



알기쉬운코딩

05

# 빅데이터/인공지능과 함께하는 직업

## 쿠팡이니까 가능한 인공지능 혁신





## 학습내용

- 미래선도 기술
- 빅데이터/인공지능 기술 및 직업군
- 빅데이터/인공지능 사례
- 빅데이터/인공지능 데모
- 모델링 종류



## 학습목표

- 미래선도 기술에 대해 설명할 수 있다.
- 빅데이터/인공지능 기술 및 직업군에 대해 설명할 수 있다.
- 빅데이터/인공지능 사례에 대해 설명할 수 있다.
- 빅데이터/인공지능 데모를 만들 수 있다.
- 모델링 종류에 대해 설명 할 수 있다.



01

# 미래선도 기술



## 1 | 최신기술

## 미래선도 기술!

가트너 주식회사  
(Gartner, Inc.)

## Top Strategic Technology Trends for 2023

 Optimize	 Scale	 Pioneer
<ul style="list-style-type: none"><li>• Digital Immune System</li><li>• Applied Observability</li><li>• AI Trust, Risk and Security Management</li></ul>	<ul style="list-style-type: none"><li>• Industry Cloud Platforms</li><li>• Platform Engineering</li><li>• Wireless Value Realization</li></ul>	<ul style="list-style-type: none"><li>• Superapps</li><li>• Adaptive AI</li><li>• Metaverse</li></ul>
Sustainable Technology		

Source: Gartner  
775990

Gartner

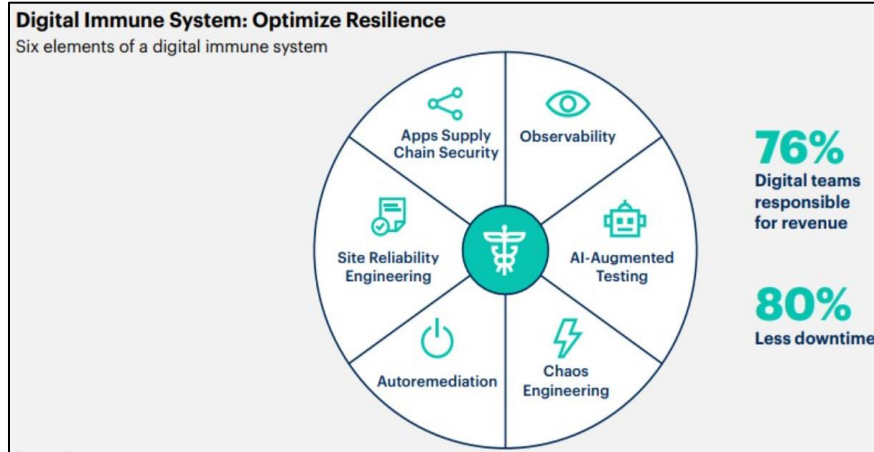
- 미국의 정보 기술 연구 및 자문 회사
- 매년 동향을 발표하며 IT계에서는 리서치를 통해 다양한 방향성을 제시

## 2 | 최신기술(Optimize)

## 1 Trend No. 1 : Digital Immune System

## 운영에 대한 데이터 기반의 인사이트, 자동화된 테스트, 애플리케이션 공급망의 보안

- 기술 : 빅데이터, 인공지능, 보안
- 예 : 2025년까지 기업이 투자 시 시스템  
다운타임 80%감축 → 수익증가



## 2 | 최신기술(Optimize)

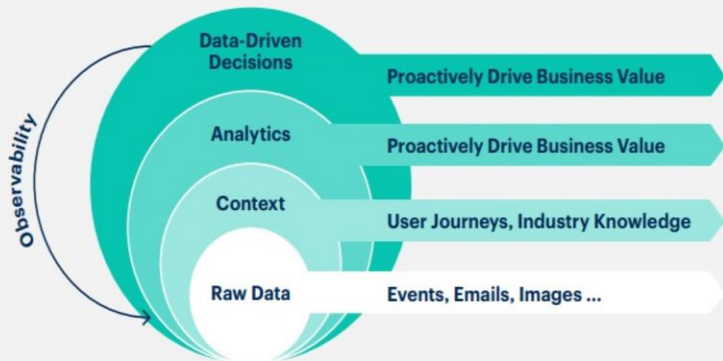
2 Trend No. 2 : **Applied Observability (\*)**

관찰 가능한 데이터 로그, 파일 다운로드 등 유저가  
특정 행동을 취할 경우 나타나는 내용을 반영

- 기술 : AI, ML, 빅데이터
- 예 : 조직데이터를 활용한 미래  
수요예측

**Applied Observability: Optimize Operations**

AI analyzes data from, and the context of, prior decision making to drive faster and more accurate future decisions.



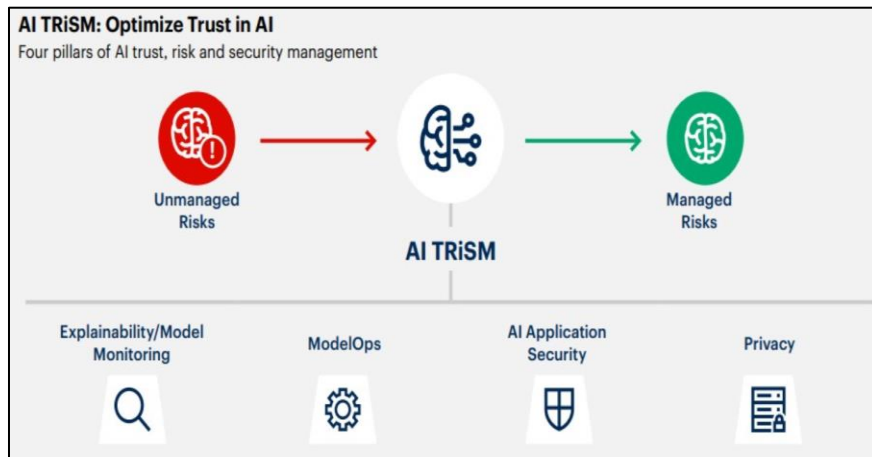


## 2 | 최신기술(Optimize)

## 3 Trend No. 3 : AI 신뢰, 리스크 및 보안관리

조직이 모델의 안정성, 신뢰성, 데이터 보호를 보장 시 높은 수준의 비즈니스 서비스를 구축할 수 있음

- 기술 : 리스크 및 보안관리
- 예 : 조직의 41%가 AI관련 개인정보 침해 또는 보안사고 경험

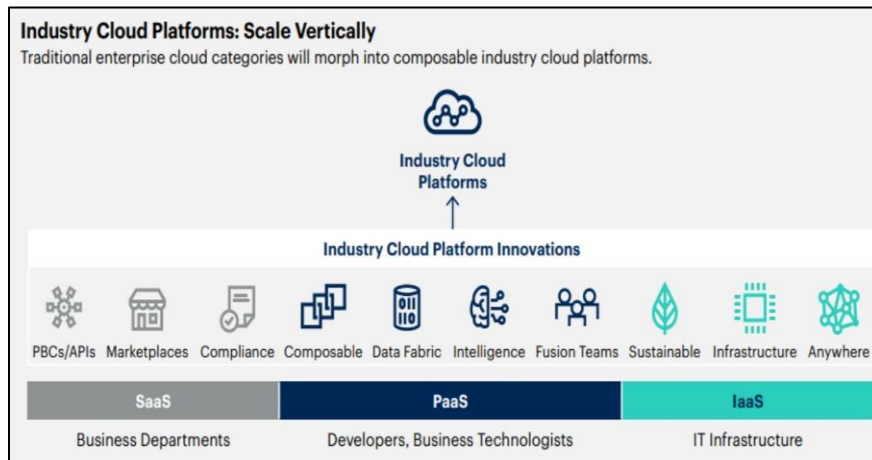


## 3 | 최신기술(Scale)

## 1 Trend No. 1 : Industry Cloud Platform

## SAAS, PAAS, IAAS를 통합하여 서비스 제공

- 기술 : SAAS, PAAS, IAAS
- 예 : 2027년까지 기업의 50% 이상이 산업 클라우드 플랫폼을 사용하여 비즈니스 가속화



## 3 | 최신기술(Scale)

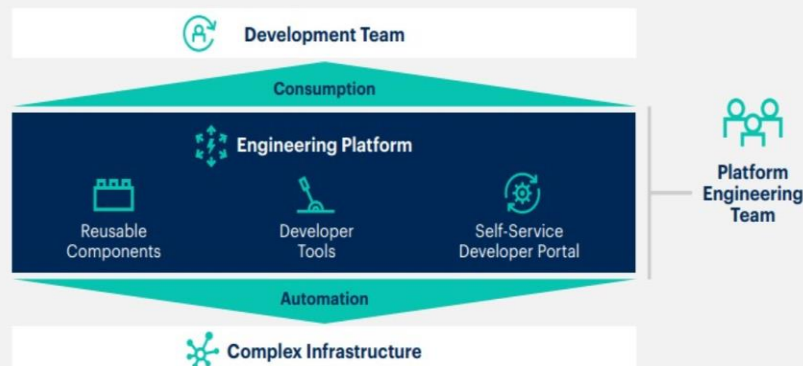
## 2 Trend No. 2 : Platform Engineering

## 소프트웨어를 제공하고 수명 주기를 관리하기 위해 셀프 서비스 내부 개발자 플랫폼 구축

- 기술 : 플랫폼 엔지니어, 웹개발 기술
- 예 : 배달의 민족 플랫폼 자동으로  
고객이 사용  
(개발/사용자 중간에 위치한  
기술 지원팀)

