

## Bill Gates: 법률안의 문

권민석 고려대학교 산업경영공학부

배지원 고려대학교 경제학과

정종락 고려대학교 일반대학원 정치외교학과

정하연 고려대학교 경영대학



# 국회 법률안 가결 예측

시의성과 논의 흐름을 중심으로

2024 K-DS 해커톤

# 목 차

1. 프로젝트 개요
  2. 시의성 분석
  3. 국회 법률안/회의록 분석
  4. 보완 방향성 및 계획
- 부록

# 프로젝트 개요

주제 선정 배경

서비스 개요

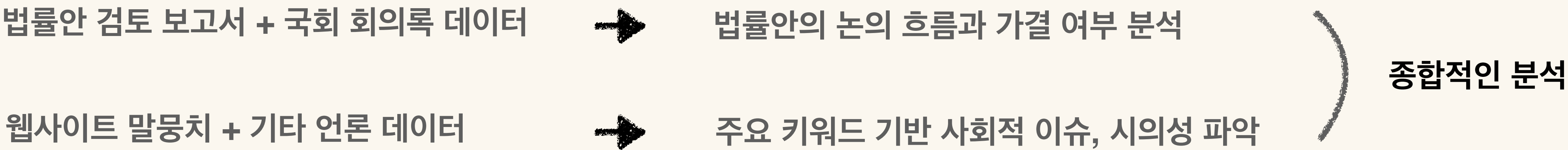
사용 데이터

## 주제 선정 배경

- 법률안 - 회의록 - 기사 등 각지에 흩어져 있는 다양한 정보를 통합, 이를 기반으로 한 체계적인 분석/이해 필요
- 입법/의정활동 전반에서 국회의원의 전문성 향상을 위한 솔루션 제공
- 국회 전문위원의 법률안 검토 등 업무 전반의 효율성 제고
- 국민/산업계에 입법, 의정활동의 흐름 전반의 열린 정보 제공

서비스 개요

“ 법률안 제안 및 분석에 활용할 수 있는  
포괄적인 데이터프레임과 분석/예측 모델 구축 ”



\* 추후 법률, 법률안, 기사 데이터 등 추가 데이터를 활용해 구체화 예정

사용 데이터

#자연어

BETA 법률안 검토 보고서 요약 데이터

분야 한국어 유형 텍스트

구축년도 : 2023 갱신년월 : 2024-06 조회수 : 2,994 다운로드 : 182 용량 : 146.30 MB

- ▶ 국회 법률안에 대한 전문위원의 검토 보고서
- ▶ 위원회, 분야, 발의자, 원문/요약, 처리결과 등
- ▶ `bill\_id`를 다른 데이터셋과 merge하기 위한 primary key로 활용

#자연어

BETA 국회 회의록 기반 지식 검색 데이터

분야 한국어 유형 텍스트

구축년도 : 2023 갱신년월 : 2024-06 조회수 : 3,620 다운로드 : 239 용량 : 313.48 MB

- ▶ 국회 회의록의 발화를 질의/답변 쌍으로 쪼개어 구성
- ▶ 회의의 위원회, 의제(법률안, 현안), 발화내용 등

#챗봇 #스마트 스피커 #AI #인공지능 대규모 한국어 학습용 데이터 #데이터 생태계 #자연어

대규모 웹데이터 기반 한국어 말뭉치 데이터

분야 한국어 유형 텍스트

구축년도 : 2021 갱신년월 : 2022-07 조회수 : 22,843 다운로드 : 1,639 용량 : 8.95 GB

- ▶ 웹데이터 기반 뉴스 보도 텍스트 데이터
- ▶ 필요한 변수의 column을 정의해 json 추출/통합
- ▶ `sentences`에서 뉴스 제목과 내용 추출
- ▶ `field`를 해당 뉴스 내용의 분야로 활용

4862715 rows x 4 columns

json으로 제공된 위 데이터셋들을 각각 하나의 csv로 merge하여 활용

사용 데이터

#자연어

BETA 법률안 검토 보고서 요약 데이터

분야 한국어 유형 텍스트

구축년도 : 2023 갱신년월 : 2024-06 조회수 : 2,994 다운로드 : 182 용량 : 146.30 MB

#챗봇 #스마트 스피커 #AI #인공지능 대규모 한국어 학습용 데이터 #데이터 생태계 #데이터 플랫폼

대규모 웹데이터 기반 한국어 말뭉치 데이터

분야 한국어 유형 텍스트

구축년도 : 2021 갱신년월 : 2022-07 조회수 : 22,843 다운로드 : 1,639 용량 : 8.95 GB

#자연어

BETA 국회 회의록 기반 지식 검색 데이터

분야 한국어 유형 텍스트

구축년도 : 2023 갱신년월 : 2024-06 조회수 : 3,620 다운로드 : 239 용량 : 313.48 MB

- ▶ `법률안 검토 보고서 요약 데이터` 의 법률안 검토 보고서 관련 데이터와 `대규모 웹데이터 기반 한국어 말뭉치 데이터` 의 기사 데이터를 관련지어 분석하기 위해
  - ▶ 각 데이터의 생성(제출, 보도) 시기를 맞추어 전처리
  - ▶ 법률안의 시의성을 확인하기 위해
  - ▶ 1년의 시간차를 기준으로 구간을 설정하여 필터링 → 국회 회의록 데이터에도 적용
- [데이터의 시간 범위]
- 기사 데이터: 2014-01-12 ~ 2021-07-01
  - 법률안 검토 보고서 데이터: 2015-01-12 ~ 2022-07-01
  - 국회 회의록 데이터: 2015-01-12 ~ 2022-07-01

# 시의성 분석

법률안 검토 보고서 데이터와 기사 데이터 연결



시의성 요인 변수의 정의

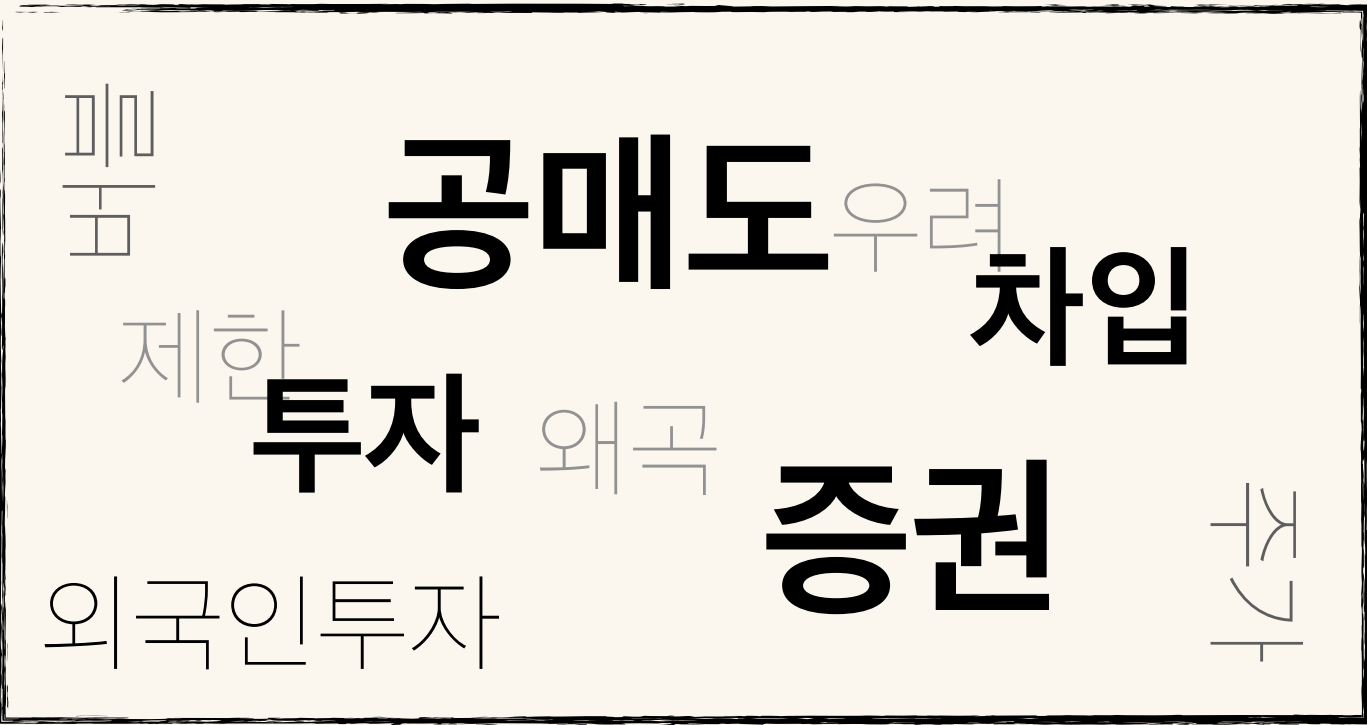
1. 직접적 연관 키워드

- ▶ 해당 키워드가 직접적으로 등장한 빈도
- ▶ 해당 키워드가 직접적으로 등장한 시점 차의 중앙값

2. 포괄적 연관 키워드

- ▶ 관련 키워드가 등장한 빈도

법률안 검토 보고서



직접적 연관 키워드

.....

