

ВІДГУК

офіційного опонента на дисертаційну роботу

Орєхова 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 обраної теми дисертаційної роботи

Дисертаційна робота Орєхова Олександра Станіславовича присвячена вирішенню важливого науково-практичного завдання – підвищенню достовірності обробки інформації з метрик коду з концептуальної моделі проєкту, необхідних для оцінювання параметра розміру Java-застосунків на ранніх стадіях розробки програмного забезпечення (ПЗ), шляхом побудови відповідних нелінійних регресійних моделей та створенню інформаційної технології для обробки інформації з метрик коду із використанням отриманих математичних моделей.

Процес створення ПЗ супроводжується плануванням ресурсів для розробки, що включає формування бюджету та розрахунок часу, необхідного для реалізації. Оцінка трудомісткості розробки ПЗ є важливим показником для планування проєкту. Серед існуючих методів найкращі оцінки трудомісткості досягаються за допомогою відповідних параметричних моделей, які враховують розмір із урахуванням факторів середовища програмного застосунку, таких як мова програмування та категорія програмного проєкту. Такий підхід дозволяє врахувати особливості розробки ПЗ для обраної мови програмування та категорії програмного проєкту. Тематика роботи спрямована на оцінювання розміру Java-застосунків за допомогою розробки відповідних математичних моделей для обробки інформації з метрик програмного коду.

Для оцінювання розміру Java-застосунків існує низка математичних моделей, однак, як показало теоретичне та експериментальне дослідження, ці моделі не забезпечують достовірного оцінювання розміру застосунків. Тому дослідження, спрямовані на підвищення достовірності обробки інформації з метрик коду для оцінювання розміру Java-застосунків шляхом побудови нелінійних регресійних моделей, є актуальними та становлять значний науковий і практичний інтерес.

Наукова новизна отриманих результатів

Наукова новизна дисертаційної роботи полягає в такому.

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, порівняно з іншими моделями має більший відсоток прогнозованих значень, менше значення середньої величини відносної похибки та меншу ширину 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 побудована, порівняно з іншими регресійними моделями, має кращі показники якості R^2 , MMRE та PRED(0,25) і 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 побудована, порівняно з іншими регресійними моделями, має більший відсоток прогнозованих значень, менше значення середньої величини відносної похибки та менші ширини інтервалу передбачення й довірчого інтервалу нелінійної регресії.

4. Отримали подальший розвиток рівняння для визначення меж 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.

Згідно з поставленим науковим завданням, дисертаційне дослідження виконано в повному обсязі, а всі поставлені завдання успішно вирішено. Отримані результати підтверджують підвищення достовірності обробки інформації з метрик коду для оцінювання розміру Java-застосунків завдяки використанню розроблених багатофакторних нелінійних регресійних моделей.

Практичне значення одержаних результатів

Практичне значення отриманих результатів полягає в розробці інструментарію ІТ для обробки інформації з метрик коду для оцінювання параметра розміру Java-застосунків, розрахунку довірчих інтервалів та інтервалів передбачення із застосуванням відповідних п'ятифакторних нелінійних регресійних моделей на основі загальної вибірки Java-застосунків, вибірки DS та ML Java-застосунків. Результати дисертаційної роботи впроваджено в таких установах та організаціях: ТОВ «Н-ІКС ДЕЛІВЕРІ» та ФОП Поліщук Андрій Борисович, що підтверджено відповідними актами впровадження. Також результати дисертаційного дослідження імплементовано в навчальний процес Національного університету кораблебудування ім. адмірала Макарова. Зокрема, під час викладання дисциплін «Керування проєктними ресурсами в інженерії програмного забезпечення», «Моделювання систем цифрового екологічного моніторингу», «Основи програмної інженерії».

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 основних положень дисертаційної роботи

Аналіз змісту дисертаційної роботи свідчить про її відповідність чинним вимогам МОН України щодо оформлення дисертаційних досліджень. Дисертаційна робота складається з вступу, чотирьох розділів, висновків, списку використаних джерел (107 найменувань) та 8 додатків. Загальний обсяг дисертації становить 223 сторінки, обсяг основного тексту складає 133 сторінки. Дисертаційна робота містить 5 рисунків і 44 таблиці. У додатках роботи наведено публікацію здобувача, де висвітлено аналіз успішності розробки проєктів ПЗ, набори даних метрик з проєктів Java-застосунків з відкритим кодом, акти впровадження отриманих результатів, перелік наукових публікацій здобувача за темою роботи,

інформацію про апробацію дослідження, інструкцію користувача та вихідний код інформаційної технології.

