УДК 004.412:519.25

ОЦІНЮВАННЯ ТРИВАЛОСТІ РОЗРОБКИ БАНКІВСЬКОГО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ З УРАХУВАННЯМ КРИТЕРІЮ КІЛЬКОСТІ КОМАНД РОЗРОБНИКІВ

Латанська Л.О., к.ф.-м.н, доц.[1]; Половинка А.В.[2]
[1,2] Національний університет кораблебудування імені адмірала Макарова
[1,2] Україна, Миколаїв
[1] liudmyla.latanska@nuos.edu.ua; [2] anton.polovynka@nuos.edu.ua

Анотація. *У даній роботі було побудовано нелінійні регресійні моделі для оцінювання тривалості розробки банківського ПЗ в розрізі кількості команд, що приймають в ній участь, використовуючи нормалізацію за допомогою багатовимірного перетворення Бокса-Кокса та логарифмічного перетворення. Зроблено висновки щодо придатності моделей.*

Ключові слова: *банківське ПЗ; тривалість розробки ПЗ; нелінійна регресійна модель; нормалізація; багатовимірне нормалізуюче перетворення Бокса-Кокса.*

Вступна частина. Питання достовірного оцінювання витрат, необхідних для розробки ПЗ, залишається однією з ключових проблем у процесі створення програмних проєктів. Банківські ІТ-проєкти мають низку специфічних особливостей, які відрізняють їх від проєктів розробки ПЗ в інших галузях. З основних можна виділити складність регуляторних вимог, складність архітектури інформаційних систем та виняткову кількість пов'язаних між собою інформаційних потоків. Також на загальні терміни проєктів впливають високі вимоги до безпеки та більш тривалий процес прийняття рішень у порівнянні зі звичайною комерційною розробкою. Питання достовірного оцінювання витрат, необхідних для розробки банківського ПЗ з урахуванням низки чинників, зокрема трудомісткості робіт та кількості команд розробки, залишається однією з ключових проблем у процесі створення даних проєктів.

Оцінювання часу розробки програмних проєктів може здійснюватися на основі моделей, у яких базовим показником є трудомісткість, таких як COCOMO II [1] та ISBSG [2]. Проте зазначені нелінійні регресійні моделі, побудовані на підставі узагальнених даних для різних типів проєктів і умов, не завжди відповідають реальним даним проєктів розробки банківського ПЗ [3]. Отже, побудова специфічної нелінійної регресійної моделі з метою підвищення достовірності оцінювання тривалості проєктів зі створення банківського ПЗ є актуальним і науково обґрунтованим завданням.

Метою дослідження є побудова нелінійних регресійних моделей для оцінювання тривалості проєктів з розробки банківського програмного забезпечення, із застосуванням нормалізуючих перетворень десяткового логарифму та Бокса-Кокса.

Основна частина. За наявності власних емпіричних даних щодо попередніх розробок, можна побудувати індивідуальну модель для прогнозування тривалості виконання майбутніх проєктів. Оскільки статистичні розподіли тривалості та трудомісткості розробки програмних продуктів, як правило, не відповідають нормальному розподілу [3], доцільним є використання нелінійних регресійних залежностей між зазначеними показниками.

Якість побудованих моделей COCOMO II та ISBSG виявилась недостатньою для їх використання для достовірної оцінки емпіричних даних тривалості та трудомісткості проєктів з розробки банківського ПЗ [3], тож було визначено необхідність побудови

нелінійних регресійних моделей із застосуванням нормалізуючих перетворень, в першу чергу десяткового логарифму та Бокса-Кокса.

Спроба побудови нелінійних регресійних моделей для оцінювання тривалості проєктів з розробки банківського ПЗ за даними всіх проєктів без їх розподілу за певними критеріями, виявилась невдалою, оскільки більшість точок даних у цьому випадку являлися викидами. Тому для оцінювання тривалості розробки у даній роботі буде враховано кількість команд розробників проєктів (однокомандні та багатокомандні проєкти).

В представленій роботі побудовано чотири нелінійні регресійні моделі для оцінювання тривалості розробки проєктів банківського ПЗ. Для побудови моделей усі 482 проєкти були розділені на дві групи в залежності від кількості команд розробників: однокомандні проєкти та проєкти з залученням більш ніж однієї команди. Кожна з цих груп також була розділена на дві виборки: проєкти для побудови моделей (60%) та проєкти для перевірки якості побудованих моделей (40%). При розбитті враховувалось, щоб у вибірку для побудови моделі входили мінімальні та максимальні значення метрик, а у тестовій вибірці були присутні значення, близькі або рівні граничним. Отриманий розподіл проєктів по групам наведено в таблиці 1.

