

РЕЦЕНЗІЯ

на дисертаційну роботу
Орехова Олександра Станіславовича
на тему

«Математичні моделі для обробки інформації з метрик коду JAVA-
застосунків для оцінювання їх розміру»,
представлену на здобуття наукового ступеня доктора філософії
в галузі знань 12 – «Інформаційні технології»
за спеціальністю 122 – «Комп'ютерні науки»

Актуальність теми дослідження

Актуальність дослідження зумовлена стрімким розвитком ІТ-індустрії та зростаючою потребою у достовірному оцінюванні розміру програмних застосунків на ранніх етапах розробки програмного забезпечення. Математичні моделі для обробки інформації з метрик коду є необхідною частиною сучасних процесів управління програмними проектами, що застосовуються для підвищення ефективності планування, точності прогнозування витрат та раціонального розподілу ресурсів. Зокрема, у продуктових ІТ-компаніях такі моделі використовуються для оцінювання обсягу робіт на етапі формування бюджету та визначення термінів випуску нових версій програмних продуктів. У сервісних та аутсорсингових компаніях вони сприяють підготовці комерційних пропозицій на пресейл-етапі, коли доступна лише концептуальна модель майбутнього проекту у вигляді UML-діаграми класів. Значного розвитку набули моделі оцінювання розміру для застосунків, що використовують технології Data Science (DS) та Machine Learning (ML), зокрема для рекомендаційних систем, систем комп'ютерного зору, обробки природної мови та прогнозової аналітики, які активно впроваджуються у галузях охорони здоров'я, фінансів, освіти та електронної комерції. Важливе значення такі моделі мають і для управління ризиками програмних проектів, оскільки достовірне оцінювання параметра розміру дозволяє своєчасно виявляти потенційні відхилення фактичних показників розробки від запланованих та вживати відповідних коригувальних заходів.

Таким чином, враховуючи що підвищення достовірності обробки інформації з метрик коду для оцінювання розміру Java-застосунків має широкий спектр застосувань для розрахунку оцінок трудомісткості розробки ПЗ в області управління ІТ проектами, дослідження в цьому напрямку є актуальним і важливим.

Дисертаційну роботу виконано у Національному університеті кораблебудування імені адмірала Макарова (НУК) відповідно до планів НДР з ініціативної теми «Математичні моделі обробки інформації з метрик коду для оцінювання параметра розміру Java-застосунків на ранніх етапах розробки програмних проєктів» (Державний реєстраційний номер: 0125U002663).

Структура та зміст дисертації

Дисертаційна робота складається з вступу, чотирьох розділів, висновків, списку використаних джерел із 107 найменувань та 8 додатків. Загальний обсяг дисертації становить 223 сторінки, обсяг основного тексту складає 133 сторінки. Дисертаційна робота містить 5 рисунків та 44 таблиці. У додатках наведено огляд успішності розробки проєктів ПЗ, набори даних загальних, DS та ML Java-застосунків, документи про впровадження результатів роботи, список публікацій здобувача за темою дисертації, відомості про апробацію результатів дисертації, методику застосування інформаційної технології та вихідний код програми.

У вступі дана характеристика науково-практичної задачі, що підлягає вирішенню, розкрито сутність і значущість дослідження, обґрунтовано його актуальність, визначено мету, завдання, наукову новизну, практичну цінність отриманих результатів, а також наведено інформацію про апробацію та публікації результатів дослідження.

Перший розділ містить аналіз існуючих методів і моделей обробки інформації з метрик коду для оцінювання розміру Java-застосунків, формування наборів даних, трансформацію метрик коду та обґрунтування необхідності проведення дослідження за обраною темою.

У другому розділі проаналізовано нормалізуючі перетворення та здійснено вибір оптимального варіанту для нормалізації шестивимірних векторів метрик коду для загальних і DS та ML Java-застосунків. Виконано одноетапну нормалізацію багатовимірних даних із використанням одновимірних та багатовимірних нормалізуючих перетворень.

