

УДК 519

DOI <https://doi.org/10.32782/pet-2023-1-3>

**Микола ГОЛОВІН**

кандидат фізико-математичних наук, доцент кафедри комп'ютерних наук та кібербезпеки, Волинський національний університет імені Лесі Українки, пр. Волі, 13, м. Луцьк, Україна, 43025

**ORCID ID:** <https://orcid.org/0000-0003-4516-4677>

**Ніна ГОЛОВІНА**

кандидат фізико-математичних наук, доцент кафедри експериментальної фізики, інформаційних та освітніх технологій, Волинський національний університет імені Лесі Українки, пр. Волі, 13, м. Луцьк, Україна, 43025

**ORCID ID:** <https://orcid.org/0000-0002-1152-1536>

**Бібліографічний опис статті:** Головін М., Головіна Н. (2023). Використання методу машинного навчання для аналізу динаміки швидкостей виконання практичних завдань з інформатики. *Фізика та освітні технології*, 1, 18–24, doi: <https://doi.org/10.32782/pet-2023-1-3>

**ВИКОРИСТАННЯ МЕТОДУ МАШИННОГО НАВЧАННЯ  
ДЛЯ АНАЛІЗУ ДИНАМІКИ ШВИДКОСТЕЙ ВИКОНАННЯ  
ПРАКТИЧНИХ ЗАВДАНЬ З ІНФОРМАТИКИ**

Карантинні обмеження раніше, військові дії тепер актуалізували дистанційне навчання. Важливими моментами стали: управління процесом навчання, фіксація та аналіз результатів навчання великих груп здобувачів освіти. Керування всяким процесом передбачає зворотній зв'язок. Актуальним перспективним напрямком забезпечення зворотного зв'язку у цьому випадку може стати аналіз відповідних статистичних розподілів результатів навчання. У цій роботі в якості наукової методології, що забезпечує згаданий вище аналіз, вибрані методи машинного навчання, а саме: дерево регресій та дерево класифікацій. Метою роботи є дослідження динаміки процесів навчання великих груп здобувачів освіти методами машинного навчання, що реалізується аналізом результатів вимірювань змін швидкостей виконання практичних завдань з інформатики. Наукова новизна цієї роботи полягає у спробі застосування методів машинного навчання до аналізу результатів практичних навчальних дій. Впродовж реалізації досліджень було проведено понад п'яти тисяч відповідних вимірів, написана програма для аналізу цих даних на мові Python з використанням бібліотеки Scikit-Learn. Ця програма представлена в роботі. За допомогою методів машинного навчання проведено аналіз результатів вимірювань швидкостей виконання завдань з інформатики. Графіки, що були отримані, мають гладку форму з незначним вигином, без екстремумів. На графіках фіксуються, значні лінійні фрагменти. У роботі здійснено порівняння отриманих результатів аналізу даних з результатами аналізу цих же даних, проведеного раніше із застосуванням методу моментів. Спостерігається співпадіння споріднених залежностей динаміки зміни швидкостей навчальних дій, отриманих за методом дерева регресій та методом моментів.

**Ключові слова:** методи машинного навчання, дерево регресій, дерево класифікацій, метод моментів, статистичні розподіли, Python, бібліотека Scikit-Learn.

**Mykola HOLOVIN**

Candidate of Physical and Mathematical Sciences, Associate Professor at the Department of Computer Science and Cybersecurity, Volyn National University named after Lesya Ukrainka, 13 Volya ave., Lutsk, Ukraine, 43025

**ORCID ID:** <https://orcid.org/0000-0003-4516-4677>

**Nina HOLOVINA**

Candidate of Physical and Mathematical Sciences, Associate Professor at the Department of Experimental Physics, Information and Educational Technologies, Volyn National University named after Lesya Ukrainka, 13 Volya ave., Lutsk, Ukraine, 43025

**ORCID ID:** <https://orcid.org/0000-0002-1152-1536>

**To cite this article:** Holovina N., Holovina N. (2023). Vykorystannia metodu mashynnoho navchannia dlia analizu dynamiky shvydkostei vykonannia praktychnykh zavdan z informatyky [Using the machine learning method to analyze the dynamics of the performance of practical tasks in computer sciences]. *Physics and Educational Technology*, 1, 18–24, doi: <https://doi.org/10.32782/pet-2023-1-3>

