



일반논문 (Regular Paper)

방송공학회논문지 제29권 제3호, 2024년 5월 (JBE Vol.29, No.3, May 2024)

<https://doi.org/10.5909/JBE.2024.29.3.352>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

토큰라이저 기반 과테말라 세관 수입 물품의 가격 경향 분석

김 신^{a)}, 윤 경 로^{a)†}

Analysis of Price Trends in Guatemalan Customs Import Declarations based on Tokenizer

Shin Kim^{a)} and Kyoungro Yoon^{a)†}

요 약

관세는 수출입 또는 통과되는 화물에 대하여 부과하는 세금을 말한다. 개인이 온라인에서 해외 물품을 구매 대행하거나 직접 구매할 경우 특정 금액 이상일 때 수입하는 물건에 대해 관세를 지불해야 한다. 관세는 신고 물품의 신고 가격과 HS 코드에 따라 결정되는데, 신고 가격을 적게 신고하거나, 기본 세율이 낮은 HS 코드를 입력하는 등 여러 속임수를 사용하여 탈세할 위험성이 존재한다. 따라서, 신고 물품에 대한 가격이 적절한지, 즉 세금을 내지 않거나 적게 내기 위해 적은 금액을 신고하지 않았는지 확인할 필요가 있다. 본 논문에서는 과테말라 세관에 수입 신고한 물품에 대한 가격 경향 분석 방법을 제안하고자 한다. 세관 수입 신고 물품에 대하여 신고 물품의 단가를 산출하고 신고 단가에 대한 전체 구간에 대해 키워드를 추출하여 특정 HS 코드의 상위 키워드를 파악하고 해당 HS 코드와의 연관성을 파악한다. 세밀한 가격 경향 분석을 위하여, 신고 단가를 다구간으로 나누어 신고 단가 구간 별 토큰라이저를 사용하여 키워드를 추출, 특정 HS 코드의 여러 가격대 별 수입 물품에 대한 키워드 경향을 파악하였다. 본 논문에서 제안한 방법으로 추출된 키워드 및 가격 경향성을 분석한 결과, 전체 구간 키워드의 경우 특정 HS 코드의 대표적인 단어 또는 대표적인 회사명 등이 추출되며, 여러 구간별 키워드의 경우 가격대에 따른 추출된 키워드의 차이가 분명한 것을 확인하였다. 또한 분석 방법 제안을 통하여 이 가능하고 신고 가격 적절성을 판단할 수 있는 기준을 제시할 수 있는 것을 확인하였다.

Abstract

Customs duties are taxes on goods imported, exported, or transited through a country. When individuals purchase items online from abroad, directly or via agents, they must pay customs duties if the purchase exceeds a certain value. These duties are calculated based on the declared value and the Harmonized System (HS) code of the goods. Risks of tax evasion exist, including underreporting the declared value or using HS codes with lower basic tax rates. It's crucial to verify the accuracy of the declared values to prevent such evasion. This paper proposes a method to analyze price trends of goods declared at Guatemala customs. By calculating unit prices and extracting keywords for various price ranges, we identify key words associated with specific HS codes. Our analysis differentiates keywords by price range, providing insights into compliance with declared values. The method allows for a comprehensive overview of customs pricing trends and offers a benchmark for evaluating the appropriateness of declared prices, enhancing the accuracy of customs enforcement.

Keyword : Customs, Customs declaration analysis, Tokenizer, Keyword extraction

Copyright © 2024 Korean Institute of Broadcast and Media Engineers. All rights reserved.

“This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered.”

I. 서론

인터넷을 통해 해외 물품에 대한 구매 대행 뿐만 아니라 사용자가 해외 물품을 직접 구입하는 일은 이제 어렵지 않은 일이 되었다. 해외 물품을 온라인에서 구입하는 경우, 이는 수입에 해당되는 행위이므로 세관에 신고해야 하며, 면세 기준보다 이상인 가격의 물품을 구입한다면 수입 물품에 대한 세금을 지불해야 한다. 관세는 이러한 수입, 수출 또는 통과 되는 화물에 대해 부과하는 세금을 의미하며, 관세는 수입, 수출 또는 통과 화물의 HS 코드, 신고 물품의 금액 등 여러 변수에 따라 달라진다. 가령, 기본 세율이 낮은 HS 코드로 신고하거나, 신고 물품의 금액을 적게 기입한다면 납세자가 국가에 내야 할 세금을 적게 낼 수 있게 된다. 하지만, 이는 엄연히 탈세 행위, 즉 범죄 행위라 할 수 있다.

수입 신고한 물건을 직접 확인하고 가격을 파악하는 것으로 탈세 행위를 적발할 수 있지만, 일일이 모든 물건에 대한 적발 여부를 사람이 직접 판단하는 것은 매우 어려울 것이다. 가령, 대한민국 전자상거래 물품 수입통관현황 중 목록 통관 건수는 2022년 기준 약 64억 건으로^[1], 하루에 약 1천7백만 건의 전자상거래 물품이 수입되므로 모든 신고 물품에 대해 적발 행위를 직접 판단하는 것은 엄청난 시간과 비용이 요구되므로 불가능하다 할 수 있다. 따라서, 세관에 신고한 수입 신고 서면 자료를 토대로 수입 물품에 대한 신고 가격이 적절인지 판단할 수 밖에 없다.

키워드 기반의 빅데이터 분석^[2,3,4,5,6]은 감정 분석, ICT 트렌드, 특정 국가의 언론, 질한 연구 분석 등 여러 분야에서 연구되고 있으나, 세관 수입 신고에 대하여 키워드 기반으로 분석된 연구는 현재까지 미비하다. 따라서, 본 논문에서는 세관 빅데이터를 분석하여 세관 수입 신고 가격 판단에 도움을 줄 수 있는 신고 데이터 분석 방법을 제안하고자 하며, 신고를 제공하고 전반적인 세관 신고 가격 경향성을 HS 코드 별

로 상위 키워드를 추출하여 해당 HS 코드의 주요 수입 물품에 대한 키워드를 이해하고, 신고 가격을 다구간으로 나눈 후 키워드를 추출하여 가격대별 주요 키워드로 신고 가격의 적절성과 전반적인 가격 경향성을 파악하고자 한다.

본 논문의 순서는 다음과 같다. 제 2장에서는 관세 신고 가격과 관련된 배경지식에 대하여 서술하며, 제 3장에서는 관세 신고 가격 분석 방법에 대해 서술한다. 제 4장에서는 관세 신고 가격 분석 결과에 대해 설명하며, 마지막으로 제 5장에서 본 논문의 결론을 짓는다.

II. 배경 지식

본 장에서는 세관 신고 분석에 있어 필요한 배경 지식에 대하여 서술하고자 한다. 먼저, HS 코드에 대하여 설명한 후 세관 관점에서의 적발 개념과, 본 논문에서 키워드를 추출하기 위해 사용한 토큰라이저의 배경지식에 대해 설명하고자 한다.

