

НЕЛІНІЙНА РЕГРЕСІЙНА МОДЕЛЬ ДЛЯ ОЦІНЮВАННЯ КІЛЬКОСТІ СТРОК КОДУ ВЕБ ЗАСТОСУНКІВ, ЩО СТВОРЮЮТЬСЯ ЗА ДОПОМОГОЮ ФРЕЙМВОРКУ DJUNGO

Приходько С.Б., д.т.н., професор¹; Шемчук С.Т.²; Карлюченко Є.К.³
^{1,2,3} Національний університет кораблебудування імені адмірала Макарова
^{1,2,3} Україна, Миколаїв

¹sergiy.prykhodko@nuos.edu.ua; ²sergey.shemchuk.t@gmail.com; ³211256@nuos.edu.ua

Анотація. В роботі на основі перетворення у вигляді десяткового логарифму побудована трьохфакторна нелінійна регресійна модель для оцінювання кількості строк коду веб застосунків, що створюються з використанням фреймворку Django.

Ключові слова: оцінювання кількості строк коду; веб застосунок; фреймворк Django; нелінійна регресійна модель.

Вступна частина. Оцінювання розміру веб застосунків як і іншого програмного забезпечення (ПЗ) є важливою задачею на ранній стадії розробки ПЗ, оскільки ця інформація використовується для прогнозування трудомісткості створення ПЗ за допомогою математичних моделей, наприклад, таких як СОСОМО II. Зараз багато веб застосунків створюються із використанням різноманітних фреймворків, які прискорюють розробку ПЗ. Одним із таких є Django (<https://www.djangoproject.com/>). Django – це безкоштовний високорівневий Python веб фреймворк з відкритим кодом.

На теперішній час відомі нелінійні регресійні моделі для оцінювання кількості строк коду веб застосунків з відкритим кодом 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 за даними веб застосунків, що створюються з використанням різних PHP-фреймворків. А Django – це на сьогодні найбільш популярний Python-фреймворк. Тому виникає потреба побудові відповідних моделей для оцінювання кількості строк коду веб застосунків, що створюються з використанням саме Python-фреймворку Django.

Метою дослідження є побудова нелінійної регресійної моделі для оцінювання кількості строк коду веб застосунків, що створюються з використанням фреймворку Django.

Основна частина. Розглянемо побудову нелінійної регресійної моделі для оцінювання кількості строк коду веб застосунків, що створюються з використанням фреймворку Django. Враховуючи на те, що Django є Python-фреймворком, метрики з 45 веб застосунків, які розміщені на сайті GitHub (<https://github.com>), були отримані за допомогою інструменту Radon (<https://radon.readthedocs.io/en/latest/index.html>) та власного скрипта (для метрик: кількість класів, методів та DIT). Як і в [3], ми брали такі метрики: розмір веб застосунку Y у тисячах рядків коду; кількість класів X_1 ; середню кількість методів на клас X_2 ; глибину дерева спадкування (DIT) X_3 . Зазначені фактори були обрані, тому що їх можна отримати з діаграми класів та між ними немає мультиколінеарності. 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.

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.

Нелінійна регресійна модель 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 0,11109, $\hat{b}_0$, $\hat{b}_1$, $\hat{b}_2$ та $\hat{b}_3$ є оцінками параметрів, які мають відповідно такі значення: -1,26991, 0,87984, 0,72781 та 0,14788. Оцінки параметрів були отримані за допомогою методу найменших квадратів. Ця модель була побудована за методом [4] із врахуванням наявності викидів за даними з 33 застосунків, які залишилися.

Створена нелінійна регресійна модель 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). 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 0,993, 0,213 і 0,667. Це свідчить про задовільну якість отриманої моделі.

Висновки. В роботі на основі десятичного логарифму була побудована нелінійна регресійна модель з трьома предикторами для оцінювання кількості строк коду веб застосунків, що створюються з використанням фреймворку Django. Побудована модель має прийнятну якість відповідного оцінювання.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] Приходько, С.Б., Приходько, Н.В., Ворона, М.В., Беловол, 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>
- [2] 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>
- [3] Prykhodko, S.B., Shutko, I.S., Prykhodko, A.S. (2021). A nonlinear regression model to estimate the size of Web apps created using the CakePHP framework. *Radio Electronics, Computer Science, Control*, 4 (59). 129-139. DOI: <http://dx.doi.org/10.15588/1607-3274-2021-4-12>
- [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., Shemchuk S.T., Karlyuchenko E.K.

A nonlinear regression model for estimating LOC of web apps created using the Django framework

Abstract. In work a nonlinear regression model with three predictors for estimating LOC of web applications created using the Django framework has been built.

Keywords: LOC estimation; Web app; Django; nonlinear regression model.