개인 맞춤형 권리락과
SNS를 활용한 주가예측 서비스

Stock CM

㈜ SI투자증권
정선일 강다현
서민규 윤정은
임대진 정혜선

INDEX

01 제안 배경

- 현재 실태
- 현 시장
- 제안 방법

02 기능설계

- 유스케이스 다이어그램
- WBS
- 시스템 흐름도

03 기능구현 / 개발 과정

- 주가예측
- SNS 데이터 수집, 전처리
- KMV모델을 통한 예상부도 확률
- 최대 예상 손실금액 VaR

04 시연영상

05 기대효과 / 향후발전

06 팀원 소개

07 참고문헌

제안배경

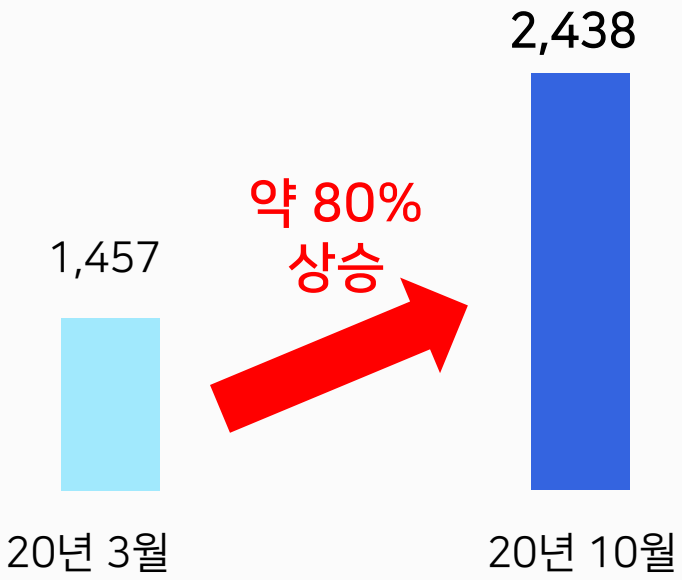
01

주식에 대한 관심 증가

지난 해, 주식에 대한 관심이 증가함에 따라 신규 가입자가 급증하고 있다.

작년 한해 KOSPI 지수 약 80% 상승

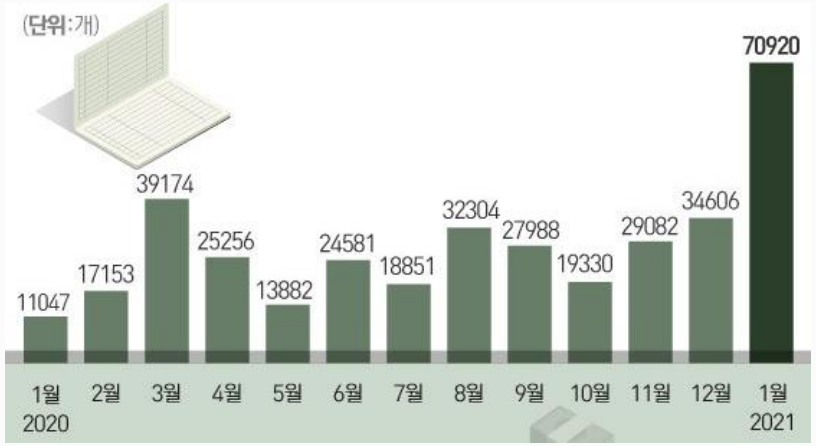
[2020년 3월 ~ 10월 KOSPI지수]



코로나 19 이후 주식에 대한 관심 증가

[2020년 월별 일평균 주식거래활동 계좌 증가 수]

출처 : 금융투자협회



새해 들어 주식 활동 계좌가 하루에 7만개로 급증
2021년 1월 **한달**동안 무려 **141만개 증가**
2020년 3월에 비해 1.8배 증가

