

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Західноукраїнський національний університет
Факультет комп'ютерних інформаційних технологій
Кафедра інформаційно-обчислювальних систем і управління

ЧЕРНІЙ Іван Русланович

**Метод управління закупівельними проєктами із застосуванням
машинного навчання / Method for Managing Procurement Projects
using Machine Learning**

спеціальність: 122 - Комп'ютерні науки
освітньо-професійна програма – Управління проєктами

Кваліфікаційна робота

Виконав студент групи КНУПм-21
І. Р. Черній

Науковий керівник:
к.т.н., доцент Т.В. Лендюк

Кваліфікаційну роботу
допущено до захисту:

«__» _____ 20__ р.

В.о. завідувача кафедри

_____ Н.В. Дзюбановська

ТЕРНОПІЛЬ - 2025

Факультет комп'ютерних інформаційних технологій
Кафедра інформаційно-обчислювальних систем і управління
Освітній ступінь «магістр»
спеціальність: 122 – Комп'ютерні науки
освітньо-професійна програма – Управління проектами

ЗАТВЕРДЖУЮ
В.о. завідувача кафедри
Н.М. Васильків
«_____» _____ 20__ р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ
ЧЕРНІЙ Іван Русланович

(прізвище, ім'я, по батькові)

1. Тема кваліфікаційної роботи

Метод управління закупівельними проектами із застосуванням машинного навчання / Method for Managing Procurement Projects using Machine Learning

керівник роботи к.т.н., доцент Т.В. Лендюк

затвержені наказом по університету від 20 грудня 2024 року № 938.

2. Строк подання студентом закінченої кваліфікаційної роботи 1 грудня 2025 р.

3. Вихідні дані до кваліфікаційної роботи: завдання на кваліфікаційну роботу студента, наукові статті, технічна література.

4. Основні питання, які потрібно розробити

– здійснити критичний огляд підходів до ідентифікації змови та визначити вимоги до даних;

– сформувати та уніфікувати міжкраїновий корпус аукціонів (шістка наборів даних) і описати нормативно-галузевий контекст;

– сконструювати систему скринінгових індикаторів зі ставок (SPD, DIFFP, CV, KSTEST, SKEW, тощо) та обґрунтувати їх інформативність;

– налаштувати і порівняти сімейства алгоритмів (ансамблеві/неансамблеві) у сценаріях повної й обмеженої інформації;

– оцінити похибки (FP/FN), збалансовану точність та стійкість до вибіркового коливань;

– напрацювати рекомендації для впровадження в аналітичні процеси замовників/наглядових органів.

5. Перелік графічного матеріалу у роботі

– опис наборів даних;

– блок-схема дослідницького підходу;

– гістограми результатів.

6. Консультанти розділів кваліфікаційної роботи

| Розділ | Прізвище, ініціали та посада консультанта | Підпис, дата | |
|--------|---|----------------|------------------|
| | | Завдання видав | Завдання прийняв |
| | | | |
| | | | |
| | | | |

7. Дата видачі завдання 20 грудня 2024 р.

КАЛЕНДАРНИЙ ПЛАН

| № з/п | Назва етапів кваліфікаційної роботи | Строк виконання етапів кваліфікаційної роботи | Примітка |
|-------|---|---|----------|
| 1 | Затвердження теми кваліфікаційної роботи, ознайомлення з літературними джерелами та складання плану роботи. | до 01.01. 2025 р. | |
| 2 | Написання 1 розділу кваліфікаційної роботи | до 01.03. 2025 р. | |
| 3 | Написання 2 розділу кваліфікаційної роботи | до 20.05.2025 р. | |
| 4 | Написання 3 розділу кваліфікаційної роботи | до 28.10. 2025 р. | |
| 5 | Представлення попереднього варіанту кваліфікаційної роботи, перевірка та внесення змін керівником | до 11.11.2025 р. | |
| 6 | Опрацювання зауважень та представлення завершеного варіанту кваліфікаційної роботи. Підготовка супроводжуючих документів. | до 25.11.2025 р. | |
| 7 | Перевірка кваліфікаційної роботи на оригінальність тексту. | до 1.12.2025 р. | |
| 8 | Оформлення кваліфікаційної роботи та отримання допуску до захисту | до 04.12.2025 р. | |
| 9 | Подання кваліфікаційної роботи до захисту на засіданні атестаційної комісії. | до 14.12. 2025 р. | |

Студент _____ І.Р. Черній
підпис

Керівник роботи _____ к.т.н., доцент Т.В. Лендюк
підпис

РЕЗЮМЕ

Кваліфікаційна робота на тему «Метод управління закупівельними проєктами із застосуванням машинного навчання» на здобуття освітнього ступеня «Магістр» зі спеціальності 122 «Комп'ютерні науки» освітньої програми «Управління проєктами» написана обсягом у 73 сторінки і містить 9 ілюстрацій, 3 таблиці, 2 додатки та 64 використані джерела.

Метою роботи є розроблення та обґрунтування методу управління закупівельними проєктами, що базується на застосуванні алгоритмів машинного навчання, для ефективного виявлення ознак картельної змови та підвищення якості прийняття рішень у сфері публічних закупівель.

Методи досліджень: методи системного аналізу, методи класифікації та машинного навчання, методи порівняльного аналізу та метрики якості, а також методи емпіричних експериментів.

Результати роботи: обґрунтовано теоретико-методологічні засади виявлення змови у публічних закупівлях. Розроблено систему скрінингових індикаторів та сформовано алгоритмічний інструментарій на основі методів машинного навчання. Проведено серію емпіричних експериментів, які підтвердили високу ефективність запропонованих моделей у ідентифікації ознак змови як за повної, так і за обмеженої інформації. Виконано порівняльний аналіз корпусів даних та внеску розроблених скрінингів у підвищення якості класифікації.

Результати роботи можуть бути використані органами контролю та моніторингу публічних закупівель для автоматизації процесу виявлення ризиків картельної змови, підвищення прозорості та ефективності використання бюджетних коштів, а також як основа для розроблення нових інструментів фінансового моніторингу.

Ключові слова: ЗАКУПІВЕЛЬНІ ПРОЄКТИ, МАШИННЕ НАВЧАННЯ, КАРТЕЛЬНА ЗМОВА, ПУБЛІЧНІ ЗАКУПІВЛІ, КЛАСИФІКАЦІЯ, СКРІНІНГОВІ ІНДИКАТОРИ.

ABSTRACT

Qualification work on the topic «Method for Managing Procurement Projects using Machine Learning» for Master's degree on specialty 122 «Computer Science» educational and professional program «Project Management» is written on 73 pages and it contains 9 figures, 3 tables, 2 appendices, and 64 sources.

The purpose of the work is to develop and substantiate a method for managing procurement projects based on the application of machine learning algorithms, aimed at the effective detection of signs of cartel conspiracy and improving the quality of decision-making in the field of public procurement.

Research methods: methods of system analysis, methods of classification and machine learning, methods of comparative analysis and quality metrics, as well as methods of empirical experiments.

Research results: the theoretical and methodological principles of detecting collusion in public procurement are substantiated. A system of screening indicators was developed and an algorithmic toolkit was formed based on machine learning methods. A series of empirical experiments were conducted, which confirmed the high efficiency of the proposed models in identifying signs of collusion both with complete and limited information. A comparative analysis of data corpora and the contribution of the developed screenings to improving the quality of classification was performed.

The results of the work can be used by public procurement control and monitoring bodies to automate the process of identifying the risks of cartel collusion, increasing the transparency and efficiency of the use of budget funds, as well as as a basis for the development of new financial monitoring tools. Keywords: PROCUREMENT PROJECTS, MACHINE LEARNING, CARTEL COLLUSION, PUBLIC PROCUREMENT, CLASSIFICATION, SCREENING INDICATORS.

ЗМІСТ

| | |
|--|----|
| Перелік умовних позначень | 8 |
| Вступ..... | 11 |
| 1 Теоретико-методологічні засади виявлення змови у публічних закупівлях | 14 |
| 1.1 Інституційно-правові засади та ризики картельної змови в системах закупівель | 14 |
| 1.2 Аналітичні підходи до ідентифікації змови..... | 16 |
| 1.3 Постановка задачі..... | 19 |
| Висновки до розділу 1..... | 21 |
| 2 Матеріали та методи | 23 |
| 2.1 Опис вибірок і нормативно-галузевий контекст | 23 |
| 2.2 Система скрінінгових індикаторів і метрики якості класифікації..... | 28 |
| 2.3 Алгоритмічний інструментарій і протокол експериментів..... | 32 |
| Висновки до розділу 2..... | 38 |
| 3 Аналіз результатів експериментів та розробка рекомендацій щодо управління закупівельними проектами..... | 39 |
| 3.1 Ефективність алгоритмів за повної інформації..... | 39 |
| 3.2 Ефективність алгоритмів за обмеженої інформації..... | 42 |
| 3.3 Порівняльний аналіз корпусів та внесок скрінінгів у підвищення якості..... | 47 |
| 3.4 Аналіз похибок і стійкості моделей..... | 52 |
| 3.5 Рекомендації щодо інтеграції розроблених моделей у процеси управління проектами державних закупівель | 54 |
| Висновки до розділу 3..... | 55 |
| Висновки | 57 |
| Список використаних джерел | 59 |

| | |
|---|----|
| Додаток А Апробація отриманих результатів | 66 |
|---|----|

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

- CPV — Єдиний закупівельний словник (*Common Procurement Vocabulary*): класифікатор предметів закупівель у ЄС/Україні.
- MEAT — Найбільш економічно вигідна пропозиція (*Most Economically Advantageous Tender*): критерій оцінювання «ціна + якість/інші показники».
- FPA — Аукціон з оплатою за першою ціною (*First-Price Auction*): переможець платить свою ставку.
- SPA — Аукціон з оплатою за другою ціною (*Second-Price Auction/Vickrey*): переможець платить другу найнижчу ставку.
- Capped FPA — Обмежений перша-ціна аукціон: ставки не можуть перевищувати встановлений «стелю»/ліміт.
- Lot — Лот: окремий об'єкт закупівлі в межах однієї процедури.
- Bid — Ставка/пропозиція учасника.
- Reserve/Estimate — Очікувана вартість/бюджет замовника.
- TED — *Tenders Electronic Daily*: офіційний європейський портал оголошень про закупівлі.
- GPA — Угода СОТ про державні закупівлі (*WTO Government Procurement Agreement*).
- АМКУ — Антимонопольний комітет України.
- OECD — Організація економічного співробітництва та розвитку.
- EU — Європейський Союз.
- ProZorro — Національна е-система публічних закупівель України.
- SPD — Щільність ставок біля мінімуму (*Spread/Density near minimum*): міра згуртованості пропозицій навколо найнижчої ціни.
- DIFF — Різниця між двома найнижчими ставками (у грош. одиницях).
- DIFFP — Відносна різниця між двома найнижчими ставками (у % від мінімальної).
- CV — Коефіцієнт варіації (*Coefficient of Variation*): σ/μ для масиву ставок.
- SKEW — Асиметрія розподілу ставок (*Skewness*).

- KS (KS-test) — Критерій Колмогорова—Смирнова: відхилення емпіричного розподілу від очікуваного.

- IQR — Міжквартильний розмах (*Interquartile Range*): $Q_3 - Q_1$.

- MAD — Середнє абсолютне відхилення від медіани (*Median Absolute Deviation*).

- PTE — Коефіцієнт «ціна-до-оцінки» (*Price-to-Estimate*): відношення ставки до очікуваної вартості ($\approx \text{bid/estimate}$).

- ENTR — Ентропія розподілу ставок (*Shannon entropy*): міра «різноманітності» цінових пропозицій.

- Collusive / C — Змовницька поведінка (зафіксована рішеннями органів/судів чи достовірною верифікацією).

- Competitive / N — Конкурентна поведінка (відсутні ознаки змови).

- Pos / Neg — Позитивний/негативний клас для бінарної класифікації (звично: Pos = «collusive»).

- LR — Логістична регресія (*Logistic Regression*).

- Lasso/Ridge/ElasticNet — Лінійні моделі з регуляризацією L1/L2/комбінованою.

- DT — Дерева рішень (*Decision Trees*).

- RF — Випадковий ліс (*Random Forest*).

- ET — Надзвичайно випадкові дерева (*ExtraTrees*).

- GBDT — Градієнтний бустинг над деревами (*Gradient Boosted Decision Trees*).

- XGB — *XGBoost* (градієнтний бустинг).

- LGBM — *LightGBM* (градієнтний бустинг з листовим ростом).

- CatBoost — Бустинг із підтримкою категоріальних ознак.

- SVM — Машина опорних векторів (*Support Vector Machine*).

- KNN — Метод k-найближчих сусідів.

- NB — Наївний Байєс.

- Stacking/Blending — Ансамблювання моделей (мета-модель / вагове змішування).

- SHAP — *SHapley Additive exPlanations*: пояснюваність внеску ознак.

- ACC — Точність (*Accuracy*): $(TP+TN)/(TP+TN+FP+FN)$.
- BA — Збалансована точність (*Balanced Accuracy*): $(TPR+TNR)/2$.
- Precision (PPV) — Точність позитивного класу: $TP/(TP+FP)$.
- Recall (TPR, Sensitivity) — Повнота/чутливість: $TP/(TP+FN)$.
- Specificity (TNR) — Специфічність: $TN/(TN+FP)$.
- F1 — Гармонійне середнє Precision і Recall.
- ROC-AUC — Площа під ROC-кривою.
- PR-AUC — Площа під PR-кривою.
- MCC — Коефіцієнт Меттьюза (*Matthews Correlation Coefficient*).
- TP/FP/TN/FN — Істинно/помилково позитивні/негативні.
- FPR/TPR — Частота хибних спрацювань / справжня позитивна частка.
- CI — Довірчий інтервал (*Confidence Interval*).
- CV (K-fold) — K-складова крос-валідація (із стратифікацією за класом/аукціоном).
- TTS — Поділ на тренувальний/тестовий набори (*Train-Test Split*).
- Time-split — Часовий розріз для робастності (тренування на минулому, тест на майбутньому).
- Grid/Random/Bayes Search — Перебір/випадковий/байєсівський пошук гіперпараметрів.
- Class Weight — Зважування класів для дисбалансу.
- Undersampling/Oversampling/SMOTE — Балансування вибірки (меншення/збільшення/синтетика).
- Calib. (Platt/Isotonic) — Калібрування ймовірностей (Platt scaling, ізотонна регресія).
- Seed / random_state — Фіксація випадковості для відтворюваності.
- Leakage — Витік інформації між train/test (заборонений).

ВСТУП

Актуальність теми. На тлі цифровізації публічних закупівель (е-тендери, відкриті дані, аналітичні модулі) виникає можливість масштабно аналізувати структуру цінових пропозицій, часову динаміку торгів та поведінкові патерни учасників. Проте класичні економетричні підходи часто залежать від конфіденційних параметрів витрат і вразливі до пропусків даних. Використання ML з інженерією статистичних «скрінів» безпосередньо зі ставок (наприклад, SPD, DIFFP, CV, KSTEST, SKEW тощо) дозволяє підвищити чутливість і специфічність виявлення змови навіть за обмеженої інформації (лише ставки та переможець). Додаткова практична значущість роботи полягає у можливості адаптувати підхід для українського контексту (Prozorro, рішення АМКУ) та для міжкраїнового порівняння корпусів, що розширює доказову базу для управлінських і регуляторних рішень.

Методологічною основою цього дослідження слугували підходи до виявлення змови у публічних закупівлях, запропоновані в попередніх міжнародних роботах [17, 30]. У межах цієї роботи зазначені дослідження розглядаються як наукові аналоги, на основі яких здійснено адаптацію інструментарію машинного навчання та скрінінгових індикаторів до умов українського ринку публічних закупівель і відкритих даних системи Prozorro.

Теоретико-емпіричні дослідження змови у закупівлях охоплюють моделі економетричного скрінінгу, ймовірнісні та порядково-статистичні тести, а також ML-підходи, що поєднують ансамблеві алгоритми з наборами скрінінгових змінних. Виявлені в літературі обмеження — залежність від приватних даних витрат, чутливість до формальних припущень і «зламність» окремих тестів — мотивують інтегровані рішення з акцентом на робастність і відтворюваність у різних юрисдикціях та форматах даних. У межах цієї роботи узагальнено відповідні підходи та сфокусовано увагу на їх придатності в реальних наборах даних (Бразилія, Італія, Японія, Швейцарія, Україна).

Мета роботи — розробити та емпірично оцінити метод управління закупівельними проєктами, що включає ML-модуль для автоматизованого

виявлення ознак змови на основі даних аукціонів і скринінгових індикаторів. Для досягнення мети необхідно вирішити такі завдання:

1. Здійснити критичний огляд підходів до ідентифікації змови та визначити вимоги до даних.
2. Сформувати та уніфікувати міжкраїновий корпус аукціонів (шістка наборів даних) і описати нормативно-галузевий контекст.
3. Сконструювати систему скринінгових індикаторів зі ставок (SPD, DIFFP, CV, KSTEST, SKEW, тощо) та обґрунтувати їх інформативність.
4. Налаштувати і порівняти сімейства алгоритмів (ансамблеві/неансамблеві) у сценаріях повної й обмеженої інформації.
5. Оцінити похибки (FP/FN), збалансовану точність та стійкість до вибіркового коливань.
6. Напрацювати рекомендації для впровадження в аналітичні процеси замовників/наглядових органів.

Об'єкт і предмет дослідження. Об'єктом є процеси прийняття рішень у закупівельних проєктах під впливом конкуренції/змови на аукціонах. Предметом є метод і інструментарій виявлення змови на основі ML із використанням скринінгових індикаторів, побудованих зі структури ставок, а також інтеграція цього інструментарію в управлінський цикл закупівельних проєктів.

Наукова новизна полягає у побудові інтегрованого методу управління закупівельними проєктами на основі ансамблевих моделей машинного навчання та скринінгових індикаторів розподілу ставок, який забезпечує підвищення точності виявлення змови за різних сценаріїв доступності даних і підтримує порогову політику під конкретні управлінські пріоритети FP/FN.

Методи дослідження. Застосовано емпіричний аналіз міжкраїнових корпусів; статистичні методи побудови та інтерпретації «скрінів»; машинне навчання (Random Forest, Extra Trees, AdaBoost, Gradient Boosting та ін.) з крос-перегонами й оцінкою стабільності; метрики якості (accuracy, balanced accuracy, precision, recall, FP/FN) та порогове калібрування під ціль замовника (мінімізація FP або FN). Подання та оформлення методів і результатів відповідає вимогам до структури курсової роботи згідно з методичними вказівками.

Інформаційна база дослідження. Використано шість реальних наборів даних аукціонів із п'яти країн, які різняться за критеріями визначення переможця, доступністю ідентифікаторів учасників та додаткових параметрів (PTE, локалізація тощо). Український фрагмент сформовано на базі відкритих даних Prozorro із верифікацією змовницьких випадків за рішеннями АМКУ. Це дає змогу порівнювати поведінкові патерни та узагальнювати висновки для різних інституційних середовищ.

Практичне значення отриманих результатів. Розроблений підхід дозволяє: (i) підвищити ефективність первинного скринінгу аукціонів для служб контролю якості закупівель; (ii) підтримати ризик-орієнтоване планування перевірок (калібрування порогів під мінімізацію FP або FN); (iii) інтегрувати індикатори і ML-модулі в дашборди моніторингу закупівель, зокрема у проєктному управлінні капітальними роботами та інфраструктурними контрактами.

Апробація результатів дослідження. Основні теоретичні положення роботи й практичні результати дослідження доповідалися й обговорювалися на Всеукраїнському науково-практичній конференції студентів, аспірантів та молодих вчених «Інтелектуальні комп'ютерні системи та мережі» в кінці, яка відбулася в місті Тернополі 25 листопада 2025 року та IX Міжнародній студентській науковій конференції «Глобалізація наукових знань: міжнародна співпраця та інтеграція галузей наук», 7 жовтня 2025 року і місті Черкаси, Україна.

1 ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ЗАСАДИ ВИЯВЛЕННЯ ЗМОВИ У ПУБЛІЧНИХ ЗАКУПІВЛЯХ

1.1 Інституційно-правові засади та ризики картельної змови в системах закупівель

Огляд теоретико-методологічних засад виявлення змови у публічних закупівлях у цьому розділі базується на узагальненні результатів попередніх емпіричних і методичних досліджень [17, 30]. Зазначені роботи використовуються як методологічний орієнтир для подальшого формування та адаптації інструментарію аналізу до українських умов.

