

Використання методів Data Mining для аналізу даних рекламної мережі

Важливою і активно досліджуваною темою в сучасних дослідженнях рекламних мереж є прогнозування та аналіз даних, що містить побудову моделей отримання даних і вибір методів їх аналізу. Центральним аспектом дослідження є використання методів Data Mining для обробки великих масивів рекламних даних. Запропоновано математичну модель та методи, що дозволяють прогнозувати навантаження на рекламні сервери і кластеризувати стан їх роботи, визначаючи потенційні збої та проблеми на ранніх стадіях. Ці моделі та методи обробки даних впроваджуються у системи керування, дозволяючи оптимізувати роботу рекламних мереж, знижуючи фінансові витрати та підвищуючи ефективність управління рекламними ресурсами. Крім цього, розроблено алгоритми для автоматичного масштабування ресурсів, що базуються на прогнозованих даних, забезпечуючи стабільність та надійність системи. Запропоновані технології та методи можуть бути використані у кадровому менеджменті та підготовці висококваліфікованих ІТ-спеціалістів, забезпечуючи їхнє відповідне навчання та розвиток відповідно до вимог сучасних рекламних мереж.

Ключові слова: інтелектуальний аналіз даних; прогнозування; рекламні мережі; машинне навчання; поліноміальна регресія.

Актуальність теми. Складність аналізу та консолідації великих обсягів і великої розмірності даних рекламних серверів, що не передбачають чіткої структури та зв'язків, в умовах щоденного збільшення обсягів інформації загострює потребу в сучасних комп'ютеризованих засобах інтелектуального аналізу та прогнозування поведінки їх роботи. Зростання популярності онлайн-сервісів та реклами в інтернеті відкриває безпрецедентні можливості для вирішення дослідницьких завдань.

Аналіз останніх досліджень та публікацій, на які спираються автори. Зокрема, інтелектуальним алгоритмам аналізу даних присвятили публікації І.І. Сугоняк [2, 3], В.Л. Левківський [1], В.О. Праздніков [2], В.Я. Круціцький [3], Н.М. Лобанчикова [1], Г.В. Марчук [1]. Аналіз цих досліджень дав можливість категоризувати бачення застосування запропонованої моделі та підібрати підходи і алгоритми для доменної області аналізу даних рекламних мереж.

Метою статті є аналіз методів і моделей для прогнозування та аналізу даних рекламних мереж, а також розробка ефективних стратегій оптимізації роботи рекламних серверів з використанням технологій Data Mining. Особлива увага приділяється вдосконаленню методів збору даних, їх обробки та аналізу для визначення неочевидних закономірностей, які можуть сприяти підвищенню ефективності рекламних кампаній і зниженню витрат на підтримку мережі.

Викладення основного матеріалу дослідження. Для створення системи прогнозування було побудовано фундаментальну систему моніторингу стану – класичну систему логування. Для того щоб створити можливість прогнозування у часі, наближеному до реального, було використано систему повідомлень Kafka та її інтеграцію з timeseries OLAP-сховищем Druid. Druid – це розподілена база даних в реальному часі, розроблена для швидкої обробки великих обсягів даних. Вона оптимізована для запитів на агрегування даних та OLAP (онлайн-аналітична обробка) запитів. Druid часто використовується в системах аналітики великих даних, де потрібна висока продуктивність та масштабованість. Системи (в цьому випадку рекламні сервери) надсилають канонічні метрики до системи моніторингу, такі як: запит, його результат, тривалість, шлях сервера, параметри та інше (рис. 1). Ці дані доступні для аналізу майже миттєво, завдяки створеній системі інтеграції передаються, індексуються з мінімальною затримкою і готові для подальшої обробки.

```
df v
1 {
2   "intervals":["1900-01-01/2100-01-01"],
3   "granularity":{"type":"period","period":"PT1H","timeZone":"UTC"},
4   "context":{"timeout":60000},
5   "queryType":"timeseries",
6   ["dataSource":"exchange_onesegment",
7   "aggregations":
8   [{"type":"doubleSum","name":"sum","fieldName":"event_count"}],
9   "filter":
10  [{"type":"selector","dimension":"event_type","value":"2"}
11 ]
```