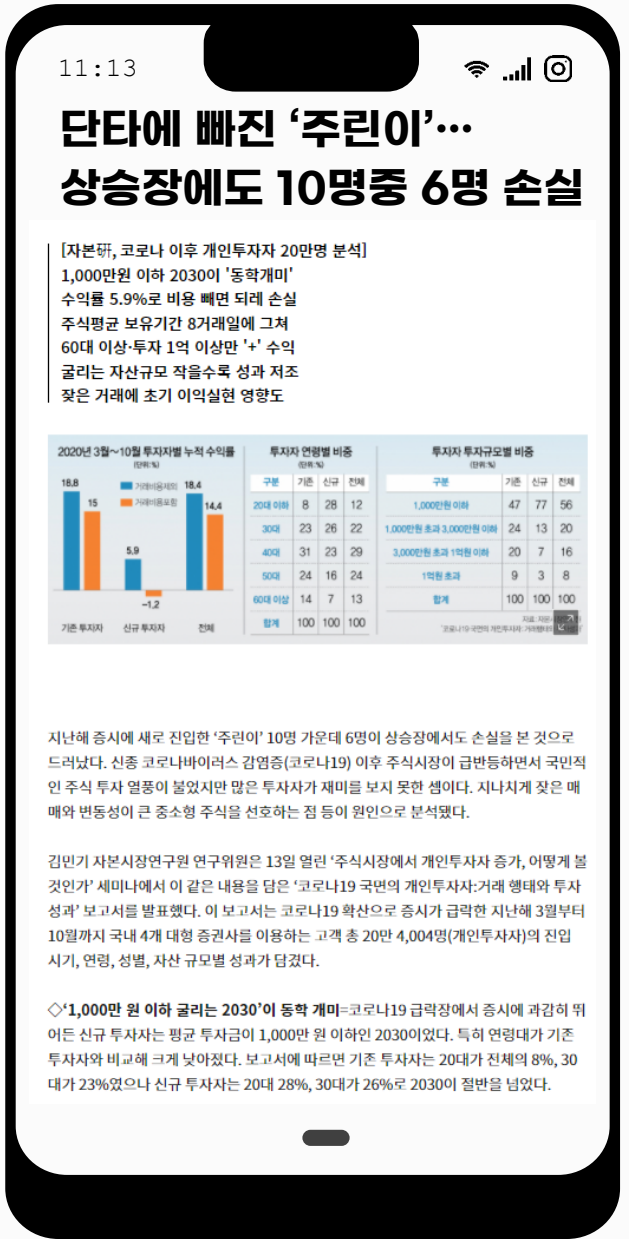
주식 입문자에겐 어려운 주식시장

주식의 진입장벽은 낮아졌지만, 낯선 주식 시장에서
신규 투자자의 성과는 -1.2% 에 불과하다.

KOSPI 지수가 **약 80% 상승** 했음에도
신규 투자자의 성과는 **-1.2%**에 불과함

[2020년 3월 ~ 10월 투자자별 누적 수익률]

(단위 : %)



출처 : 서울경제

주식 입문자에겐 어려운 주식시장

주식의 진입장벽은 낮아졌지만, 낯선 주식 시장에서
신규 투자자의 성과는 -1.2% 에 불과하다.