.....

포괄적 연관 키워드

기사 데이터

기사 1 :	_____
기사 2 :	_____
기사 3 :	_____
	⋮

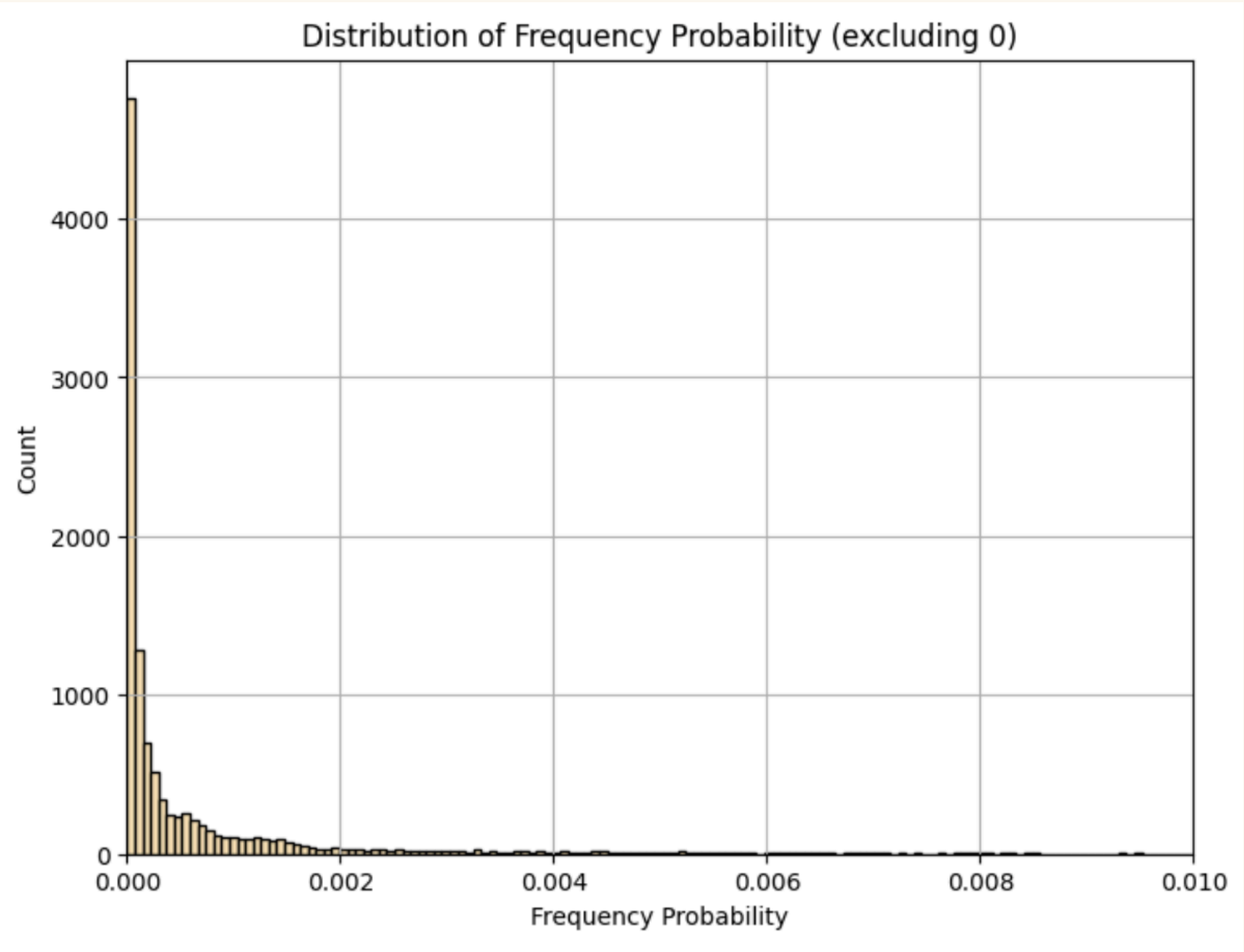
\* 법률안의 발의-검토 시점으로부터 최근 1년 간의 기사에 한하여 시의적인 연관성 파악

시의성 요인 변수를 반영한 최종 데이터

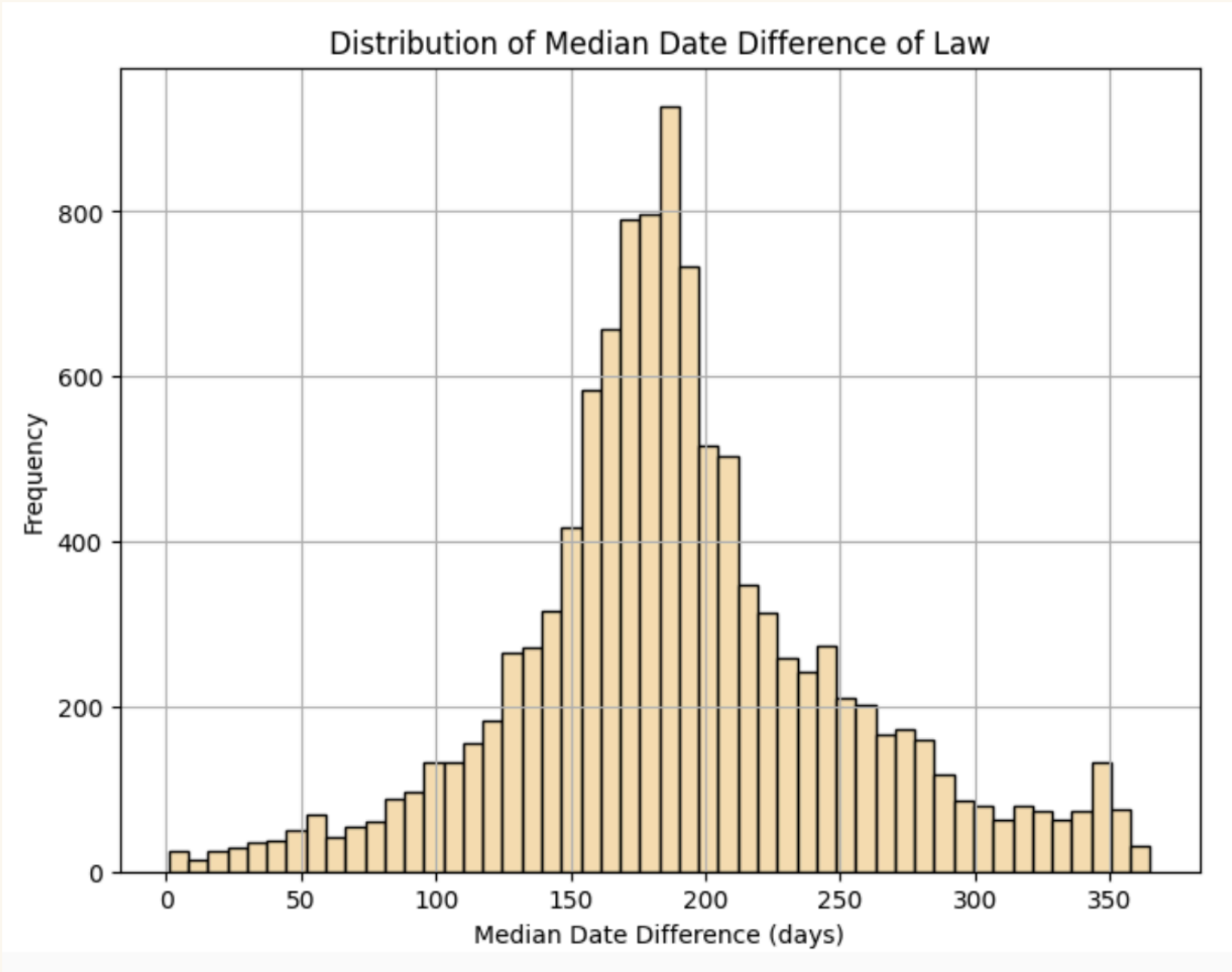
	id	date	field	terminology	text	direct_freq	direct_median	indirect_freq	non_shown	result
0	1039_9088	2017-03-02	국회에서의 증언·감정 등에 관한 법률 일부개정법률안	불출석 등의 죄, 벌금형을 삭제, 형량을 상향조정, 국정감·조사, 국회에서의 증언·...	개정안(우상호·김중로·이원욱 의원안)은 불출석 등의 죄의 형량을 현행 3년 이하의 ...	0.000000	NaN	0.221961	1	대안반영폐기
1	1040_1628	2020-09-24	민법 일부개정법률안	민법, 성적자기결정권, 손해배상 청구권	가. 소멸시효 제도의 의의 및 미성년자에 대한 특칙\r\n소멸시효는 권리자가 자신의...	0.000060	210.0	0.825995	0	대안반영폐기
2	1039_1553	2021-12-02	조세특례제한법 일부개정법률안	벤처기업, 비과세 특례, 주식매수선택권, 적격주식매수선택권, 피인수기업, 인센티브,...	□ 현행법 제16조의4는 벤처기업 적격주식매수선택권 행사이익에 대한 과세특례를 규정...	0.000947	243.0	0.913536	0	대안반영폐기
3	1040_656	2021-02-24	국회법 일부개정법률안	상임위원, 영리행위, 공직자윤리법, 이해충돌 방지 의무, 국회법, 의정활동	김남국의원안 제40조의2는 상임위원의 영리행위 등 금지행위를 그 유형별로 상세히 규...	0.000591	240.0	0.515267	0	철회
4	1040_3210	2018-12-08	조세특례제한법 일부개정법률안	환경보전시설, 투자세액공제, 조세특례심층평가, 세액공제율	□ 10건의 개정안은 환경보전시설 투자세액공제의 적용기한을 연장하고, 7건의 개정안...	0.000006	34.0	0.743927	0	대안반영폐기