1. HS 코드

HS(Harmonized Commodity Description and Coding System) 코드 또는 품목 코드는 WCO(World Customs Organization)에서 개발한 국제 상품 명명법으로, 무역 관련 통계 및 관세 수집을 용이하게 하는 것을 목적으로 한다. 전세계 공통으로 6자리의 HS 코드를 사용하고 있으며, 이후의 코드는 각 국가별로 별도로 정의하여 사용한다. HS 코드 6자리의 중 앞 2자리는 류(Chapter), 중간 2자리는 호(Heading), 뒤 2자리는 소호(Subheading)를 의미한다. 류는 상품이 속한 범주를 의미하며, 예를 들어 류의 코드가 09일 경우 커피, 향신료를 의미한다. 호는 해당 류에 속해 있는 범주 내 더 세분화된 분류 지표이다. 류가 09이면서 호가 05인 경우 향신료 중 바닐라를 의미하게 된다. 소호는 호보다 더 세분화된 분류를 나타낸다. 류와 호가 0905이면서, 소호는 10인 경우 부수지도, 그렇다고 잘게 다지지도 않은 바닐라를 의미한다. 또 다른 예로서, HS 코드 중 851713는 전자기기와 그 부분품, 녹음기, 음성 재생기, 텔레비전의 영상과 음성 기록기, 재생기와 이들이 부분품, 부

a) 건국대학교 스마트ICT융합공학과(Smart ICT Convergence Engineering Dept. Konkuk University)

‡ Corresponding Author : 윤경로(Kyoungro Yoon)

E-mail: yoonk@konkuk.ac.kr

Tel: +82-2-450-4219

ORCID: <https://orcid.org/0000-0002-1153-4038>

※ 본 논문은 과테말라 조세청 지능형 위험분석 및 BI(Business Intelligence) 기술 연구 과제의 지원 받아 작성하였음.

· Manuscript April 11, 2024; Revised May 8, 2024; Accepted May 8, 2024.

속품류(제85류)이며, 호 단위의 하위 분류는 셀룰러 통신망 용이나 그 밖의 무선통신망용의 스마트폰, 그 밖의 전화기와 같은 전화기(제8517호)이며, 소호 단위의 하위 분류로는 스마트폰이다. 즉, 851713은 스마트폰을 나타내는 HS 코드라 할 수 있다. HS 코드 분류는 관세에 있어 중요한 요소이다. 코드에 따라 기본세율이 달라질 수 있기 때문이다. 가령, 제 17류는 당류와 설탕 과자이며, 그 중 02호는 고체 상태 한정으로서 화학적으로 순수한 유당, 맥아당, 포도당 등 그 밖에 당류이거나 당 시럽, 인조 꿀, 캐러멜당을 의미한다. 소호 11은 유당을 의미하며, 소호 20은 단풍당과 단풍당 시럽을 의미한다. 하지만 유당의 기본세율은 20%, 단풍당의 기본세율은 8%로서 기본세율이 동일하지 않다. 즉, HS 코드의 소호에 따라 기본세율이 천차만별이며 그에 따라 신고 물품에 대한 HS 코드는 관세에 있어 매우 중요한 요소이다.

HS 코드 6자리는 국제 표준이며, 국가마다 세세부품목에 대하여 별도로 코드를 부여하여 관리할 수 있다. 대개 10자리에서 12자리로 HS 코드를 정의하여 관리하며, 대한민국의 경우 하위 4자리를 추가하여 10자리로 HS 코드를 관리하고 있다⁷⁾. 예를 들어, 세계 표준 HS 코드 851714는 셀룰러 통신망 용이나 그 밖의 무선 통신망용 전화기를 의미하며, 대한민국의 경우 하위 4자리 코드를 덧붙여서 관리한다. 하위 4자리 코드 1000, 즉 대한민국 HS 코드 8517141000은 그 밖의 전화기 중 코드 분할 다중접속을 사용하는 동기식 전화기를 의미하고, 8517142000은 그 밖의 전화기 중 시분할 다중접속을 사용하는 비동기식 전화기를 의미한다.

2. 적발

세관 관점에서 적발이라 하는 것은 수입 또는 수출하는 물품에 대하여 국가에 내야 할 세금을 의도적으로 적게 내려고 하다가 걸리는 행위를 의미한다. 수입하는 물건에 대하여 관세를 적게 또는 내지 않는 탈세 방법은 다음과 같은 방법이 존재한다. 첫 번째로 수입 물품에 대한 가격을 낮게 신고하는 것이다. 관세는 신고 금액에 대하여 일정 비율로 세금을 부과하기 때문에 가격을 낮게 신고한다면 그만큼 관세를 적게 낼 수 있다. 두번째 방법은 기본 세율이 낮은 HS 코드로 기입하는 것이다. 즉, 기본 세율이 낮다는 것은 기본으로 부과하는 세금이 낮다는 의미이므로 동일한 신고

금액이라 할지라도 부과되는 관세를 낮춤으로써 탈세한다. 본 논문에서는 첫 번째 방법을 방지하기 위한 가격 경향 분석 방법을 제안하고자 한다. 예를 들어, 애플 제품 중 아이폰 13은 미국 시장의 가격 기준으로 하였을 때 128GB는 599달러, 512GB는 미국 가격 기준 899달러이다. 즉, 같은 아이폰 13을 구매 대행하여 수입할 때 실제로는 512GB의 아이폰 13을 수입하지만 수입 신고할 때 128GB를 수입한다고 신고한다면 스마트폰에 대한 기본 세율 8%이므로 약 24달러 정도 절세를 할 수 있게 된다. 따라서, 수입 신고한 물품과 신고 가격이 합당한 범위 내인지 파악하기 위해서는 세관 신고 수입 물품에 대한 가격대별 키워드 분석과 같은 보조적인 분석 기법이 필요하다.

3. 토크나이저

본 논문에서 제안하는 분석 방법 중 키워드 추출 및 분석은 수입 신고 물품의 설명 칸에 기입된 문자열을 토대로 수행한다. 토크나이저는 하나의 문자열 데이터를 여러 개의 문자로 분할하는 도구로, 키워드를 추출할 때 필요하며, 다양한 방식의 토크나이저가 존재한다.

정규표현식 토크나이저⁸⁾는 정규표현식 기반으로 문자열을 토큰으로 나누는 도구다. 정규표현식은 특정 규칙과 기호를 사용하여 패턴을 정의하고 문자열에서 특정 패턴을 찾기 위한 토큰화 처리 방식을 의미한다. 따라서, 정규표현식 토크나이저를 사용한다면, 사용자가 토큰화 하고자 하는 패턴을 직접 정의할 수 있으며, 가장 단순한 토크나이저 중 하나이다. 하지만, 정규표현식 토크나이저는 특정 패턴에 의해 추출되므로 언어학적인 관점에서 패턴을 정의하고 문자열을 토큰화 하는 것은 사실상 매우 어려운 작업이라 할 수 있다.

톡톡(Toktok) 토크나이저⁹⁾는 구두점과 단어를 효과적으로 분리할 수 있는 토큰화 도구로서, 경량 토크나이저이므로 대량의 텍스트를 빠르게 처리할 수 있다. 토크톡 토크나이저는 프랑스어, 독일어 등 다양한 언어를 지원하며, 사용자 입장에서 사용하기 간편하다는 장점이 있다.

스페이스(SpaCy) 토크나이저¹⁰⁾는 자연어 처리를 위한 오픈 소스 라이브러리 기반의 고성능 토크나이저로서, 문장 구조, 어휘, 문맥 등 언어학적인 특징을 고려하여 토큰화할 수 있다. 최적화된 알고리즘을 토대로 빠른 속도 처리를

보여주며, 다양한 언어에 대해 학습하여 여러 언어에 대하여 특성을 이해하고 토큰화를 수행할 수 있다.

