DOI : 10.26565/2519-2310-2023-2-01
УДК 004.8: 342.9

ПРОБЛЕМНІ ПИТАННЯ ТЕХНОЛОГІЇ МАШИННОГО НАВЧАННЯ В ПРАВООХОРОННІЙ ДІЯЛЬНОСТІ

Дмитро Узлов[1], Володимир Струков[2], Владислав Гуділін[3], Олексій Власов[4]

[1]Харківський національний університет імені В.Н. Каразіна, майдан Свободи, 4, Харків, 61022, Україна,
e-mail: dmytro.uzlov@karazin.ua, ORCID: https://orcid.org/0000-0003-3308-424X
[2]Харківський національний університет внутрішніх справ, пр. Льва Ландау, 27, Харків, 61080, Україна,
e-mail: struk_vm@ukr.net, ORCID: https://orcid.org/0000-0003-4722-3159
[3]Харківський національний університет внутрішніх справ, пр. Льва Ландау, 27, Харків, 61080, Україна,
e-mail: vgudilin7@gmail.com, ORCID: https://orcid.org/0000-0002-3844-1448
[4]Харківський національний університет радіоелектроніки, пр. Науки, 14, Харків, 61166, Україна,
e-mail: moonreactor@gmail.com, ORCID:https://orcid.org/0000-0003-1619-0032

Надійшла до редакції 17 жовтня 2023 р. Переглянута 19 листопада 2023 р. Прийнята 20 грудня 2023 р.

Анотація:Правоохоронні органи все частіше використовують технології прогнозування та автоматизації, де основною технологією часто є застосування методів машинного навчання (ML). У статті розглядається проблема підзвітності та відповідальності правоохоронних органів і посадових осіб в контексті застосування моделей машинного навчання ML. Автори вказують, що підзвітність є ключовим елементом демократичної правоохоронної діяльності, але використання прогнозного програмного забезпечення може створювати проблеми у забезпеченні цієї підзвітності. Стаття обговорює, що застосування ML може призвести до завуалювання відповідальності та ускладнення підзвітності у «мультиагентних структурах», що об'єднують людей і обчислювальні інструменти. Особлива увагу приділяється непрозорості алгоритмів прикладних прогнозних моделей та автоматизованих систем прийняття рішень, що стає джерелом непорозумінь і обережності щодо їх використання. Автори висувають питання щодо того, як можна забезпечити ефективний контроль та повну звітність, коли ключові компоненти процесу прийняття рішень залишаються невідомими для широкої громадськості, посадових осіб та навіть розробників моделей. У статті стверджується, що важливі питання, пов'язані з моделями рішень ML, можуть бути розглянуті без детального знання алгоритму навчання, що дає змогу експертам правоохоронної діяльності, які не займаються ML, вивчати їх у формі інтелектуального контролю. Експерти, які не займаються ML, можуть і повинні переглядати навчені моделі ML. Автори надають «набір інструментів» в формі запитань про три елементи моделі прийняття рішень, які можуть бути якісно досліджені експертами, які не є спеціалістами з машинного навчання: навчальні дані, навчальна мета та антиципаційна оцінка результатів. Такий підхід розширює можливості цих експертів у вигляді об'єктивної оцінки використання моделей ML у правоохоронних завданнях. Основна ідея полягає в тому, що навіть без глибоких технічних знань експерти можуть аналізувати та переглядати моделі ML, розкриваючи їхню ефективність через призму власного досвіду. Даний підхід сприяє порозумінню використання технологій машинного навчання в рамках правоохоронної діяльності, розширюючи потенціал відповідних експертів, не пов'язаних з ML.

Ключові слова:*машинне навчання, штучний інтелект, аналіз даних.*

1.Вступ

Правоохоронні органи все частіше застосовують досягнення інформаційних технологій та штучного інтелекту, щоб намагатися передбачити події та автоматизувати обробку даних що виникають в процесі правоохоронної діяльності. У цьому правоохоронна діяльність схожа на багато інших галузей – управління авто, прогнозування погоди, вирішення заявок на кредит тощо. Прогностична аналітика підтримує управління ризиками у сфері управління безпекою [1]. Лондон, Лос-Анджелес, Мюнхен, Новий Орлеан, Філадельфія, Цюрих та Харків – це приклади міст, де поліція використовує або тестувала інтелектуальне поліцейське програмне забезпечення, яке має на меті або передбачити, де можуть статися злочини, або хто, ймовірно, вчинить злочин у майбутньому. Машинне навчання (ML) є ключовою технологією, яка лежить в основі багатьох із цих програм. Програмне забезпечення для машинного навчання може раціоналізувати трудомісткі завдання обробки даних, таких як аналіз великого обсягу документів, оприлюднених у ході розслідування, та класифікація їх за категоріями [2]. Разом з цим

© УзловД., Струков В., Гуділін В., Власов О., 2023
This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0

підзвітність поліції викликала занепокоєння що використання моделей ML робить людей неспроможними відповідати за рішення що були прийняті на їх основі [3]. Щоб спростувати подібні занепокоєння, необхідно зробити процеси прийняття рішень, що ґрунтуються на результатах використання моделей ML, доступними для контролю.