Дисертаційна робота демонструє системний і ґрунтовний підхід здобувача до отримання наукових результатів.

У першому розділі проведено детальний аналіз наявних методів і моделей обробки інформації з метрик коду для оцінювання розміру Java-застосунків, сформовано навчальні та тестові вибірки й обґрунтовано необхідність подальших досліджень.

Другий розділ присвячений аналізу нормалізуючих перетворень та вибору оптимального підходу для нормалізації шестивимірних векторів характеристик метрик коду з метою оцінювання параметра розміра.

У третьому розділі запропоновано алгоритм побудови нелінійних регресійних моделей, побудовано багатофакторні нелінійні регресійні моделі на основі перетворень Бокса-Кокса з їх інтервалами передбачення та довірчими інтервалами, а також проведено порівняння якості оцінювання параметра розміру.

В четвертому розділі розроблено інформаційну технологію для практичного використання побудованих нелінійних регресійних моделей.

Матеріал дисертаційної роботи викладено чітко та лаконічно, формулювання відповідають загальноприйнятій термінології в галузі інформаційних технологій. Отримані положення, наукові та практичні результати, узагальнення і висновки повністю відповідають темі роботи та її меті, розкривають суть вирішеної наукової задачі й завдань дослідження.

Ступінь обґрунтованості наукових положень, висновків, сформульованих у дисертації, їх достовірність

Наукові положення, висновки та рекомендації, викладені в дисертаційній роботі, є достатньо обґрунтованими, повними та відповідають сучасним вимогам наукових досліджень. Для їх отримання автор провів ґрунтовне дослідження, спираючись на вітчизняні та міжнародні літературні джерела, а також застосовуючи класичні й сучасні методи аналізу. Достовірність отриманих результатів підтверджується застосуванням математичного моделювання, статистичних методів аналізу та нормалізації даних, а також перевіркою ефективності побудованих нелінійних регресійних моделей на реальних наборах даних.

У процесі дослідження здобувач зібрав набори даних характеристик з метрик коду, визначив найбільш ефективне нормалізуюче перетворення для нормалізації багатовимірних даних, розробив методику їх застосування та експериментально довів підвищення достовірності оцінювання параметра розміру на ранніх етапах проектування ПЗ за допомогою побудованих математичних моделей. Кожен розділ

роботи містить логічні висновки, що дають змогу оцінити сутність проведених досліджень, а також практичну значущість отриманих результатів.

У висновках роботи надано загальну оцінку отриманих результатів та показано їхню взаємопов'язаність, що дає змогу застосовувати розроблені математичні моделі та методики в подальших дослідженнях і на практиці у сфері управління проектами з розробки ПЗ.

Повнота викладу результатів дослідження в опублікованих працях

Аналіз наукових публікацій, повноти представлення результатів дисертації у науковому просторі та особистого внеску здобувача у кожную з них підтверджує, що всі результати дослідження, викладені в дисертаційній роботі, отримані самостійно та повністю відображають основні положення й висновки роботи. Вони висвітлені у наукових публікаціях, представлені на науково-практичних конференціях, де були обговорені фаховою спільнотою.

Основні результати дисертаційного дослідження опубліковані у 18 наукових працях, а також отримано 1 свідоцтво на авторське право на комп'ютерну програму. З них 4 статті в наукових журналах, що на момент публікації входили до переліку наукових фахових видань України, та 1 одноосібний розділ у колективній монографії. Апробацію результатів дослідження засвідчують 3 статті конференцій, індексованих в базах даних Scopus, та 9 публікацій матеріалів міжнародних і всеукраїнських науково-практичних конференцій. Ще 1 розділ у колективній монографії у співавторстві та 1 свідоцтво на авторське право на комп'ютерну програму додатково відображають наукові результати дисертаційної роботи.