Рис. 1. Запит для отримання кількості запитів до системи погодинно

У контексті дослідження моделювання навантаження на сервери важливим аспектом є використання поліноміальної регресії для аналізу часових рядів даних. Поліноміальна регресія 3-го ступеня була обрана для моделювання добових змін кількості запитів до сервера, які наближено умовно відповідають «синусоїдальному» розподілу з піковими значеннями вдень та мінімальними вночі. Математична модель виражена таким рівнянням:

$$Q = \beta_0 + \beta_1 T + \beta_2 T^2 + \beta_3 T^3 + \varepsilon, \quad (1)$$

де Q – залежна змінна, прогнозована кількість запитів за одиницю часу, що визначена як $Q \in N, 1 \leq Q \leq S \times 1000$, позначаючи S як кількість рекламних серверів, що відповідно може обробляти 1000 запитів за секунду, T – одиниця часу, де $T \in \{0, \dots, 23\}$ що відображає години, β_i – константні коефіцієнти регресійної моделі, ε – випадковий шум або помилка моделі. Коефіцієнти моделі розраховуються за допомогою методу найменших квадратів. Вектор коефіцієнтів ($\beta_i = \text{const}$) обчислюється за формулою:

$$\beta = (X^T X)^{-1} X^T y, \quad (2)$$

де X – матриця дизайну, яка інкорпорує інтерцепт і степені часу доби. Кожен рядок відповідає одному часовому інтервалу. Кожен стовпець відповідає степеням часу доби ($1, T, T^2, T^3$), X^T – транспонована матриця X , $(X^T X)^{-1}$ – обернена матриця від добутку транспонованої матриці X на матрицю X^T у X на вектор y , який містить спостережувані значення кількості запитів, y – вектор відповідей, який містить фактичні значення кількості запитів для кожного часового інтервалу. Ця модель дозволяє кількісно оцінити вплив часу доби на кількість запитів та підтвердити теоретичні припущення про сезонні коливання в навантаженні на сервер. Регресійний аналіз було виконано за допомогою методу найменших квадратів, що забезпечив оптимальне підбирання параметрів моделі для максимально точного прогнозування. Використання методу найменших квадратів дозволило оптимізувати параметри моделі, мінімізуючи суму квадратів різниць між спостережуваними та прогнозованими значеннями.

Модель була перевірена за допомогою крос-валідації, що забезпечило оцінку її узагальнюваної здатності. Вибір полінома 3-го ступеня був обґрунтований аналізом валідаційних кривих, де було виявлено, що такий ступінь забезпечує оптимальне співвідношення між точністю апроксимації та уникненням перенавчання. В результаті модель продемонструвала високу точність прогнозування, підтверджену метриками, такими як середньоквадратична помилка (MSE) та корінь з середньоквадратичної помилки (RMSE), що свідчить про її ефективність для практичного використання у вирішенні завдань прогнозування навантаження на IT-інфраструктуру. У системі візуалізації цей запит буде відображено таким чином (рис. 2).



Рис. 2. Вигляд запиту у системі візуалізації

У цьому випадку дані відображаються за часовим поясом ЮТС-0, але можна чітко прослідкувати залежність навантаження від часу, за яким відбувається агрегація. Далі підключаємо модуль аналізу за поліноміальною регресією, який реалізований нами у вигляді бібліотеки javascript та виконується на сервері системи графічного представлення, аналізуючи результуючий набір даних (рис. 3):

