서울과학기술대학교 SEOUL NATIONAL UNIVERSITY OF SCIENCE & TECHNOLOGY

Kdata 한국데이터산업진흥원

약관 상세 분석 시스템

약간 신경 쓴 약관

꼼꼼히 읽어조 | 서민정 박지원 서태구 신승민 신영우 전종훈

CONTENTS

CHAPTER 1

주제

CHAPTER 2

기능

CHAPTER 3

결론

CHAPTER 1

주제

01. 주제 선정 배경

02. 주제 선정 목적

암호문 같은 약관

보험사 사장도 “암호문 같아”…‘깨알약관’ 언제쯤 쉬워질까

전문용어 빽빽…국민 90% “불편해”
매년 약관 이해도 평가하지만 효과 의문
정권 바뀌자 금융당국도 ‘조용’
“약관 개선 노력에 정부 적극 개입해야”

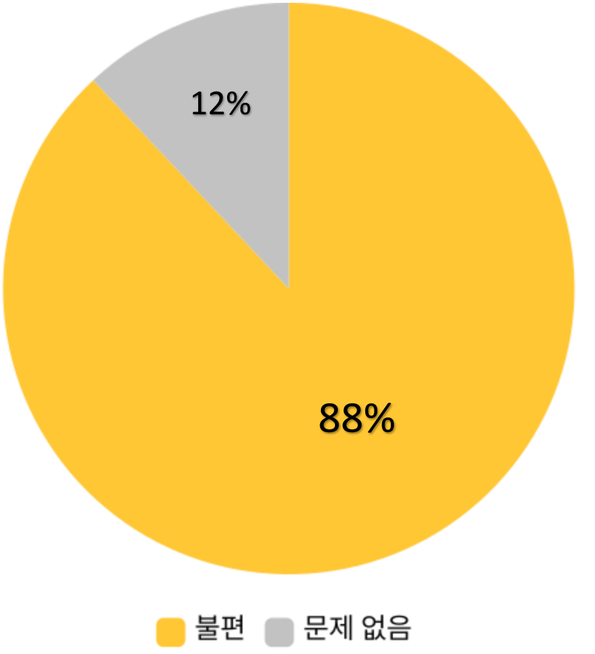
기사승인 2023-04-23 06:00:02

“보험사 사장을 지낸 저도 솔직히 **보험약관을 끝까지 읽어보지 못했다.** 위해 수 십개의 특약을 붙이고 내용을 다 담**약관은 두꺼운 암호문**” 것 아닌가하는 생각이 든다” (최종구 전 금융위원장, 2019년 2월26일 간담

‘숨어 있는 약관’ 때문에 피해를 본다

글자 크기가 10포인트 미만으로 작거나 장평·자간이 좁**가독성이 떨어지는 문제**트 미만의 글씨가 들어간 보험사도 있었다. 페이지, 목차 누락으로 전체 내용을 파악 한 밑줄과 하이라이트를 할 필요가 있다는 개선의견이 나왔다.

중요부분을 강조



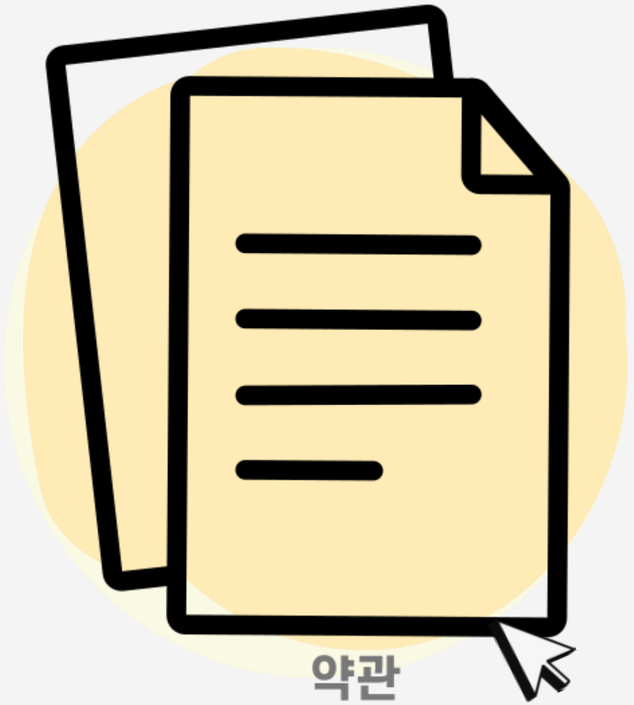
약관 및 상품 설명서 국민 인식조사 (2019)

