



DOI 10.28925/2663-4023.2026.35.1406

УДК 004.85:004.62

Лобода Юлія Геннадіївна

кандидат педагогічних наук, доцент, доцент кафедри програмної інженерії,
Національний університет «Одеська юридична академія», Одеса, Україна,
ORCID: 0000-0001-7083-552X
loboda@onua.edu.ua

Трофименко Олена Григорівна

кандидат технічних наук, доцент, доцент кафедри програмної інженерії,
Національний університет "Одеська юридична академія", Одеса, Україна,
ORCID: 0000-0001-7626-0886
trofymenko@onua.edu.ua

Ворона Михайло Владиславович

доктор філософії, викладач кафедри інформаційних управляючих систем та технологій,
Національний університет кораблебудування імені адмірала Макарова, Миколаїв, Україна,
ORCID: 0000-0003-4288-0096
mykhailo.vorona@nuos.edu.ua

Гайдаєнко Оксана Володимирівна

кандидат технічних наук, доцент, доцент кафедри інформаційних управляючих систем та технологій,
Національний університет кораблебудування імені адмірала Макарова, Миколаїв, Україна,
ORCID: 0000-0002-6614-5443
oksana.gaidaienko@nuos.edu.ua

**ВИЯВЛЕННЯ ДЕГРАДАЦІЇ МОДЕЛЕЙ МАШИННОГО НАВЧАННЯ НА ОСНОВІ СПІЛЬНОГО
АНАЛІЗУ ДРЕЙФУ ДАНИХ І МЕТРИК ЯКОСТІ**

Анотація. Моделі машинного навчання після розгортання працюють у середовищі, статистичні властивості якого можуть змінюватися, тому збереження параметрів моделі не гарантує сталої прогнозовної якості. Водночас сам факт дрейфу вхідних даних не є доказом деградації, а безпосереднє оцінювання якості часто ускладнене через затримку або часткову доступність фактичних цільових значень. Метою дослідження є розроблення методу виявлення деградації моделей машинного навчання, який поєднує аналіз змін вхідних даних із контролем прогнозовної якості та враховує доступність фактичних цільових значень під час формування рішення. Запропонований метод передбачає окреме оцінювання дрейфу вхідних ознак, зміни розподілу прогнозів і відхилення метрик якості від еталонних значень. Рівні дрейфу вхідних ознак і зміни розподілу прогнозів формуються за індексом стабільності популяції. За наявності достатньої кількості верифікованих спостережень стан моделі визначається за дрейфом вхідних ознак і зміною прогнозовної якості, що дає змогу розмежувати стабільну роботу, дрейф без підтвердженої деградації, зниження якості без зафіксованого дрейфу та їх одночасний прояв і визначити відповідний рівень реагування. За відсутності фактичних цільових значень або затримки їх надходження формується лише оцінка ризику за змінами вхідних ознак і розподілу прогнозів, яка уточнюється після отримання фактичних значень. Метод реалізовано в інформаційній системі моніторингу якості моделей машинного навчання та перевірено на задачах класифікації і регресії у контрольованих сценаріях зміни даних. Експериментальні результати показали, що виражений дрейф вхідних ознак може не супроводжуватися погіршенням прогнозовної якості, а критична деградація моделі може виникати без суттєвої зміни контрольованих вхідних розподілів. Також показано, що оцінка ризику, сформована до верифікації прогнозів, може бути уточнена після надходження фактичних значень. Запропонований підхід дає змогу розмежовувати статистичні зміни даних і фактичне погіршення прогнозовної якості та уникати необґрунтованого висновку про стабільний стан моделі за відсутності інформації для оцінювання її якості.

Ключові слова: аналіз даних; машинне навчання; дрейф даних; деградація моделі; індекс стабільності популяції; метрики прогнозовної якості; ухвалення рішень; розподіл прогнозів.



ВСТУП

Постановка проблеми. Моделі машинного навчання після розгортання функціонують у середовищі, статистичні властивості якого можуть змінюватися з часом. Розподіли вхідних ознак, склад об'єктів спостереження та залежності між ознаками і цільовою змінною не обов'язково залишаються такими самими, як у даних, використаних для навчання моделі. Унаслідок цього модель, параметри якої залишаються незмінними, може поступово втрачати прогнозну якість. У прикладних задачах це означає, що рішення продовжують ухвалюватися на підставах, які вже не відповідають дійсності. Водночас сама наявність зміни розподілу вхідних даних ще не означає погіршення роботи моделі: зсув окремих ознак може не впливати на прогнозну здатність, тоді як зниження якості може відбутися без помітної зміни вхідних розподілів.

Особливо складною є ситуація, коли фактичне значення цільової змінної стає відомим не одночасно з формуванням прогнозу, а із затримкою. Для частини прикладних задач така затримка визначається природою самого процесу: результат медичного прогнозу може бути встановлений після завершення певного періоду спостереження, а у задачах прогнозування відтоку персоналу факт звільнення стає відомим лише після настання відповідної події. Унаслідок цього виникає часовий розрив між моментом формування прогнозу та моментом, коли прогнозну якість моделі можна безпосередньо оцінити на основі фактичних цільових значень. У межах цього дослідження такий часовий розрив розглядається як затримка верифікації; для окремих задач він може вимірюватися місяцями.

За таких умов моніторинг ґрунтується на двох групах показників. Показники дрейфу дають змогу виявляти зміни статистичних властивостей доступних вхідних даних без використання фактичних цільових значень, однак не визначають, чи призвели ці зміни до втрати прогнозної якості. Метрики прогнозної якості, навпаки, безпосередньо характеризують результати роботи моделі, але можуть бути обчислені лише після надходження фактичних значень. Отже, використання кожної з цих груп показників окремо не дає достатньої підстави для однозначного визначення стану моделі.

Загалом задача моніторингу полягає у спільній інтерпретації показників дрейфу та метрик якості з урахуванням доступності фактичних цільових значень. За їх наявності необхідно розмежовувати стабільний стан моделі, дрейф вхідних даних без підтвердженої деградації, погіршення прогнозної якості без виявленого дрейфу вхідних даних та дрейф вхідних даних із підтвердженою деградацією. За їх відсутності або недостатньої кількості остаточний висновок про деградацію не може бути сформований, тому доцільним є формування попередньої оцінки ризику з її подальшим уточненням після надходження фактичних значень.

Аналіз останніх досліджень і публікацій. Проблема деградації моделей машинного навчання після розгортання розглядають у наукових працях переважно у двох напрямках: через виявлення змін у робочих даних та через контроль результатів роботи моделі. У роботі [1] систематизовано підходи до виявлення дрейфу, орієнтовані на зміну прогнозної якості моделі, та розглянуто зв'язок між дрейфом і деградацією моделі. У дослідженні [2] методи виявлення дрейфу поділено на явні (explicit, supervised), що потребують розмічених даних для обчислення й моніторингу метрик прогнозної якості, та неявні (implicit, unsupervised), які виявляють зміни за характеристиками нерозмічених вхідних даних. Отже, ці групи підходів використовують різну інформацію про стан моделі та мають різні обмеження щодо застосування в умовах її експлуатації.