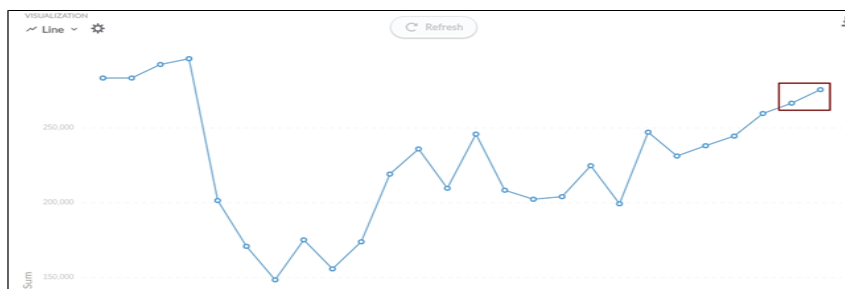


Рис. 3. Результат запиту з ввімкненою системою прогнозування

Тобто якщо результуючий набір даних закінчується на 23.00 1 серпня з результатом у 259 663 запити, то система прогнозування визначає, що о 00.00 2 серпня очікується кількість запитів у розмірі 266 330, а о 1 годині 2 серпня – 275 468. Що дає можливість прослідкувати, що трафік буде рости, але у межах допустимого навантаження на систему. Це дає нам змогу з певною точністю визначати на декілька наступних годин приблизний обсяг навантаження на систему у вигляді кількості запитів, що дає потужний інструмент аналізу з точки зору забезпечення безперебійного процесу роботи системи.

Класифікація вектора запитів, отриманих з наступного запиту до сховища даних (рис. 4):

```

1 {
2   "intervals": ["1900-01-01/2100-01-01"],
3   "granularity": "all",
4   "context": {
5     "timeout": 60000
6   },
7   "queryType": "groupBy",
8   "dataSource": "exchange_onesegment",
9   "aggregations": [
10    {
11      "type": "doubleSum",
12      "name": "sum",
13      "fieldName": "event_count"
14    }
15  ],
16  "dimensions": [
17    {
18      "type": "extraction",
19      "dimension": "time",
20      "outputName": "timestamp",
21      "extractionFn": {
22        "type": "timeFormat",
23        "format": "yyyy-MM-dd'T'HH:00:00Z",
24        "timeZone": "UTC",
25        "locale": "en-US"
26      }
27    },
28    {
29      "type": "event type",
30      "dimension": "event type"
31    }
32  ],
33  "limitSpec": [
34    {
35      "columns": [
36        {
37          "dimension": "timestamp",
38          "direction": "ascending"
39        },
40        {
41          "dimension": "event type",
42          "direction": "ascending"
43        }
44      ]
45    }
46  ]
47 }

```

Рис. 4. Приклад запиту до сховища

Результатом виконання цього запиту будуть дані зазначені далі (рис. 5):

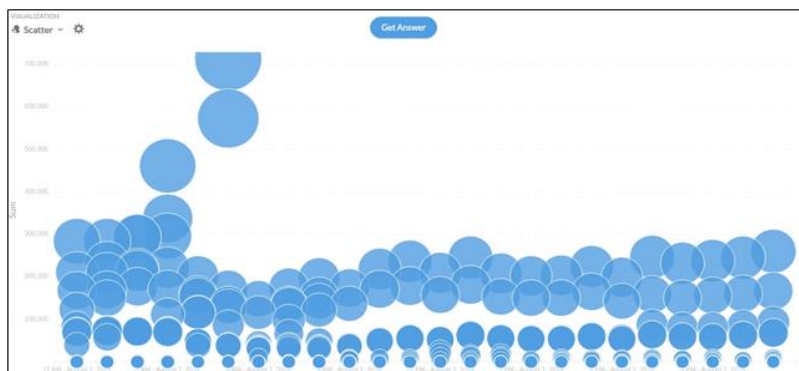


Рис. 5. Візуалізація виконаного запиту у вигляді графіка розподілення

Наведений графік може також відображатись у вигляді звичайної таблиці – в нашому випадку для аналізу даних буде краще використовувати формат звичайної таблиці, щоб розподілити дані як вектори у систему-класифікатор, тобто надати вектори для класифікації до системи k nearest neighbor (рис. 6).

