

KB증권
디지털 Idea Market 공모전

TEAM 등 차 수 열 공 민 석 민 복 기 이 윤 주

Content



M - Lucky Betting



알고리즘



DB



M - Lucky Betting

'M-Lucky Betting'을 소개해드립니다.

01

01. 주제 선정 배경

02. 소개

03. 차별점

04. UI



서비스 기획 배경

MZ

2020년 상반기를 기점으로 증권 투자에 참여하는
MZ세대(20,30대)의 비중이 폭발적으로 증가했습니다.

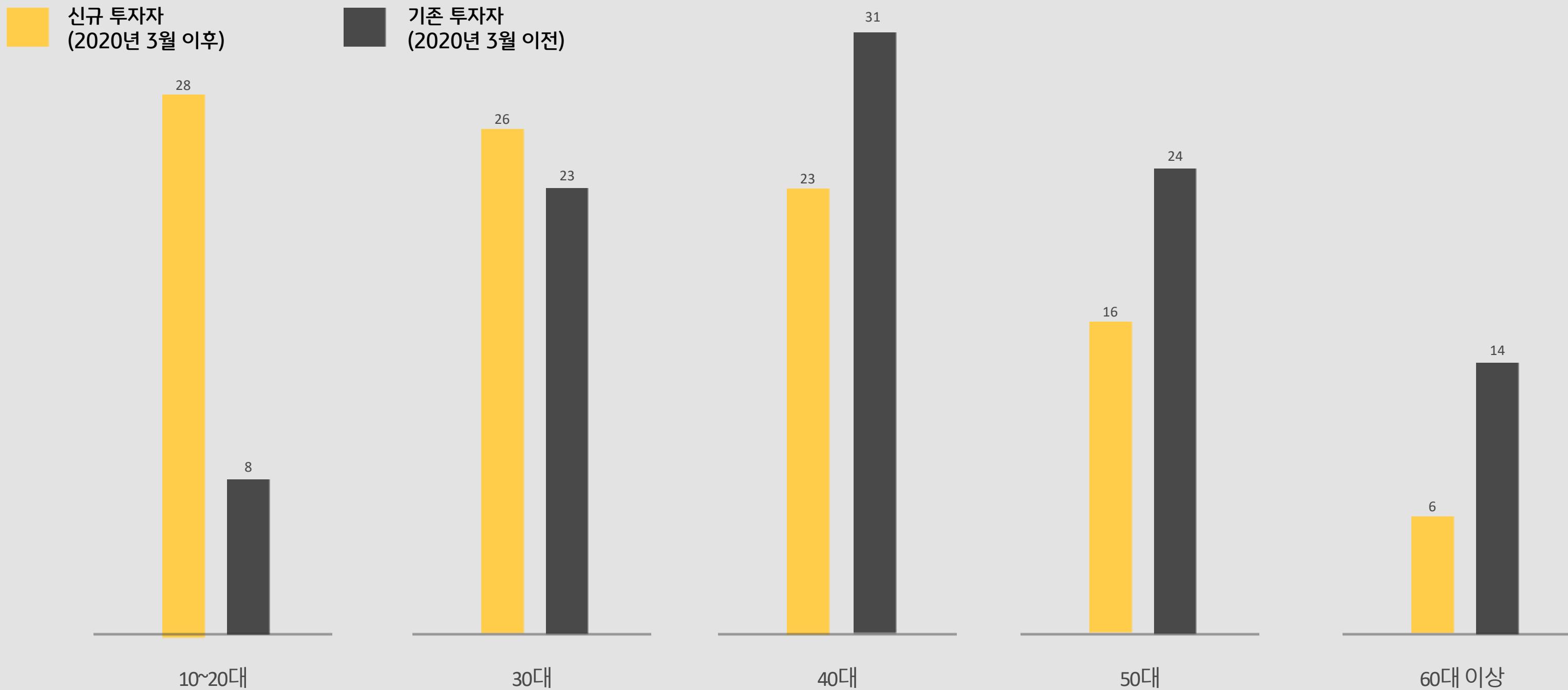
MZ세대는 향후 수십년 동안 경제활동의 주축으로 성장하며
주 고객층으로 자리잡게 될 것입니다.

