혁신성장 청년인재 집중양성 사업

AI를 활용한 금융 리포트 생성

2020년 11월 6일

인공지능 자연어처리(NLP) 기반 기업 데이터 분석 과정

AI 애널리스트
임호태, 황지상, 김인용, 강범석, 김도현

목 차

1. 프로젝트 개요 1

1.1 프로젝트 기획 배경 및 목표 1

1.2 구성원 및 역할 1

1.3 프로젝트 추진 일정 2

2. 프로젝트 현황 3

2.1 시장 분석 3

2.2 경쟁 제품 장단점 분석 5

2.3 차별화 핵심 전략 기술 6

3. 프로젝트 개발 결과 7

3.1 데이터 수집 7

3.2 분석 진행 8

3.3 분석 결과 9

3.4 활용 방안 10

4. 기대 효과 12

4.1 향후 개선 사항 12

4.2 기대 효과 12

5. 개발 후기 13

6. 참고 자료 15

1. 프로젝트 개요

1.1 프로젝트 기획 배경 및 목표

본 프로젝트는 '인공지능 자연어 처리 기반 기업 데이터 분석 과정'에서 학습한 자연어 처리를 활용해 네이버 주식 관련 기사를 수집하고 이를 전처리했다. 또한, 통계 모델인 토픽 LDA 와 딥러닝 모델인 BERT 를 활용해 향후 주가를 예측하는 금융 리포트 형태의 서비스를 제공하는 것을 목표로 진행했다. 해당 서비스를 통해 리서치 소외 지역이라 할 수 있는 약 1,500 여개의 중소기업을 발굴함으로써 건전한 투자 생태계를 만들고자 한다.

1.2 구성원 및 역할

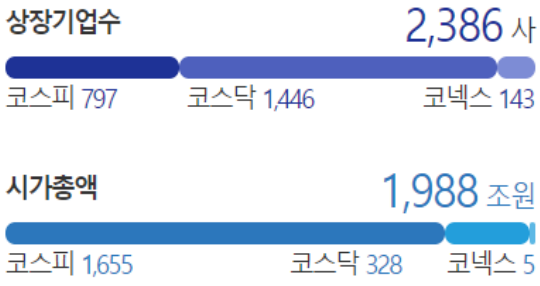
이름	전공	역할	구현 부분
임호태	경제학과, 컴퓨터공학과	팀장	프로젝트 관리, 프로젝트 자동화 구현, 웹 서비스 기획
황지상	산업경영공학과	팀원	프로젝트 기획, BERT, GPT2 학습 모델 구축
김인용	브라질학과, 통계학과	팀원	데이터 수집 및 구축, 토픽모델링 최적화
강범석	첨단로봇제어과	팀원	토픽모델링 알고리즘 구축, 데이터 구축
김도현	산업공학과	팀원	데이터 수집 및 시장분석, 알고리즘 리서치 및 구현

1.3 프로젝트 추진 일정

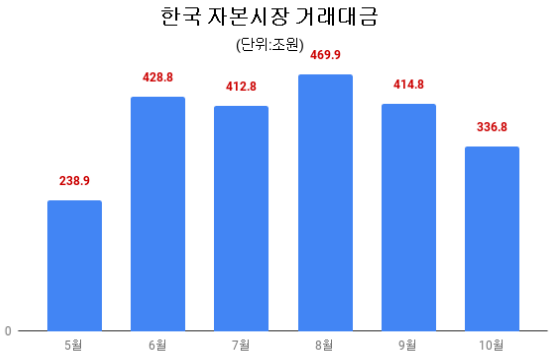
구분	기간	활동	비고
사전 기획	9/17(목) ~ 9/29(목)	프로젝트 기획	
	~10/4(일)	1. 프로젝트 주제 선정 및 분석 방향 설정 2. 수집 데이터 확정	
PJT 수행 및 완료	~10/12(월)	데이터 수집 및 구축 - 뉴스 - 주가	
	~10/30(금)	1. 토픽 모델링 참고 논문 구현 2. 자연어처리를 위한 모델링 3. 데이터 구축	
	~11/6(금)	1. 자연어처리를 위한 모델링 2. 시계열 예측 모델링 3. 웹 서비스 구현	
	~11/10(화)	1. 모델 및 성능 보완 2. 웹 서비스 개선 3. 발표 준비	
	11/11(수)	최종 발표	*최우수팀 선발*

