

주식 증가 예측 프로젝트

팬데믹이 바뀌놓은 한국 주식 시장 분석

업종간 동향과 시사점을 중심으로

박○○ 배○○ 김윤명

이○○ 황○○

TEAM MEMBER



아이디어

코드 분석

EDA

전처리

모델링

PPT



팀장

박○○



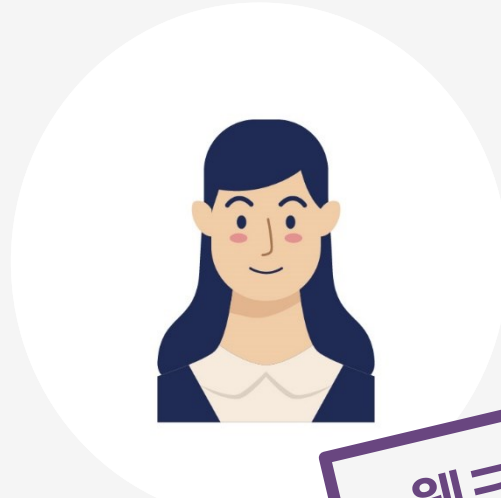
모델링

배○○



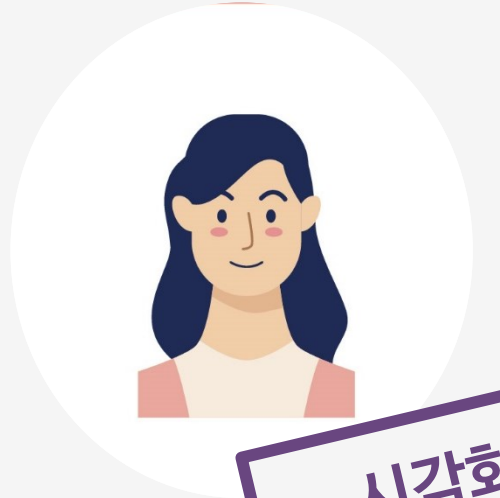
감성분석

김윤명



웹크롤링

이○○



시각화

황○○

CONTENTS INDEX

1_



주제 소개

프로젝트 분석 1

TIME-TABLE 2

진행 계획

진행 과정 3

활용 데이터 4

2_



결론

기대 효과 5

Q&A 6

3_



CHAPTER.1

