

## **РЕЦЕНЗІЯ**

на дисертаційну роботу  
Жука Івана Сергійовича

на тему «Математичні моделі і методи виявлення за публічно доступними  
даними підозрілих на фіксований результат футбольних матчів»,

представлену на здобуття ступеня доктора філософії

в галузі знань 11 Математика та статистика  
за спеціальністю 113 Прикладна математика

### **Актуальність теми дисертації**

Одним із сучасних напрямів інтелектуального аналізу даних, який знаходить застосування в самих різних галузях, є виявлення аномальної поведінки процесів і об'єктів, що спостерігаються. На відміну від традиційних статистичних застосувань, при інтелектуальному аналізі даних виявлення аномальних спостережень розглядається як виявлення новизни, оскільки вони можуть відповідати цікавим та важливим подіям. Особливим типом є контекстуальні аномалії, що спостерігаються, коли екземпляр даних є аномальним лише у певному контексті. Для визначення аномалій цього типу основним етапом є виділення контекстуальних і поведінкових атрибутів. Екземпляр даних може бути контекстуальною аномалією за даних умов, але за таких же поведінкових атрибутів вважатися нормальним в іншому контексті.

Як контекстуальні аномалії можна розглядати матчі з фіксованим результатом, що мають місце як в футболі, так і в інших ігрових видах спорту за наявності даних за весь сезон турніру. У договірних матчах, пов'язаних з вигравом за ставками, завданням є отримання результату, відмінного від очікуваного. Тому такі результати можна розглядати як аномальні. Перевагою задачі виявлення підозрілих щодо фіксованого результату матчів як контекстуальних аномалій за результатами сезону футбольного турніру у порівнянні з відомими методами спортивної аналітики є наявність загально доступної інформації щодо результатів проведених ігор всіх команд.

Особливістю задачі виявлення підозрілих щодо фіксованого результату матчів за результатами сезону футбольного турніру є те, що вхідні дані характеризуються малою кількістю дискретних числових значень, а їх закони розподілу ймовірностей є невідомими. Тому серед методів інтелектуального аналізу даних адекватним розглянутій задачі є статистичний непараметричний гістограмний метод. Але за вибірок малого об'єму, що обумовлено невеликою

кількістю команд в групах, сформованих за силою команд, гістограмний метод характеризується низькою ефективністю.

Новим перспективним напрямом пошуку аномалій у даних є математичний апарат конформних предикторів та степеневих мартингалів. Він заснований на використанні міри неконформності даних і не вимагає знання їх законів розподілу. Також цей підхід дозволяє вводити оцінки гарантованої точності для отриманих рішень.

Тому актуальною науковою задачею слід вважати розробку методів, виявлення підозрілих на фіксований результат матчів як контекстуальних аномалій на базі обробки виключно загальнодоступних публічних даних з використанням апарату конформних предикторів і степеневих мартингалів.

### **Оцінка обґрунтованості наукових результатів дисертації, їх достовірності та новизни**

Обґрунтованість наукових положень, висновків і рекомендацій, сформульованих у дисертації, обумовлена коректною постановкою науково-прикладної задачі і часткових завдань досліджень, використанням сучасних апробованих методів математичного аналізу, теорії ймовірностей і математичної статистики, машинного навчання, теорії конформних предикторів, а також методів експериментальних досліджень, таких як імітаційне моделювання.

Достовірність отриманих наукових результатів забезпечено несуперечністю отриманих у дисертації наукових результатів загальновідомим положенням розглянутої теорії, збігом окремих отриманих наукових результатів з відомими, доведенням математичних тверджень щодо властивостей степеневого мартингала, експериментальною перевіркою отриманих теоретичних положень шляхом імітаційного моделювання, а також на реальних даних.

До основних наукових результатів, отриманих в дисертації, можна віднести такі:

1. Розроблено новий метод виявлення підозрілих щодо фіксованості результату футбольних матчів, який відрізняється від відомих застосуванням конформного аномального детектора із запропонованою мірою неконформності поточного матчу, що забезпечує можливість визначення порогу прийняття рішення у відповідності до заданого значення апіорної ймовірності появи аномальних даних.

2. Розроблено новий метод виявлення підозрілих щодо фіксованості результату футбольних матчів, який відрізняється від відомих застосуванням степеневого мартингалу і правилом прийняття рішення на основі порівняння поточного значення степеневого мартингалу з попереднім, що дозволяє за рахунок зміни параметра чутливості налаштовувати степеневий мартингал на виявлення аномалій відповідного рівня і знаходити їх.