### Platform Engineering: Scale Delivery

Conceptual outline of an example platform and its principal components



## 3 | 최신클기술(Scale)

3 Trend No. 3 : **Wireless Value Realization**

기업들은 사무실 내의 와이파이, 모바일 디바이스 서비스,  
저전력 서비스 등 다양한 무선 솔루션을 사용

- 기술 : 단순 연결이 아닌 데이터를 통한  
인사이트 제공
- 예 : 네트워크가 직접 비즈니스 가치의  
원천

**Wireless-Value Realization: Scale Everywhere**

Different devices utilizing different technology create value beyond communication.

**People-Focused****Things-Focused****End-User Devices****Edge Devices****RFID****Network Demand****Wi-Fi5, Wi-Fi6, Wi-Fi7****4G, LTE, 5G****BLE, Zigbee, EnOcean****HF, UHF, VHF**

## 4 | 최신기술(Pioneer)

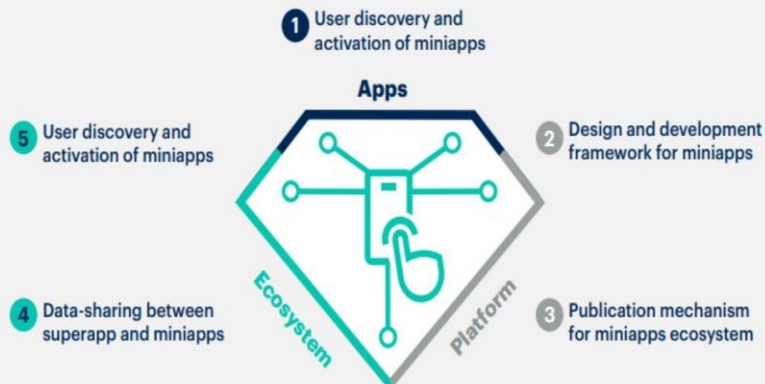
## 1 Trend No. 1 : Super Apps

## 앱, 플랫폼 및 생태계의 기능을 하나의 애플리케이션에 결합

- 기술 : 모바일 앱, 슬랙/팀즈와 같은 앱 개발 기술
- 예 : 고객 또는 직원의 사용을 위해 여러 앱을 통합

## Superapps: Pioneer Engagement

Five characteristics of a superapp

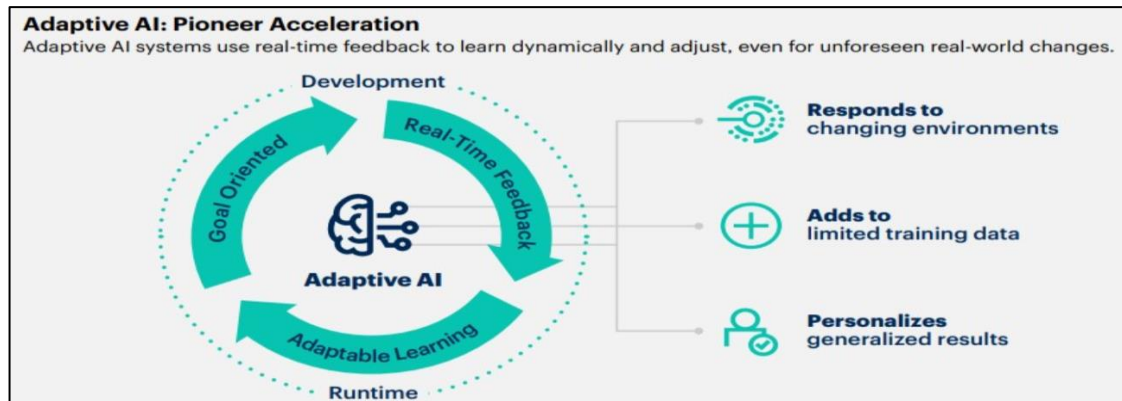


## 4 | 최신클기술(Pioneer)

## 2 Trend No. 2 : Adaptive AI

새로운 데이터를 기반으로 런타임 및 개발 환경 내에서  
모델을 지속적으로 재교육하고 학습해 초기 개발 단계 당시  
존재하지 않았거나 예측 불가능한 실제 상황의 변화에 신속하게 적응

