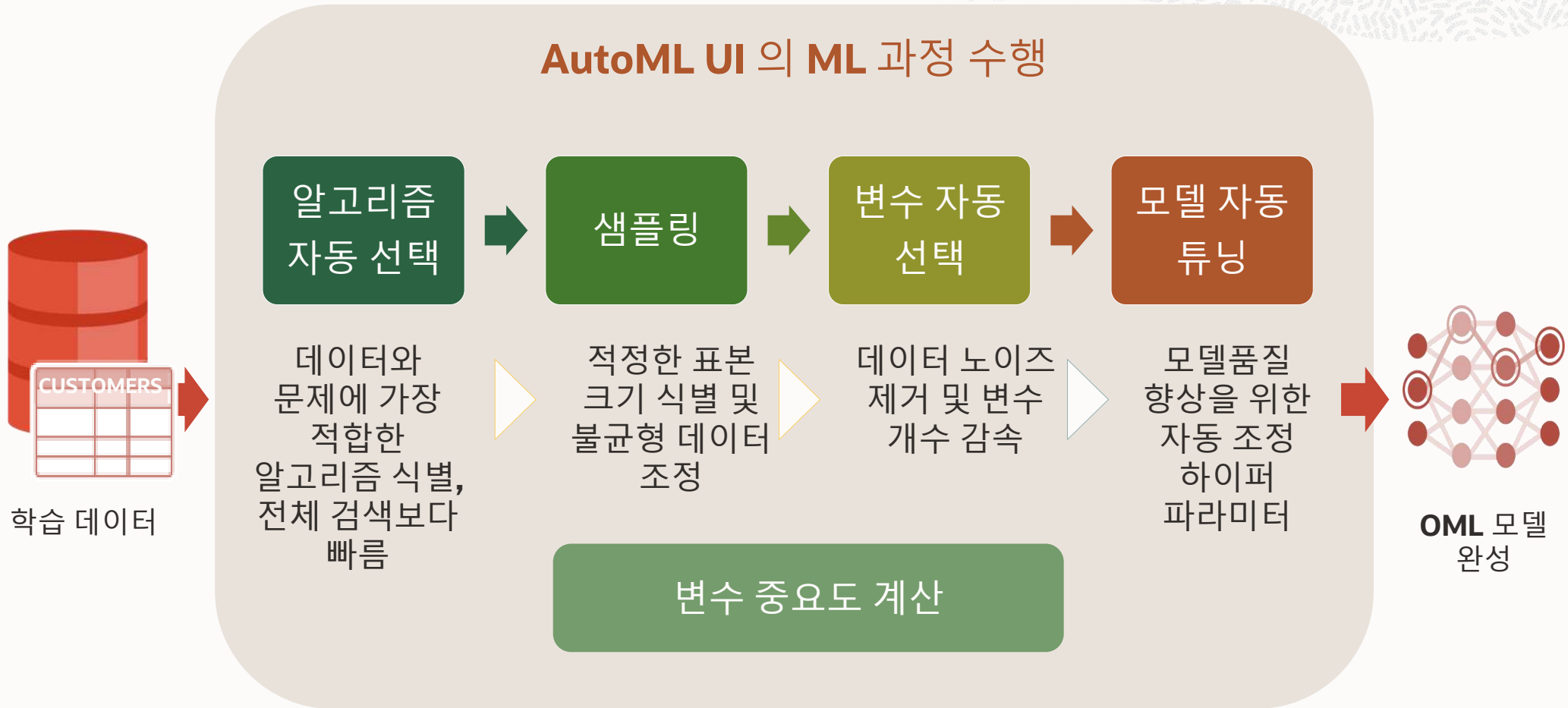


비전문가를 위한 AutoML

- 비전문가가 코딩없이 머신러닝 알고리즘을 활용하여 예측 및 배포
- 지도학습 : 회귀, 분류



비전문가를 위한 AutoML



비전문가를 위한 AutoML

자동화된 머신러닝을 지원하는 No-Code AutoML UI

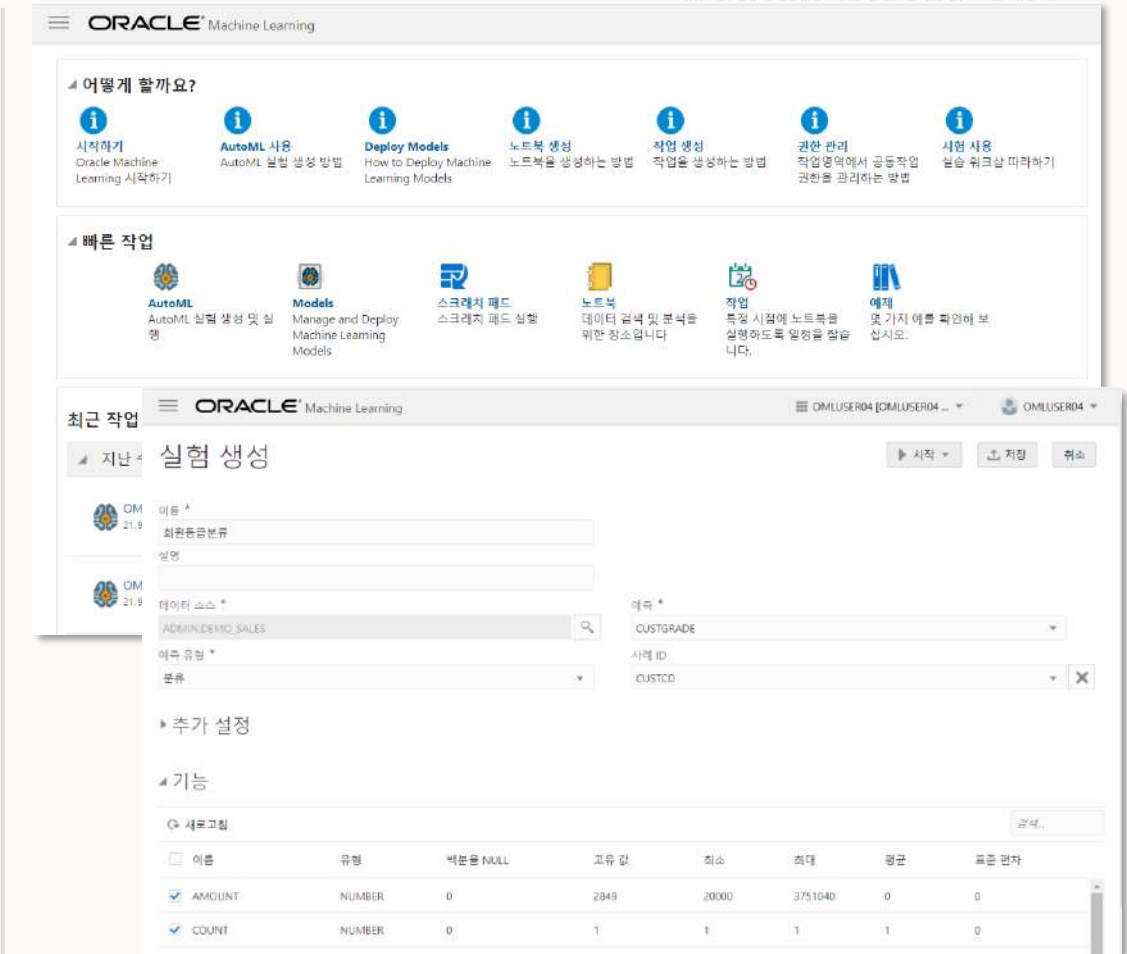
강력하고 사용하기 쉬운 UI

자동 모델 생성, 튜닝 및 배포

- 모델관리 지원
- 데이터 사이언스 생산성 강화
- **ML** 전문가가 아닌 데이터 전문가의 역량 강화

기능

- 최소 사용자 입력 : 데이터, 대상
- 모델 리더보드
- **REST** 엔드포인트를 통한 모델 배포
- 모델에서 **OML4Py** 노트북 생성



자동화된 머신러닝을 지원하는 No-Code AutoML UI

ORACLE Machine Learning
OMLUSER04 [OMLUSER04 ...]
OMLUSER04

실험 생성

▶ 시작 ⬆ 저장 취소

이름 *

회원등급예측_분류

1. AutoML 이름 입력

데이터 소스 *

ADWDEMO01.CUSTOMER_ML_MASTER_TRAINING

2. 데이터 선택

예측 *

CUST_GRD

3. 예측(대상 칼럼) 선택

예측 유형 *

분류

▶ 추가 설정

◀ 기능 **데이터 프로파일링**

↻ 새로고침 검색..

<input type="checkbox"/> 이름	유형	백분율 NULL	고유 값	최소	최대	평균	표준 편차
<input checked="" type="checkbox"/> BIRTH	NUMBER	0	47	15	61	37.08	13.33
<input checked="" type="checkbox"/> CUST_GRD	VARCHAR2	0	2				
<input checked="" type="checkbox"/> CUST_ID	VARCHAR2	0	1568				
<input checked="" type="checkbox"/> GENDER	VARCHAR2	0	2				



자동화된 머신러닝을 지원하는 No-Code AutoML UI

