

МОДЕЛИ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ КОДА ТН ВЭД: ПЕРСПЕКТИВЫ И ЭФФЕКТИВНОСТЬ ИСПОЛЬЗОВАНИЯ

Муратова Шохиста Ниматуллаевна

Начальник кафедры Таможенного института,

д.э.н., профессор

ORCID: 0000-0002-8133-9094

E-mail: shohista11@mail.ru

Аннотация

В статье рассматриваются различные модели машинного обучения для прогнозирования кода ТН ВЭД на основе введенных в грузовую таможенную декларацию описаний товаров. Коды ТН ВЭД широко используется всеми таможенными службами благодаря ряду преимуществ, включая более удобный и упрощенный подход к расчету пошлин, а также предотвращение потенциальной потери доходов. Данное исследование основано на межотраслевом процессе разработки методологии интеллектуального анализа данных. Результаты исследования показывают, что модели машинного обучения являются эффективными инструментами для прогнозирования кода ТН ВЭД на основе вводимых данных.

Ключевые слова: машинное обучение, код ТН ВЭД, прогнозные модели, таможенные службы, предотвращение потери доходов, торговля.

Annotatsiya

Maqolada bojxona yuk deklaratsiyalariga kiritilgan tovar tavsiflari asosida TIF TN kodlarini bashorat qilish uchun turli xil mashinaviy o'qitish modellari ko'rib chiqilgan. TIF TN kodlari bir qator afzalliklar, jumladan, bojlarni hisoblash va daromadning potensial yo'qotishlarining oldini olishning yanada qulay va soddallashtirilgan yondashuvi tufayli barcha bojxona xizmatlari tomonidan keng qo'llaniladi. Ushbu tadqiqot ma'lumotlarni intellektual tahlil qilish metodologiyasini ishlab chiqish uchun tarmoqlararo jarayonga asoslangan. Natijalar shuni ko'rsatdiki, mashinaviy o'qitish modellari kiritiladigan ma'lumotlarga asoslangan TIF TN kodlarini bashorat qilish uchun samarali vositalardir.

Kalit so'zlar: mashinaviy o'qitish, TIF TN kodi, prognoz modellari, bojxona xizmati, daromad yo'qotishlarining oldini olish, savdo.

Abstract

The article examines various machine learning models for predicting GN FEA codes based on product descriptions entered into customs declarations. GN FEA codes are widely used by all customs services due to a number of advantages, including a more convenient and simplified approach to calculating duties and preventing potential revenue loss. This study is based on a cross-industry process to develop a data mining methodology. The results demonstrate that machine learning models are effective tools for predicting GN FEA codes based on input data.

Keywords: machine learning, GN FEA codes, predictive models, customs services, revenue loss prevention, trade.

ВВЕДЕНИЕ

Глобализация экономики открыла ряд возможностей для достижения экономического роста и процветания. Таможенные органы являются одним из наиболее важных государственных органов в глобальной перспективе, который регулирует процессы международной торговли и услуг, такие как декларирование товаров и услуг, особенно в аспекте торговли и коммерции. Как правило, проверка описаний товаров гарантирует, что товары соответствуют государственным нормативным актам, направленным на предотвращение ненадлежащего или незаконного ввоза в страну назначения [1]. В этом смысле таможенные органы несут ответственность за обеспечение того, чтобы заявленные товары, подлежащие импорту или экспорту, были соответствующим образом классифицированы на основе описания товаров.

Традиционно задекларированные товары анализируются и проверяются инспекторами с использованием изображений, результаты которых используются в качестве основания для принятия решения [1]. Используя этот подход, можно выявить ряд слабых мест и проблем, таких как трудности с определением опасных веществ и неэффективная классификация декларируемых товаров из-за возможной неэффективности назначения кодов согласованной системы [1, 5]. Помимо этих недостатков, при применении кода Товарной номенклатуры внешнеэкономической деятельности (далее по тексту – ТН ВЭД) существуют также некоторые проблемы, связанные с достижением удовлетворительной точности. К числу этих проблем относятся сложность ТН ВЭД, пробелы в терминологии и эволюционирующий характер кода ТН ВЭД, среди прочего [2].