2. 프로젝트 현황

2.1 시장분석



<그림 1 – 상장기업 수, 시가총액>



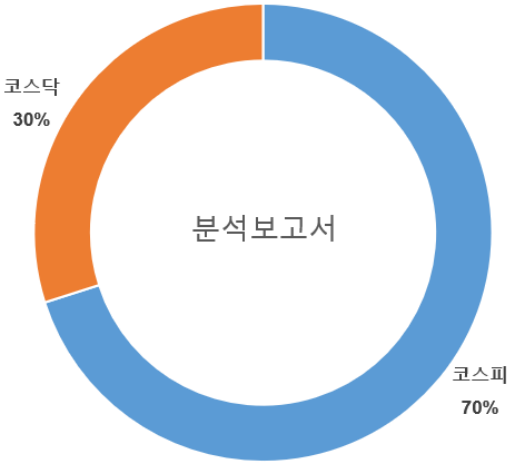
<그림 2 – 한국 자본시장 월별 거래대금 추이>

(2020.11 기준 krx 제공)

현재 국내에는 약 2,000 개가 넘는 기업이 상장되어 있으며, 월 약 400 조원의 자본이 코스피, 코스닥, 코넥스 3 종류의 시장에서 거래되고 있다. 자본시장의 거래 주체는 자신에게 더 나은 이익을 제공할 기업을 찾고 독자적인 판단으로 투자를 진행한다. 이 과정에서 투자자는 수익을 올리기 위해 기업을 분석하고 다양한 투자 인사이트를 얻기 위한 리서치를 진행한다.

최근 여러 증권사에서 AI 를 활용한 다양한 금융 서비스를 런칭하면서 개인화, 비대면, 자동화에 초점을 맞추고 있다. AI 로보어드바이저 서비스가 개인화와 자동화를 대표하는 서비스로 자리잡았다. AI 로보어드바이저를 통해 고객 개개인의 투자성향을 고려해 종목을 추천하고 자동으로 주식을 거래하는 서비스를 제공한다. IT 기업 '코스콤'에 따르면 올해 3 월 기준 로보어드바이저 가입자 수는 약 18 만 6000 명으로 지난해 말 대비 40% 증가했다. 가입 금액도 1 조 1300 억으로 전기 대비 21% 증가했다. 이렇게 금융사들은 AI 를 활용해 다양한 시도를 하고 있으며 고객들의 수요 역시 지속적으로 증가하고 있다.

리서치 분야도 예외는 아니다. 최근 한국투자증권은 'AIR(AI Research)' 서비스를 통해 국내 최초로 인공지능 기술을 적용한 리서치 서비스를 오픈하였다. 한국투자증권 리서치센터의 애널리스트들이 분석한 약 10 만건 이상의 뉴스 데이터를 기반으로 AI 가 문장과 맥락을 이해하고 분석할 수 있도록 학습이 이루어졌다.

