

「2025년 관세청 공공데이터 활용 경진대회」 아이디어 기획서

아이디어명	딥러닝 기반 자동 밀수입 의심 거래 탐지 시스템
-------	----------------------------

1) 아이디어 기획 개요

비지도 학습 기반 이상탐지 모델을 활용한 밀수입 자동 감지 시스템 설계

본 아이디어는 통관 데이터에 숨겨진 이상 거래를 자동으로 감지하여, 다양한 고위험 품목의 밀수입(밀반입)을 사전에 식별할 수 있는 시스템을 구축하는 데 목적이 있다. 비지도 학습 기반의 이상탐지 기법을 활용함으로써, 기존 정상 거래 흐름과 뚜렷하게 벗어난 수입 패턴을 사전 정의된 라벨 없이도 자동 탐지할 수 있다. 이 시스템은 중량 대비 단가, 운송 비용, 수입 국가 등 세부 통관 데이터를 기반으로, 정형화된 정상 흐름과 다른 이상 거래 신호를 감지하도록 설계된다.

1.1 주요 설계 요소

1) HSK 코드 및 품목별 수입 데이터 정제 및 분류

- 수산물 항목을 중심으로, 품목별 구조화된 데이터를 활용하여 품목별 특성과 가격대 범위를 반영

2) 중량 대비 단가, 운송 비용, 무역 국가 등을 고려한 이상값 탐지

3) 비지도 학습 기반 탐지 모델 적용

- 시계열성을 고려한 LSTM 기반 AutoEncoder 모델을 활용하여, 정상 거래의 재구성 오차를 기반으로 이상 여부 판단

4) 거시경제 요인 반영을 통한 정밀도 향상

- 국제 유가, 환율, 소비자물가지수(CPI), 생산자물가지수(PPI) 등 외부 경제 변수를 통합하여, 전체적인 가격 흐름과 시장 여건을 함께 고려

1.2 확장성과 기대효과

- 1) 본 시스템은 수산물 데이터를 중심으로 초기 개발되었으며, 향후에는 마약, 담배, 의약품 등 고위험 품목에도 동일한 방식으로 적용 가능
- 2) 관세청 등 유관 기관이 활용 시, 통관 단계에서 이상거래 자동 감지 및 실시간 알림 기능으로 운용 가능
- 3) 인력 중심의 사후 적발 방식에서 벗어나, 데이터 기반의 사전 탐지 체계로 전환함으로써 효율성과 신속성 제고
- 4) 범죄 사전 억제 효과와 더불어, 국가 차원의 수입 관리 시스템 고도화 및 불법 거래 방지에 실질적 기여

2) 아이디어 기획 세부 내용

2.1 아이디어 동기 및 목적

최근 들어 선박과 해상 화물을 이용한 다양한 형태의 밀수입 수법이 늘어나고 있다. 2024년 4월에는 멕시코발 선박에서 코카인 2톤이 적발됐고, 지난 5년간 해상 마약 밀수 적발 건수는 20건, 총 적발량은 3톤에 이르는 등 밀수입의 규모와 빈도가 모두 증가하는 추세에 있다. 특히, 수산물 분야에서는 킹크랩 5,400kg를 밀수입한 사례를 비롯해, 언더인보이싱(under-invoicing), 무관세 회피, 원산지 허위 신고 등 점점 더 지능화되고 조직적으로 이루어지는 밀수 방식이 계속해서 발견되고 있다. 또한 수산물이나 금속류처럼 가격 변동성이 큰 품목의 경우, 비정상적으로 낮은 단가로 수입되는 덤핑 의심 거래가 밀수입 수법과 결합돼 나타나는 사례도 많다. 예를 들어, 고의로 수입 단가를 낮춰 신고하는 언더인보이싱 방식이 덤핑과 유사한 형태로 나타나는 경우가 대표적이다. 이처럼 단순한 가격 경쟁을 넘어 시장 질서를 교란하고 통관 시스템의 사각지대를 노리는 거래는, 밀수입과 함께 정밀하게 탐지할

필요가 있다.

밀수입의 고도화에 대응하기 위해 관세청은 ROV(수중 비디오 촬영 장치), GPS 탐지기 등 물리적 장비를 활용하는 한편, 빅데이터 분석이나 국제 공조 체계 등 과학 기반의 대응 방식도 점차 확대하고 있다. 하지만 이러한 대응 역시 여전히 인력과 예산에 대한 의존도가 크고, 사후 단속 중심의 구조에서 벗어나지 못해, 사전 탐지의 정밀성과 실시간 대응에는 한계가 있다.

이에 본 아이디어는, 비지도 학습 기반의 이상 거래 탐지 모델을 통해 통관 단계에서 수산물 등 고위험 품목의 밀수입 가능성을 자동으로 감지할 수 있는 시스템을 설계하는 것을 목표로 한다. 이 시스템은 HSK코드 기준으로 품목별 수입 데이터를 정제한 뒤, 정상 거래의 흐름을 학습해, 중량 대비 단가, 무역 국가별 운임비 등 다양한 통관 정보와 거시경제변수를 바탕으로 사전 정의된 라벨 없이도 이상 거래를 식별할 수 있도록 구성된다.