Данное исследование направлено на то, чтобы способствовать снижению сложности определения кода ТН ВЭД и устранению пробелов путем изучения различных моделей прогнозирования, основанных на машинном обучении. Чтобы лучше оценить влияние внедрения модели прогнозирования кодов ТН ВЭД, основанной на машинном обучении, мы использовали пример таможни Республики Узбекистан. Ожидается, что интеграция технологических достижений в классификацию кодов ТН ВЭД может способствовать упрощению кода ТН ВЭД и повышению его точности.

ЛИТЕРАТУРНЫЙ ОБЗОР

Методы машинного обучения в настоящее время интегрируются в различные программы и системы, чтобы помочь организациям принимать более эффективные решения, особенно в отношении прогнозного анализа и распознавания образов. Авторы M.Mohammed, M.V.Kha, E.V. Bashie [4] полагают, что методы машинного обучения в значительной степени интегрируются в сферу безопасности, например, для улучшения распознавания лиц за счет использования большого объема данных из различных источников,

что человеку трудно сделать вручную. Кроме того, машинное обучение также интегрируется в различные системы для содействия автоматизации и повышения эффективности и точности. Как было отмечено авторами B.Zang, Y.Li, W.Xie, Z.Chen, C.Tsai, C.Laing [7] методы машинного обучения интегрируются в коды ТН ВЭД для повышения точности, интеллектуальности и автоматизации.

В исследовании L.Ding, Z.Fan, D.Chen [2] приводится описание интеграции машинного обучения с процедурами международной торговли. С точки зрения таможенных служб, интеграция машинного обучения стала жизненно важной для того, чтобы они могли оставаться конкурентоспособными в области упрощения процедур международной торговли. Различные методы машинного обучения уже применяются для получения и изучения долгосрочных и стабильных критериев категоризации текста. Например, использование ключевых слов является одним из наиболее распространенных подходов к машинному обучению благодаря использованию модели векторного пространства (Vector Space Models) [2].

Кроме того, стратегии машинного обучения также интегрируются в ТН ВЭД для улучшения процесса автоматизации. По данным KPMG International [3], машинное обучение может быть использовано для создания базы знаний для обучения и разработки набора алгоритмов на основе большого объема данных с целью получения обоснованных прогнозов. KPMG International [3] пояснил, что “Сочетание обработки естественного языка и машинного обучения позволяет автоматизировать сбор, систематизацию и анализ неструктурированных данных и преобразовать их в структурированные данные, которые могут быть использованы в налоговом заявлении”. Это говорит о том, что интеграция машинного обучения в ТН ВЭД может способствовать повышению эффективности процесса за счет повышения качества, согласованности и точности классификации кодов за счет снижения вероятности человеческих ошибок. В настоящее время разработано несколько инструментов и методик машинного обучения. Тем не менее, машинное обучение можно рассматривать как средство, обеспечивающее техническую основу для интеллектуального анализа данных.

МЕТОДОЛОГИЯ

В данной статье рассматриваются особенности использования моделей машинного обучения, как эффективные инструменты для прогнозирования кода ТН ВЭД. В ходе исследования был проведен монографический анализ внедрения модели машинного обучения для прогнозирования кода ТН ВЭД. Для понимания, того как работает модель машинного обучения, были использованы системный подход, научные методы дедукции и индукции, а также нормативно-правовые и эмпирические методы.

АНАЛИЗ И РЕЗУЛЬТАТЫ

Чтобы способствовать эффективному применению и интеграции машинного обучения в ТН ВЭД, важно понимать, как это работает. В общей

перспективе машинное обучение можно рассматривать как новую технологию извлечения знаний из данных с целью обучения и получения точных прогнозируемых результатов. Таким образом, машинное обучение можно рассматривать как непосредственно связанное с интеллектуальным анализом данных. В соответствии с этим, методы машинного обучения, интегрированные с ТН ВЭД, также могут способствовать интеллектуальному сопоставлению и классификации за счет создания интеллектуальных таможенных систем, управляемых данными [6]. Объясняется, что интеллектуальная таможня, основанная на данных, может способствовать интеллектуальному правоприменению, интеллектуальному контролю рисков и интеллектуальному сбору доходов, при этом сбор данных может привести к созданию интеллектуальных приложений.

Целью данного исследования является изучение возможности внедрения модели машинного обучения для прогнозирования кода ТН ВЭД для товаров на основе предоставленных описаний в грузовых таможенных декларациях. Для выполнения этой задачи интеллектуального анализа текста мы использовали методологию межотраслевого анализа данных. Эта методология состоит из следующих процессов, а именно: *понимания бизнеса, анализа данных, подготовки данных, моделирования данных, оценки производительности и внедрения* (Рис.1).

