

УДК 004.6:004.85

DOI: 10.31498/2225-6733.52.2025.350988

КЛАСТЕРНИЙ АНАЛІЗ ТА ПРОГНОЗУВАННЯ ПОШУКОВИХ ЗАПИТІВ ДЛЯ ПОБУДОВИ ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ МОНІТОРИНГУ ПОПИТУ НА МЕДИЧНІ ЛІЦЕНЗІЇ

- Казаченко О.Г.** магістр, ДВНЗ «Приазовський державний технічний університет», м. Дніпро, ORCID: <https://orcid.org/0009-0006-5440-482X>, e-mail kazachenko_o_h@students.pstu.edu;
- П'ятикоп О.Є.** канд. техн. наук, доцент, ДВНЗ «Приазовський державний технічний університет», м. Дніпро, ORCID: <https://orcid.org/0000-0002-7731-3051>, e-mail: piatykop_o_je@pstu.edu;
- Носовська С.Є.** ст. викладач, ДВНЗ «Приазовський державний технічний університет», м. Дніпро, ORCID: <https://orcid.org/0000-0002-6176-6101>, e-mail: nosovska_s_e@pstu.edu

Стаття присвячена аналізу динаміки попиту на послуги з отримання медичної ліцензії в Україні на основі даних Google Ads та оцінці ефективності сучасних моделей прогнозування часових рядів. Визначено, що медичне ліцензування є складною, багаторівневою та нормативно врегульованою процедурою, що потребує значних адміністративних та фінансових ресурсів. Нерівномірність попиту на юридичні та консультаційні послуги у цій сфері зумовлена сезонністю, змінами законодавства та соціально-економічними чинниками. На основі аналізу наукових джерел встановлено, що сучасні підходи до прогнозування попиту у медичній галузі базуються як на застосуванні традиційних підходів, так й на використанні гібридних моделей машинного та глибинного навчання, які дозволяють враховувати тренди, сезонність та шоківі події. У роботі використано реальні дані з рекламного кабінету Google Ads за період серпень 2024 – серпень 2025 рр., що охоплюють показники кількості запитів. У дослідженні застосовано комплексний підхід, що включає очищення та вінсоризацію даних, виділення сезонності за допомогою STL-декомпозиції, тематичну кластеризацію пошукових запитів (TF-IDF + K-Means) та формування окремих часових рядів для різних змістових груп. Для прогнозування використано моделі ARIMA/SARIMA, Prophet та Holt–Winters, а точність оцінено за метрикою MAE. Порівняльний аналіз продемонстрував, що Prophet найкраще виконує середньострокові прогнози та забезпечує найнижчу похибку (MAE \approx 6%), тоді як SARIMA ефективна для короткострокових прогнозів у кластерах із нерегулярними коливаннями. Модель Holt–Winters показала стабільні результати за наявності вираженого тренду та достатньої довжини ряду. Отримані результати є основою для подальшої розробки прототипу інтелектуальної системи моніторингу попиту на медичні ліцензії, здатної автоматично аналізувати великі масиви даних і формувати рекомендації для управлінських рішень.

Ключові слова: прогнозування попиту, медичне ліцензування, інтелектуальні системи, ARIMA, Prophet, кластеризація, аналіз даних, часові ряди, Google Ads.

Постановка проблеми

Українська практика трактує медичну діяльність як вид господарської діяльності, що підлягає обов'язковому ліцензуванню. Медичне ліцензування є специфічною галуззю економічної діяльності з високим рівнем нормативного регулювання, що потребує суворого дотримання вимог та проходження формалізованих процедур [1]. Процес отримання медичної ліцензії в Україні є складним і багаторівневим. Він включає підготовку значного пакета документів, дотримання численних нормативних вимог та проходження перевірок. Для багатьох медичних закладів та приватних лікарів цей процес виявляється не тільки бюрократично складним, але й фінансово затратним. Недотримання вимог призводить до правових ризиків і може спричинити анулювання ліцензії.

