

ситуації, яка може статися. Це може бути, наприклад, відмова обладнання, зависання важливих сервісів, вірусні і хакерські атаки, масштабні природні і техногенні аварії тощо. Організація процесів аварійного відновлення функціонування КС повинна бути стратегією всієї компанії, а не тільки власника КС (системного адміністратора). Керівники всіх підрозділів, діяльність яких залежить від послуг, що надаються КС, розробляють процедури дій у надзвичайних обставинах, а також беруть участь у розробці Плану. Кожний функціональний підрозділ «запускає» частину Плану в рамках загальних зусиль з відновлення діяльності КС. Важливим моментом у процесі складання Плану є участь вищого керівництва компанії, відповідального за надання ресурсів і визначення балансу між інвестиціями в інформаційні технології і часом простою втратою даних у разі збою. Обов'язком керівництва підприємства є (в тому числі власників процесів, власника Системи і департаменту Якості) гарантування, що відповідний документ «План забезпечення безперервності бізнесу та аварійного відновлення», розроблений і впроваджений, буде періодично перевірятися і відразу після початку впровадження його вимогам буде слідувати персонал структурних підрозділів і аналізувати відповідність вимог Плану корпоративній стратегії компанії і нормативним вимогам.

У Плані має бути відображено навіть виконання таких забезпечуючих функцій, як утримання будівель і устаткування. Ефективність виконання цих функцій може не впливати безпосередньо на роботу ІТ-інфраструктури, проте від неї буде частково залежати відновлення роботи технічних засобів.

Документ «План забезпечення безперервності бізнесу та аварійного відновлення» є не тільки технічним планом – він передбачає проведення організаційних заходів. Тому в основу Плану покладені відомості про структуру та функції компанії, кошти, необхідні для підтримки її діяльності, величину збитку від неможливості нормального функціонування, персонал, який прийме на себе управління в кризовій ситуації, і процедури, які будуть використовувати.

План включає докладні інструкції з переключення на резервні канали зв'язку, використання резервних потужностей, резервування та відновлення інформації, необхідної для роботи, перелік та методика дій з ліквідації наслідків надзвичайної ситуації, а також план тренувальних і тестувальних заходів для персоналу та керівництва.

Аналіз літературних джерел та практичні підходи дозволили автору сформулювати основні розділи Плану та короткий їх зміст.

Першим розділом має бути вступ, в якому викладено мету, завдання, основні положення документу. Формулювання повинні бути чіткими і конкретними.

Другий розділ – це оцінка надзвичайної / аварійної ситуації. Цей розділ описує чинники ризику і заплановані заходи щодо їх контролю, а саме: опис несприятливих подій і небезпеки (кризові ситуації), які можуть негативно вплинути на функціонування КС. Тут доцільним буде провести класифікацію можливих несприятливих подій і небезпек (наприклад, загрозлива, серйозна, що вимагають уваги ситуації), оцінку ймовірності їх виникнення. Коротко описати перелік критеріїв для визначення несприятливих подій і небезпек, а також їх потенційні джерела. Сценарії кризових ситуацій, визначення можливого збитку (у разі потреби). Необхідно провести й описати оцінку остаточних ризиків.

Третій розділ – діяльність у кризовій ситуації. Розділ описує необхідні ресурси і умови для аварійного відновлення. Під час його формування необхідно описати пріоритетні бізнес-процеси компанії, скласти перелік внутрішніх і зовнішніх ресурсів (наприклад, технічні засоби, програмне забезпечення, засоби зв'язку, документи); внести облікову інформацію про технічне, програмне та інше забезпечення, що необхідне для відновлення бізнесу компанії в разі кризової ситуації, контактну інформацію (наприклад, список осіб, яких необхідно сповістити про кризову ситуацію із зазначенням адрес і телефонів). У разі дії SLA-угоди (угоди про рівень надання послуги постачальником КС) – вказати координати зовнішнього підрядника.

Четвертий розділ – організаційні ресурси. Необхідно визначити необхідні ресурси та умови для аварійного відновлення.

П'ятий розділ – технічне забезпечення. Розділ описує заходи щодо створення і підтримки бази технічних засобів, що забезпечує безперебійну діяльність КС у надзвичайній / кризовій ситуації.

Шостий розділ – тестування Плану. Необхідно описати порядок проведення випробувань плану, при цьому вказати, наприклад:

- що підлягає випробуванням під час перевірки реалізації Плану;
- хто повинен проводити випробування;
- коли повинні здійснюватися випробування;
- результати випробувань.

План тестування може бути загальним або складатися з окремих частин. Деякі розділи цього плану можуть бути розділами інших загальних планів.