본 분석에서는 공공데이터 이용에 한계가 있어 수산물 데이터를 중심으로 다루었지만, 실제 운영 단계에서는 수산물에 국한되지 않고 다양한 품목군에 적용할 수 있을 것으로 기대된다.

2.2 아이디어 핵심 기술 및 특징

본 시스템의 가장 큰 특징은 라벨링된 학습 데이터나 감시 인력 없이도 자가 학습을 통해 의심 거래를 자동으로 탐지할 수 있는 것이다. 전통적인 지도학습 기반의 이상탐지 방식과 달리, 비지도 학습 기반의 딥러닝 모델을 활용하여 정상적인 통관 흐름에서 벗어난 이상 거래를 사전 정의된 기준 없이 스스로 식별할 수 있도록 설계했다.

1) LSTM 기반의 Autoencoder를 적용한 이상 탐지 모델

수입 거래는 계절성, 가격 변동성, 운송 비용 변화 등 다양한 시계열적 특성을 가지므로, 본 연구는 시계열 데이터의 장기 의존성(Long-term dependency)을 학습하는 데 강점을 지닌 LSTM(Long Short-Term Memory) 구조와, 입력 데이터를 압축·복원하며 핵심 패턴을 파악할 수 있는 AutoEncoder(AE) 모델을 결합한 LSTM 기반 AutoEncoder 모델을 활용하고자 한다. 이 모델은 정상 거래 데이터를 학습하여 패턴을 압축·복원하고, 재구성 오차가 기준 이상인 경우 이상 거래로 판단한다.

실제로 정세연·김상철(2022)의 연구 「LSTM 오토인코더를 활용한 축산 환경 시계열 데이터의 이상치 탐지: 경계값 설정에 따른 성능 비교」는 LSTM AutoEncoder를 활용한 축산 환경 시계열 데이터 이상치 탐지에서 Dynamic Threshold 방식으로 99.66%의 높은 정확도를 기록하며 본 모델의 효과를 입증한 바 있다. 이처럼 LSTM과 AutoEncoder의 결합은 정상 패턴 복원 과정에서 이상 여부에 민감하게 반응하여 수산물 수입 거래의 정상·이상 거래 구분에 적합하다. 따라서 시계열 구조와 이상 패턴 복원 능력을 동시에 반영할 수 있는 LSTM 기반 AutoEncoder를 본 연구의 최종 모델로 선정하였다.

2) 다양한 데이터 이용

모델의 학습에는 단순 가격 정보만이 아니라 다양한 통관 및 거래 정보가 활용된다. 주요 변수로는 다음과 같은 정형 데이터를 포함한다:

- 수입 국가
- HSK 세부 품목 코드
- 중량 대비 단가
- 국가별 운임비

이러한 구조화된 데이터에 국제 유가, 환율, 소비자물가지수(CPI), 생산자물가지수(PPI)와 같은 주요 거시경제 지표를 함께 반영한다. 이를 통해 단순한 거래 패턴뿐 아니라, 글로벌 경제 여건에 따른 가격 변동과 왜곡 현상도 모델이 학습하고 고려할 수 있도록 구성했다. 다음의 거시경제 변수를 채택한 배경에 대해서는 아이디어 기획 세부 내용에서 자세하게 설명할 것이다.

3) 시계열 예측 및 이상 거래 탐지 성능 검증

본 시스템은 2020년 5월부터 2025년 4월까지의 데이터를 활용해 학습용 데이터와 평가용 데이터로 구분한 뒤 모델을 구축하였다. 이를 통해 실제 시계열 예측 성능과 이상 거래 탐지 정확도를 함께 검

증할 수 있다.

특히, 관세청이 공개한 데이터는 보안 및 법적 제약으로 인해 실제 밀수입 사례나 이상 거래가 포함되어 있을 가능성이 낮으므로 일반적으로 자주 나타나는 밀수입 수법을 고려하여 직접 모의 이상치 데이터를 구축했다. 이를 위해 첫째, 단가, 유가, 국가 등을 기준으로 운송비가 비정상적으로 높은 거래를 추가하였고, 둘째, 특정 품목이나 국가의 일반적인 거래 단가와 비교했을 때 지나치게 낮은 가격으로 수입된 거래를 일부러 포함시켰다. 이는 실제 거래 금액보다 낮게 신고하는 '언더인보이싱'이 의심되는 사례를 시뮬레이션 데이터에 반영하기 위함이다. 셋째, 원래 고가에 거래되는 품목군 안에서 유독 단가가 낮게 책정된 사례도 데이터에 넣었다. 이는 같은 품목 내에서 평균적인 단가 수준과 크게 벗어난 이례적인 거래를 재현해, 정상적인 단가 분포에서 벗어난 이상 거래의 탐지를 가능하게 하고자 했다.