## USING THE MACHINE LEARNING METHOD TO ANALYZE THE DYNAMICS OF THE PERFORMANCE OF PRACTICAL TASKS IN COMPUTER SCIENCES

*Covid-19 quarantine restrictions and the war in Ukraine have actualized distance learning more than ever. In these conditions, management of the learning process, recording and analysis of learning results of large groups of education seekers became very important. Management of any process involves feedback. In this case, the analysis of relevant statistical distributions of training results can become a relevant perspective direction of providing feedback. In this work, the methods of machine learning, namely: the regression tree and the classification tree, were selected as a scientific methodology that ensures the above-mentioned analysis. The purpose of the work is to study the dynamics of learning processes of large groups of education seekers using machine learning methods, which is realized by analyzing the results of measuring changes in the speed of performing practical tasks in informatics. The scientific novelty of this work consists of an attempt to apply machine learning methods to the analysis of the results of practical educational activities. In the course of research, more than five thousand relevant measurements were taken. In addition a program was written for the analysis of these data in the Python language using the Scikit-Learn library. This program is presented in the work. With the help of machine learning methods, an analysis of the results of measuring the speed of computer science tasks was carried out. The resulting graphs have a smooth shape with a slight bend, without extremes. Significant linear fragments are fixed on the graphs. The paper compares the results of the data analysis with the results of the analysis of the same data carried out earlier using the method of moments. There is a coincidence of related dependencies of the dynamics of changes in the speed of educational actions, obtained by the regression tree method and the method of moments.*

**Key words:** machine learning methods, regression tree, classification tree, method of moments, statistical distributions, Python, Scikit-Learn library.

**Актуальність проблеми.** При впровадженні в практику освіти дистанційного навчання важливими стають аналіз результатів цього навчання та управління цим процесом. При дистанційному навчанні кількість здобувачів освіти зростає, а процес навчання набуває стохастичного характеру. Якщо при традиційному навчанні вчитель мав справу з групою, яка складалась від десятка до кількох десятків осіб, то при дистанційному навчанні керування може здійснюватись процесом, який включає дії значно більшої кількості суб'єктів навчання. Навчальний матеріал може вимагати такого зворотного зв'язку, коли сотні суб'єктів навчання виконують десятки завдань он-лайн в стислі проміжки часу. У цих умовах інформація, що надходить від суб'єктів навчання складається з тисяч відповідей. При такій лавиноподібній інформаційній складовій навчання, актуальним стає автоматизоване, комп'ютерно-орієнтоване навчання, яке технологічно спирається на імовірнісні стохастичні процеси. У цьому контексті стають цікавими статистичні розподіли результатів навчання.

**Аналіз попередніх досліджень і публікацій.** У роботах (Головін, 2011; Головін, 2018) розглядалися технологічні підвалини навчання, які базувались на аналізі статистичних роз-

поділів швидкостей навчальних дій за оригінальним методом (Головін, 2020), що в свою чергу ґрунтується на відомому методі моментів (Гласс, 1976; Крамер, 1975). У цих роботах досліджувалось, як проявляються когнітивні механізми мислення, характерні для окремих особистостей в процесі навчання великих груп студентів.

Актуальним перспективним напрямком подальшого аналізу статистичних розподілів результатів навчання можуть стати технології, пов'язані з методами машинного навчання, зокрема, методи дерева регресій та дерева класифікацій (Barros, 2011; Hugh, 1998).

**Метою роботи** є дослідження динаміки процесів навчання великих груп здобувачів освіти методами машинного навчання, що реалізується аналізом результатів вимірювань змін швидкостей виконання практичних завдань з інформатики.

**Виклад основного матеріалу дослідження.** Обробка експериментальних даних була реалізована на мові Python із застосуванням бібліотеки Scikit-Learn, яка є однією з найпопулярніших бібліотек машинного навчання на мові Python.