Прогнозну правоохоронну діяльність можна розглядати як окрему техніку під ширшою парасолькою правоохоронної діяльності, керованої аналітикою (ILP). ILP виник як практична управлінська програма для прийняття рішень, щодо правоохоронних послуг на основі об'єктивного аналізу даних [4]. Систематичний збір і аналіз розвідувальних даних мають на меті підвищити як ефективність протидії злочинності, забезпечуючи як більш точне визначення цілей, так і економічну ефективність [5]. У правоохоронній діяльності з прогнозуванням, як і в ILP, аналіз і рішення централізовані та раціоналізовані; прогностична правоохоронна діяльність підкреслює об'єктивний, науковий вибір стратегій та тактик і надає перевагу централізованому, раціоналізованому прийняттю рішень на основі аналізу даних.

2.Підзвітність правоохоронних органів

Підзвітність та відповідальність правоохоронних організацій і посадових осіб є ключовим компонентом демократичної правоохоронної діяльності, і вже давно є предметом занепокоєння дослідників і практиків правоохоронних органів [6]. З точки зору позиції правоохоронних сил у демократичній системі, підзвітність може означати політичний контроль над поліцією або співпрацю між поліцією та урядом, згідно з якою поліція повинна надавати пояснення про прийнятті рішення.

Застосування прогнозного програмного забезпечення або програмного забезпечення для автоматизації для підтримки прийняття рішень може фундаментально поставити під сумнів здатність посадових осіб та організацій звітувати про процеси прийняття рішень, а також завуалювати відповідальність у «мультиагентних структурах», що складаються з людей і обчислювальних інструментів. Непрозорість «алгоритмів» прикладних прогнозних моделей або автоматизованих систем прийняття рішень залишається основною причиною занепокоєння щодо їх використання [7]. Існує занепокоєння, що алгоритми «є непрозорими» в тому сенсі, що одержувачі вихідних даних роботи алгоритму ML*(класифікація, кластеризація, прогноз)*, рідко мають конкретне уявлення про те, як і чому конкретна класифікація, кластеризація або прогноз були отримані на основі вхідних даних [7].

Коли один або більше елементів процесу прийняття рішень незрозумілі, будь-яка з вищезгаданих концепцій підзвітності ставиться під сумнів. Модель ML, як правило, вбудована в програмне забезпечення, працює як «чорна скринька», де вхідні дані *(наприклад, геопросторові дані, щодо злочинності та/чи демографії)* обробляються у вихідні дані *(наприклад, прогноз чи класифікацію)* за допомогою обчислень, які залишаються невидимими для кінцевого користувача. Незважаючи на те, що цей процес, по суті, не є незрозумілим, він практично незрозумілий для не експертів, і може зробити основу незрозумілості щодо обґрунтування прийняття рішень.

Виникають питання: як може існувати ефективний політичний контроль над прийняттям рішень, якщо ключовий компонент у формуванні прийняття рішень фактично невідомий? Як поліція може повністю звітувати про свої рішення, якщо вони частково спиралися на аналіз, який вони самі не в змозі пояснити? В цьому сенсі ***прозорість*** розглядається як частина ідеального вирішення проблем використання ML для підзвітного прийняття рішень. Для досягнення прозорості інформація має бути доступною та зрозумілою [8]. Однак це складно, коли йдеться про напівавтоматизовані інтелектуальні системи. Разом з цим, підзвітність може бути

можливою без повної прозорості *(наприклад, розкриття вихідного коду)* шляхом розробки підзвітності в програмному забезпеченні.

3. Машинне навчання і правоохоронна діяльність

Незважаючи на те, що машини вже досить давно можуть навчатися на основі даних, за останні десятиліття машини стали здатними навчатися та досягати успіху в когнітивних завданнях, таких як позначення об'єктів на зображеннях і визначення слів за звуками. Одним із технологічних застосувань цього було автоматизоване розпізнавання номерних знаків *(APNR)*. Системи *APNR*, встановлені на правоохоронних транспортних засобах, полегшили поліцейський моніторинг правопорушників. Ці розробки відбулися завдяки поєднанню нових алгоритмів навчання *(деякі розроблені з 1950-х років і раніше)*, більшій обчислювальній потужності та розробці коду для ефективного використання обчислювальної потужності машини для вирішення проблем навчання [9]. На додаток до можливості навчання когнітивним завданням, ще однією не менш важливою розробкою ML є винайдення алгоритмів навчання, які можуть наближено створювати складні функції та вибирати важливі характеристики без перенавчання моделі відносно навчальної вибірки. Ці вдосконалення алгоритму дозволили машині навчатися з наборів даних із тисячами позначених функцій, щоб вона могла вибирати функції (змінні) і функціональну форму, яка, ймовірно, добре працюватиме під час прогнозування нових зразків.