약간 신경쓴 약관

Who
약관의 이해를 원하는
서비스 이용자

When
약관을 읽을 때

Where
웹에서



What
이해가 필요한 약관

How
조항별 유/불리 판단
키워드 분석

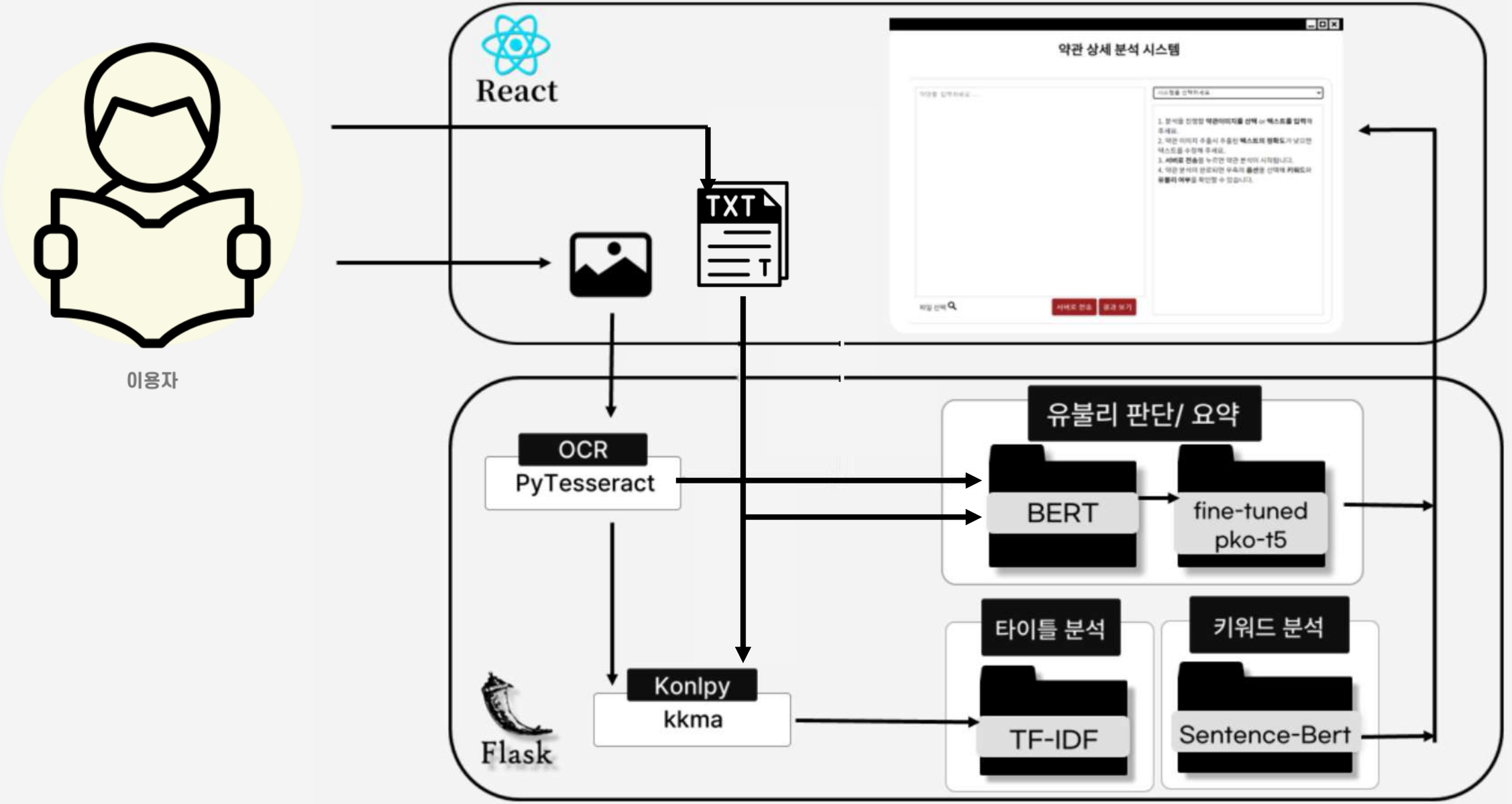
Why
약관의 명확한 이해도모 하고
정보손실을 최소화 하기 위해

CHAPTER 2

기능

01. 시스템 구조도
02. 웹 기능 방식
03. 기능별 모델
04. 웹 구현 결과

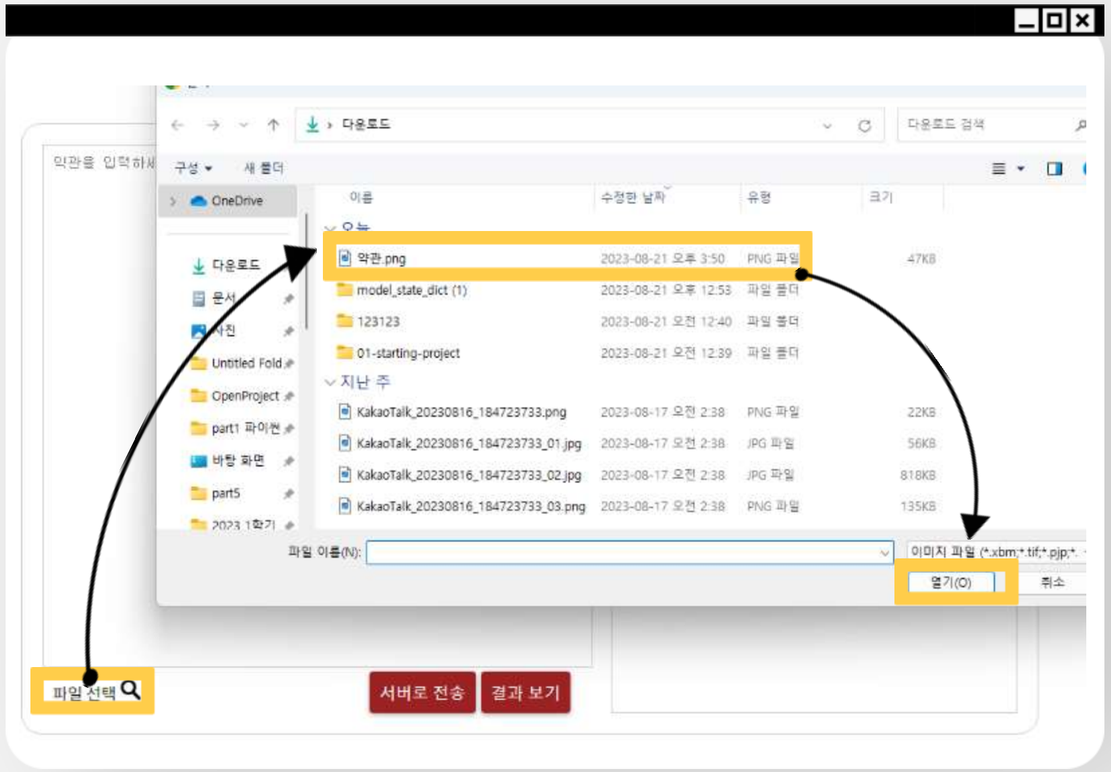
시스템 구조도

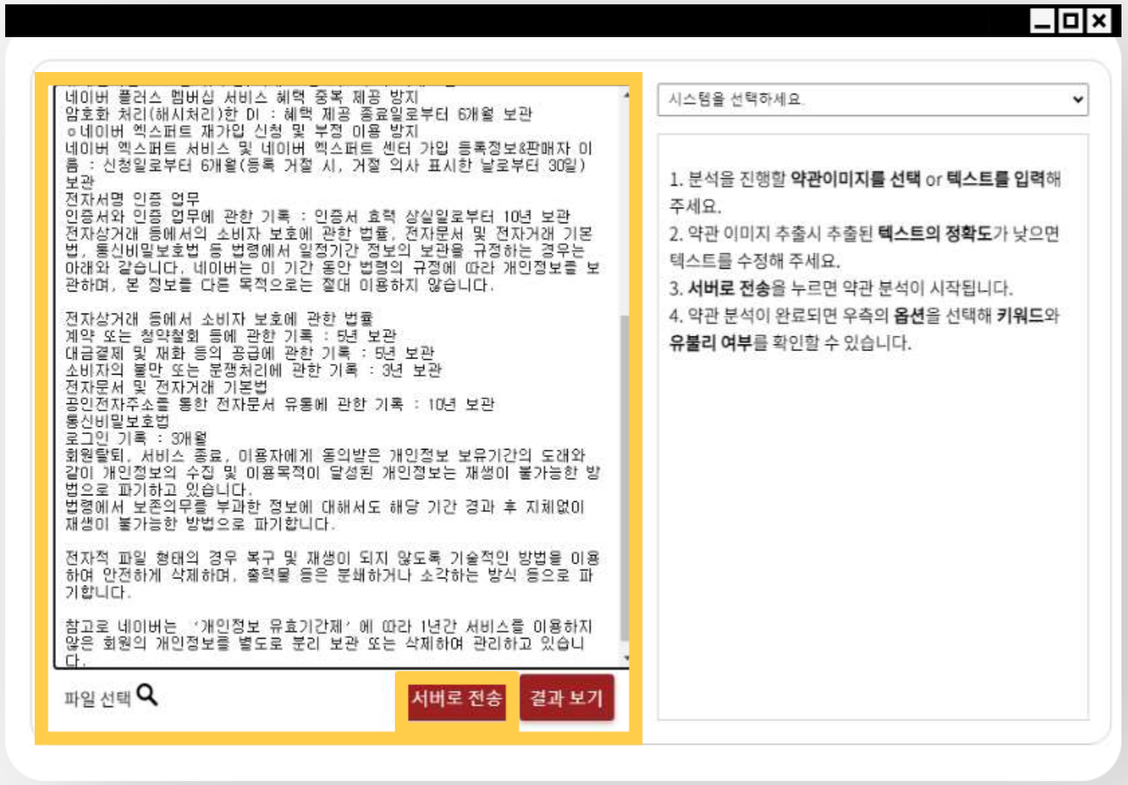

이용자
React
TXT
약관 상세 분석 시스템
OCR
PyTesseract
Konlpy
kkma
Flask
유불리 판단/ 요약
BERT
fine-tuned
pko-t5
타이틀 분석
TF-IDF
키워드 분석
Sentence-Bert

약관 상세 분석 시스템

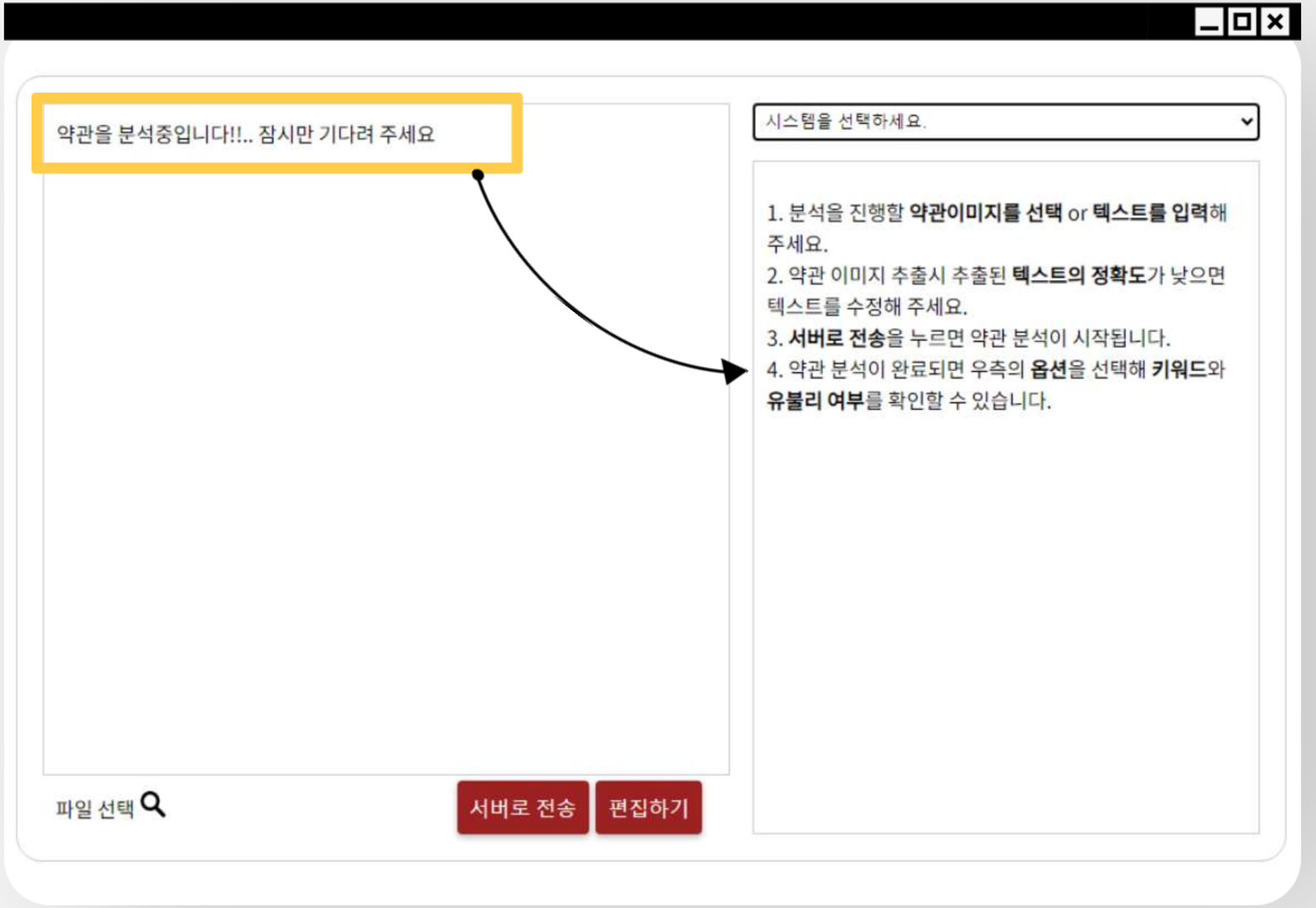

AI 약관 상세 분석 시스템
약관을 입력하세요...
시스템을 선택하세요.
1. 분석을 진행할 약관이미지를 선택 or 텍스트를 입력해 주세요.
2. 약관 이미지 추출시 추출된 텍스트의 정확도가 낮으면 텍스트를 수정해 주세요.
3. 서버로 전송을 누르면 약관 분석이 시작됩니다.
4. 약관 분석이 완료되면 우측의 옵션을 선택해 키워드와 유불리 여부를 확인할 수 있습니다.
파일 선택
서버로 전송
결과 보기

