

АНОТАЦІЯ

Ворона М.В. Математичні моделі та інформаційна технологія для оцінювання розміру програмних застосунків з відкритим кодом на РНР. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії (PhD) за спеціальністю 122 "Комп'ютерні науки" (галузь 12 – Інформаційні технології). – Національний університет кораблебудування імені адмірала Макарова, Міністерство освіти і науки України, Миколаїв, 2021.

Дисертаційна робота присвячена вирішенню важливого науково-практичного завдання підвищення достовірності оцінювання розміру програмних застосунків з відкритим кодом на РНР за рахунок побудови відповідних математичних моделей у вигляді нелінійних регресійних моделей та створенню на їх основі інструментарію інформаційної технології (ІТ) обробки інформації для оцінювання розміру програмних застосунків з відкритим кодом на РНР на ранніх стадіях розробки за метриками діаграми класів.

Актуальність цієї роботи пов'язана з наступним. По-перше, оцінка розміру програмного забезпечення (ПЗ) необхідна для оцінювання трудомісткості розробки програмного застосунку, наприклад, за допомогою такої відомої моделі як СОСОМО II. По-друге, раннє оцінювання розміру ПЗ є складним завданням, оскільки на початкових фазах створення програмного застосунку доступна обмежена інформація. По-третє, існуючі математичні моделі для оцінювання розміру програмних застосунків з відкритим кодом на РНР як правило не містять у своєму складі випадкові змінні. Все це призводить до низької достовірності оцінювання розміру програмних застосунків з відкритим кодом на РНР на початкових етапах їх розробки.

Метою дисертаційної роботи є підвищення достовірності оцінювання розміру програмних застосунків з відкритим кодом на РНР у фазах аналізу та проектування за метриками діаграми класів за допомогою нелінійних

регресійних моделей.

Робочою науковою гіпотезою дисертаційного дослідження є твердження, що підвищення достовірності оцінювання розміру програмних застосунків з відкритим кодом на РНР у фазах аналізу та проектування за метриками діаграми класів досягається за рахунок застосування багатомірних нелінійних регресійних моделей, які дозволяють описувати зазначений розмір як випадкову величину.

Для побудови вказаних нелінійних регресійних моделей пропонується використовувати відповідний метод на основі багатовимірних нормалізуючих перетворень, які дозволяють враховувати кореляцію між залежною і незалежними змінними.

Для досягнення поставленої мети потрібно вирішити такі завдання:

- проаналізувати існуючі математичні моделі для оцінювання розміру програмних застосунків з відкритим кодом на РНР;
- удосконалити трьох-факторну регресійну модель для оцінювання розміру програмних застосунків з відкритим кодом на РНР за метриками діаграми класів на основі чотиривимірною нормалізуючого перетворення;
- удосконалити рівняння для визначення ширин довірчого інтервалу та інтервалу передбачення нелінійної регресії розміру програмних застосунків з відкритим кодом на РНР за метриками діаграми класів на основі чотиривимірною нормалізуючого перетворення;
- на підставі запропонованої трьох-факторної нелінійної регресійної моделі розробити інструментарій інформаційної технології для оцінювання розміру програмних застосунків з відкритим кодом на РНР за метриками діаграми класів.

Наукова новизна одержаних результатів полягає у наступному.

1) Удосконалено трьох-факторне рівняння нелінійної регресії для оцінювання розміру програмних застосунків з відкритим кодом на РНР на початкових етапах розробки в залежності від трьох факторів (кількості класів; середньої кількості методів на клас, середнього значення метрики DIT на клас)

на основі чотиривимірного нормалізуючого перетворення Джонсона сім'ї S_B . Це дозволяє підвищити достовірність оцінювання вибіркового середнього залежної змінної трьох-факторної нелінійної регресії розміру програмних застосунків з відкритим кодом на РНР у порівнянні з використанням одновимірних нормалізуючих перетворень. Рівняння, що побудовано, в порівнянні з іншими регресійними рівняннями, має менше значення середньої величини відносної похибки та менші ширини довірчого інтервалу нелінійної регресії.

