



훈련과정 IT

(산대특) 빅데이터 분석 기반 **AI** 알고리즘 개발 과정 육성



안녕하세요



이 광 호 강사

leekh4232@gmail.com

<https://youtube.com/@hossam-codingclub>

<https://hossam.kr>

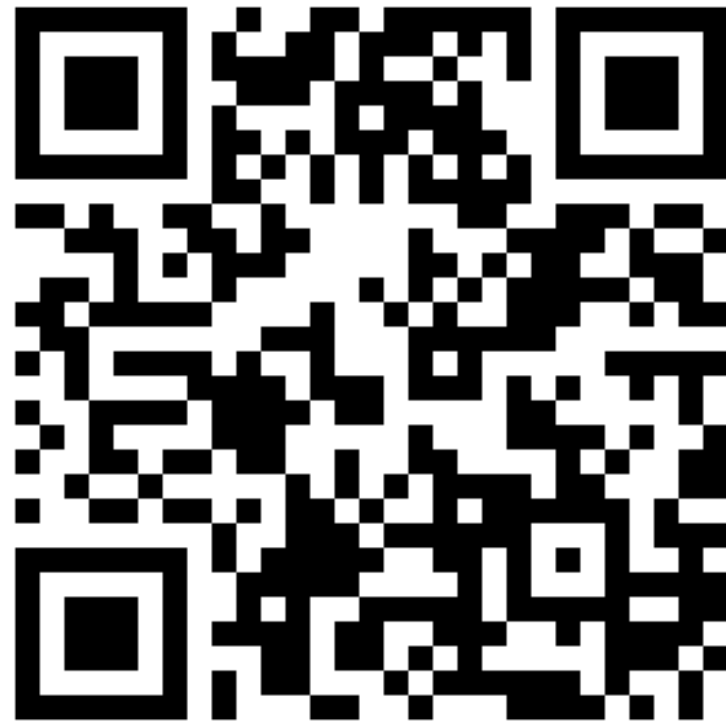


카카오톡 오픈채팅

채팅방 입장 후 대화명을 실명으로 설정해 주세요.



단톡방



1:1 채팅

Part 1

과정 개요





컴퓨터와 정보통신 기술의 발달로 매일 방대한 양의 데이터가 생산됨



머니투데이

로그인

≡ 뉴스 증권 정치 법률 유니콘팩토리 헬스·바이오 칼럼 연예 이슈 MT리포트 **속보** "아깝다" 류종일호, 日에 연장 역전패...

"데이터 폭발의 시대, 빅데이터는 선택이 아닌 필수"

"우리는 데이터 폭발의 시대를 살고 있다."

전 세계의 내로라하는 빅데이터 전문가들이 한 자리에 모였다. 이들은 24일 서울 여의도 콘래드호텔에서 개최된 '2014 키플랫폼'에 참석해 빅데이터 현황에 대해 소개했다. 빅데이터 시장을 바라보는 전문가들의 시각은 대동소이했다. 폭발하는 데이터와 함께 엄청난 기회가 열려 있다는 것이다. 하지만 이를 사업적으로 성공시키는 데는 '전략'이 필요하다는 지적이다. 그 전략이 '2014 키플랫폼'에서 공개됐다.

◇빅데이터, 유통·에너지 등 산업 속으로=알렉세이 아그레체브 리테일넥스트 설립자는 이날 '플러그인&토크'의 첫 번째 발표자로 나서 "리테일넥스트는 처음부터 기술회사가 아니라 유통회사가 돼야 한다고 생각했다"며 "빅데이터와 결합된 산업 속으로 들어가는 것이 성공의 핵심"이라고 말했다. 2007년 미국에서 설립된 리테일넥스트는 빅데이터를 활용해 유통매장의 각종 고객정보를 제공해준다.



예컨대 특정 성별의 고객이 유통매장에서 어떤 공간을 선호하는지 등을 분석해 유통매장 관리자에게 제공한다. 관련정보는 CCTV나 와이파이를 통해 수집한다. 정보수집과 분석 등 기술적인 능력이 무엇보다 중요할 수밖에 없다. 그럼에도 리테일넥스트는 창업초기 고용한 20명의 직원 중 절반을 유통 담당 직원으로 채웠다. 유통과 연계된 빅데이터 사업 과정에서 유통 전문가도 중요하다는 판단에서였다.