ORACLE Machine Learning

<- 실험

회원등급예측_분류

▶ 실험 설정 ✎ 편집

균형 조정된 정확도

리더 게시판

배지	Rename	노트북 생성	측정항목
알고리즘	모델 이름		균형 조정된 정확도
Support Vector Machine(선형)	svml_a783a0f70d		0.9962
신경망	nn_70c34c290f		0.9961
일반화 선형 모델(능형 회귀)	glm_f65c9db15c		0.9956
Support Vector Machine(가우스)	svmg_16f49f8bd3		0.9932
무작위 포리스트	rf_3de268ef68		0.9608

모델 생성시 필요한
데이터 전처리 자동화

실행 요약

- 알고리즘 선택 완료됨 ✓
- 적용형 샘플링 완료됨 ✓
- 기능 선택 완료됨 ✓
- 모델 튜닝 완료됨 ✓
- 일반화 선형 모델(능형 회귀) 완료됨 ✓
- 신경망 완료됨 ✓
- Support Vector Machine(선형) 완료됨 ✓
- Support Vector Machine(가우스) 완료됨 ✓
- 무작위 포리스트 완료됨 ✓
- 기능 예측 영향 완료됨 ✓



자동화된 머신러닝을 지원하는 No-Code AutoML UI

리더 게시판

배치 Rename 노트북 생성 측정항목 **모델이름 변경 : svml_cust_grade**

알고리즘	모델 이름	균형 조정된 정...	정확도	F1	정밀도	회수	ROC AUC
Support Vector Machine(선형)	svml_a783a0f70d	0.9962	0.9937	0.9809	0.9628	1.0000	0.9996
신경망	nn_70c34c290f	0.9961	0.9968	0.9908	0.9866	0.9955	0.9998
일반화 선형 모델(능형 회귀)	glmr_f65c9db15c	0.9956	0.9955	0.9871	0.9786	0.9959	0.9997
Support Vector Machine(가우스)	svmg_16f49f8bd3	0.9932	0.9935	0.9813	0.9702	0.9927	0.9997
무작위 포리스트	rf_3de268ef68	0.9608	0.9741	0.9196	0.9003	0.9412	0.9942

기능

새로고침

검색...