4) X-ray 이미지 기반 다중모달 시스템으로 확장 가능성

향후 확장 기능으로, 비정형 데이터인 X-ray 영상 이미지와 정형 데이터를 통합 분석하는 다중모달 기반 딥러닝 시스템도 고려할 수 있을 것이다. 물류 통관 과정에서 촬영된 X-ray 이미지를 바탕으로, 특정 업체나 품목에 대해 반복적으로 나타나는 포장 구조, 구성 밀도 등을 학습한 후, 정형 정보(HSK 코드, 수량, 무게 등)와 이미지 간 불일치가 발생하는 경우 자동으로 이상 거래로 탐지하는 방식이다. 예를 들어, 동일 HSK코드의 품목임에도 과거와 비교해 포장 밀도나 내부 구성 구조가 현저히 다를 경우, 기존 정상 패턴과의 차이를 기준으로 위험 거래로 판단하게 된다. 이와 같은 영상 기반 탐지 기술은 정형 데이터만으로는 식별하기 어려운 고차원 이상 징후를 포착하는 데 유효하며, 통관 검사 효율을 더욱 높이는 역할을 할 수 있을 것으로 기대된다.

이러한 기술적 구조를 기반으로 한 본 시스템을 통해 실제 밀수입 및 덤핑 의심 거래를 사전에 탐지하고, 고위험 품목에 대한 정밀 대응하며 궁극적으로는 데이터 기반의 통관 정책 고도화에 기여하고자 한다.

3) 관세청 공공데이터 등 데이터 활용 방안

3.1 활용 데이터 개요

본 분석에서는 데이터의 최신성과 연속성을 확보하기 위해 2020년 5월부터 2025년 4월까지의 기간에 해당하는 데이터를 수집하여 활용하였다. 이는 시계열 기반 이상치 탐지 모델 설계에 필요한 충분한 데이터 길이와 계절성 확보, 그리고 최근 추세 반영을 위한 적절한 기간 설정을 목적으로 한다. 각 데이터의 정의, 출처 및 내용은 다음과 같다.

3.2 관세청 공공데이터 활용 계획

1) 운송비 데이터

운송비는 수산물의 수입 원가를 구성하는 주요 요소 중 하나로, 해상 운송을 중심으로 분석하였다. 이를 통해 운송비의 변동이 수입 단가 및 물동량에 미치는 영향을 정밀하게 파악하고자 하였다.

내용:

- 관세청_해상수출입(수입) 운송비용_2025.04.30
- 기반 자료: 수입신고 시 운임 + 화물정보
- 대상: 국가·지역 단위 항로, FOB 조건
- 단위: 천 원/2TEU
- 적재형태: 단일 화주 적재된 FCL(Full Container Load)
- 컨테이너 종류: 40피트(2TEU) 일반 화물용(GP)
- 활용: 해상 물류비 추이 분석 및 업계 참고자료

2) 수출입 데이터

수산물 수출입 정보는 관세청 수출입신고 데이터를 바탕으로 해양수산부에서 국가별·품목별 월별 통계로 집계한 자료이다.

내용:

- 수산물은 HSK 코드 기준으로 구분되며, 월별 수입 중량(kg)과 금액(USD) 제공
- 주요 변수: 기준년월, 국가, 품목, 중량, 금액 등
- 활용 목적:
 - 수입 급증·단가 변화 감지
 - 이상 거래 탐지 (예: 밀수입 의심)
 - 외부 요인(운송비, 유가 등)과의 연관성 분석

3.3 타 기관 공공데이터 및 민간데이터 활용 계획

1) 국제 유가

국제 유가는 Dubai, Brent, WTI 등 원유의 국제 시장에서의 거래 가격을 의미하며, 주로 운송비 및 원자재 수입 원가에 큰 영향을 미치는 대표적인 외생 변수이다. 특히 한국을 포함한 아시아 국가들이 수입하는 원유의 상당수가 중동산 원유이기 때문에, 두바이유 가격은 국내 원유 수입단가 및 유가 관련 비용의 변동을 가장 잘 설명할 수 있는 지표로 사용된다. 따라서 본 연구에서는 한국의 수입 비용 및 물가에 미치는 영향을 보다 정확하게 반영하기 위해 두바이유를 국제유가의 대표 지표로 선택하였다.

근거 : 김창범(2018)의 연구 「국내경기와 가격변수가 우리나라 수산물 수입물동량에 미치는 영향」에 따르면, 유가는 수산물 수입물동량에 대해 장기 및 단기 모두에서 유의미한 양(+)의 영향을 미치는 것으로 나타났다.

2) 환율

환율은 원화 대비 외국 통화(예: 달러, 위안, 엔)의 상대 가치를 나타내며, 수입 원가 산정과 가격 책정에 직결되는 주요 거시경제 변수이다.