Окрему увагу в сучасних дослідженнях приділено співвідношенню між дрейфом даних і фактичним погіршенням роботи моделі. У роботі [3] запропоновано розрізняти зміни даних за характером їхнього впливу на прогнозну точність та визначати області низької точності, щоб повторне навчання виконувалося лише за дрейфу, який негативно впливає на модель. Для медичних систем у дослідженні [4] показано, що зміни характеристик популяції після розгортання моделі потребують окремого оцінювання їхнього впливу на прогнозні властивості. Під час аналізу часової деградації моделей у роботі [5] також встановлено, що статистичні зміни даних і втрата прогнозної якості не є тотожними явищами. Отже, сам факт зміни розподілу вхідних даних не може розглядатися як достатня підстава для висновку про деградацію моделі.

Значна частина публікацій присвячена статистичним засобам виявлення змін у робочих даних. У роботі [6] проаналізовано Population Stability Index (PSI) та показано обмеження використання фіксованих порогів цього показника без урахування обсягу вибірки. Загальні можливості методів математичної статистики для кількісного аналізу наборів даних розглянуто також у роботі [7]. Для порівняння емпіричних розподілів застосовується критерій Колмогорова–Смирнова [8]. У дослідженні [9] використано Maximum Mean Discrepancy (MMD) для виявлення змін у багатовимірних представленнях даних і локалізації спостережень, пов'язаних із дрейфом. Підхід до послідовного моніторингу відхилень у часі та формування сповіщень лише для змін, що мають практичне значення, розглянуто у роботі [10].



Наведені підходи розширюють можливості статистичного контролю, однак переважно характеризують сам факт або величину зміни даних, а не її наслідки для прогнозної якості моделі.

Інший напрям досліджень пов'язаний із використанням інформації про поведінку самої моделі. За наявності фактичних цільових значень стан моделі можна оцінювати за метриками прогнозної якості, проте в реальних системах такі значення не завжди доступні одразу після формування прогнозу. Тому для моніторингу без міток досліджуються непрямі показники, сформовані як на основі виходів прогнозної моделі, так і за допомогою окремих моделей виявлення відхилень. У роботі [11] для виявлення відхилень у робочому потоці використано зміни розподілу впевненості класифікатора. У дослідженні [12] проаналізовано методи калібрування оцінок упевненості для виявлення ненадійних прогнозів за умов зміни даних. У роботі [13] як сигнал зміни умов функціонування системи застосовано похибку відновлення автокодувальника. Такі підходи дають змогу отримувати інформацію про зміну поведінки моделі до надходження фактичних значень, однак використовувані показники залишаються непрямими оцінками її стану.

Проблема несвоєчасної доступності фактичних цільових значень сформувала окремий напрям досліджень потокового машинного навчання. У праці [14] систематизовано методи роботи з потоками даних за умов відкладених і частково доступних міток та розглянуто вплив затримки їх надходження на методи навчання з учителем і напівконтрольованого навчання. Вплив затримки верифікації на процес навчання в нестационарних середовищах досліджено у роботі [15]. Часова природа надходження фактичних цільових значень характерна і для прикладних задач прогнозування подій. Зокрема, у роботі [16] для аналізу відтоку персоналу поєднано класифікаційні моделі з аналізом виживання, що дає змогу враховувати не лише факт настання події, а й час до її виникнення. Сукупність цих досліджень свідчить, що несвоєчасна доступність фактичних цільових значень обмежує можливість безпосереднього контролю прогнозної якості та робить актуальним використання проміжних сигналів до появи фактичних результатів.

Сучасні дослідження демонструють перехід від використання окремих показників до поєднання кількох джерел моніторингової інформації. У роботі [17] запропоновано дворівневий підхід, у якому моніторинг змін даних доповнюється контролем поведінки моделі. У дослідженні [18] виявлення змін пов'язано з визначенням доцільності повторного навчання, що дає змогу уникати необґрунтованого оновлення моделі у стабільні періоди. У роботі [19] запропоновано метод оновлення клінічних прогностичних моделей, спрямований на усунення калібрувального дрейфу, з одночасним моніторингом показників роботи моделі та розподілу даних. Спільне використання засобів виявлення зсуву вхідних даних та безміткового показника, сформованого за впевненістю прогнозів, досліджено у роботі [20]. Отримані результати показують, що зсув вхідних даних не в усіх випадках супроводжується зниженням прогнозної якості. У дослідженні [21] проаналізовано моніторинг різних типів дрейфу в конвеєрах машинного навчання у виробничому середовищі та механізми реагування на виявлені зміни. Проведений аналіз свідчить, що сучасні підходи вже поєднують сигнали від даних і моделі, оцінювання впливу змін на прогнозну якість та механізми повторного навчання. Водночас у проаналізованих роботах не виявлено формалізації доступності фактичних цільових значень як параметра, що визначає спосіб формування висновку та рівень реагування, а також цілісної процедури переходу від попередньої оцінки до підтвердженого стану після верифікації.

Метою роботи є розроблення методу виявлення деградації моделей машинного навчання, який поєднує показники дрейфу вхідних даних із метриками прогнозної якості та враховує доступність фактичних цільових значень як параметр, що визначає спосіб формування рішення, а також експериментальна перевірка сформованих ним оцінок стану моделі та ризику деградації порівняно з рішеннями, отриманими за окремими показниками.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Запропонований метод ґрунтується на спільному аналізі трьох типів моніторингових сигналів: зміни розподілів вхідних даних, зміни розподілу прогнозів та зміни прогнозної якості моделі. Спосіб формування рішення залежить від доступності фактичних цільових значень. За достатньої кількості верифікованих спостережень стан моделі визначається за показниками дрейфу вхідних даних і метриками прогнозної якості. Якщо фактичні цільові значення відсутні або їх кількість недостатня для коректного обчислення метрик, формується попередня оцінка ризику на основі зміни вхідних даних і розподілу прогнозів. Після надходження достатньої кількості фактичних цільових значень попередня оцінка уточнюється.

1. Показники дрейфу вхідних даних

Для формування сигналу про зміну вхідних даних оцінюються розподіли контрольованих ознак відносно еталонної вибірки. Як основний показник використано індекс стабільності популяції (Population



Stability Index, PSI), властивості та обмеження якого проаналізовано в [6]. Для числових ознак інтервали розбиття формуються за еталонними даними і надалі залишаються незмінними для всіх періодів моніторингу; для категоріальних ознак аналізуються частки відповідних категорій.

Для j -ої ознаки у періоді t значення показника визначається як:

$$PSI_{j,t} = \sum_{i=1}^{k_j} (p_{ji} - q_{ji,t}) \ln \frac{p_{ji}}{q_{ji,t}}, \quad (1)$$

де p_{ji} – частка спостережень i -го інтервалу або категорії в еталонній вибірці; $q_{ji,t}$ – відповідна частка у поточному періоді; k_j – кількість інтервалів або категорій для j -ої ознаки. Для уникнення невизначеності логарифмічного виразу за нульових часток у програмній реалізації застосовується згладжування $\epsilon=10^{-4}$.

Оскільки модель може одночасно використовувати декілька контрольованих ознак, для формування узагальненого сигналу дрейфу необхідне агрегування отриманих значень.

У запропонованому методі використовується максимальне значення показника:

$$PSI_{max,t} = \max_{j=1,\dots,m} PSI_{j,t} \quad (2)$$

де m – кількість контрольованих вхідних ознак.

За значенням $PSI_{max,t}$ формується рівень дрейфу вхідних даних L_D . Для цього використовуються два пороги ρ_1 і ρ_2 , які розділяють відсутній, помірний та виражений рівні зміни. Початкові значення $\rho_1=0,10$ і $\rho_2=0,25$ традиційно використовуються в практиці моделювання кредитного ризику і мають евристичний характер; у роботі [6] показано обмеження застосування таких фіксованих порогів без урахування обсягу вибірки. Тому в системі вони задаються як параметри конкретної моделі та можуть бути скориговані відповідно до особливостей предметної області.