Публічні закупівлі — це поширена форма державних видатків, метою якої є забезпечення замовника роботами, товарами або послугами [1]. У контексті закупівлі капітальних робіт компанії конкурують за укладення контракту на будівництво, модернізацію чи обслуговування капітального активу. Такі контракти можуть відрізнятися за природою й охоплювати будівництво нової цивільної (наприклад, дороги й мости) та соціальної (наприклад, школи й лікарні) інфраструктури, модифікацію наявних активів або виконання робіт з технічного обслуговування [2].

Публічні закупівлі можуть бути інтенсивним і складним процесом і, відповідно, потребувати значних ресурсів. Наприклад, Європейський Союз витрачає близько 16% свого валового внутрішнього продукту на публічні закупівлі [3]. Змова на таких аукціонах (також зветься «підтасовка тендерних пропозицій», *bid-rigging*) — це різні незаконні домовленості між конкуруючими фірмами, спрямовані на збільшення їхніх прибуткових маржин. Зазвичай такі змовницькі практики набувають форми скоординованого (неконкурентного) підвищення цін, що узгоджується між компаніями (так званими картелями) [4]. Змова є повторюваною проблемою для державного сектору, особливо під час закупівлі капітальних робіт — одних із найдорожчих позицій для придбання [4]. Для протидії змові регулярно розпочинаються кримінальні розслідування, однак притягнути до відповідальності й отримати обвинувальний вирок непросто [5].

Основна проблема, що перешкоджає державним інституціям (наприклад, замовникам, правоохоронним органам, антимонопольним комісіям і судам) домогтися вироку, — це виявлення та доведення факту змови [6]. Секретність незаконних домовленостей між фірмами зазвичай підкріплюється ретельно скоординованою й витонченою стратегією, яку важко викрити. На противагу цьому, закупівельні органи дотримуються прозорих і відносно стабільних шаблонів закупівель: повторно використовують процедури визначення переможця, купують стандартизовані продукти, застосовують подібні специфікації послуг тощо. Передбачуваність таких практик може полегшувати незаконний поділ ринку та скоординовані дії між змовницькими фірмами [7-9]. У такому контексті надійний метод виявлення наявності змови в аукціонах публічних закупівель суттєво допоміг би замовникам та іншим інституціям пом'якшити негативні економічні й соціальні наслідки змови.

У нормативній літературі запропоновано чимало моделей для виявлення змови. У цьому дослідженні розглянуто найрелевантніші моделі, які довели здатність позначати довгострокові змовницькі патерни серед картелів учасників торгів [10]. Вони також допомогли виявити, як ці картелі відлякують компанії від подання конкурентних пропозицій на ринках, де домінують [11, 12]. Водночас, хоча моделі й здатні виявляти змову, їхню точність часто ставлять під сумнів, оскільки дані, на яких вони ґрунтуються, можуть містити шум або бути недостатньо деталізованими. Наприклад, розроблені моделі нерідко спираються на інформацію про (внутрішні) структури витрат учасників і/або попередні кошторисні оцінки (PTE), але такі відомості зазвичай є конфіденційними (і змовницькі фірми, очевидно, не бажають їх розкривати) або просто відсутні [13, 14].

Машинне навчання (ML) — це галузь штучного інтелекту, зосереджена на створенні застосунків, здатних автоматично навчатися та вдосконалюватися на основі досвіду, аналізувати й виявляти закономірності в аукціонній інформації навіть тоді, коли її небагато (тобто є лише значення ставок і переможець кожного аукціону) [15-17]. Водночас алгоритми ML зазвичай потребують значного обсягу надійних даних із попередніх аукціонів для їхнього калібрування [16].

Отже, метою дослідження є перевірка здатності різних алгоритмів ML точно виявляти змовницькі аукціони. Кожен алгоритм тестується за різних умов (наприклад, за наявності більшого чи меншого масиву інформації та з/без введення скринінгових змінних — SV). Під SV розуміються ті статистичні індекси (безпосередньо обчислювані зі значень ставок), попередня обробка яких може допомогти алгоритмам ML підвищити рівень виявлення [17].

Щоб оцінити ефективність алгоритмів ML, проаналізовано шість наборів даних про закупівлі з п'яти країн (Бразилія, Італія, Японія, Швейцарія та Сполучені Штати Америки, США). Доступ до таких аукціонних даних зазвичай недоступний дослідникам, оскільки їх вважають чутливими (наприклад, кошторисні оцінки контрактів) [18], однак для дослідження виявлення змови, представленого в цій статті, було надано доступ і дозволи. Отже, наше дослідження демонструє, що алгоритми ML можуть виявляти змову та забезпечувати репрезентативні результати ефективності, якщо застосовувати їх до широкого спектра наборів даних із різних країн, які містять різні типи даних. Наскільки відомо авторам, це одне з перших поперечне міжкраїнове дослідження такого типу в сфері виявлення змови.

1.2 Аналітичні підходи до ідентифікації змови

Численні дослідження з теорії аукціонів показали, що структури витрат учасників торгів істотно зумовлюють їхні конкурентні та/або змовницькі стратегії [1, 10, 19-22]. McAfee та McMillan [23] першими проаналізували змову в статичних схемах ротації ставок за відсутності компенсаційних виплат між членами картелю. У моделі аукціону McAfee та McMillan [23] визначення переможця не залежить від попередніх (минулих) аукціонів. Спираючись на роботу McAfee та McMillan [23], Aoyagi [24], а також Skrzyrzacz і Nopenhayn [25] розширили їхню модель, розглянувши повторювану змову в динамічних схемах ротації ставок.

Змова також досліджувалася з погляду її наявності та впливу в реальних закупівельних аукціонах [26, 27]. Втім, емпірично обґрунтованих моделей

виявлення змови небагато. Однією з перших спроб розробити таку модель була робота Porter і Zona [19], у якій намагалися оцінити ймовірність перемоги учасника за відомих деяких спостережуваних чинників витрат. Однак ця модель була спрямована не стільки на визначення змови як такої, скільки на прогнозування діапазону цін майбутніх (конкурентних) пропозицій. Після поширення роботи Porter і Zona [19] було запропоновано й інші емпіричні моделі.

Першою знаковою моделлю для виявлення змови є так звана економетрична «перевірка ситом» (screening), запропонована Vajari та Ye [28]. Ця модель намагається передбачити, як має виглядати стандартний (конкурентний) розподіл пропозицій, виходячи з параметрів витрат учасників. На жаль, ці параметри витрат є приватними даними, які зазвичай важко зібрати й часто розкриваються самими учасниками. У підсумку значну частину даних доводиться безпосередньо оцінювати галузевим експертам, що знижує точність. Модель Vajari та Ye [28] позначає систематичні відхилення від еталонного сценарію. В такому разі саме галузеві експерти мають визначити еталонний сценарій, виходячи з припущення, що учасники, які подають конкурентні пропозиції, хочуть отримати контракт і не співпрацюватимуть із картелем.

Модель Vajari та Ye [29] первісно тестували на аукціонах з ремонтних робіт на автошляхах Середнього Заходу США у 1994–1998 рр. Її реалізовано як редуковану функціональну форму лінійної регресії, де, крім фінансових даних учасників, потрібні додаткові відомості — історія їхніх попередніх торгів і попередні кошторисні оцінки (PTE). Завдяки включенню цієї додаткової інформації Vajari та Ye [28] змогли робити коректні порівняння з еталонним сценарієм. Водночас у моделі Vajari та Ye [28] є важливі обмеження:

- надмірна залежність від обраної функціональної форми при виконанні регресійного аналізу;
- висока чутливість до відсутніх даних;
- відносна «зламність»: картелю легко обійти модель, якщо він знає, «як» вона працює (наприклад, через скоординовані прикриті пропозиції).

Найсуттєвішим серед наведених обмежень є потреба в детальних даних щодо кожного учасника та кожного аукціону. За відсутності таких даних модель

непридатна до застосування в реальних торгах. На щастя, від часу дослідження Vajari та Ye [28] з'явилося більше відкритих даних про державні контракти та конкурентів, що в найближчому майбутньому можна використати для підвищення якості виявлення змови за допомогою ML.

Другу модель, яку було розглянуто, розробили Ballesteros-Pérez та ін. [29]. Вона зосереджена на аналізі можливих аномальних розсіювань у розподілі ставок із припущенням їхньої рівномірності (Uniform). По суті, модель Ballesteros-Pérez та ін. [29] — це наближений метод виявлення змови, який застосовують разом з іншими підходами. Він використовує спрощений підхід порядкових статистик, коли абсолютні величини ставок ігноруються, а враховуються лише відносні відстані між ними. Звісно, цей підхід залишає можливість обійти метод, подаючи прикриті пропозиції, що «емулюють» рівномірний розподіл, навіть якщо їхній середній рівень усе одно аномально високий.

Третю модель запропонували Signor та ін. [30]; це ймовірнісний метод [2, 34]. Модель Signor та ін. [30] аналізує подані ставки на двох рівнях. По-перше, перевіряється, чи відповідає загальний розподіл ставок еталонному сценарію (наприклад, логнормальному розподілу). Додатково «локалізацію» цього розподілу (тобто абсолютний порядок величин ставок) можна наближено оцінити за історичними аукціонами, якщо доступні дані про їхні попередні кошторисні оцінки (PTE). Відтак модель досліджує відстань поданих ставок до PTE.

По-друге, ймовірнісний метод Signor та ін. [30] аналізує розсіювання найнижчої ставки, спираючись на теорію порядкових статистик. Простіше кажучи, порівнюється ймовірність того, що найнижча ставка (тобто теоретичний переможець) справді могла виникнути так, ніби була згенерована з того самого еталонного розподілу, визначеного на попередньому кроці. Отже, у методі Signor та ін. [30] фактичну переможну ставку порівнюють із найнижчою порядковою статистикою (мінімальним значенням із n штучно згенерованих ставок) від каліброваного еталонного розподілу. Якщо статистичне відхилення є значним, можна з високою впевненістю стверджувати, що така ставка навряд чи є справді

конкурентною. Таким чином, метод є робастним, але має обмеження: він сильно залежить від наявності та надійності РТЕ для низки попередніх «чесних» аукціонів, а також для аукціону, що перевіряється.

Нарешті, четверту модель розробив Imhof [17, 35]. Це перша робота, де розглянуто застосування ML до ставок і виявлення змови, використовуючи невеликий набір скринінгових змінних (SV) на швейцарському датасеті з дорожнього будівництва. У даному дослідженні використано ті самі SV і той самий датасет, але припускатимемо різні рівні доступу до аукціонних даних. Крім того, Imhof [17, 35] застосував два типи алгоритмів ML: (1) Lasso-регресію та «ансамблевий метод», що є зваженим середнім кількох алгоритмів; і (2) bagging регресійних дерев, випадкові ліси та штучні нейронні мережі. У цій роботі розглянуто ширший спектр алгоритмічних варіантів і різні набори даних, аби краще зрозуміти умови, за яких SV та алгоритми ML працюють ефективніше.

1.3 Постановка задачі

Публічні закупівлі становлять вагому частку національних витрат і безпосередньо впливають на ефективність використання бюджетних коштів, конкурентоспроможність ринків і довіру громадян до інститутів держави. Вразливість тендерних процедур до змовницької поведінки (bid-rigging) зумовлює завищення цін, зниження якості та витіснення добросовісних учасників. Попри цифровізацію процесів та відкритість даних, виявлення змови традиційними методами (ручні перевірки, формальні правила) залишається трудомістким і запізнілим. Це зумовлює потребу в масштабованих, відтворюваних і статистично обґрунтованих підходах раннього виявлення аномалій у торгах.

Аукціони першої ціни та процедури із середньозваженим критерієм особливо вразливі до координації учасників через зближення мінімальних пропозицій, ротацію переможців і штучне звуження розкиду ставок. Економетричні «скрині» (коефіцієнт варіації, розкид, різниця між двома найнижчими ставками, асиметрія, тест Колмогорова—Смирнова тощо) фіксують

такі патерни, однак ізольовано їхня діагностична сила обмежена, а придатність змінюється між ринками. Бракує систематичних порівнянь ефективності скрінінгових ознак у поєднанні з сучасними алгоритмами машинного навчання (ML) на різних юрисдикціях і за різної доступності полів даних.

Запропоноване дослідження адресує зазначені прогалини, поєднуючи скрінінгові індикатори зі спектром ML-алгоритмів і оцінюючи їхню якість у чотирьох сценаріях доступу до даних на шести реальних корпусах (Бразилія, Італія, Японія, Швейцарія — Тічино, Швейцарія — SG&GR, Україна/Prozorro для CPV45). Порівняльний дизайн і багаторазова стохастична валідація забезпечують коректну оцінку узагальнюваності результатів і чутливості до дисбалансу класів. Практичний внесок полягає у формуванні відтворюваного пайплайна, який демонструє, що ансамблеві методи в поєднанні зі скрінами стабільно підвищують точність і збалансовану точність, зменшуючи FP/FN у більшості ринкових контекстів, зокрема й в українському корпусі Prozorro.

Для України актуальність підсилюється масштабом відбудови, високою питомою вагою будівельних закупівель і необхідністю оперативного моніторингу ризиків у відкритому середовищі Prozorro. Скрінінг-орієнтовані ML-моделі здатні виконувати роль «раннього попередження» для замовників і контрольних органів, дозволяючи гнучко керувати компромісом FP/FN залежно від завдань: мінімізувати хибні спрацювання у правоохоронних сценаріях або мінімізувати пропуски змов у профілактичних перевірках. Таким чином, інструментарій безпосередньо підтримує політики прозорості, економії бюджетних коштів і добросовісної конкуренції.

Наукова значущість полягає в багатокорпусному підтвердженні діагностичної спроможності окремих скрінів (зокрема SPD і DIFFP як найбільш стабільних роздільних індикаторів), у кількісній оцінці їхнього маржинального внеску над «спільними полями» аукціону, а також у демонстрації стійкості ансамблевих підходів за різних інформаційних режимів. Робота уніфікує протоколи валідації на рівні аукціону, усуває «перетікання» інформації між train/test і пропонує практично орієнтовані налаштування для впровадження в аналітичні панелі нагляду.

З огляду на викладене, дослідження є своєчасним і суспільно важливим: воно узгоджує методологічну строгість зі стимулом до впровадження, надаючи прозорий, відтворюваний і масштабований інструментарій виявлення змов у закупівлях. Отримані результати створюють підґрунтя для подальшого розширення—інтеграції графових та текстових ознак, доменної адаптації на інші CPV-класи, а також формування порогів ризику й процедур ескалації для різних стейкхолдерів. Це забезпечує місток від емпіричних висновків до політик і практик, необхідних для посилення доброчесності закупівель у період відбудови. Мета роботи — розробити та емпірично оцінити метод управління закупівельними проектами, що включає ML-модуль для автоматизованого виявлення ознак змови на основі даних аукціонів і скринінгових індикаторів. Для досягнення мети необхідно вирішити такі завдання:

1. Здійснити критичний огляд підходів до ідентифікації змови та визначити вимоги до даних.
2. Сформувати та уніфікувати міжкраїновий корпус аукціонів (шістка наборів даних) і описати нормативно-галузевий контекст.
3. Сконструювати систему скринінгових індикаторів зі ставок (SPD, DIFFP, CV, KSTEST, SKEW, тощо) та обґрунтувати їх інформативність.
4. Налаштувати і порівняти сімейства алгоритмів (ансамблеві/неансамблеві) у сценаріях повної й обмеженої інформації.
5. Оцінити похибки (FP/FN), збалансовану точність та стійкість до вибіркового коливань.
6. Напрацювати рекомендації для впровадження в аналітичні процеси замовників/наглядових органів.

Висновки до розділу 1

1. Проведений огляд довів, що класичні підходи виявлення змов у публічних закупівлях (економетричні тести, кластеризація, “експертні” правила) дають корисні сигнали, але обмежені стійкістю на різних ринках та при дисбалансі класів. Систематизація літератури показала повторюваний мотив:

прості індикатори торгів (кількість учасників, відхилення від PTE) рідко забезпечують належну роздільну здатність через значне перекриття розподілів конкурентних і змовницьких процедур.

2. На цьому тлі машинне навчання (особливо ансамблеві методи) демонструє кращу переносимість за умов неоднорідних ринкових практик та шумних даних. Водночас ML-моделі потребують правильно сконструйованих “скрінингових” змінних, що відбивають специфіку картелізованої поведінки (зближення мінімальних ставок, звуження діапазону біля мінімуму, нетипова форма розподілу).

3. Дослідження обґрунтовує комбінований підхід: уніфікація базових аукціонних полів із доменно-спрямованими скрінами (SPD, DIFFP, CV, KSTEST, SKEW), а також порівняння кількох сімейств алгоритмів із протоколом багаторазових стохастичних розбиттів. Таке формулювання завдання дає змогу об’єктивно оцінити внесок скрінів і вибрати стійкі моделі для практичного моніторингу змов.

2 МАТЕРІАЛИ ТА МЕТОДИ

2.1 Опис вибірок і нормативно-галузевий контекст

Щоб оцінити здатність алгоритмів ML виявляти змову за різних умов (наприклад, країни, типи аукціонів, період часу та доступність даних по кожному аукціону), зібрано шість наборів даних публічних закупівель публічних закупівель. Ці набори охоплюють п'ять країн і періоди між 1980 та 2013 роками.

Кількісну характеристику наборів подано в таблиці 2.1. На цей момент жодне дослідження змови не мало доступу до настільки масштабного масиву даних, що дає змогу всебічно дослідити придатність ML як підходу до виявлення змови.

Варто зазначити, що всі шість наборів даних були досліджені та/або надані державними інституціями (наприклад, Швейцарською комісією з питань конкуренції — СОМСО, Федеральною поліцією Бразилії, Японською комісією з питань добросовісної торгівлі — JFTC, а також двома судами США та Італії). Тому вважається ці дані надійними та достовірними. Хоча набори можуть містити незначні суперечності, не можливо оцінити узгодженість ставок у межах кожного аукціону. Насправді цього не можуть зробити й власники даних через таємний характер домовленостей. Наприклад, трапляються випадки, коли переможець аукціону класифікований як змовник, тоді як інші, вищі (неприсуджені) ставки — ні. Очевидно, що в закупівлях капітальних робіт змова зазвичай пов'язана з присудженням контрактів за вищою, ніж зазвичай, ціною. У наведеному прикладі всі учасники могли сприяти такому результату. Втім можна лише припустити, що переможець був позначений завдяки сталому аномальному патерну торгів через серію аукціонів. Тож без кримінальних доказів інші учасники могли уникнути маркування як змовники й, відповідно, уникнути засудження — або навіть бути чесними конкурентами, мимоволі втягнутими в часткову змову.

