УДК 004.412:519.237.5

НЕЛІНІЙНА РЕГРЕСІЙНА МОДЕЛЬ ДЛЯ ОЦІНЮВАННЯ КІЛЬКОСТІ СТРОК КОДУ ЗАСТОСУНКІВ НА KOTLIN

Приходько С.Б., д.т.н., проф.[1]; Кольцов А.В.[2]; Лопатін В.А.[3]
[1,2,3] Національний університет кораблебудування імені адмірала Макарова
[1,2,3] Україна, Миколаїв
[1]sergiy.prykhodko@nuos.edu.ua; [2]211258@nuos.edu.ua; [3]221311@nuos.edu.ua

Анотація. *В роботі на основі перетворення у вигляді десяткового логарифму побудована трьохфакторна нелінійна регресійна модель для оцінювання кількості строк коду застосунків на Kotlin.*
Ключові слова: *оцінювання кількості строк коду; програмний застосунок; Kotlin; нелінійна регресійна модель.*

Вступна частина. Оцінювання розміру програмних застосунків є важливою задачею на ранній стадії їх розробки, оскільки ця інформація використовується для прогнозування трудомісткості створення програмного забезпечення (ПЗ) за допомогою математичних моделей, наприклад, таких як COCOMO II. Зараз багато програмних застосунків створюються із використанням мови Kotlin.

На теперішний час відомі нелінійні регресійні моделі для оцінювання кількості строк коду різноманітних програмних застосунків з відкритим кодом у переважній більшості в залежності від трьох метрик, що можуть бути визначені за діаграмою класів: загальна кількість класів, середня кількість методів на клас та глибина дерева спадкування [1-3]. Але ці моделі побудовані в першу чергу для застосунків, що створюються з використанням мов програмування PHP та Java. Для застосунків на Kotlin подібні моделі відсутні. Крім того, як це буде показано далі, зазначені моделі, не завжди дають достовірні результати у разі оцінювання кількості строк коду застосунків на Kotlin. Тому виникає потреба у побудові відповідних моделей для оцінювання кількості строк коду застосунків на Kotlin.

Метою дослідження є побудова нелінійної регресійної моделі для оцінювання кількості строк коду застосунків на Kotlin.

Основна частина. Розглянемо побудову нелінійної регресійної моделі для оцінювання кількості строк коду (Y у тисячах рядків коду) застосунків на Kotlin в залежності від трьох факторів: кількість класів X_1; середня кількість методів на клас X_2; глибина дерева спадкування (DIT) X_3. Для цього програмні метрики з 54 застосунків з відкритим кодом на Kotlin, які розміщені на сайті GitHub, були отримані за допомогою інструменту CodeMR (https://plugins.jetbrains.com/plugin/10811-codemr). Зазначені фактори були обрані, тому що їх можна отримати з діаграми класів та між ними немає мультіколінеарності. Далі ми перевірили багатовимірні дані на нормальність за допомогою тесту Мардіа (Mardia's test), якій базується на багатовимірних асиметрії та ексцесі. Тест показав, що розподіл даних не є гаусівським. Тому для визначення наявності викидів у даних ми використовували тест на основі квадрату відстані Махаланобіса для нормалізованих за десятковим логарифмом даних. Для обґрунтування необхідності побудови нелінійної регресійної моделі для оцінювання кількості строк коду застосунків на Kotlin ми перевірили якість відповідної моделі з [1], яка була розроблена для застосунків з відкритим кодом на Java. Отримані значення MMRE та PRED(0,25) вказують на погану якість моделі з [1] у разі її використання

для оцінювання кількості строк коду застосунків на Kotlin. А це призводить до необхідності побудови регресійної моделі для оцінювання кількості строк коду застосунків на Kotlin.