| Timestamp: Hour | 1 | 10 | 11 | 13 | 14 | 15 | 16 | 17 | 18 | 2 | 3 | 5 | 6 | 7 | 8 | 9 |
|------------------------|--------|-----|-----|--------|---------|---------|---------|---------|---------|---------|--------|---------|---------|-----|---------|---------|
| 12 AM - August 1, 2016 | 69,896 | 119 | 328 | 69,568 | 40,310 | 123,500 | 234 | 82,955 | 283,295 | 71,358 | 562 | 211,509 | 119 | 19 | 167,645 | |
| 1 AM - August 1, 2016 | 73,951 | 205 | 341 | 73,610 | 58,494 | 169,000 | 241 | 150,515 | 283,343 | 73,324 | 582 | 209,673 | 205 | 26 | 233,980 | |
| 2 AM - August 1, 2016 | 71,584 | 244 | 426 | 71,158 | 70,609 | 214,500 | 383 | 176,718 | 292,347 | 73,260 | 809 | 218,288 | 244 | 33 | 292,829 | |
| 3 AM - August 1, 2016 | 70,734 | 255 | 458 | 70,276 | 110,654 | 338,000 | 285 | 171,284 | 296,219 | 73,325 | 744 | 222,078 | 255 | 52 | 460,880 | |
| 4 AM - August 1, 2016 | 47,523 | 1 | 158 | 358 | 47,165 | 38,330 | 117,000 | 25 | 115,426 | 201,399 | 46,885 | 383 | 154,101 | 159 | 18 | 159,590 |
| 5 AM - August 1, 2016 | 39,451 | 2 | 205 | 365 | 39,086 | 136,538 | 572,000 | 27 | 86,571 | 170,676 | 38,811 | 392 | 131,381 | 207 | 88 | 713,392 |
| 6 AM - August 1, 2016 | 33,584 | 50 | 221 | 33,363 | 9,802 | 32,500 | 73 | 36,950 | 148,474 | 31,693 | 294 | 116,438 | 50 | 5 | 43,052 | |
| 7 AM - August 1, 2016 | 39,598 | 98 | 280 | 39,318 | 32,058 | 97,500 | 387 | 66,707 | 175,305 | 37,306 | 667 | 137,258 | 98 | 15 | 133,058 | |
| 8 AM - August 1, 2016 | 35,213 | 61 | 222 | 34,991 | 49,441 | 143,000 | 369 | 31,560 | 155,654 | 33,344 | 591 | 121,710 | 61 | 22 | 197,777 | |
| 9 AM - August 1, 2016 | 39,423 | 12 | 197 | 39,226 | 590 | 6,500 | 128 | 8,123 | 173,773 | 37,118 | 325 | 136,372 | 12 | 1 | 7,090 | |
| 10 AM - August 1, 2016 | 50,011 | 11 | 194 | 49,817 | | | | 51 | 8,117 | 219,033 | 48,503 | 245 | 170,382 | 11 | | |
| 11 AM - August 1, 2016 | 55,622 | 6 | 206 | 55,416 | 2,250 | 6,500 | 104 | 2,710 | 235,733 | 56,937 | 310 | 178,557 | 6 | 1 | 9,000 | |

Рис. 6. Візуалізація виконаного запиту у вигляді таблиці

При ввімкненні класифікатора k nearest neighbor ми отримуємо дані такого вигляду (рис. 7):

