

거시경제 지표와 주식시장과의 실증분석

Team 9 마이러버(MY ROBO)
['김태환', '김병준', '김태용', '이현철']

Hackathon Project

Team 9

<주제> '주식 시장 예측'

김태환: 유의미한 변수를 데이터로만 예측해보자.

김병준 : 텍스트 마이닝을 통한 감성분석,
개인투자자의 성향 파악해보자

김태용 : 볼린저 밴드 중심으로 20일 이동평
균선을 활용해보자.

이현철 : LSTM, FFT 활용해서 예측해보자.



for 수업내용 복습 ,
데이터 분석 및 활용을 최대화

'데이터 기반 예측' 채택

Hackathon Project

Team 9

세부 주제 선정 리스트	주제 토론
1. 배당주 높은 기업들 주가 특징 분석	ex. 배당하기 전 주가가 꾸준히 오르다가 배당금이 지급된 후엔 주가가 많이 떨어진다 패턴 분석 & 패턴 보고 주가예측 배당이 높을 수록 가치가 높나? 30프로는 무관하다 40프로는 맞다 30프로는 음의 관계다.
2. 외인의 매도세가 크면 주가하락, 매수세가 크면 주가상승?!?!	외국인의 수급 동향과 주식시장과의 관계 설명변수 : 순매수 순매도 종속변수 : 코스피지수 혹은 개별종목 상관분석 회귀분석 스캐터 외국인의 영향 수치화? HINT 어플 등
3. (경제적 변수) 거래량 , 환율(외인투자가 많음), 미국주가, 경기선행지수, 비트코인시세 등	(사회적 변수) 취업률, 출산율, 자살률, 미세먼지 지수, 기호식품(술, 담배 소비량), 코로나 확진자수 분석방법: 회귀분석(Linear Regression) 패널분석? 거시경제 지표와 주식시장과의 실증분석
4. 경기국면별 스타일투자 연구	저 PER 혹은 저 PBR 가치주 성장주 정의 수축기간엔 가치주? 성장주?

〈세부 주제 선정〉

'경제적 지표들을 활용하여 KOSPI 지수 예측하기'

1)
삼성전자 주가를 중심으로 개별 기업 주가 예측 시도.
But 특정 기업에만 적용되는 예측이라면
기업 고유의 변수들이 존재(실적, 액면분할 등),
또한 모델이 틀리면 무의미한 결과물이 될 리스크도 있음.

2)
한 기업만을 90% 확률로 예측하는 것보다
여러 기업을 70% 확률로 예측하는 방향으로
범용성을 가진 모델을 만들고 싶었음.

3)
결국 한국 주식시장의 흐름과 가격은
KOSPI로 파악 가능하기 때문에
KOSPI 지수 자체를 예측하기로 결정

〈해커톤 계획〉

1. 데이터 수집(04.23)
2. 데이터 전처리(04.23~04.25)
3. 데이터 확인, 분석방법 설정(04.25~04.26)
4. 분석(04.26~04.28)
5. 분석 결과 해석(04.28)

Hackathon Project

Team 9

현재진행 과정

1- 데이터 수집

-경제 지표 데이터들 수집하기
야후 파이낸스 , 공공데이터포털등등

feature 선택

삼성전자, S&P500,거래량, 환율, GDP, 경기선행지수,경상수지 등

2. 데이터 전처리 및 탐색

<데이터 전처리>

-결측치 처리와 경기선행지수와 주가 간의 기간 조정

-코스피가 공시된 날짜에 feature들 날짜 맞추기

Hackathon Project

Team 9

3. 모델링(상관분석과 회귀분석 사용)