본 논문에서는 세관 수입 신고 물품에 대한 신고 가격 및 키워드 분석을 위해 정규표현식 토큰라이저, 토크 토큰라이저, 스페이시 토큰라이저 총 세 가지 토큰라이저를 사용하였고, 전 구간 및 다 구간에서 토큰라이저로 키워드를 추출하여 HS 코드 및 가격 구간별 주요 키워드를 비교하였다. 추가적으로, 토큰라이저에 따른 키워드 추출 성능 차이 여부를 분석하였다.

III. 제안 분석 방법

세관 신고 물품의 탈세 행위를 간파하고 올바른 신고를 유도하기 위해서는 먼저 신고 물품에 대한 신고 가격의 적절성을 판단해야 한다. 다시 말해, 수입하려고 하는 자가 신고한 물품의 가격이 타 시장 가격 대비 적절한 가격으로 신고되었는지 파악하여 탈세의 위험성을 간파해야 한다. 따라서, 본 논문에서는 신고 물품에 대한 신고 가격의 단가를 산출하고, 전체 구간 및 다구간별 키워드를 추출하여 가격대에 따른 키워드 차이 분석을 수행하였다.

본 논문에서 제안하는 신고 가격 분석 방법은 다음과 같다. 먼저, 신고 물품에 대한 신고 가격의 단가를 산출한다. 만약, 기업이 동일한 물품에 대해 여러 개를 수입하는 경우 세관에는 수입 물품에 대한 총액으로 신고하게 된다. 따라서, 동일한 물품에 대해 여러 개를 수입하는 경우를 고려하여 단가 산출 과정이 필요하다. 신고 가격의 단가는 관세에 물품의 포장 수와 물품 단위수로 나눠 계산하였으나, 0으로 기입된 데이터의 경우 나누지 않는 예외 처리를 추가하였다. 이후 신고 단가에 관한 히스토그램과 커널 밀도 추정을 산출하였고, 신고 물품 단가에 대한 전체 구간과 다구간에서 상위 키워드를 추출하였다. 신고 물품 단가의 구간은 신고 단가의 전체 구간 및 일정 간격으로 나눈 여러 구간으로 나뉘었으며, 일정 간격은 신고 수의 1할을 기준으로 하였다. 하지만, 특정 신고 건수가 특정 수 미만일 경우 신고 건수별로 구간을 설정하였으며, 기준 신고 건수는 100건으로 설정하였다. 신고 단가에 대해 구간을 나눈 후 토큰라이저를 기반으로 전 구간과 다 구간 내 키워드와 그에 대한 빈도수를 추출하였다.

본 논문에서 제안하는 신고 물품의 단가 경향 파악 및 구간별 키워드 분석 방법을 위해 과테말라 조세청에서 제공한 데이터 중 2022년 완전 수입 데이터를 토대로 분석하였다. 과테말라는 스페인어를 사용하는 국가이므로 세관 신고 물품에 대한 설명은 스페인어로 작성된다. 스페인어 기반의 문장에서 키워드를 추출하기 위하여 정규표현식 토큰라이저와 스페인어 기반 토큰화가 가능한 토크 토큰라이저, 스페이시 토큰라이저를 사용하였다.

본 논문에서 제안하는 신고 물품의 단가 및 키워드 분석 방법에 대한 개발 환경은 표 1과 같다.

표 1. 세관 신고 가격 경향 분석 환경

Table 1. Environment for customs declaration price trend analysis

Framework or Library	Version
Python	3.10
Pandas	2.1.3
NLTK	3.8.1
Spacy	3.7.4

신고 물품의 단가 및 키워드 기반 가격 분석을 위해 판다스 라이브러리 2.1.3 버전을 사용하였고, 정규표현식 토큰라이저, 토크 토큰라이저를 내포하고 있는 NLTK 라이브러리 3.8.1 버전, 스페이시 토큰라이저 3.7.4 버전을 사용하였다.

IV. 세관 신고 물품의 신고 가격 경향 분석

본 절에서는 제안하는 방법 기반으로 수입 물품의 신고 가격 경향 분석 결과를 보여준다.

데이터 분석을 위하여 HS 코드 별 신고 건에 대한 신고 가격 단가의 히스토그램 및 커널 밀도 추정을 산출하였으며, 그림 1과 그림 2는 HS 코드 중 8517130000에 대한 신고 단가의 히스토그램과 커널 밀도 추정을 보여준다. 세계 HS 코드 중 851713은 스마트폰을 의미하며, 과테말라 조세청에서는 세계 HS 코드 뒤에 0000을 붙인 후 ‘Telefonos inteligentes’(번역하면 ‘스마트 폰’)으로 신고 물품 품목을 정의하고 수입 물품에 대해 관리하고 있다. 그림 1과 그림 2에서 X축은 단위가 과테말라 화폐 단위인 께찰 단위의 신고 가격의 단가, Y축은 신고 건수를 의미한다.

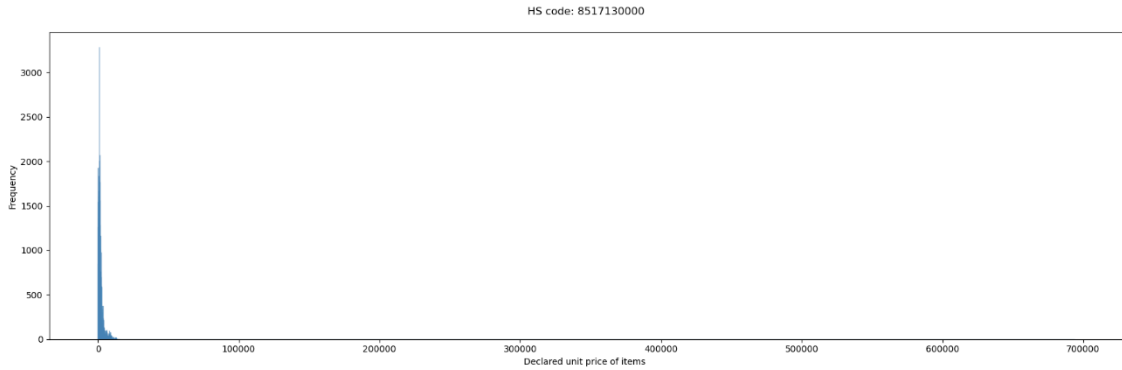


그림 1. HS 코드 8517130000(스마트 폰)의 신고 물품의 단가 기반 히스토그램
 Fig. 1. Histogram based on unit price of declared goods for HS Code 8517130000 (Smartphones)

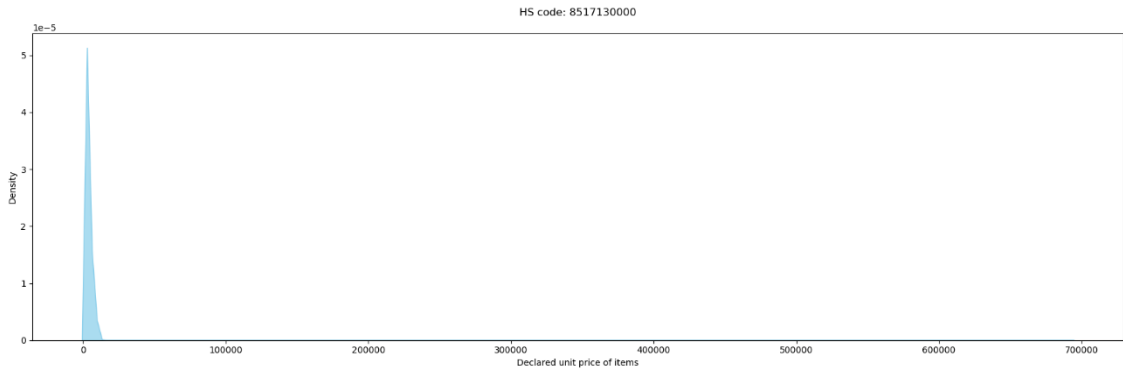


그림 2. HS 코드 8517130000(스마트 폰)의 신고 물품의 단가 기반 커널 밀도 추정
 Fig. 2. Kernel density estimation based on unit price of declared goods for HS Code 8517130000 (Smartphones)