따라서 저희는 MZ세대를 주 타겟층으로
염두에 두고 'M-able 미니'에 추가할 컨텐츠를
구상하였습니다.

MZ세대의 증권 활동에 대한 참여와 흥미를
이끌어 낼 수 있도록 하며

AI를 활용해 주식 관련 정보를 제공 할 수 있는
'M-Lucky Betting(MLB)'을 제안합니다.

주식투자 연령대별 비중(단위:%)



10 ~ 30대에서 뚜렷한 증가세 확인 가능

특징

M-LB는 라이브 방송, 커피 한 잔 아껴 주식하기 등의 트랜디한 서비스를 제공하는 M-able mini의 특색을 살릴 수 있는 참신한 주식 추천 서비스입니다.

차별점

- 01 운세 컨셉으로 종목 랭킹 및 배팅을 제공하는 M-LB는 미디어 중심의 M-able Mini 어플에 최적화된 서비스입니다
- 02 기존 영상 위주의 컨텐츠에 더불어 고정된 형식의 컨텐츠를 제공함으로써 언제나 정보를 확인할 수 있습니다

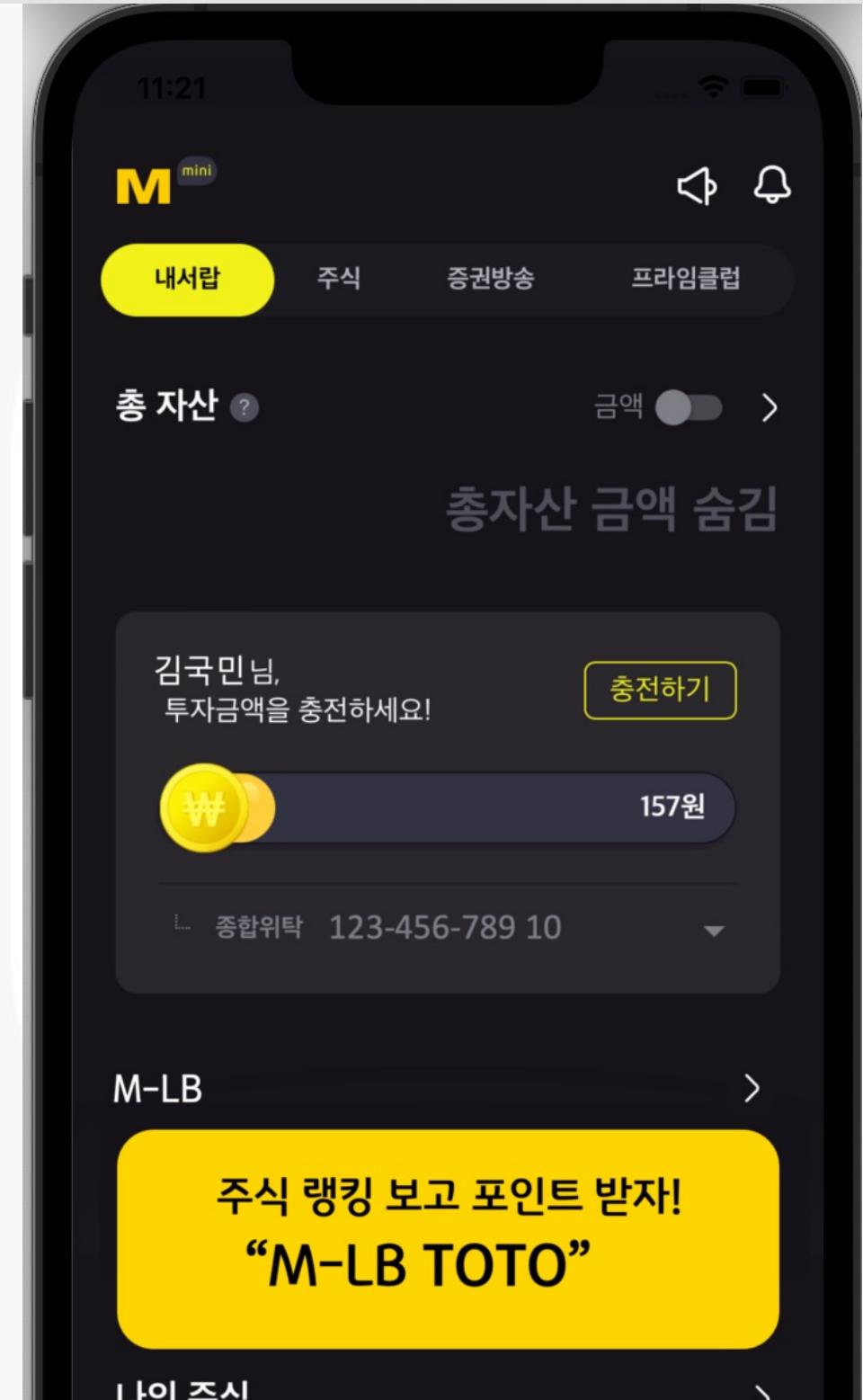
발전 가능성

- 03 사용자 참여형 서비스로 MZ세대의 유입 증가를 기대할 수 있습니다
- 04 무료 AI 종목 추천을 통해 프라임클럽 가입의 유인이 될 수 있습니다

M-Lucky Betting (M-LB)

- 오늘의 장 운세
- 종목 등락 AI 예측 서비스
- M-LB TOTO

M
mini



오늘의 장운세

10개 종목의 등락 여부에 대한 예측값을 통해

하루 장에 대한 전반적인 전망을

재치있는 문구를 통해 사용자에게 제공합니다.

전반적인 전망과 함께

매일 추천/비추천 주식을

행운/액운 주식으로 추천합니다.

11:15
오늘 예상되는 국내 증시는...

