

자체연구 2023-07

생성형 AI를 활용한 법령사무조사 자동화 기초 연구

주희진 · 최인수 · 전대욱 · 박재희

Korea
Research
Institute for
Local
Administration



저 자 주희진, 최인수, 전대욱, 박재희

연구책임자 주희진 | 한국지방행정연구원 부연구위원

연 구 진 최인수 | 한국지방행정연구원 선임연구위원

전대욱 | 한국지방행정연구원 연구위원

박재희 | 한국지방행정연구원 부연구위원

연 구 협 력 이광춘 | (주)씨지인사이드

김태훈 | KPMG

고유정 | 서울대학교 빅데이터 핀테크 과정

노희찬 | 서울대학교 빅데이터 핀테크 과정

이홍주 | 서울대학교 빅데이터 핀테크 과정

이화정 | 서울대학교 빅데이터 핀테크 과정

임동건 | 서울대학교 빅데이터 핀테크 과정

홍수민 | 서울대학교 빅데이터 핀테크 과정

연구요약

□ 연구의 배경 및 필요성

- 효율적인 사무의 수행을 위해 국가-지방사무의 배분 및 사무수행체계 현황을 파악하기 위한 노력의 일환으로 1994년부터 약 5년 단위로 한국지방행정 연구원을 중심으로 「법령상 사무총조사」가 진행되어 왔음
- 그러나 기존의 방식은 연구자의 주관적 판단에 의존하는 방식을 취하고 있어 과도하게 많은 시간과 비용의 투입, 조사 결과의 일관성·신뢰성에 대한 비판 우려 등 한계에 직면하고 있음
- 데이터 기반 행정의 대두와 윤석열 정부의 디지털플랫폼정부의 국정 기조 하에 본 연구는 생성형 AI를 활용하여 보다 효율적이고 일관성 있는 법령사무조사를 수행하기 위한 자동화 모델 개발을 목적으로 함
- 생성형 AI 전문기관을 통한 법령사무조사 자동화 모델을 개발 및 향후 생성형 AI를 활용한 법령상 사무총조사의 수행 가능성 등의 미래 연구 방향을 제시함

□ 연구의 목적 및 방법

- 본 연구의 목적은 생성형 AI를 활용한 법령사무조사 자동화 로직 등 모델을 개발하여 2022.5.1. 기준 중앙행정기관이 소관하는 법령 중 양행정기관, 지방자치단체를 사무수행 주체로 규정하는 법령을 대상으로 사무추출하여 2022년 「법령상 사무총조사」의 결과 및 자동화 모델을 통해 추출된 사무배분 결과를 비교함으로써 자동화 모델을 활용한 「법령사무조사」의 가능성 검토 및 미래 연구 방향을 제시하고 함
- 이를 위하여 본 연구는 2022.5.1. 기준 중앙행정기관이 소관하는 전체 법령 중 중앙행정기관, 지방자치단체를 사무수행 주체로 규정하는 4,339개의 법령을

대상으로 ① 생성형 AI 전문기관에의 위탁을 통한 법령사무조사 자동화 모델 개발, ② 자동화 모델의 예측 정확성 검증, ③ 2022년 5월 기준 법령 데이터에 대한 검증, ④ 2022년 5월 이후 법령 데이터에 대한 적용 및 예측, ⑤ 법령사무 조사에의 자동화 프로그램 적용 가능성 검토 등으로 구성됨

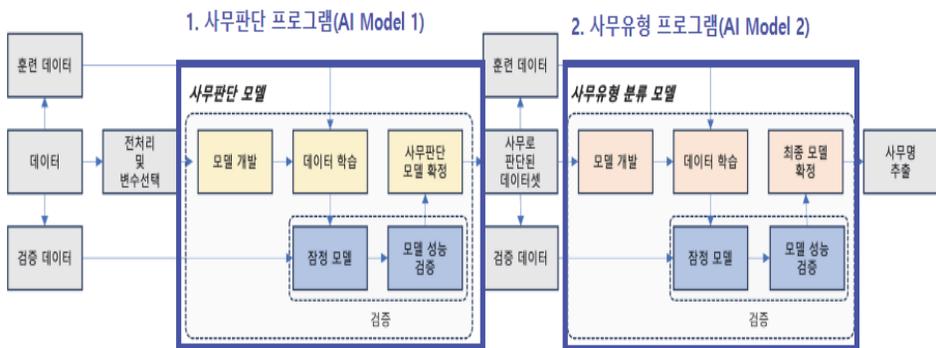
□ 법령상 사무 및 법령상 사무총조사

- 사무유형은 법령상 권한 및 책임이 귀속되는 주체를 기준으로 국가사무, 국가와 지방자치단체의 공동사무, 자치사무로 유형화할 수 있으며, 다시 16가지로 세분화하여 구분됨
 - (국가사무) 국가직접처리사무, 특별지방행정기관사무, 국가위탁사무, 시·도위임사무, 시·군·구위임사무로 구분함
 - (국가-지방자치단체 공동사무) 국가-시도-시군구 공동사무, 국가-시도 공동사무, 국가-시군구 공동사무로 구분함
 - (자치사무) 자치사무는 시·도직접처리사무, 시·도위탁사무, 시·군·구직접처리사무, 시·군·구위탁사무, 시·도-시·군·구 공동사무, 시·도-시·군·구 위임사무로 구분함
- 한국지방행정연구원은 국가와 지방자치단체간의 사무배분 현황을 파악하기 위한 기본통계 작성 및 사무정보관리시스템 구축을 위하여 「법령상사무총조사」를 수행하고 있음
 - 2022년 「법령상 사무총조사」에서 단위사무는 “모든 법의 각 조·항·호와 이와 연계된 시행령과 규칙의 각 조·항·호를 상호 연계하여, 법령상 권한과 책임 및 의무가 발생하는 처리권자가 행정목적 달성을 위하여 수행해야 하는 최소 단위의 사무”로 정의함
 - 2022.5.1. 기준 시행 중인 4,339개 법령(법 1,443개, 시행령 1,675개, 시행규칙 1,221개)으로부터 총 60,114개의 단위사무 추출

□ 법령사무조사 자동화 모델의 개발

- 법령사무조사 자동화 모델 개발은 프로그래밍 전문기관인 (주)씨지인사이드에 위탁 및 KPMG와의 업무협력을 통하여 자동화 모델 개발자 7명이 2023년 11월 27일부터 12월 27일까지 30일까지 개발에 착수하였음
- 생성형 AI를 활용한 법령사무조사 자동화 기초연구를 위한 개발 과정은 다음과 같음
 - 수집된 데이터를 바탕으로 전처리, 변수선택, 개발, 검증 및 단계를 통해 이루어짐. 데이터를 전처리하여 분석 가능한 형태인 정형 데이터로 가공한 뒤, 모델 개발 과정을 거침
 - 먼저, 사무판단 모델을 개발함. 학습을 통해 모델을 도출하며 데이터 학습으로 도출한 잠정 모델의 성능을 검증하고, 최종 모델을 선정함
 - 다음으로, 사무유형 분류 모델을 개발함
 - 이 과정을 통해서 도출된 사무라고 판단된 데이터셋만을 이용하여 위와 동일한 과정을 반복함
 - 최종적으로 모델이 확정되면, 해당 조문에 대한 사무명을 추출함

[그림] 법령사무조사 자동화 모델 개발 절차



- 사무판단모델(AI Model 1)은 Transformer 모델을 활용하여 조문의 텍스트를 토큰화하여 숫자로 변경한 후 문장 안의 맥락과 단어에 담긴 의미를 반영하여 Transformer Encoder 과정을 거침
 - 본 모델을 통하여 사무와 비사무 여부를 판단할 수 있으며, 이때 추출된 비사무는 인간의 판단 영역에서 1차적으로 제외될 수 있기 때문에 정확도가 가장 중요한 항목임
 - 비사무를 사무로 판단하는 경우, 세부적인 사무 판단과정에서 오류를 시정할 수 있는 기회가 제공될 수 있지만, 사무를 비사무로 판단하는 경우 추가적 검증 과정의 기회가 원천적으로 배제될 수 있음
 - 따라서 사무로 추출되는 정확도보다 비사무를 추출하는 정확도가 훨씬 더 중요하다고 판단하였으며, 본 과정을 수행할 수 있는 여러 모델 중 가장 적합한 모델(2-layer 신경망 with Transformer Encoder)을 선정함
- 사무유형모델(AI Model 3)은 랜덤포레스트와 Transformer Encoder를 적용한 인공지능망 모델을 결합하여 하나의 새로운 모델(양상블 방식)을 개발하여 제시함
 - 양상블 방식은 각 모델별 클래스 예측값을 조합하여, 두 모델이 모두 같은 클래스로 예측하면 해당 예측값과, '확실'이라는 지표를 함께 나타냄
 - 두 모델이 다른 클래스로 예측하면 해당 예측값과, '애매'라는 지표를 함께 나타냄

□ 자동화 모델의 예측 정확성 검증

- 2layer-인공신경망(transformer 인코더 적용) 모델을 적용한 사무판단모델(AI Model 1)은 타 모델에 비하여 가장 높은 재현율인 94.9%의 재현율을 나타냄
 - 사무판단 모델(AI Model 1)에서는 실제로 사무 처리가 필요한 조문을 놓치지 않고 정확하게 식별하는 것이 중요함
- 사무유형모델(AI Model 2)의 경우, 본 연구에서 사용한 양상블 모델은 타 모델에 비하여 가장 높은 정확도를 나타내며, 81%의 정확도를 나타냄

- 사무유형 분류 모델은 사무가 아닌지, 국가사무인지, 지방사무인지, 공동 사무인지 맞는 모델로 유형의 분류를 잘 맞추는가에 초점을 두었음

□ 법령사무조사 자동화 모델을 적용한 데이터 오류 검증

- 2022년 법령상 사무총조사에서 사용된 법령은 총 4,357개(법 1,451개, 시행령 1,683개, 시행규칙 1,123개)로 사무판단이 필요한 행은 총 848,072줄로 이루어짐
 - 조사 과정에서 발생하는 각종 오류에 대한 면밀한 검증이 필요하지만, 방대한 양으로 인하여 조사뿐만 아니라 검증에 대한 투입 비용 역시 상당한 수준으로 요구되는 실정임
 - 인공지능을 통한 자동화 모델로 하여금 학습된 내용을 바탕으로 2022년 법령상 사무총조사의 데이터 중 오류 사항을 인식하도록 명령한 후, 해당 오류 데이터가 학습 알고리즘의 미흡으로 인한 오류인지 혹은 인간의 입력 오류인지 등을 판단·검증할 수 있음
- 전체 사무판단이 필요한 844,220개의 행 중에서 인공지능이 오류가 발생했다고 추출한 데이터는 총 14,546개로 전체의 1.723%에 해당함
 - 이 중 연구진의 입력 오류에 의한 데이터 오류는 79개(0.009%)이며, AI의 학습 미흡에 의한 데이터 오류는 14,467개(1.714%)의 비중으로 나타남
 - 다만, 본 연구에서는 사무수행 주체가 상이한 특별지방행정기관의 경우는 삭제(3,852개)하고 분석하였기 때문에 844,220개의 행에 대해서만 프로그래밍을 진행함
- 인공지능이 오류로 인식한 데이터 중 99%는 인공지능에 대한 학습 미흡에 의한 것이기 때문에 이에 대해서는 후속 연구가 필요할 것임

□ 2022년 법령 데이터에 대한 자동화 모델의 적용 및 활용 : 대시보드의 제시

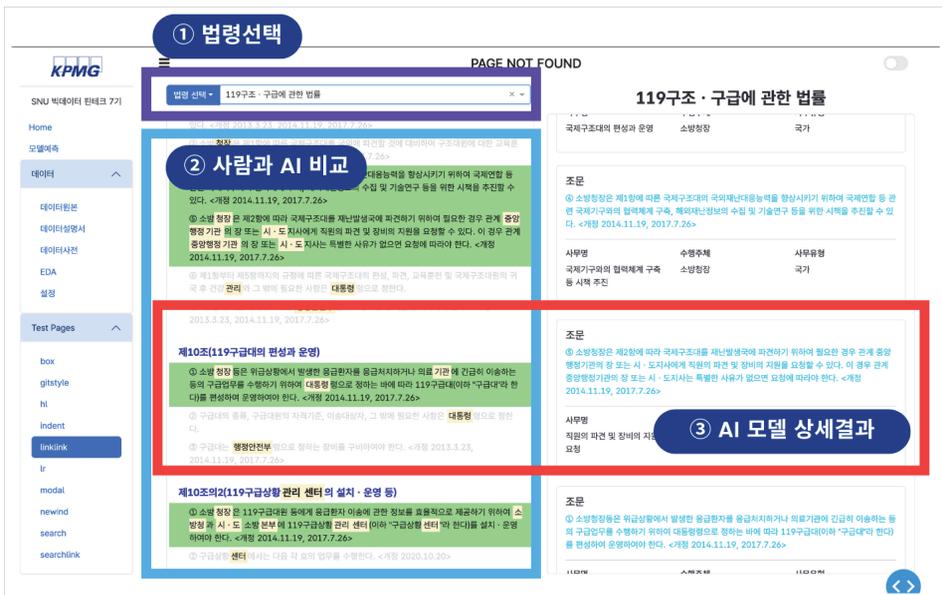
- 자동화 모델의 결과를 일반인도 손쉽게 활용할 수 있는 방안으로서 이를 활용한 대시보드를 예시로 제시하고자 함

- 인공지능을 활용한 법령사무조사 자동화 모델은 복잡한 명령어를 기반으로 한 모델 로직이기 때문에 이를 일반인이 실생활·실무에 적용하는 것은 사실상 불가능함
- 사람이 직접 조사한 결과와 AI가 판단한 내용을 비교하는 형식으로 나타내 주어 AI가 어떤 모델을 잘 분류하는지 분류를 잘 하지 못하는지 바로 파악 가능한 대시보드를 구상함

○ 본 연구에서 샘플로 제시한 대시보드를 통한 법령사무조사 자동화 모델은 사용자가 원하는 법령을 선택하면, 해당 법령의 모든 조문에 대하여 사람(연구자)의 주관적 판단에 의한 결과치와 AI에 의한 결과치가 색깔로 제시됨

- 동시에 왼쪽의 상세결과로서 AI 모델이 해당 조문에 대한 사무판단 여부 및 예측 확률, 사무명, 수행주체, 사무유형 등의 세부 정보를 제시함

[그림] 대시보드의 전체 구성



[그림] AI모델의 상세결과

사무 유형	확률				AI 판단	사무명	수행주체	실제 정답
	비사무	국가	지방	공동				
국가	0.2	0.6	0.1	0.1	확실	국제기구와의협력체계 구축	소방청장 시도 중앙행정기관	국가

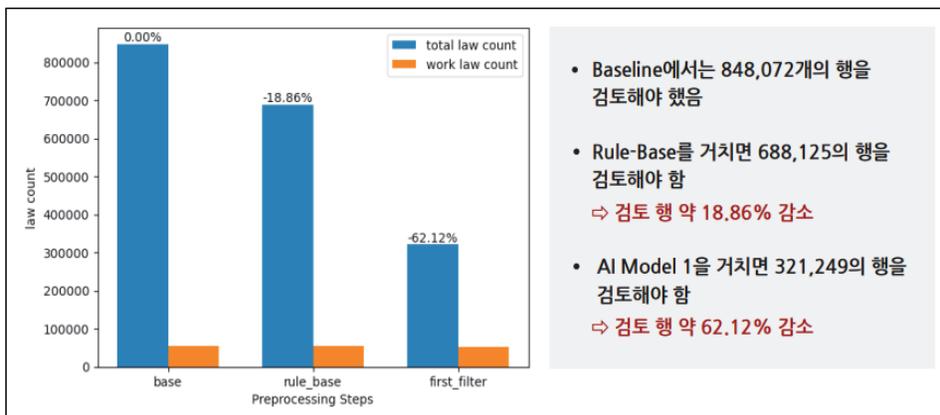
□ 자동화 모델의 타당성 검증: 2022년 5월 이후 법령 데이터의 적용

- 2022년 5월 1일 이후 법령의 제·개정 등으로 인한 변경 사항은 총 956행이며, 그 중 입력 오류 등의 결측치를 제거하여 최종 898행에 대하여 검증을 실시함
- 타당성 검증 결과, 사무판단 모델(AI Model 1)의 recall은 98.5%로 2022년 법령 데이터를 대상으로 검증한 결과와 큰 차이를 보이지 않음
- 하지만 사무유형 모델(AI Model 2)의 경우 정확도가 63.8%로 2022년의 검증 정확도(81%)에 비하여 큰 폭으로 감소한 것으로 나타남
 - 이는 2022년의 경우 전체 조문 대비 사무 비율이 약 6%로 나타난 것에 반해, 2023년의 경우 사무가 전체 조문의 약 35%를 차지하여, 이는 2022년 5월 이후의 데이터 세팅 자체가 사무로 추출된 조문으로 1차 추출된 결과로 입력이 된 것에 기인한 것으로 보임
- 그럼에도 불구하고 정확도가 큰 폭으로 감소한 것은 본 연구에서 개발한 자동화 모델이 2022년 법령 데이터에 과적합되어 있어 성능이 비교적 낮아진 것으로 보임
 - 다만, 법령 데이터 오류의 축소, 2022년 5월 이후 법령 데이터 전체에 대한 재학습 및 연도별 AI 딥러닝을 지속적으로 수행한다면 이러한 정확성의 큰 격차는 점차 축소시킬 수 있을 것으로 사료됨

□ 생성형 AI를 활용한 법령사무조사 자동화 모델 구축의 기대효과

- 「2022년 법령상 사무총조사」의 경우 2022년 5월을 기준으로 시행 중인 법령 4,339개의 법령을 대상으로 연구진이 직접 판단하는 방식으로 조사되었음
 - 각 법령의 모든 조문을 엑셀파일에 크롤링한 결과, 전체 엑셀의 행 수는 84만여 개로 각 조문의 조, 항, 호 모두에 대한 판단이 투입되어야 하였기 때문에 방대하게 많은 연구진의 노력이 투입되어야만 하였음
- 그러나 자동화 모델의 적용을 통하여 1차적인 rule-based의 단계를 통하여 약 19%의 법령 데이터는 판단 대상에서 제외할 수 있음
 - rule-based 규칙은 무조건 비사무로 판단될 수 있는 35개의 원칙을 인공지능이 확인하여 이에 해당하는 조문을 1차적으로 제거하는 방식임
 - 이를 통하여 당초 검토해야 할 전체 행 848,072개 중 18.9%의 행이 원천적으로 제거되어 688,125개의 행이 검토대상 행으로 추출될 수 있음
- 또한 사무판단 모델(AI Model 1)을 통하여 인공지능이 확실하게 비사무인 사무만을 제거하면 검토 행의 약 62.1%가 감소한 321,249개의 행만 검토가 필요한 행으로 추출됨

[그림] 자동화 모델의 성능지표-모델 적용에 따른 검토 행의 감소 규모



□ 정책적 제언

- 첫째, 디지털플랫폼정부의 구현을 위한 법령상 사무 및 사무배분 등의 현황을 파악하기 위하여 인공지능을 활용한 자동화 모델을 구축·적용할 필요가 있음
 - 자동화 모델의 구축·적용을 통하여 사람(연구진)의 노력을 획기적으로 줄일 수 있다면 빠르게 변화하는 법·행정·사회적 환경에 보다 신속하게 대응할 수 있을 것임
 - 또한 그동안 사람의 주관적 판단으로 인하여 사무총조사 결과에 대한 신뢰도 문제 역시 일정 부분 해결할 수 있을 것임
 - 다만, 사무에 관한 복잡한 판단을 포착할 수 있는 자동화 모델의 로직이 개발되기 전까지는 1차적으로 연구진이 판단한 결과를 토대로 자동화 모델이 2차 검증하는 역할에서 모델의 성숙에 따라 점차 역할을 확대해 나가는 방안을 고려할 필요가 있음
- 둘째, 정확한 법령상 사무의 실태 파악을 위하여 체계적이고 정확한 법령 데이터의 관리·보완 체계의 마련이 필요함
 - 본 자동화 모델을 통하여 소수이기는 하지만 데이터의 오류 및 기존 사무총조사 결과의 오류를 추출하였으나, 이러한 잘못된 데이터와 결과로 인하여 결국 법령상 사무의 실태를 정확하게 포착이 어려울 수 있기 때문에 이에 대한 체계적이고 지속적인 관리와 보완이 필요할 것임
- 셋째, 자동화 모델에서 기능 중심의 키워드 추출 방식의 도입을 통하여 타 분야에의 적용 확대 가능성 모색이 필요함
 - ‘규제’, 승인이 필요한 사무의 추출 등 ‘행정절차’ 등의 키워드를 중심으로 사무를 추출할 수 있는 기능을 보완한다면 이와 관련한 타 분야에까지 확대하여 적용·활용될 수 있을 것임
 - 또한 법령뿐만 아니라 조례까지 적용 범위를 확대한다면 국가사무와 지방사무 등의 실태뿐만 아니라 자치사무 등의 실태를 파악하는 데에도 광범위하게 활용될 수 있을 것임

- 넷째, 법령상 사무총조사를 위한 전문인력의 육성이 필요함
 - 광범위한 법령 데이터를 구축·관리하기 위해서는 컴퓨터 프로그래밍에 대한 이해뿐만 아니라 법령의 체계 등에 대한 이해가 뒷받침되어야 보다 정확한 판단이 이루어질 수 있음
 - 또한 제정과 개정을 반복하는 법령의 특성상 이를 관리하기 위해서는 법령 데이터를 전담하여 관리할 수 있는 전문인력 및 관련 교육 프로그램 등의 방안을 모색할 필요가 있음

목 차

제1장 | 서 론

제1절 연구의 목적 및 필요성	3
1. 연구의 배경 및 필요성	3
2. 연구의 목적	5
제2절 연구의 범위 및 방법	6
1. 연구의 범위	6
2. 연구의 방법	6
3. 연구의 흐름	7

제2장 | 이론적·제도적 검토

제1절 법령상 사무 및 법령상 사무총조사	11
1. 사무에 대한 논의	11
2. 사무배분 및 이양에 관한 연구동향	13
3. 「2022년 법령상 사무총조사」에 관한 논의	16
제2절 인공지능 기반 디지털플랫폼정부	20
1. 데이터 기반 행정과 디지털플랫폼정부의 실현	20
2. 생성형 AI의 의의와 디지털플랫폼정부의 실현 방향	23

제3장 | 법령사무조사 자동화 모델의 개발

27

제1절 자동화 모델 개발 설계	29
1. 자동화 모델 개발 개요	29
2. 자동화 모델 개발 내용	29
3. 자동화 모델 개발 절차	33

제2절 자동화 모델의 개발	34
1. 데이터 전처리 및 데이터 분할	34
2. (AI Model 1) 사무판단 모델의 개발	38
3. (AI Model 2) 사무유형 모델의 개발	46

제 4 장 | 법령사무조사 자동화 모델의 적용

제1절 자동화 모델의 예측 정확성 검증	55
1. 사무판단 및 사무유형 모델의 평가도구: Confusion Matrix	55
2. 사무판단 모델(AI Model 1)의 검증	56
3. 사무유형 모델(AI Model 2)의 검증	57
제2절 자동화 모델을 적용한 데이터 오류 검증	58
1. 개요	58
2. 인공지능을 통해 추출된 데이터 오류 사항	59
3. 인공지능을 통해 추출된 데이터 오류의 검증	62
제3절 2022년 법령 데이터에 대한 자동화 모델의 적용 및 활용	64
1. 자동화 모델을 활용한 사무판단 및 사무유형 추출 결과	64
2. 법령상 사무에 대한 자동화 모델의 활용: 대시보드의 제시	66
제4절 자동화 모델의 타당성 검증: 2022년 5월 이후 법령 데이터의 적용	72
1. 개요	72
2. 법령사무조사 자동화 모델의 타당성 검증	72

제 5 장 | 결론 및 정책적 제언

제1절 연구요약 및 정책적 제언	77
1. 연구요약	77
2. 기대효과	79
3. 정책적 제언	80
제2절 연구의 한계 및 향후 과제	83

【참고문헌】	85
【부록】	87
부록 1. 사례 적용 및 결과값 비교: 「우주손해배상법」	87
부록 2. 데이터 처리: 사무판단, 사무유형 등	90
부록 3. 법령사무조사 자동화 모델의 개발 과정	99

표 목차

[표 1-1] 2022 법령상 사무총조사 결과	7
[표 2-1] 2022 법령상 사무총조사 결과	18
[표 2-2] 디지털플랫폼정부의 구성요소	20
[표 2-3] 국내 공공분야 생성형 AI 활용 현황	24
[표 3-1] 2022 법령상 사무총조사의 사무 구분 기준	30
[표 3-2] 데이터 요약	34
[표 3-3] rule-base 규칙을 적용한 데이터 형태와 제거되는 행 개수	39
[표 3-4] 조문 텍스트의 Okt 기반 2음절 이상 명사 토큰화 예시	42
[표 3-5] 데이터를 8:2로 분할하여 학습한 데이터셋(버전1) 기준, 사용변수에 따른 모형	47
[표 4-1] 분류 방법론별 적용 결과	56
[표 4-2] 인공지능을 통해 추출된 데이터 오류 사항	59
[표 4-3] 인공지능이 추출한 데이터 오류의 검증	63