В останньому розділі з робочою назвою «Порядок перегляду плану» необхідно вказати терміни перегляду плану (наприклад, плановий, частковий, повний).

План повинен бути створений для кожної впровадженої КС у компанії і повинен охоплювати не тільки процес відновлення системи, але і будь-якої інфраструктури, необхідної для функціонування системи.

На думку автора, створення Плану не підміняє собою систему резервного копіювання, він лише надає їй додатковий функціонал у частині критично важливих інформаційних ресурсів і цілий комплекс процедур, необхідних для того, щоб забезпечити постійний доступ до інформаційних ресурсів компанії.

Висновки. З огляду на викладене забезпечення безперервності бізнесу є важливим фактором ефективного функціонування фармацевтичної компанії. Разом з тим не всі аспекти цієї проблеми знайшли досить повне вивчення. Зокрема, є необхідність пошуку більш ефективних управлінських рішень з забезпечення безперервності бізнесу, що включають: облік нових викликів і ІТ-загроз, підвищення обізнаності персоналу, сертифікації планів безперервності бізнесу та впровадження системи менеджменту безперервності бізнесу, що дасть можливість відповідати корпоративним вимогам і враховувати специфічні ризики фармацевтичної галузі. Дані тенденції приводять до появи нових підходів в управлінні, пов'язаних з забезпеченням безперервності діяльності ІТ-залежного бізнесу.

Список літератури:

1. Фабрис Ж. Валідація комп'ютеризованих систем в контексті GMP. *Аптека*. 2012. № 30 (851). С. 10–11.
2. Зборовська Т.В. Обґрунтування актуальності впровадження стандарту ISO 22301 у фармацевтичному секторі України. *Управління, економіка та забезпечення якості в фармації*. 2017. № 2 (50). С. 4–10.
3. Борсуковська В.Ю. Борсуковський Ю.В. Безперервність бізнесу: новий тренд або необхідність. *Економіка. Менеджмент. Бізнес*. 2017. № 2 (20). С. 48–53.
4. СТ-Н МОЗУ 42-4.0:2016. Лікарські засоби. Належна виробнича практика. Київ, 2016. 358 с.
5. GAMP 5. Good Automated Manufacturing Practice. International Society for Pharmaceutical Engineering. 2011. 196 p.
6. Портянко Т.М., Кучеренко В.С. Особливості управління ризиками при впровадженні комп'ютеризованих систем на фармацевтичних підприємствах. *Вчені записки ТНУ імені В.І. Вернадського. Серія: технічні науки*. 2018. № 6. Том 29 (68) Ч. 1. С. 228–333.
7. СТ-Н МОЗУ 42-4.2:2011 (ICH Q9). Лікарські засоби. Управління ризиками для якості. Київ, 2011. 36 с.
8. ISO 31000:2009. Risk management. Principles and guidelines. International Organization for Standardization. 2018. 32 p.

Portianko T.M., Kucherenko V.S. THE ROLE OF THE BUSINESS CONTINUITY PROTECTION PLAN IN THE IMPLEMENTATION OF COMPUTERIZED SYSTEMS AT PHARMACEUTICAL ENTERPRISES

Ensuring business continuity is one of the most important strategic directions for the development of any company. In accordance with the requirements of the documentation regulating the introduction of computerized systems (hereinafter, CP) at pharmaceutical enterprises, ensuring business continuity is due to the need to maintain the stability and stability of CP operation in various conditions of adverse influence of external and internal factors of technogenic and/or natural character on the business processes of the company. Patient safety, product quality and data integrity should not be adversely affected by malfunctions or breakdowns during any CP operation. Business Continuity Management includes disaster recovery, which covers the technical upgrading of the COP and contains the procedures for staff to restore critical IT services. All alarm / emergency operations must be carefully documented and approved. In order to implement the aforementioned, one shall create a document (plan) describing the steps to be taken in the event of a partial or total failure of the CP. The article describes practical approaches to creating a Business Continuity and Disaster Recovery Plan (hereinafter, the Plan), taking into account the specifics and regulatory requirements for pharmaceutical manufacturing. Measures that can range from scheduled hardware reductions to transition to the system with paper information carriers, and the like, are displayed. The article highlights how the Plan defines how companies' business processes continue to function and process data after a malfunction in the CP. Particular attention is given to sections of the Plan, which outline the steps required to restore business processes after a breach and, where appropriate, how the data obtained during a breach, the roles and responsibilities of staff, the process of information transfer, and how to define a scheme for launching the Plan into action. Its necessity is identified and substantiated where, in order to continue the business processes, an interactive input of information is used, as any electronic records or data will be synchronized immediately after the restoration of the CP. The article focuses on contingency planning, including regular testing (trial), evaluating the effectiveness of the Plan to confirm its performance under changing conditions, and its correction.