웹 기능 방식

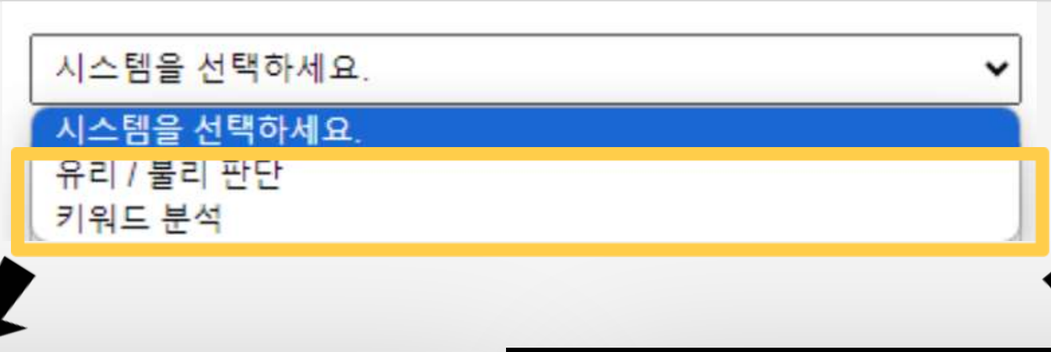

다운로드
다운로드 검색
구성
새 폴더
OneDrive
다운로드
문서
사진
Untitled Fold
OpenProject
part1 파이썬
바탕 화면
part5
이름
수정한 날짜
유형
크기
약관.png 2023-08-21 오후 3:50 PNG 파일 47KB
model_state_dict (1) 2023-08-21 오후 12:53 파일 폴더
123123 2023-08-21 오전 12:40 파일 폴더
01-starting-project 2023-08-21 오전 12:39 파일 폴더
지난 주
KakaoTalk_20230816_184723733.png 2023-08-17 오전 2:38 PNG 파일 22KB
KakaoTalk_20230816_184723733_01.jpg 2023-08-17 오전 2:38 JPG 파일 56KB
KakaoTalk_20230816_184723733_02.jpg 2023-08-17 오전 2:38 JPG 파일 818KB
KakaoTalk_20230816_184723733_03.png 2023-08-17 오전 2:38 PNG 파일 135KB
파일 이름(N):
이미지 파일 (*.xbm;*.tif;*.pjp;*.
열기(O)
취소
파일 선택
서버로 전송
결과 보기


네이버 플러스 멤버십 서비스 혜택 중복 제공 방지
암호화 처리(해시처리)한 DI : 혜택 제공 종료일로부터 6개월 보관
○ 네이버 엑스퍼트 재가입 신청 및 부정 이용 방지
네이버 엑스퍼트 서비스 및 네이버 엑스퍼트 센터 가입 등록정보&판매자 이름 : 신청일로부터 6개월(등록 거절 시, 거절 의사 표시한 날로부터 30일) 보관
전자서명 인증 업무
인증서와 인증 업무에 관한 기록 : 인증서 효력 상실일로부터 10년 보관
전자상거래 등에서의 소비자 보호에 관한 법률, 전자문서 및 전자거래 기본법, 통신비밀보호법 등 법령에서 일정기간 정보의 보관을 규정하는 경우는 아래와 같습니다. 네이버는 이 기간 동안 법령의 규정에 따라 개인정보를 보관하며, 본 정보를 다른 목적으로는 절대 이용하지 않습니다.
전자상거래 등에서 소비자 보호에 관한 법률
계약 또는 청약철회 등에 관한 기록 : 5년 보관
대금결제 및 재화 등의 공급에 관한 기록 : 5년 보관
소비자의 불만 또는 분쟁처리에 관한 기록 : 3년 보관
전자문서 및 전자거래 기본법
공인전자주소를 통한 전자문서 유통에 관한 기록 : 10년 보관
통신비밀보호법
로그인 기록 : 3개월
회원탈퇴, 서비스 종료, 이용자에게 동의받은 개인정보 보유기간의 도래와 같이 개인정보의 수집 및 이용목적이 달성된 개인정보는 재생이 불가능한 방법으로 파기하고 있습니다.
법령에서 보존의무를 부과한 정보에 대해서도 해당 기간 경과 후 지체없이 재생이 불가능한 방법으로 파기합니다.
전자적 파일 형태의 경우 복구 및 재생이 되지 않도록 기술적인 방법을 이용하여 안전하게 삭제하며, 출력물 등은 분쇄하거나 소각하는 방식 등으로 파기합니다.
참고로 네이버는 '개인정보 유효기간제' 에 따라 1년간 서비스를 이용하지 않은 회원의 개인정보를 별도로 분리 보관 또는 삭제하여 관리하고 있습니다.
파일 선택
서버로 전송
결과 보기
시스템을 선택하세요.
1. 분석을 진행할 약관이미지를 선택 or 텍스트를 입력해 주세요.
2. 약관 이미지 추출시 추출된 텍스트의 정확도가 낮으면 텍스트를 수정해 주세요.
3. 서버로 전송을 누르면 약관 분석이 시작됩니다.
4. 약관 분석이 완료되면 우측의 옵션을 선택해 키워드와 유불리 여부를 확인할 수 있습니다.

웹 기능 방식


약관을 분석중입니다!!.. 잠시만 기다려 주세요
시스템을 선택하세요.
1. 분석을 진행할 약관이미지를 선택 or 텍스트를 입력해 주세요.
2. 약관 이미지 추출시 추출된 텍스트의 정확도가 낮으면 텍스트를 수정해 주세요.
3. 서버로 전송을 누르면 약관 분석이 시작됩니다.
4. 약관 분석이 완료되면 우측의 옵션을 선택해 키워드와 유불리 여부를 확인할 수 있습니다.
파일 선택
서버로 전송
편집하기

웹 기능 방식



AI 약관 상세 분석 시스템 : 유리 / 불리 판단

회사는 이 보험의 기본계약 만기시점의 계약자적립액을 만기환급금으로 지급하며, 납입보험료 중 적립부분 순보험료(적립보험료에서 계약체결비용 및 계약관리비용을 제외한 금액)를 기준으로 공시이율을 적용한 금액으로, 향후 공시이율의 변경, 계약내용의 변경, 보험료 실제 납입일자, 중도인출 여부 등에 따라 달라질 수 있습니다.

적립되는 보험료 없이 보장담보만으로 가입하시는 경우 보험계약 만기시 지급받는 금액(만기환급금)이 없습니다.

예금자보호 이 보험은 예금자보호법에 따라 예금보험공사가 보호합니다.

예금자보호 한도는 본 보험회사에 있는 귀하의 모든 예금보호 대상금융상품의 해약환급금(또는 만기시 보험금이나 사고보험금)과 기타 지급금을 합하여 1인당 "최고 5천만원"이며, 5천만원을 초과하는 나머지 금액은 보호하지 않습니다.

지정대리청구 보험사고(치매 등) 발생으로 본인 스스로 보험금 청구가 현실적으로 어려운 상황이 발생할 경우 보험금을 대신 청구하는 자(보험금 지정 대리청구인)를 보험가입초기 또는 유지 중에 미리 지정하는 제도입니다.

※ 계약자가 자신을 위한 치매보험 가입하고 치매가 발생한 경우 계약자가 보험금을 직접 청구할 수 없어 보험금 청구가 곤란

유리 / 불리 판단

유리한 조건

- 예금보험공사가 보호하는 이 보험은 예금자보호법에 따라 예금보험공사가 보호한다.
- 지정대리청구는 보험사고 발생으로 본인 스스로 보험금 청구가 어려운 상황이 발생할 경우 보험금을 대신 청구하는 자를 보험가입초기 또는 유지 중에 미리 지정하는 제도이다.
- 이런 경우에 대비하여 보험금 대리청구인을

불리한 조건

- 예금자보호 한도는 본 보험회사에 있는 귀하의 모든 예금보호 대상금융상품의 해약환급금과 기타 지급금을 합하여 1인당 최고 5천만원이며 5천만원을 초과하는 나머지 금액은 보호하지 않습니다.

