

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ КОРАБЛЕБУДУВАННЯ
ІМЕНІ АДМІРАЛА МАКАРОВА
МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Кваліфікаційна наукова праця
на правах рукопису

Ворона Михайло Владиславович

УДК 004.412:519.237.5

ДИСЕРТАЦІЯ

МАТЕМАТИЧНІ МОДЕЛІ ТА ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ДЛЯ
ОЦІНЮВАННЯ РОЗМІРУ ПРОГРАМНИХ ЗАСТОСУНКІВ З ВІДКРИТИМ
КОДОМ НА PHP

Спеціальність 122 – Комп'ютерні науки

Галузь знань 12 – Інформаційні технології

Подається на здобуття наукового ступеня доктора філософії (PhD)

Дисертація містить результати власних досліджень. Використання ідей,
результатів і текстів інших авторів мають посилання на відповідне джерело

_____ М.В. Ворона

Науковий керівник Приходько Сергій Борисович, доктор технічних наук,
професор

Миколаїв – 2021

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 багатовимірних 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 чотиривимірного 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 чотиривимірного 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

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 нелінійними), 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 чотиривимірного нормалізуючого перетворення Джонсона сім'ї S_B . Це у порівнянні з використанням одновимірних перетворень дозволяє в багатьох випадках зменшити ширини 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. Було розроблено відповідне ПЗ *sci*-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 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.

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 РНР-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 РНР-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. Бєловол // 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 багатфакторними 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%2Fconferencii%2Finternet-conferencia_V.pdf

ABSTRACT

Vorona M.V. Mathematical models and information technology for estimating the size of open source PHP-based applications. – Manuscript of the 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 the 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 research is related to the following. First, software size (software) estimation enables 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, the existing mathematical models for estimating the size of open source software applications in PHP generally do not contain random variables. All the above causes low confidence in estimating the size of open source software applications in PHP in the initial stages of their development.

The dissertation aim is 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

at the analysis and design phases according to the metrics of the class diagram is achieved by means of multiply nonlinear regression models that describe this size as a random variable.

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

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

- to analyze the 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;

- to improve the equation for determining 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 for estimating 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 for estimating 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, enables, 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 the 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 at the analysis and design phases according to the metrics of the class diagram. The Scilab modeling system 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 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 Section 1 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 Section 2 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 Section 3 presents the construction of the three-factor nonlinear regression equation, the 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; the 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 Section 4 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 at the analysis and design phases according to the metrics of the class diagram. The Scilab modeling system was used for this purpose. The appropriate sci-language software also 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.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ	19
ВСТУП	21
РОЗДІЛ 1 АНАЛІЗ ІСНУЮЧИХ МЕТОДІВ І МОДЕЛЕЙ ДЛЯ ОЦІНЮВАННЯ РОЗМІРУ ПРОГРАМНИХ ЗАСТОСУНКІВ З ВІДКРИТИМ КОДОМ НА РНР ТА ОБҐРУНТУВАННЯ НЕОБХІДНОСТІ ПРОВЕДЕННЯ ДОСЛІДЖЕНЬ ЗА ОБРАНОЮ ТЕМОЮ	29
1.1 Аналіз існуючих методів і моделей для оцінювання розміру програмних застосунків з відкритим кодом на РНР	29
1.2 Обґрунтування необхідності проведення досліджень за обраною темою	33
1.3 Висновки за розділом 1 та постановка завдань дослідження	39
РОЗДІЛ 2 ВИБІР НОРМАЛІЗУЮЧОГО ПЕРЕТВОРЕННЯ ДЛЯ НОРМАЛІЗАЦІЇ ЧОТИРИВИМІРНИХ ДАНИХ З МЕТРИК ПРОГРАМНИХ ЗАСТОСУНКІВ З ВІДКРИТИМ КОДОМ НА РНР	42
2.1 Існуючі взаємо-зворотні нормалізуючі перетворення	42
2.2.1 Одновимірні перетворення	43
2.2.2 Багатовимірні перетворення	52

2.3 Вибір перетворення для нормалізації чотиривимірних даних з м	56
2.4 Висновки за розділом 2	67
РОЗДІЛ 3 ПОБУДОВА НЕЛІНІЙНИХ РЕГРЕСІЙНИХ МОДЕЛЕЙ ДЛЯ ОЦІНЮВАННЯ РОЗМІРУ ПРОГРАМНИХ ЗАСТОСУНКІВ З ВІДКРИТИМ КОДОМ НА РНР НА ОСНОВІ НОРМАЛІЗУЮЧИХ ПЕРЕТВОРЕНЬ	69
3.1 Побудова рівняння та довірчих інтервалів нелінійної регресії для оцінювання розміру програмних застосунків з відкритим кодом на РНР на основі чотиривимірного перетворення Джонсона сімейства S_V	70
3.2 Побудова нелінійної регресійної моделі та її інтервалів передбачення для оцінювання розміру програмних застосунків з відкритим кодом на РНР на основі чотиривимірного перетворення Джонсона сімейства S_V	78
3.3 Порівняння результатів оцінювання розміру програмних застосунків з відкритим кодом на РНР за нелінійними регресійними моделями, що побудовані на основі чотиривимірного перетворення Джонсона сімейства S_V та на основі одновимірних перетворень	93
3.4 Висновки за розділом 3	104

РОЗДІЛ 4. ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ДЛЯ ОЦІНЮВАННЯ РОЗМІРУ ПРОГРАМНИХ ЗАСТОСУНКІВ З ВІДКРИТИМ КОДОМ НА РНР	106
4.1 Створення інформаційної технології для оцінювання розміру програмних застосунків з відкритим кодом на РНР	106
4.2 Методика оцінювання розміру програмних застосунків з відкритим кодом на РНР	113
4.3 Висновки за розділом 4	126
ВИСНОВКИ.....	128
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	132
Додаток А. Акти впровадження результатів дисертаційної роботи у розробку програмних застосунків.....	146
Додаток Б. Акт впровадження результатів дисертаційної роботи в навчальний процес.....	150
Додаток В. Список публікацій здобувача за темою дисертації та відомості про апробацію результатів дисертації.....	151

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

ММП – метод максимальної правдоподібності;

ОПР – особа, яка приймає рішення;

ПЗ – програмне забезпечення;

ФЩЙ – функція щільності ймовірності;

АС – Afferent Coupling is the number of classes affected by given class;

APP – APPlication;

СВО – Coupling Between Objects is the number of classes coupled to a given class.

BFC – Base Function Components;

COCOMO – the Constructive Cost Model;

COSMIC – Common Software Measurement International Consortium;

DIT – the Depth of Inheritance Tree metric provides for each class a measure of the inheritance levels from the object hierarchy top;

EC – Efferent Coupling is the number of classes from which given class receives effects;

FiSMA – Finnish Software Measurement Association;

FPA – Function Point Analysis;

ISBSG – International Software Benchmarking Standards Group;

KLOC – thousand line of code;

KSLOC – thousand source line of code;

LB – lower bound;

MD – the Mahalanobis distance;

MRE – a magnitude of relative error;

MMRE – a mean magnitude of relative error;

PHP – hypertext preprocessor;

PRED – percentage of prediction;

SMD – the squared Mahalanobis distance;

SS_E – the residual sum of squares (сума квадратів залишків);

SS_R – the regression sum of squares (сума квадратів регресії);

SS_T – the total sum of squares (загальна сума квадратів);

UB – upper bound;

VIF – variance influence factor.

ВСТУП

Актуальність теми. Як відомо, PHP – це популярна мова сценаріїв загального призначення у світі [1] та серед програмістів в Україні [2], яка особливо підходить для веб-розробки (<https://www.php.net/>). Однак ця мова дозволяє не тільки швидко писати веб-сторінки, що динамічно генеруються, але і робити набагато більше, включаючи різні фреймворки, конвертери та інші програмні застосунки.

Задача оцінювання розміру програмного забезпечення (ПЗ) на ранній стадії програмного проекту є важливою, оскільки оцінка розміру ПЗ використовується для прогнозування трудомісткості розробки ПЗ за допомогою математичних моделей, таких як СОСОМО II [3, 4]. Для цього потрібні відповідні математичні моделі у тому числі і регресійні для оцінювання розміру ПЗ, включаючи програмні застосунки з відкритим кодом на PHP [5-10].

На теперішній час для оцінювання кількості строк коду інформаційних систем з відкритим кодом на PHP розроблені як лінійні так і нелінійні регресійні рівняння та моделі в залежності від трьох метрик концептуальної моделі даних у вигляді діаграми класів [5-9]. В [4, 5] трьох-факторне лінійне рівняння регресії для оцінювання кількості строк коду інформаційних систем з відкритим кодом на PHP побудовано на основі методів множинного лінійного регресійного аналізу. Але, як відомо, при побудові лінійних регресійних моделей необхідно виконання певних умов, зокрема похибки повинні бути розподілені за нормальним законом, що має місце лише в поодиноких випадках. А це веде до необхідності побудови нелінійних регресійних моделей, у тому числі і для оцінювання розміру ПЗ, та застосування певних методів множинного нелінійного регресійного аналізу [11].

Тому для оцінювання розміру інформаційних систем з відкритим кодом на РНР в [7-9] були запропоновані нелінійні регресійні рівняння та модель. Запропоновані нелінійні регресійні рівняння та модель були побудовані за допомогою множинного нелінійного регресійного аналізу із застосуванням чотиривимірною перетворення Джонсона сім'ї S_B на основі тих же трьох метрик діаграми класів, що і в [5, 6]: загальна кількість класів, загальна кількість зв'язків та середня кількість атрибутів на клас. Але для програмних застосунків з відкритим кодом на РНР, що не є інформаційними системами, таких як різноманітні інструменти і фреймворки, регресійні моделі можуть залежати від інших метрик [10]. А це потребує розробки відповідних математичних моделей для оцінювання розміру програмних застосунків з відкритим кодом на РНР, які не є інформаційними системами.

Зазвичай для побудови моделей нелінійної регресії використовують одновимірні нормалізуючі перетворення [12-16], що дозволяє у отриманих таким чином моделях уникнути проблеми негаусового характеру помилки [17]. Але їх застосування для побудови нелінійних регресійних моделей не завжди призводить до задовільних результатів оцінювання, насамперед за такими стандартними показниками, як середня величина відносної похибки та відсоток передбачення [7-11]. Також нелінійні регресійні моделі, що побудовані за допомогою одновимірних нормалізуючих перетворень, як правило, мають більші ширини довірчих інтервалів та інтервалів передбачення. Це призводить до необхідності використання багатовимірних нормалізуючих перетворень при побудові нелінійних регресійних моделей, у тому числі і для оцінювання розміру застосунків з відкритим кодом на РНР. Тому удосконалювати нелінійну регресійну модель для оцінювання розміру програмних застосунків з відкритим кодом на РНР ми будемо саме на основі багатовимірних нормалізуючих перетворень із використанням відповідного методу [11].

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційну роботу виконано у Національному університеті кораблебудування імені адмірала Макарова (НУК) відповідно до планів НДР з ініціативної теми «Побудова нелінійних регресійних моделей для оцінювання розміру програмних застосунків з відкритим кодом на РНР» (Державний реєстраційний номер: 0120U100035). У вказаній НДР здобувач був відповідальним виконавцем.

Мета і завдання дослідження. Метою роботи є підвищення достовірності оцінювання розміру програмних застосунків з відкритим кодом на РНР на ранніх стадіях розробки за метриками діаграми класів за допомогою нелінійних регресійних моделей.

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 на ранніх стадіях розробки за метриками діаграми класів на основі багатовимірного нормалізуючого перетворення;
- удосконалити рівняння для визначення границь довірчого інтервалу та інтервалу передбачення нелінійної регресії розміру програмних застосунків з відкритим кодом на РНР на ранніх стадіях розробки за метриками діаграми класів на основі багатовимірного нормалізуючого перетворення;
- на підставі запропонованої трьох-факторної нелінійної регресійної моделі розробити інструментарій інформаційної технології (ІТ) для оцінювання розміру програмних застосунків з відкритим кодом на РНР на ранніх стадіях розробки за метриками діаграми класів.

Об'єктом дослідження є процес оцінювання розміру програмних застосунків з відкритим кодом на РНР.

Предметом дослідження є математичні моделі для оцінювання розміру програмних застосунків з відкритим кодом на РНР.

Методи дослідження. У вирішенні поставлених завдань використано методи теорії ймовірності, математичної статистики, багатовимірного статистичного аналізу, лінійного та нелінійного регресійного аналізу. Методи теорії ймовірності, математичної статистики та багатовимірного статистичного аналізу застосовано для знаходження оцінок характеристик випадкових величин, перевірки статистичних гіпотез щодо нормальності емпіричних розподілів даних (як одновимірних так і багатовимірних), оцінювання параметрів нормалізуючих перетворень. Методи лінійного та нелінійного регресійного аналізу використано для побудови нелінійних регресійних моделей, рівнянь для визначення границь довірчого інтервалу та інтервалу передбачення нелінійної регресії для оцінювання розміру програмних застосунків з відкритим кодом на РНР.

Наукова новизна одержаних результатів полягає у наступному.

Наукова новизна одержаних результатів полягає у наступному.

1) Удосконалено 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 та 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 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, має більший відсоток прогнозування, 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 чотиривимірного нормалізуючого перетворення Джонсона сім'ї S_B . Це у порівнянні з використанням одновимірних перетворень дозволяє в багатьох випадках зменшити ширини 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. Розроблено інструментарій ІТ для оцінювання розміру програмних застосунків з відкритим кодом на РНР на ранніх стадіях розробки 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. Програму, що розроблено, можна використовувати для оцінювання розміру програмних застосунків з відкритим кодом на РНР на ранніх стадіях проектування.

Результати дисертаційної роботи, висновки та рекомендації, що містяться у роботі, схвалені та використовуються в наступних установах: ТОВ «АРТ ПОРТ», Комунальне підприємство «МІСЬКИЙ ІНФОРМАЦІЙНО-ОБЧИСЛЮВАЛЬНИЙ ЦЕНТР», ФОП Погіба О.Р. Завдяки застосуванню інформаційної технології для оцінювання розміру програмних застосунків з відкритим кодом на РНР на ранніх стадіях проектування створеної з використанням розроблених математичних моделей, які враховують реальний негаусівський характер розподілу емпіричних даних, підвищилась достовірність оцінювання розміру програмних РНР-застосунків на ранніх стадіях проектування.

Результати дисертаційної роботи також застосовуються в навчальному процесі на кафедрах програмного забезпечення автоматизованих систем, інформаційних управляючих систем та технологій Національного університету кораблебудування ім. адмірала Макарова (акт впровадження від 10.06.2021р.).

Акти про впровадження результатів дисертаційної роботи наведені у додатках А і Б.

Особистий внесок здобувача. Усі наукові результати отримано здобувачем самостійно. У працях, опублікованих у співавторстві, дисертанту належать: в [10] – трьох-факторна нелінійна регресійна модель для оцінювання розміру програмних застосунків з відкритим кодом на РНР в залежності від кількості

класів, суми кількості класів, на які впливає даний клас (Afferent Coupling) і кількості класів, з яких даний клас отримує ефекти (Efferent Coupling), та кількості методів; в [18] – трьох-факторна нелінійна регресійна модель для оцінювання розміру програмних застосунків з відкритим кодом на РНР в залежності від кількості класів, суми середньої кількості класів, на які впливає даний клас, і середньої кількості класів, з яких даний клас отримує ефекти, та середньої кількості методів; в [19] – трьох-факторні нелінійні регресійні моделі для оцінювання розміру програмних застосунків з відкритим кодом на РНР в залежності від кількості класів, середньої кількості методів на клас, суми середнього аферентного та еферентного зв'язків на клас та в залежності від кількості класів, середньої кількості методів на клас та середнього значення метрики DIT на клас на основі чотиривимірною нормалізуючого перетворення Джонсона сім'ї S_B .

Апробація результатів досліджень. Основні результати дисертації доповідалися та обговорювалися на 7 науково-технічних конференціях, зокрема:

- XV-й та XVIII-й всеукраїнських конференціях студентів і молодих науковців «Інформатика, інформаційні системи та технології» (Одеса, 2018 р., 2021 р.);

- the 2018 IEEE First International Conference on SYSTEM ANALYSIS & INTELLIGENT COMPUTING (SAIC) (Kyiv, Ukraine, 2018);

- IV-й Міжнародній науково-практичній конференції «Прикладні науково-технічні дослідження» (Івано-Франківськ, 2020 р.),

- Всеукраїнській науково-практичній інтернет конференції «Інформаційні технології: моделі, алгоритми, системи» (ITMAS – 2020) (Миколаїв, 2020 р.);

- VI-й Міжнародній науково-технічній конференції «Комп'ютерне моделювання та оптимізація складних систем» (КМОСС-2020) (Дніпро, 2020 р.);

- V-ій науково-практичній інтернет-конференції «Проблеми моделювання та розроблення інформаційних систем» (Дрогобич, 2021 р.).

Публікації. Результати роботи викладені у 11 публікаціях. Основні наукові результати дисертації опубліковано у 4 статтях, з яких одна стаття у періодичному науковому виданні інших держав, які входять до Європейського Союзу, з наукового напрямку, за яким підготовлено дисертацію здобувача, а 3 статті у наукових виданнях, включених до Переліку наукових фахових видань України. Ще 7 публікацій (5 матеріалів і праць та 2 тез конференцій) засвідчують апробацію матеріалів дисертації. Зазначимо, одні праці конференції включені до міжнародних наукометричних баз **Scopus** та **Web of Science**.

Структура і обсяг дисертації. Дисертаційна робота складається з вступу, чотирьох розділів, висновків, списку використаних джерел з 89 найменувань і 3 додатків. Загальний обсяг дисертації становить 153 сторінки, обсяг основного тексту складає 111 сторінок. Дисертаційна робота містить 12 рисунків, 22 таблиці. У додатках наведені документи про впровадження результатів роботи, список публікацій здобувача за темою дисертації та відомості про апробацію результатів дисертації.

РОЗДІЛ 1 АНАЛІЗ ІСНУЮЧИХ МЕТОДІВ І МОДЕЛЕЙ ДЛЯ ОЦІНЮВАННЯ РОЗМІРУ ПРОГРАМНИХ ЗАСТОСУНКІВ З ВІДКРИТИМ КОДОМ НА РНР ТА ОБҐРУНТУВАННЯ НЕОБХІДНОСТІ ПРОВЕДЕННЯ ДОСЛІДЖЕНЬ ЗА ОБРАНОЮ ТЕМОЮ

У першому розділі 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.

1.1 Аналіз існуючих методів і моделей для оцінювання розміру програмних застосунків з відкритим кодом на РНР

Мова програмування РНР є мовою сценаріїв загального призначення, яка широко використовується в веб-розробці [1-2], працює практично на всіх популярних операційних системах, є повністю безкоштовною. Крім того ця мова дозволяє не тільки швидко писати веб-сторінки, що динамічно генеруються, але й робити набагато більше, включаючи різні фреймворки, конвертери та інші програмні застосунки.

Однією з важливих задач з розробки програмного забезпечення (ПЗ) є оцінювання розміру ПЗ на ранніх стадії програмного проекту. Оцінка розміру ПЗ використовується для прогнозування трудомісткості розробки програмних продуктів за допомогою математичних моделей, таких як СОСОМО II [3, 4]. Для цього потрібні відповідні математичні моделі у тому числі і регресійні для

оцінювання розміру ПЗ, включаючи програмні застосунки з відкритим кодом на PHP [5-10].

Після COCOMO II було запропоновано багато варіантів розширень, таких як COCOMO II.2000 [3, 20] та COSYSMO [21]. Аналогічно, SLIM, PUTNAM [22], PRICE-S, SEER та ESTIMACS також можна віднести до такого типу моделей. Зазначені вище моделі є параметричними. До непараметричних моделей, які будуються за методами машинного навчання, слід віднести такі як нейронна мережа та її розширення [23-26].

Незважаючи на достатньо велику кількість існуючих на сьогоднішній день різноманітних методів і моделей для оцінювання розміру ПЗ [5-11, 27-33] роботи в цьому напрямі не припиняються [18, 19, 34, 35]. Це пов'язано в першу чергу з недостатньою точністю оцінювання розміру ПЗ на ранніх етапах його розробки [34, 35]. Один із шляхів вирішення цієї проблеми полягає в побудові відповідних моделей для оцінювання розміру ПЗ, яке розробляється як певною мовою програмування [5-10, 18-27, 30], так і для певного типу застосунків [5-10, 18, 28, 33]. Крім того, для оцінювання розміру ПЗ на ранніх етапах його розробки використовуються метрики UML діаграм, в першу чергу діаграми класів [5-10, 18-27, 30]. Також, як це зазначено в [10, 11], покращити точність оцінювання розміру ПЗ на ранніх етапах його розробки можна за рахунок відповідних математичних моделей, зокрема нелінійних регресійних моделей, які будуються за допомогою багатовимірних нормалізуючих перетворень та інтервалів передбачення. На відміну від одновимірних використання багатовимірних нормалізуючих перетворень дозволяє врахувати кореляцію між залежною і незалежними змінними, а застосування інтервалів передбачення нелінійної регресії при побудові нелінійної регресійної моделі дозволяє позбутися викидів. Зазначене і призводить до підвищення точності оцінювання залежної змінної за допомогою відповідної нелінійної регресійної моделі [36].

Строки коду (LOC) і функціональні точки (FP) [37, 38] найчастіше використовуються як міри розміру в існуючих методах та моделях оцінювання трудомісткості розробки ПЗ. Як відомо [38], обидва ці показники (LOC та FP) мають свої переваги та недоліки при використанні для оцінювання трудомісткості розробки ПЗ. Хоча міра на основі функціональних точок має перевагу перед строками коду у тому, що вона не залежить від технологій, що використовуються, зокрема, мови програмування, і використовується у багатьох методах, які стандартизовані у багатьох стандартах ISO, таких як IFPUG [39] МКІІ FPA [40], NESMA [41] та COSMIC [42], проте оцінювання трудомісткості розробки ПЗ вимагає врахування зазначених факторів (технології та мови програмування), які ще називають факторами навколишнього середовища [38]. Врахування вищезазначених факторів, зокрема, мови програмування, може бути забезпечене відповідними математичними моделями для оцінювання розміру на основі кількості строк коду.

На теперішній час для оцінювання кількості строк коду інформаційних систем з відкритим кодом на РНР розроблені як лінійні так і нелінійні багатофакторні регресійні рівняння та моделі в залежності від трьох метрик концептуальної моделі даних у вигляді діаграми класів [5-9]. В [4, 5] трьох-факторне лінійне рівняння регресії для оцінювання кількості строк коду інформаційних систем з відкритим кодом на РНР побудовано на основі методів множинного лінійного регресійного аналізу. Але, як відомо, при побудові лінійних регресійних моделей необхідно виконання певних умов, зокрема похибки повинні бути розподілені за нормальним законом, що має місце лише в поодиноких випадках. А це веде до необхідності побудови нелінійних регресійних моделей, у тому числі і для оцінювання розміру ПЗ, та застосування певних методів множинного нелінійного регресійного аналізу [11].

Тому для оцінювання розміру інформаційних систем з відкритим кодом на РНР в [7-9] були запропоновані нелінійні регресійні рівняння та модель. Запропоновані нелінійні регресійні рівняння та модель були побудовані за допомогою множинного нелінійного регресійного аналізу із застосуванням чотиривимірного перетворення Джонсона сім'ї S_B на основі тих же трьох метрик діаграми класів, що і в [5, 6]: загальна кількість класів, загальна кількість зв'язків та середня кількість атрибутів на клас. Але для застосунків з відкритим кодом на РНР, що не є інформаційними системами, таких як різноманітні інструменти і фреймворки, регресійні моделі можуть залежати від інших метрик [10]. А це потребує розробки відповідних математичних моделей для оцінювання розміру програмних застосунків з відкритим кодом на РНР, які не є інформаційними системами.

Зазвичай для побудови моделей нелінійної регресії використовують одновимірні нормалізуючі перетворення [12-16]. Але їх застосування для побудови нелінійних регресійних моделей не завжди призводить до задовільних результатів оцінювання, насамперед за такими стандартними показниками, як середня величина відносної похибки та відсоток передбачення [7-11]. Також нелінійні регресійні моделі, що побудовані за допомогою одновимірних нормалізуючих перетворень, як правило, мають більші ширини довірчих інтервалів та інтервалів передбачення. Це призводить до необхідності використання багатовимірних нормалізуючих перетворень при побудові нелінійних регресійних моделей, у тому числі і для оцінювання розміру застосунків з відкритим кодом на РНР. Тому удосконалювати нелінійну регресійну модель для оцінювання розміру програмних застосунків з відкритим кодом на РНР необхідно саме на основі багатовимірних нормалізуючих перетворень із використанням відповідного методу [11].

1.2 Обґрунтування необхідності проведення досліджень за обраною темою

Для обґрунтування необхідності проведення досліджень за обраною темою окрім аналізу публікацій у попередньому підрозділі перевіримо можливість використання багатofакторного лінійного регресійного рівняння для оцінювання розміру програмних застосунків з відкритим кодом на РНР в залежності від метрик, що можуть бути знайдені за діаграмою класів.

Були зібрані дані з метрик 44 програмних застосунків з відкритим кодом на РНР розташованих на сайті GitHub (<https://github.com>): фактична кількість строк коду *LOC*; кількість класів *NOC*; сума кількості класів, на які впливає даний клас, і кількості класів, з яких даний клас отримує ефекти, *AEC* та кількість методів *NOM*. Ці дані були отримані за допомогою інструменту PhpMetrics (<https://phpmetrics.org/>) та наведені у табл. 1.1.

Дані, що наведені в табл. 1.1, мають негаусівський розподіл оскільки для трьох застосунків (1, 2 та 43) значення квадрату відстані Махаланобіса MD^2 , які відповідно дорівнюють 33,67, 34,32 та 27,13, є більшими ніж величина квантиля розподілу χ^2 , що становить 14,86 для рівня значущості 0,005. Також про негаусівський розподіл чотиривимірних даних з табл. 1.1 свідчить оцінка багатовимірного ексцесу β_2 , яка визначалася за [43]. Відомо, що для m -вимірного нормального розподілу $\beta_2 = m(m + 2)$. У нашому випадку $\beta_2 = 24$. Для чотиривимірних даних з табл. 1.1 оцінка β_2 дорівнює 83,11, що майже в 3,5 рази перевищує теоретичне значення.

Також майбутні фактори (*NOC*, *AEC* та *NOM*) були перевірені на наявність мультиколінеарності. Наявність мультиколінеарності свідчить про те, що в

множинній регресійній моделі два або більше факторів пов'язані між собою або мають високий ступінь кореляції [44].

Таблиця 1.1 – Дані з метрик програмних застосунків з відкритим кодом на РНР

№	<i>LOC</i>	<i>NOC</i>	<i>AEC</i>	<i>NOM</i>	№	<i>LOC</i>	<i>NOC</i>	<i>AEC</i>	<i>NOM</i>
1	174927	2075	13332	9979	23	10044	314	1332	699
2	112048	445	1359	1153	24	15477	280	1062	952
3	82551	411	2613	3229	25	15595	115	586	916
4	12022	132	667	767	26	2323	15	74	151
5	5347	5	12	163	27	7101	25	86	235
6	601	25	44	27	28	1431	22	57	75
7	1561	25	101	56	29	37081	278	1219	1637
8	33276	216	2182	1657	30	32826	235	1511	1925
9	36028	126	326	1367	31	12219	58	531	776
10	100245	448	2829	4287	32	59618	568	3570	3393
11	4458	73	381	325	33	24864	363	4165	2031
12	2988	18	126	173	34	2362	28	175	150
13	4047	31	166	196	35	381	8	22	25
14	6688	125	477	353	36	4308	52	256	339
15	1247	2	4	36	37	3412	52	398	173
16	5966	74	359	457	38	15785	126	528	1130
17	38996	269	1720	2067	39	535	3	10	28
18	3269	37	183	189	40	31676	76	568	1072
19	35548	335	3717	1949	41	13940	251	848	856
20	8910	117	437	543	42	3334	16	141	129
21	14019	209	1821	948	43	24298	794	1780	387
22	2920	19	109	155	44	42941	282	1104	810

Як це зазвичай робиться [44], наявність мультиколінеарності будемо визначати за коефіцієнтами впливу дисперсії (VIFs) серед майбутніх предикторів (факторів) в моделі множинної лінійної регресії. Для лінійної моделі множинної

регресії з k -предикторами X_i , $i=1,2,\dots,k$, VIFs – це діагональні елементи оберненої коваріаційної матриці $k \times k$ k -предикторів [44]. Значення VIFs більше за 10 сигналізує нам про те, що дані мають проблеми з мультиколінеарністю. У разі, якщо значення VIFs знаходяться у межах від 1 до 5, то мультиколінеарність відсутня.

Коваріаційная матриця для трьох факторів (*NOC*, *AEC* та *NOM*) за даними табл. 1.1 є такою:

$$\begin{pmatrix} 1,00000 & 0,96669 & 0,91919 \\ 0,96669 & 1,00000 & 0,96698 \\ 0,91919 & 0,96698 & 1,00000 \end{pmatrix}, \quad (1.1)$$

а обернена матриця до (1.1) має вигляд

$$\begin{pmatrix} 16,1865 & -19,4018 & 3,8827 \\ -19,4018 & 38,6533 & -19,5431 \\ 3,8827 & -19,5431 & 16,3289 \end{pmatrix}. \quad (1.2)$$

Елементи на головній діагоналі оберненої матриці (1.2) – значення VIFs, більше за 10. Це вказує на те, що дані мають проблеми з мультиколінеарністю. Для подолання проблеми мультиколінеарності дані з табл. 1.1 були перетворені у такі: фактичний розмір РНР-застосунків з відкритим кодом в тисячах рядків коду Y , загальна кількість класів X_1 , сума середньої кількості класів, на які впливає даний клас, і середньої кількості класів, з яких даний клас отримує ефекти, X_2 та середня кількість методів на клас X_3 . Ці перетворені дані наведені у табл. 1.2.

