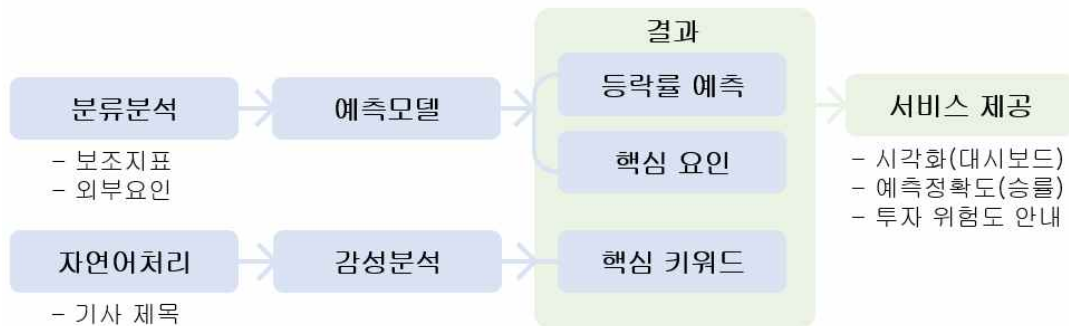


『제1회 KRX 금융 빅데이터 활용 아이디어 경진대회』 제안서

① 아이디어 명	삼성전자 주가 핵심 요인 분석	② 구분	개인 <input type="checkbox"/>	팀 <input checked="" type="checkbox"/>
③ 팀 명	한강 View	④ 팀 구분 (② '팀' 체크 시)	일반 <input checked="" type="checkbox"/>	법인 <input type="checkbox"/>
⑤ 법인 명 (④ '법인' 체크 시)	.	⑥ 법인등록번호 (④ '법인' 체크 시)	.	.

아이디어 개요

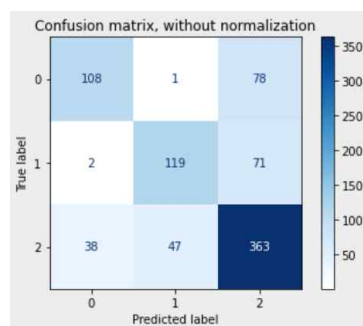


본 아이디어는 자연어처리(NLP)와 분류분석을 활용한 머신러닝 앙상블모델을 통해 여러 가지 변수의 데이터로 주가의 등락률을 학습하고 학습에 영향을 준 핵심 요인들을 산출해 사용자에게 제공하는 것입니다.

분석 방법과 머신러닝 모델은 다음과 같습니다.

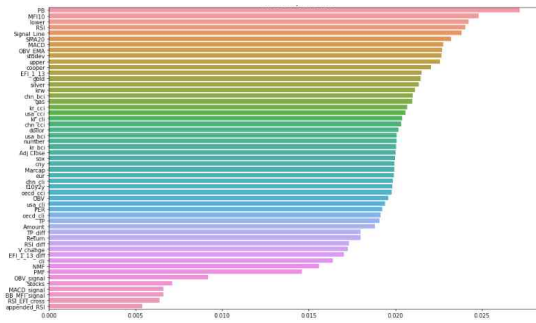
- 분류분석
 - 보조지표: 주가의 변동성과 추세의 방향을 동시에 고려해 분석에 사용합니다.
 - 외부요인: 해당 종목의 주가에 영향을 미치는 외적 요인을 고려해 분석에 사용합니다.
- 예측모델
 - 주가 등락률의 5일 뒤의 지수이동평균(EMA)을 통해 주가의 변동추세를 기초 통계량을 기준으로 3가지(상승, 하락, 보합) 경우로 분류합니다.
 - 주가에 영향을 미치는 핵심 요인을 특성 중요도 분석을 통해 파악합니다.

[등락률 예측 - 혼동행렬]



정확도: 0.71

[특성 중요도 추출]



핵심 요인 Top3

- 구리(Copper)
- 금(Gold)
- 은(Silver)

• 자연어처리 및 감성분석

- 삼성전자와 연관된 기사의 제목 데이터를 일별로 전처리합니다.
- 데이터와 수익률로 얻은 감성점수(변수)를 모델에 적용합니다.
- Wordcloud를 활용해 주가에 영향을 미치는 주요 키워드를 시각화하여 제공합니다.