파일 선택 서버로 전송 편집하기

AI 약관 상세 분석 시스템 : 키워드 분석

회사는 이 보험의 기본계약 만기시점의 계약자적립액을 **만기환급금**으로 지급하며, **납입보험료** 중 **적립부분** 순보험료(적립보험료에서 계약체결비용 및 계약관리비용을 제외한 금액)를 기준으로 공시이율을 적용한 금액으로, 향후 공시이율의 변경, 계약내용의 변경, 보험료 실제 납입일자, 중도인출 여부 등에 따라 달라질 수 있습니다.

적립되는 보험료 없이 보장담보만으로 가입하시는 경우 보험계약 만기시 지급받는 금액(**만기환급금**)이 없습니다.

예금자보호 이 보험은 예금자보호법에 따라 예금보험공사가 보호합니다.

예금자보호 한도는 본 보험회사에 있는 귀하의 모든 예금보호 **대상금융상품**의 해약환급금(또는 만기시 보험금이나 사고보험금)과 기타 지급금을 합하여 1인당 "최고 5천만원"이며, 5천만원을 초과하는 나머지 금액은 보호하지 않습니다.

지정대리청구 보험사고(치매 등) 발생으로 본인 스스로 보험금 청구가 현실적으로 어려운 상황이 발생할 경우 보험금을 대신 청구하는 자(보험금 지정 대리청구인)를 **보험가입초기** 또는 유지 중에 미리 지정하는 제도입니다.

※ 계약자가 자신을 위한 치매보험 가입하고 치매가 발생한 경우 계약자가 보험금을 직접 청구할 수 없어 보험금 청구가 곤란

키워드 분석

타이틀

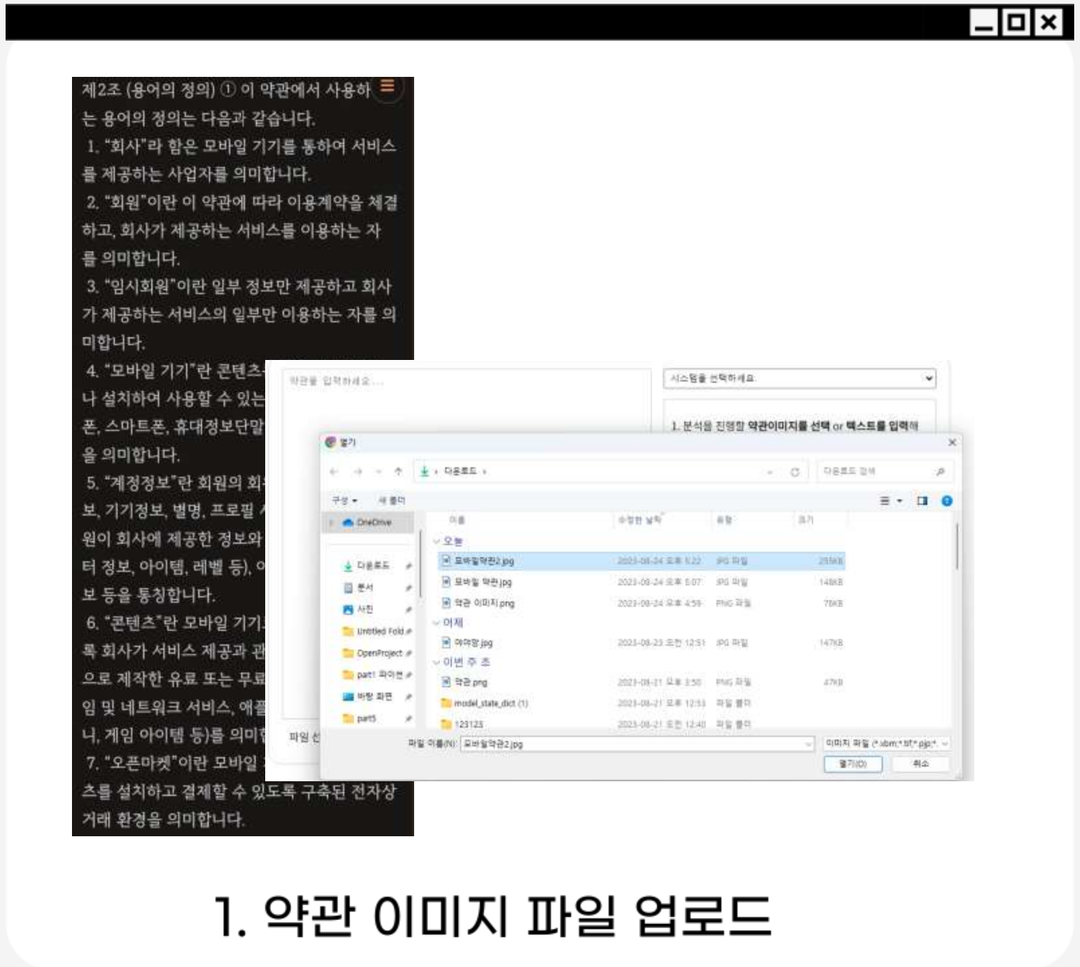
- 질병보험

키워드

- 납입보험료
- 대상금융상품
- 지정대리청구
- 적립부분
- 만기환급금
- 보험가입초기

파일 선택 서버로 전송 편집하기

01. OCR (Pytesseract)



1. 약관 이미지 파일 업로드

2. 정확도 향상을 위해 수정 가능

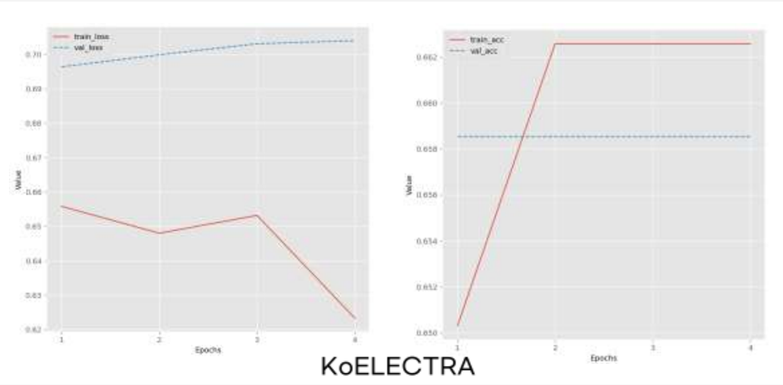
02. 유불리 판단

: 입력받은 약관에서, 문장별 유리/불리 여부를 판단

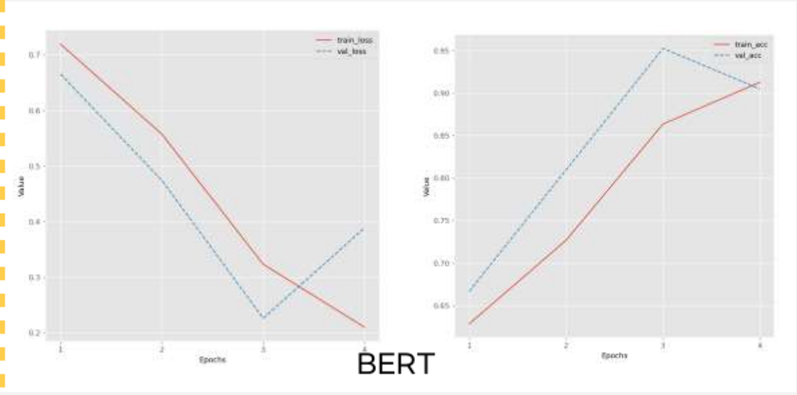
최종 선정 모델 : BERT

Why Bert?

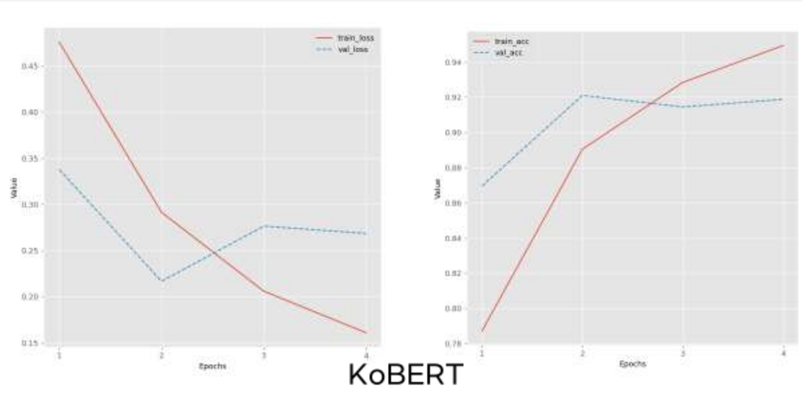
곡선형을 지켜본 후 에포크별 로스가 지속적으로 낮아지고 정확도가 올라가는 BERT, KoBERT 수행해보기로 결정



KoELECTRA



BERT



KoBERT

Why Bert?