- 기술: AI, ML 기술

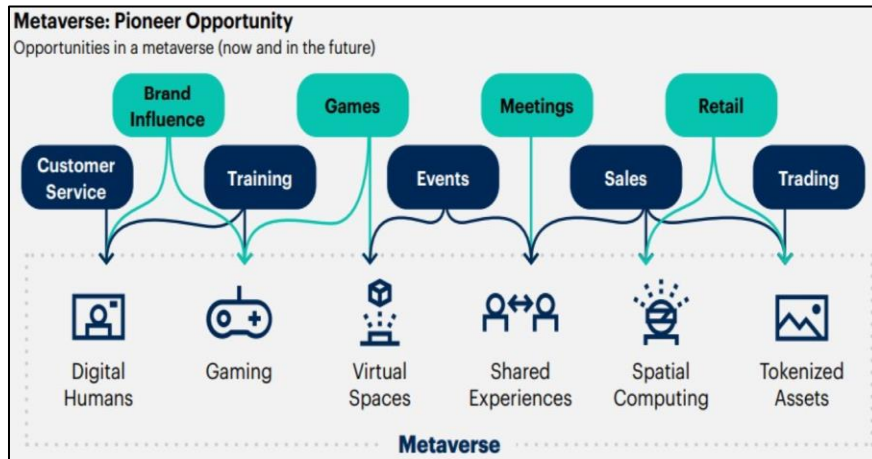


## 4 | 최신기술(Pioneer)

## 3 Trend No. 3 : Metaverse

## 메타버스는 지속적이며 향상된 몰입 경험을 제공

- 기술 : 가상환경에서 디지털화폐, NFT
- 예 : 메타버스 환경에서 아바타 구매

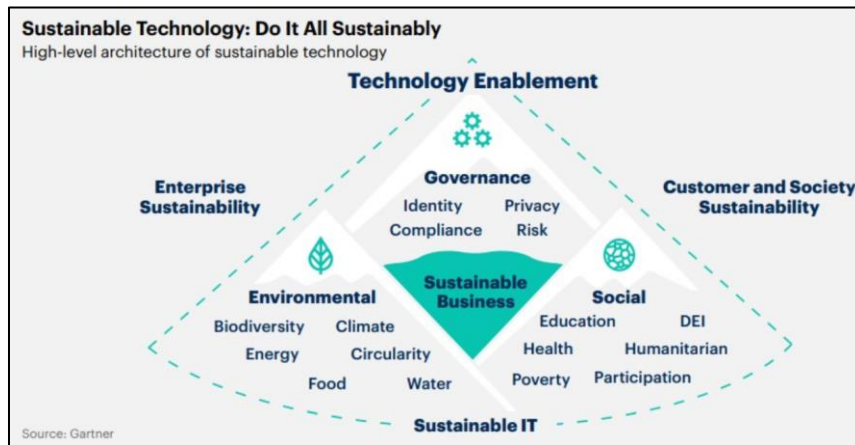


## 5 | 최신기술

## 1 Trend No. 1 : Sustainable Technology

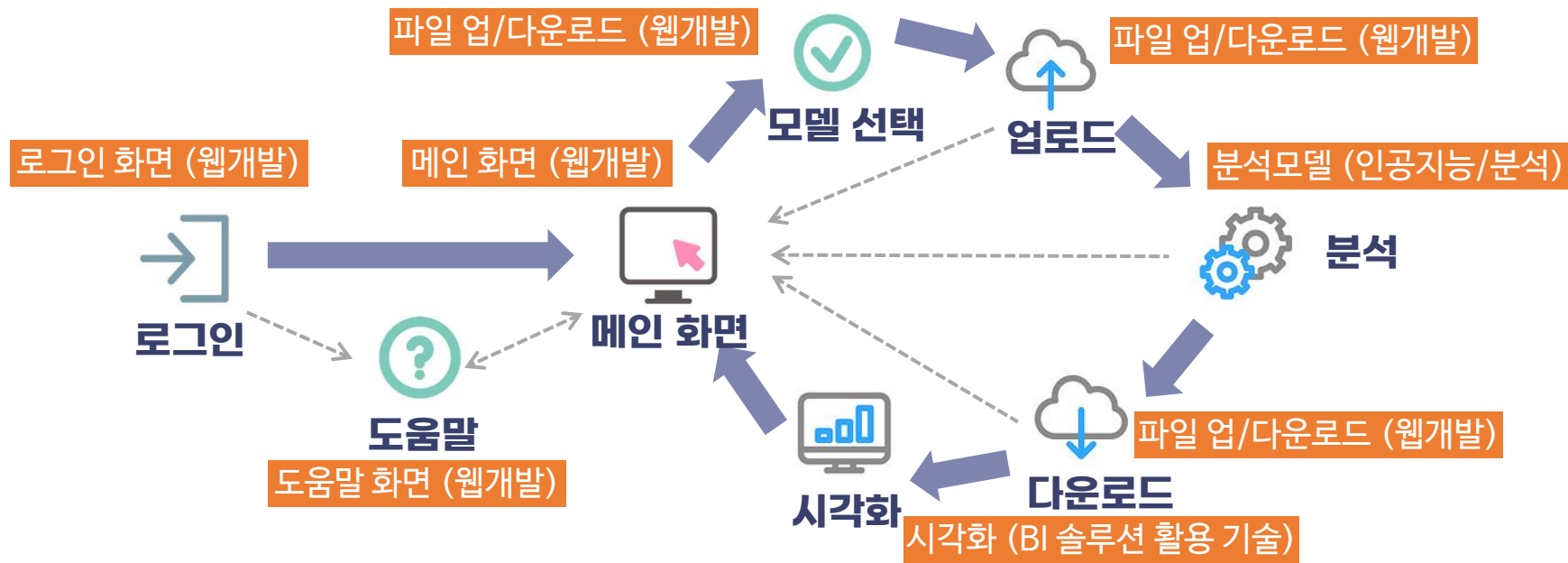
## 최근 전략기술을 모두 포함

- 기술 : 지속 가능한 기술 프레임워크
- 예 : 지속적인 IT프레임워크를 통해 데이터 분석, AI와 같은 기술들을 통해 기업의 지속 가능성 실현





## 6 | 최근 개발 트렌드



운영체제 (운영체제)

리눅스 운영체제 (데이터: 데이터베이스)

데이터 (SQL)



02

## 빅데이터/인공지능 기술 및 직업군



### 1 | 빅데이터 기술

#### 빅 데이터 개념 (3V)

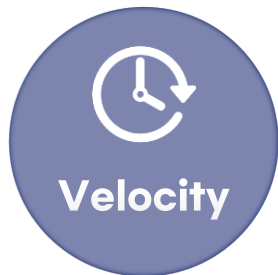


Volume

기가 단위 이하



기가 단위 이상



Velocity

시간/분



실시간



Variety

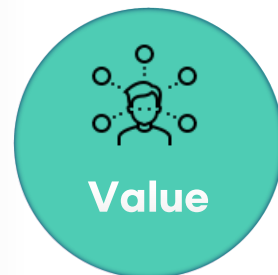
샘플 데이터



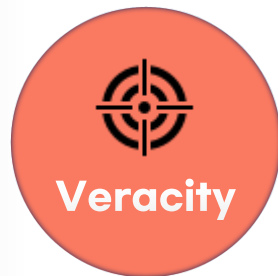
전체 데이터

(IoT device SNS, News)

#### 확장 (5V)



Value



Veracity

1 | 빅데이터 기술

## 왜 필요 할까요?

다음 주 TV 생산량은  
몇 대 인가요?



빅 데이터를 빠르게 분석하여  
적시에 정확한 정보를 제공



공급/수요  
판단 오류

일반 플랫폼  
(늦은 정보제공)



공급/수요  
밸런스 유지

빅 데이터 분석 플랫폼  
(빠른 정보제공)

### 분석 알고리즘

수요 예측

패턴 생성

이상치 제거

### 분석 데이터

과거 판매실적

연휴 일정

판매가격

1 | 빅데이터 기술

# 왜 필요 할까요?

다음 주 TV 생산량은  
몇 대 인가요?



빅 데이터를 빠르게 분석하여  
적시에 정확한 정보를 제공



공급/수요  
판단 오류

일반 플랫폼  
(늦은 정보제공)



공급/수요  
밸런스 유지

빅 데이터 분석 플랫폼  
(빠른 정보제공)

빅데이터  
플랫폼  
엔지니어

분석 알고리즘

수요 예측

패턴 생성

이상치 제거

데이터  
분석가

분석 데이터

과거 판매실적

연휴 일정

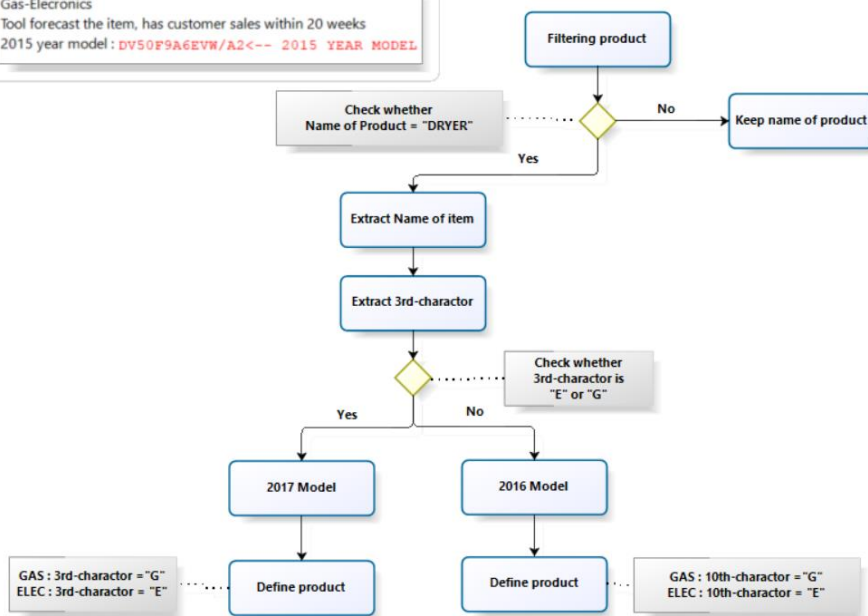
판매가격

데이터  
엔지니어

## 1 | 빅데이터 기술

### 설계서

Gas-Electronics  
Tool forecast the item, has customer sales within 20 weeks  
2015 year model : DV50F9A6EVM/A2<-- 2015 YEAR MODEL



### 1 | 빅데이터 기술

## 개발코드

### 1. Extract Name of Item

```
invalue = "DV50F9A6EVW/A2"  
modelAfter2017 = True
```

### 2. Extract 3rd-character

```
charData = invalue[2]
```

```
if ((charData == "E") ||  
    (charData == "G")):  
    modelAfter2017 = True  
else:  
    modelAfter2017 = False
```

### 3. Define Model

```
modelDefine = ""  
if modelAfter2017 == True:  
    if charData == "G":  
        modelDefine = "GAS"  
    else:  
        modelDefine = "ELEC"  
else:  
    if invalue[12] == "GAS":  
        modelDefine = "GAS"  
    else:  
        modelDefine = "ELEC"
```

```
print( "해당 모델은 {} 입니다.".format(modelDefine) )
```

해당모델은 ELEC 입니다.

