УДК 004.412:519.237.5

БАГАТОФАКТОРНА НЕЛІНІЙНА РЕГРЕСІЙНА МОДЕЛЬ ДЛЯ ОЦІНЮВАННЯ РОЗМІРУ ВЕБЗАСТОСУНКІВ, ЩО СТВОРЮЮТЬСЯ З ВИКОРИСТАННЯМ МОВИ ПРОГРАМУВАННЯ C#

Петриченко С.А.[1]; Пухалевич А.В., к.т.н.[2]
Національний університет кораблебудування імені адмірала Макарова
Україна, Миколаїв
[1] petrichenko.serzh111@gmail.com; [2] andrii.pukhalevych@nuos.edu.ua

Анотація. *В роботі розглянуто побудову багатофакторної нелінійної регресійної моделі для оцінювання розміру коду вебзастосунків, що створюються з використанням мови програмування C#. В якості факторів побудованої моделі використовувались метрики програмного забезпечення: кількість класів та кількість методів на клас. Побудована модель має високу достовірність оцінювання, та дозволяє досягти підвищення достовірності в порівнянні з існуючими моделями для оцінювання розміру коду.*

Ключові слова: *регресійна модель; розмір коду; C#-вебзастосунки; оцінювання метричних даних.*

Вступ. Створення вебзастосунків є одним з найбільш актуальних та затребуваних напрямків в галузі інженерії програмного забезпечення. Особливо популярними є вебзастосунки, розроблені з використанням мови програмування C#, завдяки своїй продуктивності, масштабованості та можливостям інтеграції з іншими технологіями. Однак, в процесі створення вебзастосунків виникають значні виклики, особливо коли мова йде про таку задачу програмної інженерії як оцінювання розміру вихідного коду. Розмір коду програмного забезпечення є критичним параметром, який визначає трудомісткість, тривалість та вартість розробки.

Існують розв'язання задачі оцінювання розміру вебзастосунків для інших мов програмування, наприклад регресійні моделі наведені в [1-2]. Однак, модель [1] призначена для мови програмування Java, а модель [2] є однофакторною, через що застосування цих моделей для оцінювання вебзастосунків на мові C# може мати низьку достовірність. Тому, побудова багатофакторної моделі для оцінювання розмірів вебзастосунків, що створюються з використанням мови C# має актуальність.

Метою дослідження є підвищення достовірності оцінювання розміру вебзастосунків, що створюються мовою C#.

Основна частина. Щоб побудувати багатофакторну нелінійну регресійну модель для оцінювання розміру вебзастосунків, що створюються з використанням мови програмування C# використовувались такі фактори:

- Кількість класів проєкту, в якості першої незалежної змінної x_1.

- Кількість методів на клас проєкту, в якості другої незалежної змінної x_2.

Загальна кількість рядків коду проєкту (в тисячах) використовувалась в якості залежної змінної Y.

Для побудови моделі використовувалися метричні дані 40 вебзастосунків, що розміщені на сайті GitHub. Метрики були отримані з використанням спеціалізованого для мови C# інструменту статичного аналізу коду Ndepend [3].

Використовуючи отримані метрики була побудована регресійна модель:

$$Y = \hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \varepsilon, \quad (1)$$

де: $\hat{b}_0$, $\hat{b}_1$, $\hat{b}_2$ – це коефіцієнти регресії, ε – це випадкова похибка.

Для визначення коефіцієнтів регресії (1) був обраний метод найменших квадратів (МНК), який використовується для оцінки параметрів моделі з метою мінімізації суми квадратів різниць між спостережуваними і передбаченими значеннями. Суть МНК полягає в тому, що для заданої моделі і набору даних МНК знаходить такі значення параметрів моделі, які мінімізують суму квадратів різниць між спостережуваними значеннями і значеннями, передбаченими моделлю. Метод спрямований на знаходження "найкращих" параметрів моделі в тому розумінні, що мінімізується середньоквадратична помилка передбачень моделі [4].