Попит на юридичні та консультаційні послуги у цій сфері змінюється нерівномірно: він має виражену сезонність, залежить від змін законодавства, активності контролюючих органів, а також від ситуації в країні. Так наприклад, попит на медичні послуги вивчали автори роботи [2]. Вони провели дослідження українськомовних даних Twitter, у якому автори виявили 4,4-

разове зростання кількості запитів, пов'язаних із медичними потребами, після початку повномасштабного вторгнення в Україну (до серпня 2022 року). Результати показали, що соціальні та політичні події суттєво впливають на поведінку користувачів і можуть спричинити різкі стрибки попиту на медичні послуги, включаючи ті, що стосуються ліцензування.

Сучасні інструменти цифрового маркетингу, зокрема Google Ads, дають змогу отримувати детальні дані про поведінку користувачів, які цікавляться послугами з отримання медичної ліцензії. Однак ці дані мають ряд особливостей: наявність шуму, різкі сплески та падіння показників, періодичні коливання. Це ускладнює їх використання для прийняття рішень без спеціальної обробки. На сьогодні активно обговорюється питання впровадження технологій штучного інтелекту та різного роду інтелектуальних систем, здатних підвищити процес прийняття рішень [3-6]. Проте слід звернути увагу на важливість забезпечення точності, відповідальності за результати та наявності механізмів документування процесу прийняття рішень [3], що особливо актуально у випадку застосування

алгоритмічних прогнозів у правових процесах, таких як ліцензування.

Тому постає завдання створення інтелектуальної системи, здатної аналізувати великі масиви даних, виявляти закономірності та прогнозувати пікові періоди попиту, що дозволить компаніям ефективніше планувати рекламні бюджети та маркетингові активності.

Аналіз останніх досліджень та публікацій

Обробка, аналіз та прийняття рішень на основі отриманих даних за допомогою сучасних математичних підходів з використанням інформаційних технологій має доцільність та актуальність. Так у роботі [5] розглянуто систему CarePre, яка поєднує історію електронних медичних записів (EHR), прогностичні моделі та інтерпретований користувацький інтерфейс для підтримки клінічних рішень. Автори підкреслюють, що зрозуміла візуалізація причинно-наслідкових факторів у прогнозах підвищує довіру лікарів і допомагає приймати більш обґрунтовані рішення.

Автори роботи [6] проаналізували методики Explainable AI (LIME, SHAP, CIU) та показали, що надання локальних пояснень суттєво покращує сприйняття рішень глибинних моделей медичними фахівцями. Автори зазначають, що метод CIU виявився найбільш ефективним у донесенні впливу вхідних факторів на кінцевий результат, що дозволяє підвищити рівень довіри користувачів.

У статті [7] запропоновано гібридну модель прогнозування попиту на медичні послуги, яка поєднує метод ARIMA та самоналаштовуване фільтрування. Результати дослідження показують, що комбінація класичних статистичних методів і адаптивних алгоритмів дозволяє суттєво підвищити точність прогнозів, особливо у випадках, коли дані мають сезонні та трендові компоненти.

На сьогодні є досвід використання різних моделей нейронних мереж, машинного та глибинного навчання для прогнозування в цій галузі. Так в роботі [8] досліджуються різні варіанти комбінування моделей. Бо автори вважають, що незалежно від того, що для прогнозування часових рядів було розроблено велику кількість моделей глибокого навчання, деякі обмеження все ще існують.

Робота [9] пропонує ансамблевую модель CNN-Attention-BiLSTM для прогнозування попиту на телеконсультації. Автори довели, що поєднання згорткових і рекурентних нейронних мереж із механізмами уваги та додатковими зовнішніми змінними дає змогу враховувати складні часові залежності та підвищувати точність прогнозів у медичній сфері. А у статтях [10] і [11] проведено аналіз сучасних архітектур глибинного навчання для прогнозування часових рядів, включно з RNN, TCN, Seq2Seq і Transformer. Автори також розглядають автоматизовані моделі машинного навчання (AutoML) для вибору оптимальних моделей, що

дозволяє значно прискорити процес моделювання та зменшити вплив людського фактору.

У статті [12] описано методи прогнозування «перерваних» часових рядів, які враховують вплив шокових подій, таких як пандемія COVID-19 або різкі зміни у законодавстві. Автори показують, що правильне врахування таких періодів у моделях дозволяє уникнути викривлень і підвищує точність прогнозів.