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

Аналіз змісту дисертаційної роботи свідчить про її відповідність чинним вимогам МОН України щодо оформлення дисертаційних досліджень. Дисертація характеризується логічною побудовою, чіткою науковою стилістикою викладу матеріалу, структурованістю та послідовністю. Всі положення засновані на сучасних методах математичного моделювання та статистичного аналізу, що підтверджено належним обґрунтуванням на всіх етапах роботи. У роботі детально висвітлено процес побудови багатофакторних нелінійних регресійних моделей для підвищення достовірності обробки інформації та оцінювання розміру Java-застосунків на ранніх етапах проектування ПЗ. Здобувачем проведено аналіз існуючих моделей та методів, формування та обробку наборів даних з метрик коду проектів ПЗ, вибір нормалізуючих перетворень, побудову багатофакторних нелінійних регресійних моделей, їх довірчих інтервалів та інтервалів передбачення, а також перевірку отриманих результатів.

Кожен етап дослідження має завершений характер, результати підкріплені статистичними розрахунками, експериментальними перевірками та відповідними джерелами. Таким чином, дисертаційне дослідження є завершеним, повністю відповідає заявленій тематиці та задовольняє вимоги, викладені в Постанові Кабінету Міністрів України № 44 від 12.01.2022 р. «Порядок 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).

Дотримання академічної доброчесності

Аналіз дисертаційної роботи та наукових публікацій автора підтвердив повну відповідність вимогам академічної доброчесності. Згідно з результатами перевірки тексту дисертації електронним сервісом Identific від 17 березня 2026 р., порушень академічної доброчесності не виявлено.

Усі використані здобувачем наукові положення та результати інших дослідників коректно оформлені з відповідними посиланнями на першоджерела, що підтверджує дотримання принципів академічної етики та коректності наукового цитування.

Рівень оволодіння здобувачем методологією наукової діяльності

Тема дисертації повністю відповідає її змісту та проведенням дослідженням. Результати дисертаційного дослідження мають високий рівень наукової обґрунтованості та чітко відповідають заявленій темі. У дисертаційній роботі послідовно викладено результати проведених досліджень.

Наукова обґрунтованість отриманих результатів базується на використанні загальновизнаних методів математичного моделювання та статистичного аналізу.

Здобуті результати підтверджують теоретичну обґрунтованість дослідження та демонструють здатність здобувача успішно застосовувати наукові методи під час побудови, аналізу й практичної реалізації математичних моделей. Робота вирізняється чіткою структурою та методологічною строгістю, що підтверджує високий рівень оволодіння здобувачем методологією наукової діяльності.

Відповідність теми дисертації та рівня наукової обґрунтованості результатам досліджень

Тема дисертації повністю відповідає її змісту та проведенням дослідженням. В роботі розглядаються нелінійні регресійні моделі обробки інформації з метрик коду для оцінювання розміру Java-застосунків, методи нормалізації даних і

побудова нелінійних регресійних моделей, що узгоджується із заявленою метою та завданнями дослідження.

Наукова обґрунтованість отриманих результатів базується на використанні сучасних методів регресійного аналізу, математичної статистики та моделювання, які пройшли апробацію в попередніх дослідженнях інших авторів. Дисертація ґрунтується на загальновизнаних методах нормалізації, зокрема одновимірних та багатовимірних перетвореннях Бокса-Кокса, а також класичних статистичних підходах, таких як розподіли хі-квадрат, розподіл Гауса, t-розподіл Стюдента.

Важливим аспектом є те, що отримані теоретичні результати підтверджені експериментальними дослідженнями, проведеними на реальних наборах даних з характеристик метрик коду Java-застосунків з відкритим кодом. Аналіз результатів підтверджує підвищення достовірності обробки інформації з метрик коду для оцінювання розміру Java-застосунків завдяки застосуванню побудованих моделей.

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

Дискусійні положення та зауваження щодо змісту дисертації

Зауваження стосуються як термінологічних моментів, так і питань практичної реалізації запропонованих підходів.

1. Основним елементом наукової новизни пунктів 2 та 3 заявлено «розподіл метрики CLASS на CLS та INFC». Проте з точки зору методології регресійного аналізу це лише декомпозиція однієї незалежної змінної на дві. Поліпшення моделі може пояснюватися не змістовною різницею між класами та інтерфейсами, а просто збільшенням кількості параметрів (з 4 до 5). Слід було провести F-тест для перевірки статистичної значущості додавання п'ятого фактора та порівняти AIC/BIC чотири- та п'ятифакторних моделей.

2. У висновках зазначено, що «рівняння для визначення границь інтервалів передбачення отримали подальший розвиток». Однак формули (3.7) та (3.8) є стандартними рівняннями для інтервалів передбачення лінійної регресії, до яких застосовано зворотне перетворення. Що саме є «подальшим розвитком» порівняно з раніше відомими рівняннями, не визначено.