< 예시에 대한 BERT, KoBERT의 유/불리 결과값 >

```
유리 조항 리스트
[]


불리 조항 리스트
['적립되는 보험료 없이 보장담보만으로 가입하시는 경우 보험계약 만기시 지급받는 금액(만기환급금)이 없습니다.', '예금자보호 한도는 본 보험회사에 있는 귀하의 모든 예금보호 대상금융상품의 해약환급금(또는 만기시 보험금이나 사고보험금)과 기타 지급금을 합하여 1인당 "최고 5천만원"이며, 5천만원을 초과하는 나머지 금액은 보호하지 않습니다.', '※계약자가 자신을 위한 치매보험 가입하고 치매가 발생한 경우 계약자가 보험금을 직접 청구할 수 없어 보험금 청구가 곤란', '이런 경우에 대비하여 보험금 대리청구인을 미리 지정하면 계약자를 대신하여 보험금을 청구할 수 있음', '이 경우 보장내용, 보험가입금액 및 납입보험료 등이 변경될 수 있습니다.']
```

KoBERT

```
유리 조항 리스트
['예금자보호 이 보험은 예금자보호법에 따라 예금보험공사가 보호합니다.', '지정대리청구 보험사고(치매 등) 발생으로 본인 스스로 보험금 청구가 현실적으로 어려운 상황이 발생할 경우 보험금을 대신 청구하는 자(보험금 지정 대리청구인)를 보험가입초기 또는 유지 중에 미리 지정하는 제도입니다.', '이런 경우에 대비하여 보험금 대리청구인을 미리 지정하면 계약자를 대신하여 보험금을 청구할 수 있음', '② 계약자가 계약내용 변경을 원하지 않는 경우 회사는 해당 계약의 계약내용 변경시점의 계약자적립액 및 미경과보험료를 지급하고, 해당계약은 더 이상 효력이 없습니다.', '③ 회사가 계약내용을 변경할 경우에는 지체없이 서면(등기우편 등), 전화(음성녹취) 또는 전자문서 등으로 보장내용 및 보험가입금액 변경내역, 보험료 수준, 특별약관 내용 변경 절차 등을 2회 이상 계약자에게 알립니다.']
```

```
불리 조항 리스트
['예금자보호 한도는 본 보험회사에 있는 귀하의 모든 예금보호 대상금융상품의 해약환급금(또는 만기시 보험금이나 사고보험금)과 기타 지급금을 합하여 1인당 "최고 5천만원"이며, 5천만원을 초과하는 나머지 금액은 보호하지 않습니다.']
```

BERT

KoBERT는 대부분을
불리한 약관으로 추출

BERT로 결정!

Bert 모델 학습 과정 전처리

STEP 01

Aihub에 있는 "법률/규정 (판결서, 약관 등) 텍스트 분석 데이터"를 활용


```
{
    "clauseField": "31",
    "ftcCnclsns": "2",
    "clauseArticle": [
        "제3조 (약관 등의 명시와 설명 및 개정)\n① 몰은 이 약관의 내용과 상호 및 대표자 성명, 영업소 소재지 주소(소비자의 불만을 처리할 수 있는 곳의 주소를 포함), 전화번호, 모사전송번호, 전자우편주소, 사업자등록번호, 통신판매업 신고번호, 개인정보관리책임자등을 이용자가 쉽게 알 수 있도록 ㅇㅇ 사이버몰의 초기 서비스화면(전면)에 게시해야 한다. 다만, 약관의 내용은 이용자가 연결화면을 통하여 볼 수 있도록 할 수 있다."
    ],
    "dvAntageous": "1",
    "comProvision": [
        "제3조 (약관 등의 명시와 설명 및 개정)\n① 몰은 이 약관의 내용과 상호 및 대표자 성명, 영업소 소재지 주소(소비자의 불만을 처리할 수 있는 곳의 주소를 포함), 전화번호, 모사전송번호, 전자우편주소, 사업자등록번호, 통신판매업 신고번호, 개인정보관리책임자등을 이용자가 쉽게 알 수 있도록 ㅇㅇ 사이버몰의 초기 서비스화면(전면)에 게시합니다. 다만, 약관의 내용은 이용자가 연결화면을 통하여 볼 수 있도록 할 수 있습니다."
    ]
}
```

	text	label
45	[CLS] 제4조(카드의 유효기한 및 재발급)\n② 카드사는 회원이 카드의 분실 및...	01.유리
29	[CLS] 제8조 (소멸시효)\n② 신유형 상품권을 구매한 날 또는 충전일로부터 6...	01.유리
24	[CLS] 제10조(수강신청의 철회)\n② 수강자가 다음 각호의 요건을 모두 갖춘 ...	01.유리
8	[CLS] 사 여부를 확인합니다. [SEP]	01.유리
5	[CLS] 제28조(가지급금의 지급) \n② 피보험자가 가지급금을 청구한 경우 보험...	01.유리
..	...	...
167	[CLS] 제2조(계약금과 피로연 식대비용의 지급)\n제1항 계약금은 전체금의 30...	02.불리
182	[CLS] 제20조(중도계약해지 및 계약해지) \n제2항 임의 해지\n가.행점 개점 ...	02.불리
192	[CLS] 제4조(변적 및 조건변경)\n1항 본 계약은 갑이 제시한 도면에 의한 구...	02.불리
152	[CLS] 공정경쟁질서 유지에 관한 상호협정에 의거 고의, 과실로 모집질서 위반, ...	02.불리
201	[CLS] 제30조(일반법령 및 기타의 적용)\n본 계약에 정한 이외의 사항 또는 ...	02.불리

STEP 02

토크나이저, 라벨인코더, 패딩, 임배딩의 과정
+ 그리드 서치

그리드 서치


```
{'input_ids': tensor([[   101,    101,   9672,  ...,      0,      0,      0],
        [   101,    101,   9672,  ...,      0,      0,      0],
        [   101,    101,   9672,  ...,      0,      0,      0],
        ...,
        [   101,    101,   9672,  ...,      0,      0,      0],
        [   101,    101,   9672,  ...,      0,      0,      0],
        [   101,    101, 107930,  ...,      0,      0,      0]]), 'token_type_ids': tensor([[0, 0, 0,  ..., 0, 0, 0],
        [0, 0, 0,  ..., 0, 0, 0],
        [0, 0, 0,  ..., 0, 0, 0],
        ...,
        [0, 0, 0,  ..., 0, 0, 0],
        [0, 0, 0,  ..., 0, 0, 0],
        [0, 0, 0,  ..., 0, 0, 0]]), 'attention_mask': tensor([[1, 1, 1,  ..., 0, 0, 0],
        [1, 1, 1,  ..., 0, 0, 0],
        [1, 1, 1,  ..., 0, 0, 0],
        ...,
        [1, 1, 1,  ..., 0, 0, 0],
        [1, 1, 1,  ..., 0, 0, 0],
        [1, 1, 1,  ..., 0, 0, 0]]), 'labels': tensor([0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 1, 0, 0, 1, 1, 0])}
```

```
Parameters: {'batch_size': 16, 'epochs': 5, 'lr': 2e-05}
Train Loss: 0.004133754118935684
Validation Loss: 0.07241602093593731
Validation Accuracy: 0.9912280701754386

Parameters: {'batch_size': 16, 'epochs': 5, 'lr': 3e-05}
Train Loss: 0.0020089423140277063
Validation Loss: 0.1049860051299906
Validation Accuracy: 0.9890350877192983

Parameters: {'batch_size': 16, 'epochs': 5, 'lr': 4e-05}
Train Loss: 0.016103669376434482
Validation Loss: 0.10641876388680223
Validation Accuracy: 0.981359649122807

Parameters: {'batch_size': 16, 'epochs': 5, 'lr': 5e-05}
Train Loss: 0.007305082956463593
Validation Loss: 0.07120562925717279
Validation Accuracy: 0.9901315789473685

Parameters: {'batch_size': 32, 'epochs': 2, 'lr': 2e-05}
Train Loss: 0.004701572275216392
Validation Loss: 0.06624312639490981
Validation Accuracy: 0.9901315789473685
```


STEP 03

문장에 대한 유/불리 확률 추출
유리리스트, 불리리스트 추출

pkl 파일


```
유리 조항 리스트
['예금자보호 이 보험은 예금자보호법에 따라 예금보험공사가 보호합니다.', '지정대리청구 보험사고(치매 등) 발생으로 본인 스스로 보험금 청구가 현실적으로 어려운 상황이 발생할 경우 보험금을 대신 청구하는 자(보험금 지정 대리청구인)를 보험가입초기 또는 유지 중에 미리 지정하는 제도입니다.', '이런 경우에 대비하여 보험금 대리청구인을 미리 지정하면 계약자를 대신하여 보험금을 청구할 수 있음', '④ 계약자가 계약내용 변경을 원하지 않는 경우 회사는 해당 계약의 계약내용 변경시점의 계약자적립액 및 미경과보험료를 지급하고, 해당계약은 더 이상 효력이 없습니다.', '④ 회사가 계약내용을 변경할 경우에는 지체없이 서면(등기우편 등), 전화(음성녹취) 또는 전자문서 등으로 보장내용 및 보험가입금액 변경내역, 보험료 수준, 특별약관 내용 변경 절차 등을 2회 이상 계약자에게 알립니다.']


불리 조항 리스트
['예금자보호 한도는 본 보험회사에 있는 귀하의 모든 예금보호 대상금융상품의 해약환급금(또는 만기시 보험금이나 사고보험금)과 기타 지급금을 합하여 1인당 "최고 5천만원"이며, 5천만원을 초과하는 나머지 금액은 보호하지 않습니다.']
```

03. 요약 기능

: 좋은 가독성을 위한 약관 조항별 요약

최종 선정 모델 : pko-t5

Why pko-t5?