그림 목차

[그림 2-1] 디지털플랫폼정부의 핵심추진과제	22
[그림 3-1] 법령사무조사 자동화 모델 개발 절차	33
[그림 3-2] 2개 이상의 소관부처명 순서 변경 처리 및 소관부처 순서가 다른 경우 처리(예시)	36
[그림 3-3] 법령구분 자료의 데이터 분할 과정	37
[그림 3-4] 데이터 분할 방식	38
[그림 3-5] 사무판단 모델(AI Model 1)의 흐름	38
[그림 3-6] 데이터셋 전체의 빈도수	43
[그림 3-7] Transformer 모델	44
[그림 3-8] 사무판단 모델(AI Model 1): Transformer Encoder 활용	44
[그림 3-9] 사무판단 모델(AI Model 1) 방법론별 적용 결과	45
[그림 3-10] 사무유형 모델(AI Model 2)의 흐름	46
[그림 3-11] 수행주체에 대한 디셔너리 작성 예시	48
[그림 3-12] 머신러닝 모델학습 적용	49
[그림 3-13] 앙상블 모델 구조(확실한 결과)	50
[그림 3-14] 앙상블 모델 구조(애매한 결과)	50
[그림 3-15] 사무유형 모델(AI Model 2) 방법론별 적용 결과	50
[그림 4-1] Confusion Matrix	55
[그림 4-2] 사무유형 분류 모델 결과	57
[그림 4-3] ‘호’ 단위에서 비사무로 추출되는 경우(rule-based 사용 규칙)	61
[그림 4-4] 대시보드 아키텍처	64
[그림 4-5] 2022년 법령에 대한 인공지능의 사무판단 및 사무유형 추출 결과(예시) ..	67
[그림 4-6] 대시보드의 전체 구성	68
[그림 4-7] 법령선택 박스	69

[그림 4-8] 원하는 법령명의 검색 박스	69
[그림 4-9] 사람과 AI의 사무판단 비교	69
[그림 4-10] 수행주체 후보들 시각화	70
[그림 4-11] AI모델의 상세결과	71
[그림 4-12] 실제 대시보드 적용 화면	71
[그림 5-1] 자동화 모델의 성능지표-모델 적용에 따른 검토 행의 감소 규모	79



제1장

서론

제1절 연구의 목적 및 필요성

제2절 연구의 범위 및 방법

제1장 서론

제1절 연구의 목적 및 필요성

1. 연구의 배경 및 필요성

□ 데이터 기반 행정의 대두 및 윤석열 정부의 디지털플랫폼 정부의 실현 계획

- 「데이터기반행정 활성화에 관한 법률」에 의거하여 중앙부처, 지방자치단체, 공공기관은 데이터 공유·분석 및 정책활용을 위한 데이터 체계를 구축 중임
- 윤석열정부는 디지털플랫폼정부의 실현을 통하여 맞춤형 정보의 선제적 제공을 위한 청사진을 발표함(2023.04.14.)
 - 디지털플랫폼정부란, 정부의 업무프로세스를 디지털 기술을 활용하여 전산화하고, 데이터를 표준화하고, 공유하는 시스템임
 - 디지털플랫폼정부는 AI를 기반으로 각 부처와 기관에 흩어져 있는 정보를 하나의 플랫폼에 통합하여 국민이 원하는 정보를 맞춤형으로 선제적 제공하는 미래 정부 모델임

※ (국정과제 11번) 모든 데이터가 연결되는 세계 최고의 디지털플랫폼정부 구현

: 대국민 선제적·맞춤형 서비스 제공, 인공지능·데이터 기반의 과학적 행정 구현, 국민·기업·정부 협력을 통한 혁신 생태계 조성 등

- 생성형 AI를 활용한 법령사무조사는 방대한 양의 법령에서 국가 및 지방자치단체의 사무를 빠르고 정확하게 추출하고 손쉽게 활용할 수 있는 시스템을 구축할 수 있다는 점에서 디지털플랫폼정부와 맥을 같이 한다고 할 수 있음

□ 효율적인 사무의 수행을 위하여 국가-지방 사무배분 및 사무수행체계 현황을 파악하기 위한 노력은 지속적으로 이어져 왔음

- 국가사무와 지방자치단체사무 등 사무수행체계의 현황을 파악하기 위하여 1994년부터 약 5년 단위로 우리 원을 중심으로 「법령상 사무총조사」가 진행되었음
 - 「법령상 사무총조사」는 국가-지방자치단체의 사무 배분 현황의 파악을 위한 기본 통계 작성을 통하여 사무정보관리시스템 구축을 위한 기초자료 제공을 목적으로 함
 - 「법령상 사무총조사」를 통한 기초자료는 국가사무의 지방이양 추진뿐만 아니라 최근 논의되고 있는 행정체제개편에 따라 전환되는 사무의 추출 등 다양한 영역에서 활용할 수 있을 것으로 사료됨
- 그러나 지금까지의 「법령상 사무총조사」는 연구자의 직접 판단에 의한 사무추출 방식으로 이루어져 여러 한계에 직면하고 있음**
- 법령상 사무배분 및 사무수행체계의 현황을 파악하기 위해 수행되고 있는 「법령상 사무총조사」는 조사의 필요성과 중요성에도 불구하고 여전히 연구진의 주관적 판단에 의존하는 방식을 취하고 있어 여러 한계에 직면하고 있음
 - 첫째, 방대한 양의 법령에 대한 사무배분의 판단은 연구진의 주관적 판단으로 이루어짐에 따라 시간과 노력이 과도하게 투입되고 있음
 - 법령은 지속적으로 제·개정되며 이에 따른 사무배분의 형태가 변동될 수 있음에도 불구하고, 시간과 비용의 문제로 인하여 이러한 변화에 능동적이고 시의성 있는 대응에 한계가 있을 수 밖에 없음
 - ※ 2022.5.1. 시행 기준 중앙행정기관 소관 전체 법령 4,760개 중 선정 제외 법령 제외 4,339개 중 2019.5.1.* 이후 제정되었거나 개정된 법령은 총 3,508개이며, 2019.5.1. 이전에 제정되어 이후 변동이 없는 법령은 831개
 - * 2022년 법령상 사무총조사 직전 조사
 - 둘째, 사무배분이 연구자의 주관적 판단에 의존한다는 점 등에서 조사 결과에 대한 신뢰성 문제가 지적될 수 있음

- 사무배분은 공통 기준이 있음에도 불구하고 사무배분은 다양한 경우의 수가 존재하며 이 경우 연구진의 주관적 판단이 투입될 수밖에 없으나 약 5,000개에 달하는 모든 법령에 대한 개별 검토는 사실상 불가능한 실정임

2. 연구의 목적

□ 생성형 AI를 활용한 법령사무조사 자동화 로직 등 모델 개발

- 생성형 AI 전문기관에의 위탁을 통한 법령사무조사 자동화 모델 개발
 - 법령상 사무를 추출하는 코딩, 로직 등 프로그래밍을 포함함
 - ※ (2023.11.28. 기준) 원내 계약체결 준비 중(계약체결 협조요청 공문 시행 완료)

□ 주요 법령에 대한 법령사무조사 자동화 모델의 시범 적용

- 2022.5.1. 기준 중앙행정기관이 소관하는 법령 중 중앙행정기관, 지방자치단체를 사무수행 주체로 규정하는 법령을 대상으로 3단계* 사무추출
 - * 3단계: 대분류 4개 유형, 중분류 6개 유형, 소분류 16개 유형

□ 자동화 모델을 활용한 「법령상 사무총조사」 수행 가능성 등 미래 연구 방향 제시

- 2022년 「법령상 사무총조사」의 결과 및 자동화 모델을 통해 추출된 사무배분 결과를 비교분석함으로써 향후 생성형 AI를 활용한 법령상 사무총조사의 수행 가능성 등의 미래 연구 방향을 제시하고자 함

제2절 연구의 범위 및 방법

1. 연구의 범위

□ 연구의 대상 범위

- 2022.5.1. 기준 중앙행정기관이 소관하는 전체 법령 중 중앙행정기관, 지방자치단체를 사무수행 주체로 규정하는 4,339개의 법령을 대상

□ 연구의 내용 범위

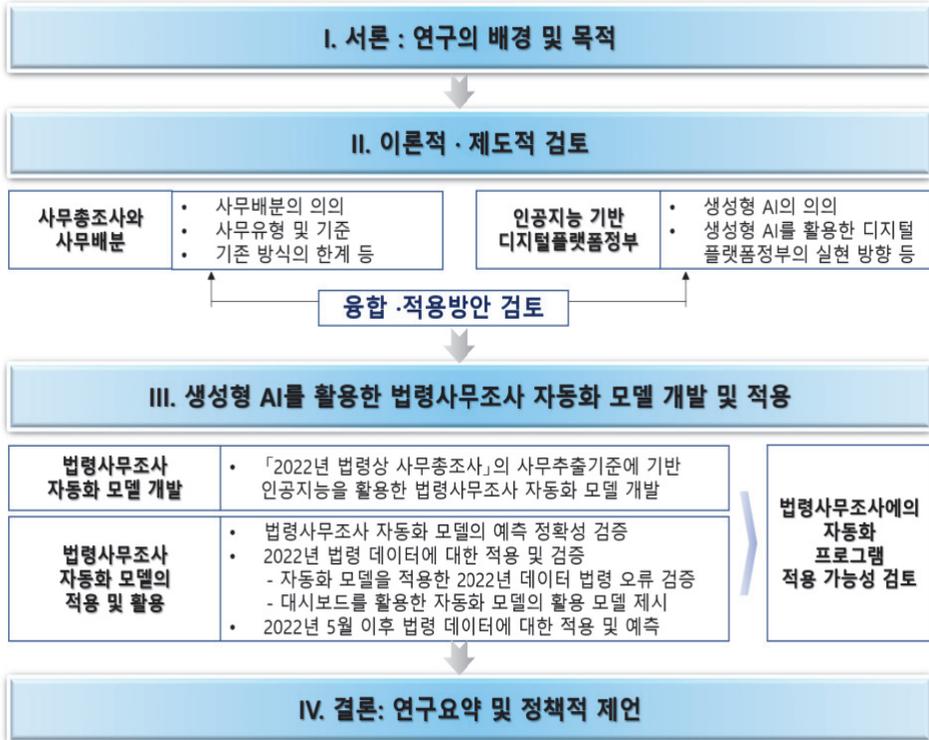
- 생성형 AI를 활용한 법령사무조사 자동화 모델 개발 및 적용
 - 2022년 법령상 사무총조사 데이터를 활용하여 인공지능 모델 개발
 - 2022년 「법령상 사무총조사」 결과와의 비교 분석
- 자동화 모델을 활용한 「법령상 사무총조사」 수행 가능성 및 미래 연구방향 제시

2. 연구의 방법

- 문헌 연구
 - 「법령상 사무총조사」 관련 선행연구의 사무추출의 및 사무분류 기준을 검토함
- 생성형 AI를 활용한 법령사무조사 자동화 모델 개발
 - 전문기관 위탁하여 법령사무조사 자동화 로직 등 모델을 개발함
- 표적집단면접조사(Focus Group Interview, FGI)
 - 생성형 AI를 통한 자동화 모델의 적용 가능성 및 미래 연구 방향 전반에 대하여 실시함

3. 연구의 흐름

[표 1-1] 2022 법령상 사무총조사 결과





제2장

이론적·제도적 검토

제1절 법령상 사무 및 법령상 사무총조사

제2절 인공지능 기반 디지털플랫폼정부

제1절 법령상 사무 및 법령상 사무총조사

1. 사무에 대한 논의¹⁾

1) 사무의 개념

- 우리나라의 경우 중앙정부와 지방자치단체 간 사무배분 실태 혹은 사무배분의 기준을 설정하는 데 있어서 단위사무를 사용하고 있으나, 현재까지 단위사무의 개념에 대한 합의된 견해 없이 진행되어 왔으며, 연구 목적에 따라 다양한 조작적 정의로 접근하고 있음(한부영·박재희, 2019)
 - 현행법상 국가, 자치단체의 권한 및 책임의 범위를 정하고 있으나, 소관 사무 또는 사무의 정의는 명확히 규정하고 있지 않음
 - 학술연구에서도 단위사무에 대한 일치된 정의 없이 사무, 단위사무, 기능, 이양, 배분과 같은 개념을 혼용하고 있음
- 단위사무는 지방자치단체가 행정 목적을 달성하기 위하여 수행하는 업무임(한부영·박재희, 2019)
 - 단위사무는 자치법규에 제시된 분장 사무이며 개인들이 직무상 처리하는 사무로 법령상 처리권자가 명확하고 동시에 행위가 수반되는 법령상 조·항·호에서 추출된 사무로 보고 있음(한부영·박재희, 2019)
 - 2022년 「법령상 사무총조사」에서 단위사무는 “모든 법의 각 조·항·호와 이와 연계된 시행령과 규칙의 각 조·항·호를 상호 연계하여, 법령상 권한과 책임 및 의무가 발생하는 처리권자가 행정목적 달성을 위하여 수행해야 하는 최소 단위의 사무”로 정의함

1) 한국지방행정연구원. (2022). 「법령상 사무총조사 연구」. 한국지방행정연구원.

- “모든 법률의 각 조·항·호를 기본 단위로 하고 이와 연계된 시행령과 규칙의 각 조·항·호를 동시에 고려하여 단위사무를 추출”하며, 가능한 한 “항”을 기준으로 각 조가 한 문장으로 이루어진 경우 1개의 단위사무로 계산하며, 호별로 별도의 수행주체가 다른 경우 예외적으로 사무를 추출하는 경우가 있음

2) 사무의 구분

□ 대분류

- 사무유형은 법령상 권한 및 책임이 귀속되는 주체를 기준으로 국가사무, 국가와 지방자치단체의 공동사무, 자치사무로 유형화함
 - 공동사무는 법령상 개념은 아니며 ① 국가와 지방자치단체 간, 지방자치단체 상호 간 협력과 이해관계를 가지며, ② 시간·업무 범위에서 동시적 책무로 규정하고 있고, ③ 사무의 성격상 공동적 비용 부담을 의무화한 사무를 의미함

□ 중분류

- 대분류의 세부유형으로 국가사무, 자치사무, 자치공동사무를 지방자치단체의 공동사무를 국가-시도-시군구 공동사무, 국가-시도 공동사무, 국가-시군구 공동사무, 시도-시군구 공동사무로 유형화함
- 사무유형은 법령상 권한 및 책임이 귀속되는 주체를 기준으로 국가사, 국가-시도-시군구사무, 국가-시도공동사무, 국가-시군구공동사무, 시도-시군구 공동사무, 시도사무, 시군구사무 총 7개 유형으로 구분함
 - 국가와 지방자치단체 공동사무를 국가-시도-시군구사무, 국가-시도 공동사무, 국가-시군구 공동사무로 구분함
 - 자치사무는 시도-시군구 공동사무, 시도사무, 시군구사무로 구분함

□ 소분류

- 사무분류의 세부 유형으로 현행 법령상 규정과 같이 위임·위탁 대상 기관을 구분하여 총 16개 유형으로 분류함

- (국가사무) 국가직접처리사무, 특별지방행정기관사무, 국가위탁사무, 시·도 위임사무, 시·군·구위임사무로 구분함
- (국가-지방자치단체 공동사무) 국가-시도-시군구 공동사무, 국가-시도 공동사무, 국가-시군구 공동사무로 구분함
- (자치사무) 자치사무는 시·도직접처리사무, 시·도위탁사무, 시·군·구 직접처리사무, 시·군·구위탁사무, 시·도-시·군·구 공동사무, 시·도-시·군·구 위임사무로 구분함

2. 사무배분 및 이양에 관한 연구동향

□ 사무배분 및 이양에 관한 선행연구 검토

- 중앙-지방 간의 사무의 이양 및 배분에 관한 선행연구는 시기별로는 특정 정권을 기준으로 이루어진 연구, 여러 정권별 사무이양의 실태에 있어 나타나는 차이점을 분석한 연구로 구분할 수 있음(권경득·우무정, 2009; 최송이·최병대, 2012; 이해영, 2014; 고광용, 2016; 최근열, 2016)
- 분석 시기에 따른 구분을 떠나, 선행연구의 내용적 측면의 공통점으로는 이양 사무의 특성과 이양의 기준점을 실증 분석하거나, 지방분권의 효과적인 추진을 위한 사무 배분 및 이양의 원칙과 기준을 제시하는 목적의 연구가 주를 이루었다는 점이 특징임
- 사무배분 및 이양에 관한 선행연구는 지방분권을 위한 중앙-지방 간의 사무배분 및 이양이 필요함을 공통적인 전제로 논의를 전개함
 - 이와 같은 전제하에, 사무의 지방이양 및 분배에 관한 실태 분석을 통하여 개선점 및 정책적 시사점을 도출하는 것이 전반적인 연구의 흐름이라고 볼 수 있음
 - 사무의 지방이양과 분배를 저해하는 요인과 그 한계점에 관한 논의를 전개함으로써 개선방안을 탐색하는 귀납적 구성이 주요 특성임

□ 정권별·정권 간 사무이양의 실태 분석 연구

- 권경득·우무정(2009)은 참여정부의 중앙사무 지방이양 실태를 분석하였음
 - 분석 결과, 중앙과 지방간의 기능배분의 실적이 저조하며, 행·재정적 측면의 포괄적인 이양 노력이 이루어지지 않았다는 점, 지방의 차별적 특성을 고려하지 않았다는 점을 문제점으로 제시함
- 최송이·최병대(2012)는 1991년부터 2010년까지의 20년 간의 중앙정부-지방정부 간 사무배분을 분석대상으로 하여 사무배분의 실태를 분석함. 이를 통하여 중앙-지방 간 역할분담의 방향성을 제시하였음
- 이해영(2014)은 이명박정부 시기의 지방분권촉진위원회 제3실무위원회의 이양대상사무 심의자료를 분석하여 국가 사무의 지방이양을 저해하는 실질적인 요인을 도출함
 - 이양사무의 발굴에 있어 체계를 합리화하여야 하며, 사무배분 원칙을 확립하여 배분의 우선순위를 명확히 할 필요가 있다는 점, 포괄적 이양을 위하여 사무이양 시 소요되는 인력 및 비용을 보다 구체적으로 산정하여야 한다는 점을 시사점으로 도출함
- 고광용(2016)은 김대중 정부, 노무현 정부, 이명박 정부의 지방 사무이양 체계 및 성과를 살펴보고 이에 따른 중앙·지방간 사무관계 변화양상을 탐색함
 - 노무현 정부에서 사무이양의 노력과 성과가 높게 나타난 반면, 이명박 정부는 사무이양 노력과 성과는 비교적 낮게 나타남. 이러한 지방이양 성과 분석을 토대로 향후 사무이양을 위한 적절한 세원·재원을 이양하는 등 정부 차원의 노력이 필요하다고 제언함
- 최근열(2016)은 역대 정부에서 추진해 온 지방분권정책과제 중 중앙권한 및 사무의 지방이양을 중심으로 이양 실태 및 문제점을 살펴보고 향후의 발전과제를 제시하였음
 - 해당 연구는 복잡화·전문화되고 있는 정부기능을 중앙정부와 지방정부가 합리적으로 분담하여 수행해야 하며, 행정기능의 효율적인 수행체제를 확립하고 국가경쟁력을 강화하는 데에 정부기능의 분담이 중요하다는 점을 강조함

□ 사무 배분 및 이양의 원칙과 기준 제시 연구

- 홍준현(2001)은 우리나라에서 1999년에 효력이 발휘된 「중앙행정권한의 지방이양촉진 등에 관한 법률」에 명시된 차등이양제도와 외국의 차등이양제도의 실태를 파악하고 차등이양제도의 유형별 장단점을 살펴봄
 - 유형별 장단점 분석을 통하여 우리나라의 차등이양제도는 포괄적 차등이양 방식과 부분적 차등이양방식의 균형점을 고민해야 한다는 함의를 도출함
- 박혜자(2002)는 지방이양추진위원회의 심의내용을 자료로 수집하여 이양의 유형과 성격, 기준, 태도 등을 실증 분석함
 - 분석 결과, 지방이양 사무는 정책적 사무보다 집행적 성격의 사무가 주였으며, 대부분 이미 위임된 기관위임사무가 재이양되었다는 결과를 도출함
- 하정봉·소진광(2007)은 지방이양실적이 저조한 점과 그 기대 수준이 미치지 못한 점을 위임 사무의 규모와 이양에 따른 지원부족, 이양 이후의 지속적인 관리 노력의 부재를 원인으로 제시하였음
 - 이에 따라, 실질적으로 지방정부에게 파급 효과가 큰 사무를 위임할 필요성과 행·재정적 지원 노력, 주기적인 점검을 통한 지속적인 관리를 제언함
- 한부영·박재희(2019)는 사무배분에 대한 새로운 원칙과 기준을 재정립하여 자치분권을 효과적으로 추진하기 위한 제언을 도출하는 것을 목적으로 연구를 진행함
 - 분석 결과, 국가와 지방자치단체 간 사무배분의 기준을 명확하게 하고, 사무 구분 기준의 적용 가능성을 제고하여 지방이양의 대상이 되는 사무를 발굴하기 위한 밑거름을 마련할 필요성을 제시함

3. 「2022년 법령상 사무총조사」²⁾에 관한 논의

1) 연구 개요

□ 연구배경 및 목적

- 윤석열 정부는 ‘대한민국 어디서나 살기 좋은 지방시대’를 주요 국정목표로 제시함
 - ‘대한민국 어디서나 살기 좋은 지방시대’는 사는 곳의 차이가 기회와 생활의 격차로 이어지는 불평등을 멈추고 수도권 쏠림과 지방소멸의 악순환을 끊어내는 지속가능한 대한민국을 목표로 설정함(대한민국정부, 2022)
- 국가-지방 기능 조정은 ‘지방시대 실현을 위한 지방분권 강화’라는 세부 과제를 달성하기 위한 주요한 정책과제에 해당함
- 이에 법령상 사무총조사는 국가-지방자치단체의 사무배분 현황 파악을 위한 기본 통계를 작성하여 사무정보관리시스템 구축에 기초자료를 제공하여 국가-지방자치단체 또는 시·도·시·군·구 간 분권 수준 측정 및 자치분권 확대를 위한 기초자료로 활용하기 위함을 목적으로 함

□ 연구대상의 선정

- (선정) 2022.5.1. 기준 중앙행정기관³⁾이 소관하는 전체 법령 중 중앙행정기관, 지방자치단체를 사무수행 주체로 규정하는 법령을 대상으로 선정함
- (선정 제외) 입법부·사법부 등 중앙행정기관이 아닌 기관⁴⁾이 소관하는 법령, 법원이 사무수행 주체인 법령, 조직 및 직제, 사무처리기준, 사법(私法)관계, 형벌 기준을 규율한 규정으로만 구성된 법령은 조사 범위에서 제외됨

2) 한국지방행정연구원. (2022). 「법령상 사무총조사 연구」. 한국지방행정연구원.

3) 「정부조직법」 제2조 제2항에 따른 중앙행정기관(4실, 1원, 18부, 5처, 18청, 6위원회)을 의미하며, 국가행정사무를 담당하기 위하여 설치된 행정기관으로서 그 관할권의 범위가 전국에 미치는 기관을 의미함

4) 국회, 감사원, 법원, 헌법재판소, 중앙선거관리위원회, 민주평화통일자문회의

□ 분석 과정

- 크롤링을 통해 추출한 법령 조문으로부터 사무를 추출하고 유형화하기 위한 작업으로 4단계 과정으로 추진함
 - (1단계) 훈련받은 연구원이 세부적 사무 판단을 위해 법 규정 및 그 하위법령, 연계 법령을 종합적으로 검토하여 법규정상의 단위사무를 추출함
 - (2단계) 법령 조문으로부터 추출된 사무에 대하여 사무 판단 근거 및 사무명의 적정성, 사무 유형에 대한 교차 검증을 실시함
 - (3단계) 사무추출 및 사무 유형 분류가 애매한 경우 전체 연구진 회의 및 전문가 자문회의를 통해 결정함
 - (4단계) 사무 구분 결과의 타당성 확보를 위해 행정안전부 및 중앙행정기관의 법제 담당 부서의 샘플 검증을 실시함
- 법령으로부터 추출된 단위사무의 총합을 계산하고 국가사무, 국가-지방자치단체 공동사무, 자치사무의 개수 및 비율을 도출함

2) 주요 연구 결과 및 한계

□ 주요 연구 결과

- 사무추출 대상 법령은 중앙행정기관 의견조회를 거쳐 「정부조직법」 제2조 제2항에 따른 중앙행정기관(4실, 1원, 18부, 5처, 18청, 6위원회) 소관의 4,357개 법령(법 1,451개, 시행령 1,683개, 시행규칙 1,223개)으로 함
 - 자치법규(조례, 규칙)를 제외하고, 국가법령정보센터에 등록된 2022.5.1. 기준 「정부조직법」 제2조 제2항에 따른 중앙행정기관 소관 법령은 총 4,760개(법 1,550개, 시행령 1,827개, 시행규칙 1,383개)이나,
 - 이 중 중앙행정기관 및 지방자치단체 사무와 관련성이 없는 403개 법령은 사무추출 제외 대상 법령으로 설정하고 타당성 검토를 실시하였으며, 4,357개 법령을 분석 대상으로 확정
 - 추가적으로 법령 폐지, 타법 폐지, 시행일 미도래 등으로 2022.5.1 기준 존재하지 않는 법령 18개 법령 제외

- 2022.5.1. 기준 시행 중인 4,339개 법령(법 1,443개, 시행령 1,675개, 시행규칙 1,221개)으로부터 총 60,114개의 단위사무 추출
 - 분석결과 총 60,114개 사무 중 국가사무 41,911개(약 69.7%), 국가-지방 자치단체 공동사무 7,135개(약 11.9%), 자치사무 11,068개(지방자치단체 간 공동사무 4,480개 포함, 18.4%)로 도출
 - 국가사무, 자치사무, 지방자치단체 간 공동사무를 국가사무와 지방자치단체 사무로 이원화할 경우, 총 사무 77,242개 중 국가사무 48,247개(기관위임 사무 799개 포함 63.5%), 자치사무 28,196개(36.5%)

[표 2-1] 2022 법령상 사무총조사 결과

사무유형	사무 수
총계	60,114(100%)
국가사무 합계	41,911(69.7%)
국가사무 소계	41,112(68.4%)
국가직접처리사무(국가소속/산하사무 포함)	33,873
특별지방행정기관사무	3,852
국가위탁사무	3,387
기관위임사무 소계	799(1.3%)
시·도위임사무	579
시·군·구위임사무	123
시·군·구재위임사무	11
시·도 및 시·군·구위임사무	86
국가-지방자치단체 공동사무 합계	7,135(11.9%)
국가·시·도·시·군·구 공동사무	5,513
국가·시·도 공동사무	1,427
국가·시·군·구 공동사무	195
자치사무 합계	11,068(18.4%)
시·도 사무 소계	3,342(5.6%)
시·도직접처리사무(시·도소속/산하사무 포함)	3,018

사무유형	사무 수
시·도위탁사무	185
시·도·시·군·구위임사무	139
지방자치단체 간 공동사무 소계	4,480(7.5%)
시·도·시·군·구 공동사무	4,480
시·군·구사무 소계	3,246(5.4%)
시·군·구직접처리사무	3,085
시·군·구위탁사무	161

출처: 한국지방행정연구원. (2022). 「법령상 사무총조사 연구」. 한국지방행정연구원.