Інформаційною базою цієї роботи були результати навчання великих груп здобува-

чів освіти, що хотіли покращити свої уміння і навички в галузі практичної роботи та відпрацьовували до автоматизму дії, які стосувались роботи з файлами та текстами. В експерименті використовувались два рівноцінних, відкаліброваних за часом виконання пакети завдань. У кожному пакеті завдань було по 15 вправ. Кожне завдання пакету можна було виконати в проміжку часу від кількох хвилин до 15–25 секунд, при сформованих уміннях і навичках. Саме ця динаміка формування умінь та навичок досліджувалась. Необхідно відмітити, що ця база вже оброблювалась методом моментів (Головін, 2011) на предмет проявлення когнітивних механізмів мислення, характерних для окремих особистостей в процесі навчання великої групи.

Завдання мали комплексний характер. У завданнях на роботу з файлами практичні дії реалізовувались в основному через файловий менеджер. Передбачалось, що виконання завдань відбувались в технологіях, що пов'язані, як з буфером обміну, так і з технологією перетаскування. Пакет завдань з текстами мав у своєму складі завдання на копіювання, перенос, знищення, форматування фрагментів текстів, розташованих у різних вікнах і різних файлах. У складі завдань були вправи, що вимагали сортування, на роботу зі списками та таблицями. Навчальні практичні дії можна було здійснювати кількома різними редакторами.

Зрозуміло, що для виконання завдань у першому і в другому пакеті були в значній мірі пов'язані спільними технологіями, хоча і виконувались різними програмними засобами.

Фіксація результатів дій була реалізована через оригінальну програму, що контролювала події в файловій системі комп'ютера.

База даних, в якій концентрувались результати навчання включала інформацію про номер групи, номер студента, номер завдання в пакеті завдань, що виконувався, номер спроби та час виконання. Зрозуміло, що в базі була також інформація про дату і час виконання, а також про те правильно чи не правильно виконане завдання.

Всього в інформаційній базі, що оброблювалась було більше 5 тисяч записів про динаміку зміни швидкості виконання завдань.

Програма, що використовувалась для аналізу динаміки зміни швидкості виконання завдань

була розроблена на мові Python з використанням бібліотеки sklearn (scikit-learn).

На початковому етапі роботи ця програма завантажує інформаційну базу про швидкість виконання завдань впродовж серії послідовних спроб. Результатом роботи програми є графік, що візуалізує динаміку навчання. Надалі саме форма кривої залежності часу (Time) виконання завдань від номеру спроби (N) має піддаватись аналізу.

Програма починається з підключення потрібних бібліотек з відповідними, необхідними для роботи програми методами. Далі з файлу «Learn.csv» читається сама інформаційна база.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as seabornInstance
from sklearn.model_selection import
train_test_split
from sklearn.linear_model import
LinearRegression
from sklearn.tree import
DecisionTreeRegressor
from sklearn import metrics
dataset = pd.read_csv('learn.csv')
```

Після завантаження інформаційної бази виводиться інформація, що характеризує цю базу. Цю роботу виконують наступні рядки.

Рядок `print(dataset.shape)` виводить кількість рядків і стовпчиків у базі, а рядок `print(dataset.describe())` – розширений опис бази.

У нашому випадку, в основній базі 2884 рядків та 5 стовпчиків. Ця база відповідає пакету завдань роботи з файлами (рис. 1, 2а, 3а, 4). База, для пакету завдань роботи з текстами, має 2304 рядків і 5 стовпчиків (рис. 2б, 3б, 4). У стовпчиках обох баз інформація про: номер групи, номер студента, номер завдання, номер спроби «N», та час виконання завдання «Time».

Корисно перед обробкою даних бачити цілісно графік статистичного розподілу. Виведемо його попередньо розташувавши спроби «N» в масиві «X», а час виконання завдань «Time» – в масиві «y».

