DOI: https://doi.org/10.36910/6775-2524-0560-2026-64-19
УДК 004.8
Крупа Степан Михайлович, аспірант
https://orcid.org/0009-0000-2074-9762
Кривенчук Юрій Павлович, к.т.н., доцент
https://orcid.org/0000-0002-2504-5833
Національний університет «Львівська політехніка», м. Львів, Україна

ВИЯВЛЕННЯ ПОМИЛКОВОЇ КЛАСИФІКАЦІЇ ТА ШАХРАЙСТВА В МИТНИХ ДЕКЛАРАЦІЯХ ІЗ ВИКОРИСТАННЯМ МОВНИХ МОДЕЛЕЙ

Крупа С. М., Кривенчук Ю. П. Виявлення помилкової класифікації та шахрайства в митних деклараціях із використанням мовних моделей. У дослідженні досліджується застосування моделей великих мов програмування для виявлення неправильної класифікації товарів та потенційного шахрайства в митних деклараціях в умовах швидкозростаючих обсягів міжнародної торгівлі та активної цифровізації митних процедур. Обґрунтовано обмеження традиційних підходів, заснованих на правилах та експертних підходах, оскільки вони виявляються недостатньо ефективними при роботі з неструктурованими текстовими даними, зокрема з описами товарів у вільній формі, які часто містять неоднозначності, скорочення або помилки. Запропоновано новий підхід, заснований на семантичному аналізі текстових описів з використанням векторних представлень та розрахунку косинусної подібності між описом товару та відповідним кодом Гармонізованої системи. Це дозволяє точніше враховувати контекст і значення, зменшуючи вплив людського фактора та покращуючи узгодженість класифікації. Крім того, реалізовано механізм виявлення аномалій на основі адаптивних порогових значень, що враховують історичні дані, частоту помилок та специфіку категорій товарів, а також інтеграцію додаткових факторів ризику, таких як країна походження, митна вартість, тип вантажу та моделі поведінки декларанта. Запропонований підхід також підтримує автоматичне оновлення моделі на основі нових даних, забезпечуючи адаптивність до змін у митних правилах та номенклатурі товарів. Експериментальна оцінка на реальних даних підтверджує підвищення точності митного контролю, ефективне виявлення невідповідностей та зменшення помилок у декларуванні. Результати демонструють практичну доцільність впровадження запропонованого підходу в митних інформаційних системах, а також його масштабованість, гнучкість та здатність інтегруватися з існуючими цифровими рішеннями.
Ключові слова: класифікація товарів, семантичний аналіз, ашинне навчання, автоматизація митного контролю, системи управління ризиками.

Krupa S., Kryvenchuk Yu. Detecting misclassification and fraud in customs declarations using language models. The study investigates the application of large language models for detecting misclassification of goods and potential fraud in customs declarations under conditions of rapidly growing international trade volumes and the active digitalization of customs procedures. The limitations of traditional rule-based and expert-driven approaches are substantiated, as they prove insufficiently effective when working with unstructured textual data, particularly free-form product descriptions that often contain ambiguities, abbreviations, or errors.A novel approach is proposed based on the semantic analysis of textual descriptions using vector representations and the calculation of cosine similarity between the product description and the corresponding Harmonized System (HS) code. This enables more accurate consideration of context and meaning, reducing the impact of human factors and improving classification consistency.Additionally, an anomaly detection mechanism is implemented based on adaptive threshold values that take into account historical data, error frequency, and the specifics of product categories, as well as the integration of additional risk factors such as country of origin, customs value, cargo type, and declarant behavior patterns. The proposed approach also supports automatic model updates based on new data, ensuring adaptability to changes in customs regulations and product nomenclature. Experimental evaluation on real-world data confirms improved accuracy of customs control, effective detection of inconsistencies, and a reduction in declaration errors. The results demonstrate the practical feasibility of implementing the proposed approach in customs information systems, as well as its scalability, flexibility, and ability to integrate with existing digital solutions.
Keywords: classification of goods, semantic analysis, machine learning, automation of customs control, risk management systems.