Третій розділ присвячений розробленим чотирьохфакторним та п'ятифакторним нелінійним регресійним моделям на основі нормалізуючих перетворень Бокса-Кокса та побудові їх інтервалів передбачення і довірчих інтервалів. Отримані нелінійні регресійні моделі порівняно між собою за критеріями якості та перевірено на незалежних тестових вибірках.

У четвертому розділі запропоновано інформаційну технологію та методику обробки інформації з метрик коду для оцінювання розміру Java-застосунків за допомогою побудованих нелінійних регресійних моделей у вигляді вебзастосунку, який розроблено мовою програмування Kotlin.

Наукова новизна отриманих результатів

В дисертаційній роботі наведено чотири пункти наукової новизни.

- 1) Удосконалено чотирьохфакторну нелінійну регресійну модель для обробки інформації з метрик коду для оцінювання параметра розміру JAVA-застосунків залежно від загальної кількості класів та інтерфейсів (CLASS), загальної кількості унікальних викликів методів у класах (RFC), середнього значення кількості зв'язків між класами (aCBO) та середнього значення кількості видимих (public та protected) методів на клас (aVMQ) на основі п'ятивимірного перетворення Бокса-Кокса, що дозволило підвищити достовірність оцінювання параметра розміру таких застосунків як негаусівської випадкової величини у порівнянні з існуючими регресійними моделями. Модель, що побудована, у порівнянні з іншими моделями має більший відсоток прогнозованих значень, менше значення середньої величини відносної похибки та менші ширини інтервалу передбачення нелінійної регресії.

2) Вперше побудовано п'ятифакторну нелінійну регресійну модель для обробки інформації з метрик коду для оцінювання параметра розміру JAVA-застосунків на основі шестивимірного перетворення Бокса-Кокса, в якій, на відміну від існуючих моделей, було розділено метрику загальної кількості класів та інтерфейсів (CLASS) на дві окремі метрики, а саме кількість класів (CLS) та кількість інтерфейсів (INFC), що дозволило підвищити достовірність обробки інформації з метрик коду для оцінювання параметра розміру в порівнянні з існуючими математичними моделями. Модель, що побудовано, в порівнянні з іншими регресійними моделями, має кращі показники якості R^2 , $MMRE$ та $PRED(0,25)$ та менші ширини інтервалу передбачення та довірчого інтервалу нелінійної регресії.

3) Вперше побудовано п'ятифакторну нелінійну регресійну модель для обробки інформації з метрик коду для оцінювання параметра розміру DS та ML JAVA-застосунків на основі шестивимірного перетворення Бокса-Кокса, в якій, на відміну від існуючих моделей, було розділено метрику загальної кількості класів та інтерфейсів (CLASS) на дві окремі метрики коду кількості класів (CLS) та інтерфейсів (INFC) та враховано особливості розробки цієї категорії застосунків, що дозволило підвищити достовірність обробки інформації з метрик коду для оцінювання параметра розміру у порівнянні із використанням одновимірного та п'ятивимірного перетворень Бокса-Кокса. Модель, що побудовано, в порівнянні з іншими регресійними моделями, має більший відсоток прогнозованих значень, менше значення середньої величини відносної похибки та менші ширини інтервалу передбачення та довірчого інтервалу нелінійної регресії.

4) Отримали подальший розвиток рівняння для визначення границь інтервалів передбачення п'ятифакторної нелінійної регресії для загальних JAVA-застосунків та DS і ML JAVA-застосунків на основі шестивимірних нормалізуючих перетворень Бокса-Кокса. Це у порівнянні з використанням п'ятивимірних перетворень Бокса-Кокса дозволяє зменшити ширини інтервалів передбачення нелінійних регресій для оцінювання параметра розміру

відповідних програмних застосунків і тим самим підвищити достовірність оцінювання зазначеного розміру як залежної випадкової величини.

Дисертаційна робота має ознаки наукової новизни в достатньому об'ємі для присудження здобувачу наукового ступеня доктора філософії.

Рівень виконання поставленого наукового завдання

Наукове завдання дисертаційної роботи складається з 8 пунктів.