Коваріаційная матриця для факторів X_1 , X_2 та X_3 за даними табл. 1.2 має вигляд

$$\begin{pmatrix} 1,00000 & 0,12465 & -0,22070 \\ 0,12465 & 1,00000 & -0,00293 \\ -0,22070 & -0,00293 & 1,00000 \end{pmatrix}, \quad (1.3)$$

а обернена матриця до (1.3) така

$$\begin{pmatrix} 1,068 & -0,133 & 0,235 \\ -0,133 & 1,016 & -0,026 \\ 0,235 & -0,026 & 1,052 \end{pmatrix}. \quad (1.4)$$

Таблиця 1.2 – Перетворені дані з метрик програмних застосунків з відкритим кодом на РНР

№	Y	X ₁	X ₂	X ₃	№	Y	X ₁	X ₂	X ₃
1	174,927	2075	6,425	4,809	23	10,044	314	4,242	2,226
2	112,048	445	3,054	2,591	24	15,477	280	3,793	3,400
3	82,551	411	6,358	7,856	25	15,595	115	5,096	7,965
4	12,022	132	5,053	5,811	26	2,323	15	4,933	10,067
5	5,347	5	2,400	32,600	27	7,101	25	3,440	9,400
6	0,601	25	1,760	1,080	28	1,431	22	2,591	3,409
7	1,561	25	4,040	2,240	29	37,081	278	4,385	5,888
8	33,276	216	10,102	7,671	30	32,826	235	6,430	8,191
9	36,028	126	2,587	10,849	31	12,219	58	9,155	13,379
10	100,245	448	6,315	9,569	32	59,618	568	6,285	5,974
11	4,458	73	5,219	4,452	33	24,864	363	11,474	5,595
12	2,988	18	7,000	9,611	34	2,362	28	6,250	5,357
13	4,047	31	5,355	6,323	35	0,381	8	2,750	3,125
14	6,688	125	3,816	2,824	36	4,308	52	4,923	6,519
15	1,247	2	2,000	18,000	37	3,412	52	7,654	3,327
16	5,966	74	4,851	6,176	38	15,785	126	4,190	8,968
17	38,996	269	6,394	7,684	39	0,535	3	3,333	9,333
18	3,269	37	4,946	5,108	40	31,676	76	7,474	14,105
19	35,548	335	11,096	5,818	41	13,940	251	3,378	3,410
20	8,910	117	3,735	4,641	42	3,334	16	8,813	8,063
21	14,019	209	8,713	4,536	43	24,298	794	2,242	0,487
22	2,920	19	5,737	8,158	44	42,941	282	3,915	2,872

Оскільки елементи на головній діагоналі оберненої матриці (1.4) – значення VIFs, менше за 5, це вказує на відсутність мультиколінеарності факторів X_1 , X_2 та X_3 .

Зазначимо, що розподіл даних з табл. 1.2 також є негаусівським тому, що для трьох застосунків (1, 2 та 5) значення MD^2 , які відповідно дорівнюють 32,57, 17,75 та 26,38, є більшими ніж величина квантиля розподілу χ^2 , що становить 14,86 для рівня значущості 0,005. Про негаусівський розподіл даних з табл. 1.2 також свідчить оцінка багатовимірного ексцесу β_2 , що дорівнює 59,21. Це значення більш ніж вдвічі перевищує теоретичне, що у нашому випадку дорівнює 24. Також про негаусівський розподіл даних з табл. 1.2 свідчить і оцінка багатовимірної асиметрії β_1 .

Квадрат відстані Махаланобісу визначався як

$$d_i^2 = (\mathbf{P}_i - \bar{\mathbf{P}})^T \mathbf{S}_P^{-1} (\mathbf{P}_i - \bar{\mathbf{P}}), \quad (1.5)$$

де $\mathbf{P}_i$ – вектор, який визначає i -ту точку багатовимірних даних, $i=1,2,\dots,N$, $\mathbf{P} = \{Y, X_1, X_2, \dots, X_k\}^T$; k – це кількість незалежних змінних (факторів) X_j , $j=1,2,\dots,k$; $\bar{\mathbf{P}}$ – вектор вибірових середніх; $\mathbf{S}_P$ – вибіркова коваріаційна матриця для компонентів вектору $\mathbf{P}$ [44]

$$\mathbf{S}_P = \frac{1}{N} \sum_{i=1}^N (\mathbf{P}_i - \bar{\mathbf{P}})(\mathbf{P}_i - \bar{\mathbf{P}})^T. \quad (1.6)$$

Оцінка багатовимірного ексцесу β_2 визначалася як [43]

$$\hat{\beta}_2 = \frac{1}{N} \sum_{i=1}^N \left\{ (\mathbf{P}_i - \bar{\mathbf{P}})^T \mathbf{S}_P^{-1} (\mathbf{P}_i - \bar{\mathbf{P}}) \right\}^2.$$

Оцінка багатовимірної асиметрії β_1 визначалася як [43]

$$\hat{\beta}_1 = \frac{1}{N^2} \sum_{i=1}^N \left\{ (\mathbf{P}_i - \bar{\mathbf{P}})^T \mathbf{S}_P^{-1} (\mathbf{P}_i - \bar{\mathbf{P}}) \right\}^3.$$

В подальшому чотиривимірні негаусівські дані, що наведені у табл. 1.2, використовуються для побудови трьох-факторної нелінійної регресійної моделі для оцінювання розміру програмних застосунків з відкритим кодом на РНР. Але спочатку для остаточного обґрунтування необхідності її побудови отримуємо лінійну регресійну модель для оцінювання розміру РНР-застосунків з відкритим кодом у вигляді

$$Y = \hat{b}_0 + \hat{b}_1 X_1 + \hat{b}_2 X_2 + \hat{b}_3 X_3 + \varepsilon, \quad (1.7)$$

де оцінки параметрів такі: $\hat{b}_0 = -3,1977$, $\hat{b}_1 = 0,0886$, $\hat{b}_2 = 0,7045$, $\hat{b}_3 = 0,6840$; ε – випадкова величина з розподілом Гаусу, $\varepsilon \sim N(0, \sigma_\varepsilon^2)$. Сума квадратів відхилень для моделі (1.7) склала 14465,3.

Оцінки параметрів знаходилися за методом найменших квадратів за емпіричними даними шляхом рішення наступної задачі математичного програмування

$$\hat{\theta} = \arg \min_{\theta} \left\{ \sum_{i=1}^N [Y_i - \hat{Y}_i]^2 \right\}, \quad (1.8)$$

де θ – вектор параметрів, що потребують оцінювання, $\theta = \{b_0, b_1, b_2, \dots, b_k\}$; $\hat{\theta}$ – вектор оцінок параметрів, $\hat{\theta} = \{\hat{b}_0, \hat{b}_1, \hat{b}_2, \dots, \hat{b}_k\}$; k – кількість факторів або незалежних змінних; N – кількість точок емпіричних даних; Y_i – i -те значення випадкової величини Y ; $\hat{Y}_i$ – i -те значення оцінки $\hat{Y}$, $\hat{Y}_i = \hat{b}_0 + \hat{b}_1 X_{1i} + \hat{b}_2 X_{2i} + \dots + \hat{b}_k X_{ki}$.

Перевірку нульової гіпотези про нормальність закону розподілу випадкової величини ε для моделі (1.7) здійснюємо за критерієм Пірсона. Для вибірки значень випадкової величини ε значення χ^2 , яке дорівнює 52,04, більше за $\chi_{кр}^2$, що становить 7,81 для 3 ступенів вільності та 0,05 рівня значущості. Тобто,

цю гіпотезу про нормальність розподілу випадкової величини ε потрібно відкинути. Це свідчить про відсутність теоретичного обґрунтування можливості використання моделі множинної лінійної регресії (1.7) і призводить до необхідності побудови трьох-факторної нелінійної регресійної моделі для оцінювання розміру програмних застосунків з відкритим кодом на РНР.

Таким чином, виникає потреба у проведенні досліджень за обраною темою.

1.3 Висновки за розділом 1 та постановка завдань дослідження

У даному розділі наведено огляд існуючих методів та моделей для оцінювання розміру програмних застосунків з відкритим кодом на РНР, здійснено обґрунтування необхідності проведення досліджень за обраною темою.

1) Зараз для оцінювання трудомісткості програмних застосунків у тому числі з відкритим кодом на РНР використовують такі моделі як СОСОМО ІІ, СОSYSМО та інші, основним фактором в яких є розмір ПЗ. А розмір ПЗ на початкових етапах розробки ще є не відомим. Тому для його оцінювання на початкових етапах створюють відповідні математичні моделі. Ці математичні моделі, як правило є лінійними і не містять у своєму складі випадкових складових. Це у свою чергу не дозволяє враховувати ту обставину, що розмір програмних застосунків є негаусівською випадковою величиною та веде до низької достовірності оцінювання розміру програмних застосунків з відкритим кодом на РНР на початкових етапах.

2) Існуючі регресійні моделі для оцінювання розміру програмних застосунків з відкритим кодом на РНР, в основному є лінійними. Використання

моделей лінійної регресії потребує виконання певних припущень, одним з яких є нормальність розподілу помилок. Але це припущення справедливе лише в поодиноких випадках. Підтверджено, що це припущення не виконується і для лінійної регресійної моделі, яка побудова за даними з метрик програмних застосунків з відкритим кодом на РНР, що були зібрані Це призводить до необхідності розробки нелінійних регресійних моделей, в тому числі для оцінювання розміру програмних застосунків з відкритим кодом на РНР розташованих на сайті GitHub (<https://github.com>). Ці дані з метрик програмних застосунків були отримані за допомогою інструменту PhpMetrics (<https://phpmetrics.org/>) та мають негаусівський розподіл.

3) Нормалізуючі перетворення часто є добрим способом побудови моделей, довірчих інтервалів та інтервалів передбачення нелінійних регресій. Причому зазвичай у якості нормалізуючого перетворення використовують десятковий логарифм. Наприклад, така відома модель як СОСОМО II була створена саме із використанням одновимірного нормалізуючого перетворення у вигляді десятичного логарифму. Але використання одновимірних нормалізуючих перетворень не дозволяє врахувати наявність кореляції між змінними при нормалізації даних, що як правило призводить до поганих результатів нормалізації і як наслідок – застосування одновимірних нормалізуючих перетворень для побудови нелінійних регресійних моделей не завжди призводить до задовільних результатів оцінювання, насамперед за такими стандартними Тому удосконалювати регресійні моделі для оцінювання розміру програмних застосунків з відкритим кодом на РНР, доцільно саме на основі багатовимірних нормалізуючих перетворень із врахуванням наявності викидів у даних.

Метою роботи є підвищення достовірності оцінювання розміру програмних застосунків з відкритим кодом на РНР на ранніх стадіях розробки за метриками діаграми класів за допомогою нелінійних регресійних моделей.

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 на ранніх стадіях розробки за метриками діаграми класів на основі багатовимірного нормалізуючого перетворення;
- удосконалити рівняння для визначення границь довірчого інтервалу та інтервалу передбачення нелінійної регресії розміру програмних застосунків з відкритим кодом на РНР на ранніх стадіях розробки за метриками діаграми класів на основі багатовимірного нормалізуючого перетворення;
- на підставі запропонованої трьох-факторної нелінійної регресійної моделі розробити інструментарій інформаційної технології (ІТ) для оцінювання розміру програмних застосунків з відкритим кодом на РНР на ранніх стадіях розробки за метриками діаграми класів.

Результати досліджень поточного розділу автором опубліковано в [10, 46-52].

РОЗДІЛ 2 ВИБІР НОРМАЛІЗУЮЧОГО ПЕРЕТВОРЕННЯ ДЛЯ НОРМАЛІЗАЦІЇ ЧОТИРИВИМІРНИХ ДАНИХ З МЕТРИК ПРОГРАМНИХ ЗАСТОСУНКІВ З ВІДКРИТИМ КОДОМ НА RHP

У другому розділі розглянуто існуючі взаємо-зворотні нормалізуючі перетворення, здійснено вибір перетворення для нормалізації чотиривимірних даних з метрик програмних застосунків з відкритим кодом на RHP з метою його подальшого застосування для побудови математичних моделей для оцінювання розміру RHP-застосунків.

2.1 Існуючі взаємо-зворотні нормалізуючі перетворення

На сьогодні відомі дві групи взаємо-зворотних нормалізуючих перетворень: одновимірні та багатовимірні. Одновимірні нормалізуючі перетворення дозволяють перетворити негаусівську випадкову величину у гаусівську або величину, яка наближено може вважатися гаусівською. Багатовимірні нормалізуючі перетворення дозволяють перетворити систему негаусівських випадкових величин у систему гаусівських випадкових величин або систему випадкових величин, які наближено можуть вважатися гаусівськими. При цьому у разі використання багатовимірних нормалізуючих перетворень для нормалізації багатовимірних даних на відміну від одновимірних враховується кореляція між випадковими величинами.

2.2.1 Одновимірні перетворення

Серед відомих взаємо-зворотних нормалізуючих перетворень слід в першу чергу виділити логарифмічне перетворення, перетворення Бокса-Кокса та перетворення Джонсона.

Логарифмічне перетворення для логнормального розподілу.

Логарифмічне перетворення [54, 55]

$$z = \ln x \quad (2.1)$$

або

$$z = \ln(1 + x) \quad (2.2)$$

використовуються для отримання значень випадкової величини z з розподілом наближеним до нормального та з незв'язаними вибірковими середнім і дисперсією, в той час як стандартне відхилення початкової випадкової величини x пропорційним середньому. Якщо коефіцієнт пропорційності k , то дисперсія величини z наближено дорівнює k^2 . Формула (2.2) краща за (2.1) в тому випадку, коли деякі значення величини x можуть дорівнювати нулю. У якості основи логарифму можна взяти 10, тоді дисперсія буде дорівнювати $0,1886k^2$, а саме перетворення буде таким:

$$z = \lg x. \quad (2.3)$$

Зворотне перетворення до перетворення (2.3) має вигляд

$$x = 10^z. \quad (2.4)$$

Перетворення (2.3) і (2.4) було застосовано для побудови таких відомих моделей як СОСОМО та ISBSG [56-58] для визначення тривалості розробки програмних проектів, а також – СОСОМО II для оцінювання трудомісткості розробки ПЗ.

Перетворення Бокса-Кокса

Перетворення Бокса-Кокса було запропоновано у 1964 р. і задається як [59]

$$x(\lambda) = \begin{cases} \frac{x^\lambda - 1}{\lambda}, & \text{якщо } \lambda \neq 0; \\ \ln(x), & \text{якщо } \lambda = 0. \end{cases} \quad (2.5)$$

Тут x – це випадкова величина, що спостерігається; λ – це невідомий параметр; $x(\lambda)$ – це перетворена випадкова величина, яка виражається у формі лінійної моделі

$$x(\lambda) = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon},$$

де $\mathbf{X}$ – невідома $n \times p$ матриця; $\boldsymbol{\theta}$ – $p \times 1$ вектор параметрів, пов'язаних з перетворенням. У контексті лінійної моделі вважається, що вектор випадкової похибки $\boldsymbol{\varepsilon}$ має багатомірний нормальний розподіл з вектором математичного сподівання $\mathbf{0}$ і коваріаційною матрицею $\sigma^2 \mathbf{I}_n$.

Перетворення (2.5) є модифікацією степеневого перетворення, представленого Тюкі (Tukey) у 1957 р. [60, 61]

$$x(\lambda) = \begin{cases} x^\lambda, & \text{якщо } \lambda \neq 0; \\ \ln(x), & \text{якщо } \lambda = 0. \end{cases}$$

Перетворення (2.5) створено з урахуванням розриву у $\lambda = 0$. Перетворення (2.5) придатне лише для додатних значень випадкової величини x . Для додатних та від'ємних значень випадкової величини x було запропоновано розширене перетворення Бокса-Кокса

$$x(\lambda) = \begin{cases} \frac{(x + \lambda_2)^{\lambda_1} - 1}{\lambda_1}, & \text{якщо } \lambda_1 \neq 0; \\ \ln(x + \lambda_2), & \text{якщо } \lambda_1 = 0. \end{cases} \quad (2.6)$$

В (2.6) $\lambda = (\lambda_1, \lambda_2)^T$. На практиці ми можемо вибрати λ_2 таким, що $x + \lambda_2 > 0$ для деяких x .

Зрозуміло, не всі дані можуть бути перетворені за степеневими перетвореннями (2.5) або (2.6) в такі, що мають нормальний розподіл. В 1969 р. Драпер (Draper) і Кокс (Cox) [62] вивчили цю проблему та прийшли до висновку, що навіть у разі, коли степеневе перетворення не може приводити розподіл до точно нормального, зазвичай оцінки параметра λ будуть приводити до розподілу, яке задовольняє певним обмеженням за першими чотирма статистичним моментам, отже буде зазвичай симетричним.

Для оцінювання параметра λ Бокс і Кокс у 1964 р. [59] запропонували використовувати метод максимальної правдоподібності, а також метод Байєса. Вони вважають, перш за все, що

$$\hat{\theta}(\lambda) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}' x(\lambda)$$

та

$$\hat{\sigma}^2(\lambda) = x'(\lambda) \left[\mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}' \right] x(\lambda) \quad (2.7)$$

є оцінками відповідно θ і σ^2 . Далі вони оцінювали λ шляхом максимізації логарифму функції правдоподібності

$$l(\lambda) = \frac{-n \ln \hat{\sigma}^2(\lambda)}{2 + \ln J(\lambda, x)},$$

де

$$J(\lambda, x) = \prod_{i=1}^n x_i^{\lambda-1}.$$

Оцінка максимальної правдоподібності λ підставляється у (2.7), яка визначає оцінки інших параметрів.

Отже для перетворення (2.5) параметр λ визначається за методом максимальної правдоподібності (ММП) шляхом максимізації логарифму функції правдоподібності [63]

$$l(\lambda) = C - \frac{n}{2} \ln \sum_{i=1}^n \frac{(x_i(\lambda) - \bar{x}(\lambda))^2}{n} + (\lambda - 1) \sum_{i=1}^n \ln(x_i), \quad (2.8)$$

де C – постійна, яку визначають з умови нормування; $\bar{x}(\lambda) = \sum_{i=1}^n x_i(\lambda) / n$.

Перетворення Джонсона.

Ідея перетворення одної випадкової величини в іншу, яка має більш просту щільність розподілу (в першу чергу нормальну), була розвинена Джонсоном. У 1949 році Джонсон представив нормалізуюче перетворення з трьох сімей, яким відповідають відповідні розподіли. В [64-66] рекомендується застосовувати саме розподіли Джонсона тому, що вони побудовані на основі нормалізуючих перетворень, які дозволяють перейти від значень негаусівської випадкової величини до нормально розподіленої випадкової величини. А історично склалося так, що велика кількість статистичних критеріїв і методів розроблені в основному тільки для випадку нормального розподілу.

Сім'ї розподілів Джонсона отримані шляхом перетворення нормованої нормально розподіленої випадкової величини z з нульовим математичним сподіванням і одиничною дисперсією. В загальному випадку перетворення Джонсона має вигляд [65]

$$z = \gamma + \eta h(x, \varphi, \lambda); \quad \eta > 0; \quad -\infty < \gamma < \infty; \quad \lambda > 0; \quad -\infty < \varphi < \infty, \quad (2.9)$$

де h – довільна функція; γ , η , λ , φ – параметри розподілу Джонсона. Слід відзначити, що γ і η – це параметри форми, λ – це параметр масштабу, φ – це параметр положення, характеризує зсув розподілу у напрямку осі абсцис.

Джонсон запропонував три різні сім'ї функцій h :

$$1) \quad h(x, \varphi, \lambda) = \ln\left(\frac{x - \varphi}{\lambda}\right), \quad x > \varphi \text{ (сім'я } S_L);$$

$$2) \quad h(x, \varphi, \lambda) = \ln\left(\frac{x - \varphi}{\lambda + \varphi - x}\right), \quad \varphi < x < \varphi + \lambda \text{ (сім'я } S_B);$$

$$3) \quad h(x, \varphi, \lambda) = \text{Arsh}\left(\frac{x - \varphi}{\lambda}\right), \quad -\infty \leq x \leq +\infty \text{ (сім'я } S_U),$$

$$\text{де } \text{Arsh}\left(\frac{x - \varphi}{\lambda}\right) = \ln\left[\frac{x - \varphi}{\lambda} + \sqrt{\left(\frac{x - \varphi}{\lambda}\right)^2 + 1}\right].$$

Якщо ввести позначення $\tilde{x} = (x - \varphi)/\lambda$, то функції h для сімей розподілів Джонсона можна записати як

$$h = \begin{cases} \ln(\tilde{x}), & x > \varphi, & \text{для сім'ї } S_L; \\ \ln[\tilde{x}/(1 - \tilde{x})], & \varphi < x < \varphi + \lambda, & \text{для сім'ї } S_B; \\ \text{Arsh}(\tilde{x}), & -\infty \leq x \leq +\infty, & \text{для сім'ї } S_U. \end{cases}$$

Зворотне до (2.9) перетворення можна визначити як

$$x = \varphi + \lambda h^{-1}(z, \gamma, \eta); \quad \eta > 0; \quad -\infty < \gamma < \infty; \quad \lambda > 0; \quad -\infty < \varphi < \infty, \quad (2.10)$$

де

$$h^{-1} = \begin{cases} e^\zeta, & \text{для сім'ї } S_L; \\ 1/(1 + e^{-\zeta}), & \text{для сім'ї } S_B; \\ (e^\zeta - e^{-\zeta})/2, & \text{для сім'ї } S_U. \end{cases}$$

$$\text{В (2.10) } \zeta = (z - \gamma)/\eta.$$

Для сім'ї S_L функція щільності ймовірності задається як

$$f_L(x) = \frac{\eta}{\sqrt{2\pi}(x-\varphi)} \exp\left\{-\frac{\eta^2}{2}\left[\frac{\gamma - \eta \ln \lambda}{\eta} + \ln(x-\varphi)\right]^2\right\},$$

де $x > \varphi$; $\eta > 0$; $-\infty < \gamma < \infty$; $\lambda > 0$; $-\infty < \varphi < \infty$.

Ця функція щільності ймовірності зростає від φ до точки максимуму і потім більш повільно спадає у разі спрямування x у нескінченність ($x \rightarrow \infty$). Функція щільності ймовірності для сім'ї S_L завжди унімодальна, асиметрія завжди додатна, а ексцес більший за 3 (значення ексцесу нормального розподілу).

Для сім'ї S_B функція щільності ймовірності задається формулою

$$f_B(x) = \frac{\eta\lambda}{\sqrt{2\pi}(x-\varphi)(\lambda+\varphi-x)} \exp\left\{-\frac{1}{2}\left[\gamma + \eta \ln\left(\frac{x-\varphi}{\lambda+\varphi-x}\right)\right]^2\right\},$$

де $\varphi < x < \varphi + \lambda$; $\eta > 0$; $-\infty < \gamma < \infty$; $\lambda > 0$; $-\infty < \varphi < \infty$.

Функції щільності ймовірності сім'ї S_B можуть бути як унімодальними, так і бімодальними. Необхідні та достатні умови для бімодальності полягають у тому, що

$$\eta < 1/\sqrt{2}, \quad |\gamma| < \eta^{-1}\sqrt{1-2\eta^2} - 2\eta \operatorname{arth}\sqrt{1-2\eta^2}.$$

Для сім'ї S_U функція щільності ймовірності визначається як

$$f_U(x) = \frac{\eta}{\sqrt{2\pi}\{(x-\varphi)^2 + \lambda^2\}} \exp\left\{-\frac{1}{2}\left[\gamma + \eta \ln\left(\frac{x-\varphi}{\lambda} + \sqrt{\left(\frac{x-\varphi}{\lambda}\right)^2 + 1}\right)\right]^2\right\},$$

де $-\infty < x < \infty$; $\eta > 0$; $-\infty < \gamma < \infty$; $\lambda > 0$; $-\infty < \varphi < \infty$.

Графіки функції щільності ймовірності сім'ї S_U мають високий порядок стику з віссю абсцис на кінцях інтервалу зміни аргументу, є унімодальними та їх моди лежать між медіаною та нулем. Тім самим ці розподіли мають додатну або

від'ємну асиметрію в залежності від того, від'ємна або додатна величина параметра γ .

Сім'ї розподілів Джонсона відрізняються багатостатністю форм і у площині асиметрії у квадраті A^2 та ексцесу ε займають значні області. На рис.2.1 представлена діаграма Джонсона (області комбінацій A^2 і ε для різних розподілів Джонсона). Ця діаграма дозволяє підібрати сім'ю розподілів Джонсона за значеннями оцінок A^2 та ε вибіркового розподілу.

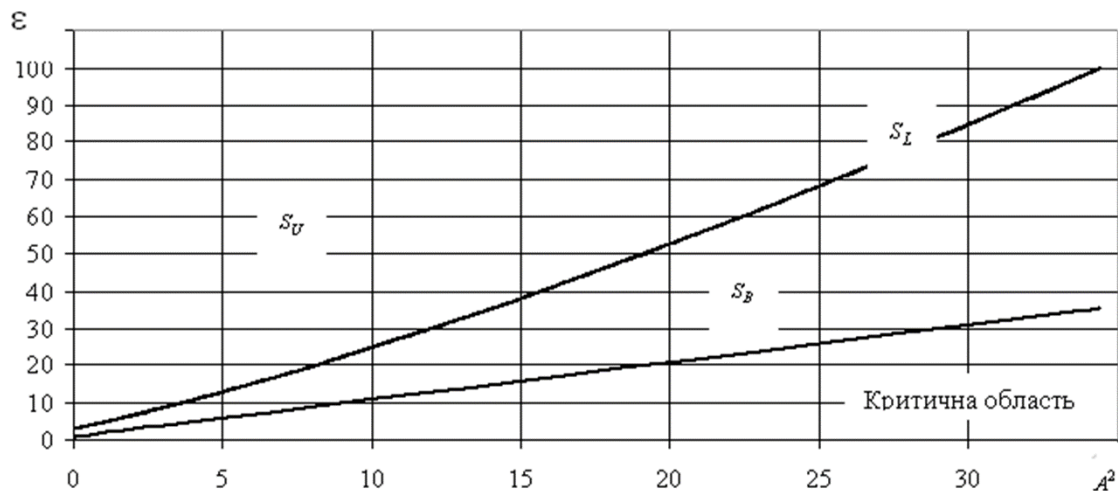


Рисунок 2.1 – Комбінації A^2 і ε для різних розподілів Джонсона [67]

Лінія верхньої границі критичної області задається залежністю

$$\varepsilon = A^2 + 1.$$

Якщо комбінація A^2 і ε знаходиться біля лінії S_L , то вибіркового розподіл можна апроксимувати розподілом із сім'ї S_L (логарифмічно нормальним розподілом). Розподіл зі значеннями A^2 і ε , які лежать вище лінії S_L , апроксимуються розподілом із сім'ї S_U , а які лежать нижче лінії S_L (до лінії критичної області) — розподілом із сім'ї S_B . Якщо комбінація A^2 і ε попадає в

критичну область, то вибірковий розподіл не можна апроксимувати розподілом із сімей Джонсона.

Лінію S_L в межах $A^2 \in [0, 27]$ можна апроксимувати наступною функцією [68]:

$$\varepsilon(A^2) = 7,2315 \cdot 10^{-6} A^8 - 6,9860 \cdot 10^{-4} A^6 + 4,5460 \cdot 10^{-2} A^4 + \\ + 1,7979 A^2 + 2,9891.$$

Лінію S_L в межах $A^2 \in [0, 40]$ можна апроксимувати такою функцією [69]:

$$\varepsilon(A^2) = 4,59 \cdot 10^{-6} A^8 - 4,8805 \cdot 10^{-4} A^6 + 4,1655 \cdot 10^{-2} A^4 + \\ + 1,8203 A^2 + 2,9658.$$

Для сім'ї розподілів S_L статистичні моменти $\alpha_1(x)$, $\mu_2(x)$, $\mu_3(x)$ і $\mu_4(x)$ можна представити у вигляді наступних виразів:

$$\alpha_1(x) = \varphi + \lambda \omega \rho;$$

$$\mu_2(x) = \lambda^2 \omega^2 \rho^2 (\omega^2 - 1);$$

$$\mu_3(x) = \lambda^3 \omega^3 \rho^3 (\omega^6 - 3\omega^2 + 2);$$

$$\mu_4(x) = \lambda^4 \omega^4 \rho^4 (\omega^2 - 1)(\omega^8 + 2\omega^6 + 3\omega^4 - 3),$$

де $\omega = \exp(0,5/\eta^2)$; $\rho = \exp(-\gamma/\eta)$.

Тоді параметри функції щільності ймовірності γ , η , λ та φ для сім'ї S_L розподілів Джонсона можна знайти із рішення наступної системи рівнянь:

$$\varphi + \lambda \omega \rho = \hat{\alpha}_1(x);$$

$$\lambda^2 \omega^2 \rho^2 (\omega^2 - 1) = \hat{\mu}_2(x);$$

$$\lambda^3 \omega^3 \rho^3 (\omega^6 - 3\omega^2 + 2) = \hat{\mu}_3(x);$$

$$\lambda^4 \omega^4 \rho^4 (\omega^2 - 1)(\omega^8 + 2\omega^6 + 3\omega^4 - 3) = \hat{\mu}_4(x).$$

Найпростішу процедуру підгонки параметрів для сім'ї розподілів S_L запропонував Уіксел і вона полягає у наступному. Спочатку знаходять $t = (\omega^2 - 1)^{0,5}$ як додатній корінь рівняння

$$t^3 + 3t - A = 0.$$

Після чого параметр зсуву φ визначається за формулою

$$\varphi = \alpha_1(x) - \frac{\sqrt{\mu_2(x)}}{t}.$$

Параметр η визначається значенням ω , а саме

$$\eta = \frac{1}{\sqrt{2 \ln \omega}}.$$

Вираз $\gamma - \eta \ln \lambda$ розглядається як єдиний параметр і обчислюється після логарифмування добутку λp .