Y = KOSPI 지수

X = 삼성전자, S&P500, 거래량, 환율, GDP, 경기선행지수, 경상수지 등

DataFrame 만들고 Linear regression 학습

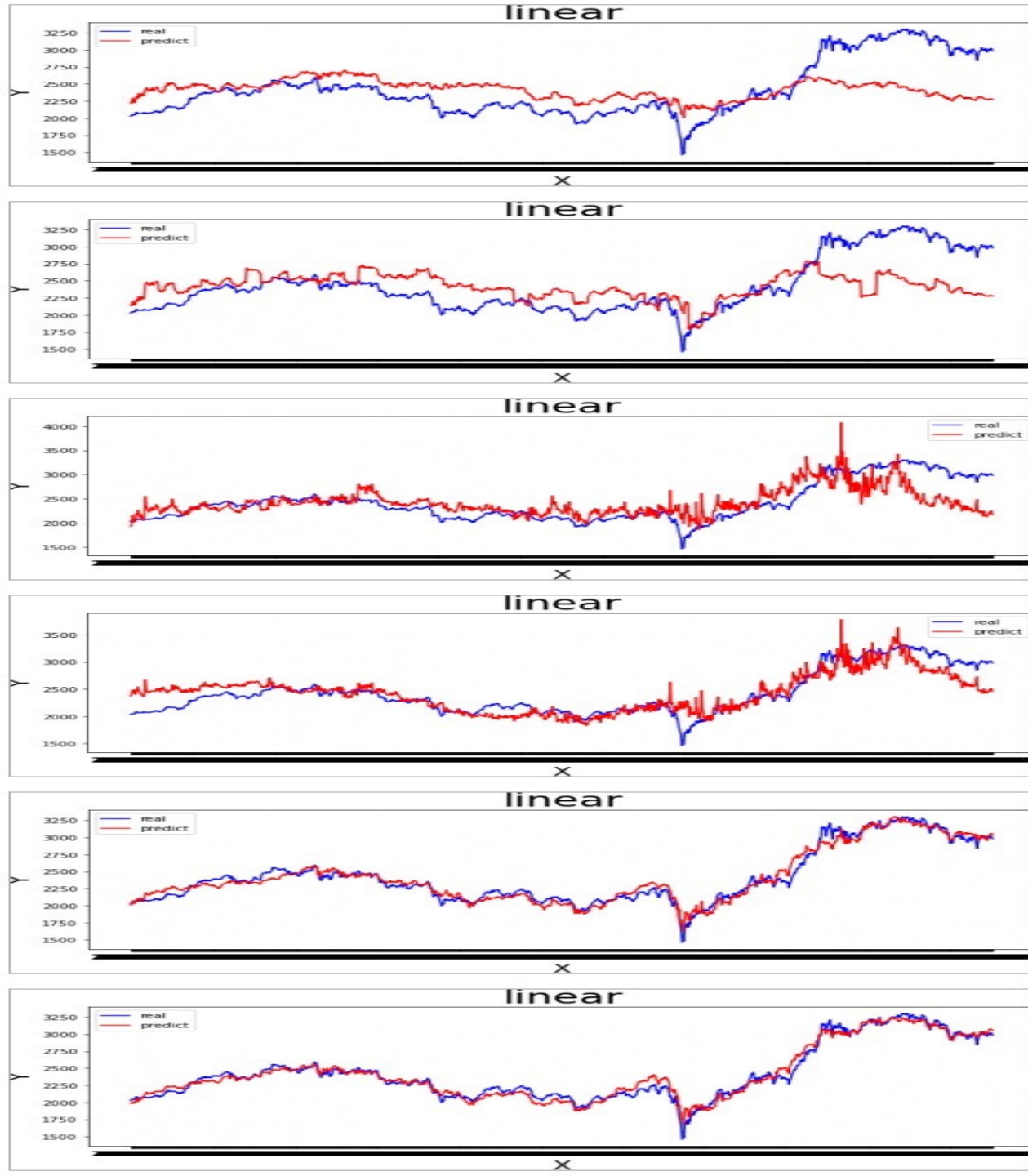
상관성 파악(heatmap, scatter matrix)

회귀 분석 (Ridge, Lasso 등)

모델 성능 높이기(feature 추가, scale, ARIMA 등)

4. 해석 및 시각화

(seaborn, matplotlib, pandas, Scikit-learn, numpy 등 활용)



X_features
[환율]

X_features
[환율, 경상수지]

X_features
[환율, 경상수지, 거래량]

X_features
[환율, 경상수지, 거래량,
경기선행지수]

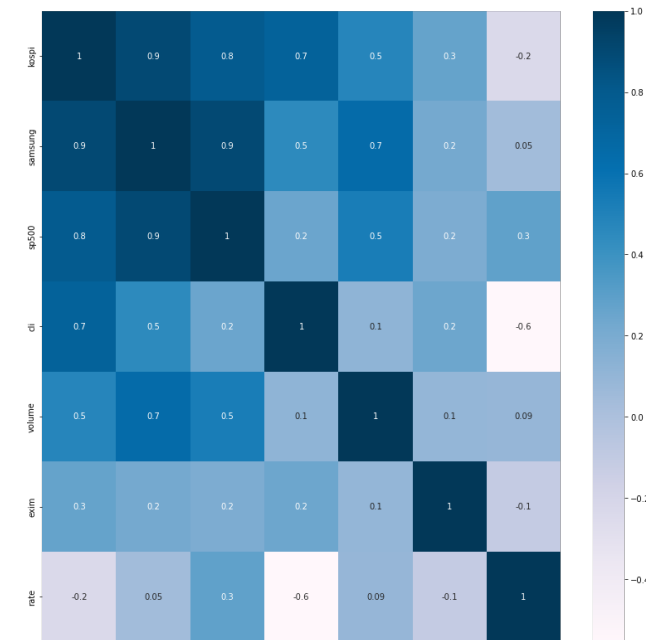
X_features
[환율, 경상수지, 거래량,
경기선행지수, s&p500]

X_features
[환율, 경상수지, 거래량,
경기선행지수, s&p500, 삼성전자]

Predict

Real(KOSPI)

**train_test_split 전
원본 csv
Linear Regression**



**KOSPI와
상관계수가 낮은 것부터
feature에 추가함**