[주요 키워드 시각화]



2017-02-17
이벤트 수익률 0.0248755096953351

날짜 및 등락률이 함께 보여집니다.

사용자는 예측모델을 통해 주가에 영향을 미치는 주요 변수들을 쉽게 파악할 수 있으며 모델이 제시한 변수들에 집중하여 특정 주식시장의 흐름을 빠르게 파악할 수 있습니다.

주가 등락을 예측모델을 위해 우선적으로 삼성전자의 주가 데이터를 사용했고, 해당 모델을 발전시켜 주요 산업 및 ETF 등 다양한 주식의 흐름을 파악하는 서비스에도 적용할 수 있습니다.

아이디어 제안 배경 및 목적

코로나-19 확산으로 전 세계 주식시장이 크게 하락한 뒤 급반등한 직후부터 ‘동학 개미 운동’으로 불리는 개인들의 주식투자 열풍이 불었습니다. 그러나 거세지는 열기와는 반대로 투자지식과 정보를 활용해 투자하는 경우는 드물어지고 있습니다. 실제로 한국거래소 리포트에 따르면, 전체 주식투자자의 48%는 자신의 투자지식 수준이 낮다고 평가했고, 전문기관 리포트 또는 직접 찾은 투자 정보를 수집, 분석하여 스스로 결정하는 경우는 7%로 매우 낮은 수치를 보입니다.

저희는 객관적인 데이터를 활용하는 투자자들의 비율이 낮은 이유로 제공되는 투자 분석 자료와 실제 투자 사이에 괴리가 있다는 점을 꼽았습니다. 주가에 영향을 미치는 요인들이 매우 많고 관련된 자료를 모두 투자에 활용하기엔 어려움이 있기 때문입니다. 이에 저희는 ①다양한 종류의 지표 및 데

이터를 활용하여, ②실제 투자에 쉽게 적용할 수 있고 도움이 되는 정보를 제공하는 서비스를 위한 예측모델을 만들었습니다.

활용 데이터

- 기간(공통): 2000.01.01.~2022.06.30

사이트	활용 데이터
야후파이낸스	삼성전자 일별 주가
인베스팅닷컴	원자재(금, 구리, 은, 천연가스, WTI유 등)
인베스팅닷컴	환율(원달러, 유로달러, 달러위안 등)
인베스팅닷컴	미국달러지수
빅카인즈	삼성전자 기사 크롤링 데이터
KRX	휘발유, 경유 일별 주가
Fred	미국장단기금리차_10년2년

상품·서비스·콘텐츠 세부내용

- 분류 모델을 사용한 주가 등락을 예측
- 연관 기사의 제목을 활용이 가능한 데이터로 변환 후(자연어처리) 주가 분석에 사용

기존 상품·서비스·콘텐츠와의 차별성

- 본 서비스는 특정 주가에 영향을 미치는 주요 변수를 제공합니다. 투자자는 제공되는 변수 정보를 활용해 주식의 흐름을 파악하고 더 확실한 투자 전략을 세울 수 있습니다.
- 등락을 분석에 자연어처리를 활용하여 주가의 변동성과 연관이 있는 이슈 및 키워드를 확인할 수 있습니다.

시장성 및 사업화 가능성

다음과 같은 정보를 사용자에게 제공하여 사업화를 검토해 볼 필요가 있습니다.

- 산출된 핵심 지표 및 키워드의 시각화(대시보드 활용 등)
- 알고리즘을 활용한 주가 상승/하락 예측정확도(승률)
 - 5일, 30일 뒤 상승/하락할 확률 등
- 예측정확도에 따른 고위험, 저위험 등의 정보
 - 사용자가 투자 위험시기를 판단하여 수익률에 도움이 될 수 있도록 함

기타(아이디어 강점 등 자유 기재)

- 분류 모델을 이용하여 보조지표 및 외부요인에 해당하는 데이터를 최대한 많이 수집해 분석했습니다.
- 기존에 적용하기 어려웠던 사회적 영향을 비정형 데이터를 활용해 차트에 반영했습니다. 본 모델에 사용한 기사 제목 데이터에서 나아가 주식 토론방 글, 기사 본문 등의 데이터를 추가로 사용할 수 있습니다.