Таблиця 1

Розподіл даних по проєктам розробки банківського ПЗ по групам за критерієм кількості команд

Група проєктів	Загальна кількість проєктів	Кількість проєктів для побудови моделі	Кількість проєктів для перевірки якості
Однокоманді	266	177	89
Багатокомандні	216	144	72

Для нормалізації даних були використані нормалізуюче перетворення за десятковим логарифмом та багатовимірне нормалізуюче перетворення Бокса-Кокса, що дозволило нормалізувати емпіричні дані та поліпшити якість подальшого аналізу, спираючись на аналогічні методики [4].

Спочатку було проведено перевірку всіх значень на наявність викидів за допомогою методу, заснованого на обчисленні квадрата відстані Махаланобіса для нормалізованих даних. На другому кроці було перевірено нормальність розподілу залишків у моделі лінійної регресії для нормалізованих даних. На третьому – виконано пошук значень, які виходять за межі інтервалу прогнозування.

Кількість видалених викидів при використанні багатовимірного перетворення Бокса-Кокса для однокомандних та багатокомандних проєктів становило відповідно 37 та 27, а при логарифмічному перетворенні – 110 та 71 відповідно. Кількість викидів при логарифмічному перетворенні становить більше половини початкових даних, з чого було зроблено висновок про неможливість подальшого використання нелінійної регресійної моделі на основі даного перетворення.

Для перетворення Бокса-Кокса на основі даних по кожній групі проєктів було побудоване нелінійне регресійне рівняння виду:

$$\hat{Y} = \psi_Y^{-1}(\hat{Z}_Y) = \psi_Y^{-1}(b_0 + b_1 * Z_1),$$

де $\hat{Y}$ – оцінка тривалості розробки проєкту, ψ_Y – перша компонента вектору ψ перетворення, Z_1 – вектор нормалізованих значень трудомісткості проєктів, b_0, b_1 – оцінки параметрів моделі, які були отримані за допомогою методу максимальної правдоподібності. Результати оцінювання якості моделей та оцінки параметрів моделей наведені у таблиці 2.

Таблиця 2

Результати оцінювання якості та оцінки параметрів моделей, побудованих для різних груп проєктів з використанням нормалізуючого перетворення Бокса-Кокса

Група проєктів	R^2	*MMRE*	*PRED*(0,25)	b_0	b_1
Однокоманді	0,8762	0,1128	0,9143 (128/140)	1,0515	0,2233
Багатокомандні	0,9654	0,1016	0,9658 (113/117)	2,2450	0,2835

Значення R^2, *MMRE* та *PRED*(0,25) для побудованих моделей свідчать про їх задовільну якість. Тобто дані моделі можна використовувати для оцінювання тривалості розробки банківського ПЗ на основі його трудомісткості.

Висновки. В роботі були побудовані нелінійні регресійні моделі для оцінювання тривалості розробки банківського ПЗ в залежності від трудомісткості розробки з розділенням за критерієм кількості команд, що приймають в ній участь. Моделі на основі нормалізуючого перетворення десяткового логарифму були визнані непридатними по причині великої кількості викидів. Моделі, що були побудовані на основі нормалізації за допомогою багатовимірного перетворення Бокса-Кокса, демонструють задовільну якість оцінювання за показниками R^2, *MMRE* та *PRED*(0,25), але кількість викидів при побудові моделі для однокомандних проєктів склала 20,9% даних, а для багатокомандних – 18,8%.

В подальшому планується для покращення якості розглянути використання інших нормалізуючих перетворень та використання в побудові моделей такого додаткового фактору, як кількість розробників. Моделі можуть бути використані для прогнозування тривалості проектів, що розробляються в банківській сфері.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Boehm, B.W. et al. (2000). *Software Cost Estimation with COCOMO II.* Upper Saddle River, NJ: Prentice Hall, 2000. 544 p. ISBN 978-0137025763.

[2] International Software Benchmarking Standards Group. (n.d.). URL: https://www.isbsg.org.

[3] Latanska, L., & Polovynka, A. (2025). Analysis of methods and models for estimating the duration of banking software development projects. *Herald of Khmelnytskyi national university. Technical sciences.* Issue 3, part 2, 2025 (353). P.140-145. https://doi.org/10.31891/2307-5732-2025-353-18

[4] Prykhodko, N.V., & Prykhodko, S.B. (2018). Constructing the Nonlinear Regression Models on the Basis of Multivariate Normalizing Transformations. *Electronic Modeling.* 2018. V.40. №6. P. 99-108. https://doi.org/10.15407/emodel.40.06.101.

Latanska L.O., Polovynka A.V.

Estimation of Banking Software Development Duration Considering the Criterion of Developer Team Count

Abstract. *This study develops nonlinear regression models for estimating the duration of banking software development, taking into account the number of teams involved in the process. The modelling approach employs normalization using a multivariate Box–Cox transformation and logarithmic transformation. Conclusions are drawn regarding the suitability and applicability of the proposed models.*

Keywords: *banking software; software development duration; nonlinear regression model; normalization; multivariate Box–Cox transformation.*