이름	중요도	유형	백분율 NULL	고유 값	최소	최대	평균	표준 편차
OFFLINESUM잡화	<div style="width: 100%;"></div>	NUMBER	0	554	0	3120480	183825.32	374623.08
OFFLINESUM남성의류	<div style="width: 100%;"></div>	NUMBER	0	1025	0	5673980	333694.49	676239.68
OFFLINECNT잡화	<div style="width: 100%;"></div>	NUMBER	0	8	0	7	0.95	1.25
OFFLINESUM여성의류	<div style="width: 100%;"></div>	NUMBER	0	997	0	7928900	450234.79	768085.61
OFFLINECNT남성의류	<div style="width: 100%;"></div>	NUMBER	0	13	0	13	2.45	2.36
BIRTH	<div style="width: 100%;"></div>	NUMBER	0	47	15	61	37.08	13.33
OFFLINECNT여성의류	<div style="width: 100%;"></div>	NUMBER	0	13	0	12	3.28	2.89
CUST_ID	<div style="width: 100%;"></div>	VARCHAR2	0	1568				



자동화된 머신러닝을 지원하는 No-Code AutoML UI

리더 게시판

배지 Rename 노트북 생성 측정항목

알고리즘	모델 이름	균형 조정된 정...	정확도	F1	정밀도	회수	ROC AUC
Support Vector Machine(선형)	svml_cust_grade	0.9962	0.9937	0.9809	0.9628	1.0000	0.9996
신경망	nn_70c34c290f						
일반화 선형 모델(능형 회귀)	glmr_f65c9db15c						
Support Vector Machine(가우스)	svmg_16f49f8bd3						
무작위 포리스트	rf_3de268ef68						

모델 세부정보 - svml_cust_grade

예측 영향 혼동 매트릭스

이름 예측 영향

- OFFLINESUM잡화
- OFFLINESUM남성의류
- OFFLINECNT남성의류
- OFFLINECNT잡화
- OFFLINESUM여성의류
- OFFLINECNT여성의류

기능

새로고침

이름	중요도	타입	0	1	2	3	4	5
OFFLINESUM잡화		NUMBER	0	8	0	7	0.95	1.25
OFFLINESUM남성의류		NUMBER	0	997	0	7928900	450234.79	768085.61
OFFLINECNT남성의류		NUMBER	0	13	0	13	2.45	2.36
BIRTH		NUMBER	0	47	15	61	37.08	13.33
OFFLINECNT여성의류		NUMBER	0	13	0	12	3.28	2.89
CUST_ID		VARCHAR2	0	1568				



자동화된 머신러닝을 지원하는 No-Code AutoML UI

Oracle Machine Learning interface showing a model deployment dialog box. The dialog box is titled "모델 배치 - svm1_a783a0f70d" and contains fields for Name (svm1_customer_grade), URI (predict_customer_grade), and Version (1.0). A red box highlights the "배치" button in the background table, with an arrow pointing to the dialog box. A red box also highlights the "공유" checkbox in the dialog box. A red arrow points from the "공유" checkbox to the text "Open API 형태로 배포".

알고리즘	모델 이름	균형	ROC AUC
Support Vector Machine(선형)	svm1_a783a0f70d	0.996	0.9996
신경망	nn_70c34c290f	0.996	0.9998
일반화 선형 모델(능형 회귀)	glm1_f65c9db15c	0.995	0.9997
Support Vector Machine(가우스)	svmg_16f49f8bd3	0.9932	0.9997
무작위 포리스트	rf_3de268ef68	0.9608	0.9942



자동화된 머신러닝을 지원하는 No-Code AutoML UI

The screenshot displays the Oracle Machine Learning AutoML UI interface. At the top, the Oracle logo and 'Machine Learning' text are visible. The main area shows a notebook titled 'predict_cust_grade' with the following code blocks:

```

Get proxy object for selected data
%python
import oml

columns = "BIRTH", "CUST_ID", "GENDER", "OFFLINECNT남성외류", "OFFLINECNT여성외류", "OFFLINECNT잡화", "OFFLINESUM남성외류", "OFFLINESUM여성외류", "OFFLINESUM잡화", "ONLINECNT남성외류", "ONLINECNT여성외류",
"ONLINECNT잡화", "ONLINESUM남성외류", "ONLINESUM여성외류", "ONLINESUM잡화", "CUST_GRD"
schema = "ADMD20081"
table = "CUSTOMER_ML_MASTER_TRAINING"

column = ','.join(columns)
query = "SELECT " + column + " FROM " + schema + "." + table

build_data = oml.sync(query=query)
%show build_data

Prepare training data
%python
import oml

X_train = build_data[:, ['BIRTH', 'CUST_ID', 'GENDER', 'OFFLINECNT남성외류', 'OFFLINECNT여성외류', 'OFFLINECNT잡화', 'OFFLINESUM남성외류', 'OFFLINESUM여성외류', 'OFFLINESUM잡화', 'ONLINECNT남성외류', 'ONLINECNT여성외류', 'ONLINECNT잡화',
'ONLINESUM남성외류', 'ONLINESUM여성외류', 'ONLINESUM잡화']]
v_train = build_data[:, 'CUST_GRD']

Build 'SUPPORT_VECTOR_MACHINES' model
%python
import oml

svm_settings = {
    "CVMC COLLECT", "CVMC COLLECT CONT", "ALGO NAME": "ALGO CHUNNOT VECTOR MACHINES", "CVMC COMPLEXTY FACTOR": "1.000", "CVMC KERNEL COEFFICIENT": "CVMC LTHRESH", "CVMC LETOUTE PARAMETER": "OML", "CVMC REGULAR": "NONE", "CVMC SCALE":
  