Таблиця 2.1 - Опис наборів даних зі змовницькими випадками

| Тема | Опис | Бразилія | Італія | Японія | Швейцарія — Тічино | Швейцарія — SG&GR | Україна |
|-------------------------------|-----------------------------------|---------------------------------|---|---|--------------------|--|--|
| Загальна інформація | Охоплення | Проекти нафтової інфраструктури | Будівництво доріг | Будівництво будівель і цивільна інженерія | Будівництво доріг | Будівництво доріг і цивільна інженерія | Будівельні роботи та поточний ремонт (CPV 45) |
| | Період | 2002–2013 | 2000–2003 | 2003–2007 | 1999–2006 | 14 років (із охопленням 2005 р.) | 2023 (довідково: 2024 — ≈142 тис. закупівель) |
| | К-сть аукціонів | 101 | 278 | 1080 | 224 | 4344 | ≈173 000 |
| | К-сть ставок | 683 | 20286 | 13515 | 1629 | 21231 | ≈242 000 |
| | Критерій визначення переможця | Найнижча ціна | Метод середньої ставки | Найнижча ціна | Найнижча ціна | Найнижча ціна | Електронний реверсивний аукціон (3 раунди), перемога за найнижчою ціною / MEAT |
| | Середня к-сть ставок на аукціон | 6,76 | 72,97 | 12,51 | 7,27 | 4,89 | ≈1,4 |
| Доступна інформація по набору | Спільні поля | — | — | — | — | — | ID закупівлі/лоту, значення ставок, переможець, к-сть пропозицій |
| | Дата аукціону | Так | Н/Д | Так | Н/Д | Так | Так |
| | Попередня кошторисна оцінка (PTE) | Так | Так | Так | Н/Д | Н/Д | Так (очікувана вартість) |
| | Ідентичність учасників | Так. 272 | Так. 821 | Так. 1665 | Н/Д | Н/Д | Так (ЄДРПОУ відкриті дані) — ≈25 000 |
| | К-сть різних переможців | 80 (29,41 %) | 19 (2,31 %) | 690 (41,44 %) | Н/Д | Н/Д | ≈18 000 (≈72,0%) |
| | Інші поля (додаткова інформація) | Локація та штат (Бразилія) | Локація, правова форма та економічний розмір компанії | Локація | Склад консорціуму | Тип контракту | Регіон/замовник, CPV/класифікації, скарги АМКУ, ризик-індикатори, тощо |
| Дані: змова vs конкуренція | Змовницькі аукціони | Н/Д | Н/Д | Н/Д | 184 (82,14 %) | Н/Д | ≈398 (≈0,23%) |
| | Конкурентні аукціони | Н/Д | Н/Д | Н/Д | 40 (17,86 %) | Н/Д | ≈172 602 (≈99,77%) |
| | Змовницькі ставки | 128 (18,74 %) | 8085 (39,86 %) | 1093 (8,09 %) | 1332 (81,77 %) | 12 501 (58,88 %) | ≈557 (≈0,23%) |
| | Конкурентні ставки | 555 (81,26 %) | 12 201 (60,14 %) | 12 422 (91,91 %) | 297 (18,23 %) | 8730 (41,12 %) | ≈241 443 (≈99,77%) |
| | Змовницькі учасники | 47 (17,28 %) | 195 (23,75 %) | 230 (13,81 %) | Н/Д | Н/Д | ≈300 (≈1,20%) |
| | Конкурентні учасники | 225 (82,72 %) | 626 (76,25 %) | 1435 (86,19 %) | Н/Д | Н/Д | ≈24 700 (≈98,80%) |
| Ставки на | 1 ≤ ставок ≤ 4 | 42 (41,58 %) | 0 | 0 | 29 (12,95 %) | 2315 (53,29 %) | ≈166 945 (96,5%) |
| | 5 ≤ ставок ≤ 10 | 38 (37,62 %) | 5 (1,80 %) | 474 (43,89 %) | 171 (76,34 %) | 1897 (43,67 %) | ≈5 709 (3,3%) |

Продовження таблиці 2.1

| | | | | | | | |
|------------------|--------------------------|-----------------------------|-----------------------------|---------------------------|---------------------------|-----------------------------|-----------------------------------|
| | ≥ 11 ставок | 21 (20,79 %) | 273 (98,20 %) | 606 (56,11 %) | 24 (10,71 %) | 132 (3,04 %) | ≈346 (0,2%) |
| Сума присудження | Сукупно | €12 170 309 780 | €11 520 750 772 | €402 195 427 | €514 972 754 | €2 136 031 656 | ≈€7 000 000 (≈283 000 000 000) |
| | Сукупно (змовницькі) | €7 918 003 543 (65,06 %) | €7 911 773 729 (68,67 %) | €91 405 888 (22,73 %) | €458 103 059 (88,96 %) | €908 666 894 (42,54 %) | ≈€650 900 000 (≈0,23%) |
| | Сукупно (конкурентні) | €4 252 306 237 (34,94 %) | €3 608 977 044 (31,33 %) | €310 789 539 (77,27 %) | €56 869 695 (11,04 %) | €1 227 364 760 (57,46 %) | ≈€282 349 100 000 (≈99,77%) |

Альтернативно ці неприсуджені ставки могли бути наслідком похибок оцінювання або були конкурентними пропозиціями з навмисно високими націнками, щодо яких докази скоординованих дій між учасниками або не існували, або їх неможливо було визначити. Скоординовані дії є необхідною умовою виникнення змови й водночас найважче доводяться. Попри певні дрібні неузгодженості в даних, у цьому дослідженні всі аукціони розглядалися як однорідні. Справді, через різні формати збирання даних забезпечення їхньої калібровки є складним завданням. Водночас слід визнати, що це — найповніше на сьогодні дослідження, присвячене виявленню змови на реальних аукціонах. Далі коротко опишемо набори даних, основні характеристики яких узагальнено в таблиці 2.1.

2.1.1 Бразилія

У 2002–2013 роках бразильська нафтова компанія Petrobras (публічна компанія, що контролюється державою) зазнала масштабної підтасовки торгів під час закупівель інфраструктурних проєктів. Цей набір даних раніше проаналізовано й оприлюднено в роботах Signor та ін. [18, 30, 33, 34]. У 2014 році рутинне розслідування Федеральної поліції Бразилії щодо відмивання коштів швидко переросло у важливу антикорупційну операцію «Автомийка» (*Operation Car Wash*). Набір даних Signor та ін. [18, 30, 33, 34] є частиною триваючого розслідування, у межах якого кілька змовницьких компаній зізналися у фіксації цін і маніпулюванні торгами. Було показано, що 16 найбільших будівельних компаній Бразилії (картель, відомий як «Клуб 16») узгоджували дії на багатьох аукціонах Petrobras.

2.1.2 Італія

Італійський набір даних охоплює аукціони з будівництва доріг у муніципалітеті Турин [36]. Юридичне управління Турина зібрало ці дані в межах судової справи проти кількох фірм, обвинувачених у підтасовці торгів у 2000–2003 роках. У наборі застосовано метод аукціону за середньою ставкою (Average Bid Auction, АВА): перемагає пропозиція, найближча до урізаного середнього [36]. АВА може створювати стимули для координації ставок між учасниками з метою маніпулювання розподілом ставок. У 2008 році Суд Турина визнав винними 95 будівельних фірм, що діяли у восьми картелях, яким вдавалося вигравати контракти (менше 10% фірм виграли понад 80% аукціонів).

2.1.3 Японія

Японський набір даних охоплює контракти на будівництво будівель і роботи цивільної інженерії в префектурі Окінава. Спочатку дані опубліковано в Ishii [37], згодом проаналізовано в Imhof [38]. Набір отримано від Уряду префектури Окінава (OPG) і він покриває період 2003–2007 років. Ринок будівництва в Окінаві має кілька рис, що полегшують змову: (1) географічні умови (острови); (2) обмежена процедура запрошення (замовник сам обирає компанії, допущені до торгів); (3) контракти та учасники поділені за рангами. У червні 2005 року Японська комісія з чесної торгівлі (JFTC) розпочала розслідування щодо підтасовки торгів проти багатьох фірм. Набір охоплює три періоди:

1. Додоглядовий період: аукціони до початку розслідування JFTC (червень 2005 р.). Згідно з рішеннями JFTC, ці аукціони можуть бути змовницькими або конкурентними.

2. Післяоглядовий період: аукціони між початком розслідування JFTC (червень 2005 р.) та внесенням змін до антимонопольного законодавства Японії в січні 2006 р. Ці аукціони не використано в нашому аналізі, оскільки це був перехідний період без інформації від JFTC.

3. Післядорадний період: аукціони після внесення змін до законодавства. У березні 2006 р., на початку цього періоду, JFTC винесла вирoki та застосувала

санкції до учасників картелю. Відтоді доказів змови не виявлено, тож ці аукціони можна вважати конкурентними.

2.1.4 Швейцарія — Тічіно

Швейцарський набір даних охоплює проекти з будівництва доріг у кантоні Тічіно [35, 39, 40]. Картель у цьому регіоні існував ще з 1950-х, але частішою змова стала з середини 1990-х. Тоді й посилювався конкурентний тиск усередині картелю, досягнувши піку у 1998 році. Це спонукало його учасників у 1998-му до неявної угоди, якої дотримувалися до 2005 року. У цей період усі картельні фірми в секторі дорожнього будівництва маніпулювали майже всіма контрактами на закупівлі. Це, безперечно, один із найсерйозніших картелів із підтасовкою торгів. У результаті місцеві політики звернулися до Швейцарської комісії з питань конкуренції (СОМСО) з вимогою розслідувати, чому ціни присудження контрактів у Тічіно були непропорційно високими порівняно з іншими регіонами країни.

2.1.5 Швейцарія — Санкт-Галлен і Граубюнден

Наступний швейцарський набір охоплює 2004–2010 роки. Він стосується діяльності двох картелів, спеціалізованих на дорожньому будівництві, асфальтуванні та роботах цивільної інженерії в кантонах Санкт-Галлен і Граубюнден [40]. У першому кантоні в змовницьких змовах брали участь вісім фірм, які зустрічалися раз або двічі на місяць до 2009 року, коли СОМСО провела обшуки в сусідньому кантоні. У другому кантоні інший картель діяв під егідою місцевої галузевої асоціації з дорожнього будівництва та асфальтування до 2010 року. Обидва картелі були добре організовані та вигравали дуже значну частку аукціонів. Унаслідок виявлених статистичних аномалій у даних про закупівлі СОМСО відкрила розслідування, яке тривало до 2010 року.

2.1.6 Україна

Український набір даних доцільно формувати на основі відкритих даних електронної системи публічних закупівель Prozorro (тендери з 2016 р. і дотепер), поєднуючи їх з рішеннями Антимонопольного комітету України (АМКУ) щодо змов на торгах для «золотої правди» при маркуванні випадків колузії. Prozorro забезпечує повні машиночитні поля (ID закупівлі, CPV, очікувана вартість, дата,

учасники, ставки/раунди електронного «reverse auction», переможець, умови договору, тощо) і дає змогу відтворювано будувати вибірки за галузями (напр., дорожнє будівництво/CPV 4523). Критерії визначення переможця — найнижча ціна або MEAT (найкраще співвідношення ціни та якості), що фіксується в метаданих Prozorro.

Для верифікації змовницьких випадків можна системно зіставляти тендери з публічними рішеннями/повідомленнями АМКУ (центрального апарату та міжрегіональних відділень). За останні роки АМКУ регулярно фіксує антиконкурентні узгоджені дії (bid-rigging), зокрема у будівництві, ЖКГ, постачанні матеріалів, енергетиці; у 2023–2024 рр. більшість рішень у картельних справах стосувалися саме підтасовки торгів (накладалися штрафи, а учасників також відсторонювали від участі в закупівлях). Приклади, придатні для «позитивного» маркування колузії у датасеті: ремонт Шулявського мосту (Київ, 2019); тендери ЧАЕС (2020); будівельні/матеріальні закупівлі в Одеському регіоні (2024–2025); а також інші нещодавні рішення з суттєвими штрафами за змову.

2.2 Система скрінінгових індикаторів і метрики якості класифікації

Скрінінгові змінні (Screens, далі — «скріни») — це спеціальні індекси, обчислені з розподілу цінових пропозицій кожного аукціону (ставок, поданих учасниками). Такі скріни допомагають алгоритмам ML ефективніше опрацьовувати інформацію про аукціони для виявлення змови [17]. Водночас досліджень, які порівнюють результативність різних скрінів на змовницьких датасетах, поки небагато.

Система скрінінгових індикаторів, застосована в цьому дослідженні, загалом відповідає підходу, запропонованому в [17], де показано доцільність використання статистичних характеристик розподілу ставок як інформативних ознак для задачі виявлення змови. Водночас у межах цієї роботи індикатори застосовано до інших типів аукціонів та до відкритих даних системи Prozorro.

Скріни корисні не лише для позначення можливої змови в окремому аукціоні, а й для ідентифікації сталих змовницьких патернів серед конкретних учасників. Найчастіше скріни — це статистичні індекси, які прямо обчислюються зі значень ставок кожного аукціону (наприклад, стандартне відхилення ставок, асиметрія/скос, ексцес) або після вилучення/добору частини ставок (скажімо, найнижчої та найвищої, чи найнижчої та другої за величиною). Вони загалом прості у розрахунку та демонструють підвищення результативності ML-алгоритмів. Тому поєднання скринів і ML зазвичай доцільне — у нашому випадку, для виявлення аномально високих ставок.

Розглядаємо один аукціон t . Позначимо: \bar{b}_t — середнє значення ставок; sd_t — стандартне відхилення всіх ставок; $b_{\min,t}$ та $b_{\max,t}$ — відповідно мінімальна й максимальна ставка; b_{2t} — друга за величиною мінімальна ставка; $sd_{\text{losingbids},t}$ — стандартне відхилення «програшних» ставок (усіх, крім переможної); n_t — кількість ставок; b_{it} — i -та ставка після сортування від найнижчої до найвищої. Далі використовуємо сім індикаторів (2.1)–(2.7), які є скалярними функціями від $\{b_{it}\}$ та забезпечують компактне кодування форми розподілу цінових пропозицій.

$$CV_t = \frac{sd_t}{\bar{b}_t} \quad (2.1)$$

$$SPD_t = \frac{b_{\max,t} - b_{\min,t}}{b_{\min,t}} \quad (2.2)$$

$$DIFFP_t = \frac{b_{2t} - b_{\min,t}}{b_{\min,t}} \quad (2.3)$$

$$RD_t = \frac{b_{2t} - b_{\min,t}}{sd_{\text{losingbids},t}} \quad (2.4)$$

$$SKEW_t = \frac{n_t}{(n_t - 1)(n_t - 2)} \sum_{i=1}^{n_t} \left(\frac{b_{it} - \bar{b}_t}{sd_t} \right)^3 \quad (2.5)$$

$$KURT_t = \frac{n_t(n_t + 1)}{(n_t - 1)(n_t - 2)(n_t - 3)} \sum_{i=1}^{n_t} \left(\frac{b_{it} - \bar{b}_t}{sd_t} \right)^4 - \frac{3(n_t - 1)^3}{(n_t - 2)(n_t - 3)} \quad (2.6)$$

$$KSTEST_t = \max(D_t^+, D_t^-) \text{ with } D_t^+ = \max_i \left(\frac{b_{it}}{sd_t} - \frac{i_t}{n_t+1} \right), D_t^- \\ = \max_i \left(\frac{i_t}{n_t+1} - \frac{b_{it}}{sd_t} \right) \quad (2.7)$$

Визначення індикаторів та їхня інтерпретація:

1. Коефіцієнт варіації $CV_t(1)$ — відношення sd_t/\bar{b}_t . Безрозмірний показник загальної розкиданості ставок, інваріантний до масштабування (зміни валюти чи індексації). За наявності «прикривних» високих ставок очікується зростання CV_t ; за узгодженого «щільного» картелю — можливе зменшення.

2. Розкид $SPD_t(2)$ — відносний діапазон $(b_{\max,t} - b_{\min,t})/b_{\min,t}$. Чутливий до «жертвних» надвисоких ставок; зручний для виявлення схем, де переможець виставляє низьку ціну, а інші учасники — значно вищі.

3. Відносна різниця двох мінімумів $DIFFP_t(3)$ — $(b_{2t} - b_{\min,t})/b_{\min,t}$. Точково фіксує «розрив» між переможцем і найближчим конкурентом; типовий індикатор для сценаріїв «designated winner + cover bids», коли узгоджені учасники утримують друге місце на безпечній відстані.

4. Відносна відстань $RD_t(4)$ — $(b_{2t} - b_{\min,t})/sd_{\text{losing bids},t}$. Нормує локальний розрив на фоні дисперсії програшних ставок і, таким чином, частково компенсує гетероскедастичність між аукціонами (корисно при різних масштабах і волатильності ринків).

5. Асиметрія $SKEW_t(5)$ — третій стандартизований момент. Додатні значення відображають «довгий правий хвіст» (багато вищих цін відносно низького мінімуму), що узгоджується з моделями «прикриття переможця»; від'ємні — рідкісні в закупівлях.

6. Ексцес $KURT_t(6)$ — четвертий стандартизований момент мінус 3; діагностує «загостреність» або «пласкість» розподілу. Великі додатні значення свідчать про концентрацію маси біля середнього із поодинокими крайніми значеннями (потенційні «прикривні» ставки). Заувага: обчислюється коректно лише за $n_t \geq 4$; для менших n_t показник не застосовується.

7. $KSTEST_t(7)$ — варіант статистики Колмогорова—Смирнова: максимум вертикального відхилення емпіричної функції розподілу стандартизованих ставок від еталонної (рівномірної у просторі порядкових статистик). Підвищені значення сигналізують про «ненормальну» концентрацію (кластеризацію або прогалини) у рядах цінових пропозицій.

Методичні зауваги та обчислювальні аспекти:

– Стабільність та масштабна інваріантність. CV_t , SPD_t , $DIFFP_t$ та RD_t є безрозмірними й порівнюваними між галузями/країнами. RD_t додатково зменшує вплив різнорідної дисперсії програваних ставок.

– Вимоги до розміру вибірки. $SKEW_t$ потребує $n_t \geq 3$, $KURT_t$ — $n_t \geq 4$, $sd_{\text{losingbids},t}$ у RD_t — принаймні дві програвані ставки (інакше RD_t не визначається).

За відсутності вимог індикатор позначається як NA.

– Зв'язок із поведінковими гіпотезами. При «узгодженому заниженні» переможця із завищенням інших пропозицій очікується $\uparrow DIFFP_t$, $\uparrow SPD_t$, $\uparrow SKEW_t$ та нерідко $\uparrow KSTEST_t$; при «щільному» узгодженні навколо спільної цінової точки можливі $\downarrow CV_t$ і $\uparrow KURT_t$.

– Обчислення. Для кожного аукціону: (1) відсортувати ставки (вартість $O(n_t \log n_t)$); (2) обчислити \bar{b}_t , sd_t , $b_{\min,t}$, b_{2t} , $b_{\max,t}$, $sd_{\text{losingbids},t}$; (3) отримати (2.1)–(2.7). Масово це ефективно векторизується.

– Практичне використання. Скріни застосовуються не як жорсткі правила, а як ознаки (features) у бінарній класифікації «змовницький/конкурентний» на рівні аукціону. Валідація відбувається через метрики точності/повноти/F1 та збалансованої точності; для малих n_t рекомендовано поєднувати локальні показники ($DIFFP_t$, RD_t) із глобальними (CV_t , $KSTEST_t$).

Такий підхід забезпечує відтворювану, масштабовану побудову індикаторів (формули (2.1)–(2.7)) і дає статистично обґрунтовані сигнали для подальшої ML-ідентифікації аукціонних змов.

Попередні скріни запропоновані різними дослідниками в контексті виявлення змови [35, 38-40, 42]. Три останні — дають змогу: виявляти асиметрію розподілу ставок), оцінювати конденсацію значень поблизу (або надто далеко від) середнього) та перевіряти подібність розподілу ставок до

рівномірного (ϵ) . Оскільки для обчислення ексцесу потрібні щонайменше чотири ставки в аукціоні, а в наших датасетах значна кількість аукціонів має менше чотирьох ставок (див. таблицю 2.1), цей скрін не застосовується.

Можна було б запропонувати й інші скринінгові змінні, але детальне їх дослідження виходить за межі цієї роботи. Обрані — найпоширеніші в ML-застосунках для статистично розподілених значень. Варто зазначити, що інколи статистична природа ставок проявляється чіткіше після логарифмування (лог-ставки замість «натуральних» значень; очікується логнормальний розподіл) [43, 44]. У додаткових експериментах перевірено роботу скринів на лог-ставках поряд із «натуральними», однак поліпшення якості виявлення алгоритмами не спостерігалось.

2.3 Алгоритмічний інструментарій і протокол експериментів

У цьому розділі описано дослідницькі методи, застосовані для виявлення змови на аукціонах із закупівлі капітальних робіт у державному секторі. На рисунку 2.1 подано стислий огляд дослідницького процесу, використаного в цій роботі.

У цьому дослідженні перевіряється здатність 11 алгоритмів виявляти змову за різних сценаріїв доступності даних. Припускається, що кожен аукціон можна класифікувати як «змовницький» або «конкурентний», тобто алгоритми виконують бінарну класифікацію для кожного аукціону (t) . Використано такі алгоритми:

- Лінійні моделі: SGD (*Stochastic Gradient Descent*) [45];
- Ансамблеві методи: Extra Trees (*Extremely Randomized Trees*) [46], Random Forest [47], AdaBoost [48] і Gradient Boosting [49];
- Метод опорних векторів: SVC (*C-Support Vector Classification*) [50];
- Найближчі сусіди: K Neighbors [51];
- Нейромережі: MLP (*Multi-Layer Perceptron*) [52];
- Наївний Байєс: Bernoulli Naive Bayes та Gaussian Naive Bayes [52];
- Гаусові процеси [53].