주식 입문자의 문제점

지나치게 잦은 매매

높은 수익률을 기대할 수 있는 저렴한 주식 선호

수익 상태인 주식을 빨리 매도해 이익실현

주식 매매 타이밍을 찾는데
어려움을 겪고 있음

국내 시장 현황

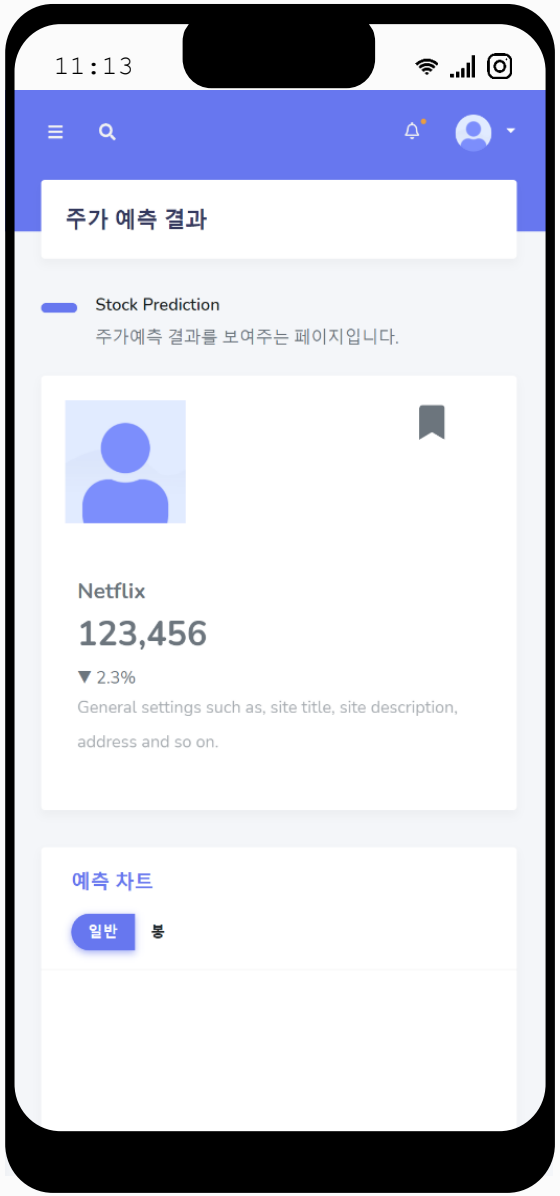
직접 국내 증권사 어플 13개 조사 및 UI/UX 분석

증권사에서 제공하는 많은 양의 정보 해석의 어려움

기존의 UI/UX를 벗어나 간소화된 어플이 트렌드

주식용어로 인한 어려움

사용하기 쉬운**UI/UX 환경 및 용어 설명 제공**



따라서 저희는 **개인맞춤형 주가 예측 서비스**를 기획했습니다.

① 개인 맞춤형 주가예측

사용자는 자신이 알고 있는 정보를 추가하여 주가 예측을 할 수 있다.