Key words: threats, emergencies, business processes, patient safety, business continuity plan, computerized system.

УДК 004.412:519.237.5
DOI

Приходько С.Б.

Національний університет кораблебудування імені адмірала Макарова

Приходько Н.В.

Національний університет кораблебудування імені адмірала Макарова

Смикодуб Т.Г.

Національний університет кораблебудування імені адмірала Макарова

ЧОТИРЬОХФАКТОРНА НЕЛІНІЙНА РЕГРЕСІЙНА МОДЕЛЬ ДЛЯ ОЦІНЮВАННЯ РОЗМІРУ JAVA-ЗАСТОСУНКІВ З ВІДКРИТИМ КОДОМ

Метою роботи є створення множинної нелінійної регресійної моделі для оцінювання розміру Java-застосунків з відкритим кодом на основі багатовимірного нормалізуючого перетворення за значеннями змінних, що можуть бути визначені за діаграмою класів. Чотирьохфакторну нелінійну регресійну модель для оцінювання розміру Java-застосунків з відкритим кодом побудовано на основі нормалізації за допомогою п'ятивимірного перетворення Джонсона для сімейства S_B негаусівського набору даних: кількості строк коду (LOC); кількості класів (Classes); кількості статичних методів (NOSM); метрики, що характеризує відсутність згуртованості методів (Lack of Cohesion of Methods, LCOM), та кількості викликів унікального методу в класі (the Response for Class, RFC) з 38 застосунків, розташованих на сайті GitHub (<https://github.com>) за допомогою інструменту СК (<https://github.com/tauricioapiche/ck>). Також нормалізацію цього набору даних було здійснено і за допомогою двох одновимірних перетворень: у вигляді десяткового логарифму та перетворення Джонсона для сімейства S_B . Використання п'ятивимірного перетворення порівняно з одновимірними дозволяє врахувати кореляцію між змінними, що призводить до покращення нормалізації даних, яка пов'язана з виконанням статистичної гіпотези щодо відповідності їх розподілу п'ятивимірному розподілу Гаусу, з подальшим підвищенням достовірності відповідного оцінювання. Виконано порівняння побудованої нелінійної моделі з лінійною регресійною моделлю і нелінійними регресійними моделями на основі десяткового логарифму і одновимірного перетворення Джонсона. Нелінійна модель, що побудована, порівняно з іншими регресійними моделями (як лінійними, так і нелінійними) має більші значення множинного коефіцієнту детермінації та відсотка прогнозування на рівні величини відносної похибки, який дорівнює 0,25, менші значення середньої величини відносної похибки та ширини інтервалу передбачення нелінійної регресії. Цей результат може бути пояснений найкращою багатовимірною нормалізацією і тим, що немає підстав відкидати нульову гіпотезу про те, що п'ятивимірний розподіл для нормалізованих даних, який нормалізується за допомогою п'ятивимірного перетворення Джонсона для сімейства S_B , є таким самим, як і п'ятивимірний нормальний розподіл.

Ключові слова: нелінійна регресійна модель, інтервал передбачення, оцінювання розміру програми, Java-застосунок, нормалізуюче перетворення, негаусівські дані.

Постановка проблеми. Сьогодні найпопулярнішою мовою програмування у світі [1] та серед програмістів в Україні [2] є Java, вперше випущена Sun Microsystems у 1995 році (<https://www.java.com>). Зараз Java використовується практично скрізь – від ноутбуків до центрів обробки даних, ігрових консолей до суперкомп'ютерів, мобільних телефонів до Інтернету.

Задача оцінювання розміру Java-застосунків з відкритим кодом, як і іншого програмного забезпечення (ПЗ) на ранній стадії розробки, є важливою, оскільки ця інформація використовується для прогнозування трудомісткості створення ПЗ

за допомогою такої відомої моделі, як СОСОМО II [3]. Це потребує відповідних моделей для оцінювання розміру ПЗ, включаючи Java-застосунки з відкритим кодом.

