

КРИВЕНЧУК ЮРІЙ

Національний університет "Львівська політехніка"

<https://orcid.org/0000-0002-2504-5833>e-mail: Yurii.P.Kryvenchuk@lpnu.ua**КРУПА СТЕПАН**

Національний університет "Львівська політехніка"

<https://orcid.org/0009-0000-2074-9762>e-mail: stepan.m.krupa@lpnu.ua

ОГЛЯД МОЖЛИВОСТІ ПОКРАЩЕННЯ АВТОМАТИЗОВАНОГО ПІДБОРУ HS КОДУ З ВИКОРИСТАННЯМ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ ОПТИМІЗАЦІЇ МИТНОГО КЛАСИФІКАЦІЙНОГО ПРОЦЕСУ

Покращення сучасних підходів до оптимізації митного класифікаційного процесу через впровадження методів машинного навчання для автоматизованого підбору HS кодів. На тлі зростання обсягів та складності міжнародної торгівлі, важливість швидкого та точного визначення класифікацій стає надзвичайно актуальною. Розглянемо вплив використання методів машинного навчання на покращення процесів класифікації товарів за системою Гармонізованої системи (HS). Зокрема, проаналізуємо переваги автоматизації підбору кодів, такі як підвищена точність, швидкість та ефективність у порівнянні з традиційних методів. Продемонструємо практичні приклади впровадження машинного навчання в митній сфері та висвітлені потенційні перспективи розвитку цього напрямку. Дослідження покажуть на важливість вдосконалення та сучасного підходу до митної класифікації товарів для оптимізації і поліпшення процесів міжнародної торгівлі. Автоматизований підбір HS кодів за допомогою машинного навчання прискорює визначення класифікації та забезпечує високу ступінь надійності у порівнянні з традиційними методами.

Ключові слова: митна класифікація, HS коди, машинне навчання, автоматизація, оптимізація процесів.

KRYVENCHUK YURI, KRUPA STEPAN

Lviv Polytechnic National University

IMPROVEMENT OF AUTOMATED HS CODE SELECTION USING MACHINE LEARNING METHODS FOR OPTIMIZATION OF THE CUSTOMS CLASSIFICATION PROCESS

Improvement of modern approaches to the optimization of the customs classification process through the introduction of machine learning methods for the automated selection of HS codes. Against the background of growing volumes and complexity of international trade, the importance of quick and accurate determination of classifications becomes extremely urgent. Let's consider the impact of using machine learning methods on improving the processes of product classification according to the Harmonized System (HS). In particular, we will analyze the advantages of automating the selection of codes, such as increased accuracy, speed and efficiency compared to traditional methods. We will demonstrate practical examples of the implementation of machine learning in the customs sphere and highlight potential prospects for the development of this direction. Research will show the importance of improvement and a modern approach to the customs classification of goods for optimization and improvement of international trade processes. In the conditions of the rapid development of international trade and the expansion of the assortment of goods, customs services are increasingly directing their efforts to the improvement and optimization of processes. One of the key directions of this improvement is the introduction of automated HS code selection, which determines the classification of goods according to the Harmonized System (HS). The importance of improvement and a modern approach to the customs classification of goods for optimization and improvement of international trade processes becomes even greater in the context of the rapid development of this sector and the expansion of the range of goods. As part of this improvement, the use of machine learning methods for automated selection of HS codes comes to the fore. These advanced techniques can improve not only the accuracy but also the efficiency of the product classification process. Automated selection of HS codes using machine learning accelerates classification determination and provides a high degree of reliability compared to traditional methods.

Keywords: customs classification, HS codes, machine learning, automation, process optimization.

Постановка проблеми

У сучасному світі, де обсяги міжнародної торгівлі швидко зростають, та товари стають більш різноманітними та складними, проблема точної та ефективної класифікації за системою Гармонізованої системи (HS) стає актуальнішою. Особливо гостро це відчувається в митницях, де митний персонал повинен швидко та точно визначати HS коди для тисяч товарів.

Головна проблема полягає в тому, що традиційні методи класифікації товарів стають недостатньо ефективними та часовитратними у зіставленні зі зростаючою складністю товарів та змінами в міжнародному законодавстві. Людські помилки та відсутність швидких адаптацій до змін у класифікаційних стандартах можуть призводити до затримок у митних процедурах та втрат ефективності.

Однією з ключових проблем є також необхідність вдосконалення процесу митної класифікації за допомогою сучасних технологій машинного навчання. Відсутність автоматизованих систем та використання лише традиційних методів обмежує швидкість та точність класифікації.

Таким чином, виникає необхідність у вдосконаленні та оптимізації митного класифікаційного процесу через впровадження методів машинного навчання для автоматизованого підбору HS кодів. Це дозволить підвищити ефективність, зменшити часові затрати та уникнути помилок, що може мати значущий вплив на покращення митних операцій та сприяти зручності для міжнародних торговців.