② 주식 매매 타이밍 알림

매도, 매수 타이밍에 어려움을 느끼는 주식 입문자에게 도움을 줄 수 있음

③ 기업별 SNS 분석

SNS 자연어 분석을 통해 기업에 대한 여론의 좋고 나쁨을 수치화

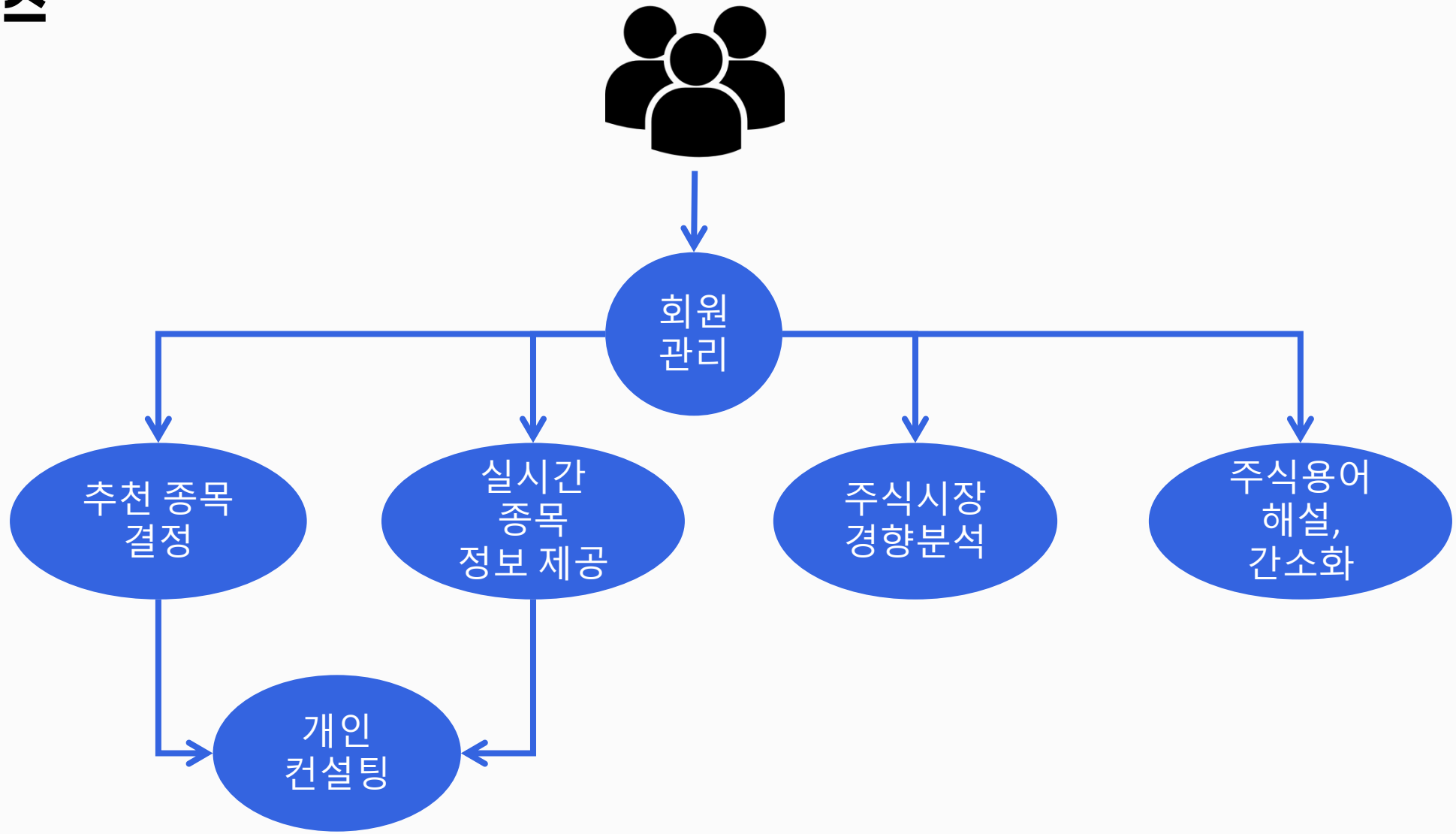
④ 사용자 000 환경제공

익숙하지 않은 주식용어 해설 및 낯선 기능에 대한 설명 제공

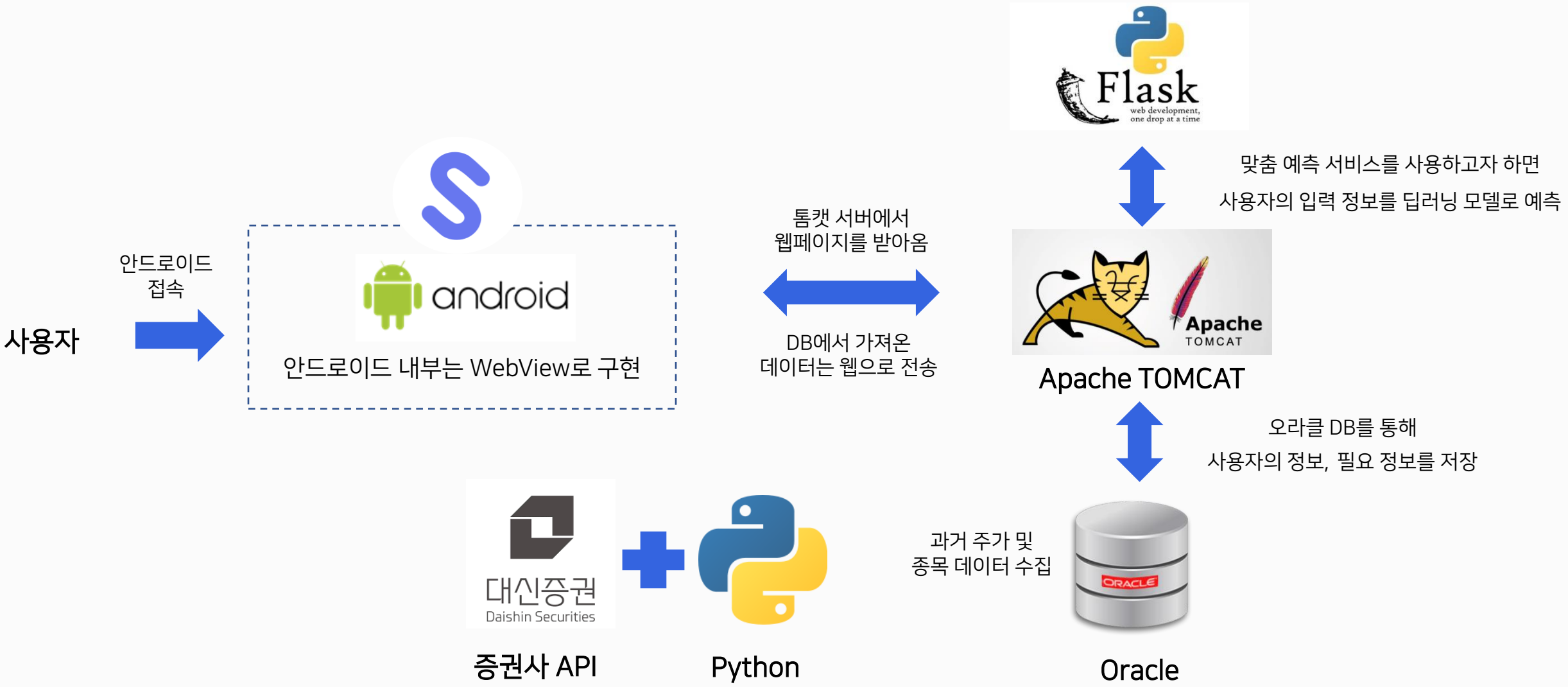
기능설계

02

유스케이스



서비스 흐름도


Flask
web development, one drop at a time
맞춤 예측 서비스를 사용하고자 하면
사용자의 입력 정보를 딥러닝 모델로 예측
톰캣 서버에서
웹페이지를 받아옴
DB에서 가져온
데이터는 웹으로 전송
Apache
TOMCAT
Apache TOMCAT
오라클 DB를 통해
사용자의 정보, 필요 정보를 저장
사용자
안드로이드
접속
android
안드로이드 내부는 WebView로 구현
대신증권
Daishin Securities
증권사 API
Python
과거 주가 및
종목 데이터 수집
ORACLE
Oracle

기능구현

03

SNS 자연어 분석을 통한 기업 점수화

데이터 수집

상장된 기업 목록 수집

대신증권 Daishin Securities

증권사 API

종목 명을 통해 SNS 검색

종목별 SNS 내용 수집

	종목	내용
0	동화약품	평택존놈 정오영\n@pt502\n·\nNov 24, 2020\n[부탁]\n오늘 식사...