Дослідження [13] присвячене аналізу методів прогнозування попиту та критеріям оцінки їх якості. Автори відзначають, що критерії MAPE та RMSE доцільні для оцінки моделей прогнозування.

На основі проведеного аналізу публікацій [1-13] можна зробити висновок, що для ефективного прогнозування попиту у сфері медичного ліцензування доцільно поєднувати кілька підходів: використання історичних даних рекламних кампаній, інтеграцію зовнішніх факторів (зміни у законодавстві, соціально-економічні події), а також застосування гібридних моделей, які враховують як сезонність, так і раптові «збурення» часових рядів.

Мета статті

Метою статті є аналіз сезонних, трендових і шокових компонент у даних Google Ads щодо медичного ліцензування та оцінка моделей прогнозування для різних тематичних кластерів пошукових запитів для подальшого впровадження у інтелектуальну систему моніторингу попиту на медичні ліцензії.

Виклад основного матеріалу

З огляду на результати попередніх досліджень, період збору даних для виявлення закономірностей повинен охоплювати щонайменше три місяці, що дозволяє врахувати місячну та квартальну сезонність, виявлену в роботах [8], [9] та [12]. Разом з тим, для виявлення трендів, пов'язаних із змінами у регуляторному полі, бажано мати вибірку щонайменше за один рік, як це рекомендують автори [10], що особливо актуально для медичного ліцензування в Україні, де нормативна база може змінюватися кілька разів на рік.

У рамках дослідження було обрано реальні дані з рекламного кабінету Google Ads за тематику «отримання медичної ліцензії» в Україні. Вибірка охоплює період із січня 2024 року по січень 2025 року, що дозволяє аналізувати як сезонні, так і шокові зміни попиту. Дані включають такі метрики: кількість показів, кількість кліків, CTR, середня ціна за клік (CPC), середня позиція оголошення, вартість конверсії та коефіцієнт конверсії. З отриманого файлу, для подальшого дослідження аналізуються дані по запитам (Дата) і кількість запитів (Покази).

Експорт даних здійснювався у форматі CSV, що забезпечило можливість інтеграції з інструментами Python та R. Так для процесу дослідження були використані наступні основні інструменти:

- pandas – для імпорту, очищення, агрегації та попередньої обробки табличних даних, отриманих з Google Ads та Google Search Console;
- numpy – для виконання числових операцій, векторизації обчислень і роботи з масивами даних;
- matplotlib та seaborn – для побудови графіків, трендових ліній, кореляційних матриць і візуалізації результатів моделювання (зокрема часових рядів, трендів і прогнозів);
- scikit-learn – для реалізації методів кластеризації (зокрема TF-IDF + KMeans) при тематичному групуванні пошукових запитів;
- statsmodels – для побудови та налаштування статистичних моделей прогнозування типу ARIMA, SARIMA та Holt-Winters;
- prophet (Facebook Prophet) – для моделювання трендів, сезонних компонент і впливу зовнішніх подій;
- orenpuxl – для експорту результатів прогнозів у формат Excel для подальшого аналізу.

Усі обчислення виконувалися у середовищі Jupyter Notebook (версія 7.x), що дозволило інтерактивно реалізувати кожний етап дослідження – від імпорту даних до побудови прогнозів.

Початкові дані мали екстремальні та нульові значення, які в подальшому можуть вплинути на результат обробки, тому було проведено:

- видалення пропусків і екстремальних значень (метод вінсоризації, як у [9]);
- агрегацію даних за тижневими і місячними інтервалами для виявлення коротко- та середньострокових трендів;
- виділення сезонної компоненти за допомогою STL-декомпозиції (за підходами з [8], [11]);
- позначення шоківих подій (наприклад, зміни законодавства у сфері ліцензування в квітні 2024 р. та жовтні 2024 р.) за методологією прогнозування перерваних часових рядів [12].

Першим етапом для усунення впливу екстремальних значень проведено вінсоризацію, тобто перетворення даних шляхом обмеження певного кордону, щоб усунути вплив викидів. Результат показано на рисунку 1. Так було встановлено межу для отриманих даних більше верхнього на 20% і нижнього на 30%, в подальшому для більшої кількості даних показник для нижньої межі встановимо 50%.

