

# 패킷 기반 민원 다차원 분석을 위한 자동 분류 모델<sup>+</sup>

## (A Study on an Automatic Classification Model for Facet-Based Multidimensional Analysis of Civil Complaints)

김 나 랑<sup>1)\*</sup>  
(Na Rang Kim)

**요약** 시민의 의견인 민원은 다양한 사람들이 여러 주제에 대하여 반복·지속적으로 실시간 쏟아지기 때문에 담당자가 이를 읽고 분석하는데 한계가 있다. 이에 본 연구에서는 빅데이터 분석을 통해 주요 현안에 대한 여론 및 요구 사항을 파악하기 위하여 정성적인 분석에 패킷을 기반으로 한 정량적인 다차원 분석을 위한 자동 분류 모델을 제안하였다. 구체적으로 첫째, 패킷 이론과 정치분석모형을 기반으로 민원 특성을 분석하고 이를 정책 단계에 활용할 수 있는 새로운 분류 프레임워크를 제시하였다. 둘째, 민원 분석 및 처리에 따른 행정 업무를 감소시키고, 시민들의 정책참여를 용이하게 하기 위해 딥러닝을 활용하여 패킷 분석 프레임에 의해 자동으로 속성을 추출하고 분류 하였다. 본 연구결과는 학문적으로 민원 빅데이터의 특성을 이해하고 분석하는데 중요한 단초를 제공하여 향후 많은 후속 연구를 창출할 수 있을 것으로 기대되며, 공공분야를 넘어 교육, 산업, 의료 등 다른 분야에서의 비정형 데이터의 계량화를 위한 가이드 라인과 다차원 분석의 활용에 대한 이론적 근거를 제시할 수 있다. 실무적으로 대용량 전자 민원에 대한 처리체계 개선 및 딥러닝을 통한 자동화로 민원처리 업무의 효율성과 신속성을 높일 수 있으며, 다른 분야의 텍스트 데이터의 처리에 활용될 수 있을 것이다.

**핵심주제어:** 패킷 이론, 민원, 자동 분류, 다차원 분석, 정치분석모형, 딥러닝

**Abstract** In this study, we propose an automatic classification model for quantitative multidimensional analysis based on facet theory to understand public opinions and demands on major issues through big data analysis. Civil complaints, as a form of public feedback, are generated by various individuals on multiple topics repeatedly and continuously in real-time, which can be challenging for officials to read and analyze efficiently. Specifically, our research introduces a new classification framework that utilizes facet theory and political analysis models to analyze the characteristics of citizen complaints and apply them to the policy-making process. Furthermore, to reduce administrative tasks related to complaint analysis and processing and to facilitate citizen policy participation, we employ deep learning to automatically extract and classify attributes based on the facet analysis framework. The results of this study are expected to provide important insights into understanding and analyzing the characteristics of big data related to citizen complaints, which can pave the way for future research in various fields beyond the public sector, such as education, industry, and healthcare, for quantifying unstructured data and utilizing multidimensional analysis. In practical terms, improving the processing system for large-scale electronic complaints and automation through deep learning can enhance the efficiency and responsiveness of complaint handling, and this approach can also be applied to text data processing in other fields.

**Keywords:** Facet Theory, Civil Complaints, Automatic Classification, Multidimensional Analysis, Political Analysis Model, Deep Learning

\* Corresponding Author: whitecoral@hanmail.net

+ 이 논문은 2020년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2020S1A5A8042164)

Manuscript received February 01, 2024 / revised February 16, 2024 / accepted February 21, 2024

1) 동아대학교 경영정보학과, 제1저자

## 1. 서론

새로운 환경변화에 대처하고 시민의 수요에 대응하며 이와 동시에 정부의 목표를 달성하기 위해 시민의 의사를 기반하여 정책 의사를 결정하는 방법은 매우 중요하다. 시민들은 다양한 입장에서 정책집행이나 조치에 대해 자신의 요구를 드러내고 관철시키기 위해 민원을 제기한다. 이러한 민원은 공공정책 과정의 자연스러운 현상으로 정책 결정이나 집행 단계에서 예상하지 못한 다양한 요구에 맞추어 정책을 시행하는데 큰 도움이 된다. 한편, 일반 시민의 거버넌스에 대한 욕구는 인터넷 사용 보편화와 각종 스마트폰기기의 보급 증가로 인해 민원접수의 증가를 불러왔다.