아그레체브 설립자는 "빅데이터 분야에서는 데이터를 분석하고 연구하는 데 상대적으로 오랜 시간이 걸리기 때문에 다른 곳에서 아이디어를 모방하기 쉽지 않다"며 "리테일넥스트가 유통 분야에 맞춘 빅데이터 모델을 만든 것처럼 독자적인 모델을 만든다면 승산이 있다"고 강조했다.



뉴욕타임즈가 하루에 실는 정보의 양은
17세기 영국의 평범한 한 사람이
평생 소비 하는 정보의 양과 비슷하다

(Wurman, S.A. (1987) "Information Anxiety"
New York: Doubleday , p.32)



페이스북에서는
하루에 4페타 바이트의 정보가 생성된

다(<https://kinsta.com/blog/facebook-statistics/>, Jan 3, 2021)

1페타바이트 = 1024테라바이트

Part 1, 훈련 과정 카테고리

 앞으로 함께 공부할 과목들 입니다.

001



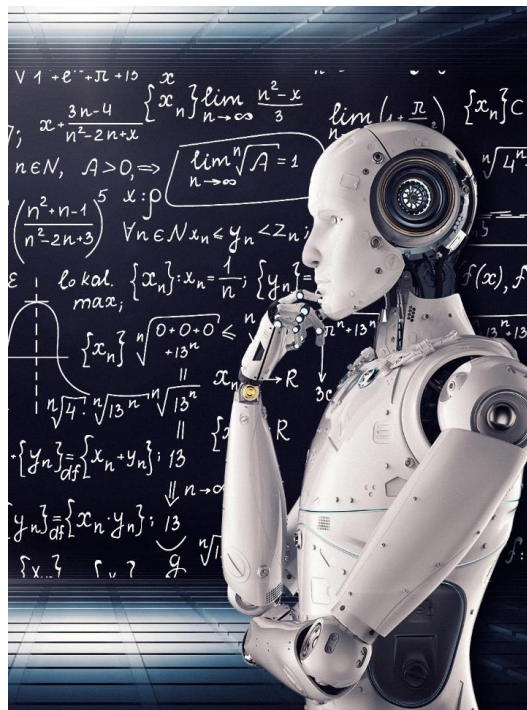
컴퓨터 사이언스 기초
파이썬 프로그래밍
데이터 수집

002



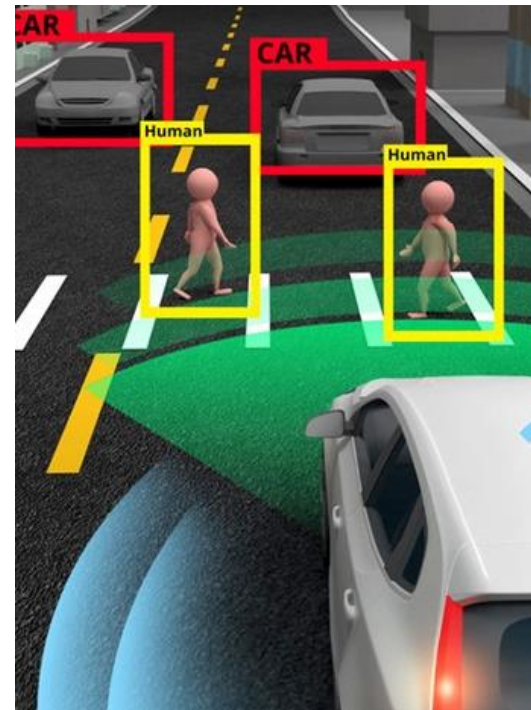
데이터 전처리
탐색적 데이터 분석
데이터 시각화

003



확증적 데이터 분석
머신러닝

004



딥러닝
텍스트 마이닝
영상 처리

STEP 1 - 통계학

전체 데이터 중에서
일부를 샘플로 추출
하여 이를 대상으로
규칙을 표현할 수 있
는 수학적식을 찾아냄
 $y = ax + b$

상관분석, 회귀분석
등

그냥 수학임

>>

STEP 2 - 빅데이터

수학적식을 내장한 프
로그래밍적 도구를
적용하여 **대량의 데
이터에 통계를 적용
함**

통계의 연장

>>