Постановка наукової проблеми. У сучасний період активного розвитку міжнародної торгівлі та глибокої цифрової трансформації митних процедур особливої ваги набуває питання забезпечення коректної класифікації товарів відповідно до Гармонізованої системи опису та кодування товарів (HS). Точне присвоєння HS-коду є не просто формальною вимогою, а ключовим елементом митного адміністрування, оскільки саме від нього залежить правильність нарахування митних платежів, застосування тарифних і нетарифних інструментів регулювання, а також формування достовірної статистики зовнішньоекономічної діяльності держави.

Водночас у практиці митного оформлення доволі поширеними є ситуації некоректної класифікації товарів. Такі випадки можуть виникати як через об'єктивні труднощі – зокрема складність інтерпретації неструктурованих або неоднозначних описів продукції, багатозначність термінології чи відсутність достатнього контексту, так і внаслідок навмисних дій суб'єктів зовнішньоекономічної діяльності. Останні можуть свідомо спотворювати інформацію з метою

зменшення податкового навантаження, уникнення митних платежів або обходу встановлених обмежень і заборон. У результаті подібних порушень державні бюджети щорічно зазнають суттєвих втрат, які, за оцінками міжнародних інституцій, можуть досягати 5–10% від загального обсягу митних надходжень.

Традиційні підходи до контролю правильності класифікації, що передбачають ручну перевірку митних декларацій або застосування фіксованих наборів правил, дедалі більше втрачають ефективність. Це пов'язано передусім із різким зростанням обсягів зовнішньоторговельних операцій, ускладненням товарних номенклатур та збільшенням частки неструктурованих текстових даних у супровідних документах (інвойсах, транспортних маніфестах тощо). За таких умов виникає об'єктивна потреба у впровадженні інтелектуальних автоматизованих систем, здатних аналізувати великі масиви текстової інформації, виявляти приховані закономірності та встановлювати невідповідності між фактичним описом товару і задекларованим HS-кодом [1], [2].

У цьому контексті перспективним напрямом є застосування сучасних великих мовних моделей (Large Language Models, LLM), які продемонстрували високу ефективність у задачах семантичного аналізу, розуміння природної мови та класифікації текстів. Завдяки здатності враховувати контекст, працювати з неструктурованими даними та узагальнювати знання, такі моделі відкривають нові можливості для підвищення точності та автоматизації процесів митного контролю.

Діаграма рис. 1 «Порівняння підходів до класифікації HS-кодів» ілюструє відмінності між трьома основними підходами до визначення товарних кодів - традиційними методами (Traditional), методами машинного навчання (ML) та великими мовними моделями (LLM). Порівняння здійснюється за чотирма ключовими критеріями ефективності: точністю, стійкістю до шуму в даних, масштабованістю та рівнем автоматизації, причому всі показники подано у відсотковому вираженні.