Аналіз останніх джерел

У статті [1] Гармонізованої системи (HS). За словами F. Altaheri and K. Shaalan., код HS Code, також відомий як Гармонізована система опису та кодування товарів, був розроблений Всесвітньою митною організацією (WCO), що базується в Брюсселі, щоб впоратися зі швидким зростанням міжнародної торгівлі в усьому світі. Відповідно до цього Weerth (2008) пояснив, що товари в митному тарифі повинні бути повністю описані в митній декларації, щоб їх можна було відповідно класифікувати. Наприклад, дерев'яний стілець можна класифікувати відповідно до матеріального стану або його функції як меблі (Weerth 2008).

Автор статті [2] вирішив показати методи машинного навчання зараз інтегруються в різні програми та системи, щоб допомогти організаціям приймати кращі рішення, зокрема щодо прогнозного аналізу та розпізнавання образів. Наприклад, методи машинного навчання значною мірою інтегруються в сферу безпеки, наприклад, для покращення розпізнавання обличчя за рахунок використання великої кількості даних із різних джерел, що людям важко виконувати вручну (Mohammed et al. 2018). Крім того, машинне навчання також інтегрується в різні системи для сприяння автоматизації та підвищення ефективності та точності. Наприклад, методи машинного навчання інтегруються в коди HS для підвищення точності, інтелекту та автоматизації (Zang et al. 2008).

У даній статті [3] один з авторів *Paul Pu Liang* розглянув дані, використані в цій роботі, наданих Дубайською митницею в рамках конкурсу зі штучного інтелекту (AI), який проводився в жовтні 2019 року. Ці дані складаються з 22 346 194 записів, у яких кожен запис має два атрибути; гармонізований код системи (Код HS) і опис введених користувачем даних. У даному розділі описано методи обробки, які ми застосували для перегляду, аналізу та підготовки вхідних даних.

Публікація [4] дослідників *Alqaryouti, O., Siyam, N., Monem, A.A., Shaalan, K* описує, стратегії машинного навчання також інтегруються в HS, щоб покращити процес автоматизації. Відповідно до KPMG International (2019), машинне навчання можна використовувати для створення бази знань для навчання та розробки набору алгоритмів на основі великої кількості даних, щоб робити обґрунтовані прогнози. KPMG International (2019, стор. 37) пояснила, що «комбінація обробки природної мови та машинного навчання дає змогу автоматизувати збір, масив і аналіз неструктурованих даних і перетворити їх у структуровані дані, які можна використовувати в додатку атах». Це свідчить про те, що інтеграція машинного навчання в HS може сприяти підвищенню ефективності процесу за рахунок покращення якості, узгодженості та точності класифікації коду завдяки зменшенню ймовірності людських помилок. Розроблено кілька інструментів і методів машинного навчання. Тим не менш, машинне навчання можна розглядати як те, що забезпечує технічну основу для інтелектуального аналізу даних (Witten et al. 2011).

У статті [5] описано, що НЛП токенізація є важливим кроком для визначення ваги (важливості) кожного слова в тексті. Загалом, для виконання токенізації широко використовуються дві основні техніки. Ці методи: (1) сума слів і (2) частота термінів - частота інвертування документа (TF-IDF). Сума слів — це просто метод підрахунку, за якого кожному слову надається значення, яке представляє частоту, з якою слово з'являється в тексті. Відповідно, використання цієї методики залежить від того, чи кількість появ кожного слова в тексті розглядається як окремі ознаки в кожній моделі (Alqaryouti та ін. 2019). Однак TF-IDF вважається більш просунутою технікою, оскільки вона використовує частоту слова, щоб визначити унікальність кожного слова в тексті. TF-IDF починається з обчислення частоти кожного слова в тексті ($F_t(w)$). Це обчислюється шляхом ділення кількості разів, коли слово (w) з'являється в тексті, на загальну кількість слів. Це позначає термін частоти в розрахунку (TF).

Мета роботи полягає в забезпеченні цінної інформації для фахівців у сфері митного класифікаційного процесу, а також внесенні вагомому вкладу в розвиток автоматизованих систем митного контролю з використанням сучасних технологій машинного навчання.

Етапи процесу покращення автоматизованого підбору митного коду

Процес вдосконалення автоматизованого підбору HS коду за допомогою методів машинного навчання для оптимізації митного класифікаційного процесу включає кілька ключових етапів:

1. Збір та Підготовка даних:

- Збирання реальних даних, таких як декларації товарів, рахунки, та інші відомості з митниць та торгових організацій.

- Очищення та підготовка даних, включаючи обробку відсутніх значень, нормалізацію та інші техніки для забезпечення якості та відповідності стандартам.