Це означає, що змінні, які використовуються в моделях машинного навчання, не обов'язково вибираються фахівцями в галузі, а скоріше самим алгоритмом машинного навчання, і що рішення приймаються не на основі теорій, розроблених людьми, а більше з точки зору того, *«що працює»* в терміни прогнозної сили ML. Не дивно, що ці нові можливості зробили моделі з машинним навчанням дедалі кориснішими для прийнятті рішень на практиці. Моделі ML були використані, наприклад, Управлінням з боротьби з серйозними шахрайствами Великобританії *(UK Serious Fraud Office)* для виявлення юридично конфіденційних матеріалів серед мільйонів розкритих документів у розслідуванні [10], а Норвезьким органом інспекції праці для прогнозування робочих місць з високим ризиком порушень для перевірки агентством.

Обговорюючи, чи використовувати ML у процесі прийняття рішень правоохоронними органами, важливо порівнювати ML не з ідеальним процесом прийняття рішень, а з прийняттям рішень людьми. Машини приймають рішення в неоптимальних середовищах на основі непереконливих, незбагненних і оманливих доказів. Щоразу, коли прийняття рішень призводить до несправедливих результатів, процеси може бути важко відстежити, і «тяжко буває просто визначити, хто повинен нести відповідальність за заподіяну шкоду» [11]. Однак це фундаментальна проблема прийняття рішень як така, а не унікальна для рішень, які приймаються або підтримуються машинами.

Люди чудово навчаються на основі когнітивних даних. Слухаючи звуки, дивлячись на обличчя та спостерігаючи за навколишнім середовищем, ми розрізняємо склади, слова, речення та значення. Ми можемо встановити зв'язок між усмішкою, саркастичним тоном, буквальним значенням речення та тим, що мав намір сказати мовець. Ми можемо читати книги та новини, розмовляти з людьми і робити складні висновки. Комп'ютери все ще не так повно використовують когнітивні дані, як це роблять люди. І в той час як люди зазвичай накопичують лише весь чуттєвий діапазон своїх переживань, то певні дані (наприклад, зображення, звук, відео певного виду) зазвичай збирають з метою навчання комп'ютерів.

Важлива відмінність між машинним і людським навчанням полягає в тому, що ML базується на відомих алгоритмах. За визначенням, алгоритм – це набір інструкцій, які описують порядок дій виконавця, щоб досягти результату розв'язання задачі за скінченну кількість дій;

система правил виконання дискретного процесу, яка досягає поставленої мети за скінченний час. Люди, звичайно, також мають процедури для вирішення проблем у скінченну кількість кроків, які часто включають повторення операції. Однак, навіть людина, яка їх використовує, не завжди може знати або розуміти ці процедури.

Ми знаємо, які алгоритми використовують машини *(ми записуємо їх на мовах програмування)*, і ми можемо контролювати дані, з яких вони навчилися (ми можемо в будь-який момент скинути їх налаштування, ввести певні навчальні дані в модель або припинити навчання). Навчання та наступні рішення що приймаються машиною, в принципі, більш прозорі ніж ті що приймаються людьми. Зрештою, ми не писали код для навчання людини, і ми мало контролюємо вхідні дані, які люди використовували у своєму навчанні та прийнятті рішення. Отже, є певна іронія в тому, що одна з головних критичних зауважень щодо використання машинного навчання при прийнятті рішень полягає в тому, що машинні рішення є непрозорими.

Одне з можливих пояснень цієї невідповідності полягає в тому, що можна відносно просто запитати людей, як вони прийшли до своїх рішень. Було б розумно очікувати, що начальник поліції пояснить факти, інтерпретації та пріоритети, що стоять за його прийняттям рішень. Набагато важче дати подібні пояснення того, чому машина змоделювала саме такі результати своєї роботи; а у багатьох випадках може бути навіть важко описати це простою мовою. Непрозорість навчання машин може, в принципі, бути нижчою, ніж у людей, але на практиці вона вища. Як люди, ми краще підготовлені до того, щоб запитувати інших людей, як вони дійшли своїх висновків, ніж допитувати модель машини.