히스토그램 및 커널 밀도 추정 산출 결과 10000계찰(한화 약 170만원) 미만의 상품이 주로 수입되는 것을 파악할 수 있다. 하지만, X 축의 범위가 700000계찰(한화 약 1억

1936만원)까지다. 단가를 산출할 때 포장 개수와 단위 개수로 나누어 산출한다. 여기서 포장 개수와 단위 개수는 물건을 포장하였을 때의 포장 개수와 포장 내 물품 단위 개수를

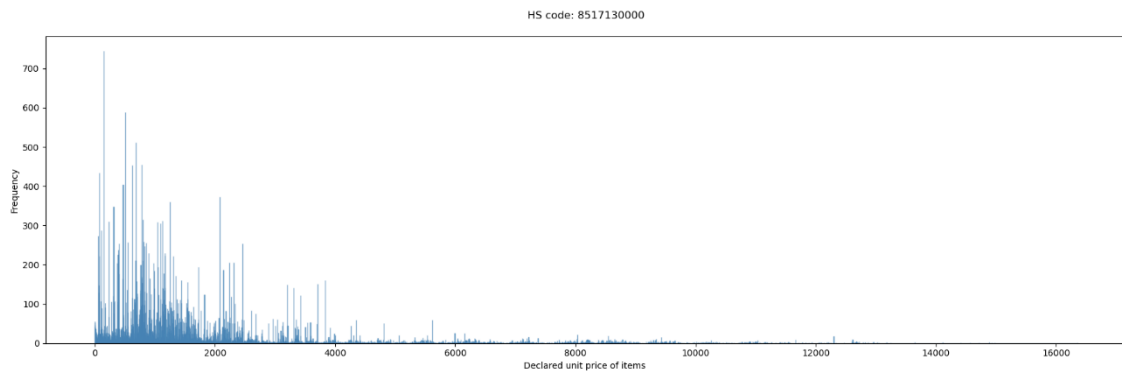


그림 3. HS 코드 8517130000(스마트 폰)의 신고 물품의 단가 기반 히스토그램(이상치 제거)
 Fig. 3. Histogram based on unit price of declared goods for HS Code 8517130000 (Smartphones), with outlier removal

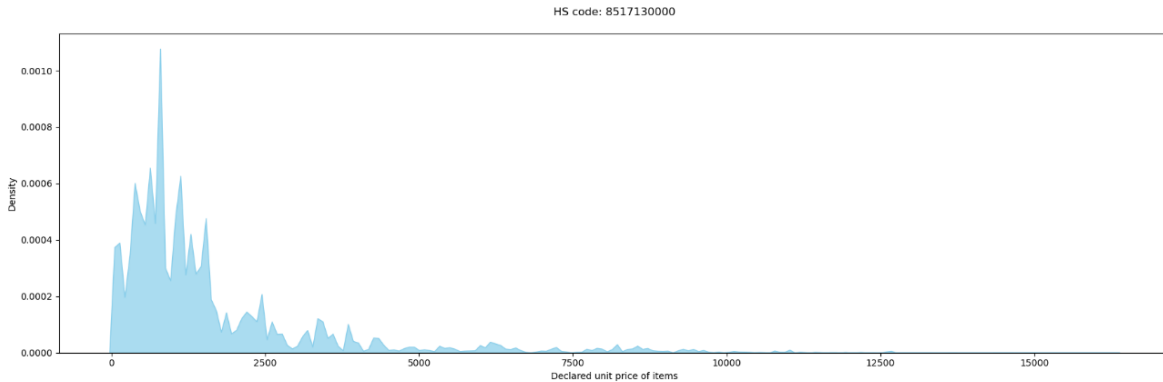


그림 4. HS 코드 8517130000(스마트 폰)의 신고 물품의 단가 기반 커널 밀도 추정(이상치 제거)
 Fig. 4. Kernel density estimation based on unit price of declared goods for HS Code 8517130000 (Smartphones), with outlier removal

의미한다. 단위 단가 산출 이후 데이터를 확인한 결과 694004.87개찰(한화 약 1억 1833만원)에 신고된 건수가 1건이 확인돼, 포장 개수 및 단위 개수가 모두 0으로 들어가 있어 발생된 것으로 추정된다. 따라서 이상치를 제거한 범위에 대하여 히스토그램과 커널 밀도 추정을 다시 산출하였으며, 그 결과는 그림 3, 그림 4와 같다.

그림 3과 그림 4를 통해 알 수 있듯이, HS 코드 8517130000에 대한 주된 신고 수입 물품에 대한 단가 범위는 0개찰 초과 4000개찰(한화 약 67만원) 미만이다. 4000개찰 이상의 금액대는 4000개찰 미만의 금액대에 비해 수입 신고 수가 적은 것을 알 수 있다.

신고 물품의 가격대별 가격 경향성 상세 분석을 위하여 HS 코드 8517130000에 대해 키워드 추출을 수행하였다. 먼저, 신고 가격대의 전 구간에 대해 세 가지 토큰라이저를 사용하여 키워드를 추출하였으며, 표 2는 8517130000(스마트폰)에 대한 세 가지 토큰라이저 기반 상위 키워드 추출 및 해당 키워드에 대한 빈도수를 보여준다.

정규표현식 토큰라이저 기반 상위 10개의 키워드는 CELULAR(휴대전화), TELEFONO(전화기), SAMSUNG, TELEFONOS(전화기의 복수형 단어), SM, BLACK, CELULARES(휴대전화의 복수형 단어), 128GB, BLUE, GALAXY이며, 빈도수는 단어 순서대로 31562번, 24310번, 14353번, 9120번, 8596번, 8223번, 8061번, 7256번, 7047번, 6930번이다.

특목 토큰라이저 기반 상위 키워드는 CELULAR(휴대전화), TELEFONO, SAMSUNG, TELEFONOS, CELULARES, BLACK, GALAXY, BLUE, SIM, DUAL이며, 빈도수는 단어 순으로 31537번, 24257번, 14338번, 9082번, 8028번, 7563번, 6930번, 6611번, 6274번, 5742번이다.

스페이스 토큰라이저 기반 상위 키워드는 CELULAR, TELEFONO, GB, SAMSUNG, TELEFONOS, CELULARES, BLACK, GALAXY, BLUE, SIM이며, 빈도수는 단어순으로 31542번, 24257번, 21573번, 14338번, 9210번, 8024번, 8000번, 6930번, 6896번, 6275번이다.