```
X = dataset['N'].values.
reshape(-1,1)
y = dataset['Time'].values.
```

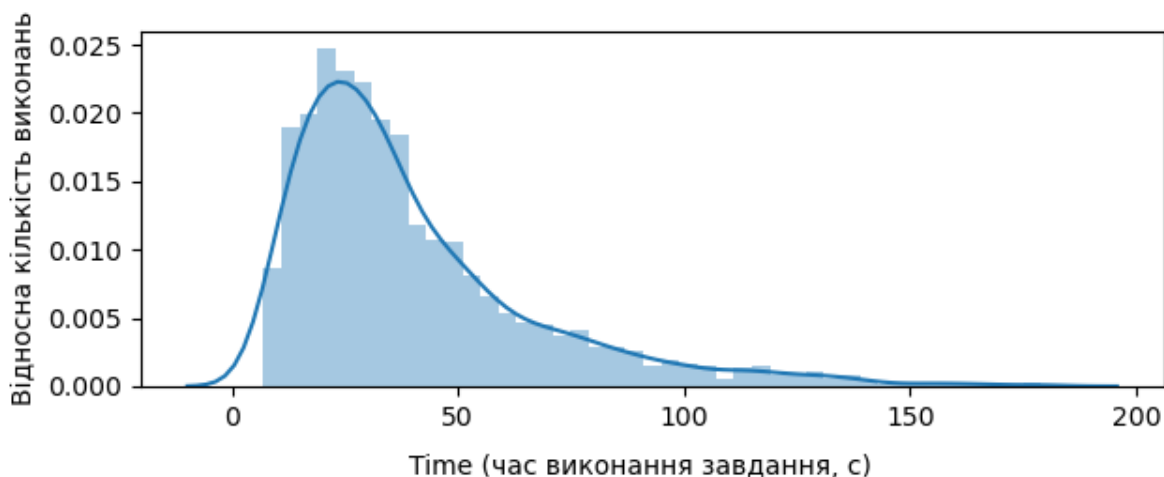


Рис. 1. Розподіл кількості виконаних завдань за часом виконання

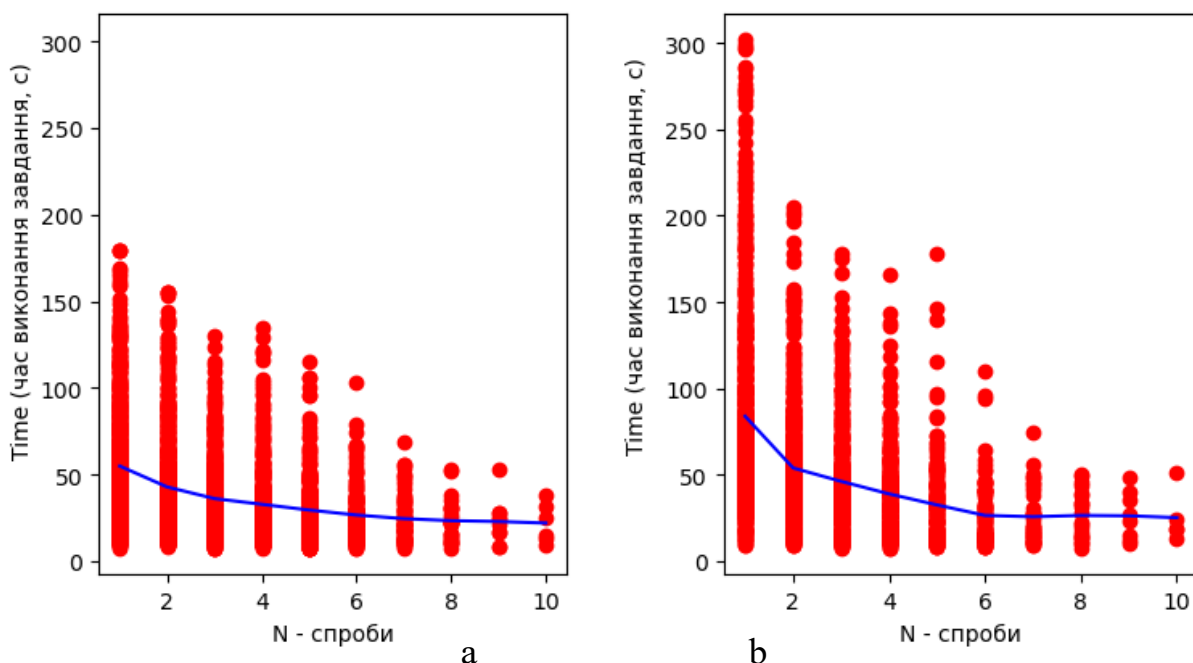


Рис. 2. Лінія динаміки навчання великої групи здобувачів освіти із освоєння пакету завдань роботи з файлами (а); з текстами (б). Лінія була визначена методом дерева регресії (DecisionTreeRegressor)