Аналіз останніх досліджень і публікацій. Натепер для оцінювання кількості строк коду інформаційних Java-систем з відкритим кодом існують як лінійні, так і нелінійні регресійні рівняння та моделі залежно від трьох метрик концептуальної моделі даних у вигляді діаграми класів [4–7]. В [4; 5] відповідне лінійне рівняння побудовано на основі методів множинного лінійного регресійного аналізу. Але, як відомо, під час

побудови лінійних регресійних моделей необхідно виконання певних умов, зокрема, похибки повинні бути розподілені за нормальним законом, що має місце лише в поодиноких випадках. А це веде до необхідності побудови нелінійних регресійних моделей, у тому числі і для оцінювання розміру ПЗ, та застосування певних методів множинного нелінійного регресійного аналізу [8].

Тому для оцінювання розміру інформаційних Java-систем з відкритим кодом в [7] було запропоновано рівняння нелінійної регресії, а в [6] – нелінійна регресійна модель. Запропоновані нелінійні регресійні рівняння та модель побудовано за допомогою множинного нелінійного регресійного аналізу із застосуванням чотиримірного перетворення Джонсона сім’ї S_B на основі таких же трьох метрик діаграми класів, що і в [4; 5]: загальна кількість класів, загальна кількість зв’язків та середня кількість атрибутів на клас. Але для Java-застосунків з відкритим кодом, що не є інформаційними системами, таких як різноманітні інструменти і фреймворки, регресійні моделі можуть залежати від інших метрик.

Зазвичай для побудови моделей нелінійної регресії використовують одновимірні нормалізуючі перетворення [9–13]. Але їх застосування для побудови нелінійних регресійних моделей не завжди призводить до задовільних результатів прогнозування, насамперед за такими стандартними показниками, як середня величина відносної похибки та відсоток передбачення [6–8]. Також нелінійні регресійні моделі, що побудовані за допомогою одновимірних нормалізуючих перетворень, зазвичай мають більші ширини довірчих інтервалів та інтервалів передбачення. Це призводить до необхідності використання багатовимірних нормалізуючих перетворень під час побудови нелінійних регресійних моделей, у тому числі і для оцінювання розміру Java-застосунків з відкритим кодом.

Постановка завдання. Метою статті є побудова чотирьохфакторної моделі нелінійної регресії та визначення нижньої і верхньої границь її інтервалів передбачення для оцінювання розміру Java-застосунків з відкритим кодом залежно від: кількості класів (Classes); кількості статичних методів (NOSM); метрики, що характеризує відсутність згуртованості методів (LCOM), та кількості викликів унікального методу в класі (RFC) на основі п’ятивимірного нормалізуючого перетворення Джонсона. Це дозволить підвищити достовірність оцінювання залежної змінної нелінійної регресії порівняно з лінійними моделями та нелі-

нійними моделями з використанням одновимірних нормалізуючих перетворень.

Виклад основного матеріалу дослідження.

Для досягнення цілі статті, що сформульована вище, ми скористалися методами, наведеними в [8, с. 100–102]. Згідно з [8, с. 100] спочатку виконується нормалізація багатовимірних негаусових даних за багатовимірним нормалізуючим перетворенням. Для побудови нелінійної регресійної моделі для оцінювання розміру Java-застосунків з відкритим кодом були зібрані дані з метрик 38 програм, розташованих на сайті GitHub (<https://github.com>): фактична кількість строк коду (в тисячах рядків коду) Y ; кількість класів (Classes) X_1 ; кількість статичних методів (NOSM) X_2 ; метрика, що характеризує відсутність згуртованості методів (LCOM) X_3 , та кількість викликів унікального методу в класі (RFC) X_4 . Ці дані були отримані за допомогою інструменту СК (<https://github.com/mauricioaniche/ck>) та наведені у табл. 1. Вибір саме цих метрик був зумовлений практичною відсутністю мультиколінеарності між ними. Наявність мультиколінеарності свідчить про те, що в множинній лінійній регресійній моделі два або більше факторів пов’язані між собою або мають високий ступінь кореляції [14].

Наявність або відсутність мультиколінеарності ми визначали за коефіцієнтами впливу дисперсії (VIFs) серед майбутніх факторів у моделі множинної лінійної регресії. Для множинної лінійної регресійної моделі з k -факторами X_i , $i = 1, 2, \dots, k$, VIFs – це діагональні елементи оберненої коваріаційної $k \times k$ матриці [15]. Значення VIFs більше за 10 часто сприймаються як сигнал, що дані мають проблеми з мультиколінеарністю. У разі, якщо значення VIFs знаходяться у межах від 1 до 5, то мультиколінеарності немає. Для X_1 , X_2 , X_3 та X_4 значення VIFs відповідно дорівнюють 4,79, 3,10, 1,12 та 5,63, що свідчить про практичну відсутність мультиколінеарності між цими факторами.

