

РЕГРЕСІЙНА МОДЕЛЬ ДЛЯ ОЦІНЮВАННЯ РОЗМІРУ ВЕБ-ЗАСТОСУНКІВ, ЩО СТВОРЮЮТЬСЯ МОВОЮ PYTHON З ВИКОРИСТАННЯМ DJANGO REST FRAMEWORK

Давлатова Д.Х.¹, Макарова Л.М., к.т.н., доцент²

^{1,2} Національний університет кораблебудування імені адмірала Макарова

^{1,2} Україна, Миколаїв

¹ didi77777@ukr.net; ² lidiia.makarova@nuos.edu.ua

Анотація. Розглянуто проблему оцінювання розміру веб-застосунків, що створюються мовою Python з використанням Django Rest Framework. Для побудови нелінійної регресійної моделі в якості нормалізуючого перетворення було використано десятковий логарифм та дані з 71 проекту з відкритим вихідним кодом, завдяки чому було підвищено достовірність оцінювання розміру відповідних веб-застосунків.

Ключові слова: нелінійна регресійна модель; веб-застосунки; логарифмічне перетворення; Python; Django Rest Framework.

Вступна частина. Успіх або невдача проекту на ранньому етапі розробки, залежить безпосередньо від ефективності оцінювання трудомісткості, яка в свою чергу залежить від розміру програми, тому одним з тих факторів, що впливають на процес розробки ПЗ, є розмір програми. Проблема отримання ефективної системи оцінювання розміру веб-застосунків – це важливе завдання, що вимагає удосконалення існуючих методів. [1].

Важливим етапом аналізу складної системи є побудова математичної моделі, що здійснюється на підставі змістовної моделі і складається з вхідних параметрів, внутрішніх параметрів та вихідних параметрів. Однією з умов побудови якісної адекватної моделі є належність вхідних даних, тому потрібна первинна обробка експериментальних даних із метою виявлення ненадійних вимірів [2]. Виявлення аномальних значень відноситься до проблеми пошуку зразків у даних, які не відповідають очікуваній поведінці.

Найбільш часто використовуваний підхід до оцінювання заснований на регресії, де модель СОСОМО є найвідомішою моделлю [3]. Багато дослідницьких робіт з оцінювання веб-зусиль також були проведені Мендесом та його співробітниками [4].

Під час проведення регресійного аналізу дослідники, зазвичай, віддають перевагу побудові лінійних моделей, які легко оцінюються, і результати яких чітко інтерпретуються. Однак дані рідко носять абсолютно лінійний характер, тому для їх адекватного опису доводиться вдаватися до побудови нелінійних залежностей.

TIOBE Software представила рейтинг найпопулярніших мов програмування на серпень 2022 року [5]. Порівняно з минулим роком Python додав до популярності 3,56%, перемістившись з другого на перше місце з показником 15,42%. Це найвищий показник популярності цієї мови програмування за час існування рейтингу. Найнижчий був зафіксований у 2003 році (0,97%), коли Python посідав 13 місце в рейтингу [6].

Існують моделі оцінювання розміру різноманітних програмних застосунків створених мовою Java, додатків реалізованих на JavaScript та веб-застосунків створених мовою PHP як із використанням фреймворків, наприклад Laravel, так і без них [6-8]. Але враховуючи те, що всі моделі безпосередньо залежать від мови програмування та її особливостей, використавши розроблені моделі для однієї мови, будуть отримані результати, що суттєво відрізнятимуться від тих, що були б отримані з використанням відповідної моделі. Тобто маємо, що для

обраної мови програмування, фреймворку та бібліотеки не було створено відповідної моделі, реалізованої для особливостей мови Python, тому є доцільним створити таку модель.

Таким чином, побудова нелінійних регресійних моделей із використанням нормалізуючих перетворень для підвищення достовірності оцінювання розміру веб-застосунків, що створюються мовою Python з використанням Django Rest Framework, і створення на їх основі інформаційної технології переробки інформації є актуальною та має практичну цінність.

Метою дослідження є підвищення достовірності оцінювання розміру веб-застосунків, що створюються мовою Python з використанням Django Rest Framework.

Основна частина. Для побудови двофакторної нелінійної регресійної моделі для оцінювання розміру веб-застосунків, що створюються мовою Python з використанням Django Rest Framework було обрано наступні фактори:

- загальна кількість рядків коду проекту в тисячах (kilolines of code, KLOC), в якості значення залежної змінної Y ;
- кількість класів проекту, в якості першої незалежної змінної X_1 ;
- кількість методів проекту, в якості другої незалежної змінної X_2 .