이에 양질의 민원 서비스 제공을 위한 민원 빅데이터를 효과적으로 관리하고 다차원적으로 분석하는 방안 마련이 절실한 상황이다. 국내에서도 민원 데이터 활용 및 연구가 활발히 이루어지고 있다. 관련 연구로 Park et al.(2015)은 공간 빅데이터를 기반으로 인천지역 환경민원 패턴을 분석하였다. Choi(2016)는 환경/위생분야 정책 수립을 위해 민원을 수집하여 주요 키워드를 도출하고, 키워드들 사이의 관계를 분석하였으며, Son and Kim(2017)은 공공민원 자동 분류 시스템 연구를, Yu et al.(2018)은 키워드 분석과 텍스트 마이닝을 통해 버스와 관련된 주요 키워드 간의 변화와 중요도를 연구하였다. 하지만 기존의 연구들은 민원들의 관계를 찾거나 딥러닝 기법을 활용한 민원 자동 분류 연구 등이 주를 이루어 다양한 민원 집행시 우선순위를 결정하여 의사결정을 할 수 있는 다차원 분석에 대한 연구는 없다.

다차원 분석에 있어 Ranganathan(1967)이 발전시킨 패킷 이론은 텍스트에 포함된 개념들을 개별단위의 계층구조로 조직할 수 있는 체계를 마련해 준다(lee et al., 2020). 패킷 이론은 복잡 주제에 대해 다양한 시각 및 관점으로 접근하는 다면적 방식을 말한다. 존재하는 모든 것들의 주제 분야를 5개의 기본 카테고리 구분하고 있는 패킷 분류 체계는 다양한 분야에서 활용되어 다차원 분석의 기반이 되고 있다.

이에 본 연구에서는 민원 빅데이터 분석을 통한 주요 현안에 대한 여론 및 요구 등을 파악하기 위하여 정성적인 분석에 패킷을 기반으로 한 정량적인 다차원 분석을 추가하여, 정책 의사결정에 유용한 지식을 창출하고 적절한 대응 방안을 모색할 수 있는 패킷 기반 자동 분류 모델을 제안하고자 한다.

## 2. 이론적 배경

### 2.1 비정형 빅데이터 분석 기술의 발달

민원과 같은 비정형 빅데이터의 증가함에 따라 학문적으로도 이를 이용하여 자료의 연관성을 규명하고, 과학적으로 관리하며, 전략을 수립하는데 다양한 분야에서 활용 및 연구되고 있다. 빅데이터 분석은 영화에 대한 리뷰 등을 이용하여 영화 관객수, 매출액 등 흥행을 예측할 수 있으며(Lee et al., 2016; Lee and Woo, 2019), 농산물 가격의 등락을 기준으로 감성사전을 구축하여 농산물 관련 온라인 뉴스의 긍정/부정을 분류할 수 있다(Kim et al., 2018). 또한 토픽모델링과 텍스트 분석을 통한 관광인식 분석(Cho, 2021; Han and An, 2022) 또는 외국인 관광객의 선호 관광지 및 관광 매력 요인을 연구할 수 있다(Lee and Kang, 2020). 이외 Choi et al.(2023)은 텍스트 마이닝과 네트워크 군집 분석을 활용하여 한국의 데이터 관련 정책 사업을 분석하였다.

활발한 빅데이터 관련 연구의 이면에는 딥러닝 등 빅데이터 분석 기법의 발전이 있었다. 특히 자연어 처리에서 심층 신경망 기법이 좋은 성능을 나타내고 있다. 이러한 신경망을 통한 딥러닝 기술의 발전으로 비정형 빅데이터의 분석 및 처리 연구가 더욱더 활발히 진행될 수 있는 것이다. 그 중에서도 CNN(Convolution Neural Network)과 RNN(Recurrent Neural Network) 계열이 더욱 효과적이라는 것이 알려지면서 다양한 곳에서 사용되고 있다(Cai et al., 2018; Karim et al., 2018). 관련 연구로 의학 텍스트를 자동 분류한 연구(Banerjee et al., 2018)와 RNN

을 이용하여 트위터 데이터에서 헤이트스피치를 감지한 연구(Pitsilis et al., 2018), LSTM을 이용한 시계열 분류 연구(Karim et al., 2018), 딥러닝 기반의 BERT 모델을 활용한 학술 문헌 자동 분류(Kim and Kim, 2022), 한글 텍스트 감정 이진 분류 모델 생성을 위한 미세 조정과 전이학습에 관한 연구(Kim, 2023) 등이 있다.

## 2.2 민원 분석

쉽게 접할 수 있는 온라인 서비스로 인하여 접수되는 민원은 해마다 늘고 있다. 2010년 이후 민원 접수 건수가 매년 21% 이상 증가하고 있고, 2017년의 2만 6533건에서 2019년에는 5만 6189건으로 그 내용 역시 복잡 다양하게 진화하고 있다. 이렇게 해마다 민원수가 폭발적으로 증가하는 반면 평균 민원 처리 기간은 늘어나고 있고, 60일(법정 처리기간) 초과 처리 민원 또한 매년 증가하는 등 민원 처리의 질이 악화되고 있다. 특히 민원 평균 처리 기간은 2017년 18.8일에서 2018년 19.8일, 2019년 23.0일로 늘었는데 이로 인해 민원 처리 만족도는 2017년 78.5점에서 2018년 76.7점, 2019년 75.9점으로 매년 떨어지고 있다(Newspim.com., 2020). 이렇게 증가하고 있는 민원에 대하여 정부는 적절하게 대응하지 못할 뿐만 아니라 방대한 민원을 파악하고 시민들의 실질적인 관심사를 분석하고 집약하는데 어려움을 겪고 있다(Jeong et al., 2017).

