УДК 004.942

РЕГРЕСІЙНА МОДЕЛЬ ОЦІНЮВАННЯ РОЗМІРУ ВЕБДОДАТКІВ, РЕАЛІЗОВАНИХ МОВОЮ JAVA

Прокопович Л.Б., к.е.н., доц.
Національний університет кораблебудування імені адмірала Макарова
Україна, Миколаїв
ORCID: https://orcid.org/0000-0003-2561-8862

Анотація. *Розглянута проблема побудови регресійної моделі оцінювання розміру вебдодатків, реалізованих мовою Java. За допомогою методів машинного навчання (*k-ближніх сусідів, бінарного дерева рішень*) отримані регресійні моделі оцінювання розміру вебдодатків. Проведено порівняння отриманих результатів, обрана найліпша модель.*
Ключові слова*: регресійна модель; k-ближніх сусідів; дерево рішень; розмір коду програми.*

Вступ. Прогнозування розміру проекту під час розробки програмного забезпечення, особливо на початковій стадії життєвого циклу, є важливим завданням по управлінню проектом. Для оцінювання розміру програмного забезпечення використовують різноманітні метрики. В свою чергу, підходи щодо розробки моделі оцінювання розміру програмного забезпечення також поділяються на кілько груп.

Питання оцінки розміру програмного забезпечення була досліджена в багатьох джерелах, серед яких можна виділити наступні роботи [1-3]. Незважаючи на раніше отримані здобутки, проблема оцінки розміру програмного забезпечення залишається актуальною та потребує подальших досліджень.

Мета роботи: підвищення достовірності оцінювання розміру вебдодатків, реалізованих мовою Java.

Основна частина. В якості первинних даних було використана інформація [4] (див. табл. 1 [4, с. 150]), де Макарова Л.М., Приходько Н.В. та Кудін О.О. розглядали питання побудови нелінійної регресійної моделі для оцінювання розміру вебдодатків, реалізованих мовою Java. В наслідок проведеного ними дослідження, були отримані лінійна та нелінійні регресійні моделі які, за показниками MMRE та PRED(0,25), не досягли прийнятних значень.

Так як моделі побудовані за допомогою методу найменших квадратів не змогли досягти задовільної точності, було прийняте рішення побудувати регресійні моделі для оцінювання розміру вебдодатків, реалізованих мовою Java, за допомогою метода k-ближніх сусідів (k-nearest neighbors algorithm) та бінарного дерева рішень. Для моделі k-ближніх сусідів в якості гіперпараметру було обрано значення n_neighbors. Для дерева рішення був обраний max_depth.

За результатами побудови моделей були сформовані таблиці (див. табл. 1 та 2) де відбувається порівняння моделей за величинами: коефіцієнту детермінації R^2, середньої помилки апроксимації A, рівня прогнозування PRED (0,25).

Таблиця 1

Порівняння моделей побудованих за допомогою метода k-ближніх сусідів

Значення n_neighbors	Коефіцієнт детермінації, R^2	Середня помилка апроксимації, A у %	PRED (0,25)
2	0,838	20,622	0,654
3	0,771	25,164	0,731
4	0,674	26,628	0,692
5	0,665	25,429	0,692

Досліджуючи порівняння регресійних моделей на основі методу k-ближніх сусідів можна стверджувати, що за значенням коефіцієнту детермінації 3 моделі виявилися придатними ($0,5<R^2<0,8$), а модель де n_neighbors дорівнює 2 – доброю ($0,8<R^2<0,95$). Проте, величина середньої помилки апроксимації за всіма моделями перевищує граничне значення (A>10%), що вказує на непридатність отриманих моделей. Значення PRED (0,25) всіх моделей за методом k-ближніх сусідів також менший за 0,75.

Таблиця 2

Порівняння моделей побудованих за допомогою метода бінарного дерева рішень

Значення глибини дерева, max_depth	Коефіцієнт детермінації, R^2	Середня помилка апроксимації, A у %	PRED (0,25)
2	0,821	27,599	0,615
3	0,882	19,155	0,769
4	0,942	12,632	0,846
5	0,967	9,860	0,923

Порівняння моделей за методом бінарного дерева рішень (див. табл. 2) показало, що за величиною коефіцієнту детермінації 3 моделі є добрими, а одна модель (при max_depth = 5) має високу точність апроксимації ($R^2>0,95$). Вказана модель, також має величину середньої помилки апроксимації яка відповідає нормативному значенню. Нажаль, величина даного показника для інших моделей дерева рішень знову перевищує граничне значення. Значення PRED (0,25) трьох моделей вище 0,75. За результатами порівняння, за всіма трьома показниками, найліпшою була визначена модель з глибиною дерева 5. Графік залежності розміру коду вебдодатків за даною моделлю наведено на рис. 1.