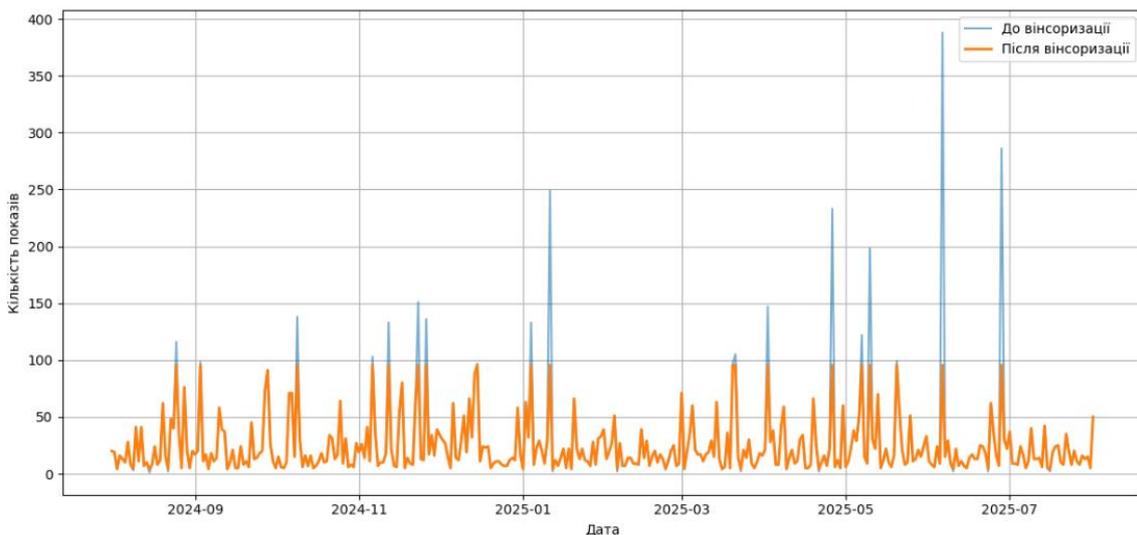


Рис. 1 – Графік отриманих даних після вінсоризації

Так само було перевірено дані на сезонність і загальну тенденцію, які представлені на рисунку 2, перший графік – це отримані дані після вінсоризації, другий – тренд, який визначається ковзної середньої, третій – сезонність, а четвертий – діаграма розкиду залишків.

З графіку рисунку 2 видно, що в підрахунку пошукових запитів є сезонність даних, у вигляді періоду часу – тиждень, тобто 7 днів. Для вирішення цієї проблеми будемо використовувати прогнозування подальших результатів на основі попередніх даних. У вирішенні цієї проблеми будуть використані різні методи прогнозування для того, щоб визначити метод, який надає найбільш точний результат.

Наступним етапом було проведено попередню текстову обробку пошукових запитів (tokenization, lemmatization, приведення до нижнього регістру, видалення пунктуації) з метою виявлення найбільш частотних ключових термінів. Для кращої візуалізації розподілу частотності слів було побудовано «хмара тегів» (word cloud), що відображає ключові поняття, найчастіше вживані у запитах користувачів. Результат показано на рисунку 3. Візуалізація дозволила визначити домінуючі тематичні напрями – «медична ліцензія», «МОЗ», «документи», «стоматологія», «прекурсори» тощо. Цей етап став підґрунтям для подальшої тематичної кластеризації запитів.