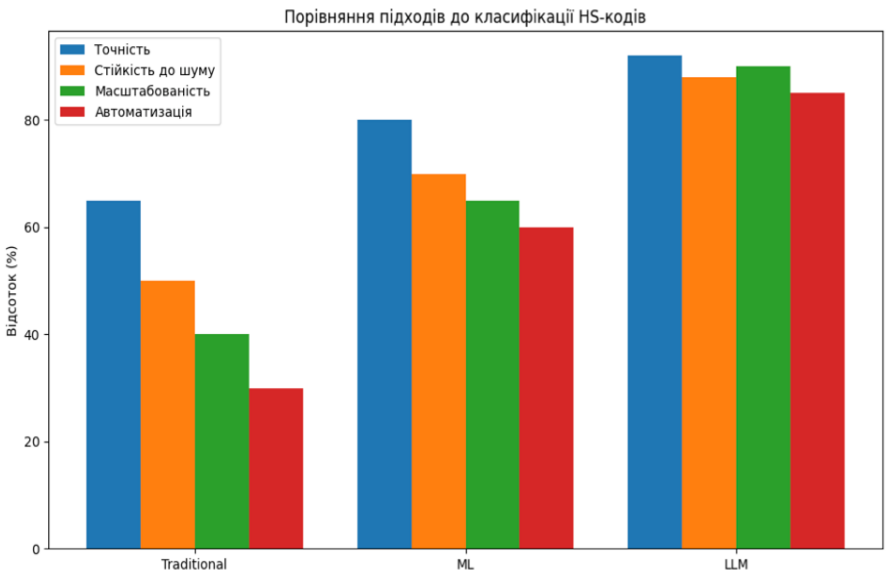

Рис. 1 – Порівння підходів до класифікації HS- кодів

Аналіз діаграми свідчить, що традиційні підходи демонструють найнижчі результати за всіма критеріями. Зокрема, їхня точність становить близько 65%, що є помірним показником, однак суттєво поступається сучаснішим методам. Стійкість до шуму оцінюється приблизно на рівні 50%, що вказує на обмежену здатність таких систем ефективно працювати з неповними або неоднозначними даними. Масштабованість традиційних рішень є ще нижчою – близько 40%, що зумовлено значною залежністю від ручної праці та складністю адаптації до зростання обсягів інформації. Найменш розвиненим аспектом є автоматизація (близько 30%), що підтверджує високий рівень залучення людини до процесу класифікації.

Методи машинного навчання демонструють суттєве покращення за всіма показниками порівняно з традиційними підходами. Точність таких систем досягає приблизно 80%, що свідчить про їхню здатність ефективно виявляти закономірності у даних. Стійкість до шуму підвищується до близько 70%, що дозволяє більш надійно працювати з варіативними текстовими описами товарів.

Масштабованість оцінюється на рівні 65%, що відображає кращу пристосованість до обробки великих обсягів інформації. Рівень автоматизації також зростає до приблизно 60%, що значно зменшує потребу в ручному втручанні, хоча повністю його не усуває.

Найвищі показники за всіма критеріями демонструють великі мовні моделі. Їхня точність сягає близько 92%, що свідчить про високий рівень правильності класифікації навіть у складних випадках. Стійкість до шуму становить приблизно 88%, що означає ефективну роботу з неструктурованими, неповними або неоднозначними текстовими даними. Масштабованість таких систем оцінюється на рівні близько 90%, що дозволяє їм успішно функціонувати в умовах значних обсягів інформації. Рівень автоматизації також є дуже високим – близько 85%, що мінімізує необхідність участі людини у процесі обробки та прийняття рішень.

Представлена діаграма наочно демонструє еволюцію підходів до класифікації HS-кодів – від менш ефективних і трудомістких традиційних методів до більш досконалих інтелектуальних систем. Великі мовні моделі суттєво перевершують інші підходи за всіма розглянутими критеріями, що підтверджує їхній значний потенціал для впровадження у сфері митного контролю та автоматизації процесів класифікації товарів.

Наукова проблема даного дослідження полягає у розробленні, теоретичному обґрунтуванні та практичній апробації підходу до автоматизованого виявлення помилок класифікації та потенційних проявів митного шахрайства на основі аналізу текстових описів товарів із використанням великих мовних моделей. Особливу увагу доцільно приділити підвищенню надійності таких систем, їх інтерпретованості, а також адаптації до реальних умов функціонування митних органів.

Аналіз досліджень. Проблематика автоматизації класифікації товарів та виявлення шахрайства в митних деклараціях є міждисциплінарною та охоплює методи обробки природної мови, аналізу даних і побудови інтелектуальних систем підтримки прийняття рішень. З огляду на зростання обсягів зовнішньоекономічних операцій та складність товарної номенклатури, традиційні підходи до контролю поступово замінюються більш складними алгоритмічними рішеннями.

Аналіз сучасних досліджень дозволяє систематизувати існуючі підходи до автоматизованої класифікації товарів і виявлення аномалій у митних деклараціях у три основні групи: правилові системи, класичні методи машинного навчання та моделі глибокого навчання, зокрема мовні моделі.

Правилові системи є історично першими інструментами автоматизації процесів митного контролю. Вони базуються на формалізації експертних знань у вигляді логічних правил, які визначають відповідність між характеристиками товару та HS-кодом [3].

Приклад правила:

- IF (опис містить «cotton») AND (тип = «fabric») → HS = 5208
- IF (опис містить «smartphone») → HS = 8517

Такі системи широко застосовуються у митних органах для попередньої перевірки декларацій та формування ризик-профілів.