이를 해결하고자 국내에서도 민원 빅데이터 활용 및 연구가 활발히 이루어지고 있다. 관련 연구로 Jin et al.(2019)은 민원 빅데이터 분석을 중심으로 생활밀착형 환경이슈에 대한 수요반영 개선 연구를 하였으며, Lee(2020)는 해양경찰 민원을 분석하여 민원서비스 개선 방안 수립을 위한 기초자료를 제안하였다. 또한 Lee et al.(2021)은 민원 데이터 분석을 통한 방송중·고 운영 개선 방안을, Kim et al.(2023)은 민원데이터 분석을 통한 멸종위기종 보호 정책 개선 방안을 제시하였다. Joo(2023)는 집단민원의 감정 분석을 이용한 공간 빅데이터 시각화 방안을 연구하였다.

한편, 통계 등의 정량적 데이터의 분석과 달

리 비정형 데이터인 민원은 정책 판단의 근거자료로 그것을 '어떻게 해석할 것인가'와 '채택할 것인지 말것인지'에 대한 판단이 요구된다. 하지만 기존의 민원 활용 및 관련 연구는 다양하고 복잡한 민원 데이터의 특성을 무시하고, 일차원적 분석인 민원 빈도와 주제별로 나누는 텍스트마이닝이 가장 많이 이루어지고 있다(Kim et al., 2018). 특히 동향 분석, 감성분석 또는 연관성분석 등 단편적인 분석만을 실시하여, 의사결정의 근거로 활용하기가 어렵다. 또한 다량의 민원 빅데이터를 대상으로 빠른 시간내에 새로운 정책 수요를 예측하고 의사결정을 지원하는데 한계가 있다.

증가하는 시민의 거버넌스 요구에 부응하여 디지털 직접 민주주의를 실현하고 공공대국민 서비스 질 확보를 위해, 비정형 빅데이터로 다양한 주제를 가진 민원의 다차원적인 분석이 필요하다. 또한 객관적인 정책 의사결정의 근거를 제시하여 정책 수요자인 시민의 정책 참여를 지원하여야 한다. 즉, 시민이 생산한 비정형 빅데이터에서 더 가치 있는 정보를 추출하고 분석하여 현상의 본질에 가장 근접한 종합적 추론을 제시해야 한다. 이에 기존의 담당자 경험이나 민원 빅데이터 양으로 선정하던 단순 방식에서 벗어나, 객관적인 지표에 의한 다차원 분석 결과를 활용해 과학적이고 효율적인 정책 개발과 정책 의사결정이 이루어 질 수 있어야 한다. 이를 위해 민원 빅데이터 처리 및 분석 방법을 개선하고 다차원 분류체계를 기준으로 딥러닝 기술을 활용한 민원 분류 및 분석에 대한 자동화가 필요하다.

## 2.3 패킷 이론

복합 주제에 대해 다면적으로 접근하는 패킷 이론은 주제 분야를 하위개념 및 속성으로 분석하고, 정보자원을 분류하거나 기술할 때 적합한 카테고리들을 조합·합성하여 다면적인 주제를 표현할 수 있게 만들어 낸다(Kim, 2005).

랑가나단은 존재하는 모든 것들의 주제 분야를 5개의 기본 카테고리로 구분하고, 주제를 형성하는 본질적인 속성인 P(Personality), 재료 및

부분인 M(Matter), 행위들을 나타내는 E(Energy), 지리적 장소 S(Space), 시간이나 시대적 속성인 T(Time)으로 정의하였다.

패킷 분류 체계는 다양한 분야에서 활용 되어 다차원 분석의 기반이 되고 있다. Rosati et al.(2004)은 패킷기반 전자 정부의 코퍼스 분석를 통해 전자정부의 분석체계를 제시하였고, Kwon et al.(2006)은 법률정보를 제공하는 온라인 상용 데이터베이스의 인터페이스를 평가하였다. Park et al.(2010)은 패킷 이론을 이용하여 지진 발생시 행동 유형을 분석하여 피난 행동 심리에 관하여 다차원 분석을 하였고, Park and Kim (2011)은 주거 공간 설정의 심리적 요인에 관한 연구를, Park et al.(2019)은 재난안전정보 패킷 분류체계를 제시하고 실증적으로 검증하였다. 정책관련 연구로 Jeong et al.(2010)은 패킷 이론을 바탕으로 한 프로토콜 분석을 통해 공적 토론에서 참가자 간의 인지 부조화를 조사하는 새로운 방법론을 제안하였다. 즉, 공공프로젝트 내에서 참가자들의 의견 부조화를 조사하기 위해 패킷 이론에 기초하여 다차원 척도분석을 하여 갈등과 논쟁의 패턴을 설명하였다. Lee et al.(2020)은 패킷 분석기법을 활용하여 부산진구의회 의사결정과정에서 나타난 발언목적, 논증성격, 갈등원인을 분석하여, 지방의회 의사결정과정의 개선을 위한 정책적 함의를 제시하였다.