3. Огляд у розділі 1 зосереджений переважно на роботах тієї самої наукової школи. Сучасні роботи з підходів машинного навчання (ML) до оцінювання розміру ПЗ (neural networks, random forest, gradient boosting) згадані лише побіжно в розділі 1.4 як альтернативи, яким надано перевагу регресії. Проте конкретне порівняння точності запропонованих моделей з ML-підходами на однакових даних відсутнє.

4. Вибірка Java-застосунків складається лише з 112 проєктів (74 навчальних і 38 тестових). Для п'ятифакторної нелінійної регресійної моделі це критично мало. Після видалення викидів (9 точок із 74 навчальних) навчальна вибірка скорочується до 65 проєктів. Слід було провести аналіз потужності (power analysis), щоб обґрунтувати достатність такого розміру вибірки для надійної оцінки 5 параметрів регресії, а також параметрів перетворення Бокса-Кокса.

5. Метод ітеративного видалення викидів за відстанню Махаланобіса потенційно призводить до надмірної підгонки на навчальній вибірці: зі 74 точок навчальної вибірки видалено 9 (12,2%) під час побудови моделі (3.17). Значна різниця між R^2 на навчальній (0,9901) та тестовій (0,9456) вибірках підтверджує цей ризик. Необхідно було виконати аналіз ступеня впливу видалення кожного викиду на параметри моделі.

6. Вибірка сформована виключно з GitHub-репозиторіїв з відкритим кодом. У дисертації відсутнє обговорення потенційних відмінностей: відкриті проєкти можуть системно відрізнятися від комерційних за структурою коду, розміром, кількістю інтерфейсів тощо. Це безпосередньо обмежує зовнішню валідність моделей для промислового застосування.

7. Дисертація обсягом 223 сторінки містить лише 5 рисунків при 44 таблицях. Складні багатовимірні процеси (алгоритм побудови моделей, архітектура SaaS-системи) були б наочнішими у вигляді діаграм. Зокрема, загальний алгоритм побудови нелінійних регресійних моделей (розділ 3.1) викладено лише в текстовій формі.

8. Ряд тверджень у тексті підтверджується виключно власними конференційними матеріалами автора (джерела [27]–[32]), частина з яких є тезами обсягом 3–4 сторінки. Для фундаментальних теоретичних тверджень слід посилатися на незалежні рецензовані джерела.

Незважаючи на виявлені недоліки, дисертація відповідає вимогам до наукових досліджень, а отримані результати мають наукову новизну та практичну значущість.

Висновок

Дисертаційна робота Орехова Олександра Станіславовича є логічно завершеним, самостійно виконаним дослідженням, яке присвячене підвищенню достовірності обробки інформації з метрик коду Java-застосунків для оцінювання їх розміру за рахунок побудови багатфакторних лінійних регресійних моделей. У роботі детально проаналізовано існуючі математичні моделі та методи, обґрунтовано необхідність підвищення достовірності обробки інформації з негаусівськими даними метрик коду, а також розроблено нелінійні регресійні

моделі для підвищення достовірності та якості обробки інформації для оцінювання параметра розміру Java-застосунків.

Зміст дисертації повністю відповідає поставленим завданням і меті дослідження. Представлені наукові положення є новими та підтвердженими відповідними експериментами. Отримані результати узгоджуються із сучасними підходами до математичного моделювання та статистичного аналізу. За кількістю та змістом наукові публікації здобувача відповідають вимогам МОН України щодо присудження ступеня доктора філософії. Текст дисертації написаний науковим стилем, логічно структурований, містить необхідну кількість висновків і наукову новизну. Кожен розділ має обґрунтовані та логічно побудовані висновки.

Таким чином, на підставі викладеного можна зробити висновок, що дисертаційне дослідження Орехова Олександра Станіславовича 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 Постановою Кабінету Міністрів України № 44 від 12.01.2022 р. «Порядок 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».

Офіційний опонент

професор кафедри інженерії
програмного забезпечення
Національного університету
«Одеська політехніка»,
доктор технічних наук, професор

Віра ЛЮБЧЕНКО

Вчений секретар Вченої ради
Національного університету
«Одеська політехніка»



Лада ПРОКОПОВИЧ