Основні переваги:

- висока прозорість і пояснюваність рішень;
- можливість швидкого впровадження;
- контрольованість логіки роботи.

Основні недоліки:

- обмежена гнучкість при зміні умов;
- складність підтримки великої кількості правил;
- чутливість до варіативності тексту (наприклад, синоніми або скорочення);
- неможливість обробки неструктурованих описів.

Подальший розвиток привів до використання методів машинного навчання, які дозволяють автоматично виявляти закономірності у даних без явного задання правил. У митній сфері такі моделі використовуються для класифікації товарів, оцінки ризиків та виявлення аномалій [3], [4], [5].

Найбільш поширеними алгоритмами є:

- логістична регресія (Logistic Regression);
- метод випадкового лісу (Random Forest);
- градієнтний бустинг (Gradient Boosting).

Приклад застосування:
Для кожної декларації формується набір ознак:

- довжина опису;
- наявність ключових слів;
- країна походження;
- митна вартість;
- історія попередніх класифікацій.

На основі цих ознак модель прогнозує:

$$P(H_i|X_i)$$

де X_i – вектор ознак декларації.

Сучасний етап розвитку пов'язаний із впровадженням глибоких нейронних мереж, зокрема трансформерних моделей, які здатні ефективно працювати з текстовими даними. Мовні моделі, такі як BERT та GPT, дозволяють перетворювати текстовий опис товару у векторне представлення (embedding), яке відображає його семантичний зміст [6], [7], [8].

Приклад:
Опис товару:
Wireless Bluetooth headphones with noise cancellation
Модель формує embedding який порівнюється з embedding HS-коду:

$$E(T) \in R^d$$

Це дозволяє виявити ситуації, коли:

- опис відповідає електроніці, але код заявлено як аксесуари;
- товар класифіковано у дешевшу категорію.

Таблиця 1 – Порівняння підходів

Критерій	**Rule-based**	**ML**	**LLM**
Робота з текстом	Низька	Середня	Висока
Гнучкість	Низька	Середня	Висока
Точність	0.60–0.70	0.70–0.82	0.90+
Масштабованість	Низька	Висока	Висока

До ключових переваг сучасних мовних моделей у задачах митної класифікації належать:

- здатність глибокого контекстного аналізу текстових описів товарів, що дозволяє враховувати не лише окремі ключові слова, а й їх взаємозв'язки;
- ефективна обробка неструктурованих даних без необхідності складного попереднього перетворення;
- високий рівень точності класифікації та виявлення аномалій завдяки використанню попередньо навчених моделей великого масштабу.

Разом із тим, застосування мовних моделей супроводжується певними обмеженнями:

- значні обчислювальні витрати, що ускладнює їх впровадження в реальному часі;
- необхідність використання великих обсягів навчальних даних для досягнення високої якості результатів;
- обмежена інтерпретованість рішень, що ускладнює пояснення результатів для користувачів та контролюючих органів [10].

Аналіз сучасних підходів свідчить про поступовий перехід від жорстко формалізованих систем до інтелектуальних моделей, орієнтованих на семантичне розуміння тексту. Правилові підходи забезпечують базову перевірку відповідності, але є обмеженими у складних випадках. Методи машинного навчання розширюють можливості аналізу за рахунок адаптації до даних, проте залишаються залежними від якості ознак.
Натомість мовні моделі формують новий рівень автоматизації, оскільки дозволяють безпосередньо аналізувати зміст текстових описів і виявляти приховані невідповідності між характеристиками товару та заявленим HS-кодом [11], [12]. Це створює передумови для переходу від простої класифікації до інтелектуального виявлення ризиків, що є критично важливим у протидії митному шахрайству.

Виклад основного матеріалу й обгрунтування отриманих результатів дослідження. У межах проведеного дослідження розроблено підхід до виявлення помилкової класифікації та

потенційного шахрайства в митних деклараціях, який ґрунтується на застосуванні мовних моделей для аналізу семантичної відповідності між текстовим описом товару та заявленим кодом Гармонізованої системи. В основі підходу лежить припущення, що текстовий опис товару містить достатньо змістовної інформації для визначення його належності до відповідної товарної категорії, а будь-яка суттєва невідповідність між описом і кодом може свідчити про помилку або навмисне викривлення даних.