`non\_shown`: 지난 1년간 키워드가 직접적으로 등장한 적이 없으면 1, 있으면 0

## 시의성 요인 변수의 분포



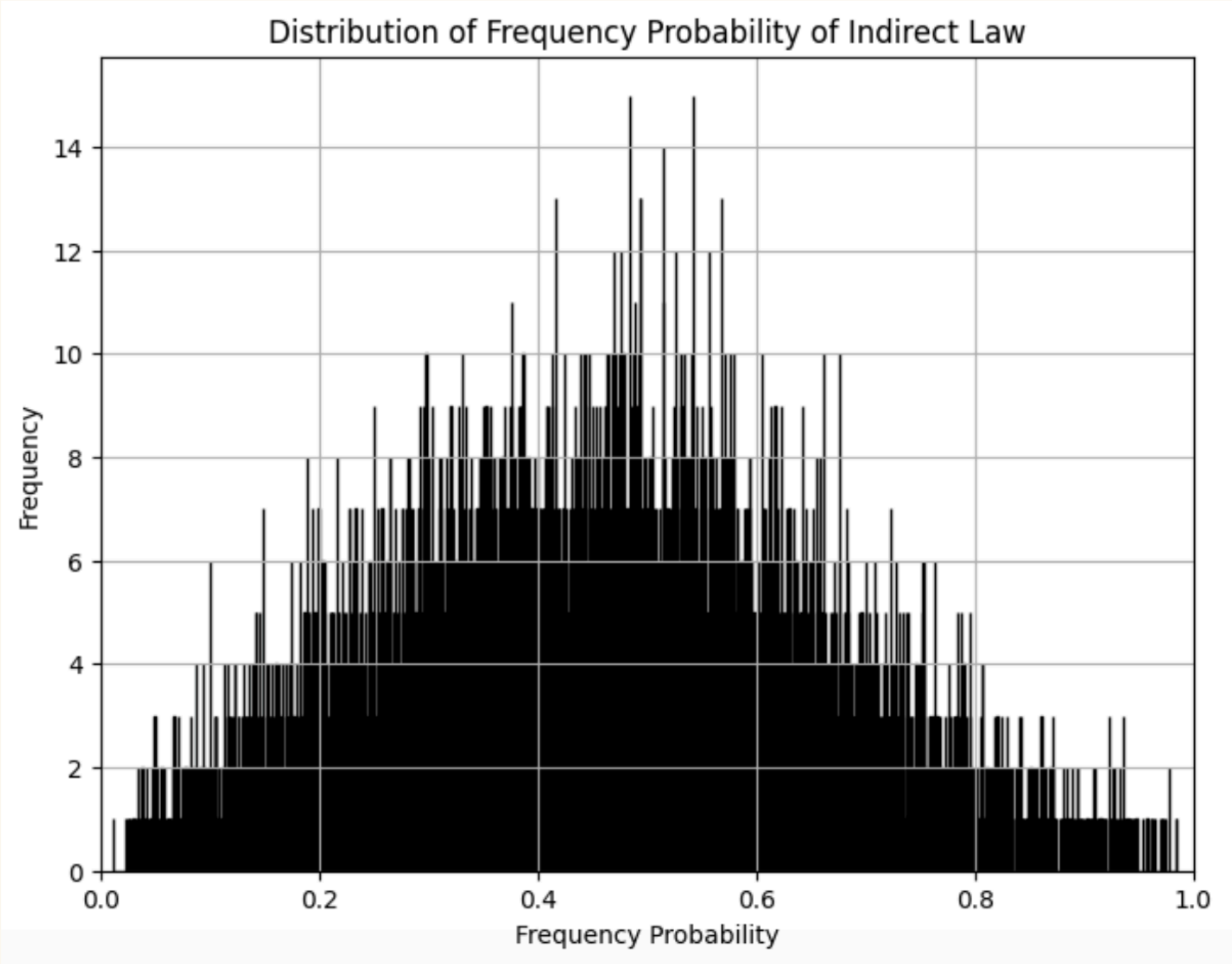
지난 1년 간 직접적 연관 키워드가 등장한  
뉴스기사 빈도 히스토그램



지난 1년 간 직접적 연관 키워드가 등장한  
시점 차이 일수의 중앙값 히스토그램

\* 법률안 검토 보고서 제출일 기준

## 시의성 요인 변수의 분포

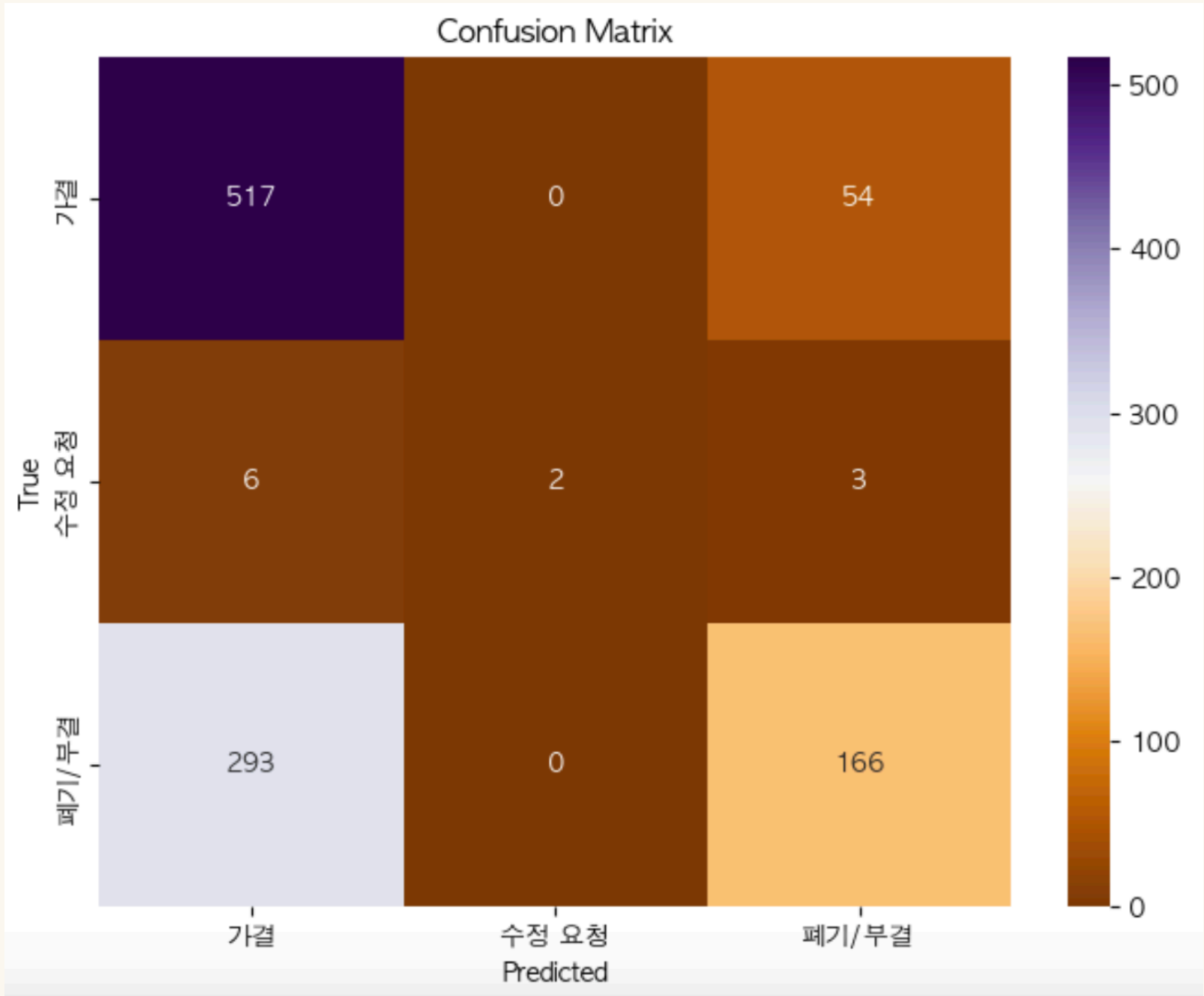


지난 1년 간 포괄적 연관 키워드가 등장한 뉴스기사 빈도 히스토그램

\* 법률안 검토 보고서 제출일 기준

분류/예측 분석: 데모 모델링 결과

	Precision	Recall	F1-Score	Support
가결	0.63	0.91	0.75	571
폐기/부결	1.00	0.18	0.31	11
수정 요청	0.74	0.36	0.49	459
Accuracy	-	-	0.66	1041
Macro Avg.	0.79	0.48	0.51	1041
Weighted Avg.	0.69	0.66	0.63	1041