Перевірка гіпотези за допомогою критерію Пірсона показала, що розподіл випадкової величини є не відповідає гаусівському розподілу. Тому немає теоретичного обґрунтування для використання лінійної регресії. В цьому випадку можна побудувати нелінійну регресійну модель з використанням нормалізуючих перетворень. Одним з таких нормалізуючих перетворень є десятковий логарифм. За допомогою нормалізованих даних була побудована лінійна регресійна модель у вигляді:

$$Z_y = \hat{b}_0 + \hat{b}_1 Z_1 + \hat{b}_2 Z_2 + \varepsilon, \quad (2)$$

де: $Z_y = log_{10}(Y)$, $Z_1 = log_{10}(X_1)$, $Z_2 = log_{10}(X_2)$.

При побудові моделі (2) була проведена обробка емпіричних даних націлена на виявлення викидів. Для цього були використані так методи, як перевірка відстані квадрату Махаланобіса для кожного рядка, перевірка гіпотези про нормальний розподіл випадкової величини ε за критерієм Пірсона, та перевірка виходу значень залежної змінної за інтервал передбачення. Окремі значення вибірки, які не відповідали критичним показникам, були ідентифіковані як викиди. Викиди були видалені з вибірки, а знаходження коефіцієнтів моделі проводилось повторно, до поки не були видалені усі викиди.

Загалом було видалено 8 викидів, для 32 рядків даних, що залишилися, оцінки коефіцієнтів були наступні: $\hat{b}_0 = 1{,}5805$, $\hat{b}_1 = 0{,}7562$, $\hat{b}_2 = 0{,}6147$.

На основі лінійної моделі для нормалізованих значень було побудовано багатофакторну нелінійну регресійну модель:

$$Y = 10^{\varepsilon + b_0} x_1^{b_1} x_2^{b_2}. \quad (3)$$

Для моделі (3) значення коефіцієнта детермінації R^2 = 0,96, середнє значення модуля відносної похибки MMRE = 0,13 і значення рівня прогнозування PRED(0,25) = 0,97.

Порівнюючи з раніше відомою моделлю [2], для якої MMRE = 1,39, вдається досягти підвищення достовірності за значенням MMRE на 1000%.

Висновки. В роботі було побудовано багатофакторну нелінійну регресійну модель для оцінювання розміру вебзастосунків, що створюються з використанням мови програмування C#. Побудована модель має високу достовірність оцінювання. В порівнянні з моделлю [2], побудована модель дозволяє досягти підвищення достовірності за значенням MMRE на 1000%.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Макарова, Л.М., Приходько, Н.В., Кудін, О.О. (2019). Побудова нелінійної регресійної моделі для оцінювання розміру веб-додатків, реалізованих мовою Java. *Науковий журнал «ВІСНИК ХНТУ»*, 2 (69), 148-152.

[2] Латанська, Л.О., Макарова, Л.М., Нікітіна, О.Ю., Нікітін, О.В. (2022). Математичне моделювання в задачах оцінювання розміру прикладного програмного забезпечення з відкритим кодом на C#. *Computer Science and Applied Mathematics, 66-74.* https://doi.org/10.26661/2413-6549-2022-1-08

[3] NDepend (n.d.). *100+ .NET and C# Code Metrics.* Retrieved October 15, 2023 from https://www.ndepend.com/docs/code-metrics

[4] Павленко, П.М. (2013). *Основи математичного моделювання систем і процесів : навч. посіб*.

Petruchenko Serhii, Pukhalevych Andrii

A multivariate nonlinear regression model for estimating the size of web applications created using the C# programming language

Abstract. *The article considers the construction of a multivariate nonlinear regression model for estimating the code size of web applications created using the C# programming language. The following metrics were used as factors: number of classes and number of methods per class.*

Keywords: *regression model; code size; C#-applications; metric data estimation.*