Згідно з [16], п’ятивимірні дані для змінних Y , X_1 , X_2 , X_3 та X_4 , що наведені в табл. 1, мають негаусівський розподіл, оскільки для п’яти застосунків (2, 10, 12, 37 та 39) значення квадрату відстані Махаланобіса MD^2 , які, відповідно, дорівнюють 30,17, 20,21, 34,02, 28,28 та 21,47, є більшими, ніж величина квантіля розподілу χ^2 , що становить 16,75 для рівня значущості 0,005. Значення MD^2 , що є більшими за 16,75, в табл. 1 виділені напівжирним шрифтом. Також про негаусівський розподіл п’ятивимірних даних для змінних Y , X_1 , X_2 , X_3 та X_4 з табл. 1 свідчить оцінка багатовимірною ексцесу β_2 , яка визначалася за [16].

Відомо, що для m -вимірною нормального розподілу $\beta_2 = m(m + 2)$. У нашому випадку $\beta_2 = 35$. Для цих п'ятивимірних даних оцінка β_2 дорівнює 100,77, що майже у 3 рази перевищує теоретичне значення.

Також у табл. 1 наведені значення нормалізованих змінних з метрик Java-застосунків з відкритим кодом, які були отримані за допомогою п'ятивимірною перетворення Джонсона сімейства S_B , компоненти якого визначаються, як і в [8]:

$$Z_j = \gamma_j + \eta_j \ln \frac{X_j - \phi_j}{\phi_j + \lambda_j - X_j}, \quad (1)$$

де γ_j, η_j, ϕ_j та λ_j – параметри перетворення Джонсона, $j=1,2,3,4, j=1,2,3,4$.

Значення нормалізованої залежної змінної Z_Y також визначається за (1) з тою різницею, що в (1) замість $Z_j, X_j, \gamma_j, \eta_j, \phi_j$ та λ_j потрібно підставити, відповідно, $Z_Y, Y, \gamma_Y, \eta_Y, \lambda_Y$ та λ_Y .

Для даних з табл. 1 оцінки параметрів п'ятивимірною перетворення Джонсона сімейства S_B такі: $\hat{\gamma}_Y = 1,39056, \hat{\gamma}_1 = 1,0380, \hat{\gamma}_2 = 1,18671, \hat{\gamma}_3 = 2,44690, \hat{\gamma}_4 = 1,82065, \hat{\eta}_Y = 0,455264, \hat{\eta}_1 = 0,373337, \hat{\eta}_2 = 0,388760, \hat{\eta}_3 = 0,352541, \hat{\eta}_4 = 0,519554, \hat{\phi}_Y = 0,875, \hat{\phi}_1 = 13,500, \hat{\phi}_2 = 137,252, \hat{\phi}_3 = 137,252, \hat{\phi}_4 = 77,9181, \hat{\lambda}_Y = 868,212, \hat{\lambda}_1 = 4276,548, \hat{\lambda}_2 = 4627,30, \hat{\lambda}_3 = 26483781,6$ і $\hat{\lambda}_4 = 160289,7$.

Далі для нормалізованих даних будемо лінійну регресійну модель

$$Z_Y = \hat{Z}_Y + \varepsilon = \hat{b}_0 + \hat{b}_1 Z_1 + \hat{b}_2 Z_2 + \hat{b}_3 Z_3 + \hat{b}_4 Z_4 + \varepsilon, \quad (2)$$

де ε – випадкова величина з розподілом Гаусу, $\varepsilon \sim N(0, \sigma_\varepsilon^2)$; оцінки параметрів для даних з табл. 1 є такими: $\hat{b}_0 = 0, \hat{b}_1 = 0,094587, \hat{b}_2 = 0,178654, \hat{b}_3 = 0,090668, \hat{b}_4 = 0,654896$. Оцінки параметрів моделі (2) визначалися за методом найменших квадратів. За даними з табл. 1, сума квадратів відхилень для моделі (2) склала 1,0378.

Потім будемо нелінійну регресійну модель за [8, с. 101]

$$Y = \hat{\phi}_Y + \hat{\lambda}_Y \left[1 + e^{-(\hat{Z}_Y + \varepsilon - \hat{\gamma}_Y) / \hat{\eta}_Y} \right]^{-1}, \quad (3)$$

де відповідні складники визначаються за (2). За даними з табл. 1, сума квадратів відхилень для моделі (3) склала 25 414,8.

Побудована модель (3) була перевірена за множинним коефіцієнтом детермінації R^2 , середньою величиною відносної помилки MMRE і відсотком прогнозованих результатів, для яких величини відносної помилки MRE менші за 0,25, PRED(0,25).