주제 소개

프로젝트 분석

국내 주식 시장의 과도기 그리고 예측

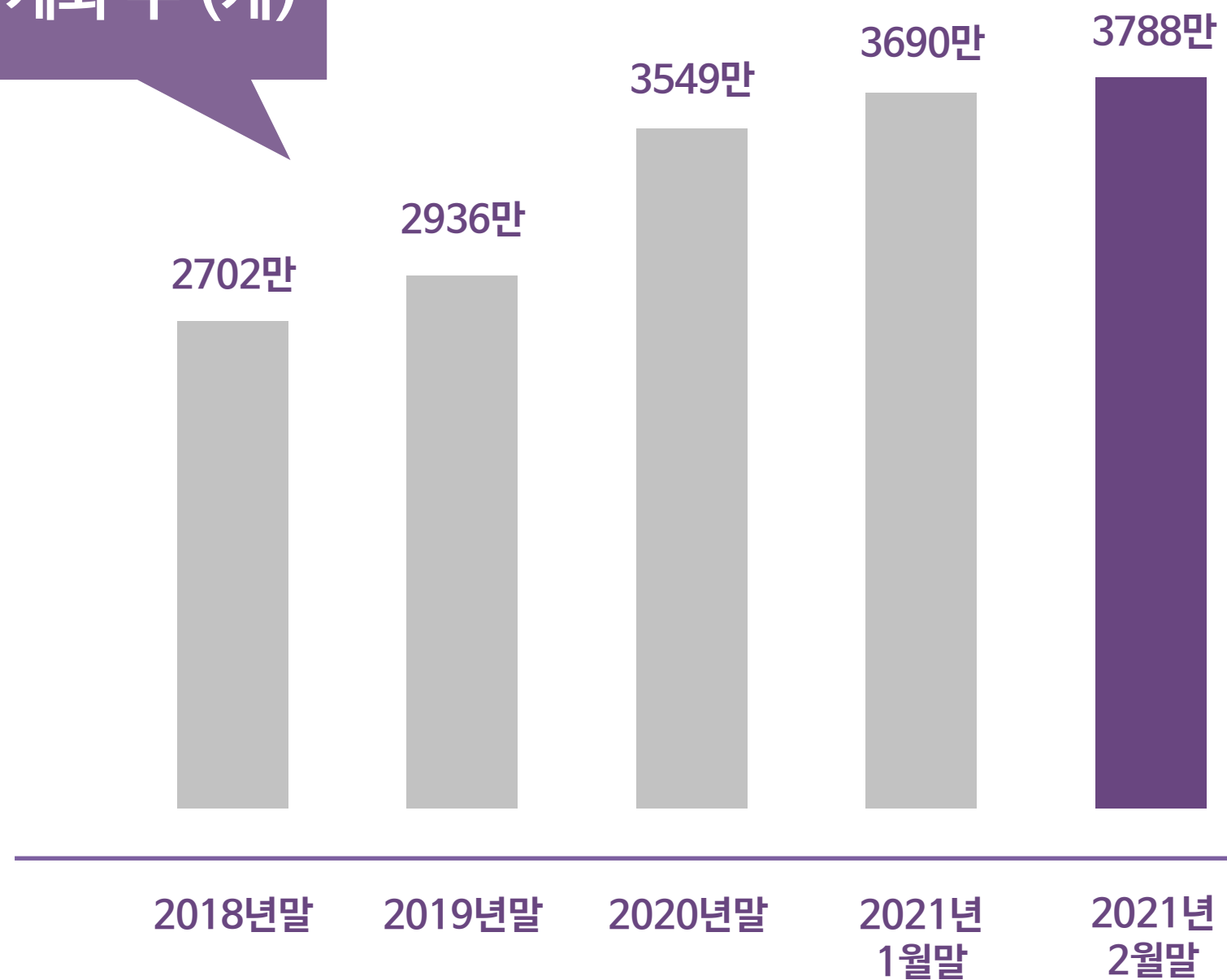
저금리 그리고 테크 시대에 MZ세대는 새로운 대안으로 주식에 관심을 갖기 시작한지 오래이며, 현재 국내 주식 투자 인구는 천 만명을 넘었을 것이란 관측이 나오고 있다. 최근 미국의 테이퍼링과 같은 현상으로 인해 주식 시장이 어떻게 변화할지 모르는 과도기에 주식 종가 예측은 유의미한 분석이 될 수 있을 것이다.



주식 계좌 수와 투자 인구

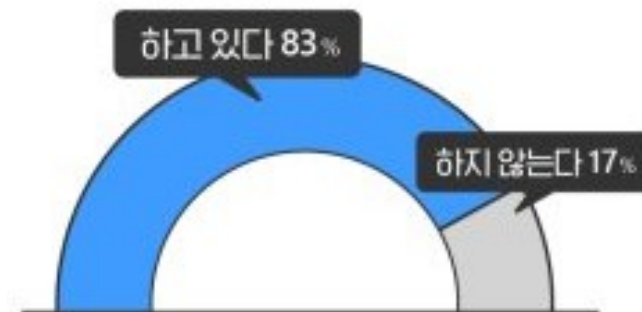
(단위 : 개) 자료: 금융투자협회, 예탁결제원

주식 거래
활동계좌 수 (개)



MZ세대 직장인 83%, “지금은 투자 중”