Для сім'ї розподілів S_U статистичні моменти $\alpha_1(y)$, $\mu_2(y)$, $\mu_3(y)$ і $\mu_4(y)$ можна представити у вигляді наступних виразів:

$$\begin{aligned} \alpha_1(y) &= -\omega sh\Omega; \\ \mu_2(y) &= \frac{1}{2}(\omega^2 - 1)(\omega^2 ch2\Omega + 1); \\ \mu_3(y) &= -\frac{1}{4}\omega^2(\omega^2 - 1)^2 \{\omega^2(\omega^2 + 2)sh3\Omega + 3sh\Omega\}; \\ \mu_4(y) &= \frac{1}{8}(\omega^2 - 1)^2 \{\omega^4(\omega^8 + 2\omega^6 + 3\omega^4 - 3)ch4\Omega + \\ &\quad + 4\omega^4(\omega^2 + 2)ch2\Omega + 3(2\omega^2 + 1)\}, \end{aligned} \tag{2.11}$$

де $y = (x - \varphi)/\lambda$; $\omega = \exp(0,5/\eta^2)$; $\Omega = \gamma/\eta$; $sh\Omega = 0,5(e^\Omega - e^{-\Omega})$;
 $ch\Omega = 0,5(e^\Omega + e^{-\Omega})$.

Оцінки параметрів перетворення Джонсона краще визначати за ММП

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} l(\boldsymbol{\theta}), \quad (2.12)$$

де $l(\boldsymbol{\theta})$ – логарифмічна функція правдоподібності. Для перетворення Джонсона сім'ї S_U логарифмічну функцію правдоподібності можна записати як [67]

$$l(\boldsymbol{\theta}) = N \ln(\eta) - \frac{N \ln(2\pi)}{2} - \frac{1}{2} \sum_{i=1}^N \ln[(x_i - \varphi)^2 + \lambda^2] - \\ - \frac{1}{2} \sum_{i=1}^N \left[\gamma + \eta \operatorname{Arsh} \left(\frac{x_i - \varphi}{\lambda} \right) \right]^2.$$

Для перетворення Джонсона сім'ї S_B , логарифмічну функцію правдоподібності можна записати як [47]

$$l(\boldsymbol{\theta}) = N \ln(\eta\lambda) - \frac{N \ln(2\pi)}{2} - \sum_{i=1}^N \ln(x_i - \varphi) - \sum_{i=1}^N \ln(\varphi + \lambda - x_i) - \\ - \frac{1}{2} \sum_{i=1}^N \left[\gamma + \eta \ln \frac{x_i - \varphi}{\varphi + \lambda - x_i} \right]^2. \quad (2.13)$$

Зауважимо, при оцінюванні параметрів γ , η , λ та φ за ММП потрібно враховувати відповідні обмеження, що накладаються на ці параметри для обраної сім'ї розподілів Джонсона.

В подальшому при застосуванні функції (2.13) замість x_i у (2.13) слід підставляти або Y_i , або X_{ji} .

2.2.2 Багатовимірні перетворення

Серед відомих взаємо-зворотних багатовимірних нормалізуючих перетворень слід в першу чергу виділити перетворення Бокса-Кокса та перетворення Джонсона.

Багатовимірне перетворення Джонсона

Багатовимірне перетворення Джонсона записується як [70, 71]

$$\mathbf{z} = \boldsymbol{\gamma} + \boldsymbol{\eta} \mathbf{h}[\boldsymbol{\lambda}^{-1}(\mathbf{x} - \boldsymbol{\varphi})], \quad (2.14)$$

де $\mathbf{z}$ – вектор нормалізованих випадкових величин, $\mathbf{z} = (z_1, z_2, \dots, z_m)^T$; $\mathbf{h}[(y_1, y_2, \dots, y_m)^T] = \{h_1(y_1), h_2(y_2), \dots, h_m(y_m)\}^T$; $h_i(\cdot)$ – функція для одновимірного перетворення Джонсона для i -ої компоненти, $i \in [1, m]$; $\boldsymbol{\gamma}$, $\boldsymbol{\eta}$, $\boldsymbol{\varphi}$ та $\boldsymbol{\lambda}$ – параметри розподілу Джонсона, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_m)^T$, $\boldsymbol{\eta} = \text{diag}(\eta_1, \eta_2, \dots, \eta_m)$, $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \dots, \varphi_m)^T$, $\boldsymbol{\lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m)$; $\mathbf{x}$ – вектор випадкових величин, що нормалізуються, $\mathbf{x} = (x_1, x_2, \dots, x_m)^T$.

Зазначимо, що $\mathbf{z}$ має багатомірний розподіл з нульовим вектором математичного сподівання $\mathbf{m}$ та коваріаційною матрицею $\boldsymbol{\Sigma}$

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho_{z_{12}} & \dots & \rho_{z_{1m}} \\ \rho_{z_{21}} & 1 & \dots & \rho_{z_{2m}} \\ \dots & \dots & 1 & \dots \\ \rho_{z_{m1}} & \rho_{z_{m2}} & \dots & 1 \end{pmatrix},$$

де $\rho_{z_{kl}} = \frac{M\{(z_k - m_{z_k})(z_l - m_{z_l})\}}{\sqrt{D_{z_k} D_{z_l}}}$; $M\{\cdot\}$ – операція математичного сподівання;

$$m_{z_k} = M\{z_k\}; D_{z_k} = M\{(z_k - m_{z_k})^2\}; k \in [1, m]; l \in [1, m].$$

Перетворенню (2.14) відповідає зворотне багатовимірне перетворення Джонсона

$$\mathbf{x} = \boldsymbol{\varphi} + \boldsymbol{\lambda} \mathbf{h}^{-1}[\boldsymbol{\eta}^{-1}(\mathbf{z} - \boldsymbol{\gamma})]. \quad (2.15)$$

На теперішній час прикладів застосування багатомірних взаємно-зворотних перетворень Джонсона (2.14) і (2.15) не так вже багато. Найбільш відомим – є двомірний розподіл Джонсона, якій був представлений ще у 1949 році [70], і отримав подальшого розвитку в [71, 72]. В [73-75] було

запропоновано використовувати багатомірні взаємно-зворотні перетворення Джонсона (2.14) і (2.15) для нормалізації та побудови математичних моделей нелінійних СДС. Пізніше багатомірні взаємно-зворотні перетворення Джонсона (2.14) і (2.15) було запропоновано використовувати для визначення викидів у багатовимірних негаусівських даних [45, 76, 77], для побудови нелінійних регресійних моделей [11] та для виявлення викидів у нелінійних регресійних моделях [78].

Оцінки параметрів багатовимірних перетворень також краще визначати за ММП

$$\hat{\theta} = \arg \max_{\theta} l(\mathbf{P}, \theta), \quad (2.16)$$

де $\mathbf{P}$ – це негаусівський випадковий вектор, $\mathbf{P} = \{Y, X_1, X_2, \dots, X_k\}^T$, який в загальному випадку перетворюють у гаусівський випадковий вектор $\mathbf{T}$, $\mathbf{T} = \{Z_Y, Z_1, Z_2, \dots, Z_k\}^T$, за допомогою багатовимірного нормалізуючого перетворення [11]

$$\mathbf{T} = \psi(\mathbf{P}), \quad (2.17)$$

яке має зворотне перетворення

$$\mathbf{P} = \psi^{-1}(\mathbf{T}). \quad (2.18)$$

Тут k – це кількість незалежних змінних (факторів).

У разі використання перетворення Джонсона сім'ї S_B компоненти вектору $\mathbf{T}$ визначаються як

$$Z_j = \gamma_j + \eta_j \ln \frac{X_j - \varphi_j}{\varphi_j + \lambda_j - X_j}, \quad \varphi_j < X_j < \varphi_j + \lambda_j, \quad (2.19)$$

де Z_j – це стандартна гаусівська змінна, $Z_j \sim N(0,1)$; γ_j , η_j , φ_j та λ_j – це параметри перетворення Джонсона сім'ї S_B , $\eta_j > 0$, $\lambda_j > 0$, $j=1,2,\dots,k$.

Компонента Z_Y визначається аналогічно (2.17) з тою різницею, що замість Z_j , X_j , γ_j , η_j , ϕ_j та λ_j слід підставити відповідно Z_Y , Y , γ_Y , η_Y , ϕ_Y та λ_Y .

Для чотиривимірного перетворення Джонсона сім'ї S_B логарифмічну функцію правдоподібності можна записати як

$$\begin{aligned} l(\mathbf{P}, \boldsymbol{\theta}) = & N \ln(\eta_Y \lambda_Y) - N \sum_{j=1}^k \ln(\eta_j \lambda_j) - \frac{N(k+1) \ln(2\pi)}{2} - \\ & - \frac{N}{2} \ln[\det(\mathbf{S}_T)] - \sum_{i=1}^N \ln[(Y_i - \phi_Y)(\phi_Y + \lambda_Y - Y_i)] - \\ & - \sum_{j=1}^k \sum_{i=1}^N \ln[(X_{ji} - \phi_j)(\phi_j + \lambda_j - X_{ji})] - \frac{1}{2} \sum_{i=1}^N \mathbf{T}_i^T \mathbf{S}_T^{-1} \mathbf{T}_i, \end{aligned} \quad (2.20)$$

де $k=4$; $\boldsymbol{\theta} = \{\gamma_Y, \eta_Y, \phi_Y, \lambda_Y, \gamma_1, \gamma_2, \dots, \gamma_k, \eta_1, \eta_2, \dots, \eta_k, \phi_1, \phi_2, \dots, \phi_k, \lambda_1, \lambda_2, \dots, \lambda_k\}$; $\mathbf{S}_T$ – це вибіркова коваріаційна матриця для компонентів вектору $\mathbf{T}$

$$\mathbf{S}_T = \frac{1}{N} \sum_{i=1}^N \mathbf{T}_i \mathbf{T}_i^T. \quad (2.21)$$

В (2.21) враховане те, що математичні сподівання компонентів вектору $\mathbf{T}$ для перетворення Джонсона дорівнюють нулю.

Багатовимірне перетворення Бокса-Кокса.

У разі використання перетворення Бокса-Кокса компоненти вектору $\mathbf{T}$ багатовимірного нормалізуючого перетворення (2.17) визначаються як

$$Z_j = x(\lambda_j) = \begin{cases} (X_j^{\lambda_j} - 1)/\lambda_j, & \text{if } \lambda_j \neq 0; \\ \ln(X_j), & \text{if } \lambda_j = 0. \end{cases} \quad (2.22)$$

де Z_j – це гаусівська змінна; λ_j – це параметри перетворення Бокса-Кокса, $j=1,2,\dots,k$. Компонента Z_Y визначається аналогічно (2.22) з тою різницею, що замість Z_j , X_j та λ_j слід підставити відповідно Z_Y , Y та λ_Y .

Для чотиривимірного перетворення Бокса-Кокса логарифмічну функцію правдоподібності можна записати як [63]

$$l(\mathbf{P}, \boldsymbol{\theta}) = (\lambda_Y - 1) \sum_{i=1}^n \ln(Y_i) + \sum_{j=1}^k (\lambda_j - 1) \sum_{i=1}^n \ln(X_{ji}) - \frac{N}{2} \ln[\det(\mathbf{S}_T)], \quad (2.23)$$

де $k=4$; $\boldsymbol{\theta} = \{\lambda_Y, \lambda_1, \lambda_2, \dots, \lambda_k\}$; $\mathbf{S}_T$ – це вибіркова коваріаційна матриця для компонентів вектору $\mathbf{T}$

$$\mathbf{S}_T = \frac{1}{N} \sum_{i=1}^N (\mathbf{T}_i - \bar{\mathbf{T}})(\mathbf{T}_i - \bar{\mathbf{T}})^T. \quad (2.24)$$

В (2.24) враховане те, що математичні сподівання компонентів вектору $\mathbf{T}$ для перетворення Бокса-Кокса в загальному випадку не дорівнюють нулю.

2.3 Вибір перетворення для нормалізації чотиривимірних даних з метрик РНР-застосунків

Здійснимо вибір перетворення для нормалізації чотиривимірних даних (фактичний розмір РНР-застосунків з відкритим кодом в тисячах рядків коду Y , загальна кількість класів X_1 , середня кількість методів на клас X_2 , та сума середньої кількості класів, на які впливає даний клас, і середньої кількості класів, з яких даний клас отримує ефекти, X_3), що наведені у табл.1.2. Але спочатку перевіримо чотиривимірні дані на наявність багатовимірних викидів. Перед цим необхідно перевірити нормальність цих багатовимірних даних, оскільки добре відомі статистичні методи (наприклад, виявлення багатовимірних викидів на основі квадрату відстані Махаланобісу) використовуються для виявлення викидів у багатовимірному наборі даних за умови, що дані є гаусівськими.

В роботі використано тест на нормальність багатовимірних даних,

заснований на багатовимірних асиметрії β_1 та ексцесу β_2 [43]. Згідно з цим тестом розподіл чотиривимірних даних не є гаусівським, оскільки статистика тесту на багатовимірну асиметрію цих даних, яка дорівнює 61,97, перевищує значення квантилю розподілу χ^2 , що становить 40,00 для 20 ступенів свободи та рівня значущості 0,005. Аналогічно, статистика тесту для багатовимірного ексцесу, що дорівнює 31,35, перевищує значення квантилю розподілу Гауса, який становить 30,46 для середнього і вибіркової дисперсії, що дорівнюють відповідно 24 і 4,36, та рівня значущості 0,005. Ці результати вказують на необхідність подальшого застосування методу для визначення багатовимірних викидів у багатовимірних негаусівських даних на основі багатовимірних нормалізуючих перетворень. Згідно з [77], дані рядка 2 є багатовимірним викидом тому, що для даних рядка 2 квадрат відстані Махаланобісу, який дорівнює 19,71 за (2.22), більше за значення квантилю розподілу χ^2 , що становить 14,86 для 4 ступенів свободи та рівня значущості 0,005.

Зазначимо, квадрат відстані Махаланобісу для нормалізованих даних визначався як

$$d_i^2 = (\mathbf{T}_i - \bar{\mathbf{T}})^T \mathbf{S}_T^{-1} (\mathbf{T}_i - \bar{\mathbf{T}}), \quad (2.25)$$

де $\bar{\mathbf{T}}$ – вектор вибірових середніх; $\mathbf{S}_T$ – вибіркова коваріаційна матриця для компонентів вектору $\mathbf{T}$ знаходиться за (2.24). Зауважимо, у разі використання перетворення Джонсона для нормалізації даних формулу (2.25) можна переписати як

$$d_i^2 = \mathbf{T}_i^T \mathbf{S}_T^{-1} \mathbf{T}_i, \quad (2.26)$$

а вибіркову коваріаційну матрицю $\mathbf{S}_T$ в (2.26) визначати за (2.21).

Далі 43 рядки даних з табл.1.2 (за виключенням даних рядка 2) знову були проаналізовані на наявність багатовимірних викидів. Згідно з [77], дані 43 рядків даних з табл.1.2 (за виключенням даних рядка 2) не містять багатовимірних

викидів тому, що для всіх 43 рядків даних значення квадрату відстані Махаланобісу менше за значення квантилю розподілу χ^2 для 4 ступенів свободи та рівня значущості 0,005. Тому для обґрунтування вибору перетворення для нормалізації чотиривимірних даних використовуємо далі 43 рядки даних з табл.1.2 (за виключенням даних рядка 2). Для нормалізації зазначених чотиривимірних даних використовуємо три одновимірні перетворення (десятькового логарифму, Бокса-Кокса та Джонсона) і два чотиривимірних (Бокса-Кокса та Джонсона). Для визначення якості нормалізації за зазначеними перетвореннями використовуємо тест перевірки на нормальність багатовимірних даних, заснований на вимірах багатовимірних асиметрії β_1 та ексцесу β_2 [43]. Результати нормалізації 43 рядків нормалізованих даних з табл.1.2 (за виключенням даних рядка 2) за допомогою одновимірного перетворення у вигляді десятикового логарифму (2.3) наведені в табл.2.1.

Далі, було застосовано тест для перевірки нормальності чотиривимірних даних з табл.2.1, заснований на багатовимірних асиметрії β_1 та ексцесу β_2 [43]. Згідно з цим тестом розподіл чотиривимірних даних не є гаусівським, оскільки статистика тесту на багатовимірну асиметрію цих даних, яка дорівнює 71,30, перевищує значення квантилю розподілу χ^2 , що становить 40,00 для 20 ступенів свободи та рівня значущості 0,005. Аналогічно, статистика тесту для багатовимірного ексцесу, що дорівнює 30,13, перевищує значення квантилю розподілу Гауса, який становить 29,44 для середнього і вибіркової дисперсії, що дорівнюють відповідно 24 і 4,465, та рівня значущості 0,005.

Ці результати вказують нам на незадовільну нормалізацію зазначених вище багатовимірних даних за допомогою одновимірного перетворення у вигляді десятикового логарифму (2.3).

Результати нормалізації 43 рядків нормалізованих даних з табл.1.2 (за

виключенням даних рядка 2) за допомогою одновимірного перетворення Бокса-Кокса (2.5) наведені в табл.2.2.

Таблиця 2.1 – Результати нормалізації за допомогою одновимірного перетворення у вигляді десяткового логарифму

№	Z_Y	Z_1	Z_2	Z_3	№	Z_Y	Z_1	Z_2	Z_3
1	2,243	3,317	0,682	0,808	23	1,002	2,497	0,348	0,628
2	-	-	-	-	24	1,190	2,447	0,531	0,579
3	1,917	2,614	0,895	0,803	25	1,193	2,061	0,901	0,707
4	1,080	2,121	0,764	0,704	26	0,366	1,176	1,003	0,693
5	0,728	0,699	1,513	0,380	27	0,851	1,398	0,973	0,537
6	-0,221	1,398	0,033	0,246	28	0,156	1,342	0,533	0,413
7	0,193	1,398	0,350	0,606	29	1,569	2,444	0,770	0,642
8	1,522	2,334	0,885	1,004	30	1,516	2,371	0,913	0,808
9	1,557	2,100	1,035	0,413	31	1,087	1,763	1,126	0,962
10	2,001	2,651	0,981	0,800	32	1,775	2,754	0,776	0,798
11	0,649	1,863	0,649	0,718	33	1,396	2,560	0,748	1,060
12	0,475	1,255	0,983	0,845	34	0,373	1,447	0,729	0,796
13	0,607	1,491	0,801	0,729	35	-0,419	0,903	0,495	0,439
14	0,825	2,097	0,451	0,582	36	0,634	1,716	0,814	0,692
15	0,096	0,301	1,255	0,301	37	0,533	1,716	0,522	0,884
16	0,776	1,869	0,791	0,686	38	1,198	2,100	0,953	0,622
17	1,591	2,430	0,886	0,806	39	-0,272	0,477	0,970	0,523
18	0,514	1,568	0,708	0,694	40	1,501	1,881	1,149	0,874
19	1,551	2,525	0,765	1,045	41	1,144	2,400	0,533	0,529
20	0,950	2,068	0,667	0,572	42	0,523	1,204	0,906	0,945
21	1,147	2,320	0,657	0,940	43	1,386	2,900	-0,312	0,351
22	0,465	1,279	0,912	0,759	44	1,633	2,450	0,458	0,593

Оцінки параметрів одновимірних перетворень Бокса-Кокса (2.5) для

змінних Y , X_1 , X_2 та X_3 визначалися за ММП з логарифмічною функцією правдоподібності (2.23) і дорівнюють відповідно 0,054251, 0,098545, 0,243665 та 0,190194.

Таблиця 2.2 – Результати нормалізації за допомогою одновимірного перетворення Бокса-Кокса

№	Z_Y	Z_1	Z_2	Z_3	№	Z_Y	Z_1	Z_2	Z_3
1	5,960	11,392	1,913	2,232	23	2,458	7,735	0,884	1,663
2	-	-	-	-	24	2,953	7,534	1,426	1,517
3	4,987	8,215	2,678	2,217	25	2,962	6,049	2,701	1,909
4	2,662	6,271	2,197	1,897	26	0,862	3,104	3,100	1,865
5	1,755	1,744	5,488	0,953	27	2,068	3,788	2,981	1,393
6	-0,502	3,788	0,078	0,597	28	0,362	3,613	1,429	1,044
7	0,451	3,788	0,891	1,599	29	3,992	7,521	2,218	1,707
8	3,860	7,087	2,638	2,905	30	3,844	7,231	2,747	2,233
9	3,957	6,196	3,233	1,042	31	2,681	4,993	3,617	2,754
10	5,235	8,372	3,012	2,207	32	4,577	8,810	2,240	2,200
11	1,557	5,340	1,801	1,941	33	3,511	7,992	2,139	3,105
12	1,128	3,344	3,019	2,355	34	0,880	3,944	2,074	2,193
13	1,452	4,087	2,328	1,977	35	-0,940	2,308	1,313	1,115
14	2,002	6,183	1,181	1,525	36	1,520	4,831	2,376	1,862
15	0,222	0,717	4,196	0,741	37	1,269	4,831	1,397	2,485
16	1,875	5,361	2,291	1,842	38	2,976	6,196	2,900	1,647
17	4,053	7,464	2,641	2,225	39	-0,615	1,160	2,968	1,353
18	1,223	4,337	2,002	1,868	40	3,801	5,402	3,717	2,450
19	3,940	7,849	2,199	3,052	41	2,832	7,344	1,430	1,370
20	2,322	6,077	1,861	1,498	42	1,244	3,188	2,721	2,696
21	2,839	7,032	1,828	2,678	43	3,483	9,447	-0,659	0,873
22	1,103	3,416	2,740	2,072	44	4,171	7,546	1,203	1,558

Для перевірки нормальності чотиривимірних даних з табл.2.2 було

застосовано тест, заснований на багатовимірних асиметрії β_1 та ексцесу β_2 [43]. Згідно з цим тестом розподіл чотиривимірних даних не є гаусівським, оскільки статистика тесту на багатовимірну асиметрію цих даних, яка дорівнює 42,91, перевищує значення квантилю розподілу χ^2 , що становить 40,00 для 20 ступенів свободи та рівня значущості 0,005. Статистика тесту для багатовимірного ексцесу, що дорівнює 27,40, менше значення квантилю розподілу Гауса, який становить 29,44 для середнього і вибіркової дисперсії, що дорівнюють відповідно 24 і 4,465, та рівня значущості 0,005.

Ці результати вказують нам на незадовільну нормалізацію зазначених вище багатовимірних даних за допомогою одновимірного перетворення Бокса-Кокса (2.5). Хоча слід зазначити, що у порівнянні з десятковим логарифмом, одновимірне перетворення Бокса-Кокса краще нормалізує дані з табл.1.2.

Результати нормалізації 43 рядків нормалізованих даних з табл.1.2 (за виключенням даних рядка 2) за допомогою чотиривимірного перетворення Бокса-Кокса наведені в табл.2.3.

Оцінки параметрів чотиривимірного перетворення Бокса-Кокса (2.22) для змінних Y , X_1 , X_2 та X_3 визначалися за методом максимальної правдоподібності і дорівнюють відповідно -0,0040926, 0,024790, 0,312664 та 0,111095.

Для перевірки нормальності чотиривимірних даних з табл.2.3 було застосовано тест заснований на багатовимірних асиметрії β_1 та ексцесу β_2 [43]. Згідно з цим тестом розподіл чотиривимірних даних не є гаусівським, оскільки статистика тесту на багатовимірну асиметрію цих даних, яка дорівнює 46,04, перевищує значення квантилю розподілу χ^2 , що становить 40,00 для 20 ступенів свободи та рівня значущості 0,005. Статистика тесту для багатовимірного ексцесу, що дорівнює 28,27, менше значення квантилю розподілу Гауса, який становить 29,44 для середнього і вибіркової дисперсії, що дорівнюють відповідно

24 і 4,465, та рівня значущості 0,005.

Таблиця 2.3 – Результати нормалізації за допомогою чотиривимірного перетворення Бокса-Кокса

№	Z_Y	Z_1	Z_2	Z_3	№	Z_Y	Z_1	Z_2	Z_3
1	5,110	8,409	2,028	2,066	23	2,296	6,179	0,909	1,567
2	-	-	-	-	24	2,724	6,047	1,491	1,437
3	4,374	6,491	2,895	2,053	25	2,732	5,035	2,921	1,785
4	2,474	5,191	2,346	1,775	26	0,841	2,801	3,386	1,746
5	1,671	1,642	6,309	0,919	27	1,952	3,351	3,246	1,324
6	-0,510	3,351	0,078	0,583	28	0,358	3,213	1,495	1,004
7	0,445	3,351	0,917	1,510	29	3,587	6,039	2,369	1,606
8	3,480	5,750	2,849	2,637	30	3,466	5,846	2,975	2,067
9	3,558	5,138	3,542	1,003	31	2,490	4,272	3,998	2,510
10	4,564	6,591	3,282	2,045	32	4,054	6,868	2,394	2,039
11	1,490	4,527	1,903	1,814	33	3,192	6,347	2,281	2,803
12	1,092	2,996	3,291	2,172	34	0,858	3,474	2,207	2,032
13	1,394	3,584	2,495	1,845	35	-0,967	2,134	1,369	1,071
14	1,893	5,129	1,226	1,444	36	1,456	4,151	2,549	1,744
15	0,221	0,699	4,698	0,721	37	1,224	4,151	1,459	2,284
16	1,780	4,542	2,453	1,726	38	2,744	5,138	3,152	1,553
17	3,636	6,001	2,852	2,060	39	-0,626	1,114	3,232	1,288
18	1,182	3,777	2,127	1,749	40	3,431	4,572	4,118	2,254
19	3,545	6,254	2,348	2,759	41	2,621	5,922	1,495	1,304
20	2,177	5,055	1,970	1,419	42	1,201	2,870	2,944	2,462
21	2,626	5,712	1,933	2,447	43	3,170	7,261	-0,644	0,845
22	1,069	3,055	2,967	1,928	44	3,731	6,056	1,250	1,474

Ці результати вказують нам на незадовільну нормалізацію зазначених вище чотиривимірних даних за допомогою чотиривимірного перетворення Бокса-Кокса (2.22). Хоча слід зазначити, що у порівнянні з десятковим логарифмом, чотиривимірне перетворення Бокса-Кокса краще нормалізує дані з

табл.1.2. Результати нормалізації 43 рядків нормалізованих даних з табл.1.2 (за виключенням даних рядка 2) за допомогою одновимірного перетворення Джонсона сім'ї S_B (2.9) наведені в табл.2.4.

Таблиця 2.4 – Результати нормалізації за допомогою одновимірного перетворення Джонсона сім'ї S_B

№	Z_Y	Z_1	Z_2	Z_3	№	Z_Y	Z_1	Z_2	Z_3
1	2,614	2,629	-0,352	0,501	23	0,060	0,851	-1,366	-0,286
2	-	-	-	-	24	0,323	0,775	-0,841	-0,491
3	1,500	1,033	0,418	0,480	25	0,328	0,209	0,440	0,051
4	0,169	0,295	-0,065	0,036	26	-0,835	-1,045	0,835	-0,009
5	-0,319	-1,760	2,981	-1,433	27	-0,148	-0,729	0,718	-0,674
6	-1,876	-0,729	-2,058	-2,435	28	-1,159	-0,808	-0,837	-1,254
7	-1,099	-0,729	-1,358	-0,375	29	0,882	0,770	-0,044	-0,225
8	0,809	0,606	0,379	1,706	30	0,800	0,661	0,487	0,503
9	0,863	0,266	0,964	-1,257	31	0,179	-0,213	1,332	1,366
10	1,691	1,093	0,748	0,466	32	1,226	1,263	-0,022	0,456
11	-0,430	-0,072	-0,465	0,096	33	0,619	0,948	-0,123	2,330
12	-0,676	-0,932	0,756	0,682	34	-0,825	-0,659	-0,190	0,445
13	-0,489	-0,596	0,067	0,144	35	-2,527	-1,446	-0,952	-1,123
14	-0,184	0,261	-1,081	-0,480	36	-0,450	-0,280	0,116	-0,013
15	-1,257	-2,470	1,868	-1,942	37	-0,593	-0,280	-0,870	0,886
16	-0,253	-0,063	0,030	-0,040	38	0,335	0,266	0,638	-0,308
17	0,917	0,749	0,381	0,491	39	-2,006	-2,132	0,706	-0,734
18	-0,620	-0,488	-0,262	-0,004	40	0,777	-0,047	1,426	0,830
19	0,854	0,894	-0,063	2,133	41	0,259	0,704	-0,837	-0,709
20	-0,012	0,220	-0,404	-0,520	42	-0,608	-1,005	0,461	1,253
21	0,262	0,585	-0,438	1,220	43	0,604	1,522	-2,539	-1,607
22	-0,690	-0,898	0,480	0,276	44	0,984	0,780	-1,060	-0,433

Вибір сім'ї S_B був обумовлений тим, що значення оцінок асиметрії у квадратах для змінних Y , X_1 , X_2 та X_3 дорівнюють 9,637, 18,502, 8,179 та 0,678, а

значення оцінок ексцесу для змінних Y , X_1 , X_2 та X_3 дорівнюють 14,836, 25,544, 15,045 та 3,251 відповідно. А як можна побачити з рис.2.1, всі ці комбінації значень оцінок асиметрії у квадраті та ексцесу для кожної змінної знаходяться в області S_B .

Оцінки параметрів одновимірних перетворень Джонсона (2.9) сім'ї S_B для змінних Y , X_1 , X_2 та X_3 для 43 рядків даних з табл.1.2 (за виключенням даних рядка 2) визначалися за методом максимальної правдоподібності (2.12) із логарифмічною функцією правдоподібності (2.13) і наведені у табл.2.5.

Таблиця 2.5 – Оцінки параметрів одновимірних перетворень Джонсона сім'ї S_B для змінних Y , X_1 , X_2 та X_3 для 43 рядків даних

Індекс j	$\hat{\gamma}_j$	$\hat{\eta}_j$	$\hat{\phi}_j$	$\hat{\lambda}_j$
Y	1,75572	0,55912	0,281	212,274
1	2,11464	0,59365	0,696	2946,829
2	14,63054	2,00122	-1,693	11606,864
3	0,82028	0,93873	1,394	12,099

Для перевірки нормальності чотиривимірних даних з табл.2.4 було застосовано тест, заснований на багатовимірних асиметрії β_1 та ексцесу β_2 [43]. Згідно з цим тестом розподіл чотиривимірних даних наближено є гаусівським, оскільки статистика тесту на багатовимірну асиметрію цих даних, яка дорівнює 39,21, не перевищує значення квантилю розподілу χ^2 , що становить 40,00 для 20 ступенів свободи та рівня значущості 0,005, а статистика тесту для багатовимірного ексцесу, що дорівнює 29,03, також менше значення квантилю розподілу Гауса, який становить 29,44 для середнього і вибіркової дисперсії, що

дорівнюють відповідно 24 і 4,465, та рівня значущості 0,005. Ці результати вказують нам на задовільну нормалізацію зазначених вище багатовимірних даних за допомогою одновимірного перетворення Джонсона (2.9) сім'ї S_B .