Ця непрозорість, хоч і зрозуміла, викликає занепокоєння, оскільки може призвести до «позбавлення від відповідальності» людей у змішаних системах *«чоловік-машина»* [8]. Вихідні дані для машини можуть здаватися *«де-суб'єктивованими»* і, таким чином, інтерпретуватися кінцевими користувачами, як більш об'єктивні, ніж вони є насправді, бо на справді, вони цілком залежать від даних навчальної вибірки, яку формує людина, а значить тут є фактор суб'єктивності. В цьому сенсі, може бути корисним структурувати обговорення між експертами з ML та іншими профільними експертами навколо трьох елементів, які відображають цей тип перевірки:

1. Вимоги к даним, які використовуються для навчання нейронної мережі за допомогою ML;
2. Мета навчання нейронної мережі;
3. Як результати впливають на подальші навчальні дані.

Ці елементи не експерти з ML можуть зрозуміти та оцінити.

Корисним припущенням для експертів, які не займаються ML, під час обговорення моделей ML є припущення, що алгоритм навчання, обраний експертом ML, є оптимальним для досягнення встановленої мети за допомогою заданих даних. Незважаючи на те, що це припущення багато разів хибне, воно має перевагу, оскільки робить більшу частину складності машинного навчання, наприклад знання того, як функціонують рекурентні нейронні мережі, неактуальними. Вважається, що це припущення може знизити планку для нефахівців щодо вступу в дискусію з експертами з ML і сприяти плідній дискусії.

Оптимальний у цьому контексті не є нормативним терміном і існує ключова відмінність між поняттями оптимального та доброго. Обчислення та статистика пропонують можливість економічно ефективного тестування величезної кількості можливих моделей. Метою алгоритму ML є визначення оптимальних параметрів для досягнення визначеної мети навчання, нехтуючи такими речами, як етичні проблеми, пов'язані з поліцейською діяльністю, якщо вони явно не реалізовані та запрограмовані [12]. Оптимізація означає вибір параметрів, які роблять

найточніші прогнози, враховуючи використані дані та навчання, щоб досягти найкращої продуктивності.

4. Питання про справедливість і обґрунтованість: інструментарій

Суспільство зацікавлено в запобіганні злочинності та ефективній поліцейській діяльності, але також зацікавлено в тому, щоб стратегії правоохоронних органів, включаючи рішення щодо розгортання та стеження, були ефективними, чесними та справедливими. Це вимагає розуміння, оцінки та управління [3].

Загалом кажучи, рішення можна критикувати з огляду на два різні питання: обґрунтованість рішення та справедливість рішення. Щоб розглянути валідність (*відповідність*) моделі, ми запитуємо: чи призвело рішення до запланованого результату? Щоб оцінити валідність, рецензенту потрібно буде розглянути: - чи модель навчання відображає фактичну ефективність на основі узгодженого показника ефективності, або сама метрика ефективності вимірює те, що ми мали намір виміряти. Оскільки цілі навчання можуть бути досить абстрактними та суперечливими (*наприклад, ціль зменшення рівня злочинності*), обсяг питань валідності, ймовірно, буде за межами предметної області для розуміння програмістів і статистиків. Однак, навіть досить «вузькі» питання, такі як упередженість відбору в навчальних даних, може бути легше викрита експертами, котрі не займаються ML і які можуть знати, наприклад, як збираються відповідні відомості. Так наприклад – інформація, швидше за все, буде зафіксована поліцейськими, якщо вони вважатимуть її корисною для успішного розкриття або запобігання злочину.

Перевірка справедливості рішення, прийнятого на основі «людської» або машинної моделі, передбачає запитання, чи були запланований результат і засоби його досягнення хорошими? Оцінка справедливості є нормативним завданням. У цьому контексті це означає, що мета навчання, процес, який покращує навчання, і засоби для досягнення успіху в навчанні визначені демократично легітимним шляхом. Забезпечення можливості відкритих і демократичних дебатів є як вимогою, так і частиною вирішення проблеми справедливості.

Далі наведено набір питань, які неексперти можуть поставити розробникам моделей з машинним навчанням, сподіваючись отримати зрозумілі відповіді. Відповіді у формі «проте ми врахували це в нашій моделі» вимагають рішень моделювання, які можна було б висловити явно, і ці рішення повинні бути застереженням для всіх, хто використовує модель. Інструментарій поділено на розділи з питаннями про дані, про навчання та про антиципаційну оцінку результатів. Мета інструментарію – надати можливість експертам, які не займаються ML, вести дебати з експертами з ML.

5. Дані для навчання MLмоделей

Злочини частіше за все фіксуються поліцією і лише зареєстровані злочини стають даними про злочини. Таким чином, статистика злочинності проходить процес відбору. Перша стадія процесу – законодавча; це коли певні діяння криміналізовані. В подальшому дані накопичуються в правоохоронних інформаційних системах. Дані категоризуються, частково структуруються та захищаються. Суспільство має доступ лише до частини відомостей про злочини. Більша частина даних є закритою від суспільства. Збір даних є суб'єктивним і залежіть від суб'єкта що їх збирає (*специфічного підрозділу*). Частина злочинів є латентною (прихованою) і не попадає в системи поліцейського обліку через те, що деякі злочини не повідомляються або не розкриваються громадськістю та поліцією.