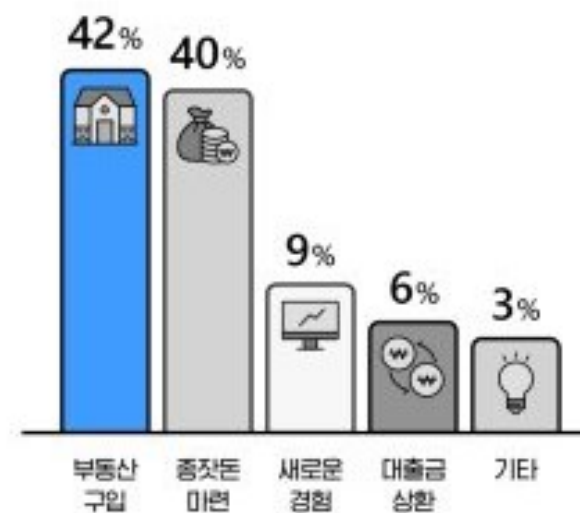
G 투자여부



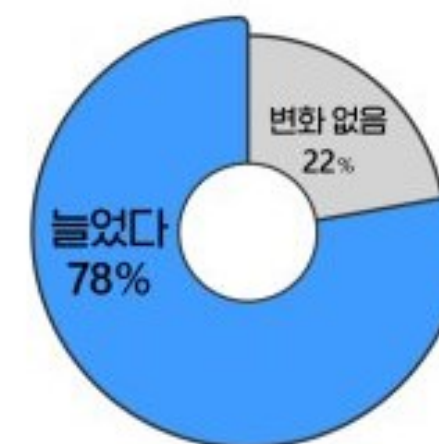
G 투자방법 (복수응답)



G 투자목적



G 보험에 대한 관심



MZ세대가 이처럼 직접 투자에 관심을 갖게 된 계기 중 가장 큰 특징은 바로 저금리 시대이기 때문이다. 저금리 기조가 이어지고 더 이상 근로소득으로는 계층 사다리 올라타기가 불가능하다는 풍조가 확산되면서 비교적 자산 모으기에 관심이 적었던 2030세대에서도 재테크가 필수로 자리를 잡고 있다.