Результати нормалізації 43 рядків нормалізованих даних з табл.1.2 (за виключенням даних рядка 2) за допомогою чотиривимірного перетворення Джонсона (2.14) сім'ї S_B наведені в табл.2.6.

Таблиця 2.6 – Результати нормалізації за допомогою чотиривимірного перетворення Джонсона сім'ї S_B

№	Z_Y	Z_1	Z_2	Z_3	№	Z_Y	Z_1	Z_2	Z_3
1	2	3	4	5	6	7	8	9	10
1	2,614	2,619	-0,364	0,499	23	0,060	0,862	-1,358	-0,269
2	-	-	-	-	24	0,323	0,784	-0,848	-0,472
3	1,500	1,048	0,410	0,478	25	0,328	0,206	0,433	0,061
4	0,169	0,294	-0,078	0,046	26	-0,835	-1,066	0,834	0,002
5	-0,319	-1,759	3,052	-1,426	27	-0,148	-0,748	0,715	-0,653
6	-1,876	-0,748	-2,012	-2,571	28	-1,159	-0,828	-0,844	-1,239
7	-1,099	-0,748	-1,351	-0,357	29	0,882	0,780	-0,057	-0,210
8	0,809	0,612	0,370	1,679	30	0,800	0,668	0,480	0,501
9	0,863	0,264	0,966	-1,242	31	0,179	-0,225	1,344	1,344
10	1,691	1,109	0,746	0,465	32	1,226	1,282	-0,034	0,456
11	-0,430	-0,081	-0,477	0,105	33	0,619	0,961	-0,136	2,309
12	-0,676	-0,952	0,753	0,676	34	-0,825	-0,678	-0,203	0,445
13	-0,489	-0,615	0,055	0,152	35	-2,527	-1,460	-0,957	-1,105
14	-0,184	0,259	-1,083	-0,461	36	-0,450	-0,293	0,104	-0,001
15	-1,257	-2,366	1,897	-1,978	37	-0,593	-0,293	-0,876	0,874
16	-0,253	-0,072	0,018	-0,028	38	0,335	0,264	0,634	-0,291

Продовження табл.2.6

1	2	3	4	5	6	7	8	9	10
17	0,917	0,758	0,373	0,490	39	-2,006	-2,091	0,702	-0,713
18	-0,620	-0,505	-0,275	0,007	40	0,777	-0,055	1,441	0,819
19	0,854	0,906	-0,076	2,107	41	0,259	0,711	-0,844	-0,687
20	-0,012	0,217	-0,417	-0,500	42	-0,608	-1,026	0,453	1,233
21	0,262	0,591	-0,450	1,201	43	0,604	1,543	-2,452	-1,610
22	-0,690	-0,919	0,473	0,280	44	0,984	0,789	-1,062	-0,414

Оцінки параметрів чотиривимірного перетворення Джонсона (2.14) сім'ї S_B для змінних Y , X_1 , X_2 та X_3 визначалися за методом максимальної правдоподібності (2.16) з логарифмічною функцією правдоподібності (2.20) і наведені у табл.2.7.

Таблиця 2.7 – Оцінки параметрів чотиривимірного перетворення Джонсона сім'ї S_B для змінних Y , X_1 , X_2 та X_3 для 43 рядків даних

Індекс j	$\hat{\gamma}_j$	$\hat{\eta}_j$	$\hat{\phi}_j$	$\hat{\lambda}_j$
Y	1,75572	0,55912	0,281	212,274
1	2,19844	0,61039	0,239	3116,897
2	15,30976	2,10956	-2,072	11606,864
3	0,79841	0,89377	1,494	11,822

Для перевірки нормальності чотиривимірних даних з табл.2.6 було застосовано тест, заснований на багатовимірних асиметрії β_1 та ексцесу β_2 [43]. Згідно з цим тестом розподіл чотиривимірних даних наближено є гаусівським, оскільки статистика тесту на багатовимірну асиметрію цих даних, яка дорівнює 38,79, менше значення квантиля розподілу χ^2 , що становить 40,00 для 20

ступенів свободи та рівня значущості 0,005, а статистика тесту для багатовимірного ексцесу, що дорівнює 28,73, також менше значення квантиля розподілу Гауса, який становить 29,44 для середнього і вибіркової дисперсії, що дорівнюють відповідно 24 і 4,465, та рівня значущості 0,005. Ці результати вказують нам на добру нормалізацію зазначених вище багатовимірних даних за допомогою чотиривимірного перетворення Джонсона (2.14) сім'ї S_B .

У якості нормалізуючого перетворення було обрано чотиривимірне нормалізуюче перетворення Джонсона сім'ї S_B , за допомогою якого краще наведені у табл.1.2. І як буде показано у подальшому застосування зазначеного чотиривимірного перетворення у порівнянні з одновимірними перетвореннями при побудові нелінійних регресійних моделей приводить до кращих показників якості відповідних моделей.

2.4 Висновки за розділом 2

У даному розділі розглянуто відомі нормалізуючі перетворення та методи оцінювання їх параметрів. Розглянуто одновимірні та багатомірні нормалізуючі перетворення Джонсона та Бокса-Кокса та підходи для їх підбору. Здійснено вибір перетворення для нормалізації чотиривимірних даних з метрик РНР-застосунків.

1. Розглянуто відомі взаємо-зворотні нормалізуючі перетворення, які дозволяють перетворити випадкові величини з негаусівським розподілом у такі, що мають гаусівський розподіл або наближений до нього. Показано, що існуючі нормалізуючі перетворення можна поділити на одновимірні та багатовимірні перетворення, перетворення для окремих розподілів, перетворення Джонсона і

перетворення Бокса-Кокса.

2. Розглянуто багатовимірне перетворення Джонсона та підхід для його підбору з урахуванням кореляцій між випадковими величинами. Для оцінювання параметрів чотиривимірного перетворення Джонсона сім'ї S_B пропонується використовувати метод максимальної правдоподібності.

3. Здійснено вибір перетворення для нормалізації чотиривимірних даних з метрик РНР-застосунків. У якості нормалізуючого перетворення було обрано чотиривимірне нормалізуюче перетворення Джонсона сім'ї S_B , за допомогою якого краще здійснюється нормалізація чотиривимірних даних з метрик РНР-застосунків у порівнянні з одновимірними перетвореннями у вигляді десяткового логарифму, Джонсона сім'ї S_B і Бокса-Кокса та чотиривимірним перетворенням Бокса-Кокса.

Результати досліджень поточного розділу автором опубліковано в [10, 46-52].

РОЗДІЛ 3 ПОБУДОВА МАТЕМАТИЧНИХ НЕЛІНІЙНИХ МОДЕЛЕЙ ДЛЯ ОЦІНЮВАННЯ РОЗМІРУ ПРОГРАМНИХ ЗАСТОСУНКІВ З ВІДКРИТИМ КОДОМ НА РНР НА ОСНОВІ НОРМАЛІЗУЮЧИХ ПЕРЕТВОРЕНЬ

У третьому розділі 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.

Для побудови трьох-факторних нелінійних регресійних моделей використовуємо метод, заснований на багатовимірних нормалізуючих перетвореннях та інтервалах передбачення [36]. Суть цього методу [36] є такою. На першому етапі початкові дані перевіряються на наявність викидів і, якщо останні знайдено, то вони відкидаються. Для цього використовується критерій на основі квадрату відстані Махаланобіса для нормалізованих даних із 0,005 рівнем значущості. На другому етапі будується нелінійна регресійна модель із використанням відповідного методу на основі нормалізуючих перетворень [11]. На третьому етапі визначаються границі інтервалу передбачення нелінійної регресії для рівня значущості, що дорівнює 0,05, за відповідним методом [11]. На останньому четвертому етапі перевіряють, чи є серед даних, за якими будувалася нелінійна регресійна модель такі, що виходять за границі інтервалу передбачення. Та, якщо відповідні дані знайдено, то вони відкидаються, і надалі

повторюємо знову всі етапи, починаючи з першого, для нових даних. Якщо таких викидів не було, то повторення етапів завершується, відповідна нелінійна регресійна модель побудована. Як і в [11, 36], у якості багатовимірного нормалізуючого перетворення застосовуємо чотиривимірне перетворення Джонсона для сімейства S_B .

3.1 Побудова рівняння та довірчих інтервалів нелінійної регресії для оцінювання розміру програмних застосунків з відкритим кодом на РНР на основі чотиривимірного перетворення Джонсона сімейства S_B

Для побудови рівняння та довірчих інтервалів нелінійної регресії для оцінювання розміру програмних застосунків з відкритим кодом на РНР були використані дані, що наведені у табл. 1.2.

На першому етапі початкові дані були перевірені на наявність багатовимірних викидів. Це було зроблено у підрозділі 2.3. Згідно з [36], дані рядка 2 з табл. 1.2 є багатовимірним викидом тому, що для даних рядка 2 квадрат відстані Махаланобісу, який дорівнює 19,71, більше за значення квантилю розподілу χ^2 , що становить 14,86 для 4 ступенів свободи та рівня значущості 0,005. Дані рядка 2 відкидаються. На цьому перша ітерація завершується і далі знов переходимо до етапу 1 але вже використовуємо 43 рядки даних з табл.1.2 (за виключенням даних рядка 2).

На другій ітерації згідно з [36], як було показано у підрозділі 2.3, дані 43 рядків даних з табл.1.2 (за виключенням даних рядка 2) не містять багатовимірних викидів тому, що для всіх 43 рядків даних значення квадрату відстані Махаланобісу менше за значення квантилю розподілу χ^2 для 4 ступенів

свободи та рівня значущості 0,005. Тому далі переходимо до побудови нелінійного регресійного рівняння.

Для цього спочатку для нормалізованих за багатовимірним перетворенням Джонсона сімейства S_B даних з табл.2.6 будуємо лінійне регресійне рівняння

$$\hat{Z}_Y = \hat{b}_0 + \hat{b}_1 Z_1 + \hat{b}_2 Z_2 + \hat{b}_3 Z_3, \quad (3.1)$$

де параметри рівняння (3.1) оцінювалися методом найменших квадратів і їх оцінки є такими: $\hat{b}_0 = 0$, $\hat{b}_1 = 1,072221$, $\hat{b}_2 = 0,503937$, $\hat{b}_3 = -0,045683$. Сума квадратів відхилень для рівняння (3.1) склала 1,7085.

Потім за (3.1) на основі багатовимірного перетворення Джонсона сімейства S_B будуємо нелінійне регресійне рівняння

$$\hat{Y} = \hat{\phi}_Y + \hat{\lambda}_Y \left[1 + e^{-(\hat{Z}_Y - \hat{\gamma}_Y)/\hat{\eta}_Y} \right]^{-1}, \quad (3.2)$$

де $Z_j = \gamma_j + \eta_j \ln \frac{X_j - \phi_j}{\phi_j + \lambda_j - X_j}$, $\phi_j < X_j < \phi_j + \lambda_j$, $j = 1, 2, 3$. Сума квадратів

відхилень для рівняння (3.2) з параметрами багатовимірного перетворення Джонсона сімейства S_B , що наведені в табл.2.7, склала 1713,038.

Далі на третьому етапі були визначені границі інтервалу передбачення нелінійної регресії для рівня значущості, що дорівнює 0,05, за відповідним методом [5]. Згідно з [11], границі інтервалу передбачення нелінійної регресії визначаються як

$$\psi_Y^{-1} \left(\hat{Z}_Y \pm t_{\alpha/2, v} S_{Z_Y} \left\{ 1 + \frac{1}{N} + (\mathbf{z}_X^+)^T \left[(\mathbf{Z}_X^+)^T \mathbf{Z}_X^+ \right]^{-1} (\mathbf{z}_X^+) \right\}^{1/2} \right), \quad (3.3)$$

де ψ_Y^{-1} – зворотне до нормалізуючого перетворення ψ_Y для залежної змінної Y ; $\hat{Z}_Y$ – результат передбачення лінійного регресійного рівняння для значень компонент вектору $\mathbf{z}_X = \{Z_1, Z_2, \dots, Z_k\}$, що отриманий за

нормалізованими даними; $\mathbf{Z}_X^+$ – матриця центрованих регресорів, яка містить значення $Z_{1i} - \bar{Z}_1, Z_{2i} - \bar{Z}_2, \dots, Z_{ki} - \bar{Z}_k$; $t_{\alpha/2, \nu}$ – квантиль t -розподілу Стюдента в залежності від рівня значимості $\alpha/2$ та кількості ступенів свободи ν ;

$$S_{Z_Y}^2 = \frac{1}{\nu} \sum_{i=1}^N (Z_{Y_i} - \hat{Z}_{Y_i})^2, \nu = N - k - 1; (\mathbf{Z}_X^+)^T \mathbf{Z}_X^+ - k \times k \text{ матриця}$$

$$(\mathbf{Z}_X^+)^T \mathbf{Z}_X^+ = \begin{pmatrix} S_{Z_1 Z_1} & S_{Z_1 Z_2} & \dots & S_{Z_1 Z_k} \\ S_{Z_1 Z_2} & S_{Z_2 Z_2} & \dots & S_{Z_2 Z_k} \\ \dots & \dots & \dots & \dots \\ S_{Z_1 Z_k} & S_{Z_2 Z_k} & \dots & S_{Z_k Z_k} \end{pmatrix},$$

$$\text{де } S_{Z_q Z_r} = \sum_{i=1}^N [Z_{q_i} - \bar{Z}_q][Z_{r_i} - \bar{Z}_r], q, r = 1, 2, \dots, k.$$

В (3.3) знаки мінус та плюс відповідають рівнянням нижньої та верхньої границь інтервалів передбачення нелінійних регресій, що побудовані на основі нормалізуючих перетворень (як одновимірних, так і багатовимірних).

На другій ітерації матриця $(\mathbf{Z}_X^+)^T \mathbf{Z}_X^+$ для нормалізованих за багатовимірним перетворенням Джонсона сімейства S_B даних з табл.2.6 є такою

$$(\mathbf{Z}_X^+)^T \mathbf{Z}_X^+ = \begin{pmatrix} 43,00 & -15,99 & 15,63 \\ -15,99 & 43,00 & 7,44 \\ 15,63 & 7,44 & 43,00 \end{pmatrix}.$$

Для нелінійного регресійного рівняння (3.2) з оцінками параметрів, що наведені в табл.3.1, виявилось: значення Y для двох РНР-застосунків (35 і 44) виходять за визначені межі інтервалу передбачення. В табл.3.1 ліва границя інтервалу передбачення, що отриманий на другій ітерації за [11], позначена як LB_2 , а права – як UB_2 . Всього було 4 таких ітерацій, після яких залишилося 40 РНР-застосунків (1, 3-6, 8-34, 36-43). В табл.3.1 ліва границя інтервалу передбачення, що отриманий на четвертій ітерації за [11], позначена як LB_4 , а

права – як UB_4 . На четвертій ітерації матриця $(Z_X^+)^T Z_X^+$ для нормалізованих за багатовимірним перетворенням Джонсона сімейства S_B даних є такою

$$(Z_X^+)^T Z_X^+ = \begin{pmatrix} 43,00 & -15,99 & 15,63 \\ -15,99 & 43,00 & 7,44 \\ 15,63 & 7,44 & 43,00 \end{pmatrix}.$$

Таблиця 3.1 – Границі інтервалу передбачення нелінійної регресії, що отримані на другій та четвертій ітераціях за допомогою чотиривимірного перетворення Джонсона сім'ї S_B

№	Y	LB ₂	UB ₂	LB ₄	UB ₄
1	2	3	4	5	6
1	174,927	141,219	193,978	137,430	221,703
2	112,048	-	-	-	-
3	82,551	36,445	105,427	40,098	89,814
4	12,022	7,023	28,295	8,508	21,425
5	5,347	2,569	12,792	3,733	10,881
6	0,601	0,470	1,301	0,458	1,106
7	1,561	0,580	1,754	-	-
8	33,276	15,512	58,221	18,403	45,611
9	36,028	16,803	63,803	19,485	49,153
10	100,245	50,521	127,947	55,464	117,672
11	4,458	2,607	10,705	3,298	8,440
12	2,988	1,533	6,159	1,995	5,132
13	4,047	1,629	6,446	2,069	5,257
14	6,688	2,963	12,392	3,723	9,633
15	1,247	0,555	1,758	0,675	1,718
16	5,966	4,003	16,475	5,000	12,711
17	38,996	22,098	74,708	25,110	59,510
18	3,269	1,533	6,015	1,931	4,902

Продовження табл.3.1

1	2	3	4	5	6
19	35,548	17,166	64,386	20,036	50,228
20	8,910	4,797	19,821	5,881	15,004
21	14,019	7,879	32,283	9,622	24,565
22	2,920	1,361	5,296	1,735	4,408
23	10,044	6,670	27,764	8,062	20,813
24	15,477	9,091	36,423	10,679	27,061
25	15,595	9,176	36,212	11,029	27,574
26	2,323	1,433	5,636	1,844	4,699
27	7,101	2,285	9,420	2,918	7,490
28	1,431	0,722	2,396	0,817	2,008
29	37,081	17,071	61,480	19,367	47,085
30	32,826	20,645	71,044	23,650	56,401
31	12,219	8,111	33,723	10,298	26,543
32	59,618	37,982	108,120	40,874	91,235
33	24,864	17,585	66,332	20,438	51,732
34	2,362	1,199	4,582	1,502	3,815
35	0,381	0,396	0,858	-	-
36	4,308	2,922	11,992	3,697	9,415
37	3,412	1,275	5,081	1,620	4,217
38	15,785	12,274	47,116	14,481	35,991
39	0,535	0,430	1,020	0,416	0,925
40	31,676	12,426	48,388	15,259	38,290
41	13,940	8,126	33,053	9,605	24,503
42	3,334	1,061	4,112	1,350	3,499
43	24,298	9,401	41,759	11,033	30,214
44	42,941	7,620	31,153	-	-

Як свідчать дані табл.3.1, на четвертій ітерації викиди відсутні тому, що всі 40 значень Y знаходяться в межах границь інтервалу передбачення нелінійної

регресії. Тому на цьому процес побудови рівняння нелінійної регресії завершується. Остаточні рівняння нелінійної регресії (3.2) побудовано з оцінками параметрів $\hat{b}_0 = 0$, $\hat{b}_1 = 1,072221$, $\hat{b}_2 = 0,503937$, $\hat{b}_3 = -0,045683$ та тими, що наведені в табл.3.2. При цьому вибіркова коваріаційна матриця S_N

$$S_N = \begin{pmatrix} 1,0000 & 0,7067 & 0,5491 & 0,7808 \\ 0,7067 & 1,0000 & 0,8516 & 0,6374 \\ 0,5491 & 0,8516 & 1,0000 & 0,8161 \\ 0,7808 & 0,6374 & 0,8161 & 1,0000 \end{pmatrix}.$$

Таблиця 3.2 – Оцінки параметрів чотиривимірного перетворення Джонсона сім'ї S_B для змінних Y , X_1 , X_2 та X_3 для 40 рядків даних

Індекс j	$\hat{\gamma}_j$	$\hat{\eta}_j$	$\hat{\phi}_j$	$\hat{\lambda}_j$
Y	2,33151	0,67007	0,132	330,648
1	2,50296	0,63891	-0,092	4465,000
2	16,23225	2,26378	-2,515	11714,039
3	0,62893	0,81468	1,560	11,133

Сума квадратів відхилень для рівняння (3.2) на четвертій ітерації остаточно склала 1179,663, що на 45% менше у порівнянні з сумою квадратів відхилень для рівняння (3.2) на другій ітерації.

Були також визначені нижні (LB) і верхні (UB) межі довірчих інтервалів нелінійних регресій за [11] на основі відповідно одновимірного і багатовимірного перетворень Джонсона та одновимірного перетворення у вигляді десяткового логарифму за даними 40 застосунків для рівня значущості 0,05. Ці межі наведені у табл.3.3. Для визначення нижньої і верхньої границь довірчих інтервалів нелінійної регресії побудовано відповідні рівняння

$$\hat{Y}_{CI} = \psi_Y^{-1} \left(\hat{Z}_Y \pm t_{\alpha/2, v} S_{Z_Y} \left\{ \frac{1}{N} + (\mathbf{z}_X^+)^T \left[(\mathbf{Z}_X^+)^T \mathbf{Z}_X^+ \right]^{-1} (\mathbf{z}_X^+) \right\}^{1/2} \right). \quad (3.4)$$

В (3.4) позначення такі ж самі, що і в (3.3). Зауважимо, що в табл.3.3 ще наведені оцінки $\hat{Y}$ та значення MRE за даними 40 застосунків за рівнянням (3.2) з параметрами, які отримані як для чотиривимірного, так і для одновимірного перетворення Джонсона сім'ї S_B .

Таблиця 3.3 – Оцінки $\hat{Y}$ та межі довірчих інтервалів нелінійних регресій

№	одновимірне перетворення				чотиривимірне перетворення			
	$\hat{Y}$	LB	UB	MRE	$\hat{Y}$	LB	UB	MRE
1	2	3	4	5	6	7	8	9
1	180,329	168,582	190,329	0,0309	180,555	163,001	197,773	0,0322
3	63,985	56,929	71,558	0,2249	61,130	54,899	67,895	0,2595
4	13,341	12,194	14,592	0,1097	13,550	12,566	14,607	0,1271
5	5,981	4,358	8,241	0,1186	6,372	4,855	8,365	0,1918
6	0,737	0,637	0,873	0,2256	0,696	0,566	0,866	0,1581
8	30,585	26,130	35,678	0,0809	29,254	25,655	33,306	0,1209
9	32,066	26,292	38,881	0,1100	31,287	26,448	36,908	0,1316
10	85,458	75,633	95,769	0,1475	82,721	73,506	92,676	0,1748
11	4,936	4,430	5,504	0,1073	5,271	4,801	5,787	0,1823
12	3,060	2,658	3,532	0,0242	3,187	2,810	3,617	0,0667
13	3,106	2,786	3,469	0,2324	3,286	2,979	3,625	0,1881
14	5,483	4,778	6,298	0,1801	5,986	5,326	6,727	0,1050
15	1,000	0,838	1,217	0,1981	1,060	0,862	1,312	0,1497
16	7,675	7,028	8,381	0,2864	7,979	7,409	8,593	0,3375
17	40,800	36,552	45,430	0,0463	39,136	35,542	43,043	0,0036
18	2,873	2,570	3,217	0,1212	3,064	2,770	3,392	0,0626
19	33,787	27,905	40,674	0,0495	32,077	27,279	37,619	0,0976
20	8,973	8,053	10,000	0,0071	9,409	8,601	10,292	0,0560

Продовження табл.3.3

1	2	3	4	5	6	7	8	9
21	15,356	13,280	17,739	0,0954	15,444	13,709	17,389	0,1016
22	2,629	2,330	2,972	0,0998	2,753	2,464	3,076	0,0573
23	12,321	10,541	14,393	0,2267	12,999	11,415	14,795	0,2942
24	16,679	14,579	19,063	0,0777	17,087	15,273	19,104	0,1040
25	17,660	16,010	19,468	0,1324	17,530	16,146	19,025	0,1240
26	2,800	2,475	3,175	0,2056	2,931	2,620	3,280	0,2618
27	4,448	3,937	5,031	0,3736	4,667	4,207	5,179	0,3427
28	1,231	1,078	1,417	0,1396	1,266	1,100	1,459	0,1155
29	31,222	27,790	35,011	0,1580	30,493	27,602	33,655	0,1777
30	38,496	34,458	42,907	0,1727	36,953	33,551	40,655	0,1257
31	17,131	14,432	20,300	0,4020	16,618	14,419	19,135	0,3600
32	65,439	58,287	73,098	0,0976	62,220	55,923	69,050	0,0436
33	34,840	28,289	42,604	0,4012	32,898	27,513	39,205	0,3231
34	2,257	1,980	2,582	0,0444	2,380	2,106	2,692	0,0076
36	5,622	5,130	6,164	0,3051	5,896	5,452	6,377	0,3687
37	2,419	2,036	2,889	0,2910	2,600	2,221	3,047	0,2380
38	23,450	20,733	26,485	0,4856	22,996	20,738	25,480	0,4568
39	0,693	0,628	0,775	0,2953	0,607	0,528	0,702	0,1341
40	25,394	21,781	29,530	0,1983	24,365	21,396	27,711	0,2308
41	14,908	12,939	17,161	0,0694	15,411	13,703	17,322	0,1055
42	2,092	1,757	2,507	0,3725	2,159	1,834	2,545	0,3524
43	17,313	13,124	22,737	0,2875	18,381	14,615	23,054	0,2435

Для одновимірного перетворення Джонсона сім'ї S_B рівняння нелінійної регресії (3.2) побудовано з оцінками параметрів $\hat{b}_0 = 0$, $\hat{b}_1 = 1,12050$, $\hat{b}_2 = 0,543689$, $\hat{b}_3 = -0,032126$ та тими, що наведені в табл.3.4. Сума квадратів відхилень для рівняння (3.2) з параметрами, що отримані на основі одновимірного перетворення Джонсона сім'ї S_B для 40 застосунків склала 1026,022.

Таблиця 3.4 – Оцінки параметрів одновимірного перетворення Джонсона сім'ї S_B для змінних Y , X_1 , X_2 та X_3 для 40 рядків даних

Індекс j	$\hat{\gamma}_j$	$\hat{\eta}_j$	$\hat{\phi}_j$	$\hat{\lambda}_j$
Y	1,8721	0,59488	0,368	228,904
1	2,17311	0,6136	0,221	3113,525
2	16,21968	2,18868	-2,241	14392,198
3	0,68671	0,9045	1,381	11,734

Зауважимо, що як можна побачити з табл.3.3, у випадку чотиривимірного перетворення Джонсона сім'ї S_B у порівнянні з одновимірним для 20 застосунків (6, 13-15, 17, 18, 22, 25, 27, 28, 30-34, 37-39, 42,43) з 40 значення MRE менше. Значення MMRE менше у випадку чотиривимірного перетворення Джонсона сім'ї S_B у порівнянні з одновимірним: 0,1753 проти 0,1808.

3.2 Побудова нелінійної регресійної моделі та її інтервалів передбачення для оцінювання розміру програмних застосунків з відкритим кодом на РНР на основі чотиривимірного перетворення Джонсона сімейства S_B

Для побудови рівняння та довірчих інтервалів нелінійної регресії для оцінювання розміру програмних застосунків з відкритим кодом на РНР використовувалися дані, що наведені у табл. 1.2.

На першому етапі початкові дані були перевірені на наявність багатовимірних викидів. Це було зроблено у підрозділі 2.3. Згідно з [77], дані рядка 2 з табл. 1.2 є багатовимірним викидом тому, що для даних рядка 2 квадрат

відстані Махаланобісу, який дорівнює 19,71, більше за значення квантилю розподілу χ^2 , що становить 14,86 для 4 ступенів свободи та рівня значущості 0,005. Дані рядка 2 відкидаються. На цьому перша ітерація завершується і знову переходимо до етапу 1 але вже використовуємо 43 рядки даних з табл.1.2 (за виключенням даних рядка 2).

На другій ітерації згідно з [77], як було показано у підрозділі 2.3, дані 43 рядків даних з табл.1.2 (за виключенням даних рядка 2) не містять багатовимірних викидів тому, що для всіх 43 рядків даних значення квадрату відстані Махаланобісу менше за значення квантилю розподілу χ^2 для 4 ступенів свободи та рівня значущості 0,005. Тому далі переходимо до побудови нелінійного регресійної моделі.

Для цього спочатку для нормалізованих за багатовимірним перетворенням Джонсона сімейства S_B даних з табл.2.6 будуємо лінійну регресійну модель

$$Z_Y = \hat{Z}_Y + \varepsilon = \hat{b}_0 + \hat{b}_1 Z_1 + \hat{b}_2 Z_2 + \hat{b}_3 Z_3 + \varepsilon, \quad (3.5)$$

де ε – випадкова величина з розподілом Гаусу, $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ з оцінкою середньоквадратичного відхилення $\hat{\sigma}_\varepsilon = 0,1525$; параметри моделі (3.5) оцінювалися методом найменших квадратів і їх оцінки є такими ж як для рівняння (3.1): $\hat{b}_0 = 0$, $\hat{b}_1 = 1,072221$, $\hat{b}_2 = 0,503937$, $\hat{b}_3 = -0,045683$. Сума квадратів відхилень для моделі (3.5) склала 1,7085.

Потім за (3.5) на основі багатовимірного перетворення Джонсона сімейства S_B будуємо нелінійну регресійну модель для оцінювання розміру РНР-застосунків у тисячах строк коду в залежності від факторів X_1 , X_2 та X_3

$$Y = \hat{\phi}_Y + \hat{\lambda}_Y \left[1 + e^{-(\hat{Z}_Y + \varepsilon - \hat{\gamma}_Y) / \hat{\eta}_Y} \right]^{-1}, \quad (3.6)$$

де $\hat{Z}_Y = \hat{b}_0 + \hat{b}_1 Z_1 + \hat{b}_2 Z_2 + \hat{b}_3 Z_3$; ε – випадкова величина з розподілом Гаусу, $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ з оцінкою середньоквадратичного відхилення $\hat{\sigma}_\varepsilon = 0,1525$; Z_j – випадкова величина з розподілом Гаусу, $Z_j \sim N(0,1)$, $Z_j = \gamma_j + \eta_j \ln \frac{X_j - \phi_j}{\phi_j + \lambda_j - X_j}$, $\phi_j < X_j < \phi_j + \lambda_j$, $j=1,2,3$. Сума квадратів відхилень для моделі (3.6) з параметрами багатовимірної перетворення Джонсона сімейства S_B , що наведені в табл.2.7, склала 1179,663.

Зауважимо, нелінійна регресійна модель (3.6) для оцінювання розміру РНР-застосунків у тисячах строк коду в залежності від факторів X_1 , X_2 та X_3 отримана за чотири ітерації, при чому було чотири викиди. Тобто ця модель побудована за 40 точками даних.