#### 2.4 Political Analysis Model

민원을 기반으로한 정책에 대한 정치적 분석은 정보의 수집과 분류, 실현가능성에 대한 판단과정을 거친다. Meltsner(1972)는 실현가능성이란 정책이 정부나 정치체계 내에서 어떻게 수용되고 실행될 수 있는지를 의미하며, 실현가능성을 고려한 정책 분석이 정부와 정책 제안자에게 실용적 가치를 제공할 수 있다고 하였다. 특히 정책 분석과 관련된 중요한 주제인 실현가능성에 대해 논의하고, 정책 제안이나 변경사항을 실행 가능하게 만들기 위해 정치적인 측면을 고려해야 한다고 강조하였다. 이를 위해 정책 분석의 이론적인 측면과 현실적인 정치적 상황 간의 관계를 탐구하며, 실행 가능한 정책을 개발

하는 데 필요한 접근 방법을 제안하였다. Political Analysis Model은 정책사례를 정치적 관점에서 분석할 경우 필요한 체크 항목을 제시하고 있어, 정책 대안을 정치적 실행가능성의 측면에서 분석하는데 도움을 준다는데 의의가 있다(Ban, 2016). 정책 분석에서 정치적 쟁점을 이해하고 분석하기 위한 Political Analysis Model의 핵심 요소는 정책관여자(Actors), 동기(Motivation), 신념(Beliefs), 자원(Resource), 장소(Site), 교환(Exchange)로 구성된다. 정책관여자(Actors)는 정치적 결정에 관여하는 주체들을 나타내며 이들은 정부 관료, 정치인, 이해관계자, 시민단체, 기업, 시민 등 다양한 그룹과 개인을 포함할 수 있다. 동기(Motivation)는 정책 관여자들이 특정 정책이나 결정에 대해 왜 노력하고, 어떤 이유로 행동하는지를 나타내며, 정치적 영향력을 키우거나 유지하기 위한 관여자의 관심, 목표, 욕망을 포함한다. 신념(Beliefs)은 정책관여자들의 신념 체계와 가치관을 나타내며, 관여자들의 입장과 정책 선호도에 영향을 미친다. 이에 정책관여자들의 이념적 입장과 태도를 이해하는 것이 중요하다. 자원(Resource)은 정책관여자들이 정책 실행을 위해 사용할 수 있는 자금, 인력, 지식, 정보, 기술 등의 자원을 의미한다. 장소(Site)는 정책 실행과 관련된 물리적 및 제도적 환경을 나타내며, 정책 실행이 이루어지는 위치나 공간뿐만 아니라 제도적 규칙과 절차도 포함한다. 교환(Exchange)은 정책 관여자들 간의 상호작용과 협력을 의미하며 정보 교환, 협상, 양해 감정, 지원 등을 말한다. 특히 교환과 협력은 정책의 성공 또는 실패에 영향을 미치는 요소이다.

관련 연구로 국내에서는 천성산 터널 갈등사례를 중심으로 Meltsner의 정치분석모형을 적용한 Kim(2008)의 연구와 표준화된 정책분석의 질적인 분석 틀을 구축하고, 이를 활용하여 대학구조개혁의 정치적 특성과 쟁점을 분석한 Ban(2016)의 연구가 있다.

### 3. 패킷 기반 민원 분석 프레임워크 및 자동화 분류

3.1 패킷 기반 민원 분석 프레임워크

본 연구에서는 Ranganathan의 패킷 이론과 Meltsner의 Political Analysis Model을 기반으로 민원에 대한 다차원적이고 표준화된 분류 틀을 구축하고, 이를 활용하여 민원 특성을 분석하여 정책 단계에 활용하고자 Table 1과 같이 하나의 민원에서 여러 측면을 검토하기 위한 패

킷 기반 민원 분석 프레임워크를 구성하였다. 먼저 Ranganathan의 패킷 이론의 기본 범주 중 본질적인 속성인 P(Personality), 재료 및 부분인 M(Matter), 행위들을 나타내는 E(Energy), 지리적 장소 S(Space), 시간 속성인 T(Time)를 Meltsner의 Political Analysis Model의 Belief/Motivation, Actors/Resource, Exchange, Space와 각각 연결시키고 이를 다시 민원의 특