Використовуючи зібрані дані із зазначених вище метрик веб застосунків була побудована лінійна регресійна модель $Y = \hat{b}_0 + \hat{b}_1 X_1 + \hat{b}_2 X_2 + \hat{b}_3 X_3 + \varepsilon$, де $\hat{b}_0$, $\hat{b}_1$, $\hat{b}_2$ та $\hat{b}_3$ є оцінками параметрів. Причому змінна ε повинна бути випадковою величиною з розподілом Гаусу. Нульова гіпотеза про нормальність розподілу ε була перевірена за критерієм Пірсона. Ця гіпотеза була відкинута для рівня значимості 0,05. Те, що розподіл ε не є гаусівським вказує на відсутність теоретичного обґрунтування застосування лінійної регресійної моделі та на необхідність побудови нелінійної регресійної моделі, яка у нашому випадку має вигляд $Y = 10^{\varepsilon+\hat{b}_0} X_1^{\hat{b}_1} X_2^{\hat{b}_2} X_3^{\hat{b}_3}$, де ε – гаусівська випадкова величина з нульовим математичним сподіванням та середньоквадратичним відхиленням, оцінка якого дорівнює 0,08605, $\hat{b}_0$, $\hat{b}_1$, $\hat{b}_2$ та $\hat{b}_3$ є оцінками параметрів, які мають відповідно такі значення: -2,11196, 1,02339, 0,65293 та 0,04783. Оцінки параметрів були отримані за допомогою методу найменших квадратів. Ця модель була побудована за методом [4] із врахуванням наявності викидів за даними з 51 застосунку, які залишилися після видалення викидів.

Створена нелінійна регресійна модель була перевірена за відомими показниками R^2, MMRE та PRED(0,25). Допустимі значення MMRE і PRED(0,25) складають не більше 0,25 і не менше 0,75 відповідно. Значення R^2, MMRE і PRED(0,25) для створеної моделі дорівнюють відповідно 0,9166, 0,1548 та 0,8039. Це свідчить про її добру якість.

Висновки. В роботі на основі десяткового логарифму була побудована нелінійна регресійна модель з трьома предікторами для оцінювання кількості строк коду застосунків на Kotlin. Побудована модель має добру якість відповідного оцінювання.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Приходько, С.Б., Приходько, Н.В., & Смикодуб, Т.Г. (2020). Чотирьохфакторна нелінійна регресійна модель для оцінювання розміру Java-застосунків з відкритим кодом. *Науковий журнал «Вчені записки Таврійського національного університету імені В. І. Вернадського. Серія: Технічні науки», 2 (70),* 157-162. DOI: https://doi.org/10.32838/2663-5941/2020.2-1/25

[2] Приходько, С.Б., & Приходько, Н.В., Ворона, М.В., Бєловол, І.О. (2021). Нелінійна регресійна модель для оцінювання розміру Web-застосунків, що створюються з використанням фреймворку Laravel. *Інформаційні технології та комп'ютерна інженерія, 1 (50),* 115-121. DOI: https://doi.org/10.31649/1999-9941-2021-50-1-115-121

[3] Prykhodko, S., Shutko, I., & Prykhodko, A. (2021). Early LOC estimation of Web apps created using Yii framework by nonlinear regression models. *WSEAS Transactions on Computers*, *20,* 321-328. DOI: http://dx.doi.org/10.37394/23205.2021.20.35

[4] Prykhodko, S., & Prykhodko, N. (2022). A modified technique for constructing nonlinear regression models based on the multivariate normalizing transformations. *Selected Papers of the VIII International Scientific Conference "Information Technology and Implementation" (IT&I-2021). Workshop Proceedings (IT&I-WS-2021), Kyiv, Ukraine, December 1-3, 2021. CEUR Workshop Proceedings, 3179,* 156-166. http://ceur-ws.org/Vol-3179/Paper_15.pdf

Prykhodko S.B, Koltsov A.V., Lopatin V.A.

A nonlinear regression model for estimating LOC of applications in Kotlin

Abstract. *In work a nonlinear regression model with three predictors for estimating LOC of applications in Kotlin has been built.*

Keywords: *LOC estimation; software application; Kotlin; nonlinear regression model.*