В таких умовах, формування даних для навчання моделей ML зустрічається з проблемами:

- репрезентативність вибірки;
- актуальність;
- неупередженість даних.

Упередженість правоохоронних практик інколи можуть впливати на дані, створені поліцією. Дослідження *Human Rights Data Analysis Group* наводить показовий приклад [13]. Дослідження змоделювало прогнози правоохоронної діяльності з використанням алгоритму ML*«PredPol»* [14] на основі правоохоронних даних щодо боротьби з наркотиками в Окленді, Каліфорнія, а потім порівняло прогнози з моделями вживання наркотиків, оціненими на основі даних національного опитування про вживання наркотиків і здоров'я. Було виявлено, що за результатами роботи алгоритму *«PredPol»* «темношкірі люди африканського походження будуть об'єктом поліції по боротьбі з наркотиками приблизно вдвічі частіше», незважаючи на оцінки, які показують приблизно однакові рівні вживання наркотиків [13]. Люди з низьким рівнем доходу та не білошкірі, окрім темношкірих людей африканського походження, також будуть непропорційною мішенню, тобто надмірною цілю поліції.

Цей приклад упередженості показує, як вхідні дані, які використовуються для навчання машин і людей, можуть призвести до недійсних моделей і несправедливої практики. У цьому випадку недійсною моделлю або переконанням є те, що націлювання на житлові райони темношкірих людей є розумним способом поведінки поліції, незважаючи на те, що моделі вживання наркотиків свідчать про те, що в житлових районах темношкірих людей не повинно бути вищих випадків вживання наркотиків. Результатом є несправедлива правоохоронна практика, згідно з якою темношкірі громадяни та райони піддаються більшому нагляду, ніж білошкірі громадяни, незважаючи на відсутність об'єктивної основи в расових моделях злочинів, пов'язаних із наркотиками.

Таким чином, неексперти з ML, повинні задати наступні питання стосовно даних, що планується використовувати для навчання моделі ML:

Блок А:

- які вхідні дані використовуються?
- який набір використовувався для навчання моделі?
- який набір використовується для тестування продуктивності?
- коли, ким, як і де були зібрані дані?

Блок Б:

- чи є іменовані ознаки (змінні)?
- якщо так, то які вони і які найбільше впливають на результати?
- які операції з ними відбуваються та як вимірюються результати?

Блок В:

- - чи охоплюють вхідні дані функції *(прямо чи опосередковано)*, що не використовуються для прийняття рішення? Наприклад, чи пов'язані будь-які вхідні характеристики зі статтю таким чином, що модельні рішення відрізняються, якщо ви чоловік чи жінка?

Блок Г:

- чи дані репрезентативні, як впливають на результат роботи моделі? Наприклад, чи була модель перевірена в умовах, де вона застосована?
- які найбільш очевидні відмінності між умовами навчання та поточною роботою моделі?
- чи потрібно вносити якісь корективи для окремих груп даних чи результатів?

Блок Ґ:

- як збираються дані? Наприклад, чи їх збирали з наміром використовувати для таких рішень?
- чи знаємо ми про будь-які упередження відбору *(або через задум, або через практичні проблеми)* щодо збору даних?
- хто збирає дані?

6 Мета навчання. Постановка питань

Будь-яке навчання має мету. У моделях ML цілі можуть бути більш або менш явними. Незалежно від того, чи є навчання з вчителем, або ні, можна й доцільно запитати, якою є головна мета навчання та яке конкретне правило чи вимірювання використовується як еталон для визначення того, чи навчається модель.

Моделі ML оптимізуються відповідно до конкретних цілей навчання, які необхідно реалізувати та виміряти [15]. Оскільки деякі типи результатів легше виміряти, ніж інші, моделям ML властива упередженість щодо вибору навчальних цілей, які найлегше виміряти. Результати, які вже були виміряні, наприклад, місце арешту, стають привабливішими, ніж невиміряні результати, такі як реакція громадян на тактику поліції. Коли властива упередженість переноситься з машинних моделей на фактичне прийняття рішень, наслідки можуть бути різноманітними, як показує дослідження *HRDAG* [13].