Table 1 Facet-based Complaint Analysis Framework

Ranganathan Facet Analysis Model		Meltsner's Political Analysis Model		Complaint Category
Facet Basic Category	Characteristics	Main Variable	Key Questions	
P (Personality)	Fundamental attributes that constitute a subject.	Belief	What is the vision, ideology, and values that the policy aims to promote?	Facet P Subject, Type, Characteristic, - Conflict, - Public Interest, - Interest, - Repeatability, - Complexity, - Urgency, - Objectivity
		Motivation	What are the objectives and major directions sought through the policy?	
M (Matter)	Primary materials or raw materials of all substances used or required in human life	Resource	What are the resources and influence required for policy decision-making and implementation?	Facet M Agent, Stakeholder, - Local Residents, - Local Organization, - Interest Groups, Companies
		Actors	Who are the key stakeholders involved in the policy?	
E (Energy)	All activities or actions in the universe, types of activities and processes, reference materials for activities, etc.	Exchange	What institutional mechanisms are in place for policy decision-making, implementation, and evaluation?	Facet E Cause, Result, - Satisfaction, Public Data, Policy, - Laws, Ordinances, Rules
S (Space)	Space in the universe, countries, regions	Site	What are the primary spaces and tools for policy decision-making and implementation?	Facet S Geographic Data, - Administrative Division, - Address, - Map
T (Time)	Time, Season	-	-	Facet T Temporal Data - Time, Season

성(유형, 주제, 이해관계, 원인 및 결과, 시간 및 지리적 정보 등)에 맞추어 구성하였다.

선행연구를 바탕으로 제시된 패킷 기반 민원 분석 프레임워크를 딥러닝을 통한 자동화 분류를 실험하고자 Table 2와 같이 Facet P에서 부산시 민원 특성에 맞게 하나의 민원에서 5가지 패킷 A, B, C, D, E 로 이루어진 자동화 분류 프레임워크를 구성하였다. 각 패킷에 대한 정의와 속성의 분류 근거는 다음과 같다. 패킷 A는 민원유형을 분류하기 위한 것이고, 패킷 B는 공익성을 패킷 C는 위급성을 패킷 D는 이해관계 충돌성을, 패킷 E는 객관성으로 제시하였다.

Table 2 Facet Classification Framework Based on Complaint Characteristics

	Definition	Facet Classification		
		1	2	3
Facet A	Type of Complaint	Inquiry	Complaint	Suggestion
Facet B	Public Interest	High	Neutral	Low
Facet C	Urgency	High	Neutral	Low
Facet D	Stakeholder Conflict	High	Neutral	Low
Facet E	Objectivity	High	Neutral	Low

### 3.2 학습 및 분석 절차

이 연구의 학습 및 분석 절차는 Fig. 1과 같이 구성되어 있다.

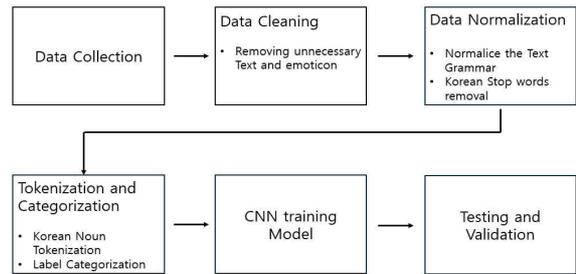


Fig. 1 CNN Training Process

민원 데이터는 정보공개 사이트(open.go.kr)에서 청구신청을 통해 12,906개를 수집한 후(2016년 1월 1일부터 2020년 12월 21일), 여러 게시판 중 ‘부산시에 바란다 게시판’의 8,293개 민원을 대상으로 데이터 정제과정을 거쳤다. Table 3과 같이 앞서 작성한 패킷 분석 프레임워크를 기반으로 3명의 연구자가 민원 속성을 패킷 프레임워크에 따라 분류하여 머신러닝 트레이닝 셋을 작성하였다.

Table 3 Example of Facet Classification

Title	Content	Facet	Facet	Facet	Facet	Facet
		A	B	C	D	E
COVID-19 social distancing sauna	I don't understand the difference in safety between clubs or entertainment venues and bathhouses. Is it possible to strengthen the policies regarding bathhouses? I believe they are more dangerous than entertainment venues due to the high concentration of elderly individuals.	3	1	1	1	3
Busan city bus crew in a vulnerable state to COVID-19	A single city bus transports between 700 and 1200 passengers in a day. Considering that a disposable mask is used once every 3 to 4 days as part of the COVID-19 protective equipment for bus drivers, do you think it makes sense? If not, I would like to request the installation of protective barriers for bus drivers' seats inside city buses, as well as the installation of barriers extending to the side driver's entry door, in order to minimize airborne transmission between passengers and staff.	3	1	1	2	2

### 3.3 패킷 기반 프레임워크 학습 및 분석결과

기계 학습의 대표적인 방법인 딥러닝은 문서 분류 및 클러스터링, 문서 요약, 웹 마이닝 및 감성 분석과 같은 다양한 텍스트 마이닝 문제에 대한 효과적인 해결책으로 부상하였다. 이에 딥러닝 기반의 CNN(Convolution Neural Network)과 RNN(Recurrent Neural Network)계열을 이용한 연구가 늘어났다.

