



A Design and Implementation of Differentiated Logistic Data Analytic Platform by Investigating Existing Similar Service Functions

Keejun Han, Hoon Jung, Dong-gil Na*

Intelligent Convergence Research Laboratory, ETRI

ABSTRACT

Despite the rapid growth of the non-face-to-face economy, customized marketing information and sales strategy provision services for small and medium-sized sellers are insufficient, highlighting the need to design and establish an advanced logistic data analytic platform is essential. Therefore, this study analyzed whether there are areas where existing functions can be improved or new functions can be added using big data and AI functions through analysis of domestic logistic analytic platforms. As a result, it seemed that existing services needed to establish data-sharing strategies and further machine learning solutions to investigate multiple internal and external data sources. In addition, in order to secure differentiation from existing services, a machine learning algorithm was designed to predict the weight and price of targeted items based on bill of landing data. Throughout the experiment, we observed that while LSTM does not work well for overall data, the linear regression and MLP models showed good overall performance. It confirmed the models could be used to provide the weight and price prediction services for countries and items targeted to export. The models built in this study are dockerized as a public API service in the future so that researchers and developers can use them actively.

© 2022 KKITS All rights reserved

KEYWORDS : Logistic bigdata analysis, Data mining, Linear regression, MLP, LSTM

ARTICLE INFO: Received 23 February 2022, Revised 27 March 2022, Accepted 8 April 2022.

*Corresponding author is with the Intelligent Convergence Research Laboratory, ETRI, 218 Gajeong-ro, Yuseong-Gu, Daejeon, 34129, Korea.

E-mail address: dgna@etri.re.kr

1. 서론

코로나-19에 따른 비대면 경제의 갑작스런 성장은 명품, 가전제품 역직구 시장등의 급격한 성장과 더불어 기존 유통물류 분야의 패러다임을 급격하게 변화시키고 있다. 서울연구원이 내놓은 연구에 따르면 2019년 한해에만 오프라인 소비는 8조원이 줄어든 반면 온라인 소비는 4조원이 증가하였으며 비대면 소비를 경험한 이들 중 80.1% 이상이 코로나의 종식 후에도 비대면으로 쇼핑 행태를 지속적으로 변화할 것이라고 응답했다. 또한 통계청에서 2022년 발표한 자료에 따르면, 2021년 온라인 쇼핑 총 거래액은 192조 8946억원 으로 전년대비 21%나 증가하였으며 이러한 금액은 관련 통계가 작성된 2001년 이래로 가장 큰 수치로 기록되었다[1,2].

특히나 주목할 부분은 K-콘텐츠의 전세계적 유행으로 인한 역직구 시장의 성장이다. 역직구란 해외 소비자가 국내 온라인 쇼핑몰에서 상품을 구매하는 행태로 관세청의 조사에 따르면 2019년 1000만건을 시작으로 2020년 2000만건을 넘어섰으며 그 금액 역시 2016년 2억 6900만 달러에서 2020년 8억 달러로 3배 이상 급성장하였다. 화장품과 같이 한류의 영향이 큰 품목들이 수출 품목 상위를 차지하였으며 전자상거래를 통한 성장은 더욱더 커질 것으로 전망되었다[3].

이러한 비대면 경제의 급격한 성장에도 불구하고 중소셀러를 위한 맞춤형 마케팅 정보 및 판매 전략 제공 서비스는 미흡한 실정이다. 한국무역협회에 따르면 국내 중소기업의 3.9%만이 해외진출을 계획하고 있는 상황이며 국내 총수출액 5,269억 달러 중 B2C 전자상거래를 통한 수출은 총수출 대비 약 0.03%에 그치는 것으로 나타났다. 또한 대기업에 비해 정보력이 약하고 자본이 빈약한 중소셀러들은 바이어 발굴, 마케팅 비용, 물류비용, 환율 및 반품문제, 해외현지 정부규제 등이 수출 전략을

수립하기 위한 큰 장애물이라고 조사되었다. 향후 2030년까지 세계 경제의 약 80%가 글로벌 시장화 될 것으로 전망되는 가운데 이러한 저조한 중소기업의 해외시장 진출 계획을 돕기위한 수출입 데이터 등의 물류 데이터 기반 해외 시장정보 제공 시스템 구축을 위한 설계는 필수적인 상황이다[4].

방대한 양의 데이터와 이를 기반으로 한 빅데이터 및 인공지능 기술의 발전은 실시간으로 수집되는 엄청난 양의 데이터를 분석하여 의미 있는 결과를 도출하고 이를 마케팅 전략 수립에 활용하거나 기업 전반의 의사결정 및 실행에 활용하는 데이터 기반 경영의 도입을 촉진시켰다. 데이터 기반 마케팅은 다양한 출처의 데이터를 결합하고 정교한 데이터 가공 및 분석을 통해 인사이트를 도출하여 이에 기반한 개인화 마케팅 전략을 통해 장기적인 관점에서 기존 고객과의 거래를 유지하고 새로운 고객의 실질적인 구매를 창출하는 것을 목표로 한다. 이러한 데이터 기반 마케팅의 궁극적인 지향점은 개인화 (Personalization)로 데이터를 활용한 기존의 세분화 마케팅에서 개개인의 특성을 반영하여 맞춤형 전략과 메시지를 수립 실행하는 개인화 마케팅으로 진화하고 있으며 이를 위한 맞춤형 광고, 추천 서비스에 대한 연구 및 개발이 최근 활발하게 진행되고 있다[5,6].

이에, 본 연구에서는 중소기업의 해외시장 진출 계획을 돕기 위한 물류 빅데이터 기반 해외 시장 정보 제공 시스템 구축을 위하여 기존의 물류 데이터를 제공하는 국내 플랫폼들에 대한 분석을 통해 빅데이터 및 AI 기능을 활용하여 기존의 기능을 개선 하거나 새로운 기능을 추가할 수 있는 부분이 있는지에 대해 분석하였다. 현재 수출입 및 관련 무역 정보를 제공하는 서비스들은 수출 기업들을 지원하는 공공기관들을 중심으로 활발하게 운영되고 있으나 대부분 1차적인 수출입 통계 데이터 제공에 그치고 있고 이러한 데이터를 추가적으

로 가공하고 결합하여 AI 및 빅데이터에 기반한 새로운 인사이트를 제공하는 기능들의 대한 제공은 미흡한 상황이다. 이를 개선하기 위해 본 연구에서는 각 서비스가 제공하고 있는 기능들에 대한 분류 및 분석을 수행하였고, 이를 통해 기존의 해외 시장정보 제공 플랫폼들의 활용도를 증대시킬 수 방안으로 선하증권 (Bill of Lading) 기반의 유통물류 데이터를 활용하여 대한민국의 주요 수출처인 베트남, 미국, 중국 시장에 대한 품목별 물량 및 가격 예측 기능을 머신러닝 기법을 활용하여 구축하였다.

궁극적으로 본 연구를 통한 유통물류 기반의 AI 및 빅데이터 연구 분야 기여점은 다음과 같다.

- 비대면, 온라인, 한류 확산에 따른 한국 상품의 선호도 증가에 따라 해외 진출을 희망하는 중소기업들을 지원할 수 있는 해외 시장정보 제공 플랫폼 서비스에 대한 시스템 설계를 제안한다.
- 기존 해외 시장정보 제공 플랫폼 서비스들의 기능을 분석한 결과 AI 및 빅데이터 분석 기반 기능이 미흡하여 서비스 활성화 방안을 위해 이에 대한 기능 추가가 필요함을 밝혔다.
- 이에 선하증권 기반 데이터를 바탕으로 수출 대상 품목에 대한 가격 및 물량 예측 알고리즘을 설계하였고, 그 성능을 검증하였다. 또한, 본 연구를 통해 개발된 모델의 활발한 재사용을 위해 Open API 설계 및 구축 전략에 대해 제시하였다.

본 논문의 구성은 다음과 같다. 제 2장에서는 기존 국내 해외 시장정보 제공 플랫폼 서비스들에 대한 소개 및 기능 분류 결과에 대하여 소개하고, 3장에서는 물량 및 가격 예측을 위한 학습 모델을 구축하고, 이에 대한 성능 검증을 통해 수출 중소기업들을 위해 해당 기능이 실제로 제공 가능함을 검증하였다. 마지막으로 4장에서는 본 연구에 대한

고찰을 통해 연구의 의의 및 향후 연구에 대하여 논의하였다.