Практика перетворення даних, щоб вони виглядали нормально за допомогою певного нормалізуючого перетворення широко використовується через її простоту. Хоча, можливо, це не найелегантніше рішення проблеми, часто ця техніка працює добре як швидке та прагматичне рішення. Тож було обрано десятковий логарифм в якості нормалізуючого перетворення для побудови регресійної моделі для оцінювання розміру веб-застосунків, що створюються мовою Python з використанням Django Rest Framework. У логарифмічному перетворенні кожна змінна x замінюється на $\lg(x)$ з основою 10 (1) і зворотним до нього є перетворення (2):

$$z = \lg(x), \quad (1)$$

$$x = 10^z. \quad (2)$$

Багатофакторна лінійна регресійна модель в загальному випадку має вигляд:

$$Z_y = \widehat{Z_y} + \varepsilon = b_0 + b_1 Z_{x1} + b_2 Z_{x2} + \dots + b_k Z_{xk} + \varepsilon, \quad (3)$$

де: ε – це випадкова похибка, розподіл якої в загальному випадку залежить від незалежних змінних, але математичне очікування якої рівне нулю.

Нелінійна регресійна модель будується за перетвореним у вигляді десяткового логарифму та лінійним рівнянням (3):

$$Y = 10^{\varepsilon + b_0} X_1^{b_1} X_2^{b_2} \cdot \dots \cdot X_k^{b_k}. \quad (4)$$

Щоб визначити коефіцієнти рівняння регресії b застосовують різні методи (графічний, метод середніх), проте найбільшого поширення набув метод найменших квадратів (МНК). Суть МНК полягає в мінімізації суми квадратів відхилень значень пояснюваної змінної, що спостерігаються, від її розрахункових значень. За допомогою МНК було знайдено коефіцієнти $b_0 = -1,6796$, $b_1 = 0,0534$, $b_2 = 0,9571$ та з використанням формули (3) побудовано лінійну регресійну модель для нормалізованих даних:

$$Z_y = -1,6796 + 0,0534 Z_{x1} + 0,9571 Z_{x2} + \varepsilon.$$

Оскільки було виконано побудову лінійної регресії для нормалізованих даних необхідно впевнитись, що залишки лінійної регресійної моделі підкорюються нормальному закону розподілу. Перевірку залишків на нормальність також проводимо на основі критерію Пірсона.

Розрахували критичне та розрахункове значення критерію Пірсона: $\chi^2_{кр} = 9,49$ та $\chi^2 = 0,22$ відповідно, з чого можемо зробити висновок, що залишки розподілені за нормальним законом, тобто можемо будувати дану лінійну регресію і за рівнянням (4) можемо перейти до нелінійної регресійної моделі:

$$Y = 10^{\varepsilon - 1,6796X_1^{0,0534}X_2^{0,9571}}.$$

Для нелінійної моделі було отримано значення коефіцієнту детермінації $R^2 = 0,8936$, середньої величини відносної похибки $MMRE = 0,2342$ і рівню прогнозування $PRED(0,25) = 0,6290$.

Висновки. Аналізуючи параметри якості двофакторної нелінійної регресійної моделі дійшли висновку, що можливо логарифмічне перетворення дає не найкращі результати і для нормалізації рядів даних можна було б використати перетворення Бокса-Кокса, перетворення Джонсона або багатовимірні перетворення та незважаючи на те, що значення рівня прогнозування $PRED(0,25)$ не дістало до необхідного мінімально допустимого значення 0,75, загалом по іншим параметрам було отримано гарні результати і отримана нелінійна регресійна модель має високу якість оцінювання.

Отже, двофакторну нелінійну регресійну модель, що розроблена на основі обраних незалежних змінних, а саме значення кількості класів X_1 та кількості методів X_2 проекту, можна використовувати для виконання прогнозу значень результативного показника розміру веб-застосунків у KLOC, тому на підставі саме цієї моделі можемо виконувати розробку програми для прогнозування.

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. Вісник Херсонського національного технічного університету, № 2(69), 145–153.
- [2] Павленко, П.М. (2013). Основи математичного моделювання систем і процесів : навч. посіб. Київ: Книжкове вид-во НАУ.
- [3] Boehm, B.W. (1981). Software engineering economics. New Jersey : Prentice Hall.
- [4] Kitchenham, B., & Mendes, E. (2009). Why Comparative Effort Prediction Studies may be Invalid. International Conference on Predictor Models in Software Engineering, 1-5.
- [5] The software quality company TIOBE. (n.d.). TIOBE Index for September 2022. Retrieved September 30, 2022, from <https://www.tiobe.com/tiobe-index/>.
- [6] The software quality company TIOBE. (n.d.). The Python Programming Language. Retrieved September 30, 2022, from <https://www.tiobe.com/tiobe-index/python/>.

Davlatova D.H., Makarova L.M.

A regression model to estimate the size of web applications created in Python using the Django Rest Framework

Abstract. The problem of estimating the size of web applications created in Python using the Django Rest Framework is considered. Decimal logarithm and data from 71 open source projects were used as a normalizing transformation to build a nonlinear regression model, which increased the reliability of the estimation of the size of the respective web applications.

Keywords: non-linear regression model; web applications; logarithmic transformation; Python; Django Rest Framework.