딥러닝 모델중 컴퓨터 비전을 위해 고안된 CNN 모델이 semantic parsing, search query retrieval, sentence modeling, 그리고 전통적인 자연어 처리에 있어서 우수한 성능을 보이고 있다. Kim(2014)는 문장 단위(sentence level)분류 작업을 위해 word2vec 알고리즘을 통해 미리 훈련된(pre-trained) 단어의 벡터 표현을 활용한 CNN을 제안하였다. CNN은 문서 분류 뿐 아니라 자연어 처리 및 텍스트마이닝의 여러 분야에서 다른 전통 분류기보다 더 높은 분류 정확도를 보여주고 있다(Zhang et al., 2015). 이에 본 연구에서는 코딩을 통해 만들어진 학습 데이터를 기반으로 CNN모델을 통한 학습을 진행하였다.

Table 4 Dataset

No	Category	Data
1	Pre-processing Before	8,293
2	Pre-processing After	5,849

데이터는 초기에 8,293개의 데이터로 구성되어 있었는데, 데이터를 정렬하고 유효하지 않거나 비어 있는 데이터를 제거한 후 데이터가 5,849개로 줄어들었다. Table 4는 데이터세트 세부사항을 설명하고 보여준다. 데이터에는 5개의 패킷이 있으며 각 패킷은 주로 상, 중, 하의 계층을 기반으로 한다. 실험에 사용된 머신러닝은 Kim(2014)의 모델을 기반으로 한 CNN이다. 모델의 정확도는 77%이며, Table 5는 CNN을 이용한 실험과 10-Fold Validation을 이용한 검증 결과이다.

Table 5 CNN Classification Performance

Classifier	Accuracy Mean (10 -Validation)	Related Research
CNN	77%	Kim Yoon (2014)

Table 6은 CNN 알고리즘을 이용한 패킷 기반 민원 자동 분류의 결과이다. 패킷 A인 민원 유형은 질의가 59%, 신고 및 고충이 29%, 건의 및 제안이 12%로 대부분 단순 질의가 많은 것으로 나타났다. 또한 패킷 B의 공익성은 중립이 44%, 사익이 39%, 공익이 17%로 나타나 공익을 위한 것보다는 사익 또는 중립이 더 많은 부분을 차지하였다. 패킷 C의 위급성은 위급하지 않음이 64%, 중립이 27%, 위급이 8%로 민원이 위급한 것은 낮게 나타났다. 패킷 D의 이해관계 충돌성은 하가 41%, 상이 34%, 중립이 25%이며, 패킷 E의 객관성은 중립이 45%, 주관적이 39%, 증거를 제시한 객관성은 17%로 나타났다. 종합적으로 볼 때 부산시 민원의 성격은 질의 유형이 많고, 비공익적이며 위급하지 않고 이해관계 충돌성이 낮으며 주관적인 민원이 많은 것으로 나타났다.

Table 6 Classification Result

Facet Types		Classification	
Facet A	Type of Complaint	Inquiry	3279(59%)
		Complaint	1602(29%)
		Suggestion	676(12%)
Facet B	Public Interest	High	970(17%)
		Neutral	2587(44%)
		Low	2292(39%)
Facet C	Urgency	High	483(8%)
		Neutral	1601(27%)
		Low	3765(64%)
Facet D	Stakeholder Conflict	High	1989(34%)
		Neutral	1469(25%)
		Low	2391(41%)
Facet E	Objectivity	High	984(17%)
		Neutral	2607(45%)
		Low	2258(39%)

#### 4. 결 론

새로운 환경변화에 대처하고 국민의 수요에 대응하며 이와 동시에 정부의 목표를 달성하기 위해 정책 의사를 결정하는 방법은 매우 중요하다. 특히 정책 의사결정의 근원이 되는 민원은 다양한 사람들이 여러 주제에 대하여, 반복된 발언과 지속적인 의견을 방대하게 실시간으로 쏟아내기 때문에 이를 소수의 공무원이 전부 읽고 다각적으로 분석하기가 어렵다.

본 연구에서는 빅데이터 분석을 통한 주요 현안에 대한 여론 및 요구 등을 파악하기 위하여 정성적인 분석에 패킷을 기반으로 한 정량적인 다차원 분석을 추가하여, 의사결정에 유용한 지식을 창출하고 적절한 대응 방안을 모색하고자 패킷 기반 자동 분류 모델을 제안하였다.

이를 위해 이에 기존의 단편적인 민원 분석에서 벗어나 시민의 목소리가 담겨져 있으며 다양한 주제를 가진 민원 빅데이터를 패킷 이론의 틀을 활용하여 새로운 다차원 분석 체계를 마련하고자 연구하였다. 선행연구를 통해 패킷 이론과 Meltsner의 Political Analysis Model을 기반으로 민원 특성을 분석하고 이를 정책 단계에 활용할 수 있는 새로운 분류 프레임워크를 제시하였다. 또한 민원 분석 및 처리에 따른 행정 업무를 감소시키고, 시민들의 정책 참여를 용이하게 하기 위해 딥러닝을 활용하여 패킷 분석 프레임에 의해 자동으로 속성을 추출하고 분류 하였다.