표 2. HS 코드 8517130000의 토큰라이저 종류에 따른 신고 가격 단가 전체 구간의 상위 10개 키워드 및 빈도수
 Table 2. Top 10 keywords and frequencies for HS Code 8517130000, according to type of tokenizer, across all price ranges of declared unit prices

Tokenizer type	Top 10 keywords(Frequency)
Regular expression tokenizer	CELULAR(31562), TELEFONO(24310), SAMSUNG(14353), TELEFONOS(9120), SM(8596), BLACK(8223), CELULARES(8061), 128GB(7256), BLUE(7047), GALAXY(6930)
Toktok tokenizer	CELULAR(31537), TELEFONO(24257), SAMSUNG(14338), TELEFONOS(9082), CELULARES(8028), BLACK(7563), GALAXY(6930), BLUE(6611), SIM(6274), DUAL(5742)
Spacy Tokenizer	CELULAR(31542), TELEFONO(24257), GB(21573), SAMSUNG(14338), TELEFONOS(9120), CELULARES(8024), BLACK(8000), GALAXY(6930), BLUE(6896), SIM(6275)

표 3. HS 코드 8517130000에 대한 토크나이저 별 특정 빈에 대한 키워드 산출 결과

Table 3. Results of keyword extraction for a specific bin according to tokenizer for HS Code 8517130000

Bin	Min value	Max value	Top 20 keywords and their frequency by Regular expression tokenizer	Top 20 keywords and their frequency by toktok tokenizer	Top 20 keywords and their frequency by spacy tokenizer
6	804.68 GTQ (137,333 KRW)	938.39 GTQ (160,153 KRW)	64GB(1329), DS(972), 32GB(661), LTE(482), A13(468), A03S(447), SIM(444), DUAL(439), A23(335), ST(255), A03(226), REDMI(218), XIAOMI(215), IPHONE(211), EU(209), US(209), WHITE(208), GRAY(179), MOTOROLA(163), PLUS(150)	64GB(906), DS(862), LTE(479), SM(474), A13(468), A03S(447), A23(335), ST(255), A03(226), REDMI(218), XIAOMI(215), IPHONE(211), US(209), WHITE(208), 128GB(204), GRAY(178), 32GB(177), EU(177), MOTOROLA(163), PLUS(150)	DS(947), LTE(479), SM(474), A13(468), A03S(447), DUAL(439), A23(335), ST(255), A03(226), REDMI(218), XIAOMI(215), IPHONE(211), US(209), WHITE(207), GRAY(178), EU(176), MOTOROLA(163), PLUS(150), INTELIGENTES(135), REALME(129)
16	2146.15 GTQ (366,280 KRW)	2279.05 GTQ (388,962 KRW)	LITE(456), XIAOMI(445), SIM(391), DUAL(388), 5G(377), EU(279), XIA(209), EUR(202), DS(201), PINK(201), INTELIGENTE(163, Intelligent in Spanish), A53(145), WHITE(139), LTE(126), X4(119), 256GB(115), SPEC(105), PRO(97), A525M(91), 11T(91)	LITE(456), XIAOMI(445), 5G(377), EU(279), 128GB(242), XIA(207), EUR(202), PINK(201), DS(183), INTELIGENTE(163, Intelligent in Spanish), A53(145), WHITE(138), LTE(126), X4(119), SPEC(105), PRO(96), 11T(91), US(83), A525M(80), Pro(79)	LITE(456), XIAOMI(445), DUAL(388), G(377), EU(279), XIA(207), EUR(202), DS(201), PINK(201), INTELIGENTE(163, Intelligent in Spanish), A53(145), WHITE(138), LTE(126), X4(119), SPEC(105), PRO(96), T(92), A525(88), US(83), M(80)
26	3488.40 GTQ (595,360 KRW)	3619.98 GTQ (617,816 KRW)	XIAOMI(172), IPHONE(92), 2107113SG(90), PRO(73), 256GB(58), INTELIGENTES(53), WHITE(46), 11T(43), EU(42), 8GB(42), GREY(41), METEORITHE(40, Miswritten of Meteorite), MOONLIGHT(40), APPLE(39), GRADE(35), USED(29), 64GB(27), AB(24), COLOR(20), GB(11)	XIAOMI(172), IPHONE(92), 2107113SG(90), PRO(73), INTELIGENTES(53), 128GB(46), WHITE(46), 11T(43), EU(42), GREY(41), METEORITHE(40, Miswritten of Meteorite), MOONLIGHT(40), APPLE(39), GRADE(35), USED(29), 64GB(27), AB(24), COLOR(20), 256GB(12), GB(11)	XIAOMI(172), IPHONE(92), 2107113SG(90), PRO(73), INTELIGENTES(53), WHITE(46), T(43), EU(42), GREY(41), METEORITHE(40, Miswritten of Meteorite), MOONLIGHT(40), APPLE(39), GRADE(35), USED(29), AB(24), COLOR(20), B(11), MEDIA(9), NOCHE(9), Pro(8)
36	4829.52 GTQ (824,247 KRW)	4953.05 GTQ (845,330 KRW)	256GB(37), IPHONE(32), COLOR(22), S21(18), FE(18), S22(17), NEGRO(16, Black in Spanish), ULTRA(13), 64GB(13), CEL(11), LIB(11), PRO(9), LTE(9), GB(8), SAMSM(7), DS(7), BLANCO(7, White in Spanish), MAX(7), Samsung(7), Galaxy(7)	IPHONE(32), 256GB(23), COLOR(22), S21(18), FE(18), S22(16), NEGRO(16, Black in Spanish), 128GB(13), ULTRA(13), 64GB(13), CEL(11), LIB(11), PRO(9), LTE(9), GB(8), DS(7), BLANCO(7), MAX(7), Samsung(7), Galaxy(7)	IPHONE(32), COLOR(22), S21(18), FE(18), S22(17), NEGRO(16, Black in Spanish), ULTRA(13), CEL(11), LIB(11), G(10), PRO(9), LTE(9), DS(7), BLANCO(7, White in Spanish), MAX(7), Samsung(7), Galaxy(7), Ultra(7), XIAOMI(6), US(6)
46	6174.40 GTQ (1,053,776KRW)	6299.43 GTQ (1,075,114 KRW)	IPHONE(56), APPLE(25), LAE(18), MIDNIGHT(15), SIM(15), FREE(15), BES(10), CE001APL70(9), FLIP(8), STARLIGHT(6), RED(6), N985FZKKGTO(4), SAMSM(4), F711(4), GB(4), CEL(4), LIB(4), Z(4), GREEN(4), MOBILE(3)	IPHONE(56), APPLE(25), 128GB(24), MIDNIGHT(15), CE001APL70(9), FLIP(8), STARLIGHT(6), RED(6), GB(4), CEL(4), LIB(4), Z(4), GREEN(4), MOBILE(3), GTO(3), CE000APL99(3), Apple(3), iPhone(3), Pro(3), Max(3)	IPHONE(56), APPLE(25), LAE(18), MIDNIGHT(15), SIM(15), FREE(15), BES(10), CE001APL70(9), FLIP(8), STARLIGHT(6), RED(6), N985FZKKGTO(4), SAMSM(4), F711(4), GB(4), CEL(4), LIB(4), Z(4), GREEN(4), MOBILE(3)
56	7514.52 GTQ (1,282,492 KRW)	7642.59 GTQ (1,304,349 KRW)	5G(6), S22(5), MUESTRAS(3, Samples in Spanish), SET(3), PHONE(2), NOTTINGHAM(2), 256GB(2), GREEN(2), ULTRA(2), Celular(2), Samsung(2), Galaxy(2), Ultra(2), 256gb(2), ABAA(1), 5109ABAC(1), APPLE(1), IPAD(1), PRO(1), M1(1)	5G(6), S22(4), MUESTRAS(3, Samples in Spanish), SET(3), PHONE(2), GREEN(2), 128GB(2), ULTRA(2), Celular(2), Samsung(2), Galaxy(2), Ultra(2), 256gb(2), ABAA(1), 5109ABAC(1), APPLE(1), IPAD(1), PRO(1), M1(1), CHIP(1)	G(6), S22(5), MUESTRAS(3, Samples in Spanish), PHONE(2), GREEN(2), ULTRA(2), Celular(2), Samsung(2), Galaxy(2), Ultra(2), gb(2), ABAA(1), 5109ABAC(1), APPLE(1), IPAD(1), PRO(1), M1(1), CHIP(1), MID(1), LTE(1)