2. Визначення задачі та параметрів:

- Визначення конкретної задачі машинного навчання, такої як класифікація товарів за HS кодами.

- Визначення вхідних параметрів (ознак) та виходу (класифікаційний код).

3. Вибір моделі:

- Вибір алгоритму машинного навчання, який найбільше відповідає вибраній задачі (наприклад, нейронні мережі, дерева рішень, або метод опорних векторів).

- Оцінка та порівняння різних моделей з метою визначення найефективнішого в контексті задачі.

4. Тренування моделі:

- Розділення даних на тренувальний та тестовий набори.

- Підгонка обраної моделі на тренувальних даних для вивчення патернів та адаптації до специфіки класифікації HS кодів.

5. Оцінка та налаштування:

- Валідація моделі за допомогою тестового набору для визначення точності та інших метрик ефективності.

- Налаштування параметрів моделі для покращення її результатів та запобігання перенаванчю.

6. Інтеграція у систему:

- Впровадження оптимізованої моделі в систему автоматизованого підбору HS коду.

- Інтеграція з існуючими митними та логістичними системами для покращення класифікаційного процесу.

7. Моніторинг та підтримка:

- Систематичний моніторинг роботи моделі та її ефективності в реальних умовах.

- Постійна підтримка та оновлення моделі відповідно до змін у митному законодавстві чи структурі товарів.

Цей цикл процесу машинного навчання допомагає створити ефективний та оптимізований митний класифікаційний процес за допомогою автоматизованого підбору HS кодів.

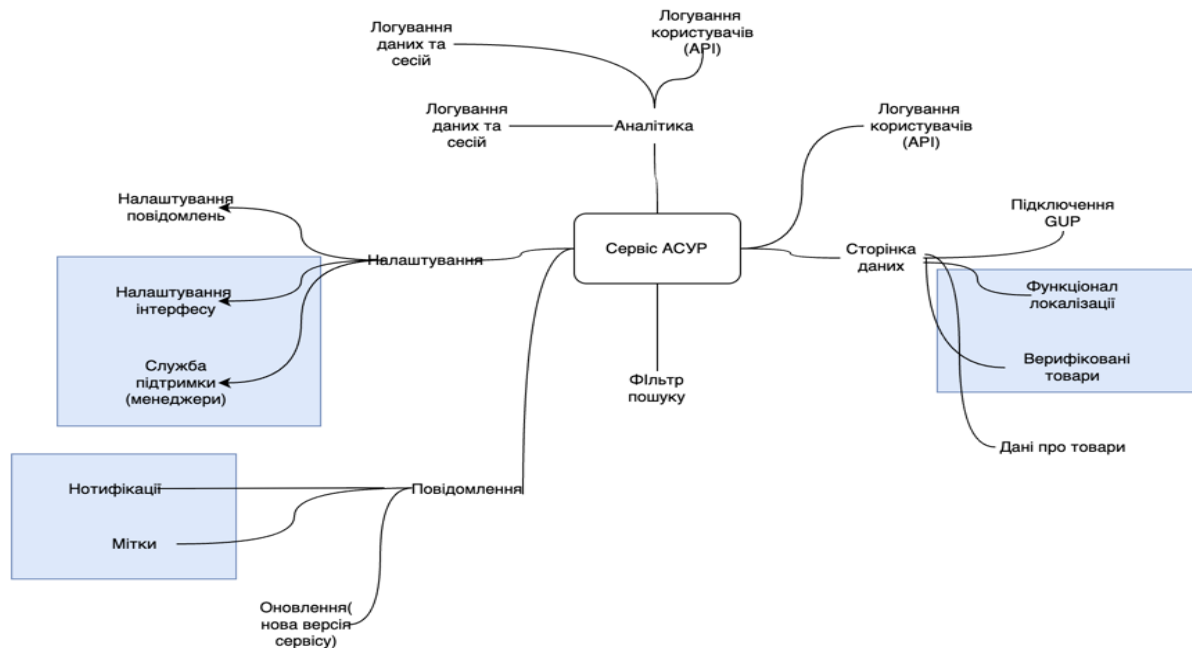


Рис. 1. Архітектура сервісу АСУР

На рис.1 представлена архітектура сервісу автоматизованої системи управління ресурсами (АСУР), на якій зображено різні інформаційні блоки. Цю архітектуру можна оптимізувати за допомогою нових функціональних модулів, які виділені на зображенні. Щодо оновленої системи митного оформлення, то в режимі тестування АСУР отримує декларацію із системи NCTS та надає можливість фіксувати виконання митних формальностей та завершувати контроль, натомість польська система HUZAR цього не підтримує. Впровадження навчання машинного коду (Machine Learning) для менеджерів у системі АСУР (Автоматизована Система Управління Ризиками) може суттєво поліпшити ефективність та точність процесу автоматизованого підбору митного коду. Впровадження системи моніторингу для стеження за ефективністю моделі та виявлення відхилень, а також інформування менеджерів про нові зміни та розширення нових товарів у базі. Важливою формою нововведення буде забезпечення регулярного оновлення моделі на основі нових даних та змін у митному законодавстві. Ключовим моментом буде виступати проведення тестів для оцінки ефективності та точності моделі перед впровадженням в реальні умови.