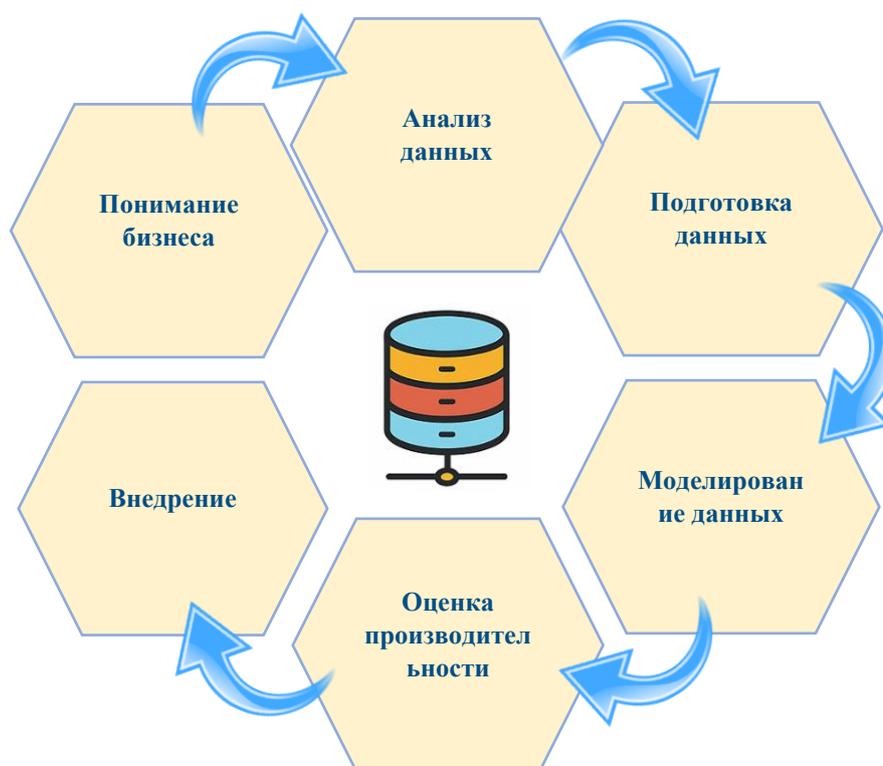


Рис.1. Методология межотраслевого анализа данных¹

¹Авторская разработка

Понимание бизнеса в основном связано с важностью решаемой проблемы, которая заключается в снижении потерь доходов за счет построения модели машинного обучения для правильного прогнозирования кода ТН ВЭД. Важность этого направления работы заключается в сокращении потерь в доходах, вызванных неправильной классификацией товаров. Такая неправильная классификация приводит к потере таможенных пошлин. Такая неправильная классификация обычно связана либо с возможностью неправильного определения пользователем требуемого кода ТН ВЭД, либо, в другой ситуации, с манипуляциями пользователей с целью уменьшения суммы пошлины, которую они обязаны уплатить.

Подготовка данных. Данные, использованные в этой работе, предоставлены таможенной службой Республики Узбекистан, состоящие из двух атрибутов: код ТН ВЭД, и описание введенных в грузовую таможенную декларацию данных. Здесь мы опишем методы обработки, использованных для проверки, анализа и подготовки входных данных.

Анализ данных. В ходе анализа предоставленных данных мы обратили внимание на несколько основных факторов, которые необходимо учитывать при разработке модели, основанной на машинном обучении. Эти факторы в основном связаны с качеством описания товаров и количеством ожидаемых классов (меток). Что касается описания, то было замечено несколько проблем. Например, в некоторых описаниях отсутствуют какие-либо ценные ключевые слова. Другими словами, орфографические ошибки в этих описаниях настолько серьезны, что человеческий разум не может их понять. Кроме того, мы заметили, что одно и то же описание использовалось несколько раз для описания совершенно разных товаров. Что касается ожидаемого количества классов, то предоставленные данные содержат почти 8000 меток. И это подчеркивает проблему, связанную с созданием любой модели машинного обучения для прогнозирования метки на основе предоставленного описания. Кроме того, было замечено, что они сильно различаются по количеству записей, связанных с каждым классом. Например, некоторые классы содержат около 10 записей, в то время как другие классы могут содержать до 700 записей. Этот фактор необходимо учитывать, чтобы избежать какой-либо предвзятости в модели машинного обучения.

