

Вміння працювати в команді та координувати дії з іншими фахівцями з різних областей правоохоронної діяльності.

3.4. Інші вимоги до рівня професійної компетентності:

Здатність до постійного навчання та оновлення знань у сфері інформаційних технологій та кримінального аналізу;

Дисциплінованість та відповідальність у виконанні посадових обов'язків;

Здатність до прийняття швидких та обґрунтованих рішень у ситуаціях, що вимагають оперативності;

Високий рівень професійної етики та усвідомлення важливості дотримання законності та прав громадян;

Етичність та здатність до збереження конфіденційної інформації;

Креативність.

Олейніков Олег Анатолійович,

начальник відділу програмно-технічного забезпечення слідчої та оперативно-розшукової діяльності Управління інформаційних технологій Державного бюро розслідувань

МОЖЛИВОСТІ ТА ПЕРЕВАГИ ВИКОРИСТАННЯ КЛАСИФІКАЦІЙНИХ МОДЕЛЕЙ ШТУЧНОГО ІНТЕЛЕКТУ ПІД ЧАС РОЗСЛІДУВАННЯ КРИМІНАЛЬНИХ ПРАВОПОРУШЕНЬ

Сучасні методи аналізу великих обсягів даних вимагають спеціалізованих інструментів. Традиційні текстові та табличні процесори, як MS Excel, обмежені у своїх можливостях обробки великих даних через великі витрати на апаратні ресурси та максимальну межу в 1 048 576 рядків, що зумовлено обмеженою архітектурою формату.

Типові завдання аналітика правоохоронного органу найчастіше перевищують зазначену межу та потребують вирішення питання масштабування. Використання офісних додатків на межі максимального обсягу створюють ризики відмови програмного забезпечення або невиправдано значного часу опрацювання.

У зв'язку з цим, Data Science надає перевагу більш пристосованим засобам, як Python та R. Створені спеціалізовані процесори даних, такі як Jupyter Notebook та DataSpell, абстрагують користувача від класичного програмування,

надаючи гнучкість та відсутність жорстких обмежень на обсяг даних. Ці інструменти, у поєднанні з бібліотеками на кшталт *pandas* та *polars*, забезпечують ефективне виконання завдань, зазвичай вирішуваних традиційними текстовими та табличними процесорами.

Додатково з'являється можливість використовувати фрагменти програмного коду для виконання специфічних аналітичних задач або перевикористання вже існуючих сценаріїв обробки. За допомогою середовищ опрацювання стало можливим впровадження у повсякденну аналітичну діяльність засобів штучного інтелекту. Моделі машинного навчання, на відміну від побудови глибоких нейронних мереж, інтуїтивно зрозумілі та не потребують надлишкових апаратних ресурсів для тренування.

Основним принципом підходу застосування моделей машинного навчання є відмова від створення алгоритму. Натомість для створення моделі надається зразок даних з промаркованим результатом (тренувальна вибірка). Для класифікаційних моделей типовим є маркування «ІСТИНА» (True) або «ХИБА» (False). Під час опрацювання тренувальних даних модель найкращим чином знаходить закономірності у даних та створює внутрішній порядок прийняття рішень. За умови надання достатньо різноманітної тренувальної вибірки, тренувана модель набуває здатності до узагальнення та можливості роботи з даними, які раніше їй були невідомі. Таким чином тренувана модель може бути перевикористана у аналогічних випадках.

Для прикладу пропонується розглянути модель виявлення помилок у написанні анкетних даних осіб. Дані про перетин державного кордону містять помилки, допущені операторами введення. Так, наприклад помилки «Олейнік – Олейник», «Ямпільський - Йампільський» мають місце при опрацюванні значних зрізів даних.

При перевірці великого датасету за участю 88000 осіб, виявлено ознаки схожості у ~2000 осіб (однакова дата народження та схоже написання імені та прізвища). Опрацювання такої кількості значень ручним способом вимагає значних часових витрат та підвищує ризик пропуску, а відмова від нормалізації призводить до втрати зв'язків під час аналізу зв'язків з використанням графічних засобів (наприклад *i2 Analyst Notebook*).

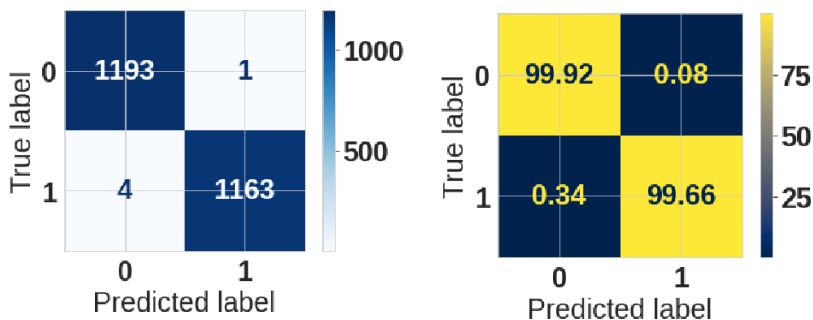
Для вирішення проблеми виявлення помилкових написань підготовлена тренувальна вибірка з промаркованими статусами для пар схожих осіб:

ПІ_особа1	ПІ_особа2	С_особа1	С_особа2	... інші ознаки	Статус
Саргалиєва	Саргалиєва	Жіноча	Жіноча	...	Одна особа
Іванов	Іванова	Чоловіча	Жіноча		Різні особи
...	...	...	...	...	...