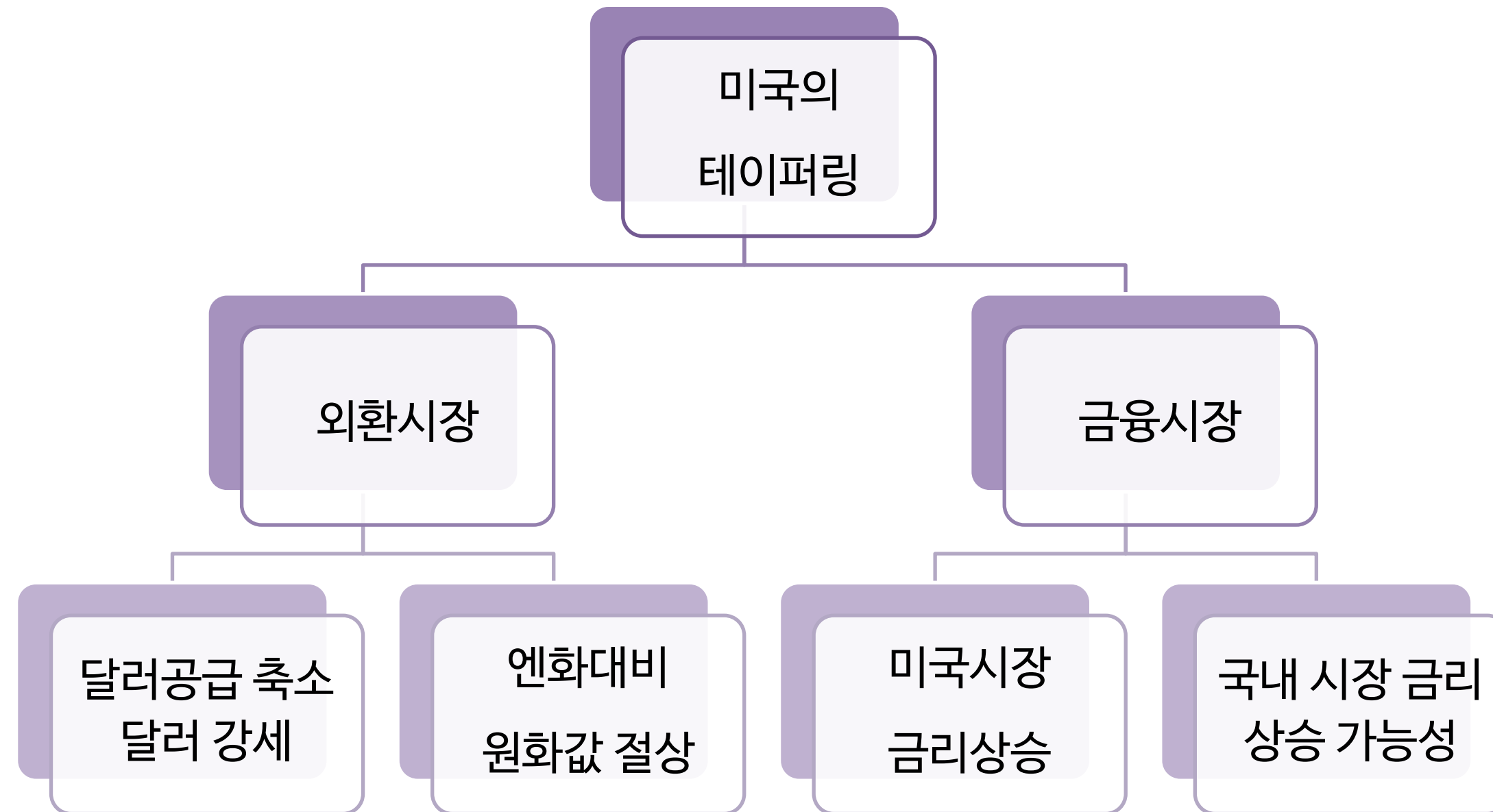
양적완화란?

중앙은행이 국채나 다른 금융상품을 사들임(매입)으로써 통화를 시중에 직접 공급해 경기를 부양하는 통화정책

테이퍼링이란?

정부가 경제 위기를 대처하기 위해 취했던 양적 완화의 규모를 점진적으로 축소해 나가는 것

INFO



테이퍼링이 주가에 미치는 영향?

경기가 불황인 상태에서 테이퍼링은 저금리로 인하여 발생하는 유동성의 증가와 돈의 가치가 하락하게 되고 주식의 가격이 급격히 올라가 주가 상승에 영향을 미치게 되며, 이로인해 경기가 개선되거나 회복단계에서는 테이퍼링은 금리 상승으로 인해 신흥국 투자 자금을 회수하게 되고 주가 하락까지 이어지는 영향을 받을 수도 있게 됩니다.

STANDARD

KOSPI 200

1~3년

산업별 분석

- 시장을 대표할 수 있는 대형주 위주의 주가지수인 KOSPI 200을 기준으로 분석
- 팬데믹 전후 1~3년을 기준으로, 예측 구간을 다양하게 설정
- 산업별로 sector를 나누어 유의미한 분석 결과를 도출

CHAPTER.1

주제 소개

TIME-TABLE

TIMETABLE

			1/6	1/7
			사전조사	
1/10	1/11	1/12	1/13	1/14
기획 발표 준비			데이터 수집	
1/17	1/18	1/19	1/20	1/21
코드분석 & EDA & 전처리				
1/24	1/25	1/26	1/27	1/28
모델링 & 전처리				
1/31	2/1	2/2	2/3	2/4
설 연휴				
2/7	2/8	2/9		
코드 수정 및 확인		발표 준비		