2) Отримало подальший розвиток рівняння для визначення границь довірчого інтервалу нелінійної регресії розміру програмних застосунків з відкритим кодом на РНР на основі чотиривимірного нормалізуючого перетворення Джонсона сім'ї S_B . Це у порівнянні з використанням одновимірних перетворень дозволяє в багатьох випадках зменшити ширини довірчого інтервалу нелінійної регресії розміру програмних застосунків з відкритим кодом на РНР і тим самим підвищити достовірність оцінювання вибіркового середнього залежної змінної нелінійної регресії, а саме – зазначеного розміру.

3) Удосконалено трьох-факторну нелінійну регресійну модель для оцінювання розміру програмних застосунків з відкритим кодом на РНР в залежності від кількості класів; середньої кількості методів на клас та середнього значення метрики DIT на клас на основі чотиривимірного нормалізуючого перетворення Джонсона сім'ї S_B , що дозволяє підвищити достовірність оцінювання розміру програмних застосунків з відкритим кодом на РНР саме як негаусівської випадкової величини у порівнянні з використанням одновимірних нормалізуючих перетворень. Модель, що побудовано, в порівнянні з іншими регресійними моделями (як лінійними, так і нелійними), має більший відсоток прогнозованих значень, менше значення середньої величини відносної похибки та менші ширини інтервалу передбачення нелінійної регресії.

4) Отримало подальший розвиток рівняння для визначення границь

інтервалу передбачення трьох-факторної нелінійної регресії розміру програмних застосунків з відкритим кодом на РНР на основі чотиридимірного нормалізуючого перетворення Джонсона сім'ї S_B . Це у порівнянні з використанням одновимірних перетворень дозволяє в багатьох випадках зменшити ширину інтервалу передбачення нелінійної регресії розміру програмних застосунків з відкритим кодом на РНР і тим самим підвищити достовірність оцінювання зазначеного розміру як залежної випадкової величини.

Практичне значення одержаних результатів полягає у наступному. Розроблено інструментарій ІТ для оцінювання розміру програмних застосунків з відкритим кодом на РНР у фазах аналізу та проектування за метриками діаграми класів. Для цього використовувалася система моделювання Scilab (<https://www.scilab.org/>). Було розроблено відповідне ПЗ scі-мовою для пакету Scilab 6.0.0 та методику оцінювання розміру програмних застосунків з відкритим кодом на РНР, яка розрахована на використання зазначеного ПЗ. Програму, що розроблено, можна використовувати для оцінювання розміру програмних застосунків з відкритим кодом на РНР за трьома метриками діаграми класів.

У вступі дисертації розкрита сутність науково-практичного завдання та його значущість, обґрунтовано необхідність проведення дослідження, подана загальна характеристика дисертації в такій послідовності: актуальність теми; зв'язок роботи з науковими програмами, планами, темами; мета і завдання дослідження; наукова новизна і практичне значення одержаних результатів; особистий внесок здобувача; апробація результатів дисертації та публікації.

У першому розділі дисертації виконано аналіз існуючих методів і моделей для оцінювання розміру програмних застосунків з відкритим кодом на РНР та здійснено обґрунтування необхідності проведення досліджень за обраною темою.

У другому розділі дисертації розглянуто існуючі взаємо-зворотні нормалізуючі перетворення, здійснено вибір перетворення для нормалізації

чотиривимірних даних з метрик програмних застосунків з відкритим кодом на РНР.

У третьому розділі дисертації побудовано трьох-факторне нелінійне регресійне рівняння, нелінійну регресійну модель, інтервали передбачення та довірчі інтервали нелінійної регресії для оцінювання розміру програмних застосунків з відкритим кодом на РНР на основі чотиривимірного перетворення Джонсона сімейства S_B ; здійснено порівняння результатів оцінювання розміру програмних застосунків з відкритим кодом на РНР за нелінійними регресійними моделями, що побудовані на основі чотиривимірного перетворення Джонсона сімейства S_B та на основі одновимірних перетворень.

У четвертому розділі дисертації запропоновано ІТ для оцінювання розміру програмних застосунків з відкритим кодом на РНР. Розроблено інструментарій ІТ для оцінювання розміру програмних застосунків з відкритим кодом на РНР у фазах аналізу та проектування за трьома метриками діаграми класів. Для цього використовувалася система моделювання Scilab. Було розроблено відповідне ПЗ sci-мовою для пакету Scilab 6.0.0 та методичку оцінювання розміру програмних застосунків з відкритим кодом на РНР, яка розрахована на використання зазначеного ПЗ.