STEP 3 - 머신러닝

이미 도출되어 있는
통계학의 수학적식을
찾아내는 것이 아닌
x와 y의 수많은 쌍을
이용하여 수식을 도
출함

숫자뿐 아니라 문자,
**이미지, 영상 등에도
적용 가능**

>>

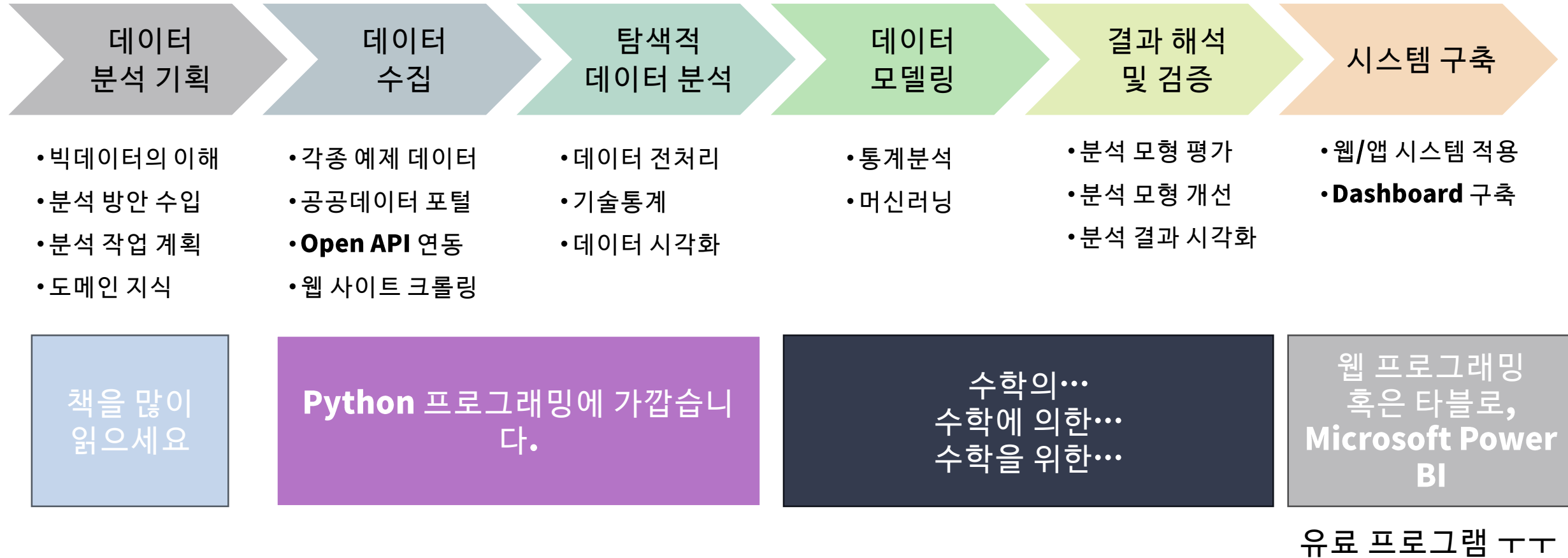
STEP 4 - 인공지능

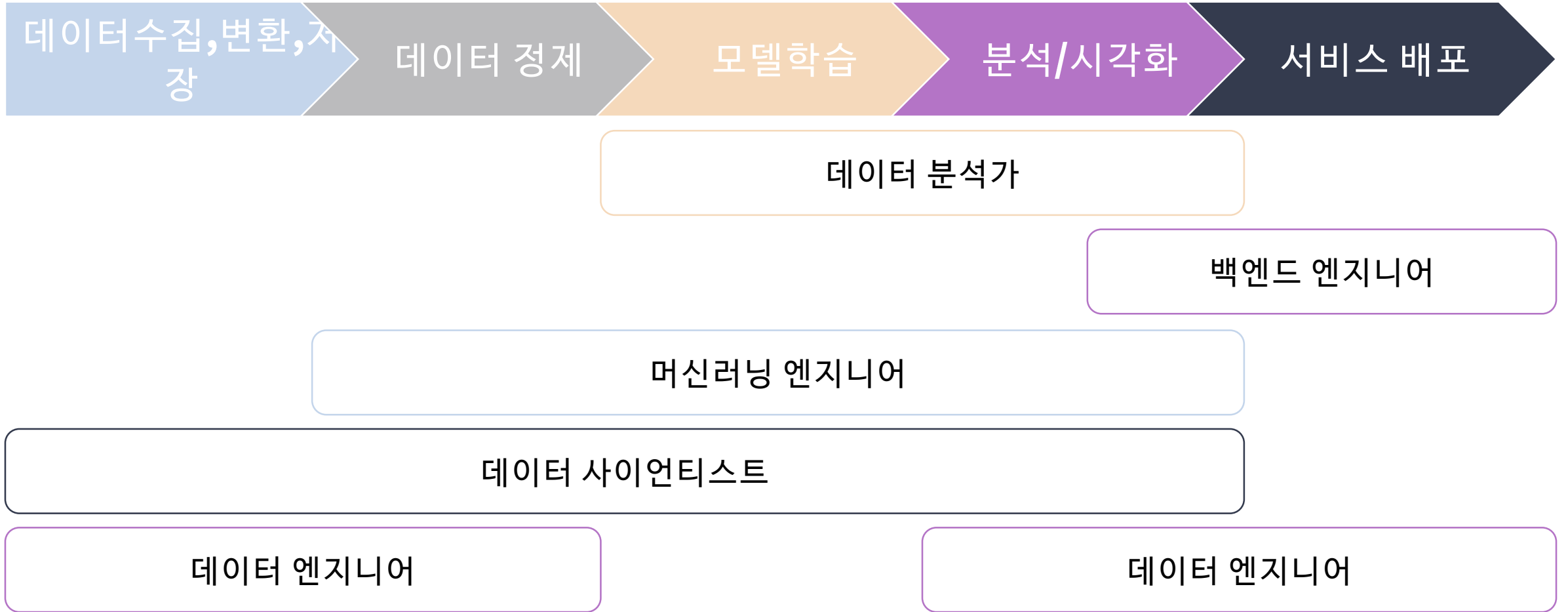
스스로 문제를 파악
하고 사고할 수 있는
소프트웨어.

사람과 비슷한 사고
를 하는 단계.

초심자도 도전해 볼 수 있음
하지만 어려움...

넘사벽





Part 2

훈련 내용 맛보기

Part 2, 데이터 분석 수업 예시

이런걸 합니다.

```
1 model = ols("dist ~ speed", data=df)
2 model.fit()
3 fit.summary()
```