□ 연구의 한계

- (내재적 한계) 조사 대상인 법조문 자체가 가진 한계로 인하여 단위 사무의 비중이 일정하지 않거나 분류가 어려운 사무가 존재함
 - 또한, 법률·시행령·시행규칙만을 조사 대상으로 하고 있기 때문에 조례 및 규칙과 같은 자치사무에 관한 법령은 조사대상에 포함되지 않음
- (조사의 일관성 확보에 관한 한계) 지금까지 수행된 법령상 사무총조사는 연구자의 판단에 의해 사무추출을 하는 방식으로 이루어졌으며, 사무 판단에 있어서도 연구자 간의 일관성을 확보하기 어려운 한계가 존재함
 - 조사의 주체 및 방법이 상이함
 - 조사 결과의 연도별 비교에 있어서는 보수적 접근이 필요함
- (과도한 투입비용의 한계) 「법령상 사무총조사」는 조사의 필요성과 중요성에도 불구하고 내재적 한계 및 조사의 일관성 확보에 관한 한계로 인하여 비효율적인 조사 진행이라는 문제가 발생함
 - 방대한 양의 법령에 관한 사무 판단을 위하여 과도한 시간과 노력이 투입됨
 - 그럼에도 불구하고, 연구자 간의 사무 판단 일관성 저하로 인한 조사 결과의 신뢰성 문제가 제기될 위험이 있음
 - 또한 법령의 변화에 대한 능동적이고 시의성 있는 대응에 관한 한계 발생

제2절 인공지능 기반 디지털플랫폼정부

1. 데이터 기반 행정과 디지털플랫폼정부의 실현

1) 디지털플랫폼정부의 개념

- 디지털플랫폼정부란 “모든 데이터가 연결되는 ‘디지털플랫폼’ 위에서 국민, 기업, 정부가 함께 사회문제를 해결하고, 새로운 가치를 창출하는 정부”를 의미함 (디지털 플랫폼정부위원회 보도자료, 2022.09.02.)
- 디지털플랫폼정부의 구성요소는 디지털플랫폼 공급자(정부), 생산자(기업 및 개발자), 이용자(국민), 생태계임

[표 2-2] 디지털플랫폼정부의 구성요소

구성요소	내용
디지털플랫폼 공급자	정부가 디지털플랫폼을 구축하고 운영, 데이터를 제공하며 플랫폼의 운영방식에 대한 조정과 결정을 통하여 플랫폼 생태계 활성화 역할을 담당
생산자	기업 및 개발자는 정부에서 제공하는 API, SDK, 클라우드, 공공데이터 등을 활용하여 정부 서비스를 제안하고 혁신하는 역할을 수행
이용자	대표적인 서비스의 이용자는 국민이며, 의견 제고를 통하여 디지털 플랫폼 정부의 기능 강화에 참여
생태계	국민의 편익을 보장하기 위하여 정부 행정을 효율화하며 민간 혁신을 통하여 국민, 정부, 민간 모두에 가치를 창출함. 운영상 수집되는 모든 데이터는 사회문제 해결에 기여

출처: 한국지능정보사회진흥원(2022). 디지털 플랫폼 정부의 개념과 특징, 인용

- 주효진 외(2022)는 디지털플랫폼정부는 정책의 중첩성 및 복잡성을 개선하면서 외부 환경 변화에 기민하게 대응하고 다양한 조직의 협력적 관계 구축 및 관계의 유연성을 통해 구성원의 디지털 역량 강화에 주력해야 하며, 다양성 실현을 통한 사회통합적 가치 추구를 지향함과 동시에 업무의 생산성 및 성과를 견인하는 수단적 가치를 달성해야 함을 제안하고 있음

- 이는 기술발전에 따른 정부의 서비스 제공방식의 변화로 인해 촉발됨(강송희·김숙경, 2020; 주효진 외, 2022)
- 아날로그 정부(Analogue Government): 정부 중심의 전통적인 서비스 공급 방식으로, 비용절감, 효율성 증대, 생산성 중시 등의 특징을 가짐
- 전자정부(E-Government): 주요 가치로는 투명성을 추구하며, 사용자 중심의 서비스 공급방식에 해당함. 서비스 개선을 위한 사용자의 요구 예측이 필요함
- 디지털정부(Digital Government): 이전 서비스 공급방식인 전자정부와 비교하여 개방성이 보다 확대되었으며, 사용자 주도의 접근방식을 추구함. 인공지능(AI), 사물인터넷(IoT)등의 다양한 신기술을 도입하여 플랫폼 및 네트워크 조직을 조성하는 것이 주요 특징임
- OECD(2019)는 사회 전반의 변화를 반영함으로써 기존의 전자정부(E-Government)에서 디지털정부(Digital Government)로 거버넌스 구조를 전환할 필요가 있음을 강조함
- 디지털정부는 정부 운영에 있어 플랫폼(platform)전략의 접목을 요구함. 공공 데이터의 개방 및 공유와 민간부문의 자발적 참여를 통하여 서비스의 질을 제고하고 공공가치를 창출하는 등 사회문제 해결 생태계를 지속하기 위한 노력의 일환으로 ‘플랫폼으로서의 정부(Government as a Platform)’로의 전환 또한 고려되어야 함(주효진 외, 2022)

2) 윤석열 정부의 디지털플랫폼정부 구현⁵⁾

- 윤석열 정부는 국정비전의 달성을 위하여 4대 기본부문⁶⁾과 미래, 지방시대 등 6대 국정목표를 설정함

5) 대한민국정부(2022). 윤석열정부120대 국정과제

6) 정치·행정, 경제, 사회, 외교·안보

- 디지털플랫폼정부의 구현은 윤석열 정부가 제시한 15개의 국정과제 중 1개에 해당함. “디지털플랫폼정부 혁신 생태계를 조성함으로써 정부가 서비스를 직접 제공하기 보다는, 국민과 함께 혁신하고 민관이 함께 성장하는 공동기반을 마련” 하는 것을 주요 내용으로 함(대한민국정부, 2022)
- (법제도적 측면) 스마트하고 유능한 정부를 표방하며 모든 데이터가 연결되는 세계 최고의 디지털플랫폼정부 구현을 위하여 2020년 12월 10일 「데이터기반 행정 활성화에 관한 법률」을 제정함
 - 데이터를 기반으로 한 행정의 활성화에 필요한 사항을 정함으로써 객관적이고 과학적인 행정을 통하여 공공기관의 책임성, 대응성 및 신뢰성을 높이고 국민의 삶의 질을 향상시키는 것을 목적으로 하여 7) 중앙부처, 지방자치단체, 공공기관은 데이터의 공유 및 분석을 통한 정책 활용을 위하여 데이터 체계를 구축하고 고도화하는 노력을 기울이고 있음

[그림 2-1] 디지털플랫폼정부의 핵심추진과제



출처: 디지털플랫폼정부위원회(n.d.). <https://www.dpg.go.kr>

- 7) 「데이터기반행정 활성화에 관한 법률」 제1조(목적) 이 법은 데이터를 기반으로 한 행정의 활성화에 필요한 사항을 정함으로써 객관적이고 과학적인 행정을 통하여 공공기관의 책임성, 대응성 및 신뢰성을 높이고 국민의 삶의 질을 향상시키는 것을 목적으로 한다.

- (조직 측면) 「데이터기반행정 활성화에 관한 법률」 제19조에 의거하여, 공공기관의 장은 해당 기관의 데이터기반행정 활성화에 관한 업무를 총괄하는 책임관을 임명해야 함
- 또한, 2022년 디지털플랫폼정부 실현을 위한 주요 정책 등 관련 사항을 심의·조정하는 기관으로 대통령 소속 위원회인 디지털플랫폼정부위원회를 설치함
- (예산 및 재정적 측면) 디지털플랫폼정부 구현을 위하여 2023년 4,192억원 대비 121% 증가한 9,262억원을 2024년 디지털플랫폼정부 예산으로 편성함 (세계일보, 2023)

2. 생성형 AI의 의의와 디지털플랫폼정부의 실현 방향

1) 생성형 AI 개념

- 구글은 “Generative AI FAQs”라는 문건에서 생성형 AI를 “제공된 예시들에서 학습한 내용을 바탕으로 텍스트, 이미지, 음악, 코드와 같은 새로운 콘텐츠를 생성할 수 있는 기계학습(ML) 모델의 한 종류”라고 정의하였음(Google, 2023.05.)
- 맥킨지(Mckinsey)는 생성형 AI에 대하여 “신경망과 어텐션(Attention) 메커니즘을 사용하여 새로운 콘텐츠를 텍스트, 사진, 이미지, 비디오 및 3D 등 다양한 양식으로 제공”하는 것으로 정의함(Tobias et al., 2023.04.)
- Gozalo-Brizuela&GarridoMerchan (2023)는 생성형 AI를 “전문가 시스템 처럼 기존 데이터를 단순히 분석하거나 그에 따라 행동하는 것이 아닌 참신한 콘텐츠를 생성할 수 있는 인공지능”이라고 설명함
- 이를 종합하면, 생성형 AI란 제공된 데이터를 통해 학습한 내용을 바탕으로 단순한 분석이 아닌 텍스트, 사진, 이미지, 비디오 및 3D 등 다양한 형태의 새로운 콘텐츠를 생성할 수 있는 인공지능을 의미함

2) 생성형 AI관련 국내 동향

□ 국내 정책 동향

- 현재, 우리나라는 생성형 AI에 관한 필요성을 중요하게 인식하는 동시에 관련 정책에 관하여는 규제 관점에서 접근하고 있음
 - 2023년 6월에 국가정보원이 발표한 ‘ChatGPT 등 생성형 AI 활용 가이드 라인’이 생성형 AI에 대한 대표적인 규제정책임(국가정보원, 2023.06.29.)
 - 생성형 AI로 인해 초래되는 대표적인 보안 위협을 잘못된 정보, 생성형 AI 모델 악용, 유사 AI 모델 서비스 빙자, 데이터 유출, 플러그인 취약점, 확장 프로그램 취약점, API 취약점 등 8가지로 구분함
- 더하여, 2023년 6월 산업통상자원부는 AI 윤리 국가표준(KS)을 제정·발표함(산업통상자원부, 2023.06.14.)
 - 2021년부터 산·학·연 전문가들이 참여하는 ‘AI 윤리 표준화 포럼’에서 논의와 검증 과정을 거쳐 표준을 개발함. 해당 표준은 최근 챗GPT 등 생성형 AI 서비스의 확산과 함께 AI의 윤리적인 사용이 제기되는 시점에서 AI 제품·서비스 개발 시에 필요한 윤리적 고려항목을 제시하고 자체 점검할 수 있는 체크리스트로 활용됨(산업통상자원부, 2023.06.14.)

□ 공공데이터를 활용한 생성형 AI 서비스 국내 사례

- 국내 공공분야 생성형 AI 활용 현황은 아래의 [표 2-3]과 같음

[표 2-3] 국내 공공분야 생성형 AI 활용 현황

기관	내용
디지털플랫폼 정부위원회	공공 데이터의 AI 학습이 수월하도록 회의록과 보도자료를 비롯한 정부 부처 문서에 개방형 포맷인 ODT(Open Document Text)를 도입(디지털플랫폼 정부위원회, 2022.12.30.)
문화체육관광부, 국립국어원	‘한국어를 잘하는 K-챗GPT’ 개발을 지원하기 위해 올해 25종 약 1억 2천만 어절의 고품질 한국어 말뭉치를 구축하고 2027년까지 10억 어절로 구축을 확대할 계획

기관	내용
경상북도	전국 지자체 최초로 대형 언어모델 기반 AI 챗봇 '챗경북' 서비스 시범 운영 (2023.3.7.)
경기도	ChatGPT를 행정 도구로 활용함으로써 창업 및 일자리 창출을 지원하는 목표를 세우고, AI 콜센터 도입을 비롯해 발달장애인 대상 예술교육 및 전시회 추진 등이 담긴 '경기GPT 추진 계획' 발표(2023.3.21.)

출처: 박상명, 조은결(2023), 「ChatGPT, 국내외 공공 부문 도입 동향과 지자체 활용방안」, 내용 재구성

- 공공데이터를 활용한 생성형 AI 서비스는 주로 공공데이터포털⁸⁾을 통해 공급된 데이터를 통하여 민간을 주체로 개발되고 있음
- 공공데이터를 활용한 민간의 생성형 AI 서비스 제공 국내 사례를 구체적으로 살펴보면 다음과 같음
 - 시각장애인 맞춤형 인공지능 기반 편리한 장보기 서비스⁹⁾
 - (활용데이터) 상품이미지, 식품 원재료 정보, 바코드연계제품정보 등
 - (서비스 내용) 저시력 시각장애인을 위한 상품 인식 서비스를 제공. 인접 매장 안내, 카메라 촬영을 통한 상품정보 제공함
 - 원격협진 시스템¹⁰⁾
 - (활용데이터) 전국 병원 의원코드, 의료기관 상세정보 서비스, 병원원 평가정보 등
 - (서비스 내용) AI기반 매칭을 통한 1차 의료기관 협진 플랫폼을 구축하여 질 좋은 의료서비스를 제공함
 - 실종자 객체인식 통합 플랫폼¹¹⁾
 - (활용데이터) 지역별 치매노인 현황, 스톱킹 신고 현황, 성범죄자 출소

8) 공공데이터포털(n.d.) <https://www.data.go.kr> (검색일: 2023.12.20.)

9) <https://www.data.go.kr/tcs/eds/ctm/selectContestData.do?contestId=40105>

10) <https://www.data.go.kr/tcs/eds/ctm/selectContestData.do?contestId=40103>

11) <https://www.data.go.kr/tcs/eds/ctm/selectContestData.do?contestId=40098>

후 학교 및 보육시설 인근 거주 현황, 국가 재난상황 감염병 현황, 전국 CCTV통합관제센터 현황

- (서비스 내용) 초고령사회 치매노인 실종 예방, 성 스토킹 범죄 및 이상동기 범죄 사전 예방을 목적으로 AI 인공지능 실종자 객체인식 통합 플랫폼 적용을 모두가 안전한 사회 구축함
- 국내외 공공·민간의 생성형 AI 활용 현황을 살펴봄으로써 다음과 같은 특성을 도출함
- 첫째, 공공영역이 생성형 AI의 개발과 학습에 활용되는 데이터 제공 역할을 수행하고 있음
- 둘째, 생성형 AI 서비스 생산자는 공공 및 민간 모두에 해당함
- 마지막으로, 제공하는 생성형 AI기반 서비스의 종류가 다양하며, 수요 범위는 일반국민에 걸쳐 장애인, 노인, 학습자, 교육자 등 특정된 대상을 포함함



제3장

법령사무조사 자동화 모델의 개발

제1절 자동화 모델 개발 설계

제2절 자동화 모델의 개발

제1절 자동화 모델 개발 설계

1. 자동화 모델 개발 개요

- 본 연구는 한국지방행정연구원에서 2022년에 수행한 「2022년 법령상 사무총조사」에서 수행한 데이터를 활용하여 생성형 AI를 통한 사무판단, 사무유형 예측을 위한 모형을 개발하는 것을 목적으로 함
 - AI를 활용한 자동화 모델 개발을 통하여 법령상 사무총조사의 자동화 조사 가능성 및 실현가능성을 모색하고자 함
- 법령사무조사 자동화 모델 개발은 프로그래밍 전문기관인 (주)씨지인사이드에의 위탁 및 KPMG와의 업무협력을 통하여 자동화 모델 개발자 7명이 2023년 11월 27일부터 12월 27일까지 30일까지 개발에 착수하였음

2. 자동화 모델 개발 내용

- 엑셀의 형태로 크롤링 된(텍스트화 된) 법령정보를 컴퓨터가 이해할 수 있는 언어로 벡터화 함
 - 텍스트 데이터 특성상, 텍스트를 컴퓨터가 이해할 수 있는 언어로 벡터화 작업이 필요함. 이를 위해 단순히 okt로 하는 방법도 사용하고, 딥러닝 모델은 RoBERTa 토큰라이저를 사용하여 (문맥을 이해할 수 있도록) 조문 텍스트를 벡터화함

- 법률, 시행령, 시행규칙 등 법령상 조문의 내용이 법령상 사무인지 여부를 판단함
 - 법령상 조문은 조와 항 수준까지 사무 여부를 판단함
 - 2022년 법령상 사무총조사에서는 조문 중 호의 경우는 원칙적으로 사무로 판단하지 않았지만 예외적으로 사무로 인정되는 경우까지 상정하였으나, 본 연구에서는 호 단위의 사무는 일괄적으로 판단하지 않음
 - 사무 여부를 판단은 사무(1)와 비사무(0)로 분류함
- 사무로 판단된 경우, 사무유형을 분류함
 - 2022년 법령상 사무총조사에서는 대분류(3가지)와 중분류(7가지), 소분류(16가지)로 세분화하여 구분하였으며, 본 연구에는 이와 동일한 분류를 AI가 이끌어낼 수 있도록 프로그래밍의 개발을 원칙적으로 목표함
 - 다만, 본 연구가 법령사무조사의 자동화에 대한 기초연구이기 때문에 현재의 여건과 수준 하에서 어느 수준까지 사무유형을 분류해 낼 수 있는지 확인하는 것을 보다 우선적으로 확인하고자 함
 - 즉, 사무유형 배분의 중요도를 고려하였을 때, 사무의 세무 유형으로의 구분 가능성 자체보다는 정확도가 가장 높은 유형 분류는 어느 수준인지 까지 확인하는 것을 보다 중점적으로 확인하고자 함

[표 3-1] 2022 법령상 사무총조사의 사무 구분 기준

대분류	중분류	소분류	세부내용
국가 사무	국가 사무	1. 국가직접처리사무 (국가소속/산하기관 사무 포함)	000장관, 000부의 사무 및 그 소속·산하 기관의 사무 소속·산하 기관은 수행 주체란에 해당 기관명 명시 소속·산하 기관이 직접 수행하거나 위임받아 수행하는 경우
		2. 특별지방행정기관사무	개별 법령 및 행정권한의 위임 위탁 규정을 통해 특별 기관이 수행하는 사무 수행 주체란에 해당 기관명 명시
		3. 국가위탁사무	원칙적으로 국가(000장관, 000부)의 사무이나, 법령의 규정에 따라 법인, 민간단체 등에 위탁하는(위탁할 수 있는) 사무는 수행주체란에 사무수행기관명 명시

대분류	중분류	소분류	세부내용
	위임사무	4. 시도위임사무	원칙적으로 국가(000장관, 000부)의 사무이나, 법령의 규정에 따라 시도에 위임하는(위임할수 있는) 사무
		5. 시군구위임사무	원칙적으로 국가(000장관, 000부)의 사무이나, 법령의 규정에 따라 시군구에 위임하는(위임할수 있는) 사무
		6. 시군구재위임사무	원칙적으로 국가(000장관, 000부)의 사무이나, 법령의 규정에 따라 시도에 위임하는(위임할수 있는) 사무 중 다시 시군구에 재위임하는(재위임할 수 있는) 사무
		7. 시도 및 시군구위임사무	원칙적으로 국가(000장관, 000부)의 사무이나, 법령의 규정에 따라 시도 및 시군구 지방자치단체)에 공동으로 위임하는(위임할수 있는) 사무
국가-지방자치단체 공동사무		8. 국가+시도+시군구 공동사무	국가와 지방자치단체(국가, 시도 및 시군구)가 공동으로 수행하는 사무 ※ 법조문의 호·목, 하위법령 등을 검토하여 대상별 분리가 불가능한 경우에만 공동사무로 분류, 분리 가능한 경우 각각 사무수행주체별로 사무 추출·유형 분류
		9. 국가+시도 공동사무	국가 및 시도가 공동으로 수행하는 사무 ※ 법조문의 호·목, 하위법령 등을 검토하여 대상별 분리가 불가능한 경우에만 공동사무로 분류, 분리 가능한 경우 각각 사무수행주체별로 사무 추출·유형 분류
		10. 국가+시군구 공동사무	국가 및 시군구가 공동으로 수행하는 사무 ※ 법조문의 호·목, 하위법령 등을 검토하여 대상별 분리가 불가능한 경우에만 공동사무로 분류, 분리 가능한 경우 각각 사무수행주체별로 사무 추출·유형 분류
지방 자치 단체 사무	시·도 사무	11. 시·도직접처리사무 (시도소속/산하사무 포함)	시·도(시·도지사) 및 그 소속/산하기관이 직접 수행하는 사무 소속·산하 기관의 경우 비교란에 해당 기관명 명시
		12. 시·도위탁사무	원칙적으로 시·도의 사무이나, 법령의 규정에 따라 법인, 민간단체 등에 위탁하는(위탁할 수 있는) 사무 비교란에 사무수행기관명 명시

대분류	중분류	소분류	세부내용
	시·군·구 사무	13. 시·군·구직접처리사무 (읍·면·동, 시·군·구 소속·산하사무 포함)	시·군·구(시장·군수·구청장), 읍·면·동 및 그 소속·산하 기관이 직접 수행하는 사무 소속·산하 기관의 경우 비교란에 해당 기관명 명시
		14. 시·군·구위탁사무	원칙적으로 시군구의 사무이나, 법령의 규정에 따라 법인, 민간단체 등에 위탁하는(위탁할 수 있는) 사무 비교란에 사무수행기관명 명시
	지방자치단체 공동사무	15. 시·도-시·군·구 공동사무	지방자치단체가 수행하도록 규정되어 있거나 시·도 및 시·군·구가 공동으로 수행하는 사무 ※ 법조문의 호·목, 하위법령 등을 검토하여 대상별 분리가 불가능한 경우에만 공동사무로 분류, 분리 가능한 경우 각각 사무수행주체별로 사무 추출·유형 분류
	지방자치단체 위임 사무	16. 시·도-시·군·구 위임사무	원칙적으로 시·도의 사무이나, 법령의 규정에 따라 시·군·구로 위임하는(위임할 수 있는) 사무

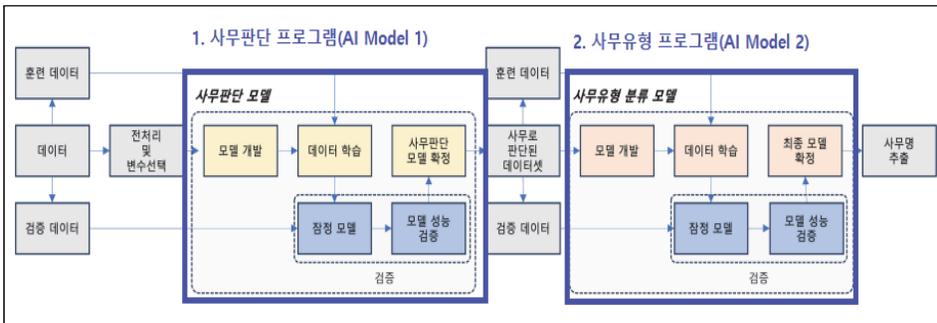
※ ‘특별자치시장·특별자치도지사·광역시장·특별시장·도지사’ → 시·도사무
 ‘특별자치시장·특별자치도지사·시장·군수·구청장’ → 시장·군수·구청장사무
 ‘특별자치시장, 특별자치도지사 또는 시장·군수·구청장’ → 시장·군수·구청장사무

- 추출된 사무에 대하여 AI가 사무명을 부여함
 - 추출된 사무에 대하여 챗GPT 등의 생성형 AI를 통하여 사무명을 추출함
- 2022년 5월 기준의 법령으로 학습된 자동화 모델을 토대로 2023년 5월 이후 제·개정된 법령에 대하여 사무판단, 사무유형 등을 예측함
 - 2023년 법령 데이터는 법제연구원에서 2023년에 수행한 「2023년 법령상 사무총조사」 데이터를 협조받아 분석하였음

3. 자동화 모델 개발 절차

- 위의 내용을 토대로 생성형 AI를 활용한 법령사무조사 자동화 기초연구를 위한 개발 과정은 다음과 같음
 - 수집된 데이터를 바탕으로 전처리, 변수선택, 개발, 검증 및 단계를 통해 이루어짐. 데이터를 전처리하여 분석 가능한 형태인 정형 데이터로 가공한 뒤, 모델 개발 과정을 거침
 - 먼저, 사무판단 모델을 개발함. 학습을 통해 모델을 도출하며 데이터 학습으로 도출한 잠정 모델의 성능을 검증하고, 최종 모델을 선정함
 - 다음으로, 사무유형 분류 모델을 개발함
 - 이 과정을 통해서 도출된 사무라고 판단된 데이터셋만을 이용하여 위와 동일한 과정을 반복함
 - 최종적으로 모델이 확정되면, 해당 조문에 대한 사무명을 추출함