Додатковим засобом перевірки зміни числових ознак є критерій Колмогорова–Смирнова [8]. У реалізованому методі він застосовується до числової ознаки j^* , для якої отримано максимальне значення індексу стабільності популяції:

$$D_t = \max_x |F_{j^*,t}^{pot}(x) - F_{j^*,t}^{et}(x)|, \quad j^* = \arg \max_j PSI_{j,t} \quad (3)$$

Статистика D_t характеризує максимальне розходження між емпіричними функціями розподілу еталонної та поточної вибірок. Разом із нею фіксується результат перевірки статистичної значущості за заданим рівнем α . Критерій Колмогорова–Смирнова виконує додаткову діагностичну функцію і не бере участі у визначенні рівня L_D , який формується на основі $PSI_{max,t}$.

2. Показники прогнозової якості

Другий тип сигналу формується за метриками прогнозової якості та може бути використаний лише тоді, коли для достатньої кількості спостережень відомі фактичні цільові значення. Для задач класифікації система підтримує частку правильних відповідей (аccuracy), точність (precision), повноту (recall), F-міру (F1-score) та площу під характеристичною кривою (ROC AUC), а для задач регресії – середню абсолютну похибку (MAE), корінь із середньоквадратичної похибки (RMSE) і коефіцієнт детермінації (R^2). Для кожної зареєстрованої метрики задається еталонне значення та напрям її інтерпретації.

Оскільки для одних метрик погіршенню відповідає зменшення значення, а для інших його збільшення, відносне погіршення r -ої метрики у періоді t визначається з урахуванням напрямку:

$$d_{r,t} = \begin{cases} \frac{M_{r,0} - M_{r,t}}{|M_{r,0}|}, & \text{якщо більше значення відповідає кращій якості;} \\ \frac{M_{r,t} - M_{r,0}}{|M_{r,0}|}, & \text{якщо менше значення відповідає кращій якості,} \end{cases} \quad (4)$$

де $M_{r,0}$ – еталонне значення метрики, а $M_{r,t}$ – її значення у поточному періоді.

За такого визначення додатне значення $d_{r,t}$ означає погіршення прогнозової якості незалежно від типу показника.

Якщо для моделі зареєстровано декілька метрик, узагальнена величина деградації визначається за найбільшим зафіксованим відносним погіршенням:

$$\Delta_t = \max_r (0, \max d_{r,t}). \quad (5)$$

Від'ємні значення, що відповідають покращенню метрики порівняно з еталонним рівнем, не розглядаються як деградація. За величиною Δ_t та порогоми δ_1 і δ_2 формується рівень зміни прогнозової якості L_M . Порогові значення задаються окремо для кожної моделі, оскільки допустима величина погіршення залежить від її призначення та наслідків помилкового прогнозу.

3. Зміна розподілу прогнозів

За відсутності фактичних цільових значень або їх недостатньої кількості безпосереднє оцінювання прогнозової якості неможливе. Для отримання додаткового сигналу в такій ситуації використовується



розподіл виходів моделі. Для класифікаційних моделей аналізуються прогнозовані ймовірності класу, для регресійних – прогнозовані числові значення. Використання власних виходів моделі як непрямого сигналу досліджувалося раніше [11], [12], [20]. У запропонованому методі для кількісного оцінювання зміни їх розподілу застосовано той самий показник PSI, що й для вхідних ознак. Це дає змогу отримувати додаткову інформацію про зміну результатів прогнозування до надходження достатньої кількості фактичних цільових значень.

PSI для розподілу виходів моделі $PSI_{\text{прогн},t}$ обчислюється за формулою (1), але не включається до $PSI_{\text{max},t}$ вхідних ознак. Тим самим L_D характеризує зміни вхідних даних, а L_P – зміну розподілу прогнозів. Рівень L_P формується за порогами ρ_1 і ρ_2 подібно до рівня L_D .

Еталонний розподіл виходів формується за позаблоковими прогнозами, отриманими під час кросвалідації навчальної вибірки. Той самий вектор прогнозів використовується для визначення еталонних метрик прогнозової якості, що забезпечує узгодженість компонентів еталонного стану та виключає оцінювання прогнозів на спостереженнях, використаних для навчання відповідної моделі.

Значення L_P обчислюється незалежно від наявності фактичних цільових значень і зберігається у протоколі моніторингу. За наявності достатньої кількості верифікованих спостережень цей показник має діагностичне значення, але не впливає на остаточне визначення стану моделі: рішення формується за L_D і L_M .

4. Урахування доступності фактичних цільових значень

Щоб урахувати доступність фактичних цільових значень, введемо змінну A_t , яка визначає можливість оцінити прогнозу якість у періоді t :

$$A_t = \begin{cases} 1, & n_t^{\text{вер}} \geq n_{\min} \text{ та зареєстровані метрики можуть бути коректно обчислені;} \\ 0, & \text{в іншому випадку,} \end{cases} \quad (6)$$

де $n_t^{\text{вер}}$ – кількість верифікованих спостережень у поточному періоді; $n_{\min}$ – мінімальна кількість верифікованих спостережень, необхідна для безпосереднього оцінювання прогнозової якості. У проведених експериментах використано $n_{\min}=30$, у системі цей поріг є параметром налаштування.

Якщо $A_t=1$, рівень L_M набуває одного зі значень 0, 1 або 2 залежно від величини Δ_t . Якщо $A_t=0$, рівень зміни прогнозової якості вважається невизначеним $L_M=\perp$.

Невизначене значення принципово не отожднюється з $L_M=0$. Відсутність можливості оцінити прогнозу якість не означає відсутності її погіршення. Зокрема, за $L_D=0$ і $L_M=\perp$ ознаки дрейфу не виявлено, а зміна прогнозової якості не може бути безпосередньо оцінена. Узагальнене тлумачення рівнів L_D , L_P та L_M наведено в табл. 1.

Таблиця 1

Рівні моніторингових сигналів

Сигнал	Рівень	Умова	Змістове тлумачення
L_D	0	$PSI_{\text{max},t} < \rho_1$	суттєвих змін вхідних розподілів не виявлено
L_D	1	$\rho_1 \leq PSI_{\text{max},t} < \rho_2$	помірна зміна вхідних даних
L_D	2	$PSI_{\text{max},t} \geq \rho_2$	виражена зміна вхідних даних
L_P	0	$PSI_{\text{прогн},t} < \rho_1$	суттєвої зміни розподілу прогнозів не виявлено
L_P	1	$\rho_1 \leq PSI_{\text{прогн},t} < \rho_2$	помірна зміна розподілу прогнозів
L_P	2	$PSI_{\text{прогн},t} \geq \rho_2$	виражена зміна розподілу прогнозів
L_M	0	$\Delta_t < \delta_1, A_t=1$	суттєвого зниження прогнозової якості не виявлено
L_M	1	$\delta_1 \leq \Delta_t < \delta_2, A_t=1$	помірне зниження прогнозової якості
L_M	2	$\Delta_t \geq \delta_2, A_t=1$	суттєве зниження прогнозової якості
L_M	$\perp$	$A_t=0$	прогнозна якість безпосередньо не оцінюється

5. Правило визначення стану моделі

Спосіб ухвалення рішення визначається значенням A_t . Якщо $A_t=1$, стан моделі встановлюється за рівнями дрейфу вхідних даних і зміни прогнозової якості:

$$A_t=1: f(L_D, L_M) \rightarrow (S, R) \quad (7)$$

де S – підтверджений діагностичний стан моделі; R – відповідний рівень реагування.