1	동화약품	982_writer(이경태)\n@dqrudxo88\n·\nNov 24, 2020\n...
2	동화약품	21세기노비-ㅅ\n@80ksyo\n·\nNov 24, 2020\nSK·동화약품 등 ...
3	동화약품	평택존놈 정오영\n@pt502\n·\nNov 24, 2020\n[알림]\n이번 주까...
4	동화약품	부업아빠\n@kjahok\n·\nNov 24, 2020\n동화약품 임상 주가 전망 ...
...	...	...
9400	프레스티지바이오파마	더퍼스트경제TV\n@TheFirstEconomy\n·\nFeb 5\n#프레스티지바이...
9401	프레스티지바이오파마	더퍼스트경제TV\n@TheFirstEconomy\n·\nFeb 5\n[특징주]프레스...
9402	프레스티지바이오파마	김원준\n@kimwj1\n·\nJan 28\n유망 제약바이오 IPO, 신축년에도 쏟...
9403	프레스티지바이오파마	Lewis Lee\n@Onsdad\n·\nJan 27\n2021년 첫 공모주 투자 ...
9404	프레스티지바이오파마	서울경제신문\n@sedaily_com\n·\nJan 26\n프레스티지바이오파마 23...

160792 rows × 2 columns

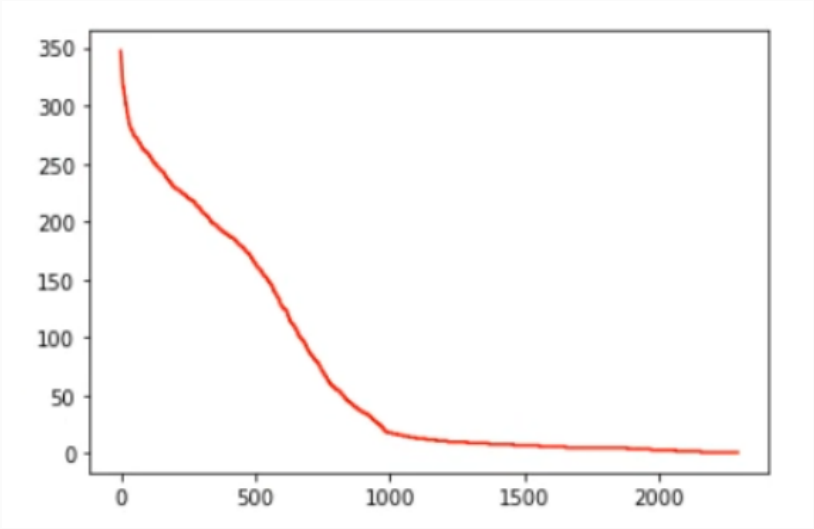
약 3,000개의 종목 내용 수집하여
16만개의 데이터 크롤링

SNS 자연어 분석을 통한 기업 점수화

데이터 전처리

1. 약 3000개의 기업 중 700개의 기업 선정

[기업별 트위터수]



트윗수가 약 100개 미만인 데이터는
신뢰도가 낮다고 판단하여 제거

2. 불용어 제거

정규 표현식 이란?