[그림 3-1] 법령사무조사 자동화 모델 개발 절차



제2절 자동화 모델의 개발

1. 데이터 전처리 및 데이터 분할

□ 데이터 전처리

- 본 개발에 활용된 데이터는 2022년 5월 기준으로 한 법령의 내용을 엑셀로 크롤링한 방식을 적용하고 있기 때문에 데이터 데이터 타입을 통일하고, 결측치를 처리하는 등의 과정이 필요함
 - 설명변수 : 소관부처명, 법령명, 조번호, 항번호, 호번호, 조문제목, 조문
 - 반응변수 : 사무판단, 사무명, 사무유형, 사무유형(소분류)

[표 3-2] 데이터 요약

구분	연번	변수명	변수타입	설명
	1	연번	수치형	
사전변수	2	소관부처명	문자형	소관부처명
	3	법령명	문자형	
	4	법령구분	명목형	1 : 법령 2 : 시행령 3 : 시행규칙
	5	조번호	문자형	
	6	항번호	문자형	
	7	호번호	문자형	
	8	조문제목	문자형	
	9	조문	문자형	
	사후변수	10	사무판단	명목형
11		사무판단 근거	문자형	
12		사무명	문자형	
13		수행주체	문자형	
14		사무유형	명목형	1 : 국가, 2 : 시도, 3 : 지방, 4 : 국가-시도-시군구, 5 : 국가-시도, 6 : 국가-시군구, 7 : 시도-시군구

구분	연번	변수명	변수타입	설명
	15	위임사무판단	명목형	1 : 위임사무
	16	위임근거규정	문자형	
	17	수임기관	문자형	
	18	특행기관	명목형	1 : 수임기관이 특행기관
	19	재위임사무판단	명목형	1 : 재위임사무
	20	재위임근거규정	문자형	
	21	재수임기관	문자형	
	22	위탁사무판단	명목형	1 : 위탁사무
	23	위탁근거규정	문자형	
	24	수탁기관	문자형	
	25	사무유형(소분류)	명목형	16개 분류
	26	기타	문자형	

○ 분석 대상 데이터는 값이 존재하지 않고 대신 데이터의 부재를 나타내는 값(결측치, null)로 입력되어 있으나, 본 연구에서 활용하고자 하는 머신러닝이나 딥러닝 등의 모델은 이러한 결측값을 처리하지 못하기 때문에 다른 값으로 대체해야 함

- 설명변수(소관부처명, 법령명, 조번호, 항번호, 호번호, 조문제목, 조문)가 모두 결측치인 행은 삭제처리(총 8행)
- 법령명은 있으나, 소관부처가 결측치인 행의 법령명들은 소관부처 별도 입력 (36개 행)
- 2개 이상의 소관부처명에 대해 순서를 동일하게 변경하였음
 - 예컨대, 소관부처명이 동일하지만, 순서가 달라 별개의 소관부처로 인식하고 있기 때문에 교육부, \n과학기술정보통신부 / 과학기술정보통신부, \n교육부 총 11개 행을 → 과학기술정보통신부, 교육부로 통일시키는 작업을 거침

[그림 3-2] 2개 이상의 소관부처명 순서 변경 처리 및 소관부처 순서가 다른 경우 처리(예시)

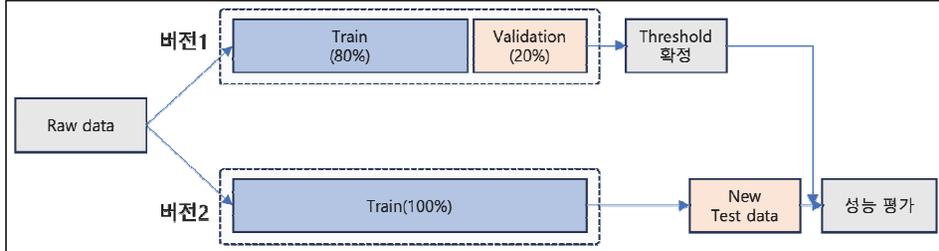
<pre>def blank_preprocessing(df): df.loc[df['소관부처명']=="교육부,Wn 고용노동부", "소관부처명"] = '고용노동부,교육부' df.loc[df['소관부처명']=="과학기술정보통신부, Wn 교육부", "소관부처명"] = '과학기술정보통신부,교육부' df.loc[df['소관부처명']=="교육부,Wn 과학기술정보통신부", "소관부처명"] = '과학기술정보통신부,교육부' return df</pre>	<table border="0"> <tr> <td>교육부,Wn 고용노동부</td> <td style="text-align: right;">109</td> </tr> <tr> <td>고용노동부,교육부</td> <td style="text-align: right;">769</td> </tr> <tr> <td>과학기술정보통신부,교육부</td> <td style="text-align: right;">133</td> </tr> <tr> <td>교육부,Wn 과학기술정보통신부</td> <td style="text-align: right;">7</td> </tr> <tr> <td>과학기술정보통신부, Wn 교육부</td> <td style="text-align: right;">4</td> </tr> </table>	교육부,Wn 고용노동부	109	고용노동부,교육부	769	과학기술정보통신부,교육부	133	교육부,Wn 과학기술정보통신부	7	과학기술정보통신부, Wn 교육부	4
교육부,Wn 고용노동부	109										
고용노동부,교육부	769										
과학기술정보통신부,교육부	133										
교육부,Wn 과학기술정보통신부	7										
과학기술정보통신부, Wn 교육부	4										

- 법령구분과 소관부처가 모두 결측치인 행은 국가법령정보센터 데이터 확인 후, 해당 법령명과 일치하는 법령구분(법령, 시행령, 시행규칙) 결측값 입력 (총 9개 행)
 - 법령구분의 결측치에 대하여 법령이 “법률”에 해당하는 것을 확인 후 법령구분 1로 처리(8개 행)
 - 조문과 조문제목이 모두 결측치인 행은 0으로 처리(59개 행)
- 앞서 논의한 결측치 처리 및 사무판단, 사무판단 근거, 사무유형 등의 구체적 처리 과정은 [부록1]에 첨부함

□ 데이터 분할

- 이 모형을 개발하는 과정에서 데이터 분할과 분할하지 않는 방법을 모두 사용하여, 목적에 따라 달리 사용하였음
 - 버전1 : 데이터 분할 후 학습 → 검증용 데이터로 최적의 임계값 확정
 - 버전2 : 전체 데이터 학습 → 버전1에서 설정한 임계값 적용 → 새로운 데이터 테스트

[그림 3-3] 법령구분 자료의 데이터 분할 과정



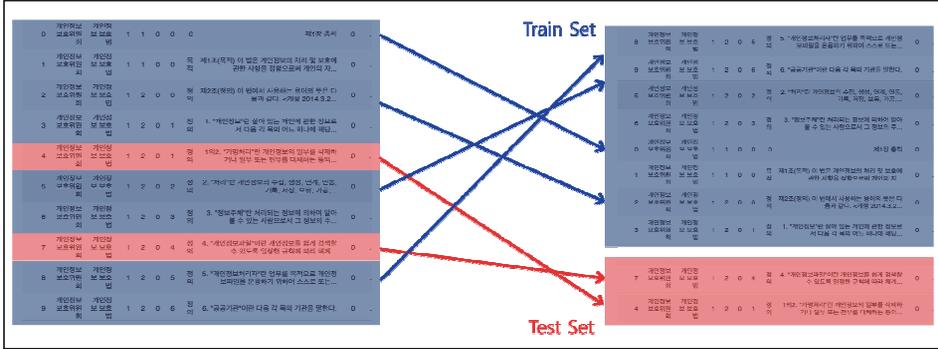
○ 버전 1: 데이터 분할 및 임계값 설정

- 전체 데이터셋을 학습용(Training)과 검증용(Validation)으로 8:2의 비율로 분할하고, 모델의 성능을 평가하기 위하여 최적의 임계값을 설정함
 - 데이터 분할은 모델이 학습 데이터에 과적합되는 것을 방지하고, 모델의 예측 성능을 더 신뢰할 수 있게 검증할 수 있도록 함
 - 임계값은 모델이 어떤 샘플을 특정 클래스(예: 사무 판단)에 속한다고 예측할지를 결정하는 기준이 됨

○ 버전 2: 전체 데이터 학습 및 새로운 데이터 테스트

- 전체 데이터 학습: 이후, 전체 데이터셋(100%)을 사용하여 모델을 재학습함. 이 과정은 모델이 가능한 한 많은 데이터로부터 학습하여 일반화 성능을 최대화 하도록 하기 위함임
- 새로운 테스트 데이터에 적용: 마지막 단계에서는 새로운 테스트 데이터셋에 모델을 적용함. 개정된 법안이나, 새로 시행된 법들이 이에 해당함
- 이 때, 2번 단계에서 결정된 임계값을 사용하여 예측을 수행함. 이렇게 하면 모델이 이전에 보지 못한 데이터에 대해 어떻게 작동하는지 평가할 수 있음
- 이 프로세스를 적용하는 이유는 기존에 학습하지 않았던 검증용 데이터 (20%)를 모두 활용하여 재학습함으로써, 새로운 또는 개정된 법안에 대한 예측 정확도를 높일 수 있고, 새롭게 수집된 데이터를 테스트 데이터로 활용하여 모델이 실제 개정 조문에 어떻게 적용되는지 평가할 수 있기 때문임

[그림 3-4] 데이터 분할 방식



2. (AI Model 1) 사무판단 모델의 개발

[그림 3-5] 사무판단 모델(AI Model 1)의 흐름



1) 확실한 ‘비사무’ 데이터 추출: 규칙 기반 변수인 rule-based 변수

- 본 연구를 통해서 머신러닝 및 딥러닝과 같은 인공지능 모델을 활용하였을 때, 인간의 라벨링 작업량을 얼마나 줄일 수 있는지에 대해서 탐구하기 위해서 명확히 사무가 아닌 조문들은 사전에 필터링할 수 있는 방법을 모색함
 - 규칙 기반의 ‘rule_based’ 컬럼은 전체 법령 중 사무가 아닌 조문을 사전에 필터링하기 위한 과정임

- 즉, 전체 법령 중 비사무인 조문들의 규칙의 발견을 통하여 향후 법령사무 조사를 위한 자동화 연구에서 인간의 판단을 확실하게 제외할 수 있는 ‘비사무’를 추출하기 위한 과정임
- 본 과정을 통하여 35개의 비사무 규칙을 생성하였으며, 이에 따른 사항을 ‘rule-based 변수’로 변환함
 - ‘rule_based’ 컬럼은 각 행이 사무가 아닌 것으로 확실히 판단될 경우 0으로 라벨링하였고, 이를 통해 총 159,947개의 행을 사무가 아닌 것으로 분류하였음
 - 여기서 0은 ‘rule_based’의 규칙에 따라 명확히 사무가 아닌 것으로 판단되는 조문을 나타냄

【표 3-3】 rule-base 규칙을 적용한 데이터 형태와 제거되는 행 개수

연번	규칙	형태	정규표현식	해당되는 행 개수
1	조문에 제 ~조(조문제목) 혹은 제 ~조로 구성된 행	제3조(개인정보 보호 원칙) 제9조	r'^제.*\)\$' r'제\d+조\$'	79,451
2	조문이 제 ~장, 제 ~절, 제~편	제1장 총칙 / 제2절 개인정보의 처리 제한	^제\d+(장 절 편)'	12,932
3	조문제목이 '목적' 또는 '정의'인 행		-	15,609
4	조문이 삭제 < > 혹은 삭제< >인 행	③ 삭제 <2020.2.4>	-	20,211
5	호 또는 조문에 '목적'으로만 구성되어 있는 행	1.목적 / 목적	-	591
6	호에 '명칭'을 포함하거나 조문이 '명칭'으로만 구성되어 있는 행	1.취수시설의 명칭 / 명칭	-	3,936
7	호가 '있는 경우'로 끝나는 행	(2.)* 다른 법률에 특별한 규정이 있는 경우	r'.*있는 경우\$'	2,578
8	호가 '없게 된 경우'로 끝나는 행	(1.) 장기간 심신장애로 인하여 직무를 수행할 수 없게 된 경우	r'.*없게 된 경우\$'	694
9	호에 '거짓이나 그 밖의 부정한 방법으로 지정을 받은 경우'가 포함된 행	1. 거짓이나 그 밖의 부정한 방법으로 지정을 받은 경우	r'\d+\..*거짓이나 그 밖의 부정한 방법으로 지정을 받은 경우'	280

연번	규칙	형태	정규표현식	해당되는 행 개수
10	호에 '의사를 밝히는 경우'가 포함된 행	(5.) 위원 스스로 직무를 수행하는 것이 곤란하다고 의사를 밝히는 경우	r'.*의사를 밝히는 경우'	481
11	호에 '인정되는 경우'가 포함된 행	3. 직무태만, 품위손상이나 그 밖의 사유로 위원으로 적합하지 않다고 인정되는 경우	r'\d+\..*인정되는 경우'	1,614
12	호에 '친족이었던 경우'가 포함된 행	(2.) 위원이 해당 사안의 당사자와 친족이거나 친족이었던 경우	r'.*친족이었던 경우'	248
13	호에 '대리인이었던 경우'가 포함된 행	(4.) 위원이나 위원이 속한 법인이 심의 안건 청구인의 대리인이거 나 대리인이었던 경우	r'.*대리인이었던 경우'	196
14	호에 '인증을 받은 경우'가 포함된 행	거짓이나 그 밖의 부정한 방법으로 인증을 받은 경우	r'\d+\..*인증을 받은 경우'	123
15	호가 '후견인'으로 끝나는 행	(5.) 피성년후견인 및 피한정후견인	r'.*후견인\$'	300
16	호에 '소재지'가 포함된 행	3. 주된 사무소의 소재지	r'\d+\..*소재지'	2,096
17	호에 '정관'이 포함된 행	9. 정관의 변경에 관한 사항	r'\d+\..*정관'	1879
18	호에 '계획서'가 포함된 행	3. 위해예방계획서	r'\d+\..*계획서'	2,426
19	호에 '서류'가 포함된 행	2. 제1호의 사실을 증명할 수 있는 서류	r'\d+\..*서류'	8,084
20	호가 '상호'로 끝나는 행	3. 법인명 또는 상호	r'\d+\..*상호\$'	157
21	호에 '증명서'가 포함된 행	8. 최종학교졸업증명서 또는 학력을 증명하는 서류	r'\d+\..*증명서'	2,393
22	호에 '경매'가 포함된 행	1. 「민사집행법」에 따른 경매	r'\d+\..*경매'	232
23	호에 '등록증'이 포함된 행	2. 주민등록증 사본 1부	r'\d+\..*등록증'	1,009
24	호가 '주소'로 끝나는 행	(3.) 소유자의 성명 및 주소	r'.*주소\$'	1,156
25	호가 '절차'로 끝나는 행	(1.) 국민의 권리·의무와 직접 관계되는 경찰행정 및 수사절차	r'.*절차\$'	575
26	호에 '출연금'이 포함된 행	1. 국가의 출연금 또는 보조금	r'\d+\..*출연금'	385
27	호가 '정도'로 끝나는 행	(1.) 고의 또는 손해 발생의 가능성을 인식한 정도	r'.*정도\$'	352
28	호에 '대표자'가 포함된 행	1. 상호 및 대표자의 성명	r'\d+\..*대표자'	1,888
29	호에 '공고의 방법'이 포함된 행	5. 공고의 방법	r'\d+\..*공고의 방법'	184

연번	규칙	형태	정규표현식	해당되는 행 개수
30	호에 '대통령령으로 정하는 사항'이 포함된 행	7. 그 밖에 대통령령으로 정하는 사항	r'\d+\.\ .*대통령령 으로 정하는 사항'	947
31	호에 '되지 아니한 자'가 포함된 행	6. 파산선고를 받고 복권되지 아니한 자	r'\d+\.\ .*되지 아니한 자'	215
32	호에 '그 밖에 필요한 사항'이 포함된 행	4. 그 밖에 필요한 사항	r'\d+\.\ .*그 밖에 필요한 사항'	161
33	조문에 '협회는 법인으로 한다'가 포함된 행	(②) 협회는 법인으로 한다.	r'.*협회는 법인으로 한다'	113
34	호에 이사회가 포함되어 있고 사항으로 끝나는 행	(5.) 이사회에 관한 사항	r'.*이사회.*사항\$'	337
35	호에 회계가 포함되어 있고 사항으로 끝나는 행	(7.) 재정 및 회계에 관한 사항	r'.*회계.*사항\$'	416

2) 사무판단 모델의 개발

□ 파생변수의 생성

- 딥러닝 모델은 그 내부 구조에 숨겨진 여러 계층(hidden layer) 때문에, 모델이 어떤 원리로 특정 결과를 도출하는지 완전히 이해하는 데 한계가 있음
- 본 연구에서는 '사무'와 같은 업무특성을 담은 조문은 특정 단어의 빈도가 중요한 역할을 할 것이라는 가설을 설정하고, 이를 바탕으로 통계적 기법을 적용한 파생변수를 생성하였음
- 이를 통해 업무와 관련된 주요 키워드의 중요성을 파악하고 이를 머신러닝 모델의 입력 변수로 활용함

□ 단어의 토큰화 : Okt(Open Korea Text)

- 한국어로 작성된 조문에 사용된 단어의 빈도수를 알기 위해서는 텍스트를 명사 단위로 분할해야 하는 작업이 필요함

- Okt는 한국어 처리를 위한 자연어(NLP) 라이브러리 중 하나로 트위터에서 개발된 오픈 소스 프로젝트인 KoNLPy(코엔엘파이)의 일부임
 - 한국어 텍스트의 토큰화(tokenization), 형태소 분석(morphological analysis), 품사 태깅(part-of-speech tagging) 등 다양한 기능을 제공함
- 본 연구에서는 조문 텍스트를 형태소 단위로 분리한 뒤, okt.nouns 메소드를 사용하여 명사만 추출한 후, 2음절 이상인 명사들만 추출하여 '조문_명_동' 변수를 생성함
- '각', '외'와 같은 한국어에서 한 글자 명사는 종종 맥락상 중요하지 않거나 모호할 수 있고 두 글자 이상의 명사는 보통 더 구체적이고 명확한 정보를 제공하기 때문임
 - 토큰화 결과는 '조문_명_동'이라는 칼럼에 저장함

[표 3-4] 조문 텍스트의 Okt 기반 2음절 이상 명사 토큰화 예시

조문	조문_명_동
국가와 지방자치단체는 개인정보의 목적 외 수집, 오용·남용 및 무분별한 감시·추적 등에 따른 피해를 방지하여 인간의 존엄과 개인의 사생활 보호를 도모하기 위한 시책을 강구하여야 한다.	['국가', '지방자치단체', '개인정보', '목적', '수집', '오용', '남용', '무분별', '감시', '추적', '피해', '방지', '인간', '존엄', '개인', '사생활', '보호', '도모', '시책', '강구', '여야']
제6조(다른 법률과의 관계) 개인정보 보호에 관하여는 다른 법률에 특별한 규정이 있는 경우를 제외하고는 이 법에서 정하는 바에 따른다. <개정 2014.3.24>	['다른', '법률', '관계', '개인정보', '보호', '다른', '법률', '규정', '경우', '제외', '개정']

□ 변수를 생성하기 위한 단어 파악 및 dictionary 생성

- 생성된 단어들을 한 행별 중복되는 단어는 1번만 count 하도록 set 함수를 사용하여 처리하고, 모든 명사에 대해서 빈도수를 계산한 dictionary를 생성하였음
 - set 자료형은 Python에서 중복을 허용하지 않는 데이터 구조로, 각 행에서 중복된 명사를 하나만 남기고 나머지를 제거하는 데 사용됨
- 이렇게 각 행에서 중복을 제거한 후, 모든 행의 명사들을 하나의 리스트에 모으고, 이 리스트를 사용하여 각 명사의 전체적인 출현 빈도를 계산함

- 최종적으로 생성된 딕셔너리는 각 명사를 키(key)로, 해당 명사가 데이터셋 전체에서 나타난 빈도수를 값(value)으로 가지며 형태는 다음과 같음

[그림 3-6] 데이터셋 전체의 빈도수

```

{'경우': 177678,
 '개정': 124778,
 '다음': 93838,
 '호의': 93357,
 '사항': 93190,
 '해당': 92086,
 '장관': 75453,
 '사업': 59113,
 '이하': 52596,
 ...

```

□ 생성한 변수 추가

- 위의 작업을 바탕으로 새롭게 생성된 4가지 변수들은 다음과 같다.
 - score_length : ['조문_명_동']에 포함된 단어들의 개수
 - 가설 : 조문이 길수록(토큰화되는 명사가 많을 수록)사무가 많을 것이다.
 - score_count : ['조문_명_동']에 포함된 단어들 중 사무판단이 1에 있는 단어들의 갯수
 - score_ratiomean : ['조문_명_동']에 포함된 단어들의 사무판단의 1인 비율의 평균
 - score_ratiomulti : ['조문_명_동']에 포함된 단어들 중 사무판단의 1인 비율 상위 3개의 곱
 - 가설 : 조문의 '사무'가 성립하려면 '_가 _을 _한다.'와 같은 형태로 사무가 되려면 최소 단어가 3개는 들어가야 할 것이라는 가정으로 생성한 변수

□ 사무판단 모델: Transformer Encoder 활용

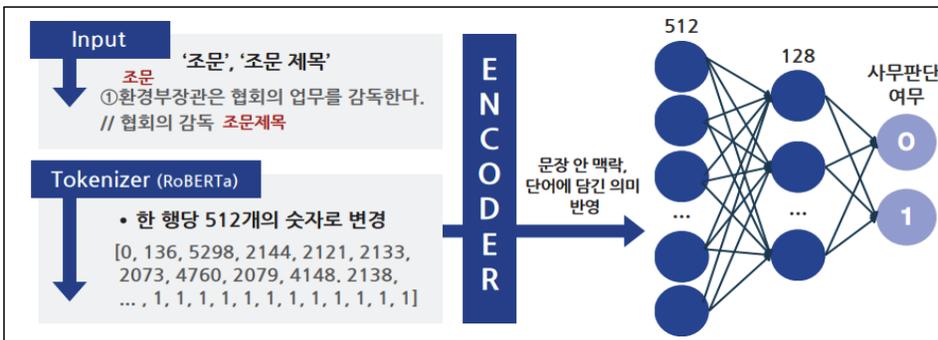
- Transformer는 딥러닝 방법의 하나로서, 인공지능망이 스스로 학습하며 성능 개선을 진행하는 방식을 취함
 - Transformer 모델은 문장 안의 맥락과 단어에 담긴 의미 등을 학습하는 신경망을 구성하는 방식을 택함

[그림 3-7] Transformer 모델

Transformer 설명	Transformer 활용 사례
<ul style="list-style-type: none"> • 구글이 2017년 공개한 언어 처리 모델 • 딥러닝의 방법 중 하나로서, 인공지능망이 스스로 학습하며 성능 개선 진행 • Transformer 모델은 문장 안의 맥락과 단어에 담긴 의미 등을 학습하는 신경망 구성 방식 <p>예시) She is eating a green apple.</p> <p style="margin-left: 40px;">높은 attention 낮은 attention</p>	<ul style="list-style-type: none"> OpenAI 'ChatGPT' <ul style="list-style-type: none"> • GPT: Generative Pre-trained Transformer • ChatGPT는 GPT-3을 텍스트 생성 및 답변 제공에 용이하도록 fine-tuning한 것 네이버 'Papago' <ul style="list-style-type: none"> • BERT(Bidirectional Encoder Representations from Transformers)를 기계번역에 사용하여 품질 향상 Facebook 'RoBERTa' <ul style="list-style-type: none"> • RoBERTa(Robustly Optimized BERT Approach) 모델을 만들어 사용자 감성분석, 스팸 탐지, 뉴스 분류 등에 사용

- Transformer 모델을 활용한 사무판단 모델(AI Model 1)은 조문의 텍스트를 토큰화하여 숫자로 변경한 후 문장 안의 맥락과 단어에 담긴 의미를 반영하여 Transformer Encoder 과정을 거침

[그림 3-8] 사무판단 모델(AI Model 1): Transformer Encoder 활용



□ 모델의 성능 비교

- 본 모델을 통하여 사무와 비사무 여부를 판단할 수 있으며, 이때 추출된 비사무는 인간의 판단 영역에서 1차적으로 제외될 수 있기 때문에 정확도가 가장 중요한 항목임
 - 비사무를 사무로 판단하는 경우, 세부적인 사무 판단과정에서 오류를 시정할 수 있는 기회가 제공될 수 있지만, 사무를 비사무로 판단하는 경우 추가적 검증 과정의 기회가 원천적으로 배제될 수 있음
 - 따라서 사무로 추출되는 정확도보다 비사무를 추출하는 정확도가 훨씬 더 중요하다고 판단하였으며, 본 과정을 수행할 수 있는 여러 모델 중 가장 적합한 모델(2-layer 신경망 with Transformer Encoder)을 선정함