```

OML4Py 형태의
노트북 생성

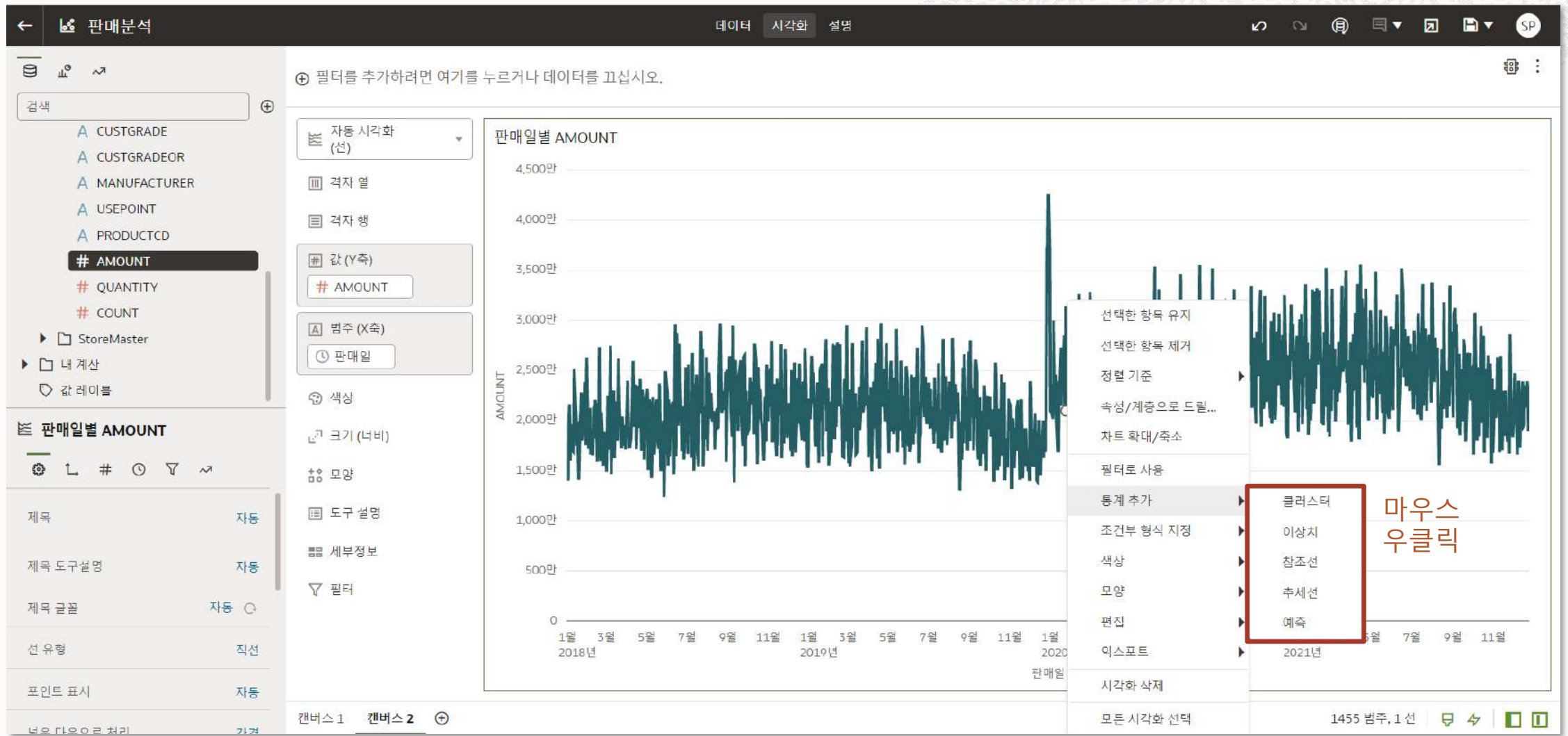


클릭기반 OAC ML

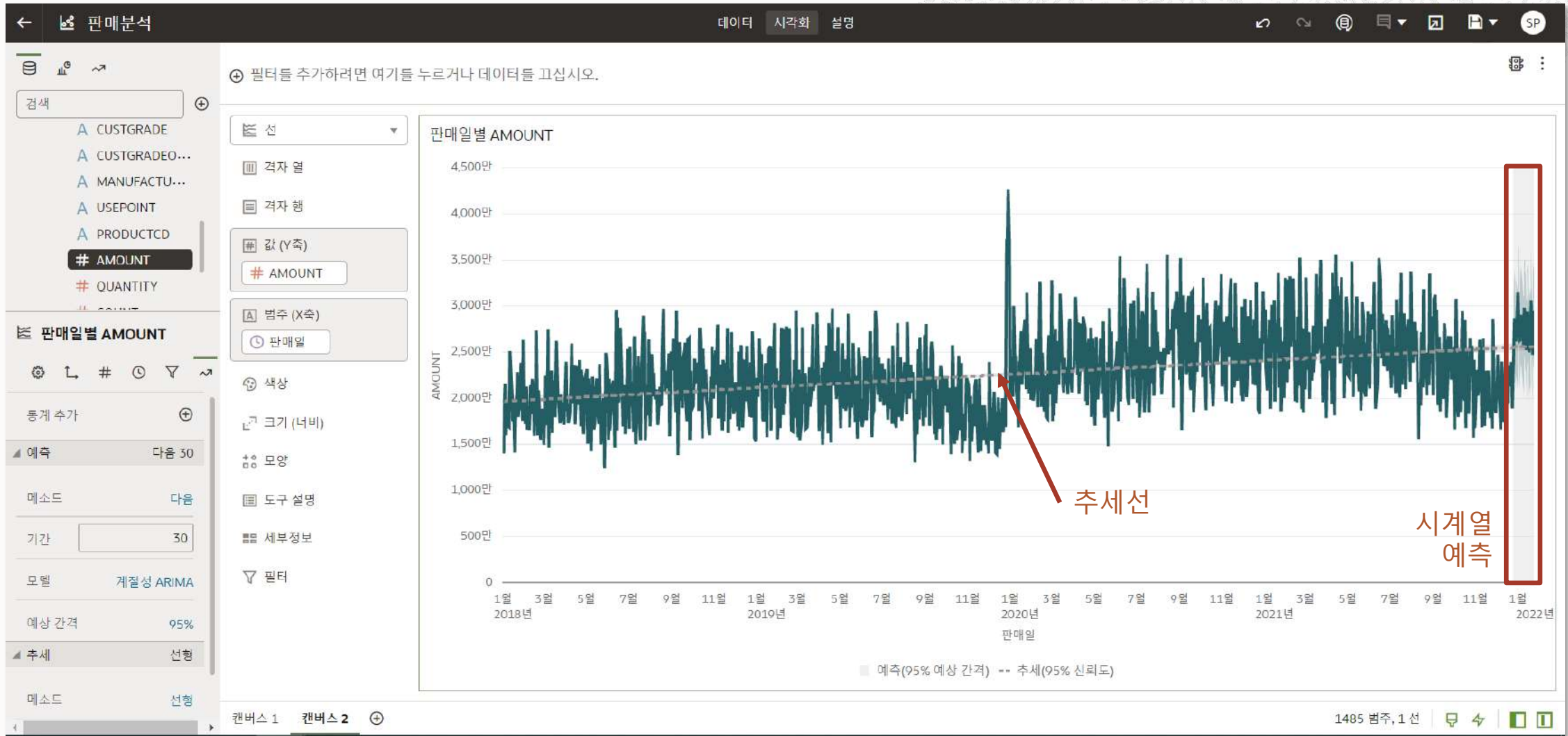
- 원클릭 머신러닝
- 클릭기반 **OAC ML**
(데이터 추출, 전처리, 모델생성, 평가 및 적용)
- **DB** 모델을 **OAC**에서 활용
- **DB** 머신러닝 알고리즘 활용