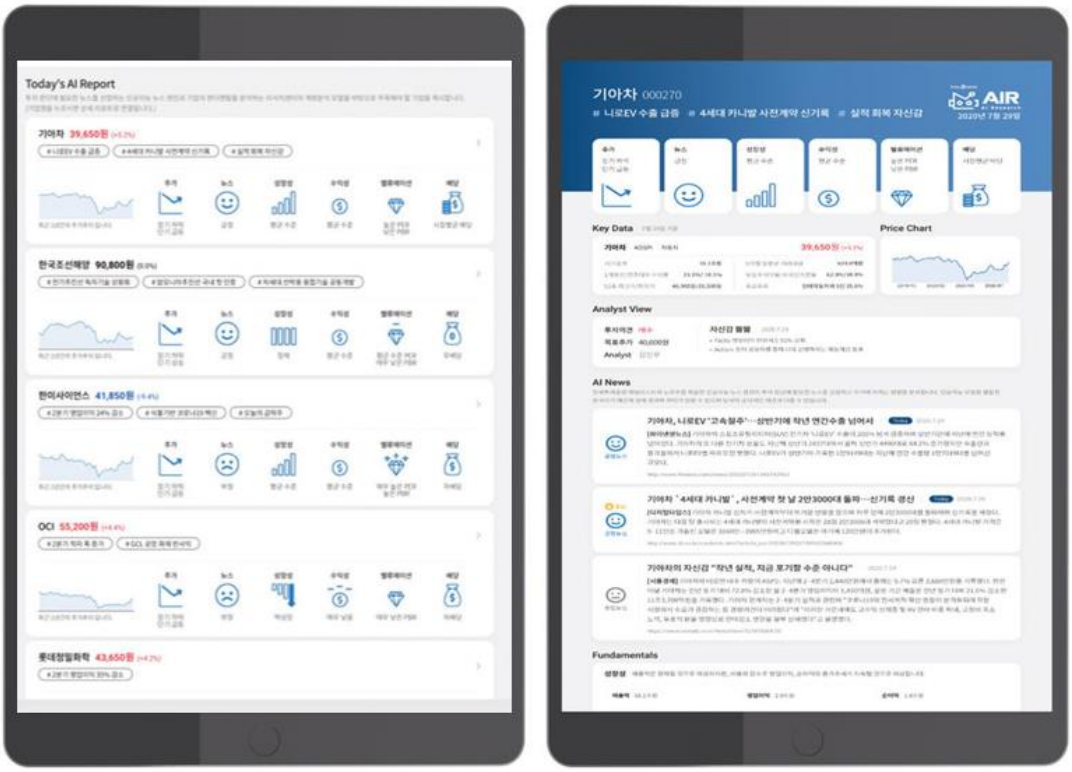


<그림 3 – 네이버 10 월 종목분석보고서 비율>

기존의 금융 리서치의 경우, 대기업을 위주로 진행되었다. 대부분의 증권사는 인력 부족 등을 이유로 대형주 분석에만 집중하는 모습을 보였다. 이로 인해 중소형주는 대형주 대비 상대적으로 리서치에서 소외되는 양상을 보였다. 그림 3 을 보면 전체 2000 여개가 넘는 상장사 가운데 애널리스트 보고서가 나오는 코스닥 종목은 약 30% 불과하다는 결과를 확인할 수 있다. 이를 통해 금융 리서치가 편향적으로 발간되었음을 알 수 있다.

한국투자증권의 AIR 는 이러한 문제를 해결하고 잠재력 있는 중소형주를 분석해 건전한 투자 생태계를 만들어 가기 위한 일환으로 개발되었음을 알 수 있다. 실제로 지난 7 월, AIR 서비스 출시 이후 3 개월간 619 개의 종목에 대해 1052 개의 리포트가 발간되었다. 이 가운데 중소형주는 360 종목으로 한국투자증권을 제외한 모든 증권사가 분석한 343 개의 중소형주보다 많은 분석을 제공하였다. 최근에는 'AIR US'를 런칭해 국내뿐 아니라 글로벌 서비스를 제공하기 위해 노력하고 있다. 이를 통해 스몰캡 종목분석은 AI 에게 맡기고 기존의 애널리스트는 테마 리포트나 거시 분석 등 롱 페이퍼에 집중하여 보고서의 퀄리티 향상에 기여할 것이라고 밝혔다.

2.2 기존 서비스의 장단점 분석

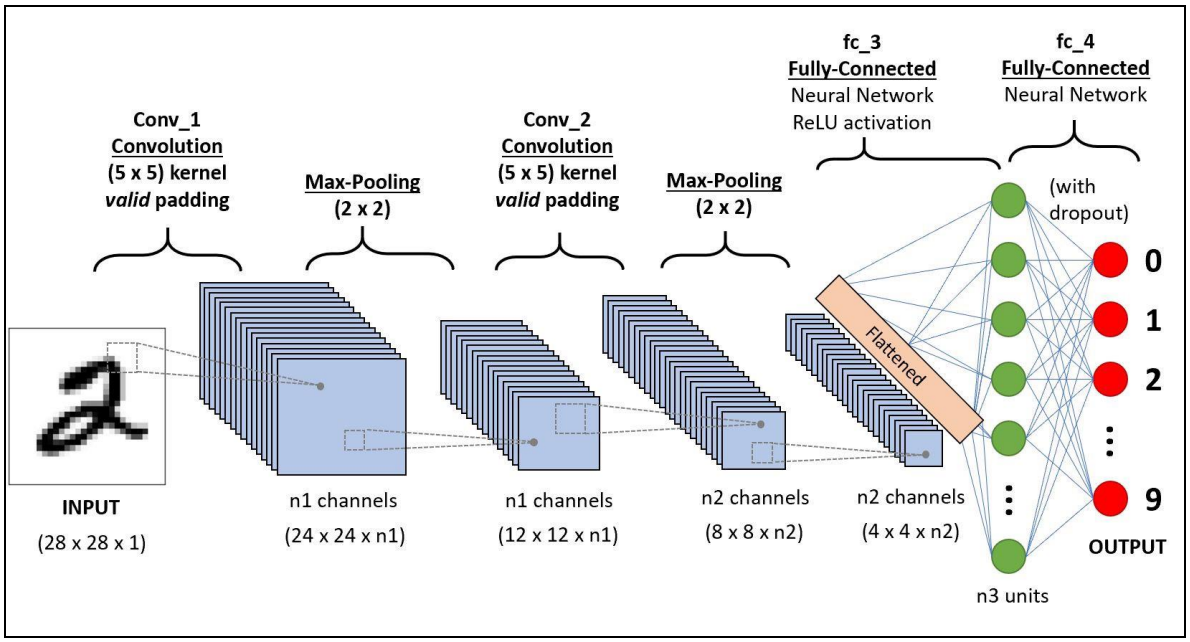


<그림 4 – 한국투자증권의 AIR>

국내에서 유일하게 AI를 이용한 리서치 서비스를 제공하고 있는 한국투자증권의 AIR의 경우, 상장기업 가운데 리서치 보고서가 한 번도 제공되지 않던 약 1500개의 소외 종목의 리서치가 제공 가능하다. 이는 잠재적 기업을 발굴해낼 수 있다는 기존 서비스의 장점으로 작용한다. 또한 실제 퀀트 애널리스트들을 중심으로 약 10만 건의 높은 퀄리티의 데이터를 이용해 모델링하였다는 점 역시 긍정적으로 작용한다.