2. 관련 서비스 분석 및 시사점

국내의 무역투자자와 관련한 정보를 제공하는 서비스들은 크게 1) 통계청에서 제공하는 수출입 통계 데이터와 같은 외부 데이터를 제공하거나 2) 이를 기반으로 서비스 자체적으로 추가적으로 수집한 데이터, 3) 이러한 정형, 비정형 데이터에 기반한 AI 및 빅데이터 기능 크게 세가지로 분류할 수 있다.

국내의 여러 주요 서비스들 중 유료 회원 가입을 통한 추가적인 기능을 제공하는 서비스들을 제공하고 수출 중소기업이 활용할 수 있는 서비스들을 분석한 결과 본 연구에서는 서비스 분석 대상으로 KOTRA의 무역투자빅데이터[7], 한국무역통계진흥원(TRASS)이 제공하는 한국무역통계정보포털[8], 중소기업진흥공단(KOSME)에서 제공하는 고비즈 코리아[9], 한국무역협회(KITA)에서 제공하는 K-STAT[10]을 선정하였다. 해당 서비스들은 국가의 위탁을 받은 공공기관에서 직접 운영하는 서비스들로 신뢰도 높은 관세청 등의 데이터를 기반으로 하여 데이터의 신뢰성이 높고 지속적인 운영을 통한 업데이트가 이루어지고 있다. 또한 대부분의 수출 중소기업들이 영세한 업체임을 고려할때 초기 수출 전략 수립을 위해 해당 서비스들의 무료 기능을 활용할 것으로 기대되어, 각 서비스들의 유료 기능들은 분석 대상에서 제외하였다.

KOTRA의 무역투자빅데이터는 크게 무역투자통계, 품목별유망시장, 국가별시장정보, 기업별 맞춤형 정보 기능을 제공하고 있다. <표 1>에서 <표 4>는 각 기능별 세부 기능 및 해당 기능이 내부, 외부 DB 기반으로 제공되는지, 혹은 AI 및 빅데이터 전략을 통해 제공되는지에 대한 분류를 나타낸 표이

다. 기능별 제공하는 데이터가 외부에서 가져온 데이터인지 명확하지 않은 경우에는 내부 데이터로 분류하였다.

표 1. KOTRA의 무역투자통계 서비스 기능분석
Table 1. Function analysis of trade and investment statistics service by KOTRA

무역투자통계		
분류	기능	세부기능
내부DB	한국수출현황	품목별 한국발 전세계 수출현황
		품목별 수출 상위 5개 지역
		품목별 전년대비 교역증가 TOP 5 교역지역 상위 5개국
	글로벌수출현황	지정 지역 간 수출입 현황
		지정 지역 간 수출 품목 TOP 5
		지정 지역 간 수출품목 분포
	외국인직접투자통계	외국인직접 투자금액
		외국인직접 투자 상위 5개국
		전년대비 투자 증가율 상위 5개국

표 2. KOTRA의 품목별 유망시장 서비스 기능분석
Table 2. Function analysis of promising market by item service by KOTRA

품목별 유망시장		
기능	분류	세부기능
품목별 유망시장 정보	내부 DB	품목별 교역 순위 상위 5개 지역
		품목별 상위 5개 지역 대비 한국 점유율
		품목별 최근 5개년 수출 신고액 추이

유망시장상세 정보	AI/빅 데이터	유망시장 상위 30개의 상세 수출입 통계	
		해당 품목 유망시장 순위	
	내부 DB	해당 품목 상위 수입 10개 지역	
		한국 기준 수출상위 5개 지역	
		조회 유망시장의 해당 품목 수출입 통계 그래프	
		조회 유망시장으로의 수출기업 수	
		연도별 해당품목 수출입현황 통계	
		조회 유망시장의 해당품목 수입액 현황	
	AI 분석	내부 DB	해당 품목 관련 뉴스 및 상품정보
			UN Comtrade 및 관세청 통계 기반 해당 품목 통계 변화 동향 리포트
		외부 DB	해당 품목 관련 전시 정보
			해당 품목 바이오퍼 현황 (바이코리아 기반)
해당 품목 관련 수입규제 현황			
해당 품목 관련 뉴스 및 상품정보			

표 3. KOTRA의 국가별 시장정보 서비스 기능분석
Table 3. Function analysis of market information by country service by KOTRA

국가별 시장정보		
기능	분류	세부기능
World Bank 지표	외부 DB	기업환경평가접수
		법인설립소요일
		실업률
		1인당 GDP (PPP)
		물류 성과 지수
		항만 컨테이너 물동량
		항공 화물 수송량
		GDP
		경제성장률
	내부 DB	호조 품목

KOTRA 지표	외부 DB	부진 품목
		호부진 품목 상세정보
		호부진 품목에 대한 수출입 통계
	내부 DB	바이오피 수 (바이코리아 기반)
언론사 뉴스	내부 DB	수출금액 (관세청 기준)
		수출기업 수 (관세청 기준)
		해당 시장 관련 뉴스 및 보고서
	AI/ 빅데이터	해당 시장 관련 무역 사기 사례
		해당 시장 수입 규제 품목 정보
		해당 시장 관련 주요 키워드
AI 분석	내부 DB	해당 시장 관련 주요 키워드 트렌드
	AI/ 빅데이터	주요 품목 전년동기 대비 증감률
		해당 시장 관련 수출 동향 리포트

		수출통계 정보
		조희 기업 수출품목 별 각 8개의 유망시장 및 관련 수출통계 정보
		인공지능이 분석한 유망시장 정보
		잠재파트너 추천정보 (키워드 연관도 기반)
		사업 추천 (KOTRA 및 정부과제관련 무역지원 사업 추천)
잠재 파트너 정보	내부 DB	KOTRA DB 기반 잠재 파트너 정보
	외부 DB	외부 DB기반 잠재 파트너 정보

표 4 KOTRA의 기업별 맞춤정보 서비스 기능분석
Table 4. Function analysis of customized information for each company service by KOTRA

기업별 맞춤정보		
기능	분류	세부기능
기업 개요	내부 DB	기업 정보 (KOTRA 가입 정보 기반)
		기업 현황 정보
		사업 참여 이력
		해외기업 상담 이력
		바이코리아 등록 상품 정보
	KOTRA 상담 이력	
	AI/빅데이터	기업 주요 키워드
유망 시장 추천	내부 DB	유사기업 (조희 기업의 10대 품목 중 1개 이상의 동일 품목을 수출하는 기업) 통계 정보
	AI/빅데이터	조희 기업 수출품목 데이터 바탕 3대 유망시장 및 관련

무역투자빅데이터의 무역투자통계 기능은 한국에서 특정한 국가로 수출된 품목에 대한 전체 및 개별 통계 기능을 제공한다. 이때, 품목은 약 10자리 정도의 국제적으로 통용되는 상품분류 체계인 국제통일상품 분류체계인 HSCODE (Harmonized Community Description and Coding System)로 나타낼 수 있다. 국제협약에 따라 HSCODE는 10자리까지 사용할 수 있으며, 6자리까지는 국제 공통, 7-10 자리까지는 각 나라에서 세분화하여 고유의 숫자를 부여할 수 있다. HSCODE에 따라 관세율, 수입요건 등이 결정되기 때문에 무역투자통계 정보를 제공하는데 가장 중요한 항목 중 하나라고 볼 수 있다.

국가별 시장정보는 기능은 KOTRA 내부가 가지고 있는 시장정보 및 세계은행에서 제공하는 국가별 주요 지표를 제공한다. 추가적으로 언론사에서 제공하는 각 국가의 수출입에 영향을 미치는 뉴스 및 보고서들을 제공한다. 해당 시장 관련 주요 키워드는 각 문서 내 단어 빈도수를 기준으로 추출되어 사용자에게 제공되며 이러한 빈도수를 기준으로 최근 4분기 동안의 해당 시장 관련 해외뉴스의 주요 키워드 트렌드를 제공하는 기능을 보유하

고 있다. 이러한 기능은 현재 해당 시장에서 주목 받고 있는 단어들에 대한 정보를 제공하는 효과가 있으나, 일반적인 단어 빈도수 기준의 정보인 TF (Term frequency) 정보만 활용하고 단어의 대표성을 측정하는데 주로 활용되는 IDF (Inversed document frequency)가 사용되지 않아 문서 대표성이 높은 단어 추출이 어려운 한계가 있다. <표 5>는 미주 시장의 2020년 분기별 주요 트렌드 단어들을 나타낸다. 이 중 ‘government’, ‘police’ 와 같은 단어들은 문서 전체적으로 등장하는 빈도수는 높지만 반대로 문서 대표성이 적은 명사들로 해당 시장의 트렌드를 정교하게 반영하기에는 한계가 존재하는 것을 볼 수 있다.