Комбінації L_D і L_M утворюють чотири змістові класи станів: стабільний стан; дрейф вхідних даних без підтвердженої деградації; погіршення прогнозової якості без виявленого дрейфу вхідних даних; дрейф вхідних даних із підтвердженою деградацією. Значення 1 і 2 для L_D та L_M характеризують ступінь



вираженості відповідного сигналу, але не утворюють нових діагностичних класів. Відповідність комбінацій L_D і L_M станам моделі та рекомендованим діям наведено в табл. 2.

Таблиця 2

Інтерпретація стану моделі за наявності достатньої кількості фактичних цільових значень

$L_D \setminus L_M$	0	1	2
0	стабільний стан; штатний моніторинг	деградація без виявленого дрейфу; можлива зміна залежності; повторне навчання на актуалізованих даних	критична деградація; рекомендація щодо невідкладного повторного навчання
1	дрейф без підтвердженої деградації; посилений контроль	дрейф із підтвердженою деградацією; підготовка до повторного навчання	критична деградація; рекомендація щодо невідкладного повторного навчання
2	виражений дрейф без підтвердженої деградації; посилений контроль	дрейф із підтвердженою деградацією; підготовка до повторного навчання	критична деградація; рекомендація щодо невідкладного повторного навчання

Наявність дрейфу за $L_M = 0$ не є підставою для твердження про деградацію моделі. У такому разі зміна вхідних даних розглядається як випереджальний сигнал, що потребує посиленого контролю. Значення $L_M = 2$ відповідає критичному погіршенню прогнозної якості незалежно від того, чи виявлено одночасно дрейф вхідних даних. Комбінація $L_D = 0, L_M = 1$ означає погіршення якості без зафіксованої зміни вхідних розподілів; це може бути пов'язано зі зміною залежності між ознаками та цільовою змінною або з іншими змінами, які не відображаються моніторингом $P(X)$. Для такого стану передбачено повторне навчання на актуалізованих даних із фактичними цільовими значеннями.

Якщо $A_t = 0$, підтверджений стан S не визначається. Оцінка формується за рівнями дрейфу вхідних даних і зміни розподілу прогнозів:

$$A_t = 0: \quad g(L_D, L_P) \rightarrow (U, R) \quad (8)$$

де U – діагностична характеристика спостережуваних змін, R – рівень ризику.

Оскільки L_D і L_P не характеризують фактичну прогнозну якість безпосередньо, навіть найвищий рівень R не означає підтверджену деградацію і не є підставою для автоматичного повторного навчання.

Рівень попереднього ризику визначається за максимальним із L_D і L_P . Якщо обидва сигнали одночасно є ненульовими, ризик підвищується до найвищого рівня:

$$R = \begin{cases} 0, & L_D = 0, L_P = 0; \\ 1, & \max(L_D, L_P) = 1 \text{ і лише один сигнал є ненульовим}; \\ 2, & \max(L_D, L_P) = 2 \text{ або } (L_D \geq 1 \wedge L_P \geq 1). \end{cases} \quad (9)$$

Одночасна зміна вхідних розподілів і розподілу прогнозів є сильнішим непрямим сигналом, ніж зміна лише одного з них. Проте за відсутності достатньої кількості фактичних цільових значень вона не підтверджує погіршення прогнозної якості. Відповідність комбінацій L_D і L_P рівням ризику наведено в табл. 3.

Таблиця 3

Оцінювання ризику за відсутності достатньої кількості фактичних цільових значень

$L_D \setminus L_P$	0	1	2
0	спостережуваних змін не виявлено; якість не оцінюється; $R=0$	зміна розподілу прогнозів; $R=1$	виражена зміна розподілу прогнозів; $R=2$
1	помірний дрейф входів; $R=1$	одночасна зміна входів і прогнозів; $R=2$	одночасна зміна входів і прогнозів; $R=2$
2	виражений дрейф входів; $R=2$	одночасна зміна входів і прогнозів; $R=2$	виражена зміна входів і прогнозів; $R=2$

Рекомендовані дії залежать від доступності фактичних цільових значень. За підтвердженого суттєвого погіршення прогнозної якості система рекомендує повторне навчання. Якщо фактичних цільових значень недостатньо, навіть $R=2$ означає високий ризик, а рекомендованими діями є пріоритетне отримання фактичних цільових значень за можливості та посилений моніторинг. За $R=1$ формується сповіщення про відхилення, за $R=0$ моніторинг продовжується без висновку про стабільність прогнозної якості.



Після надходження достатньої кількості фактичних цільових значень L_M стає визначеним, і стан моделі S обчислюється за правилом $f(L_D, L_M)$. Значення L_D і L_P , отримані для відповідного періоду до верифікації, зберігаються у протоколі. Це дає змогу зіставити оцінку ризику, сформовану до верифікації, із підтвердженим станом після надходження фактичних значень.

6. Програмна реалізація методу

Метод реалізовано в інформаційній системі моніторингу якості моделей машинного навчання, побудованій за клієнт-серверною архітектурою. Серверну частину реалізовано засобами FastAPI з використанням PostgreSQL для зберігання даних моніторингу, клієнтську частину – засобами React; компоненти системи розгортаються у контейнерному середовищі. Для кожної моделі задаються еталонні значення метрик та еталонні розподіли, контрольовані ознаки, порогові значення для визначення рівнів дрейфу, зміни розподілу прогнозів і зміни прогнозової якості, рівень статистичної значущості, мінімальна кількість верифікованих спостережень і розклад перевірок. Для кожної перевірки зберігаються L_D , L_P , L_M , значення A_t та сформований стан або рівень ризику. Невизначений рівень L_M зберігається як відсутнє значення, а не як нуль.

У системі окремо відображаються динаміка PSI вхідних ознак, зміна розподілу прогнозів і метрики прогнозової якості. Історія перевірок дає змогу простежити перехід від попередньої оцінки ризику до підтвердженого стану моделі після надходження достатньої кількості фактичних цільових значень.

7. Експериментальна перевірка

Експериментальну перевірку проведено на трьох задачах, що охоплюють класифікацію та регресію: прогнозування серцевого захворювання на наборі Heart Failure Prediction обсягом 918 записів [22], прогнозування відтоку персоналу на наборі IBM HR Analytics Employee Attrition обсягом 1470 записів [23] та прогнозування вартості медичного страхування на наборі Medical Cost Personal обсягом 1338 записів [24]. Кожен набір поділено на навчальну та експлуатаційну частини у співвідношенні 70:30. Контрольовані сценарії дрейфу сформовано зваженим відбором із поверненням з експлуатаційної частини: зміною ймовірностей потрапляння спостережень до вибірки періоду цілеспрямовано змінювалися розподіли окремих ознак або збільшувалася частка складних для моделі випадків. Нових записів не породжувалося; окремі наявні спостереження могли повторно потрапляти до вибірки періоду. Періоди містили 350-400 спостережень. Така схема дає змогу перевірити різні комбінації L_D , L_P , L_M , а також роботу системи до та після надходження достатньої кількості фактичних цільових значень.