Ці показники зазвичай використовуються для оцінювання якості прогнозування за допомогою регресійних моделей і в інженерії програмного забезпечення [17; 18]. Допустимі значення MMRE і PRED(0,25) складають не більше 0,25 і не менше 0,75 відповідно. Допустиме значення R^2 приблизно таке ж, як для PRED(0,25).

Для моделі (3), що була побудована за даними з табл. 1 на основі п'ятивимірною перетворення Джонсона сімейства S_B , значення R^2 , MMRE та PRED(0,25) мають, відповідно, такі значення: 0,971, 0,173 та 0,789, що вказує на добру її якість стосовно оцінювання розміру Java-застосунків з відкритим кодом.

Для порівняння моделі (3) для оцінювання розміру Java-застосунків з відкритим кодом в подальшому було також побудовано лінійну та нелінійну регресійні моделі на основі одновимірною Джонсона. Для цього ми також використовуємо значення змінних Y, X_1, X_2, X_3 та X_4 , що наведені в табл. 1.

Лінійна регресійна модель для оцінювання розміру Java-застосунків з відкритим кодом має вигляд

$$Y = \hat{b}_0 + \hat{b}_1 X_1 + \hat{b}_2 X_2 + \hat{b}_3 X_3 + \hat{b}_4 X_4 + \varepsilon, \quad (4)$$

де оцінки параметрів такі: $\hat{b}_0 = 5,9139, \hat{b}_1 = 0,015127, \hat{b}_2 = 0,015127, \hat{b}_3 = 6,0916 \cdot 10^{-6}, \hat{b}_4 = 0,0039244$.

Сума квадратів відхилень для моделі (4) склала 26 204,7, що всього на 3,1% більше за цю суму для моделі (3). Також невелика різниця між значеннями R^2 для моделей (3) і (4): всього 0,1%. Значення двох інших показників лінійної моделі (4) – MMRE і PRED(0,25), що дорівнюють 0,744 і 0,526 відповідно, вказують на незадовільну її якість з оцінками параметрів, що були отримані за значеннями змінних Y, X_1, X_2, X_3 та X_4 з табл. 1.

Перевірку нульової гіпотези про нормальність закону розподілу випадкової величини ε для моделі (4) здійснюємо за критерієм Пірсона. Для вибірки значень випадкової величини ε значення χ^2 , яке дорівнює 81,32, більше за $\chi_{\alpha, \nu}^2$, що становить 7,81 для 3 ступенів вільності та 0,05 рівня значущості. Тобто цю гіпотезу про нормальність розподілу випадкової величини ε потрібно відкинути. Це свідчить взагалі про відсутність теоретичного обґрунтування використання моделі лінійної регресії (4) і призводить до необхідності застосування нелінійної регресійної моделі для оцінювання розміру Java-застосунків з відкритим кодом.

Нелінійна регресійна модель для оцінювання розміру Java-застосунків з відкритим кодом на

основі одновимірного перетворення Джонсона сімейства S_B також має вигляд (3), але оцінки параметрів, що отримані за значеннями змінних Y, X_1, X_2, X_3 та X_4 з табл. 1, дещо інші: $\hat{\gamma}_Y = 1,39056$, $\hat{\gamma}_1 = 1,0380$, $\hat{\gamma}_2 = 1,18671$, $\hat{\gamma}_3 = 1,17761$, $\hat{\gamma}_4 = 1,66557$, $\hat{\eta}_Y = 0,455264$, $\hat{\eta}_1 = 0,373337$, $\hat{\eta}_2 = 0,388760$, $\hat{\eta}_3 = 0,209433$, $\hat{\eta}_4 = 0,468450$, $\hat{\phi}_Y = 0,875$, $\hat{\phi}_1 = 13,500$, $\hat{\phi}_2 = 2,500$, $\hat{\phi}_3 = 174,50$, $\hat{\phi}_4 = 139,989$, $\hat{\lambda}_Y = 868,212$, $\hat{\lambda}_1 = 4276,548$, $\hat{\lambda}_2 = 4627,30$, $\hat{\lambda}_4 = 153563,5$ і $\hat{\lambda}_4 = 153563,5$. Також інші і оцінки параметрів моделі (2): $\hat{b}_0 = 0$, $\hat{b}_1 = 0,153713$, $\hat{b}_2 = 0,228632$, $\hat{b}_3 = 0,0670969$, $\hat{b}_4 = 0,586594$.