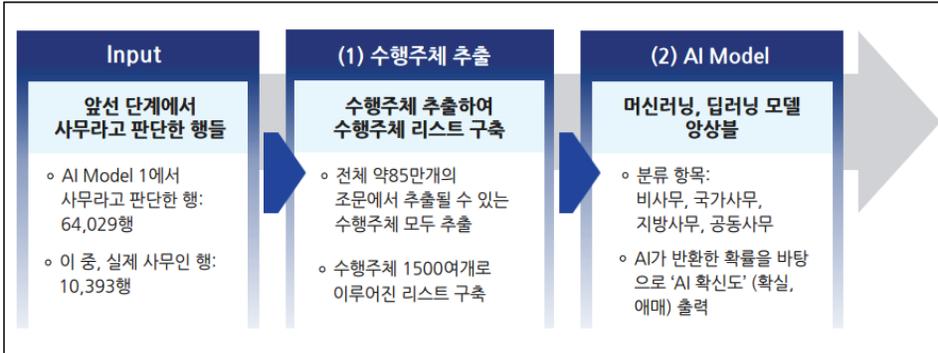
[그림 3-9] 사무판단 모델(AI Model 1) 방법론별 적용 결과

*train : test = 8 : 2 split, test set에 대한 결과

모델명	Recall	모델명	Recall
Random Forest	93.9%	2-layer 신경망	94.3%
LightGBM	94.7%	2-layer 신경망 with Transformer Encoder	94.9%
Ensemble	94.7%		

3. (AI Model 2) 사무유형 모델의 개발

[그림 3-10] 사무유형 모델(AI Model 2)의 흐름



1) 사무유형 모델의 선정

- 사무유형 분류 모델은 사용한 변수의 종류에 따라 달리 설정함
 - 랜덤포레스트(Random Forest)는 분류(Classification) 및 회귀(Regression) 문제에 널리 사용되는 앙상블 학습(Ensemble Learning) 기법 중 하나로서, 이 알고리즘의 핵심은 여러 의사결정 트리(Decision Trees)를 결합하여 하나의 모델을 형성하는 것임
 - 2-layer neural network는 간단한 인공 신경망 구조로, 입력 레이어(input layer), 은닉 레이어(hidden layer), 출력 레이어(output layer)로 이루어진 모델임. 이 모델은 다양한 기계 학습 작업에 사용될 수 있음
 - Transformer Encoder+2 Layer Neural Network 모델은 Transformer의 Encoder 부분만을 사용하고, 이를 통해 추출된 정보를 2-layer 신경망으로 처리하는 구조임
 - Transformer Encoder는 각 단어의 위치와 주변 단어들과의 관계를 고려하여 텍스트의 의미를 더욱 잘 이해할 수 있음. 그 후, 이렇게 추출된 정보는 2-layer 신경망을 통해 분류 및 예측 과정에 활용됨

- 앙상블 모형(Ensemble Model)은 여러 개의 기본 모형(Base Models)을 결합하여 하나의 예측 모델을 만드는 방법으로 개별 모형의 강점을 활용하고 약점을 보완함으로써 전반적인 예측 성능을 향상시키는 데 목적이 있음
 - 위에서 사용한 4개의 모형에서 나오는 1(사무)의 확률값의 평균을 가지는 새로운 결합 모형을 생성함

[표 3-5] 데이터를 8:2로 분할하여 학습한 데이터셋(버전1) 기준, 사용변수에 따른 모형

변수	학습 시 사용한 변수							모형 번호
	변수			새로 생성한 변수				
모형	법령 구분	조문 제목	조문	score_length	score_count	score_ratiomean	score_ratiomulti	
RF	0			0	0	0	0	(1)
LightGBM	0			0	0	0	0	(2)
2-layer-nn		0	0					(3)
transformer encoder + 2-layer-nn		0	0					(4)
ensemble	0	0	0	0	0	0	0	(5)

2) 사무유형 모델의 개발

□ 파생변수의 생성

- 사무유형은 해당 사무의 수행주체에 따라 결정되기 때문에 사무유형의 구분을 위해 수행주체 관련 파생변수를 생성할 필요가 있음
- 수행주체 관련 파생변수를 생성하기에 앞서 수행주체 및 수행주체의 사무유형 별 분포를 파악하기 위한 디서너리를 작성함
 - subject_list
 - 조문 내에서 수행주체를 추출하기 위해 조문 내에서 수행주체가 가능하다고 생각되는 주체들을 모아 조문의 내용을 기반으로 1,562 종류의 수행주체를 찾아 저장함

- subject_dictionary
 - 사무유형의 분류를 위해서 각 수행주체의 사무유형별 분포를 파악하기 위해 만들었음
 - 해당 수행주체가 조문에 있는 비사무인 항의 개수, 국가인 항의 개수, 지방인 항의 개수, 공동인 항의 개수를 세서 다음과 같이 만들었음
 - 예를 들면, 광주광역시라는 주체를 해당 데이터 셋에서 가지는 경우가, 비사무인 경우는 12개의 행이, 국가사무인 경우는 3개의 행이, 지방사무의 경우 5개의 행이 공동사무인 경우 1개로 파악이 된다면 다음과 같이 1,562종류의 모든 수행주체에 대해서 딕셔너리를 작성함

[그림 3-11] 수행주체에 대한 딕셔너리 작성 예시

예시 :
 { '광주광역시': [12, 3, 5, 1] }

- 위의 과정을 거쳐 생성된 파생변수 5개는 다음과 같음
 - subject_nan : 수행주체들이 나온 사무유형이 비사무인 항의 개수
 - subject_n : 수행주체들이 나온 사무유형이 국가인 항의 개수
 - subject_r : 수행주체들이 나온 사무유형이 지방인 항의 개수
 - subject_p : 수행주체들이 나온 사무유형이 공동인 항의 개수
 - score_subject_len : 조문에서 뽑아낸 총 수행주체의 개수
- 생성된 파생변수를 통하여 머신러닝 모델학습에 이용함

[그림 3-12] 머신러닝 모델학습 적용

예 1 :

③ 금융위원회는 제 1 항에 따른 인가를 할 때 다음 각 호의 기준에 적합한지를 심사하여야 한다.

-> subject_nan : 408

-> subject_n : 324

-> subject_r : 1

-> subject_p : 0

예 2 :

금융위원회는 적기시정조치에 관한 권한을 대통령령으로 정하는 바에 따라 금융감독원 원장(이하 ""금융감독원장""이라 한다)에게 위탁할 수 있다.

-> subject_nan : 408 + 75

-> subject_n : 324 + 37

-> subject_r : 1 + 0

-> subject_p : 0 + 1

subject dictionary

{금융위원회: [408, 324, 1, 0]}

 금융감독원: [75, 37, 0, 1]}

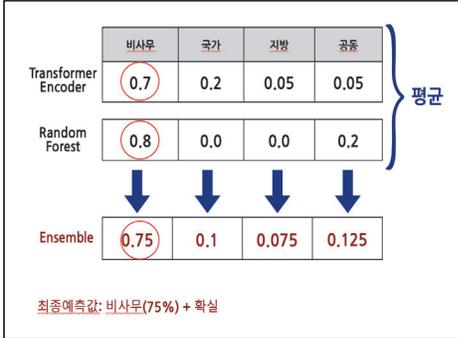
...

}

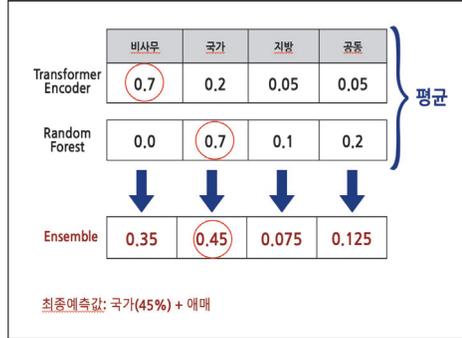
□ 사무유형 모델: 머신러닝, 딥러닝 모델 앙상블

- 사무유형 모델은 랜덤포레스트와 Transformer Encoder를 적용한 인공지능망 모델을 결합하여 하나의 새로운 모델(앙상블 방식)을 만들었음
 - 앙상블 방식은 각 모델별 클래스 예측값을 조합하여, 두 모델이 모두 같은 클래스로 예측하면 해당 예측값과, '확실'이라는 지표를 함께 나타냄
 - 두 모델이 다른 클래스로 예측하면 해당 예측값과, '애매'라는 지표를 함께 나타냄

[그림 3-13] 앙상블 모델 구조(확실한 결과)



[그림 3-14] 앙상블 모델 구조(애매한 결과)



□ 모델의 성능 비교

- 4가지 분류 항목(비사무, 국가사무, 지방사무, 공동사무)을 잘 맞추는 것이 초점을 둬

$$Accuracy = \frac{\text{올바르게 분류된 데이터 수}}{\text{전체 데이터 수}}$$

[그림 3-15] 사무유형 모델(AI Model 2) 방법론별 적용 결과

*train : test = 8 : 2 split, test set에 대한 결과

모델명	Accuracy	모델명	Accuracy
Random Forest	88.2%	2-layer 신경망 with Transformer Encoder	63.3%
2-layer 신경망	59.6%	Ensemble	88.7%

- 다만, 본 연구의 당초 목표로서 법령상 사무를 다양한 유형(중분류, 소분류)으로 분류한 결과, 대분류까지의 분류에 비하여 정확도가 현저하게 낮아지는 바 본 연구에서는 대분류까지만 구분하기로 함

- 대분류에 비하여 중분류·소분류 시 정확도가 현저하게 떨어지는 이유로는 중·소분류를 위한 위임사무 등의 판단에서 결측치가 존재하여 이에 대한 전처리가 필요하였으나, 짧은 연구기간 내에 이를 인공지능에게 학습시키기에는 시간적 한계가 존재함
- 예컨대, 위임사무의 판단은 법령에서 1차적으로 판단하는 것이 아닌, 법률에서 사무로 판단된 사항에 대하여 법률의 타 조문, 대통령령 혹은 시행규칙에서 별도로 '권한의 위임' 조문을 통해 2차적으로 위임사무 여부를 판단해야 하기 때문에 이에 대한 인공지능에의 학습은 더 복잡한 과정이 소요됨
- 따라서 추후 이러한 복잡한 과정의 알고리즘에 대한 연구가 보완되어야 할 것으로 판단됨



제4장

법령사무조사 자동화 모델의 적용

제1절 자동화 모델의 예측 정확성 검증

제2절 자동화 모델을 적용한 데이터 오류 검증

제3절 2022년 법령 데이터에 대한 자동화
모델의 적용 및 활용

제4절 자동화 모델의 타당성 검증: 2022년
5월 이후 법령 데이터의 적용

제1절 자동화 모델의 예측 정확성 검증

1. 사무판단 및 사무유형 모델의 평가도구: Confusion Matrix

- Confusion Matrix는 분류 모델을 평가하는 데 사용되는 주요 도구로, 실제 결과치와 예측된 결과치의 매칭 결과를 시각적으로 나타내는 도구로서 실제 클래스와 예측된 클래스의 매칭을 시각적으로 나타냄
 - 이진 분류 문제에서 실제 클래스를 Positive와 Negative로 구분하며, 분류 모델은 이를 바탕으로 샘플들을 분류함
 - Confusion Matrix는 4가지 상황을 구분함. 여기서 Positive는 ‘사무인’ 경우를 의미함

[그림 4-1] Confusion Matrix

		Actual	
		Positive	Negative
Predicted	Positive	TP	FP
	Negative	FN	TN

- 분류 모델의 성능 평가에는 정밀도(Precision), 재현율(Recall), F1-score, 정확도(Accuracy), ROC-AUC 등의 지표가 활용됨

2. 사무판단 모델(AI Model 1)의 검증

- 본 연구에서 모델 평가 시 고려되는 여러 지표 중에서, 특히 재현율(Recall)이 중요하다.
 - 재현율(Recall)은 실제로 Positive인 샘플 중 모델이 True Positive로 정확히 예측한 비율을 나타냄
 - 재현율(Recall) = $TP / (TP + FN)$
- 사무판단 모델(AI Model 1)에서는 실제로 사무 처리가 필요한 조문을 놓치지 않고 정확하게 식별하는 것이 중요하다
 - 실제 ‘사무’인 조문을 ‘사무가 아님’으로 잘못 분류하는 경우, 즉 거짓 음성(False Negative)이 발생하는 경우, 해당 조문은 사무판단과 사무유형판단 과정에서 배제될 위험이 있기 때문임
 - 이러한 상황은 과업 감소 목적에 부합하지 않으며, 사무에 대해서 처리 기회를 놓치는 경우가 발생함
 - 이에 따라, 본 연구에서는 재현율이 높은 모델을 선택하고, 이를 통해 실제 ‘사무’인 조문을 정확하게 식별하는 것을 목표로 함
- 본 연구에서 사용한 2layer-인공신경망(transformer 인코더 적용) 모델은 타 모델에 비하여 가장 높은 재현율인 94.9%의 재현율을 나타냄

[표 4-1] 분류 방법론별 적용 결과

모델명	Precision	Recall	F1-score	accuracy	threshold	FN개수
RandomForest	0.178	0.939	0.3	0.718	0.006	661
LightGBM	0.195	0.947	0.323	0.745	0.026	574
2-layer nn	0.135	0.943	0.237	0.608	0.231	618
transformer encoder	0.153	0.949	0.263	0.658	0.251	554
양상블	0.178	0.947	0.3	0.715	0.211	575

3. 사무유형 모델(AI Model 2)의 검증

- 본 모델 평가 시 고려되는 여러 지표 중에서, 특히 정확도(Accuracy)를 가장 중요시함
 - 정확도(Accuracy)는 모든 샘플 중 모델이 정확히 예측한 비율을 나타냄
 - $\text{정확도(Accuracy)} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$
- 사무유형 분류 모델은 사무가 아닌지, 국가사무인지, 지방사무인지, 공동사무인지 맞히는 모델로 유형의 분류를 잘 맞추는가에 초점을 둠
 - 이는 앞에서 진행한 사무판단 모델이 더 완전해진 상태의 데이터로 본 모델을 훈련시킨다면 사무가 아닌 데이터는 대부분 없게 될 것이므로 국가, 지방, 공동을 서로 혼동하지 않는 것이 중요하다고 생각됨
 - 따라서 이 모델의 결과가 사무인데 비사무라고 여기는 행들이 많더라도 의미 있다고 할 수 있음
- 사무유형 모델의 경우, 본 연구에서 사용한 앙상블 모델은 타 모델에 비하여 가장 높은 정확도를 나타내며, 81%의 정확도를 나타냄

[그림 4-2] 사무유형 분류 모델 결과

1. RandomForest					2. transformer encoder + 2-layer nn					3. 앙상블							
예측 \ 실제	0	1	2	3	Accuracy	예측 \ 실제	0	1	2	3	Accuracy	예측 \ 실제	0	1	2	3	Accuracy
0	9,792	900	209	130	0.79	0	7758	2315	814	144	0.63	0	9996	772	161	102	0.81
1	942	477	5	17		1	794	467	150	30		1	980	442	4	15	
2	276	4	77	3		2	153	114	86	7		2	277	3	76	4	
3	194	11	0	65		3	152	80	33	5		3	207	10	0	53	

- Accuracy가 가장 높은 앙상블 Model을 최종적으로 채택함

제2절

자동화 모델을 적용한 데이터 오류 검증

1. 개요

- 2022년 법령상 사무총조사에서 사용된 법령은 총 4,357개(법 1,451개, 시행령 1,683개, 시행규칙 1,123개)로 사무판단이 필요한 행은 총 848,072줄로 이루어짐
 - 이에 대하여 훈련받은 연구원이 세부적 사무 판단을 위해 법 규정 및 그 하위법령, 연계 법령을 종합적으로 검토하여 법규정상의 단위사무를 추출함
 - 훈련받은 연구원은 공통의 기준을 토대로 사무를 추출하였으나, 여러 명의 연구원에 의하여 이루어진 주관적 판단의 차이, 입력 오류 등의 오류가 발생할 수 있음
- 따라서 조사 과정에서 발생하는 각종 오류에 대한 면밀한 검증이 필요하지만, 방대한 양으로 인하여 조사뿐만 아니라 검증에 대한 투입 비용 역시 상당한 수준으로 요구되는 실정임
- 본 연구를 통해 개발한 법령사무조사 자동화 모델은 2022년 법령상 사무총조사의 데이터를 학습시킨 인공지능 모델로서 이를 통하여 기존의 2022년 법령상 사무총조사의 데이터 오류를 검증할 수 있음
 - 인공지능을 통한 자동화 모델로 하여금 학습된 내용을 바탕으로 2022년 법령상 사무총조사의 데이터 중 오류 사항을 인식하도록 명령한 후, 해당 오류 데이터가 학습 알고리즘의 미흡으로 인한 오류인지 혹은 인간의 입력 오류인지 등을 판단·검증할 수 있음

2. 인공지능을 통해 추출된 데이터 오류 사항

- 인공지능을 통해 추출된 데이터 오류 사항은 다음 <표 4-1>과 같이 8가지 유형이 추출됨

[표 4-2] 인공지능을 통해 추출된 데이터 오류 사항

(단위: 개)

구성요소	내용	개수	검증 결과
사무유형	코드북 ^주 상 사무유형(중분류)과 다른 방식으로 입력	2	• 입력 오류 • 상세 유형으로 재분류 필요
사무유형(소분류)	사무판단이 1인 경우 중, 사무유형(소분류)은 미입력되어 있는 경우	6	• 입력 오류 • 상세 유형으로 재분류 필요
중복조문 관련	동일한 조문이 2개 이상인 경우	13,549	• 복수 사무 및 수행주체의 경우 • AI 추가 학습 필요
호번호, 사무판단	호에서 사무가 추출됨	822	• 예외적으로 호에서 사무가 있는 경우 • AI 추가 학습 필요
위임사무판단	위임사무에 대하여 0/1이 아닌 01로 입력된 경우	2	• 입력 오류
위임사무판단, 위임근거규정	위임사무(1로 입력)이지만, 위임근거규정이 없는 경우	68	• 입력 오류 • 추후 근거규정 삽입 필요
위임사무판단, 위탁사무판단	위임사무와 동시에 위탁사무인 경우 (두 경우 모두 1로 입력)	96	• 사람의 개별 판단 필요 • AI 추가 학습 필요
재위임사무판단	재위임사무판단에서 입력방식이 숫자가 아닌 경우	1	• 입력 오류

주: 국가, 시도, 시군구, 국가-시도-시군구, 국가-시도, 국가-시군구, 시도-시군구

□ 코드북에서 공통기준으로 삼은 사무유형과 다른 방식으로 입력된 경우

- 코드북 상 사무유형(중분류)은 국가, 시도, 시군구, 국가-시도-시군구, 국가-시도, 국가-시군구, 시도-시군구 등 7가지 유형으로 구분되어 있음
- 그럼에도 불구하고 2개의 데이터는 '공동', '전국' 등으로 입력되어 있어 데이터 오류로 추출됨

- 이러한 경우 입력 오류로 보이며, 추후 상세 유형으로 재분류가 필요할 것으로 보임

□ 소분류의 사무유형이 공란으로 입력된 경우

- 사무로 판단되어(1로 입력) 이에 대한 대분류, 중분류, 소분류에 대한 판단이 이루어져야 함에도 불구하고 소분류가 입력되어 있지 않은 경우로 6개의 데이터가 추출됨
- 이러한 경우 입력 오류로 보이며, 추후 상세 유형으로 재분류가 필요할 것으로 보임

□ 동일한 조문이 2개 이상 반복되는 경우

- 하나의 조문에 하나의 사무가 추출되는 것이 일반적인 형태이지만, 예외적으로 하나의 조문의 복수의 사무 혹은 복수의 수행주체인 경우가 있음
 - 예컨대, 하나의 조문에 독립된 복수의 사무가 있는 경우 혹은 동일한 사무이지만 수행주체가 명백하게 달라 별도의 사무로 판단할 필요가 있는 경우 2022년 법령상 사무총조사에서는 예외적으로 해당 데이터를 처리하기 위하여 동일한 조문을 중복하여 입력하는 방식을 택함
 - 이에 인공지능은 13,549개의 행을 데이터 오류로 인식하여 추출함
- 본 연구에서 개발한 자동화 모델에서 이러한 경우에 대해서는 학습이 완료되지 않았기 때문에 오류로 추출하였기 때문에 이에 대한 추가적인 보완이 필요함

□ 조문 중 ‘호’에서 사무가 추출된 경우

- 원칙적으로 법령상 사무총조사는 조문 중 ‘조’와 ‘항’까지만 사무를 추출하였음
 - 그러나 일부의 조문에서는 ‘호’ 단위에서 독립된 사무가 추출되는 경우가 있기 때문에 예외적으로 사무로 추출하였음
- ‘호’ 단위에서의 사무판단 여부는 인간의 주관적 사고 체계에서 판단이 이루어지는 영역이므로 인공지능에게 학습시킬 수 있는 알고리즘 개발에 보완이 필요함

- 다만 본 연구에서는 이로 인한 오류를 최소화하기 위하여 사무판단 모델(AI Model 1)에서 rule-based 변수를 통하여 ‘호’ 단위에서 절대 사무가 되지 않는 경우의 수를 추출하여 프로그래밍함
 - 예컨대, 호에 ‘서류’, ‘대표자’, ‘인정되는 경우’, ‘있는 경우’ 등이 포함되어 있는 경우 사무가 아닌 대표적인 기준으로 생성한 바 있음

[그림 4-3] ‘호’ 단위에서 비사무로 추출되는 경우(rule-based 사용 규칙)

호가 다음의 규칙을 가지는 경우 (31,631행)	
<p>○ 호에 ‘서류’가 포함 (8,084행)</p> <p>3. 제63조제1항에 따른 관계 물품·서류 등의 제출</p> <p>2. 개인정보처리자가 조정을 거부하였거나 조정결과를 수락하지 아니하였음을 증명하는 서류</p>	<p>○ 호에 ‘대표자’가 포함 (1,888행)</p> <p>1. 평가기관의 명칭·주소 및 전화번호와 대표자의 성명</p> <p>7. 정관 및 대표자와 임원의 명단(법인인 경우에 한한다)</p>
<p>○ 호에 ‘인정되는 경우’가 포함 (1,614행)</p> <p>5. 정보주체 또는 그 법정대리인이 의사표시를 할 수 없는 상태에 있거나 주소불명 등으로 사전 동의 를 받을 수 없는 경우로서 명백히 정보주체 또는 제3자의 급박한 생명, 신체, 재산의 이익을 위하여 필요하다고 인정되는 경우</p> <p>3. 정보주체 또는 그 법정대리인이 의사표시를 할 수 없는 상태에 있거나 주소불명 등으로 사전 동의 를 받을 수 없는 경우로서 명백히 정보주체 또는 제3자의 급박한 생명, 신체, 재산의 이익을 위하여 필요하다고 인정되는 경우</p>	<p>○ 호가 ‘있는 경우’로 종결 (2,578행)</p> <p>1. 위원 또는 그 배우자나 배우자였던 자가 해당 사안의 당사자가 되거나 그 사건에 관하여 공동의 권리자 또는 의무자의 관계에 있는 경우</p> <p>5. 위원이나 위원이 속한 공공기관·법인 또는 단체 등이 조연 등 지원을 하고 있는 자와 이 해관계가 있는 경우</p>

□ 위임사무의 판단이 다른 방식으로 입력된 경우

- 원칙적으로 위임사무의 경우 위임일 경우 1로 입력되어야 하지만, 2개의 데이터 에서는 01로 입력되어 있었음
- 이러한 경우 입력 오류로 보이며, 개별적으로 재입력함

□ 위임사무로 판단되었으나 이에 대한 위임근거규정이 없는 경우

- 원칙적으로 위임사무의 판단은 법률 등에서 사무로 판단된 사무 중에서(1차 입력) 법령에서 별도의 조문 혹은 대통령령·시행규칙에서 ‘권한의 위임’으로 별도로 명시된 경우, 다시 해당 사무로 돌아가 위임사무로 입력(2차 입력)하는 방식으로 판단됨

- 따라서 추후 검증 과정의 용이성을 위하여 2022년 법령상 사무총조사에서는 데이터 입력 시 위임의 근거규정을 입력하도록 하였으나, 근거규정이 누락된 2개의 위임사무를 인공지능이 추출함
- 본 오류는 입력의 오류로서 추후 근거규정이 삽입이 필요할 것임

□ 위임사무인 동시에 위탁사무인 경우

- 위임사무로 판단됨과 동시에 위탁사무로 판단된 경우로 두 사무 판단에서 모두 1로 입력된 경우임
 - 앞서 언급한 바와 같이 위임·위탁사무의 경우 다른 사무와 달리 법령의 전체 체계에서 2단계의 판단과 입력 과정을 거쳐야 함
 - 그러나 본 연구에서는 시간적 한계로 이러한 과정에 대한 판단 과정은 수행하지 못하였으며, 이에 대한 추가 연구가 필요할 것으로 사료됨
- 또한 본 연구에서 인공지능이 추출한 96개의 해당 사항에 대해서는 사람인 연구자가 개별적인 판단이 필요한 영역으로 사료됨

□ 재위임에 대한 사무판단에서 입력 방식이 숫자가 아닌 경우

- 재위임사무 역시 입력은 숫자로 입력하도록 되어 있으나 1개의 데이터는 숫자가 아닌 문자로 입력이 된 경우임
- 이러한 경우 입력 오류로 사료됨

3. 인공지능을 통해 추출된 데이터 오류의 검증

- 이상의 논의를 종합하여 볼 때, 법령사무조사 자동화 모델에서 인공지능이 2022년 법령사무총조사의 데이터 중 오류라고 추출한 항목에 대한 검증은 다음과 같음

- 전체 사무판단이 필요한 844,220개의 행 중에서 인공지능이 오류가 발생했다고 추출한 데이터는 총 14,546개로 전체의 1.723%에 해당함
 - 이 중 연구진의 입력 오류에 의한 데이터 오류는 79개(0.009%)이며, AI의 학습 미흡에 의한 데이터 오류는 14,467개(1.714%)의 비중으로 나타남
 - 다만, 본 연구에서는 사무수행 주체가 상이한 특별지방행정기관의 경우는 삭제(3,852개)하고 분석하였기 때문에 844,220개의 행에 대해서만 프로그래밍을 진행함
- 인공지능이 오류로 인식한 데이터 중 99%는 인공지능에 대한 학습 미흡에 의한 것이기 때문에 이에 대해서는 후속 연구가 필요할 것임