Коли навчальна ціль є спірною або реалізуються далеко від ідеальної моделі, то прогнози таких моделей ML слід застосовувати з обережністю. Надзвичайний приклад можна знайти у By та Чжана [16], які стверджують, що їхня модель ML може автоматично ідентифікувати злочинців лише за характеристиками обличчя та «емпірично встановити достовірність автоматизованого висновку про злочинність за обличчям, незважаючи на історичні суперечки навколо цієї лінії дослідження». Тут модель не відокремлює злочинців від НЕ злочинців, а скоріше фотографії засуджених і підозрюваних із серії фотографій документів, отриманих з мережі Інтернет. Самі автори погоджуються з критиками, які стверджують, що різниця в соціально-економічному статусі в двох наборах може пояснити, чому моделі вдається розділити набори [16]. В цьому випадку мета щодо розпізнавання злочинців за обличчям є хибною з точки зору експертів в області криміналістики, і скоріш відображує розподіл неблагополучної частини населення та, відповідно, благополучної.

Існують два очевидних «занепокоєння» при розгляді використання моделі ML у процесах прийняття рішень:

- чи реалізована ціль закладена в моделі ML, та забезпечена ефективність у порівнянні з більш загальною та всеохоплюючою метою навчання?
- чи не створює операційна мета небажані побічні ефекти?

Крім того, варто окреслити ще одне занепокоєння, котре полягає в тому, що модель ML оптимізує багато, але не всі аспекти головної навчальної цілі [17]. При розробці моделі ML та оцінювання даних, які використовуються для навчання, деякі аспекти можуть бути втрачені. Обговорюючи головну мету процесуML та те, якими мають бути основні цілі навчання, можна визначити елементи, стосовно яких модель ML не оптимізується, і вжити відповідних заходів. Коли модель ML оптимізує лише деякі з встановлених цілей, необхідно бути обережними стосовно того, щоб модель ML вирішувала дії безпосередньо [18].

Таким чином, неексперти з ML, повинні задати наступні питання стосовно мети навчання моделі ML:

- яка основна мета навчання? Наприклад, чого б суспільство, хотіли досягти, приймаючи ці рішення?

- які конкретні правила та метрики використовуються як еталонні для визначення того, чи навчається модель? Наприклад, що таке залежна(і) змінна(и)?
- яке правило подібності використовується?
- які параметри навчання мають більшу вагу на роботу моделі ніж інші?
- як це правило реалізується та вимірюється?
- чи є згода щодо мети навчання?
- чи є конкретна ціль навчання повним описом того, чого ML має досягти?
- оптимізація дій або прийняття рішень щодо цієї навчальної цілі відніме зусиль або знову допоможе почати активно працювати?

7.Антиципаційна оцінка результатів, постановка питань

Наші моделі, машинні чи розумові, впливають на світ, коли ми використовуємо їх для прийняття рішень. У поліцейській діяльності головне, звичайно, зробити певний вплив на соціум. Прогнозний аналіз призначений для проактивної діяльності, «щоб визначити ймовірні цілі для втручання поліції». Рішення, дії, аналізи, політика, а також територіальний та історичний контексти сприяють формуванню сучасних концепцій у практиці правоохоронної діяльності. На відміну від, скажімо, фізики, поліцейські рішення впливають на соціальні системи. Ми використовуємо антиципаційний підхід, щоб позначити це розуміння.

Основна суть цього підходу полягає в тому, що експерти в поліцейській діяльності, які мають величезний досвід в питаннях організації процесів функціонування правоохоронної системи та результатів її роботи, спеціальний досвід *(слідчі, оперативні робітники, криміналісти та інші спеціалісти)*, мають так би мовити колективний розум, що сформований як результат сумаризації навчання та досвіду, кожного мозку спеціаліста на даних, якими він оперує на протязі професійної діяльності. Тобто при постановці задачі для машинного навчання, вони заздалегідь передбачають результат в рамках їх компетенції.

> *Вираз «колективний розум» використовується вже кілька десятиліть, але став важливим і популярним із приходом нових комунікаційних технологій. Він може викликати асоціації з груповою свідомістю або надприродними явищами, але технічно орієнтовані люди зазвичай розуміють під цим отримання нового знання з об'єднаних уподобань, поведінки та уявлень певної групи людей.*
>
> *Звичайно, колективний розум був можливим і до появи Інтернету. Для того, щоб збирати дані від розрізнених груп людей, об'єднувати їх та аналізувати, Всесвітня павутина зовсім не потрібна. До найважливіших форм подібних досліджень входять соціологічні опитування та переписи. Отримання відповідей від великої кількості людей дозволяє робити про групу такі статистичні висновки, які на основі поодиноких даних зробити неможливо. Породження нових знань виходячи з даних, отриманих від незалежних респондентів, – це і є суть колективного розуму*[19].

Різниця між природною нейронною мережею людини та штучною машини полягає в тому, що людина вчиться довгий проміжок часу на обмежених наборах даних, а машина короткий час на великих. Але людина має більш якісний набір даних, тому що постійно отримує зворотний зв'язок на протязі всього часу навчання з іншими спеціалістами. Мозок людини не здатний обробляти занадто великі обсяги даних, а машина здатна, але алгоритм обробки є алгоритмом роботи людського мозку. Спеціалісти з правоохоронної діяльності, можуть оцінити результати роботи моделі машинного навчання в правоохоронних задачах на основі свого досвіду, що не можуть зробити спеціалісти з машинного навчання.