원클릭 머신러닝



원클릭 고급분석 : 추세 및 시계열 예측



클릭기반 OAC ML : 데이터 추출 및 모델 알고리즘 선택

The screenshot displays the Oracle Analytics Cloud (OAC) interface. On the left, a navigation menu lists various data and analysis tasks. The main workspace shows a 'Customer...' dataset with a '데이터 추가' (Add Data) menu open. A red box highlights the '바이너리 분류' (Binary Classification) option. A secondary dialog box titled '2. 분류 모델 학습 스크립트 선택' (2. Select Classification Model Script) is open, listing several machine learning algorithms. 'CART for model training' is selected and highlighted with a blue bar. Below the dialog, a table of data is visible.

ab CUST_ID	ab GENDER	ab BIRTH	ab CUST_GRD	ab ONLINESUM...	ab ONLINESUM...	ab ONLINESUM...	ab ONLINESUM...	ab ONLINESUM...	ab OFFLINECNT...	ab OFFLINECNT...	ab OFFLINECNT...
C00258	여	23	VIP	0	0	0	0	0	0	3	4
C02789	남	35	VIP	0	0	0	0	0	4	5	5
C04088	남	20	VIP	0	22800	0	0	0	2	8	8
C02244	여	49	VIP	0	0	0	0	0	3	5	5
C09466	여	53	VIP	0	0	0	0	0	2	4	4
C06973	남	22	VIP	0	0	0	0	0	3	4	4
C07254	여	59	VIP	0	0	0	0	0	1	5	5
C00491	남	19	VIP	0	0	0	0	0	1	7	7
C03252	여	46	VIP	0	0	0	0	0	2	3	3
C00707	남	49	VIP	0	0	0	0	0	4	5	5
C05504	남	50	VIP	0	0	0	1043040	1447750	1385000	0	6
C05540	남	40	VIP	0	0	0	1371130	2170130	1110000	0	6



클릭기반 OAC ML : 예측값 선택 및 모델 생성

모델 생성시 필요한 데이터 전처리 자동화

바이너리 분류자 학습

모델 학습 스크립트 CART for model training

* Target CUST_GRD

target, the target(label) to learn/predict

Positive Class in Target:

Positive class in the target value. Default is Yes.

Minimum Node Size:

Maximum Depth:

Maximum Confidence:

sh CUST_ID	sh GENDER	ss BIRTH	sh CUST_GRD	ss ONLINESUM...	ss ONLINESUM...	ss ONLINESUM...	ss OFFLINESUM...	ss OFFLINESUM...	ss OFFLINESUM...	ss ONLINECNT...	ss ONLINECNT...	ss ONLINECNT...	ss OFFLINECNT...	ss OF...
C06258	여	23	VIP	0	0	0	590170	1389260	977860	0	0	0	3	4
C02709	남	35	VIP	0	0	0	960920	3234610	979150	0	0	0	4	5
C04088	남	20	VIP	0	22800	0	521520	3537060	987040	0	1	0	2	8
C02244	여	49	VIP	0	0	0	825600	1232050	990470	0	0	0	3	5
C09466	여	53	VIP	0	0	0	1061880	1630680	991730	0	0	0	2	4
C06973	남	22	VIP	0	0	0	682890	1907700	999880	0	0	0	3	4
C07254	여	59	VIP	0	0	0	649440	1543530	1031400	0	0	0	1	5
C00491	남	19	VIP	0	0	0	488720	2002120	1077900	0	0	0	1	7
C05252	여	46	VIP	0	0	0	1197480	608590	1164800	0	0	0	2	3
C00707	남	49	VIP	0	0	0	596030	828930	1293400	0	0	0	4	5
C05594	남	50	VIP	0	0	0	1043040	1447750	1385000	0	0	0	1	6



클릭기반 OAC ML : 모델 결과 평가

Customer_Grade_CART
바이너리 분류 모델

저장
닫기

일반 사항

CUST_GRD에 대한 양성 레이블 VIP

품질

F1 값

0.73

액세스

모델 정확도

89%

세부정보

정밀도

61%

관련 항목

재현율

90%

가양성 비율

11%

예측 값

	VIP	OTHER	합계
VIP	45	5	50 (16%)
OTHER	29	235	264 (84%)
합계	74 (24%)	240 (76%)	314 (100%)

정확한 예측 0 100%

잘못된 예측 0 100%

혼동행렬 및 모델 정확도 평가

Customer_Grade_CART
바이너리 분류 모델

저장
닫기

일반 사항

상설 데이터

품질

Customer_Grade_CART.Statistics

액세스

Customer_Grade_CART.Hitmap

세부정보

Customer_Grade_CART.Confusion Matrix

관련 항목

Customer_Grade_CART.CARTree

학습 데이터

표시할 항목이 없습니다.