CHAPTER.2

진행 계획

진행 과정

PROJECT PROCESS



EDA

데이터의 의미 찾기

- 통계 및 상관도 분석
- 관련 도메인 이해
- 정제 대상 데이터 확인

전처리

시장 data & text data

- features 범위 선정
- 정규화 등 정제 작업
- 감성 분석 > 이진 분류

modeling

다양한 알고리즘 사용

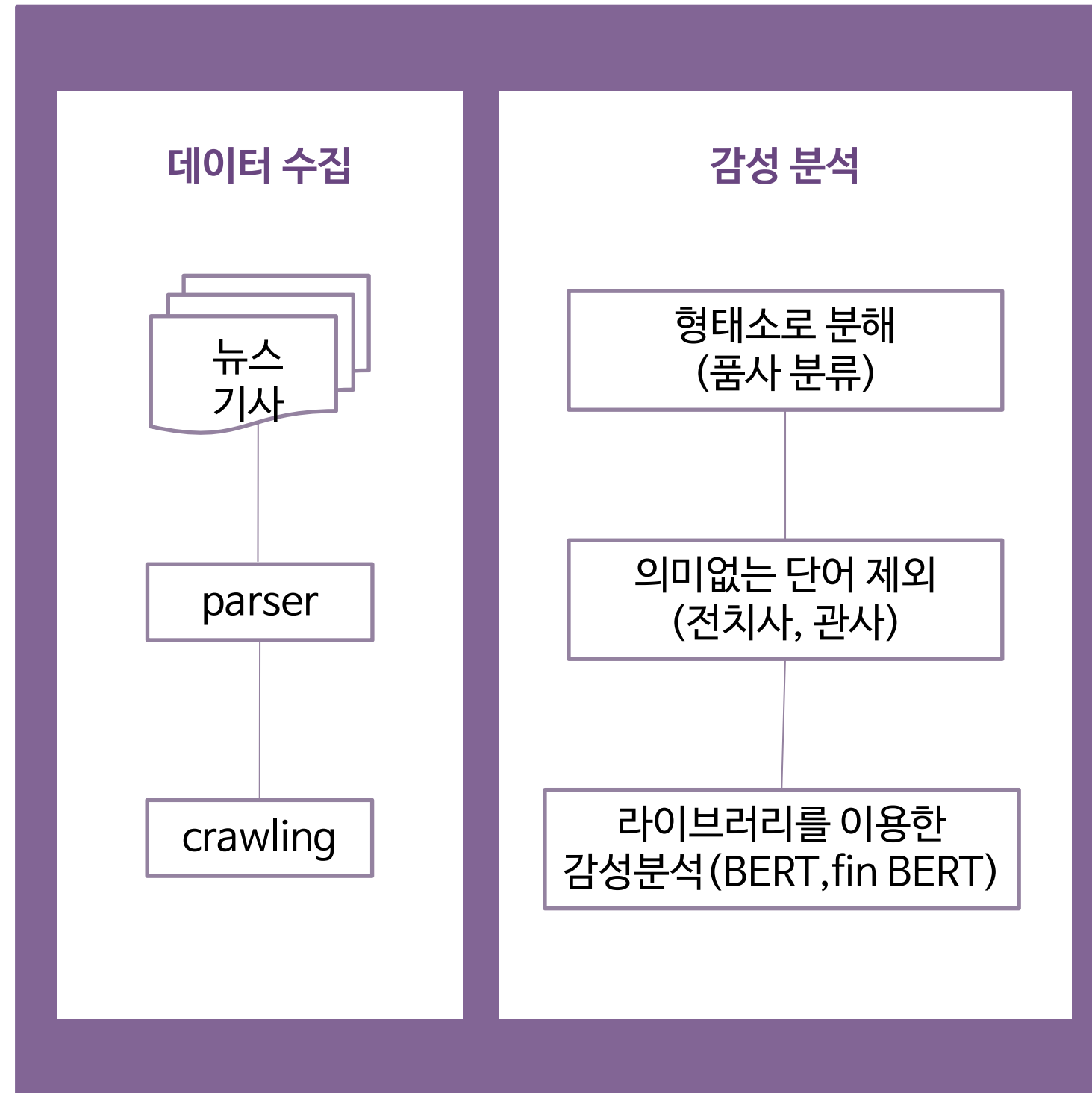
- ANN, SVM, NB
- GA, FA, DNN
- RA, HA 중 선별