Для моделі прогнозування серцевого захворювання навчальна частина містила 642 спостереження, експлуатаційна – 276. Еталонні значення площі під характеристичною кривою та F-міри, отримані за п'ятиблоковою кросвалідацією, становили відповідно 0,9076 і 0,8571. У періоді 1, що відповідає стабільному потоку, перевірку спочатку виконано без фактичних цільових значень. За $L_D=0$, $L_P=0$ і $L_M=\perp$ система сформувала висновок «якість не оцінено». Після надходження фактичних цільових значень для 400 спостережень того самого періоду отримано $L_M=0$, а стан уточнено до стабільного. Цей результат демонструє принципову відмінність між невизначеним рівнем прогнозової якості та підтвердженою відсутністю її погіршення.

У періоді 2 зваженим відбором було збільшено частку старших пацієнтів. Максимальне значення PSI серед вхідних ознак становило 0,5367 за ознакою Age, що відповідає вираженому дрейфу вхідних даних. До надходження фактичних цільових значень отримано $L_D=2$ та $L_P=1$, унаслідок чого система сформувала високий попередній ризик. Після верифікації прогнозів для 400 спостережень метрики не підтвердили погіршення якості: площа під характеристичною кривою становила 0,9278 проти еталонного значення 0,9076, а F-міра – 0,9234 проти 0,8571. Тому $L_M=0$, а остаточний стан визначено як дрейф вхідних даних без підтвердженої деградації. Цей сценарій демонструє перехід від попередньої оцінки ризику до підтвердженого стану моделі після надходження достатньої кількості фактичних цільових значень.

У періоді 3 також зафіксовано виражений дрейф вхідних даних: максимальне значення PSI становило 0,5026 за ознакою MaxHR. Зафіксоване відхилення метрик не перевищило нижнього порога, тому $L_M=0$, і цей випадок також віднесено до дрейфу вхідних даних без підтвердженої деградації.

Окремі сценарії підтвердили, що дрейф вхідних даних і погіршення прогнозової якості можуть виникати незалежно. Для моделі прогнозування серцевого захворювання у періоді 4, сформованому зростанням частки складних для моделі випадків, отримано $L_D=0$ та $L_M=2$, при цьому максимальне значення PSI серед вхідних ознак становило лише 0,0715 за ознакою MaxHR, тобто залишалося нижчим за нижній поріг 0,10. Площа під характеристичною кривою знизилася з 0,9076 до 0,5949, а F-міра – з 0,8571 до 0,6316. Система визначила критичну деградацію за відсутності суттєвого дрейфу контрольованих вхідних ознак. Динаміку PSI вхідних ознак для цієї моделі наведено на рис. 1. На рис. 1 видно, що найбільші значення PSI вхідних ознак спостерігалися у періодах 2 і 3, тоді як у періоді критичної деградації вони знизилися нижче порога суттєвого дрейфу. Отже, динаміка вхідних розподілів у цьому сценарії не відтворює динаміку прогнозової якості.

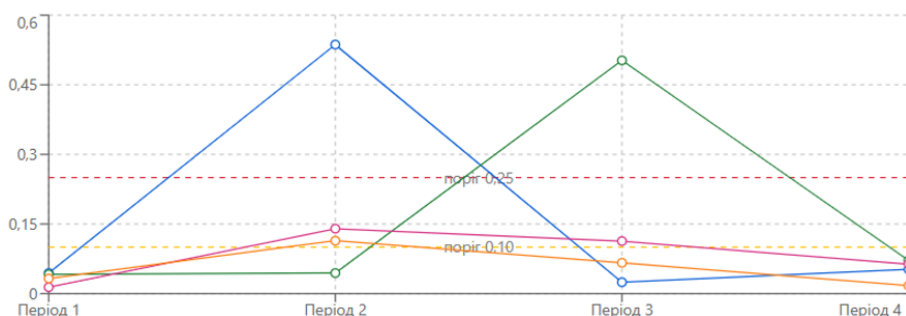


Рис. 1. Динаміка PSI вхідних ознак моделі прогнозування серцевого захворювання

Аналіз розподілу прогнозів дав додатковий сигнал у частині випадків, коли фактичних цільових значень було недостатньо. Динаміку зміни розподілу прогнозів для моделі прогнозування серцевого захворювання наведено на рис. 2.



Рис. 2. Динаміка PSI розподілу прогнозів

Як видно з рис. 2, у періоді критичного погіршення якості без суттєвого дрейфу вхідних ознак PSI розподілу прогнозів становив 0,2373, що відповідає помірному рівню зміни. Водночас для двох інших моделей у сценаріях критичної деградації без дрейфу вхідних ознак значення цього показника становили 0,0497 та 0,0475 і не перевищували нижнього порога. Отже, зміна розподілу прогнозів доповнює інформацію про стан моделі за відсутності достатньої кількості фактичних цільових значень, але не замінює оцінювання прогнозової якості після надходження достатньої кількості фактичних цільових значень.

Для наочного зіставлення змін вхідних розподілів із погіршенням прогнозової якості на рис. 3 наведено динаміку максимального PSI вхідних ознак і максимальної відносної деградації метрик у періодах, для яких прогнозу якість було оцінено.

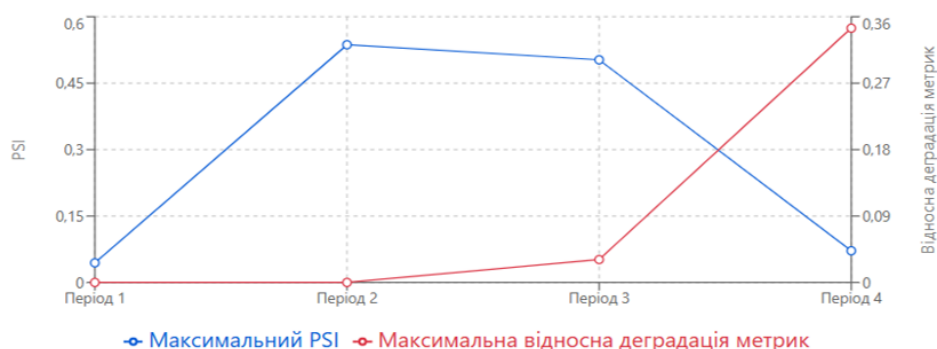


Рис. 3. Зіставлення максимального PSI і відносної деградації метрик прогнозової якості

Як видно з рис. 3, у періоді 4 максимальний PSI вхідних ознак знижується до 0,0715, тоді як відносна деградація метрик різко зростає. Цей результат показує, що контроль лише вхідних розподілів недостатній для висновку про стан прогнозової якості.

Протилежний випадок отримано для моделі прогнозування відтоку персоналу. За суттєвої зміни розподілу ознаки понаднормової роботи максимальне значення PSI серед вхідних ознак становило 0,5742,



а рівні сигналів дорівнювали $L_D=2$ і $L_P=2$. Водночас площа під характеристичною кривою практично не змінилася відносно еталонного рівня – 0,7590 проти 0,7580, а F-міра зросла з 0,4378 до 0,4768. Тому $L_M=0$, і система сформулила попередження про дрейф вхідних даних без підтвердженої деградації. Отриманий результат додатково демонструє недоцільність ототожнення вираженого статистичного зсуву з обов'язковою втратою прогнозної якості.

Для регресійної моделі прогнозування вартості медичного страхування отримано випадок одночасного помірної дрейфу та погіршення якості. У відповідному періоді $L_D=1$, $L_P=1$ і $L_M=1$; максимальна відносна деградація метрик становила 0,1441, а максимальне значення PSI серед вхідних ознак – 0,1048 для ознаки куріння. PSI розподілу прогнозів становив 0,1373. Сценарій віднесено до класу «дрейф вхідних даних із підтвердженою деградацією». Значення 1 і 2 характеризують ступінь вираженості відповідних сигналів і не утворюють окремих класів стану.