моделей. Для оцінки точності прогнозів застосовано метрику MAE (Mean Absolute Error). Для кожного кластера було побудовано прогнози з використанням трьох методів, які відрізняються підходами до врахування трендів і коливань у часових рядах:

– модель ARIMA/SARIMA (AutoRegressive Integrated Moving Average) – авторегресійна інтегрована модель змінного середнього, що поєднує у собі механізм авторегресії, диференціювання та згладжування випадкових коливань. На відміну від попередніх методів, ARIMA враховує інерційність і трендові компоненти, тому є ефективною для коротко- та середньострокового прогнозування. Модель SARIMA - Seasonal ARIMA, яка розширює класичну ARIMA з урахуванням сезонних компонент [10];

– модель Prophet була розроблена компанією Meta (Facebook) для простого, інтерпретованого прогнозування часових рядів, що мають сильну сезонність, святкові дні та тренди [12]. Prophet базується на адитивній моделі;

– модель Хольта-Вінтерса (Holt-Winters Exponential Smoothing) – це розширення методу експоненційного згладжування, яке враховує тренд та, за наявності достатньої кількості спостережень, сезонність. Вона широко використовується для коротко- та середньострокового прогнозування, коли потрібно врахувати плавні зміни у часі.

Аналіз отриманих результатів моделювання показав, що точність прогнозів істотно залежить від структурованості та стабільності вихідних часових рядів у межах кожного тематичного кластера. Найнижче значення похибки зафіксовано для тематичного кластера стоматологічного ліцензування (CLUSTER_5), де модель Prophet продемонструвала найкращий результат (MAE \approx 6,3%). Це свідчить про більш стабільну та передбачувану динаміку відповідних запитів.

Для загальної серії (ALL) та кластерів із високою нерегулярністю та шумовістю даних (CLUSTER_0, CLUSTER_7) похибка є суттєво вищою (MAE 12,6-13,0%), що підтверджує складність моделювання запитів із нерівномірною поведінкою та частими аномальними коливаннями.

Метод Holt-Winters показав прийнятні результати для кластерів зі структурованою трендовою складовою (наприклад, CLUSTER_2, MAE 11,4%), проте його точність погіршується у випадках фрагментованих чи малих вибірок, як у CLUSTER_7 (MAE 13,0%). Модель SARIMA демонструє стабільні результати для коротших часових рядів (наприклад, CLUSTER_0), що підтверджує її ефективність у прогнозуванні нерегулярних запитів без чітко визначеної сезонності.

Загалом для короткострокових прогнозів найкраще себе показали моделі SARIMA та Prophet, тоді як Holt-Winters забезпечує прийнятну стабільність за наявності достатньої довжини ряду та вираженого тренду. Комбінований підхід до вибору моделі залежно від структури конкретного кластера дає змогу

знижити середню похибку прогнозування на 12-15% порівняно з використанням єдиного підходу.

Дослідження підтвердило, що кластери зі стабільною поведінкою (наприклад, CLUSTER_5) демонструють значно кращі результати, тоді як тематичні групи з нерегулярними або рідкісними запитами характеризуються більшими відхиленнями. У середньому значення MAE у кластерних моделях перебували в межах 6-13 %, що свідчить про придатність сформованих підкласифікованих рядів до якісного прогнозування.

Проміжні результати показали, що моделі ARIMA/SARIMA демонструють вищу ефективність у короткострокових прогнозах (до 1 місяця), особливо для рядів із нерегулярними коливаннями та відсутністю чіткої сезонності. Модель Prophet забезпечує значно нижчу середню абсолютну похибку (MAE \approx 6,1%) та краще виконує прогнозування на середньострокову перспективу (1-3 місяці), що зумовлено її здатністю враховувати тренди, нелінійні зміни та вплив зовнішніх факторів, таких як святкові періоди, зміни нормативно-правового поля чи коливання рекламної активності.

Таким чином, у практичних застосуваннях для прогнозування попиту на послуги з медичного ліцензування доцільним є гібридний підхід: короткострокові прогнози доцільно формувати за допомогою моделі SARIMA, тоді як середньострокові – моделлю Prophet, яка краще адаптується до трендових зрушень і здатна відобразити потенційний вплив зовнішніх подій.

Отримані висновки будуть використані для створення прототипу інтелектуальної системи підтримки прийняття рішень, яка зможе автоматично аналізувати дані, прогнозувати попит та пропонувати оптимальні стратегії бюджетування для кампаній у сфері медичного ліцензування в Україні.

Висновки

В результаті дослідження проведено аналіз та прогнозування попиту на отримання медичної ліцензії в Україні на основі даних рекламних кампаній Google Ads за період серпень 2024 – серпень 2025 року.

