인공지능융합실습 프로젝트 최종 발표

★ Finalyst 재무오피니언 분석 서비스

Finalyst 기업 재무 이슈 분석 서비스

CONTENTS

1 Background

2 Business Modeling

3 Product - Phase 1

4 Modeling & Prototype

Project Background

리서치 & 서류작업 간소화

리서치와 서류작업 간소화 니즈

금융권 종사자들이 가장 많이 시간을 쓰는 업무는 리서치와 서류작업이다



1. 기업 리서치



2. PPT/보고서 작성



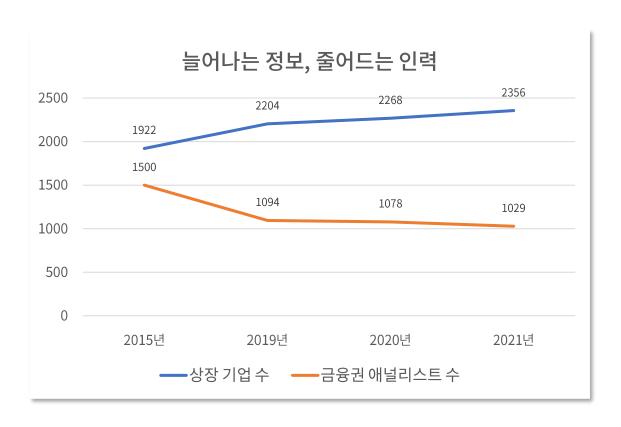
3. 발표

Project Background

금융정보에 대한 수요 공급의 간극

리서치와 서류작업 간소화 니즈

금융권 종사자들이 가장 많이 시간을 쓰는 업무는 리서치와 서류작업이다



정보의 증대에 비해 부족한 애널리스트 수

- 1. 증가하는 상장회사 수 및 시장규모
- 2. 금융투자협회 기준 60개 증권사 1,051명의 애널리스트 ('22.12.11 기준)
- 3. 금융환경의 변화(HTS, MTS)로 리서치의 수익성 악화 (비용부서화)
- 4. 수익성 악화로 인한 인력감축 → 업무과중
- 5. 비용절감 및 효율성 제고과정 필요



CONTENTS

1 Background

2 Business Modeling

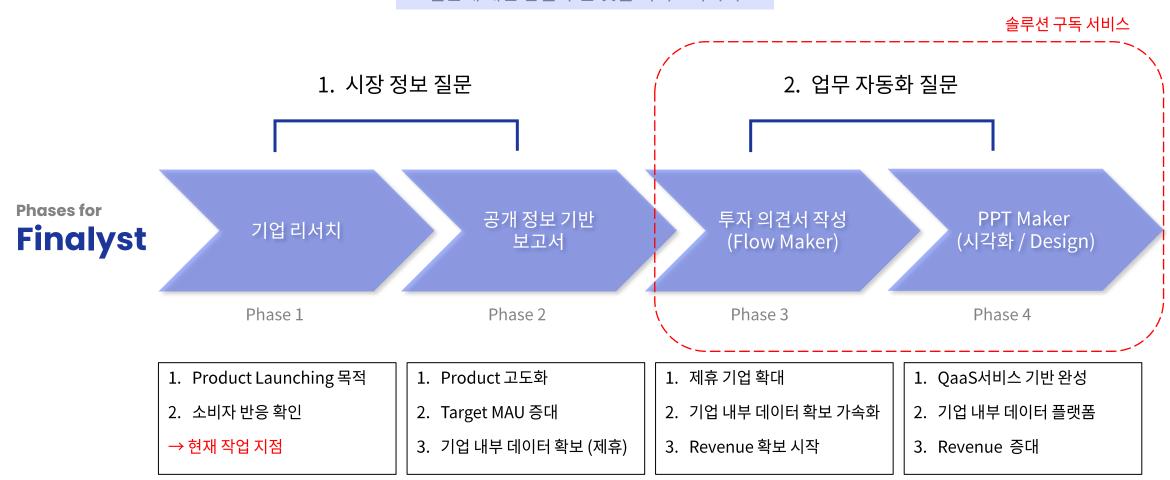
3 Product - Phase 1

4 Modeling & Prototype

QAAS - Query as a Service

QaaS (Query as a Service)