| Timestamp: Hour | 1 | 10 | 11 | 13 | 14 | 15 | 16 | 17 | 18 | 2 | 3 | 5 | 6 | 7 | 8 | 9 | class |
|------------------------|--------|----|-----|-----|--------|----|----|-----|-----|---------|--------|-----|---------|-----|----|----|-------|
| 12 AM - August 1, 2016 | 44,294 | | 98 | 306 | 44,122 | 17 | 17 | 177 | 79 | 188,319 | 56,961 | 480 | 143,933 | 98 | 17 | 17 | 4 |
| 1 AM - August 1, 2016 | 45,682 | | 164 | 300 | 45,500 | 24 | 24 | 164 | 130 | 186,789 | 58,694 | 462 | 142,051 | 164 | 24 | 24 | 4 |
| 2 AM - August 1, 2016 | 44,998 | | 205 | 385 | 44,771 | 31 | 31 | 272 | 163 | 192,957 | 58,039 | 654 | 147,422 | 205 | 31 | 31 | 3 |
| 3 AM - August 1, 2016 | 40,476 | | 212 | 414 | 40,240 | 49 | 49 | 198 | 166 | 178,138 | 54,583 | 610 | 136,795 | 212 | 49 | 49 | 4 |
| 4 AM - August 1, 2016 | 28,176 | 1 | 144 | 343 | 27,980 | 18 | 18 | 25 | 114 | 124,513 | 35,022 | 367 | 96,765 | 145 | 18 | 18 | 4 |
| 5 AM - August 1, 2016 | 21,290 | 2 | 147 | 304 | 21,120 | 71 | 71 | 26 | 67 | 96,012 | 27,412 | 327 | 75,122 | 148 | 71 | 71 | 1 |
| 6 AM - August 1, 2016 | 17,348 | | 41 | 207 | 17,269 | 5 | 5 | 69 | 33 | 79,224 | 21,013 | 276 | 62,750 | 41 | 5 | 5 | 1 |
| 7 AM - August 1, 2016 | 16,450 | | 72 | 248 | 16,341 | 15 | 15 | 292 | 51 | 74,857 | 19,933 | 537 | 59,229 | 72 | 15 | 15 | 2 |
| 8 AM - August 1, 2016 | 15,893 | | 51 | 202 | 15,827 | 20 | 20 | 286 | 28 | 71,962 | 19,051 | 477 | 56,811 | 51 | 20 | 20 | 2 |
| 9 AM - August 1, 2016 | 21,386 | | 12 | 190 | 21,325 | 1 | 1 | 100 | 9 | 96,599 | 24,637 | 288 | 76,513 | 12 | 1 | 1 | 1 |
| 10 AM - August 1, 2016 | 28,816 | | 9 | 188 | 28,751 | | | 50 | 7 | 129,787 | 34,335 | 238 | 102,124 | 9 | | | 1 |
| 11 AM - August 1, 2016 | 33,918 | | 6 | 204 | 33,857 | 1 | 1 | 94 | 3 | 149,004 | 43,071 | 298 | 115,189 | 6 | 1 | 1 | 2 |

Рис. 7. Візуалізація виконаного запиту у вигляді таблиці з класифікатором

Де тренувальний сет даних для алгоритму (рис. 8) визначений таким чином, дані розподіляються у вигляді векторів параметрів згрупованих по кількості подій за певну годину та їх типу, і кожна з годин може класифікуватися за певними ознаками, що дають користувачам системи (менеджерами) погодинне сигналізування функціонування всієї аналізованої системи:

```
const trainingSet = [
  [73951, 0, 205, 341, 73610, 58494, 169000, 241, 150515, 283343, 73324, 582, 209673, 205, 26, 233980],
  // category 1 - malformed 150515+283343 - 0.53 % of errors for stats
  [71584, 0, 244, 426, 71158, 70609, 214500, 383, 176718, 292347, 73260, 809, 218288, 244, 33,292829 ],
  // category 1 - malformed 176718+292347 - 0.60 % of errors for stats

  [73951, 0, 205, 341, 73610, 58494, 169000, 241, 150515, 283343, 73324, 582, 209673, 205, 26, 233980],
  // category 2 - useless traffic 73324+283343 - 25% of timeouts

  [ 39451, 2, 205, 365, 39086, 136538, 572000, 27, 86571, 170626, 38811, 392, 131381, 207, 88, 713392 ],
  // category 3 - effective wins 88 + 205 - 42 % of impressions

  [ 33584, 0, 50, 221, 33363, 9802, 32500, 73, 36950, 148474, 31693, 294, 116438,50, 5 ,43052 ]
  // category 4 - malformed traffic - to small inv requests due to bid-requests
]
```

Рис. 8. Візуалізація виконаного запиту у вигляді таблиці з класифікатором

Категорії визначені таким чином:

- відношення запитів до помилок більше 50 % – категорія 1;
- відношення запитів до виходу часу – більше 25 %;
- відношення перемоги у аукціонах до фактичних показів реклами – більше 40 %;
- відношення запитів на торги до фактичної кількості запитів за інвентарем менше 40 %.