다른 한국어요약 모델들(KoBART, KoBERT, KoELECTRA 등)과의 요약 성능과 러닝타임을 고려, pko-t5-base 기반 pretrained 모델 선정

Input Sentence: 네이버 주식회사(이하 '네이버(주)')는 www.naver.com을 비롯한 네이버(주) 도메인의 웹사이트 및 응용프로그램(어플리케이션, 앱)을 통해 다양한 정보의 제공을 중개하거나 해당 정보 제공자와의 연결을 매개하고 있습니다
KoBERT Sentence: : (이하'(주)') 는 www., (주) 및 (, 앱) 을 정보. 봅 큐 셈 번호 꼴 깔 셈 구장에서 침 침 층 깃 깃 베 셈 _스스로 붕 u 셈 _스스로 횟 닌 쇄 u OC 깃 벌 번호 벌 번호 벌 번호 벌 번호 벌 번호 업소 번호 섹 하우
KoBART Sentence: 네이버 주식회사는 www.naver.com을 비롯한 네이버(주) 도메인의 웹사이트 및 응용프로그램(어플리케이션, 앱)을 통해 다양한 정보의 제공을 중개하거나 해당 정보 제공자와의 연결을 매개하고 있으며 네이버 주식회사는
pko-t5-base Sentence: 네이버 주식회사(이하 '네이버(주)')는웹사이트 및 응용프로그램(앱),
pko-t5-base pre-trained Sentence: 네이버 주식회사는 네이버 도메인의 웹사이트 및 응용프로그램을 통해 다양한 정보의 제공을 중개하거나 해당 정보 제공자와의 연결을 매개하고 있습니다.

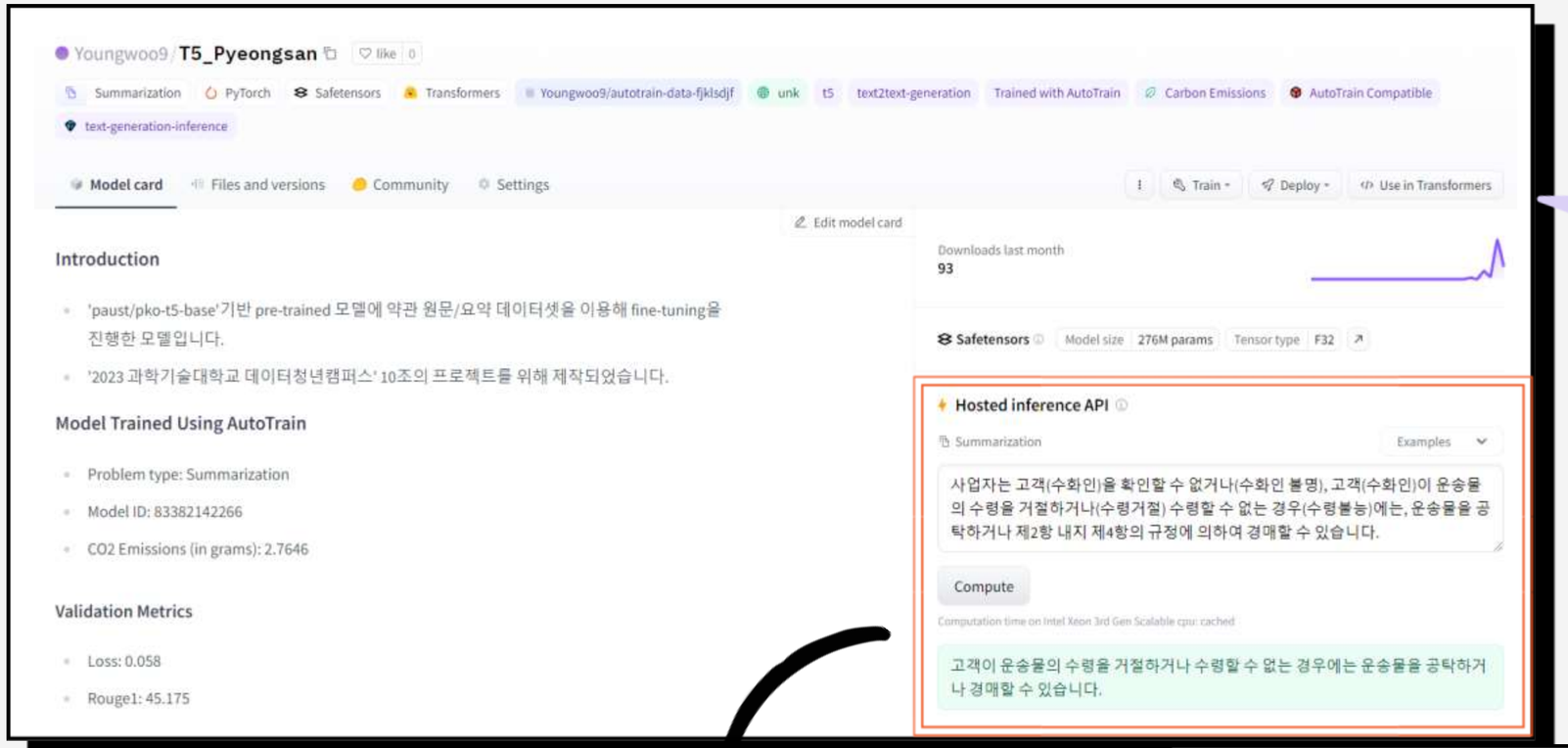
<모델별 예시 문장>

요약 모델 과정

	Total Sentences	Total Summaries
2	summarize: 협회는 법령에 따른 개인정보 보유이용기간 또는 정보주체로부터 개인정보를 수집시에 동의받은 개인정보 보유,이용	협회는 개인정보 보유 및 이용기간 내에서 개인정보를 처리 및 보유합니다.
3	summarize: 회사는 보험금의 청구에서 정한 서류를 접수한 때에는 접수증을 교부하고, 그 서류를 접수받은 후 지체없이 지급할 ㅂ	회사는 서류를 접수한 후 접수증을 교부하고 지급할 보험금을 결정하며 결정되면 5일 이내에 이를 지급
4	summarize: 수탁자는 신탁계약에 의하여 사업을 진행함에 있어 일체의 신탁사무처리비용의 조달의무를 부담하지 않는다.	수탁자는 신탁사무처리비용의 조달의무를 부담하지 않는다.
5	summarize: 몰은 이 약관의 내용과 상호 및 대표자 성명, 영업소 소재지 주소(소비자의 불만을 처리할 수 있는 곳의 주소를 포함),	몰은 약관의 내용과 정보를 이용자가 쉽게 알 수 있도록 00 사이버몰의 초기 서비스화면에 게시합니다
6	summarize: 제항의 규정에도 불구하고 은행은 접근매체의 갱신 또는 대체발급 등을 위하여 이용자의 동의를 얻은 경우로서 다음	은행은 접근매체의 갱신 또는 대체발급 등을 위하여 이용자의 동의를 얻은 경우로 다음 각 호에 해당하
7	summarize: 고객은 회사와 분쟁이 발생하는 경우 회사의 민원처리기구에 그 해결을 요구하거나 금융감독원, 협회 등에 분쟁조정	고객은 회사와 분쟁이 발생할 경우 회사의 민원처리기구에 그 해결을 요구하거나 금융감독원 협회 등에
8	summarize: 회사는 피보험자에게 다음 사항 중 어느 하나의 사유가 발생한 경우에는 보험수익자에게 약정한 보험금(보험금 지급	회사는 피보험자에게 다음 사항 중 어느 하나의 사유가 발생한 경우에는 보험수익자에게 약정한 보험금
9	summarize: 회사는 고객이 제항의 방법으로 거래를 위탁하고자 하는 경우 고객본인임을 확인할 수 있는 방법에 의하여야 한다.	회사는 고객이 제항의 방법으로 거래를 위탁하고자 하는 경우 고객본인임을 확인할 수 있는 방법에 의
	summarize: 을이 주식의 위탁판매업무를 수행함에 있어서는 관련법령에 위배되지 않는 한 본 계약이 적용되며 관련법령이 변경	을이 주식의 위탁판매업무를 수행함에 있어서는 관련법령에 위배되지 않는 한 본 계약이 적용되며 관