```

reshape(-1,1)
plt.figure(figsize=(15,10))
plt.tight_layout()
seabornInstance.
distplot(dataset['Time'])
plt.xlabel('Time (час виконання
завдання, с)')
plt.ylabel('Відносна кількість
виконань')
plt.show()

```

У нашому випадку графік має вигляд наближений до нормального (рис. 1). Це свідчить про статистичний, стохастичний, імовірнісний характер вхідних даних. Розділимо вхідні дані кожного пакету експериментальних даних на два набори: навчальний набір 65% та тестовий набір 35%. Змінна `test_size` відповідальна за цей поділ.

```

X_train, X_test, y_train, y_test = train_test_
split(X, y, test_size=0.35, random_state=0)

```

Після поділу даних на навчальні та тестові набори, було проведено відповідне навчання. Для цього був створений екземпляр класу DecisionTreeRegressor, що імпортований з бібліотеки sklearn. Проведено тренування викликом методу fit () з використанням відповідних вхідних даних.

```
regressor =
DecisionTreeRegressor(random_state=0)
regressor.fit(X,y)
plt.scatter(X, y, color = 'red')
plt.plot(X, regressor.predict(X),
color = 'blue')
plt.title('Залежність часу виконання
завдання від спроби')
plt.xlabel('N - спроби')
plt.ylabel('Time (час виконання
завдання, c)')
plt.show()
```

У результаті роботи цього програмного фрагменту з інформаційними базами, що відповідні динаміці процесу навчання, а саме, засвоєння роботи: з файлами (основна база) та текстами (альтернативна база, для порівняння) отримаємо відповідні графіки на рис. 2а та рис. 2б.

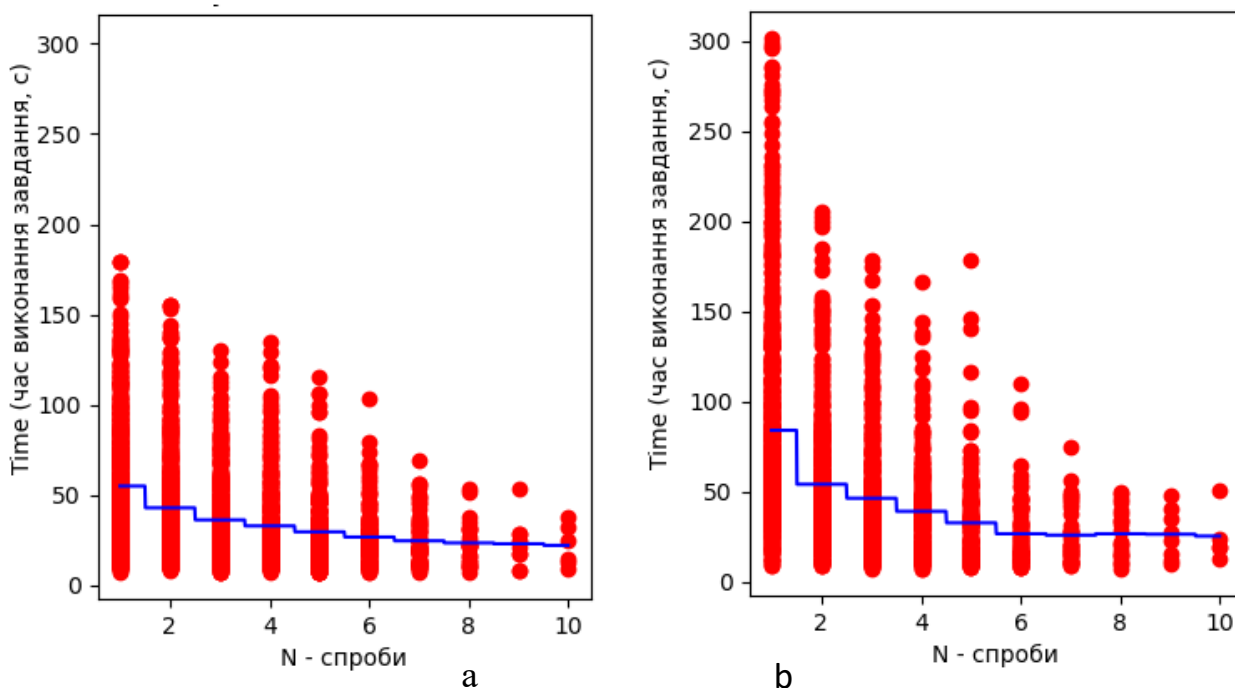
Вертикальні ряди точок вказують на результати швидкості навчальних дій окремих здобувачів освіти в окремих спробах у процесі освоєння роботи відповідно з файлами (рис. 2а) та текстами (рис. 2б).

Видно, що графіки отримані методом дерева регресії мають, з одного боку, різну динаміку змін, а з іншого, багато спільного в сенсі форми відповідних кривих.

Графіки мають гладку форму з незначним вигином, без екстремумів. На обох графіках фіксуються, значні лінійні фрагменти. Обидва графіки в області кінцевих спроб виконання завдань виходять на полицьки. В обох випадках рівень полицьок складає 24 – 22 секунди на виконання окремого завдання в спробі.

Відмінність графіків полягає у різному положенні точок, що відповідні першій спробі виконати завдання. Відповідно різна динаміка змін від початкової до кінцевої спроби.

Лінійний фрагмент лінії дерева регресії спостерігається на рис. 2а в проміжку від 3 до 10 спроби. На рис. 2б можна спостерігати дві лінійні області від 2 до 6 спроби та від 6 до 10. Аналогічні особливості можна побачити на лініях класифікацій представлених відповідно на рис. 3а та рис. 3б. Ці графіки є результатом роботи наступного програмного фрагменту.



**Рис. 3. Лінія динаміки навчання великої групи здобувачів освіти із освоєння пакету завдань роботи з файлами (а); з текстами (б). Лінія була визначена методом дерева класифікацій**