1) *Провести аналіз існуючих математичних моделей для обробки інформації з метрик коду для оцінювання розміру загальних та DS і ML JAVA-застосунків, із використанням відповідних вибірок проєктів з відкритим кодом.*

Перший пункт виконано повністю. Було проведено аналіз існуючих математичних моделей, що використовуються для обробки інформації з метрик коду Java-застосунків для оцінювання їх розміру. Обґрунтовано неможливість їх застосування на вибірках метрик коду сучасних Java-застосунків. Отримані результати узагальнено у першому розділі дисертації.

2) *Сформувати навчальні та тестові вибірки з метрик коду JAVA-застосунків та обґрунтувати вибір незалежних факторів для побудови нелінійних регресійних моделей обробки інформації з метрик коду для оцінювання їх розміру.*

Другий пункт виконано повністю. Було сформовано набори вибірок метрик коду з проєктів Java-застосунків з відкритим кодом. Підібрано набори метрик коду, які доступні з UML-діаграми класів для оцінювання розміру Java-застосунків на етапі проєктування ПЗ для подальшої побудови нелінійних регресійних моделей.

3) *Дослідити отримані метрики на відсутність мультиколінеарності за критерієм VIFs, з урахуванням ітеративної обробки багатовимірних даних.*

Третій пункт виконаний в повному обсязі. Обраний набір метрик перевірено на наявність мультиколінеарності між незалежними змінними за критерієм VIFs та проведено їх трансформацію для зменшення мультиколінеарності між ними.

4) Обрати перетворення для нормалізації шестивимірних векторів характеристик метрик коду JAVA-застосунків на основі аналізу існуючих нормалізуючих перетворень та методів оцінювання їх параметрів.

Було розглянуто одновимірні та багатовимірні нормалізуючі перетворення, які застосовуються для наближення розподілу даних до нормального. Здійснено оцінку параметрів цих перетворень та проведено вибір оптимального підходу для нормалізації багатовимірних векторів характеристик метрик коду Java-застосунків.

5) Удосконалити чотирьохфакторну нелінійну регресійну модель для обробки інформації з метрик коду для оцінювання розміру JAVA-застосунків із використанням п'ятивимірного нормалізуючого перетворення Бокса-Кокса на основі загальної вибірки метрик коду JAVA-застосунків з відкритим кодом, які можна отримати з концептуальної моделі проєкту.

На основі нормалізованих п'ятивимірних векторів характеристик метрик коду розроблено чотирьохфакторну нелінійну регресійну модель та її інтервали передбачення і довірчі інтервали. Це дозволило підвищити достовірність обробки інформації з метрик коду для оцінювання параметра розміру в порівнянні з існуючими математичними моделями. Опис моделі включено в третій розділ дисертації, що засвідчує виконання цього завдання.

6) Побудувати п'ятифакторну нелінійну регресійну модель для обробки інформації з метрик коду для оцінювання розміру JAVA-застосунків із застосуванням шестивимірного нормалізуючого перетворення Бокса-Кокса на основі загальної вибірки метрик коду JAVA-застосунків з відкритим кодом, які можна отримати з концептуальної моделі проєкту.

В третьому розділі розроблено п'ятифакторну нелінійну регресійну модель в якій, на відміну від існуючих моделей, метрику загальної кількості класів та інтерфейсів розділено на дві окремі метрики, а саме фактична кількість класів та фактична кількість інтерфейсів, що дозволило підвищити достовірність обробки інформації з метрик коду для оцінювання параметра розміру в порівнянні з існуючими математичними моделями. Ця модель має кращі показники оцінок

критеріїв якості та менші ширини інтервалів передбачення та довірчого інтервалу нелінійної регресії.

7) Побудувати п'ятифакторну нелінійну регресійну модель для обробки інформації з метрик коду для оцінювання розміру JAVA-застосунків із застосуванням шестивимірного нормалізуючого перетворення Бокса-Кокса на основі вибірки метрик коду DS та ML JAVA-проектів з відкритим кодом, які можна отримати з концептуальної моделі проекту.