## 1 | 빅데이터 기술

### 사례

#### 기술

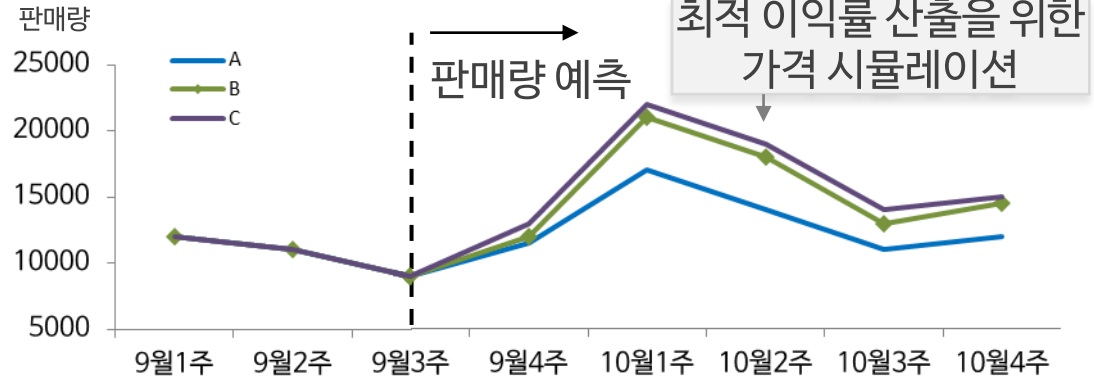
자동화 / 시간단축  
1주 → 1시간

#### 효과



어려운  
가격결정 회의

간단한  
이슈확인 회의



시뮬레이션	가격할인 비율	이익률
A	20%	17%
B	30%	23%
C	40%	21%

“혁신적인 경영관리 환경 구축가능”





03

## 빅데이터/인공지능 사례



## 1 | 판매/마케팅 분야

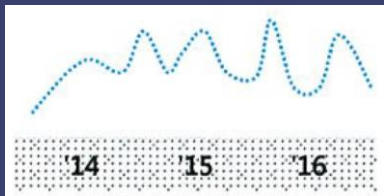
### 가치

- 수요판매 예측을 통해 판매 극대화 및 선제적 시장 대응
- 정확한 판매량 예측을 통한 프로모션 효과분석

### 활용기술

- 이벤트(프로모션/날씨/휴일 등) 기반 예측
- 최신 판매 트렌드 학습 기반 예측

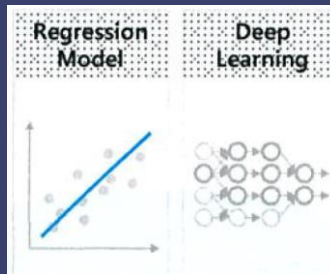
### 과거 실적 & 프로모션 등의 자료



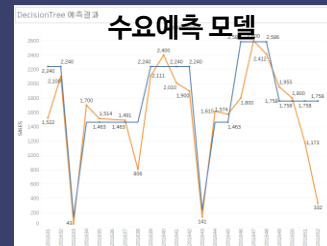
### 프로모션 정보



### 머신러닝/ 딥러닝



### 예측



## 2 | 금융분야

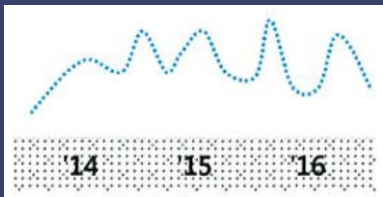
### 가치

- 부정거래 추적을 통한 사기예방 (금융결제원)
- 부정거래 자동감지

### 활용기술

- 사기유형 패턴 기반 예측
- 강화학습을 통한 사기패턴 참조데이터 강화

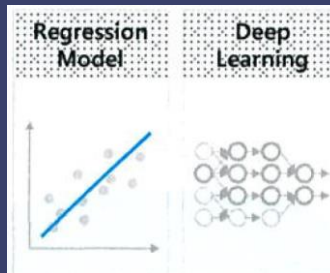
### 타 은행간 거래 실적정보



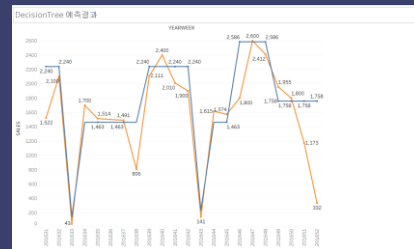
### 외부정보



### 통계적 분석 / 머신러닝



### 예측



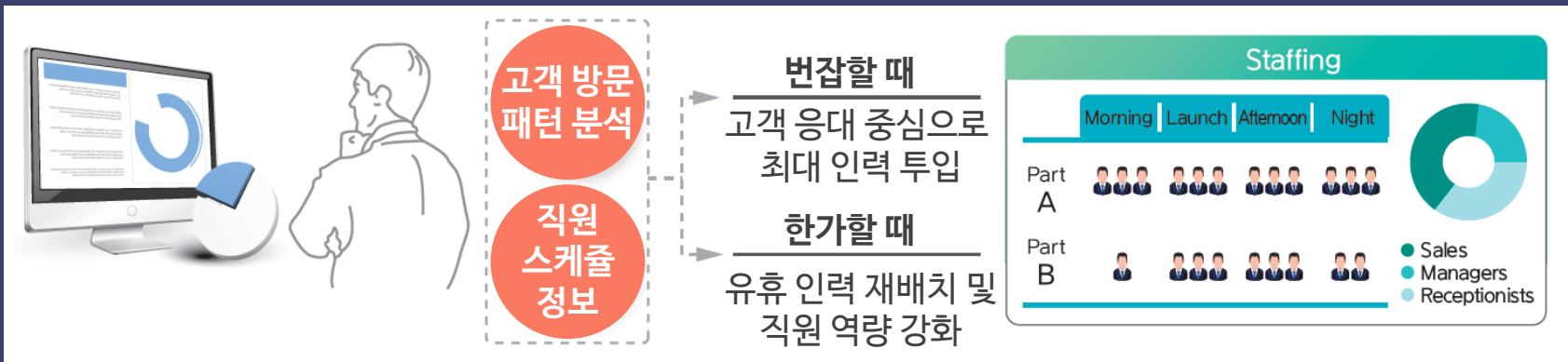
## 3 | 매장별 방문고객 예측

### 가치

- 방문객 수 예측결과 및 인건비, 구매전환율 예측
- 고객 방문패턴에 맞춘 직원 업무투입 최적화

### 활용기술

- 방문객수 예측 알고리즘
- 직원업무 투입 최적화 알고리즘



## 4 | 인프라 이상징후 감지

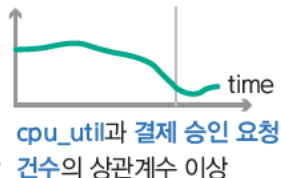
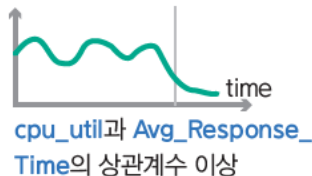
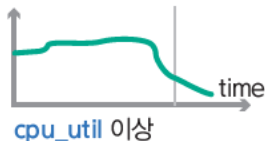
## 가치

- 빅데이터 분석에 기반한 장애 분석 및 예방 체계구축

## 활용기술

- 로그분석 기반 이상징후 감지  
: 단어 및 패턴 출현회수 측정 및 정상범위 산출
- 상관분석기반 이상징후 감지(시계열 분석)

구분	매트릭
DB	mem_usage
WEB-WAS	평균 응답 시간
WEB-WAS	메시지 서비스 개수
로그	Denied 출현 회수
Application 특화	결재 승인 요청 건수



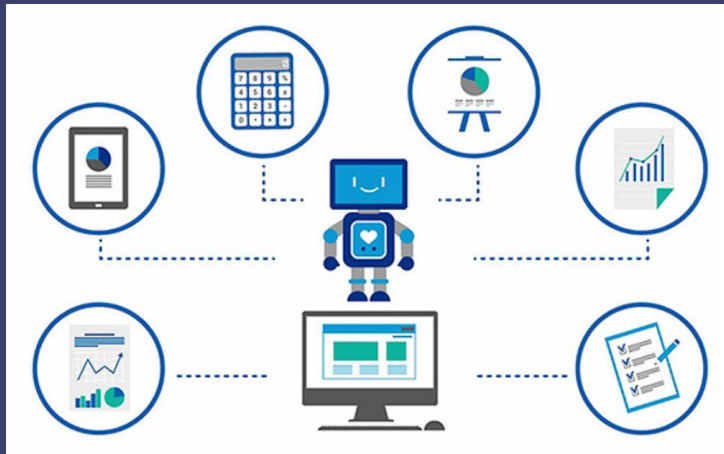
## 5 | 자동화 기술(초자동화)

