

УДК 681.518.54

Овчаренко М.А., аспірант спеціальності 126 Інформаційні системи та технології
Науковий керівник: **Каштан В.Ю.,** к.т.н., доцент, доцент кафедри Інформаційних технологій та комп'ютерної інженерії

Іванько А.М., аспірант спеціальності 126 Інформаційні системи та технології
Науковий керівник: **Гнатушенко В.В.,** д.т.н., професор, завідувач кафедри Інформаційних технологій та комп'ютерної інженерії

(Національний технічний університет «Дніпровська політехніка», м. Дніпро, Україна)

ДОСЛІДЖЕННЯ АЛГОРИТМІВ КЛАСТЕРИЗАЦІЇ ДЛЯ АНАЛІЗУ ПОВЕДІНКИ КЛІЄНТІВ У КОНТАКТ-ЦЕНТРАХ

Контакт-центри забезпечують обслуговування клієнтів та управління їхніми запитами. Внаслідок цифрової трансформації відбувається обробка великих обсягів даних, що потребує застосування аналітичних методів для їхнього аналізу. Аналітика великих даних (Big Data Analytics, BDA) сприяє виявленню закономірностей у поведінці клієнтів, що дозволяє формувати ефективні стратегії управління взаємодією з ними [1]. Одним із фундаментальних напрямів аналізу клієнтських даних є сегментація, що передбачає розподіл клієнтів на окремі групи відповідно до спільних характеристик [2]. У контексті контакт-центрів сегментація сприяє персоналізації обслуговування, оптимізації розподілу операторських ресурсів та підвищенню рівня задоволеності клієнтів. Алгоритми машинного навчання забезпечують автоматизацію цього процесу та підвищують точність і адаптивність моделей кластеризації.

Модель сегментації споживачів на основі K-середніх була запропонована в [3] і реалізована за допомогою RStudio. Набір даних у форматі CSV аналізували за допомогою різних функцій R. Візуалізації були розроблені, щоб зрозуміти демографічні дані клієнтів. Для ідентифікації кластерів використовували методику ліктя, метод середнього силуету та статистику розриву. Автори роботи [4] запропонували метод кластеризації ансамблю, який поєднує DBSCAN, K-means, MiniBatch K-means і алгоритм середнього зсуву для сегментації споживачів. Запропонований підхід був протестований на вибірці з 35 000 записів, де досягнуто показника силуету 0,72.

У цій роботі зосереджено увагу на аналізі алгоритмів кластеризації без навчання з учителем для сегментації клієнтів контакт-центру на основі їхніх характеристик, зокрема, з використанням моделі RFM (давність, частота, грошова цінність). Запропоновано провести порівняльний аналіз ефективності таких алгоритмів, як K-середніх, модель суміші Гаусса (GMM), просторова кластеризація додатків із щільністю (DBSCAN), агломеративна кластеризація та кластеризація з використанням ієрархій (BIRCH).

Для реалізації аналізу використовувалися інструменти мови програмування Python, зокрема бібліотеки Pandas (для обробки даних), NumPy (для математичних операцій), StatsModels (для статистичного аналізу), Matplotlib та Seaborn (для візуалізації), Scikit-Learn (для реалізації алгоритмів машинного навчання) і Dear (для еволюційних алгоритмів).

Клієнтські сегменти були класифіковані відповідно до їхнього RFM-рейтингу:

– найкращі клієнти з оцінкою $> 4,5$ – активні та лояльні клієнти, що часто взаємодіють із контакт-центром;

– клієнти високої цінності з оцінкою $> 4,0$ – часті користувачі послуг, які демонструють високий рівень задоволеності;

– клієнти середньої цінності з оцінкою $> 3,0$ – періодично взаємодіють із контакт-центром, можуть потребувати додаткової уваги;

–клієнти з низькою цінністю з оцінкою $> 1,6$ – мають нерегулярну історію взаємодій, можуть втратити інтерес до послуг;

–втрачені клієнти з оцінкою $< 1,6$ – рідко звертаються або повністю припинили взаємодію.

Оцінка продуктивності алгоритмів кластеризації здійснювалася за допомогою Silhouette Score [3], результати наведено в табл.1.

Таблиця 1

Кількісний аналіз алгоритмів кластеризації без нагляду

Алгоритм	Silhouette Score
K-means	0.62
BIRCH	0.62
DBSCAN	0.57
Агломеративна кластеризація	0.62
Модель суміші Гауса (GMM)	0.77

Результати з табл.1 демонструють, що алгоритм GMM перевершує інші методи, оскільки досягає найвищого коефіцієнта силуету (0.77), що свідчить про чіткішу структуру кластерів. Інші алгоритми показали однакові значення (0.62), крім DBSCAN, який дещо поступається (0.57). Це пояснюється чутливістю DBSCAN до вибору параметрів, а також тим, що він краще підходить для аналізу даних зі складною щільністю кластерів.

Проведене в роботі дослідження підтверджує ефективність використання машинного навчання у контакт-центрах для вдосконалення аналітики клієнтів та підвищення рівня обслуговування. Подальші дослідження будуть присвячені на розширення параметрів сегментації та інтеграцію нових факторів, таких як аналіз тону голосу клієнта. Це дозволить покращити точність прогнозування та ефективність обробки даних.

Список використаних джерел:

1. Lekhwar S., Yadav S., Singh A. Big data analytics in retail. In Information and Communication Technology for Intelligent Systems: Proceedings of ICTIS 2018, Volume 2; Springer: Singapore, 2019; pp. 469–477.
2. Gwozdz W., Steensen Nielsen K., Müller T. An environmental perspective on clothing consumption: Consumer segments and their behavioral patterns. Sustainability 2017, 9, 762.
3. Mathew A., Scholar P., Jobin T. Role of Big Data Analysis and Machine Learning in Ecommerce-Customer Segmentation. In Proceedings of the National Conference on Emerging Computer Applications (NCECA), Online, 17 June 2021; p. 189.
4. Hicham N., Karim S. Analysis of Unsupervised Machine Learning Techniques for an Efficient Customer Segmentation using Clustering Ensemble and Spectral Clustering. Int. J. Adv. Comput. Sci. Appl. 2022, 13, 122–130.