Для визначення нижньої і верхньої границь інтервалів передбачення нелінійної регресії побудовано відповідні рівняння

$$\hat{Y}_{PI} = \psi_Y^{-1} \left(\hat{Z}_Y \pm t_{\alpha/2, v} S_{Z_Y} \left\{ 1 + \frac{1}{N} + (\mathbf{z}_X^+)^T \left[(\mathbf{Z}_X^+)^T \mathbf{Z}_X^+ \right]^{-1} (\mathbf{z}_X^+) \right\}^{1/2} \right). \quad (3.7)$$

В (3.7) позначення такі ж самі, що і в (3.3).

Для побудови інших нелінійних регресійних моделей виду (3.6) для оцінювання розміру РНР-застосунків у тисячах строк коду в залежності від інших факторів дані з табл.1.2 були доповнені значеннями ще однієї метрики X_4 – це середнє значення DIT (глибина дерева успадкування) на клас. Ці доповнені дані з метрик програмних застосунків з відкритим кодом на РНР наведені в табл.3.5.

Для побудови трьох-факторних нелінійних регресійних моделей для оцінювання розміру РНР-застосунків з відкритим кодом в залежності від різних факторів також було застосовано метод покращення нелінійних регресійних моделей на основі багатовимірних нормалізуючих перетворень та інтервалів передбачення [36].

Таблиця 3.5 – Доповнені дані з метрик програмних застосунків з відкритим кодом на PHP

№	Y	X ₁	X ₂	X ₃	X ₄
1	2	3	4	5	6
1	174,927	2075	4,809	6,425	1,25
2	112,048	445	2,591	3,054	1,10
3	82,551	411	7,856	6,358	1,45
4	12,022	132	5,811	5,053	1,39
5	5,347	5	32,600	2,400	1,20
6	0,601	25	1,080	1,760	2,00
7	1,561	25	2,240	4,040	1,10
8	33,276	216	7,671	10,102	1,37
9	36,028	126	10,849	2,587	1,29
10	100,245	448	9,569	6,315	1,24
11	4,458	73	4,452	5,219	1,29
12	2,988	18	9,611	7,000	1,19
13	4,047	31	6,323	5,355	1,30
14	6,688	125	2,824	3,816	1,71
15	1,247	2	18,000	2,000	2,00
16	5,966	74	6,176	4,851	1,20
17	38,996	269	7,684	6,394	1,54
18	3,269	37	5,108	4,946	1,81
19	35,548	335	5,818	11,096	1,19
20	8,910	117	4,641	3,735	1,27
21	14,019	209	4,536	8,713	1,39
22	2,920	19	8,158	5,737	1,18
23	10,044	314	2,226	4,242	2,28
24	15,477	280	3,400	3,793	1,15

Продовження табл.3.5

1	2	3	4	5	6
25	15,595	115	7,965	5,096	1,30
26	2,323	15	10,067	4,933	1,17
27	7,101	25	9,400	3,440	1,26
28	1,431	22	3,409	2,591	1,54
29	37,081	278	5,888	4,385	1,77
30	32,826	235	8,191	6,430	1,34
31	12,219	58	13,379	9,155	1,27
32	59,618	568	5,974	6,285	1,25
33	24,864	363	5,595	11,474	1,08
34	2,362	28	5,357	6,250	1,36
35	0,381	8	3,125	2,750	2,33
36	4,308	52	6,519	4,923	1,28
37	3,412	52	3,327	7,654	1,13
38	15,785	126	8,968	4,190	1,09
39	0,535	3	9,333	3,333	2,00
40	31,676	76	14,105	7,474	1,00
41	13,940	251	3,410	3,378	1,28
42	3,334	16	8,063	8,813	1,31
43	24,298	794	0,487	2,242	1,08
44	42,941	282	2,872	3,915	2,25

Трьох-факторні моделі нелінійної регресії були побудовані за такими метриками діаграми класів: кількість класів X_1 , середня кількість методів на клас X_2 , сума середнього аферентного та еферентного зв'язків на клас X_3 , середнє значення DIT на клас X_4 . Таким чином, було побудовано три таких моделі в залежності від трьох різних факторів: першу – в залежності від X_1 , X_2 та X_3 , другу – в залежності від X_1 , X_2 та X_4 , і третю – в залежності від X_1 , X_3 та X_4 . Перша трьох-факторна нелінійна регресійна модель (3.6) в залежності від факторів X_1 , X_2 та X_3 була побудована раніше.

Для побудови двох інших трьох-факторні моделей нелінійної регресії спочатку було перевірено наші чотиривимірні дані на наявність багатовимірних відхилень. Але перед цим автором було перевірено нормальність цих багатовимірних даних, оскільки добре відомі статистичні методи (наприклад, виявлення багатовимірних викидів на основі квадрату відстані Махаланобісу) використовуються для виявлення викидів у багатовимірному наборі даних за умови, що дані є гаусовими. Автором застосовано тест на нормальність багатовимірних даних, заснований на вимірах багатовимірних асиметрії та ексцесу [43]. Згідно з цим тестом розподіл чотиривимірних даних не є гаусовим, оскільки статистика тесту на багатовимірну асиметрію цих даних перевищує значення квантилю розподілу χ^2 , що становить 45,31 для 20 ступенів свободи та 0,001 рівень значущості. Аналогічно, статистика тесту для багатовимірного ексцесу перевищує значення квантилю Гауса, який становить 30,46 для середнього і вибіркової дисперсії, що дорівнюють відповідно 24 і 4,36, та 0,001 рівня значущості. Ці результати вказують нам на необхідність подальшого застосування методу для визначення багатовимірних викидів у багатовимірних негаусових даних на основі багатовимірних нормалізуючих перетворень. Що й було зроблено далі згідно з [36].

Також майбутні фактори були перевірено на наявність мультиколінеарності. Наявність мультиколінеарності визначено за коефіцієнтами впливу дисперсії (VIFs) серед майбутніх факторів в моделі множинної лінійної регресії. Для моделі множинної лінійної регресії з k -факторами X_i , $i = 1, 2, \dots, k$, VIFs – це діагональні елементи оберненої коваріаційної матриці $k \times k$ k -факторів [44]. Значення VIFs більше за 10 часто сприймаються як сигнал, що дані мають проблеми з мультиколінеарністю. Для наших даних значення VIFs знаходяться у межах від 1 до 5, тому мультиколінеарності немає.

Нелінійна регресійна модель для оцінювання розміру РНР-застосунків у тисячах строк коду в залежності від факторів X_1 , X_2 та X_4 також має вигляд (3.6) лише з тією різницею, що в (3.6) індекс 3 потрібно замінити на 4, а відповідні оцінки є такими: $\hat{\sigma}_\varepsilon = 0,1533$, $\hat{b}_0 = 0$, $\hat{b}_1 = 1,050805$, $\hat{b}_2 = 0,490140$, $\hat{b}_4 = -0,011783$, $\hat{\gamma}_Y = 1,99105$, $\hat{\gamma}_1 = 2,230711$, $\hat{\gamma}_2 = 18,2818$, $\hat{\gamma}_4 = 8,560105$, $\hat{\eta}_Y = 0,612132$, $\hat{\eta}_1 = 0,618562$, $\hat{\eta}_2 = 2,443858$, $\hat{\eta}_4 = 1,710109$, $\hat{\phi}_Y = 0,113140$, $\hat{\phi}_1 = -0,214048$, $\hat{\phi}_2 = -3,303081$, $\hat{\phi}_4 = 0,900$, $\hat{\lambda}_Y = 240,494$, $\hat{\lambda}_1 = 3099,972$, $\hat{\lambda}_2 = 17083,878$ і $\hat{\lambda}_4 = 62,6318$.

Зауважимо, нелінійна регресійна модель (3.6) для оцінювання розміру РНР-застосунків у тисячах строк коду в залежності від факторів X_1 , X_2 та X_4 отримана за дві ітерації, при чому було два викиди. Тобто ця модель побудована за 42 точками даних.

Нелінійна регресійна модель для оцінювання розміру РНР-застосунків у тисячах строк коду в залежності від факторів X_1 , X_3 та X_4 також має вигляд лише з тією різницею, що в (3.6) індекси 2 і 3 потрібно замінити на 3 і 4, а відповідні оцінки є такими: $\hat{\sigma}_\varepsilon = 0,3882$, $\hat{b}_0 = 0$, $\hat{b}_1 = 0,885605$, $\hat{b}_3 = -0,0227126$, $\hat{b}_4 = -0,190975$, $\hat{\gamma}_Y = 1,474052$, $\hat{\gamma}_1 = 2,364047$, $\hat{\gamma}_3 = 0,875184$, $\hat{\gamma}_4 = 6,287214$, $\hat{\eta}_Y = 0,519960$, $\hat{\eta}_1 = 0,728907$, $\hat{\eta}_3 = 0,953635$, $\hat{\eta}_4 = 1,631979$, $\hat{\phi}_Y = 0,210403$, $\hat{\phi}_1 = -7,908353$, $\hat{\phi}_3 = 1,450813$, $\hat{\phi}_4 = 0,900$, $\hat{\lambda}_Y = 181,2127$, $\hat{\lambda}_1 = 2739,2845$, $\hat{\lambda}_3 = 12,3043$ і $\hat{\lambda}_4 = 20,7263$.

Зауважимо, нелінійна регресійна модель виду (3.6) для оцінювання розміру РНР-застосунків у тисячах строк коду в залежності від факторів X_1 , X_3 та X_4 отримана за дві ітерації, при чому був лише один викид. Тобто ця модель побудована за 43 точками даних.

Для порівняння точності оцінювання розміру РНР-застосунків за допомогою побудованих трьох-факторних нелінійних регресійних моделей в залежності від різних факторів були використані відомі показники точності прогнозування, такі як множинний або вибірковий коефіцієнт детермінації R^2 , середня величина відносної похибки MMRE та відсоток прогнозування на рівні величини відносної помилки 0,25, PRED(0,25). Ці показники зазвичай використовуються для оцінювання якості прогнозування за допомогою регресійних моделей і в інженерії програмного забезпечення [79, 80]. Допустимі значення MMRE і PRED(0,25) складають не більше 0,25 і не менше 0,75 відповідно. Прийнятне значення R^2 приблизно також ж дорівнює 0,75.

Вибірковий коефіцієнт детермінації R^2 визначається як

$$R^2 = 1 - \frac{SS_E}{SS_T},$$

де SS_E – сума квадратів залишків (the residual sum of squares) або сума квадратів помилок (the error sum of squares), $SS_E = \sum_{i=1}^N (Y_i - \hat{Y}_i)^2$; SS_T – загальна сума квадратів (the total sum of squares), $SS_T = \sum_{i=1}^N (Y_i - \bar{Y})^2$. Нагадаємо, що $SS_T = SS_R + SS_E$. Тут SS_R – сума квадратів регресії (the regression sum of squares), $SS_R = \sum_{i=1}^N (\hat{Y}_i - \bar{Y})^2$.

Коефіцієнт детермінації R^2 інтерпретується як частка мінливості у спостережуваній залежній змінній Y , що пояснюється моделлю лінійної регресії. Значення R^2 вказує наскільки емпіричні дані підтверджують математичну модель або, чи краща наша модель за вибіркове середнє. З формули для R^2 , $0 \leq R^2 \leq 1$, значення R^2 вважається прийнятним, якщо воно більше за 0,75. Велике значення R^2 свідчить про те, що модель успішно пояснює мінливість Y . Коли R^2 малий, це

може свідчити про те, що потрібно знайти альтернативну модель, яка може врахувати більшу мінливість Y .

Нелінійна регресійна модель, що побудована навколо кількості класів, середньої кількості методів на клас, суми середнього аферентного та еферентного зв'язків на клас, має наступні значення R^2 , MMRE та PRED(0,25): 0,9726, 0,1753 та 0,7500 відповідно. Значення R^2 та MMRE для цієї моделі є кращими у порівнянні з іншими моделями.

Значення величини відносної похибки MRE обчислюється як

$$MRE_i = \left| \frac{(Y_i - \hat{Y}_i)}{Y_i} \right|.$$

Середня величина відносної похибки MMRE визначається як

$$MMRE = \frac{1}{N} \sum_{i=1}^N MRE_i.$$

Зазвичай $MMRE \leq 0,25$ вважається прийнятною точністю моделі.

Значення PRED(0,25) обчислюється як

$$PRED(0,25) = \frac{1}{N} \sum_{i=1}^N \begin{cases} 1 & \text{if } MRE_i \leq 0,25 \\ 0 & \text{otherwise.} \end{cases}.$$

Зазвичай $PRED(0,25) \geq 0,75$ вважається прийнятною точністю прогнозування за допомогою регресійних моделей.

Нелінійна регресійна модель, що побудована навколо кількості класів, середньої кількості методів на клас, суми середнього аферентного та еферентного зв'язків на клас, має наступні значення R^2 , MMRE та PRED(0,25): 0,9726, 0,1753 та 0,7500 відповідно. Ця модель має на 4,4% менше значення MMRE та приблизно однакові значення R^2 та PRED(0,25) (з різницею у 0,3% та 1,6% відповідно), що модель, в якій третім фактором є сума середнього аферентного та еферентного зв'язків на клас.

Нелінійна регресійна модель, що побудована навколо кількості класів, середньої кількості методів на клас, середнього значення DIT на клас, має наступні значення R^2 , MMRE та PRED(0,25): 0,9754, 0,1831 та 0,7381 відповідно. Ця модель має на 4% більше значення MMRE та приблизно однакові значення R^2 та PRED(0,25) (з різницею у 0,3% та 1,6% відповідно), що і попередня модель, в якій третім фактором є сума середнього аферентного та еферентного зв'язків на клас.

Нелінійна регресійна модель виду (3.6), що побудована навколо кількості класів, суми середнього аферентного та еферентного зв'язків на клас, середнього значення DIT на клас має найгірші значення R^2 , MMRE та PRED(0,25), які дорівнюють 0,8646, 0,5659 та 0,2558 відповідно. Значення MMRE та PRED(0,25) для цієї моделі вказують на погану точність оцінювання залежної випадкової величини Y . Лише значення R^2 є добрим і говорить про можливість використання моделі виду (3.6) в залежності від факторів X_1 , X_3 та X_4 для оцінювання вибіркового середнього величини Y , коли ε прямує у нуль.

Підкреслимо, що для нелінійної регресійної моделі, що побудована в залежності від факторів X_1 , X_2 та X_4 , а саме – кількості класів, середньої кількості методів на клас, середнього значення DIT на клас, ширина інтервалу передбачення кількості тисяч строк коду для PHP-застосунків з розміром понад 39 KLOC (thousand lines of code) є суттєво меншою у порівнянні з моделлю, що побудована в залежності від факторів X_1 , X_2 та X_3 . Так, для застосунку 1 ($X_1=2075$, $X_2=4,809$ та $X_3=6,425$) за даними, що наведені в табл. 3.1, для моделі, що побудована в залежності від факторів X_1 , X_2 та X_3 , маємо такий інтервал передбачення залежної змінної Y : [137,430; 221,703]. Причому, актуальне значення Y для цього застосунку (а це Symfony-master), яке було отримано за допомогою PhpMetrics, дорівнює 174,927 KLOC. Для цього ж застосунку 1 за значеннями факторів $X_1=2075$, $X_2=4,809$ та $X_4=1,25$ для моделі, що побудована в

залежності від факторів X_1 , X_2 та X_4 , маємо такий інтервал передбачення залежної змінної Y : [145,090; 199,384], ширина якого на 36% менша за ширину відповідного інтервалу для моделі, що побудована в залежності від факторів X_1 , X_2 та X_3 . Подібний результат отримано й для довірчого інтервалу. Так, для значень зазначених вище факторів застосунку 1 для моделі, що побудована в залежності від факторів X_1 , X_2 та X_4 , маємо такий довірчий інтервал вибіркового середнього змінної Y : [163,733; 186,529], ширина якого на 34% менша за ширину відповідного інтервалу для моделі, що побудована в залежності від факторів X_1 , X_2 та X_3 , якій є таким [163,001; 197,773]. Зменшення ширин відповідних інтервалів також вказує на підвищення точності оцінювання розміру РНР-застосунків з відкритим кодом понад 39 KLOC за допомогою нелінійної регресійної моделі, що побудована в залежності від кількості класів, середньої кількості методів на клас, середнього значення DIT на клас.

Границі інтервалів передбачення нелінійних регресій, що отримані за допомогою чотиривимірного перетворення Джонсона сім'ї S_B , для різних факторів наведені в табл.3.6.

Таблиця 3.6 – Границі інтервалів передбачення нелінійних регресій, що отримані за допомогою чотиривимірного перетворення Джонсона сім'ї S_B , для різних факторів

№	Y	X_1, X_2 та X_3		X_1, X_2 та X_4		X_1, X_3 та X_4	
		LB	UB	LB	UB	LB	UB
1	2	3	4	5	6	7	8
1	174,927	137,430	221,703	145,090	199,384	131,233	179,433
2	112,048	-	-	-	-	21,467	145,475
3	82,551	40,098	89,814	37,651	86,061	11,406	113,470
4	12,022	8,508	21,425	8,079	21,923	3,153	51,520

Продовження табл.3.6

1	2	3	4	5	6	7	8
5	5,347	3,733	10,881	3,629	11,315	0,403	5,560
6	0,601	0,458	1,106	0,418	1,070	0,494	8,470
7	1,561	-	-	0,620	1,721	0,992	18,992
8	33,276	18,403	45,611	19,023	48,367	5,072	77,904
9	36,028	19,485	49,153	17,319	44,701	3,431	58,778
10	100,245	55,464	117,672	52,208	109,244	16,779	130,046
11	4,458	3,298	8,440	3,232	8,997	1,886	33,260
12	2,988	1,995	5,132	2,023	5,687	0,657	11,190
13	4,047	2,069	5,257	2,018	5,595	0,857	14,811
14	6,688	3,723	9,633	3,528	10,020	2,241	40,164
15	1,247	0,675	1,718	0,611	1,713	0,273	2,036
16	5,966	5,000	12,711	4,812	13,330	2,202	38,677
17	38,996	25,110	59,510	23,566	58,865	6,065	82,986
18	3,269	1,931	4,902	1,819	5,112	0,667	11,347
19	35,548	20,036	50,228	22,093	55,002	11,227	116,883
20	8,910	5,881	15,004	5,547	15,291	3,266	53,730
21	14,019	9,622	24,565	9,940	26,756	4,981	74,550
22	2,920	1,735	4,408	1,717	4,808	0,703	11,996
23	10,044	8,062	20,813	7,596	22,155	4,562	75,321
24	15,477	10,679	27,061	10,369	28,220	11,086	114,011
25	15,595	11,029	27,574	10,437	27,915	3,029	49,811
26	2,323	1,844	4,699	1,800	5,076	0,626	10,326
27	7,101	2,918	7,490	2,718	7,599	0,784	13,529
28	1,431	0,817	2,008	0,753	2,013	0,574	9,055
29	37,081	19,367	47,085	17,502	45,988	5,375	78,490
30	32,826	23,650	56,401	22,572	56,135	6,459	84,484
31	12,219	10,298	26,543	10,599	28,807	1,414	27,348
32	59,618	40,874	91,235	39,801	89,095	22,472	141,885
33	24,864	20,438	51,732	23,201	58,127	15,816	132,787
34	2,362	1,502	3,815	1,497	4,125	0,734	12,473
35	0,381	-	-	0,306	0,720	0,304	2,855

Продовження табл.3.6

1	2	3	4	5	6	7	8
36	4,308	3,697	9,415	3,552	9,850	1,377	24,629
37	3,412	1,620	4,217	1,712	4,894	1,684	31,592
38	15,785	14,481	35,991	13,614	36,260	4,942	75,357
39	0,535	0,416	0,925	0,385	0,954	0,279	2,117
40	31,676	15,259	38,290	15,254	42,265	3,667	66,202
41	13,940	9,605	24,503	9,146	24,857	7,827	96,065
42	3,334	1,350	3,499	1,419	3,918	0,519	8,473
43	24,298	11,033	30,214	12,239	35,169	-	-
44	42,941	-	-	-	-	4,085	69,572

Межі довірчих інтервалів нелінійних регресій, що отримані за допомогою чотиривимірного перетворення Джонсона сім'ї S_B , для різних факторів наведені в табл.3.6. В табл.3.7 ще наведені оцінки $\hat{Y}$ та значення MRE за рівнянням (3.2) з параметрами, які отримані для чотиривимірного перетворення Джонсона сім'ї S_B за даними 40 застосунків для факторів X_1 , X_2 та X_3 та за даними 42 застосунків для факторів X_1 , X_2 та X_4 .

Таблиця 3.7 – Оцінки $\hat{Y}$ та межі довірчих інтервалів нелінійних регресій з різними факторами

№	X_1, X_2 та X_3				X_1, X_2 та X_4			
	$\hat{Y}$	LB	UB	MRE	$\hat{Y}$	LB	UB	MRE
1	2	3	4	5	6	7	8	9
1	180,555	163,001	197,773	0,0322	175,761	163,733	186,529	0,0048
3	61,130	54,899	67,895	0,2595	58,514	51,727	65,882	0,2912
4	13,550	12,566	14,607	0,1271	13,394	12,305	14,574	0,1141
5	6,372	4,855	8,365	0,1918	6,418	4,914	8,376	0,2002

Продовження табл.3.7

1	2	3	4	5	6	7	8	9
6	0,696	0,566	0,866	0,1581	0,653	0,543	0,792	0,0868
7	-	-	-	-	1,017	0,823	1,263	0,3487
8	29,254	25,655	33,306	0,1209	30,818	27,824	34,084	0,0739
9	31,287	26,448	36,908	0,1316	28,237	25,252	31,525	0,2162
10	82,721	73,506	92,676	0,1748	78,033	69,916	86,615	0,2216
11	5,271	4,801	5,787	0,1823	5,393	4,909	5,925	0,2098
12	3,187	2,810	3,617	0,0667	3,382	2,978	3,843	0,1319
13	3,286	2,979	3,625	0,1881	3,351	3,046	3,688	0,1720
14	5,986	5,326	6,727	0,1050	5,950	5,175	6,842	0,1103
15	1,060	0,862	1,312	0,1497	1,007	0,805	1,266	0,1928
16	7,979	7,409	8,593	0,3375	8,030	7,320	8,808	0,3460
17	39,136	35,542	43,043	0,0036	37,984	33,404	43,062	0,0260
18	3,064	2,770	3,392	0,0626	3,039	2,666	3,466	0,0703
19	32,077	27,279	37,619	0,0976	35,490	32,136	39,129	0,0016
20	9,409	8,601	10,292	0,0560	9,243	8,463	10,094	0,0374
21	15,444	13,709	17,389	0,1016	16,444	14,947	18,080	0,1730
22	2,753	2,464	3,076	0,0573	2,862	2,519	3,253	0,0199
23	12,999	11,415	14,795	0,2942	13,065	10,576	16,106	0,3008
24	17,087	15,273	19,104	0,1040	17,261	15,188	19,595	0,1153
25	17,530	16,146	19,025	0,1240	17,218	15,797	18,758	0,1041
26	2,931	2,620	3,280	0,2618	3,011	2,621	3,461	0,2964
27	4,667	4,207	5,179	0,3427	4,541	4,084	5,049	0,3606
28	1,266	1,100	1,459	0,1155	1,216	1,065	1,391	0,1500
29	30,493	27,602	33,655	0,1777	28,819	24,794	33,399	0,2228
30	36,953	33,551	40,655	0,1257	36,256	32,663	40,170	0,1045
31	16,618	14,419	19,135	0,3600	17,636	15,513	20,025	0,4433
32	62,220	55,923	69,050	0,0436	61,227	55,134	67,749	0,0270
33	32,898	27,513	39,205	0,3231	37,443	32,884	42,507	0,5059
34	2,380	2,106	2,692	0,0076	2,473	2,234	2,740	0,0471
35	-	-	-	-	0,456	0,386	0,542	0,1960
36	5,896	5,452	6,377	0,3687	5,919	5,444	6,436	0,3740

Продовження табл.3.7

1	2	3	4	5	6	7	8	9
37	2,600	2,221	3,047	0,2380	2,883	2,448	3,399	0,1550
38	22,996	20,738	25,480	0,4568	22,486	19,799	25,498	0,4245
39	0,607	0,528	0,702	0,1341	0,591	0,501	0,702	0,1052
40	24,365	21,396	27,711	0,2308	25,783	20,913	31,629	0,1860
41	15,411	13,703	17,322	0,1055	15,194	13,650	16,900	0,0899
42	2,159	1,834	2,545	0,3524	2,345	2,098	2,623	0,2965
43	18,381	14,615	23,054	0,2435	21,015	16,683	26,347	0,1351

Отримані результати свідчать про наступне. До факторів, які найбільшим чином впливають на точність оцінювання розміру РНР-застосунків з відкритим кодом, насамперед потрібно віднести кількість класів. У найбільшому впливі цього фактору можна пересвідчитися порівнявши за абсолютною величиною значення оцінок параметрів b_j : значення $\hat{b}_1$ є найбільшим для всіх моделей. Другим за впливом фактором є середня кількість методів на клас. Зазначимо, що не врахування цього фактору у моделі, що була побудована в залежності від факторів X_1 , X_3 та X_4 призводить до поганої точності оцінювання залежної випадкової величини Y на що вказують значення MMRE та PRED(0,25) для цієї моделі. Тому для підвищення точності оцінювання розміру РНР-застосунків з відкритим кодом, кількість строк якого буде перевищувати 39 KLOC, на наш погляд, може використовуватися нелінійна регресійна модель, що побудована в залежності від факторів X_1 , X_2 та X_4 , а саме – кількості класів, середньої кількості методів на клас, середнього значення DIT на клас. А у разі, коли кількість строк коду РНР-застосунку буде менше за 39 KLOC може бути застосована нелінійна регресійна модель, що побудована в залежності від факторів X_1 , X_2 та X_3 , а саме – кількості класів, середньої кількості методів на клас, суми середнього аферентного та еферентного зв'язків на клас. У випадку, коли заздалегідь не

відомо, чи буде розмір РНР-застосунку більше або менше за 39 KLOC, то, на нашу думку, слід використовувати трьох-факторну нелінійну регресійну модель, що побудована в залежності від кількості класів, середньої кількості методів на клас та середнього значення DIT на клас тому, що ця модель має менші ширини інтервалів передбачення для більшої кількості даних та приблизно однакові значення R^2 та PRED(0,25), що і попередня модель, в якій третім фактором є сума середнього аферентного та еферентного зв'язків на клас.

3.3 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

Результати оцінювання розміру програмних застосунків з відкритим кодом на РНР за нелінійною регресійною моделлю (3.5), що побудована на основі чотиривимірного перетворення Джонсона сімейства S_B , були порівняні з нелінійними регресійними моделями на основі одновимірних нормалізуючих перетворень. У якості таких одновимірних нормалізуючих перетворень використовуються перетворення у вигляді десяткового логарифму (2.3), перетворення Бокса-Кокса (2.5) та Джонсона (2.7) сімейства S_B .

Трьох-факторна нелінійна регресійна модель для оцінювання розміру РНР-застосунків з відкритим кодом була побудована із використанням одновимірного нормалізуючого перетворення у вигляді десяткового логарифму (2.3). Ця модель має такий вигляд:

$$Y = 10^{\varepsilon + b_0} X_1^{b_1} X_2^{b_2} X_3^{b_3}, \quad (3.8)$$

де Y – розмір РНР-застосунків з відкритим кодом у тисячах рядків коду; X_1 – кількість класів; X_2 – середня кількість методів; X_3 – сума середньої кількості класів, на які впливає даний клас і середньої кількості класів, з яких даний клас отримує ефекти; ε – випадкова величина з розподілом Гаусу, $\varepsilon \sim N(0, \sigma_\varepsilon)$, з оцінкою середньоквадратичного відхилення $\hat{\sigma}_\varepsilon = 0,091431$; оцінки параметрів $\hat{b}_0 = -1,7564669$, $\hat{b}_1 = 1,000896$, $\hat{b}_2 = 1,2170157$, $\hat{b}_3 = -0,2120602$ визначалися за методом найменших квадратів.

При побудові моделі за емпіричними даними з метрик 44 РНР-застосунків, що наведені в табл.1.2, використання методу [36] призводить до відкидання чотирьох рядків даних – це рядки 2, 7, 43 і 44, які є багатовимірними викидами.

Порівняння моделі (3.8), що була побудована за 40 рядками даних з табл.1.2 після відкидання чотирьох рядків 2, 7, 43 і 44, здійснювалося з трьома іншими нелінійними регресійними моделями створеними на основі одновимірних перетворень Бокса-Кокса і Джонсона для сімейства S_B та чотиривимірною перетворення Джонсона для сімейства S_B .

Трьох-факторна нелінійна регресійна модель для оцінювання розміру РНР-застосунків з відкритим кодом на основі одновимірною перетворення Бокса-Кокса (2.5), що була побудована за 40 рядками даних табл.1.2 після відкидання чотирьох рядків 2, 7, 43 і 44, є такою:

$$Y = \left[\hat{\lambda}_Y (\hat{b}_0 + \hat{b}_1 Z_1 + \hat{b}_2 Z_2 + \hat{b}_3 Z_3 + \varepsilon) + 1 \right]^{1/\hat{\lambda}_Y}, \quad (3.9)$$

де ε – випадкова величина з розподілом Гаусу, $\varepsilon \sim N(0, \sigma_\varepsilon)$, з оцінкою $\hat{\sigma}_\varepsilon = 0,246528$; оцінки параметрів $\hat{b}_0 = -3,6386816$, $\hat{b}_1 = 0,71215362$, $\hat{b}_2 = 1,101478$ і $\hat{b}_3 = -0,07283967$ визначалися за методом найменших квадратів; Z_1 , Z_2 і Z_3 – це

відповідно нормалізовані за одновимірним перетворенням Бокса-Кокса величини X_1 , X_2 і X_3 , яке має вигляд (2.3)

$$Z_j = \begin{cases} (X_j^{\lambda_j} - 1)/\lambda_j, & \text{якщо } \lambda_j \neq 0; \\ \ln(X_j), & \text{якщо } \lambda_j = 0. \end{cases}, \quad (3.10)$$

В (3.10) $j=1,2,3$. Змінна Z_Y визначається аналогічно (3.10) тільки з тією різницею, що Z_j , X_j та λ_j потрібно замінити відповідно на Z_Y , Y , та λ_Y . Оцінки параметрів $\hat{\lambda}_Y = 0,0422406$, $\hat{\lambda}_1 = 0,0983839$, $\hat{\lambda}_2 = 0,0631603$ та $\hat{\lambda}_3 = 0,2649154$ визначалися за методом ММП шляхом максимізації логарифму функції правдоподібності (2.8).