[표 4-3] 인공지능이 추출한 데이터 오류의 검증

구 분		검증 결과		
		개수(개)	비율(%)	
			전체 중	오류 중
전체		844,220 ^주	100	-
오류 데이터로 인식	(인간의) 입력 오류	79	0.009	0.543
	AI 학습 미흡	14,467	1.714	99.457
	계	14,546	1.723	100

주: 전체 사무 848,072개의 행 중 특별지방행정기관에 대한 사무(3,852개)는 제외함

제3절 2022년 법령 데이터에 대한 자동화 모델의 적용 및 활용

1. 자동화 모델을 활용한 사무판단 및 사무유형 추출 결과

- 제3장의 사무판단 및 사무유형 모델의 적용 결과는 다음 <그림 4-4>와 같음
 - ‘조문’ 바로 뒤에 명시된 ‘사무판단’ 및 ‘사무유형(대분류)’은 2022년 법령상 사무총조사의 실제 판단 결과임
 - ‘사무판단 예측’은 자동화 모델을 통하여 인공지능이 해당 조문이 사무인지 여부를 판단한 결과임
 - ‘사무판단 예측 확률’은 인공지능이 사무로 판단할 가능성으로서, 인공지능이 해당 조문을 사무로 판단할 예측 확률임
 - ‘수행주체’는 조문 상에 나와 있는 사무의 수행주체임
 - ‘사무유형(대분류) 결과’는 인공지능이 해당 조문의 사무유형을 분류한 결과 값임(0=비사무, 1=국가사무, 2=지방사무, 3=공동사무)
 - ‘비사무~공동 확률’은 해당 조문의 사무가 각각의 사무유형으로 분류될 가능성을 수치화 한 값임

[그림 4-4] 대시보드 아키텍처



- <그림 4-5>의 예시로 설명하자면, 본 그림에서 제시된 조문 중 실제 모두 사무가 아닌 것으로 판단된 항목이며, 대체로 실제 판단과 인공지능의 판단 결과가 일치하는 것으로 나타남
- 다만, 인공지능을 통한 예측 결과 「수도법 시행규칙」의 ‘규제의 재검토’를 제외한 항목은 모두 일치하는 것으로 나타남
 - ‘규제의 재검토’의 경우 조문으로만 판단하면 사무로 추출될 것으로 보이지만, 실제 법령상 사무총조사시 해당 조문은 일괄적으로 사무판단의 예외 사항으로 정함 (예외적 사항)
 - 이에 인공지능은 조문의 형태로 보았을 때에는 61.4%의 확률로 사무로 추출함
 - 하지만 타 법률에 대한 학습 결과를 반영하였을 경우, 약 79.4%로 비사무로 추출될 확률임을 명시하여 이에 대한 연구자의 추가적인 판단이 필요하다는 것을 제시함
- 자세한 결과값 비교를 위하여 「우주손해배상법」을 대상으로 주 개의 결과값 (2022년 법령상사무총조사, 자동화 모델)을 비교한 결과는 [부록 1]에 첨부하였음
 - 「우주손해배상법」을 대상으로 결과값을 비교한 결과, 총 20개의 조문 중 6개의 조문에 대해서는 사무판단의 오류가 발생함
 - 실제로는 사무가 아님에도 불구하고 자동화 모델은 사무라고 인식함
 - 그러나 그 외의 조문에 대해서는 사무판단 및 사무유형에 대해서 「2022년 법령상사무총조사」와 동일하게 판단하였음

2. 법령상 사무에 대한 자동화 모델의 활용: 대시보드의 제시

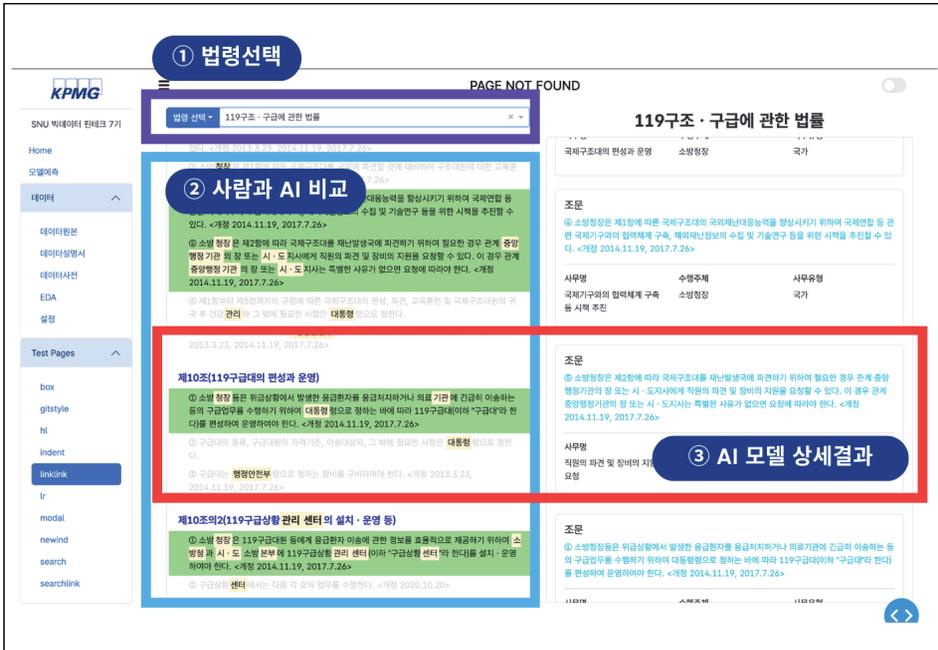
□ 개요

- 인공지능을 활용한 법령사무조사 자동화 모델은 복잡한 명령어를 기반으로 한 모델 로직이기 때문에 이를 일반인이 실생활·실무에 적용하는 것은 사실상 불가능함
- 이하에서는 자동화 모델의 결과를 일반인도 손쉽게 활용할 수 있는 방안으로서 이를 활용한 대시보드를 예시로 제시하고자 함
 - 사람이 직접 조사한 결과와 AI가 판단한 내용을 비교하는 형식으로 나타내주어 AI가 어떤 모델을 잘 분류하는지 분류를 잘 하지 못하는지 바로 파악 가능한 대시보드를 구상함
 - 추가로 AI가 다음과 같이 판단한 근거를 제시함으로써 연구자의 추가적인 판단이 필요한 영역을 확인할 수 있도록 함

□ 대시보드의 구성 개괄

- 사용자가 원하는 법령을 선택하면, 해당 법령의 모든 조문에 대하여 사람 (연구자)의 주관적 판단에 의한 결과치와 AI에 의한 결과치가 색깔로 제시됨
- 동시에 왼쪽의 상세결과로서 AI 모델이 해당 조문에 대한 사무판단 여부 및 예측 확률, 사무명, 수행주체, 사무유형 등의 세부 정보를 제시함

[그림 4-6] 대시보드의 전체 구성



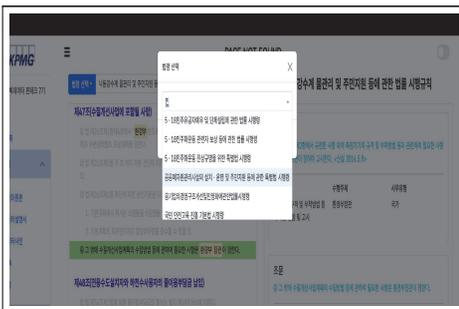
① 법령의 선택

- 법령을 선택하여 원하는 법령만 볼 수 있는 기능임
 - 왼쪽의 '법령선택' 박스를 클릭하면 다음과 같이 법령, 시행령, 시행규칙별로 선택하여 볼 수 있음
 - 또한, 직접 원하는 법령명을 검색할 수도 있음

[그림 4-7] 법령선택 박스



[그림 4-8] 원하는 법령명의 검색 박스



② 사람(연구자) 및 인공지능의 결과치 비교

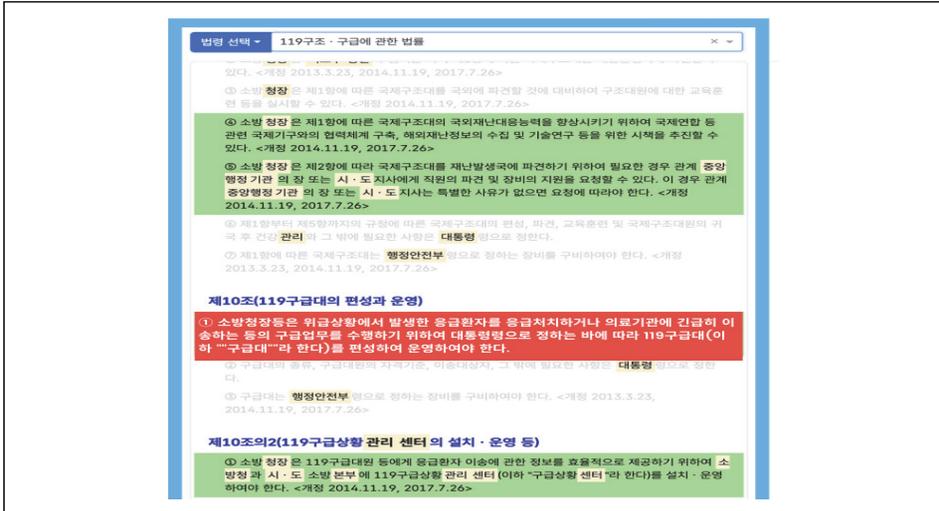
- AI로 예측한 조문과 실제 조문값을 비교하여, 알아보기 쉽게 색깔로 표시하였음
 - 초록색 : ai와 사람이 모두 사무라고 판단한 조문
 - 빨간색 : ai가 비사무라고 예측했으나, 실제로는 사무인 조문
 - 주황색 : ai가 사무라고 예측했으나, 실제로는 비사무인 조문
 - 회색 : ai와 사람이 모두 비사무라고 판단한 조문

[그림 4-9] 사람과 AI의 사무판단 비교

		사무판단 예측값	
		NO=0	YES=1
사무판단 실제값	NO=0	사람판단: 비사무 AI 판단: 비사무	사람판단: 비사무 AI 판단: 사무
	YES=1	사람판단: 사무 AI 판단: 비사무	사람판단: 사무 AI 판단: 사무

- 대시보드에서 제공되는 법령의 각 조문에 하이라이팅하여 시각화함으로써 사용자가 하여금 결과치를 보다 직관적으로 비교할 수 있도록 함

[그림 4-10] 수행주체 후보들 시각화



③ 인공지능의 결과치에 대한 상세 결과

- AI가 다음과 같이 판단한 근거를 제공하는 기능으로 구성 내용은 다음과 같음
 - 사무유형(대분류) 예측, 각 유형별 확률, AI판단의 신뢰성 정도, 사무명, 수행주체, 실제 사람이 분류한 정답으로 구성된 표를 제공하게 됨
- 사무유형(대분류) 예측, 각 유형별 확률, AI 판단의 신뢰성 정도는 Rule-Based, 사무판단모델, 사무유형(대분류) 분류 모델을 거쳐서 나온 결과와 확률값들을 표 안에 표현해 줌
 - AI 판단의 신뢰성은 ‘확실’, ‘애매’로 나타냄
 - 사무명은 현대 대시보드 상에서는 이미 라벨링된 데이터를 표시해주었지만 openai의 chatgpt api를 사용하여 조문제목과 조문을 보고 자동으로 사무명을 생성해낼 수 있도록 구현할 수 있음

- 수행주체는 조문 내에 수행주체가 될 가능성이 있는 전부 표현해 줌

[그림 4-11] 시모델의 상세결과

사무 유형	활동				AI 판단	사무명	수행주체	실제 정답
	비사무	국가	지방	공동				
국가					확실	국제기구와의협력체계 구축	소방청장 시도 중앙행정기관	국가
	0.2	0.6	0.1	0.1				

[그림 4-12] 실제 대시보드 적용 화면

The screenshot shows a dashboard with the following elements:

- Navigation Menu:** Home, Prediction.
- Search Bar:** 119구조 · 구급에 관한 법률
- News Article 1:**
 - Topic:** 소방청장은 기본계획에 따라 매년 연도별 구조·구급 집행계획(이하 '집행계획')의 편성을 수립·시행하여야 한다. <개정 2014.11.19, 2017.7.26>
 - AI Analysis:**

AI 사무확률	AI 유형 예측	예측 확률	AI 수준	실제유형	사무명
PASS	사무 아닌	비사무: 38.21%	확실	국가 사무	119구급상관 관련센터의 설치·운영 등
- News Article 2:**
 - Topic:** 소방청장은 기본계획 및 집행계획을 수립하기 위하여 필요한 경우에는 관계 중앙행정기관·지방 또는 시·도 지사에 관한 자료의 제출을 요청할 수 있다. 이 경우 자료제출을 요청받은 관계 중앙행정기관·지방 또는 시·도 지사는 특별한 사유가 없으면 이에 따라야 한다. <개정 2014.11.19, 2017.7.26>
 - Section:** 제7조(시·도 구조·구급집행계획의 수립·시행)
 - AI Analysis:**

AI 사무확률	AI 유형 예측	예측 확률	AI 수준	실제유형	사무명
PASS	공동 사무	공동: 44.22%	확실	사무 아닌	
- News Article 3:**
 - Topic:** 소방청장은 소방본부장 또는 소방서장 이하 '소방장관'이라 함은 위급상황에서 응급구조의 명령 등을 신속하고 안전하게 구조하는 업무를 수행하기 위하여 대통령령으로 정하는 바에 따라 119구조대의 "구조대"라 한다.
 - Section:** 제8조(119구조대의 편성과 운영)
 - AI Analysis:**

AI 사무확률	AI 유형 예측	예측 확률	AI 수준	실제유형	사무명
PASS	사무 아닌	비사무: 67.07%	확인필요	공동 사무	119구급차의 운용

제4절

자동화 모델의 타당성 검증: 2022년 5월 이후 법령 데이터의 적용

1. 개요

- 본 연구에서 활용한 법령 데이터는 한국지방행정연구원에서 2022년에 수행한 「2022년 법령상 사무총조사」를 통하여 구축한 데이터로서 2022년 5월 1일 기준으로 시행되었던 법령임
 - 2022년 5월 이후의 데이터는 2023년에 법제연구원에서 「법령상 사무총조사」를 수행함
- 본 절에서는 2022년 법령 데이터를 토대로 학습한 인공지능을 활용한 자동화 모델을 법제연구원의 자료 협조를 받아 2023년 법령 데이터를 적용하여 자동화 모델의 타당성을 검증하고자 함

2. 법령사무조사 자동화 모델의 타당성 검증

- 2022년 5월 1일 이후 법령의 제·개정 등으로 인한 변경 사항은 총 956행이며, 그 중 입력 오류 등의 결측치를 제거하여 최종 898행에 대하여 검증을 실시함
- 타당성 검증 결과, 사무판단 모델(AI Model 1)의 recall은 98.5%로 2022년 법령 데이터를 대상으로 검증한 결과와 큰 차이를 보이지 않음
- 하지만 사무유형 모델(AI Model 2)의 경우 정확도가 63.8%로 2022년의 검증 정확도(81%)에 비하여 큰 폭으로 감소한 것으로 나타남
 - 이는 2022년의 경우 전체 조문 대비 사무 비율이 약 6%로 나타난 것에 반해, 2023년의 경우 사무가 전체 조문의 약 35%를 차지하여, 이는 2022년 5월 이후의 데이터 세팅 자체가 사무로 추출된 조문으로 1차 추출된 결과로 입력이 된 것에 기인한 것으로 보임

- 그럼에도 불구하고 정확도가 큰 폭으로 감소한 것은 본 연구에서 개발한 자동화 모델이 2022년 법령 데이터에 과적합되어 있어 성능이 비교적 낮아진 것으로 보임
 - 다만, 법령 데이터 오류의 축소, 2022년 5월 이후 법령 데이터 전체에 대한 재학습 및 연도별 AI 딥러닝을 지속적으로 수행한다면 이러한 정확성의 큰 격차는 점차 축소시킬 수 있을 것으로 사료됨



제5장

결론 및 정책적 제언

제1절 연구요약 및 정책적 제언

제2절 연구의 한계 및 향후 과제

제1절 연구요약 및 정책적 제언

1. 연구요약

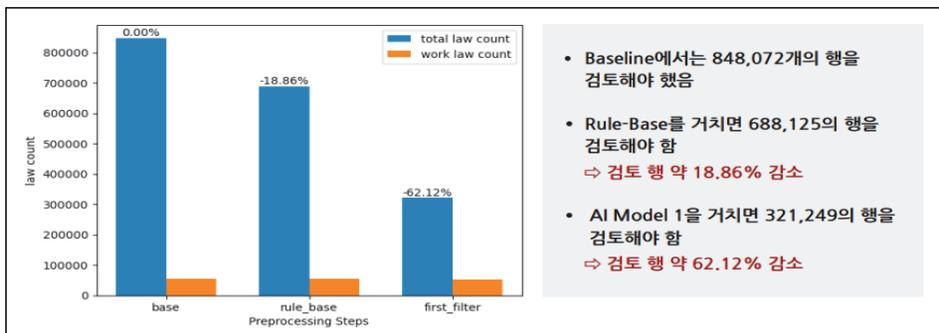
- 효율적인 사무의 수행을 위해 국가-지방사무의 배분 및 사무수행체계 현황을 파악하기 위한 노력의 일환으로 1994년부터 약 5년 단위로 한국지방행정연구원을 중심으로 「법령상 사무총조사」가 진행되어 왔음
 - 그러나 기존의 방식은 연구자의 주관적 판단에 의존하는 방식을 취하고 있어 과도하게 많은 시간과 비용의 투입, 조사 결과의 일관성·신뢰성에 대한 비판 등의 한계에 직면하고 있음
- 데이터 기반 행정의 대두와 윤석열 정부의 디지털플랫폼정부의 국정 기조 하에 본 연구는 생성형 AI를 활용하여 보다 효율적이고 일관성 있는 법령사무조사를 수행하기 위한 자동화 모델 개발을 목적으로 하며, 구체적인 연구내용은 다음과 같음
 - 첫째, 법령사무조사 자동화 모델의 개발
 - 둘째, 2022년 5월 기준 법령 데이터에 대한 법령사무조사 자동화 모델의 적용
 - 자동화 모델의 예측 정확성 검증
 - 자동화 프로그램을 적용한 법령 데이터의 오류 검증
 - 2022년 법령 데이터에 대한 자동화 모델의 적용 및 활용 방안 제시
 - 셋째, 2022년 5월 이후 법령 데이터를 적용한 자동화 모델의 타당성 검증
 - 넷째, 「법령상 사무총조사」에 대한 자동화 모델의 적용 가능성 검토 및 미래 방향 제시

- 법령사무조사의 자동화 모델은 머신러닝, 딥러닝 모델 등의 인공지능을 활용한 사무판단 모델(AI Model 1)과 사무유형 모델(AI Model 2)로 구분하여 제시함
 - 사무판단 모델(AI Model 1)은 Transformer Encoder을 활용한 2-layer 신경망 모델을 통하여 94.9%의 의 재현율을 달성하였음
 - 사무유형 모델(AI Model 2)은 랜덤포레스트(Random Forest: RF) 모델과 Transformer Encoder+2-layer neural network 모델을 결합한 앙상블 모델을 적용하여 81%의 정확도를 달성하였음
- 이와 같은 법령사무조사 자동화 모델을 적용하여 2022년 5월 기준의 법령 데이터의 오류를 검증한 결과, 인공지능은 전체 조문 행(844,220개) 중 14,546개의 오류 데이터를 추출해 냄
 - 인공지능이 오류 데이터로 추출한 데이터(14,546개, 100%) 중 인간의 입력 오류에 의한 오류는 79개로 약 0.5%에 불과한 반면, AI 학습 미흡 등의 이유로 추출된 데이터 오류는 14,467개로 약 99.5%를 차지하는 것으로 나타남
 - 이러한 오류는 본 연구가 자동화 모델 개발의 초기이기 때문에 향후 보다 면밀한 수준의 모델 개발 및 지속적인 딥러닝을 통해 보정할 수 있는 것으로 사료됨
- 한편, 본 자동화 모델을 2022년 5월 이후 제·개정된 법령 데이터에 적용한 결과 특히 사무유형 모델에서 정확도가 63.8%로 감소하여 본 모델의 타당도에 대한 추가적인 보완이 필요한 것으로 나타남
 - 이는 2022년 5월 이후의 법령 데이터의 불완전성 등의 원인도 작용할 것으로 보이지만 본 모델이 2022년에 수행된 「법령상 사무총조사」의 단년도 데이터만을 통하여 개발되었다는 점 등을 통해 2022년에 과적합되어 성능이 낮아진 것으로 보임
 - 따라서 법령 데이터의 오류 축소, 2022년 5월 이후 법령 데이터 전체에 대한 재학습 및 다년도 법령 데이터에 대한 지속적 딥러닝 등을 통하여 모델의 성능을 개선할 수 있을 것으로 보임

2. 기대효과

- 본 연구는 「2022년 법령상 사무총조사」의 단년도 법령 데이터만을 가지고 자동화 모델을 개발하였음에도 불구하고 획기적으로 인간의 노력이 투입되는 비용을 줄일 수 있는 것으로 나타남
- 「2022년 법령상 사무총조사」의 경우 2022년 5월을 기준으로 시행 중인 법령 4,339개의 법령을 대상으로 연구진이 직접 판단하는 방식으로 조사되었음
 - 각 법령의 모든 조문을 엑셀파일에 크롤링한 결과, 전체 엑셀의 행 수는 84만여 개로 각 조문의 조, 항, 호 모두에 대한 판단이 투입되어야 하였기 때문에 방대하게 많은 연구진의 노력이 투입되어야만 하였음
- 그러나 자동화 모델의 적용을 통하여 1차적인 rule-based의 단계를 통하여 약 19%의 법령 데이터는 판단 대상에서 제외할 수 있음
 - rule-based 규칙은 무조건 비사무로 판단될 수 있는 35개의 원칙을 인공지능이 확인하여 이에 해당하는 조문을 1차적으로 제거하는 방식임
 - 이를 통하여 당초 검토해야 할 전체 행 848,072개 중 18.9%의 행이 원천적으로 제거되어 688,125개의 행이 검토대상 행으로 추출될 수 있음
- 또한 사무판단 모델(AI Model 1)을 통하여 인공지능이 확실하게 비사무인 사무만을 제거하면 검토 행의 약 62.1%가 감소한 321,249개의 행만 검토가 필요한 행으로 추출됨

[그림 5-1] 자동화 모델의 성능지표-모델 적용에 따른 검토 행의 감소 규모



3. 정책적 제언

- 이상의 논의를 토대로 정책적 제언을 제시하면 다음과 같음
- 첫째, 디지털플랫폼정부의 구현을 위한 법령상 사무 및 사무배분 등의 현황을 파악하기 위하여 인공지능을 활용한 자동화 모델을 구축·적용할 필요가 있음
 - 그동안의 「법령상 사무총조사」는 방대한 법령 정보로 인하여 5년 주기로 이루어졌으며, 사람(연구진)이 직접 모든 법조문을 일일이 판단하는 방식으로 이루어져 시의성과 객관성·신뢰성 등의 문제를 내포하고 있었음
 - 특히 법령의 제·개정은 빠르게 이루어지고 있음에도 불구하고 현황 파악을 위한 조사는 약 5년을 주기로 이루어지고 있어 빠른 법령의 변화를 제대로 포착하지 못한다는 점에서 정확한 정보로 사무의 실태를 파악하는데 한계가 있었음
 - 그러나 자동화 모델의 구축·적용을 통하여 사람(연구진)의 노력을 획기적으로 줄일 수 있다면 빠르게 변화하는 법·행정·사회적 환경에 보다 신속하게 대응할 수 있을 것임
 - 또한 그동안 사람의 주관적 판단으로 인하여 사무총조사 결과에 대한 신뢰도 문제 역시 일정 부분 해결할 수 있을 것임
 - 그동안 사무총조사에 참여하는 다수의 연구진의 개인적·주관적인 판단에 의존하고 있어 기존의 방식은 사무총조사 결과에 대한 신뢰성과 일관성에 대한 우려가 지속적으로 제기되어 왔음
 - 사무총조사의 결과에 대한 재검증에 대한 필요성은 매번 강조되었음에도 불구하고 방대한 양의 법령 데이터로 인하여 전체 조사 결과에 대한 재검증은 현실적으로 쉽지 않은 상황이었음
 - 이에 1차적으로 연구진이 판단한 결과를 자동화 모델을 통하여 재검증함으로써 결과의 신뢰도를 높일 수 있을 것으로 기대됨

- 다만, 사무에 관한 복잡한 판단을 포착할 수 있는 자동화 모델의 로직이 개발되기 전까지는 1차적으로 연구진이 판단한 결과를 토대로 자동화 모델이 2차 검증하는 역할에서 모델의 성숙에 따라 점차 역할을 확대해 나가는 방안을 고려할 필요가 있음
- 둘째, 정확한 법령상 사무의 실태 파악을 위하여 체계적이고 정확한 법령 데이터의 관리·보완 체계의 마련이 필요함
 - 본 자동화 모델을 통하여 소수이기는 하지만 데이터의 오류 및 기존 사무총조사 결과의 오류를 추출하였음
 - 이러한 잘못된 데이터와 결과로 인하여 결국 법령상 사무의 실태를 정확하게 포착이 어려울 수 있기 때문에 이에 대한 체계적이고 지속적인 관리와 보완이 필요할 것임
 - 따라서 구축된 데이터에 대한 정확성 검증 등의 체계적인 관리 체계 마련이 필요할 것임
- 셋째, 자동화 모델에서 기능 중심의 키워드 추출 방식의 도입을 통하여 타 분야에의 적용 확대 가능성 모색이 필요함
 - 본 연구의 자동화 모델은 1차적으로 사무인지 여부 및 대분류 차원에서의 사무유형의 구분까지 가능한 모델임
 - 이와 더불어 사무는 법령을 근거로 이루어지고 있기 때문에 ‘규제’, 승인이 필요한 사무의 추출 등 ‘행정절차’ 등의 키워드를 중심으로 사무를 추출할 수 있는 기능을 보완한다면 이와 관련한 타 분야에까지 확대하여 적용·활용될 수 있을 것임
 - 또한 법령뿐만 아니라 조례까지 적용 범위를 확대한다면 국가사무와 지방사무 등의 실태뿐만 아니라 자치사무 등의 실태를 파악하는 데에도 광범위하게 활용될 수 있을 것임
- 넷째, 법령상 사무총조사를 위한 전문인력의 육성이 필요함
 - 프로그램 개발자의 법령에 대한 이해 부족은 법령상 사무를 AI가 정확하게 포착하는 로직 개발에 가장 큰 문제점으로 지적됨

