УДК 004.412:519.237.5

МАТЕМАТИЧНА МОДЕЛЬ ДЛЯ РАННЬОГО ОЦІНЮВАННЯ КІЛЬКОСТІ РЯДКІВ КОДУ В ОСТАННІХ ВЕРСІЯХ ВЕБЗАСТОСУНКІВ, ЯКІ РОЗРОБЛЯЮТЬСЯ ЗА ДОПОМОГОЮ ФРЕЙМВОРКІВ PHP

Приходько С.Б., д.т.н., проф.[1]; Пухалевич А.В., к.т.н., доц.[2]; Галиць В.О.[3]
[1,2,3] Національний університет кораблебудування імені адмірала Макарова
[1,2,3] Україна, Миколаїв
[1]sergiy.prykhodko@nuos.edu.ua; [2]andrii.pukhalevych@nuos.edu.ua; [3]251184@nuos.edu.ua

Анотація. *В роботі на основі перетворення у вигляді десяткового логарифму побудовано математичну модель для раннього оцінювання кількості рядків коду останніх версій вебзастосунків, які розробляються за допомогою фреймворків PHP, що залежить від кількості класів їх попередніх версій. Ця математична модель складається з трьох рівнянь еліпса прогнозування для нормалізованих метрик кількості рядків коду і класів попередніх версій та трьох однофакторних нелінійних регресійних моделей для оцінювання кількості рядків коду останніх версій вебзастосунків за кількістю класів попередніх версій.*
Ключові слова: *математична модель; раннє оцінювання; кількість рядків коду; вебзастосунок; фреймворк; PHP; еліпс прогнозування; нелінійна регресійна модель.*

Вступна частина. Оцінювання розміру різноманітних 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 вебзастосунків створюються із використанням фреймворків PHP.

На теперішний час відомі нелінійні регресійні моделі для оцінювання кількості строк коду різноманітних програмних застосунків у переважній більшості в залежності від метрик, що можуть бути визначені за діаграмою класів. А це не дозволяє здійснювати відповідне оцінювання на самому першому етапі розробки ПЗ. Тому в роботах [1-3] були запропоновані нелінійні регресійні моделі для оцінювання кількості рядків коду останніх випусків програмних застосунків за метриками їх попередніх версій. Але ці моделі не розраховані для вебзастосунків. Тому виникає потреба у побудові математичних моделей для оцінювання кількості рядків коду останніх версій вебзастосунків за метриками їх попередніх версій.

Метою дослідження є побудова математичної моделі для оцінювання кількості рядків коду останніх версій вебзастосунків, які розробляються за допомогою фреймворків PHP, за метриками їх попередніх версій.

Основна частина. Розглянемо побудову математичної моделі для оцінювання кількості рядків коду останніх версій вебзастосунків, які розробляються за допомогою фреймворків PHP, в залежності від кількості класів їх попередніх версій. Подібну математичну модель у вигляді двох нелінійних регресійних моделей було запропоновано в [2]. Для побудови цих двох нелінійних регресійних моделей в [2] було використано відповідний метод [4], який базується на нормалізуючих перетвореннях та визначенні наявності викидів у нелінійній регресійній моделі. У роботі ми також застосували цей метод [4] тому, що в процесі побудови нелінійних регресійних моделей також спостерігалися викиди.

Для побудови нелінійних регресійних моделей ми використовували дані з 288 версій 5 застосунків з відкритим кодом, які були розроблені із використанням таких трьох фреймворків PHP: CakePhp, Laravel та Symfony. Дані були отримані за допомогою інструменту PhpMetrics (https://www.phpmetrics.org/) за кодом застосунків, що розміщений на сайті GitHub.

Використовуючи зібрані дані із зазначених вище метрик вебзастосунків спочатку була побудована лінійна регресійна модель $Y = \hat{b}_0 + \hat{b}_1 X_1 + \varepsilon$, де $\hat{b}_0$ та $\hat{b}_1$ є оцінками параметрів, а змінна ε має 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 випадку нормалізуючого перетворення у вигляді десяткового логарифму має вигляд $Y = 10^{\varepsilon+\hat{b}_0} X_1^{\hat{b}_1}$, де ε – гаусівська випадкова величина з нульовим математичним сподіванням та певним середньоквадратичним відхиленням, оцінка якого $\hat{\sigma}_\varepsilon$. Ми побудували три нелінійні регресійні моделі зазначеного вигляду.

В результаті побудови першої нелінійної регресійної моделі за методом [4] було 88 викидів. Тобто перша модель була побудована за 200 точками даних. Для побудови другої моделі ми використовували 88 точок даних, які були викидами для першої моделі. В результаті побудови другої нелінійної регресійної моделі за методом [4] було 32 викиди, які ми застосували для створення третьої моделі. В результаті побудови третьої нелінійної регресійної моделі за методом [4] було 6 викидів. Тобто третя модель була побудована за 26 точками даних. Оцінки параметрів цих трьох моделей наведені в таблиці 1.