```
X_grid = np.arange(min(X), max(X),
0.01)
X_grid = X_grid.reshape((len(X_grid),
1))
plt.scatter(X, y, color = 'red')
plt.plot(X_grid, regressor.
predict(X_grid), color = 'blue')
plt.title('Залежність часу виконання
завдання від спроби')
plt.xlabel('N - спроби')
plt.ylabel('Time (час виконання
завдання, c)')
plt.show()
```

Автори роботи намагались калібрувати пакет завдань роботи з файлами та текстами таким чином, щоб ці два пакети можна було б використовувати разом. Однак, виявилась цікава особливість: обидва пакети можна вважати однаково каліброваними тільки на кінцевому етапі навчання. В обох випадках на виконання завдань в кінцевих спробах необхідний однаковий час – в середньому порядку двадцяти секунд.

На початкових спробах ці пакети не можна вважати однаково каліброваними. Так, у першій спробі при роботі з файлами в середньому досягається швидкість дій 55 секунд на одне завдання. У випадку роботи з текстами швидкість була на рівні 85 секунд на одне завдання в першій спробі. Це суттєво більше ніж в першому випадку. Метод машинного навчання дозволяє порівнювати фактичні дані з прогнозованими. Такий аналіз проведено і він дав хорошу кореляцію.

Отримані дані корелюють з аналізом, що був проведений іншим методом у роботі (Головін, 2011), де аналізувались ті ж самі експериментальні дані, що і в поточній роботі. Досліджувалась динаміка зміни моментів розподілів швидкостей дій в різних спробах. Тут необхідно пояснити, що статистичний контур, зображений на рис. 1 може бути представлений сукупністю гаусоподібних контурів, змішених один відносно одного в шкалі швидкостей виконання завдань. Зрозуміло, що контур розділу, відповідний початковій спробі, буде знаходитись в області повільних дій, а контур кінцевої спроби – в області більш швидких дій. Саме динаміка зміни положень цих контурів і їх форми досліджу-

валась у згаданій роботі. Зокрема, перший момент – центр тяжіння розподілу, змінювався від спроби до спроби так, як показано на рис. 4.

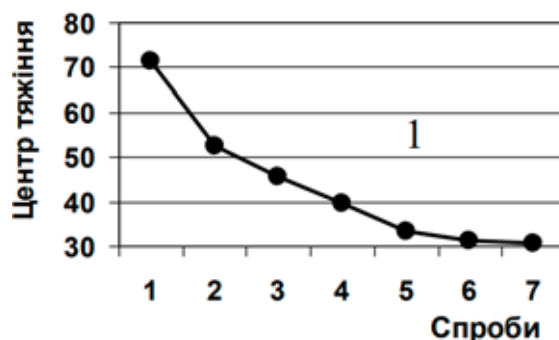


Рис. 4. Динаміка зміни першого моменту, центру тяжіння розподілу, від спроби до спроби

При порівнянні рис. 2b та рис. 4. видно, що спостерігаються співпадіння характерних особливостей цих залежностей. Кут залежностей при переході від спроби 1 до спроби 2 більший, ніж при переході від спроби 2 до спроби 3,4,5. Видно також, що спостерігається співпадіння лінійностей цих залежностей на проміжку 2-5 спроб.

#### Висновки і перспективи подальших досліджень

Проведено експеримент та аналіз застосування методів машинного навчання: лінійна регресія, регресія дерева рішень для дослідження результатів вимірювань швидкостей виконання завдань з інформатики. Порівняння прогнозованих та фактичних даних дало хорошу кореляцію.