evaluation

검증 및 평가

- Holdout
- K-fold CV
- LOOCV

감성 분석



감성분석 의미

텍스트에 들어있는 의견이나 감성, 평가, 태도 등의 주관적인 정보를 컴퓨터를 통해 분석하는 과정

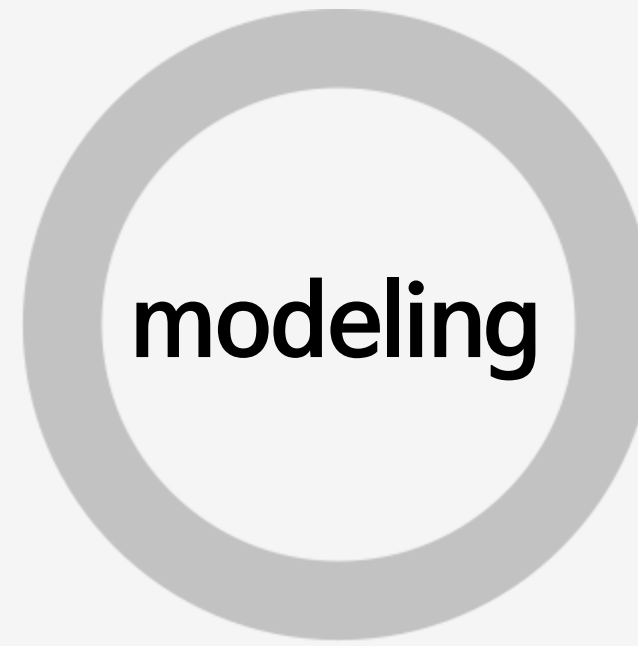
감성분석 장점 및 기대효과

뉴스 콘텐츠의 분석결과를 투자정보로 변환하여 투자의사결정을 지원함으로써 주가예측에 영향을 줄 것으로 기대함

BERT/fin BERT 특징점

- Pre-trained된 BERT 모델의 가중치들은 언어에 대한 많은 정보를 담고 있어 원하는 문제를 해결하기 위한 fine-tuning만 하면 되므로 훈련시간이 짧음
- 적은 데이터셋만 있어도 좋은 성능을 낼 수 있음

MODELING



Artificial Neural Network
(ANN)

Support Vector Machine
(SVM)

Naïve Bayes(NB)

Genetic Algorithms(GA)

Fuzzy Algorithms
(FA)

Deep Neural Network
(DNN)

Regression Algorithms
(RA)

Hybrid Approaches(HA)

MODELING

ANN: 인공신경망이라고 불리며 사람의 신경망 원리와 구조를 모방하여 만든 기계학습 알고리즘

DNN: ANN 기법의 변형으로 모델 내 은닉층을 많이 늘려서 학습의 결과를 향상시키는 방법

SVM: 기계 학습의 분야 중 하나로 패턴 인식, 자료 분석을 위한 지도 학습 모델(분류나 회귀분석 등에 이용)

NB(Naive Bayes): 특성들 사이의 독립을 가정하는 베이즈 정리를 적용한 확률 분류기(텍스트 분류 등에 이용)