OLS Regression Results

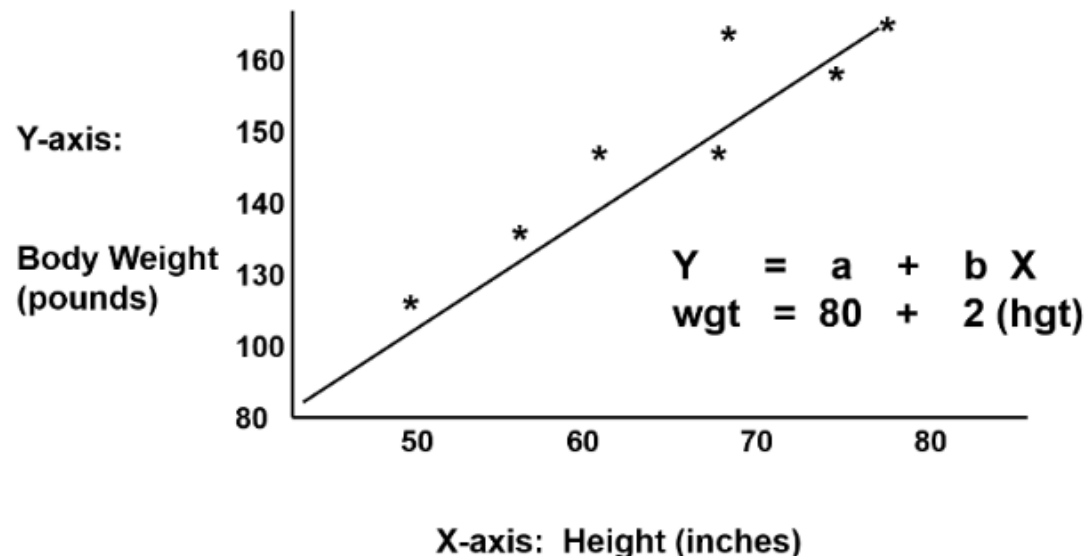
```
=====
Dep. Variable:      dist  1 R-squared:      0.651
Model:              OLS  2 Adj. R-squared:  0.644
Method:             Least Squares  3 F-statistic: 89.57
Date:               Mon, 24 Jul 2023  4 Prob (F-statistic): 1.49e-12
Time:               15:05:35  5 Log-Likelihood: -206.58
No. Observations:   50  6 AIC: 417.2
Df Residuals:       48  7 BIC: 421.0
Df Model:           1  8
Covariance Type:    nonrobust  9
=====
```

	coef	std err	t	P> t	[0.025	0.975]
Intercept	-17.5791	6.758	-2.601	0.012	-31.168	-3.990
speed	3.9324	0.416	9.464	0.000	3.097	4.768

```
=====
Omnibus:            8.975    Durbin-Watson:       1.676
Prob(Omnibus):      0.011    Jarque-Bera (JB):    8.189
Skew:               0.885    Prob(JB):            0.0167
Kurtosis:           3.893    Cond. No.            50.7
=====
```

종속변수	독립변수	비표준화 계수		표준화 계수	t	유의 확률	VIF
		B	표준오차	베타			
코로나19이후 매출액 차이	주말 매출액	0.429	0.031	1.48	13.637*	0.000	2.56
	청년층 매출액	0.4	0.032	1.33	12.638*	0.000	2.59
	생활인구 매출액	-0.045	0.021	-0.87	-1.256*	0.031	1.09

$R(613)$, $R^2(.611)$, $F(317.3)$, 유의확률(1.12e-203), Durbin-Watson(1.925)

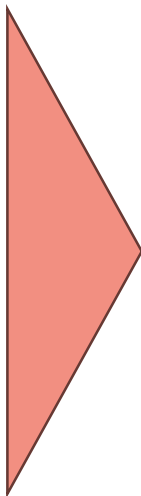


종류	모형	
단순회귀	$Y = \beta_0 + \beta_1 X + \varepsilon$	독립변수가 1개이며 종속변수와의 관계가 직선
다중회귀	$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$	독립변수가 k개이며 종속변수와의 관계가 선형 (1차 함수)
로지스틱 회귀	$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$	종속변수가 범주형(2진변수)인 경우에 적용되며, 단순 로지스틱 회귀 및 다중, 다항 로지스틱 회귀로 확장할 수 있음
다항회귀	K=2이고 2차 함수인 경우 $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{12} X_1 X_2 + \varepsilon$	독립변수와 종속변수와의 관계가 1차 함수 이상인 관계(단 k=1이면 2차 함수 이상)
곡선회귀	2차 곡선인 경우 $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$ 3차 곡선인 경우 $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \varepsilon$	독립변수가 1개이며 종속변수와의 관계가 곡선
비선형회귀	$Y = \alpha e^{-\beta X} + \varepsilon$	회귀식의 모양이 미지의 모수들의 선형관계로 이뤄져 있지 않은 모형