### 가치

- 사람이 수동으로 작업하는 부분을 컴퓨터가 담당

### 활용기술

- 브라우저 자동화 기술



웹 크롤링									
								0	1
0								디미나나 후랑상제 구운계란60구, 1개, 2,100g	10,300
1								오복유물 HACCP인증 구운계란 2만60구, 60구, 2만	13,900
2								참나루촌 무염통계란, 30개입, 1.2kg(편입)	6,200
3								문한 먹반식 구운계란 30구 1만, 30개입, 1.2kg	7,900
4								맛군 흑흑 물깃 구운 계란, 30개입, 1박스	7,900
5								강동한 간이버어 있는 흑흑한 반숙계란, 50g, 30개입	16,900
6								[계란사육] 먹반식 구운계란 구운란 60구 (2개), 2700g	11,900
7								진주탈 오마이 포켓 머주리알 5p, 25g, 10개입	9,440

공공데이터 수집									
연도	월	전국 PIR	서울 PIR	부산 PIR	대구 PIR	인천 PIR	광주 PIR	대전 PIR	울산 PIR
0	2004	3	4.21	4.89	3.95	3.73	4.65	2.81	4.68
1	2004	4	4.39	5.59	3.91	3.88	4.59	2.92	3.83
2	2004	5	4.19	5.14	4.90	3.83	4.78	3.41	4.19
3	2004	6	4.09	4.38	4.20	3.77	4.30	2.83	4.19

대기오염 나쁨 위치		관측소위치	
0	중구	서울특별시 중구 덕수궁길 15시청서소문별관 3동	
1	청계천로	서울 중구 청계천로 184(청계천4가사거리 남강빌딩 앞)	
2	용산구	서울 용산구 한남대로 136서울특별시중부기독교육원	
3	강변북로	서울 성동구 강변북로 257한강사업본부 옆	
4	홍릉로	서울 동대문구 홍릉로 1(청량리전철역 사거리 SC제일은행 앞)	



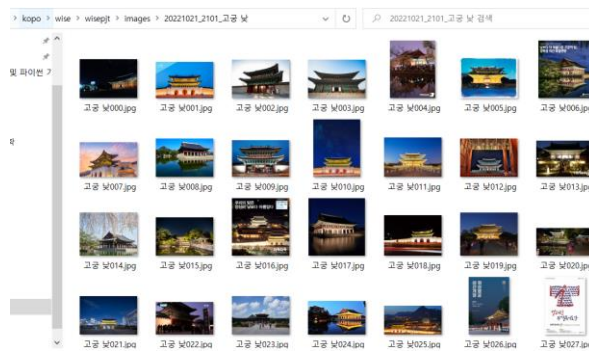
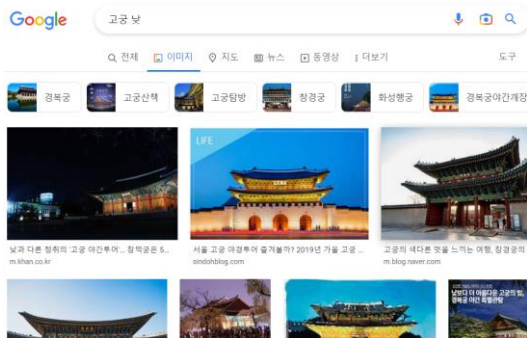
04

## 빅데이터/인공지능 데모



## 1 | 데모(초자동화)

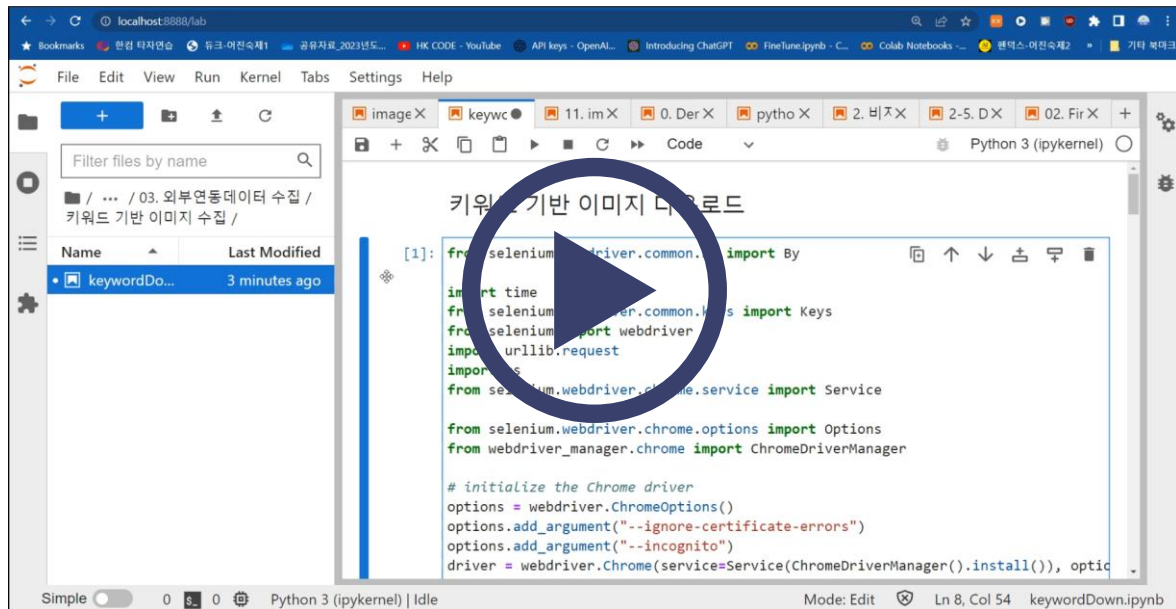
- 1 구글 이미지 웹에 접속한다.
- 2 키워드를 입력한다.
- 3 검색된 이미지를 다운로드 받는다.





## 1 | 데모(초자동화)

## 키워드 기반 이미지 다운로드 실습



The screenshot displays a Jupyter Notebook environment with the following components:

- File Explorer (Left):** Shows a directory structure with a file named `keywordDo...` modified 3 minutes ago.
- Code Editor (Center):** Contains a Python script for downloading images based on keywords. The code includes imports for Selenium WebDriver, time, urllib, and Service. It also shows the initialization of a Chrome driver with specific options.
- Output (Right):** Displays the title `키워드 기반 이미지 다운로드` (Keyword-based Image Download).
- Bottom Bar:** Indicates the current mode is `Simple` and the kernel is `Python 3 (ipykernel)`. The status bar shows `Mode: Edit`, `Ln 8, Col 54`, and the filename `keywordDown.ipynb`.

```
[1]: from selenium.webdriver.common import By
import time
from selenium.webdriver.common.keys import Keys
from selenium import webdriver
import urllib.request
import os
from selenium.webdriver.chrome.service import Service

from selenium.webdriver.chrome.options import Options
from webdriver_manager.chrome import ChromeDriverManager

# initialize the Chrome driver
options = webdriver.ChromeOptions()
options.add_argument("--ignore-certificate-errors")
options.add_argument("--incognito")
driver = webdriver.Chrome(service=Service(ChromeDriverManager().install()), optio
```

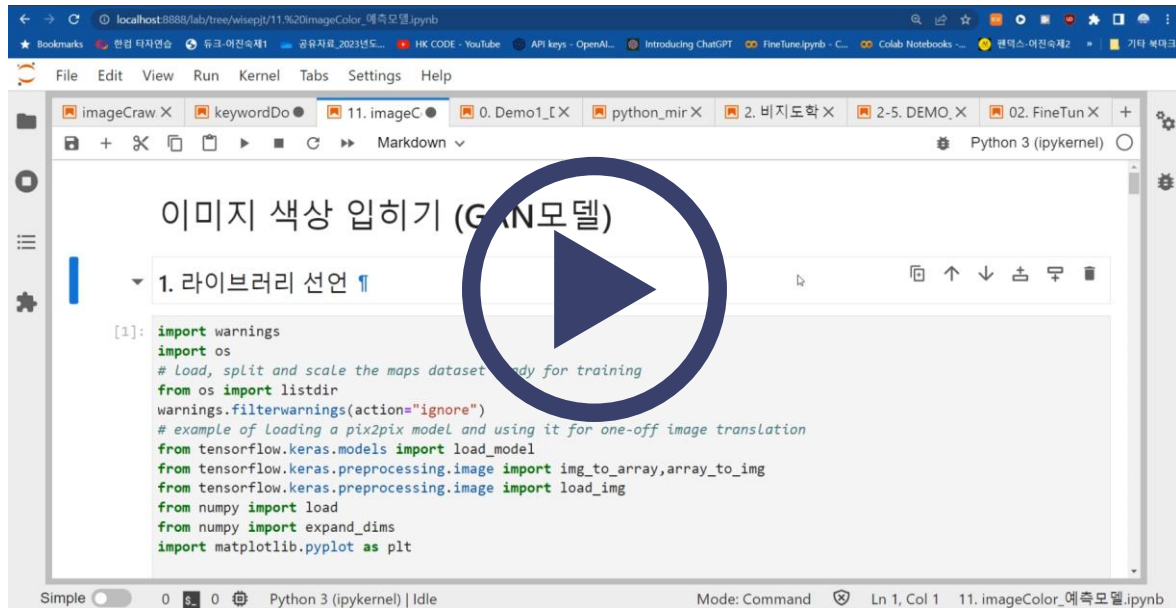
## 2 | 데모(인공지능)

- 1 흑백 컬러 이미지 쌍을 준비한다.
- 2 기계를 학습 시킨다.
- 3 아무 흑백이미지나 던지면 컬러로 변경해준다.