타겟 변수 | 가결, 폐기/부결, 수정요청 - 3개 범주 정의  
모델 | Random Forest Classifier (Multi-class)

추후 데이터 불균형 해소, 추가 데이터 확보, 변수 추가 등의 과정을 통해 모델 성능 향상 예정

# 국회 법률안/회의록 분석

법률안 검토 보고서와 국회 회의록 데이터 연결



법률안 검토 보고서 데이터의 구조

term	session	date	law	result	text	proposer	token_freq
21	378.0	2020-05-29	자본시장과 금융투자업에 관한 법률 일부개정법률안	임기만료폐기	(1) 개정안은 현재 자본시장법에서 제한적으로 허용하고 있는 차입공매도를 전면 금지...	조경태의원 등 13인	{'공매도': 18, '시장': 9, '투자': 6, '전면': 5, '금지': 5...
21	378.0	2020-05-29	자본시장과 금융투자업에 관한 법률 일부개정법률안	임기만료폐기	(1) 개정안은 유사투자자문 계약의 해지·해제시 부당한 위약금 청구 및 대금 환급 ...	김병기의원 등 12인	{'투자자': 30, '유사': 23, '업': 22, '소비자': 20, '환급'...
21	382.0	2020-12-09	자본시장과 금융투자업에 관한 법률 일부개정법률안	대안반영폐기	□ 각 개정안 부칙의 시행일을 살펴보면, 홍성국의원안 및 박용진의원안은 공포 후 3...	김병욱의원 등 12인	{'원안': 8, '개월': 3, '각': 2, '경과': 2, '날로': 2, '...'...
21	378.0	2020-05-29	자본시장과 금융투자업에 관한 법률 일부개정법률안	임기만료폐기	(1) 온라인소액투자(이하 “크라우드펀딩”)에 대한 적격투자자의 투자한도와 관련하여...	송희경의원 등 12인	{'투자': 34, '금액': 26, '한도': 25, '소득': 18, '연간'...
21	378.0	2020-05-29	자본시장과 금융투자업에 관한 법률 일부개정법률안	임기만료폐기	(1) 개정안은 유사투자자문업자의 허위·과장 광고를 금지하고 이를 위반할 경우 1천...	자상욱의원 등 14인	{'투자자': 23, '광고': 19, '유사': 12, '업': 12, '업자'...
...	...	...	...	...	...	...	...
21	378.0	2020-05-29	자본시장과 금융투자업에 관한 법률 일부개정법률안	임기만료폐기	(1) 개정안은 법령을 위반한 전문사모투자중개업자에 대한 등록취소(안 제420조 및...	유동수의원 등 11인	{'투자': 7, '중개업': 7, '이하': 4, '명칭': 3, '사용': 3,...
21	378.0	2020-05-29	자본시장과 금융투자업에 관한 법률 일부개정법률안	임기만료폐기	(1) 개정안은 상장회사의 경우 주주총회와 관련하여 전자투표(상법 제368조의4) ...	김성원의원 등 11인	{'의결권': 21, '전자': 18, '행사': 17, '투표': 13, '권유'...
21	378.0	2020-05-29	자본시장과 금융투자업에 관한 법률 일부개정법률안	임기만료폐기	(1) 개정안은 공매도 규제 위반에 대한 제재를 강화하려는 것으로, ① 형사처벌(1...	조경태의원 등 13인	{'공매도': 16, '부과': 8, '행위': 8, '제재': 7, '처벌': 7...
21	378.0	2020-05-29	자본시장과 금융투자업에 관한 법률 일부개정법률안	임기만료폐기	한 인적·물적 요건 등 등록요건을 규정하려는 것임.\r\n(2) 자본시장법에 따르면...	유동수의원 등 11인	{'투자': 14, '중개업': 11, '중소기업': 9, '금융': 7, '요건'...
21	378.0	2020-05-29	자본시장과 금융투자업에 관한 법률 일부개정법률안	임기만료폐기	(1) 개정안은 파생상품등에 대한 금융투자업자의 투자권유 과정에 서의 불완전판매를 ...	박대출의원 등 14인	{'파생상품': 21, '투자': 16, '투자자': 16, '권유': 13, '상'...

json을 입력받아 필요한 변수를 column으로 정의하여 선별해 데이터프레임 구성

이후 각 법률안 검토 보고서의 ‘주요 키워드-빈도 수’ 추출. 기존 raw 데이터의 terminology와 함께 활용

- ▶ 동일한 법률에 대한 여러 검토 보고서가 여러 시기에 걸쳐 다수 존재 (동일한 법률안 명칭-다른 내용, 혹은 다른 발의자 등)
- ▶ raw 데이터의 날짜와 국회의 회기를 mapping하여 `session` column 정의
- ▶ 검토보고서 명칭에 함께 기재되어 있던 발의자를 분리하여 `proposer` column 정의

국회 회의록 질의응답 데이터: 맥락 기반 데이터 통합

\* 상세 코드는 부록 참고

json에서 데이터 분석/연결에 필요한 변수만을 column으로 정의해 데이터 프레임 형성. 이후 `context\_full` column 생성

`agenda\_group` [회의 id - agenda]별로 질의응답 셋(`context` - `answer.comment`)을 모두 포함한 리스트 형태

- ▶ 개별 행에서 회의의 맥락(context)과 해당 질문에 대한 답변(answer.comment)을 하나의 흐름으로 파악하기 위해 두 요소 결합
- ▶ 각 `agenda\_group`에 포함된 행 내용을 모두 결합, `context\_full`column을 생성해 일련의 ‘발언-답변’ 쌍 구성
- ▶ 이를 통해 각 회의 중 특정 안건에 대한 논의 전개 과정을 연속적으로 분석 가능

parliament\_df.info()

<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 29398 entries, 0 to 29397  
Data columns (total 19 columns):  
# Column Non-Null Count Dtype  
---  
0 id 29398 non-null int64  
1 date 29398 non-null object  
2 conference\_number 29398 non-null int64  
3 meeting\_name 29398 non-null object  
4 question\_number 29398 non-null int64  
5 main\_committee 29398 non-null object  
6 sub\_committee 14422 non-null object  
7 committee\_name 29398 non-null object  
8 meeting\_number 29398 non-null object  
9 session\_number 29398 non-null object  
10 agenda 29398 non-null object  
11 law 12480 non-null object  
12 qna\_type 29398 non-null object  
13 context 29398 non-null object  
14 context\_learn 29398 non-null object  
15 questioner\_ID 29398 non-null object  
16 answerer\_ID 29398 non-null object  
17 answer.comment 29398 non-null object  
18 answer.keyword 26912 non-null object  
dtypes: int64(3), object(16)  
memory usage: 4.3+ MB

parliament\_wrap.info()

✓ 0.0s  
<class 'pandas.core.frame.DataFrame'>  
Index: 3961 entries, 44851 to 52664  
Data columns (total 6 columns):  
# Column Non-Null Count Dtype  
---  
0 date 3961 non-null object  
1 meeting\_name 3961 non-null object  
2 main\_committee 3961 non-null object  
3 agenda 3961 non-null object  
4 agenda\_index 3961 non-null object  
5 context\_full 3961 non-null object  
dtypes: object(6)  
memory usage: 216.6+ KB

parliament\_df.head(10)

id	date	conference_number	meeting_name	question_number	main_committee	sub_committee	committee_name	meeting_number	session_number	agenda	law	qna_type	context	context_learn	questioner_ID	answerer_ID
0	47897	2017-12-27	특별위원회	1	4차산업혁명특별위원회	None	4차산업혁명특별위원회	제355회	제2차	1. 4차 산업혁명 관련 제1차 공청회	NaN	추출형	민주당의 강령수입이다. 세 본 열광할 않았습니다. 이 경향 본부 장님과 이 정인...	중소벤처기업부가 먼저 정책의 컨트롤 타워로서 효과적으로 기능하기 위해서는 우선, 영...	842	
1	47897	2017-12-27	특별위원회	2	4차산업혁명특별위원회	None	4차산업혁명특별위원회	제355회	제2차	1. 4차 산업혁명 관련 제1차 공청회	NaN	추출형	한 가지만 더 여부가 보였습니다. 200만 개 일 자리 열광 하십시오. 좋은 일자 리?... 이경상 본 부장님, 정리를 잘 해 주셔서 감사하고, R&D 예산의 할당 및 운영 과정에 있어 규제 통 근본 개선 요구...	벤처 특별법이 시 200만 개 일 자리 열광 하 과정에서 누락하 이 324만 개리...	842	
2	47897	2017-12-27	특별위원회	4	4차산업혁명특별위원회	None	4차산업혁명특별위원회	제355회	제2차	1. 4차 산업혁명 관련 제1차 공청회	NaN	추출형	이경상 본 부장님, 정리를 잘 해 주셔서 감사하고, R&D 예산의 할당 및 운영 과정에 있어 규제 통 근본 개선 요구...	부처 간 및 부서 간의 간막이는 사라지고, R&D 예산의 할당 및 운영 과정에 있어 혁신을 받...	3199	