Трьох-факторна нелінійна регресійна модель для оцінювання розміру РНР-застосунків з відкритим кодом на основі одновимірного перетворення Джонсона для сімейства S_B , що була побудована за 40 рядками даних табл.1.2 після відкидання чотирьох рядків 2, 7, 43 і 44, має вигляд (3.6) з тією різницею, що відповідні оцінки є такими: $\hat{\sigma}_\varepsilon = 0,246528$, $\hat{b}_0 = 0$, $\hat{b}_1 = 1,0520157$, $\hat{b}_2 = 0,5060942$, $\hat{b}_3 = -0,05267307$, $\hat{\gamma}_Y = 1,774797$, $\hat{\gamma}_1 = 2,274354$, $\hat{\gamma}_2 = 14,2434$, $\hat{\gamma}_3 = 0,75683$, $\hat{\eta}_Y = 0,557969$, $\hat{\eta}_1 = 0,608866$, $\hat{\eta}_2 = 1,92648$, $\hat{\eta}_3 = 0,967899$, $\hat{\phi}_Y = 0,280692$, $\hat{\phi}_1 = 0,547627$, $\hat{\phi}_2 = -0,732189$, $\hat{\phi}_3 = 1,31032$, $\hat{\lambda}_Y = 215,6003$, $\hat{\lambda}_1 = 3272,1597$, $\hat{\lambda}_2 = 11660,785$ та $\hat{\lambda}_3 = 12,2236$.

Трьох-факторна нелінійна регресійна модель для оцінювання розміру РНР-застосунків з відкритим кодом на основі чотиривимірного перетворення Джонсона для сімейства S_B , що була побудована за 40 рядками даних табл.1.2 після відкидання чотирьох рядків 2, 7, 43 і 44, має вигляд (3.6) з тією різницею, що відповідні оцінки є такими: $\hat{\sigma}_\varepsilon = 0,1351$, $\hat{b}_0 = 0$, $\hat{b}_1 = 1,057585$, $\hat{b}_2 = 0,505529$, $\hat{b}_3 = -0,0420$, $\hat{\gamma}_Y = 4,84396$, $\hat{\gamma}_1 = 3,77629$, $\hat{\gamma}_2 = 15,3975$, $\hat{\gamma}_3 = 0,7440$, $\hat{\eta}_Y = 0,68696$,

$\hat{\eta}_1 = 0,662314$, $\hat{\eta}_2 = 2,08781$, $\hat{\eta}_3 = 0,882698$, $\hat{\phi}_Y = 0,041111$, $\hat{\phi}_1 = -0,092048$,
 $\hat{\phi}_2 = -1,27715$, $\hat{\phi}_3 = 1,52426$, $\hat{\lambda}_Y = 10053,641$, $\hat{\lambda}_1 = 22683,338$, $\hat{\lambda}_2 = 12428,790$ і
 $\hat{\lambda}_3 = 11,8020$.

Побудовані моделі (3.6), (3.8) і (3.9) були перевірені за сумою квадратів помилок SS_E , 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). Показники якості моделей при визначенні викидів за десятковим логарифмом наведені в табл.3.8 та свідчать про те, що найкращою за MMRE є модель (3.2), побудована за допомогою чотиривимірного перетворення Джонсона для сімейства S_B , найкращою за PRED(0,25) є моделі (3.2) і (3.8), найкращою за SS_E і R^2 є модель (3.2). При цьому діапазони зміни факторів X_1 , X_2 і X_3 є такими: від 2 до 2075; від 1,08 до 32,60 та від 1,760 до 11,474 відповідно.

Таблиця 3.8 – Показники якості нелінійних регресійних моделей при визначенні викидів за одновимірним перетворенням у вигляді десятичного логарифму

№	Перетворення	SS_E	R^2	MMRE	PRED(0,25)
1	десятичного логарифму	1206,09	0,9722	0,1719	0,7750
2	одновимірне Бокса-Кокса	4412,72	0,8981	0,1849	0,6750
3	одновимірне Джонсона S_B	938,97	0,9849	0,1819	0,7250
4	чотиривимірне Джонсона S_B	1005,14	0,9838	0,1576	0,7750

Показники якості нелінійних регресійних моделей при визначенні викидів за одновимірним перетворенням Бокса-Кокса наведені в табл.3.9. Зазначимо, при побудові моделі (3.9) на основі одновимірного перетворення Бокса-Кокса за

емпіричними даними з метрик 44 РНР-застосунків, що наведені в табл.1.2, використання методу [36] призводить до відкидання чотирьох рядків даних – це рядки 2, 40, 43 і 44.

Також нелінійні регресійні моделі (3.2) і (3.8) для оцінювання розміру РНР-застосунків з відкритим кодом були побудовані за 40 рядками даних табл.1.2 після відкидання чотирьох рядків 2, 40, 43 і 44. При цьому діапазони зміни факторів X_1 , X_2 і X_3 є такими: від 2 до 2075; від 1,08 до 32,60 та від 1,760 до 11,474 відповідно.

Таблиця 3.9 – Показники якості нелінійних регресійних моделей при визначенні викидів за одновимірним перетворенням Бокса-Кокса

№	Перетворення	SS_E	R^2	MMRE	PRED(0,25)
1	десятього логарифму	1064,96	0,9826	0,1864	0,7000
2	одновимірне Бокса-Кокса	5268,32	0,8791	0,1787	0,6750
3	одновимірне Джонсона S_B	1075,37	0,9824	0,1846	0,7000
4	чотиривимірне Джонсона S_B	1026,71	0,9832	0,1824	0,7250

Показники якості зазначених моделей з табл.3.9 свідчать про те, що найкращою за MMRE є модель (3.9) на основі одновимірного перетворення Бокса-Кокса, найкращою за іншими показниками є модель (3.6) на основі чотиривимірного перетворення Джонсона для сімейства S_B .

Показники якості нелінійних регресійних моделей при визначенні викидів за одновимірним перетворенням Джонсона для сімейства S_B наведені в табл.3.10. Зазначимо, при побудові моделі (3.6) на основі одновимірного перетворенням Джонсона для сімейства S_B за емпіричними даними з метрик 44 РНР-застосунків, що наведені в табл.1.2, використання методу [36] призводить до відкидання 15

рядків даних – це рядки 2, 6, 7, 12, 15, 18, 22, 26-28, 34, 35, 39, 42 і 44. Також моделі (3.8) і (3.9) для оцінювання розміру РНР-застосунків з відкритим кодом були побудовані за 29 рядками даних з табл.1.2 після відкидання зазначених 15 рядків даних. При цьому діапазони зміни факторів X_1 , X_2 і X_3 є такими: від 5 до 2075; від 0,487 до 32,60 та від 2,242 до 11,474 відповідно.

Таблиця 3.10 – Показники якості нелінійних регресійних моделей при визначенні викидів за одновимірним перетворенням Джонсона сімейства S_B

№	Перетворення	SS_E	R^2	MMRE	PRED(0,25)
1	десятьового логарифму	2046,55	0,9550	0,2135	0,4138
2	одновимірне Бокса-Кокса	82214,33	-	0,2290	0,5517
3	одновимірне Джонсона S_B	1367,31	0,9782	0,1586	0,8621
4	чотиривимірне Джонсона S_B	617,88	0,9901	0,1164	0,9310

Дані в табл.3.10 вказують на те, що найкращою за всіма показниками є модель (3.6) на основі чотиривимірного перетворення Джонсона для сімейства S_B .

Показники якості нелінійних регресійних моделей при визначенні викидів за чотиривимірним перетворенням Джонсона для сімейства S_B наведені в табл.3.11. Підкреслимо, при побудові моделі (3.6) на основі чотиривимірного перетворення Джонсона для сімейства S_B за емпіричними даними з метрик 44 РНР-застосунків, що наведені в табл.1.2, як це було зазначено раніше, використання методу [36] призводить до відкидання чотирьох рядків даних – це рядки 2, 7, 35 і 44.

Також моделі (3.8) і (3.9) для оцінювання розміру РНР-застосунків з відкритим кодом були побудовані за 40 рядками даних з табл.1.2 після відкидання

зазначених чотирьох рядків даних. Показники якості з табл.3.11 свідчать про те, що найкращою за MMRE є модель (3.6) на основі чотиривимірного перетворення Джонсона для сімейства S_B , найкращою за PRED(0,25) також є модель (3.6), найкращою за SS_E і R^2 є модель (3.6) на основі одновимірного перетворення Джонсона.

Таблиця 3.11 – Показники якості нелінійних регресійних моделей при визначенні викидів за чотиривимірним перетворенням Джонсона сімейства S_B

№	Перетворення	SS_E	R^2	MMRE	PRED(0,25)
1	десятьового логарифму	23675,98	0,4496	0,5248	0,3500
2	одновимірне Бокса-Кокса	27446,55	0,3620	0,4530	0,2500
3	одновимірне Джонсона S_B	1026,02	0,9836	0,1808	0,7500
4	чотиривимірне Джонсона S_B	1179,66	0,9726	0,1753	0,7500
5	чотиривимірне Бокса-Кокса	8490,14	0,8026	0,3725	0,3000

Зазначимо, для моделі (3.6) на основі чотиривимірного перетворенням Джонсона для сімейства S_B , як і для інших, що побудовані за емпіричними даними з метрик 40 РНР-застосунків, діапазони зміни факторів X_1 , X_2 і X_3 є такими: від 2 до 2075; від 0,487 до 32,60 та від 1,760 до 11,474 відповідно. При цьому ці діапазони є такими ж як і для початкових даних з 44 РНР-застосунків на відміну від попередніх трьох випадків застосування одновимірних перетворень для побудови нелінійних регресійних моделей за методом [36] із врахуванням наявності викидів.

Також модель (3.6) в залежності від факторів X_1 , X_2 та X_3 на основі чотиривимірного перетворення Джонсона для сімейства S_B , є найкращою за

ширинами інтервалів передбачення та довірчих інтервалів для більшої кількості даних у порівнянні з нелінійними регресійними моделями, що побудовані на основі як одновимірних перетворень, так і чотиривимірного перетворення Бокса-Кокса. Границі інтервалів передбачення нелінійних регресій, що отримані для різних перетворень в залежності від факторів X_1 , X_2 та X_3 наведені в табл.3.12.

Таблиця 3.12 – Границі інтервалів передбачення нелінійних регресій, що отримані для різних перетворень в залежності від факторів X_1 , X_2 та X_3

№	перетворення Джонсона S_B				перетворення Бокса-Кокса			
	чотиривимірне		одновимірне		чотиривимірне		одновимірне	
	LB	UB	LB	UB	LB	UB	LB	UB
1	2	3	4	5	6	7	8	9
1	137,430	221,703	150,412	201,033	177,149	273,486	247,883	312,483
3	40,098	89,814	39,759	95,620	57,680	149,631	65,268	185,240
4	8,508	21,425	7,652	23,018	8,087	28,471	6,660	30,363
5	3,733	10,881	3,219	11,290	3,007	13,306	2,355	13,866
6	0,458	1,106	0,556	1,090	0,239	0,603	0,190	0,480
8	18,403	45,611	17,643	50,908	22,146	72,983	21,514	88,366
9	19,485	49,153	18,285	53,762	21,925	74,724	17,719	77,939
10	55,464	117,672	55,452	120,511	83,707	189,944	96,164	226,359
11	3,298	8,440	2,880	8,614	2,403	8,670	1,793	8,348
12	1,995	5,132	1,828	5,309	1,135	4,056	0,887	4,033
13	2,069	5,257	1,866	5,350	1,245	4,380	0,928	4,136
14	3,723	9,633	3,164	9,645	2,980	10,941	2,249	10,729
15	0,675	1,718	0,693	1,594	0,330	0,991	0,333	1,293
16	5,000	12,711	4,425	13,375	3,944	14,161	2,988	13,923

Продовження табл.3.12

1	2	3	4	5	6	7	8	9
17	25,110	59,510	24,267	65,156	32,262	97,245	32,493	117,782
18	1,931	4,902	1,737	4,933	1,177	4,125	0,866	3,830
19	20,036	50,228	19,372	56,227	26,349	85,193	28,632	111,353
20	5,881	15,004	5,141	15,672	5,111	18,465	3,978	18,691
21	9,622	24,565	8,715	26,652	10,012	35,784	9,235	42,544
22	1,735	4,408	1,599	4,503	0,969	3,365	0,749	3,274
23	8,062	20,813	6,965	21,611	8,380	30,330	7,755	36,299
24	10,679	27,061	9,501	28,770	11,742	41,209	10,917	49,031
25	11,029	27,574	10,147	30,160	10,873	37,674	8,900	39,864
26	1,844	4,699	1,693	4,817	1,027	3,583	0,800	3,532
27	2,918	7,490	2,602	7,770	1,832	6,618	1,340	6,220
28	0,817	2,008	0,830	1,978	0,435	1,347	0,329	1,182
29	19,367	47,085	18,228	51,386	23,804	76,170	23,339	92,132
30	23,650	56,401	22,794	61,913	29,660	90,949	28,895	108,066
31	10,298	26,543	9,664	29,792	9,618	35,018	7,521	35,891
32	40,874	91,235	40,797	97,370	59,881	153,400	74,001	198,315
33	20,438	51,732	19,881	58,103	27,659	88,973	30,774	117,909
34	1,502	3,815	1,392	3,842	0,855	2,955	0,647	2,791
36	3,697	9,415	3,271	9,797	2,624	9,422	1,945	9,019
37	1,620	4,217	1,464	4,192	1,001	3,624	0,741	3,395
38	14,481	35,991	13,503	39,563	15,340	52,139	12,683	55,510
39	0,416	0,925	0,540	0,982	0,219	0,488	0,214	0,576
40	15,259	38,290	14,548	42,892	16,058	55,217	12,793	57,028
41	9,605	24,503	8,466	25,883	10,189	36,336	9,112	41,890
42	1,350	3,499	1,286	3,596	0,722	2,539	0,581	2,571
43	11,033	30,214	9,381	31,248	9,384	37,504	9,823	51,308

Границі довірчих інтервалів нелінійних регресій, що отримані для різних перетворень в залежності від факторів X_1 , X_2 та X_3 наведені в табл.3.13.

Таблиця 3.13 – Границі довірчих інтервалів нелінійних регресій, що отримані для різних перетворень в залежності від факторів X_1 , X_2 та X_3

№	перетворення Джонсона S_B				перетворення Бокса-Кокса			
	чотирирівимірне		однорівимірне		чотирирівимірне		однорівимірне	
	LB	UB	LB	UB	LB	UB	LB	UB
1	2	3	4	5	6	7	8	9
1	163,001	197,773	168,582	190,329	212,589	249,303	275,556	301,336
3	54,899	67,895	56,929	71,558	85,656	110,118	102,710	135,561
4	12,566	14,607	12,194	14,592	13,796	16,956	12,711	16,253
5	4,855	8,365	4,358	8,241	4,259	9,388	3,596	9,049
6	0,566	0,866	0,637	0,873	0,292	0,447	0,227	0,347
8	25,655	33,306	26,130	35,678	34,585	48,966	36,679	55,814
9	26,448	36,908	26,292	38,881	32,994	52,049	29,349	50,180
10	73,506	92,676	75,633	95,769	117,174	150,296	139,773	181,366
11	4,801	5,787	4,430	5,504	4,013	5,157	3,301	4,469
12	2,810	3,617	2,658	3,532	1,774	2,537	1,501	2,290
13	2,979	3,625	2,786	3,469	2,021	2,643	1,638	2,257
14	5,326	6,727	4,778	6,298	4,870	6,679	4,045	5,921
15	0,862	1,312	0,838	1,217	0,431	0,703	0,463	0,837
16	7,409	8,593	7,028	8,381	6,758	8,282	5,698	7,297
17	35,542	43,043	36,552	45,430	51,059	65,378	56,300	75,026
18	2,770	3,392	2,570	3,217	1,903	2,495	1,517	2,097
19	27,279	37,619	27,905	40,674	40,244	58,875	47,142	74,157
20	8,601	10,292	8,053	10,000	8,564	11,096	7,413	10,099
21	13,709	17,389	13,280	17,739	16,138	22,679	16,395	24,775
22	2,464	3,076	2,330	2,972	1,529	2,073	1,278	1,828
23	11,415	14,795	10,541	14,393	13,514	19,132	13,778	20,991
24	15,273	19,104	14,579	19,063	19,045	26,060	19,573	28,452
25	16,146	19,025	16,010	19,468	18,334	22,862	16,776	21,816
26	2,620	3,280	2,475	3,175	1,628	2,200	1,375	1,966
27	4,207	5,179	3,937	5,031	2,987	4,012	2,383	3,416
28	1,100	1,459	1,078	1,417	0,619	0,890	0,482	0,726

Продовження табл.3.13

1	2	3	4	5	6	7	8	9
29	27,602	33,655	27,790	35,011	38,146	49,924	41,152	56,489
30	33,551	40,655	34,458	42,907	47,207	60,572	50,574	67,630
31	14,419	19,135	14,432	20,300	15,193	22,628	13,069	21,205
32	55,923	69,050	58,287	73,098	88,555	113,545	113,985	149,059
33	27,513	39,205	28,289	42,604	41,932	62,084	50,107	79,791
34	2,106	2,692	1,980	2,582	1,322	1,851	1,073	1,591
36	5,452	6,377	5,130	6,164	4,455	5,523	3,651	4,748
37	2,221	3,047	2,036	2,889	1,513	2,334	1,196	2,006
38	20,738	25,480	20,733	26,485	24,983	33,094	22,982	32,071
39	0,528	0,702	0,628	0,775	0,268	0,358	0,276	0,385
40	21,396	27,711	21,781	29,530	25,425	36,118	22,433	34,078
41	13,703	17,322	12,939	17,161	16,441	23,013	16,260	24,254
42	1,834	2,545	1,757	2,507	1,058	1,667	0,907	1,547
43	14,615	23,054	13,124	22,737	13,731	26,184	15,436	33,951

Також модель (3.6) в залежності від факторів X_1 , X_2 та X_4 на основі чотиривимірного перетворення Джонсона для сімейства S_B , є найкращою за ширинами інтервалів передбачення та довірчих інтервалів для більшої кількості даних у порівнянні з нелінійними регресійними моделями, що побудовані на основі як одновимірних перетворень, так і чотиривимірного перетворення Бокса-Кокса.

Кращі показники оцінювання розміру РНР-застосунків з відкритим кодом за нелінійною регресійною моделлю (3.6) на основі чотиривимірного нормалізуючого перетворення Джонсона сім'ї S_B , як було показано раніше, можна в першу чергу пояснити кращою багатовимірною нормалізацією, яка перевірялася за відомим критерієм на основі багатовимірних ексцесу та асиметрії.

3.4 Висновки за розділом 3

У третьому розділі 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.

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 (як лінійними, так і нелінійними), 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 нелінійними), 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 чотиривимірного нормалізуючого перетворення Джонсона сім'ї S_B . Це у порівнянні з використанням одновимірних перетворень дозволяє в багатьох випадках зменшити ширини інтервалу передбачення нелінійної регресії розміру програмних застосунків з відкритим кодом на РНР і тим самим підвищити достовірність оцінювання зазначеного розміру як залежної випадкової величини.

Результати досліджень поточного розділу автором опубліковано в [10, 18, 19, 46-53].

РОЗДІЛ 4. ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ДЛЯ ОЦІНЮВАННЯ РОЗМІРУ ПРОГРАМНИХ ЗАСТОСУНКІВ З ВІДКРИТИМ КОДОМ НА РНР

У четвертому розділі запропоновано інформаційну технологію (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 (кількості класів; середньої кількості методів на клас, середнього значення метрики DIT на клас). Для цього використовувалася система моделювання Scilab. Було розроблено відповідне ПЗ scі-мовою для пакету Scilab 6.0.0. Розроблено методику оцінювання розміру програмних застосунків з відкритим кодом на РНР, яка розрахована на використання відповідного ПЗ.

4.1 Створення інформаційної технології для оцінювання розміру програмних застосунків з відкритим кодом на РНР

Термін “інформаційна технологія” має багато визначень. Так, в [81- 83] наголошують на системі процесів, методів і способів переробки інформації. Інше визначення цього терміну, яке прийняте в даній роботі, наведено в [84, 85], де ІТ визначається як комплекс методів, способів і засобів, що забезпечує збереження, обробку, передачу і відображення інформації. Розглянемо детальніше мету створення, вирішувані задачі та відповідний комплекс ІТ для оцінювання розміру програмних застосунків з відкритим кодом на РНР.

Мета створення інформаційної технології.

Метою створення ІТ для оцінювання розміру програмних застосунків з відкритим кодом на РНР є підвищення достовірності оцінювання розміру РНР-застосунків за рахунок використання відповідних математичних моделей, які наведені в розділах 2 та 3 дисертації.

Задачі, які вирішуються за допомогою інформаційної технології.

Інформаційна технологія для оцінювання розміру програмних застосунків з відкритим кодом на РНР призначена для вирішення наступних задач:

- оцінювання середнього 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);
- оцінювання границь довірчого інтервалу середнього розміру програмних застосунків з відкритим кодом на РНР в залежності від трьох факторів (кількості класів; середньої кількості методів на клас, середнього значення метрики DIT на клас) для заданої довірчої ймовірності;
- моделювання розміру програмних застосунків з відкритим кодом на РНР на початкових етапах розробки як випадкової величини в залежності від трьох факторів (кількості класів; середньої кількості методів на клас, середнього значення метрики DIT на клас) для заданої довірчої ймовірності;
- оцінювання границь інтервалу передбачення нелінійної регресії розміру програмних застосунків з відкритим кодом на РНР в залежності від трьох факторів (кількості класів; середньої кількості методів на клас, середнього значення метрики DIT на клас) для заданої довірчої ймовірності;
- збереження у файлах знайдених середнього 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), границь довірчого інтервалу середнього розміру програмних застосунків з відкритим кодом на РНР в залежності від трьох факторів (кількості класів; середньої кількості методів на клас, середнього значення метрики DIT на клас) для заданої довірчої ймовірності, розміру програмних застосунків з відкритим кодом на РНР на початкових етапах розробки як випадкової величини в залежності від трьох факторів (кількості класів; середньої кількості методів на клас, середнього значення метрики DIT на клас) для заданої довірчої ймовірності, границь інтервалу передбачення нелінійної регресії розміру програмних застосунків з відкритим кодом на РНР в залежності від трьох факторів (кількості класів; середньої кількості методів на клас, середнього значення метрики DIT на клас) для заданої довірчої ймовірності;

- перегляд ОПр (особа, яка приймає рішення) результатів попередніх оцінювань та моделювань.

Математичне забезпечення інформаційної технології.

До математичного забезпечення ІТ для оцінювання розміру програмних застосунків з відкритим кодом на РНР належать математичні моделі для оцінювання розміру програмних застосунків з відкритим кодом на РНР (одновимірне перетворення Джонсона (2.9) та багатовимірне перетворення Джонсона для сімейства S_B , які наведені в 2 розділі; нелінійна регресійна модель для оцінювання розміру програмних застосунків з відкритим кодом на РНР (3.6) в залежності від кількості класів; середньої кількості методів на клас, середнього значення метрики DIT на клас, яка була побудована в 3 розділі; рівняння нижньої та верхньої границь довірчого інтервалу (3.4) та інтервалу передбачення (3.7) нелінійної регресії розміру програмних застосунків з відкритим кодом на РНР в залежності від кількості класів; середньої кількості методів на клас, середнього

значення метрики DIT на клас, які були побудовані в 3 розділі), а також метод виключень [86, 87] для генерації значень випадкової величини із заданим законом розподілу, що використовується для моделювання розміру програмних застосунків з відкритим кодом на РНР.

Метод виключень [86, 87] зазвичай використовують для моделювання значень випадкової величини із заданим законом розподілу, який заданий щільністю ймовірності $f(x)$. Для цього спочатку генерується пара випадкових чисел $x_i \in (a, b)$ та $y_i \in (0, f_{\max})$ з рівномірним законом розподілу, де $f_{\max}$ – це максимальне значення щільності ймовірності $f(x)$ в діапазоні $x_i \in (a, b)$. Якщо $y_i \leq f(x_i)$, то значення y_i береться як поточне значення випадкової величини із законом розподілу, який заданий щільністю ймовірності $f(x)$. Інакше – значення x_i і y_i відкидаються і генерується наступна пара значень.

Інформаційне забезпечення інформаційної технології.

До інформаційного забезпечення ІТ для оцінювання розміру програмних застосунків з відкритим кодом на РНР входять файли даних, що призначені для збереження результатів оцінювань та моделювань розміру програмних застосунків з відкритим кодом на РНР.

Програмне забезпечення інформаційної технології.

Програмним забезпеченням ІТ для оцінювання розміру програмних застосунків з відкритим кодом на РНР є комп'ютерна програма з візуальним графічним інтерфейсом "Оцінювання розміру програмних застосунків з відкритим кодом на РНР". Для запуску та роботи з програмним забезпеченням необхідна операційна система Windows Vista/7/8/10 (32 або 64 біти), система моделювання Scilab версії не нижче за 6.0.0., компілятор C (Visual Studio 2010

або Visual Express 2010) для компіляції зовнішніх модулів С або С++ та використання моделі в Xcos.

Розглянемо детальніше комп'ютерну програму з візуальним інтерфейсом «Оцінювання розміру програмних застосунків з відкритим кодом на РНР», яка була розроблена в рамках даного дослідження як програмний засіб ІТ для оцінювання розміру програмних застосунків з відкритим кодом на РНР з метою програмної реалізації математичного забезпечення цієї ІТ.

В якості мови програмування для розробки програми було обрано Sci-мову [88, 89]. Sci-мова є структурованою мовою програмування високого рівня. Дозволяє працювати з елементарними і великим числом спеціальних функцій, має потужні засоби роботи з матрицями, поліномами, також проводити чисельні обчислення і рішення задач лінійної алгебри, оптимізації і моделювання, потужні статистичні функції, а також засіб для побудови та роботи з графіками. Sci-мова використовується у системі моделювання Scilab. Scilab є програмним пакетом для наукових обчислень, що вільно розповсюджується. Він є певною альтернативою комерційним пакетам моделювання MATLAB/Simulink та MATRIXx/SystemBuild [88]. Одна із відомих версій Scilab – це Scilab 6.0.0 (<http://www.scilab.org>). Scilab має схожу з MATLAB мову програмування – Sci-мову [88, 89].

Переваги Sci-мови – безкоштовне для комерційного використання, достатньо велика кількість вбудованих функцій, не потребує багато обчислювальних ресурсів, працює на більшості ОС. Для розробки програми для оцінювання розміру програмних застосунків з відкритим кодом на РНР недоліків не виявлено.

В програмі, що була розроблена для оцінювання розміру програмних застосунків з відкритим кодом на РНР, використовується динамічне створення інтерфейсних компонентів, яке полягає у тому, що при виконанні програми

можуть створюватися і знищуватися ті чи інші елементи управління (командні кнопки, радіокнопки, мітки та таке інше), а їх властивостям присвоюються відповідні значення. В програмі були використані такі інтерфейсні компоненти: вікно редагування, командні кнопки та радіокнопки. Графічне вікно програми із зазначеними інтерфейсними компонентами наведено на рис.4.1.

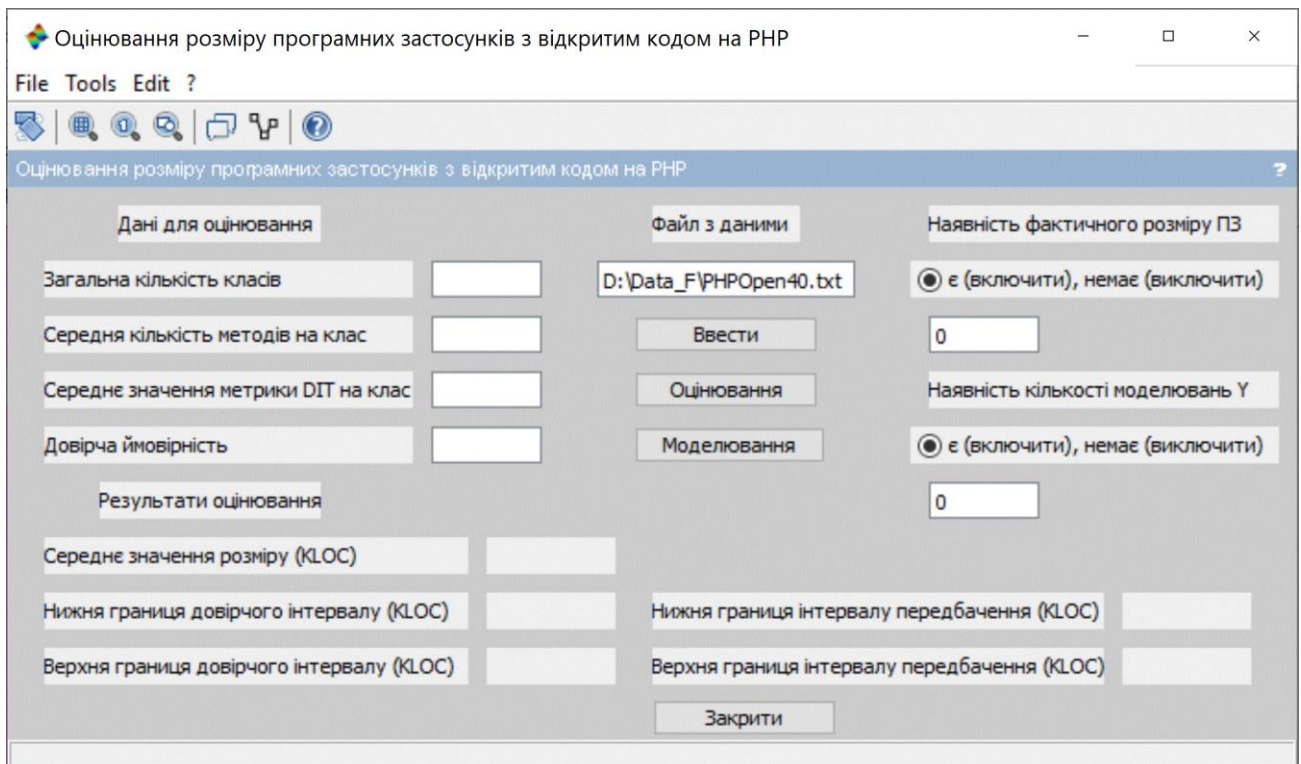


Рисунок 4.1. – Графічне вікно програми із інтерфейсними компонентами

Детальне використання відповідного програмного забезпечення – комп'ютерної програми з візуальним інтерфейсом "Оцінювання розміру програмних застосунків з відкритим кодом на PHP" буде розглянуто у наступному підрозділі.