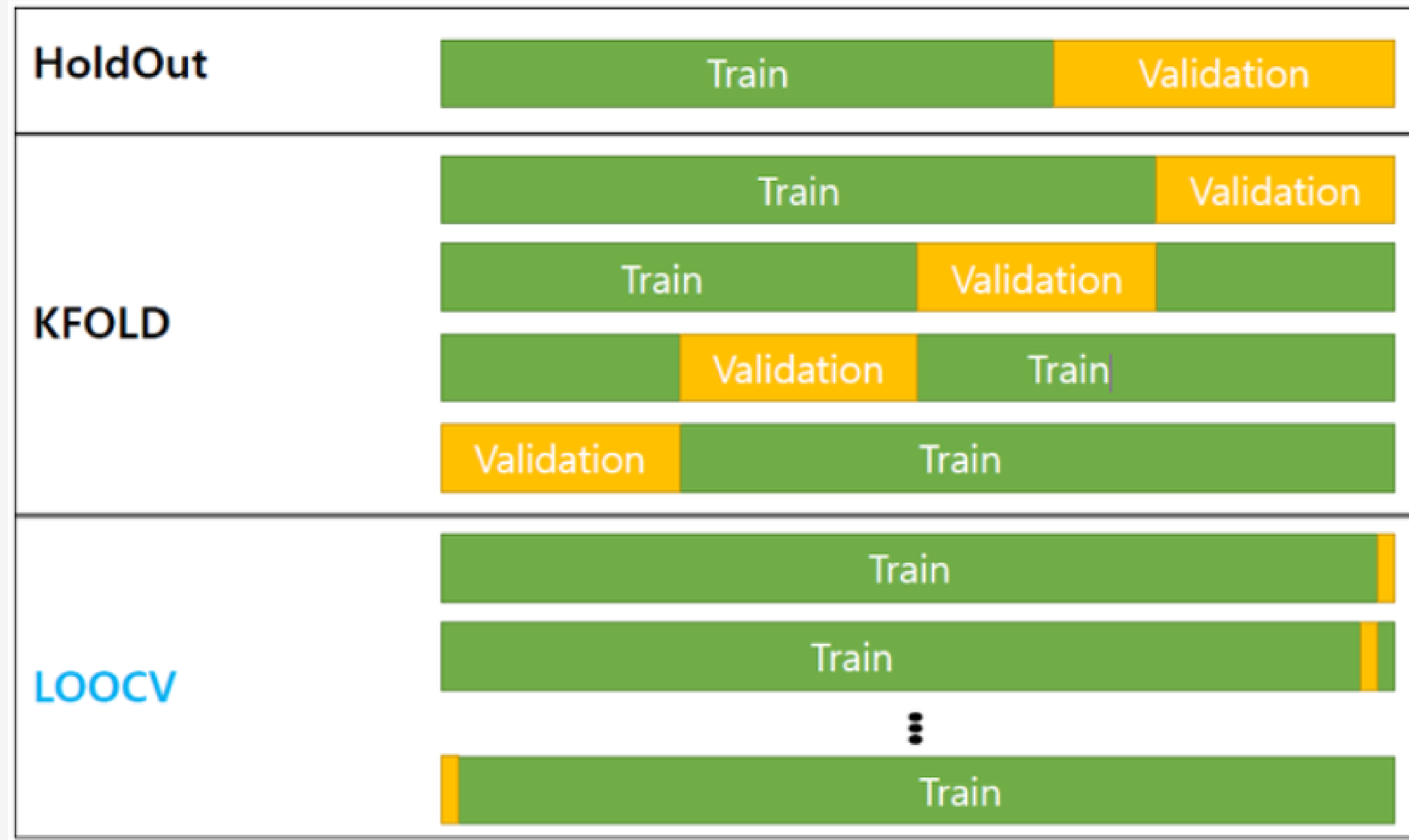
GA(Genetic Algorithms): 생물학적 진화를 모방한 자연 선택 과정을 기반으로 하여 제약 및 비제약 최적화 문제를 풀 수 있는 방법

FA(Fuzzy Algorithms): 모호한 표현을 명확하게 수치화하는 퍼지 이론을 기반으로 한 알고리즘 중 하나

RA(Regression Algorithms): 통계적 프로세스의 집합인 회귀분석 알고리즘

HA(Hybrid Approaches): 두 개의 서로 다른 머신러닝 기술들을 융합시키는 알고리즘

PROJECT PROCESS



Feature Selection ”