하지만, 서비스의 런칭 기간이 짧아 실제로 투자했을 경우 그 성능을 알기 어렵다는 단점이 존재한다. 넓은 커버리지를 가지는 대신 높은 수준의 분석이 불가능하다는 것 역시 단점 가운데 하나이다. 서비스를 통해 분석 종목의 추후 양상을 파악하기 어렵다는 단점이 있다. 또한, AI를 활용해 실제 리포트를 생성하는 것까지는 이어지지 않아 해당 종목이 선정된 이유를 파악하기 어렵다.

2.3 차별화 핵심 전략 기술



<그림 5 – CNN 모델>

한국투자증권의 AIR 서비스의 경우, 뉴스 데이터를 3 진분류해 딥러닝 모델인 CNN 으로 분석을 진행했다. 반면 AI 애널리스트 서비스의 경우, 자연어 처리에 적합한 딥러닝 모델인 'HuggingFace'의 'BERT'를 활용해 뉴스 데이터를 분석한다. 또한 뉴스 기사 전처리 과정에서 발생 가능한 띄어쓰기 오류를 개선하기 위해 데이터 전처리 과정에서 맞춤법을 검사한다. 전처리한 데이터를 LDA 알고리즘을 이용해 뉴스 기사에서 가장 핵심이 되는 주제를 선정하고 이를 사용자에게 제공한다. 추가적으로 자연어 생성 모델인 'OpenAI'의 'GPT2'를 활용해 선정된 토픽을 바탕으로 짧은 보고서를 생성하는 것이 서비스의 목표이다.

3. 프로젝트 개발 결과

3.1 데이터 수집

서비스를 개발하기 위해 범용적인 사용을 고려하지 않고 하나의 종목을 특정했다. 대상 종목으로는 대표성과 데이터의 양을 고려해 삼성전자로 선택했다. 2005 년 1 월 1 일부터의 삼성전자 주가를 수집하되 기간 중에 있던 액면분할을 반영해 데이터를 수정했다. 액면분할이 되기 이전의 주식 종가를 50 으로 나누고, 거래량에 50 을 곱하는 과정을 진행했다. 뉴스 데이터의 경우 '삼성전자'가 들어간 뉴스 가운데 2005 년 1 월 1 일부터 최근까지의 뉴스를 수집했다. 경제에 특화된 뉴스를 수집하고자 연합뉴스, 연합인포맥스, 아시아경제, 이데일리, 머니투데이, 헤럴드경제, 파이낸셜뉴스, 한국경제, 매일경제 총 9 개의 뉴스사만 선택하여 수집했다.

프로토타입으로 서비스를 제작한 이후, 범용적인 사용을 위해 보고서를 작성하고자 하는 기업의 이름이 입력됐을 때, 데이터를 수집할 수 있도록 변경할 예정이다.

3.2 분석 진행

3.2.1 LDA 토픽 모델링

LDA 토픽 모델링을 이용해 설정한 기간동안 수집한 대량의 텍스트 데이터 중 가장 핵심적인 주제를 추출해낸다. 해당 주제가 언급된 뉴스만을 데이터로 이용하기 위해 LDA 토픽 모델링을 진행하였다.

LDA 토픽 모델링을 진행하기 위해 Coherence Score 와 Perplexity Score 를 구해 기간별 최적의 토픽 개수를 정한다. 이렇게 구한 토픽 개수만큼의 토픽들에 해당하는 기사를 추출한다. 기간을 일정하게 분할해 거래량이 상위 5%에 속하는 날짜만을 추출해 거래량에 큰 영향을 주었던 주제들을 뽑아내 해당 종목의 주요 토픽으로 설정한다.

수집된 토픽들의 FVE(Fraction Of Volume Explained) 스코어를 이용해 토픽들의 유효성을 검증한다. 사전에 추출한 거래량 상위 날짜에서 토픽들의 출현 빈도를 구해 FVE 가 높은 순서대로 토픽들을 다시 추출한다. 이상치를 제거하기 위해 하위 1%의 토픽은 제거해 총 세 번의 과정으로 토픽을 추출한다.