- 복잡한 문자열을 처리할 때 사용하는 기법
- re(regular expression) 모듈 사용

정규 표현식을 이용하여 트위터 내용의
외국어, 특수문자, 숫자 불용어 제거

SNS 자연어 분석을 통한 기업 점수화

데이터 분석

[Konlpy.Okt() 사용]

```
from konlpy.tag import Okt

Okt = Okt()

text = Okt.nouns("아 셀트리온은 위험기업일것 같아 투자를 잘못한거같은데... ")

text
```

```
['셀트리온', '위험', '기업', '투자']
```

- Okt : 트위터에서 만든 한국어 처리기 Tkt 기반
- 처리속도 및 본 데이터와 적합성을 고려하여 선택

[감성사전 구축]

```
num= 0
for i in range(len(data.iloc[:,3])):
    try:
        #okt.nouns로 추출한 단어 스플릿
        vv_word_list = data.iloc[i,3].split()
        score = 0
        num+=1
        if num%10000 == 0:
            print(num/10000,"만번째 실행중")

        if len(vv_word_list) !=0:
            for j in range(len(vv_word_list)):
                vv_word = vv_word_list[j]
                score_sum_df = scoredict[scoredict['단어'] == vv_word] ## df[조건식]
                temp= score_sum_df['라벨링']
                #단어마다 감성스코어를 확인후 점수합산
                if len(temp) !=0:
                    score += int(temp)
    except:

        #오류나는 원인 분석하니 트윗중 불용어를 제외하면
        #내용이없는 경우(7~8%정도) data['추출단어']가 빌값이라 오류났음
        #이런것들은 0점처리
        print(num,"번째에서 오류")
        num+=1
        score = 0
    data.iloc[i,4]= score
```

약 80,000개의 단어 중 **빈도수 100회 이상인 3,000개의 단어** 에 대해 감성수치 라벨링

SNS 자연어 분석을 통한 기업 점수화

아쉬운 점 & 어려웠던 점

① Facebook, Twitter, Instagram등 주요 SNS에서 데이터를 가져오기 위하여

공식API, 깃허브에 올라온 파이썬 라이브러리사용

BUT

SNS의 **AntiScrap정책** 으로 Selenium, BeautifulSoup을 활용한 제한적인 방법 밖에 활용 못함

② 영화평점이나 좋아요, 싫어요 등 긍정 부정이 나뉠만한 척도가 없어

현실적으로 **수십만 개의 트위터에 대한 라벨링 불가능**으로 인해 AI모델학습을 하지 못한 아쉬움이 있다.

KMV 모형 을 통해 부도확률 예측하기

머튼의 옵션가격 모형을 이용하여 KMV모형 생성

① 자기자본가치의 변동성 σE 구하기

$$V_E = V_A N(d_1) - V_D e^{-r_f T} N(d_2)$$

$$d_1 = \frac{\ln(V_A/X) + (r_f + \sigma_A^2/2)T}{\sigma_A \sqrt{T}}$$

$$d_2 = d_1 - \sigma_A \sqrt{T}$$

```
def d1(A, sigmaA) -> float:
    return (math.log(A/B)+(r+sigmaA ** 2/2)*T)/(sigmaA * math.sqrt(T))

def d2(A, sigmaA) -> float:
    return d1(A, sigmaA) - sigmaA * math.sqrt(T)

def Ve(A, sigmaA) -> float:
    return A * stats.norm.cdf(d1(A, sigmaA)) - np.exp(-r * T) * B* stats.norm.cdf

def sigmaE2(A, sigmaA) -> float:
    return (A/E) * stats.norm.cdf(d1(A, sigmaA)) * sigmaA
```