Моделирование данных. Вышеупомянутые факторы были приняты во внимание при очистке данных. В этом разделе мы представим процедурные шаги, которые были выполнены во время очистки данных. Эти шаги можно резюмировать следующим образом:

- (1) – «удалите дублирование»,
- (2) – «удалите знаки препинания»,
- (3) – «удалите неформальные слова»,
- (4) – «удалите цифры»,
- (5) – «удалите лемматизацию (приведение слово к лемме – ее нормальной форме)».

Оценка производительности. Ожидается, что наличие классов с очень большими различиями в количестве записей окажет значительное влияние на поведение при обучении. Соответственно, необходимо решить эту проблему выборки, чтобы обеспечить согласованность с точки зрения поведения классификации. Чтобы решить эту проблему, мы указали минимальные и максимальные записи, представленных в наборе данных. Таким образом, мы выполнили выборку с понижением, чтобы получить 30 записей в качестве промежутка между минимальным и максимальным размером класса. В этой работе из-за ограниченности вычислительных возможностей используемой машины мы случайным образом выбрали только 5000 записей в качестве входных данных. Кроме того, после выполнения всех этапов предварительной обработки в качестве входных данных для окончательного набора данных сохраняются 17 070 записей.

Внедрение. В обработке текстов на естественном языке (Natural Language Processing – NLP) символизация является важным шагом для определения веса (важности) каждого слова в тексте. В целом, для выполнения символизации широко используются два основных метода:

1. Набор слов.
2. Преобразование частоты использования термина в частоту использования документа.

Набор слов – это просто метод подсчета, при котором каждому слову присваивается значение, представляющее частоту, с которой это слово встречается в тексте. Соответственно, использование этого метода зависит от того, рассматривается ли количество встречаемости каждого слова в тексте как отдельные признаки в каждой модели.

Теперь представляем и обсуждаем модели машинного обучения, которые используются в данной работе, и соответствующую производительность каждой модели. Эти эксперименты были выполнены с использованием Python 3.5 для получения спецификаций моделей.

В целом, любая модель машинного обучения работает, используя обучающий набор для изучения поведения данных. В то время как тестовый набор – это невидимые данные, которые используются для измерения точности поведения модели машинного обучения. При перекрестной проверке эксперимент проводится n раз, и каждый раз данные делятся на обучающий и тестовый наборы. В каждом подходе выбранный набор тестов должен отличаться от того, который используется для мониторинга производительности в течение n -ных прогонов, чтобы лучше анализировать производительность модели.

Основными составляющими этой работы являются использование следующих моделей машинного обучения для оценки предложенного исследовательского вопроса и ответа на него. Это модели машинного обучения: Decision Tree, Random Forest, Linear Support Vector Machine и Adaboost. В этих моделях машинного обучения мы будем использовать модель прогнозирования

при двух настройках эксперимента. При первой настройке мы проверим, способна ли модель машинного обучения правильно предсказать весь код ТН ВЭД. В то время как при второй настройке мы проверим способность прогнозировать только товарную позицию. Товарная позиция обозначает первые четыре цифры кода ТН ВЭД, которые представляют раздел (2 цифры) и группу (2 цифры). Раздел и группа представляют тип товара (например, чай или изделия из драгоценных металлов).

ВЫВОДЫ И ПРЕДЛОЖЕНИЯ

Обобщая вышеизложенное, отметим, что значимость данного исследования заключается в том, что оно направлено на расширение эмпирических знаний, особенно для устранения пробелов в использовании методов машинного обучения при совершенствовании кода ТН ВЭД. Как показывают результаты исследования, модели машинного обучения полезны для прогнозирования кода ТН ВЭД в рамках описаний товаров. Однако это исследование также не лишено ограничений. Поскольку был применен подход тематического исследования, исследование ограничено наборами данных, полученных от первичных документов, поданных для таможенного оформления для прогнозирования кода ТН ВЭД.