근거 : 차혜경(2007)의 연구 「수입물가의 환율전가 상승요인분석」에 따르면, 환율이 수입물가에 미치는 영향은 시간이 지날수록 점차 커지고 있으며, 특히 단기와 장기 모두에서 환율 전가율이 상승하는 경향을 보였다.

3) 소비자물가지수 (CPI)

소비자물가지수는 가계가 구입하는 상품 및 서비스의 평균 가격 변동을 측정한 지수로, 국민 생활 물가 수준을 대표하는 경제 지표이다.

4) 생산자물가지수 (PPI)

생산자물가지수는 국내 생산자가 공급하는 상품의 도매 가격 변동을 나타내는 지표로, 산업 원가 및 공급 측 물가 수준을 반영한다.

근거 : 이상환, 정승익, 송건호(2025)의 연구 「머신러닝을 이용한 수산물 가격 예측 연구」에 따르면, 연어 가격은 소비자물가지수(CPI)와 생산자물가지수(PPI) 모두와 강한 양(+)의 상관관계를 가지는 것으로 나타났다. 이는 수산물 가격이 거시경제 지표의 변동에 민감하게 반응한다는 점을 보여준다.

본 연구에 활용된 수출입, 운송비 데이터, 국제 유가, 생산자물가지수, 소비자물가지수, 환율 등은 모두 공공기관에서 지속적으로 제공되는 자료로, 정기적인 업데이트와 높은 접근성을 기반으로 향후 데이터 획득의 지속성이 매우 높다. 또한 이러한 데이터들은 다양한 산업 및 경제 분야에서 활용될 수 있어 분석의 범용성 또한 확보된다.

데이터 가공 과정에서는 다음과 같은 전처리를 수행하였다.

먼저, 수출입 데이터의 HS코드는 기본적으로 10자리 체계를 기준으로 하며, 일부 9자리 코드는 앞에 '0'을 붙여 모두 10자리로 통일하였다.

수출입구분코드는 'E'(수출)와 'I'(수입)로 구성되어 있으나, 본 분석에서는 국내로 반입된 수산물에 초점을 맞추기 위해 'I'(수입) 항목만을 사용하였다.

해상 운송비 데이터에서는 국가명이 '중국', '일본', '미국', '베트남', '유럽연합'으로만 제공되므로, 분석 대상 국가는 이들로 한정하였고, 유럽연합 소속 국가(예: 그리스, 네덜란드 등)는 모두 '유럽연합'으로 재분류하였다. 또한, 미국의 경우 해상 운송비가 동부와 서부로 구분되어 제공되므로, 해당 수치를 평균 내어 하나의 '미국' 값으로 통합하였다.

최종 데이터에는 기준년월, 국가코드, HSK 품목코드, 수출입구분코드 및 수출입구분명, 국가명, 경제권명, 수산물수출입품목명, 당월 수출입 중량과 미화금액, 당해 누계 수출입 중량과 미화금액, 데이터 기준일자, 파일 기준년월, 지역, HSK 품목코드 자리수, 기간, 그리고 운송비 칼럼만을 포함하였다. 이후 환율, 국제유가, 소비자물가지수, 생산자물가지수 등 주요 거시경제 변수를 순차적으로 병합하였다.

4) 아이디어 실현 가능성

4.1 아이디어의 구현 및 실현 가능성

본 아이디어는 관세청 및 유관기관이 보유한 수산물 수입 통계, 국가별 운송비, 환율, 국제유가 등 다양한 정형 데이터를 활용하며, 데이터 수집과 접근이 가능한 상태여서 실현 가능성이 매우 높다. 기존 이상탐지 알고리즘(AutoEncoder, Isolation Forest, LOF 등)을 기반으로 개발 가능하며, 수입 거래 패턴에 맞춘 룰 기반 탐지 로직도 함께 적용할 수 있다. 데이터 분석 인프라는 Python의 Pandas, scikit-learn, PyOD 등 오픈소스 라이브러리만으로 구축할 수 있고, 과거 수입 데이터만으로도 충분한 시뮬레이션이 가능하다. 기술적으로는 단기적으로 이상 거래 탐지 모델에서 의심 케이스 알림 시스템 연동으로 발전 가능하며, 중장기적으로는 관세청 통관 시스템 내 모듈 형태로 내재화해 자동화할 수 있다. 또한 기존 통계 기반 단속과 병행 운용이 가능해 기존 시스템을 대체하지 않고도 적용할 수 있다는 점에서 높은 실현성을 가진다.