parliament\_wrap

id	date	meeting_name	main_committee	agenda	agenda_index	context_full
0	44851	2015-11-03	특별위원회	국민안전혁신특별위원회	1. 현안보고가. 산업통상자원부나. 원자력안전위원회	[두 번째로 신고리 3호기 건설현장 질식사고, 보고서에 나와 있는데 보고를 생략해...
1	44857	2015-12-01	특별위원회	공무원연금개혁특별위원회	1. 업무보고가. 인사혁신처나. 기획재정부다. 보건복지부	[후선은 다음에 저희가 다시 볼 때는 세 부처 전부, 기본적으로 공무원연금과 관련...
2	44858	2015-12-07	특별위원회	국민안전혁신특별위원회	1. 현안보고가. 환경부나. 고용노동부	[경기 고양 일산 동구 유은혜입니다. 이기관 장관님께 질의하겠습니다. 말씀하...
3	44863	2015-12-09	특별위원회	군인관계선및병영문화혁신특별위원회	1. '민.관.군 병영문화혁신위원회, 권고안에 대한 국방부 검토결과 보고	[장관님, 국방부는 지난 21일 정병 신재경사 등 검사규칙을 개정해서 현역 입영...
4	44875	2015-02-05	소위원회	농림축산식품해양수산위원회	1. 흉의 날 제정에 관한 법률안(김춘진 의원 대표발의)(계속) 2. 농어업.농어촌...	[저도 지금 어디 청문회 증인할 것 같아 오다니까 지금 어떤 내용들이 어떻게 전개...
...	...	...	...	...	...	...
3956	52640	2022-12-28	소위원회	정무위원회	1. 하도급거래 공정화에 관한 법률 일부개정법률안(김경만 의원 대표발의)(의안번호 ...	[상생협력법에, 이번에는 의뢰된 자료 혹시 여기서 좀 볼 수 있나요? 정리된 자료...
3957	52642	2023-01-03	특별위원회	연금개혁특별위원회	1. 민간자문위원회 중간 보고의 건	[특직연금 관련해서 중 궁극한데요, 실제로 특직연금의 가입률도 낮지만 실제로 연금으...
3958	52664	2023-01-30	소위원회	문화체육관광위원회	1. 자연유산의 보존 및 활용에 관한 법률안(이상현 의원 대표발의)(의안번호 210...	[지난번에 환경부와도 문화재청이 합의를 했는데 그래도 그 내용을 한번 짧아 주...
3959	52664	2023-01-30	소위원회	문화체육관광위원회	10. 콘텐츠산업 진흥법 일부개정법률안(이상현 의원 대표발의)(의안번호 210983...	[장관만요. 이것 소위에서 논의한 것이 언제지요? 전반기 때문입니다, 21년도. ...
3960	52664	2023-01-30	소위원회	문화체육관광위원회	9. 예술인 복지법 일부개정법률안(이병훈 의원 대표발의)(의안번호 2115500)	[황보승희입니다. 확인하고 싶은 것이 이미 광역문화재단 14곳을 협력 협회 단체...

3961 rows x 7 columns

Bill Gates: 법률안의 문 | 국회 법률안 가결 예측

2024 K-DS 해커톤



## 국회 회의록 데이터 정제 및 법률안 검토 보고서와의 병합

### 1. ‘날짜’와 ‘위원회’로 연결

#### 1) 공통되는 기간의 범위 설정

- ▶ 국회회의록 날짜 범위: 1996-07-12 00:00:00 ~ 2023-12-06 00:00:00
- ▶ 법률안검토보고서 날짜 범위: 2015-01-12 00:00:00 ~ 2023-07-27 00:00:00
- ▶ 공통 날짜 범위(뉴스 데이터 기간과 통일): 2015-01-12 00:00:00 ~ 2022-07-01 00:00:00

#### 2) 위원회 이름 형식 통일

	committee_name	main_committee	sub_committee
29348	환경노동위원회-고용노동법안심사소위원회	환경노동위원회	고용노동법안심사소위원회
29349	환경노동위원회-고용노동법안심사소위원회	환경노동위원회	고용노동법안심사소위원회
29350	환경노동위원회-고용노동법안심사소위원회	환경노동위원회	고용노동법안심사소위원회
29351	환경노동위원회-고용노동법안심사소위원회	환경노동위원회	고용노동법안심사소위원회
29352	환경노동위원회-고용노동법안심사소위원회	환경노동위원회	고용노동법안심사소위원회
29353	환경노동위원회-고용노동법안심사소위원회	환경노동위원회	고용노동법안심사소위원회
29354	환경노동위원회-환경법안심사소위원회	환경노동위원회	환경법안심사소위원회
29355	환경노동위원회-환경법안심사소위원회	환경노동위원회	환경법안심사소위원회
29356	환경노동위원회-환경법안심사소위원회	환경노동위원회	환경법안심사소위원회
29357	환경노동위원회-환경법안심사소위원회	환경노동위원회	환경법안심사소위원회
29358	환경노동위원회-환경법안심사소위원회	환경노동위원회	환경법안심사소위원회

- ▶ 법률안 검토 보고서 데이터의 `committee`는 ‘위원회’ 형식으로 코딩되어 있는 반면, 국회 회의록 데이터에는 ‘위원회 - 소위원회’ 등의 형식 다수 존재
- ▶ 따라서 ‘committee\_name’을 ‘main\_’과 ‘sub\_’로 분리하여 ‘main\_committee’를 기준으로 법률안과 연결

## 국회 회의록 데이터 정제 및 법률안 검토 보고서와의 병합

### 2. 법률안 검토보고서와 연관된 국회 회의록 검색 함수

#### 1) 법률안 검토보고서 데이터 내 unique한 ‘bill\_id’를 기준으로 검색

- ▶ 동일한 법률안(‘field’ 및 ‘title’)을 대상으로 한 검토 보고서가 다수, 따라서 특정 법률안 검토보고서(단일 날짜)를 기준으로 필터링
- ▶ 해당 보고서와 직접적으로 연관된 회의 내용(맥락)만을 고려하기 위한 제약조건 설정

#### 2) 사용자가 지정하는 기간(default; days = 30)을 기준으로 날짜 범위로 1차 필터링

- ▶ 예시) days = 30이면, 법률안 검토보고서 등록날짜 전후 30일에 이루어진 회의록 중 탐색

#### 3) 법률안 검토보고서의 담당위원회의 회의록으로 2차 필터링

- ▶ 날짜로 필터링된 행 중, 법률안 검토보고서의 ‘committee’와 회의록의 ‘main\_committee’가 동일한 행을 추출하여,  
  
해당 법률안 검토보고서 논의와 가장 연관될 ‘담당위원회’의 회의 내용을 선별
- ▶ 단, 담당위원회에서 논의된 회의 기록이 없을 경우, 메시지와 함께 해당 날짜 범위에 있는 행을 출력
- ▶ 추가로, 지정된 기간에 논의된 회의 기록이 없을 경우, 기간을 1.5배 늘려 재검색하는 함수도 설계

국회 회의록 데이터 정제 및 법률안 검토 보고서와의 병합

2. 법률안 검토보고서와 연관된 국회 회의록 검색 함수

함수 사용 예시

```
# 사용 예시:
bill_id = 'PRC_M2K0C0Y7W0V7R1Q4Z4L3M0T8A9S4H5' # 확인하고 싶은 bill_id
indices, matching_rows = get_indices_within_days_by_bill_id_extend(bill_id, law_df, sub_wrapped)

# 일치하는 행들을 출력
matching_rows
```

✓ 0.0s

	date	meeting_name	main_committee	agenda	agenda_index	context_full
2756	2020-09-18	소위원회	행정안전위원회	2. 자연재해대책법 일부개정법률안(정부 제출)	50222-1	['본부장님, 방재신기술은 누가 인증을 하는 겁니까? 방재신기술은 행안부에서 인증을...
2844	2020-10-13	국정감사	행정안전위원회	감사개시	50468-1	['대구 달서구병 김용판 위원입니다. 국민의 생명과 재산을 지켜주는 119는 골든타...
2845	2020-10-13	국정감사	행정안전위원회	감사개시	50470-1	['서울 성북갑의 김영배입니다. 새마을운동의 정성현 이사장님께 몇 가지 여쭙겠습니다...
2877	2020-10-15	국정감사	행정안전위원회	감사개시	50522-1	['수고들 많으십니다. 서울청에서는 안전한 지하철 환경 조성하기 위해서 지하철에서 ...
2909	2020-10-19	국정감사	행정안전위원회	감사개시	50576-1	['남부청장님, 전자발찌 그것 믿을 수 있습니까? 일단은 관리 주무부서가 법무부로 ...
2948	2020-10-20	국정감사	행정안전위원회	감사개시	50639-1	['서울 강동을 출신 더불어민주당 이해식입니다. 네 분 지사님들 이렇게 한꺼번에 뵙...
2971	2020-10-22	국정감사	행정안전위원회	감사개시	50685-1	['그다음 그래프 보여 주시지요. (영상자료를 보며) 좀 전에 인터뷰에도 나왔지만 ...
2974	2020-10-23	국정감사	행정안전위원회	감사개시	50697-1	['네 분 청장님들 반갑습니다. 국감 준비하시느라고 수고하셨고요. 간단한 것 조금 ...
2975	2020-10-23	국정감사	행정안전위원회	감사개시	50698-1	['경기 광명읍의 양기대 위원입니다. 세 분 청장님들 늘 치안과 코로나19 방역으로...