맑음! 너는 다~ 투자 계획이 있구나

오늘의 행운 주식은...

"현대차"

오늘의 액운 주식은...

"안랩"

오늘의 Hot 키워드 랭킹🔥

순위	종목	순위	종목
1	현대차	6	쌍용차
2	한화	7	배럴
3	베셀	8	신흥
4	카카오	9	덴티움
5	두산	10	안랩

M-LB TOTO하기

누적 정확도

Go back

종목 등락 AI 예측 서비스

네이버 증권 뉴스 기사 데이터 분석을 통해

오늘의 운세 형식으로 화제가 되는 종목을 추천하고,

하루 단위 상위 10개 종목에 대해

AI가 예측한 등락 정보를 제공합니다.

11:15
오늘 예상되는 국내 증시는..

맑음! 너는 다~ 투자 계획이 있구나

오늘의 행운 주식은...

"현대차"

오늘의 액운 주식은...

"안랩"

오늘의 Hot 키워드 랭킹 🔥

순위	종목	순위	종목
1	현대차	6	쌍용차
2	한화	7	배럴
3	베셀	8	신흥
4	카카오	9	덴티움
5	두산	10	안랩

M-LB TOTO하기

누저 전화도

M-LB TOTO

사용자가 오늘의 상위 10개 종목에 대해

[Up, down, stay]의 선택을 합니다.

이를 'M-LB'가 예측한 등락 여부와 비교하고

정답률이 우수한 사용자에게 혜택을 제공해

사용자의 참여를 유도할 수 있습니다.

M-LB TOTO

이곳에 전화번호 8자리를 입력해주세요. 입력

당신의 예측을 보여주세요

순위	종목	▲▼↔	순위	종목	▲▼↔
1	현대차	ON!	6	쌍용차	ON!
2	한화	ON!	7	배럴	ON!
3	베셀	ON!	8	신흥	ON!
4	카카오	ON!	9	덴티움	ON!
5	두산	ON!	10	안랩	ON!