Технічне забезпечення інформаційної технології.

Технічним забезпеченням ІТ для оцінювання розміру програмних застосунків з відкритим кодом на РНР є таке. Програма має експлуатуватись на обладнанні з параметрами не нижчими за наступні: процесор Pentium IV (або еквівалентний) з інструкціями SSE2, оперативна пам'ять 2 Гб (1 Гб мінімум), простір на жорсткому диску 600 Мб мінімум.

Методичне забезпечення інформаційної технології.

Методичним забезпеченням інформаційної технології переробки інформації для оцінювання розміру програмних застосунків з відкритим кодом на РНР є методика оцінювання розміру програмних застосунків з відкритим кодом на РНР з використанням відповідного програмного забезпечення – комп'ютерної програми з візуальним інтерфейсом "Оцінювання розміру програмних застосунків з відкритим кодом на РНР", розробленої в рамках даної роботи. Вигляд головного вікна програми показаний на рис.4.1.

Впровадження інформаційної технології.

Інформаційну технологію для оцінювання розміру програмних застосунків з відкритим кодом на РНР було впроваджено у виробничі процеси ТОВ «АРТ ПОРТ», ФОП Погіба О.В., Комунальне підприємство «МІСЬКИЙ ІНФОРМАЦІЙНО-ОБЧИСЛЮВАЛЬНИЙ ЦЕНТР» (акти впровадження наведені у додатку А). Впровадження пропонованої ІТ для оцінювання розміру програмних застосунків з відкритим кодом на РНР дозволило підвищити достовірність оцінювання розміру програмних застосунків. Також, математичне забезпечення та інструментарій ІТ для оцінювання розміру програмних застосунків з відкритим кодом на РНР було впроваджено в навчальний процес кафедр програмного забезпечення автоматизованих систем та інформаційних управляючих систем та технологій Національного університету

кораблебудування імені адмірала Макарова шляхом використання в робочій навчальній програмі з дисциплін «Емпіричні методи програмної інженерії» та «Інтелектуальний аналіз даних» (акт впровадження від 10.06.2021, Додаток Б). Акти впровадження результатів дисертаційної роботи наведені в додатках А і Б.

4.2 Методика оцінювання розміру програмних застосунків з відкритим кодом на РНР

Методика оцінювання розміру програмних застосунків з відкритим кодом на РНР розрахована на використання відповідного програмного забезпечення – комп'ютерної програми з графічним інтерфейсом «Оцінювання розміру програмних застосунків з відкритим кодом на РНР», розробленої в рамках даної роботи.

Після запуску програми відкривається її графічне вікно «Оцінювання розміру програмних застосунків з відкритим кодом на РНР», що наведено на рис.4.1. Далі у цьому графічному вікні потрібно ввести дані для оцінювання. При цьому вводять такі дані: загальну кількість класів; середню кількість методів на клас, середнє значення метрики DIT на клас та довірчу ймовірність, з якою оцінюють довірчі інтервали (3.4) та інтервали передбачення (3.7) нелінійної регресії розміру програмних застосунків з відкритим кодом на РНР в залежності від загальної кількості класів; середньої кількості методів на клас, середнього значення метрики DIT на клас. Графічне вікно «Оцінювання розміру програмних застосунків з відкритим кодом на РНР» після введення значень загальної кількості класів; середньої кількості методів на клас, середнього значення метрики DIT на клас та довірчої ймовірності у вікна редагування наведено на рис.4.2. Окрім введення значень загальної кількості класів; середньої кількості

методів на клас, середнього значення метрики DIT на клас та довірчої ймовірності у вікна редагування ці значення можна також ввести через текстовий файл, ім'я якого зазначається у відповідному вікні редагування. При цьому значення загальної кількості класів; середньої кількості методів на клас, середнього значення метрики DIT на клас та довірчої ймовірності у текстовому файлі потрібно записати наступним чином:

2075 4.8092 1.25 0.95

Рисунок 4.2. – Графічне вікно програми після введення даних для оцінювання розміру програмних застосунків з відкритим кодом на PHP

Для введення даних потрібно натиснути командну кнопку «Ввести». Далі потрібно зазначити варіант оцінювання розміру програмних застосунків з відкритим кодом на PHP.

1-й варіант – оцінювання здійснюється без інформації про фактичний розмір ПЗ, при цьому відносна похибка MRE оцінювання не розраховується, можна оцінити похибку оцінки середнього розміру за границями довірчого інтервалу. Для цього варіанту оцінювання потрібно виключити верхню радіокнопку. Графічне вікно «Оцінювання розміру програмних застосунків з відкритим кодом на РНР» після виключення верхньої радіокнопки наведено на рис.4.3.

Рисунок 4.3. – Графічне вікно програми після виключення верхньої радіокнопки

Далі потрібно натиснути командну кнопку «Оцінювання». Після чого у графічному вікні програми «Оцінювання розміру програмних застосунків з відкритим кодом на РНР» виводяться результати оцінювання:

- Середнє значення розміру (у тисячах строк коду, KLOC) або значення $\hat{Y}$ за нелінійним рівнянням регресії (3.2) в залежності від загальної кількості класів; середньої кількості методів на клас, середнього значення метрики DIT на клас;

- Нижня границя довірчого інтервалу нелінійної регресії за рівнянням (3.4) або нижня границя довірчого інтервалу середнього значення розміру (у тисячах строк коду, KLOC) в залежності від загальної кількості класів; середньої кількості методів на клас, середнього значення метрики DIT на клас;

- Верхня границя довірчого інтервалу нелінійної регресії за рівнянням (3.4) або верхня границя довірчого інтервалу середнього значення розміру (у тисячах строк коду, KLOC) в залежності від загальної кількості класів; середньої кількості методів на клас, середнього значення метрики DIT на клас;

- Нижня границя інтервалу передбачення нелінійної регресії за рівнянням (3.7) або нижня границя інтервалу передбачення значення розміру (у тисячах строк коду, KLOC) в залежності від загальної кількості класів; середньої кількості методів на клас, середнього значення метрики DIT на клас;

- Верхня границя інтервалу передбачення нелінійної регресії за рівнянням (3.7) або верхня границя інтервалу передбачення значення розміру (у тисячах строк коду, KLOC) в залежності від загальної кількості класів; середньої кількості методів на клас, середнього значення метрики DIT на клас.

Графічне вікно «Оцінювання розміру програмних застосунків з відкритим кодом на PHP» після натискання командної кнопки «Оцінювання» за 1-им варіантом наведено на рис.4.4.

Оцінювання розміру програмних застосунків з відкритим кодом на PHP

File Tools Edit ?

Оцінювання розміру програмних застосунків з відкритим кодом на PHP

Дані для оцінювання		Файл з даними	Наявність фактичного розміру ПЗ	
Загальна кількість класів	2075	D:\Data_F\PHPOpen40.txt	☐ є (включити), немає (виключити)	
Середня кількість методів на клас	4.8092	Ввести	0	
Середнє значення метрики DIT на клас	1.25	Оцінювання	Наявність кількості моделювань Y	
Довірча ймовірність	0.95	Моделювання	☒ є (включити), немає (виключити)	
Результати оцінювання		0		
Середнє значення розміру (KLOC)	=175.761			
Нижня границя довірчого інтервалу (KLOC)	=163.733	Нижня границя інтервалу передбачення (KLOC)	=145.090	
Верхня границя довірчого інтервалу (KLOC)	=186.529	Верхня границя інтервалу передбачення (KLOC)	=199.384	
Закрити				

Рисунок 4.4. – Графічне вікно програми після натискання командної кнопки «Оцінювання» за 1-им варіантом

2-й варіант – оцінювання здійснюється з інформацією про фактичний розмір ПЗ, при цьому відносна похибка MRE оцінювання розраховується, також можна оцінити похибку оцінки середнього розміру за границями довірчого інтервалу. Для цього варіанту оцінювання потрібно включити верхню радіокнопку.

Графічне вікно «Оцінювання розміру програмних застосунків з відкритим кодом на PHP» після включення верхньої радіокнопки та введення фактичного розміру програмного застосунку (у тисячах строк коду, KLOC) наведено на рис.4.5.

Оцінювання розміру програмних застосунків з відкритим кодом на PHP

File Tools Edit ?

Оцінювання розміру програмних застосунків з відкритим кодом на PHP

Дані для оцінювання		Файл з даними	Наявність фактичного розміру ПЗ	
Загальна кількість класів	2075	D:\Data_F\PHPOpen40.txt	☒ є (включити), немає (виключити)	
Середня кількість методів на клас	4.8092	Ввести		174.927
Середнє значення метрики DIT на клас	1.25	Оцінювання		Наявність кількості моделювань Y
Довірча ймовірність	0.95	Моделювання	☒ є (включити), немає (виключити)	
Результати оцінювання				0
Середнє значення розміру (KLOC)	=175.761			
Нижня границя довірчого інтервалу (KLOC)	=163.733	Нижня границя інтервалу передбачення (KLOC)		=145.090
Верхня границя довірчого інтервалу (KLOC)	=186.529	Верхня границя інтервалу передбачення (KLOC)		=199.384
Закрити				

Рисунок 4.5. – Графічне вікно програми після включення верхньої радіокнопки та введення фактичного розміру програмного застосунку

Далі потрібно натиснути командну кнопку «Оцінювання». Після чого у графічному вікні «Оцінювання розміру програмних застосунків з відкритим кодом на PHP» виводяться результати оцінювання:

- Середнє значення розміру (у тисячах строк коду, KLOC) або значення $\hat{Y}$ за нелінійним рівнянням регресії (3.2) в залежності від загальної кількості класів; середньої кількості методів на клас, середнього значення метрики DIT на клас;
- Нижня границя довірчого інтервалу нелінійної регресії за рівнянням (3.4) або нижня границя довірчого інтервалу середнього значення розміру (у

тисячах строк коду, KLOC) в залежності від загальної кількості класів; середньої кількості методів на клас, середнього значення метрики DIT на клас;

- Верхня границя довірчого інтервалу нелінійної регресії за рівнянням (3.4) або верхня границя довірчого інтервалу середнього значення розміру (у тисячах строк коду, KLOC) в залежності від загальної кількості класів; середньої кількості методів на клас, середнього значення метрики DIT на клас;

- Нижня границя інтервалу передбачення нелінійної регресії за рівнянням (3.7) або нижня границя інтервалу передбачення значення розміру (у тисячах строк коду, KLOC) в залежності від загальної кількості класів; середньої кількості методів на клас, середнього значення метрики DIT на клас;

- Верхня границя інтервалу передбачення нелінійної регресії за рівнянням (3.7) або верхня границя інтервалу передбачення значення розміру (у тисячах строк коду, KLOC) в залежності від загальної кількості класів; середньої кількості методів на клас, середнього значення метрики DIT на клас;

- Відносна похибка MRE оцінювання розміру програмних застосунків з відкритим кодом на РНР.

Графічне вікно програми після натискання командної кнопки «Оцінювання» за 2-им варіантом наведено на рис.4.6.

Після проведення оцінювання розміру програмних застосунків з відкритим кодом на РНР можна виконати моделювання розміру Y як випадкової величини за нелінійною регресійною моделлю (3.6) в залежності від загальної кількості класів; середньої кількості методів на клас, середнього значення метрики DIT на клас. Для цього у графічному вікні «Оцінювання розміру програмних застосунків з відкритим кодом на РНР» потрібно включити другу радіокнопку та ввести у відповідне вікно редагування кількість значень випадкової величини Y , моделювання якої потрібно здійснити.

Оцінювання розміру програмних застосунків з відкритим кодом на PHP

File Tools Edit ?

Оцінювання розміру програмних застосунків з відкритим кодом на PHP

Дані для оцінювання		Файл з даними	Наявність фактичного розміру ПЗ	
Загальна кількість класів	2075	D:\Data_F\PHPOpen40.txt	☒ є (включити), немає (виключити)	
Середня кількість методів на клас	4.8092	Ввести	174.927	
Середнє значення метрики DIT на клас	1.25	Оцінювання	Наявність кількості моделювань Y	
Довірча ймовірність	0.95	Моделювання	☒ є (включити), немає (виключити)	
Результати оцінювання		0		
Середнє значення розміру (KLOC)	= 175.761	Відносна похибка MRE=0.0048		
Нижня границя довірчого інтервалу (KLOC)	= 163.733	Нижня границя інтервалу передбачення (KLOC)	= 145.090	
Верхня границя довірчого інтервалу (KLOC)	= 186.529	Верхня границя інтервалу передбачення (KLOC)	= 199.384	
Закрити				

Рисунок 4.6. – Графічне вікно програми після натискання командної кнопки «Оцінювання» за 2-им варіантом

Графічне вікно програми після включення другої радіокнопки та введення кількості значень випадкової величини Y , моделювання якої потрібно здійснити, наведено на рис.4.7.

Далі у графічному вікні «Оцінювання розміру програмних застосунків з відкритим кодом на PHP» потрібно натиснути на командну кнопку «Моделювання». Після чого у графічному вікні «Оцінювання розміру програмних застосунків з відкритим кодом на PHP» буде виведено повідомлення про кількість значень випадкової величини Y , що змодельовані.

Оцінювання розміру програмних застосунків з відкритим кодом на PHP

File Tools Edit ?

Оцінювання розміру програмних застосунків з відкритим кодом на PHP

Дані для оцінювання		Файл з даними	Наявність фактичного розміру ПЗ	
Загальна кількість класів	2075	D:\Data_F\PHPOpen40.txt	☐ є (включити), немає (виключити)	
Середня кількість методів на клас	4.8092	Ввести	0	
Середнє значення метрики DIT на клас	1.25	Оцінювання	Наявність кількості моделювань Y	
Довірча ймовірність	0.95	Моделювання	☒ є (включити), немає (виключити)	
Результати оцінювання		5		
Середнє значення розміру (KLOC)	= 175.761			
Нижня границя довірчого інтервалу (KLOC)	= 163.733	Нижня границя інтервалу передбачення (KLOC)	= 145.090	
Верхня границя довірчого інтервалу (KLOC)	= 186.529	Верхня границя інтервалу передбачення (KLOC)	= 199.384	
Закрити				

Рисунок 4.7. – Графічне вікно програми після включення другої радіокнопки та введення кількості значень випадкової величини Y при її моделюванні

Графічне вікно «Оцінювання розміру програмних застосунків з відкритим кодом на PHP» після моделювання випадкової величини Y наведено на рис.4.8.

The screenshot shows a software window titled "Оцінювання розміру програмних застосунків з відкритим кодом на PHP". The window contains several sections for data entry and results display.

Дані для оцінювання (Evaluation Data):

Загальна кількість класів	2075
Середня кількість методів на клас	4.8092
Середнє значення метрики DIT на клас	1.25
Довірча ймовірність	0.95

Файл з даними (Data File): D:\Data_F\PHPOpen40.txt

Наявність фактичного розміру ПЗ (Actual size of the application): ☐ є (включити), немає (виключити)

Наявність кількості моделювань Y (Number of simulations Y): ☒ є (включити), немає (виключити)

Результати оцінювання (Evaluation Results):

Середнє значення розміру (KLOC)	= 175.761	Моделювань Y виконано 5	
Нижня границя довірчого інтервалу (KLOC)	= 163.733	Нижня границя інтервалу передбачення (KLOC)	= 145.090
Верхня границя довірчого інтервалу (KLOC)	= 186.529	Верхня границя інтервалу передбачення (KLOC)	= 199.384

Buttons: Ввести, Оцінювання, Моделювання (highlighted), Закрити.

Рисунок 4.8. – Графічне вікно програми після моделювання випадкової величини Y

На рис.4.8 представлено ситуацію, коли відповідне моделювання було здійснено після оцінювання за 1-им варіантом, коли перша радіокнопка виключена. Можлива ситуація, коли моделювання випадкової величини Y здійснюється після оцінювання за 2-им варіантом у разі включення першої радіокнопки. Графічне вікно «Оцінювання розміру програмних застосунків з відкритим кодом на PHP» після моделювання випадкової величини Y у разі включення першої радіокнопки наведено на рис.4.9.

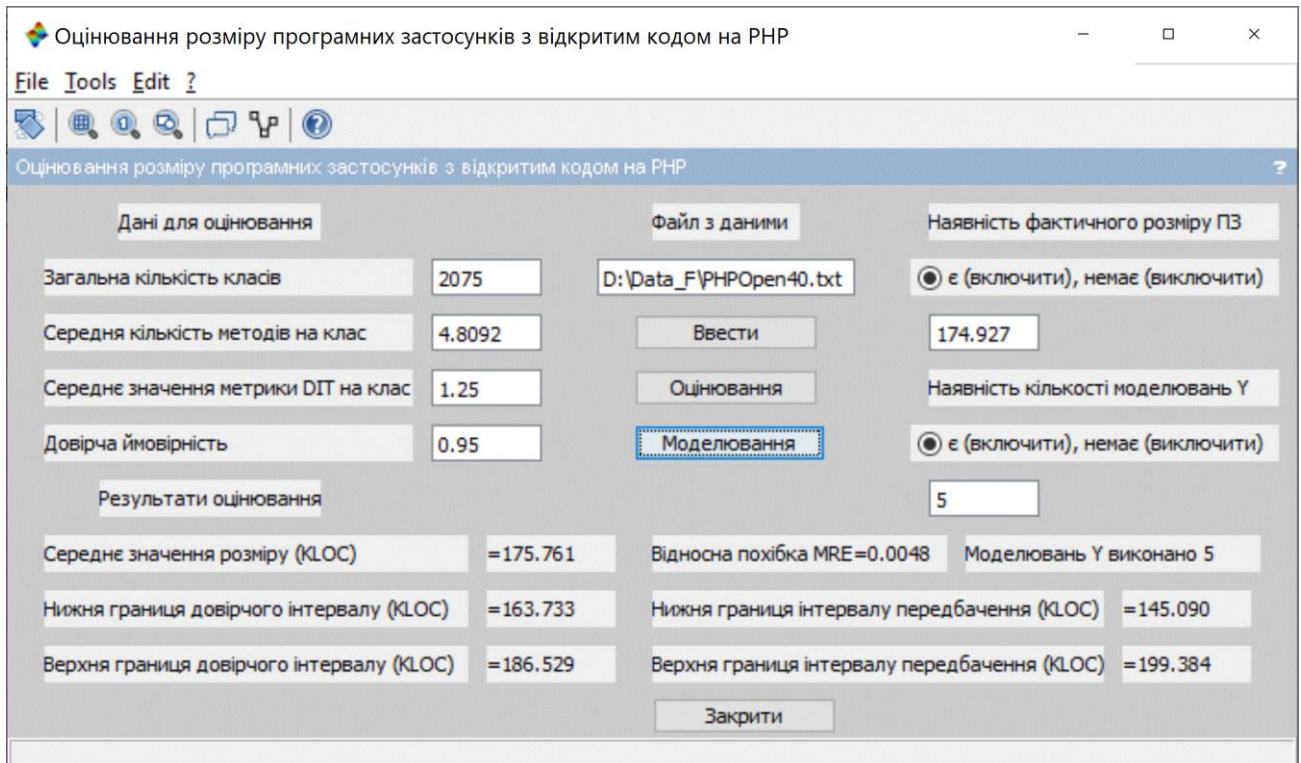


Рисунок 4.9. – Графічне вікно програми після моделювання випадкової величини Y у разі включення першої радіокнопки

Після виконання відповідного оцінювання та моделювання можна здійснити вихід з програми. Для цього у графічному вікні «Оцінювання розміру програмних застосунків з відкритим кодом на PHP» потрібно натиснути на командну кнопку «Закрити». Графічне вікно «Оцінювання розміру програмних застосунків з відкритим кодом на PHP» перед натисканням командної кнопки «Закрити» наведено на рис.4.10.

Після натискання командної кнопки «Закрити» відбувається закриття як самої програми, так і всіх графічних вікон, які були створені під час її роботи.

Оцінювання розміру програмних застосунків з відкритим кодом на PHP

File Tools Edit ?

Оцінювання розміру програмних застосунків з відкритим кодом на PHP

Дані для оцінювання		Файл з даними	Наявність фактичного розміру ПЗ
Загальна кількість класів	2075	D:\Data_F\PHPOpen40.txt	☒ є (включити), немає (виключити)
Середня кількість методів на клас	4.8092	Ввести	174.927
Середнє значення метрики DIT на клас	1.25	Оцінювання	Наявність кількості моделювань Y
Довірча ймовірність	0.95	Модельовання	☒ є (включити), немає (виключити)
Результати оцінювання		5	
Середнє значення розміру (KLOC)	=175.761	Відносна похибка MRE=0.0048	Моделювань Y виконано 5
Нижня границя довірчого інтервалу (KLOC)	=163.733	Нижня границя інтервалу передбачення (KLOC)	=145.090
Верхня границя довірчого інтервалу (KLOC)	=186.529	Верхня границя інтервалу передбачення (KLOC)	=199.384
Закрити			

Рисунок 4.10. – Графічне вікно програми перед натисканням командної кнопки «Закрити»

На рис.4.2-4.10 наведені результати оцінювання розміру PHP-застосунку Symfony-master, дані метрик якого були використані для побудови нелінійної регресійної моделі (3.6) та відповідних рівнянь (3.2), (3.4) і (3.7). Тому ми отримали при оцінюванні розміру цього PHP-застосунку малу відносну похибку 0,0048 або 0,48%. Нагадаємо, що актуальне значення розміру Y PHP-застосунку Symfony-master дорівнює 174,927 KLOC. Для цього ж застосунку за значеннями факторів $X_1=2075$, $X_2=4,809$ та $X_4=1,25$ для моделі (3.6), що побудована в залежності від факторів X_1 , X_2 та X_4 , маємо такий інтервал передбачення залежної змінної Y : [145,090; 199,384], ширина якого на 36% менша за ширину відповідного інтервалу для моделі, що побудована в залежності від факторів X_1 ,

X_2 та X_3 і є такою [137,430; 221,703]. Подібний результат ми маємо і для довірчого інтервалу.

При застосуванні інструментарію ІТ для оцінюванні розміру РНР-застосунків, дані метрик яких не були використані для побудови нелінійної регресійної моделі (3.6) та відповідних рівнянь (3.2), (3.4) і (3.7) значення відносної похибки значно гірші але є прийнятними. Так, результати оцінювання розміру РНР-застосунку серверної частини складського обліку порталу «Rikku Нуре», які наведено на рис.4.11 є такими.

Оцінювання розміру програмних застосунків з відкритим кодом на РНР

File Tools Edit ?

Оцінювання розміру програмних застосунків з відкритим кодом на РНР

Дані для оцінювання		Файл з даними	Наявність фактичного розміру ПЗ	
Загальна кількість класів	553	D:\Data_F\PHPOpen40.txt	☒ є (включити), немає (виключити)	
Середня кількість методів на клас	4.08	Ввести	32.726	
Середнє значення метрики DIT на клас	1.81	Оцінювання	Наявність кількості моделювань Y	
Довірча ймовірність	0.95	Моделювання	☐ є (включити), немає (виключити)	
Результати оцінювання			0	
Середнє значення розміру (KLOC)	=40.843	Відносна похибка MRE=0.2480		
Нижня границя довірчого інтервалу (KLOC)	=34.671	Нижня границя інтервалу передбачення (KLOC)	=25.167	
Верхня границя довірчого інтервалу (KLOC)	=47.861	Верхня границя інтервалу передбачення (KLOC)	=63.445	
Закрити				

Рисунок 4.11. – Графічне вікно програми з результатами оцінювання розміру РНР-застосунку серверної частини складського обліку порталу «Rikku Нуре»

При оцінюванні розміру цього РНР-застосунку відносна похибка складає 0,2480 або 24,80% і є прийнятною (менша за 25%). Актуальне значення розміру

У PHP-застосунку серверної частини складського обліку порталу «Rikky Нуре» дорівнює 32,726 KLOC. Для цього ж застосунку за значеннями факторів $X_1=553$, $X_2=4,08$ та $X_4=1,81$ для моделі (3.6), що побудована в залежності від факторів X_1 , X_2 та X_4 , маємо такий інтервал передбачення залежної змінної Y : [25,167; 63,445], ширина якого на 2,9% більша за ширину відповідного інтервалу для моделі, що побудована в залежності від факторів X_1 , X_2 та X_3 ($X_3=4,33$) Подібний результат ми маємо і для довірчого інтервалу.

4.3 Висновки за розділом 4

В даному розділі було створено інформаційну технологію для оцінювання розміру програмних застосунків з відкритим кодом на PHP на ранніх стадіях розробки за трьома метриками діаграми класів із застосуванням математичних моделей, побудованих з використанням чотиривимірного нормалізуючого перетворення Джонсона сімейства S_B .

1) Створено інформаційну технологію для оцінювання розміру програмних застосунків з відкритим кодом на PHP за трьома метриками діаграми класів (загальна кількість класів; середня кількість методів на клас, середнє значення метрики DIT на клас), яка дозволяє підвищити достовірність оцінювання розміру зазначених програмних застосунків на ранніх стадіях розробки за метриками діаграми класів. Основою побудови пропонованої інформаційної технології послужили математичні моделі, розроблені автором в даному дослідженні.

2) Розроблено методику оцінювання розміру програмних застосунків з відкритим кодом на PHP, яка розрахована на використання відповідного інструментарію IT – програмного забезпечення, що було створено. Застосування

цього ПЗ до низки програмних застосунків з кодом на РНР показало його працездатність: оцінки розміру більшості РНР-застосунків мали відносну похибку менше ніж 25%.

3) Запропоновано в майбутніх математичних моделях, що будуть створюватися для оцінювання розміру РНР-застосунків, враховувати ще технологію розробки цих застосунків.

Результати досліджень поточного розділу автором опубліковано в [10, 18, 19, 45-52].

ВИСНОВКИ

В дисертаційній роботі вирішено важливе науково-практичне завдання підвищення достовірності оцінювання розміру програмних застосунків з відкритим кодом на РНР на ранніх стадіях розробки за метриками діаграми класів за рахунок побудови відповідних математичних моделей у вигляді нелінійних регресійних моделей, що побудовані з використанням чотиривимірного нормалізуючого перетворення Джонсона сімейства S_B , та створено на їх основі інструментарій інформаційної технології для оцінювання розміру програмних застосунків з відкритим кодом на РНР на ранніх стадіях розробки за метриками діаграми класів.

В процесі проведення дисертаційних досліджень одержані наступні результати.

1) Проаналізовано існуючі методи і моделі для оцінювання розміру програмних застосунків, у тому числі з відкритим кодом на РНР. З'ясовано, що існуючі відповідні методи базуються в основному на математичних моделях, які, як правило є лінійними і не містять у своєму складі випадкових складових. Це у свою чергу не дозволяє враховувати ту обставину, що розмір програмних застосунків є негаусівською випадковою величиною, та веде до низької достовірності оцінювання розміру програмних застосунків з відкритим кодом на РНР на початкових етапах.

2) Існуючі регресійні моделі для оцінювання розміру програмних застосунків з відкритим кодом на РНР, в основному є лінійними. Використання моделей лінійної регресії потребує виконання певних припущень, одним з яких є нормальність розподілу помилок. Підтверджено, що це припущення не виконується для лінійної регресійної моделі, яка побудова за даними з метрик

програмних застосунків з відкритим кодом на РНР, що були зібрані. Це призводить до необхідності розробки нелінійних регресійних моделей, в тому числі для оцінювання розміру програмних застосунків з відкритим кодом на РНР розташованих на сайті GitHub (<https://github.com>). Ці дані з метрик програмних застосунків були отримані за допомогою інструменту PhpMetrics (<https://phpmetrics.org/>) та мають негаусівський розподіл.

3) Нормалізуючі перетворення часто є добрим способом побудови моделей, довірчих інтервалів та інтервалів передбачення нелінійних регресій. Але використання одновимірних нормалізуючих перетворень не дозволяє врахувати наявність кореляції між змінними при нормалізації даних, що як правило призводить до поганих результатів нормалізації і як наслідок – застосування одновимірних нормалізуючих перетворень для побудови нелінійних регресійних моделей не завжди призводить до задовільних результатів оцінювання, насамперед за такими стандартними показниками, як удосконалювати регресійні моделі для оцінювання розміру програмних застосунків з відкритим кодом на РНР доцільно саме на основі багатовимірних нормалізуючих перетворень із врахуванням наявності викидів у даних.

4) Здійснено вибір перетворення для нормалізації чотиривимірних даних з метрик РНР-застосунків. У якості нормалізуючого перетворення було обрано чотиривимірне нормалізуюче перетворення Джонсона сім'ї S_B , за допомогою якого краще здійснюється нормалізація чотиривимірних даних з метрик РНР-застосунків у порівнянні з одновимірними перетвореннями у вигляді десяткового логарифму, Джонсона сім'ї S_B і Бокса-Кокса та чотиривимірним перетворенням Бокса-Кокса.

5) Удосконалено трьох-факторні рівняння нелінійної регресії та нелінійна регресійна модель 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 та розміру програмних застосунків з відкритим кодом на РНР саме як негаусівської випадкової величини у порівнянні з використанням одновимірних нормалізуючих перетворень та чотиривимірного нормалізуючого перетворення Бокса-Кокса. Рівняння та модель, що побудовано, в порівнянні з іншими регресійними рівняннями та моделями (як лінійними, так і нелінійними), мають менше значення середньої величини відносної похибки, менші ширини довірчого інтервалу та інтервалу передбачення нелінійної регресії.

6) Отримало подальший розвиток рівняння для визначення границь довірчого інтервалу та інтервалу передбачення трьох-факторної 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 та інтервалу передбачення нелінійної регресії розміру програмних застосунків з відкритим кодом на РНР і тим самим підвищити достовірність оцінювання вибіркового середнього залежної змінної нелінійної регресії та зазначеного розміру як залежної випадкової величини.

