УДК 004.412:519.237.5

ТРИФАКТОРНА НЕЛІНІЙНА РЕГРЕСІЙНА МОДЕЛЬ ДЛЯ ОЦІНЮВАННЯ РОЗМІРУ КОДУ ІНТЕРНЕТ МАГАЗИНІВ НА PHP

Пухалевич А.В., к.т.н[1]; Алікіна Д.І.[2]
[1,2] Національний університет кораблебудування імені адмірала Макарова
[1,2] Україна, Миколаїв
[1] andrii.pukhalevych@nuos.edu.ua; [2] 181267@nuos.edu.ua

Анотація. *В роботі через перетворення з десяткового логарифму побудована трифакторна нелінійна регресійна модель для оцінювання розміру коду інтернет магазинів на PHP. За показниками R^2, PRED, MMRE побудована модель має високу якість.*

Ключові слова: *оцінювання розміру коду; PHP; інтернет магазин; нелінійна регресійна модель.*

Вступна частина. Інтернет магазини стають все популярнішими, тому розробникам цього виду програмного забезпечення постійно доводиться вирішувати таку задачу програмної інженерії, як оцінювання розміру коду інтернет магазинів. Це потрібно для правильного прогнозування ресурсів затрачених на розробку таких сайтів. Так як більшість сайтів у світі працює на PHP [1], розробка математичної моделі для оцінювання розміру інтернет магазинів на PHP та створення програмного забезпечення для її реалізації є актуальною темою дослідження. Тому що існуючі моделі, такі як [2], побудовані для інших видів програмного забезпечення на мові програмування PHP, через що може знижуватись достовірність оцінювання розміру коду інтернет магазинів на PHP.

Метою дослідження є підвищення достовірності оцінювання розміру інтернет магазинів на PHP.

Основна частина. Розглянемо побудову трифакторної нелінійної регресійної моделі для оцінювання розміру коду інтернет магазинів на PHP. Метрики 40 інтернет магазинів були отримані за допомогою утиліти PhpMetrics (https://phpmetrics.org/), які розміщені на сайті GitHub. Для побудови моделі для оцінювання кількості строк коду (Y у тисячах рядків коду) інтернет магазинів на PHP залежно від трьох факторів: кількість класів X1; середня кількість методів на клас X2; глибина дерева спадкування (DIT) X3. Вибір вищевказаних факторів було здійснено через відсутність мультиколінеарності між ними.

Далі було перевірено багатовимірні дані на нормальність закону розподілу. Було виявлено, що розподіл даних не є гаусівським, далі необхідно здійснити нормалізацію даних. Після чого для визначення наявності викидів було використано тест на основі квадрату відстані Махаланобіса для нормалізованих за десятковим логарифмом даних.

Після нормалізації з даних було видалено вісім тем, так як відстань Махаланобіса перевищувала критичне значення.

Для 32 тем було побудовано трифакторну лінійну регресійну модель для нормалізованих значень:

$log_{10}(Y) = b_0 + b_1 log_{10}(X_1) + b_2 log_{10}(X_2) + b_3 log_{10}(X_3) + \varepsilon$, де b_0, b_1, b_2, b_3 є оцінками параметрів, а ε – випадкова величина.

Методом найменших квадратів було отримано наступні значення оцінок параметрів лінійної регресії b_0 = -1,89, b_1 = -0,92, b_2 = 1,35, b_3 = -0,15.

Перевірка за критерієм Пірсона показала, що розподіл випадкової величини ε є гаусівським.

Трифакторна нелінійна регресійна модель для оцінювання розміру коду інтернет магазинів на PHP в залежності від факторів X_1, X_2 та X_3:

$$Y = 10^{b_0+b_1 log_{10}(X_1)+b_2 log_{10}(X_2)+b_3 log_{10}(X_3)+\varepsilon}.$$

Після перетворення, модель є такою:

$$Y = 10^{b_0+\varepsilon} {X_1}^{b_1} {X_2}^{b_2} {X_3}^{b_3}.$$

Побудовану модель було перевірено за множинним коефіцієнтом детермінації R^2, середньою величиною відносної помилки MMRE та відсотком оцінок кількостей рядків коду, для яких величина модулю відносної помилки не більша 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) відповідно дорівнюють 0,98, 0,12, 0,90 для побудованої моделі, що свідчить про її високу якість.

В порівнянні із моделлю [2], у якої з емпіричними даними Wordpress тем на PHP MMRE = 1,13, побудована модель дозволяє підвищити достовірність оцінювання розміру вебзастосунків на PHP. Підвищення якості за MMRE склало 834%.

Висновки. В результаті роботи було побудовано модель для оцінювання розміру коду інтернет магазинів на PHP, яка має високі характеристики R^2, MMRE і PRED(0,25). У порівнянні з моделлю для оцінювання розміру коду WordPress тем на PHP [2], побудована модель має на 834% менше значення модуля відносної похибки.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] *W3Techs. (n.d.). Usage statistics and market share of PHP for websites. Retrieved October 15, 2023*, https://w3techs.com/technologies/details/pl-php

[2] Дорогань, І.Ю., & Пухалевич, А.В. (2022). Математична модель для оцінювання розміру Wordpress тем на PHP. *Інформаційні технології: моделі, алгоритми, системи (ITMAS - 2022): Матеріали III Всеукраїнської науково-практичної інтернет конференції (26-28 жовтня 2022 р.)*, 33-34. http://itconf.nuos.edu.ua/2022/proceedings

Pukhalevych A.V, Alikina D.I.

A three factor nonlinear regression model for estimating the code size of online stores on PHP

Abstract. *In the work, a three-factor nonlinear regression model was built for estimating the size of the code of online stores in PHP using a transformation from a decimal logarithm. According to R^2, PRED, MMRE, the built model is of high quality.*

Keywords: *code size estimation; PHP; Internet-shop; non-linear regression model.*