제출하기 (10개 모두 선택해주세요!)

어제의 TOTO 결과 확인하기

어제의 TOTO 랭킹 채점 결과

순위	종목	유저	서비스	순위	종목	유저	서비스
1	현대차	✗	✗	6	쌍용차	✗	○
2	한화	✗	○	7	배럴	○	○
3	베셀	✗	○	8	신흥	✗	○



알고리즘

서비스를 구성하는 알고리즘에 대해 소개해드리겠습니다

02

01. 이슈 키워드 추출

02. Word2Vec

03. 언어모델



01. 이슈 키워드 추출

- 토큰화(Tokenizing)
 - 키워드 추출



토큰화(Tokenizing)

- 토큰화란?
 - 텍스트 데이터에 대한 정보를 단위(Token)별(형태소 별, 문장 별 등)로 나누는 작업
- KoNLPy 모듈에서 제공하는 Okt를 사용해서 뉴스 기사를 토큰화

토큰화 예시

토큰화 전	토큰화 후
뉴욕증시, 서방 제재 우려 속에 상승 출발	['뉴욕증시', '서방', '제재', '우려', '속', '상승', '출발']
BofA, 로지텍 49% 추가 상승 가능... "게임 확장 수혜"	['로지텍', '추가', '상승', '가능', '게임', '확장', '수혜']

- KoNLPy : 대표적인 한글 자연어 처리 라이브러리.
- Okt(Open Korean Text) :
KoNLPy에서 제공하는 형태소 분석기들 중 대표적인 하나.

키워드 추출

1. 토큰화 된 하루치 기사 제목 데이터에서 명사만을 추출

2. 불용어 처리

- 직접 구축한 불용어 사전을 기반으로
불용어 처리 작업(stopword_dictionary.xlsx)
- EX) [증권, 투자, 거래, 주가] 등 금융 텍스트에 자주 등장하는 단어
+) 의미를 알 수 없는 단어들

3. 단어의 등장 빈도수를 기준으로 상위 10개 키워드 추출

키워드 추출 결과
(2022.03.24)

키워드	빈도수
안랩	39
삼성전자	37
블록딜	32
삼성	22
두산	20
박용만	17
최현만	16

02. Word2Vec (Word – To – Vector)



Word2Vec

(Word-to-Vector)

Word2Vec 결과
(2022.03.24)

- Word2Vec이란?
 - '단어의 벡터화'라는 의미로 자연어 데이터에 대한 분석기법
 - 단어를 벡터(수치)화 함으로서 단어에 대한 연산이 가능해진다.
- Word2Vec모델
 - 약 한달치의 네이버 증권 기사 데이터를 모델에 학습시켜 사용
 - word2vec.model
- 추출된 키워드를 모델의 입력값으로 넣어
키워드와 가장 연관성이 높은 10가지 종목명을 추출

키워드	종목명
안랩	안랩
삼성전자	삼성전자
블록딜	APS홀딩스
삼성	한화
두산	두산
박용만	오리콤
최현만	한양증권

03. 언어 모델





Korean BERT (Bidirectional Encoder Representations from Transformers)

SKT에서 만든 한국어 버전의 BERT 모델

기존 BERT의 한국어 성능 한계를 극복하기 위해 개발

한국어의 불규칙한 언어 변화의 특성을 반영하기 위해 데이터 기반 토큰화(Tokenization) 기법을 적용하여 기존 대비 27%의 토큰만으로 2.6% 이상의 성능 향상을 이끌어냄