표 5. 무역투자빅데이터 서비스에서 추출한 2020년 분기별 미주 시장의 분기별 트렌드 단어
Table 5. Quarterly trending keywords in 2020 extracted from Trade Investment Bigdata service

분기	트렌드 단어
20.1	government, police, virus, president, minister
20.2	government, lockdown, virus, health, home
20.3	government, health, president, pandemic, hospital
20.4	government, president, election, party, pandemic

추가적으로, 해당시장의 국내 주요 품목 수출액에 대하여 수출 동향 리포트를 만들어 주는 AI 서비스가 존재하지만, 수출 호조 및 부진 품목에 대한 통계 정보를 자연어 정보로 변환한 정도의 한계를 지니고 있다.

품목별 유망시장은 과년도 품목별 수출통계데이터를 기반으로 품목별로 수출 호조를 나타내는 해외 시장에 대한 세부 정보를 제공한다. 해당 품목 유망시장 순위와 UN Comtrade 및 관세청 통계 기반 해당 품목 통계 변화 통합 리포트 역시 과년도



그림 1. 등록된 기업 정보에서 추출된 기업 주요 키워드
Figure 1. Seller keywords extracted from registered seller information

품목별 수출통계데이터 결과를 제한적인 자연어 정보로 변환한 정도의 AI 서비스를 제공한다.

기업별 맞춤정보는 수출희망기업이 KOTRA에 회원 가입시 제공한 정보들을 기반으로 KOTRA와 연계된 통계 정보들을 제공하거나 해당 정보에서 추출된 텍스트 및 취급 HSCODE를 기반으로 기업의 주요 키워드 정보를 활용하여 품목별 유망시장을 추천하는 서비스를 제공한다.

특히 무역투자빅데이터에서는 <그림 1>과 같이 추출된 기업의 주요 키워드를 바탕으로 가장 유사한 키워드를 가진 해외기업 리스트를 제공하는 잠재파트너 추천정보 서비스를 제공한다. 이는 중소 셀러들의 수출 전략 수립에 가장 큰 장애물 중 하나인 새로운 바이어 발굴에 대해 도움을 줄 것으로 예상되지만, 기업과 바이어의 대표 키워드가 정확하게 매칭되어야만 추천이 가능하여 각 단어가 정확하게 매칭되지 않더라도 단어의 유사성을 학습하여 추천해주는 심도 깊은 추천 알고리즘 활용의 필요성을 시사한다.

TRASS 역시 무역통계정보를 바탕으로 수출 중소 셀러를 대상으로 수출입기업의 무역경쟁력 제고를 위해 한국무역통계정보포털 서비스를 운영하고 있다. <표 6>에서 <표 9>는 한국무역통계정보포

털의 각 기능별 세부 기능 및 기능별 분류를 나타낸 것이다. 기능별 제공하는 데이터가 외부에서 가져온 데이터인지 명확하지 않은 경우에는 내부 데이터로 분류하였다. 기타 서비스와 유사한 기능을 제공하는 기능들을 제외하고 살펴본 한국무역통계정보포털 서비스의 차별점은 국가별, 공항, 항구별 화물통계를 제공하여 가격뿐만 아니라 중량에 대한 정보 역시 제공하고 있다는 점이다. 이는 실제 운송 과정에서 부과되는 여러 비용에 대한 정확한 산정이 필요한 수출 업체들에게 도움이 되는 정보이지만, 역시나 과년도에 화물통계 데이터의 기반한 통계 정보 제공에 그치고 있어 자동 운임 및 물량 예측등의 보다 적극적인 AI 기술의 활용이 필요함을 시사한다.

표 6. TRASS의 수출입통계 서비스 기능분석
Table 6. Function analysis of import/export statistics service by TRASS

수출입통계		
분류	기능	세부기능
내부 DB	총괄	한국 연도별 월별/연도별 수출입 통계정보
	기본	품목별 수출입 통계
		국가별 수출입 통계
		대륙/경제권별 수출입 통계
		기타 수출입 통계
상세	기본통계 복수 비교 정보	

표 7. TRASS의 테마통계 서비스 기능분석
Table 7. Function analysis of theme statistics service by TRASS

테마통계		
분류	기능	세부기능
내부 DB	중소기업 맞춤형 통계	기업유형별 개수/수출금액 통계
		기업유형별 품목 동향 통계
		품목별 수출 유망 국가
		급증감 품목

	기업무역 활동 통계	활동성 높은 기업
		생존/존속 기업
		성장/가젤 기업
		기업 유형별 수출 기여도
		기업무역활동 분석보고서
외부 DB	FTA 통계	FTA별 수출/수입 통계
		품목별 FTA 통계
		품목별 FTA 변화 추이
		국내 지역별 FTA 활용 통계

표 8. TRASS의 시장분석 서비스 기능분석
Table 8. Function analysis of market analysis service by TRASS

시장분석		
분류	기능	세부기능
내부 DB	무역변화 추이	무역수지 흑/적자국
		무역수지 흑/적자 품목
		국내지역별 무역수지
AI/빅데이터	핫템 키워드	검색 상위 품목 키워드

표 9. TRASS의 화물통계 서비스 기능분석
Table 9. Function analysis of cargo statistics service by TRASS

화물 통계		
분류	기능	세부기능
내부 DB	총괄	연도별 수출입 중량 통계정보
		국가별 중량 통계정보
	국가별	국가별 수출입 통계
		공항/항구별

TRASS에서 검색을 수행하게 되면 해당 검색어에 대한 정보를 수집하여 검색 기간별 검색 상위 품목에 대한 키워드를 <그림 2>와 같이 보여준다. 이는 선택된 기간 별로 사용자들이 많은 관심을 갖는 품목에 대하여 파악할 수 있다. 하지만 검색 키워드의 분포는 상위 소수의 검색어가 전체 검색 비중의 절대 다수를 차지하고 있는 지프의 법칙을 따르기 때문에 검색의 빈도수만 고려하는 것은 기간에 상관없이 상위 검색어가 항상 결과에 포함되는 결과를 낳는다. 이에 기간에 따라 변화하는 사

용자의 검색 트렌드를 파악하기 위해 검색의 빈도수와 더불어 기간별로 빈도의 변화수가 극심한 검색어들을 파악하여 그에 대한 가중치를 부여하는 정보검색 (Information Retrieval) 기술을 고도화하는 것이 필요함을 시사한다[11].

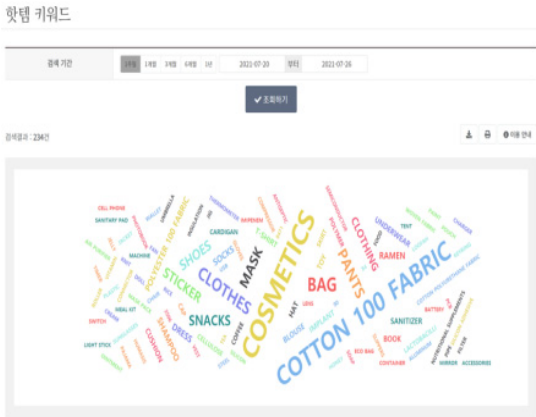


그림 2. 검색 키워드 클라우드
Figure 2. Search keywords cloud

KOSME가 운영하는 온라인 수출 플랫폼 고비즈 코리아는 2021년 한해에만 온라인 수출 3,084건 지원 및 총 1억 2800만달러라는 수출 실적을 기록한 국내 대표 B2B 플랫폼으로 약 32만명의 해외 바이어가 해당 플랫폼을 통해 역직구 품목들을 검색하고 구입하고 있는 것으로 파악된다[12].