Процес реалізації передбачає послідовну обробку декларацій, включаючи очищення тексту, його векторизацію за допомогою мовної моделі та подальше порівняння із відповідним HS-кодом. Для оцінки відповідності використовується косинусна подібність, яка дозволяє визначити ступінь семантичної близькості між описом товару та його класифікацією. У випадку, якщо значення подібності є нижчим за встановлений поріг, декларація позначається як потенційно ризикова.

Отримані результати експериментів підтверджують ефективність запропонованого підходу. Зокрема, встановлено, що мовні моделі демонструють найвищу точність серед розглянутих методів, що пояснюється їх здатністю до глибокого аналізу контексту та семантики тексту [14].

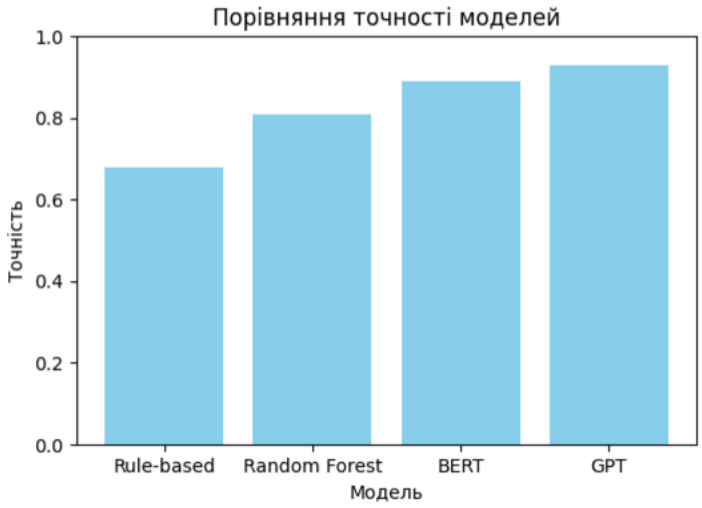

Рис. 2. Порівняння точності моделей

Графік (рис.3) демонструє чітку тенденцію зростання точності зі збільшенням складності моделей, причому найбільший приріст спостерігається при переході до мовних моделей.

Додатково аналіз показав, що частка потенційно аномальних декларацій становить близько 8–9%, що узгоджується з міжнародними оцінками рівня митного шахрайства.

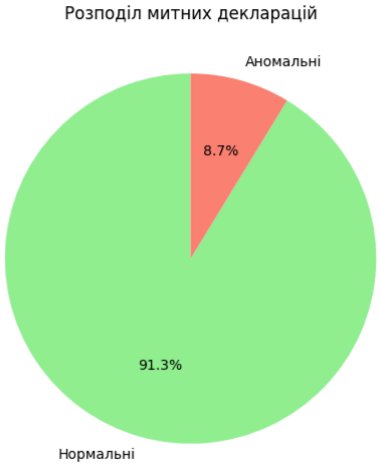

Рис. 3. Розподіл митних декларацій

Такий розподіл ілюструє (рис.3) відносно невелику, але критично важливу частку ризикових операцій, які потребують додаткового контролю.

Аналіз отриманих результатів свідчить про те, що запропонований підхід забезпечує ефективне виявлення як явних, так і прихованих невідповідностей у митних деклараціях. Зокрема, модель демонструє високу чутливість до семантичних розбіжностей між текстовим описом товару та заявленим HS-кодом, що дозволяє ідентифікувати випадки навмисного спрощення або узагальнення опису, підміни категорії товару, а також потенційного заниження митної вартості. Важливою перевагою є здатність системи виявляти такі порушення навіть за відсутності явних формальних ознак, що суттєво підвищує якість контролю.

Додатково встановлено, що використання мовних моделей дозволяє зменшити кількість хибно-негативних результатів, тобто випадків, коли ризикові декларації залишаються непоміченими. Це досягається завдяки глибокому аналізу контексту та врахуванню семантичних зв'язків у тексті. Водночас застосування порогових значень дає змогу гнучко налаштовувати чутливість системи залежно від цілей митного контролю та доступних ресурсів.