8. Зіставлення з рішеннями за окремими показниками

Для оцінювання внеску окремих компонентів запропонованого методу сформовані ним рішення зіставлено з висновками, отриманими лише за PSI вхідних ознак або лише за метриками прогнозування якості. Результати для трьох досліджуваних моделей наведено в табл. 4.

Таблиця 4

Зіставлення рішень за окремими показниками та запропонованим методом

Період	Сценарій	PSI вхідних ознак	Метрики прогнозування якості	Запропонований метод
Прогнозування серцевого захворювання				
1	стабільний потік	змін не виявлено	до верифікації – недоступні; після – якість збережено	якість не оцінено → стабільний стан
2	дрейф без деградації	дрейф	до верифікації – недоступні; після – якість збережено	високий ризик → попередження
3	дрейф без деградації	дрейф	якість збережено	попередження
4	деградація без дрейфу	змін не виявлено	критична деградація	критичний стан
Прогнозування відтоку персоналу				
1	стабільний стан	змін не виявлено	якість збережено	стабільний стан
2	дрейф без деградації	дрейф	якість збережено	попередження
3	деградація без дрейфу	змін не виявлено	критична деградація	критичний стан
Прогнозування вартості медичного страхування				
1	стабільний стан	змін не виявлено	якість збережено	стабільний стан
2	дрейф; фактичні значення недоступні	дрейф	не оцінюються	високий ризик
3	дрейф із деградацією	дрейф	помірна деградація	попередження
4	деградація без дрейфу	змін не виявлено	критична деградація	критичний стан

У табл. 4 наведено одинадцять періодів моніторингу. Для двох періодів моделі прогнозування серцевого захворювання перевірку виконували двічі, до і після надходження достатньої кількості фактичних цільових значень, тому загальна кількість виконаних перевірок становила тринадцять. За достатньої кількості фактичних цільових значень система визначала підтверджений стан моделі, а за їх відсутності або недостатньої кількості – рівень ризику.

Зіставлення показує, що PSI і метрики прогнозування якості характеризують різні аспекти стану моделі. PSI відображає зміни контрольованих вхідних розподілів, але не дає сигналу в сценаріях деградації без дрейфу. Метрики прогнозування якості дають змогу виявити погіршення після надходження достатньої кількості фактичних цільових значень, однак до цього моменту вони недоступні. Запропонований метод поєднує ці випадки в межах спільної процедури: за недостатньої кількості фактичних цільових значень формується рівень ризику, а після верифікації – підтверджений стан моделі. Це дає змогу не ототожнювати неможливість оцінювання прогнозування якості зі стабільним станом моделі та статистичний дрейф із підтвердженою деградацією.

Критерій Колмогорова-Смирнова обчислювався в одинадцяти з тринадцяти перевірок. У двох випадках ознака з максимальним значенням PSI була категоріальною: понаднормова робота в моделі прогнозування відтоку персоналу та куріння в моделі прогнозування вартості медичного страхування, тому не застосовувався. У першому періоді моделі прогнозування серцевого захворювання для ознаки Age отримано $D=0,1194$ зі статистично значущою відмінністю, тоді як PSI цієї ознаки становив 0,0442. Для першого періоду моделі прогнозування відтоку персоналу для ознаки MonthlyIncome отримано $D=0,1234$ зі статистично значущою відмінністю за $PSI=0,0576$. В обох випадках PSI залишався нижчим за поріг 0,10.



Загалом у п'яти з одинадцяти перевірок, у яких критерій був застосовним, він вказував на статистично значущу відмінність у сценаріях без суттєвого дрейфу вхідних даних.

Отримані результати показують, що статистична значущість відмінності між вибірками не обов'язково відповідає практично суттєвій зміні розподілу, що підтверджує доцільність визначення рівня L_D за PSI.

Експеримент мав демонстраційний характер: на тринадцяти перевірках не оцінювали статистичні характеристики точності виявлення деградації або частоту помилкових спрацьовувань; для цього потрібне окреме дослідження на тривалих потоках реальних даних.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

У роботі розроблено метод виявлення деградації моделей машинного навчання, який поєднує контроль змін вхідних даних із оцінюванням прогнозної якості та враховує доступність фактичних цільових значень під час формування рішення. Це дає змогу відокремити статистичну зміну робочих даних від фактичного погіршення результатів прогнозування та не ототожнювати неможливість оцінювання прогнозної якості зі стабільним станом моделі. Для випадків, коли доступна достатня кількість фактичних цільових значень, визначено відповідність між характером виявлених змін і станом моделі: стабільний стан, дрейф вхідних даних без підтвердженої деградації, погіршення прогнозної якості без виявленого дрейфу вхідних даних та дрейф вхідних даних із підтвердженою деградацією. Якщо фактичних цільових значень недостатньо для коректного обчислення метрик, остаточний висновок про деградацію не формується. У такій ситуації оцінюється ризик на основі доступних непрямих сигналів, а після надходження достатньої кількості фактичних значень попередня оцінка ризику уточнюється до підтвердженого стану моделі.

Метод реалізовано в інформаційній системі моніторингу якості моделей машинного навчання та перевірено на задачах класифікації і регресії. Експериментальні результати показали, що дрейф вхідних даних і погіршення прогнозної якості можуть виникати незалежно. Зафіксовано випадки вираженої зміни вхідних розподілів без зниження якості, а також критичного погіршення прогнозування за відсутності суттєвого дрейфу контрольованих ознак. Ці результати підтверджують недостатність використання лише одного типу показників для визначення стану моделі.

Окремо перевірено формування рішення за відкладених міток. Для стабільного потоку відсутність фактичних цільових значень спочатку не давала змоги оцінити прогнозну якість, а після надходження їх достатньої кількості було підтверджено стабільний стан моделі. В іншому сценарії до верифікації було сформовано високий ризик через одночасні зміни вхідних даних і розподілу прогнозів, однак після надходження достатньої кількості фактичних значень погіршення якості не підтвердилося. Отже, попередня оцінка не замінює безпосереднього контролю прогнозної якості, а забезпечує можливість реагування в період до надходження фактичних результатів.

Порівняння з рішеннями за окремими показниками показало, що індекс стабільності популяції характеризує зміну вхідних розподілів, але не визначає її вплив на прогнозну якість, тоді як метрики прогнозної якості можуть бути використані лише після надходження достатньої кількості фактичних цільових значень. Критерій Колмогорова-Смирнова в запропонованому методі використовується як додатковий діагностичний засіб, оскільки статистично значуща відмінність між вибірками не завжди відповідає практично суттєвій зміні розподілу.