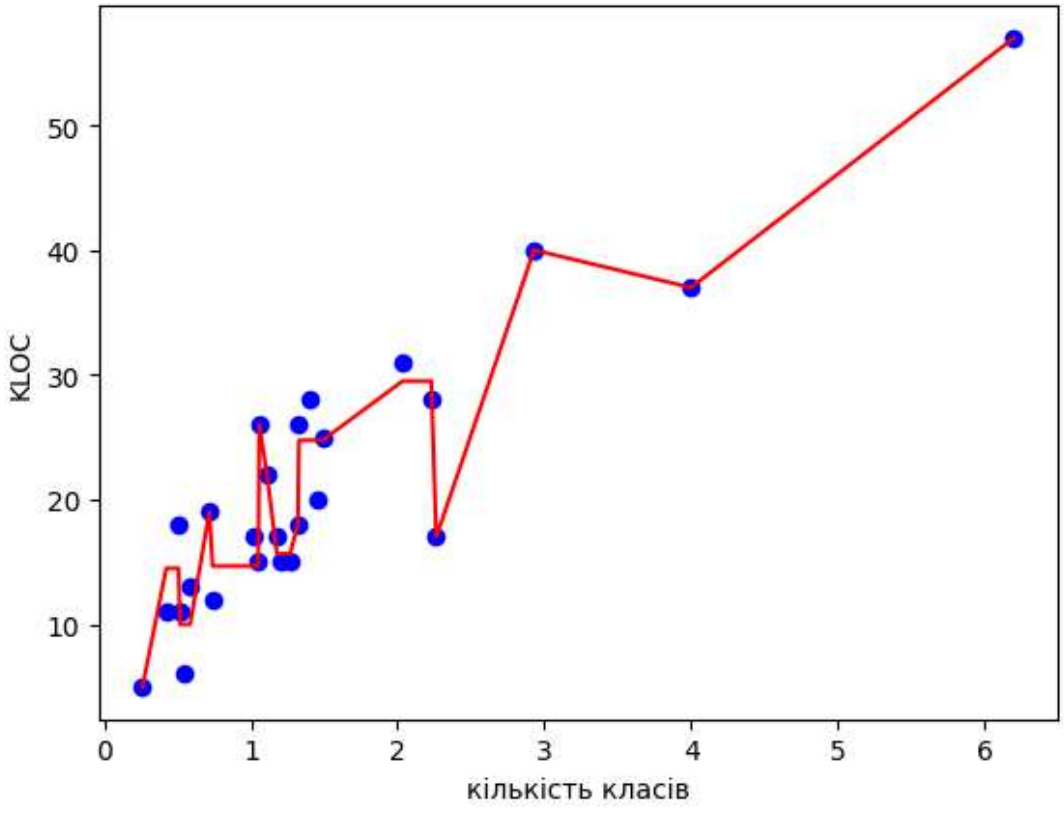

Рисунок 1 - Графік залежності розміру коду вебдодатків за моделлю дерева рішень

Отриману модель можна представити у вигляді схеми, яка наведена на рис. 2

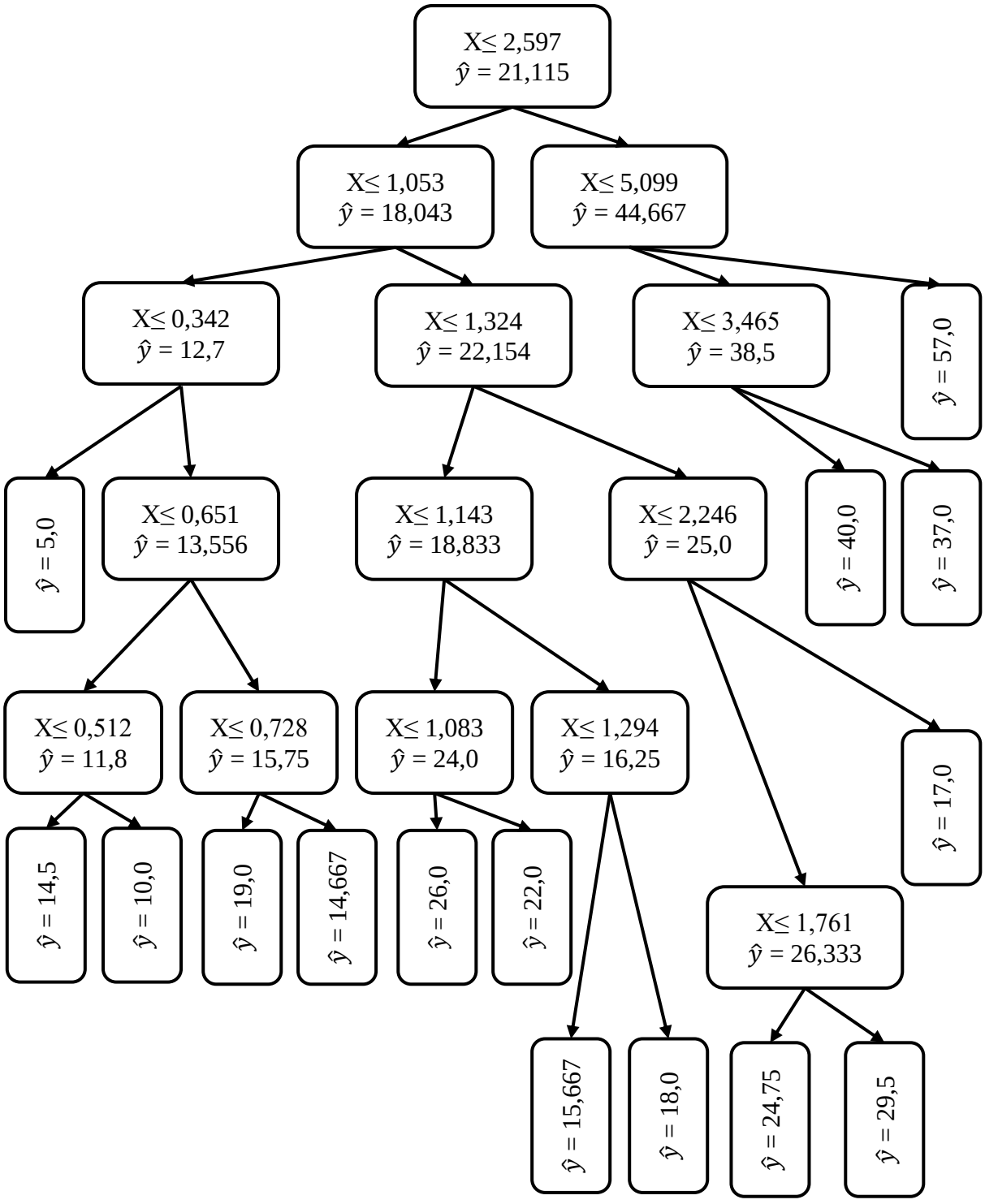

Рисунок 2 - Регресійне дерево для оцінювання розміру вебдодатків, реалізованих мовою Java (глибина дерева дорівнює 5)

Висновки. В результаті проведеного дослідження була побудована регресійна модель оцінювання розміру вебдодатків, реалізованих мовою Java. Перспективи подальших досліджень полягають в розробці програмного забезпечення для автоматизації процесу розрахунків за запропонованою моделлю.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Prykhodko, S.B., Prykhodko, K.S., & Smykodub, T.G. (2022). A statistical estimation of the coupling between object metric for open-source apps developed in Java. *Herald of Advanced Information Technology*, 5, 175–184. https://ela.kpi.ua/bitstream/123456789/59970/1/8-11_K.pdf

[2] Кулаковська І.В. (2019). Нелінійне множинне регресійне рівняння для оцінювання розміру програмного забезпечення з відкритим кодом типу mp3players на JAVA. *Комп'ютерно-інтегровані технології: освіта, наука, виробництво*, 37, 72-80. http://cit-journal.com.ua/index.php/cit/article/view/42/93

[3] Макарова, Л.М., Латанська, Л.О., Нікітін, О.В., & Нікітіна, О.Ю. (2022). Математичні моделі для оцінювання розміру програмного забезпечення для крос-платформної розробки мобільних застосунків із використанням платформи Xamarin. *Вчені записки ТНУ імені В.І. Вернадського. Серія: Технічні науки*, 33 (72), 150-156. https://tech.vernadskyjournals.in.ua/journals/2022/1_2022/23.pdf

[4] Макарова, Л.М., Приходько, Н.В., & Кудін, О.О. (2019). Побудова нелінійної регресійної моделі для оцінювання розміру веб-додатків, реалізованих мовою Java. *Вісник ХНТУ*, 2(69), 145-153. https://core.ac.uk/download/pdf/287189291.pdf

Leonid Prokopovich

A regression model for estimating the size of web-applications implemented in Java

Abstract*. The problem of building a regression model for estimating the size of web-applications implemented in Java is considered. Using machine learning methods (k-nearest neighbors, binary decision tree), regression models for estimating the size of web-applications are obtained. The obtained results are compared and the best model is selected.*

Key words*: regression model; k-nearest neighbors; decision tree; program code size.*