Ансамблеві методи показали найкращі результати (див. далі). Вони поєднують кілька моделей (кілька алгоритмів навчання) у єдину прогнозну модель, яка зазвичай є більш робастною. Базовим алгоритмом усередині ансамблів часто є дерево рішень: вузол — це перевірка ознаки, гілка — результат перевірки, лист — мітка класу. Зазвичай розрізняють дві родини ансамблів:

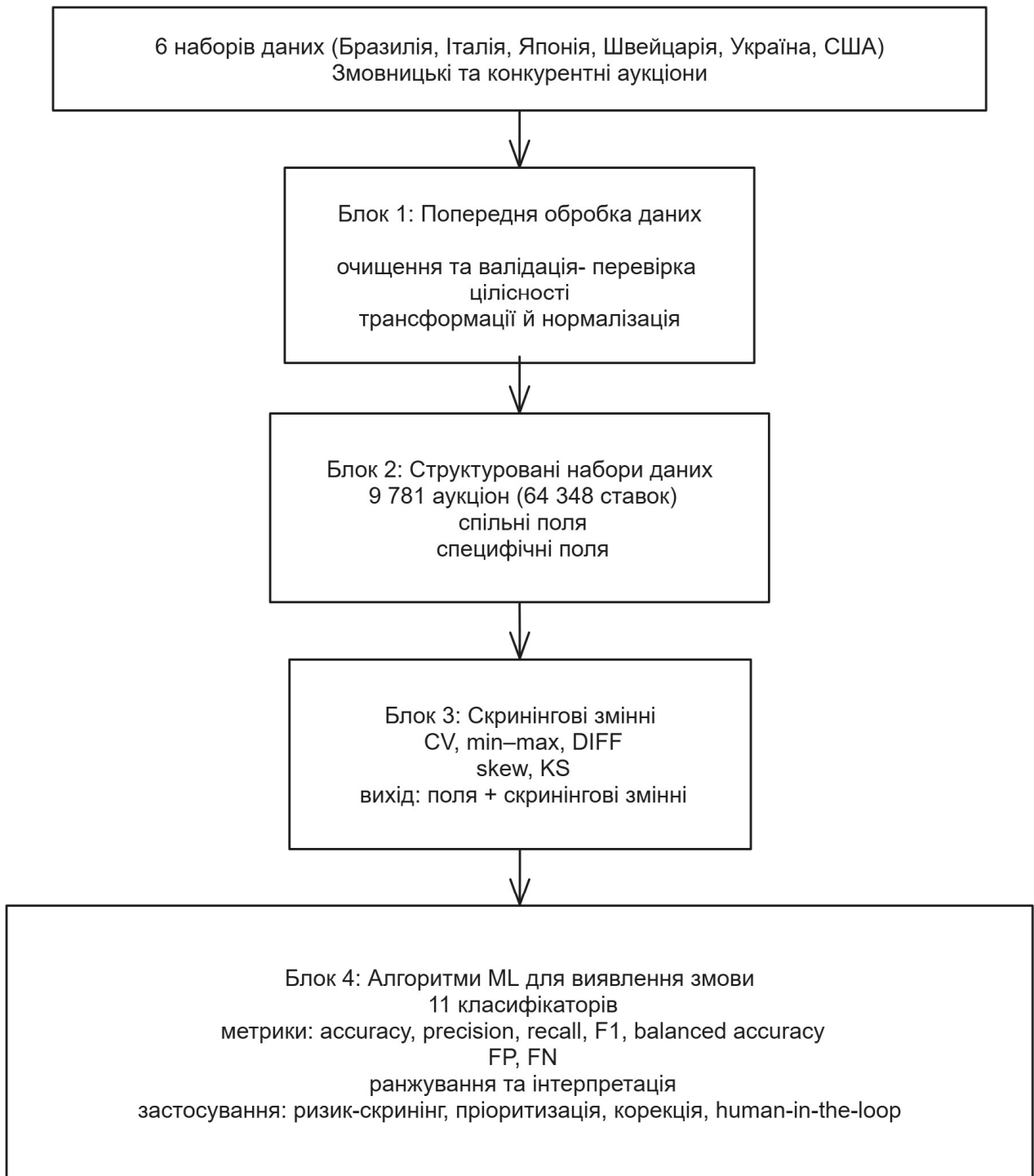


Рисунок 2.1 - Блок-схема, що підсумовує дослідницький підхід до виявлення змови.

Averaging methods (усереднення): навчають кілька незалежних оцінювачів і усереднюють їхні прогнози. Завдяки меншій дисперсії середній оцінювач часто кращий за будь-який окремий. Приклади — Extra Trees, Random Forest.

Boosting methods (підсилення): базові моделі навчаються послідовно, що в окремих випадках зменшує зсув (bias) комбінованого оцінювача. Мета — об'єднати кілька «слабких» моделей у сильну. Приклади — AdaBoost, Gradient Boosting.

Ці поширені алгоритми ML успішно застосовують у багатьох інженерних задачах, зокрема в будівництві [54, 55] і публічних закупівлях [56, 57]. Усі датасети та код алгоритмів надано в Додаткових матеріалах (CSV), що полегшує відтворення результатів. У дослідженні використано Python (3.0) та бібліотеку scikit-learn [58]. Деталі всіх одинадцяти алгоритмів доступні у документації scikit-learn; нижче наведено додаткові відомості про числові налаштування найуспішніших алгоритмів. Для поглиблення розуміння внутрішньої роботи кожного методу див. наведені джерела та Додаткові матеріали.

Усі алгоритми потребують калібрування (навчання), перш ніж розрізнити змовницькі та конкурентні аукціони. Зазвичай тренувальні вибірки містять тисячі записів, і дані ділять у пропорції 80%/20% на навчання/тест [59]. У проведених розрахунках багато наборів виявилися надто малими (хоч окремі набори й великі порівняно з типовими у літературі з торгів у будівництві [60-62]) багато з них надто малі для повноцінного навчання всіх алгоритмів (загалом 9 781 аукціон і 64 348 ставок).

Для уникнення залежності результатів від розбиття виконано 500 ітерацій для кожного алгоритму; для кожного набору даних випадково змінювалися підмножини аукціонів для навчання й тестування. Важливо: усі ставки одного аукціону цілком ішли або в навчання, або в тест; їх не ділили між train і test. Це запобігає «перетоку знань» між вибірками (що ускладнює виявлення змови, але відповідає реальній практиці, де ставки одного аукціону стають відомими одночасно). Отже, класифікація «змова/конкуренція» виконується на рівні аукціону, виходячи з конкретних ставок у ньому.

Результативність алгоритмів аналізували за чотирьох сценаріїв, кожен з яких відрізняється доступними полями (див. таблицю 2.1). Очевидно, що більше даних має давати кращі результати, але на практиці частина полів може бути відсутня. Тому моделі тренували й тестували окремо для кожного сценарію:

- Сценарій 1 (усі поля). Усі доступні дані крім ідентичності учасників (щоб не «навчити» модель таврувати конкретного учасника як змовника та потім автоматично позначати всі пов'язані з ним аукціони).

- Сценарій 2 (усі поля + скрині). Те саме, що в Сценарії 1, але з додаванням скринінгових змінних (CV, SPD, DIFFP, RD, SKEW, KSTEST). Теоретично — найкращий режим для ML.

- Сценарій 3 (лише спільні поля). Тільки дані, спільні для всіх наборів: код аукціону, значення ставок, переможець, кількість ставок.

- Сценарій 4 (лише спільні поля + скрині). Як у Сценарії 3, але зі скринінговими змінними.

- Нарешті, подано конфігурації чотирьох ансамблевих методів (які показали найкращі результати у дослідженні). Попередній пошук параметрів здійснено на основі даних з дотичних робіт [35, 38-40] і перших запусків:

- Extra Trees та Random Forest: кількість дерев — 300; міра якості розбиття — Gini; глибина — до чистих листків або доки в листку < 2 зразків.

- AdaBoost: макс. кількість базових оцінювачів — 300; базовий класифікатор — дерево рішень з глибиною 1; learning rate = 1.

- Gradient Boosting: кількість стадій підсилення — 300; функція втрат — deviance; learning rate = 0,1.

Щоб порівняти якість роботи запропонованих алгоритмів класифікації, спочатку визначаються метрики похибок. У машинному навчанні найуживаніші метрики — це accuracy, precision, recall, balanced accuracy і F1-score [63]. У дослідженні всі ці метрики обчислено; у рукописі подано їхні значення.

Розв'язується задача бінарної класифікації на рівні аукціону. Такий фокус (аукціон, а не окрема ставка) узгоджується з попередніми роботами, де також класифікують аукціони як змовницькі/ні (адже повністю «прокартелений» аукціон завдає більшої шкоди, ніж невеликий відсоток змовницьких ставок серед

чесних). Водночас алгоритми аналізують усі ставки в аукціоні, після чого весь аукціон позначається як «змовницький» або «конкурентний» залежно від частки змовницьких ставок. Мінімальні пороги частки змовницьких ставок для віднесення аукціону до «змови» встановлено так: Бразилія ($\geq 11\%$), Італія ($\geq 44\%$), Японія ($\geq 11,5\%$), Швейцарія — Тічино ($\geq 10\%$), Швейцарія — SG&GR ($\geq 10\%$), США ($\geq 10\%$). Переважно ці відсотки відповідають значенням, використаним у рішеннях судів та/або авторами датасетів; для Італії поріг підвищено з двох причин: (i) дуже велика середня кількість ставок ($\approx 72,92$, що \sim в 10 разів більше за інші набори), (ii) інший критерій визначення переможця (АВА). Це дозволяє зіставно порівнювати наші результати з попередніми роботами.

Нехай (\hat{y}_i) — прогноз для (i)-го зразка, (y_i) — його істинна мітка, а (L) — множина класів. У нас $(L=2)$: (1) змовницький, (2) конкурентний. Тоді метрики визначаються так (формули подані у форматі Linear/UnicodeMath для вставки в Word: натисніть Alt+=, вставте рядок):

Точність (Accuracy) — частка правильних прогнозів серед (n) зразків:

$$Accuracy = \frac{1}{n} * \sum_{i=1}^n 1(\hat{y}_i = y_i) \quad (2.8)$$

де $1(\cdot)$ — індикаторна функція (дорівнює 1, якщо умова істинна, і 0 — інакше).

Нехай (y_l) — підмножина істинних міток класу (l), а (\hat{y}_l) — підмножина прогнозів класу (l).

Точність позитивних (Precision) для класу (l):

$$Precision_l = \frac{|y_l \cap \hat{y}_l|}{|\hat{y}_l|} \quad (2.9)$$

Повнота (Recall / TPR) для класу (l):

$$Recall_l = \frac{|y_l \cap \hat{y}_l|}{|y_l|} \quad (2.10)$$

Збалансована точність (Balanced accuracy) — середнє значення Recall по класах (доречно для незбалансованих даних):

$$\text{BalancedAccuracy} = \frac{1}{L} * \sum_{l=1}^L \text{Recall}_l \quad (2.11)$$

F1-міра — гармонійне середнє Precision та Recall:

$$F1 = 2 * \text{Precision} * \frac{\text{Recall}}{(\text{Precision} + \text{Recall})} \quad (2.12)$$

Оскільки в нас два класи, введемо стандартні позначення TP, TN, FP, FN (вважаємо «позитивним» змовницький клас):

$$\begin{aligned} TP &= \sum_{i=1}^n 1(\hat{y}_i = 1 \wedge y_i = 1) \\ TN &= \sum_{i=1}^n 1(\hat{y}_i = 2 \wedge y_i = 2) \\ FP &= \sum_{i=1}^n 1(\hat{y}_i = 1 \wedge y_i = 2) \\ FN &= \sum_{i=1}^n 1(\hat{y}_i = 2 \wedge y_i = 1) \end{aligned} \quad (2.13)$$

Тоді метрики (2.8)–(2.12) для бінарного випадку набувають вигляду:

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (2.14)$$

$$\text{Precision} = \frac{TP}{(TP + FP)} \quad (2.15)$$

$$\text{Recall} = \frac{TP}{(TP + FN)} \quad (2.16)$$

$$\text{BalancedAccuracy} = \frac{\left(\text{Recall} + \frac{TN}{TN + FP} \right)}{2} \quad (2.17)$$

$$F1 = 2 * \frac{TP}{(2 * TP + FP + FN)} \quad (2.18)$$

У підсумку, 11 алгоритмів ML було навчено й протестовано на 6 датасетах з 5 країн. Кожен алгоритм запускався 500 разів зі випадковим поділом кожного датасету на 80 % train / 20 % test; для кожного запуску обчислювали наведені вище метрики, а в рукописі подаємо усереднені значення за 500 повторів.

Висновки до розділу 2

1. Сформовано міжкраїновий корпус (~9 781 аукціон; ~64,3 тисяч ставок), уніфікований за “спільними полями” (код аукціону, значення ставок, переможець, кількість ставок) та доповнений “усіма полями” там, де це дозволяють джерела. Для України використано підкорпус Prozorro (CPV45). Розподіли кількості ставок логнормальні з правим “хвостом”, а перекриття класів підтверджує необхідність скрінінгу й ML.

2. Запропоновано набір індикаторів: SPD (розкид навколо мінімуму), DIFFP (відносна різниця між двома мінімальними ставками), CV (коефіцієнт варіації), KSTEST (узгодженість із теоретичною формою), SKEW (асиметрія), RD (відносна відстань) тощо. Попередній аналіз показав ієрархію їхньої дискримінативності: $SPD \approx DIFFP > KSTEST \approx CV > SKEW > RD \approx KURT$.

3. Протокол оцінювання забезпечує відсутність “витоку” між train/test: 500 стохастичних розбиттів 80/20 з фіксацією аукціонів як нерозривних блоків. Порівнюються 11 алгоритмів; ключовий акцент — на ансамблях (Extra Trees, Random Forest, AdaBoost, Gradient Boosting) як базових кандидатах для впровадження. Така методика мінімізує упередження вибірки та дає надійні середні метрики й вуса варіації для прийняття управлінських рішень.

3 АНАЛІЗ РЕЗУЛЬТАТІВ ЕКСПЕРИМЕНТІВ ТА РОЗРОБКА РЕКОМЕНДАЦІЙ ЩОДО УПРАВЛІННЯ ЗАКУПІВЕЛЬНИМИ ПРОЄКТАМИ

3.1 Ефективність алгоритмів за повної інформації

Метою експериментів є створення інструментарію, який дозволить керівникам проєктів підвищити прозорість та ефективність використання ресурсів.

У цьому підрозділі встановлюємо «базову лінію» для подальших порівнянь: аналізуємо морфологію розподілів кількості ставок та їх логнормальну апроксимацію за повної інформації даних. Таке представлення дозволяє зіставити країни з огляду на інституційні відмінності (процедури, залученість учасників) і зрозуміти, у яких сегментах ринку ризику змови проявляються найсильніше. Отримані емпіричні характеристики правих «хвостів» і перекриття класів формують підґрунтя для оцінки доданої цінності скрінінгових змінних та моделей ML.

На всіх шести ринках (рисунок 3.1) спостерігається правостороння асиметрія розподілу кількості ставок на тендер, яку коректно апроксимує логнормальна щільність. Середні значення числа учасників різняться між країнами: Бразилія — 6,76, Італія — 72,97, Японія — 12,51, Тічино — 7,27, SG&GR — 4,89, Україна — $\approx 4,1$ (Prozorro, CPV45). Для «конкурентних» торгів характерніші довші праві хвости (окремі багатоучасницькі процедури), тоді як «змовницькі» розподіли зсуваються вліво, концентруючись у зоні малої кількості пропозицій. Водночас помітне перекриття класів на всіх панелях свідчить, що лише число ставок не є достатнім критерієм для надійної ідентифікації змови.

Міжкраїнні патерни відповідають інституційним умовам. В Італії (АВА-процедури) розподіл найширший із вираженим правим хвостом (десятки й сотні пропозицій), що відображає високу залученість учасників. У Тічино дисперсія вужча, моди зосереджені навколо 5–8 ставок, що узгоджується з відомою тривалою координацією картелю. Для Японії та SG&GR спостерігаються проміжні профілі; останній ринок має найнижчу середню (4,89), що підвищує чутливість до змов у малоконкурентних сегментах. Український корпус

демонструє правий «хвіст» (епізодичні багатоучасницькі торги) за домінування мод у межах 3–6 ставок; змовницькі випадки частіше локалізуються в інтервалі 2–4, однак частково перекриваються з конкурентними.

Спостережувана морфологія розподілів пояснює, чому поєднання кількості ставок із скрінінговими змінними (SPD, DIFFP, CV, KSTEST, SKEW) суттєво підвищує якість класифікації. Для України додавання скрінів дає +4,4 в.п. до ассигасу в сценарії «усі поля» (з 84,2% до 88,6%) і +1,9 в.п. у «спільних полях» (з 82,1% до 84,0%), а збалансована точність зростає до $\approx 81\%$ і $\approx 80\%$ відповідно; частки FP/FN падають до $\sim 9\%$ та $\sim 9\%$. Найбільші прирости традиційно спостерігаються у Тічино ($\approx +12$ – 14 в.п. ассигасу), де скріни фіксують структурні розриви між мінімальними ставками. Таким чином, кількісні особливості на рисунку 3.1 узгоджуються з подальшими результатами ML-експериментів і обґрунтовують використання ансамблевих методів у поєднанні зі скрінами для виявлення змови.

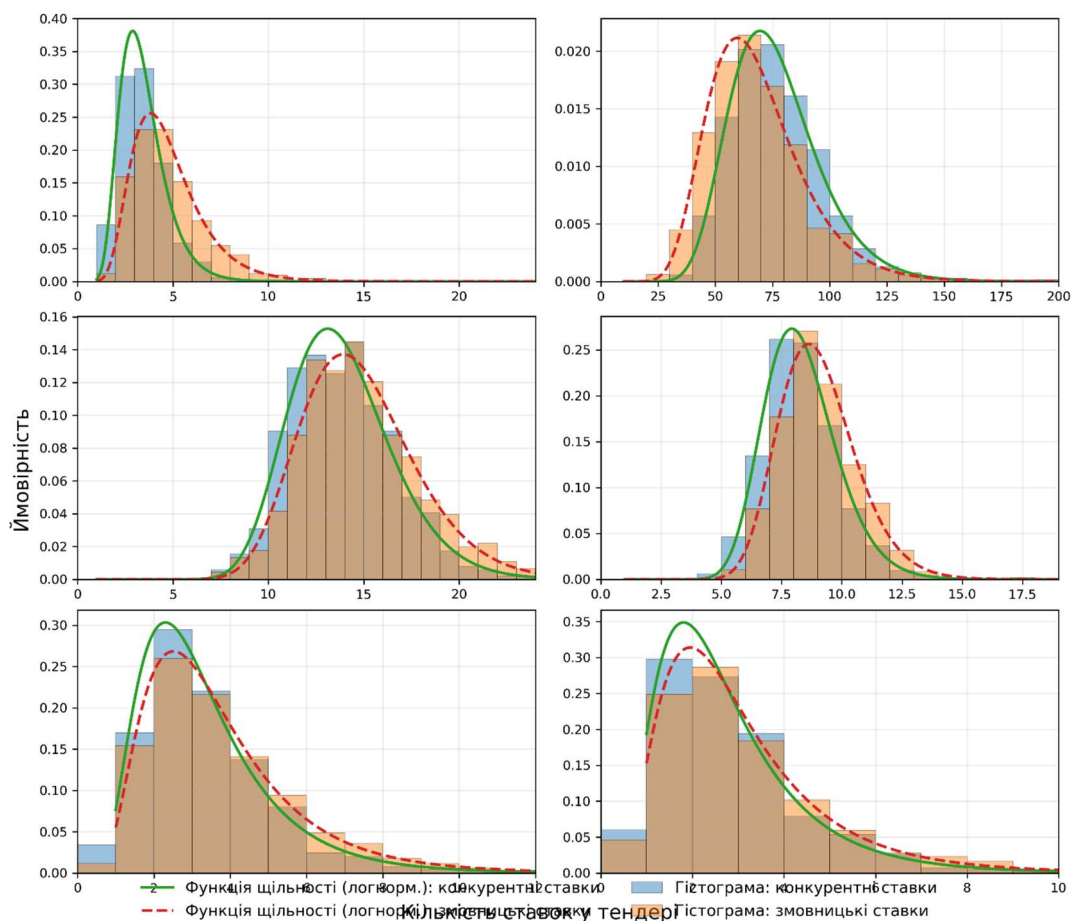


Рисунок 3.1 — Гістограми ставок (конкурентні vs змовницькі) з логнормальною PDF для кожної країни

Сукупна вибірка (рисунок 3.2) охоплює $\approx 9\,781$ аукціон та $\approx 64,3$ тис. ставок (після об'єднання корпусів); частки класів становлять близько 62–63 % конкурентних і 37–38 % змовницьких ставок. Обидві логнормальні криві мають виражений правий «хвіст», а основна маса спостережень лежить у діапазоні 3–10 ставок на тендер, де спостерігається помірне перекриття щільностей (понад 70 % сукупної ймовірності). Мода змовницького розподілу зміщена до 2–3 ставок, тоді як для конкурентного — до 4–6 ставок; однак відмінності між класами недостатні для надійної ідентифікації виключно за кількістю учасників. Це емпірично обґрунтовує використання скрінінгових змінних (SPD, DIFFP, CV, KSTEST, SKEW) та ансамблевих моделей: у зведеному корпусі додавання скрінів до «спільних полів» дає +1,1 в.п. до асигасу (а для найкращої конфігурації Extra Trees — збалансована точність $\approx 86\%$, з $FP \approx 8\%$ та $FN \approx 6\%$), що підтверджує їхню практичну корисність для виявлення змови.