학습 스크립트

CART for model scoring

모델 예측에 주요 영향을 미치는 변수

26 Copyright © 2021, Oracle and/or its affiliates | Confidential: Internal/Restricted/Highly Restricted 9/16/2021

클릭기반 OAC ML : 예측할 데이터에 모델 적용

The screenshot displays the Oracle Analytics Cloud (OAC) interface. On the left, a sidebar contains various data manipulation options. The main workspace shows a '데이터 추가 - Customer...' dialog with a '모델 적용' (Apply Model) icon highlighted in a red box. A red arrow points from this icon to the '모델 선택' (Select Model) dialog, which lists several models. The 'Customer_Grade_CART' model is highlighted in a red box. Below the dialog, a data table is visible with columns: ab CUST_ID, ab GENDER, 99 BIRTH, ab CUST_LEV, ab CUST_GRD, and 99 ONLINESUM... The table contains three rows of data.

ab CUST_ID	ab GENDER	99 BIRTH	ab CUST_LEV	ab CUST_GRD	99 ONLINESUM...
C00415	남	55	VIP	VIP	0
C02339	여	20	VIP	VIP	0
C02879	여	31	VIP	VIP	0



클릭기반 OAC ML : 예측결과 확인 및 활용

← CustGrade_Predict
데이터 시각화 설명

검색

Customer_ML_Result

- A PredictedValue
- # PredictionConfidenc...
- A PredictionGroup
- # OFFLINESUM집합
- A CUST_ID
- A GENDER
- A BIRTH
- A CUST_LEV
- A CUST_GRD
- # ONLINESUM집합
- # ONLINESUM남성의류
- # ONLINESUM여성의류

Predict_False, CUST_ID, ...

제목 자동

제목 도구설명 자동

제목 글꼴 자동

범례 위치 자동

필터를 추가하려면 여기를 누르거나 데이터를 고십시오.

데이터

행

- A CUST_ID
- A PredictedValue
- A CUST_GRD
- # PredictionCon...

색상

- A Predict_False

크기

모양

도구 설명

필터

Predict_False, CUST_ID, PredictedValue, CUST_GRD, PredictionConfidencePercentage

CUST_ID	PredictedValue	CUST_GRD	PredictionConfidencePercentage	Predict_False
C00092	OTHER	OTHER	0.98	0
C00117	OTHER	OTHER	0.98	0
C00119	OTHER	OTHER	0.98	0
C00131	OTHER	OTHER	0.98	0
C00154	VIP	OTHER	0.91	1
C00173	VIP	VIP	0.91	0
C00178	OTHER	OTHER	0.98	0
C00194	VIP	OTHER	0.91	1
C00237	VIP	VIP	0.91	0
C00283	OTHER	OTHER	0.98	0
C00285	OTHER	OTHER	0.98	0
C00307	OTHER	OTHER	0.98	0
C00315	OTHER	OTHER	0.98	0
C00523	VIP	VIP	0.91	0
C00325	OTHER	OTHER	0.98	0
C00327	VIP	OTHER	0.91	1
C00393	OTHER	OTHER	0.98	0
C00413	VIP	VIP	0.91	0

Predict_False 0 1

Predict_False#

105

Predict_False#, CUST_GRD, PredictedValue별 Predict_False#

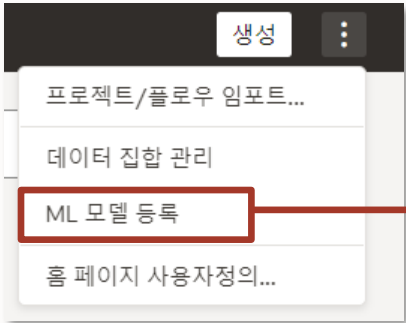
CUST_GRD	OTHER	VIP
	Predict_False#	Predict_False#
OTHER	0	101
VIP	4	0

Predict_False# 0 101

426 행, 5 열



DB 모델을 OAC에서 활용 : AutoML 에서 생성된 SVMML모델 등록



← 등록할 모델 선택

유형	이름	생성 날짜
📄	SVML_3D74C2D892	2021년 04월 2
📄	SVML_5A265D5412	2021년 06월 2
📄	SVML_658CE1072F	6일 전
📄	SVML_95A92A049D	2021년 06월 2
📄	SVML_B636EAF6BE	2021년 05월 1
📄	SVML_BD7B0105C5	2021년 06월 2
📄	SVML_CCFC06142B	2021년 08월 3
📄	SVML_CUST_GRADE	오전 4:45
📄	SVML_EFF7A605D1	오전 4:56

이름

설명

▶ 모델 정보

- ▶ 모델 클래스 CLASSIFICATION
- ▶ 알고리즘 SUPPORT_VECTOR_MACHINES
- ▶ DB 모델 이름 SVML_CUST_GRADE
- ▶ DB 모델 설명
- ▶ DB 모델 소유자 OMLUSER04
- ▶ 생성 날짜 오전 4:45
- ▶ 대상 CUST_GRD
- ▶ 입력 열
- ▶ 출력 열
- ▶ 매개변수



DB 모델을 OAC에서 활용 : ML 결과 확인

필터들 추가하려면 여기를 누르거나 데이터들 고집시오.