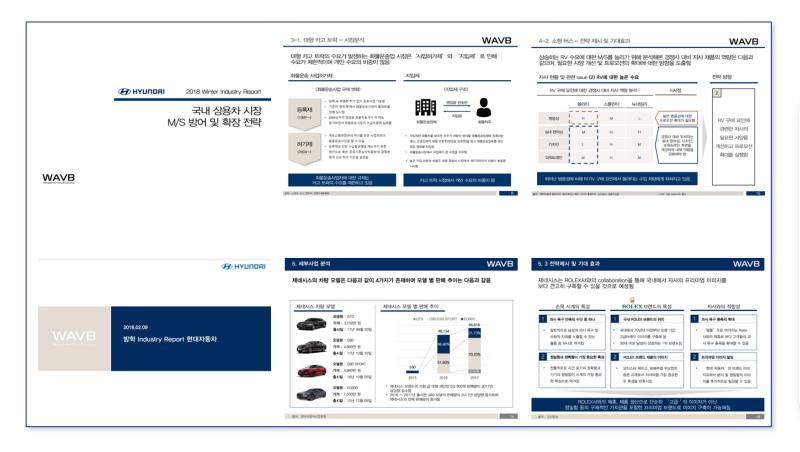
질문에 대한 답을 주는 것을 서비스화하자



Business Aim

Business Aim - 기업 내부 데이터 확보

전문가들의 의견 사항을 데이터화



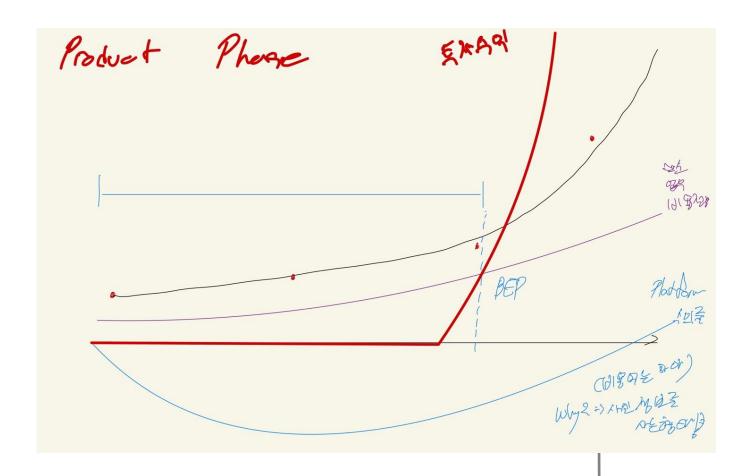
정보의 증대에 비해 부족한 애널리스트 수

- 1. 금융권/컨설팅: 피피티 형태 보고서 많음
- 2. 기업에 대한 투자 의견, 이슈 해석 정보 多
- 3. Due Diligence가 종료되면 공개 가능
- 4. 작성 시 시간 소요 多, but 재활용 잘 안됨
- 5. 비정형 형태이기 때문에 DB화도 어려움
- 6. DT도 잘 안되어 있으며 진입장벽 높음

Business Aim

Business Aim - 기업 내부 데이터 확보

전문가들의 의견 사항을 데이터화



Revenue Earning References

- 1. 혁신의 숲:
 - (seed StartUp)
 - 2. BM: 제휴 / Fund / 엑셀러레이팅
- 2. TheVC
 - 1. BM: Premium 구독
- 3. FnGuide
 - 1. BM: Premium 구독

Business Aim

Business Aim - 기업 내부 데이터 확보

전문가들의 의견 사항을 데이터화

Aim - Product Maker

- 1. 전문가 금융 정보 활용 데이터 수집
- 2. 리서치 / 업무 간편화
- 3. 투자 / 경영 활용
 - 1. Start-Up: Seed 투자 & 액셀러레이팅
 - 2. 중소기업: 저평가 기업 운영
 - 3. 상장사: 주식 투자 & M&A
- 4. 영업 이익 (구독료 / 제휴 수수료)