약관 특화 요약 모델을 얻고자 모델의 input에 맞도록 pair dataset을 구축해 fine-tuning 을 진행

기능별 모델 | OCR | Bert | pko-t5 | Konply | TF-IDF | Sentence-Bert



hugging face에서 pko-t5 기반 pre-trained model을 fine-tuning

최종 fine-tuned 모델의 요약 예시

	fine tuned 모델
예시 문장	네이버 주식회사(이하 '네이버(주)')는 www.naver.com을 비롯한 네이버(주) 도메인의 웹사이트 및 응용프로그램(어플리케이션, 앱)을 통해 다양한 정보의 제공을 중개하거나 해당 정보 제공자와의 연결을 매개하고 있습니다.
결과	네이버 주식회사는 도메인의 웹사이트 및 응용프로그램을 통해 다양한 정보의 제공을 중개하거나 해당 정보 제공자와의 연결을 매개하고 있습니다.
예시 문장	지정대리청구 보험사고(치매 등) 발생으로 본인 스스로 보험금 청구가 현실적으로 어려운 상황이 발생할 경우 보험금을 대신 청구하는 자(보험금 지정 대리청구인)를 보험가입 초기 또는 유지 중에 미리 지정하는 제도입니다.
결과	지정대리청구는 보험사고 발생으로 본인 스스로 보험금 청구가 어려운 상황이 발생할 경우 보험금을 대신 청구하는 자를 보험가입초기 또는 유지 중에 미리 지정하는 제도이다.

타이틀, 키워드 분석

명사 추출

명사 추출 방법: Konply

Why Konply?

Konlply 내에 있는 Kkma 형태소 분석기는
세부 품사를 구분하여 명사를 추출하기에
매우 효과적

```python
from konlpy.tag import Kkma

doc = '''위탁자 (이하 갑 이라 함) 은 별지1 기재의 소유부동산 (이하 신탁부동산
이라 함)의 관리 및 보전을 위하여 수탁자인 000 (이하 을 이라 함) 과
다음과 같이 부동산관리신탁계약(이하 신탁계약 이라 함)을 체결하기로 한다.

제 1 조 (신탁등기)
① 갑은 신탁계약 체결 즉시 신탁부동산을 을에게 인도하고 을은 이를 인수하여 신
탁을 원인으로 한 소유권이전등기 및 신탁등기를 한다.
② 갑은 본 계약 체결과 동시에 신탁등기에 필요한 등기권리증 인감증명서 등 제
반서류를 을에게 교부하여야 한다.
③ 제1항의 등기에 필요한 제비용은 갑이 부담키로 한다.

제 2 조 (수익자)

① 본 계약 있어서의 수익자는 별첨 특약에서 정한다.
② 갑은 을의 승낙을 얻어 수익자를 새로 지정하거나 변경할 수 있다. '''

kkma = Kkma()
nouns = kkma.nouns(doc)

filtered_nouns = list(filter(lambda noun: len(noun) > 1 and not any(char.isdigit() for char in noun), nouns))

grouped_nouns = [filtered_nouns[i:i+5] for i in range(0, len(filtered_nouns), 5)]

for group in grouped_nouns:
    print(group)
```

```
['위탁자', '이하', '별지', '기재', '소유']
['소유부동산', '부동산', '신탁', '신탁부동산', '관리']
['보전', '수탁자', '다음', '부동산관리신탁계약', '계약']
['신탁계약', '체결', '신탁등기', '등기', '즉시']
['인도', '인수', '원인', '소유권', '소유권이전등기']
['이전', '동시', '필요', '등기권리증', '권리증']
['인감', '인감증명서', '증명서', '반서류', '서류']
['교부', '제비', '부담', '부담키', '수익자']
['별첨', '특약', '승낙', '지정', '변경']
```

04. 타이틀 추출

최종 선정 모델

TF-IDF

단어 빈도와 역문서 빈도를 곱하여
단어의 중요도를 계산하는 방법

Why

TF-IDF?

문서 간의 공통 단어의
가중치를 고려하여
유사도를 계산하는 것이므로
비슷한 종류의 약관 분류에 효과적

기능별 모델 | OCR | Bert | pko-t5 | Konply | TF-IDF | Sentence-Bert

STEP 01

input

STEP 02

1. 벡터화
2.코사인 유사도 비교
3. 약관의 내용이 어떤 내용인지에 대한 타이틀 출력 모델

STEP 03

output

이미지 입력

위탁자 (이하 갑 이라 함) 은 별지1 기재의 소유부동산 (이하 신탁부동산 이라 함)의 관리 및 보전을 위하여 수탁자인 000 (이하 을 이라 함) 과 다음과 같이 부동산관리신탁계약(이하 신탁계약 이라 함)을 체결하기로 한다.

제 1 조 (신탁등기)
① 갑은 신탁계약 체결 즉시 신탁부동산을 을에게 인도하고 을은 이를 인수하여 신탁을 원인으로 한 소유권이전등기 및 신탁등기를 한다.
② 갑은 본 계약 체결과 동시에 신탁등기에 필요한 등기권리증 인감증명서 등 제반서류를 을에게 교부하여야 한다.
③ 제1항의 등기에 필요한 제비용은 갑이 부담키로 한다.

제 2 조 (수익자)

① 본 계약 있어서의 수익자는 별첨 특약에서 정한다.
② 갑은 을의 승낙을 얻어 수익자를 새로 지정하거나 변경할 수 있다.

제 3 조 (신탁의 목적 및 업무범위)
① 을은 신탁부동산의 소유권의 보존과 관리는 물론 임대 등의 부동산 사업을 행하여 그 수익을 수익자에게 교부하는 것을 목적으로 한다.
② 전항의 목적을 위하여 수행하는 업무의 범위는 별첨 특약에서 정하기로 한다.

TF-IDF

```
문서 1
apple: 0.5264
is: 0.5264
fruit: 0.5264
====================
문서 2
also: 0.5264
is: 0.5264
fruit: 0.5264
orange: 0.6677
====================
문서 3
is: 0.5264
fruit: 0.5264
banana: 0.6677
tropical: 0.6677
====================
```

보험 약관명 추출

05. 키워드 분석

: 부분 약관에 대해 사용자에게 흐름 및 큰 틀을 이해 시킬 수 있는 보조 도구

최종 선정 모델

Sentence-Bert

한국어를 지원하며 문장을 의미 있는벡터로 변환하는 키워드 추출에 특화

Why
Sentence-Bert

키워드 추출을 할 때
TextRank, Genism, RAKE와 같은 라이브러리를 사용가능

하지만 Sentence Transformers를 사용해야하는 이유는
1) 다른 라이브러리에 비해 문맥을 고려하여
2) 문장 내 단어 간의 상호작용을 잘 반영하고
3) 단어 수준 임베딩보다 더 의미론적으로 풍부한 임베딩을 제공하기 때문

Why
max sum similarity

의미적 관련성을 고려

키워드 추출 모델 작동 원리

01 임베딩 방식

Sentence Transformer 라이브러리 사용하여 벡터화,
후보 키워드들의 조합 중에서 최적의 조합 찾음

02 max sum similarity 알고리즘 원리

문서와 유사도가 높은 명사들을 조합한 후
코사인 유사도의 합이 가장 큰 조합을 키워드로 출력

```
    distances_candidates = cosine_similarity(candidate_embeddings, candidate_embeddings)

    words_idx = list(distances.argsort()[0][-nr_candidates:])
    words_vals = [nouns[index] for index in words_idx]
    distances_candidates = distances_candidates[np.ix_(words_idx, words_idx)]

    min_sim = np.inf
    candidate = None
    for combination in itertools.combinations(range(len(words_idx)), top_n):
        sim = sum([distances_candidates[i][j] for i in combination for j in combination if i != j])
        if sim < min_sim:
            candidate = combination
            min_sim = sim

    return [words_vals[idx] for idx in candidate]

top_keywords = max_sum_sim(doc_embedding, candidate_embeddings, nouns, top_n=6, nr_candidates=10)
print(top_keywords)
```

```
Downloading (...)925a9/.gitattributes: 100% 690/690 [00:00<00:00, 31.8kB/s]
Downloading (...)_Pooling/config.json: 100% 190/190 [00:00<00:00, 9.06kB/s]
Downloading (...)1a515925a9/README.md: 100% 3.99k/3.99k [00:00<00:00, 305kB/s]
Downloading (...)515925a9/config.json: 100% 550/550 [00:00<00:00, 35.9kB/s]
Downloading (...)ce_transformers.json: 100% 122/122 [00:00<00:00, 8.30kB/s]
Downloading pytorch_model.bin: 100% 265M/265M [00:00<00:00, 310MB/s]
Downloading (...)nce_bert_config.json: 100% 53.0/53.0 [00:00<00:00, 3.92kB/s]
Downloading (...)cial_tokens_map.json: 100% 112/112 [00:00<00:00, 5.71kB/s]
Downloading (...)925a9/tokenizer.json: 100% 466k/466k [00:00<00:00, 21.8MB/s]
```

```
주요 키워드:
['납입보험료', '대상금융상품', '지정대리청구', '적립부분', '만기환급금', '보험가입초기']
```