3.2.2 알고리즘 자동화

Scrapy 와 apscheduler 패키지를 활용해 일정 시간 마다 그날의 뉴스 데이터를 수집한다. 기사 내에 존재하는 불필요한 기호와 특수문자 텍스트를 제거하고 맞춤법을 검사하는 전처리 작업을 거쳐 기사 내에서 유의미한 텍스트 만을 추출한다. 이 전체 과정을 프로그램이 주기성을 두고 자동으로 실행한다.

3.2.3 예측 모델

예측 모델은 'XGBoost' 모델을 이용한다. XGBoost 는 'Gradient Boosting updated'를 이용한 모델이다. Gradient Boosting 은 'Adaboost' 와 같이 각 반복마다 데이터셋의 가중치를 수정하는 것이 아닌 이전 예측 모델에서 계산된 Cost 를 경사하강법을 통해 새로운 예측 모델을 학습시킨다. 본 프로젝트에서는 예측을 위한 모델로 'XGBoost' 종류 가운데 하나인 'XGBRegressor'을 이용함으로써 회귀를 통한 예측을 진행했다.

XGBRegressor 모델에 주식 데이터와 뉴스 텍스트 데이터를 설명변수로 예측에 활용한다. 주식 데이터의 경우, 일별 종목의 종가와 거래량, 시가총액을 사용하였다. 수집 기간 도중 액면분할 등으로 대상 종목의 주가가 변동했을 경우, 수정된 주가를 반영한다.

뉴스 텍스트 데이터의 경우, 전체 뉴스에서 LDA 토픽 모델링에 통해 뉴스를 추려낸다. 뉴스가 발간된 일자의 주가를 기준으로 일주일 뒤의 주가를 비교한다. 이때 일주일 뒤의 주가가 5% 이상 상승했을 경우 1 을, 5% 이하 하락했을 경우, 0 을 라벨링한다. -5%에서 +5%까지의 구간은 Grey Area 라고 판단한다. 이를 통해 주가 변동에 유의미한 뉴스만을 학습데이터로 이용하고자 하였다. 라벨링한 데이터를 자연어 딥러닝 모델 가운데 하나인 'Bert' 모델을 통해 Fine-Tunning 하였다. 'Google'의 'BERT base multilingual cased'의 경우 한국어 텍스트를 분석하는데 성능의 한계를 보이기 때문에 'SKT'에서 사전학습한 'KoBERT'를 이용하였다.

'BERTClassification'을 이용해 뉴스별로 예측을 진행한다. 이를 일별로 합산해 -1 과 1 사이로 정규화 시켜 XGBRegressor 의 설명변수로 이용한다.

3.2.4 웹 서비스 구현

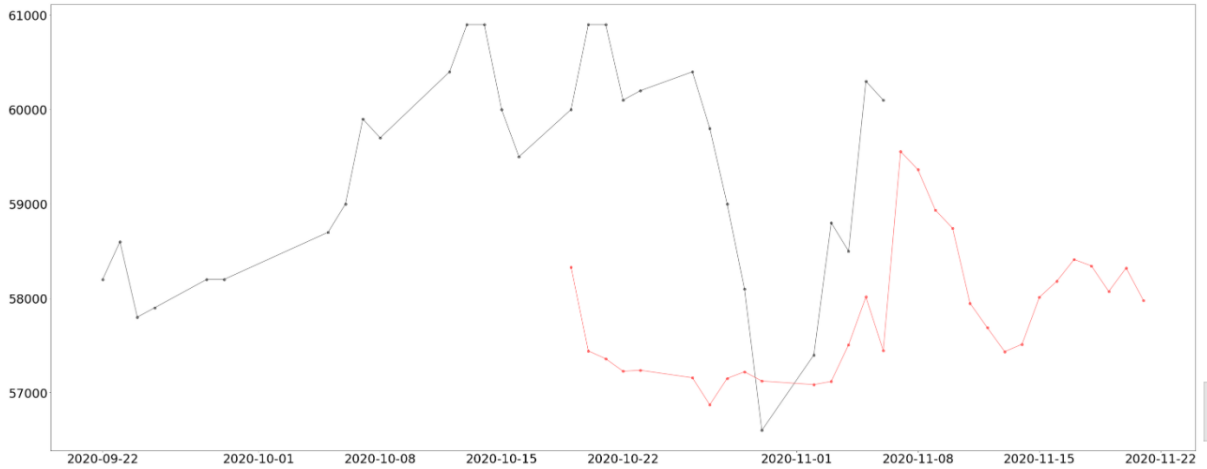
웹 서비스를 구현하기 위해 웹 프레임워크 가운데 하나인 플라스크를 이용한다. 해당 웹 페이지에는 지정 기간의 대상 종목의 종가와 코스피를 그래프로 제시한다. 그래프 아래에 대상 종목의 거래량에 많은 영향을 미친 핵심 키워드를 제시한다. 또한, 직전 분기에서 거래량에 많은 영향을 미친 키워드를 추려내고 해당 키워드 기반의 뉴스를 요약해서 사용자에게 제공하고자 한다.