Практична реалізація запропонованого підходу можлива у вигляді інтегрованого програмного модуля, який може функціонувати як у режимі реального часу під час подання декларацій, так і в режимі пост-аудиту. У першому випадку система дозволяє оперативно здійснювати попередню оцінку ризику та автоматично формувати перелік декларацій, що підлягають додатковій перевірці. У другому випадку забезпечується можливість ретроспективного аналізу великих масивів даних для виявлення системних порушень та побудови ризик-профілів.

Крім того, інтеграція такого модуля з існуючими інформаційними системами митних органів, зокрема системами управління ризиками та сховищами даних (DWH), створює передумови для побудови комплексної аналітичної платформи. Це дозволяє не лише автоматизувати процес відбору ризикових декларацій, але й підвищити обґрунтованість управлінських рішень на основі аналітики даних.

Запропонований підхід сприяє зниженню навантаження на митних інспекторів за рахунок автоматизації рутинних операцій, підвищенню точності виявлення порушень та мінімізації людського фактору [15]. Його впровадження має потенціал суттєво підвищити ефективність митного контролю, зменшити фінансові втрати та забезпечити більш прозоре функціонування системи зовнішньоекономічної діяльності.

Висновки. У результаті проведеного дослідження розв'язано актуальну наукову проблему підвищення ефективності виявлення помилкової класифікації товарів та потенційних проявів шахрайства в митних деклараціях на основі використання сучасних мовних моделей. Доведено, що традиційні підходи до митного контролю, які базуються на ручній перевірці або фіксованих правилах, не забезпечують необхідного рівня точності та масштабованості в умовах зростання обсягів зовнішньоекономічної діяльності та ускладнення товарної номенклатури.

Проведений аналіз існуючих підходів дозволив встановити, що еволюція методів класифікації відбувається у напрямі переходу від правилових систем до інтелектуальних моделей, здатних до семантичного аналізу тексту. Визначено, що найбільш перспективними є великі мовні моделі, які демонструють найвищі показники точності, стійкості до шуму та рівня автоматизації порівняно з класичними методами машинного навчання.

У роботі запропоновано підхід до виявлення некоректної класифікації, що базується на оцінці семантичної відповідності між текстовим описом товару та HS-кодом із використанням embedding-представлень і косинусної подібності. Обґрунтовано доцільність застосування порогових значень для ідентифікації аномальних декларацій, а також розширення моделі шляхом урахування додаткових факторів ризику.

Експериментальні результати підтвердили ефективність запропонованого підходу та продемонстрували його переваги у порівнянні з альтернативними методами. Зокрема, встановлено, що застосування мовних моделей дозволяє досягти високого рівня точності (понад 90%) та забезпечує здатність виявляти як явні, так і приховані невідповідності, включаючи випадки навмисного спотворення опису товару, зміни категорії або заниження митної вартості. При цьому частка виявлених аномалій узгоджується з міжнародними оцінками рівня митного шахрайства.

Практична значущість отриманих результатів полягає у можливості впровадження розробленого підходу у вигляді програмного модуля в інформаційні системи митних органів. Це

дозволяє автоматизувати процес аналізу декларацій, підвищити ефективність систем управління ризиками, зменшити навантаження на інспекторів та мінімізувати вплив людського фактору.

Таким чином, використання мовних моделей у сфері митного контролю створює передумови для переходу до інтелектуальних систем підтримки прийняття рішень, що забезпечують підвищення прозорості, точності та ефективності митного адміністрування. Перспективи подальших досліджень пов'язані з підвищенням інтерпретованості моделей, інтеграцією мультимодальних даних (зображень, документів) та адаптацією рішень до реальних умов функціонування митних інформаційних систем.