Ключові слова: оцінювання розміру, програмний застосунок, РНР, трьох-факторна нелінійна регресійна модель, довірчий інтервал, інтервал передбачення, нормалізуюче перетворення, негаусівські дані, інформаційна технологія.

Список публікацій здобувача за темою дисертації:

1) в яких опубліковані основні наукові результати дисертації:

- статті у періодичних наукових виданнях інших держав, які входять до Організації економічного співробітництва та розвитку та/або

Європейського Союзу, з наукового напрямку, за яким підготовлено дисертацію здобувача:

1. **Ворона, М.** Порівняння результатів оцінювання розміру PHP-застосунків з відкритим кодом за нелінійними регресійними моделями / **М. Ворона** // Deutsche internationale Zeitschrift für zeitgenössische Wissenschaft, German International Journal of Modern Science, № 6, 2021. – С. 43-47. – ISSN 2701-8369. DOI: <https://doi.org/10.24412/2701-8369-2021-6-1-43-47> (“Deutsche internationale Zeitschrift für zeitgenössische Wissenschaft” є періодичним науковим виданням, Німеччина; стаття включена до міжнародних наукометричних баз: Google Scholar, Index Copernicus International)

- статті у наукових виданнях, включених до переліку наукових фахових видань України:

2. Приходько, С.Б. Трьохфакторна нелінійна регресійна модель для оцінювання розміру PHP-застосунків з відкритим кодом / С. Б. Приходько, Н. В. Приходько, Т. А. Фаріонова, **М. В. Ворона** // Науковий журнал «Вчені записки Таврійського національного університету імені В. І. Вернадського. Серія: Технічні науки» Том 31 (70) № 1, 2020. – С. 124-131. DOI: <https://doi.org/10.32838/2663-5941/2020.1-1/23> (**Index Copernicus International**)

3. Приходько, С.Б. Нелінійна регресійна модель для оцінювання розміру Web-застосунків, що створюються з використанням фреймворку Laravel / С. Б. Приходько, Н. В. Приходько, **М. В. Ворона**, І. О. Беловол // Інформаційні технології та комп'ютерна інженерія. – 2021. – Т.50. – № 1. – С. 115-121. – ISSN 1999-9941. DOI: <https://doi.org/10.31649/1999-9941-2021-50-1-115-121>

4. Приходько, С.Б. Оцінювання розміру PHP-застосунків з відкритим кодом за нелінійними регресійними моделями з різними факторами / С. Б. Приходько, **М. В. Ворона** // Збірник наукових праць Національного університету кораблебудування імені адмірала Макарова. – 2021. – №1. – С. 92-98. ISSN: 2311-3405. DOI: [https://doi.org/10.15589/znp2021.1\(484\).13](https://doi.org/10.15589/znp2021.1(484).13)

2) які засвідчують апробацію матеріалів дисертації:

5. **Ворона, М.В.** Оцінювання трудомісткості розробки мобільних додатків за рівнянням нелінійної регресії / **М. В. Ворона**, К. О. Книрик, С. Б. Приходько // Інформатика, інформаційні системи та технології: тези доповідей п'ятнадцятої всеукраїнської конференції студентів і молодих науковців. Одеса, 27 квітня 2018 р. – Одеса, 2018. – С. 95-96.

6. Prykhodko, N. The Non-Linear Regression Model to Estimate the Part of NPLS in the Whole Loan Portfolio of Ukrainian Banks / Natalia Prykhodko, Sergiy Prykhodko, **Mykhaylo Vorona** // Proceedings of the 2018 IEEE First International Conference on SYSTEM ANALYSIS & INTELLIGENT COMPUTING (SAIC), 08-12 October, 2018, Kyiv, Ukraine. – P. 261-265. – ISBN: 978-1-5386-7195-5. DOI: <https://doi.org/10.1109/SAIC.2018.8516899> (включена до міжнародних наукометричних баз: **Scopus, Web of Science, Google Scholar**)

7. Приходько, С.Б. Оцінювання розміру JAVA та PHP-застосунків з відкритим кодом за багатофакторними нелінійними регресійними моделями / С. Б. Приходько, Т. Г. Смикодуб, **М. В. Ворона**, Є. Ю. Беркунський // Прикладні науково-технічні дослідження: матеріали IV міжнар. наук.-практ. конф., 1-3 квіт. 2020 р. – Академія технічних наук України. – Івано-Франківськ: Симфонія форте, 2020. – С. 36-38. – ISBN 978-966-284-132-9 http://ukrtsa.org.ua/media/theses/4-prykladni-naukovo-tekhnichni-doslidzhennia/ATSU2020_T1_P36.pdf