Отримані графіки мають експоненціальну гладку форму з незначним вигином, без екстремумів. На обох графіках фіксуються, значні лінійні фрагменти. Обидва графіки в області кінцевих спроб виконання завдань виходять на полицьки. В обох випадках рівень полицьок складає 24 – 22 секунди на виконання окремого завдання в спробі. Аналіз процесу навчання великої групи здобувачів освіти зручно здійснювати через оцінку трансформації форми кривих динаміки навчання. Так, на початковому етапі навчання нахил кривої більший, ніж в центральній частині. Це свідчить про більш швидке набуття умінь та навичок на початковому етапі навчання. Вихід на полицьку в кінцевих спробах свідчить про сформовані уміння

та навички. Центральна лінійна частина кривої, що охоплює найбільшу кількість спроб, відображає типову швидкість процесу навчання на рутинному його етапі.

Наведені вище статистичні дослідження дають змогу виявити найзагальніші закономірності навчання, які проявляються тільки як консолідований результат діяльності великої групи учнів. Цей результат не може бути отриманий під час аналізу навчальних дій окремих індивідумів.

Спостерігається співпадіння споріднених залежностей динаміки зміни швидкостей навчальних дій, отриманих за методом дерева регресій та методом моментів.

Проведені дослідження відкривають перспективу для подібної обробки результатів навчання у більш складних розділах інформатики, наприклад, в програмуванні. Тут можна досліджувати швидкість формування стандартних програмних конструкцій, зокрема таких як вкладені цикли та розгалуження.

#### ЛІТЕРАТУРА:

1. Головін М.Б. Дослідження процесів навчання на основі аналізу моментів статистичних розподілів швидкостей навчальних дій (на матеріалах вивчення інформатики). *Психологічні перспективи*. Луцьк, 2011. В. 18. С. 62–72. URL: <https://evnuir.vnu.edu.ua/bitstream/123456789/1743/1/Holovin.pdf>
2. Головін М.Б., Головіна Н. А., Головіна Н. М. Модельний розгляд пізнавальних процесів супутніх навчальному програмуванню. *Психологічні перспективи*. Луцьк, 2018. В. 31. С. 57 – 70. URL: [http://nbuv.gov.ua/UJRN/Ppst\\_2018\\_31\\_7](http://nbuv.gov.ua/UJRN/Ppst_2018_31_7)
3. Головін М.Б., Головіна Н. М., Гузачов Д.М., Головіна Н.А. Метод моментів як інструмент комп'ютерної діагностики навчальної діяльності. *Комп'ютерно-інтегровані технології: освіта, наука, виробництво*. Луцьк: Вид-во ЛНТУ, 2020. В. 38. С. 67-78. URL: <https://evnuir.vnu.edu.ua/bitstream/123456789/19700/1/moment.pdf>
4. Гласс Дж. *Статистические методы в педагогике и психологии*. М.: Прогресс, 1976. 494 с.
5. Крамер Г. *Математические методы статистики*. М.: Мир, 1975. 648 с.