Fine-Tuning

네이버 금융 뉴스 탭에서 실시간 속보 제목을 52000개 크롤링

크롤링한 기사 제목을 라벨링

국내증시 **좋음은 2, 보통 또는 중립은 1, 나쁨은 0**으로 라벨링

라벨링한 데이터를 KoBERT 모델에 학습시켜
fine-tuning하였습니다

주식 종목 추천에 필요한 하루치 크롤링 데이터를 라벨링하는
알고리즘에 이렇게 fine-tuning한 KoBERT 모델을 활용하였습니다



Test Accuracy

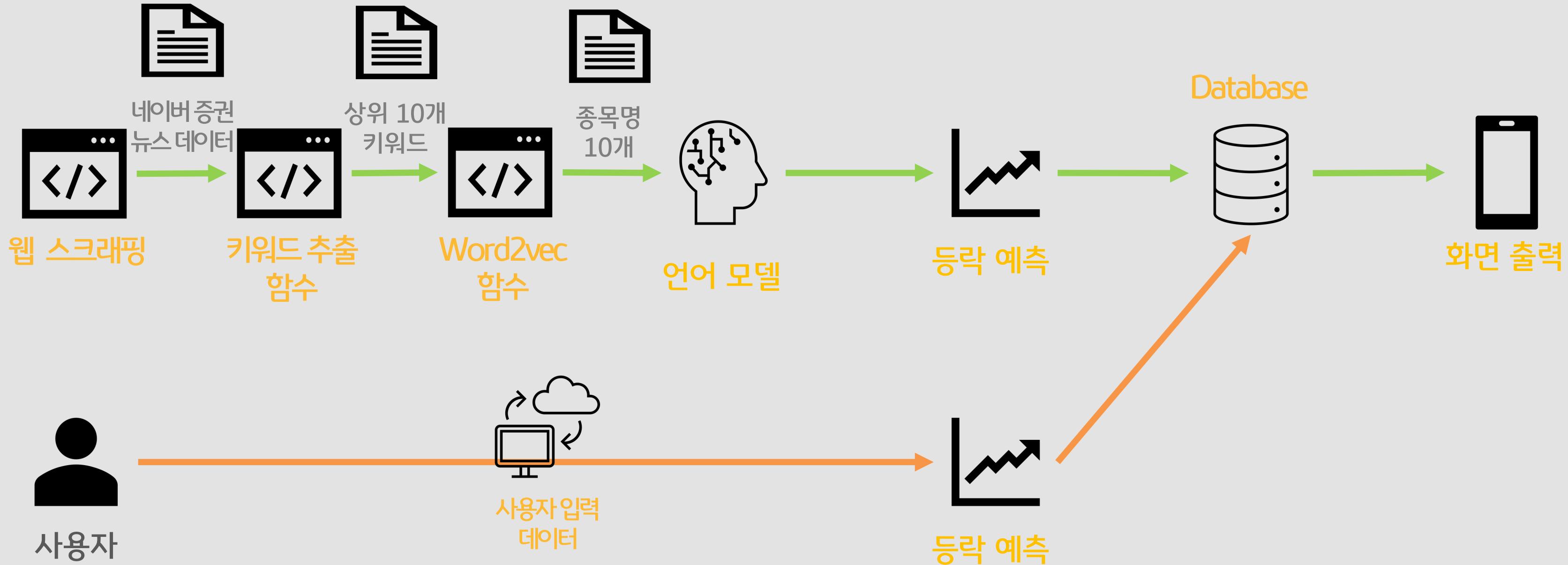
Test 데이터로 성능을 측정한 결과 약 85%의 정확도를 보였습니다

이 때, 학습 및 성능 측정에 사용된 파라미터는 다음과 같습니다

```
max_len = 200  
batch_size = 16  
warmup_ratio = 0.1  
num_epochs = 5  
max_grad_norm = 1  
log_interval = 200  
learning_rate = 5e-5
```