4.2 아이디어의 시장성 및 실제 사업 가능성

본 시스템은 고수익을 노린 밀수입이 증가하는 현 상황에서, 기존 수작업 위주의 단편적 감시 방식에서 벗어나 다변량 기반 이상 거래 탐지 로직을 적용하여 조기 위험관리를 가능하게 한다. 데이터 기반 자동화 탐지를 통해 업무 효율성 증대, 인력 부담 완화, 운영비 절감 효과도 기대할 수 있다. 나아가 관세청뿐만 아니라 식품의약품안전처, 세관, 해양경찰 등 다양한 정부 기관과 민간 물류·보안 기업 등으로의 사업 확장 가능성이 크다. 이에 따라 관련 기술 및 서비스의 상용화와 확산을 통해 신규 시장 창출과 관세 행정 혁신의 계기를 마련할 수 있을 것으로 기대된다.

4.3 아이디어 구체적 세부 내용(앱/웹 등 서비스 구현 예시 화면 및 설명)

1. 데이터 및 분석 개요

총 데이터 수: 62,955건

Train 데이터 기간: 2020년 5월부터 2024년 4월까지 (약 4년, 51,622건)

Test 데이터 기간: 2024년 5월부터 2025년 4월까지 (1년, 11,333건)

Train/Test 데이터를 나눈 기준은 시점 기준의 시계열 일반화 성능 평가를 위함이다. 2024년 5월 1일을 기준으로 분리한 이유는, 최근 데이터(2024년 5월~2025년 4월)를 실제 운영 환경에 대응하는 검증 데이터로 삼기 위해서이며, 과거 4년간의 데이터를 학습에 활용함으로써 충분한 계절성 및 추세 학습이 가능하도록 구성하였다.

2. 이상치 생성

본 프로젝트에서는 2024년 5월부터 2025년 4월까지 1년간의 테스트 데이터에 대해 총 100건의 인위적 이상치를 생성하여 이상 탐지 모델의 성능 검증에 활용하였다. 이상치는 크게 운송비 이상치와 단가 이상치로 구분하며, 실제 밀수입에서 발생 가능한 다양한 비정상 거래 유형을 반영하였다.

- 운송비 이상치 (총 20건)

- 1) 단가 대비 높은 운송비 (10건)** : 실제 단가 대비 운송비를 1.5배로 인위 조작하여, 정상적인 거래 대비 과도한 운송비 지출 패턴을 생성하였다.
- 2) 유가와 무관하게 급등한 비 (5건)** : 운송비를 유가의 20~30배 범위 내에서 무작위로 높여, 운송비가 유가와 상관없이 비정상적으로 변동하는 상황을 구현하였다.
- 3) 지역별 kg당 운송비 급등 (5건)** : 지역 및 월별 kg당 평균 운송비를 기준으로 3배 수준의 운송비를 설정하여, 특정 지역에서의 과도한 운송비 지출을 반영하였다.

- 단가 이상치 (총 80건)

- 1) 단가 인위 하락 (50건)** : 정상 단가 대비 15~30% 하락 30건, 30~70% 하락 10건, 70~80% 하락 10건으로 분류하여 언더인보이싱 가능성을 모사하였다.
- 2) 저가 품목 코드에 고가 단가 적용 (20건)** : 하위 20% 저가 품목 코드에 상위 20% 고가 품목의 평균 단가를 부여하여, 품목 코드와 단가가 일치하지 않는 이상 거래를 생성하였다.
- 3) 고가 품목 코드에 저가 단가 적용 (10건)** : 상위 20% 고가 품목 코드에 하위 20% 저가 품목 평균 단가를 적용하여, 단가와 품목 코드 간 불일치를 통한 이상치를 구성하였다.

이와 같은 다각적 이상치 제작 방식을 통해 실제 밀수입 사례에서 나타날 수 있는 다양한 비정상 거래 유형을 모사함으로써, 모델의 이상 탐지 능력과 현장 적용 가능성을 종합적으로 평가할 수 있다.

3. 데이터 모델링 과정

이상치 탐지 모델 개발 과정에서는 다양한 실험을 통해 최적의 데이터 구성과 모델 구조를 탐색하였다. 먼저, 이상치 판단 기준인 Threshold 설정 방식은 MSE(Mean Squared Error)를 기반으로 상위 5% 또는 10%를 이상치로 판단하는 직관적인 방법, IQR 기반 방식, 선행 논문에서 제안된 Dynamic Threshold 방식, 그리고 Threshold 튜닝 방식을 비교실험하였다.

데이터 시계열 구성 방식은 단순 3차원 배열 방식과 슬라이딩 윈도우 기법을 적용해 품목별로 12개월 단위 시퀀스를 구성하는 방식을 실험하였으며, LSTM의 장점을 살리기 위해 후자인 슬라이딩 윈도우 방식이 선택되었다. 입력 피처 구성에서는 수치형 변수만 사용하는 경우와 HSK 코드를 임베딩 레이어를 통해 포함시키는 경우를 비교하였다.

시계열 길이의 경우, 6개월과 12개월 중 12개월이 더 안정적인 성능을 보여 선택하였으며, LSTM 구조는 단방향이 양방향보다 더 나은 성능을 보였다. 이외에도 Dropout, Batch Normalization, 뉴런 수와 층 수 조정 등 다양한 구조 변경 실험을 병행하였다.