Aim - Users

- 1. 리서치 / 업무 간편화 (사용성 강화)
 - 1. 리서치
 - 2. 자료정리
 - 3. 의견 정리
 - 4. PPT 작성

CONTENTS

1 Background

2 Business Modeling

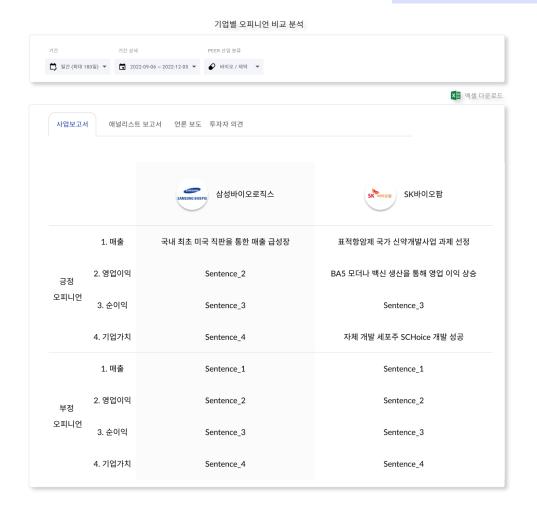
3 Product – Phase 1

4 Modeling & Prototype

결과물 형태

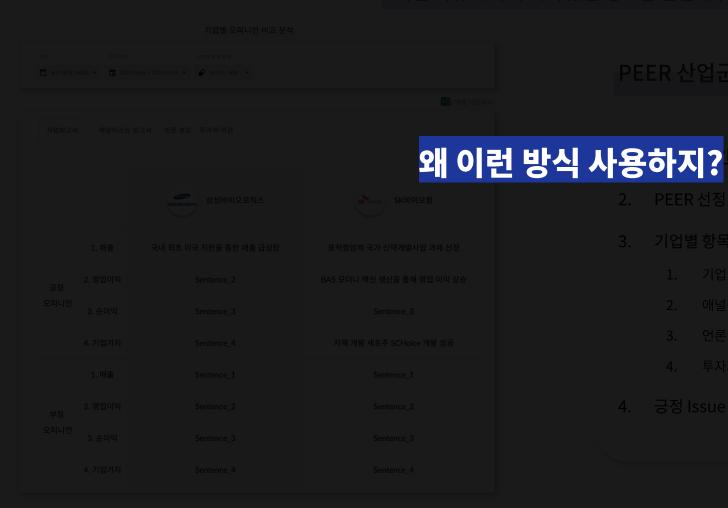
Phase 1 - 대상 기업 Issue 오피니언 정리

기업 이슈에 대해 의미 있는 정보를 선별해주자



PEER 산업군을 고려한 기업별 이슈 비교

- 1. 리서치 기간 설정
- 2. PEER 선정
- 3. 기업별 항목에 대한 이해관계자별 의견 정리
 - 1. 기업 자체 의견
 - 2. 애널리스트 의견
 - 3. 언론 의견
 - 4. 투자자 의견
- 4. 긍정 Issue / 부정 Issue에 따른 비교



Research Problem

기업 리서치 Problem - 무엇이 중요한지 알기 어렵다

금융시장 정보의 특징

금융 정보의 특징

- 1. 다양한 분야의 수많은 기업들
- 2. 각 사업분야별 이슈의 다양성
- 3. 사업보고서 등 공시자료의 해석이 어려움
- 4. 급변하는 시장상황, 그에 따른 방대한 양의 정보 (뉴스, 애널리스트 리포트 등)