8. **Ворона, М.В.** Математичне моделювання оцінювання розміру Java та PHP-застосунків з відкритим кодом за регресійними моделями / **М. В. Ворона**, С. Б. Приходько, Т. Г. Смикодуб // Інформаційні технології: моделі, алгоритми, системи (ITMAS – 2020): Матеріали Всеукраїнської науково-практичної інтернет конференції (26-28 жовтня 2020 р.). – Миколаїв: НУК імені адмірала Макарова, 2020. – С. 6-9.

9. Приходько, С.Б. Математичне моделювання оцінювання розміру JAVA та PHP-застосунків з відкритим кодом за багатофакторними нелінійними регресійними моделями [Електронний ресурс] / С. Б. Приходько, **М. В. Ворона**, Т. Г. Смикодуб // Комп'ютерне моделювання та оптимізація

складних систем (КМОСС-2020): матеріали VI Міжнародної науково-технічної конференції (м. Дніпро, 4-6 листопада 2020 року) / ДВНЗ УДХТУ. – С. 57-58. – DOI: <https://doi.org/10.32434/CMOCS-2020> – Режим доступу: <http://orgconf.com/infolist/kmoss2020.pdf>

10. **Ворона, М.В.** Оцінювання трудомісткості розробки мобільних додатків за рівнянням нелінійної регресії / **М. В. Ворона**, А. С. Приходько, І. С. Шутко // Інформатика, інформаційні системи та технології: тези доповідей вісімнадцятої всеукраїнської конференції студентів і молодих науковців. Одеса, 23 квітня 2021 р. – Одеса, 2021. – С. 111-113.

11. Приходько, С.Б. Нелінійні регресійні моделі для оцінювання розміру Web-застосунків, що створюються за допомогою PHP-фреймворків / С. Б. Приходько, **М. В. Ворона**, А. С. Приходько, І. С. Шутко // Проблеми моделювання та розроблення інформаційних систем: матеріали V науково-практичної інтернет-конференції (Дрогобич, 23 квітня 2021 року). – Дрогобич: ДДПУ ім. І. Франка, 2021. – С. 28-31. https://docs.google.com/viewer?url=http%3A%2F%2Fist.dp.ua%2Fpublic%2Fdocuments%2F/conferencii%2F/internet-conferencia_V.pdf

ABSTRACT

Vorona M.V. Mathematical models and information technology for estimating the size of open source PHP-based applications. – Manuscript qualification scientific work.

Thesis for the degree of philosophy doctor in specialty 122 "Computer Science" (field 12 – Information technology). – Admiral Makarov National University of Shipbuilding, Mykolaiv, 2021.

The dissertation is devoted to solving an important scientific and practical problem of increasing the confidence of estimating the size of open source PHP-based applications by building appropriate mathematical models in the form of nonlinear regression models and creating information technology (IT) tool for estimating the size of open source PHP-based applications in the early stages of development according to the metrics of the class diagram.

The relevance of this work is related to the following. First, software size (software) estimation is needed to estimate the effort of software application development, for example, using a well-known model such as COCOMO II. Second, early estimation of software size is a difficult task, as limited information is available in the initial stages of software development. Third, existing mathematical models for estimating the size of open source software applications in PHP generally do not contain random variables. All this leads to low confidence in estimating the size of open source software applications in PHP in the initial stages of their development.

The dissertation aims to increase the confidence of estimating the size of open source PHP-based applications at the early stage using nonlinear regression models.

The working scientific hypothesis of the dissertation research is that increasing the confidence of estimating the size of open source PHP-based applications in the phases of analysis and design according to the metrics of the class

diagram is achieved through multiply nonlinear regression models that describe this size as a random variable.

To construct the above nonlinear regression models, it is proposed to use an appropriate method based on multivariate normalizing transformations.