국회 회의록 데이터 정제 및 법률안 검토 보고서와의 병합

3. 최종 데이터프레임(소위원회) - bill\_id별 연관 소위원회 회의록 추출

- related\_agenda\_index : 관련된 회의록의 agenda\_index를 리스트로 저장
- unique\_agendas : 관련된 회의록의 고유한 agenda 명을 리스트로 저장
- directly\_related : 담당위원회의 회의록인지 여부를 확인해 직접적으로 관련된 회의록이면 1, 아니면 0으로 표시

	bill_id	date	id	title	directly_related	related_agenda_index	unique_agendas
0	PRC_M2K0C0Y7W0V7R1Q4Z4L3M0T8A9S4H5	2020-09-24	1040_1594	119구조·구급에 관한 법률 일부개정법률안(김승수의원 등 14인)	1	[50222-1, 50468-1, 50470-1, 50522-1, 50576-1, ...]	[2. 자연재해대책법 일부개정법률안(정부 제출), 감사개시]
1	PRC_X2G0B0M7K0W1Y1Y1S2Y1Y2B0T4O5X1	2020-09-24	1040_1648	119구조·구급에 관한 법률 일부개정법률안(박덕흠의원 등 12인)	1	[50222-1, 50468-1, 50470-1, 50522-1, 50576-1, ...]	[2. 자연재해대책법 일부개정법률안(정부 제출), 감사개시]
2	PRC_X2G0B0M7K0W1Y1Y1S2Y1Y2B0T4O5X1	2020-09-24	1040_1648	119구조·구급에 관한 법률 일부개정법률안(박덕흠의원 등 12인)	1	[50222-1, 50468-1, 50470-1, 50522-1, 50576-1, ...]	[2. 자연재해대책법 일부개정법률안(정부 제출), 감사개시]
3	PRC_O2R0O0B8L3J1D1V1T5T1X5T5X1U6O4	2020-12-09	1040_924	119구조·구급에 관한 법률 일부개정법률안(박완주의원 등 19인)	1	[50424-1, 50457-1, 50461-1, 50498-1, 50498-2, ...]	[1. 2021년도 예산안(계속)가. 중앙선거관리위원회 소관 나. 소방청 소관다. 경...
4	PRC_O2R0O0B8L3J1D1V1T5T1X5T5X1U6O4	2020-12-09	1040_924	119구조·구급에 관한 법률 일부개정법률안(박완주의원 등 19인)	1	[50424-1, 50457-1, 50461-1, 50498-1, 50498-2, ...]	[1. 2021년도 예산안(계속)가. 중앙선거관리위원회 소관 나. 소방청 소관다. 경...
...	...	...	...	...	...	...	...
16235	PRC_X1M6A0C8Y2S9K1I9G2R1H5Y3G0O6T9	2017-03-30	1039_10822	후천성면역결핍증 예방법 일부개정법률안(최도자의원 등 10인)	1	[47094-1, 47096-1, 47096-2, 47096-3, 47096-4, ...]	[1. 국민건강보험법 일부개정법률안(윤소하 의원 대표발의)(의안번호 521)(계속)...
16236	ARC_V1Z4Y1O2J2S9J1K7A0I4N0O3Q8E3L0	2016-03-03	1040_9254	휴면예금관리재단의 설립 등에 관한 법률 전부개정법률안	0	[46090-1, 46091-1, 46092-1, 46093-1, 46094-1, ...]	[1. 군인사법 일부개정법률안(안규백 의원 대표발의) 2. 군인사법 일부개정법률안(...
16237	PRC_Q2R0D0P9Z1H5V1O3U0N8M3P9B9S1G8	2023-05-25	1040_723	희귀질환관리법 일부개정법률안(강훈식 의원 등 11인)	0	[]	[]
16238	PRC_Q2R0D0P9Z1H5V1O3U0N8M3P9B9S1G8	2023-05-25	1040_723	희귀질환관리법 일부개정법률안(강훈식 의원 등 11인)	0	[]	[]
16239	PRC_E1A8U1C2B2N8P1B1E4A1Z3N2E9I7W2	2019-04-05	1039_6731	희귀질환관리법 일부개정법률안(최도자의원 등 10인)	1	[49046-1, 49046-2, 49048-1, 49048-2, 49048-3, ...]	[1. 의료법 일부개정법률안(박인숙 의원 대표발의)(의안번호 14397)(계속) 2...

16240 rows × 7 columns



국회 회의록 데이터 정제 및 법률안 검토 보고서와의 병합

4. 최종 데이터프레임(본회의) - 법률안명, ‘terminology’를 이용해 법률안 검토 보고서 - 본회의 발화와 연결

▶ 법률안 검토 보고서의 ‘terminology’를 이용해 안건을 파악하고, 해당 안건을 다루고 있는 회의록의 발화와 연결 (`matching\_agenda\_index` column 정의)

id	term	session	date	law	result	text	terminology	token_freq	matching_agenda_index
1040_656	21	384.0	2021-02-24	국회법 일부개정법률안	철회	김남국의원안 제40조의2는 상임위원의 영리행위 등 금지행위를 그 유형별로 상세히 규...	상임위원, 영리행위, 공직자윤리법, 이해충돌 방지 의무, 국회법, 의정활동	{‘금지’: 14, ‘행위’: 8, ‘상임’: 7, ‘직무’: 5, ‘재산’: 5,...}	45257-1
1039_11096	21	378.0	2020-05-29	국립묘지의 설치 및 운영에 관한 법률 일부개정법률안	임기만료폐기	(1) 개정안은 「일제강점하 반민족행위 진상규명에 관한 특별법」에 따라 친일 반민족행...	일제강점하 반민족행위 진상규명에 관한 특별법, 친일반민족행위자, 국립 묘지 안장, 국...	{‘안장’: 18, ‘국립묘지’: 12, ‘민족’: 10, ‘사람’: 8, ‘친일’...	46180-1
1039_9909	21	378.0	2020-05-29	보조금 관리에 관한 법률 일부개정법률안	임기만료폐기	현행법은 지방자치단체에 대한 보조금은 대상사업의 범위와 기준보조율을 대통령령으로 정...	지방자치단체, 보조금, 기준보조율, 대통령령, 국회, 헌법	{‘보조’: 6, ‘확정’: 5, ‘지방자치단체’: 3, ‘율’: 3, ‘정’: 3...}	44926-1
1037_1368	21	378.0	2020-05-29	국유재산법 일부개정법률안	임기만료폐기	현행 국유재산법 제57조는 일반재산을 국유재산관리기금의 운용계획에 따라 국유재산관리...	국유재산법, 국유재산관리기금, 행정재산, 일반재산, 용도폐지	{‘국유재산’: 6, ‘재산’: 5, ‘행정재산’: 5, ‘개발’: 3, ‘용도’...	52147-1
1039_1931	21	396.0	2022-04-30	공소청법안	대안반영폐기	검찰제도는 수사·소송·재판과 관련한 법치국가적 통제장치로서의 성격을 가지고 있다고 ...	검찰제도, 수사권, 기소권, 형사소송제도, 영미법, 대륙법, 소추권한, 형사 절차체계	{‘수사’: 41, ‘경찰’: 36, ‘검사’: 35, ‘사법’: 13, ‘범죄’...	49409-1
...	...	...	...	...	...	...	...	...	...
1039_6229	21	378.0	2020-05-29	의료법 일부개정법률안	임기만료폐기	최근 강북삼성병원 의사 사망사건 등 의료인 폭행 사고가 사회적 문제로 대두됨에 따라...	보건복지부장관, 진료환경 안전, 실태조사, 정책 수립 기반을 마련, 보건의 료기본법,...	{‘실사’: 17, ‘진료’: 9, ‘환경’: 8, ‘의료’: 7, ‘정책’: 6,...}	49896-1
1037_1073	20	358.0	2018-03-30	산업안전보건법 일부개정법률안	대안반영폐기	개정안은 안 제26조의2제2항과 제3항을 위반한 자에게 500만원 이하의 과태료를 ...	500만원 이하, 과태료, 산업안전보건법, 사용자, 사업주	{‘과태료’: 7, ‘부과’: 4, ‘것임’: 3, ‘안전’: 3, ‘보 건’: 3,...}	46187-1
1037_2868	21	405.0	2023-04-27	의료법 일부개정법률안	대안반영폐기	개정안은 이러한 문제의식에서, 평가인증기구의 인증을 받기 전 신실 간호 학 교육과정에...	간호학과, 교육과정, 보건복지부장관, 응시자격, 교육부 소관 사항, 교육부	{‘인증’: 12, ‘신실’: 9, ‘교육과정’: 7, ‘평가’: 5, ‘입 학’: ...}	46416-1
1039_2723	21	378.0	2020-05-29	공무원의 노동조합 설립 및 운영 등에 관한 법률 일부개정법률안	임기만료폐기	□ 국공립대학 조교는 「교육공무원법」에 따른 교육공무원이고, 교육공무원은 「국가공무...	교육공무원법, 국가공무원법, 지방공무원법, 공무원의 노동조합 설립 및 운영 등에 관...	{‘공무원’: 43, ‘교원’: 21, ‘노조’: 20, ‘직’: 15, ‘대 학’...	46959-1
1040_292	21	406.0	2023-05-25	회귀질환관리법 일부개정법률안	대안반영폐기	가. 개정안의 내용\r\n□ 개정안은 현행 매년 5월 23일로 지정되어 있는 ‘회귀...	회귀질환 극복의 날, 지방자치단체, 조례, 회귀질환지원센터, 세계 회귀질 환의 날, ...	{‘회귀질환’: 61, ‘센터’: 15, ‘유전상담’: 14, ‘날’: 12, ‘조’...	49025-1