정규화 방식은 MinMaxScaler와 RobustScaler를 모두 시도했다.

추가적으로 Isolation Forest, LOF, EllipticEnvelope 등 전통적인 이상치 탐지 모델들과 룰 기반 방식도 실험해보았으나, 딥러닝 기반 Autoencoder보다 성능이 낮아 채택되지 않았다.

이러한 실험을 바탕으로 최종적으로 두 가지 모델을 도출하였다. 모델 A는 MinMaxScaler 기반의 LSTM Autoencoder에 상위 10% 기준을 적용한 방식이며, 모델 B는 RobustScaler를 사용하고 HSK 코드 임베딩을 포함하며 IQR 기준을 활용한 방식이다.

```
[ ] time_steps = 12
def build_sequences(df, time_steps):
    sequences = []
    end_indices = []
    for code, group in df.groupby('HSK품목코드'):
        group = group.sort_values('기준년월')
        data = group[features].values
        indices = group.index.values
        if len(data) >= time_steps:
            for i in range(len(data) - time_steps + 1):
                window = data[i:i+time_steps]
                end_idx = indices[i] + time_steps - 1
                sequences.append(window)
                end_indices.append(end_idx)
    return np.array(sequences), end_indices

train_df = df[df['기준년월'] < '2024-05-01']
test_df = df[df['기준년월'] >= '2024-05-01']

X_train, train_end_idx = build_sequences(train_df, time_steps)
X_test, test_end_idx = build_sequences(test_df, time_steps)
test_df_seq = test_df.loc[test_end_idx].reset_index(drop=True)

def create_lstm_autoencoder(input_shape):
    model = models.Sequential()
    model.add(layers.LSTM(64, activation='relu', input_shape=input_shape, return_sequences=False))
    model.add(layers.RepeatVector(input_shape[0]))
    model.add(layers.LSTM(64, activation='relu', return_sequences=True))
    model.add(layers.TimeDistributed(layers.Dense(input_shape[1])))
    model.compile(optimizer='adam', loss='mse')
    return model

input_shape = X_train.shape[1:]
model = create_lstm_autoencoder(input_shape)
model.fit(X_train, X_train, epochs=50, batch_size=64, validation_data=(X_test, X_test), verbose=0)
```

[그림 1] 모델 A

[그림 2] 모델 B

두 모델의 예측 결과를 실제 이상치 데이터와 비교한 결과, 모델 A는 운송비 이상치 20건 중 14건, 단가 이상치 80건 중 49건을 정확히 탐지하였다. 모델 B는 운송비 이상치 20건 중 11건, 단가 이상치 80건 중 58건을 탐지하였다. 즉, 모델 A는 운송비 이상치 탐지에서, 모델 B는 단가 이상치 탐지에서 더 우수한 성능을 보였다.

이처럼 다양한 실험과 조합을 통해 도출된 두 모델은 실제 이상 거래 탐지에 있어 높은 재현율을 보여주며, 실무 적용 가능한 이상치 탐지 시스템으로서의 가능성을 입증하였다.

이상치 탐지 모델을 바탕으로 다음과 같은 밀수입 위험 품목 예측 서비스를 제안한다.

[밀수입 위험 품목 예측 시스템]

분석 기준 설정

분석 기간: 2024-01

품목명(HSK코드):