## 2 | 데모(인공지능)

## 이미지 색상 입히기 실습



localhost:8888/lab/tree/wisepjt/11.%20imageColor\_예측모델.ipynb

File Edit View Run Kernel Tabs Settings Help

imageCraw X keywordDo 11. imageC 0. Demo1\_L X python\_mir X 2. 비지도학 X 2-5. DEMO\_X 02. FineTun X

Markdown Python 3 (ipykernel)

## 이미지 색상 입히기 (GAN모델)

### 1. 라이브러리 선언

```
[1]: import warnings
import os
# load, split and scale the maps dataset ready for training
from os import listdir
warnings.filterwarnings(action="ignore")
# example of loading a pix2pix model and using it for one-off image translation
from tensorflow.keras.models import load_model
from tensorflow.keras.preprocessing.image import img_to_array, array_to_img
from tensorflow.keras.preprocessing.image import load_img
from numpy import load
from numpy import expand_dims
import matplotlib.pyplot as plt
```

Simple 0 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 11. imageColor\_예측모델.ipynb



05

## 모델링 종류



## 1 | 분석 프로세스

## 데이터 레이크 환경

데이터 탐색  
및 준비

## 데이터 웨어하우스 환경



장기 저장용 테이블 형태

분석 및 시각화용  
데이터

## 데이터 분석/시각화



## 데이터 분석

통계적 접근  
(Statistics)

A

## 통계방식 활용

- Time-Series
- Clustering using quantile
- Moving average

통계 기법을 활용하여 문제를 풀어내는 방법

(머신러닝) 지도 학습  
(Supervised Learning)

B

회귀  
(Regression)

- Linear Regression
- Decision Tree

C

분류  
(Classification)

- SVM (Support Vector Machine)
- Logistic Regression

답에 영향을 주는 특성을 기계에 학습시킨 후 미래에 특성만을 활용하여 답을 예측하는 방법

(머신러닝) 비지도 학습  
(Unsupervised Learning)

D

클러스터링  
(Clustering)

- K-means
- Gaussian Mixture Model

답이 없는 상황에서 주어진 데이터를 특성에 맞게 가장 잘 설명하는 방법



컴퓨터가 이해하는 코드는  
어느 바보나 다 짤 수 있다.  
훌륭한 프로그래머는  
사람이 이해할 수 있는 코드를 짤다  
- 마틴 파울러 (Martin Fowler)



## 3 | 엔지니어의 역할

순서	대상	엔지니어의 역할
1	데이터 탐색	분석 작업을 위한 필요 데이터를 정의하고 유/무 확인
2	연동 데이터 정의	저장된 데이터의 시스템이름, 저장타입(RDB/FILE), 용량, 획득주기 등을 문서화한다.
3	데이터 정제	비정상 데이터를 정제한다. (또는 유효데이터라 하더라도 데이터를 사용할 수 있도록 가공한다.)
4	분석모델 적용	문제를 해결하기 위한 최적의 분석모델을 적용한다. (데이터 사이언티스트)
5	검증 및 테스트	분석모델의 정확도를 검증한다.
6	분석모델 튜닝	개선점을 모델에 반영한다.
7	배포	프로그래머 없이 주기적으로 작동하도록 시스템화 한다.
8	유지보수	VOC를 접수 받아 모델을 보완/개선 한다.



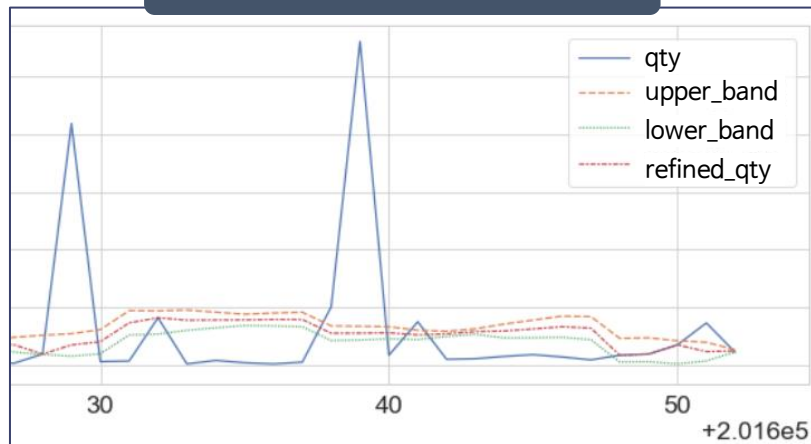
## 4-1 | 데모(통계적 접근)

## [개념] - 통계함수 활용 A

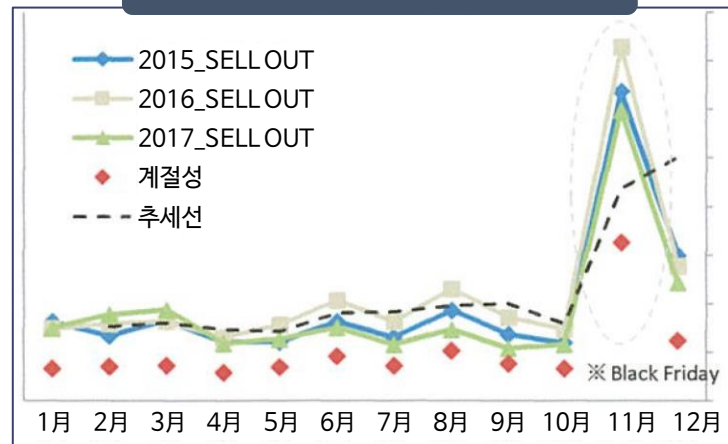
$$FCST\ QTY = \left[ \begin{array}{c} \text{기본예측모델} \\ \text{판매량} \cdot \text{계절성} \\ \text{지수} \cdot \text{지수} \end{array} \right] \cdot \left[ \begin{array}{c} \text{추가모델} \\ \text{PRICE} \cdot \text{제품} \cdot \text{지역} \cdot \text{MKT} \\ \text{AGE} \cdot \text{EVENT} \cdot \text{비용효과} \end{array} \right]$$

할인행사 기간, 날씨, Cannibalization, ...

## 계절성 지수 (주차별 효과)



## 판매량 예측



## 4-1 | 데모(통계적 접근)

PRODUCT GROUP	PRODUCT	ITEM	YEARWEEK	SALES
REF	2DOOR	REF56A01	201801	50
REF	2DOOR	REF56A01	201802	100
REF	1DOOR	REF56B02	201801	30
REF	1DOOR	REF56B02	201802	60

