

НЕЛІНІЙНА РЕГРЕСІЙНА МОДЕЛЬ ОЦІНЮВАННЯ РОЗМІРУ TELEGRAM БОТІВ НА PYTHON

Шакула О.В.¹; Латанська Л.О., к.ф.-м.н., доцент²

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

^{1,2} Україна, Миколаїв

¹ shakula13@gmail.com; ² llatanskaya@gmail.com

Анотація. На ресурсі GitHub зібрано емпіричні дані для побудови нелінійної регресійної моделі оцінювання розміру Telegram ботів на Python та виконано їх передобробку. Побудовано нелінійну регресійну модель для прогнозування розміру програмного забезпечення обраного типу та оцінено рівень довіри та прогнозу.

Ключові слова: передобробка даних, нелінійна регресійна модель; оцінювання розміру Telegram ботів на Python; рівень довіри та прогнозу.

Вступна частина. Розрахунок витрат – одне з найскладніших завдань в управлінні проектом. Процес оцінювання витрат на створення програмного забезпечення включає оцінку розміру програмного продукту, який розроблятиметься, оцінку необхідних зусиль і ресурсів, написання попередніх планів робіт і, нарешті, розрахунок повної вартості проекту.

Одним з основних показників, які впливають на вартість розробки, є розмір. Існує ряд методів та моделей для визначення розміру програмного забезпечення. Досить часто для прогнозування розміру використовуються регресійні моделі. Відомо, що розмір програми в значній мірі залежить від особливостей мови програмування, методології розробки, платформи розробки та інших факторів. Тому існуючі методи та моделі прогнозування розміру програмного продукту призначені для вузького класу задач. І тому для підвищення точності оцінювання розміру програмного забезпечення певного класу необхідно удосконалювати існуючі та розробляти нові методи та моделі, що робить задачу удосконалення регресійної моделі для оцінювання розміру Telegram ботів на Python актуальною [1].

Метою роботи є побудова нелінійної регресійної моделі для оцінювання розміру Telegram ботів на Python з використанням нормалізуючого перетворення десятковий логарифм.

Основна частина. Для досягнення поставленої мети було виконано наступну роботу:

– Зібрано емпіричні дані.

На ресурсі GitHub було зібрано інформацію по Telegram ботам на Python, а саме: кількість класів та кількість строк коду для 40 проектів.

– Виконано передобробку емпіричних даних.

За допомогою критерію згоди Пірсона χ^2 визначено, що емпіричні дані не являються нормально розподіленими. Тому виконано нормалізацію даних з використанням нормалізуючого перетворення десятковий логарифм.

З нормалізованих даних вилучено дублікати та викиди. При аналізі даних на викиди використано квадрат відстані Махаланобіса та допоміжну статистику, яку порівняно з критичним значенням Фішера $F_{k,N-k,\alpha}$ (N – кількість елементів у вибірці, k – кількість параметрів, α – рівень значущості). Ті елементи вибірки, для яких статистика перевищувала критичне значення Фішера, являлися викидами. Їх прибирали з вибірки і заново повторювали розрахунок. Було знайдено 2 викиди [2].

– Виконано специфікацію моделі, тобто обрано вид функціональної залежності.

В якості функціональної залежності обрано лінійну регресійну модель для очищених нормалізованих даних та нелінійну регресійну модель для вихідних даних.

– Виконано оцінювання параметрів моделі.

Побудовано лінійну регресійну модель для очищених нормалізованих даних та через зворотне перетворення отримано нелінійну регресійну модель для вихідних даних. Параметри лінійної регресії було визначено методом найменших квадратів [3].

– Виконано оцінювання рівня довіри до отриманої моделі

Побудовано довірчі інтервали для лінійної та, через зворотне перетворення, для нелінійної моделей.

– Виконано оцінювання прогнозних можливостей моделі.

Побудовано інтервали прогнозування для лінійної та нелінійної моделей. Обчислено коефіцієнт детермінації R^2 , середню величину відносної похибки MMRE та рівень прогнозування PRED(0,25) [4].

– Виконано порівняння лінійної моделі для вихідних даних у припущенні про нормальність їх розподілу та отриманої нелінійної моделі для вихідних даних по R^2 , MMRE та PRED(0,25).

По всім показникам побудована нелінійна модель є кращою, ніж лінійна, та значення R^2 , MMRE та PRED(0,25) знаходяться у допустимих межах (відповідно $>0,7$; $<0,3$ та $>0,7$). Таким чином, нелінійна модель більше підходить для прогнозування розміру Telegram ботів на Python.

Висновки. Побудовано нелінійну регресійну модель оцінювання розміру Telegram ботів на Python з використанням нормалізуючого перетворення десятковий логарифм. Виконано порівняння отриманої моделі з лінійною, побудованою у припущенні про нормальність розподілу емпіричних даних. В подальшому планується розробити програмне забезпечення для автоматизації процесу розрахунків за запропонованою нелінійною моделлю.

REFERENCES

- [1] Кантор, М. (2002). *Управление программными проектами. Практическое руководство по разработке успешного программного обеспечения*. М.: Издательский дом «Вильямс».
- [2] Макарова, Л.М., Приходько, Н.В., Кудін, О.О. (2019). Побудова нелінійної регресійної моделі для оцінювання розміру веб-додатків, реалізованих мовою Java. *Вісник Херсонського національного технічного університету*, 2 (69), 145-153.
- [3] Приходько, Н.В., Приходько, С.Б. (2018). Нелінійна регресійна модель для оцінювання розміру програмного забезпечення інформаційних систем на базі VB. *Інформаційні технології та комп'ютерна інженерія*, 3, 37-42.
- [4] Домбровский, В.В. (2004). *Эконометрика*. М.: Нов.учебник.

Oleksandr Shakula, Liudmyla Latanska

Nonlinear regression model for estimating the size of Telegram bots on Python

Abstract. Empirical data collected on the GitHub resource to build a nonlinear regression model for estimating the size of Telegram bots on Python and performed them preprocessing. A nonlinear regression model for predicting the size of the selected type of software is constructed and the level of trust and forecast is estimated.

Keywords: data preprocessing; nonlinear regression model; estimating the size of Telegram bots in Python; the level of trust and forecasting.