Сума квадратів відхилень для моделі (3) з оцінками параметрів, що отримані на основі одновимірного перетворення Джонсона сімейства S_B , склала 28 398,6, що на 11,7% більше за цю суму для моделі (3) з оцінками параметрів, що отримані на основі п'ятивимірного перетворення Джонсона сімейства S_B . Також гірші значення для моделі (3) з оцінками параметрів, що отримані на основі одновимірного перетворення Джонсона сімейства S_B , мають три інші показники – R^2 , MMRE і PRED(0,25), що дорівнюють 0,968, 0,176 і 0,711 відповідно. Хоча слід зазначити, що значення MMRE і PRED(0,25) значно кращі, ніж у лінійної моделі (4).

Для визначення нижньої і верхньої границь інтервалів передбачення нелінійних регресій ми використовували відповідний метод, запропоно-

ваний у [8, с. 101]. Нижні (LB) і верхні границі (UB) інтервалів передбачення нелінійних регресій для моделі (3), як для одновимірного, так і для п'ятивимірного перетворення Джонсона сімейства S_B , наведені в табл. 2.

Для моделі (3) з оцінками параметрів, що були отримані за даними з табл. 1 з 38 Java-застосунків на основі п'ятивимірного перетворення Джонсона сімейства S_B , ширини інтервалу передбачення нелінійної регресії менші для 32 рядків даних (всі, окрім рядків 5, 7, 8, 26, 29 та 36) порівняно з одновимірним перетворенням Джонсона. Причому різниця у ширині інтервалу передбачення для даних з рядка 2 складає 58%.

Кращі показники оцінювання розміру Java-застосунків з відкритим кодом за моделлю нелінійної регресії на основі п'ятивимірного нормалізуючого перетворення Джонсона сімейства S_B можна передусім пояснити кращою багатовимірною нормалізацією, яка перевірялася за відомими критеріями [15]. Так, якщо за критерієм на основі квадрата відстані Махаланобіса гіпотеза про нормальність багатовимірного закону розподілу нормалізованих за допомогою п'ятивимірного нормалізуючого перетворення Джонсона сімейства S_B даних для 38 застосунків з табл. 1 приймається для рівня значущості 0,0001, то у разі застосування одновимірного перетворення та без нього – відкидається.

Таблиця 2

Межі інтервалів передбачення нелінійних регресій

№	Одновимірне		П'ятивимірне		№	Одновимірне		П'ятивимірне	
	LB	UB	LB	UB		LB	UB	LB	UB
1	1,20	3,12	1,24	2,99	20	1,52	5,08	1,55	4,57
2	45,99	316,74	41,88	212,74	21	1,13	2,78	1,15	2,57
3	14,75	79,99	13,12	63,11	22	11,42	59,01	10,96	49,27
4	78,51	319,04	91,23	322,20	23	27,24	135,89	28,78	125,71
5	126,11	441,94	162,79	480,73	24	13,18	68,27	12,43	56,43
6	34,58	167,74	36,16	153,74	25	93,53	366,15	114,60	384,17
7	54,64	250,17	68,06	267,52	26	72,33	311,51	88,90	329,42
8	0,98	1,62	1,09	2,09	27	1,78	6,49	1,67	5,17
9	5,39	27,67	5,39	23,72	28	61,97	271,63	71,82	274,32
10	64,20	283,07	66,32	260,91	29	125,62	436,82	154,86	471,30
11	59,02	260,37	66,05	255,55	30	33,39	161,86	34,43	146,74
12	743,90	849,30	748,40	847,83	31	13,04	68,27	12,65	58,63
13	12,75	65,93	12,74	57,90	32	12,83	67,57	12,35	56,64
14	97,62	371,75	107,86	363,15	33	6,23	31,73	6,42	28,24
15	22,21	112,77	21,81	97,25	34	12,59	65,26	12,67	57,00
16	10,12	52,56	10,04	45,71	35	30,23	148,45	32,53	139,25
17	1,34	4,39	1,35	3,91	36	43,17	212,42	51,75	226,62
18	3,52	18,84	3,89	18,04	37	72,18	328,61	72,31	296,22
19	2,34	10,93	2,37	9,49	38	208,17	587,82	232,28	581,96

Висновки. Удосконалено чотирьохфакторну модель нелінійної регресії та визначення нижньої і верхньої границь її інтервалів передбачення для оцінювання розміру Java-застосунків з відкритим кодом залежно від кількості класів; кількості статичних методів; метрики, що характеризує відсутність згуртованості методів, та кількості викликів унікального методу в класі на основі п'ятивимірного нормалізуючого перетворення Джонсона сімейства S_B . Це дозволяє підвищити достовірність

оцінювання залежної змінної нелінійної регресії порівняно з використанням одновимірних нормалізуючих перетворень. Модель, що побудовано, порівняно з іншими регресійними моделями має більший відсоток прогнозування, менші середні величини відносної похибки та ширини інтервалу передбачення нелінійної регресії. В майбутньому планується використання інших наборів даних для побудови нелінійної регресійної моделі для оцінювання розміру Java-застосунків з відкритим кодом.