고비즈 코리아의 경우 전반적인 물류 통계 데이터에 대한 정보 제공 이외에도 중소기업을 위한 재정 및 비재무 프로그램, 컨설팅, 교육, 마케팅 및 글로벌 협력 프로그램을 포함한 자문 프로그램을 운영하고 있으며 이러한 컨설팅 및 지원사업 프로그램은 본 연구의 범위에는 벗어나기에 해당 기능들은 제외하고 순수한 시장정보 제공과 관련된 기능 부분들만 분석을 수행하였다. 고비즈 코리아의 경우 이러한 시장정보 제공은 마케팅서비스의 일환으로 제공되고 있으며 해당 서비스의 세부 기능

및 기능별 분류 결과는 <표 10>과 같다.

해외 바이어 동향은 영문 고비즈코리아 서비스를 방문한 바이어들이 많이 클릭한 품목들에 대한 정보를 수집하여 품목별, 제품별로 실시간 동향 리스트를 제공한다. 해외 구매 오피 정보의 경우 고비즈 코리아로 요청된 바이어의 오피를 게시판 형태를 통해 공고하고, 이에 관심있는 수출 업체를 중개해주는 서비스이다.

표 10. KOSME의 고비즈 코리아 기능분석
Table 10. Function analysis of Gobiz Korea by KOSME

마케팅 서비스		
분류	기능	세부기능
내부DB	글로벌 전자상거래 동향	시장분석보고서
		월동향보고서
	해외 바이어 동향	실시간 동향
		국가별 동향
품목별 동향		
	제품별 동향	
	업체별 동향	
	해외 구매 오피 정보	오피 공고

고비즈 코리아가 타 서비스 대비 갖는 데이터 관점의 차별화는 고비즈 코리아에서 제공하는 시장분석보고서, 월동향보고서, 오피 공고와 같은 기능의 경우 일차적으로 수집되고 가공된 데이터들을 기반으로 한 실질적인 지식들을 제공하고 있다는 점이다. 이는 뉴스 및 수출입 등의 단순한 사실의 나열인 데이터들로부터 통계 정보 및 단어 그룹들을 제공함을 통해 의미있는 데이터인 정보 (Information)를 제공하는데 그친 타 서비스에 비하여 복수개의 정보에 대한 교차 분석 및 관계 정보 파악을 통해 사용자에게 실질적으로 유용한 정보를 제공하는 지식 (Knowledge)을 제공한다는 점이다. 이는 측정된 형태의 값으로 이루어진 데이터, 데이터의 일차 가공을 통해 의미있는 데이터들을 파악하는 정보, 정보의 조합을 통해 새로운 가치를 제공하는 지식, 제공된 가치를 통해 새로운 지식들

을 창출해내는 지혜의 단계로 각 모듈들이 연결되어 있는 DIKW 피라미드에서 타 서비스 대비 보다 고차원적인 가치인 지식 단계의 기능을 제공하고 있음을 의미한다[13].

이러한 지식들은 실제 중소기업들의 수출 전략 수립에 있어 유용한 인사이트를 제공하지만 반대로 정보의 가공과 교차 분석에 있어 많은 인력과 비용을 요구한다. 이는 서비스의 유지에 있어 지식공학자의 지속적인 개입을 요구하기 때문에 운영비용의 상승과 점진적으로 쌓이는 정보 및 지식 관리의 어려움을 야기한다. 이러한 한계를 극복하기 위하여 비정형 형태의 텍스트 데이터로부터 주요한 부분만을 추출하여 요약하는 자동 요약(Automatic summerization) 기법이나 자동 문장 생성에 널리 활용되는 전이 학습 (Transfer Learning) 기반의 언어 모델 등의 AI 기술을 활용한 서비스 운영의 효율화가 필요하다[14-16].

표 11. K-STAT의 마케팅 서비스 기능분석
Table 11. Function analysis of marketing service in K-STAT

마케팅 서비스		
분류	기능	세부기능
내부DB	국내 수출입 통계	총괄수출수입통계
		동향분석보고서
		결계형태별 수출입 통계
		수출형태별 수출입 통계
		수출종류별 수출
	가공단계 수출입 통계	가공단계별 수출수입통계
		각 가공단계 품목별 수출수입통계
		각 가공단계별 국가 수출수입통계
	품목 수출입 통계	품목별 수출수입 통계
		품목별 특정국가에 대한 수출입 통계
		전년대비 투자 증가율 상위 5개국
	국가 수출입 통계	국가별 수출수입 통계
		한 품목에 대한 국가별 수출입통계

대륙/경제권 수출입	대륙/경제권 수출입통계
	대륙/경제권 내 품목 수출입통계
	선택된 대륙의 국가별 수출수입통계
지자체 수출입	지자체 수출입통계
	지자체별 품목 수출수입통계
	지자체별 대륙권 수출수입통계
	지자체별 국가 수출수입통계
	지자체 가공단계별 수출수입통계
항구/공항 수출입	항구/공항 수출입
	항구/공항 품목별 수출입
	항구/공항별 국가 수출입
FTA 체결국 수출입	FTA 체결국 수출입
	FTA 체결국 품목별 수출입

표 12. K-STAT의 해외무역통계 서비스 기능분석
Table 12. Function analysis of function trade statistics service in K-STAT

해외무역통계		
분류	기능	세부기능
내부DB	해외무역관련통계	해외무역수출입
		국가별 수출입
		각국 10대 품목
		한국 10대 품목
		부품 소재

KITA에서 제공하는 K-STAT은 한국을 포함한 세계 61개국의 무역통계를 서비스하고 있으며 국내 최대 규모의 무역통계 데이터를 보유하고 있는 수출입 데이터 제공 서비스이다. <그림 6>은 K-STAT의 각 기능별 세부 기능 및 기능별 분류를 나타낸 것이다. 기능별 제공하는 데이터가 외부에서 가져온 데이터인지 명확하지 않은 경우에는 내부 데이터로 분류하였다. 기타 서비스와 유사한 기능을 제공하는 기능들을 제외하고 살펴본 K-STAT의 차별점은 방대한 양의 통계 데이터를 바탕으로 완제품

이 아닌 가공단계의 수출입 통계 데이터, 대륙/경제권 수출입, 지자체 수출입, FTA 체결국 수출입 등의 보다 세분화된 형태의 통계 정보를 제공한다 는 점이다. 또한 2019년 한일 무역 분쟁으로 촉발된 소재, 부품, 장비 품목에 대한 국산화 수요 및 수급 분석에 활용할 수 있는 부품 소재 관련 통계 정보를 제공하여 정부 정책 및 중소기업의 수출 및 부품 수급 전략을 수립하는데 활용될 수 있다. 그러나 K-STAT은 이러한 통계 정보를 제공하는데 그치고 있어 이를 활용하여 중소 셀러들이 수출 전략을 수립하기 위해서는 추가적인 분석 및 검토가 필요하다.

〈표 13〉은 분석 대상 서비스의 기능별 분류를 요약한 표이다. 분석 대상 서비스 모두 각 서비스를 운영하는 주체들이 보유한 내부 통계 데이터를 기반으로 서비스를 구축 및 운영하고 있음을 파악할 수 있다. 이는 각 서비스가 제공하는 데이터를 가공하여 정보를 제공할 수는 있으나, 서로 다른 정형/비정형 데이터를 융합하고 이를 기반한 빅데이터 분석을 통한 가치있는 지식을 창출하여 중소 수출 셀러들에게 새로운 인사이트를 도출하기 위한 서비스 개선이 필요함을 의미한다. 또한 전체 기능 중에서 내부 및 외부의 통계 데이터의 제공이 약 94%에 이르는 반면, AI나 빅데이터 기반의 기능을 제공하는 것은 6%에 그치고 있고 실제 기능 구현에 활용된 기술도 빈도수 세기 기반의 기초적인 활용에 그치고 있어 머신러닝 및 딥러닝 기술을 활용한 AI 및 빅데이터 기능의 성능 개선에 대한 시도가 필요하다.

특히, 서론에서 언급한 바와 같이 중소 셀러의 경우 수출하고자 하는 품목에 대한 가격 예측 및 물량 예측을 통해 보다 선제적으로 인기있는 품목들은 선 매입하고, 이를 적절한 시기에 대상 국가에 수출하는 의사 결정이 매우 중요하다. 이러한 기능은 기존 분석 대상 서비스에서 제공하고 있지

않기 때문에 수집 가능한 유통 물류 데이터를 통해 해당 기능을 제공하면 수출 중소 셀러들의 수출 전략 수립에 큰 도움이 될 것으로 판단된다.