- 광범위한 법령 데이터를 구축·관리하기 위해서는 컴퓨터 프로그래밍에 대한 이해뿐만 아니라 법령의 체계 등에 대한 이해가 뒷받침되어야 보다 정확한 판단이 이루어질 수 있음
 - 실제로 한국지방행정연구원은 「2022년 법령상 사무총조사」 수행 당시 본격적인 조사에 착수하기 전에 참여 연구진에 대한 사전 교육을 한 달 이상 진행하여 법령의 체계, 내용 등에 대한 이해도를 높이는 과정을 거쳤음
- 또한 제정과 개정을 반복하는 법령의 특성상 이를 관리하기 위해서는 법령 데이터를 전담하여 관리할 수 있는 전문인력이 필요할 것임
- 따라서 법령 데이터에 대한 전담인력과 이에 대한 교육 프로그램 등의 마련 방안을 모색할 필요가 있음

제2절 연구의 한계 및 향후 과제

- 본 연구는 효율적인 법령상 사무의 조사를 위하여 인공지능을 활용한 자동화 모델을 개발하였고, 모델의 정확성과 타당성, 연구진의 투입 노력의 감소 등 구체적인 성과 도출을 통하여 법령사무조사 자동화 모델의 활용 가능성을 검증하였음
- 그러나 본 연구에서 제시한 모델은 매우 초기적인 단계의 모델인바, 다음과 같이 연구의 한계 및 향후 과제를 제시할 수 있음
- 첫째, 사무유형의 구분을 대분류 수준까지만 제시함
 - 「법령상 사무총조사」는 사무의 유형 구분의 수준으로서 3가지의 대분류, 7가지의 중분류 및 16가지의 소분류로 세분화하여 제시함
 - 대분류는 국가사무, 지방사무, 공동사무까지만 제시하고 있기 때문에 본 자동화 모델을 통한 위임사무나 특별지방행정사무 등에 대한 세부적인 판단은 불가능함
 - 그러나 사무배분의 실태를 면밀하게 파악하기 위해서는 보다 세분화된 사무 구분이 필요하기 때문에 이에 대한 향후 보완이 필요할 것임
- 둘째, 법령 데이터의 한계가 있음
 - 2022년 5월 기준의 법령 데이터는 여러 가지 입력 오류로 결측치를 제거한 후 모델링을 하였기 때문에 이에 대한 후속 보완작업이 필요할 것임
 - 또한 수행주체 판단의 특이성으로 인하여 특별지방행정 사무는 모델링 과정에서 원천적으로 제거하였기 때문에 향후 연구에서는 이에 대한 보완이 필요할 것임
 - 한편, 2022년 5월 이후의 법령 데이터는 「2023년 법령상 사무총조사」를 수행한 법제연구원의 협조를 받아 제공받았으나, 2022년 5월 법령 데이터의 경우 전체 조문 중 사무 비율이 약 6%인데 반해 해당 데이터는 약 35%의 사무 비율이 추출된 것으로 봤을 때, 사무에 해당하는 법령 데이터를 2차적으로 추출하였을 가능성이 있으므로 이에 대한 보완이 필요할 것임

- 셋째, 짧은 연구기간과 모델 개발자의 법령에 대한 이해도 부족의 한계가 있음
 - 본 연구는 2023년 11월 27일부터 12월 31일까지 약 한 달 동안 진행된 연구이기 때문에 타 연구에 비해서도 상당히 짧은 기간 이루어진 연구임
 - 또한 자동화 모델을 구축하기 위하여 프로그래밍을 전문으로 하는 기관과의 협업을 통하여 진행되었으나 짧은 연구 기간으로 인하여 모델 개발자의 법령 이해도가 낮은 상태에서 연구가 진행되었고, 이에 따라 착수 초기 법령에 대한 이해를 위하여 상당 시간을 소요하였음
 - 따라서 향후 자동화 모델의 고도화를 위해서는 모델 개발자에 대한 사전적인 법령의 이해 증진 및 장기간의 연구를 통하여 보다 엄밀한 자동화 모델을 구축할 필요가 있음

참고문헌

- 고광용. (2016). 중앙·지방정부간 사무이양 체계 및 성과에 관한 연구. 입법과 정책, 8(1), 57-81.
- 권경득·우문정. (2009). 참여정부 지방분권정책의 실태분석-중앙사무의 지방이양을 중심으로-. 한국지방자치학회보, 21(2), 5-28.
- 박혜자. (2002). 지방이양과 중앙-지방정부간 사무배분체계의 변화. 현대사회와 행정연구, 13(3), 125-146.
- 이혜영. (2014). 국가 사무의 지방 이양의 한계에 관한 연구: 지방분권촉진 실무위원회의 심의 내용을 중심으로. 현대사회와 행정, 24(2), 199-222.
- 주호진·최희용·최윤희. (2022). 디지털플랫폼정부와 정부혁신- 정부 역할 및 기능 재정립을 중심으로. 지방정부연구, 26(3), 307-327.
- 최근열. (2016). 중앙권한 및 사무의 지방이양 실태 및 발전과제. 한국지방자치연구, 18(3), 25-45.
- 최송이·최병대. (2012). 중앙-지방정부간 역할분담에 대한 추이분석: 1991년 이후 지난 20 년간의 사무배분을 중심으로. 한국지방자치학회보, 24(3), 1-24.
- 하정봉·소진광. (2007). 중앙정부 사무의 지방이양에 대한 실태분석: 44개 기능에 대한 공무원 인식조사를 중심으로. 도시행정학보, 20(2), 29-55.
- 홍준현. (2001). 중앙사무의 지방이양에 있어서 차등이양제도의 도입방향. 한국지방자치학회보, 13(3), 5-24.
- 강송희·김숙경. (2020). GovTech와 공공 생태계 혁신. 경기: 소프트웨어정책연구소.
- 한국지방행정연구원. (2022). 법령상 사무총조사 연구. 한국지방행정연구원.
- 한부영·박재희. (2019). 국가와 지방자치단체 간 사무배분 원칙과 기준 재정립 방안 연구. 한국지방행정연구원 연구보고서 2019-08.

- 국가정보원. (2023.06.29.) 보도자료: 국정원, '챗GPT 등 생성형 AI 활용 보안 가이드 라인' 배포 (검색일: 2023.12.20.)
- 대한민국 정부. (2022). 윤석열정부 120대 국정과제
- 디지털플랫폼정부위원회. (2022). 보도자료: 대통령 직속 디지털플랫폼정부위원회 출범. 2022.09.02. (검색일: 2023.12.20.)
- 산업통상자원부. (2023.06.14.). 보도자료: 인공지능(AI) 윤리 국가표준(KS) 첫 제정 (검색일: 2023.12.20.)
- 세계일보. (2023). "2024년은 디지털플랫폼정부 구현 원년"... 예상 9262억원 '두배 경충' (검색일: 2023.09.04.)
- Google. (2023.05.). Generative AI FAQs.
- Gozalo-Brizuela, R., and E. C. Garrido-Merchan. (2023). "ChatGPT is not all you need. A State of the Art Review of large Generative AI models," arXiv preprint arXiv:2301.04655, 1-22.
- OECD. (2019). Digital Government Review of Sweden: Towards A Data-Driven Public Sector.
- Tobias, H. et al. (2023.04.). "Exploring opportunities in the generative AI value chain," McKinsey Digital, 1-10.

부 록

[부록 1] 사례 적용 및 결과값 비교: 「우주손해배상법」

우주손해배상법					결과값 비교						
					2022년 법령상 사무총조사		자동화모델				
조	항	호	조문 제목	조문	사무 여부	사무 유형	사무여부		사무유형(확률)		
							여부	예측확률	국가	지방	공동
1	0	0	목적	이 법은 우주손해가 발생한 경우의 손해배상 범위와 책임한계 등을 정하여 피해자를 보호하고 우주개발 사업의 건전한 발전에 기여하는 것을 목적으로 한다.	0	0	0.514	0.000	0.000	0.000	
2	0	0	정의	이 법에서 사용하는 용어의 정의는 다음과 같다.	0	0	0.022	0.000	0.000	0.000	
2	0	1		“우주물체”란 「우주개발진흥법」 제2조 제3호에 따른 우주물체를 말한다.	0	0	0.007	0.000	0.000	0.000	
2	0	2		“우주물체 발사자”란 「우주개발진흥법」 제8조에 따라 우주물체를 예비등록 또는 등록한 자나 같은 법 제11조에 따라 우주발사체 발사허가를 받은 자를 말한다.	0	0	0.027	0.000	0.000	0.000	
2	0	3		“우주물체 발사”란 「우주개발진흥법」 제11조제1항에 따라 허가를 받은 자가 우주물체를 발사하는 것을 말하며, 발사준비·시험발사 및 성공하지 못한 발사를 포함한다.	0	0	0.061	0.000	0.000	0.000	
2	0	4		“우주손해”란 우주물체의 발사·운용 등으로 인하여 발생된 제3자의 사망·부상 및 건강의 손상과 같은 인적 손해와 재산의 파괴·훼손·망실과 같은 물적 손해를 말한다.	0	0	0.053	0.000	0.000	0.000	
3	1	0	국제협약과의 관계	정부는 「우주물체에 의하여 발생한 손해에 대한 국제책임에 관한 협약」에 따라 정부가 외국정부에 대하여 손해배상을 한 경우에는 우주물체 발사자에 대하여 구상할 수 있다.	0	0	0.061	0.000	0.000	0.000	

우주손해배상법					결과값 비교						
					2022년 법령상 사무총조사		자동화모델				
조	항	호	조문 제목	조문	사무 여부	사무 유형	사무여부		사무유형(확률)		
							여부	예측확률	국가	지방	공동
3	2	0		이 법은 대한민국 국민, 대한민국의 법령에 따라 설립된 법인·단체 또는 대한민국 정부가 입은 우주손해의 배상을 금지하거나 제한하는 국가의 개인·법인·단체 또는 정부에 대하여는 그 적용을 배제 또는 제한할 수 있다.	1	국가	1	0.770	0.270	0.044	0.051
4	1	0		우주손해가 발생한 경우에는 해당 우주물체 발사자가 그 손해를 배상할 책임이 있다. 다만, 국가간의 무력충돌, 적대행위, 내란 또는 반란으로 인한 우주손해와 우주공간에서 발생한 우주손해의 경우에는 고의 또는 과실이 있는 경우에 한한다.	0		1	0.612	0.155	0.056	0.074
4	2	0	무과실 책임 및 집중 등	제3자의 고의 또는 과실로 인하여 생긴 우주손해를 제1항에 따라 배상한 우주물체 발사자는 그에 대하여 구상할 수 있다. 다만, 그 손해가 우주물체 발사 등에 제공될 자재의 공급이나 역무(노무를 포함한다. 이하 같다)의 제공에 의하여 생긴 때에는 해당 자재의 공급이나 역무를 제공한 자나 그 종업원의 고의 또는 중대한 과실이 있을 때에 한하여 구상할 수 있다.	0		0	0.193	0.000	0.000	0.000
				우주손해에 대하여는 「제조물책임법」을 적용하지 아니한다.	0		1	0.636	0.100	0.043	0.048
5	0	0	손해 배상 책임 한도액	우주물체 발사자가 배상하여야 하는 책임 한도는 2천억원으로 한다.	0		1	0.625	0.093	0.045	0.047
6	1	0	손해 배상 책임 보험의 가입	「우주개발진흥법」 제11조에 따라 우주발사체의 발사허가를 받고자 하는 자는 손해배상을 목적으로 하는 책임보험에 가입하여야 한다.	0		1	0.544	0.089	0.051	0.052

우주손해배상법					결과값 비교						
					2022년 법령상 사무총조사		자동화모델				
조	항	호	조문 제목	조문	사무 여부	사무 유형	사무여부		사무유형(확률)		
							여부	예측확률	국가	지방	공동
6	2	0		제1항에 따라 가입하여야 하는 보험금액은 제5조에 따른 손해배상책임 한도액의 범위 안에서 우주물체의 특성, 기술의 난이도, 발사장 주변 여건 및 국내외 보험시장 등을 고려하여 과학기술정보통신부장관이 정하여 고시한다.	0		1	0.646	0.102	0.037	0.068
7	1	0		정부는 우주손해가 발생한 경우에 피해자의 구조 및 피해의 확대 방지에 필요한 조치를 시행하여야 한다.	0		0	0.054	0.000	0.000	0.000
7	2	0	정부의 조치	정부는 제4조제1항에 따라 우주물체 발사자가 배상하여야 할 손해배상액이 제6조 제2항의 보험금액을 초과하는 경우에 이 법의 목적을 달성하기 위하여 필요하다고 인정할 때에는 우주물체 발사자에 대하여 필요한 지원을 할 수 있다.	0		0	0.030	0.000	0.000	0.000
7	3	0		정부가 제2항의 지원을 할 때에는 국회의 의결에 의하여 허용된 범위 안에서 한다.	1	국가	1	0.583	0.273	0.153	0.107
8	1	0	권리 행사의 기간	이 법에 따른 손해배상청구권은 피해자 또는 그 법정대리인이 그 손해 및 제4조제1항에 따라 손해배상책임을 지는 자를 안 날부터 1년 이내에 행사하지 아니하면 시효로 인하여 소멸한다.	0		0	0.007	0.000	0.000	0.000
8	2	0		이 법에 따른 손해배상청구권은 우주손해가 발생한 날부터 3년이 경과한 경우에는 행사하지 못한다.	1	국가	1	0.630	0.134	0.038	0.045
9	0	0	규제의 재검토	과학기술정보통신부장관은 제8조에 따른 손해배상청구권의 권리행사 기간에 대하여 2016년 1월 1일을 기준으로 2년마다(매 2년이 되는 해의 1월 1일 전까지를 말한다) 폐지, 완화 또는 유지 등의 타당성을 검토하여야 한다.	1	국가	1	0.708	0.238	0.043	0.051

주: 노란색 음영표시는 결과 불일치 조항임

[부록 2] 데이터 처리: 사무판단, 사무유형 등

□ 결측치에 따른 전처리

- 설명변수(소관부처명, 법령명, 조번호, 항번호, 호번호, 조문제목, 조문)가 모두 결측치인 행은 삭제처리함 (총 8행)

[그림 부-1] 설명변수(소관부처명, 법령명, 조번호, 항번호, 호번호, 조문제목, 조문)가 모두 결측치인 행의 삭제처리 (총 8행)

	소관부 처명	법령 명	법령 구분	조번 호	항번 호	호번 호	조문 제목	조문	사무 판단	사무판 단근거	...	수임기관	특별 기관	재위임 사무판 단	재위임근 거규정	재수임 기관	위탁사 무판단	위탁근거 규정	수탁기 관	사무유형 (소분류)	기타
101191	NaN	NaN	1	NaN	NaN	NaN	NaN	NaN	1.0	NaN	...	NaN	0.0	0.0	NaN	NaN	0.0	NaN	NaN	시도-시군 구공통사무	NaN
623132	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	NaN	...	NaN	NaN	0.0	NaN	NaN	0.0	NaN	NaN	국가직접처 리사무	NaN
781030	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.0	위임/위 탁근거	...	NaN	NaN	0.0	NaN	NaN	0.0	NaN	NaN	NaN	NaN
837934	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	NaN	...	국립환경 과학원장	0.0	0.0	NaN	NaN	0.0	NaN	NaN	국가직접처 리사무	NaN
840880	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	NaN	...	NaN	NaN	0.0	NaN	NaN	0.0	NaN	NaN	시도-시군 구공통사무	NaN
854636	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	NaN	...	시도지사	0.0	0.0	NaN	NaN	1.0	시행령 45조2항	국립공 원공단	시도위임사 무	NaN
854677	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	NaN	...	시도지사	0.0	0.0	NaN	NaN	1.0	시행령 45조2항	국립공 원공단	시도위임사 무	NaN
854775	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1.0	NaN	...	시도지사	0.0	0.0	NaN	NaN	1.0	시행령 45조2항	국립공 원공단	시도위임사 무	NaN

8 rows × 25 columns

- 동일한 법령에 대해서는 동일한 소관부처를 가져야 함. 이에 따라 동일한 법령에 대해서는 소관부처명이 동일하도록, 각 법령에 해당하는 소관부처명을 채워 넣는 처리를 함
- 법령명은 있으나, 소관부처가 결측치인 행의 법령명들은 소관부처를 채워 넣는 처리를 함 (36개 행)

[그림 부-2] 법령명은 있으나, 소관부처명이 결측치인 경우(36개 행)

소관부처명	법령명	법령구분	조번호	항번호	호번호	조문제목	조문	사무관단	사무판단근거	...	수인기관	특행기관	재위원사무관단	재위원근거규정	재수일기관	위탁사무판단	위탁근거규정	수탁기관	사무유형(소분류)	기타
217829	NaN	건설산업기본법	1.0	17	2	NaN	건설업의양도등 ② 제1항제1호에 따라 건설업양도신고를 하려는 자가 「국가를 당사자로 하는 계약에 ...	0.0	다른 법률과의 관계	...	NaN	NaN	0.0	NaN	NaN	0.0	NaN	NaN	NaN	NaN
461624	NaN	보건범죄단속에관한특별조치법시행령	2	1	NaN	NaN	목적 제1조(목적) 이 영은 보건범죄단속에관한특별조치법(이하 "법"이라 한다)의 시행에 ...	0.0	목적	...	NaN	NaN	0.0	NaN	NaN	0.0	NaN	NaN	NaN	NaN
461625	NaN	보건범죄단속에관한특별조치법시행령	2	2	NaN	NaN	용어의정의 제2조(용어의 정의) 이 영에서 "주된 성분"이라 함은 의약품의 효능에 직접적인 영...	0.0	정의	...	NaN	NaN	0.0	NaN	NaN	0.0	NaN	NaN	NaN	NaN
461626	NaN	보건범죄단속에관한특별조치법시행령	2	3	NaN	NaN	특정의약품 제3조(특정의약품) 법 제3조제1항제1호의 규정에서 "대통령령으로 정하는 의약품"은...	0.0	정의	...	NaN	NaN	0.0	NaN	NaN	0.0	NaN	NaN	NaN	NaN
461627	NaN	보건범죄단속에관한특별조치법시행령	2	3	NaN	1.	특정의약품 1. 항생물질과 그 제제중 내복제 및 주사제	0.0	호	...	NaN	NaN	0.0	NaN	NaN	0.0	NaN	NaN	NaN	NaN

[그림 부-3] 소관부처 결측치의 처리

'건설산업기본법' 1 개행 -> 국토교통부
 '보건범죄단속에관한특별조치법시행령' 24 개행 -> 보건복지부
 '항로표지법' 5 개행 -> 해양수산부
 '수산자원관리법' 2 개행 -> 해양수산부
 '연안관리법' 3 개행 -> 해양수산부
 '야생생물 보호 및 관리에 관한 법률' 1 개행 -> 환경부

○ 2개 이상의 소관부처명에 대해 순서를 동일하게 변경하였음

[그림 부-4] 2개 이상의 소관부처명 순서 변경 처리

```
def blank_preprocessing(df):
    df.loc[df['소관부처명']=='교육부, ₩n 고용노동부',
'소관부처명'] = '고용노동부, 교육부'
    df.loc[df['소관부처명']=='과학기술정보통신부, ₩n 교육부',
'소관부처명'] = '과학기술정보통신부, 교육부'
    df.loc[df['소관부처명']=='교육부, ₩n 과학기술정보통신부',
'소관부처명'] = '과학기술정보통신부, 교육부'
    return df
```

- 소관부처명이 동일하지만, 순서가 달라 별개의 소관부처로 인식하고 있기 때문에 교육부, \n과학기술정보통신부 / 과학기술정보통신부, \n교육부 총 11개 행을 → 과학기술정보통신부, 교육부로 통일시키는 작업을 거침
- 또한 고용노동부, 교육부로 동일하게 통일시킴

[그림 부-5] 소관부처의 순서가 다른 경우 처리

교육부, Wn 고용노동부	109
고용노동부, 교육부	769
과학기술정보통신부, 교육부	133
교육부, Wn 과학기술정보통신부	7
과학기술정보통신부, Wn 교육부	4

- 법령구분과 소관부처가 모두 결측치인 행은 국가법령정보센터 데이터 확인 후, 해당 법령명과 일치하는 법령구분(법령, 시행령, 시행규칙) 결측값을 채워 넣음 (총 9개 행)

[그림 부-6] 법령구분과 소관부처가 모두 결측치인 행 처리

소관부처명	법령명	법령구분	조번	행번	호번	표제	표준	사무명	사무명	...	수령기관	특정기관	재위임사	재위임사	재수령기관	위탁사	위탁기관	수탁기관	사무유형	기타	
14168	NaN	NaN	NaN	48	3	0	과태료	제1항 및 제2항에 따른 과태료는 다음 표정의 시·도지사는 소속 시·군·구의 직권으로 ...	1.0	NaN	..	지방고용노동관리서	1.0	NaN	NaN	NaN	0.0	NaN	NaN	특별지방행정기관사무	NaN
198519	NaN	NaN	NaN	NaN	NaN	NaN	NaN	① 제10조제1항제3호의 서약규제를 관할하는 시·도지사는 소속 시·군·구의 직권으로 ...	1.0	NaN	..	NaN	NaN	0.0	NaN	NaN	0.0	NaN	NaN	시도지정사무	NaN
686791	NaN	NaN	1	32	2.0	NaN	과태료	② 제1항에 따른 과태료는 대통령령으로 정하는 내역에 따라 해당수신부담금 또는 지참사...	1.0	NaN	..	특별자치도지사	0	0.0	NaN	NaN	0.0	NaN	NaN	시군구위임사무	NaN
708300	NaN	NaN	1	7	2	NaN	해양공간관리계획의 수립 등	② 제1항에도 불구하고 시·도지사는 물 이상의 시·도에 걸쳐 관리계획을 수립할 경우...	1.0	NaN	..	NaN	NaN	0.0	NaN	NaN	0.0	NaN	NaN	국가직접사무	NaN
708831	NaN	NaN	1	6	1	NaN	실적조사	③ 해당근로자는 다음 각 호의 사항을 수행하기 위하여 해당기관을 통보할 수 있다.	1.0	NaN	..	NaN	NaN	0.0	NaN	NaN	1.0	법제3조 제7항	해양조사기관	시도위탁사무	NaN
756079	NaN	NaN	1	74	6.0	NaN	감독 등	④ 주무부장관은 제1항부터 제5항까지의 규정에 따른 공고 또는 운영회에 대한 감독 ...	1.0	NaN	..	NaN	NaN	0.0	NaN	NaN	0.0	NaN	NaN	시군구직립사무	NaN
859679	NaN	NaN	NaN	18	1	NaN	고농도 미세먼지 비상저감조치	① 시·도지사는 환경부장관이 정하는 기간 동안 소미세먼지 배출 농도가 한 정해일 경우...	1.0	NaN	..	수도권대기환경청	1.0	0.0	NaN	NaN	0.0	NaN	NaN	특별지방행정기관사무	NaN
859692	NaN	NaN	NaN	20	1	NaN	비상저감조치 절차의 보고 등	① 시·도지사가 비상저감조치를 발령한 때에는 그 발령일부터 30일 이내에 환경부장관으로...	1.0	NaN	..	유역환경청, 지방환경청, 수도권대기환경청	1.0	0.0	NaN	NaN	0.0	NaN	NaN	특별지방행정기관사무	NaN
859755	NaN	NaN	NaN	26	1	NaN	자료제출 검사 등	① 환경부장관 또는 시·도지사는 이 법의 시행에 필요하다고 인정하는 경우 다음항목으로...	1.0	NaN	..	NaN	NaN	0.0	NaN	NaN	0.0	NaN	NaN	시도지정사무	NaN

[그림 부-7] 국가법령정보센터 데이터 확인 후 결측값 처리

```

idx = 14168
df.loc[idx, '소관부처명'] = '고용노동부'
df.loc[idx, '법령명'] = '근로자퇴직급여 보장법'
df.loc[idx, '법령구분'] = 1

idx = 198519
df.loc[idx, '소관부처명'] = '국토교통부'
df.loc[idx, '법령명'] = '택수운송사업의 발전에 관한 법률'
df.loc[idx, '법령구분'] = 1
df.loc[idx, '조번호'] = 11
df.loc[idx, '항번호'] = 1
df.loc[idx, '조문제목'] = '감차계획의 수립 및 시행 등'

idx = 686791
df.loc[idx, '소관부처명'] = '해양수산부'
df.loc[idx, '법령명'] = '수산업·어촌 공익기능 증진을 위한 직접지불제도 운영에 관한 법률'

idx = 708300
df.loc[idx, '소관부처명'] = '해양수산부'
df.loc[idx, '법령명'] = '해양공간계획 및 관리에 관한 법률'

idx = 708831
df.loc[idx, '소관부처명'] = '해양수산부'
df.loc[idx, '법령명'] = '해양폐기물 및 해양오염퇴적물 관리법'

idx = 766079
df.loc[idx, '소관부처명'] = '행정안전부'
df.loc[idx, '법령명'] = '세마들급고법'

idx = 859679
df.loc[idx, '소관부처명'] = '환경부'
df.loc[idx, '법령명'] = '미세먼지 저감 및 관리에 관한 특별법'

idx = 859692
df.loc[idx, '소관부처명'] = '환경부'
df.loc[idx, '법령명'] = '미세먼지 저감 및 관리에 관한 특별법'

idx = 859755
df.loc[idx, '소관부처명'] = '환경부'
df.loc[idx, '법령명'] = '미세먼지 저감 및 관리에 관한 특별법'
    
```