3. Розроблено новий метод виявлення підозрілих щодо фіксованості результату футбольних матчів, який відрізняється від відомих застосуванням інтегрального мартингалу і правилом прийняття рішення на основі порівняння поточного значення інтегрального мартингалу з попереднім, що дає змогу виявляти аномальні матчі без налаштування параметрів.

4. Отримала подальший розвиток імітаційна модель футбольного сезону, яка на відміну від існуючих враховує розбиття матчів на класи за контекстуальними атрибутами «сила команди» і «тип гри» — домашня або виїзна, що забезпечує моделювання договірних матчів з фіксованим результатом, які мають аномальний характер.

5. Удосконалено метод кластеризації на основі Гаусівських сумішей в частині регуляризації недіагональних елементів коваріаційних матриць, що дало змогу зменшити чутливість до початкових умов і отримувати кластери еліпсоподібної форми, які враховують неочевидні зв'язки між точками набору даних.

Дисертаційні дослідження були виконані здобувачем на кафедрі прикладної математики КПІ ім. Ігоря Сікорського в рамках науково-дослідної роботи № 2310п «Інформаційно-аналітична система для математичного моделювання та управління соціальними ризиками з застосуванням у техніці та медицині» (номер державної реєстрації — 0120U102216) під керівництвом завідувача кафедри прикладної математики КПІ ім. Ігоря Сікорського, д.т.н., професора Чертова Олега Романовича.

Отже, в дисертаційній роботі поставлене наукове завдання розробки методів виявлення підозрілих на фіксований результат матчів з використанням апарату конформних предикторів і степеневих мартингалів на базі обробки виключно загальнодоступних публічних даних виконано повністю, здобувач повною мірою оволодів методологією наукової діяльності.

**Оцінка змісту дисертації, її завершеність та дотримання принципів академічної доброчесності**

За своїм змістом дисертаційна робота здобувача Жука І. С. повністю відповідає Стандарту вищої освіти зі спеціальності 113 Прикладна математика та напрямкам досліджень відповідно до освітньої програми 46343 Прикладна математика.

Дисертаційна робота є завершеною науковою працею і свідчить про наявність особистого внеску здобувача у науковий напрям інтелектуального аналізу даних для пошуку аномалій з використанням конформних предикторів і степеневих мартигалів.

Розглянувши звіт подібності за результатами перевірки дисертаційної роботи на текстові співпадіння, можна зробити висновок, що дисертаційна робота Жука Івана Сергійовича є результатом самостійних досліджень здобувача і не містить елементів фальсифікації, компіляції, фабрикації, плагіату та запозичень. Використані ідеї, результати і тексти інших авторів мають належні посилання на відповідне джерело.

### **Мова та стиль викладення результатів**

Дисертаційна робота написана українською мовою. Вона складається з вступу, чотирьох розділів, висновків, списку літератури та додатків. Загальний обсяг дисертації складає 165 сторінок.

У вступі наведено загальні характеристики роботи, обґрунтовано її актуальність і зв'язок з науковими програмами, планами, темами. Сформульована мета та завдання досліджень, наукова новизна і практичне значення одержаних результатів. Наведено дані щодо особистого внеску здобувача, апробацію результатів роботи і публікації.

У першому розділі розглянуто проблематику договірних футбольних матчів і обґрунтовано актуальність задачі виявлення футбольних матчів, підозрілих на фіксований результат. Проведено аналіз математичних методів футбольної аналітики, таких як прогнозування результату матчу, аналізу ставок та дій учасників матчу протягом всієї гри щодо виявлення підозрілих на фіксований результат матчів і відмічені їх недоліки. Розглянуто задачу виявлення підозрілих щодо фіксованого результату матчів за результатами сезону футбольного турніру і показано, що вона відноситься до класу задач виявлення контекстних аномалій. Проведено аналіз методів інтелектуального аналізу даних для пошуку аномалій. Особливу увагу присвячено методам класифікації на основі конформних предикторів, які запропоновано використати для підвищення ефективності виявлення потенційно підозрілих

договірних матчів. Сформульовано мету і наукові завдання дисертаційного дослідження.

Другий розділ роботи присвячено формалізації опису вхідних даних, групування команд та розробці імітаційної моделі футбольного сезону з матчами з фіксованим результатом. З використанням методів  $K$ -середніх та Гаусівських сумішей проведено групування команд сезону за силою. За допомогою визначених контекстуальних атрибутів на основі реальних даних розроблено імітаційну модель футбольного сезону та проведено її перевірку за критерієм Колмогорова-Смирнова. Розроблено алгоритм моделювання договірних матчів, пов'язаних із заробітком на ставках.