알고리즘 도식화



03

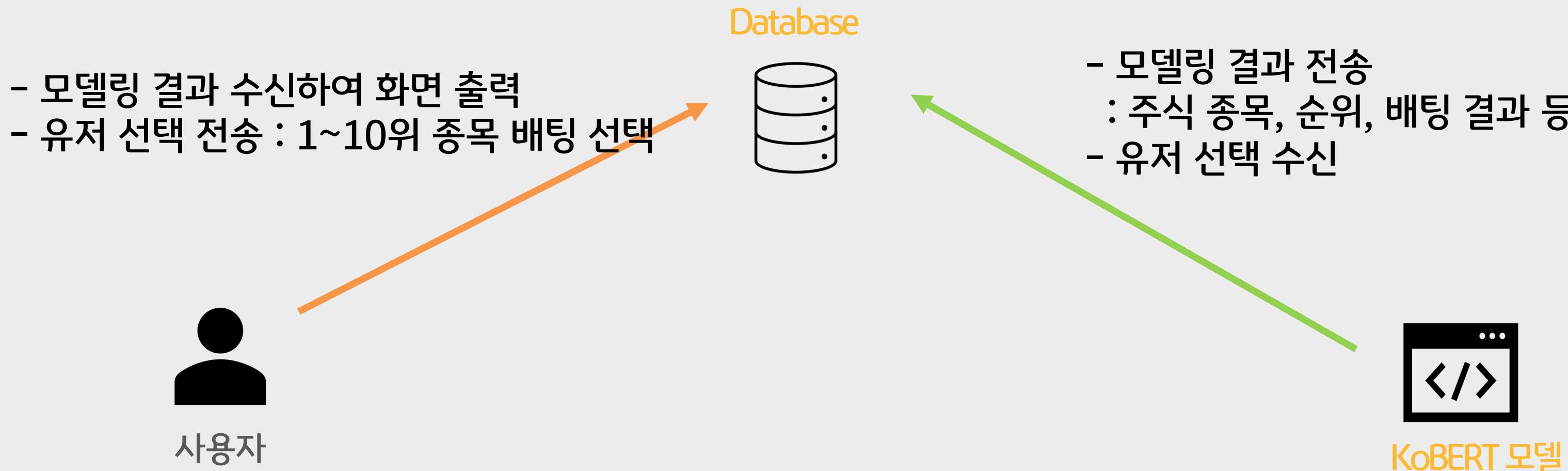
DB

사용자 참여 어플 개발을 위해 사용한
실시간 데이터베이스입니다
모델과 사용자의 소통을 원활히 할 수 있습니다.



데이터베이스

- **FireBase**
 - 구글이 제공하는 백엔드 (사용자 화면과 서비스를 이어주는 분야) 구현 서비스
- **FireBase의 Real Time DataBase**
 - FireBase의 Real Time Data Base를 통해 사용자-어플 입출력 실시간 소통



Reference

<https://www.koreascience.or.kr/article/CFKO201835372170725.pdf>

「공공이슈 추출을 위한 뉴스 빅데이터 분석 시스템」(2018) - 동국대학교 컴퓨터 공학과 [김승주 외 6인]

<https://www.koreascience.or.kr/article/CFKO202025036019264.pdf>

「BERT를 활용한 문장 감성 분석 연구」(2020.11) - 성균관대학교 소프트웨어대학 [이한범 외 2인]

<https://github.com/SKTBrain/KoBERT>

「KoBERT 깃허브」

<https://s3.us-west-2.amazonaws.com/secure.notion-static.com/4f4b613e-c395-400c-b51f-d680b81dc566/>

「경제 텍스트 데이터를 활용한 키워드 분석방안 연구」 - [김한준, 조새롬, 김동찬]

<https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE09348126>

「BERT를 활용한 뉴스 감성분석과 거시경제지표 조합을 이용한 주가지수 예측」 - (2020) [장은아, 최회련, 이홍철]

<https://www.koreascience.or.kr/article/CFKO201725864428451.pdf>

「딥러닝 기반의 뉴스 분석을 활용한 주제별 최신 연관단어 추출 기법」 - (2017) [김성진, 김건우, 이동호]

마치며



GitHub URL : https://github.com/shirtnjean/2022_kbSecurities_DIM

YouTube : https://youtu.be/qAXV0Vwf_04

어플 다운로드 : https://drive.google.com/file/d/1j6HDiSUKqE8Pg8jdlCCbs5_qNugzEaqb/view?usp=sharing

THANK YOU

End of Document