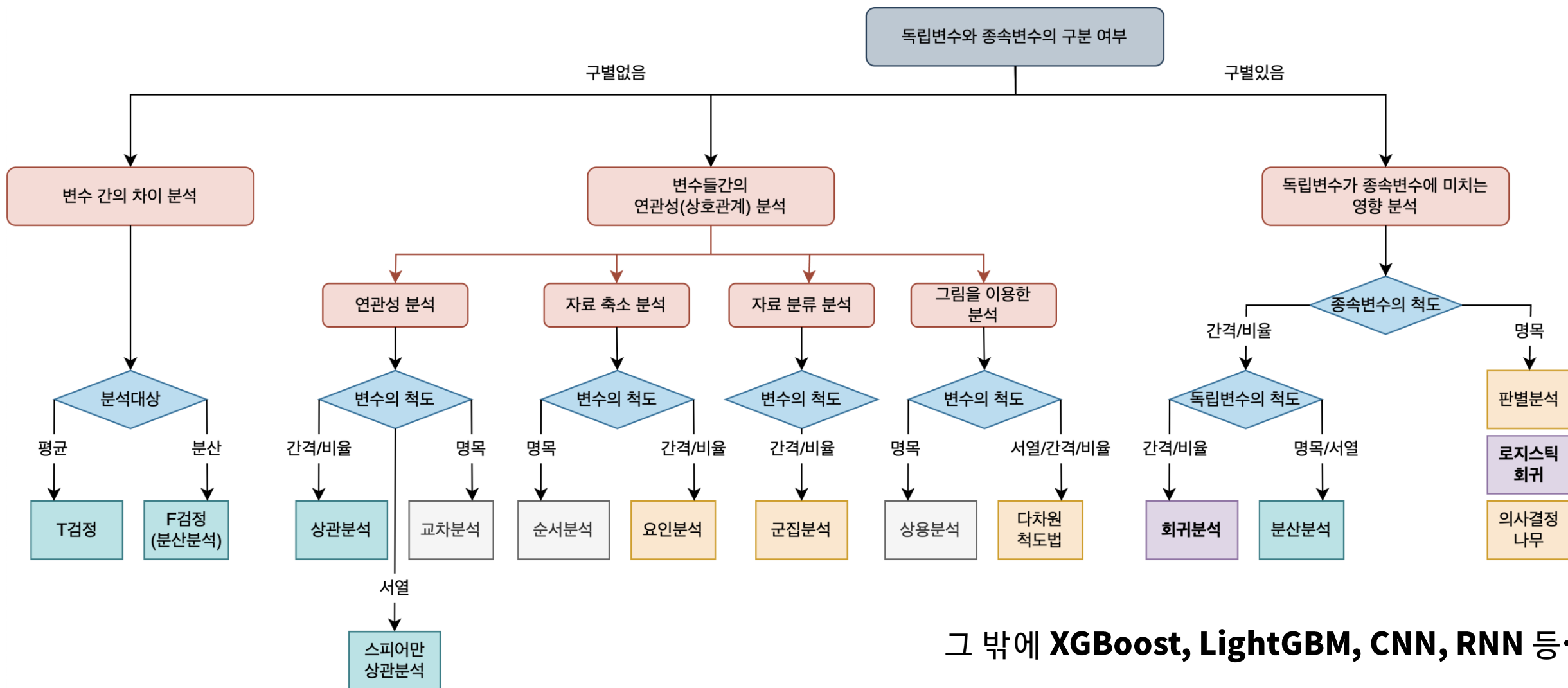


이런걸 합니다.



결국은 숫자의 집합으로 변환한 후
다시 통계학을 적용합니다.

[illegible]



Part 2

당부의 말



1

마음 단단히 먹으세요. 어렵습니다.

2

전공 수준의 이론(수학적 접근)을
다루지는 못합니다.

3

때때로 컴퓨터 성능 때문에
오랜 시간이 걸릴 수 있습니다.

4

지각, 결석, 조퇴로 인한
훈련 공백은 본인의 몫입니다.

5

훈련기간이 깁니다.
1주 이상 슬럼프가 올 수도 있어요.

6

총 **4**번의 과제(프로젝트)가
있습니다.

Part 3

훈련생 소개



Thank you