<그림 6 - 종목 분석 보고서 예시>

3.3 분석 결과

최종 사용토픽		
T_INDEX	allow_topic	corr
68	[심의위, 검찰, 부회장, 수사]	0.282797
95	[평택, 공장, 낸드, 라인]	0.174342
79	[출시, 모델, 갤럭시, 카메라]	0.125227

LDA 토픽 모델링을 통해 지난 분기에 대한 토픽을 추출한 결과, 다음 도표와 같은 결과를 얻었다. 가장 상관관계가 높은 토픽으로는 '심의위, 검찰, 부회장, 수사' 가 추출되었다. 해당 토픽을 통해 가장 유사도가 높은 기사를 재추적한 결과, '연합뉴스'의 8 월 30 일자 보도 기사인 '삼성 합병·승계 수사 주초 결론…이재용 기소 유력'을 찾을 수 있었다. 이는 삼성전자의 주가 거래량에 실제로 큰 영향을 미치는 기사임을 확인할 수 있었다. 해당 방법을 이용해 지난 분기에 거래량에 큰 영향을 미친 주요 토픽들을 대상으로 뉴스를 요약하여 사용자에게 제공하고자 하였다.



<그림 7 – 삼성전자 주가 30 일 예측 예시>

XGBoost 를 활용한 주가 움직임 예측 결과는 위와 같다. XGBoost 의 예측 정확도를 확인하기 위해 MAPE(Mean Absolute Percentage Error), 우리 말로 평균절대비오차를 이용했다.

$$\mathrm{MAPE} = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

<그림 8 – MAPE 공식>

MAPE 를 이용해 오차를 계산한 결과, 403.06%의 Score 를 확인할 수 있었다. 지나치게 큰 오차가 나온 이유는 현재 설명변수로 투입된 것이 대상 종목 종가, 거래량 그리고 시가 총액 이렇게 세 개에 불과하기 때문이다. 세 가지의 설명 변수로는 향후 추세를 파악하는데 한계가 있기 때문에 이를 해결하기 위해 뉴스 데이터를 활용할 계획이다. 현재 BERT 모델의 경우, 학습 중이기에 학습이 끝나는 대로 데이터를 추가해서 다시 모델링을 진행할 계획이다.

다음의 결과를 종합적으로 고려했을 때, AI 애널리스트 프로그램은 삼성전자의 4Q 전망을 부정적으로 예측하여 '보류' 시그널로 결론지었다.

3.4 활용 방안

본 프로젝트는 국내 주식 가운데 가장 대표성을 지니는 삼성전자를 선택하여 하나의 종목만을 분석 대상으로 진행했다. 하지만 종목 분석 보고서 자동 생성을 위한 MVP 모델[1]을

[1] 최소 기능 제품 : 초기에 최소한의 기능을 구현한 제품으로 정식 출시 전 피드백을 반영하기 위한 목적으로 설계

제시하는데 의미가 있다. 뉴스와 주가를 데이터로 활용한 예측을 통해 해당 데이터를 이용하는 것이 유의미하다는 것을 증명하였다.

또한, 사용자가 종목명을 입력하면 해당 종목에 대한 데이터 수집부터 분석 보고서까지 받아 볼 수 있도록 프로젝트를 설계했다. 필요한 데이터의 수집부터 최종적인 웹서비스 구현까지 자동화를 적용하여 사용자의 편의성을 높일 것으로 기대된다.

증권사가 발행한 종목 분석 보고서만을 참고한 기존 방법의 한계에서 벗어나 정보 접근성과 다양성을 제고해 투자자가 언제든지 원하는 종목을 분석할 수 있도록 했다. 이를 통해 투자자에게 주식 거래의 정보 편향성을 줄이고 폭넓은 거래의 기회를 제공할 수 있다. 기존의 대형주 위주의 투자에서 포트폴리오의 다각화를 통해 변동성과 투자 리스크를 낮출 수 있을 것으로 기대된다.

제로금리와 부동산 시장의 경색으로 주식시장에 돈이 몰리면서 일명 빚투, 동학개미운동 등의 개인 투자 열풍이 불고 있다. 또한 공매도 금지가 내년 3 월까지 연장된 것에 이어 최근 주식 양도소득세의 기준이 10 억으로 유지되면서 당분간 활황세가 계속될 것이라는 전망이 우세하다.