금융 정보 활용 핵심

- 1. 빠른 의사결정
- 2. 적절한 기업선정(비교정보)
- 3. 중요한 이슈의 판단
- 4. 수익성

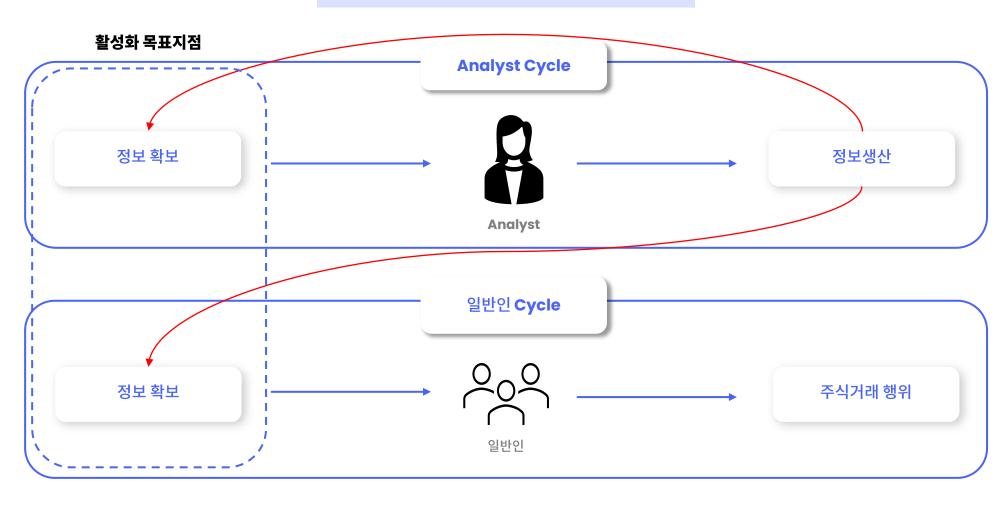
결론

- 1. Peer그룹에 대한 비교정보도 제공해야 함
- 2. 특정 기업 정보 중 무엇이 중요한지 판단해야 함
- 3. 중요 정보 중 기업에게 호재인지 악재인지 판단 필요
- 4. 의사결정과정을 간소화 필요

Goal

금융 정보 확보를 용이하게 해주자

기업 이슈에 대해 의미 있는 정보를 선별해주자



Product Funnel

Finalyst (Finance + Analyst)

복잡하고 방대한 양의 금융정보를 간편하게 처리

Phasel - Funnel Finalyst

기업명 입력

대상기업 및 Peer 정보 확보

Data 요약 정리

분류별 오피니언 출력

정보 수집

- 1. 기업의 공시정보(연내)
- 2. 네이버 뉴스(최근 3개월)
- 3. 애널리스트보고서
- 4. 종목토론방 반응

모델

Kor – FinBERT

440,067개의 뉴스기사 11,237개의 리포트 총 6,379,315개의 문장으로 pre-trained

Fine tuning

4,827개의 긍정 / 부정 dataset 사용

키워드 출력 (단일기업)

	긍정	부정
공시정보		
뉴스		
보고서		
종목토론방		

CONTENTS

1 Background

2 Business Modeling

3 Product - Phase 1

4 Modeling & Prototype

Modeling

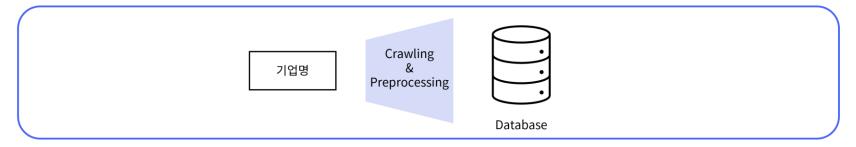
Model Pipeline

Model Pipeline

Flow

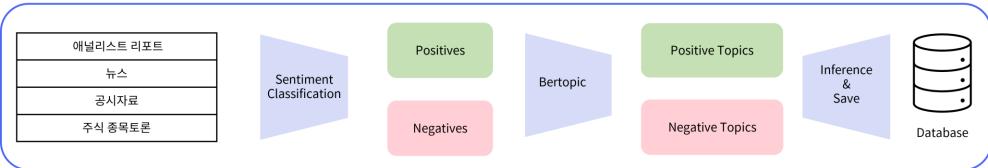
Phase1

Crawling



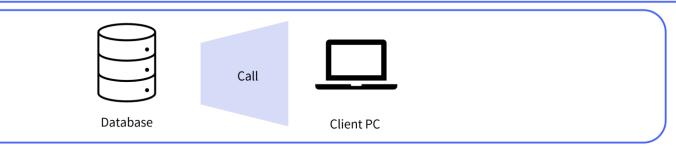
Phase2

Modeling



Phase3

Product







Data Crawling

	사업보고서	애널리스트 보고서	언론 브리핑	투자자 의견
Source	DART 정기보고서	한경 컨센서스 애널리스트 보고서	네이버 뉴스 제목 및 본문 요약	네이버 주식 종목토론방
Method	OpenDartReader API	Pdfminer3 API	BeautifulSoup API random_user_agent API	BeautifulSoup API random_user_agent API
Period	2021.12. ~ 2022.09. (2021.12 사업보고서 2022.03 분기보고서 2022.06 반기보고서 2022.09 분기보고서)	2021.12.11. ~ 2022.12.11. (사이트 내 업로드 날짜 기준)	2021.12.11. ~ 2022.12.11. (뉴스업로드날짜기준)	2021.12.11. ~ 2022.12.11. (댓글등록날짜기준)
Sentence Split	'다.' 기준으로 분리	'다.' 기준으로 분리	제목, 본문 요약문을 각각 한 문장으로 취급	댓글 하나를 한 문장으로 취급