어떤 feature를
선택하는가에 따라
분석 결과의 정확도나
insight가 확연히 달라짐

Explanatory Power ”

Raw data를
얼마나 의미 있게
가공하는가에 따라
설명력의 크기가 좌우됨

Model Performance ”

고도의 분석 기술이 있더라도
데이터가 잘 정제되지 않으면
모델 성능을 원하는 만큼
향상시킬 수 없음

CHAPTER.2

진행 계획

활용 데이터



정형 데이터

* 경제주요지표 데이터 마이닝
: 인베스팅닷컴, FRED 등 참고



비정형 데이터

* 뉴스기사와 주식종목토론방 등에서
수집한 Text data를 감성 분석
: BERT, fin-BERT 이용하여 이진분류

FEATURES

환율

S&P 500

장-단기 금리차

CHAPTER.3

결론

기대 효과

주식시장에 대한 보다 높은 이해

”

- 주가의 시차값 뿐만 아니라 경제 전반의 지표를 변수로 하여 주가를 예측함으로써 주식시장에 영향을 주는 요인들을 파악 가능
- 체계적 위험과 비체계적 위험의 크기를 비교분석하여 주식시장 전체 및 개별 종목의 변동성과 위험도를 평가
- 궁극적으로 주식시장에 대한 심도 있는 이해가 가능

팬데믹 이후 주가에 영향을 미치는 핵심변수 확인

”

- 팬데믹과 같은 비일상적인 상황에서 주가 변동을 가져오는 유의미한 feature들을 파악 가능
- 팬데믹 이후 2020년 우리나라 증권시장은 가장 큰 변동성을 보였음
- 팬데믹이라는 특수 상황에서 주가 결정 요인은 일상적인 상황과는 다른 양상이 펼쳐질 가능성이 높음

정성분석을 통한 ‘기대심리’

”

- 개인투자자를 합리성 (rationality)에서 분리시키는 행태적 편의(behavioral bias)는 개인투자자의 직접투자 성과에 영향
- 개인투자자는 정보열 위에 있는데다 합리적 투자의사결정을 방해하는 여러 행태적 편의에 노출
- 행동적 관점에서 이러한 행태적 편의를 모델에 편입시켜 보다 정교한 예측이 가능할 것이라고 기대됨

투자 의사 결정에 도움 제공

”

- 개별 종목, 나아가 산업별 주가 예측 정보를 제공함으로써 개인의 투자의사결정에 도움을 제공
- 개별 투자자는 미래 주가의 변동 방향과 크기에 대한 정보를 얻게 되고 이를 바탕으로 주식 매매에 대한 결정의 근거로 삼을 수 있음

발표 내용에 관해
궁금한 점이 있으시다면
자유롭게 질문해주세요!

Q & A

Holdout 검증

- 가장 기본적인 모델 검증 방법으로, 전체 데이터셋에서 테스트 데이터를 분리하고 남은 학습 데이터의 일부를 검증 데이터셋으로 또 분리하는 방법
- 즉, 전체 데이터를 3개 (학습 데이터, 검증 데이터, 테스트 데이터)로 나누는 것
- 이 방식은 가장 간단하지만 학습 데이터에 손실이 있기 때문에 데이터가 적은 경우에 사용하기 힘들고, 검증을 한 번 밖에 진행할 수 없다는 단점 존재

K-fold CV

- 홀드아웃 방식은 학습 데이터에 손실이 있고 검증을 한 번 밖에 할 수 없다는 단점이 있음.
- 이를 개선하기 위한 교차검증(Cross validation, CV) 방식을 사용.
- 교차 검증은 최소 2번 이상의 검증을 진행하므로 각 검증 결과치의 평균을 모델의 검증 결과로 사용

LOOCV

- LOOCV(Leave-one-out cross validation)는 K-fold 교차 검증의 극단적인 형태로, 학습 데이터셋이 극도로 작을 때 사용할 수 있는 교차검증 방법

감사합니다!