증권사의 분석에서 소외된 중소 기업 입장에서도 투자자에게 정보를 제공할 수 있다. 해당 정보를 통해 중소 기업이 투자 받을 수 있는 확률을 높여 원활한 유동성의 확보가 가능할 것으로 기대된다. 특히 '한겨레 신문'에 따르면, 최근 코로나로 어려워진 경제상황에서 유동비율이 100%를 밑도는 기업이 186 곳인 것으로 나타났다. 특히, 이는 중소 기업과 중견 기업에 집중되었다. 전체 상장사 가운데 13%에 해당하는 기업들이 유동성 위기를 겪고 있는 만큼 중소기업에 대한 투자의 활성화가 절대적으로 필요하다. 이런 상황 속에서 AI 애널리스트는 정보 비대칭성을 해소함으로써 중소 기업 투자의 기회를 제공하고 유동성 양극화를 완화하는 효과가 있을 것으로 예상된다. 이런 시기에 AI 애널리스트가 큰 기여를 할 수 있을 것으로 기대된다.

추후, 더 적합한 자연어처리와 요약 모델을 적용해 성능이 향상되고 직관적인 종목 분석 보고서의 생성이 가능하다. 이를 통해 종목 분석 뿐만 아니라 환율이나 파생 상품 등 다양한 금융 상품에 적용을 통해 전반적인 금융 서비스 만족도 제고에 기여할 것이다.

4. 기대 효과

4.1 향후 개선 사항

- LDA 알고리즘 상관관계 개선
- GPT 모델 개선
- BERT 모델 개선을 통한 정확도 향상
- XGBoost 모델 개선을 통한 정확도 향상

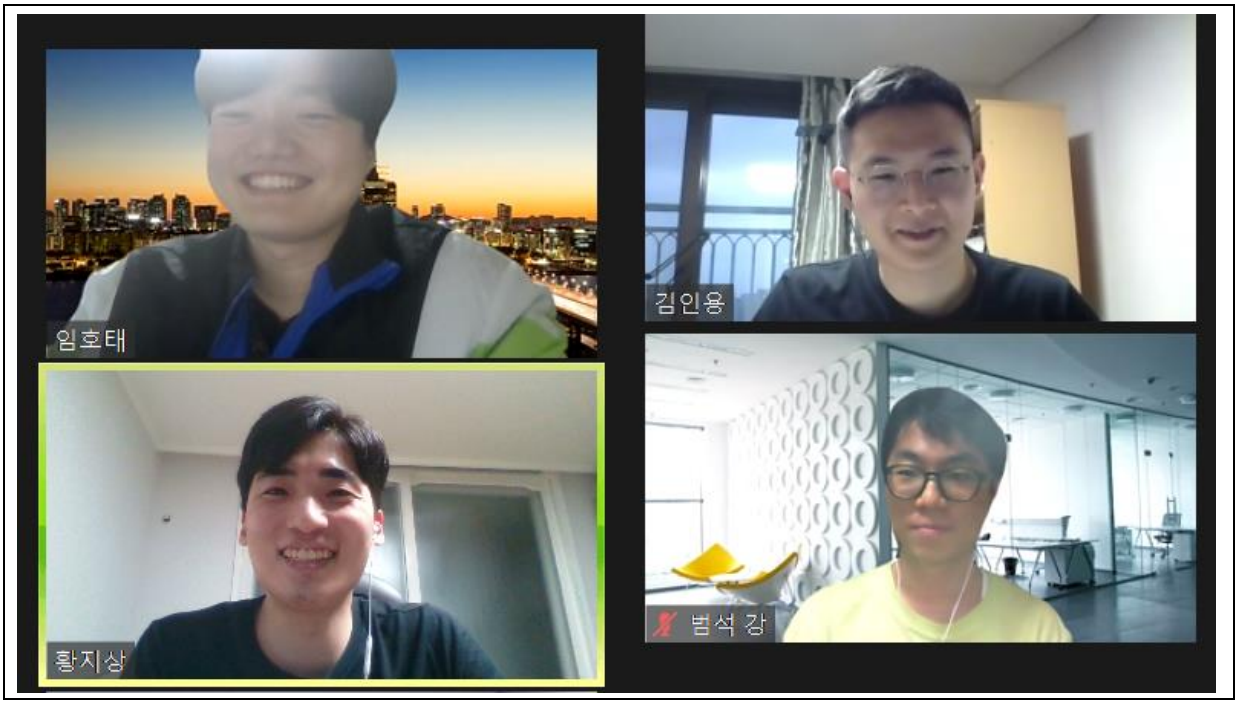
4.2 기대 효과

첫째, 인적자원의 부족으로 인해 대기업 위주로 주식 상품을 분석하는 편향적인 상황을 극복할 수 있다. AI 애널리스트 서비스를 통해 기존에는 시도할 엄두가 나지 않았던 중소형주 분석에 시간 자원을 투입할 수 있을 것으로 기대된다. 이를 통해 중소기업으로 투자가 활성화됨을 기대할 수 있다.