Data Crawling and Preprocessing

Data Crawling Examples

사업보고서	"당사는 업계 평균대비 배 향상된 세포 생존력과 개선된 증식력을 보여주는 자체 개발 세포주 에스초이스SCHOice 개발에 성공하였고 개발 단계별로 표준화된 프로세스를 통해 개발기간을 혁신적으로 단축시킨 위탁 개발 플랫폼 에스셀러레이트SCellerate를 런칭하였습니다." "삼성바이오로직스는 바이오의약품을 위탁생산하는 CMO 사업 세포주개발 공정 제형 분석법 세포주 개발초기 임상까지 개발서비스를 제공하는 CDO사업 바이오시밀러 개발 상업화에 중점을 두고 사업을 영위 중입니다."
애널리스트 리포트	"국내 바이오 회사 최초로 미국 직판을 위해 선제적으로 투자를 했기 영업망을 다지는 시기가 지나면 비용의 안정화와 처방수의 J커브를 기대해볼 수 있다"
언론 브리핑	"BA5 모더나 백신 긴급사용승인 삼성바이오로직스 생산" "SK바이오팜 표적항암제 국가신약개발사업 과제 선정" "SK바이오팜 세노바메이트 엔도그룹에 기술수출캐나다서 상업화"
투자자 의견	"미국에서 희귀의약품으로 지정했으면 상한번 쳐야지" "SK바이오팜 항암 신약 국가신약개발사업 과제 선정" "단일가 6만원대 실화냐"



Preprocessing

전처리 필요성

- 1. 투자자 의견 등에는 은어, 속어 등 다수 포함
- 2. 다른 소스에서도 무의미한 패턴 다수 존재

Text Preprocessing

- 1. 불용어 제거
- 2. 특수문자 제거
 - 3. 공백 제거

POS Tagging Based Preprocessing

(konlpy OKT Tokenizer)

- 1. 6글자 이상의 'Number' tagged token 제거
 - 2. 'Verb' tagged token 없는 데이터 제거
- 3. 'Puctuation', 'Number', 'Alpha' tagged token 로만 이루어진 데이터 제거
 - 4. 4,000자 이상 데이터 제거



Model - Fin-BERT

Fin-BERT 특징

- 1. 문장의 긍/부정 판단
- 2. **금융 관련 어휘에 특화**
- 3. **높은 확장성** (by finetuning)

Model Pipeline

Fin-BERT

Positive	Negative
현대바이오, '폴리탁셀' 코로나19 치료 가능성에 19% 급등	영화관株 '코로나 빙하기' 언제 끝나나"CJ CGV 올 4000억 손실 날수도"
이수화학, 3분기 영업익 176억전년比 80%↑	C쇼크에 멈춘 흑자비행대한항공 1분기 영업적자 566억
"GKL, 7년 만에 두 자릿수 매출성장 예상"	'1000억대 횡령·배임' 최신원 회장 구속 SK네트웍스 "경 영 공백 방지 최선"
위지윅스튜디오, 콘텐츠 활약에 사상 첫 매출 1000억원 돌파	부품 공급 차질에기아차 광주공장 전면 가동 중단
삼성전자, 2년 만에 인도 스마트폰 시장 점유율 1 위 '왕좌 탈환'	현대제철, 지난해 영업익 3,313억원···전년比 67.7% 감소

Model Pipeline

BERTopic Examples

공시자료

뉴스정보

종목토론방

삼성 바이오

```
Positive Topic Modeling ...
OMP: Info #276: omp_set_nested routine deprecated, please use omp_set_max_active_levels instead.
               -1_체결 _순차 _기술 _지속
              0_개 발 _세 포 _에 스 _연 구 개 발
               1_공장_부상_손상_원가
         13 2_치료_서비스_분야_자가면역질환
              3_바이오_밀러_의약품_시장
              4_사용권_리스_분기_에너지
               5 계약 상승 이익 증가
               6 환경 관리 위해 녹색
             7_사 외 이 사 _위 원 회 _사 업 _임 원
Negative Topic Modeling ...
       9 -1_원리금_능력_등급_지급
         13 0 감사 압수수색 혐의 법률
         14 1_위험_유동성_어려움_의미
         14 2 항소 기업 직원 유상증자
         24 3_위험_신용_거래_금융
    5 29 5_자산_부채_증가_분기
```

```
Positive Topic Modeling ...
OMP: Info #276: omp_set_nested routine deprecated, please use omp_set_max_active_levels instead.
    -1 162 -1_선물_급등_징조_실적
         120 0_쌈 바 _손 실 _복 구 _삼 바
48 1_삼 바 _아 주 _적 정 _정 주
          42 2_세계_연간_호재_종목
          36 3_수급_가격_주식_공매도
          34 4_목표_제시_바이오_직스
          28 5_바이오_직스_미국_치매
          23 6_오늘_어제_매수_성공
          22 7_정 상 _슈 팅 _순 이 익 _최 소
          20 9_폭락_나스닥_폭발_
Negative Topic Modeling ...
    -1 14 -1_개미_박스_하향_이탈
                 0_뉴 스 ___
         31 1_선물_급락_보미_긴급
         21 2 미국 오늘밤 막판 급락
    3 410 3_하락_삼바_공매도_바이오
```