# 보완 방향성 및 계획

## 1. 분석에 활용할 데이터와 변수 보완/추가

- ▶ 법률 조항, 법률안, 법안, 회의록 등 추가 데이터를 확보해 활용
- ▶ 텍스트 형태 데이터에서 feature 추출

예시) 국회 회의록 발화 긍/부정 감성 분석, ‘소위원회 - 본회의’ 회의록 간의 유사도 점수 산출 등

## 2. 모델 비교/평가를 통한 성능 개선

- ▶ 추가 데이터 및 변수와 함께 Random Forest 이외의 여러 분류/예측 모델 시도

## 3. 법률안 가결/부결에 영향을 미치는 변수 파악, 중요도 분석 등을 통한 설명력 향상

- ▶ 예측력/정확도 뿐만 아니라 feature 간의 관계나 특성을 포착하여 모델의 설명력 확보

# 감사합니다.

## Bill Gates: 법률안의 문

권민석 고려대학교 산업경영공학부

배지원 고려대학교 경제학과

정종락 고려대학교 일반대학원 정치외교학과

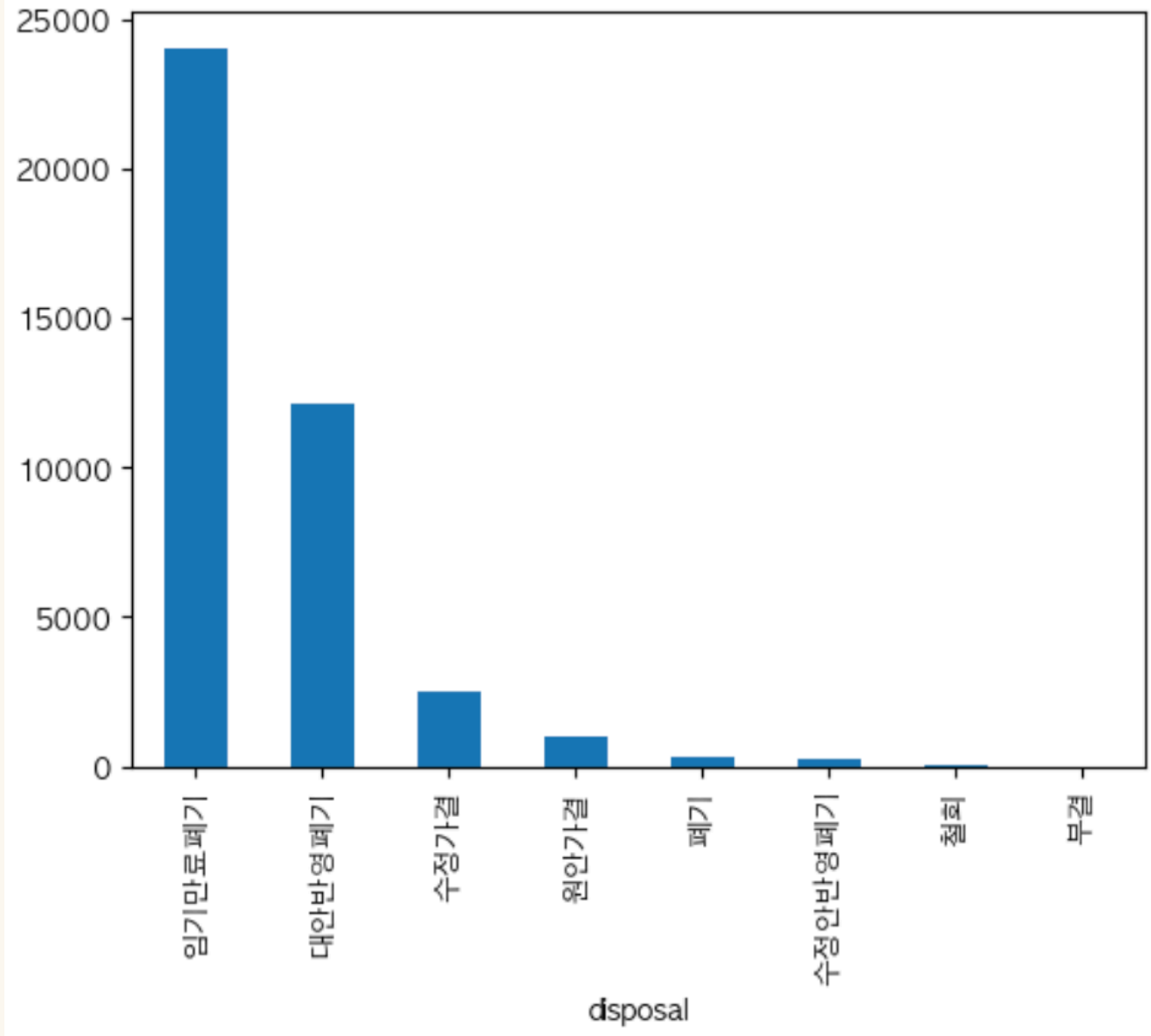
정하연 고려대학교 경영대학



2024 K-DS 해커톤

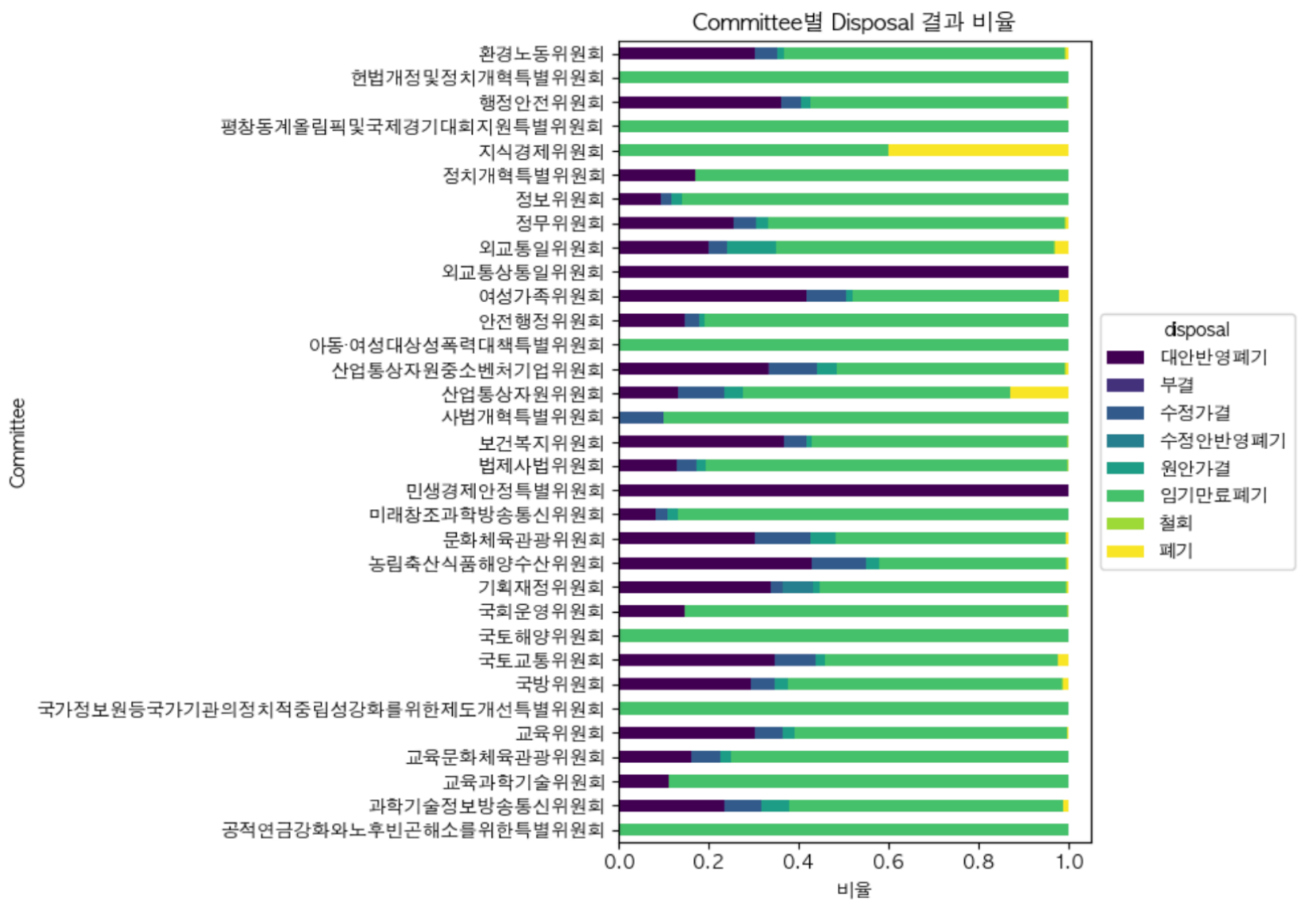


법률안 검토 보고서 가결 여부 분포



위원회별 법률안 검토 보고서 가결 여부 테이블 및 시각화

	disposal	대안반영폐기	부결	수정가결	수정안반영폐기	원안가결	임기만료폐기	철회	폐기
committee									
공적연금강화와노후빈곤해소를위한특별위원회	0	0	0	0	0	0	4	0	0
과학기술정보방송통신위원회	254	0	87	0	67	656	0	13	
교육과학기술위원회	1	0	0	0	0	8	0	0	
교육문화체육관광위원회	186	0	74	0	27	861	0	0	
교육위원회	289	0	58	0	25	578	2	2	
국가정보원등국가기관의정치적중립성강화를위한제도개선특별위원회	0	0	0	0	0	2	0	0	
국방위원회	221	0	38	0	22	458	2	9	
국토교통위원회	1730	0	458	0	116	2595	1	116	
국토해양위원회	0	0	0	0	0	2	0	0	
국회운영위원회	114	0	2	0	0	670	2	0	
기획재정위원회	1157	0	87	231	57	1873	8	14	
농림축산식품해양수산위원회	1348	0	384	0	95	1307	3	13	
문화체육관광위원회	269	0	112	0	49	459	1	4	
미래창조과학방송통신위원회	31	0	11	0	8	335	0	0	
민생경제안정특별위원회	18	0	0	0	0	0	0	0	
법제사법위원회	348	0	122	0	55	2180	4	1	
보건복지위원회	2034	0	266	0	80	3146	8	0	
사법개혁특별위원회	0	0	4	0	0	37	0	0	
산업통상자원위원회	64	0	52	0	19	293	0	64	
산업통상자원중소벤처기업위원회	643	3	208	0	84	981	8	11	
아동·여성대상성폭력대책특별위원회	0	0	0	0	0	13	0	0	
안전행정위원회	201	0	48	0	15	1124	1	0	
여성가족위원회	213	0	44	0	8	234	0	11	
외교통상통일위원회	2	0	0	0	0	0	0	0	
외교통일위원회	125	0	27	0	69	390	2	18	
정무위원회	627	1	124	0	66	1626	4	15	
정보위원회	4	0	1	0	1	37	0	0	
정치개혁특별위원회	37	0	0	0	0	182	0	0	
지식경제위원회	0	0	0	0	0	3	0	2	
평창동계올림픽및국제경기대회지원특별위원회	0	0	0	0	0	1	0	0	
행정안전위원회	1302	0	157	0	72	2066	6	0	
헌법개정및정치개혁특별위원회	0	0	0	0	0	15	0	0	
환경노동위원회	925	0	147	0	48	1906	4	19	



## 국회회의록 데이터 재구조화 상세 코드

```
# 'agenda'의 첫 문자가 같은 행을 그룹으로 묶기 위한 함수
def get_agenda_group(agenda):
    return agenda.split(".")[0]

# 'agenda_group' 생성
parliament_df['agenda_group'] = parliament_df['agenda'].apply(get_agenda_group)

# 그룹화하여 context와 answer.comment를 각각 리스트로 결합
parliament_wrap = parliament_df.groupby(['id', 'agenda_group']).agg({
    'context': lambda x: list(x), # context를 리스트로 결합
    'answer.comment': lambda x: list(x), # answer.comment를 리스트로 결합
    'agenda': 'first',
    'date': 'first',
    'meeting_name': 'first',
    'main_committee': 'first'
}).reset_index()

# context와 answer.comment를 리스트로 결합하여 context_full 생성
parliament_wrap['context_full'] = parliament_wrap.apply(lambda row: row['context'] + row['answer.comment'], axis=1)

# agenda_index 부여: id 별로 1, 2, 3 순서로 인덱스 생성
parliament_wrap['agenda_index'] = parliament_wrap.groupby('id').cumcount() + 1
parliament_wrap['agenda_index'] = parliament_wrap['id'].astype(str) + '-' + parliament_wrap['agenda_index'].astype(str)

# 필요한 칼럼만 선택
parliament_wrap = parliament_wrap[['id', 'date', 'meeting_name', 'main_committee', 'agenda', 'agenda_index', 'context_full']]
```

## 국회 회의록 데이터 정제 및 법률안 검토 보고서와의 병합

\* ‘날짜’와 ‘위원회’로 연결