Антиципаційний підхід допоможе вирішити три головних занепокоєння щодо застосування ML для прийняття рішень:

- по-перше, оцінити дані що застосовуються за їх повнотою, репрезентативністю, актуальністю та відповідності постановці задачі;
- по-друге, оцінити навчальні шаблони, та шаблони отримані в результаті роботи щодо наявності причинно-наслідкових зав'язків;
- по-третє, оцінити чи не суперечать отримані результати роботи алгоритмів машинного навчання історичній практиці.

Неексперти в ML але в експерти в поліцейській діяльності, повинні отримати відповіді на питання:

- чи може рішення машини, вплинути на пізніші дані навчання?
- чи модель машини представляє причинно-наслідковий зв'язок, чи це прагматичне рішення?
- чи модель спирається на кореляції, які, ймовірно, лише покращують ефективність через історичну практику?
- чи результати не суперечать історичній практиці?

8. Висновки

1. Оскільки поліцейські департаменти прагнуть одночасно запобігти, як заподіянню шкоди, так і ощадливо витрачати ресурси, то вони все частіше впроваджують проактивну політику і методи. Однак використання інструментів прогнозування вимагає ретельного розгляду і спільної експертизи, як експертами в ML, так і експертами з правоохоронної діяльності, що не є експертами з ML. Тільки сумісна робота, стосовно даних, цілей і конструктивної моделі використання ML, від початку подачі даних до етапу досягнення цілей, дасть можливість правильно оцінити результати роботи, використовувати результати в прийнятті рішень та створювати коректну ініціативу, щодо нормативного врегулювання використання моделей ML в поліцейській діяльності. Питання про мету використання технології ML у правоохоронній діяльності є, як моральними, так і політичними.

2. Головна мета полягає в тому, щоб розширити можливості нетехнічних експертів і зацікавлених сторін та заохотити їхню участь у роботі, щодо можливостей застосування ML у поліцейській діяльності, а також у процесах розробки профільної моделі ML. Така участь є не лише технічно й морально необхідною, але й можливою.

3.Сформовано перелік питань, які доцільно розглянути під час проведення сумісних обговорень і відповідних тематичних експертиз, щодо можливостей застосування технологій ML в поліцейській діяльності та пов'язаних із нею сферах боротьби зі злочинністю. В цьому сенсі питання, щодо коректності формування вихідних даних, мети навчання та того, як саме використовувані модельні рішення впливають на подальші дані, є 3-ма головними векторами можливих досліджень, які неексперти можуть переосмислити та завчасно скоригувати.

References

[1] Halterlein, J. and Ostermeier, L. (2018). 'Special Issue: Predictive Security Technologies'. European Journal for Security Research 3(2): 91–94.DOI: 10.1007/s41125-018-0034-z

[2] Hughes, D. (2017). 'Robot Investigators 'Could Be Used to Examine Documents in Criminal Cases'. The Independent (14 December 2017). https://www.independent.co.uk (accessed 20 November 2018).

[3] Bennett Moses, L. and Chan, J. (2016). 'Algorithmic Prediction in Policing: Assumptions, Evaluation, and Accountability'. Policing and Society 28 (7): 1–17.DOI: 10.1080/10439463.2016.1253695

[4] Ratcliffe, J. H. (2016). Intelligence-Led Policing. Abingdon, Oxon; New York: Routledge. DOI: 10.4324/9781315717579

[5] Innes, M. and Sheptycki, J. W. (2004). 'From Detection to Disruption: Intelligence and the Changing Logic of Police Crime Control in the United Kingdom'. International Criminal Justice Review 14(1): 1–24.DOI: 10.1177/105756770401400101

[6] Goldstein, J. (1960). 'Police Discretion Not to Invoke the Criminal Process: Low-Visibility Decisions in the Administration of Justice'. The Yale Law Journal 69(4): 543–594.