SK 바이오

```
Positive Topic Modeling ...

MP: Info #276: omp_set_nested routine deprecated, please use omp_set_max_active_levels instead.

Topic Count Name
0 -1 55 -1 밀러 회사 지속 기술
1 0 11 0 개발 세포 에스 연구개발
2 1 22 1 28 나는 상소 실원가
3 2 36 2 개약 상송 이익 증가
4 3 13 3.처로 서비스 분야 자가면역질환
5 4 44 4 나이오 사용된 규모 시항
6 5 20 5 사임이사 사업 위원회 임원
7 6 19 6 환경 근라 의위에 녹색

Negative Topic Modeling ...

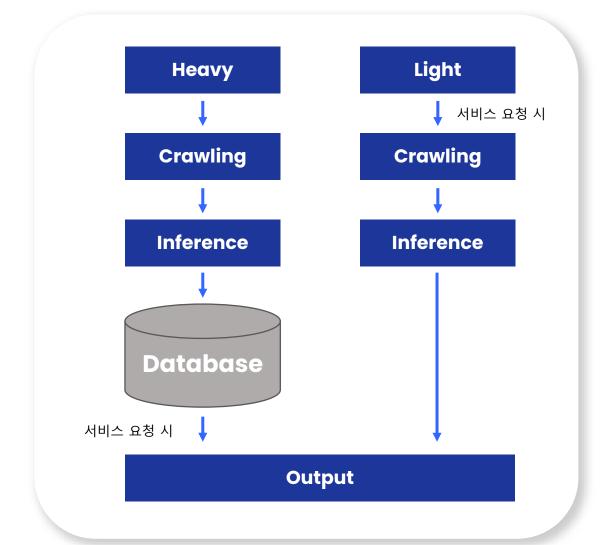
Negative Topic Modeling ...

Topic Count Name
0 13 0.감사 입수수색 현의 법률
1 1 19 1.위험 유동성 총족 어려움
2 2 5 12 고자산 부채리스 단위
3 3 24 3.위험 신용 기례 급용
4 4 14 4.기업 항소 직원 유상성 중자
```