이에 다음장에서는 기존 통계 기반의 서비스들과의 차별성 확보를 위해 기존의 수출입 통계 정보를 바탕으로 국가별 물량 및 가격의 예측치를 판단하는 머신 러닝 알고리즘을 설계하고 이에 대한 성능을 측정한다. 이를 통해 기존 서비스 분석을 통해 분석된 개선을 위한 제안들의 실현가능을 검증한다. 또한, 서비스에 종속적인 데이터 분석에 대한 한계를 개선하기 위해 구축된 물량 및 가격 예측 모델을 Open API로 제공하기 위한 시스템 설계를 제안한다.

표 13. 분석 대상 서비스의 기능별 분류
Table 13. Functional classification for the investigated services.

서비스명	통계 데이터		AI, 빅데이터 기반
	내부DB 기반 개수(비율)	외부DB 기반 개수(비율)	
무역투자빅 데이터	38 (58%)	16 (25%)	11 (17%)
한국무역통 계정보포털	21 (81%)	4 (15%)	1 (4%)
고비즈 코리아	8 (100%)	0 (0%)	0 (0%)
K-STAT	31 (100%)	0 (0%)	0 (0%)
평균 비율	84%	10%	6%

3. 물량 및 가격 예측 모델

물량 및 가격 예측과 같은 머신러닝 모델은 데이터의 특성 및 전처리, 정답셋의 활용 가능 여부에 따라 적용 가능한 기법이 다양하다. 본 연구에서 활용하는 선하증권 데이터는 수출일을 통해 품목의 흐름을 시간순으로 파악할 수 있고, 물량 및 금액에 대한 데이터가 존재하여 이를 학습 데이터

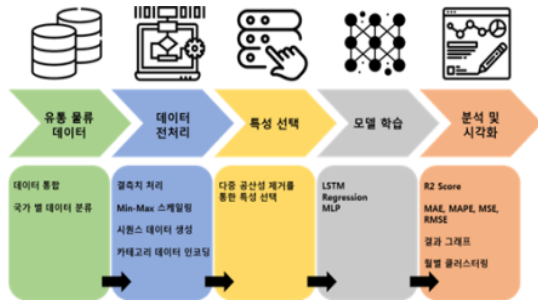


그림 3. 물량 및 가격 예측 모델 구축을 위한 프로세스
Figure 3. A process for building a quantity and price prediction model

로 활용이 가능하다. 이에 본 연구에서는 전통적인 회귀 모델, 데이터의 순서가 예측모델 성능에 영향을 미치는지를 파악하기 위한 LSTM, 마지막으로 딥러닝모델의 효용성을 알아보기 위한 MLP 모델을 대상 모델로 선정하였다.

<그림 3>은 중소 수출 셀러가 수출 전략을 수립하기 위해 필수적인 품목별 물량 및 가격 예측 모델을 구축하기 위한 전체적인 프로세스를 나타낸 것이다. 각 단계별 수행된 작업에 대한 요약은 아래와 같다.

- 유통 물류 데이터: 원본 데이터에 대하여 필요한 새로운 특성들을 추가하고 서비스를 제공하고자 하는 국가별로 데이터를 분류하였다.
- 데이터 전처리: 데이터셋 내의 모든 결측치를 처리하고, Min-Max 스케일링을 적용하여 데이터 표준화를 수행하였다. LSTM 모델의 입력으로 활용할 수 있도록 데이터 셋을 시퀀스 데이터셋으로 바꾸주었으며 HSCODE와 같은 범주형(카테고리) 데이터는 One-hot Encoder를 사용하여 벡터화 시켜주었다.
- 특성 선택: 회귀 문제를 해결하기 위해서 변수들 간의 상관관계를 없애기 위해 다중공선성을 확인하여 전체 변수의 VIF(Variance Inflation Factor)가 10이하가 될 때까지 VIF가 가장 높은

변수를 하나씩 제거하여 특성을 선택하였다.

- 모델 학습: 특성 선택이 완료된 데이터를 기반으로 예측 모델을 학습하였다.
- 분석 및 시각화: 예측된 결과를 바탕으로 결정 계수(R2 score)와 MAE, MAPE, MSE, RMSE와 같은 예측 오차에 대한 평가 지표를 활용하여 결과를 분석하였으며 이를 시각화하였다.

본 실험은 Intel i7 CPU, NVIDIA 1080 Ti x2 GPU, RAM 32GB, SDD 512GB/HDD 3.4TB, Ubuntu 16.04 버전 OS 사양의 서버에서 수행되었으며, Python 3.5 및 Flask 기반으로 학습 모델 및 API를 구축하였다. 이에 추후 다른 환경에서도 구축된 API를 쉽게 활용하기 위하여 도커 1.13.1 버전을 활용하여 도커 이미지로 구축하였다.

3.1 데이터 분석 및 전처리

본 연구에서 활용한 데이터는 각 국의 관세청 및 그 유관기관에 신고하는 선하증권의 정보로 수출자, 수입자, 도착일자, 제품 상세, 물량, 선적지 및 양하지 등의 내용을 포함하고 있어 대상 국가에 대한 품목 물량 및 가격 예측 모델을 학습시키는데 활용가능한 데이터로, 2020년 1월부터 12월까지의 1,026,522개의 열로 구성되어 있다. 데이터에 대한 속성은 <표 14>와 같다.

표 14. 데이터셋 속성
Table 14. A description of the dataset

	변수명	설명
1	index	인덱스
2	date	수출일
3	hs_code	품목 HSCODE
4	mti_code	품목 mti code
5	country	수출대상국가
6	quantity	물량
7	weight	중량(kg)
8	price	금액(원)

9	online_quantity	전자상거래 물량
10	online_weight	전자상거래 중량 (kg)
11	online_price	전자상거래 금액 (원)

본 연구에서는 이중 가장 활발하게 역직구 되는 품목인 향료, 화장품 (HSCODE:33), 의류, 신발, 모자류 (HSCODE:61~65), 전자기기 (HSCODE: 85)를 선정하였으며 한류의 영향이 큰 중국, 베트남과 가장 큰 역직구 시장인 미국에 대한 51,004, 49,010, 47,772개의 데이터를 활용하였다. 이는 전체 데이터의 약 14% 수준이다.

모델 학습에 앞서 결측치 혹은 수기로 입력된 값의 오류들을 수정하기 위하여 다음과 같은 전처리 과정을 수정하였다. Quantity 필드의 값이 결측치 혹은 0 이하일 경우, 해당 트랜잭션을 삭제하였다. 해당 값들은, 세관 기입 상의 누락으로 인한 것으로 데이터를 학습하거나 평가하는데 있어 부정적 영향을 끼친다고 판단하여 제거하였다. Price 필드의 값이 0 이하일 경우, 해당 트랜잭션을 삭제하였다. 해당 값들은, 세관 기입 상의 누락으로 인한 것으로 데이터를 학습하거나 평가하는데 있어 부정적 영향을 끼친다고 판단하여 제거하였다. 이러한 전처리 과정들을 통하여 우리가 얻은 국가별 데이터의 최종 트랜잭션은 <그림 4>와 같다.

우리는 선하증권 정보 외에 수출입 변동에 직접적인 영향을 미치는 추가적인 메타 데이터를 활용하여 제안하는 모델이 여러 가지 특징들을 고려하도록 설계하였다. 4가지 대표적인 지표들은 다음과 같다.

- 소비자물가지수 (CPI) : 한 가정이 소비하기 위해 구입하는 재화와 용역의 평균 가격을 측정하는 지표로, 우리가 살아가기 위한 상품이나 서비스의 가격이 얼마나 오르고 내렸는지를 확인하는데 활용함.
- 생산자물가지수 (PPI) : 국내 생산자가 국내 시

장에 공급하는 재화의 요금을 경제에서 차지하는 중요도를 고려하여 평균을 낸 종합적인 가격수준을 지수화한 통계이다. 이는 경기 동향을 파악하는 지표로 이용함.