To achieve this aim we need to solve the following tasks:

- analyze existing mathematical models to evaluate the size of open source PHP-based applications;

- to improve the three-factors regression model for estimating the size of mobile application according to the metrics of the class diagram based on four-variate normalizing transformation;

- improve the equation to determine the widths of the confidence interval and the prediction interval of the nonlinear regression of the size of open source PHP-based applications according to the metrics of the class diagram based on four-variate normalizing transformation;

- to develop information technology tools to estimate the size of open source PHP-based applications according to the metrics of the class diagram based on the proposed three-factors nonlinear regression model.

The scientific novelty of the obtained results is as follows.

- 1) The three-factor nonlinear regression equation to estimate the size of open source PHP-based applications at the initial stages of development depending on three factors (number of classes; average number of methods per class, average value of DIT metric per class) based on the four-variate normalizing Johnson transformation of the S_B family has been improved. This allows increasing the confidence of estimating the sample mean of the dependent variable nonlinear regression of the size of open source PHP-based applications in comparison with the use of univariate normalizing transformations. The constructed equation, in comparison with other regression equations, has a smaller value of the mean magnitude of the relative error and a smaller width of the confidence interval of nonlinear regression.

2) The equation for determining the confidence interval of the nonlinear regression of the size of developing open source PHP-based applications based on the four-variate normalizing Johnson transformation of the S_B family has been further developed. This, in comparison with the use of univariate transformations, allows in many cases to reduce the width of the confidence interval of nonlinear regression of size of open source PHP-based applications and thus increase the confidence of estimating the sample mean dependent variable of nonlinear regression of this size.

3) A three-factor nonlinear regression model for estimating the size of open source PHP-based applications depending on number of classes, average number of methods per class, and average value of DIT metric per class based on the four-variate normalizing Johnson transformation of the S_B family, which increases the confidence of estimating the size of open source PHP-based applications as a non-Gaussian random variable using univariate normalizing transformations has been improved. The constructed model, in comparison with other regression models (both linear and nonlinear), has a higher percentage of predicted values, a smaller value of the mean magnitude of the relative error and a smaller width of the prediction interval of nonlinear regression.

4) The equation for determining the boundaries of the interval for predicting three-factor nonlinear regression of the size of open source PHP-based applications based on the four-variate Johnson transformation of the S_B family has been further developed. This, in comparison with the use of univariate transformations, allows in many cases to reduce the width of the prediction interval of nonlinear regression of the size of open source PHP-based applications and thus increase the confidence of estimating this size as the dependent random variable.

The practical significance of the obtained results is as follows. IT tools have been developed to estimate the size of open source PHP-based applications in the phases of analysis and design according to the metrics of the class diagram. The Scilab modeling system (<https://www.scilab.org/>) was used for this purpose. Appropriate software in sci-language has been developed for the Scilab 6.0.0

package. The developed program can be applied to evaluate the size of open source PHP-based applications using three metrics of the class diagram.

The introduction of the dissertation reveals the essence of the scientific and practical task and its significance, substantiates the need for research, presents the general characteristics of the dissertation in the following sequence: relevance of the topic; connection of work with scientific programs, plans, topics; purpose and objectives of the study; scientific novelty and practical significance of the obtained results; personal contribution of the applicant; approbation of dissertation and publication results.

In the first section of the dissertation, the analysis of existing methods and models for estimating the size of open source PHP-based applications is performed and the necessity of researching the chosen topic is substantiated.

In the second section of the dissertation, the existing mutually inverse normalizing transformations are considered, the choice of transformation for normalization of four-dimensional data from metrics of open source PHP-based applications is made.

The third section of the dissertation constructs a three-factor nonlinear regression equation, a nonlinear regression model, prediction intervals, and confidence intervals of nonlinear regression to estimate the size of open source PHP-based applications based on the four-variate Johnson transformation of the S_B family; a comparison of the results of estimating the size of open source PHP-based applications by nonlinear regression models based on the four-variate Johnson transformation of the S_B family and the basis of univariate transformations.

In the fourth section of the dissertation, IT is proposed to estimate the size of open source PHP-based applications. IT tools have been developed to estimate the size of open source PHP-based applications in the phases of analysis and design according to the metrics of the class diagram. The Scilab modeling system was used for this purpose. Appropriate sci-language software has been developed for the Scilab 6.0.0 package.

Keywords: size estimation, software application, PHP, three-factor nonlinear regression model, confidence interval, prediction interval, normalizing transformation, non-Gaussian data, information technology.