토큰라이저를 사용하여 신고 물품에 대한 상위 키워드를 추출한 결과, 토큰라이저에 상관없이 추출 키워드의 차이는 대동소이한 것을 알 수 있다. 정규표현식 토큰라이저를 사용하여 분석한 경우, 불분명한 의미의 ‘SM’이라는 키워드가 포함되어 있는데, 이는 삼성 갤럭시 모델명이 ‘SM’으로 시작하기 때문이다. 그 밖은 키워드로 이해할 만한 것이 추출되었다. 스페이시 토큰라이저의 경우 ‘GB’ 키워드가 추출되었는데, ‘128GB’와 같이 용량을 나타내는 복합 키워드를 ‘128’, ‘GB’로 분할하는 것으로 파악된다.

표 3은 정규표현식 토큰라이저, 토큰 토큰라이저, 스페이시 토큰라이저 기반 일부 구간별 단가의 최솟값, 최댓값, 추출된 키워드 및 그에 대한 빈도수를 보여준다. 구간별 키워드는 전 구간에서 추출된 키워드를 제외하고 빈도수가 높은 상위 20개를 보여준다. 일부 구간은 특정 크기의 간격으로 추출하였으며, 토큰라이저 간 비교가 용이하기 위해 이상치를 제외한 구간 중 매우 극 일부 구간만을 추렸다. 표 3에서는 일부 구간 중 6개의 빈을 확인한 결과, HS 코드 8517130000, 스마트폰에 해당하는 수입 물품은 대개 스페인어가 아닌 영어로 입력되어 대부분의 키워드가 영어로 추출된 것을 파악할 수 있다. XIAOMI와 같은 회사명 또는 IPHONE과 같은 특정 제품명이 주로 상위 키워드에서 추출되며, 그 밖에 스마트폰의 색상, 용량, 모델번호와 같은 키워드도 주로 추출되는 것을 알 수 있다. 색상과 관련된 키워드는 주로 영어로 표현하고 있지만, 드물게 스페인어로 표현되고 있는 것을 파악할 수 있다. 가령, 전 구간 키워드인 BLACK, BLUE 뿐만 아니라 WHITE, GREEN 등의

색상을 여러 구간에서 확인할 수 있으며, BLANCO(흰색), VERDE(녹색) 또한 여러 구간에서 나타나는 것을 알 수 있다. 또한, 낮은 수의 빈일수록 상대적으로 중저가의 브랜드 또는 저 사양 모델의 스마트폰이 주된 키워드인 반면, 높은 수의 빈일수록 고가의 브랜드 또는 고사양 모델의 스마트폰과 관련된 키워드가 추출되는 것을 알 수 있다. 따라서, 고가의 브랜드나 고사양의 스마트폰을 저가로 신고할 경우, 분석된 가격 경향을 통해 저가 신고의 가능성을 쉽게 파악할 수 있게 된다.

정규표현식 토큰라이저 또는 토큰 토큰라이저를 사용하여 구간별 키워드를 추출하였을 때, 구간 별 스마트 폰 사양 중 용량에 대한 키워드 차이가 구간별로 있는 것을 확인할 수 있다. 하지만 스페이시 토큰라이저는 앞서 언급하였던 것처럼 ‘128GB’와 같이 용량을 표기하였을 때 ‘128’, ‘GB’로 세부분할을 수행하기 때문에 산출된 키워드 기반으로 용량을 파악할 수 없다. 또한, 세 토큰라이저 기반으로 추출한 키워드를 분석한 결과, 토큰라이저 별 키워드 추출 결과가 대동소이한 것으로 토큰라이저에 따른 키워드 추출 성능 차이가 거의 나지 않는 것을 표를 통해 확인할 수 있다.

다른 분석 결과를 확인하기 위해 두번째 HS 코드로 8703226991를 선택하였다. 해당 코드는 자동차에 해당되는 HS 코드로, 870322는 실린더 용량이 1,000CC 초과 1,500CC 이하인 것을 의미하는 코드이며, 최하위 4자리의 6991은 ‘Con capacidad de transporte inferior o igual a 5 personas, incluido el conductor’, 번역하면 ‘운전자 포함 5명 이하의 수송 능력을 가진 차량에 대한 코드’를 의미한다.

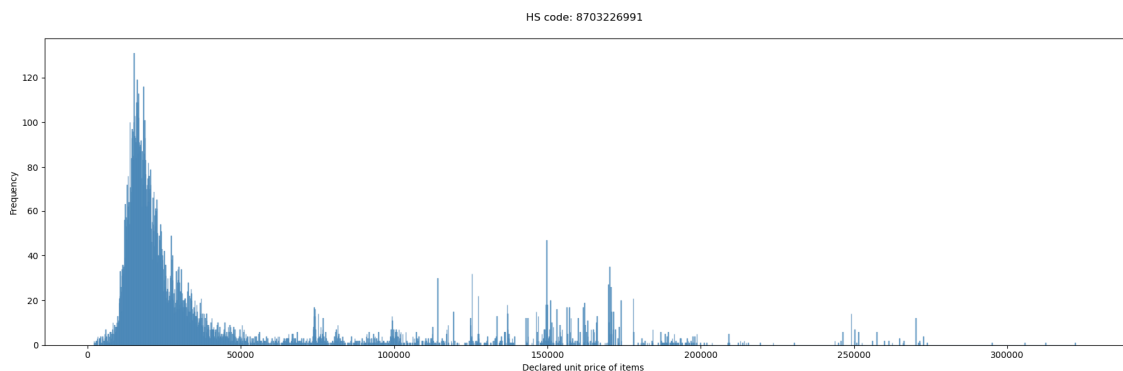


그림 5. HS 코드 8703226991(5인승 자동차)의 신고 물품의 단가 기반 히스토그램
 Fig. 5. Histogram based on the unit price of declared goods for HS Code 8703226991 (5-Seater automobiles)

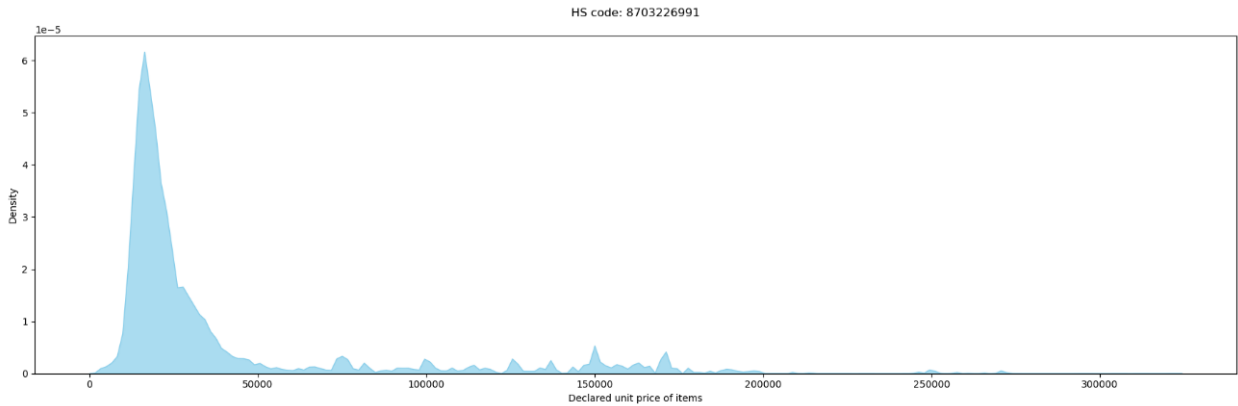


그림 6. HS 코드 8703226991(5인승 자동차)의 신고 물품의 단가 기반 커널 밀도 추정
 Fig. 6. Kernel density estimation based on the unit price of declared goods for HS Code 8703226991 (5-Seater automobiles)

그림 5, 그림 6은 원 데이터에 대한 히스토그램 및 커널 밀도 추정 산출 결과이다.