В рамках дослідження було вперше побудовано п'ятифакторну нелінійну регресійну модель для обробки інформації з метрик коду для оцінювання параметра розміру Java-застосунків, що враховує особливості розробки категорії DS та ML. Ця модель, в порівнянні з іншими регресійними моделями, має більший відсоток прогнозованих значень, менше значення середньої величини відносної похибки та менші ширини інтервалу передбачення та довірчого інтервалу нелінійної регресії. Завдання виконано повністю.

8) Розробити інформаційну технологію (SaaS програмне рішення) та методику для обробки інформації з метрик коду, отриманих з UML-діаграми класів, для оцінювання параметра розміру JAVA-застосунків із використанням побудованих п'ятифакторних нелінійних регресійних моделей.

На основі розроблених нелінійних регресійних моделей було створено інформаційну технологію, що дозволяє здійснювати обробку інформації з метрик коду для оцінювання розміру Java-застосунків на ранніх етапах проектування. До розробленого ПЗ додається методика проведення обчислень. Завдання виконано повністю.

Таким чином, поставлені у дисертаційній роботі завдання досягнуті.

Рівень оволодіння здобувачем методологією наукової діяльності

У дисертаційній роботі здобувач послідовно та ґрунтовно підходить до отримання наукових результатів. У першому розділі здійснено детальний аналіз існуючих методів і моделей обробки інформації з метрик коду Java-застосунків для оцінювання їх розміру, сформовано навчальні і тестові вибірки метрик коду

загальних Java-застосунків, DS та ML Java-застосунків, обґрунтовано необхідність подальших досліджень. Другий розділ присвячено дослідженню нормалізуючих перетворень та вибору оптимального підходу для нормалізації шестивимірних векторів метрик коду загальних Java-застосунків, DS та ML Java-застосунків. У третьому розділі розроблено чотирьохфакторну нелінійну регресійну модель та п'ятифакторні нелінійні регресійні моделі на основі вибірок метрик коду Java-застосунків, побудовано їх інтервали передбачення та довірчі інтервали, проведено порівняння цих моделей за критеріями якості для регресій. Четвертий розділ містить розробку інформаційної технології та ПЗ з використанням мови програмування Kotlin для розрахунку параметра розміру Java-застосунків та методику для обробки відповідної інформації.

Отримані результати підтверджують методологічну обґрунтованість дослідження, а також демонструють вміння здобувача застосовувати наукові підходи до побудови, аналізу та практичної реалізації ймовірнісних моделей.

Таким чином, у роботі забезпечено логічний, послідовний та методологічно обґрунтований підхід до дослідження обраної теми, що свідчить про достатній рівень оволодіння здобувачем методологією наукової діяльності.

Практичне значення одержаних результатів

Завдяки розробці інструментарію ІТ для оцінювання розміру Java-застосунків на основі інформації з метрик коду, використовуючи мову програмування Kotlin та Spring Framework, вдалося підвищити достовірність оцінювання параметра розміру на ранніх етапах проектування ПЗ для мови програмування Java. Це підтверджується актами впровадження результатів дисертаційної роботи в ТОВ «Н-ІКС ДЕЛІВЕРІ», ФОП Поліщук Андрій Борисович та при викладанні студентам Навчально-наукового інституту комп'ютерних наук та управління проектами (ННІКНУП) Національного університету кораблебудування імені адмірала Макарова таких дисциплін: «Керування проєктними ресурсами в інженерії програмного забезпечення» в ОПП «Інженерія програмного забезпечення» для 2-го магістерського рівня ВО,

спеціальності 121 (F2) «Інженерія програмного забезпечення», галузі знань 12 (F) «Інформаційні технології»; «Основи програмної інженерії» в ОПП «Інженерія програмного забезпечення» для 1-го бакалаврського рівня ВО, спеціальності 121 (F2) «Інженерія програмного забезпечення», галузі знань 12 (F) «Інформаційні технології»; «Моделювання систем цифрового екологічного моніторингу» в ОПП «Цифровий екологічний моніторинг» для 2-го магістерського рівня ВО, спеціальності E2 «Екологія», галузі знань E «Природничі науки, математика та статистика».