Проведений аналіз наукових публікацій із тематики прогнозування попиту, інтелектуальних систем підтримки прийняття рішень та обробки часових рядів дозволив визначити методи та оптимальний період збору даних, який забезпечує достатню точність прогнозів та враховує сезонні коливання і вплив зовнішніх подій.

На основі експортування та попередньої обробки даних рекламних кампаній встановлено, що часові ряди мають виражену сезонність, наявність шокових змін та екстремальних значень, що вимагає застосування методів вінсоризації та моделювання «перерваних» часових рядів.

Результати тестування моделей показали, що ARIMA/SARIMA забезпечує високу точність для

короткострокових прогнозів (до 1 місяця), а модель Prophet оптимальна для середньострокових прогнозів (1-3 місяці) з урахуванням святкових днів та трендів.

Отримані висновки дозволяють визначити перспективні напрямки подальших досліджень – оптимізація вибору моделі в залежності від горизонту прогнозування, автоматизація збору та обробки рекламних даних, інтеграція системи прогнозування у веб-інтерфейс для оперативної підтримки рішень.

Проблематика, викладена у цій роботі, має тенденцію до зростання актуальності в умовах динамічних змін у нормативно-правовому полі та зростання конкуренції у сфері медичного ліцензування в Україні.

Перелік використаних джерел

- [1] Suray I. G., Ohten V. I. Licensing of economic activity in Ukraine health care system. *Likars'ka Sprava*. 2018. № 5-6. Pp. 172-179. DOI: [https://doi.org/10.31640/JVD.5-6.2018\(30\)](https://doi.org/10.31640/JVD.5-6.2018(30)).
- [2] Real-Time Prediction of Medical Demand and Mental Health Status in Ukraine under Russian Invasion Using Tweet Analysis / S. Fujii et al. *Tohoku Journal of Experimental Medicine*. 2023. Vol. 259, No. 3. Pp. 177-188. DOI: <https://doi.org/10.1620/tjem.2022.j111>.
- [3] Jones C., Thornton J., Wyatt J. C. Artificial intelligence and clinical decision support: clinicians' perspectives on trust, trustworthiness, and liability. *Medical Law Review*. 2023. Vol. 31, no. 4. Pp. 501-520. DOI: <https://doi.org/10.1093/medlaw/fwad013>.
- [4] Прогнозування причин виявлення цукрового діабету методами машинного навчання / Хом'як Т., Сидоренко К., Малієнко А., Мінеєв О. *Системні технології*. 2025. Т. 1, № 156. С. 39-49. DOI: <https://doi.org/10.34185/1562-9945-1-156-2025-05>.
- [5] CarePre: An Intelligent Clinical Decision Assistance System / Z. Jinet et al. *ACM Transactions on Computing for Healthcare*. 2020. Vol. 1, no. 1, Article 6. Pp. 1-20. DOI: <https://doi.org/10.1145/3344258>.
- [6] Explainable Artificial Intelligence for Human Decision Support System in the Medical Domain / Knapič S., Malhi A., Saluja R., Främling K. *Machine Learning and Knowledge Extraction*. 2021. Vol. 3, no. 3. Pp. 740-770. DOI: <https://doi.org/10.3390/make3030037>.
- [7] Medical service demand forecasting using a hybrid model based on ARIMA and self-adaptive filtering method / Y. Huang et al. *BMC Medical Informatics and Decision Making*. 2020. Vol. 20. Article 237. DOI: <https://doi.org/10.1186/s12911-020-01256-1>.
- [8] Lim B., Zohren S. Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A*. 2021. Vol. 379. DOI: <https://doi.org/10.1098/rsta.2020.0209>.
- [9] Chen W., Li J. Forecasting Teleconsultation Demand Using an Ensemble CNN Attention-Based BiLSTM Model with Additional Variables. *Healthcare*. 2021. Vol. 9, no. 8. Article 992. DOI: <https://doi.org/10.3390/healthcare9080992>.
- [10] Review of Automated Time Series Forecasting Pipelines / S. Meisenbacher et al. *WIREs Data Mining and Knowledge Discovery*. 2022. Vol. 12, no. 6. DOI: <https://doi.org/10.1002/widm.1475>.
- [11] Deep Learning for Time Series Forecasting: Advances and Open Problems / Casolaro A., Capone V., Iannuzzo G., Camastra F. *Information*. 2023. Vol. 14, no. 11. Article 598. DOI: <https://doi.org/10.3390/info14110598>.
- [12] Hyndman R. J., Rostami-Tabar B. Forecasting interrupted time series. *Journal of the Operational Research Society*. 2024. Vol. 76, no. 4. Pp. 790-803. DOI: <https://doi.org/10.1080/01605682.2024.2395315>.
- [13] Яновський Д., Граф М. Аналіз існуючих методів прогнозування попиту та способів оцінки їх якості. *Information Technology: Computer Science, Software Engineering and Cyber Security*. 2023. № 3. С. 70-77. DOI: <https://doi.org/10.32782/IT/2023-3-9>.