분석 대상국: 전체 / 운송비 이상치 / 단가 이상치 / X-ray 이미지 이상

분석 대상품: 운송비 이상, 단가 이상

이상 범위 수준: 상위 0.5%

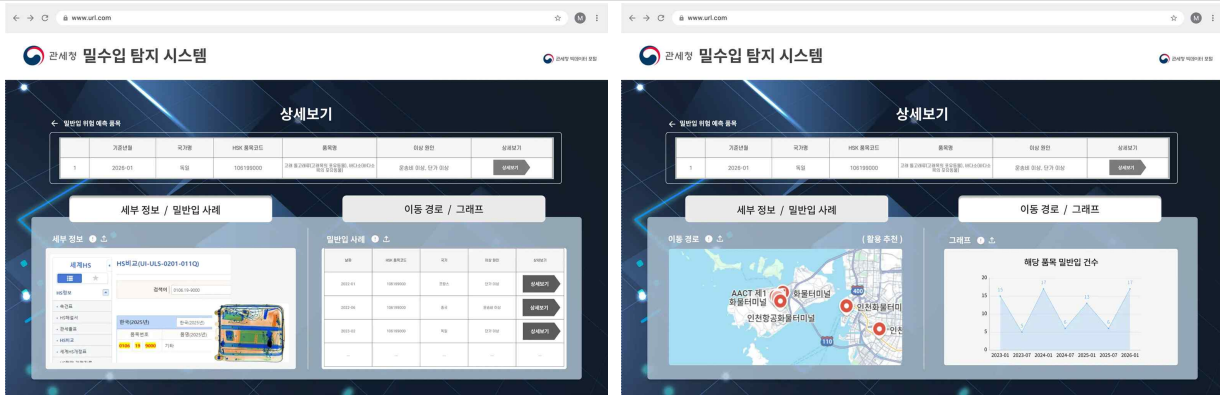
가격 범위: 0원 ~ 999원

일반입 위험 예측 품목

번호	기준년월	국가명	HSK 품목코드	품목명	이상 원인	상세보기
1	2024-01	독일	106199000	2차 날고리(하)고리(하)의 모(단)형: HGS199000	운송비 이상, 단가 이상	상세보기
2	2024-01	일본	106199000	가라 부유물	운송비 이상	상세보기
...
21	2024-01	오스트리아	2501009010	식염	단가 이상	상세보기
22	2024-01	오스트리아	2501009010	식염	운송비 이상	상세보기

[그림 3] 일반입 위험 예측 품목 조회 : 밀수입 의심 거래 예측 가능

- 특정 기간, 분석 대상국, 품목(HSK코드), 예측 기준, 이상 범위 수준, 가격 범위 등을 설정
- 설정한 기준을 바탕으로 일반입 위험 예측 품목 조회 가능



-> 밀반입 위험 예측 품목 상세 조회 : 밀수입 의심 품목 상세 조회 가능

[그림 3-1] 해당 품목의 정보 조회 가능

- 해당 품목의 세부 정보 : 관세율, 화물 X-ray 이미지
- 과거 밀반입 사례 : 밀수입으로 분류된 원인 (ex. 단가/운송비 이상) 조회 가능

[그림 3-2] 해당 품목과 관련된 추가 정보 조회 (이후 확장 및 적용 가능)

- 이동 경로 : 주요 수입국, 경유지 정보 제공
- 시각화 그래프 : 과거 경향성 파악 가능

5) 기존 서비스와 차별성 및 독창성

본 딥러닝 기반 자동 밀수입 의심 거래 탐지 시스템은 기존 관세청의 인공지능 기반 탐지 방식과 비교해 구조적·기술적으로 뚜렷한 차별성을 지닌다. 현재의 관세 행정 시스템은 주로 과거 적발 사례를 바탕으로 한 지도학습 방식에 의존하고 있어, 새로운 수법이 등장할 경우 대응력이 떨어지고 데이터 편향의 한계 또한 존재한다. 반면 본 시스템은 사전 라벨이 없는 비지도 학습 기반으로 작동함으로써, 과거에 존재하지 않았던 형태의 밀수입 거래에도 선제적으로 탐지할 수 있는 유연한 구조를 갖추고 있다.

또한 기존 AI 모델이 물품명이나 설명 등 텍스트 정보 중심으로 위험도를 판단하는 데 반해, 본 시스템은 거래 단가, 중량단가, 운송비, 관세율, HSK 코드 등 다양한 수치형 데이터를 중심으로 탐지 모델을 구성한다. 이로 인해 단순한 품목 판별을 넘어 가격 조작(언더인보이싱)이나 고관세 회피와 같은 정교한 밀수 전략까지 식별할 수 있으며, 시계열 분석 기법을 통해 시간 흐름에 따른 수입 단가의 이상 변화까지 반영함으로써 더욱 고도화된 판별을 할 수 있다.

본 시스템은 다음과 같이 기존의 방식과 구분되는 독창성과 창의적 요소를 지닌다:

1. 지도학습 대비 유연성 높은 비지도 학습 구조

- 과거 사례에 국한되지 않고 새로운 유형의 밀수 수법에 선제적으로 대응 가능
- 데이터 편향 및 과적합의 위험을 낮춰 다양한 상황에 유연하게 적용 가능

2. 정형 수치 데이터를 활용한 다변량 이상 거래 탐지

- 단가, 중량, 운송비, 관세율 등 다양한 수치형 정보를 기반으로 탐지
- 고도화된 정량적 판별을 통해 단순 텍스트 분석보다 신뢰도 향상

3. 범용성과 확장성을 갖춘 플랫폼 구조

- 본 분석에서 적용된 수산물 분야를 넘어 농산물, 생활용품, 담배, 의약품 등 다양한 품목과 국가에 적용 가능
- 장기적으로는 다양한 목적에 활용 가능한 통합형 탐지 시스템까지 발전 가능

4. 비정형·시계열 패턴까지 대응하는 정교한 이상 탐지 체계

- LSTM 기반의 Autoencoder 알고리즘을 통해 명확한 규칙에 부합하지 않는 거래도 탐색 가능

-
- 단발성 이상뿐만 아니라 점진적·복합적 패턴 변화까지 포착 가능
 - 계절성과 시기별 단가 변동 등 시계열적 특징을 반영해 더욱 정교한 탐지 가능

5. AI 기반 자동화로 인한 효율성과 비용 절감 효과

- 기존의 인력 중심 대응 방식 대비 분석 속도와 정확도를 높일 수 있으며, 인적 자원 부담을 경감
- 운영 효율성을 높여 지속 가능한 관세행정 운영 기반 마련

이러한 점에서 본 시스템은 단순히 기존 탐지 방식을 보완하는 수준을 넘어서, 관세 행정 전반의 대응 방식을 전환할 수 있는 기반 기술로 작용할 수 있을 것이다. 궁극적으로 공공 부문이 이러한 기술을 적극적으로 도입할 경우, 행정 효율성 향상뿐만 아니라 AI 기반 업무 혁신의 실질적인 사례로 이어질 수 있을 것이다. 이는 향후 다른 공공 영역에서의 AI 기술 확산과 민간 기술 발전을 촉진하는 계기가 될 수 있을 것이다.