- 환율 (EX) : 국가 간 통화의 교환 비율로 해외 직구, 역직구시 통화의 대외가치를 나타내는 의미를 갖음.
- 금리 (LR) : 금리는 빌려 준 돈이나 예금에 붙는 이자 또는 비율을 의미함.
- 소득교역조건지수 (TTI) : 상품추출입에 따른 가격상의 유리도와 불리도가 어느 정도인지를 알아보고 이를 무역정책에 반영하기 위해 작성하는 지표로, 순상품 교역조건에 수출물량지수를 곱한 값을 의미함.

3.2 특성 선택

원-핫 인코딩(one-hot encoding)은 이름에서 명시하는 대로, 하나의 1과 수많은 0 값으로 특정한 데이터를 표현하는 인코딩을 의미한다. 이는 일반적으로 벡터로 표현되어지는데, 이렇게 표현된 벡터를 one-hot vector라고 한다. 본 연구에서, 우리는 주어진 HSCODE를 수치 데이터가 아니라, 카테고리 정보를 갖는 데이터로 간주하여 학습을 진행

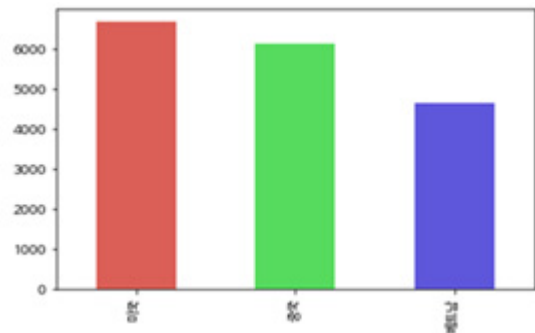


그림 4. 실험 데이터셋 사이즈
Figure 4. Experiment dataset size

하였다. 이는 HSCODE가 수치적인 값을 의미하지 않기 때문이다. 다르게 말하면, 두 개의 HSCODE의 산술연산은 그 어떤 의미도 갖지 않는다. 이를 수행하기 위하여 HSCODE에서의 클래스 정보를 의미하는 앞자리 숫자를 이용하여, one-hot vector를 생성한다. 구체적으로, 최상단 4자리 (호), 그리고 최상단 6자리 (소호)를 이용하여 실험을 수행하였다. 이전 4장에서 언급한 전처리를 통해 얻은 데이터에서, 우리가 존재하는 모든 자리(10자리)를 사용할 경우, 총 901 개의 독립적인 hs_code group가 생성되었으며, 각 hs_code 당 평균 19.26 개의 값을 가진다는 것을 확인할 수 있었다. 최상단 4자리만을 사용할 경우, 우리는 77개의 독립적인 hs_code group이 만들어질 수 있으며, 평균 개수는 226.17 개이다. 만약 우리가 6자리를 고려할 경우, 449개의 hs_code group이 생성되며, 평균 개수는 38.79개이다. 회귀분석에서의 다중 공선성은 모형의 일부 독립 변수가 다른 독립변수와 상관되어 있을 때 발생하는 조건을 의미한다. 다중 공선성이 큰 값을 가지는 것은, 회귀 계수의 분산을 증가시켜 모형을 불안정하고 해석하기 어렵게 만들기 때문에 문제가 된다. 독립 변수(특성)간의 다중 공선성을 측정하기 위해 독립 변수의 상관 구조를 검사할 수 있는데 이때 분산 팽창 인수(Variance Inflation Factor, VIF)를 검사할 수 있다. VIF는 독립 변수가 상관관계에 있는 경우 추정된 회귀 계수의 분산이 증가하는 정도를 측정한다. 모든 VIF가 1이면 다중 공선성이 없지만 일부 VIF가 1보다 크면 예측 변수 간에 완전한 상관관계가 있을 수 있으며 보통 VIF 10 이상인 경우에 회귀 계수가 잘못 추정되었다고 할 수 있다. 본 연구에서는 모든 독립변수가 10 이하의 VIF 값을 가질 때 까지 가장 높은 VIF를 갖는 독립변수를 하나씩 제거하는 방법으로 독립 변수를 선택하였다. 최종적으로 선택된 독립변수는 다음과 같다. 각 feature 데이터 값의 분포가 넓게 퍼

져있어서 이를 보완하고 전체 데이터를 일관성 있게 다루기 위해 MinMaxScaler를 사용하여 정규화하였다. <표 15>는 이러한 특성 선택을 통해 최종적으로 선택된 특성들이다.

표 15. 최종 선택된 특성
Table 15. A list of the chosen features

	VIF 값	변수명	설명
1	1.003	quantity	물량
2	1.012	weight	중량 (kg)
3	1.055	price	가격 (원)
4	2.519	online_quantity	전자상거래 물량
5	7.621	online_weight	전자상거래 중량 (kg)
6	4.763	online_price	전자상거래 금액 (원)
7	5.693	TTI	소득교역조건지수
8	3.487	IR	금리
9	4.118	Kor_PPI	한국 생산자 물가지수
10	1.7	PPI	생산자물가지수

3.3 모델 학습

본 연구에서는 전통적인 회귀 모델, 데이터의 순서가 예측모델 성능에 영향을 미치는지를 파악하기 위한 LSTM, 마지막으로 딥러닝모델의 효율성을 알아보기 위한 MLP 모델을 대상 모델로 선정하였다. 또한, 학습된 모델이 사용자 PC 및 서버 환경에서 곧바로 동작할 수 있도록 지원하기 위해 가상의 리눅스 환경에

3.3.1 회귀 모델 (Linear Regression)

회귀 (Regression) 모델은 선형 방정식을 관찰 데이터를 통하여 둘 이상의 독립 변수와 하나의 종속 변수간의 관계를 모델링하는데 널리 사용된다. 회귀 모델은 대표적으로 회귀 계수 결합 형태에 따라 Linear/Non-linear regression, 독립 변수 개수에

따라 Simple/Multiple regression, 종속 변수 개수에 따라 Univariate / Multivariate regression으로 구분된다. 본 연구에서는 그중 Multivariate Linear Regression을 사용하였다. Linear Regression은 머신러닝에서 가장 일반적인 회귀분석 유형이라고 할 수 있는 회귀 분석 모델이며 이는 독립 변수와 종속 변수로 구성된다. 이러한 선형 회귀 모델은 학습과 예측이 매우 빠르며, 특성이 많을 경우에 학습이 용이하다는 장점을 갖고 있다.

3.3.2 LSTM(Long Short-term Memory)

Sequence model 이란 연속적인 입력으로부터, 연속적인 출력을 생성할 수 있는 모델이다. 주로 음성인식 혹은 자연어 처리에서 널리 사용되어지고 있다 [17]. Recurrent Neural Network (RNN)은 히든 노드가 directed edge로 연결되어 순환구조를 이루는 인공신경망의 한 종류로 순차적인 데이터를 처리를 위해 제안된 모델로 필요에 따라 유연한 구조를 만들 수 있는 장점이 있다. RNN은 관련 정보와 그 정보를 실제로 사용하는 지점의 거리가 먼 경우에 vanishing gradient 라는 문제가 발생할 수 있다. 이는 역전파시에 gradient 가 줄어, 학습 능력이 크게 저하되는 것을 의미한다. 이러한 RNN의 단점을 해결하고자, LSTM은 이러한 RNN의 단점을 개선하고자 제안된 모델로, 오래된 정보를 잊지 않고 기억할 수 있는 long-term dependencies 를 지니고 있다. 본 연구에서는 price와 quantity 각각 두 가지 값을 target으로 예측하는 모델을 사용하였다. layer의 구조는 한 개의 LSTM layer와 depth 8 linear layer로 이루어져있고 각 layer의 상세 정보는 <표 16>와 같다.

3.3.3 MLP (Multi-layer Perceptron)

표 16. LSTM과 MLP의 Layer 구성
Table 16. Layer settings for LSTM and MLP

Layer	Width	역할
1	300	input layer
2	200	hidden layer
3	100	hidden layer
4	70	hidden layer물량
5	50	hidden layer
6	30	hidden layer
7	10	hidden layer
8	1	output layer

MLP 란 여러 개의 퍼셉트론 뉴런을 다층으로 쌓은 다층신경망 구조를 의미한다. 이는 입력층과 은닉층, 출력층과 같이 3가지 이상의 layer로 이루어져 있으며 전방향 신경망으로도 불린다. 이러한 구조를 통하여 단순한 선형문제 뿐만 아니라 비선형문제도 해결할 수 있다는 것이 MLP의 핵심이다 [18, 19]. MLP에서 은닉층의 출력값에 대한 판별을 수행할 수 없기 때문에, 우리는 출력층에서 발생하는 오차값을 이용하여 은닉층으로 오차값을 역전파(back propagation)시켜서 weights를 업데이트한다. 본 연구에서는 price와 quantity 각각 두 가지 값을 target으로 예측하는 모델을 만들었고, 각 layer의 상세 정보는 <표 16>와 같다.