Подальше дослідження може бути спрямоване на перевірку методу на триваліших потоках даних, аналіз послідовних і поступових змін розподілів та оцінювання стійкості сформованих рішень у довготривалому моніторингу.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Bayram, F., Ahmed, B. S., & Kassler, A. (2022). From concept drift to model degradation: An overview on performance-aware drift detectors. *Knowledge-Based Systems*, 245, 108632. <https://doi.org/10.1016/j.knsys.2022.108632>
2. Sethi, T. S., Kantardzic, M. (2017). On the reliable detection of concept drift from streaming unlabeled data. *Expert Systems with Applications*, 82, 77–99. <https://doi.org/10.1016/j.eswa.2017.04.008>
3. Dong, S., Wang, Q., Sahri, S., Palpanas, T., & Srivastava, D. (2024). Efficiently Mitigating the Impact of Data Drift on Machine Learning Pipelines. *Proceedings of the VLDB Endowment*, 17(11), 3072–3081. <https://doi.org/10.14778/3681954.3681984>
4. Sahiner, B., Chen, W., Samala, R. K., & Petrick, N. (2023). Data drift in medical machine learning: implications and potential remedies. *The British Journal of Radiology*, 96,



20220878. <https://doi.org/10.1259/bjr.20220878>
5. Vela, D., Sharp, A., Zhang, R., Nguyen, T., Hoang, A., & Pianykh, O. S. (2022). Temporal quality degradation in AI models. *Scientific Reports*, 12, 11654. <https://doi.org/10.1038/s41598-022-15245-z>
 6. Potgieter, C. J., Van, Zyl C., Schutte, W. D., & Lombard, F. (2026) The population resemblance statistic: a chi-square measure of fit for banking. *Annals of Operations Research*. 361, 413–435. <https://doi.org/10.1007/s10479-025-07024-6>.
 7. Чепурна, О. Є., Черевко, Є. В., Лобода, Ю. Г., Кухаренко, С. В. (2025). Застосування математичної статистики для аналізу великих даних в алгоритмах обробки інформації. *Вчені записки Таврійського національного університету імені В.І. Вернадського. Серія: Технічні науки*, 36(75)(4), 313–320. <https://doi.org/10.32782/2663-5941/2025.4.2/41>
 8. Massey, F. J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, 46(253), 68–78. <https://doi.org/10.1080/01621459.1951.10500769>
 9. Khaki, S., Aditya, A. A., Karnin, Z., Ma, L., Pan, O., & Chandrashekar, S. M. (2023). Uncovering drift in textual data: An unsupervised method for detecting and mitigating drift in machine learning models. *arXiv*. <https://doi.org/10.48550/arXiv.2309.03831>.
 10. Heinrichs, F. (2023). Monitoring machine learning models: Online detection of relevant deviations. *arXiv*. <https://doi.org/10.48550/arXiv.2309.15187>
 11. Ackerman, S., Farchi, E., Raz, O., Zalmanovici, M., & Dube, P. (2020). Detection of data drift and outliers affecting machine learning model performance over time. *arXiv*. <https://doi.org/10.48550/arXiv.2012.09258>.
 12. Bauer, J. C., Trattnig, S., Vieltorf, F., & Daub, R. (2026). Handling data drift in deep learning-based quality monitoring: evaluating calibration methods using the example of friction stir welding. *Journal of Intelligent Manufacturing*, 37, 759–774. <https://doi.org/10.1007/s10845-025-02569-6>.
 13. Shen, G., Li, S., Zhang, Y., Zhou, H., & Li, M. (2025). Dynamic Monitoring Method of Polymer Injection Molding Product Quality Based on Operating Condition Drift Detection and Incremental Learning. *Polymers*, 17(22), 3025. <https://doi.org/10.3390/polym17223025>
 14. Gomes, H. M., Grzenda, M., Mello, R., Read, J., Le Nguyen, M. H., & Bifet, A. (2023). A survey on semi-supervised learning for delayed partially labelled data streams. *ACM Computing Surveys*, 55(4), Article 75, 1–42. <https://doi.org/10.1145/3523055>
 15. Castellani, A., Schmitt, S., & Hammer, B. (2022). Stream-based active learning with verification latency in non-stationary environments. In: *Artificial Neural Networks and Machine Learning – ICANN 2022* (pp. 260–272). https://doi.org/10.1007/978-3-031-15937-4_22
 16. Лобода, Ю. Г., Трофименко, О. Г., Логінова, Н.І. (2025). Інтегрована система аналізу відтоку персоналу на основі поєднання класифікаційних моделей та аналізу виживання. *Таврійський науковий вісник. Серія: Технічні науки*, 3, 228–238. <https://doi.org/10.32782/tnv-tech.2025.3.25>
 17. Miller, T., & Durluk, I. (2026). When models fail: Trustworthy anomaly detection under distributional drift via dual-layer monitoring of data and AI behaviour. *Applied Sciences*, 16(11), 5293. <https://doi.org/10.3390/app16115293>
 18. Kidane, L., Townend, P., Metsch, T., & Elmroth, E. (2026). Efficient retraining of machine learning algorithms in cloud management systems. *Future Generation Computer Systems*, 182, 108478. <https://doi.org/10.1016/j.future.2026.108478>
 19. Chi, S., Tian, Y., Wang, F., Zhou, T., Jin, S., & Li, J. (2022). A novel lifelong machine learning-based method to eliminate calibration drift in clinical prediction models. *Artificial Intelligence in Medicine*, 125, 102256. <https://doi.org/10.1016/j.artmed.2022.102256>
 20. Guan, H., & Zhou, L. (2026). Detecting performance degradation under data shift in pathology vision-language model. *arXiv*. <https://doi.org/10.48550/arXiv.2601.00716>
 21. Баутіна, М. (2025). Моніторинг дрейфу моделей машинного навчання в промислових пайплайнах: методи, метрики та аспекти розгортання. *Інформаційні технології та суспільство*, 1(16), 8–15. <https://doi.org/10.32689/maup.it.2025.1.1>.
 22. Fedesoriano. (2021). *Heart failure prediction dataset* [Data set]. Kaggle. <https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>
 23. Pavansubhash. (n.d.). IBM HR analytics employee attrition & performance [Data set]. Kaggle. <https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset>
 24. Choi, M. (n.d.). Medical cost personal datasets [Data set]. Kaggle. <https://www.kaggle.com/datasets/mirichoi0218/insurance>

**Yuliia G. Loboda**

PhD, Associate Professor, Associate Professor at the Department of Software Engineering,
National University "Odesa Law Academy", Odesa, Ukraine,
ORCID: 0000-0001-7083-552X
loboda@onua.edu.ua

Olena G. Trofymenko

PhD, Associate Professor, Associate Professor at the Department of Software Engineering,
National University "Odesa Law Academy", Odesa, Ukraine,
ORCID: 0000-0001-7626-0886
trofymenko@onua.edu.ua

Mykhaylo V. Vorona

PhD, lecturer at the Department of Information Control Systems and Technology,
Admiral Makarov National University of Shipbuilding, Mykolaiv, Ukraine,
ORCID: 0000-0003-4288-0096
mykhailo.vorona@nuos.edu.ua

Oksana V. Haidaienko

PhD, Associate Professor at the Department of Information Control Systems and Technology,
Admiral Makarov National University of Shipbuilding, Mykolaiv, Ukraine,
ORCID: 0000-0002-6614-5443
oksana.gaidaienko@nuos.edu.ua

DETECTING MACHINE LEARNING MODEL DEGRADATION BASED ON JOINT ANALYSIS OF DATA DRIFT AND PERFORMANCE METRICS

Abstract. Machine learning models operate in environments whose statistical properties may change after deployment; therefore, keeping the model parameters unchanged does not guarantee stable predictive performance. At the same time, the presence of input data drift alone does not provide evidence of model degradation, while direct performance evaluation is often complicated when ground truth is delayed or only partially available. The aim of this study is to develop a method for detecting machine learning model degradation that combines the analysis of changes in input data with predictive performance monitoring and accounts for the availability of ground truth when making monitoring decisions. The proposed method separately evaluates input feature drift, changes in the distribution of predictions, and deviations of performance metrics from their reference values. The Population Stability Index is used to determine the levels of input feature drift and change in the distribution of predictions. When a sufficient number of verified observations is available, the model state is determined from input feature drift and changes in predictive performance, which makes it possible to distinguish stable operation, drift without confirmed degradation, deterioration in predictive performance without detected drift, and the simultaneous occurrence of both, and to determine the corresponding response level. When the number of verified observations is insufficient, only a risk assessment based on changes in input features and in the distribution of predictions is produced. This assessment is refined once ground truth becomes available for a sufficient number of observations. The method was implemented in an information system for monitoring machine learning model performance and evaluated on classification and regression tasks under controlled scenarios of data change. The experimental results showed that pronounced input feature drift may occur without deterioration in predictive performance, whereas critical model degradation may arise without substantial changes in the monitored input distributions. The results also demonstrated that a risk assessment produced before the predictions are verified can be refined once ground truth becomes available. The proposed approach makes it possible to distinguish statistical changes in data from actual deterioration in predictive performance and to avoid an unjustified conclusion that the model is stable when the information required to evaluate its performance is unavailable.