Висновок

Підсумовуючи цю статтю було розглянуто важливий аспект оптимізації митного класифікаційного процесу через впровадження методів машинного навчання для автоматизованого підбору HS коду. Використання сучасних технологій в області машинного навчання надає великий потенціал для покращення ефективності та точності визначення класифікаційних кодів товарів. Результати досліджень вказують на те, що застосування цих методів призводить до значного збільшення автоматизації процесу, зменшення ймовірності помилок та оптимізації часових та фінансових ресурсів, пов'язаних з митним оформленням. Адаптивність моделей машинного навчання дозволяє їм ефективно працювати в умовах змінного торговельного середовища. Цей дослід допомагає підкреслити важливість інтеграції новітніх технологій в митницю для підвищення ефективності та адаптивності у глобальному торговельному вирі. Висновки статті сприяють розвитку та впровадженню інновацій у митну сферу, спрощуючи та прискорюючи класифікаційні процеси та сприяючи підвищенню конкурентоспроможності митних органів на міжнародній арені.

References

1. F. Altaheri and K. Shaalan, "Exploring Machine Learning Models to Predict Harmonized System Code" in *Information Systems. EMCIS 2019. Lecture Notes in Business Information Processing*, Switzerland:Springer Nature Switzerland, pp. 291-303, 2020.
2. Mohammed, M., Khan, M.B., Bashier, E.B.M.: *Machine Learning-Algorithms and Applications*. CRC Press, Boca Raton (2020)
3. Paul Pu Liang, Yiwei Lyu, Xiang Fan, Zetian Wu, Yun Cheng, Jason Wu, Leslie Chen, Peter Wu, Michelle A Lee, Yuke Zhu, et al. Multibench: Multiscale benchmarks for multimodal representation learning. arXiv preprint arXiv:2107.07502, 2021.
4. Alqaryouti, O., Siyam, N., Monem, A.A., Shaalan, K.: Aspect-based sentiment analysis using smart government review data. *Appl. Comput. Inform.* (2019, Accepted Manuscript)
5. Alqaryouti, O., Khwileh, H., Farouk, T., Nabhan, A., Shaalan, K.: Graph-based keyword extraction. In: Shaalan, K., Hassaniien, A.E., Tolba, F. (eds.) *Intelligent Natural Language*
6. Alqaryouti, O., Khwileh, H., Farouk, T., Nabhan, A., Shaalan, K.: Graph-based keyword extraction. In: Shaalan, K., Hassaniien, A.E., Tolba, F. (eds.) *Intelligent Natural Language*.
7. Mohammed, M., Khan, M.B., Bashier, E.B.M.: *Machine Learning-Algorithms and Applications*. CRC Press, Boca Raton (2018)
8. GHS), 5th Revised edn. United Nations, Geneva Shearer, C.: The CRISP-DM model: the new blueprint for data mining. *J. Data Warehous.* 5(4), 13–22
9. Alqaryouti, O., Siyam, N., Monem, A.A., Shaalan, K.: Aspect-based sentiment analysis using smart government review data. *Appl. Comput. Inform.*
10. Zang, B., Li, Y., Xie, W., Chen, Z., Tsai, C., Laing, C.: An ontological engineering approach for automating inspection and quarantine at airports. *J. Comput. Syst. Sci.* 74, 196–210.
11. Youyi, W.: Exploration of Data-Driven Intelligent Customs (2017). https://www.eiseverywhere.com/file_uploads/c7e054aa02ad13907d6ad513ea57b8d_session3-YouyiWu.pdf
12. Weerth, C.: Basic Principles of Customs Classifications under the Harmonized System. *Global Trade Customs J.* 3(2), 61–67
13. Mohammed, M., Khan, M.B., Bashier, E.B.M.: *Machine Learning-Algorithms and Applications*. CRC Press, Boca Raton
14. Luppess, J.: *Classifying short text for the harmonized system with convolutional neural networks*. Master Thesis, Radboud University
15. VAlqaryouti, O., Siyam, N., Monem, A.A., Shaalan, K.: Aspect-based sentiment analysis using smart government review data. *Appl. Comput. Inform.*
16. Mohammed, M., Khan, M.B., Bashier, E.B.M.: *Machine Learning-Algorithms and Applications*. CRC Press, Boca Raton (2020)
17. Luppess, J.: *Classifying short text for the harmonized system with convolutional neural networks*. Master Thesis, Radboud University (2019)
18. Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021.