Відповідність теми дисертації та рівня наукової обґрунтованості результатам досліджень

Тема дисертації повністю відповідає проведенню у ній дослідженням. У роботі досліджуються математичні моделі обробки інформації з метрик коду Java-застосунків для оцінювання їх розміру, методи нормалізації даних та побудова нелінійних регресійних моделей, що узгоджується з метою та завданнями дисертації.

Рівень наукової обґрунтованості результатів базується на використанні сучасних методів математичного моделювання та статистичного аналізу, що пройшли відповідну апробацію в попередніх дослідженнях. Робота спирається на загальновизнані методи нормалізації, зокрема багатовимірне перетворення Бокса-Кокса, а також класичні статистичні підходи, такі як розподіли χ -квадрат, Гауса та Стюдента в алгоритмі побудови нелінійних регресійних моделей.

Важливим аспектом є те, що отримані теоретичні результати підтверджені експериментальними дослідженнями, проведеними на реальних наборах даних з метрик коду загальних Java-застосунків, DS та ML Java-застосунків. Аналіз результатів підтверджує підвищення достовірності обробки інформації для оцінювання розміру за рахунок використання побудованих моделей.

Таким чином, дисертаційна робота містить достатній рівень наукової обґрунтованості отриманих результатів і відповідає заявленій тематиці дослідження.

Відповідність дисертації встановленим вимогам, затвердженим МОН України, та її завершеність

Дисертаційна робота є завершеним науковим дослідженням, у якому чітко визначено мету і завдання, а також обґрунтовано вибір методологічних та методичних підходів. Сформульовані в дисертації положення, висновки та практичні рекомендації є повними, науково обґрунтованими і логічно узгоджуються з метою та поставленими завданнями.

Робота написана українською мовою. Автор демонструє вміння послідовно й логічно викладати суть проблеми, аргументовано обґрунтовувати запропоновані рішення. Отримані результати є достовірними, добре узгоджуються із сучасними науковими підходами та не викликають сумнівів.

Представлена дисертаційна робота відповідає вимогам Міністерства освіти і науки України, що висуваються до дисертацій на здобуття наукового ступеня доктора філософії.

Повнота викладу результатів дослідження в опублікованих працях

Оцінка наукових публікацій, повноти опублікування результатів дисертації та особливого внеску здобувача у всі наукові публікації, зараховані за темою дисертації, показала, що результати дослідження, викладені в дисертаційній роботі, отримані автором самостійно та повністю відображають основні положення та висновки роботи. Вони повно висвітлені у публікаціях, доповідалися та обговорювалися на науково-практичних конференціях.

Основні результати дисертаційної роботи викладені у 19 публікаціях. Основні наукові результати дисертації опубліковано у 4 статтях у наукових журналах, що на момент публікації входили до переліку наукових фахових видань України та в 1 одноосібному розділі у колективній монографії. Ще 12 публікацій засвідчують апробацію матеріалів дисертації, з яких 3 індексуються в базі даних Scopus (за винятком видань з держав, визнаних Верховною Радою України державами-агресорами), 1 – розділ в колективній монографії у

співавторстві та 1 – свідоцтво про реєстрацію авторського права на комп'ютерну програму, що додатково відображає наукові результати дисертації.

Автором виконані вимоги пп. 8 та 9 Постанови Кабінету Міністрів України № 44 від 12.01.2022 р. «Про затвердження Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії» (із змінами, внесеними згідно з Постановами КМ № 341 від 21.03.2022, № 502 від 19.05.2023, № 507 від 03.05.2024) щодо кількості публікацій, відповідності опублікованих результатів тим, що містяться у дисертації, апробації її основних положень.

Дотримання академічної доброчесності

Проведений аналіз дисертаційної роботи та публікацій автора засвідчив повну відповідність вимогам академічної доброчесності. Згідно аналізу звіту щодо перевірки тексту дисертації електронним сервісом Identific від 17.03.2026 р., порушень академічної доброчесності не було встановлено.