Predict_False, CUST_ID, Prediction, CUST_GRD, PredictionProbability

CUST_ID	Prediction	CUST_GRD	PredictionProbability	Predict_False
C09925	OTHER	OTHER	1.00	0
C09814	OTHER	OTHER	1.00	0
C09759	OTHER	OTHER	1.00	0
C09735	OTHER	OTHER	1.00	0
C09603	VIP	VIP	1.00	0
C09595	OTHER	OTHER	1.00	0
C09577	VIP	VIP	1.00	0
C09569	OTHER	OTHER	1.00	0
C09394	OTHER	OTHER	1.00	0
C09383	OTHER	OTHER	1.00	0
C09342	VIP	VIP	1.00	0
C09334	OTHER	OTHER	1.00	0
C09254	OTHER	OTHER	1.00	0
C09236	OTHER	OTHER	1.00	0
C09203	OTHER	OTHER	1.00	0
C09196	OTHER	OTHER	1.00	0
C09161	OTHER	OTHER	1.00	0
C09144	OTHER	OTHER	0.93	0

Predict_False#

3

Predict_False#, CUST_GRD, Prediction별 Predict_False#

	OTHER	VIP
CUST_GRD	Predict_False#	Predict_False#
OTHER	0	3
VIP		0

예측값이 VIP인 고객 대상으로 VIP 프로모션 캠페인 기획

Predict_False# 0 3

426 행, 5 열



DB 머신러닝 알고리즘 활용 : 장바구니 분석

데이터베이스 분석 선택

분석 작업	설명
Text Tokenization	Create text tokens from an existing de...
Sampling Data	The Oracle database ora_hash clause i...
Frequent Itemsets	Generate frequent item sets using ass...
Dynamic Clustering	This script performs dynamic clusterin...
Time Series	Time Series is a data mining function t...
Dynamic Anomaly Detection	This script performs dynamic anomaly...
Un-pivoting Data	The DB Unpivot clause allows to trans...

DB 알고리즘

- 텍스트 마이닝
- 샘플링
- 장바구니분석
- 클러스터
- 시계열
- 이상치
- Un-Pivoting



DB 머신러닝 알고리즘 활용 : 장바구니 분석

SALES_TRANS_CUST_ASSOCIATE
레이블 표시 100%

검색

- 데이터 추가
- 조인
- 행 결합
- 필터
- 집계
- 데이터 집합 저장
- Essbase 큐브 생성
- 열 추가
- 열 선택
- 열 이름 바꾸기
- 열 변환
- 열 병합
- 열 분할
- Bin
- 그룹
- 분기
- 누적 값
- 시계열 예측
- 감정 분석
- 수치 예측 학습
- 다중 분류자 학습
- 클러스터링 학습
- 바이너리 분류자 학습

SALES_TR... 데이터베이스 분석 데이터 저장

매개변수

- * Transaction Id Column: CUSTCD
Column that represents Transactions in your dataset
- * Item Value Column: PRODUCTNAME
Column that represents the Items to group into Itemsets
- * Maximum Itemset Size: 4
Maximum size of the generated frequent item set
- * Minimum Itemset Size: 2
Minimum size of the generated frequent item set
- * Minimum Support Percent: 0.25
The minimum percentage of total dataset transactions that contain this itemset (support)
- Top N Frequent Itemset: 50
Specifies number of top (ordered by support) items

Itemsetid	NumberOfItems	Item1	Item2	SupportPercentage	NumberOfOccurrences	TotalCountOfTransactions
1500	2	(모노 화이트 (OFF-WHITE)) 오브 화이트 SS 반팔티 (OMAA02755006 1020)	여성 체크 모피코트	0.32	32	10,014
1508	2	Essex Pants 106256	여성 체크 모피코트	0.32	32	10,014
1510	2	Fox Umbrellas 네이비 반자 동 Hand Made 원목 장우산 DBUM9E002	기본 카라 홈잡 자켓 RNEWJK060	0.34	34	10,014
1513	2	Fox Umbrellas 네이비 반자 동 Hand Made 원목 장우산 DBUM9E002	여성 체크 모피코트	0.39	39	10,014
1522	2	GG 자카드 패턴 니트 스카 프 / 133483 3G200 2764	여성 체크 모피코트	0.31	31	10,014
1525	2	Gloverall Gloverall Morris 모리스 Duffie Coat 원성	여성 체크 모피코트	0.35	35	10,014
1527	2	JOHNNY HATES JAZZ (JF16-J002-BK)	여성 체크 모피코트	0.34	34	10,014
1531	2	Karen Kane Piper Pant 10600837	여성 체크 모피코트	0.35	35	10,014
1534	2	LAB 화이트 실크 혼방 길라 네 반팔 롱소버 (PC9151XL61)	여성 체크 모피코트	0.37	37	10,014
1539	2	Straight Ank 10752673	여성 체크 모피코트	0.36	36	10,014

이 단계에는 데이터 미리보

distinct Items

1 transaction

= 1 Item-set



DB 머신러닝 알고리즘 활용 : 고객군별 장바구니 결과를 추천아이템으로 활용

← 제목 없음
데이터 시각화 설명

필터를 추가하려면 여기를 누르거나 데이터를 고집시오.

SALES_TRANS_CUST_ASS...

- ItemsetId
- SupportPercentage
- NumberOfItems
- NumberOfOccurrences**
- TotalCountOfTransacti...
- Item1
- Item2
- Item3
- Item4