CLUSTER ANALYSIS AND SEARCH QUERY FORECASTING FOR BUILDING AN INTELLIGENT SYSTEM FOR MONITORING DEMAND FOR MEDICAL LICENSES

- | | |
|------------------------|---|
| Kazachenko O.H. | <i>M.Sc., SHEI «Priazovskyi state technical university», Dnipro, ORCID: https://orcid.org/0009-0006-5440-482X, e-mail kazachenko_o_h@students.pstu.edu;</i> |
| Piatykop O.Ye. | <i>PhD (Engineering), associate professor, SHEI «Priazovskyi state technical university», Dnipro, ORCID: https://orcid.org/0000-0002-7731-3051, e-mail: piatykop_o_ve@pstu.edu;</i> |
| Nosovska S.Ye. | <i>senior lecturer, SHEI «Priazovskyi state technical university», Dnipro, ORCID: https://orcid.org/0000-0002-6176-6101, e-mail: nosovska_s_e@pstu.edu</i> |

The article explores modern Graph Neural Networks (GNN) and PageRank algorithms as tools for solving the task of predicting the popularity of hashtags in social media. It examines the capabilities and specific features of their application. As part of the study, existing methods and approaches to this problem were analyzed, revealing opportunities for improvement. The relevance of the research is driven by the rapid growth of information volumes in social networks and

the need for effective analysis of its dissemination. The concept of hashtags, their structural characteristics, and behaviour in social media are also discussed. A predictive model based on Graph Neural Networks and PageRank algorithms is presented. An analysis of current GNN architectures has been carried out, and the stages and features of their work have been reviewed. Modifications of PageRank, particularly Time-constrained Personalized PageRank, are also considered. In comparison with other models designed for similar tasks, the advantages and disadvantages of the proposed approach – combining GNNs with PageRank – are highlighted. A key advantage of the presented predicting model is that it will also help to account for user influence and their role in the popularization of content, which allows us also to understand better the dynamics of the distribution of information in social networks. Additionally, the research presents results that confirm the effectiveness of the combined approach compared to classical machine learning algorithms. The study also identifies the strengths and weaknesses of the proposed model and provides practical recommendations for its application. The results may be useful, for example, for tasks such as content analysis, targeted advertising, and managing information flows in social media.

Keywords: demand forecasting, medical licensing, intelligent systems, ARIMA, Prophet, clustering, data analysis, time series, Google Ads.