Keywords: data analysis; machine learning; data drift; model degradation; Population Stability Index; performance metrics; decision-making; prediction distribution.



REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Bayram, F., Ahmed, B. S., & Kassler, A. (2022). From concept drift to model degradation: An overview on performance-aware drift detectors. *Knowledge-Based Systems*, 245, 108632. <https://doi.org/10.1016/j.knosys.2022.108632>
2. Sethi, T. S., Kantardzic, M. (2017). On the reliable detection of concept drift from streaming unlabeled data. *Expert Systems with Applications*, 82, 77–99. <https://doi.org/10.1016/j.eswa.2017.04.008>
3. Dong, S., Wang, Q., Sahri, S., Palpanas, T., & Srivastava, D. (2024). Efficiently Mitigating the Impact of Data Drift on Machine Learning Pipelines. *Proceedings of the VLDB Endowment*, 17(11), 3072–3081. <https://doi.org/10.14778/3681954.3681984>
4. Sahiner, B., Chen, W., Samala, R. K., & Petrick, N. (2023). Data drift in medical machine learning: implications and potential remedies. *The British Journal of Radiology*, 96, 20220878. <https://doi.org/10.1259/bjr.20220878>
5. Vela, D., Sharp, A., Zhang, R., Nguyen, T., Hoang, A., & Pinykh, O. S. (2022). Temporal quality degradation in AI models. *Scientific Reports*, 12, 11654. <https://doi.org/10.1038/s41598-022-15245-z>
6. Potgieter, C. J., Van, Zyl C., Schutte, W. D., & Lombard, F. (2026) The population resemblance statistic: a chi-square measure of fit for banking. *Annals of Operations Research*. 361, 413–435. <https://doi.org/10.1007/s10479-025-07024-6>.
7. Chepurna, O. Ye., Cherevko, Yu. G., Loboda, Yu. H., & Kukharenko, S. V. (2025). Application of mathematical statistics for big data analysis in information processing algorithms. *Scientific Notes of V. I. Vernadsky Taurida National University. Series: Technical Sciences*, 36(75)(4), 313–320. <https://doi.org/10.32782/2663-5941/2025.4.2/41>
8. Massey, F. J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, 46(253), 68–78. <https://doi.org/10.1080/01621459.1951.10500769>
9. Khaki, S., Aditya, A. A., Karnin, Z., Ma, L., Pan, O., & Chandrashekar, S. M. (2023). Uncovering drift in textual data: An unsupervised method for detecting and mitigating drift in machine learning models. *arXiv*. <https://doi.org/10.48550/arXiv.2309.03831>.
10. Heinrichs, F. (2023). Monitoring machine learning models: Online detection of relevant deviations. *arXiv*. <https://doi.org/10.48550/arXiv.2309.15187>
11. Ackerman, S., Farchi, E., Raz, O., Zalmanovici, M., & Dube, P. (2020). Detection of data drift and outliers affecting machine learning model performance over time. *arXiv*. <https://doi.org/10.48550/arXiv.2012.09258>.
12. Bauer, J. C., Trattnig, S., Vieltorf, F., & Daub, R. (2026). Handling data drift in deep learning-based quality monitoring: evaluating calibration methods using the example of friction stir welding. *Journal of Intelligent Manufacturing*, 37, 759–774. <https://doi.org/10.1007/s10845-025-02569-6>.
13. Shen, G., Li, S., Zhang, Y., Zhou, H., & Li, M. (2025). Dynamic Monitoring Method of Polymer Injection Molding Product Quality Based on Operating Condition Drift Detection and Incremental Learning. *Polymers*, 17(22), 3025. <https://doi.org/10.3390/polym17223025>
14. Gomes, H. M., Grzenda, M., Mello, R., Read, J., Le Nguyen, M. H., & Bifet, A. (2023). A survey on semi-supervised learning for delayed partially labelled data streams. *ACM Computing Surveys*, 55(4), Article 75, 1–42. <https://doi.org/10.1145/3523055>
15. Castellani, A., Schmitt, S., & Hammer, B. (2022). Stream-based active learning with verification latency in non-stationary environments. In: *Artificial Neural Networks and Machine Learning – ICANN 2022* (pp. 260–272). https://doi.org/10.1007/978-3-031-15937-4_22
16. Loboda, Yu. G., Trofymenko, O. G., & Lohinova, N. I. (2025). Integrated employee attrition analysis system based on the combination of classification models and survival analysis. *Tavriyskyi Naukovyi Visnyk. Series: Technical Sciences*, 3, 228–238. <https://doi.org/10.32782/tnv-tech.2025.3.25>
17. Miller, T., & Durlík, I. (2026). When models fail: Trustworthy anomaly detection under distributional drift via dual-layer monitoring of data and AI behaviour. *Applied Sciences*, 16(11), 5293. <https://doi.org/10.3390/app16115293>
18. Kidane, L., Townend, P., Metsch, T., & Elmroth, E. (2026). Efficient retraining of machine learning algorithms in cloud management systems. *Future Generation Computer Systems*, 182, 108478. <https://doi.org/10.1016/j.future.2026.108478>
19. Chi, S., Tian, Y., Wang, F., Zhou, T., Jin, S., & Li, J. (2022). A novel lifelong machine learning-based method to eliminate calibration drift in clinical prediction models. *Artificial Intelligence in Medicine*, 125, 102256. <https://doi.org/10.1016/j.artmed.2022.102256>
20. Guan, H., & Zhou, L. (2026). Detecting performance degradation under data shift in pathology vision-language model. *arXiv*. <https://doi.org/10.48550/arXiv.2601.00716>



21. Bautina, M. (2025). Monitoring machine learning model drift in production pipelines: Methods, metrics, and deployment aspects. *Information Technology and Society*, 16(1), 8–15. <https://doi.org/10.32689/maup.it.2025.1.1>
22. Fedesoriano. (2021). *Heart failure prediction dataset* [Data set]. Kaggle. <https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>
23. Pavansubhash. (n.d.). IBM HR analytics employee attrition & performance [Data set]. Kaggle. <https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset>
24. Choi, M. (n.d.). Medical cost personal datasets [Data set]. *Kaggle*. <https://www.kaggle.com/datasets/mirichoi0218/insurance>

Отримано редакцією журналу / Received: 28.06.26

Прорецензовано / Revised: 15.07.26

Схвалено до друку / Accepted: 26.10.26



This work is licensed under Creative Commons Attribution-noncommercial-sharelike 4.0 International License.