Висновки та перспективи подальших досліджень. Застосування розроблених моделей і методів прогнозування та аналізу рекламних мереж демонструє значний потенціал для оптимізації ефективності рекламних серверів. Використання сучасних підходів Data Mining дозволяє не тільки покращити якість рекламного націлювання (targeting), але й знизити витрати, асоційовані з ресурсними потребами мережі. Передбачається, що подальше удосконалення алгоритмів і підвищення точності прогнозів може сприяти ранньому виявленню та вирішенню проблем у роботі серверів, підвищуючи таким чином загальну надійність системи. Майбутні дослідження зосередяться на розширенні функціональності аналітичних інструментів, інтеграції з іншими технологіями та поглибленні аналізу поведінкових моделей користувачів, що має призвести до нових відкриттів у галузі рекламних технологій.

Список використаної літератури:

1. Levkivskiy V. Research of algorithms of Data Mining / V.Levkivskiy, N.Lobanchykova, D.Marchuk // The International Conference on Sustainable Futures: Environmental, Technological, Social and Economic Matters. – 2020. – Vol. 166. – pp. 1–6. DOI: 10.1051/e3sconf/202016605007.
2. Праздніков В.О. Моделі та методи машинного навчання для розпізнавання фейкового контенту / В.О. Праздніков, І.І. Сугоняк // Технічна інженерія. – 2023. – № 2 (92). – С. 131–136. DOI: 10.26642/ten-2023-2(92)-131-136.

3. Круціцький В.Я. Оцінка ефективності використання інструментів NLP та систем AI для аналізу рекламних оголошень у системах обміну інтернет-рекламою / В.Я. Круціцький, І.І. Сугоняк // Технічна інженерія. – 2023. – № 1 (91). – С. 161–165. DOI: 10.26642/ten-2023-1(91)-161-165.
4. Shukla M. Extrapolation Using Regression: Challenges and Solutions / M.Shukla, A.Majumdar. – 2020 [Electronic resource]. – Access mode : https://www.mphasis.com/content/dam/mphasis-com/global/en/home/innovation/next-lab/Mphasis_Extrapolation_using_Regression_Challenges_and_Solutions_Whitepaper.pdf.
5. Що таке лінійна регресія / robot_dreams. – 2021 [Електронний ресурс]. – Режим доступу : <https://robotdreams.cc/uk/blog/437-shcho-take-liniyna-regresiya>.
6. FITTING OF A POLYNOMIAL USING LEAST SQUARES METHOD // Neutrium. – 2019 [Electronic resource]. – Access mode : <https://neutrium.net/mathematics/least-squares-fitting-of-a-polynomial/>.
7. Survey of Improving K-Nearest-Neighbor for Classification / J.Liangxiao, C.Zhihua, W.Dianhong, J.Siwei. – 2007 [Electronic resource]. – Access mode : https://www.researchgate.net/publication/4300177_Survey_of_Improving_K-Nearest-Neighbor_for_Classification.
8. Ying L. An improved k-nearest neighbor algorithm and its application to high resolution remote sensing image classification / L.Ying, C.Bo. – 2009. – DOI: 10.1109/GEOINFORMATICS.2009.5293389.
9. Bronstein A. K-Nearest Neighbors Algorithm / A.Bronstein // Medium. – 2019 [Electronic resource]. – Access mode : <https://medium.com/@adi.bronstein/a-quick-introduction-to-k-nearest-neighbors-algorithm-62214cea29c7>.
10. Apache Druid (part 1): A Scalable Timeseries OLAP Database System [Electronic resource]. – Access mode : <https://anskarl.github.io/post/2019/druid-part-1/>.