LV1 제품이 속한 제품군

REF 냉장고

LV2 ITEM이 속한 제품

2DOOR

1DOOR

상위 계층의 주차별 효과

LV3 실제 구매한 ITEM

REF56A01

REF56A02

REF56B01

REF56B02

실 판매제품 예측

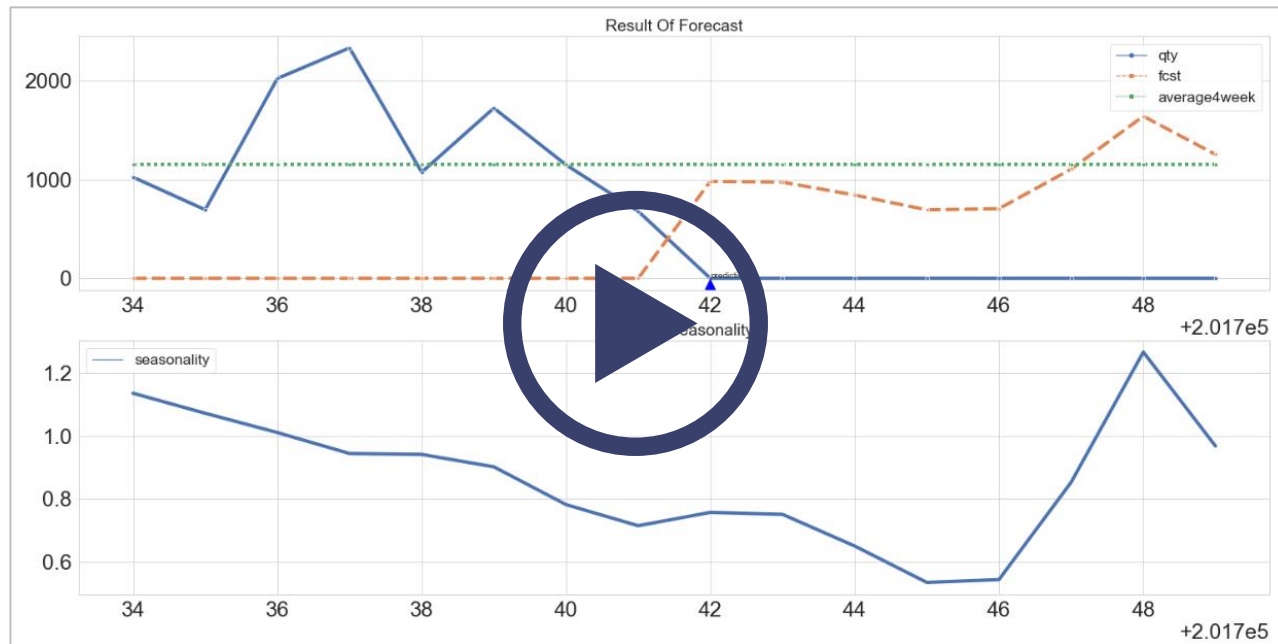
## 4-1 | 데모(통계적 접근)

## [세부내용] - 통계함수 활용

## 이동평균법을 활용한 계절성 지수 산출방법

주차		1	2	3	4	5	6	7	8	9	...	45	46	47	48	49	50	51	52
1	실제 거래량	513589	438251	420290	458431	482381	570412	620392	571781	569545	...	621592	644949	681873	110366 5	835398	752613	863189	119717 0
2	추세 거래량	500536	509442	516120	-	-	-	-	-	실제거래량에 대한 이동평균 (13)			-	-	-	-	-	831470	866666
3	변동률	7818	7275	6858	-	-	-	-	-	거래추세량에 대한 표준편차 (5)			-	-	-	-	-	36844	30520
4	정상 범위	508354 ~ 492718	516716 ~ 502167	522978 ~ 509262	-	-	-	-	-	상한: 거래추세량 + 변동률 하한: 거래추세량 - 변동률			-	-	-	-	-	868315 ~ 794626	897187 ~ 836146
5	정제된 거래량	508355	502167	509262	-	실제거래량 > 상한가 → 상한가 실제거래량 < 하한가 → 하한가 실제거래량											-	868315	836146
6	스무딩 처리	506955	508139	509330	-	이동평균 (3)											-	811030	830280
7	계절성 지수	1.013/ 1.003	0.86/ 0.99	0.83/ 1	-	계절성 지수 = 실제 거래량 / 스무딩 처리 계절성 지수 = 정제된 거래량 / 스무딩 처리											-	1.45/ 1.07	0.98/ 1.007

## 4-1 | 데모(통계적 접근)

[시연] - 통계함수 활용 **A**

구분	대상
개발언어	Python
개발 플랫폼	Anaconda
개발환경	Jupyter notebook
파일명	Minimal Forecasting

**예측 모델 데모  
(통계적 접근)**

4-2 | 데모(지도학습)



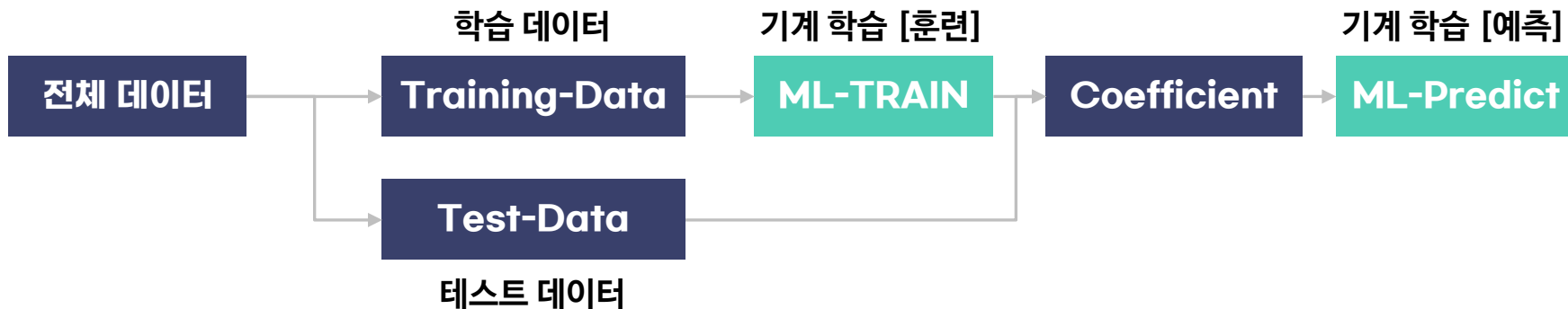
Features(휴일정보)



Label(판매량)



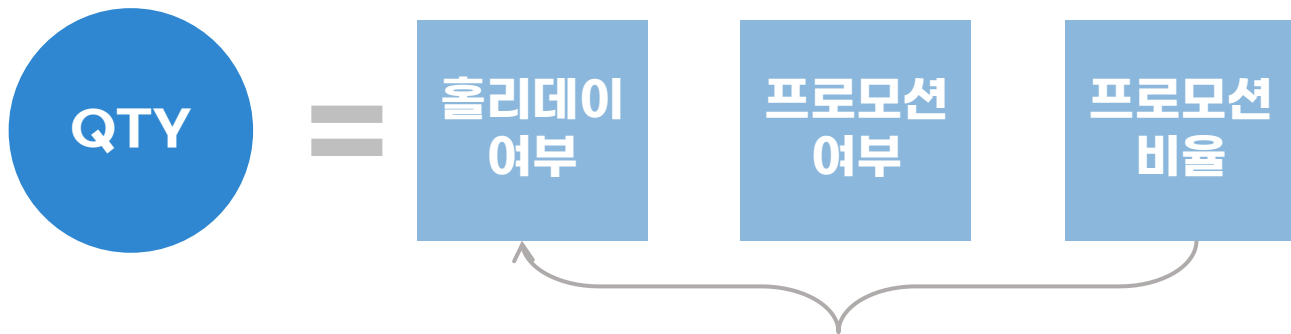
Features(프로모션 정보)



## 4-2 | 데모(지도학습-Regression)

## [개념] - 회귀방식 (Decision Tree)

과거 거래량에 영향을 주는 특성(프로모션, 홀리데이)과의 관계를 학습한 후  
미래의 특성 값만을 활용하여 거래량 예측



## 4-2 | 데모(지도학습-Regression)

## [개념] - 회귀방식 (Decision Tree)

과거 거래량에 영향을 주는 특성(프로모션, 홀리데이)과의 관계를 학습한 후  
미래의 특성 값만을 활용하여 거래량 예측

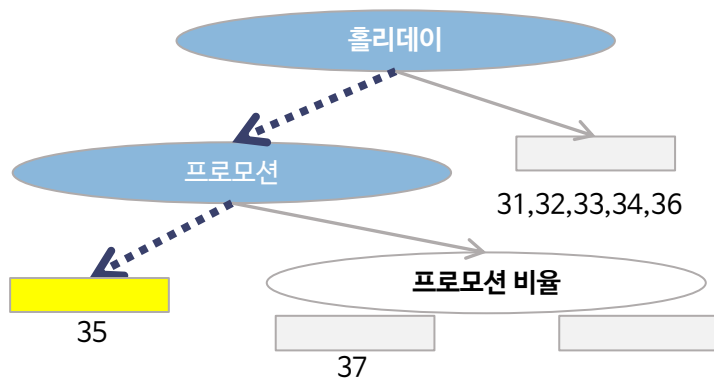
## 데이터 탐색

미래

과거

주차 정보	대상	홀리데이 여부	프로모션 여부	프로모 션 비율	거래량
38	A	NO	NO	-	?
37	A	NO	YES	10	32
36	A	YES	YES	30	52
35	A	NO	NO	-	12
34	A	YES	NO	-	14
33	A	YES	YES	50	24
32	A	YES	NO	-	12
31	A	YES	NO	-	16