본 연구 결과는 학문적으로 민원 빅데이터의 특성을 이해하고 분석하는데 중요한 단초를 제공하여 향후 많은 후속 연구를 창출할 수 있을 것으로 기대되며, 공공분야를 넘어 교육, 산업, 의료 등 다양한 분야에서의 비정형 데이터의 계량화를 위한 가이드 라인과 다차원 분석의 활용에 대한 이론적 근거를 제시할 수 있다. 실무적으로 대응량 전자 민원에 대한 처리체계 개선 및 딥러닝을 통한 자동화로 민원처리 업무의 효율성과 신속성을 높일 수 있으며, 다른 분야의 텍스트 데이터의 처리에 활용될 수 있을 것이다. 또한, 실사구시적 연구로 실제 정책에 적용하고 검증하여 정책 개발시 담당자 경험이나 민원의 양으로 선정하던 단순 방식에서 벗어나 과

학적이고 효율적인 정책 개발을 지원하며 정책 개발의 근거를 제시하고 정책 수요자인 시민의 정책참여를 용이하게 할 것으로 기대된다. 그러나 제시된 패킷 기반 민원 분석 프레임워크를 실증 분석하고자 5가지 패킷만을 이용하여 실험 하였으므로 분석에 한계가 있을 수 있다. 향후 민원의 성격에 따라 다양한 패킷을 조합하여 분석할 것을 제안한다.

#### References

- Banerjee, I., Ling, Y., Chen, M. C., Hasan, S. A., Langlotz, C. P., and Moradzadeh. (2018). Comparative Effectiveness of Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) Architectures for Radiology Text Report Classification. *Artificial intelligence in medicine*.
- Ban, S. J. (2016). Politics of Universities Reorganization Evaluation Policy. *The Politics of Education*, 23(1), 59-89.
- Cai, J., Li, J., Li, W., and Wang, J. (2018). Deep Learning Model Used in Text Classification. *In 2018 15th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, 123-126.
- Cho, Y. (2021). Awareness and resolution of touristification in Korea using social media big data analytics. *Global Business & Finance Review (GBFR)*, 26(1), 68-78.
- Choi, H. O. (2016). Study on Selecting Priority Criteria Utilizing Civil Complaint Data in the Field of Environment and Sanitation. *Environmental policy*, 24(2), 45-57.
- Choi, S. J., Shin, G. Y., and Oh, Y. H. (2023). Analyzing data-related policy programs in Korea using text mining and network cluster analysis. *Journal of Korea Society of Industrial Information Systems*, 28(6), 63-81.

- Han, J. H., and An, K. S. (2022). Comparison of Perceptions of Wellness Tourism in Korea Before and After COVID-19: Results of Social Big Data Analysis. *Global Business & Finance Review*, 27(2), 1-13.
- Jeong, H., Hatori, T., and Kobayashi, K. (2010). A Protocol Analysis of Public Debate using Facet Theory. *In Geographical Marginality as a Global Issue*, 77.
- Jin, D. Y., Kang, S. W., Han, K. J., Kim, J. H., Kim, D. Y., and Kang, S. A. (2019). A Study on Improving Reflection of Demands for Settling Environmental Issues in Daily Lives: Focusing on the Analysis of Civil Complaint Big Data. *KEI Research Report*, 2019(9), 1-100.
- Joo, Y. J. (2023). A Study on the Visualization of Geospatial Big Data using Sentiment Analysis of Collective Civil Complaints. *Journal of the KAGIS*, 26(1), 11-20.
- Karim, F., Majumdar, S., Darabi, H., and Chen, S. (2018). LSTM Fully Convolutional Networks for Time Series Classification. *IEEE Access*, 6, 1662-1669.
- Kim, H. J., Lee, T. H., Yu, S. E., and Kim, N. N. (2018). A Study on Text Mining Methods to Analyze Civil Complaints - Structured Association Analysis. *Journal of Korea Society of Industrial Information Systems*, 23(3), 15-30.
- Kim, I., and Kim, S. (2022). Automatic Classification of Academic Articles Using BERT Model Based on Deep Learning. *Journal of the Korean Society for Information Management*, 39, 293-310.
- Kim, J. S. (2023). A Study on Fine-Tuning and Transfer Learning to Construct Binary Sentiment Classification Model in Korean Text. *Journal of Korea Society of Industrial Information Systems*, 28(5), 15-30.
- Kim, M. S., Yang, H. J., Tien Wing An, J., Joo, J. M., and Chae, H. J. (2018). Sentiment Analysis of News Articles on Agricultural Products Using Convolutional Neural Network. *KIISE Transactions on Computing Practices*, 24(11), 623-627.
- Kim, N., Nam, H., and Park, Y. (2023). Proposals for Improving Endangered Species Protection Policies through Analysis of Citizen Complaint Data. *Proceedings of the Korean Society of Environmental Ecology*, 2023(2), 87-88.
- Kim, S. (2008). Meltsner's political analysis model as a case study method for problem structuring: Focused on the conflict case of Mountain Cheonseong Tunnel. *The Korean Journal of Local Government Studies*, 11(4), 151-170.
- Kim, T. S. (2005). Understanding Classification. Bibliographic Information Processing Society.
- Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification. *In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1746-1751.
- Kwon, K. S., Choi, Y. S., and Hwang, E. Y. (2006). A study on Interface Evaluation of On-line Commercial Database of Legal Information. *Proceedings of the Korean Society for Information Management Conference*, 49-56.
- Lee, E., Gu, Y., and Kim, S. (2021). Analysis on the Inquiries of Open Secondary School for Its Operational Improvement. *Research Institute of Education Korea University*, 34(4), 75-106.
- Lee, H. J., and Kang, Y. O. (2020). Understanding Tourist's Region of Attraction and Image of City through Social Network Data Analysis. *Journal of the Korean Urban Geographical Society*, 23(1), 101-114.
- Lee, I., and Woo, J. (2019). Movie attendance and sales forecast model through big data analysis. *The Korea Journal of BigData*, 4(2), 185-194.