7) Створено 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 методику оцінювання розміру програмних застосунків з відкритим кодом на РНР, яка розрахована на використання зазначеного ПЗ.

8) Результати досліджень впроваджено в ТОВ «АРТ ПОРТ», Комунальне підприємство «МІСЬКИЙ ІНФОРМАЦІЙНО-ОБЧИСЛЮВАЛЬНИЙ ЦЕНТР», ФОП Погіба О.Р., що підтверджується відповідними актами, які наведені у додатку А. Також результати дослідження впроваджено у навчальний процес кафедр програмного забезпечення автоматизованих систем та інформаційних управляючих систем та технологій Національного університету кораблебудування імені адмірала Макарова при викладанні дисциплін «Емпіричні методи програмної інженерії» та «Інтелектуальний аналіз даних».

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. TIOBE Index for June 2021. June Headline: Python has never been so close to position #1 before. URL: <https://www.tiobe.com/tiobe-index/> (дата звернення: 29.06.2021)
2. Рейтинг мов програмування 2021: частка Python зменшується, а TypeScript обійшов C++. URL: <https://dou.ua/lenta/articles/language-rating-jan-2019/> (дата звернення: 29.06.2021)
3. Boehm, B.W. Software Cost Estimation with COCOMO II / B. W. Boehm, C. Abts, A. W. Brown, S. Chulani, B. K. Clark, E. Horowitz, R. Madachy, D. J. Reifer, B. Steece. – Upper Saddle River, NJ: Prentice Hall PTR, 2000. – 544 p.
4. Yahya, M.A. Effects of software process maturity on COCOMO II's effort estimation from CMMI perspective / M. A. Yahya, R. Ahmad, S. P. Lee // Proceedings from RIVFCCT'08: IEEE International Conference on Research, Innovation and Vision for the Future in Computing and Communication Technologies, Ho Chi Minh City, Vietnam, 2008. – P. 255-262. DOI: <https://doi.org/10.1109/RIVF.2008.4586364>
5. Tan, H.B.K. Estimating LOC for information systems from their conceptual data models / H. B. K. Tan, Y. Zhao, H. Zhang // Software Engineering : the 28th International Conference (ICSE '06), Shanghai, China, May 20-28, 2006 : proceedings. – P. 321-330. DOI: <https://doi.org/10.1145/1134285.1134331>
6. Tan, H.B.K. Conceptual data model-based software size estimation for information systems / H. B. K. Tan, Y. Zhao, H. Zhang // Transactions on Software Engineering and Methodology. – 2009. – Vol. 19. – Issue 2. – October 2009. – Article No. 4. DOI: <https://doi.org/10.1145/1571629.1571630>
7. Prykhodko, S.B. Constructing the non-linear regression equation to estimate the software size of open source PHP-based information systems /

S. B. Prykhodko, N. V. Prykhodko, T. G. Smykodub, A. V. Spinov // Проблемы информационных технологий. – 2018. – № 1 (023). – С. 118-125. – ISSN 1998-7005

8. Приходько, Н.В. Нелінійне регресійне рівняння для оцінювання розміру програмного забезпечення інформаційних систем з відкритим кодом на PHP / Н. В. Приходько, С. Б. Приходько // Вісник Східноукраїнського національного університету імені Володимира Даля : наук. журнал / відп. ред. д.т.н., доцент Скарга-Бандурова І.С. ; СХУ ім. В. Даля. – Сєверодонецьк : СХУ ім. В. Даля, 2018. – № 6 (247). – С. 135-139.

9. Prykhodko, S. Estimating the Software Size of Open-Source PHP-Based Systems Using Non-Linear Regression Analysis / Sergiy Prykhodko, Natalia Prykhodko, Lidiia Makarova // Proceedings of International Conference “Advanced Computer Information Technologies” (ACIT-2018). CEUR Workshop Proceedings. – 2019. – Vol.2300. – Ceske Budejovice, CZECH REPUBLIC. CEUR-WS.org – P. 199-202. ISSN 1613-0073. <http://ceur-ws.org/Vol-2300/Paper48.pdf>

10. 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>

11. Prykhodko, N.V. Constructing the non-linear regression models on the basis of multivariate normalizing transformations / N. V. Prykhodko, S. B. Prykhodko // Electronic modeling. – 2018. – Т.40. – № 6. – P. 99-108. – ISSN 0204–3572. – DOI: <https://doi.org/10.15407/emodel.40.06.101>

12. Bates, D.M. Nonlinear Regression Analysis and Its Applications / D. M. Bates, D. G. Watts. – New York: John Wiley & Sons, 1988. – 384 p. DOI: <https://doi.org/10.1002/9780470316757>

13. Seber, G.A.F. Nonlinear Regression / G. A. F. Seber, C. J. Wild. – New York: John Wiley & Sons, 1989. – 768 p. DOI: <https://doi.org/10.1002/0471725315>
14. Ryan, T.P. Modern regression methods / T. P. Ryan. – New York: John Wiley & Sons, 1997. – 529 p. DOI: 10.1002/9780470382806
15. Drapper, N.R. Applied regression analysis / N. R. Drapper, H. Smith. – New York: John Wiley & Sons, 1998. – 736 p.
16. Johnson, R.A. Applied Multivariate Statistical Analysis / R. A. Johnson, D. W. Wichern. – Pearson Prentice Hall, 2007. – 800 p.
17. Заболотній, С.В. Поліноміальні адаптивні процедури регресійного аналізу із використанням моделей негаусових помилок на основі статистик вищих порядків / С. В. Заболотній, О. М. Ткаченко // «Computational Intelligence» (Results, Problems and Perspectives): Proceedings of the International Conference, May 16-18, 2017, Kyiv-Cherkasy, Ukraine; Vitaliy Ye. Snytyuk (Editor). – К.: VPC “Kyiv University”, 2017. – С. 113-114.
18. 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>
19. Приходько, С.Б. Оцінювання розміру РНР-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)
20. Dillibabu, R. Cost estimation of a software product using COCOMO II.2000 model – a case study / R. Dillibabu, K Krishnaiah // International Journal of

Project Management, – 2005. – Vol. 23, Issue 4, May. – P. 297-307. DOI: <https://doi.org/10.1016/j.ijproman.2004.11.003>

21. Valerdi, R. COSYSMO: A constructive systems engineering cost model coming of age / R. Valerdi, B. W. Boehm, D. J. Reifer // in Proc. INCOSE Int. Symp., Vol. 13, No. 1. Hoboken, NJ, USA: Wiley, 2003. – P. 70–82.

22. Putnam, L.H. A general empirical solution to the macro software sizing and estimating problem / L. H. Putnam // IEEE Trans. Softw. Eng., Vol. SE-4, No. 4, Jul. 1978. – P. 345–361.

23. Singh, Y. ANN model for predicting software function point metric / Y. Singh, P. K. Bhatia, O. Sangwan // ACM SIGSOFT Softw. Eng. Notes, Vol. 34, No. 1, Jan. 2009. – P. 1–4.

24. de Barcelos Tronto, I.F. An investigation of artificial neural networks based prediction systems in software project management / I. F. De Barcelos Tronto, J. D. S. da Silva, N. Sant’Anna // J. Syst. Softw., Vol. 81, No. 3, Mar. 2008. – P. 356-367.

25. Kumar, K.V. Software development cost estimation using wavelet neural networks / K. V. Kumar, V. Ravi, M. Carr, N. R. Kiran // J. Syst. Softw. – 2008. – Vol. 81. – No. 11. – P. 1853-1867.

26. Pospieszny, P. An effective approach for software project effort and duration estimation with machine learning algorithms / P. Pospieszny, B. Czarnacka-Chrobot, A. Kobylinski // J. Syst. Softw. – 2018. – Vol. 137. – No. 3. – P. 184-196. DOI: <https://doi.org/10.1016/j.jss.2017.11.066>

27. Kaczmarek, J. Size and effort estimation for applications written in Java / J. Kaczmarek, M. Kucharski // Information and Software Technology. – 2004. – Vol. 46 (9). – P. 589-601. DOI: <https://doi.org/10.1016/j.infsof.2003.11.001>

28. Lind, K. CompSize: Automated Size Estimation of Embedded Software Components / K. Lind, R. Heldal, T. Harutyunyan, T. Heimdahl // Proceedings from Joint Conference of the 21st International Workshop on Software Measurement and the

6th International Conference on Software Process and Product Measurement, Nara, Japan, 2011. – P. 86-95. DOI: <https://doi.org/10.1109/IWSM-MENSURA.2011.49>.

29. Zifen, Y. An improved software size estimation method based on object-oriented approach / Y. Zifen // Proceedings from EEESYM'12: IEEE Symposium on Electrical & Electronics Engineering, Kuala Lumpur, Malaysia, 2012. – P. 615-617. DOI: <https://doi.org/10.1109/EEESym.2012.6258733>.

30. Kiewkanya, M. Constructing C++ software size estimation model from class diagram / M. Kiewkanya, S. Surak // Computer Science and Software Engineering : 13th International Joint Conference, Khon Kaen, Thailand, July 13-15, 2016 : proceedings. – P. 1-6. DOI: <https://doi.org/10.1109/JCSSE.2016.7748880>.

31. Cheng, Z. Esse: An early software size estimation method based on autoextracted requirements features / Z. Cheng, T. Shensi, M. Wenkai, Z. Yang, X.Yong, S. Beijun // Proceedings from Internetware'16: the 8th Asia-Pacific Symposium on Internetware, New York, NY, USA, 2016. – P. 112–115.

32. Zaw, T. The Measurement of Software Size based on Generation Model using COSMIC FSM / T. Zaw, Hlaing, S. Z., M. Myint Lwin, K. Ochimizu // Proceedings from ICSEC'19: 23rd International Computer Science and Engineering Conference, Phuket, Thailand, 2019. – P. 373-378. DOI: <https://doi.org/10.1109/ICSEC47112.2019.8974688>.

33. Neyveli, V.R.N. An Approach to Estimate the Size of Web Application Using IFML User Interface Model / V. R. N. Neyveli, S. S. Sivakumar, D. Arunagiri, C. Arumugam, A. M. Veeramani // Proceedings from AICAI'19: Amity International Conference on Artificial Intelligence, Dubai, United Arab Emirates, 2019. – P. 292-295. DOI: <https://doi.org/10.1109/AICAI.2019.8701268>.

34. Zhang, K. Efficiency Improvement of Function Point-Based Software Size Estimation with Deep Learning Model / K. Zhang, X. Wang, J. Ren, C. Liu // IEEE

Access. – 2020. – Vol. 9. – P. 107124-107136. DOI: <https://doi.org/10.1109/ACCESS.2020.2998581>.

35. Daud, M. Improving the Accuracy of Early Software Size Estimation Using Analysis-to-Design Adjustment Factors (ADAFs) / M. Daud, A. A. Malik // IEEE Access. – 2021. – Vol. 9. – P. 81986-81999. DOI: <https://doi.org/10.1109/ACCESS.2021.3085752>.

36. Prykhodko, S. Mathematical Modeling of Non-Gaussian Dependent Random Variables by Nonlinear Regression Models Based on the Multivariate Normalizing Transformations / S. Prykhodko, N. Prykhodko // In: Shkarlet S., Morozov A., Palagin A. (eds) Mathematical Modeling and Simulation of Systems (MODS'2020). MODS 2020. Advances in Intelligent Systems and Computing. 2021. – Vol. 1265. – P. 166-174. – Springer, Cham. DOI: https://doi.org/10.1007/978-3-030-58124-4_16

37. Albrecht, A. Function, Source Lines of Code, and Development Effort Prediction: A Software Science Validation / A. Albrecht, J. E. Gaffney // IEEE Transactions on Software Engineering. – 1983. – Vol. SE-9. – No. 6. – P. 639-648. DOI: <https://doi.org/10.1109/TSE.1983.235271>

38. Trendowicz, A. Software Project Effort Estimation. Foundations and Best Practice Guidelines for Success / A. Trendowicz, R. Jeffery. – Springer International Publishing, 2014. – 493 p. DOI: <https://doi.org/10.1007/978-3-319-03629-8>

39. Function Point CPM IFPUG. Release. 4.1., IFPUG, Westerville, OH, USA, 1999.

40. Symons, C.R. Software Sizing and Estimating: MK II FPA (Function Point Analysis) / C. R. Symons. – Hoboken, NJ, USA: Wiley, 1991. – 218 p.

41. Niessink, F. Predicting maintenance effort with function points / F. Niessink, H. Van Vliet // 1997 Proceedings International Conference on Software Maintenance, 1997. – P. 32–39. DOI: <https://doi.org/10.1109/ICSM.1997.624228>

42. Meli, R. On the applicability of COSMIC-FFP for measuring software throughout its life cycle / R. Meli, A. Abran, V. T. Ho, S. Oligny // Proceedings of the 11th European Software Control and Metrics Conference, 2000. – P. 18–20.

43. Mardia, K.V. Measures of multivariate skewness and kurtosis with applications / K. V. Mardia // Biometrika. – 1970. – Vol. 57. – P. 519–530. DOI: <https://doi.org/10.1093/biomet/57.3.519>

44. Chatterjee, S. Regression analysis by example / S. Chatterjee, B. Price. – New York: John Wiley & Son, 1977. – 228 p.

45. Prykhodko, S. Detecting Outliers in Multivariate Non-Gaussian Data on the basis of Normalizing Transformations / S. Prykhodko, N. Prykhodko, L. Makarova, K. Pugachenko // Electrical and Computer Engineering : the 2017 IEEE First Ukraine Conference (UKRCON) «Celebrating 25 Years of IEEE Ukraine Section», Kyiv, Ukraine, May 29 – June 2, 2017 : proceedings. – P. 846-849. DOI: <https://doi.org/10.1109/UKRCON.2017.8100366>

46. **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.

47. 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>

48. Приходько, С.Б. Оцінювання розміру JAVA та PHP-застосунків з відкритим кодом за багатofакторними 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-prikladni-naukovo-tekhnichni-doslidzhennia/ATSU2020_T1_P36.pdf

49. **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.

50. 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>

51. **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.

52. 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

модельовання 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

53. **Ворона, М.** Порівняння результатів оцінювання розміру РНР-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>

54. Монтгомери, Д.К. Планирование эксперимента и анализ данных: Пер. с англ. – Л.: Судостроение, 1980. – 384 с.

55. Поллард, Дж. Справочник по вычислительным методам статистики / Пер. с англ. В. С. Занадворова; Под ред. и с предисл. Е. М. Четыркина. – М.: Финансы и статистика, 1982. – 344 с.

56. Oligny, S. An empirical assessment of project duration models in software engineering [Text] / S. Oligny, P. Bourque, A. Abran. // In Proc. 8th European Software Control and Metrics Conference ESCOM. – Berlin, 1997.

57. Oligny, S. Exploring the relation between effort and duration in software engineering projects [Text] / S. Oligny, P. Bourque, A. Abran, B. Fournier // In proc. of the World Computer Congress, Aug. 2000. – P. 175-178.

58. Project Estimation Tools That Use ISBSG Data [Електронний ресурс] // International Software Benchmarking Standards Group. – Режим доступу : <https://www.isbsg.org/project-estimation-tools/>

59. Box, G.E.P. An analysis of transformations / G.E.P. Box, D.R. Cox // Journal of the Royal Statistical Society (B). – 1964. – No.2 (26). – P. 211-252.

60. Tukey, J.W. The comparative anatomy of transformations / J. W. Tukey // Annals of Mathematical Statistics. – 1957. – 28. – P. 602-632.

61. Sakia, R.M. The Box-Cox transformation technique: a review / R. M. Sakia // The Statistician. – 1992. – 41. – P. 169-178.

62. Draper, N.R. On distributions and their transformation to normality / N. R. Draper, D. R. Cox // Journal of the Royal Statistical Society (B). – 1969. – No.2 (31). – P.472-476.

63. Johnson, R.A. Applied Multivariate Statistical Analysis / R. A. Johnson, D.W. Wichern. – Pearson Prentice Hall, 2007. – 800 p.

64. Плескунин, В.И. Теоретические основы организации и анализа выборочных данных в эксперименте / В. И. Плескунин, Е. Д. Воронина. – Л.: Изд-во Ленингр. ун-та, 1979. – 232 с.

65. Приходько, С.Б. Методичні вказівки до виконання лабораторних робіт з дисципліни “Обробка експериментальних даних на ЕОМ” / С. Б. Приходько. – Миколаїв: НУК, 2005. – 52 с.

66. Коваленко, І.І. Сучасні методи статистичного аналізу даних [Текст]: [навч. посіб.] / І. І. Коваленко, С. Б. Приходько, Л. О. Латанська; Нац. ун-т кораблебудування ім. адмірала Макарова. – Миколаїв: НУК, 2011. – 189 с. – ISBN 978-966-321-118-3.

67. Методичні вказівки та завдання до виконання лабораторних робіт з дисципліни "Обробка експериментальних даних на комп'ютері" / С. Б. Приходько, Л. М. Макарова, К. С. Пугаченко. – Миколаїв: НУК, 2018. – 76 с.

68. Приходько, С.Б. Аналитическая зависимость для выбора распределения Джонсона семейства S_L / С. Б. Приходько, Л. Н. Макарова // Вестник ХНТУ. – Херсон: ХНТУ, 2012. – № 2 (45). – С. 101-104. – ISSN 1998-7005

69. Приходько, С.Б. Аналитическая зависимость для выбора семейства распределений Джонсона / С. Б. Приходько, Л. М. Макарова, А. С. Приходько // Проблемы информационных технологий. – 2016. – № 2 (020). – С. 105-110. – ISSN 1998-7005

70. Кендалл, М. Теория распределений / М. Кендалл, А. Стюарт // Пер. с англ. / Под ред. А. Н. Колмогорова. – М.: Наука. Гл. ред. физ.-мат. лит., 1966. – 588 с.

71. Stanfield, P.M. Multivariate input modeling with Johnson distributions / Paul M. Stanfield, James R. Wilson, Gary A. Mirka, Naomi F. Glasscock, Jennie P. Psihogios, Joseph R. Davis // Proceedings of the 28th Winter simulation conference WSC'96, 8-11 December 1996, Coronado, CA, USA, ed. S.Andradóttir, K.J.Healy, D.H.Withers, and B.L.Nelson. – IEEE Computer Society Washington, DC, USA, 1996. – P. 1457-1464. – ISBN: 0-7803-3383-7.

72. Wilson, J.R. Modeling dependencies stochastic simulation inputs / James R. Wilson // Proceedings of the 29th Winter simulation conference WSC'97, 7-10 December 1997, Atlanta, GA, USA, ed. S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson. – IEEE Computer Society Washington, DC, USA, 1997. – P. 47-52. – ISBN: 0-7803-4278-X.

73. Приходько, С.Б. Багатовимірні перетворення для нормалізації математичних моделей нелінійних стохастичних диференціальних систем / С. Б. Приходько // Тези доповідей міждерж. науково-методич. конф. “Проблеми математичного моделювання” (27-29 травня 2009 р., м. Дніпродзержинськ). – Дніпродзержинськ: ДДТУ, 2009. – С. 108-110.

74. Приходько, С.Б. Застосування багатовимірних перетворень для побудови математичних моделей нормалізованих сигналів нелінійних стохастичних диференціальних систем / С. Б. Приходько // Четверта науково-практична конф. з міжнародною участю “Математичне та імітаційне

модельовання систем. МОДС '2009". Тези доповідей. – Київ. – 2009. – 22-26 червня 2009 р. – С. 267-270.

75. Приходько, С.Б. Багатовимірні перетворення для нормалізації математичних моделей нелінійних стохастичних диференціальних систем / С. Б. Приходько // Науковий журнал “Математичне моделювання”. – Дніпродзержинськ: ДДТУ, 2009. – № 1 (20) – С. 12-15.

76. Prykhodko, S. Statistical anomaly detection technique based on normalizing transformations for multivariate non-Gaussian data / Sergiy Prykhodko, Natalia Prykhodko, Lidiia Makarova, Kateryna Pugachenko // «Computational Intelligence» (Results, Problems and Perspectives): Proceedings of the International Conference, May 16-18, 2017, Kyiv-Cherkasy, Ukraine; Vitaliy Ye. Snytyuk (Editor). – K.: VPC “Kyiv University”, 2017. – P. 59-60.

77. Prykhodko, S. Application of the Squared Mahalanobis Distance for Detecting Outliers in Multivariate Non-Gaussian Data / Sergiy Prykhodko, Natalia Prykhodko, Lidiia Makarova, Andrii Pukhalevych // Proceedings of 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET), Lviv-Slavske, Ukraine, February 20-24, 2018, P. 962-965. – DOI: <https://doi.org/10.1109/TCSET.2018.8336353>

78. Prykhodko, S. Outlier Detection in Non-Linear Regression Analysis Based on the Normalizing Transformations / Sergiy Prykhodko, Natalia Prykhodko, Lidiia Makarova, Andrii Pukhalevych // Proceedings of the 15th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET), IEEE, Lviv-Slavske, 2020. – P. 407-410. <https://doi.org/10.1109/TCSET49122.2020.235464>

79. Foss, T. A simulation study of the model evaluation criterion MMRE / T. Foss, E. Stensrud, B. Kitchenham, I. Myrtveit // IEEE Transactions on software engineering. – 2003. – 11(29). – P. 985–995.

80. Port, D. Comparative studies of the model evaluation criterions MMRE and PRED in software cost estimation research / D. Port, M. Korte // In: Proceedings of the 2nd ACM-IEEE International Symposium on Empirical Software Engineering and Measurement. – ACM, New York, 2008. – P. 51–60.

81. ДСТУ 2226-93. Автоматизовані системи. Терміни та визначення [Текст]. – Надано чинності 1994-07-01. – Київ : Держстандарт України, 1994. – 92 с.

82. ДСТУ 2481-94. Системи оброблення інформації. Інтелектуальні інформаційні технології. Терміни та визначення [Текст]. – Надано чинності 1995-01-01. – Київ : Держстандарт України, 1994. – 72 с.

83. A Dictionary of Physics (7 ed.) [Text] / Edited by J. Law, R. Rennie. – Oxford University Press, 2015. – 672 p. – ISBN 978-0198714743.

84. Словарь по кибернетике (2-е издание) [Текст] / Под редакцией В. С. Михалевича. – Киев : Главная редакция УСЭ им. М. П. Бажана, 1989. – 751 с. – ISBN 5-88500-008-5.

85. ГОСТ 7.0-99. Система стандартов по информации, библиотечному и издательскому делу. Информационно-библиотечная деятельность, библиография. Термины и определения [Текст]. – Введен 2000-07-01. – М. : Издательство стандартов, 1999. – 23 с.

86. Крайников, А.В. Вероятностные методы в вычислительной технике: Учебное пособие для вузов по спец. ЭВМ [Текст] / А. В. Крайников, Б. А. Курдилов, А. Н. Лебедев [и др.]; Под ред. А. Н. Лебедева и Е. А. Чернявского. – М. : Высшая школа, 1986. – 312 с.

87. Ситнік, В.Ф. Імітаційне моделювання: Навч. посібник [Текст] / В. Ф. Ситнік, Н. С. Орленко. – К.: КНЕУ, 1998. – 232 с.

88. Campbell, S.L. Modeling and Simulation in Scilab/Scicos [Text] / Stephen L. Campbell, Jean-Philippe Chancelier, Ramine Nikoukhah. – Springer, 2005. – 313 p.

89. Алексеев, Е.Р. Scilab: Решение инженерных и математических задач [Текст] / Е. Р. Алексеев, О. В. Чеснокова, Е. А. Рудченко. – М.: ALT Linux; БИНОМ. Лаборатория знаний, 2008. – 269 с.

**ДОДАТОК А. АКТИ ВПРОВАДЖЕННЯ РЕЗУЛЬТАТІВ
ДИСЕРТАЦІЙНОЇ РОБОТИ У РОЗРОБКУ ПРОГРАМНИХ
ЗАСТОСУНКІВ**



тел.: (0512) 67-00-97, e-mail: artemenko_max@artport.pro

<http://artport.pro/>

**Товариство з
обмеженою
відповідальністю
«Арт Порт»**

67832, Одеська обл., Овідіопольський район,
смт Великодолинське, вул. Паркова, б. 28, кв. 7
п/р: UA27 326610 00000 26002053229766
МРУ ПАТ КБ "ПРИВАТБАНК", ЄДРПОУ 40835892

Вих. № АП-1/1004 від 04.06.2021 р.

АКТ

впровадження результатів дисертаційної роботи на здобуття наукового
ступеня доктора філософії (PhD) в галузі знань 12 «Інформаційні технології»
за спеціальністю 122 "Комп'ютерні науки"

Ворони Михайло Владиславовича

**«Математичні моделі та інформаційна технологія для оцінювання
розміру програмних застосунків з відкритим кодом на РНР»**

Ми, що нижче підписалися, директор ТОВ «АРТ ПОРТ» Артьоменко Максим Сергійович, керівник проєкту Клясс Євген Олегович, розробник проєкту Глумаков Олександр Володимирович, цим Актом засвідчуємо, що результати дисертації Ворони Михайло Владиславовича, а саме: інструментарій інформаційної технології для оцінювання розміру програмних застосунків з відкритим кодом на РНР на ранніх стадіях розробки за метриками діаграми класів було впроваджено для створення РНР-застосунків «Регіональна торгово-промислова палата Миколаївської області» та «Приватний кабінет клієнта компанії «АРТ ПОРТ», що розроблялися ТОВ «АРТ ПОРТ».

Впровадження зазначеного вище результату дозволило підвищити достовірність оцінювання розміру вище зазначених РНР-застосунків на ранніх стадіях їх розробки.

Директор ТОВ «АРТ ПОРТ»

Керівник проєкту
Розробник



М.С. Артьоменко

Є.О. Клясс

О.В. Глумаков

ЗАТВЕРДЖУЮ

ФОП Погіба Олексій Романович



«03» червня 2021 р.

АКТ

впровадження результатів дисертаційної роботи на здобуття наукового ступеня
доктора філософії (PhD) в галузі знань 12 «Інформаційні технології»
за спеціальністю 122 "Комп'ютерні науки"
Ворони Михайло Владиславовича
«Математичні моделі та інформаційна технологія для оцінювання розміру
програмних застосунків з відкритим кодом на PHP»

Я, Погіба Олексій Романович, цим Актом засвідчуємо, що результати дисертації Ворони Михайло Владиславовича, а саме: інструментарій інформаційної технології для оцінювання розміру програмних застосунків з відкритим кодом на PHP на ранніх стадіях розробки за метриками діаграми класів було впроваджено для PHP-застосунків: серверної частини та внутрішньої системи складського обліку порталу «Rikky Hure».

Впровадження зазначеного вище результату дозволило підвищити достовірність оцінювання розміру зазначених PHP-застосунків на ранніх стадіях розробки нових модулів та оновлень.



Погіба Олексій Романович



УКРАЇНА
МИКОЛАЇВСЬКА МІСЬКА РАДА
Комунальне підприємство
«Міський інформаційно-обчислювальний центр»

вул. Адміральська, 20, м. Миколаїв, 54027, тел. (0512) 70-93-71
E-mail: miacc@mk.gov.ua Код ЄДРПОУ 31554081

10.06.2021 № 1-1/10-06/2021 На № _____ від _____

АКТ

впровадження результатів дисертаційної роботи на здобуття наукового ступеня доктора
філософії (PhD) в галузі знань 12 «Інформаційні технології»
за спеціальністю 122 "Комп'ютерні науки"

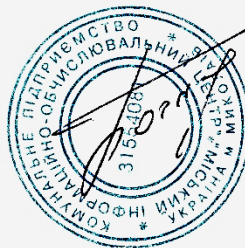
Ворони Михайло Владиславовича

**«Математичні моделі та інформаційна технологія для оцінювання розміру
програмних застосунків з відкритим кодом на РНР»**

Я, що нижче підписався, директор КП «Міський інформаційно-обчислювальний центр» Богаченко Євгеній Валерійович, цим Актом засвідчую, що результати дисертації Ворони Михайло Владиславовича, а саме: інструментарій інформаційної технології для оцінювання розміру програмних застосунків з відкритим кодом на РНР на ранніх стадіях розробки за метриками діаграми класів було впроваджено для створення РНР-застосунків ІС «Система керування світлодіодними табло для зупинок громадського транспорту міста Миколаєва».

Впровадження зазначеного вище результату дозволило підвищити достовірність оцінювання розміру вище зазначених РНР-застосунків на ранніх стадіях їх розробки.

Директор КП «Міський
інформаційно-обчислювальний центр»



Є.В. Богаченко

ДОДАТОК Б. АКТ ВПРОВАДЖЕННЯ РЕЗУЛЬТАТІВ ДИСЕРТАЦІЙНОЇ РОБОТИ В НАВЧАЛЬНИЙ ПРОЦЕС

«ЗАТВЕРДЖУЮ»

Ректор Національного
університету кораблебудування
імені адмірала Макарова

Канд. техн. наук, проф.

Грушляков Є.І.

«10.06» 2021 р.



АКТ ВПРОВАДЖЕННЯ

Результатів роботи Ворони Михайла Владиславовича «Математичні моделі та інформаційна технологія для оцінювання розміру програмних застосунків з відкритим кодом на PHP».

Результати дисертаційної роботи Ворони М.В. використовуються при викладанні студентам Навчально-наукового інституту комп'ютерних наук та управління проектами наступних дисциплін:

- «Емпіричні методи програмної інженерії» спеціальності 121 «Інженерія програмного забезпечення», освітньої програми «Інженерія програмного забезпечення»;
- «Інтелектуальний аналіз даних» спеціальності 122 «Комп'ютерні науки», освітньої програми «Комп'ютерні науки» та спеціальності 126 «Інформаційні системи та технології», освітньої програми «Інформаційні системи та технології».

Заступник директора
Навчально-наукового інституту
комп'ютерних наук та
управління проектами,
к.ф.-м.н., доцент

Латанська Л.О.

ДОДАТОК В. СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА ЗА ТЕМОЮ ДИСЕРТАЦІЇ ТА ВІДОМОСТІ ПРО АПРОБАЦІЮ РЕЗУЛЬТАТІВ ДИСЕРТАЦІЇ

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 РНР-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 РНР-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>. <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.