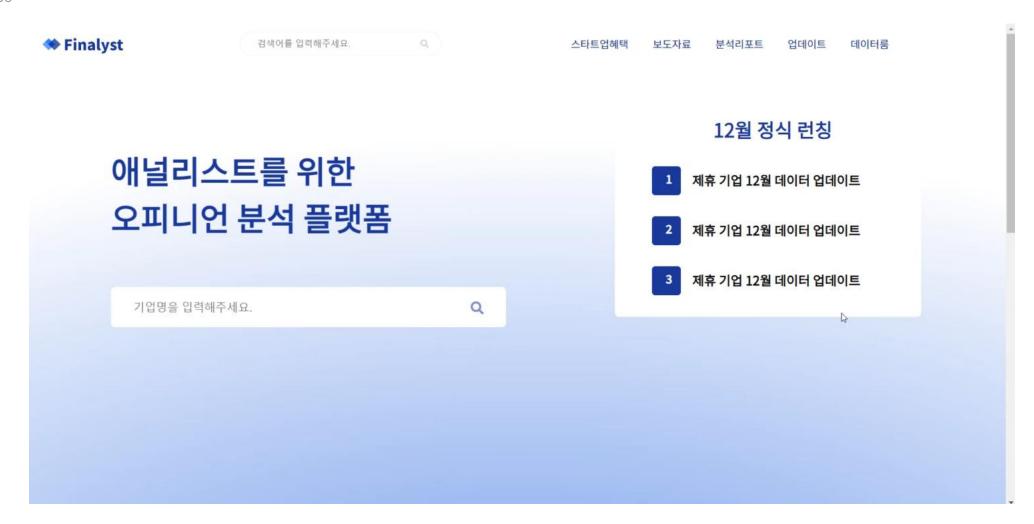
데이터 관리 및 구성

- 1. DB: db.sqlite3 사용
- 2. Crawling, Inference가 오래 걸림
 → DB에 데이터 저장 후, 요청이 들어오면 출력
- 3. DB에 저장되는 기업 → KOSPI 상장사 (Heavy)
- 4. 저장되지 않는 경우 (상장사가 아닌 경우)
 - → 적은 데이터 규모
 - → 상대적으로 적은 시간으로 출력 가능
- 5. Crawling Data 일별 수집 00시 기준
 - 최종 크롤링 위치 이후부터 업데이트
- 6. 누적해서 커지는 규모의 DB관리 대책 필요
- 7. 모델 경량화 방식 고민 필요



Prototype

Prototype



Prototype

Prototype

Open Service

1. 기업에 대한 이슈를 Topic화해서 제공

2. Topic의 긍정, 부정 여부 제공

3. Peer와 비교 분석





인공지능융합실습 프로젝트 최종 발표

Thank You!

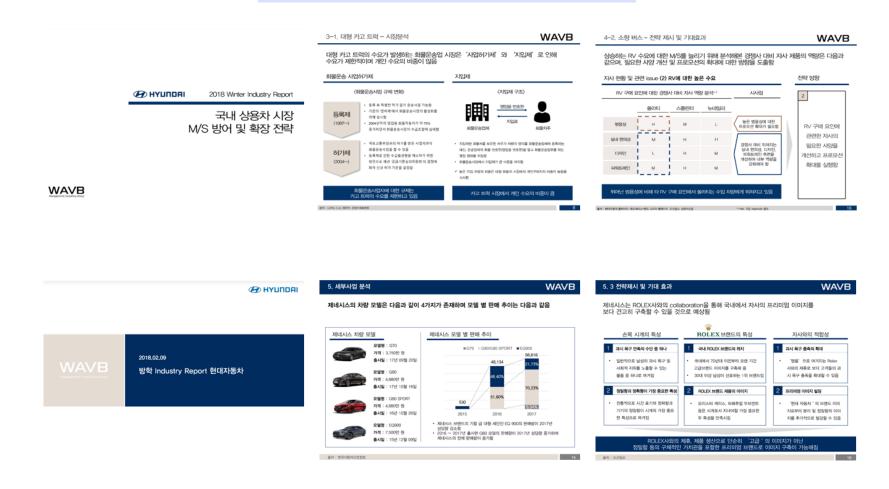
Finalyst | 기업 재무 이슈 분석 서비스

Appendix

금융/컨설팅 데이터 - 장표

장표 데이터 - Flow & 의견

전문가들의 의견 사항을 데이터화



Appendix - Business Model

Considerations

Targets (Subscribe)

- 1. 증권사 리서치 부서(애널리스트)
- 2. 투자에 대해 어느 정도 관심이 있는 사람
- 3. 금융 요약정보 텍스트를 바탕으로 주체적 이해가능자

Services For Subscribers

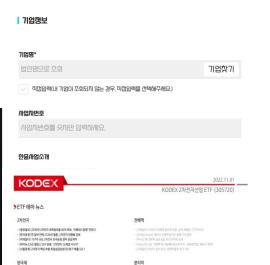
- 1. ETF 분석
 - 1) ETF에 포함된 각 기업에 대한 이슈분석 & 공통이슈 제공
 - 2) 각 기업에 대한 긍/부정 Prob 가중합 → ETF 평가
- 2. 기업 및 각 Topic에 대한 긍정/부정 점수 제공
- 3. 산업별 분석 제공
- 4. 목적에 따른 비교군 제공양상 변화 ex) 유사 기업(기본옵션), 동일 Thema, 위험 Hedging 등
- 5. 비상장기업과 제휴를 통해 정보제공(유료플랜)

References





스타트업 등록하기



[디지털역달리] 왜 안의제 5사, 3분기 역대급 실적, 해외 전출 본격회

Appendix - Business Model

Considerations

Why 주체적 이용자?