내 계산

- # of Sets

값 레이블

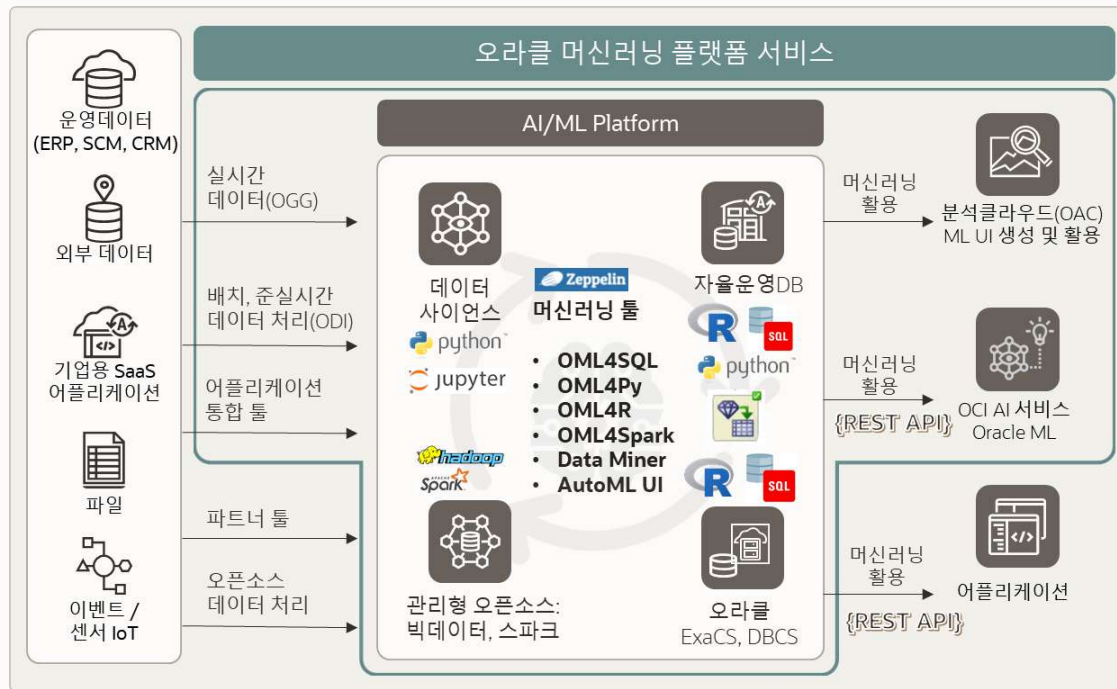
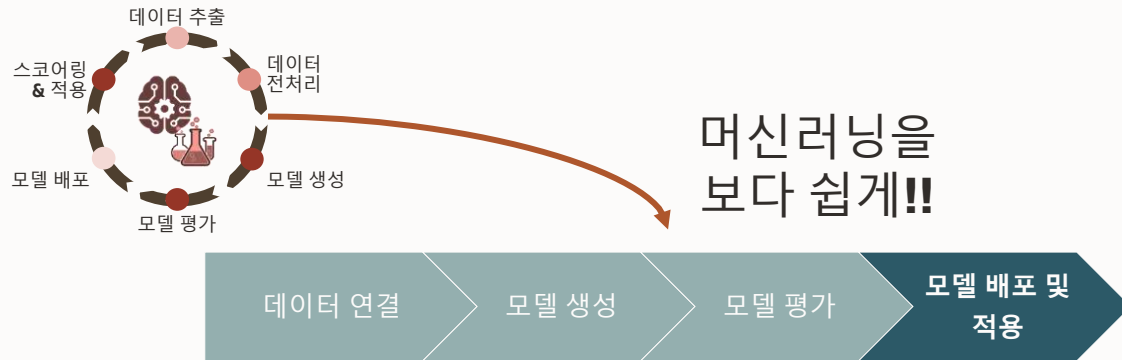
트리맵

Item1, Item2, NumberOfOccurrences													
여성 체크 모피코트	솔카라 롱 원피스...	Fox Um...	스판정...	블루 하...	모던 스...	라쿤트...	그린 Ha...	극산 옥...	구스다...	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트
화이트 슬릿 데님 미디 스커트 (...)	캐시미어 100% 신상품 베이지...	카키 헤리티지 애니 다운 BF7X38...	통브라운 네이비/카라/...	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트
트리아나 동남방...	제이제... 신수레... JJ지고르 자수 라...	이지오 캐시미어 혼방 세...	올프 디테일 GG 실크 스카프 ...	블랙 Hand Mad...	기본 카라 롱겹 자켓 R...	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트
케이엘 노카라... KWJKJ...	일포르소 일포르소 다크그린 통통더...	오일릴리 스트라... 패턴 블라우...	여성스... 색시 멩크 칼라 모피코트 들...	블랙 Hand Mad...	두블럭 폭스퍼 사선릴...	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트
				다블 자기드 GG ...	씨저스 ...	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트
				킨록 라이트... 루즈핏 ...	시그니...	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트	여성 체크 모피코트

SupportPercentage, ItemsetId, NumberOfItems, NumberOfOccurrences, Item1, Item2

ItemsetId	NumberOfItems	NumberOfOccurrences	Item1	Item2	SupportPercentage
1624	2	48	여성 체크 모피코트	올프 디테일 GG 실크 스카프 / 545936 4G001 1068	0.48
1612	2	40	스판정장/통/여름/~3XL 빅사이즈/남성정장/양복	여성 체크 모피코트	0.40
1513	2	39	Fox Umbrellas 네이비 반자동 Hand Made 원목 장우산 DBUM9E002	여성 체크 모피코트	0.39
1633	2	39	여성 체크 모피코트	케이엘 노카라자켓스커트장상세트 KWJKJD0700KWSKJD0700(울산모다)	0.39
1625	2	38	여성 체크 모피코트	오일릴리 스트라이프 패턴 블라우스(OW9SRBL025)	0.38
1620	2	38	여성 체크 모피코트	카키 헤리티지 애니 다운 BF7X380034	0.38

클릭으로 가능한 오라클 머신러닝



- 데이터 추출이 필요 없음!!
→ 데이터 이동시간 단축 및 비용 감소
- 데이터 전처리 자동화
- 클릭만으로 예측 모델 생성
- 모델평가에 필요한 결과값 제공
- 손쉬운 배포 및 적용
- 머신러닝에 필요한 다양한 알고리즘 제공
- 주요 ML 처리 단계 자동화



ORACLE

Our mission is to help people see data in new ways,
discover insights, unlock endless possibilities.