З наявних даних сформовано ознаки схожості (features) для кожної з пар:

Опис ознаки	Назва ознаки	Пара осіб 1	Пара осіб 2	Інші пари
Маркований результат (чи одна особа)	target	0	1	...
Однакова стать	same_sex	1	1	...
Однаковий тип документа	same_doc_type	1	1	...
Однаковий номер документа	same_doc_num	0	0	...
Однакова країна походження	same_country	0	1	...
Співпадіння написання українською	same_ukr_full	0	0	...
Співпадіння написання англійською	same_eng_full	0	0	...
Зважена різниця написання прізвища	ukr_sname_diff	0.33	0.90	...
Зважена різниця написання імені	ukr_fname_diff	0.44	0.85	...

Після тренування з допомогою навчальної вибірки, модель машинного навчання може аналізувати нові дані, не потребуючи їх попереднього маркування. Вона здійснює автоматичні передбачення, визначаючи, чи є подана пара даних помилковим дублікатом однієї особи, чи належить вона до різних осіб. Це дозволяє швидко і точно ідентифікувати помилки.



Матриця помилок.

True label – дійсні значення, *Predicted label* – робота моделі

Результати валідації моделі вказують на достатню ефективність. Серед 2361 пар схожих анкет правильно визначено 2356 пар тотожних осіб.

Для задачі визначення тотожності осіб використана класифікаційна модель Decision Tree Classifier (Дерево рішень). Для схожих задач можливе використання більш надійних моделей, таких як: ансамблеві моделі – Random Forest (Випадковий ліс), моделі, що використовують градієнтний спуск виправлення помилки – lgbm (Light Gradient Boosting Machine), XGBoost (Extreme Gradient Boost).

Переваги класифікаційних моделей:

- самостійно формують кращий алгоритм для великої кількості факторів;
- придатні до перевикористання у аналогічних випадках;
- можливі для впровадження малими командами чи окремими аналітиками;
- не використовують зовнішні сервіси.

Недоліки класифікаційних моделей:

- потребують матеріал для навчання;
- залежать від помилок учителя;
- обмежені у передбаченні кількісних значень;
- менш ефективні ніж моделі глибокого навчання.

Наразі існують повні посібники для впровадження методів машинного навчання, в тому числі з наочною демонстрацією роботи моделей з використанням найбільш популярного програмного пакету sklearn: www.scikit-learn.org. Всесвітньою практикою набуття професійних навичок вважається участь у змаганнях аналітиків на платформі www.kaggle.com.

Використання класифікаційних моделей машинного навчання мають широкі можливості застосування у типових аналітичних задачах аналітика правоохоронного органу. Зокрема: визначення призначення колонок під час опрацювання таблиць з несталою структурою, визначення скорінгу використання конспіративних засобів зв'язку серед абонентів радіомоніторингу, визначення переліку господарських товариств з підвищеним ризиком до бюджетних злочинів, виділення найбільш перспективних об'єктів аналізу до оперативної перевірки, інші задачі, що потребують аналізу одночасно значної кількості ознак з неочевидною кореляцією.

Впровадження моделей машинного навчання та сучасних середовищ опрацювання даних може значно підвищити швидкість та якість роботи аналітика правоохоронного органу, розширити спектр виконуваних задач, отримати нові підходи до здобуття аналітичної інформації.

Олексин Христина Любомирівна,
курсант навчально-наукового інституту № 3
Національної академії внутрішніх справ
Науковий керівник:
Мостепанюк Людмила Олександрівна,
доцент кафедри кримінального права
Національної академії внутрішніх справ,
кандидат юридичних наук, доцент

ТЕНДЕНЦІЇ КРИМІНАЛЬНО-ПРАВОВОГО ЗАХИСТУ ПРАЦІВНИКІВ ПРАВООХОРОННИХ ОРГАНІВ

Як відомо, Конституцією України було проголошено захист прав і свобод людини і громадянина, а також забезпечення державного суверенітету і територіальної цілісності України, зокрема, належна охорона економічної та інформаційної системи, як важливих складових в нормальному функціонуванні державного апарату. Повне та досконале виконання цих простих завдань є підґрунтям для розбудови суспільства та держави і подальшої ефективної інтеграції нашої країни до Європейського Союзу. До того ж варто враховувати, що для виконання цих завдань невід'ємною складовою є ефективна діяльність правоохоронних органів, що потребує належного кримінально-правового захисту, особливо під час дії військового стану, який був введений на території України Указом Президента від 24.02.2022 р. Враховуючи вище зазначене, державою було взято