6) 기대 효과

본 시스템은 수입 통관 과정 전반에 AI 기반의 자동화 탐지 체계를 도입함으로써, 기존의 인력 중심 사후 적발 위주의 밀수 대응 체계를 데이터 기반의 선제적 대응 구조로 전환하는 데 핵심적인 역할을 할 수 있다. 이를 통해 관세청과 유관 기관의 밀수 대응 체계는 기술적으로 한층 정밀하고 신속한 구조로 고도화될 수 있으며, 자원 운용 효율성과 행정 체계 전반의 역량 향상에도 기여할 수 있을 것이다.

특히, 본 시스템은 AI 기반의 비지도 학습 모델을 활용해 감시 인력과 물리적 장비에 대한 의존도를 낮추면서도 실시간 탐지 및 이상거래 알림 기능을 제공할 수 있다. 이는 반복적인 수작업 판독을 줄이고, 이상 거래를 조기에 식별하고 대응할 수 있으므로 관세청의 인력 부담을 크게 줄이고 행정 효율을 높이는 효과를 줄 수 있을 것이다. 또한 다음과 같은 사회적 가치를 창출할 수 있을 것이다.

1. 통관 행정의 정밀도 및 효율성 향상

- 중량 대비 단가, 운송 방식별 운임, 수입 경로 등 정형 통관 데이터를 분석
- 국제유가, 환율, 소비자물가지수(CPI), 생산자물가지수(PPI) 등 거시경제 변수를 통합하여 경제 상황에 따른 가격 왜곡 가능성에 대해서도 반영 가능
- 시계열 기반 LSTM AutoEncoder를 통해 거래 흐름의 변화까지 고려한 탐지 가능

2. 대응 품목의 확장성과 시스템 유연성 확보

- 다양한 품목군에 대한 적용을 염두에 둔 모델 구조
- 품목별 HS코드 구조, 거래 이력, 운임·단가 기준 등을 반영한 유연한 모델 설계
- 품목별 이상 탐지 기준을 동적으로 반영함으로써 다품목 통합 감시 시스템 구현 가능

본 시스템은 수산물 데이터를 기반으로 개발되었지만, 동일한 구조를 활용해 마약, 담배, 의약품, 귀금속 등 다양한 고위험 품목으로도 확장 적용이 가능하다. 이를 통해 특정 품목에 한정된 단속을 넘어, 수입 물류 전반에 대한 포괄적인 감시 체계를 마련할 수 있을 것이다.

3. 기술 일자리 창출 및 산업적 파급 효과

- 데이터 분석, 머신러닝 모델링, 리스크 진단 등의 전문 인력 수요 확대
- 민간 기술 기업 및 스타트업의 시장 진입 기회 제공
- 관세·무역 특화 데이터 기반 서비스 산업 생태계 형성

본 시스템은 나아가 산업과 고용 전반에도 긍정적인 파급 효과를 유발할 수 있을 것으로 기대된다. 자동 탐지 시스템의 상용화와 확대 적용 과정에서, AI 모델 개발, 데이터 분석, 시스템 운영 유지보수

등 다양한 기술 기반 직무에 대한 수요가 발생할 수 있다. 특히 공공 부문 중심의 수입 통제 시스템에 간 기술 솔루션이 도입된다면, 민간 스타트업과 기술 기업들이 참여할 수 있는 데이터 기반 관세 솔루션 시장이 형성되고, 스타트업 육성에도 도움이 될 수 있을 것이다.

이처럼, 본 시스템은 관세 행정의 구조적인 전환, 데이터 산업의 성장, 일자리 창출, 공정 무역 질서 확보 등 여러 측면에서 실질적인 효과를 기대할 수 있다. 이러한 긍정적인 효과는 밀수입에 따른 세수 손실을 줄이고, 고위험 거래로 인한 시장 불안을 예방하는 데 도움을 주며, 궁극적으로 지속 가능하고 효율적인 관세 행정 환경을 만드는 데 기여할 수 있을 것이다.

※ **기획서 작성 시 유의사항**

- 자유양식으로 작성하되, **분량은 10 페이지 이내로 작성**
- 작성시 도표, 이미지 영상 등 활용 가능

※ **AI 활용 시 최대 5점 가점 부여**