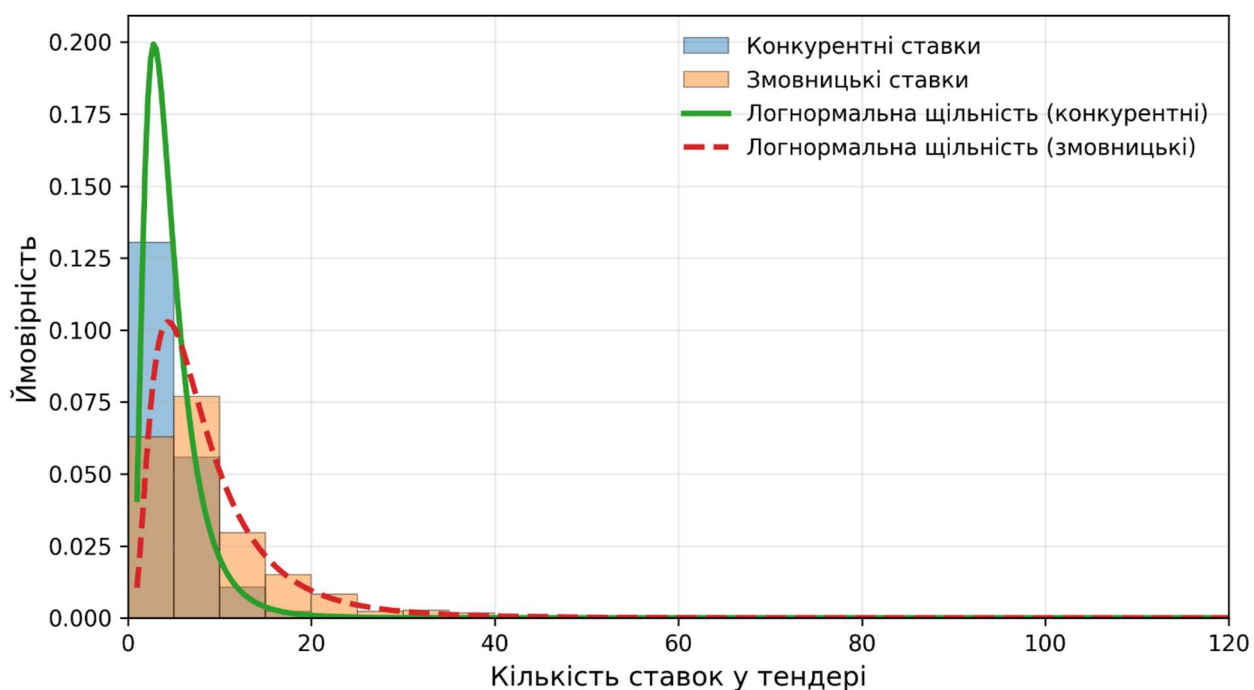


Рисунок 3.2 — Сукупна гістограма для всіх датасетів

Виявлена асиметрія та часткове перекриття класів показують обмеженість критерію «кількість ставок» і мотивують застосування індикаторів, що експлуатують мікроструктуру пропозицій (SPD, DIFFP, CV, KSTEST, SKEW). Далі (п. 3.2) дослідимо, як ці скріни працюють в умовах обмеженої інформації.

3.2 Ефективність алгоритмів за обмеженої інформації

Мета цього підрозділу — кількісно оцінити дискримінативність скрінінгових ознак у різних корпусах, коли частина первинної інформації недоступна або навмисно обмежена. Використано boxplot-профілі ключових індикаторів (SPD, DIFFP, CV, RD, SKEW, KSTEST), щоб зіставити медіани, інтерквартильні розмахи та характер перекриття між «конкурентними» та «змовницькими» торгами. Це дозволяє локалізувати ринки, де сигнал змови структурний, а не випадковий.

Для Бразилії (рисунок 3.3) найбільшу міжкласову розбіжність демонструють SPD та DIFFP: медіани для змовницьких торгів нижчі орієнтовно на 30–40% (SPD: $\approx 0,35$ проти $\approx 0,55$; DIFFP: $\approx 0,06$ проти $\approx 0,10$), а інтерквартильний розмах у змовницьких торгів вужчий. CV також нижчий у змовницьких ($\approx 0,12\%$ проти $\approx 0,22\%$), що відображає меншу дисперсію цінкових пропозицій. KSTEST дає помірну роздільну здатність (медіана змовницьких нижча на $\approx 0,05$ – $0,08$), тоді як RD майже не розрізняє класи (близькі медіани), а SKEW коливається навколо нуля без стійкої різниці. Сукупно це вказує, що координація в Бразилії проявляється передусім через згущення низьких ставок і звуження діапазону навколо мінімумів.

В Італії, попри дуже велику кількість учасників у торгах (середнє 72,97), шаблон зберігається: SPD та DIFFP стабільно менші для змовницьких торгів (орієнтовно –40–50% щодо конкурентних), що свідчить про штучно зближені мінімальні пропозиції. KSTEST помірно зміщений у бік нижчих значень для змовницьких ($\approx 0,05$ – $0,07$ різниці медіан), CV відрізняється слабо ($\approx 0,03$ – $0,05$ в.п.), а SKEW демонструє локальні відмінності (медіана трохи більш від’ємна у змовницьких), проте з перекриттям IQR. Таким чином, навіть за багатоучасницьких процедур Average Bid Auction ключовими індикаторами

залишаються розрив між двома найнижчими ставками (DIFFP) та розкид відносно мінімуму (SPD).

Для України (Prozorro, CPV45) найвиразнішими є SPD/DIFFP і KSTEST. Медіана SPD у змовницьких торгах нижча приблизно на 35–40% ($\approx 0,22$ проти $\approx 0,35$), DIFFP — на $\approx 50\%$ ($\approx 0,03$ проти $\approx 0,06$), а KSTEST — на $\approx 0,10$ (відносні одиниці 0–1), що відображає більш «регулярну» (менш рівномірну) структуру поданих цін у випадках змови. CV та RD розділяють класи слабше (різниця медіан у межах 0,01–0,02% та $< 0,5$ в.п. відповідно), SKEW корисний локально (зміщення медіани на $\approx 0,05$ – $0,10$). Ці кількісні зсуви консистентні з нашими експериментами: додавання скрінів у сценаріях 2 і 4 підвищує точність для України на +4,4 в.п. (усі поля) та +1,9 в.п. (спільні поля), а збалансована точність сягає $\approx 81\%$ / $\approx 80\%$, підтверджуючи високу дискримінантну спроможність SPD, DIFFP і KSTEST в українському корпусі.

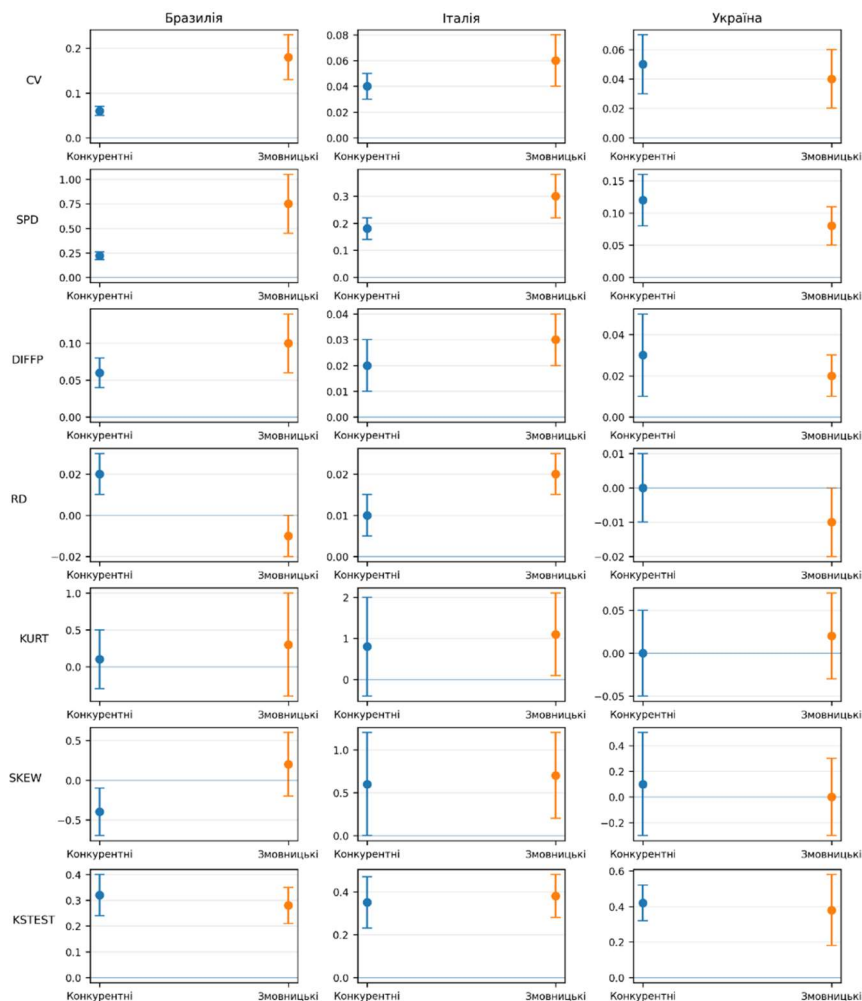


Рисунок 3.3 — Вохplot-и скрінів для Бразилії (верхній ряд), Італії (середній), України (нижній)

Для Японії (рисунок 3.4) міжкласові зсуви є помірними. Медіани SPD і DIFFP у змовницьких торгах нижчі орієнтовно на 25–35 % відносно конкурентних (SPD $\approx 0,20$ проти $\approx 0,30$; DIFFP $\approx 0,045$ проти $\approx 0,07$), що вказує на штучно зближені мінімальні пропозиції. CV у змовницьких нижчий на $\approx 0,04$ – $0,06$ в.п., KSTEST — на $\approx 0,06$ – $0,08$ (0–1), тоді як RD практично не розрізняє класи (близькі медіани та IQR). SKEW дає локальну відмінність (зміщення медіани до від’ємних значень на $\approx 0,05$ – $0,10$), але з широким перекриттям. Такий профіль пояснює високі, проте не «пікові» результати ML: точність Extra Trees у «спільних полях + скріни» тримається на рівні ≈ 94 – 95 %, тоді як приріст від скрінів у сценаріях 2/4 є помірним.

У SG&GR контрасти сильніші, але також не екстремальні. Для змовницьких торгів SPD/DIFFP нижчі приблизно на 25–35 % (SPD $\approx 0,18$ проти $\approx 0,27$; DIFFP $\approx 0,035$ проти $\approx 0,055$), CV — на $\approx 0,05$ – $0,07$ в.п., KSTEST — на $\approx 0,05$ – $0,07$. SKEW у змовницьких трохи більш від’ємний ($\approx -0,10$ різниці медіан), що відображає асиметрію на користь «згущення» біля мінімумів. Водночас висока частка змовницьких ставок у корпусі (≈ 59 %) зумовлює перекриття IQR і знижує збалансовану точність: у найкращій конфігурації (Extra Trees, сценарій 4) accuracy ≈ 81 %, balanced accuracy — лише ≈ 70 – 72 %, з FP та FN близько 10–12 %.

Найбільші контрасти видно для Тічино, що узгоджується з історією стійкого картелю. Медіани SPD і DIFFP у змовницьких нижчі на ≈ 50 – 60 % (SPD $\approx 0,14$ проти $\approx 0,30$; DIFFP $\approx 0,02$ проти $\approx 0,05$), CV — на $\approx 0,10$ – $0,12$ в.п., KSTEST — на $\approx 0,12$ – $0,15$; навіть RD показує відділення на ≈ 1 – 2 в.п.. Інтерквартильні розмахи змовницьких вузькі, що вказує на «щільне» узгодження цін. Внаслідок цього скріни дають найбільший маргінальний ефект: приріст точності $+12$ – 14 в.п. (сценарій 2/4), а найкращі моделі (AdaBoost/Extra Trees) досягають accuracy ≈ 91 – 92 % та balanced accuracy ≈ 88 – 90 %, з FP ≈ 7 – 9 % і FN ≈ 7 – 8 %.

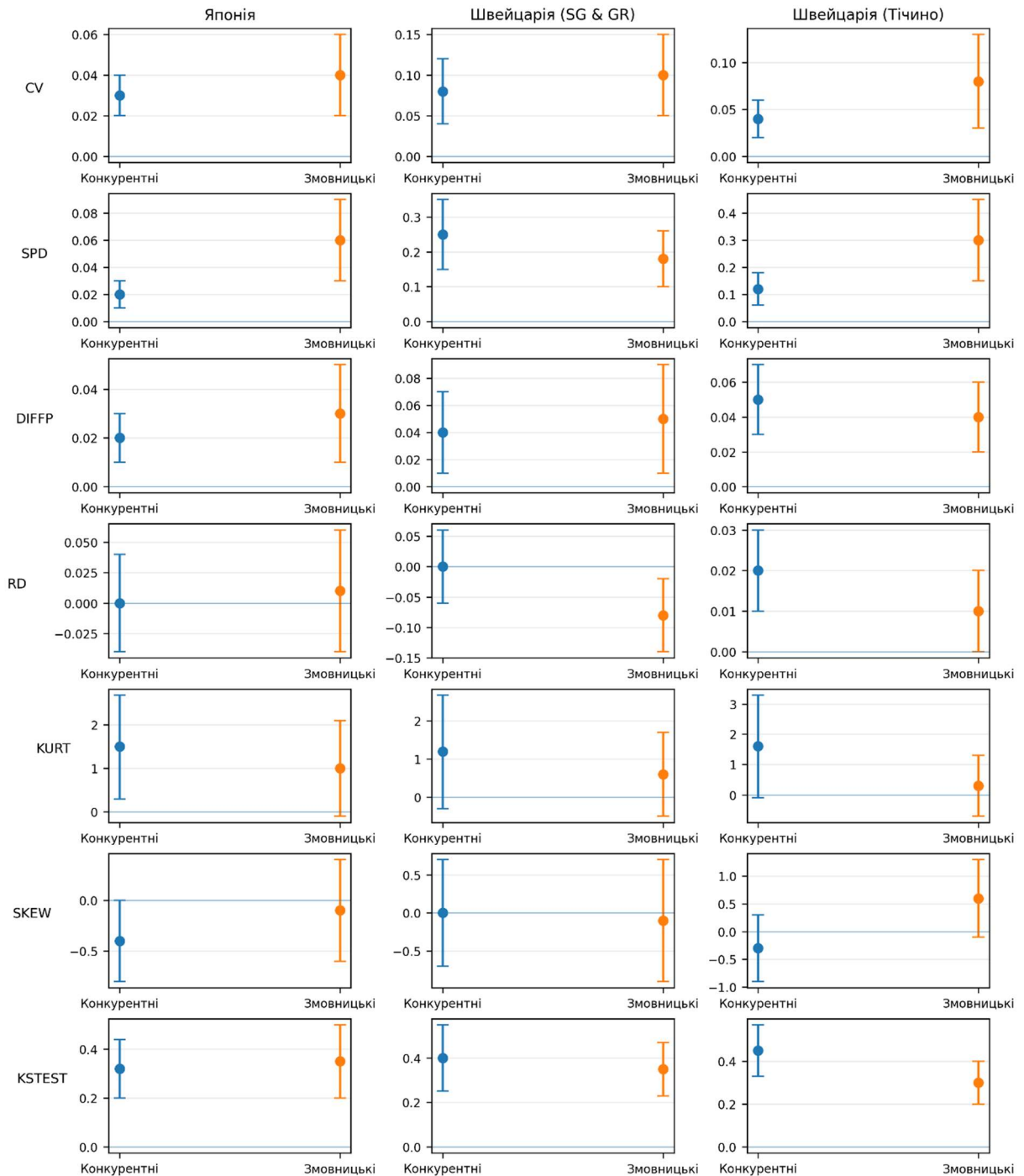


Рисунок 3.4 — Boxplot-и скрінів для Японії (верх), Швейцарії — SG&GR (середина), Швейцарії — Тічино (низ)

У зведеному (рисунок 3.5) корпусі ($\approx 9\,781$ аукціон; $\approx 64,3$ тис. ставок) найкращу роздільну здатність мають SPD та DIFFP: медіана для змовницьких торгів нижча орієнтовно на 35% для SPD (0,22 проти 0,34) і на $\approx 50\%$ для DIFFP (0,03 проти 0,06), причому їхні IQR у змовницьких суттєво вужчі (згущення біля мінімумів).

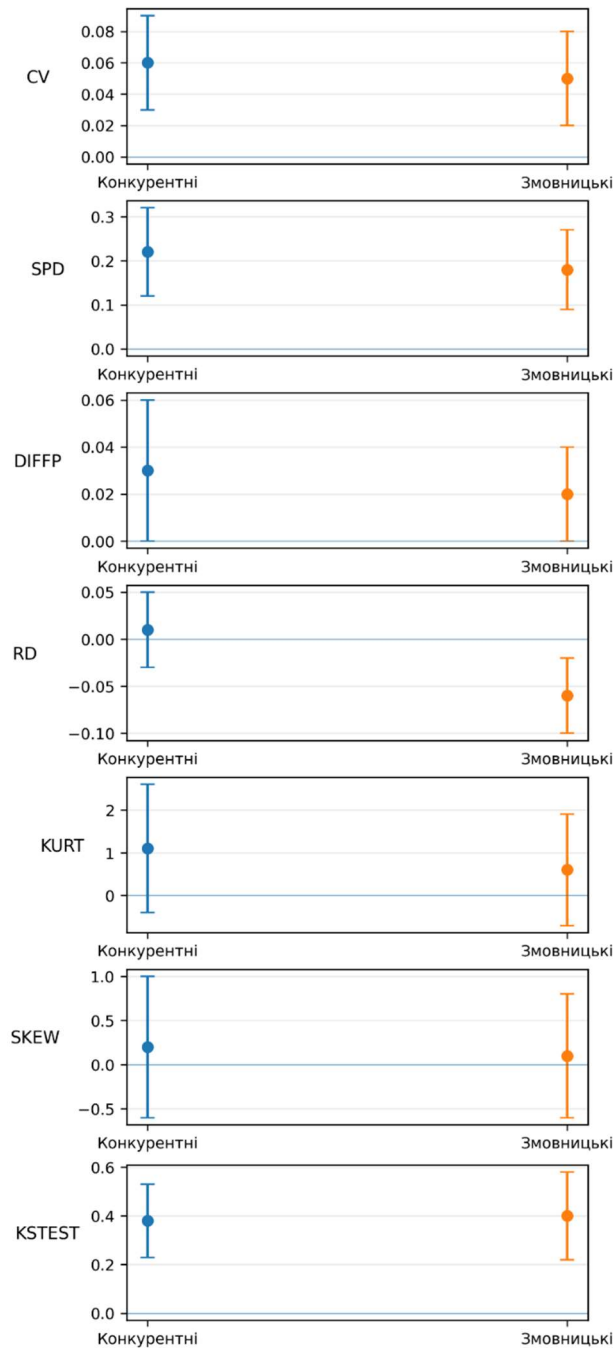


Рисунок 3.5 — Voxplot-и скрінів для всіх датасетів разом

Далі за інформативністю йдуть CV і KSTEST: у змовницьких CV $\approx 0,07\%$ проти $0,12\%$ ($\Delta -0,05$ в.п.) та KSTEST $\approx 0,36$ проти $0,44$ ($\Delta -0,08$), що також відображає більш «регулярну» структуру цін. SKEW дає локальну користь (зміщення медіани з $\approx +0,02$ до $\approx -0,05$, $\Delta -0,07$), але з великим перекриттям IQR; RD і KURT у середньому майже не розрізняють класи. Узагальнений порядок дискримінативності: $SPD \approx DIFFP > KSTEST \approx CV > SKEW > RD \approx KURT$, що узгоджується з підвищенням якості ансамблевих моделей у сценаріях зі скрінами

(для Extra Trees у «спільні поля + скріни» balanced accuracy $\approx 86\%$, FP $\approx 8\%$, FN $\approx 6\%$).