References:

1. Levkivskiy, V., Lobanchykova, N. and Marchuk, D. (2020), «Research of algorithms of Data Mining», *The International Conference on Sustainable Futures: Environmental, Technological, Social and Economic Matters*, Vol. 166, pp. 1–6, doi: 10.1051/e3sconf/202016605007.
2. Prazdnikov, V.O. and Suhoniak, I.I. (2023), «Modeli ta metody mashynnoho navchannia dlia rozpoznavannia feikovooho kontentu», *Tekhnichna inzheneriia*, No. 2 (92), pp. 131–136, doi: 10.26642/ten-2023-2(92)-131-136.
3. Krutsitskiy, V.Ya. and Suhoniak, I.I. (2023), «Otsinka efektyvnosti vykorystannia instrumentiv NLP ta system AI dlia analizu reklamnykh oholoshen u systemakh obminu internet-reklamoiu», *Tekhnichna inzheneriia*, No. 1 (91), pp. 161–165, doi: 10.26642/ten-2023-1(91)-161-165.
4. Shukla, M. and Majumdar, A. (2020), *Extrapolation Using Regression: Challenges and Solutions*, [Online], available at: https://www.mphasis.com/content/dam/mphasis-com/global/en/home/innovation/next-lab/Mphasis_Extrapolation_using_Regression_Challenges_and_Solutions_Whitepaper.pdf
5. «Shcho take liniina rehresiya» (2021), *robot_dreams*, [Online], available at: <https://robotdreams.cc/uk/blog/437-shcho-take-liniyna-regresiya>
6. «FITTING OF A POLYNOMIAL USING LEAST SQUARES METHOD» (2019), *Neutrium*, [Online], available at: <https://neutrium.net/mathematics/least-squares-fitting-of-a-polynomial/>
7. Liangxiao, J., Zhihua, C., Dianhong, W. and Siwei, J. (2007), «Survey of Improving K-Nearest-Neighbor for Classification», [Online], available at: https://www.researchgate.net/publication/4300177_Survey_of_Improving_K-Nearest-Neighbor_for_Classification
8. Ying, L. and Bo, C. (2009), «An improved k-nearest neighbor algorithm and its application to high resolution remote sensing image classification», doi: 10.1109/GEOINFORMATICS.2009.5293389.
9. Bronstein, A. (2019), «K-Nearest Neighbors Algorithm», *Medium*, [Online], available at: <https://medium.com/@adi.bronstein/a-quick-introduction-to-k-nearest-neighbors-algorithm-62214cea29c7>
10. «Apache Druid (part 1): A Scalable Timeseries OLAP Database System» (2019), [Online], available at: <https://anskarl.github.io/post/2019/druid-part-1/>

Вольський Ростислав Аркадійович – асистент кафедри комп’ютерних наук Державного університету «Житомирська політехніка».

<https://orcid.org/0009-0009-1336-9492>.

Наукові інтереси:

- розподілений пошук, системи зберігання інформації;
- архітектура систем, розподілені системи, системи консенсусу;
- машинне навчання, штучний інтелект, інтелектуальний аналіз даних;
- алгоритми, оптимізація.

Сугоняк Інна Іванівна – кандидат технічних наук, доцент, завідувач кафедри комп’ютерних наук Державного університету «Житомирська політехніка».

<https://orcid.org/0000-0002-0484-4839>.

Наукові інтереси:

- системний аналіз та теорія оптимальних рішень;
- проектування сховищ даних;
- інтелектуальний аналіз даних.

Volskyi R.A., Suhoniak I.I.

Data mining methods for advertising network data analysis

The important and actively researched topic in modern studies of advertising networks is the forecasting and analysis of data, which includes the construction of data retrieval models and the selection of methods for their analysis. The research's central aspect is the use of Data Mining methods for processing large arrays of advertising data. A mathematical model and methods have been proposed that allow for predicting the load on advertising servers and clustering the state of their operation, identifying potential failures and problems at early stages. These models and data processing methods are implemented in management systems, allowing for the optimization of advertising network operations, reducing financial costs, and enhancing the efficiency of advertising resource management. In addition, algorithms for automatic scaling of resources based on predicted data have been developed, ensuring system stability and reliability. The proposed technologies and methods can be used in human resource management and the training of highly qualified IT specialists, providing their appropriate training and development in accordance with the requirements of modern advertising networks.

Keywords: data mining; forecasting; advertising networks; machine learning; polynomial regression.

Стаття надійшла до редакції 02.04.2024.