С другой стороны, информация, представленная в исследовании, может быть использована для будущих исследований, также может быть использована для изучения и определения стратегий, которые положительно повлияют на эффективность использования кода ТН ВЭД. Вопрос о внедрении моделей машинного обучения в качестве прогностических систем и их соответствующей эффективности может быть дополнительно рассмотрен в будущем.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Che J., Xing Y., Zhang L. A comprehensive solution for deep-learning based cargo inspection to discriminate goods in containers. – IEEE Xplore, 2018. URL: https://www.researchgate.net/publication/329744175_A_Comprehensive_Solution_f_or_Deep-Learning_Based_Cargo_Inspection_to_Discriminate_Goods_in_Containers
2. Ding L., Fan Z., Chen D. Auto-categorization of HS Code using background net approach. // Procedia Computer Science, 2015. – P.1462-1471. URL: <https://www.sciencedirect.com/science/article/pii/S1877050915023510?via%3Dihub>
3. KPMG International: Transforming the tax function through technology, 2018. URL: <https://home.kpmg/content/dam/kpmg/au/pdf/2018/transforming-the-taxfunction-au.pdf>
4. Mohammed M., Kha M.B., Bashie E.B. Machine Learning-Algorithms and Applications. // CRC Press, Boca Raton, 2017. URL: https://api.pageplace.de/preview/DT0400.9781498705394_A27230551/preview-9781498705394_A27230551.pdf
5. United Nations: Globally Harmonized System of Classification and Labelling of Chemicals (GHS). // 5th Revised ed. United Nations, Geneva, 2013. URL:

https://unece.org/DAM/trans/danger/publi/ghs/ghs_rev05/English/ST-SG-AC10-30-Rev5e.pdf

6. Youyi W. Exploration of Data-Driven Intelligent Customs, 2017. URL: https://www.eiseverywhere.com/file_uploads/c7e054aa02ad13907d6ad513ea57b8d_session3-YouyiWu.pdf

7. Zang B., Li Y., Xie W., Chen Z., Tsai C., Laing C. An ontological engineering approach for automating inspection and quarantine at airports. // Journal Computer System Science, №74, 2008. – P.196-210. URL: https://www.sciencedirect.com/science/article/pii/S0022000007000499?pes=vor&utm_source=wiley&getft_integrator=wiley



Marketing

ilmiy, amaliy va ommabop jurnali

Muharrir: Xakimov Ziyodulla Axmadovich
Ingliz tili muharriri: Tursunov Boburjon Ortiqmirzayevich
Rus tili muharriri: Kaxramonov Xurshidjon Shuxrat o'g'li
Musahhah: Karimova Shirin Zoxid qizi
Sahifalovchi va dizaynerlar: Sadikov Shoxrux Shuxratovich
Abidjonov Nodirbek Odijon o'g'li

2025-yil, noyabr, 11-son

© Materiallar ko'chirib bosilganda "Marketing" ilmiy, amaliy va ommabop jurnali manba sifatida ko'rsatilishi shart. Jurnalda bosilgan material va reklamalardagi dalillarning aniqligiga mualliflar mas'ul. Tahririyat fikri har vaqt ham mualliflar fikriga mos kelavermasligi mumkin. Tahririyatga yuborilgan materiallar qaytarilmaydi.

Mazkur jurnalda maqolalar chop etish uchun quyidagi havolalarga murojaat qilish mumkin. Ilmiy maqola, ommabop maqola, reklama, hikoya va boshqa ilmiy-ijodiy materiallar yuborishingiz mumkin.

Materiallar va reklamalar pullik asosda chop etiladi.

Elektron pochta: info@marketingjournal.uz
Bot: [@marketinjournalbot](https://t.me/@marketinjournalbot)
Tel.: +998977838464, +998939266610
Jurnalning rasmiy sayti: <https://marketingjournal.uz>

Marketing jurnali O'zbekiston Respublikasi Oliy ta'lim, fan va innovatsiyalar vazirligi huzuridagi **Oliy attestatsiya komissiyasi rayosatining 2024-yil 04-oktabrdagi 332/5 sonli qarori** bilan milliy ilmiy nashrlar ro'yxatiga kiritilgan



"Marketing" ilmiy, amaliy va ommabop jurnali 2024-yil 15-martdan O'zbekiston Respublikasi Prezidenti Administratsiyasi huzuridagi Axborot va ommaviy kommunikatsiyalar agentligi tomonidan **C-5669517** reyestr raqami tartibi bo'yicha ro'yxatdan o'tkazilgan. **Litsenziya raqami: №240874**



"Marketing" ilmiy, amaliy va ommabop jurnalining xalqaro darajasi: **9710**. GOCT 7.56-2002 "Seriya nashrlarning xalqaro standart raqamlanishi" davlatlararo standartlari talablari. **Berilgan ISSN tartib raqami: 3060-4621**