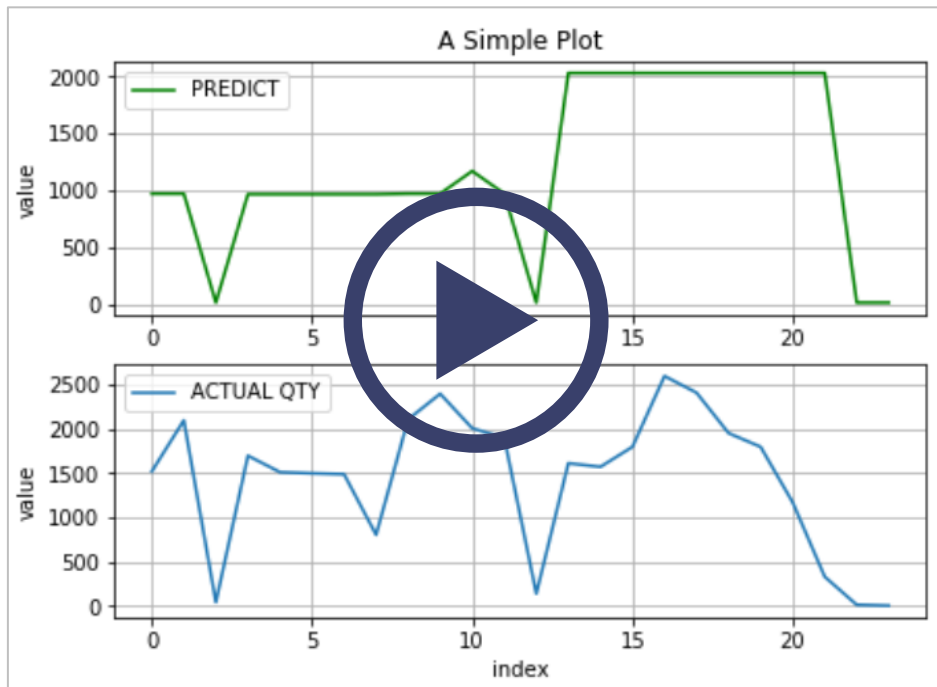
## 모델 설계



38주 Holiday 아니면서 프로모션 없음 → 12

## 4-2 | 데모(지도학습-Regression)

## [시연] - 회귀방식 (Decision Tree) B



구분	대상
개발언어	Python
개발 플랫폼	Anaconda
개발환경	Jupyter notebook
파일명	4-0. DecisionTree

예측 모델 데모  
(지도학습)



## 4-2 | 데모(지도학습-Classification)

## [개념/시연] - 분류방식 (SVM)

데이터의 집단을 분류 시 최적의 분류 선을 생성 한 후  
미래 데이터에 대해서 기존에 학습한 내용을 바탕으로 예측

## 데이터 탐색

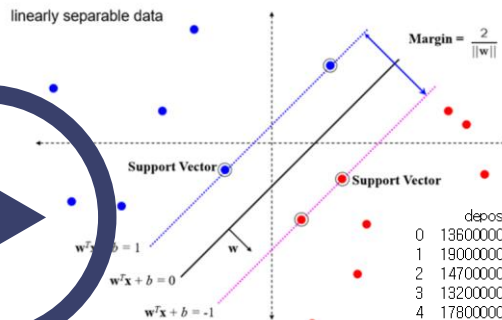
## 모델 설계

미래

과거

고객 ID	주식	부동산	대출	고객등급
C0001	1200	1500	0	
C0001	1000	1200	0	VIP
C0002	500	600	100	DIA
C0003	200	100	100	NORMAL
C0004	500	2000	0	VIP
C0005	100	50	50	NORMAL
C0006	100	100	30	NORMAL
C0007	700	600	0	DIA

linearly separable data



```

deposit  stock  label  predict
0  1360000000  35000000  normal  normal
1  1900000000  61000000   vip     vip
2  1470000000  36000000   normal  normal
3  1320000000  77000000   diamond diamond
4  1780000000  44000000   diamond diamond

```

정답률 = 0.9841666666666666

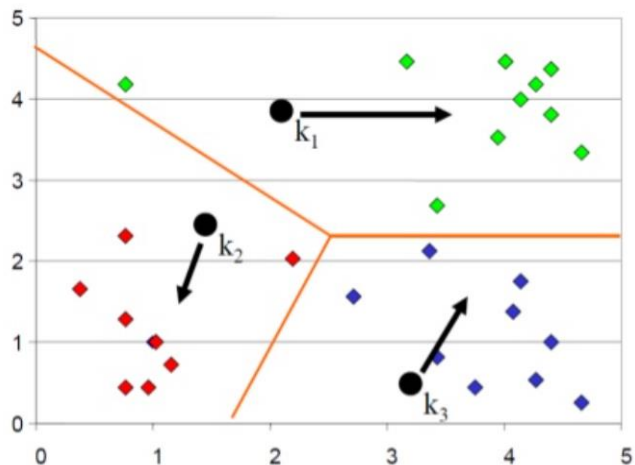
리포트 =

	precision	recall	f1-score	support
diamond	0.97	1.00	0.99	3492
normal	1.00	0.96	0.98	1758
vip	1.00	0.97	0.98	750
avg / total	0.98	0.98	0.98	6000

## 4-3 | 데모(비 지도 학습)

## [개념] - 비지도 학습 (K-Means Clustering)

데이터를 내부 집단간의 유사성이 높고 외부 집단과는 유사성이 낮도록 분류하는 방법



## STEP 1

- 클러스터의 개수를 정함
- K개의 초기 centroid(중심)를 random하게 선택

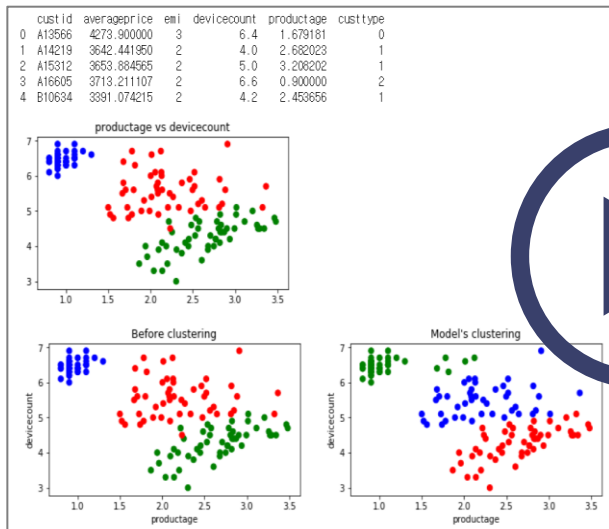
## STEP 2

- 각각의 객체를 자기자신에서 가장 가까운 centroid에 할당
- 같은 centroid에 할당된 객체의 평균을 구함

## STEP 3

- 각 centroid의 평균을 중심으로 다시 Step2를 반복 (클러스터의 멤버가 변하지 않은 경우 멈춘다.)

## 4-3 | 데모(비 지도 학습)

[시연]- 분류 (K-Means Clustering) D

**고객분류 데모**  
(비지도 학습)



**학생분류 데모**  
(비지도 학습)

구분	대상
개발언어	Python
개발환경	Jupyter notebook
시각화	Jupyter notebook
파일명	학생분류



# 정리하기



## 1 2023년도 미래선도 기술

- 초자동화 및 AI 중심 내용에 집중

## 2 빅데이터/인공지능 기술 및 직업군

- 빅데이터플랫폼엔지니어 / 데이터엔지니어 / 분석가 차이 이해

# 정리하기

## 3 빅데이터/인공지능 사례

- 플랫폼엔지니어 / 데이터엔지니어 / 분석가 차이 이해

순번	구분	활용	가치	비고
1	판매/마케팅	수요예측	선제적 시장대응	
2	금융	부정거래 탐지	원활한 금융시장 형성	
3	인사	투입인력 분배	원활한 인력운영 가능	
4	IT 인프라	장애분석 / 애방체계	안정적 인프라 운영	
5	자동화	브라우저 자동화기술	수동작업 자동화	

## 4 빅데이터/인공지능 데모

- 실 구현 가능한 모델에 대한 이해