6. Barros R. C., Cerri R., Jaskowiak P. A., Carvalho A. C. P. L. F. A bottom-up oblique decision tree induction algorithm. *Proceedings of the 11th International Conference on Intelligent Systems Design and Applications (ISDA 2011)*. 2011. С. 450–456. ISBN 978-1-4577-1676-8. doi:10.1109/ISDA.2011.6121697
7. Hugh A. Chipman, Edward I. George, Robert E. McCulloch. Bayesian CART model search. *Journal of the American Statistical Association*. 1998. Т. 93, В. 443. С. 935–948. doi:10.1080/01621459.1998.10473750
8. Бібліотека Scikit-Learn [https://scikit-learn.org/dev/whats\\_new/v0.20.html](https://scikit-learn.org/dev/whats_new/v0.20.html)

#### REFERENCES:

1. Holovin M.B. (2011). Doslidzhennia protsesiv navchannia na osnovi analizu momentiv statystychnykh rozpodiliv shvydkostei navchalnykh dii (na materialakh vvychennia informatyky) [The study of learning processes based on the analysis of the moments of statistical distributions of the speed of learning actions (on the materials of the study of informatics)]. *Psykhologichni perspektyvy*. Lutsk, 2011. V. 18. S. 62 – 72. URL: <https://evnuir.vnu.edu.ua/bitstream/123456789/1743/1/Holovin.pdf>
2. Holovin M.B., Holovina N. A., Holovina N. M. (2018). Modelnyi rozghliad piznavalnykh protsesiv suputnykh navchalnomu prohramuvanniu [Model consideration of cognitive processes associated with educational programming]. *Psykhologichni perspektyvy*. Lutsk. V. 31. S. 57–70. URL: [http://nbuv.gov.ua/UJRN/Ppst\\_2018\\_31\\_7](http://nbuv.gov.ua/UJRN/Ppst_2018_31_7)
3. Holovin M.B., Holovina N. M., Huzachov D.M., Holovina N.A. (2020). Metod momentiv yak instrument komp'uternoї diahnostryky navchalnoi diialnosti [The method of moments as a tool for computer diagnostics of educational activity]. *Komp'uterno-intehrovani tekhnolohii: osvita, nauka, vyrobnytstvo*. Lutsk: Vyd-vo LNTU, 2020. V. 38. S. 67-78. URL: <https://evnuir.vnu.edu.ua/bitstream/123456789/19700/1/moment.pdf>
4. Hlass Dzh. *Statisticheskie metody v pedagogike i psihologii* [Statistical methods in pedagogy and psychology]. М.: Prohress, 1976. 494 s.
5. Kramer H. *Matematicheskie metody statistiki* [Mathematical methods of statistics]. М.: Myr, 1975. 648 s.
6. Barros R. C., Cerri R., Jaskowiak P. A., Carvalho A. C. P. L. F. A bottom-up oblique decision tree induction algorithm. *Proceedings of the 11th International Conference on Intelligent Systems Design and Applications (ISDA 2011)*. 2011. S. 450–456. ISBN 978-1-4577-1676-8. doi:10.1109/ISDA.2011.6121697
7. Hugh A. Chipman, Edward I. George, Robert E. McCulloch. Bayesian CART model search. *Journal of the American Statistical Association*. 1998. Т. 93, V. 443. S. 935–948. doi:10.1080/01621459.1998.10473750
8. Biblioteka Scikit-Learn [https://scikit-learn.org/dev/whats\\_new/v0.20.html](https://scikit-learn.org/dev/whats_new/v0.20.html)