신고 가격 단가의 최댓값이 약 350,000(한화 5907만 원)인 것으로 보아 단가 산출에 있어서 문제가 발생하지 않은 것으로 판단하고, 이상치는 별도로 제거하지 않았다. 표 4는 세 토큰라이저 별 전 구간의 상위 키워드 10개를 보여준다.

HS 코드 8703226991에 대하여 신고 단가의 전 구간에서 추출한 키워드는 완벽히 동일하지만, 키워드에 대한 빈도가 매우 조금의 차이가 나는 것을 확인할 수 있다. 주요 키워드는 AUTOMOVIL(자동차의 스페인어), USADO(중고의 스페인어), VIN(Vehicle identification number, 차대번호), TOYOTA, CAMIONETA(SUV의 스페인어), HONDA, MODELO(모델의 스페인어), YARIS, SPORT, FIT 순이다. TOYOTA, HONDA 회사명의 자동차가 주로 수입되는 것을 알 수 있으며, 중고 SUV가 주로 수입되는

차종으로 볼 수 있고, 그 밖에 YARIS, SPORT, FIT과 같은 소형차도 주로 수입되는 차의 모델이라 할 수 있다. 표 5는 동일한 HS 코드에 대한 토큰라이저 별 특정 빈에서 추출한 키워드와 그의 빈도수를 보여준다. 가격 구간의 빈의 선정 기준은 58번째 빈을 기준으로 하여 200개의 빈을 간격으로 하였으며, 해당 빈에 추출된 값이 없을 경우 제일 가까운 빈을 기준으로 하여 추출 결과를 비교하였다.

키워드 분석 결과 58번째 빈을 제외한 모든 빈에서 추출된 키워드가 토큰라이저에 상관없이 동일한 것을 파악하였다. 낮은 순서의 빈의 경우 USADA와 같은 중고에 대한 키워드 M2005, M2009와 같은 오래된 연식을 나타내는 키워드가 추출되었으며, 높은 단가를 포함하고 있는 빈일수록 VOLKSWAGEN, AUDI, BMW 순으로 고가의 자동차 차량을 파는 회사명이 나오는 것을 확인하였고, 탈세 목적으로 저가 신고를 하는 경우 탈세 의심을 할 수 있을 것으로 보인다.

표 4. HS 코드 8703226991의 토큰라이저 종류에 따른 신고 가격 단가 전체 구간의 상위 10개 키워드 및 빈도수
 Table 4. Top 10 Keywords and frequencies for HS Code 8703226991, according to type of tokenizer, across all price ranges of declared unit prices

Tokenizer type	Top 10 keywords(Frequency)
Regular expression tokenizer	AUTOMOVIL(9419), USADO(4497), VIN(3874), TOYOTA(3527), CAMIONETA(2456), HONDA(1986), MODELO(1812), YARIS(1683), SPORT(1412), FIT(921)
Toktok tokenizer	AUTOMOVIL(9419), USADO(4497), TOYOTA(3527), VIN(3176), CAMIONETA(2456), HONDA(1986), MODELO(1812), YARIS(1683), SPORT(1412), FIT(921)
Spacy Tokenizer	AUTOMOVIL(9419), USADO(4497), VIN(3831), TOYOTA(3527), CAMIONETA(2456), HONDA(1986), MODELO(1811), YARIS(1683), SPORT(1412), FIT(921)

표 5. HS 코드 8703226991에 대한 토큰라이저 별 특정 빈에 대한 키워드 산출 결과

Table 5. Results of keyword extraction for a specific bin according to type of tokenizer for HS Code 8703226991

Bin	Min value	Max value	Top 20 keywords and their frequency by Regular expression tokenizer	Top 20 keywords and their frequency by toktok tokenizer	Top 20 keywords and their frequency by spacy tokenizer
58	16713.85 GTQ (2,819,445KRW)	16964.17 GTQ (2,863,385 KRW)	MAZDA(26), SCION(25), CAMIONETILLA(24, SUV in Spanish), XB(11), USADA(11, Used in Spanish), Usadas(11, Used in Spanish), MAZDA2(10), CHEVROLET(6), M2007(6), GRIS(6, Grey in Spanish), S(6), POLICROMADO(5, Polychrome in Spanish), M2005(5), M2009(5), SERIE(5), POL(5), AÑO(4, Year in Spanish), XA(4), TOURING(3), NEGRO(3, black in Spanish)	MAZDA(26), SCION(25), CAMIONETILLA(24, SUV in Spanish), XB(11), USADA(11, Used in Spanish), Usadas(11, Used in Spanish), MAZDA2(10), CHEVROLET(6), M2007(6), GRIS(6, Grey in Spanish), POLICROMADO(5, Polychrome in Spanish), M2005(5), M2009(5), SERIE(5), POL(5), AÑO(4, Year in Spanish), XA(4), TOURING(3), NEGRO(3), M2012(3)	MAZDA(26), SCION(25), CAMIONETILLA(24, SUV in Spanish), XB(11), USADA(11, Used in Spanish), Usadas(11, Used in Spanish), MAZDA2(10), CHEVROLET(6), M2007(6), GRIS(6, Grey in Spanish), S(6), POLICROMADO(5, Polychrome in Spanish), M2005(5), M2009(5), SERIE(5), POL(5), AÑO(4, Year in Spanish), XA(4), TOURING(3), NEGRO(3)
258	41939.86 GTQ (7,078,891 KRW)	42147.64 GTQ (7,114,107 KRW)	BALENO(2)	BALENO(2)	BALENO(2)
458	117590.15 GTQ (19,848,061 KRW)	117772.02 GTQ (19,878,759 KRW)	VEHICULO(10, Vehicle in Spanish), VOLKSWAGEN(10), JETTA(10)	VEHICULO(10, Vehicle in Spanish), VOLKSWAGEN(10), JETTA(10)	VEHICULO(10, Vehicle in Spanish), VOLKSWAGEN(10), JETTA(10)
655	167233.93 GTQ (28,227,444KRW)	167233.93 GTQ (28,227,444 KRW)	AUDI(1), Q2(1), 4X2(1), SERIE(1), GAGBCW(1)	AUDI(1), Q2(1), 4X2(1), SERIE(1), GAGBCW(1)	AUDI(1), Q2(1), 4X2(1), SERIE(1), GAGBCW(1)
862	219510.14 GTQ (37,051,154 KRW)	219510.14 GTQ (37,051,154 KRW)	BMW(1), 118i(1)	BMW(1), 118i(1)	BMW(1), 118i(1)
1063	270175.2 GTQ (45,602,918 KRW)	270175.2 GTQ (45,602,918 KRW)	X1(12), sDrive(12), 18i(12), MARCA(12, make in Spanish), BMW(12), M(12)	X1(12), sDrive(12), 18i(12), MARCA(12, make in Spanish), BMW(12), M(12)	X1(12), sDrive(12), 18i(12), MARCA(12, make in Spanish), BMW(12), M(12)