Використанні здобувачем окремі наукові положення і результати інших науковців мають посилання на відповідні джерела.

Дискусійні положення та зауваження щодо змісту дисертації

Попри значну наукову цінність роботи, у дисертаційному дослідженні є деякі аспекти, які могли б бути покращені.

1) У першому розділі наведено огляд існуючих лінійних та нелінійних регресійних моделей для оцінювання розміру Java-застосунків, однак їх аналіз має переважно описовий характер. Було б доцільно подати порівняльний аналіз існуючих моделей у вигляді узагальнюючої таблиці з чітким зазначенням використаних факторів, обсягу вибірок, показників якості (R^2 , $MMRE$, $PRED(0,25)$) та ширини інтервалів передбачення, що дозволило б більш наочно обґрунтувати наукову новизну запропонованих у дисертації моделей.

2) У роботі слід було приділити увагу багатовимірному нормалізуючому перетворенню Джонсона. Це перетворення гарно працює з даними, що мають значні відхилення від нормального розподілу. Використання цього підходу дозволило б оцінити його ефективність у порівнянні з іншими нормалізуючими перетвореннями та, можливо, отримати кращі результати.

3) Для запропонованої інформаційної технології не наведено опис процесу тестування та валідації коректності обчислень, які виконуються розробленим ПЗ. Зокрема, у роботі відсутня інформація щодо покриття SaaS-рішення модульними та інтеграційними тестами, які б підтверджували достовірність програмної реалізації побудованих нелінійних регресійних моделей. Доцільно було б доповнити роботу описом методики тестування програмного коду та результатів валідації обчислень, що дозволило б підтвердити коректність роботи розробленої ІТ та достовірність отриманих оцінок параметра розміру Java-застосунків.

4) У дисертаційній роботі відсутній опис методики імпорту та експорту побудованих нелінійних регресійних моделей у складі розробленого SaaS-рішення. Було б доцільно навести формат збереження параметрів моделей та процедуру їх інтеграції до інформаційної технології, що дозволило б спростити процес оновлення моделей та розширення системи новими регресійними залежностями у майбутніх дослідженнях.

Висловлені зауваження та рекомендації не впливають на загальну позитивну оцінку дисертаційної роботи. Робота є результатом наукового дослідження, що має високу наукову та практичну цінність. Отримані результати свідчать про оригінальність мислення автора та його внесок у розвиток наукової думки.

Висновок

Дисертаційна робота Орєхова Олександра Станіславовича є оригінальним науковим дослідженням, що вирізняється високим рівнем теоретичної обґрунтованості та практичної значущості. Отримані результати є новими і

достовірними, а висновки – логічними та обґрунтованими. Публікації автора, присвячені темі дисертації, підтверджують актуальність обраної проблематики та свідчать про наукову зрілість дослідника.

Дисертаційна робота на тему «Математичні моделі для обробки інформації з метрик коду JAVA-застосунків для оцінювання їх розміру» відповідає вимогам Наказу Міністерства освіти і науки № 40 від 12.01.2017 р. «Про затвердження вимог до оформлення дисертації» (із змінами, внесеними згідно з Наказом Міністерства освіти і науки № 759 від 31.05.2019) та Постанові Кабінету Міністрів України № 44 від 12.01.2022 р. «Про затвердження Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії» (із змінами, внесеними згідно з Постановами КМ № 341 від 21.03.2022, № 502 від 19.05.2023, № 507 від 03.05.2024), а її автор Орехов Олександр Станіславович, заслуговує на присудження наукового ступеня доктора філософії, за спеціальністю 122 – «Комп'ютерні науки», галузі знань 12 – «Інформаційні технології».

Рецензент

доцент кафедри програмного забезпечення
автоматизованих систем,
кандидат технічних наук, доцент



Лідія МАКАРОВА

Підпис Макарової Л.М. засвідчую:
начальник відділу кадрів
Національного університету кораблебудування
імені адмірала Макарова



Анатолій ЗУБАРЄВ