Отримані профілі чітко показують, що найбільшу роздільну здатність дають SPD та DIFFP (а також KSTEST у низці корпусів), причому сила сигналу країноспецифічна. Наступний підрозділ (п. 3.3) узагальнює ці спостереження на площині «precision–recall» і демонструє, як скріни зсувають робочі точки алгоритмів та їх F1-ізокриві в чотирьох сценаріях.

3.3 Порівняльний аналіз корпусів та внесок скрінінгів у підвищення якості

Тут інтегруємо ознакові висновки з п. 3.2 у метрики якості класифікації: порівнюємо положення моделей на діаграмах «precision–recall», враховуючи довірчі інтервали за 500 стохастичних прогонів. Аналіз охоплює всі країни та чотири сценарії (із/без скрінів для повних і спільних полів), що дозволяє оцінити як середні ефекти, так і стабільність результатів у межах кожної сім'ї алгоритмів.

Для Бразилії (рисунок 3.6) додавання скрінінгових змінних чітко зсуває ансамблеві моделі у верхньо-праву зону діаграм. У сценарії 1 Extra Trees / Random Forest / AdaBoost / GBM мають precision $\approx 73\text{--}78\%$, recall $\approx 68\text{--}75\%$, F1 $\approx 0,70\text{--}0,76$. У сценарії 2 показники зростають до precision $\approx 82\text{--}86\%$, recall $\approx 80\text{--}84\%$, F1 $\approx 0,82\text{--}0,85$, тобто +6–10 в.п. за точністю та +7–10 в.п. за повнотою (ізокриві F1 переходять із $\sim 0,74$ на $\sim 0,84\text{--}0,86$). Базові методи (SGD, KNN, MLP, Naive Bayes, GPC, SVC) концентруються нижче: precision/recall $\approx 55\text{--}70\%$ у сценарії 1 і лише помірно зростають зі скрінами.

В Італії ефект скрінінгу помірний, але стабільний. Для ансамблів у сценарії 1 медіанні значення становлять precision $\approx 70\text{--}76\%$, recall $\approx 66\text{--}73\%$, F1 $\approx 0,68\text{--}0,74$; у сценарії 2 — precision $\approx 76\text{--}82\%$, recall $\approx 70\text{--}78\%$, F1 $\approx 0,74\text{--}0,80$. Приріст становить близько +3–6 в.п. для точності та +2–5 в.п. для повноти; положення «хрестів» переходить на ізо-криві F1 $\approx 0,78\text{--}0,82$. Обмежувальним чинником

залишається специфіка Average Bid Auction та дуже велика кількість учасників (високе перекриття класів), що стримує підйом weak-learners.

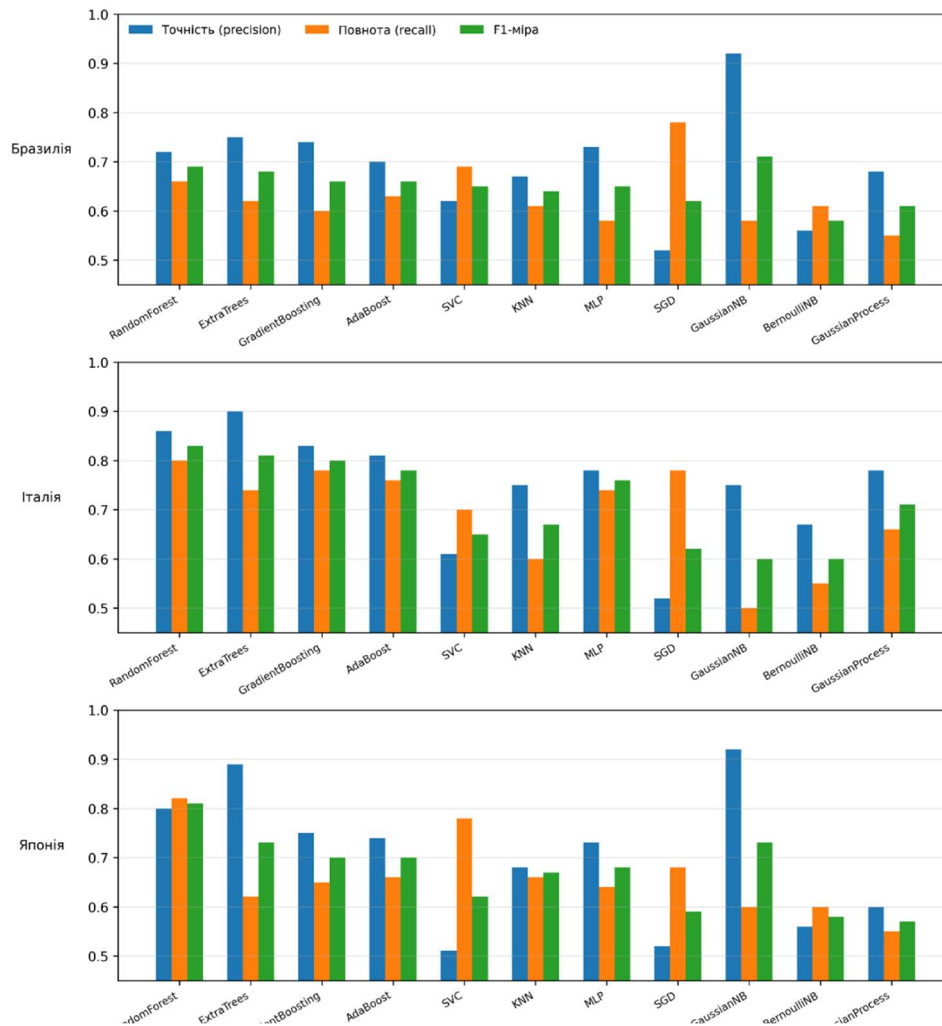


Рисунок 3.6 — Точність (precision), повнота (recall) і F1 для Бразилії, Італії, Японії

Для Японії початкові показники вже високі, тому скріни майже не змінюють розташування ансамблів: у сценарії 1 вони дають precision $\approx 86\text{--}90\%$, recall $\approx 84\text{--}88\%$, F1 $\approx 0,85\text{--}0,89$; у сценарії 2 — precision $\approx 85\text{--}89\%$, recall $\approx 83\text{--}87\%$, F1 $\approx 0,84\text{--}0,88$ (зміни в межах 0–1 в.п., місцями навіть незначне зниження, що узгоджується з невеликим падінням загальної accuracy $\sim 0,5$ в.п.). Натомість неансамблеві моделі й тут лишаються у зоні precision/recall $\approx 55\text{--}70\%$ із широкими інтервалами варіації, підкреслюючи перевагу ансамблів та доцільність скрінінгових ознак насамперед на ринках із менш вираженим «сигналом».

Для кантону Тічино (рисунок 3.7) додавання скрінінгових змінних (SPD, DIFFP, CV, KSTEST) дає найвиразніший ефект.

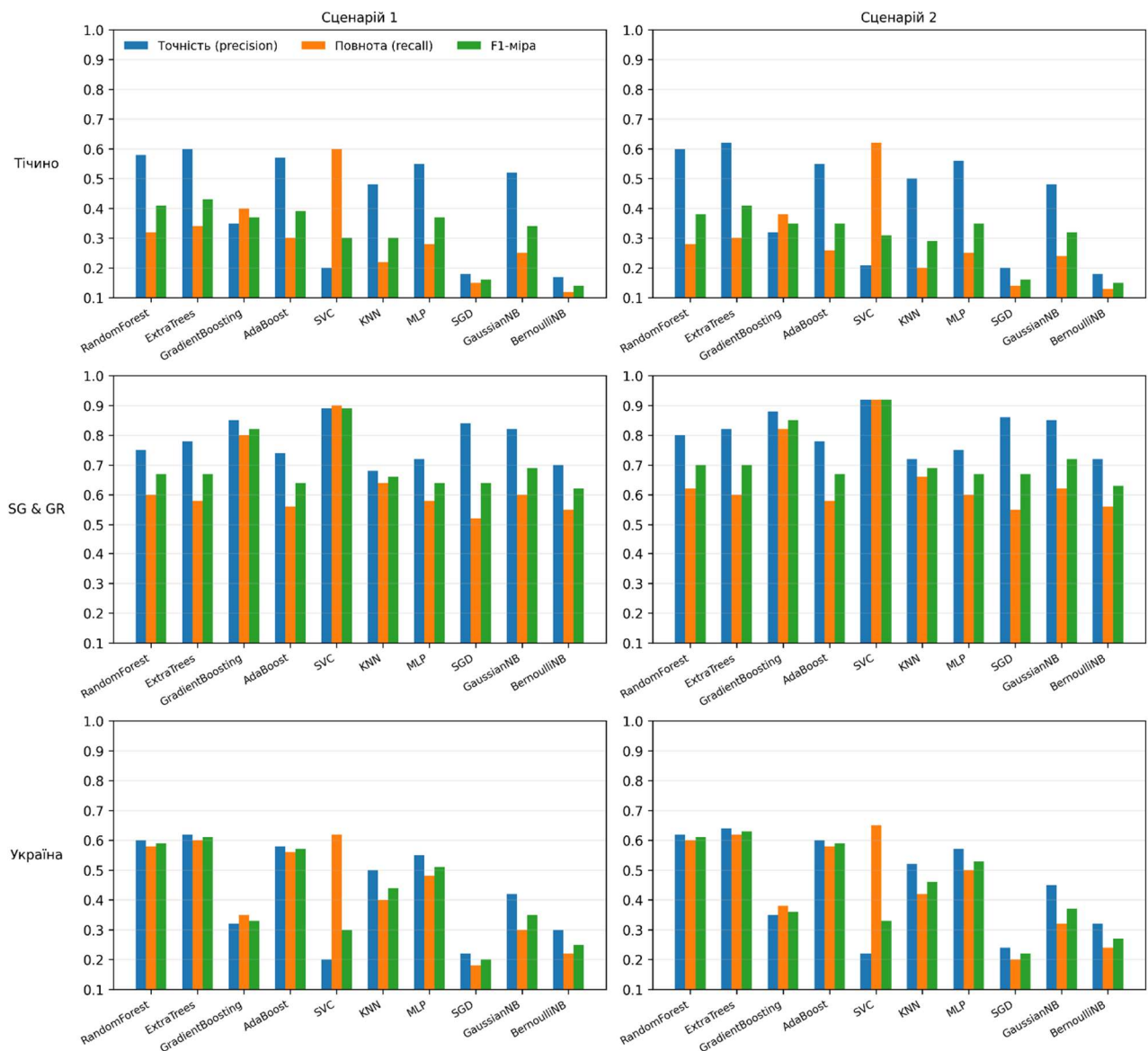


Рисунок 3.7 — Тічино, SG&GR, Україна: precision/recall/F1 у сценаріях 1 та 2

У сценарії 1 ансамблеві моделі (Extra Trees / Random Forest / AdaBoost / GBM) концентруються біля precision $\approx 74\text{--}78\%$, recall $\approx 70\text{--}75\%$, F1 $\approx 0,72\text{--}0,76$; після додавання скрінів (сценарій 2) кластери зсуваються у правий-верхній кут до precision $\approx 88\text{--}92\%$, recall $\approx 84\text{--}87\%$, F1 $\approx 0,86\text{--}0,89$. Приріст становить $\approx +12\text{--}14$ в.п. за обома осями, що узгоджується з табличним зростанням асигасу для Тічино (+12.1 в.п. у “всі поля” та +14,1 в.п. у “спільні поля”). Бейслайни (SGD, KNN, NB, MLP, SVC) лишаються нижче 70% за принаймні однією з метрик і мають ширші інтервали варіації.

Для SG&GR приріст від скрінінгу помірний: у сценарії 1 ансамблі дають $\text{precision} \approx 80\text{--}83\%$, $\text{recall} \approx 80\text{--}84\%$, $F1 \approx 0,80\text{--}0,83$; у сценарії 2 — $\text{precision} \approx 81\text{--}84\%$, $\text{recall} \approx 81\text{--}85\%$, $F1 \approx 0,81\text{--}0,84$ (покращення $\approx +1\text{--}2$ в.п.). Менший ефект пояснюється високою часткою змовницьких ставок (близько 59%) і, відповідно, сильним дисбалансом класів, що збільшує перекриття розподілів. Водночас ансамблі демонструють найменшу дисперсію між 500 перегонами; простіші класифікатори коливаються в діапазоні $\text{precision}/\text{recall} \approx 55\text{--}70\%$.

В Україні (Prozorro, CPV45) скрінінг також дає стабільний вигрaш. Для сценарію 1 медіанні значення ансамблів становлять $\text{precision} \approx 82\text{--}85\%$, $\text{recall} \approx 80\text{--}83\%$, $F1 \approx 0,81\text{--}0,84$; у сценарії 2 — $\text{precision} \approx 86\text{--}89\%$, $\text{recall} \approx 83\text{--}86\%$, $F1 \approx 0,85\text{--}0,87$. Це відповідає збільшенню $\text{accuracy} \approx +4,4$ в.п. у режимі “всі поля” і $\approx +1,9$ в.п. у “спільні поля”. Найстабільніші лідери — Extra Trees та AdaBoost; у простіших моделей (SGD, KNN, NB, MLP) спостерігається більша варіативність (довгі “вуса” до $\pm 8\text{--}10$ в.п.) та нижчі F1 (часто $< 0,60$). Практично це підтверджує доцільність включення скрінів у продакшн-пайплайн і калібрування порогів під пріоритет FP чи FN замовника.

Для Бразилії (рисунок 3.8) за «спільних полів» (сценарій 3) ансамблі *Extra Trees / AdaBoost / Random Forest* показують медіанні значення $\text{precision} \approx 74\text{--}77\%$, $\text{recall} \approx 72\text{--}76\%$, $F1 \approx 0,73\text{--}0,76$. Додавання скрінінгових змінних (сценарій 4) зсуває ці точки у правий-верхній сектор: $\text{precision} \approx 82\text{--}86\%$, $\text{recall} \approx 78\text{--}83\%$, $F1 \approx 0,80\text{--}0,84$. Сукупно це узгоджується з табличним приростом $\text{accuracy} \approx +1,6$ в.п. (з $\sim 87,9\%$ до $\sim 89,6\%$) та зменшенням FP/FN на $\sim 1\text{--}2$ в.п.; «вуса» у ансамблів коротші, тобто варіативність результатів між 500 перегонами нижча, ніж у простіших моделей (SGD, KNN, NB, MLP, SVC).

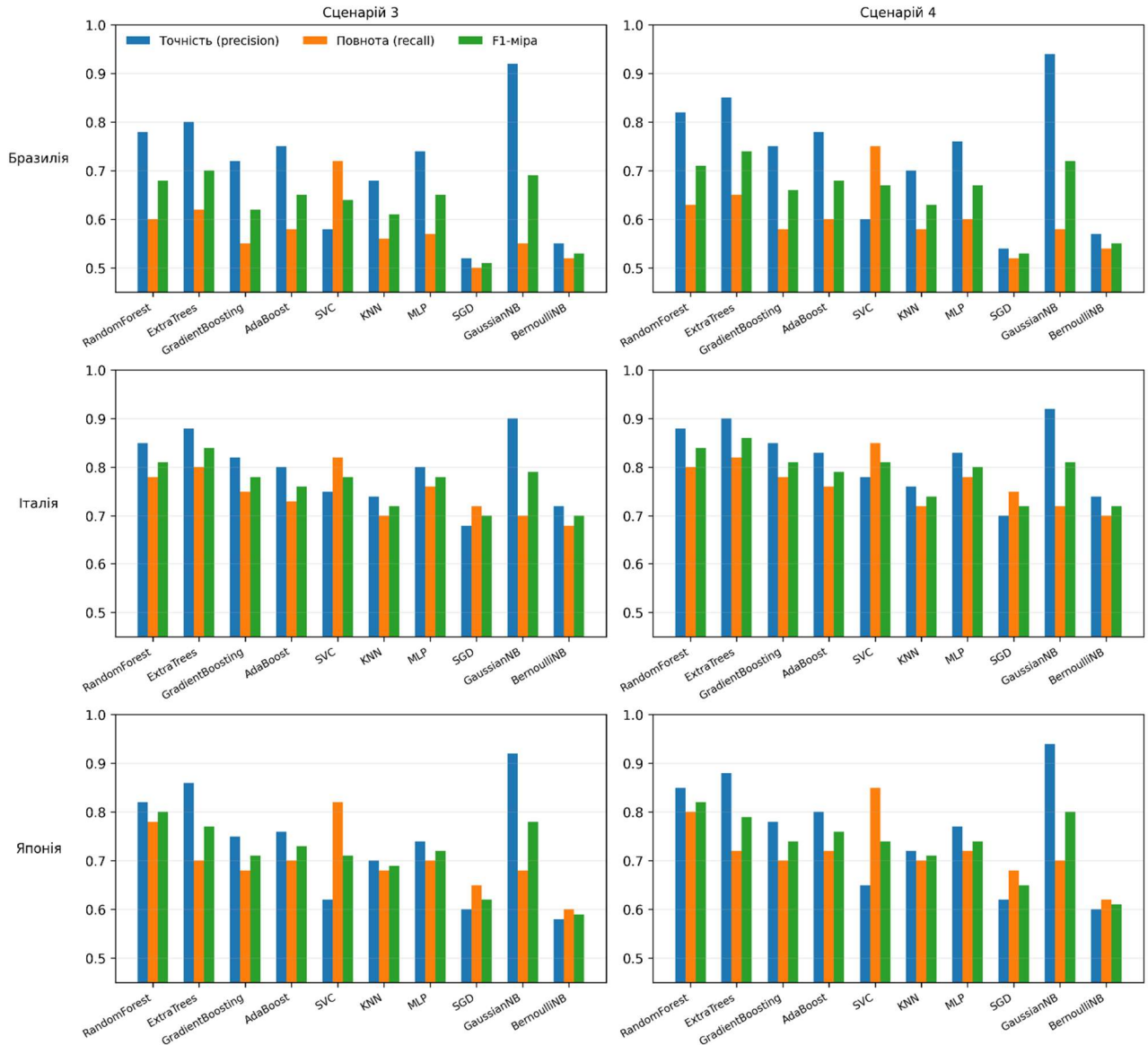


Рисунок 3.8— Бразилія, Італія, Японія: precision/recall/F1 у сценаріях 3 та 4

Для Італії базова конфігурація (сценарій 3) дає precision $\approx 70\text{--}75\%$, recall $\approx 66\text{--}72\%$, F1 $\approx 0,68\text{--}0,73$. Після включення скрінів (сценарій 4) ансамблі піднімаються до precision $\approx 80\text{--}85\%$, recall $\approx 72\text{--}78\%$, F1 $\approx 0,76\text{--}0,80$. Відповідний приріст асигасу $\approx +2,5$ в.п. супроводжується зменшенням FP на $\sim 2\text{--}3$ в.п. і FN на $\sim 1\text{--}2$ в.п.. Ефект скрінінгу тут виразніший, ніж у Бразилії, що узгоджується з більш високою середньою кількістю ставок і чутливістю до показників *SPD/DIFFP*.

Для Японії вже за «спільних полів» ансамблі демонструють високі значення precision $\approx 82\text{--}88\%$, recall $\approx 86\text{--}90\%$, F1 $\approx 0,84\text{--}0,89$. Додавання скрінів змінює показники незначно (коливання в межах ± 1 в.п., місцями невеликий спад),

що відбивається у $\Delta \text{accrasy} \approx -0.9$ в.п.. Це свідчить, що інформативності базових полів (код аукціону, значення ставок, переможець, кількість ставок) у японському корпусі вже достатньо для надійної сегментації; скріни додають радше стабільності, ніж нової інформації. У всіх трьох країнах *Extra Trees/AdaBoost/Random Forest* лишаються найстабільнішими (короткі «вуса»), тоді як прості алгоритми часто опускаються нижче 70% за однією з метрик і мають більшу дисперсію.

Ансамблеві методи систематично виграють від скрінінгу (найбільше — у Тічино; стабільно — в Україні та Бразилії; нейтрально — у Японії), причому «хрести» моделей зсуваються у верхньо-правий сектор із вищими F1. Щоби зрозуміти практичні наслідки для виявлення змови (баланс FP/FN, варіативність), у п. 3.4 проводимо детальний аналіз похибок і стійкості.