텍스트에서의 가장 중요한 상위 6개 단어

CHAPTER 3

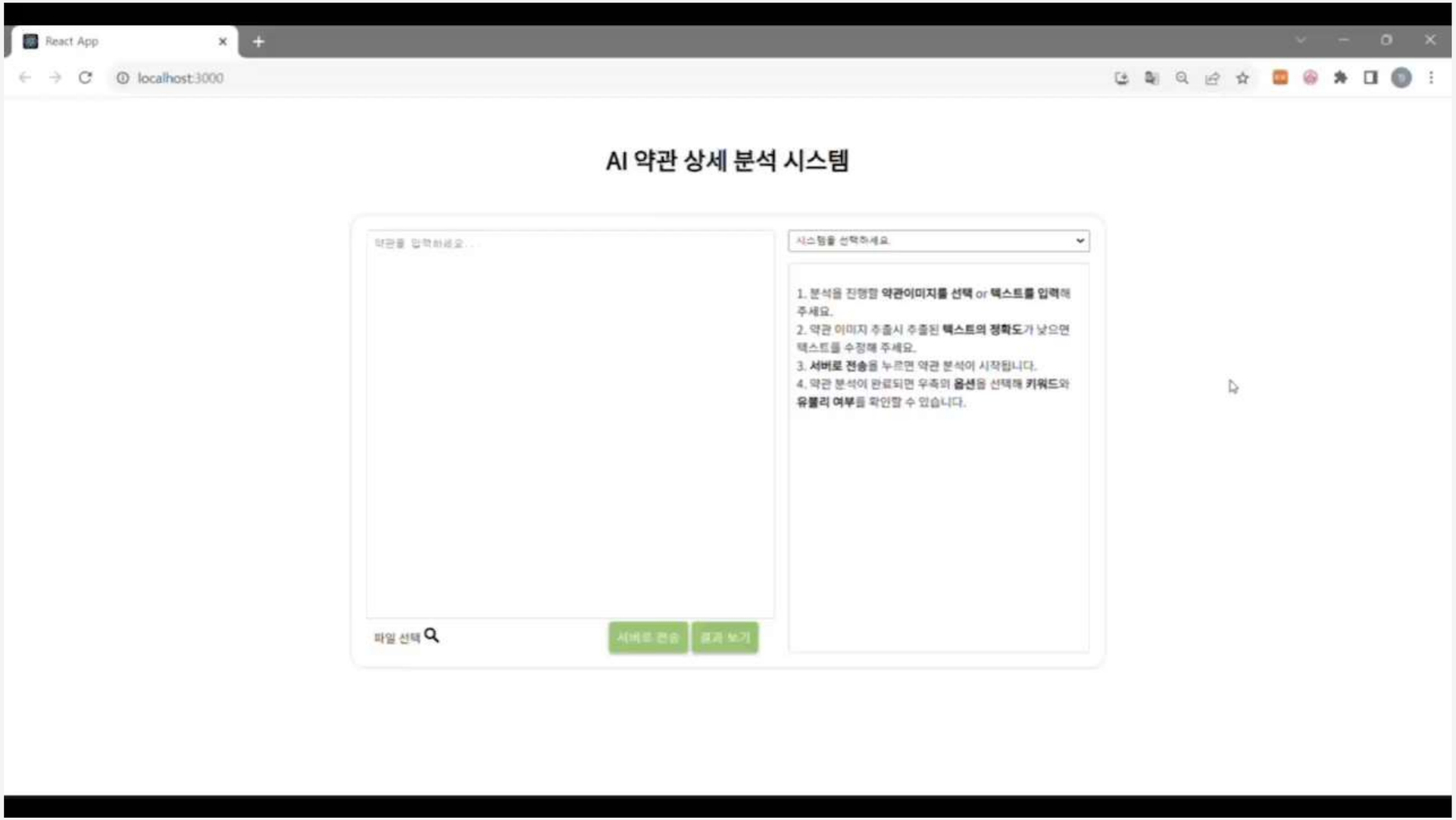
결론

01. 영상

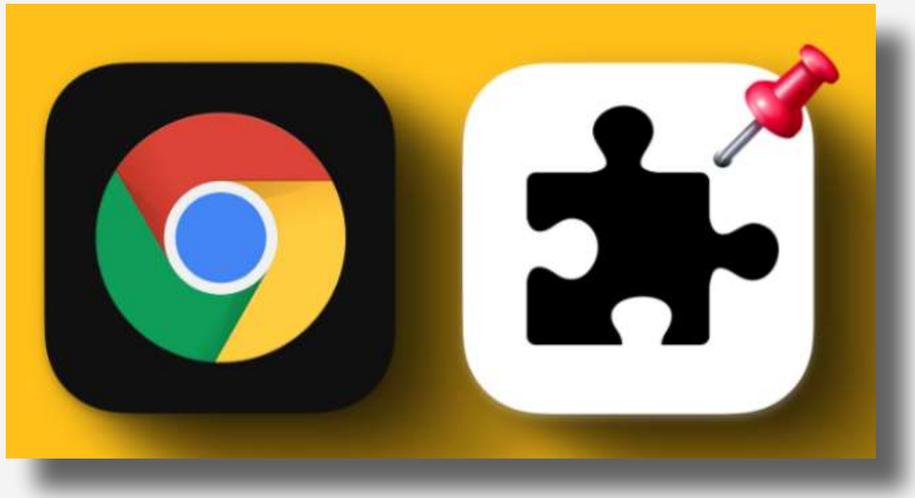
02. 향후 계획 및 기대 효과

03. 한계 및 개선 방안

웹 영상


React App
localhost:3000
AI 약관 상세 분석 시스템
시스템을 선택하세요.
1. 분석을 진행할 약관이미지를 선택 or 텍스트를 입력해 주세요.
2. 약관 이미지 추출시 추출된 텍스트의 정확도가 낮으면 텍스트를 수정해 주세요.
3. 서버로 전송을 누르면 약관 분석이 시작됩니다.
4. 약관 분석이 완료되면 우측의 옵션을 선택해 키워드와 유불리 여부를 확인할 수 있습니다.
파일 선택

향후 계획

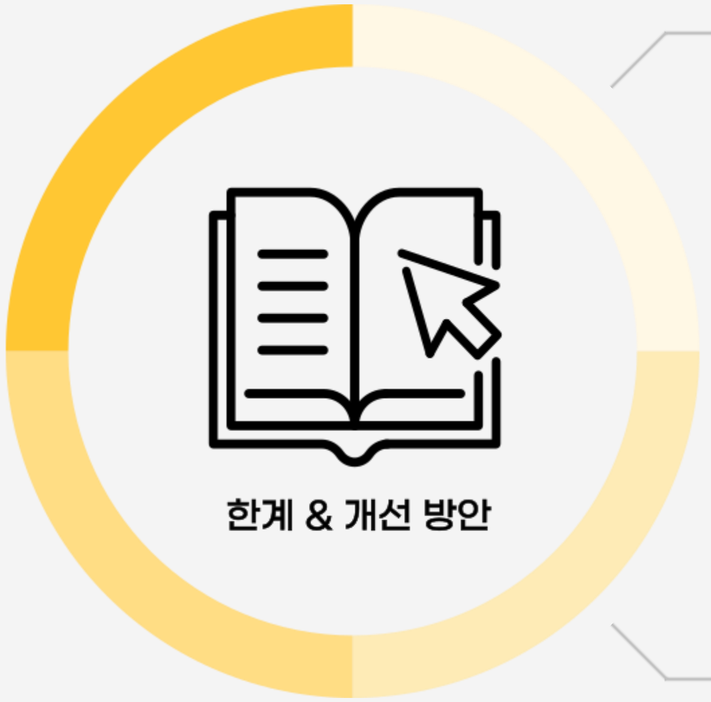


구글 확장 프로그램 개발 및 배포

기대 효과

- 사용자 권리 보호
- 불이익 초소화
- 시간 절약 및 편의 증대

한계 및 개선사항



키워드/ 타이틀

공통: TF-IDF 분석 방식 및 빈도 분석은 단어의 문맥을 고려x, 의미 이해능력 없어 제한적
→ 단어의 문맥 파악 필요!
BERT,GPT등과 같은 프리트레인드 언어 모델을 fine-tuning하여 사용

키워드 추출 모델: 불용어 처리 기준이 모호, 최적의 추출 단어수가 6개가 아닐 수 있음
→ 더 상세한 불용어 처리 및 최적의 추출 단어 수 모색 필요

유/불리 판단

"제 n조" 형태가 아닌 조항이 실제로는 많음
→ 문자열 슬라이싱 과정에서 전처리 단계 필요

요약

"."으로 끝나는 문장에 대해서만 요약
→ 성능 개선을 위한 충분한 양의 Pair Dataset 확보 필요

Thank You

데이터 청년 캠퍼스