LSTM과 MLP에서 설정된 모델 하이퍼파라미터는 아래와 같이 동일하게 설정하였다.

- dropout=0.5
- activation function=ReLU
- optimizer=Adam,
- epoch=earlyStopping
- patience=30
- learning rate=0.0005
- decay factor=0.005

3.4 성능 평가

3.4.1 평가 지표

표 17. 실험 결과
Table 17. Experiment results

		R2 score		MAE	
		가격	물량	가격	물량
Regression	미국	-1.97464	0.25441	24.768189	2.54840
	중국	-3.57235	0.34289	54.833740	2.983888
	베트남	-3.20943	0.26897	33.328325	2.695516
LSTM	미국	-0.17736	-0.00183	0.08108	0.00347
	중국	-0.04452	-0.00003	0.08955	0.02336
	베트남	-0.14399	-0.04438	0.07368	0.00604
MLP	미국	-10.89518	-0.26407	37.860767	3.087846
	중국	-1.42455	-2.89233	36.319016	5.717388
	베트남	-0.99625	-2.66475	20.179517	5.229362

본 실험에서는 평가 지표로 MAE (Mean of absolute error)와 R2 Score를 활용하였다. 생성된 모델로 얻을 수 있는 예측 오차는 실제값과의 차이에 따라 양수 혹은 음수로 나오기 때문에, 이를 그대로 합산할 경우 크기를 산정하기 어렵다는 문제가 있다. MAE는 이를 해결하기 위해 오차에 절댓값을 취하고 데이터 수로 나눈 것으로 그 계산을 위한 식은 아래와 같다.

$$MAE = \frac{\sum_{i=1}^n |Y_i - \hat{Y}_i|}{n} \quad (1)$$

결정 계수(Coefficient of Determination 또는 R2 Score)는 추정된 선형 모형이 주어진 자료에 적합한 정도를 재는 척도이다. 선형 회귀 모델을 최적화시킨 후, 모델이 데이터에 얼마나 적합한지 확인하는 통계 방법이다. 일반적으로 0~1 사이의 값을 가지며, 상관관계가 높을수록 1에 가까워진다. 모델이 전체를 평균(mean)으로 근사하는 모델보다 성능이 안 좋다면 음의 값을 가질 수 있다. 이 지표를 통해 모델의 예측정확도를 설명하는 지표로 사용할 수 있다. R2 score의 계산을 위한 식은 아래와 같다.

$$R2 = 1 - \frac{SSR}{SST} = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (2)$$

다른 성능 지표의 경우 데이터의 스케일에 따라 값의 편차가 크므로 절댓값만으로 성능을 판단하기 어렵지만 결정계수(R2)의 경우 상대적 성능이기 때문에 종속변수에 대한 설명변수의 설명력을 직관적으로 알 수 있는 장점이 있다.

3.4.2 실험 결과

<표 17>은 각각 화장품, 의류, 전자제품에 대한 R2 score와 MAE 결과를 box-cox로 변환하여 표기한 결과이다. 이는 예측 모델을 적은수로 표현하게 되면 축약되고 함축된 표현으로 인해 학습을 하는데 문제가 생길 수 있어 4자리보다는 6자리의 HSCODE를 사용하였다.

실험 결과 Price와 Quantity를 나타내는 값은 작을수록 선호되어지고, R2 값을 클수록 선호되어진다. 또한 전반적으로 Price와 Quantity를 예측한 값의 예러가 작을수록 R2 값이 커지는 것을 확인할 수 있으며, regression 모델이 전반적으로 가장 나

은 성능을 보여주는 것을 확인할 수 있었다.

전반적으로, 우리는 LSTM이 전반적으로 전체 데이터들에 대해 잘 작동하지 않는다는 것을 확인할 수 있었다. 이는, 학습에 사용된 데이터가 12개월로 굉장히 짧고 전체 데이터의 개수가 너무 제한적이기 때문이다. 이에 반해, Regression과 MLP는 전반적으로 나은 성능을 보여주었다. R2 score와 MAE 분석 결과, 선하증권 데이터에 기반한 가격 및 물량 예측 알고리즘의 구현은 가능하며, 그 성능 역시 실제 활용 가능성을 알 수 있었다. 또한 선하증권 데이터를 보유하고 있지 않다고 하더라도 특정 품목에 대한 순차적 물량 및 가격 데이터를 보유한 경우 Regression과 MLP 모델을 활용하여 대상 국가 및 품목에 대한 물량 및 가격 예측 서비스를 제공하여 중소 수출 셀러의 수출 전략 수립에 활용할 수 있다.

3.4.3 도커라이징

2장에서는 우리는 기존의 유통물류정보를 제공하는 서비스들이 주로 내부 데이터를 기반으로 서비스를 운영하고 있어 다양한 정형, 비정형 데이터의 융합 분석을 통해 새로운 인사이트를 찾는 빅데이터 기반의 서비스 운영의 어려움을 존재함을 밝혔다. 이러한 어려움을 경감하기 위하여 우리는 본 연구를 통해 개발된 물량 및 가격 예측 모델 및 전처리 관련된 모델에 대하여 Python과 Flask를 활용하여 프로토타입 API를 구축하여 전체 프로세스에 대하여 하나의 컨테이너로 제공할 수 있도록 도커라이징 작업을 수행하여 다른 운영 환경에서도 안정적으로 운영가능한 API 제공이 가능하다. 최종적으로 도커 컨테이너를 통해 제공되는 API 리스트는 <표 18>과 같다. 제공되는 API 중 클러스터링 부분은 서론에 정의된 본 연구의 범위를 넘어서는 부분이기 때문에 본 연구에서는 생략하며

후속 연구에서 다뤄질 예정이다.

표 18. API 리스트
Table 18. A list of APIs

	서비스명	URL
1	재학습 기능	/api/retrain?parameter={검색구분}
2	이상치 탐색	/api/anomaly/y={}?parameter={검색구분}
3	클러스터링	/api/clustering?parameter={nation}
4	가격 예측	/api/prediction/price?parameter={nation, date}
5	물량 예측	/api/prediction/quantity?parameter={nation, date}

4. 결 론

본 연구에서는 중소기업의 해외시장 진출 계획을 돕기 위한 물류 빅데이터 기반 해외 시장정보 제공 시스템 구축을 위하여 기존의 물류 데이터를 제공하는 국내 플랫폼들에 대한 분석을 통해 빅데이터 및 AI 기능을 활용하여 기존의 기능을 개선하거나 새로운 기능을 추가할 수 있는 부분이 있는지에 대해 분석하였다. 이를 위해 KOTRA의 무역투자빅데이터[7], 한국무역통계진흥원(TRASS)이 제공하는 한국무역통계정보포털[8], 중소벤처기업진흥공단(KOSME)에서 제공하는 고비즈 코리아[9], 한국무역협회(KITA)에서 제공하는 K-STAT[10]을 분석 대상으로 선정하였다. 그 결과 분석 대상 서비스 모두 각 서비스를 운영하는 주체들이 보유한 내부 통계 데이터를 기반으로만 서비스가 구축되어 있어 데이터의 공유 전략 수립이 필요하며, 제공되는 AI나 빅데이터 기반의 기능 역시 빈도수 세기 기반의 기초적인 활용에 그치고 있어 머신러닝 및 딥러닝 기술을 활용한 AI 및 빅데이터 기능의 성능 개선에 대한 시도가 필요함을 확인하였다.