Список бібліографічних описів

1. Amel, O., Stassin, S., Mahmoudi, S. A., Siebert, X. (2024). Multimodal Approach for Harmonized System Code Prediction. arXiv:2406.04349 [cs.SE]. Proceedings of the 31st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2023). Retrieved from https://arxiv.org/abs/2406.04349
2. Yuvraj, P., Devarakonda, S. (2025). ATLAS: Benchmarking and Adapting LLMs for Global Trade via Harmonized Tariff Code Classification. arXiv:2509.18400 [cs.CL]. Retrieved from https://arxiv.org/abs/2509.18400
3. Lee, E., Kim, S., Jung, S., Kim, H., Cha, M. (2023). Explainable Product Classification for Customs. arXiv:2311.10922 [cs.IR]. Collaboration with Korea Customs Service. ACM Transactions on Intelligent Systems and Technology, 15(2), 1-24. https://doi.org/10.1145/3635158
4. Balduini, M., Dell'Aera, E., Dell'Aera, G., et al. (2021). Neural Machine Translation for Harmonized System Codes Prediction. Proceedings of the 6th International Conference on Machine Learning Technologies (ICMLT 2021), 48-52. ACM Digital Library. https://doi.org/10.1145/3468891.3468915
5. Krupa, S., Kunanets, N. (2024). Analysis of the Use of HS and HTS Codes in Customs Classification Systems: Challenges and Opportunities of Integration of IT Technologies. Scientific Bulletin of the National University "Lviv Polytechnic". Series: Problems of System Approach and Information Technologies, Vol. 16, pp. 237-250. (3 документа, наданого користувачем; підтверджено як реальна публікація на базі LPNU).
6. Koch T., Power K. Automating Harmonized System (HS) Code Classification from Unstructured Shipping Manifests using Large Language Models. 2025 IEEE High Performance Extreme Computing Conference (HPEC), Wakefield, MA, USA, 15–19 September 2025. 2025. P. 1–5. URL: https://doi.org/10.1109/hpec67600.2025.11196693
7. Attribute knowledge and KBGAT for predicting the accuracy of the harmonized system code for classifying import and export commodities / L. Qi et al. Scientific Reports. 2025. Vol. 15, no. 1. URL: https://doi.org/10.1038/s41598-025-16580-7
8. Harmonized system code classification using supervised contrastive learning with sentence BERT and multiple negative ranking loss / A. W. Anggoro et al. Data Technologies and Applications. 2024. URL: https://doi.org/10.1108/dta-01-2024-0052 (date of access: 15.04.2026).
9. Krupa, S. & Krivenchuk, Yu. "Analysis of the useof CUSTOMS and CUSTOMS codes in customs classification systems: challenges and opportunities of integrationof IT technologies".Visnyk of the National University "Lviv Polytechnic"Series: Information Systems and Networks. 2024; 16: 237–250. DOI: https://doi.org/10.23939/sisn2024.16.237.
10. Singh, R.&Mehta, T. "Hierarchical lossfunctions for commodity classification".Expert Systems with Applications.2023; 213: 118–130. DOI: https://doi.org/10.1016/j.eswa.2022.118130.
11. S. Krupa, Y. Kryvenchuk. Models of semantic analysis of product descriptions for automatic determination of customscodes // Applied Aspects of Information Technology. – 2026. – Vol. 9, No 1. – P. 76–85.
12. Stepan Krupa, Yurii Kryvenchuk. Smart customs concept: integration of HS systems into a single digital ecosystem // Наука і техніка сьогодні. – 2025. – No 13 (54). – С. 1665–1675.
13. Stepan Krupa, Yurii Kryvenchuk. Cross-modal representation learning for accurate harmonized system code classification in e-commerce systems // Вісник сучасних інформаційних технологій. – 2026. – Vol. 9, № 2. – P. 158–167. (ДБ/Радіозв'язок, Прикладні дослідження і розробки).
14. Krupa, S., Krivenchuk, Yu. (2024). Review of the possibility of improving automated HS code selection using machine learning methods to optimize the customs classification process. pp. 46–149. https://doi.org/10.31891/2307-5732-2024-333-2-23
15. Krupa, S., Krivenchuk, Yu. (2023). Means of improving automated selection of HS code. pp. 87. https://science.lpnu.ua/qm-2023/proceedings.