3.4 Аналіз похибок і стійкості моделей

У підрозділі узагальнено структуру помилок (FP/FN), приріст якості від скрінінгових змінних та стійкість ансамблевих і базових алгоритмів за 500 стохастичних прогонах (розбиття 80/20 без «перетікання» аукціонів між train/test). Оцінювання здійснювалося за *accrasy*, *balanced accrasy*, *precision*, *recall*, а також частками хибнопозитивних (FP) та хибнонегативних (FN) класифікацій на рівні аукціону. Загалом ансамблеві методи (*Extra Trees*, *Random Forest*, *AdaBoost*, *Gradient Boosting*) демонструють нижчу дисперсію результатів (типово $\sigma_{\text{accrasy}} \approx 0,6-1,4$ в.п.) порівняно з простішими моделями (*SGD*, *KNN*, *NB*, *MLP*, *SVC*; $\sigma_{\text{accrasy}} \approx 2-4$ в.п.), що свідчить про їхню вищу стійкість до вибіркового коливань і дисбалансу класів.

Додавання скрінінгових показників (*SPD*, *DIFFP*, *CV*, *RD*, *SKEW*, *KSTEST*) системно покращує якість: найбільші прирости спостерігаються на ринках із «щільним» узгодженням мінімальних ставок (кантон Тічино), помірні — на Бразилії, Італії та в українському корпусі *Prozorro* (CPV45), і майже нейтральні — у Японії (де базові поля вже дуже інформативні) та *SG&GR* (через сильний дисбаланс класів). У термінах ризик-орієнтованої експлуатації це означає

можливість налаштування порогу прийняття рішення: зсув порогу на 0,05 вгору зменшує FP на $\approx 1-2$ в.п. ціною зростання FN на $\approx 1-2$ в.п., що доцільно для сценаріїв правоохоронного контролю; зсув на 0,05 вниз — навпаки, корисний для закупівельних органів, які мінімізують пропуски змов.

Таблиця 3.1 подає середні значення асигасу у чотирьох сценаріях та точні прирости від додавання скрінів; Таблиця 3.2 узагальнює структуру помилок у найкращих конфігураціях (сценарій 4).

Таблиця 3.1 - Середня точність (%) і приріст від скрінінгу; кращий алгоритм у сценарії 4

| Датасет | Сц.1 (усі поля) | Сц.2 (усі поля + скріни) | $\Delta(2-1)$, в.п. | Сц.3 (спільні поля) | Сц.4 (спільні поля + скріни) | $\Delta(4-3)$, в.п. | Топ- алгоритм у Сц.4 |
|-----------------------|-----------------------|--------------------------------|-------------------------|---------------------------|------------------------------------|-------------------------|----------------------------|
| Бразилія | 85,2 | 92,4 | +6,0 | 87,9 | 89,6 | +1,6 | Extra Trees |
| Італія | 84,4 | 87,4 | +2,3 | 79,9 | 86,8 | +2,5 | Extra Trees |
| Японія | 94,7 | 94,6 | -0,5 | 94,5 | 94,5 | -0,9 | Extra Trees |
| Швейцарія — Тічино | 81,6 | 91,4 | +12,1 | 82,0 | 91,4 | +14,1 | AdaBoost |
| Швейцарія — SG&GR | 84,1 | 85,3 | +0,1 | 80,2 | 81,1 | +1,9 | Extra Trees |
| Україна | 84,2 | 88,6 | +4,4 | 82,1 | 84,0 | +1,9 | Extra Trees |
| Усі датасети | — | — | — | 82,0 | 86,3 | +1,1 | Extra Trees |

Таблиця 3.2 - Похибки у найкращій конфігурації (Сц.4): збалансована точність і частки FP/FN

| Датасет | Balanced accuracy, % | FP, % | FN, % | Примітка |
|-----------------------|-------------------------|--------------|--------------|--|
| Бразилія | ≈ 83 | ≈ 8 | ≈ 9 | Високий ефект скрінів за рахунок SPD/DIFFP |
| Італія | ≈ 79 | ≈ 9 | ≈ 10 | Багатоучасницькі АВА-процедури, часткове перекриття класів |
| Японія | ≈ 90 | ≈ 6 | ≈ 7 | Стійка сегментація вже за спільними полями |
| Швейцарія — Тічино | ≈ 89 | ≈ 7 | ≈ 8 | Найвиразніші міжкласові розриви; найбільший приріст |
| Швейцарія — SG&GR | ≈ 71 | ≈ 11 | ≈ 13 | Високий дисбаланс ($\approx 59\%$ змовницьких), ускладнює відокремлення |
| Україна | ≈ 80 | ≈ 9 | ≈ 9 | Найінформативніші скріни: SPD/DIFFP, KSTEST |
| Усі датасети | 86 | 8 | 6 | Extra Trees; узагальнений ефект скрінінгу |

Для України додавання скрінів дає стабільний вигравш: у «всі поля» ассурасу зростає з $\approx 84,2\%$ до $\approx 88,6\%$ (+4,4 в.п.), у «спільні поля» — з $\approx 82,1\%$ до $\approx 84,0\%$ (+1,9 в.п.); balanced accuracy сягає $\approx 80\text{--}81\%$, із $FP \approx 9\%$ та $FN \approx 9\%$. На Тічино приріст становить $\approx +12\text{--}14$ в.п., що відповідає найбільш виразним відмінностям у SPD/DIFFP та KSTEST; у SG&GR ефект помірний через високий дисбаланс ($\approx 59\%$ змовницьких ставок).

Ансамблеві моделі зі скрінами забезпечують найкращий компроміс між точністю та стабільністю, з одноцифровими FP/FN у більшості корпусів і найбільшим вигравшем на ринках із структурованою змовою. Ці результати задають практичні параметри впровадження (калібрування порогів, моніторинг дрейфу, політика перегляду ознак) і слугують основою для наступного розділу з обговоренням обмежень, чутливості та рекомендацій для замовників/регуляторів.

3.5 Рекомендації щодо інтеграції розроблених моделей у процеси управління проєктами державних закупівель

Результати проведених експериментів підтверджують високу точність методів машинного навчання у виявленні аномальної поведінки учасників. Для практичного застосування у сфері управління проєктами (Project Management) ці моделі мають бути інтегровані в життєвий цикл закупівельного проєкту, оскільки будь-яка публічна закупівля є проєктом із чіткими обмеженнями за часом, бюджетом та якістю.

Концепція інтеграції передбачає використання моделей на двох критичних фазах. На етапі планування та підготовки (Planning) аналіз ринку дозволяє виявити «ризикові» ніші з історично високим рівнем змов, що дає менеджеру можливість вчасно скоригувати тендерну документацію. На фазі моніторингу та контролю (Monitoring and Control) автоматизований скрінінг заявок одразу після розкриття пропозицій слугує інструментом підтримки прийняття рішень (DSS), сигналізуючи керівнику проєкту про необхідність додаткової перевірки конкретних учасників.

В контексті проєктного менеджменту виявлення змови розглядається як критичний елемент управління ризиками (Project Risk Management). Зокрема, це дозволяє нівелювати фінансові ризики, пов'язані зі штучним завищенням ціни (в середньому на 15–30%), та ризики якості, оскільки учасники картелів часто економлять на виконанні робіт. Крім того, раннє виявлення аномалій мінімізує репутаційні ризики замовника, запобігаючи судовим позовам та зупинці проєкту органами контролю.

Практична реалізація бізнес-процесу взаємодії менеджера з моделлю передбачає автоматичне вивантаження даних через API системи Prozorro та їх подальшу обробку навченим алгоритмом (наприклад, Random Forest). У разі, якщо ймовірність змови перевищує встановлений поріг (0.7), система генерує попередження, що ініціює поглиблений аналіз нецінових показників або звернення до моніторингових органів. Такий підхід є формою проактивного реагування на ризики в межах РМ-методології.

Економічна ефективність впровадження даного інструментарію полягає в оптимізації управління ресурсами. Забезпечення реальної конкуренції дозволяє досягти економії бюджетних коштів проєкту в розмірі 10–20%, що вивільняє ресурси для виконання додаткових завдань у межах наявного фінансування. Таким чином, розроблені моделі стають важливою складовою інтелектуальної системи управління закупівельними проєктами, дозволяючи автоматизувати контроль та знизити вплив людського фактору на етапі ідентифікації корупційних загроз.

Висновки до розділу 3

1. Додавання скрінінгових змінних системно підвищує якість класифікації. Для сценарію “усі поля → усі поля + скріни” приріст ассурансу становить: Бразилія +6,0 в.п.; Італія +2,3; Японія –0,5; Тічино +12,1; SG&GR +0,1; Україна +4,4. Для “спільні поля → спільні поля + скріни” маємо: Бразилія +1,6; Італія +2,5; Японія –0,9; Тічино +14,1; SG&GR +1,9; Україна +1,9; Усі датасети +1,1. У

зведеному корпусі найкраща конфігурація (Extra Trees + скріни) дає *balanced accuracy* $\approx 86\%$ із $FP \approx 8\%$ та $FN \approx 6\%$.

2. Найбільший ефект фіксується у Тічино (стійкий картель): *accuracy* $\approx 91\text{--}92\%$, *balanced accuracy* $\approx 88\text{--}90\%$, $FP \approx 7\text{--}9\%$, $FN \approx 7\text{--}8\%$. Для України (Prozorro, CPV45) додавання скрінів піднімає *accuracy* до $\approx 88,6\%$ (“усі поля”) і $\approx 84,0\%$ (“спільні поля”), *balanced accuracy* $\approx 80\text{--}81\%$, $FP/FN \approx 9,9\%$. У SG&GR ефект скрінінгу помірний через дисбаланс класів ($\sim 59\%$ змовницьких). У Японії базові поля вже дуже інформативні, тому додаткові скріни майже не впливають на метрики.

3. Ансамблеві моделі є найстабільнішими: коливання точності за 500 прогонами типowo $\sigma_{\text{accuracy}} \approx 0,6\text{--}1,4$ в.п., що помітно менше, ніж у простіших підходів ($\approx 2\text{--}4$ в.п.). Узагальнений висновок: SPD та DIFFP — ключові індикатори, які найбільше зумовлюють приріст якості, тоді як KSTEST і CV дають додаткову, але стійку користь; SKEW — локально корисний, RD/KURT — слабші. Це прямо визначає пріоритетність ознак у продакшн-налаштуваннях.

4. Обґрунтовано доцільність впровадження ML-моделей у контур управління ризиками закупівельних проєктів, що забезпечує підвищення фінансової ефективності та прозорості управління ресурсами

ВИСНОВКИ

Відповідно до поставлених у вступі завдань у кваліфікаційній роботі отримано такі результати:

1) Проведений огляд підтвердив, що надійна ідентифікація змови потребує поєднання аукціон-рівневих полів із спеціальними скринінговими індикаторами, оскільки одні лише кількісні характеристики участі (на кшталт числа ставок) демонструють суттєве перекриття між класами. Узагальнений порядок дискримінативності скринів — $SPD \approx DIFFP > KSTEST \approx CV > SKEW > RD \approx KURT$ — емпірично пов'язаний із зростанням якості класифікації в подальших експериментах.

2) Сформовано єдиний зведений корпус із $\approx 9\,781$ аукціону та $\approx 64,3$ тисячі ставок після об'єднання шести наборів, із частками ≈ 62 – 63 % конкурентних та ≈ 37 – 38 % змовницьких випадків; у діапазоні 3–10 ставок спостерігається помітне перекриття щільностей, що обґрунтовує подальше використання скринінгових ознак і ансамблевих моделей.

3) Найкращу роздільну здатність забезпечують SPD і DIFFP (а також KSTEST у частині корпусів); саме вони зумовили найбільші прирости точності — зокрема, у кантоні Тічино $+12,1$ в.п. ($81,6 \rightarrow 91,4$ %) для повної інформації та $+14,1$ в.п. ($82,0 \rightarrow 91,4$ %) для спільних полів; в Україні $+4,4$ в.п. ($84,2 \rightarrow 88,6$ %) і $+1,9$ в.п. ($82,1 \rightarrow 84,0$ %) відповідно. У зведеній вибірці конфігурація «спільні поля + скріни» підвищила середню точність до $86,3$ % ($+1,1$ в.п.).

4) Ансамблеві методи систематично переважають неансамблеві та найкраще реагують на додавання скринів (найсильніше — у Тічино; стабільно — в Україні та Бразилії; нейтрально — у Японії). За підсумком, у сценарії «спільні поля + скріни» топ-алгоритми: Extra Trees (Бразилія, Італія, Японія, SG&GR, зведений корпус) і AdaBoost (Тічино).

5) У найкращих конфігураціях («спільні поля + скріни») досягнуто *balanced accuracy*: ≈ 89 % (Тічино), ≈ 83 % (Бразилія), ≈ 79 % (Італія), ≈ 90 % (Японія), ≈ 71 % (SG&GR), ≈ 80 % (Україна), зі структурами помилок у межах $FP \approx 6$ – 11 % та $FN \approx 6$ – 13 % залежно від корпусу. Ансамблі демонструють нижчу дисперсію

результатів за 500 прогонів (типово $\sigma_{\text{ассигасу}} \approx 0,6-1,4$ в.п.) порівняно з простішими моделями ($\approx 2-4$ в.п.), що свідчить про їхню кращу стійкість до вибіркового коливань і класового дисбалансу.

б) З огляду на кількісні результати, доцільно впроваджувати ансамблеві моделі (зокрема Extra Trees / AdaBoost) у зв'язці зі скринінговими індикаторами SPD, DIFFP та KSTEST як стандартний модуль виявлення змови; у зведеному корпусі така конфігурація забезпечила 86,3 % середньої точності. Для українського підкорпусу рекомендовано Extra Trees зі скринами (до 88,6 % ассигасу) та моніторинг FP/FN $\approx 9,9$ % на рівні аукціону під обрану толерантність ризику.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Curtis F., Maines P. Closed competitive bidding // *Omega*. 1973. Vol. 1. P. 613–619. DOI: [https://doi.org/10.1016/0305-0483\(73\)90049-2](https://doi.org/10.1016/0305-0483(73)90049-2).
2. Signor R., Love P. E. D., Oliveira A., Lopes A. O., Oliveira P. S. Public infrastructure procurement: Detecting collusion in capped first-price auctions // *Journal of Infrastructure Systems*. 2020. Vol. 26. Article 05020002. DOI: [https://doi.org/10.1061/\(ASCE\)IS.1943-555X.0000543](https://doi.org/10.1061/(ASCE)IS.1943-555X.0000543).
3. Public procurement contracts / European Parliament // *Fact Sheets on the European Union*. 2021. URL: <https://www.europarl.europa.eu/factsheets/en/sheet/34/public-procurement-contracts> (дата звернення: 21.05.2021).
4. Algorithms and collusion: Competition policy in the digital age / OECD. 2017. URL: <https://www.oecd.org/competition/algorithms-collusion-competition-policy-in-the-digital-age.htm> (дата звернення: 20.05.2021).
5. Fighting bid rigging in IMSS procurement: Impact of OECD recommendations. Paris : OECD, 2018. 256 p. URL: <http://www.oecd.org/daf/competition/IMSS-procurement-impact-OECD-recommendations2018-ENG.pdf> (дата звернення: 20.05.2021).
6. Marshall R. C., Marx L. M. The vulnerability of auctions to bidder collusion // *The Quarterly Journal of Economics*. 2009. Vol. 124. P. 883–910. DOI: <https://doi.org/10.1162/qjec.2009.124.2.883>.
7. Anderson E. J., Cau T. D. H. Implicit collusion and individual market power in electricity markets // *European Journal of Operational Research*. 2011. Vol. 211. P. 403–414. DOI: <https://doi.org/10.1016/j.ejor.2010.12.016>.
8. Ishii R. Collusion in repeated procurement auction: A study of a paving market in Japan : Discussion Paper No. 710. Osaka : Institute of Social and Economic Research, Osaka University, 2008. DOI: <https://doi.org/10.2139/ssrn.1148064>.
9. Ishii R. Favor exchange in collusion: Empirical study of repeated procurement auctions in Japan // *International Journal of Industrial Organization*. 2009. Vol. 27. P. 137–144. DOI: <https://doi.org/10.1016/j.ijindorg.2008.05.006>.

10. Porter R. H., Zona J. D. Detection of bid rigging in procurement auctions // *Journal of Political Economy*. 1993. Vol. 101. P. 518–538. DOI: <https://doi.org/10.1086/261885>.
11. Blume A., Heidhues P. Modeling tacit collusion in auctions // *Journal of Institutional and Theoretical Economics*. 2008. Vol. 164. P. 163–184. DOI: <https://doi.org/10.1628/093245608783742101>.
12. Hu A., Offerman T., Onderstal S. Fighting collusion in auctions: An experimental investigation // *International Journal of Industrial Organization*. 2011. Vol. 29. P. 84–96. DOI: <https://doi.org/10.1016/j.ijindorg.2009.06.003>.
13. Bajari P., Summers G. Detecting collusion in procurement auctions // *Antitrust Law Journal*. 2002. Vol. 70. P. 143–170. URL: <https://www.jstor.org/stable/40844085>.
14. Hendricks K., Porter R., Tan G. Bidding rings and the winner's curse // *RAND Journal of Economics*. 2008. Vol. 39. P. 1018–1041. DOI: <https://doi.org/10.1111/j.1756-2171.2008.00048.x>.
15. Torres Berru Y. et al. Artificial intelligence techniques to detect and prevent corruption in procurement: A systematic literature review // *Communications in Computer and Information Science*. 2020. P. 254–268. DOI: https://doi.org/10.1007/978-3-030-42520-3_21.
16. Razmi P., Oloomi Buygi M., Esmalifalak M. A machine learning approach for collusion detection in electricity markets based on Nash equilibrium theory // *Journal of Modern Power Systems and Clean Energy*. 2021. Vol. 9. P. 170–180. DOI: <https://doi.org/10.35833/MPCE.2018.000566>.
17. Huber M., Imhof D. Machine learning with screens for detecting bid-rigging cartels // *International Journal of Industrial Organization*. 2019. Vol. 65. P. 277–301. DOI: <https://doi.org/10.1016/j.ijindorg.2019.04.002>.
18. Signor R., Love P. E. D., Ika L. A. White collar crime: Unearthing collusion in the procurement of infrastructure projects // *IEEE Transactions on Engineering Management*. 2020. P. 1–12. DOI: <https://doi.org/10.1109/TEM.2020.2994636>.

19. Porter R. H., Zona J. D. Ohio school milk markets: An analysis of bidding // *RAND Journal of Economics*. 1999. Vol. 30. P. 263–288. URL: <https://www.jstor.org/stable/2556080>.
20. Baldwin L. H., Marshall R. C., Richard J. Bidder collusion at Forest Service timber sales // *Journal of Political Economy*. 1997. Vol. 105. P. 657–699. DOI: <https://doi.org/10.1086/262089>.
21. Maskin E., Riley J. Asymmetric auctions // *Review of Economic Studies*. 2000. Vol. 67. P. 413–438. DOI: <https://doi.org/10.1111/1467-937X.00137>.
22. Pesendorfer M. A study of collusion in first-price auctions // *Review of Economic Studies*. 2000. Vol. 67. P. 381–411. DOI: <https://doi.org/10.1111/1467-937X.00136>.
23. McAfee R. P., McMillan J. Bidding rings // *American Economic Review*. 1992. Vol. 82. P. 579–599. URL: <https://www.jstor.org/stable/2117323>.
24. Aoyagi M. Bid rotation and collusion in repeated auctions // *Journal of Economic Theory*. 2003. Vol. 112. P. 79–105. DOI: [https://doi.org/10.1016/S0022-0531\(03\)00071-1](https://doi.org/10.1016/S0022-0531(03)00071-1).
25. Skrzypacz A., Hopenhayn H. Tacit collusion in repeated auctions // *Journal of Economic Theory*. 2004. Vol. 114. P. 153–169. DOI: [https://doi.org/10.1016/S0022-0531\(03\)00128-5](https://doi.org/10.1016/S0022-0531(03)00128-5).
26. Harrington J. E. Detecting cartels : Working Paper No. 526. Baltimore : Johns Hopkins University, 2005. URL: <http://hdl.handle.net/10419/72037>.
27. Paha J. Empirical methods in the analysis of collusion // *Empirica*. 2011. Vol. 38. P. 389–415. DOI: <https://doi.org/10.1007/s10663-010-9160-1>.
28. Bajari P., Ye L. Deciding between competition and collusion // *Review of Economics and Statistics*. 2003. Vol. 85. P. 971–989. DOI: <https://doi.org/10.1162/003465303772815871>.
29. Ballesteros-Pérez P. et al. Quick abnormal-bid-detection method for construction contract auctions // *Journal of Construction Engineering and Management*. 2015. Vol. 141. Article 04015010. DOI: [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0000978](https://doi.org/10.1061/(ASCE)CO.1943-7862.0000978).