V. 결론

본 논문에서는 세관 수입 신고 물품에 대한 신고 가격의 단가를 산출하고, 전체 구간 및 다구간에서 키워드를 추출하여 구간별 키워드 비교 분석 방법에 대해 제안하였다. 과테말라 조세청의 수출입 신고 데이터 중 2022년 완전 수입 데이터를 기반으로 분석을 수행하였으며, 특정 HS 코드의 전 구간 및 특정 구간에서 추출한 키워드와 그의 빈도수를 직접 비교분석 하였다. 키워드 추출을 위해 정규표현식 토큰라이저, 토크 토큰라이저, 스페이시 토큰라이저를 사용하였으며, 추가적으로 토큰라이저 별 추출 결과를 비교 분석하였다.

스마트폰에 해당되는 HS 코드 8517130000에 대한 가격 경향 분석 결과, 낮은 가격대의 빈에서는 비교적 저가의 회사명, 저 사양, 중고 등에 대한 키워드가 주로 추출되었으며, 높은 가격대의 빈에서는 상대적으로 고가의 브랜드, 고

사양과 관련된 키워드가 추출되었다. 토큰라이저 별 추출 결과를 비교하자면, 스페이시 토큰라이저의 경우 ‘256GB’와 같은 용량 키워드에서 ‘256’, ‘GB’와 같은 형태로 토큰화 하나, 정규표현식 토큰라이저와 토크 토큰라이저는 스페이시 토큰라이저와 달리 하나의 토큰으로 인식하는 특징이 있다. 하지만, 토큰라이저 별 키워드 추출 결과는 대동소이 성능 차이가 거의 없는 것을 확인하였다.

운전자를 포함하여 최대 5인을 수송할 수 있는 자동차에 해당되는 HS 코드 8703226991의 경향 분석 결과, 낮은 가격대의 빈에서는 연식이 오래된 중고 차량과 관련된 키워드가 주로 추출되었으며, 높은 가격대의 빈에서는 VOLKSWAGEN, AUDI, BMW와 같은 고가의 차량을 파는 회사명 및 고가의 차량 모델명이 추출되었다. 또한, 토큰라이저 별 추출되는 키워드에 있어 거의 차이가 없었음을 확인하였다.

본 논문에서 제안하는 방법은 분석 결과 세관 수입 신고 물품명과 신고 가격에 대한 경향성을 파악할 수 있는 것을

확인하였으며, 가격 경향성 분석 결과는 고가의 브랜드가 저가로 신고될 경우 탈세의 의심을 할 수 있는 기준이 될 수 있음을 확인하였다. 따라서, 이전보다 쉬워질 것으로 예상하며, 그러나 본 논문에서 제안하는 분석 방법은 동일한 스페인어에 대한 대소문자 단복수형을 구분하여 키워드를 별도로 추출하는 문제점이 있다. 따라서, 동일한 단어에 대해 대소문자, 단수형 및 복수형의 형태와 무관하게 단수형 키워드로 추출하는 방법을 향후 연구로서 연구 및 개발할 예정이다.

참 고 문 헌 (References)

- [1] Indicator Service | e-National Indicators, https://www.index.go.kr/unity/potal/main/EachDtlPageDetail.do?idx_cd=2457 (Accessed Mar. 13, 2024)
- [2] J. Kim, and B. Jeon, "Analyzing Global Startup Trends Using Google Trends Keyword Big Data Analysis: 2017~2022." *Journal of platform technology*, Vol 11, No.4, pp. 19-34, August, 2023. doi: <https://doi.org/10.23023/JPT.2023.11.4.019>
- [3] G. Lee, and S. Choi, "Analysis of Work-Related Musculoskeletal Disorders Research Trends Using Keyword Frequency Analysis and CONCOR Technique." *Journal of the Korea Society of Computer and Information*, Vol. 28, No. 8, pp. 137-144, August, 2023. doi: <https://doi.org/10.9708/jksci.2023.28.08.137>
- [4] C. Lee, and M. Moon, "Keyword and Emotional Analysis Diary Service Using KoNLPy and KoBERT." *Proceedings of the Korean Society of Computer Information Conference*, Jeju, Korea, pp. 501-502, 2022.
- [5] K. S. Lee, G. M. Lim, and H. M. Cho, "Searching of Keyword using XAI based Gambling Site Classification. Proceeding of Korea Computer Congress 2022, Jeju. Korea, pp. 1291-1293, 2022.
- [6] D. H. Kim, "A study on ICT trend analysis based on ICT keyword big data." *Proceedings of Symposium of the Korean Institute of communications and Information Sciences*, Yeosu, Korea, pp. 742-743, 2021.
- [7] Customs Legislation Information Portal(CLIP), <https://unipass.customs.go.kr/clip/index.do> (Accessed Mar. 13, 2024)
- [8] NLTK :: nltk.tokenize.regexp module, <https://www.nltk.org/api/nltk.tokenize.regexp.html> (Accessed Mar. 13, 2024)
- [9] NLTK :: nltk.tokenize.toktok module, <https://www.nltk.org/api/nltk.tokenize.toktok.html> (Accessed Mar. 13, 2024)
- [10] Tokenizer - spaCy documentation, <https://spacy.io/api/tokenizer> (Accessed Mar. 13, 2024)

저 자 소 개



김 신

- 2015년 8월 : 건국대학교 정보통신대학 컴퓨터공학부 졸업(학사)
- 2017년 2월 : 건국대학교 대학원 컴퓨터정보통신공학과 졸업(석사)
- 2023년 2월 : 건국대학교 대학원 컴퓨터정보통신공학과 졸업(박사)
- 2023년 3월 ~ 2024년 2월 : 건국대학교 스마트ICT융합공학과 박사후연구원
- 2024년 3월 ~ 현재 : 건국대학교 스마트ICT융합공학과 학술연구교수
- ORCID : <https://orcid.org/0000-0001-8492-3758>
- 관심분야 : 멀티미디어검색, 영상처리, 메타버스, 데이터분석, 인공지능



윤 경 로

- 1987년 2월 : 연세대학교 전자전산기공학과 졸업(학사)
- 1989년 12월 : University of Michigan, Ann Arbor, 전기전산기공학과 졸업(석사)
- 1999년 5월 : Syracuse University, 전산학과 졸업(박사)
- 1999년 6월 ~ 2003년 8월 : LG전자기술원 책임연구원/그룹장
- 2003년 9월 ~ 현재 : 건국대학교 컴퓨터공학과스마트ICT융합공학과 교수
- 2017년 10월 ~ 현재 : 국립전파연구원 멀티미디어부호화 전문위원회 대표전문위원
- 2019년 10월 ~ 현재 : IEEE-SA Interfacing Cyber and Physical World Working Group 의장
- ORCID : <https://orcid.org/0000-0002-1153-4038>
- 관심분야 : 스마트미디어시스템, 멀티미디어검색, 영상처리, 인공지능, 멀티미디어/메타데이터 처리, 메타버스