[7] Burrell, J. (2016). 'How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms'. Big Data & Society 3(1): 1–12.DOI: 10.1177/2053951715622512
[8] Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S. and Floridi, L. (2016). 'The Ethics of Algorithms: Mapping the Debate'. Big Data & Society 3(2): 1–21.DOI: 10.1177/2053951716679679
[9] Strukov V.M., Uzlov D.Yu. et al. Information Technologies in Law Enforcement. Part 1. High-Tech Trends in Law Enforcement of Foreign Countries. Kharkiv: LLC 'DISA PLUS', 2020. [In Ukrainian]
[10] Hughes, D. (2017). 'Robot Investigators 'Could Be Used to Examine Documents in Criminal Cases'. The Independent (14 December 2017). https://www.independent.co.uk (accessed 20 November 2018).
[11] Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S. and Floridi, L. (2016). 'The Ethics of Algorithms: Mapping the Debate'. Big Data & Society 3(2): 1–21.DOI: 10.1177/2053951716679679
[12] Norwegian Board of Technology (2018). Artificial Intelligence: Opportunities, Challenges and a Plan for Norway. Oslo: Norwegian Board of Technology. DOI: 10.5617/nmi.9950
[13] Lum, K. and Isaac, W. (2016). 'To Predict and Serve?'. Significance 13(5): 14–19.DOI: 10.1111/j.1740-9713.2016.00960.x
[14] Mohler, G. O., Short, M. B., Malinowski, S. et al. (2015). 'Randomized Controlled Field Trials of Predictive Policing'. Journal of the American Statistical Association 110(512): 1399–1411.DOI: 10.1080/01621459.2015.1077710
[15] Zhang, Lemoine, B. and Mitchell, M. (2018). Mitigating Unwanted Biases with Adversarial Learning. Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, ACM. pp. 335–340.DOI: 10.1145/3278721.3278779
[16] Wu, X. and Zhang, X. (2016). 'Responses to Critiques on Machine Learning of Criminality Perceptions' (Addendum of arXiv: 1611.04135). ArXiv: 1611.04135 [Cs]. http://arxiv.org/abs/1611.04135 (accessed 8 January 2019).
[17] Beck, C. and McCue, C. (2009). 'Predictive Policing: What Can we Learn from Wal-Mart and Amazon about Fighting Crime in a Recession?'. Police Chief 76 (11): 18–24
[18] Sklansky, D. A. (2008). Democracy and the Police. Stanford, CA: Stanford University Press. DOI: 10.1515/9780804763226
[19] Toby Segaran (2008) Programming Collective Intelligence. Published by O'Reilly Media Inc., ISBN-10:0-596-52932-5.

Submitted October 17, 2023; Revised November 19, 2023; Accepted December 20, 2023

Authors:
Dmytro Uzlov, Acting Dean of the Faculty of Computer Science, Candidate of Technical Sciences, Associate Professor, V. N. Karazin Kharkiv National University, Ukraine.
E-mail:dmytro.uzlov@karazin.ua
ORCID: https://orcid.org/0000-0003-3308-424X

Volodymyr Strukov, professor of the Department of Cybersecurity and DATA technologies, Candidate of Technical Sciences, Associate Professor, Kharkiv National University of Internal Affairs, Ukraine.
E-mail:struk_vm@ukr.net
ORCID: https://orcid.org/0000-0003-4722-3159

Vladyslav Hudilin, master's degree (cybersecurity), Kharkiv National University of Internal Affairs, Ukraine
E-mail:vgudilin7@gmail.com
ORCID: https://orcid.org/0000-0002-3844-1448

Oleksii Vlasov, Ph.Dcadidate, Kharkiv National University of Radio Electronics, Ukraine.
E-mail:moonreactor@gmail.com
ORCID: https://orcid.org/0000-0003-1619-0032

Problematic issues of machine learning technology in law enforcement.

Abstract. Law enforcement agencies increasingly use predictive and automation technologies where the core technology is often a machine learning (ML) model. The article considers the problem of accountability and responsibility of law enforcement agencies and officials connected with using of ML models. The authors point out that accountability is a key element of democratic law enforcement, but using of the predictive software can create challenges in ensuring that accountability. The article discusses how the application of ML can lead to obfuscation of responsibility and complicating accountability in «multi-agent structures» that combine humans and computational tools. Special attention is paid to the opacity of predictive algorithms and automated decision-making systems. It becomes a source of misunderstandings and caution regarding their use. The authors raise questions about how effective oversight and full reporting can be ensured when key components of the decision-making systems remain unknown to the general public, officials, and even developers of the models. The paper argues that important questions related to ML decision models can be solved without detailed knowledge of the machine learning algorithms, allowing non-ML law enforcement experts to study them in a form of intelligent control. Non-ML experts can and should review trained ML models. The authors provide a «toolkit» in the form of questions about three elements of the ML-based decision models that can be qualitatively explored by non-machine learning experts: training data, training goal, and anticipatory outcome evaluation. This approach expands the capabilities of these experts in the form of an objective assessment of the use of ML models in law enforcement tasks. This will allow them to evaluate effectiveness of the models through the prism of their own experience. The basic idea is that even without deep technical knowledge, law enforcement experts can analyze and review ML models. This approach promotes understanding of the use of machine learning technologies in law enforcement, expanding the potential of non-ML law enforcement experts.

Keywords*: Machine Learning, Artificial Intelligence, Data Analysis.*