이에 본 연구에서는 기존 서비스들과의 차별성

확보를 위해 수출입 통계 정보를 바탕으로 국가별 물량 및 가격의 예측치를 판단하는 머신 러닝 알고리즘을 설계하고 성능 평가를 통해 본 연구에 활용한 데이터에 특성과 맞지 않은 LSTM을 제외하고 Regression과 MLP 모델을 활용하여 해당 예측 모델을 구축할 수 있음을 확인하였다. 마지막으로, 본 연구에서 활용한 전처리 및 모델의 공유를 통해 국내 전반의 유통물류정보 서비스의 AI 및 빅데이터 기능 활성화를 위하여 도커를 통한 API 서비스를 구축하였다. 구축된 API는 ETRI 내부의 다른 API 서비스들과 함께 2022년 하반기 Open API 플랫폼 형태로 외부에 오픈이 예정되어 있어 추후 빅데이터 연구자 및 기업들이 직접 활용할 수 있을 것으로 기대된다.

References

- [1] Contactless economy, <https://www.si.re.kr/node/65175>, Jan. 2022.
- [2] Online shopping trends in 2021 and overseas direct sales and purchase trends in the forth quarter, https://kostat.go.kr/portal/korea/kor_nw/1/12/3/index.board?bmode=read&bSeq=&aSeq=416587&pageNo=1&rowNum=10&navCount=10&currPg=&searchInfo=&sTarget=title&sTxt, Feb. 2022.
- [3] H. H. Yoon, Cross border purchase was the largest ever led by K-beauty, https://biz.chosun.com/site/data/html_dir/2020/10/18/2020101800759.html, Jan. 2022.
- [4] Difficulty in the process of pioneering overseas market, https://kosis.kr/statHtml/statHtml.do?orgId=453&tblId=DT_393001_066&conn_path=12, Jan. 2022.
- [5] S. J. Choi, *Bigdata based advertising techniques*, Korea Life Insurance Association, Seoul, 2019.
- [6] K. J. Han, M. Yi, and J. E. Kim, *Search personalization in folksonomy by exploiting multiple and temporal aspects of user profiles*, IEEE Access, Vol. 7, pp. 65610-95619, 2019.
- [7] Trade investment big data, <http://www.kotra.or.kr/bigdata/bhrcMarket>, Jul. 2021.
- [8] Korea trade statistics information portal, <https://www.bandtrass.or.kr/index.do>, Jul. 2021.
- [9] Gobiz Korea, <https://kr.gobizkorea.com/kruser/main.do>, Jul. 2021.
- [10] K-Stat, <https://stat.kita.net/newMain.screen>, Jul. 2021.
- [11] K. J. Han, J. Y. Park, D. H. Hong, J. S. Shin, and M. Y. Yi, *P-download: A new personalization approach for a content-based search system*, Proceedings of International Conference on Emerging Databases, pp. 1-12, 2013.
- [12] Maximum performance of expert platform 'GoBiz Korea', <https://www.mk.co.kr/news/business/view/2021/12/1147948/>, Dec. 2021.
- [13] H. J. V. Meter, *Revising the DIKW pyramid and the real relationship between data, information, knowledge, and wisdom*, Law, Technology and Humans, Vol. 2, No. 2, pp. 69-80, 2020.
- [14] K. J. Han, E. G. Lee, H. W. Jae, and M. Y. Yi, *Introducing experiential knowledge platform: a smart decision supporter for field experts*, Proceedings of International

- Conference on Big Data and Smart Computing, pp. 404-407, 2016.
- [15] P. Suhan, Z. Li, and J. Dai, *An improved TextRank keywords extraction algorithm*, Proceedings of the ACM Turing Celebration Conference-China, pp. 1-7, 2019.
- [16] H. Zhao, J. Lu, and J. Cao, *A short text conversation generation model combining BERT and context attention mechanism*, International Journal of Computational Science and Engineering, Vol. 23, No. 2, pp. 136-144, 2020.
- [17] Y. R. Oh, K. Y. Park, H.B. Jeon, and J. G. Park, *Automatic proficiency assessment of Korean speech read aloud by non-natives using bidirectional LSTM-based speech recognition*, ETRI Journal, Vol. 42, No. 5, pp. 761-772, 2020.
- [18] A. Farhadipour, H. Veisi, M. Asgari, and M. A. Keyvanrad, *Dysarthric speaker identification with different degrees of dysarthria severity using deep belief networks*, ETRI Journal, Vol. 40, No. 5, pp. 643-652, 2018.
- [19] M. G. Kim, and M. H. Lee, *A Study on rainfall estimation of high resolution using machine learning*, Journal of Knowledge Information Technology and Systems, Vol. 12, No. 2, pp. 295-303, 2017.
- [20] J. H. Moon, and Y. S. Won, *An investigation of personality traits and self-esteem in workers*, Journal of Knowledge Information Technology and Systems, Vol. 16. No. 5, pp. 1023-1033, 2021.

기존 유사 서비스 기능 분석을 통한 차별화된 물류 데이터 분석 플랫폼 설계 및 구현

한기준¹, 정훈², 나동길³

¹ETRI 지능화용융합연구소 연구원

²ETRI 우정물류기술연구센터 센터장

³ETRI 지능화용융합연구소 책임연구원

요 약

비대면 경제의 급격한 성장에도 불구하고 중소셀러를 위한 맞춤형 마케팅 정보 및 판매 전략 제공 서비스는 미흡한 실정이며, 중소기업의 해외시장 진출 계획을 돕기위한 수출입 데이터 등의 물류 데이터 기반 해외 시장정보 제공 시스템 구축을 위한 설계는 필수적인 상황이다. 이에 본 연구에서는 기존의 물류 데이터를 제공하는 국내 플랫폼들에 대한 분석을 통해 빅데이터 및 AI 기능을 활용하여 기존의 기능을 개선하거나 새로운 기능을 추가할 수 있는 부분이 있는지에 대해 분석하였고, 그 결과 기존의 서비스들은 여러 내부 및 외부 데이터들을 활용하여 데이터 공유 전략과 추가적인 머신러닝 솔루션을 도입할 필요성이 있음을 보였다. 또한 기존 서비스와의 차별성 확보를 위하여 수출입 통계 데이터에 기반으로 대상 품목의 물량 및 가격의 예측치를 판단하는 머신 러닝 알고리즘을 설계하였다. 실험을 통해 LSTM 모델은 전체 데이터에 대해 잘 동작하지 않지만 선형 회귀 및 MLP 모델은 전반적으로 양호한 성능을 보였다. 실험을 통해, 본 연구에서 제안된 모델은 수출 대상 국가 및 품목에 대한 물량 및 가격 예측 서비스를 제공하는 데 사용할 수 있음 확인하였다. 본 연구에서 구축된 전처리 프로세스 및 모델은 향후 공공 API 서비스로 도거화하여 연구자 및 개발자들이 적극적으로 활용할 수 있도록 하였다.

감사의 글

이 논문은 산업통상자원부가 지원한 '지식서비스산업기술개발사업'으로 지원을 받아 수행된 연구 결과입니다. [과제명: 수출 중소셀러를 위한 해외 시장정보 분석 시스템 개발 / 과제고유번호: 20014772]



Keejun Han received the bachelor's degree in the Department of Information technology from the Australian National University in 2010. He received the M.S. degree and the Ph.D. degree in the Graduate School of Knowledge Service Engineering from KAIST in 2013 and 2019, respectively. He is currently a researcher in the Intelligent Convergence Research Laboratory at ETRI. Before joining ETRI, he was a staff engineer at Samsung Electronics. He is a member of the KKITS.

E-mail address: keejun@etri.re.kr

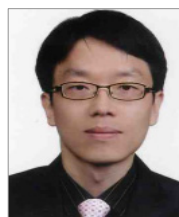
Research Laboratory at ETRI. He is a member of the KKITS.

E-mail address: dgna@etri.re.kr



Hoon Jung received the Ph.D. degree in industrial engineering from University of Missouri in 2001. He joined Electronics and Telecommunications Research Institute (ETRI) in 2002. He is in charge of Drone Delivery Research and Development as head of the Postal Logistics Technology Research Center since 2017.

E-mail address: hoonjung@etri.re.kr



Dong-gil Na received the bachelor's degree in the Department of Industrial Engineering from the Chonbuk National University in 1998. He received the M.S. degree and the Ph.D. degree in the Graduate School of Industrial Engineering from Chonbuk National University in 2000 and 2004, respectively. From 2005, he was a senior researcher at ETRI. He is currently a principal researcher in the Intelligent Convergence