Список літератури:

1. TIOBE Index for April 2020. URL: <https://www.tiobe.com/tiobe-index/> (дата звернення: 03.04.2020)
2. Рейтинг мов програмування 2019: JavaScript майже зрівнялася з Java, популярність Go знижується URL: <https://dou.ua/lenta/articles/language-rating-jan-2019/> (дата звернення: 03.04.2020)
3. Boehm B.W., Abts C., Brown A.W., Chulani S., Clark B.K., Horowitz E., Madachy R., Reifer D.J., Steece B. Software Cost Estimation with COCOMO II. Upper Saddle River, NJ : Prentice Hall PTR, 2000. 544 p.
4. Tan H.B.K., Zhao Y., Zhang H. Estimating LOC for information systems from their conceptual data models. *Proceedings of the 28th International Conference on Software Engineering (ICSE '06)*. (May 20-28, 2006, Shanghai, China). Shanghai, 2006. P. 321–330.
5. Tan H.B.K., Zhao Y., Zhang H. Conceptual data model-based software size estimation for information systems. *Transactions on Software Engineering and Methodology*. 2009. Vol. 19. Issue 2. October 2009. Article No. 4.
6. Prykhodko N.V., Prykhodko S.B. The non-linear regression model to estimate the software size of open source Java-based systems. *Radio Electronics, Computer Science, Control*. 2018. No. 3 (46). P. 158–166. DOI: 10.15588/1607-3274-2018-3-17.
7. Prykhodko S.B., Prykhodko N.V., Mandra A.V. Building the nonlinear regression equations to estimate the software size of Java-based information systems. *Materials of the VII International scientific-practical conference on Information Control Systems and Technologies*. (17th-18th September, 2018, Odessa). Odessa, Astroprint, 2018. P. 222–224.
8. Prykhodko N.V., Prykhodko S.B. Constructing the non-linear regression models on the basis of multivariate normalizing transformations. *Electronic modeling*. 2018. Vol. 40. No. 6. P. 101-110. DOI: 10.15407/emodel.40.06.101.
9. Bates D.M., Watts D. G. Nonlinear regression analysis and its applications. New York: John Wiley & Sons, 1988. 384 p.
10. Seber G.A.F., Wild C.J. Nonlinear regression. New York : John Wiley & Sons, 1989. 768 p.
11. Ryan T.P. Modern regression methods. 2nd Edition. New York : John Wiley & Sons, 2008. 672 p.
12. Drapper N.R., Smith H. Applied regression analysis. New York : John Wiley & Sons, 1998. 736 p.
13. Johnson R.A., Wichern D.W. Applied multivariate statistical analysis. – Pearson Prentice Hall, 2007. 800 p.
14. Chatterjee S., Price B. Regression analysis by example. New York: John Wiley & Son, 1977. 228 p.
15. Olkin I., Sampson A.R. Multivariate Analysis: Overview. *International encyclopedia of social & behavioral sciences* / N. J. Smelser, P. B. Baltes (eds.) 1st edn. Elsevier, Pergamon, 2001. P. 10240–10247.
16. Mardia K.V. Measures of multivariate skewness and kurtosis with applications. *Biometrika*. 1970. Vol. 57. P. 519–530. DOI: 10.1093/biomet/57.3.519.
17. Foss T., Stensrud E., Kitchenham B., Myrvtveit I. A simulation study of the model evaluation criterion MMRE. *IEEE Transactions on software engineering*. 2003. 11(29). P. 985–995.
18. Port D., Korte M. Comparative studies of the model evaluation criterions MMRE and PRED in software cost estimation research. *Proceedings of the 2nd ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*. ACM, New York, 2008. P. 51–60.

Prykhodko S.B., Prykhodko N.V., Smykodub T.G. FOUR-FACTOR NON-LINEAR REGRESSION MODEL TO ESTIMATE THE SIZE OF OPEN SOURCE JAVA-BASED APPLICATIONS

The goal of the work is the creation of the multiple non-linear regression model for estimating the size of open source Java-based applications based on the multivariate normalizing transformation. A four-factor non-linear regression model to estimate the size of open source Java-based applications is constructed on the basis of the Johnson five-variate normalizing transformation for S_B family of the non-Gaussian data set from 38 applications hosted on GitHub (<https://github.com>). The data set was obtained using the CK tool