- Lee, K. H. (2020). A Study on the Analysis of Civil Complaints in Maritime Police. *Korean Association of Maritime Police Science*, 10(4), 1-18.
- Lee, S., Jung, C., and Kim, J. (2016). Sentiment analysis on movie reviews through building a modified sentiment dictionary by movie genre. *Journal of Intelligence and Information Systems*, 22(2), 97-113.
- Lee, T., Hong, S., and Jeong, H. (2020). An Automatic Analysis of Public Debate of Local Assembly by Utilizing a Facet Analysis. *The Korea Association for Policy Studies*, 29(1), 47-69.
- Meltsner, A. J. (1972). Political feasibility and policy analysis. *Public administration review*, 859-867.
- Newspim.com. (2020). [2020 National Assembly Audit] National Human Rights Commission of Korea Sees a Sharp Increase in Complaints of 'Negligent Handling'... More Than Doubled in Two Years. <https://www.newspim.com/news/view/20201015000111> (Accessed on November. 12th, 2023)
- Park, H. J., Choi, B. G., Na, Y. W., and Kim, S. P. (2015). Pattern Analysis of Environmental Complaints Using Spatial Big Data In Incheon. *Ksce 2015 Convention 2015 Civil Expo & Conference*, 35-36.
- Park, J., and Kim, S. (2011). A Study on Psychological Factors in Establishing Living Space Based on Facet Theory, *Proceedings of the Korean Institute of Architects Conference*, 31(1), 17-18.
- Park, J., Choi, J., and Kim, S. (2010). A Study on Psychological Behavior at Earthquake Evacuation using the Facet Theory - Focusing on the MDS Analysis Technique. *Journal of the Architectural Institute of Korea*, 26(8), 71-78.
- Park, T. Y., Kim, S. J., and Oh, H. J. (2019). Empirical Verification of the Disaster Safety Information Facet Classification Scheme. *Journal of the Korean Society for Hazard Mitigation*, 19(3), 85-95.
- Pitsilis, G. K., Ramampiaro, H., and Langseth, H. (2018). Applied Effective Hate-speech Detection in Twitter Data using Recurrent Neural Networks. *Intelligence*, 4730-4742.
- Ranganathan, S. R. (1967). *Prolegomena to library classification*. Asia Publishing House (New York).
- Rosati, L., Lai, M. E., and Gnoli, C. (2004). Faceted classification for public administration. *In Semantic Web Applications and Perspectives (SWAP)-1st Italian Semantic Web Workshop*.
- Son, N. R., and Kim, S. Y. (2017). Complaints Statistics and Department of Automated Classifications System through Public Complaints Big Data Analysis. *Journal of Korean Institute of Next Generation Computing*, 13(1), 22-35.
- Yu, S. E., Hong, S. G., Lee, T. H., and Kim, N. R. (2018). A Pattern Analysis of Bus Civil Complaint in Busan City Using the Text Network Analysis. *Korean Computers and Accounting Review*, 16(2), 19-43
- Zhang, X., Zhao, J., and LeCun, Y. (2015). Character-level convolutional networks for text classification. *Advances in neural information processing systems*, 28.



**김 나 랑 (Na Rang Kim)**

- 정회원
- 부산대학교 문헌정보학과 문헌정보학사
- 동아대학교 경영정보학과 경영학석사
- 동아대학교 경영정보학과 경영

학박사

- (현재) 동아대학교 경영정보학과 조교수
- 관심분야: 에듀테크, 메타버스, 인공지능, 지역혁신, Co-creation, 경영정보시스템, SNA, 딥러닝