### 1) 공통 날짜 범위 설정

```
# 날짜 범위 추출
par_start = partial_par['date'].min()
par_end = partial_par['date'].max()

law_data_start = lawdata['date'].min()
law_data_end = lawdata['date'].max()

# 공통 범위 계산
common_start = max(par_start, law_data_start)
common_end = min(par_end, law_data_end)
```

### 2) 위원회명 형식 통일

```
# '-'를 기준으로 쪼개서 두 개의 새로운 칼럼을 만들
# n=1; 최대 한 번만 -를 기준으로 문자열을 나누겠다는 의미
parliament_df[['main_committee', 'sub_committee']] = parliament_df['committee_name'].str.split('-', n=1, expand=True).copy()
```

## 법률안 검토보고서와 연관된 국회 회의록 검색 함수 1

```
# 특정 법안의 bill_id에 따라 30일 내외(혹은 사용자 지정 범위)의 날짜로 필터링한 후, 'committee'가 같은지 확인하는 함수
def get_indices_within_days_by_bill_id(bill_id, law_df, sub_wrapped, days=30):
    # 입력된 bill_id에 해당하는 law_df 행의 'date'와 'committee' 값 가져오기
    law_row = law_df.loc[law_df['bill_id'] == bill_id]
    law_date = law_row['date'].values[0]
    law_committee = law_row['committee'].values[0]

    # 'date'가 law_date 기준으로 지정된 범위 내에 있는 행들 필터링
    mask = (sub_wrapped['date'] >= law_date - pd.Timedelta(days=days)) & (sub_wrapped['date'] <= law_date + pd.Timedelta(days=days))
    sub_filtered = sub_wrapped[mask]

    # 필터링된 행 중에서 'committee'가 같은 행을 다시 필터링
    sub_committee_filtered = sub_filtered[sub_filtered['main_committee'] == law_committee]

    # 'committee'가 같은 데이터가 없으면 메시지를 출력하고 날짜 범위에 해당하는 행만 반환
    if sub_committee_filtered.empty:
        print(f"담당 위원회에서 전후 {days}일 안에 논의된 기록이 없습니다.")
        return sub_filtered.index.tolist(), sub_filtered
    else:
        return sub_committee_filtered.index.tolist(), sub_committee_filtered

# 사용 예시:
# bill_id = 'want_to_find' # 확인하고 싶은 bill_id
# indices, matching_rows = get_indices_within_days_by_bill_id(bill_id, law_df, sub_wrapped)
```

## 법률안 검토보고서와 연관된 국회 회의록 검색 함수 2 - 연장된 범위 내 재탐색

```
# 특정 법안의 bill_id에 따라 30일 내외(혹은 사용자 지정 범위)의 날짜로 필터링한 후, 'committee'가 같은지 확인하는 함수
def get_indices_within_days_by_bill_id(bill_id, law_df, sub_wrapped, days=30):
    # 입력된 bill_id에 해당하는 law_df 행의 'date'와 'committee' 값 가져오기
    law_row = law_df.loc[law_df['bill_id'] == bill_id]
    law_date = law_row['date'].values[0]
    law_committee = law_row['committee'].values[0]

    # 'date'가 law_date 기준으로 지정된 범위 내에 있는 행들 필터링
    mask = (sub_wrapped['date'] >= law_date - pd.Timedelta(days=days)) & (sub_wrapped['date'] <= law_date + pd.Timedelta(days=days))
    sub_filtered = sub_wrapped[mask]

    # 필터링된 행 중에서 'committee'가 같은 행을 다시 필터링
    sub_committee_filtered = sub_filtered[sub_filtered['main_committee'] == law_committee]

    # 'committee'가 같은 데이터가 없으면 메시지를 출력하고 날짜 범위에 해당하는 행만 반환
    if sub_committee_filtered.empty:
        print(f"담당 위원회에서 전후 {days}일 안에 논의된 기록이 없습니다.")
        return sub_filtered.index.tolist(), sub_filtered
    else:
        return sub_committee_filtered.index.tolist(), sub_committee_filtered

# 사용 예시:
# bill_id = 'want_to_find' # 확인하고 싶은 bill_id
# indices, matching_rows = get_indices_within_days_by_bill_id(bill_id, law_df, sub_wrapped)
```

## bill\_id 별 연관 소위원회 회의록 최종 데이터 프레임 EDA