Таблиця 1

Модель	$\hat{b}_0$	$\hat{b}_1$	$\hat{\sigma}_\varepsilon$	R^2	*MMRE*	*PRED*(0,25)
Модель 1	1,94774	0,89402	0,04084	0,9434	0,0834	1,0000
Модель 2	1,42201	0,95581	0,02798	0,9898	0,0496	1,0000
Модель 3	2,59869	0,63021	0,11008	0,8002	0,1941	0,7692

Якість трьох побудованих нелінійних регресійних моделей було перевірено за коефіцієнтом детермінації R^2, середньою величиною відносної помилки *MMRE* та відсотком прогнозованих результатів *PRED*, для яких величини відносної помилки менші за 0,25, *PRED*(0,25). В таблиці 1 наведені значення цих показників якості трьох побудованих нелінійних регресійних моделей. Значення показників якості з таблиці 1 свідчать про добру якість трьох побудованих нелінійних регресійних моделей. Великі значення R^2 вказують на те, що більшість значень *Y* знаходяться біля лінії регресії.

Для вибору певної з трьох побудованих нелінійних регресійних моделей для оцінювання кількості рядків коду останніх версій вебзастосунків, які розробляються за допомогою фреймворків PHP, пропонується використовувати три рівняння еліпса прогнозування для нормалізованих метрик кількості рядків коду і класів попередніх версій та інтервали їх зміни. Кожне рівняння відповідає своїй нелінійній регресійні моделі. Застосування зазначених рівнянь еліпса прогнозування обумовлено тим, що інтервали зміни вказаних метрик мали перетини при побудові трьох нелінійних регресійних моделей. Вибір певної нелінійної регресійної моделі здійснюється таким чином. Задається кількість рядків коду і класів попередньої версії вебзастосунку, для якого треба оцінити кількість рядків коду його наступної версії. За цими даними для кожної моделі перевіряється, чи значення правої частини відповідного для неї рівняння еліпса прогнозування (квадрату відстані Махаланобіса) не

перевищує 10,60 (значення правої частини), а також вони (ці дані) не виходять за відповідні інтервали. Якщо зазначена умова виконується для певної моделі, то вона обирається для оцінювання кількості рядків коду останнього версії вебзастосунку. У разі, якщо зазначена умова не виконується для жодної моделі, то для цього вебзастосунку ми не можемо застосувати побудовану математичну модель.

Висновки. В роботі удосконалено математичну модель для раннього оцінювання кількості рядків коду останніх версій вебзастосунків, що складається з трьох рівнянь еліпса прогнозування для нормалізованих метрик кількості рядків коду і класів попередніх версій та трьох однофакторних нелінійних регресійних моделей для оцінювання кількості рядків коду останніх версій вебзастосунків за кількістю класів попередніх версій. Ця математична модель дозволяє підвищити достовірність раннього оцінювання кількості рядків коду останніх версій вебзастосунків, які розробляються за допомогою фреймворків PHP, у порівнянні з існуючими моделями.

Використання інструментів ШІ. Під час підготовки матеріалів було використано Gemini Deep Research для пошуку інформації. Усі результати перевірено та відредаговано авторами. Наукові положення, висновки та результати є власним інтелектуальним внеском авторів.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Гунченко, Ю., Мещеряков, В., Рачинська, А., & Приходько, С. (2025). Регресійні моделі для раннього оцінювання кількості рядків коду різних випусків застосунків з відкритим кодом. *Прикладні питання математичного моделювання, 2* (8), 86-92. https://doi.org/10.32782/mathematical-modelling/2025-8-2-9

[2] Гунченко, Ю., Мещеряков, В., Приходько, С., & Рачинська, А. (2025). Раннє оцінювання кількості рядків коду останніх випусків застосунків з відкритим кодом за метриками їх перших випусків. *Information Technology: Computer Science, Software Engineering and Cyber Security*, *4*, 72-78. https://doi.org/10.32782/IT/2025-4-9

[3] Gunchenko, Y. O., Meshcheryakov, V. I., Prykhodko, S. B., & Rachinska, A. L. (2026). Estimating the number of lines of code of the latest releases of open-source applications using nonlinear regression models. *Applied Aspects of Information Technology, 1* (9), 122–133. https://doi.org/10.15276/aait.09.2026.09

[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, Halyts V.O.

Mathematical model for early estimation of the number of lines of code in the latest versions of web applications developed using PHP frameworks

Abstract. *In this work, a mathematical model for early estimation of the number of lines of code of the latest versions of web applications developed using PHP frameworks, which depends on the number of classes of their previous versions, is built based on the decimal logarithm transformation. This mathematical model consists of three ellipse prediction equations for normalized metrics of the number of lines of code and classes of previous versions, and three univariate nonlinear regression models for estimating the number of lines of code of the latest versions of web applications by the number of classes of previous versions.*

Keywords: *mathematical model; early estimation; number of lines of code; web application; framework; PHP; prediction ellipse; non-linear regression model.*