둘째, 인공지능을 이용했기에 애널리스트의 주관이 포함되지 않은 보고서를 고객에게 제공할 수 있다. 지난 분기의 매출에 영향을 끼친 주제들을 키워드로 제공함으로써 종목별로 지난 분기의 키워드를 파악할 수 있다. 고객들에게 분기 별 키워드를 함께 제공함으로써 수많은 기업들의 정보를 어려움 없이 접할 수 있는 효과를 기대할 수 있다.

셋째, 생성모델을 이용해 주가 변동에 유의미했던 키워드를 문장으로 바꿔준다. 이를 사용자가 분석 보고서를 작성하는 첫 문장으로 사용할 수 있다. 첫 문장을 사용자에게 제공해줌으로써 작성하는 보고서에서 진행되어야 하는 내용을 작성하는데 도움을 얻을 수 있을 것으로 기대된다.

5. 개발 후기



성명	후기
임호태	금융사에서 진행되고 있는 다양한 IT 서비스들을 리서치하며 내가 할 수 있는 부분은 어떤 것이 있으며, 어떤 공부를 추가적으로 해야겠다는 방향성을 잡을 수 있어서 좋은 기회였다고 생각합니다. 향후에도 디지털 금융 분야에서 많은 사람들의 자산을 안정적으로 관리할 수 있도록 돕는 일을 하고 싶습니다.
황지상	자연어 처리에 사용되는 최신 딥러닝 모델을 이용하게 된 점은 뜻 깊었다. 하지만 한국어 자연어 처리에 적용할 독보적으로 뛰어난 모델이 없음에 만족스러운 결과가 나오지 못했다. 이를 해결하기 위해서는 국가와 기업에서 자연어 처리 모델에 대해 많은 관심을 갖고 개발할 필요가 있다. 해당 프로젝트는 주가 분석을 위한 툴로 BERT 를 적용했다. 이를 허깅페이스의 AlBERT, RoBERTa 나 스탠포드 연구팀의 ELECTRA 같은 최신 모델로 바꿔보는 것도 좋은 결과물을 만들 수단으로 예상된다. 이를 위해서는 잘 정제된 한국어 텍스트 데이터를 구축하기 위해 데이터를 전처리하는 것이 더욱 강조될 것이다.

김인용	한국어 자연어처리 라이브러리가 다양하지 않았던 것이 성능에 영향을 미친 것 같아 아쉽다. 프로젝트를 통해 한국어 자연어처리 추가 개발 필요성을 느꼈다. 또한 시간이 부족해 만족스러운 결과를 뽑지 못했다. 그렇지만 금융 데이터와 한국어 텍스트를 이용해 새로운 인사이트를 도출할 수 있었기에 의미있는 경험이었다. 추후, 실제 활용할 수 있도록 보완하겠다.
강범석	프로젝트에서 LDA 토픽 모델링을 활용하기 위해 논문을 참조하여 구현해보고 적용시키는 과정이 매우 유익했다. 그러나 LDA 토픽 모델링 결과를 다른 모델과 접목시키는데 어려움이 있었다. 또한 프로젝트를 진행하면서 한국어 자연어처리를 위해 사용할 수 있는 라이브러리가 적었고 시간이 부족하여 뛰어난 결과를 도출하지 못한 점이 아쉬웠다. 추후에 자연어 처리를 위한 다양한 모델을 활용하여 성능을 향상시킬 수 있도록 하겠다.
김도현	실제 증권사 보고서처럼 자연스러운 맥락으로 구성된 AI 보고서를 만들수는 없었지만, 해당 프로젝트를 점차적으로 보완해 나간다면 증시 뿐만 아니라 다양한 분야에 적용이 가능할 것이라 생각한다. 따라서 추후에는 현재 자연스럽지 않은 문장들에 대해 좀더 증시에 맞는 간단한 워딩으로 처리될 수 있도록 하여 개선해 나아가고자 한다.

6. 참고 자료

6.1 참고 논문

- "Language Models are Unsupervised Multitask Learners", Alec Radford et al, OpenAI, 2019.02
- "High quality topic extraction from business news explains abnormal financial market volatility", Ryohei Hisano et al, JSPS Grants-in-Aid for Scientific Research, 2012.10

6.2 참고 패키지

- "KoBERT", https://github.com/SKTBrain/KoBERT
- "KoGPT2", https://github.com/SKT-AI/KoGPT2
- "KoGPT2-chatbot", https://github.com/haven-jeon/KoGPT2-chatbot
- "py-hanspell", https://github.com/ssut/py-hanspell
- "pyeunjeon", https://github.com/koshort/pyeunjeon