- 1. 기업정보를 '탐색'하는 시점에서 어느 정도 투자에 대해 관심
- 2. 전문가의 경우 리서치 과정에서 이용할 수 있는 간편한 툴
- 3. 일반인의 경우 복잡한 환경에서 '무엇부터 봐야'하는지 제공

Why 구독모델?

- 1. 금융정보는 그 자체로 가치를 지님
- 2. AI를 통해 다양한 기업에 대한 탐색을 돕기 위한 서비스 → 정보에 대한 접근권한을 주는 것이 효과적
- 3. 국내 금융정보 제공은 대체로 구독모델을 통해 이루어짐 ϵ x) FnGuide, 한경컨센서스 등
- 4. 초기 시장진입을 위해서는 무료플랜이 필요

From 상장, To 비상장

From 상장

- 1. 애널리스트 리포트는 공중을 대상으로 투자의견 제시
- 2. 일반인 입장에서 주 투자대상은 상장기업임
- 3. 초기 성장은 상장사 정보제공을 바탕으로

To 비상장

- 1. 비상장기업/스타트업: 정보가 적음
- 2. 정보를 직접 전달받는 과정이 중요
- 3. 따라서, 제휴 등의 방법으로 정보 수집
- 4. 정보에 대해 지불용의가 높은 대상에 한정해서만 공개

Appendix - Expected Benefit

Considerations

Analyst Benefit

- 1. 증권사 리서치 부서를 대상
- 2. 애널리스트 등 금융전문가의 노동생산성 증가
- 3. 초기 프로세스 자동화를 통한 기회비용 절감
- 4. 일반투자자 인식정보 제공, 이슈와 괴리 확인

일반사용자 Benefit

- 1. 금융정보 이해력 증진
- 2. 정보를 바탕으로 합리적 투자 가능성

Side Effect

- 1. 이슈/평가간 딜레이를 줄임 → 적절한 자원배분
- 2. 다양한 기업에 대한 주가정보성 제고
- 3. 주식매매 활성화 → 증권사 수수료 증가

Appendix - Complementary

Considerations

Complementary

Legal Issues

- 1. Analyst Report의 저작권: 발행 증권사 소유
 → 재생산에 대한 법률적 검토 필요
- FnGuide 등에서 리포트와 그 요약 등을 제공하고 있음. 이와 관련하여, 해당 서비스와의 법리적 차이 가 무엇일지 면밀한 검토가 필요함

Model

- 1. 긍정/부정에 대한 분류 정확도는 높음
- 2. 단어 인식에 대한 문제 ex) 셀트리 온 램시마: '시마'로 인식
- 3. 모델 경량화 방식 고민 필요

Compliance notice

- 본 조사분석자료의 애널리스트는 7월 29일 현재 위 조사분석자료에 언급된 종목의 지분을 보유하고 있지 않습니다.
- 당사는 7월 29일 현재 위 조사분석자료에 언급된 종목의 지분을 1% 이상 보유하고 있지 않습니다.
- 본 조사분석자료에는 외부의 부당한 압력이나 간섭 없이 애널리스트의 의견이 정확하게 반영되었음을 확인합니다.
- 본 조사분석자료는 당사의 저작물로서 모든 저작권은 당사에게 있습니다.
- 본 조사분석자료는 당사의 동의 없이 어떠한 경우에도 어떠한 형태로든 복제, 배포, 전송, 변형, 대여할 수 없습니다.
- 본 조사분석자료에 수록된 내용은 당사 리서치센터가 신뢰할 만한 자료 및 정보로부터 얻어진 것이나, 당사는 그 정확성이나 완전성을 보장할 수 없습니다. 따라서 어떠한 경우에도 본 자료는 고객의 주식투자의 결과에 대한 법적 책임소재에 대한 증빙자료로 사용될 수 없습니다.
- 본 조사분석자료는 기관투자가 등 제3자에게 사전 제공된 사실이 없습니다.

Positive Topic Modeling ...

OMP: Info #276: omp_set_nested routine deprecated, please use omp_set_max_active_levels instead.

Negative Topic Modeling ...

Topic Count Name 0 -1 31 -1_주가_이익_목표_영업

https://medifonews.com → mobile → article ▼

셀트리온 '램시마SC', 세계적 권위 유럽류마티스학회서 호평

2022. 6. 3. — 램**시마**SC는 TNF- α (종양괴사인자) 억제제 시장에서 폭넓게 쓰이는 인플릭시맙을 피하주사(SC) 제형으로 개발한 바이오베터(Biobetter) 제품으로, 전 세계 ...