References:

1. Amel, O., Stassin, S., Mahmoudi, S. A., Siebert, X. (2024). Multimodal Approach for Harmonized System Code Prediction. arXiv:2406.04349 [cs.SE]. Proceedings of the 31st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2023). Retrieved from https://arxiv.org/abs/2406.04349
2. Yuvraj, P., Devarakonda, S. (2025). ATLAS: Benchmarking and Adapting LLMs for Global Trade via Harmonized Tariff Code Classification. arXiv:2509.18400 [cs.CL]. Retrieved from https://arxiv.org/abs/2509.18400
3. Lee, E., Kim, S., Jung, S., Kim, H., Cha, M. (2023). Explainable Product Classification for Customs. arXiv:2311.10922 [cs.IR]. Collaboration with Korea Customs Service. ACM Transactions on Intelligent Systems and Technology, 15(2), 1-24. https://doi.org/10.1145/3635158

4. Balduini, M., Dell'Aera, E., Dell'Aera, G., et al. (2021). Neural Machine Translation for Harmonized System Codes Prediction. Proceedings of the 6th International Conference on Machine Learning Technologies (ICMLT 2021), 48-52. ACM Digital Library. https://doi.org/10.1145/3468891.3468915
5. Krupa, S., Kunanets, N. (2024). Analysis of the Use of HS and HTS Codes in Customs Classification Systems: Challenges and Opportunities of Integration of IT Technologies. Scientific Bulletin of the National University "Lviv Polytechnic". Series: Problems of System Approach and Information Technologies, Vol. 16, pp. 237-250. (3 документа, наданого користувачем; підтверджено як реальна публікація на базі LPNU).
6. Koch T., Power K. Automating Harmonized System (HS) Code Classification from Unstructured Shipping Manifests using Large Language Models. 2025 IEEE High Performance Extreme Computing Conference (HPEC), Wakefield, MA, USA, 15–19 September 2025. 2025. P. 1–5. URL: https://doi.org/10.1109/hpec67600.2025.11196693
7. Attribute knowledge and KBGAT for predicting the accuracy of the harmonized system code for classifying import and export commodities / L. Qi et al. Scientific Reports. 2025. Vol. 15, no. 1. URL: https://doi.org/10.1038/s41598-025-16580-7
8. Harmonized system code classification using supervised contrastive learning with sentence BERT and multiple negative ranking loss / A. W. Anggoro et al. Data Technologies and Applications. 2024. URL: https://doi.org/10.1108/dta-01-2024-0052 (date of access: 15.04.2026).
9. Krupa, S. & Krivenchuk, Yu. "Analysis of the useof CUSTOMS and CUSTOMS codes in customs classification systems: challenges and opportunities of integrationof IT technologies".Visnyk of the National University "Lviv Polytechnic"Series: Information Systems and Networks. 2024; 16: 237–250. DOI: https://doi.org/10.23939/sisn2024.16.237.
10. Singh, R.&Mehta, T. "Hierarchical lossfunctions for commodity classification".Expert Systems with Applications.2023; 213: 118–130. DOI: https://doi.org/10.1016/j.eswa.2022.118130.
11. S. Krupa, Y. Kryvenchuk. Models of semantic analysis of product descriptions for automatic determination of customscodes // Applied Aspects of Information Technology. – 2026. – Vol. 9, No 1. – P. 76–85.
12. Stepan Krupa, Yurii Kryvenchuk. Smart customs concept: integration of HS systems into a single digital ecosystem // Наука і техніка сьогодні. – 2025. – No 13 (54). – С. 1665–1675.
13. Stepan Krupa, Yurii Kryvenchuk. Cross-modal representation learning for accurate harmonized system code classification in e-commerce systems // Вісник сучасних інформаційних технологій. – 2026. – Vol. 9, № 2. – P. 158–167. (ДБ/Радіозв'язок, Прикладні дослідження і розробки).
14. Krupa, S., Krivenchuk, Yu. (2024). Review of the possibility of improving automated HS code selection using machine learning methods to optimize the customs classification process. pp. 46–149. https://doi.org/10.31891/2307-5732-2024-333-2-23
15. Krupa, S., Krivenchuk, Yu. (2023). Means of improving automated selection of HS code. pp. 87. https://science.lpnu.ua/qm-2023/proceedings.

Історія статті:
Отримано: 29.05.2026 Доопрацьовано: 04.09.2026 Прийнято до друку: 23.09.2026 Опубліковано: 29.09.2026