Третій розділ займає центральне місце в роботі за отриманими науковими результатами і носить теоретичний характер. В ньому розроблено метод виявлення підозрілих щодо фіксованості результату футбольних матчів за наявності даних про весь сезон з використанням конформного аномального детектора. Доведені нові властивості степеневого мартингалу, які в подальшому використані при розробці методів виявлення підозрілих щодо фіксованості результату футбольних матчів з використанням степеневого та інтегрального мартингалів.

Четвертий розділ має експериментальний характер і присвячений аналізу і оцінці ефективності розроблених методів виявлення підозрілих щодо фіксованості результату футбольних матчів за наявності даних про весь сезон шляхом імітаційного моделювання. Розроблені методи також перевірено на реальних даних сезону 2014–2015 рр. Серії *B* Італії.

У висновках приведено вичерпний перелік основних отриманих результатів та рекомендацій до їх застосування.

Робота написана грамотно і достатньо проілюстрована графіками й рисунками. Основні результати та положення дисертації характеризуються єдністю змісту, а стиль викладу матеріалів досліджень забезпечує доступність їх сприйняття.

Дисертаційна робота оформлена відповідно до вимог наказу МОН України від 12 січня 2017 р. № 40 «Про затвердження вимог до оформлення дисертації».

### **Оприлюднення результатів дисертаційної роботи**

Основні результати дисертаційної роботи достатньо повно відображено у 8 наукових працях, в тому числі 3 статті, 2 з яких опубліковані у виданні, що входить до міжнародної наукометричної бази SCOPUS з квантилем Q3, 5 тез доповідей на науково-технічних конференціях, що цілком відповідає вимогам,

які ставляться до опублікування результатів дисертацій на здобуття ступеня доктора філософії. Усі основні результати дисертаційного дослідження, представлені до захисту, одержані автором особисто.

### **Недоліки та зауваження до дисертаційної роботи**

1. Потребує пояснення, чому для визначення груп команд за силою використано лише два контекстуальних атрибута «кількість набраних очок» та «різниця між забитими і пропущеними м'ячами команд».

2. В роботі йде мова про вибірки малого об'єму, але не сформульовано означення малої вибірки.

3. В роботі розроблено метод виявлення підозрілих на фіксований результат матчів на основі степеневого мартингалу, показники ефективності якого залежать від параметрів чутливості. Але при цьому методики вибору значення цього параметру не наведено.

4. Статистичне моделювання було проведено на одному модельному сезоні. Було б бажано провести таке моделювання на наборі модельних сезонів, створених для різних реальних сезонів, що дозволило б більш точно оцінити ефективність запропонованих методів виявлення підозрілих щодо фіксованого результату матчів.

5. В роботі не вказано, чи можуть бути використані запропоновані моделі і методи для виявлення підозрілих щодо фіксованого результату матчів, спрямованих на досягнення спортивного результату.

Вважаю, що висловлені зауваження не є визначальними і не зменшують загальну наукову новизну та практичну значимість результатів та не впливають на позитивну оцінку дисертаційної роботи.

### **Висновок про дисертаційну роботу**

Вважаю, що дисертаційна робота здобувача ступеня доктора філософії Жука Івана Сергійовича на тему «Математичні моделі і методи виявлення за публічно доступними даними підозрілих на фіксований результат футбольних матчів» виконана на високому науковому рівні, не порушує принципів академічної доброчесності та є закінченим науковим дослідженням, сукупність теоретичних та практичних результатів якого розв'язує наукове завдання, що має істотне значення для галузі знань 11 Математика та статистика. Дисертаційна робота за актуальністю, практичною цінністю та науковою новизною повністю відповідає вимогам чинного законодавства України, що

передбачені в пп. 6–9 «Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії», затвердженого Постановою Кабінету Міністрів України від 12 січня 2022 р. № 44.

Здобувач Жук Іван Сергійович заслуговує на присудження ступеня доктора філософії в галузі знань 11 Математика та Статистика за спеціальністю 113 Прикладна математика.

**Рецензент:**

завідувач кафедри  
інформаційної безпеки,  
д.т.н., проф.

/  /

Дмитро ЛАНДЕ

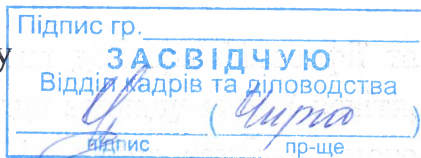
М.П.



«    »

20

року



**Примітки:**

1. Текст документа друкується без підкреслень та підрядкових символів
2. Підпис особи засвідчується печаткою за місцем основної роботи.