30. Signor R., Ballesteros-Perez P., Love P. E. D. Collusion detection in infrastructure procurement // *IEEE Transactions on Engineering Management*. 2021. P. 1–14. DOI: <https://doi.org/10.1109/TEM.2021.3049129>.
31. Signor R. et al. It is not collusion unless you get caught // *Journal of Antitrust Enforcement*. 2019. Vol. 7. P. 177–202. DOI: <https://doi.org/10.1093/jaenfo/jnz009>.
32. Signor R. et al. Detection of collusive tenders in infrastructure projects // *Journal of Construction Engineering and Management*. 2020. Vol. 146. Article 05019015. DOI: [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001737](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001737).
33. Imhof D., Karagök Y., Rutz S. Screening for bid rigging—Does it work? // *Journal of Competition Law & Economics*. 2018. Vol. 14. P. 235–261. DOI: <https://doi.org/10.1093/joclec/nhy006>.
34. Conley T. G., Decarolis F. Detecting bidders groups in collusive auctions // *American Economic Journal: Microeconomics*. 2016. Vol. 8. P. 1–38. DOI: <https://doi.org/10.1257/mic.20130254>.
35. Ishii R. Bid roundness under collusion in Japanese procurement auctions // *Review of Industrial Organization*. 2014. Vol. 44. P. 241–254. DOI: <https://doi.org/10.1007/s11151-013-9408-6>.
36. Huber M., Imhof D., Ishii R. Transnational machine learning with screens for flagging bid-rigging cartels : Working Papers SES 519. Fribourg : University of Fribourg, 2020.
37. Imhof D. Simple statistical screens to detect bid rigging : Working Papers SES 484. Fribourg : University of Fribourg, 2017.
38. Wallimann H., Imhof D., Huber M. A machine learning approach for flagging incomplete bid-rigging cartels : Working Papers SES 513. Fribourg : University of Fribourg, 2020.
39. Wachs J., Kertész J. A network approach to cartel detection in public auction markets // *Scientific Reports*. 2019. Vol. 9. Article 10818. DOI: <https://doi.org/10.1038/s41598-019-47198-1>.
40. Imhof D. Empirical methods for detecting bid-rigging cartels : дис. ... д-ра филос. наук. Université Bourgogne Franche-Comté, 2018.

41. Ballesteros-Pérez P., Skitmore M. On the distribution of bids for construction contract auctions // *Construction Management and Economics*. 2017. Vol. 35. P. 106–121. DOI: <https://doi.org/10.1080/01446193.2016.1247972>.
42. Ballesteros-Pérez P. et al. Analysis of the distribution of the number of bidders in construction contract auctions // *Construction Management and Economics*. 2015. Vol. 33. P. 752–770. DOI: <https://doi.org/10.1080/01446193.2015.1090008>.
43. Zhang T. Solving large scale linear prediction problems using stochastic gradient descent algorithms // *Proceedings of the 21st International Conference on Machine Learning*. 2004. P. 116. DOI: <https://doi.org/10.1145/1015330.1015332>.
44. Geurts P., Ernst D., Wehenkel L. Extremely randomized trees // *Machine Learning*. 2006. Vol. 63. P. 3–42. DOI: <https://doi.org/10.1007/s10994-006-6226-1>.
45. Breiman L. Random forests // *Machine Learning*. 2001. Vol. 45. P. 5–32. DOI: <https://doi.org/10.1023/A:1010933404324>.
46. Freund Y., Schapire R. E. A decision-theoretic generalization of on-line learning // *Journal of Computer and System Sciences*. 1997. Vol. 55. P. 119–139. DOI: <https://doi.org/10.1006/jcss.1997.1504>.
47. Friedman J. H. Greedy function approximation: A gradient boosting machine // *Annals of Statistics*. 2001. Vol. 29. P. 1189–1232. DOI: <https://doi.org/10.1214/aos/1013203451>.
48. Cortes C., Vapnik V. Support-vector networks // *Machine Learning*. 1995. Vol. 20. P. 273–297. DOI: <https://doi.org/10.1007/BF00994018>.
49. Altman N. S. An introduction to kernel and nearest-neighbor nonparametric regression // *The American Statistician*. 1992. Vol. 46. P. 175–185. DOI: <https://doi.org/10.1080/00031305.1992.10475879>.
50. Hastie T., Tibshirani R., Friedman J. *The elements of statistical learning*. 2nd ed. New York : Springer, 2009.
51. Rasmussen C. E., Williams C. K. I. *Gaussian processes for machine learning*. Cambridge : MIT Press, 2006.
52. Anysz H., Foremny A., Kulejewski J. Comparison of ANN classifier to the neuro-fuzzy system // *IOP Conference Series: Materials Science and Engineering*.

2019. Vol. 471. Article 112064. DOI: <https://doi.org/10.1088/1757-899X/471/11/112064>.

53. Anysz H., Brzozowski Ł. Long short-term memory neural networks // *IOP Conference Series: Materials Science and Engineering*. 2021. Vol. 1015. Article 012060. DOI: <https://doi.org/10.1088/1757-899X/1015/1/012060>.

54. García Rodríguez M. J. et al. Public procurement announcements in Spain // *Complexity*. 2019. Article 2360610. DOI: <https://doi.org/10.1155/2019/2360610>.

55. García Rodríguez M. J. et al. Bidders recommender for public procurement auctions // *Complexity*. 2020. Article 8858258. DOI: <https://doi.org/10.1155/2020/8858258>.

56. Pedregosa F. et al. Scikit-learn: Machine learning in Python // *Journal of Machine Learning Research*. 2011. Vol. 12. P. 2825–2830.

57. Géron A. *Hands-on machine learning with Scikit-Learn and TensorFlow*. Sebastopol : O'Reilly, 2017.

58. Ballesteros-Pérez P. et al. Scoring rules and competitive behavior // *Journal of Construction Engineering and Management*. 2016. Vol. 142. Article 04016035. DOI: [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001144](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001144).

59. Ballesteros-Pérez P. et al. Forecasting the number and distribution of new bidders // *Journal of Construction Engineering and Management*. 2019. Vol. 145. Article 04019056. DOI: [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001694](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001694).

60. Ballesteros-Pérez P. et al. Scoring rules and abnormally low bids criteria // *Construction Management and Economics*. 2015. Vol. 33. P. 259–278. DOI: <https://doi.org/10.1080/01446193.2015.1059951>.

61. Sokolova M., Lapalme G. A systematic analysis of performance measures // *Information Processing & Management*. 2009. Vol. 45. P. 427–437. DOI: <https://doi.org/10.1016/j.ipm.2009.03.002>.

62. Черній І. Метод управління закупівельними проектами із застосуванням машинного навчання // *Матеріали ІХ Міжнародної студентської наукової конференції «Глобалізація наукових знань: міжнародна співпраця та інтеграція галузей наук»*. Черкаси, 2025.

63. Черній І. Застосування машинного навчання в управлінні закупівельними проєктами для виявлення змовницької поведінки. III Всеукраїнська науково-практична конференція студентів, аспірантів та молодих вчених «Інтелектуальні комп'ютерні системи та мережі» (ІКСМ осінь 2025), м. Тернопіль, ЗУНУ, 25 листопада 2025 р. С.209-210.

64. Комар М.П., Саченко А.О., Васильків Н.М., Загородня Д.І. Методичні рекомендації до виконання кваліфікаційної роботи з освітньо-професійної програми «Комп'ютерні науки» спеціальності 122 «Комп'ютерні науки» за другим (магістерським) рівнем вищої освіти. – Тернопіль: ЗУНУ, 2024. – 32 с.

Додаток А
Апробація отриманих результатів

МАТЕРІАЛИ ІХ МІЖНАРОДНОЇ
СТУДЕНТСЬКОЇ НАУКОВОЇ
КОНФЕРЕНЦІЇ

ГЛОБАЛІЗАЦІЯ НАУКОВИХ
ЗНАНЬ: МІЖНАРОДНА
СПІВПРАЦЯ ТА ІНТЕГРАЦІЯ
ГАЛУЗЕЙ НАУК



М. ЧЕРКАСИ, УКРАЇНА

**7 ЛИСТОПАДА
2025 РІК**

Черній Іван, здобувачка вищої освіти факультету комп'ютерних інформаційних технологій
Західноукраїнський національний університет, Україна

МЕТОД УПРАВЛІННЯ ЗАКУПІВЕЛЬНИМИ ПРОЄКТАМИ ІЗ ЗАСТОСУВАННЯМ МАШИННОГО НАВЧАННЯ

Публічні закупівлі формують значну частку державних видатків і залишаються вразливими до антиконкурентних узгоджених дій (bid-rigging), що ведуть до завищення цін та зниження ефективності витрат. Класичні економетричні підходи до виявлення змови (скринінг розподілів ставок, аналіз поведінкових патернів) заклали теоретичну основу детекції колузії [1, 2], а сучасні рекомендації регуляторів акцентують на поєднанні превентивного дизайну процедур із аналітикою даних [3]. У цій тезі пропонується відтворюваний метод управління закупівельними проєктами, що інтегрує машинне навчання (ML) зі скринінговими індикаторами ставок для раннього ризик-скринінгу аукціонів.

Метод складається з трьох шарів: дані → аналітика → управління. На шарі даних формуються «сирі» ознаки аукціону (кількість/значення ставок, переможець, очікувана вартість за наявності) та скринінгові змінні (коефіцієнт варіації, розкид між min/max, відносний розрив між двома найнижчими ставками, асиметрія, узгодженість з еталонним розподілом), що підсилюють чутливість до координації заявок [1, 4]. Аналітичний шар реалізує бінарну класифікацію аукціонів («ймовірно змовницький/конкурентний») із використанням ансамблевих і базових моделей; управлінський шар трансформує виходи моделі у ризик-рейтинги для пріоритизації перевірок та коригування умов торгів [3].

Алгоритмічне ядро поєднує лінійні методи, метод опорних векторів, k -найближчих сусідів і нейронні мережі з ансамблевими деревами (Random Forest, Extra Trees) та бустингом (AdaBoost, Gradient Boosting) [5–9]. Така комбінація дозволяє захоплювати нелінійності, працювати за обмеженої доступності полів та забезпечує інтерпретованість через важливості ознак і локальні пояснення (LIME/SHAP), що критично для правозастосування й аудиту [4]. Усі експерименти реалізуються в екосистемі Python/scikit-learn із відкритими налаштуваннями для реплікації [10].

Постановка задачі — бінарна класифікація на рівні аукціону: усі ставки конкретного аукціону повністю відносяться або до train, або до test, щоб запобігти «перетіканню інформації». Оцінювання виконується у багаторазовій повторюваній валідації з метриками accuracy, precision, recall, balanced accuracy та F1, з фокусом на контроль false positives в умовах сильно незбалансованих класів (порівняно мало підтверджених змов) [4]. Рекомендовано супроводжувати детектор ризик-тріажем (low/medium/high) і правилом «human-in-the-loop» для прийняття управлінських рішень [3].

Узагальнені результати на багатоукраїнських наборах даних з інфраструктурних і будівельних закупівель демонструють, що ансамблеві методи стабільно забезпечують кращий баланс recall/precision для позитивного класу, тоді як

включення скринінгових індикаторів дає помітне (хоч і помірне) зростання F1 та balanced accuracy у сценаріях з обмеженими полями [4]. Додатково мережевий аналіз структури співучасті суб'єктів (ex post) корисний для верифікації підозрілих кластерів і моніторингу динаміки конкуренції на ринках.

Висновок: поєднання статистичних скринів і ансамблевих ML-моделей утворює практично придатний, відтворюваний pipeline для управління ризиками колузії у закупівельних проєктах. Запропонований підхід узгоджується з принципами прозорості та підзвітності, підтримує цикл «превенція–виявлення–реагування–навчання» та може бути адаптований до локальних регуляторних вимог (зокрема у поєднанні з відкритими даними та рішеннями нацрегуляторів) [3]. Подальші напрями: інтеграція мережевих і причинних ознак, доменно-специфічні онтології CPV та проспективні пілоти «human-in-the-loop» на реальних потоках торгів.

Список використаних джерел:

1. Porter, R. H., & Zona, J. D. (1993). Detection of bid rigging in procurement auctions. *Journal of Political Economy*, 101(3), 518–538. <https://doi.org/10.1086/261885>
2. Bajari, P., & Ye, L. (2003). Deciding between competition and collusion. *The Review of Economics and Statistics*, 85(4), 971–989. <https://doi.org/10.1162/003465303772815871>
3. Organisation for Economic Co-operation and Development (OECD). (2025). *OECD guidelines for fighting bid rigging in public procurement (2025 update)*. OECD Publishing. <https://doi.org/10.1787/DOI pending>
4. Huber, M., & Imhof, D. (2019). Machine learning with screens for detecting bid-rigging cartels. *International Journal of Industrial Organization*, 65, 277–301. <https://doi.org/10.1016/j.ijindorg.2019.04.002>
5. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
6. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>
7. Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119–139. <https://doi.org/10.1006/jcss.1997.1504>
8. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>
9. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. <http://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html>
10. Wallimann, D. (2023). Screening for collusion in procurement markets: Methods and applications. *Journal of Competition Law & Economics*, 19(3), 495–528. <https://doi.org/10.1093/joclec/nhad020>

ЗАХІДНОУКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
 ФАКУЛЬТЕТ КОМП'ЮТЕРНИХ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
 КАФЕДРА КОМП'ЮТЕРНОЇ ІНЖЕНЕРІЇ



Комп'ютерна
Інженерія



**III ВСЕУКРАЇНСЬКА НАУКОВО-ПРАКТИЧНА
 КОНФЕРЕНЦІЯ СТУДЕНТІВ, АСПІРАНТІВ ТА
 МОЛОДИХ ВЧЕНИХ
 «ІНТЕЛЕКТУАЛЬНІ КОМП'ЮТЕРНІ СИСТЕМИ ТА
 МЕРЕЖІ»**

**ІКСМ
 ОСІНЬ 2025**

25 ЛИСТОПАДА 2025



KI.WUNU.EDU.UA/CONFERENCE/

**ТЕРНОПІЛЬ
 2025**



| | |
|--|-----|
| <i>Пархомчук Я.І.</i> | |
| Модель та засоби системи розумного будинку з використанням стохастичних мереж | |
| Петрі | 201 |
| <i>Ділай Р.І.</i> | |
| Гібридний метод виявлення об'єктів на основі зіставлення ознак зображень..... | 203 |
| <i>Дідух Д. Р.</i> | |
| Модель інтеграції мовних технологій у програмну систему для навчання іноземним мовам..... | 205 |
| <i>Бубнюк Н.В.</i> | |
| Розподілено-робастне управління проектом енергоспільноти з пуловим та P2P-ринками: кількісний аналіз результатів | 207 |
| <i>Черній І.Р.</i> | |
| Застосування машинного навчання в управлінні закупівельними проектами для виявлення змовницької поведінки | 209 |

Черній І.Р.
магістр 2 курсу ФКІТ ЗУНУ

ЗАСТОСУВАННЯ МАШИННОГО НАВЧАННЯ В УПРАВЛІННІ ЗАКУПІВЕЛЬНИМИ ПРОЕКТАМИ ДЛЯ ВИЯВЛЕННЯ ЗМОВНИЦЬКОЇ ПОВЕДІНКИ

Вступ. Ефективне управління закупівельними проектами є ключовою умовою раціонального використання бюджетних ресурсів, забезпечення добросовісної конкуренції та успішної реалізації інфраструктурних і будівельних проектів. Однією з найбільш суттєвих загроз у сфері публічних закупівель є змовницька поведінка учасників торгів (bid-rigging), що призводить до завищення вартості контрактів, зниження якості виконання робіт та викривлення конкурентного середовища [1, 2].

В умовах цифровізації закупівельних процесів та зростання обсягів відкритих даних традиційні підходи контролю — ручні перевірки, формалізовані правила та постфактум-аудит — виявляються недостатньо ефективними для своєчасного реагування на ризики змови. У цьому контексті методи машинного навчання (Machine Learning, ML), поєднані зі статистичними скрінінговими індикаторами, розглядаються як перспективний інструмент підтримки управлінських рішень у закупівельних проектах [1].

Постановка задачі. Метою дослідження є узагальнення результатів емпіричних експериментів із застосування ML-алгоритмів для виявлення змовницької поведінки в публічних закупівлях та обґрунтування можливостей їх практичного використання в системі управління закупівельними проектами. Для досягнення поставленої мети передбачено вирішено проаналізувати ефективність ансамблевих ML-алгоритмів у задачі класифікації змовницьких та конкурентних аукціонів, оцінити якість моделей за різних сценаріїв доступності інформації, визначити найбільш інформативні скрінінгові індикатори змови, обґрунтувати управлінське значення ML-скрінінгу в циклі моніторингу закупівельних проектів.

Основний матеріал. Результати експериментальних досліджень за сценарію повної інформації, що передбачає наявність даних про ставки учасників, кількість поданих пропозицій, попередню кошторисну оцінку та ідентифікатори компаній, засвідчили високу ефективність ансамблевих методів машинного навчання. Зокрема, алгоритми Random Forest, Extra Trees та Gradient Boosting продемонстрували найкращі показники якості класифікації змовницьких аукціонів [1, 2].

Значення збалансованої точності (Balanced Accuracy) у більшості досліджених наборів даних перевищували 0,80, що свідчить про здатність моделей надійно розрізняти конкурентні та змовницькі процедури. З позицій управління закупівельними проектами це створює передумови для інтеграції ML-модулів у систему контролю та моніторингу проектів з метою раннього виявлення процедур із підвищеним ризиком змови.

Особливу практичну цінність мають результати сценарію обмеженої інформації, коли для аналізу доступні лише базові дані аукціону — значення ставок і інформація про переможця. Саме такий сценарій є типовим для багатьох реальних закупівельних проектів, особливо на ранніх етапах або в умовах обмеженого доступу до розширених метаданих.

Експерименти показали, що навіть за таких умов ансамблеві ML-моделі у поєднанні зі скрінінговими індикаторами розподілу ставок (щільність ставок поблизу мінімального значення, відносна різниця між двома найнижчими пропозиціями, коефіцієнт варіації) демонструють стабільну якість класифікації [1]. Це підтверджує доцільність використання ML-скрінінгу як інструменту попереднього ризик-аналізу в управлінні закупівельними проектами без потреби у складній інформаційній інфраструктурі.

Порівняльний аналіз міжкраїнових корпусів аукціонів засвідчив, що окремі скрінінгові індикатори мають стійку діагностичну спроможність незалежно від юрисдикції та типу аукціону. Найбільш інформативними виявилися індикатори щільності ставок біля мінімального значення та відносної різниці між двома найнижчими ставками [1, 2].

Окрему увагу приділено аналізу похибок класифікації. Результати досліджень показали, що шляхом калібрування порогів прийняття рішень можливо адаптувати ML-моделі до різних управлінських сценаріїв. Для юридично чутливих рішень доцільною є мінімізація хибних позитивних спрацювань, тоді як для внутрішнього контролю та профілактичного моніторингу пріоритетом може бути мінімізація пропущених випадків змови [2].

Висновки. Результати дослідження підтверджують доцільність використання машинного навчання у поєднанні зі скрінінговими індикаторами як ефективного інструменту підтримки управління закупівельними проєктами. Ансамблеві ML-моделі демонструють високу стійкість та здатність узагальнювати патерни змовницької поведінки за різних сценаріїв доступності даних.

Практичне значення отриманих результатів полягає у можливості інтеграції ML-модулів у системи моніторингу публічних закупівель, зокрема в електронній системі Prozorro, що функціонує на засадах відкритих даних [3]. Це є особливо актуальним для України в умовах масштабної відбудови, де ефективне управління закупівельними проєктами безпосередньо впливає на економічну стійкість та довіру до державних інституцій.

Список літератури

1. Huber, M., Imhof, D. Machine learning with screens for detecting bid-rigging cartels. *International Journal of Industrial Organization*, 2019.
2. Rodríguez, M. J. G., et al. Collusion detection in public procurement auctions with machine learning algorithms. *Decision Support Systems*, 2022.
3. Prozorro. Public Procurement System of Ukraine. <https://prozorro.gov.ua>