② 자산의 시장가치, 자산가치의 변동성

```
error1 = sigmaE
error2 = E
A = 0
sigmaA = 0
breakr = False
E = int(E)
for A1 in tqdm(range(int(E - (E/2)), E + E)):
    for sigmaA1 in range(1, 100):
        sigmaA1 = sigmaA1 / 100
        if abs(sigmaE - sigmaE2(A1, sigmaA1)) < error1:
            if  abs(E-Ve(A1, sigmaA1)) < error2:
                error1 = abs(sigmaE - sigmaE2(A1, sigmaA1)) # 오차
                error2 = abs(E-Ve(A1, sigmaA1)) # 오차
                A = A1
                sigmaA = sigmaA1
                if error1 < 0.001 and error2 < 1:
                    breakr = True
                    break
    if breakr == True:
        break
```

주식회사 세동
SAE DONG CO., LTD.

```
1 print(A, sigmaA)
```

```
260 0.39
```

자산의 시장가치 : 260
자산가치의 변동성 : 0.39

KMV 모형 을 통해 부도확률 예측하기

머튼의 옵션가격 모형을 이용하여 KMV모형 생성

③ 부도거리

$$DD = \frac{\ln(\frac{V_A}{DP}) + (\mu - \sigma_A^2)T}{\sigma_A\sqrt{T}} \quad (4)$$

σ_A : Variability of corporate asset values
μ : Growth rate of return on asset
T : Liability redemption period
DD : Distance to Default
DP : Default Point
V_A : Corporate asset value

```
DP = B + 0.5*LTD # 부도점
# 2020년도 명목 GDP: 1,933.2조원 / 2019년도 명목 GDP:1924.5조원
# 명목 경제성장률 = (2020년도 명목 GDP - 2019년도 명목 GDP)/2020년도 명목 GDP * 100
mu = ((1933.2-1924.5)/1933.2) * 100 # 자산의 기대수익률
DD = (math.log(A/DP) + (mu - (sigmaA**2)/2)*T) / (sigmaA*math.sqrt(T)) # 부도거리
DD
```

④ 표준정규분포를 이용하여 예상 부도 확률

$$EDF = N(-DD)$$

```
1 # 예상 부도확률
2 stats.norm.cdf(-DD)
```

```
0.9586117888119846
```

세동 예상 부도 확률 : 0.9586

×
시연영상

04

기대효과 및 향후발전

05

Stock CM의 기대효과

01 주식 초보자에게 손실의 위험성을 낮춰준다.

딥러닝 기반 주가 예측 서비스로 주식 초보자의 무분별한 주식 투자로 인한 경제적 손실의 위험성을 낮춰준다.

02 간편한 UI로 주식에 쉽게 접근할 수 있다.

주식 초보자들도 이해하기 쉽게 제작된 UI로 주식 투자의 접근성을 높였다.

03 주가예측 종목추천으로 사용자의 수익성 상승

딥러닝 기반 주가 예측으로 종목을 추천하고 주식 초보자들에게 좀 더 나은 수익을 도울 수 있다.

Stock CM의 향후발전

01 사용자의 투자 성향파악으로 상세한 서비스 제공

사용자의 투자 성향 파악으로 좀 더 상세한 개인 맞춤 주식 커스터마이징 서비스를 제공한다.

02 주식에 영향을 미치는 환경요소 확장으로 성능 강화

주식에 영향을 미치는 많은 환경요소를 현재에 그치지 않고 더 확장하여 주가예측의 성능을 강화시킨다.

03 자동 매매 서비스

위의 모든 기능들이 충족된다면 Stock CM 어플 내에서 자동으로 매매할 수 있는 서비스를 구현한다.

정선일

PEE DO RI

API 데이터 수집
딥러닝 모델 설계
DB 제어

정혜선

PEE DO RI

재무 재표 크롤링
퀀트 포트폴리오 작성
웹 기능 구현

윤정은

식품영양학과

UI / UX 작성
웹 디자인
프레젠테이션 제작

강다현

정보보호학과

UI / UX 작성
웹 디자인
프레젠테이션 제작

임대진

PEE DO RI

웹 서버 구축
DB 구축
웹 기능 구현

서민규

PEE DO RI

SNS 데이터 크롤링
SNS 데이터 분석
웹 기능구현

참고 문헌

표지 일러스트 : https://kr.freepik.com/vectors/people

THANK YOU