○ 법령구분의 결측치 처리에 관해서는, 법령이 “법률”에 해당하는 것을 확인 후 법령구분 1로 처리함(8개 행)

- (6개 행) 법령구분인 Nan인 경우
- (2개 행) 법령구분이 공백인 경우

[그림 부-8] 법령구분이 Nan인 경우 (6개 행)

소관부처명	법령명	법령구분	조번호	항번호	조문제목	조문	사무관 단	사무원 단근거	...	수입기관	특별기관	제외대상 여부	제외대상 사유	제수입 기관	위탁사 여부	위탁사 계약	수탁기관	사무유형(소분류)	기타			
217769	국토교통법	건설산업기본법	NaN	9	4	NaV	건설업 등록 등			④ 식재 <2016.2.3>	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN		
544789	산업통상자원부	경제자유구역의 지정 및 운영에 관한 특별법	NaN	27.4	0.0	8	경제자유구역을 지정하는 경우			8 제2조외개 따른 핵심전략사업 선정 요청 및 투자유치 촉진을 위한 특별법	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN		
544790	산업통상자원부	경제자유구역의 지정 및 운영에 관한 특별법	NaN	27.4	0.0	9	경제자유구역을 지정하는 경우			9 그 밖의 대통령령으로 정하는 업무	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN		
859679	환경부	미세먼지 저감 및 관리에 관한 특별법	NaN	18	1	NaV	기능도 개선한 지 비상저감조치			① 시·도지사는 환경부장관이 정하는 기간 동안 초미세먼지 연속 농도가 한계부일 경우...	1.0	NaN	...	수도권대기환경청	0.0	0.0	NaN	NaN	0.0	NaN	특별지정 행정기관	
859692	환경부	미세먼지 저감 및 관리에 관한 특별법	NaN	20	1	NaV	비상저감조치 절차에 보고 등			① 시·도지사가 비상저감조치를 발령한 때에는 그 발령일부터 30일 이내 환경부장관...	1.0	NaN	...	유역환경청, 지방환경청, 수도권대기환경청	1.0	0.0	NaN	NaN	0.0	NaN	NaN	특별지정 행정기관
859755	환경부	미세먼지 저감 및 관리에 관한 특별법	NaN	26	1	NaV	자료제공: 감사 등			① 환경부장관 또는 시·도지사는 이 법의 시행에 필요하다고 인정하는 경우 대통령령...	1.0	NaN	...	NaN	NaN	0.0	NaN	NaN	0.0	NaN	NaN	시도지정 행정기관

[그림 부-9] 법령구분이 공백인 경우 “ ” (2개 행)

소관부처명	법령명	법령구분	조번호	항번호	호번호	조문제목	조문	사무관단	사무관단	수임기관	특별기관	대위업무단	대위업무규정	제수임기관	위탁사무단	위탁근거규정	수탁기관	사무유형(소분류)	기타	
771451	대한안전보험산업 시행령		13	3.0	NaN	대한안전보험산업 시행령	대한안전 보험기관에서 장이 지체 할부동산등 또는 정보시스템을 제공한정보통신서비스	0.0	정차	..	NaN	NaN	0.0	NaN	NaN	0.0	NaN	NaN	NaN	NaN
802021	행정안전부 행정기관의 조직과 정무원에 관한 통칙		32	7	NaN	조직인단	행정안전부당국은 제청 및 제청에 따른 업무를 효율적으로 수행하기 위하여 전문...	0.0	방범	..	NaN	NaN	0.0	NaN	NaN	0.0	NaN	NaN	NaN	NaN

2 rows x 25 columns

○ 법령이 “시행령”에 해당하는 것을 확인 후 법령구분 2로 처리함. 또한, 법령구분 자료형을 정수형으로 통일함

[그림 부-10] 법령구분 자료형의 정수형 처리

```
# 법령구분 자료형 통일
df['법령구분'] = df['법령구분'].astype('int64')
```

[표 부-1] 법령구분 자료형의 정수형 처리의 전후 비교

정수형 처리 전		정수형 처리 후	
1	365485	1	365485
2	319828	2	319828
3	176398	3	176398
Name: 법령구분, dtype: int64		Name: 법령구분, dtype: int64	

○ 조문과 조문제목이 모두 결측치인 행 처리(59개 행): 모두 0으로 처리함

[그림 부-11] 조문과 조문제목이 모두 결측치인 행 처리(59개 행)

소관부처명	법령명	법령구분	조번호	항번호	호번호	조문제목	조문	사무관단	사무관단	수임기관	특별기관	대위업무단	대위업무규정	제수임기관	위탁사무단	위탁근거규정	수탁기관	사무유형(소분류)	기타	
12711	고용노동부 노동조합 및 노동관계조정법		1	40	0	0	NaN	NaN	NaN	NaN	..	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
13543	고용노동부 남녀고용평등과 일-가정 양립 지원에 관한 법률 시행규칙		3	1	0	0	NaN	NaN	NaN	NaN	..	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
32488	고용노동부 장애인고용촉진 및 직업재활법		1	27	NaN	NaN	NaN	NaN	NaN	NaN	..	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
33401	고용노동부 장애인고용촉진 및 직업재활법 시행규칙		3	22	NaN	NaN	NaN	NaN	NaN	NaN	..	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
33428	고용노동부 임금채권보장법		1	3.2	0	0	NaN	NaN	NaN	NaN	..	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

□ 사무판단 및 사무판단 근거 칼럼

○ ‘사무판단’ 칼럼

- 표기방식 동일 (nan → 0, “1”, “2”, “3” → 문자형을 숫자int형으로 변환)
- 오분류 행 삭제처리
 - 사무가 아님에도 사무 유형이 분류된 경우
 - 사무임에도 사무 유형이 분류되지 않은 경우
- 결측값 처리
 - 경우1: 사무 유형이 분류된 경우 => ‘1’로 채움
 - 경우2: 사무 유형이 분류되지 않은 경우 => ‘0’으로 채움
- 사무판단 보류행 제거 (0 1)
- 오분류 처리 및 보류항을 제거하였다.

○ 사무판단근거

- 사무판단 공백처리, 사무판단 근거: (삭제, 판단 불가) → np.nan 변환

□ 사무유형, 중복조문, 중복조문 등

○ 사무유형 처리

- 공백제거 후 표기방식을 통일함
- 사무유형은 총 7개로 라벨링이 되어야 하는데(국가, 시도, 시군구, 국가-시도, 시도-시군구, 국가-시도-시군구, 국가-시군구) 사무유형이 ‘공동’, ‘정부’로 오분류된 경우를 발견하여, 연구원분들의 재확인 절차를 거쳐 ‘국가-시도’, ‘국가’로 재라벨링함

[그림 부-12] 사무유형의 재라벨링 진행

```
def change_law_type(df):
    df.loc[df['사무유형']=='국가', '사무유형'] = '국가'
    df.loc[df['사무유형']=='시도', '사무유형'] = '시도'
    df.loc[df['사무유형']=='시군구', '사무유형'] = '시군구'
    df.loc[df['사무유형']=='국가Wn 시도', '사무유형'] = '국가-시도'
    df.loc[df['사무유형']=='국가', '사무유형'] = '국가'
    df.loc[df['사무유형']=='Wn 국가-시도-시군구', '사무유형'] =
'국가-시도-시군구'

    df.loc[df['사무유형']=='공동', '사무유형'] = '국가-시도'
    df.loc[df['사무유형']=='정부', '사무유형'] = '국가'
    return df
```

○ 중복 조문 제거

- 동일한 조문에 대해서 수행주체가 다르거나 사무명이 다를 경우, 이에 따라서 사무의 분류가 달라짐
- 기존 데이터에서는 동일한 조문이 중복으로 나타나는 경우가 많았음. 하지만 이 과정을 수행하기 위해서는 중복사무 조문의 식별과 사무의 개수 판단 작업이 필요함
- 이를 위해서는 고도화된 모델과 접근방식이 필요할 것으로 보이며, 향후 연구에서 개선해야 할 부분임
- 우선 본 연구에서는 중복되는 조문은 제거를 하고 진행함
- 중복 조문을 제거하는 방식은 ['소관부처명', '법령명', '법령구분', '조번호', '항번호', '호번호', '조문제목', '조문'] 칼럼이 동일한 행을 중복사무로 간주하여 이를 추출하여 삭제하였다. 총 13,549개의 중복 조문이 삭제됨

○ 위임사무판단 처리

- 자료형을 마찬가지로 통일함
- '0 1'과 같이 오분류된 행은 → 2로 재 라벨링함

[그림 부-13] 위임사무판단 처리

```
def change_delegated_tasks(df):

df.loc[(df['위임사무판단']==0.0)|(df['위임사무판단']=='0')|(df['위임사무판단']=='0.0')|(df['위임사무판단']==0.0)|(df['위임사무판단'].isna()), '위임사무판단'] = 0
df.loc[(df['위임사무판단']==1.0)|(df['위임사무판단']=='1'), '위임사무판단'] = 1
df.loc[(df['위임사무판단']=='0 1'), '위임사무판단'] = 2

df['위임사무판단'] = df['위임사무판단'].astype(int)
return df
```

○ 기타 변수 자료형 통일

- [특행기관, 재위임 사무판단, 재위임 근거 규정, 재수임기관, 위탁사무판단] 칼럼에 대해서도 동일하게 자료형을 정수형으로 변경함

○ 대분류 칼럼 추가

- 따로 대분류 칼럼이 원본 데이터에는 없어서 대분류 칼럼을 추가함

[그림 부-14] 법령구분상 대분류 칼럼 추가 처리

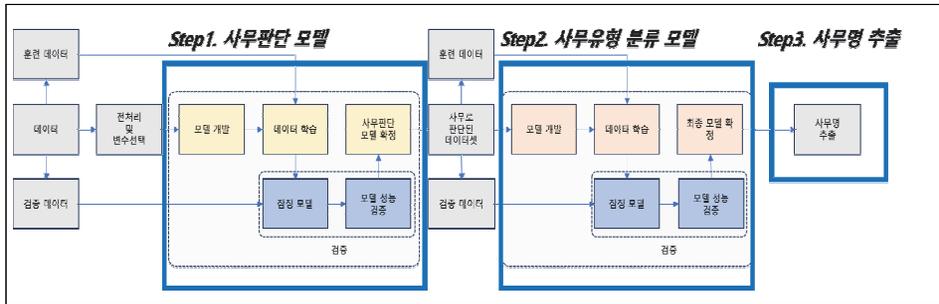
```
def make_large(df):  
    # 사무유형 대분류 추가  
    df['사무유형(대분류)'] = np.nan  
    scale_n = ['국가직접처리사무', '특별지방행정기관사무',  
              '국가위탁사무', '시도위임사무', '시군구위임사무',  
              '시군구재위임사무', '시도및시군구위임사무']  
    scale_p = ['국가-시도-시군구공동사무', '국가-시도공동사무',  
              '국가-시군구공동사무']  
    scale_r = ['시도직접처리사무', '시도위탁사무',  
              '시군구직접처리사무', '시군구위탁사무', '시도-시군구공동사무',  
              '시도-시군구위임사무']  
    for ii in df.index:  
        minitask = str(df.loc[ii, '사무유형(소분류)'])  
        if minitask == 'nan':  
            df.loc[ii, '사무유형(대분류)'] = '0'  
        elif minitask in scale_n:  
            df.loc[ii, '사무유형(대분류)'] = '국가'  
        elif minitask in scale_p:  
            df.loc[ii, '사무유형(대분류)'] = '공동'  
        elif minitask in scale_r:  
            df.loc[ii, '사무유형(대분류)'] = '지방'  
    return df
```

[부록 3] 법령사무조사 자동화 모델의 개발 과정

1. 인공지능 모델 개발 과정

□ 전체 프로세스

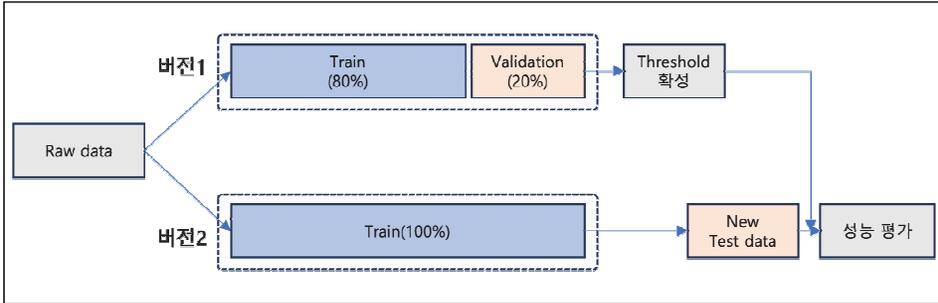
[그림 부-15] 사무판단의 전체 프로세스



- 수집된 데이터를 바탕으로 전처리, 변수선택, 개발, 검증 및 단계를 통해 이루어짐. 데이터를 전처리하여 분석 가능한 형태인 정형 데이터로 가공한 뒤, 모델 개발 과정을 수행함
- 먼저, 사무판단 모델을 개발함. 학습을 통해 모델을 도출하며 데이터 학습으로 도출한 잠정 모델의 성능을 검증하고, 최종 모델을 선정함
- 다음으로, 사무유형 분류 모델을 개발함
- 이 과정을 통해서 도출된 사무라고 판단된 데이터셋만을 이용하여 위와 동일한 과정을 반복함
- 최종적으로 모델이 확정되면, 해당 조문에 대한 사무명을 추출함

□ 데이터 분할

[그림 부-16] 법령구분 자료의 데이터 분할 과정



- 이 모형을 개발하는 과정에서 데이터 분할과 분할하지 않는 방법을 모두 사용하여, 목적에 따라 달리 사용하였음
- 아래와 같은 프로세스로 진행됨
 - 버전1 : 데이터 분할 후 학습 → 검증용 데이터로 최적의 임계값 확정
 - 버전2 : 전체 데이터 학습 → 버전1에서 설정한 임계값 적용 → 새로운 데이터 테스트

□ 데이터 분할 과정

- 버전 1: 데이터 분할 및 임계값 설정
 - 데이터 분할: 전체 데이터셋을 학습용(Training)과 검증용(Validation)으로 8:2의 비율로 분할한다. 이는 모델이 학습 데이터에 과적합되는 것을 방지하고, 모델의 예측 성능을 더 신뢰할 수 있게 검증할 수 있도록 함
 - 임계값(Threshold) 설정: 8:2로 분할한 검증 데이터셋에서 모델의 성능을 평가하여 최적의 임계값을 설정함. 이 임계값은 모델이 어떤 샘플을 특정 클래스(예: 사무 판단)에 속한다고 예측할지를 결정하는 기준이 됨
- 버전 2: 전체 데이터 학습 및 새로운 데이터 테스트
 - 전체 데이터 학습: 이후, 전체 데이터셋(100%)을 사용하여 모델을 재학습함.

[표 부-2] 전체 데이터 개수와 사무판단 열 비율

구분	사무판단 열의 행 개수			비율(%)	
	전체	0값	1값	사무가 아닐 확률	사무일 확률
사무판단					
행 개수	848,072	793,526	54,546	93.6%	6.4%

- 종속변수의 분포가 균일하지 않은 경우에 단순분할 방식으로 데이터세트를 분할하면 특정값이 학습용 또는 검증용 데이터에 편중될 수 있기 때문에 이 경우에 모형학습이 제대로 이루어지지 않아, 모형의 성능이 감소할 수 있음
- 따라서 본 연구에서는 데이터세트를 분할할 때, 계층분할(Stratified split) 방식을 사용하였음. 사무판단이 1인 것과 0인 것의 비율이 6.4%로 동일하도록 분할하는 방식임. 분할한 데이터 개수는 [표 부-3]과 같음

[표 부-3] 데이터 세트 분할 결과

구분	데이터 구분	사무판단 열 개수(비율)		
		소계	0값	1값
사무판단	전체	848,072 (100%)	793,526 (93.6%)	54,546 (6.4%)
	학습용	678,457 (100%)	634,820 (93.6%)	43,637 (6.4%)
	검증용	169,615 (100%)	158,706 (93.6%)	10,909 (6.4%)

2. 사무판단 모델의 개발: 모델별 비교

□ 분석모형의 종류는 사용한 변수의 종류에 따라 달리 설정함

[표 부-4] 데이터를 8:2로 분할하여 학습한 데이터셋(버전1) 기준, 사용변수에 따른 모형

변수	학습 시 사용한 변수							모델 번호
	변수			새로 생성한 변수				
모형	법령 구분	조문 제목	조문	score_ length	score_ count	score_ ratiomean	score_ ratiomulti	
RF	0			0	0	0	0	(1)
LightGBM	0			0	0	0	0	(2)
2-layer-nn		0	0					(3)
transformer encoder + 2-layer-nn		0	0					(4)
ensemble	0	0	0	0	0	0	0	(5)

(1) Randomforest

- 랜덤포레스트(Random Forest)는 분류(Classification) 및 회귀(Regression) 문제에 널리 사용되는 앙상블 학습(Ensemble Learning) 기법 중 하나임
- 이 알고리즘의 핵심은 여러 의사결정 트리(Decision Trees)를 결합하여 하나의 모델을 형성하는 것임
- (사용변수) 앞서 생성한 명사빈도 기반 파생변수인 score_length, score_count, score_ratiomean, score_ratiomulti와 법령구분 변수를 사용함

(2) lightGBM

- LightGBM (Light Gradient Boosting Machine)은 그래디언트 부스팅 기반의 머신 러닝 모델로, 효율적인 학습과 예측 속도, 높은 성능을 제공하는 모델 중 하나임

- LightGBM은 Microsoft에서 개발한 오픈 소스 프로젝트로, 특히 대용량 데이터셋과 카테고리형 특징을 다루는 데 강점을 가지고 있음
- (사용변수) 랜덤포레스트와 동일하게 score_length, score_count, score_ratiomean, score_ratiomulti와 범령구분 변수를 사용함
- 파라미터
 - num_leaves=31
 - learning_rate=0.05
 - n_estimators=100

(3) 2 layer neural network

- 2-layer neural network는 간단한 인공 신경망 구조로, 입력 레이어(input layer), 은닉 레이어(hidden layer), 출력 레이어(output layer)로 이루어진 모델임. 이 모델은 다양한 기계 학습 작업에 사용될 수 있음
- (사용변수) 텍스트데이터인 조문과 조문제목 변수를 결합한 뒤, RoBERTa 토큰나이저를 활용하여 벡터화 과정을 거친 후 이를 input값으로 활용함

[그림 부-18] 2 layer neural network의 구조 및 파라미터

```

* 신경망 구조: input(512), hidden(128), output(2)
* batch_size = 16
* num_epochs = 5
* scheduler = get_linear_schedule_with_warmup(optimizer,
num_warmup_steps=0, num_training_steps=total_steps)
* class_weights = [0.06, 1] #imbalance 데이터 가중치
부여(1에 큰 가중치 부여)
* optimizer = AdamW(model.parameters(), lr=1e-5)

```

(4) transformer encoder + 2 layer neural network

- Transformer Encoder와 2 Layer Neural Network를 결합한 모델로 Transformer의 Encoder 부분만을 사용하고, 이를 통해 추출된 정보를 2-layer 신경망으로 처리하는 구조임
- Transformer Encoder는 각 단어의 위치와 주변 단어들과의 관계를 고려하여 텍스트의 의미를 더욱 잘 이해할 수 있음. 그 후, 이렇게 추출된 정보는 2-layer 신경망을 통해 분류 및 예측 과정에 활용됨
- (사용변수) 2-layer neural network 모델과 동일하게 조문과 조문제목 변수를 결합한 뒤, RoBERTa 토큰나이저를 활용하여 벡터화 과정을 거친 후 이를 input값으로 활용하였음

[그림 부-19] transformer encoder + 2 layer neural network의 구조 및 파라미터

```

○ 모델 구조 & 파라미터
  * 신경망 구조: encoder + input(512), hidden(128),
  output(2)
    * (Encoder 구조)
      * attention_heads = 8
      * feedforward_size = 1024 (512 -> 1024 ->
  512)
      * dropout_rate=0.1
    * batch_size = 16
    * num_epochs = 6
    * scheduler = get_linear_schedule_with_warmup(optimizer,
  num_warmup_steps=0, num_training_steps=total_steps)
    * class_weights = [0.06, 1] #imbalance 데이터 가중치
  부여(1에 큰 가중치 부여)
    * optimizer = AdamW(model.parameters(), lr=1e-5)
  
```

(5) 앙상블 모델

- 앙상블 모형(Ensemble Model)은 여러 개의 기본 모델(Base Models)을 결합하여 하나의 예측 모델을 만드는 방법임
- 이러한 방식은 개별 모델의 강점을 활용하고 약점을 보완함으로써 전반적인 예측 성능을 향상시키는 데 목적이 있음
- 위에서 사용한 4개의 모델에서 나오는 1(사무)의 확률값의 평균을 가지는 새로운 결합 모델을 설계함

3. 사무유형 모델의 개발: 모델별 비교

- 분석모형의 종류는 사용한 변수의 종류에 따라 달리 설정함

[표 부-5] 데이터를 8:2로 분할하여 학습한 데이터셋 기준, 사용변수에 따른 모형

변수	학습 시 사용한 변수								모델 번호
	변수		새로 생성한 변수						
모형	법령명	조문 제목	subject_list	subject_nan	subject_n	subject_r	subject_p	score_subject_len	
RF	0			0	0	0	0	0	(1)
transformer encoder + 2-layer-nn	0	0	0						(2)
ensemble	0	0	0	0	0	0	0	0	(3)

(1) RandomForest

- (사용변수)앞서 생성한 수행주체 기반 파생변수인 subject_nan, subject_n, subject_r, subject_p, subject_len과 법령명을 사용함

- 법령명은 전처리 과정을 통해 법률, 법률에 대한 시행령, 법률에 대한 시행규칙은 같은 범으로 인지하라고 처리함
- (파라미터) 우리의 데이터가 매우 unbalanced 되어있는 상황에서 각각 데이터의 개수에 따라 비중을 두어 가중치를 부여함
 - class_weights = [0.2969, 2.2733, 9.0880, 12.1487]
 - imbalance 데이터 가중치 부여(1에 큰 가중치 부여)

(2) [2-layer + encoder 모델 구조]

- (변수) Input = [법령명 + 조문제목 + 수행주체]
 - 토큰화 모델: 'klue/roberta-large' + (128 padding 적용)

[그림 부-20] 2-layer + encoder 모델 구조

```

○ 모델구조
  * 신경망 구조: encoder + input(128), hidden(64), output(4)
    * (Encoder 구조)
      * attention_heads = 8
      * feedforward_size = 1024   (128 -> 1024 -> 128)
      * dropout_rate=0.1
    * batch_size = 16
    * num_epochs = 6
    * scheduler = get_linear_schedule_with_warmup(optimizer,
num_warmup_steps=0, num_training_steps=total_steps)
    * class_weights = [0.2969, 2.2733, 9.0880, 12.1487]   #
imbalance 데이터 가중치 부여(1에 큰 가중치 부여)
    * optimizer = AdamW(model.parameters(), lr=1e-5)

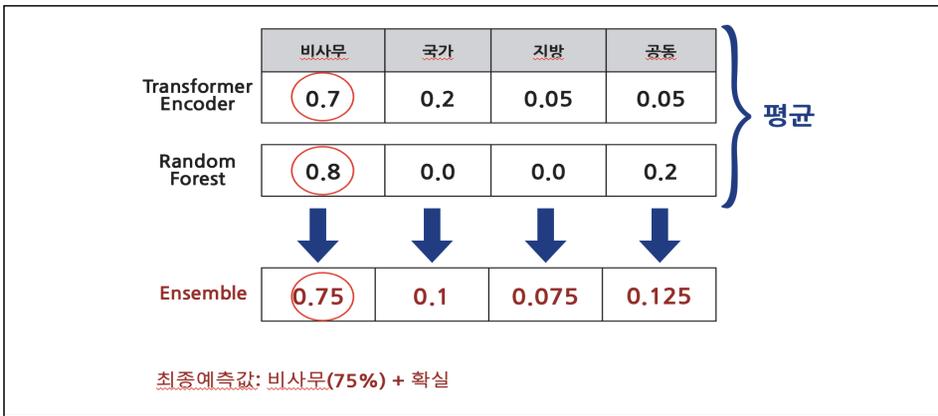
```

(3) 앙상블

- 위에서 사용한 랜덤포레스트와 transformer encoder를 적용한 인공지능망 모델을 결합하여 하나의 새로운 모델을 만들었음

- 앙상블 방식은 각 모델별 클래스 예측값을 조합하여, 두 모델이 모두 같은 클래스로 예측하면 해당 예측값과, '확실'이라는 지표를 함께 나타냄
- 앙상블 모형으로 인해 만들어진 각 클래스별 확률값은 두 모델의 확률값의 평균을 사용함

[그림 부-21] 앙상블 모델 구조(확실한 결과)



- 두 모델이 다른 클래스로 예측하면 해당 예측값과, '애매'라는 지표를 함께 나타냄

[그림 부-22] 앙상블 모델 구조(애매한 결과)