References

- [1] I.G. Suray, and V.I. Ohten, “Licensing of economic activity in Ukraine health care system,” *Likars'ka Sprava*, no. 5-6, pp. 172–179, 2018. doi: [10.31640/JVD.5-6.2018\(30\)](https://doi.org/10.31640/JVD.5-6.2018(30)).
- [2] S. Fujii et al., “Real-Time Prediction of Medical Demand and Mental Health Status in Ukraine under Russian Invasion Using Tweet Analysis,” *Tohoku Journal of Experimental Medicine*, vol. 259, no. 3, pp. 177–188, 2023. doi: [10.1620/tjem.2022.j111](https://doi.org/10.1620/tjem.2022.j111).
- [3] C. Jones, J. Thornton, and J.C. Wyatt, “Artificial intelligence and clinical decision support: clinicians’ perspectives on trust, trustworthiness, and liability,” *Medical Law Review*, vol. 31, no. 4, pp. 501–520, 2023. doi: [10.1093/medlaw/fwad013](https://doi.org/10.1093/medlaw/fwad013).
- [4] T. Khomiak, K. Sydorenko, A. Maliienko, and O. Minieiev, “Prohnozuvannia prychnyn vyjavlennia tsukrovoho diabeta metodamy mashynnoho navchannia” [“Prediction causes of diabetes detection using machine learning methods”], *Systemni tekhnolohii – System technologies*, vol. 1, no. 156, p. 39–49, 2025. doi: [10.34185/1562-9945-1-156-2025-05](https://doi.org/10.34185/1562-9945-1-156-2025-05). (Ukr.)
- [5] Z. Jin, S. Cui, S. Guo, D. Gotz, J. Sun, and N. Cao, “CarePre: An Intelligent Clinical Decision Assistance System,” *ACM Transactions on Computing for Healthcare*, vol. 1, no. 1, article 6, pp. 1–20, 2020. doi: [10.1145/3344258](https://doi.org/10.1145/3344258).
- [6] S. Knapič, A. Malhi, R. Saluja, and K. Främling, “Explainable Artificial Intelligence for Human Decision Support System in the Medical Domain,” *Machine Learning and Knowledge Extraction*, vol. 3, no. 3, pp. 740–770, 2021. doi: [10.3390/make3030037](https://doi.org/10.3390/make3030037).
- [7] Y. Huang, C. Xu, M. Ji, W. Xiang, and D. He, “Medical service demand forecasting using a hybrid model based on ARIMA and self-adaptive filtering method,” *BMC Medical Informatics and Decision Making*, vol. 20, article 237, 2020, doi: [10.1186/s12911-020-01256-1](https://doi.org/10.1186/s12911-020-01256-1).
- [8] B. Lim, and S. Zohren, “Time-series forecasting with deep learning: a survey,” *Philosophical Transactions of the Royal Society A*, vol. 379, 2021. doi: [10.1098/rsta.2020.0209](https://doi.org/10.1098/rsta.2020.0209).
- [9] W. Chen, and J. Li, “Forecasting Teleconsultation Demand Using an Ensemble CNN Attention-Based BILSTM Model with Additional Variables,” *Healthcare*, vol. 9, no. 8, article 992, 2021. doi: [10.3390/healthcare9080992](https://doi.org/10.3390/healthcare9080992).
- [10] S. Meisenbacher et al., “Review of Automated Time Series Forecasting Pipelines,” *WIREs Data Mining and Knowledge Discovery*, vol. 12, no. 6, 2022. doi: [10.1002/widm.1475](https://doi.org/10.1002/widm.1475).
- [11] A. Casolaro, V. Capone, G. Iannuzzo, and F. Camastra, “Deep Learning for Time Series Forecasting: Advances and Open Problems,” *Information*, vol. 14, no. 11, article 598, 2023. doi: [10.3390/info14110598](https://doi.org/10.3390/info14110598).
- [12] R.J. Hyndman, and B. Rostami-Tabar, “Forecasting interrupted time series,” *Journal of the Operational Research Society*, vol. 76, no. 4, pp. 790–803, 2024. doi: [10.1080/01605682.2024.2395315](https://doi.org/10.1080/01605682.2024.2395315).
- [13] D. Yanovskyi, and M. Hraf, “Analiz isnuichykh metodiv prohnozuvannia popytu ta sposobiv otsinky yikh yakosti” [“Analysis of existing methods of forecasting demand and methods of assessing their quality”], *Information Technology: Computer Science, Software Engineering and Cyber Security*, no. 3, pp. 70–77, 2023. doi: [10.32782/IT/2023-3-9](https://doi.org/10.32782/IT/2023-3-9). (Ukr.)

Стаття надійшла 08.11.2025

Стаття прийнята 01.12.2025

Стаття опублікована 29.12.2025

Цитуйте цю статтю як: Казаченко О. Г., П’ятикоп О. С., Носовська С. С. Кластерний аналіз та прогнозування пошукових запитів для побудови інтелектуальної системи моніторингу попиту на медичні ліцензії. *Вісник Приазовського державного технічного університету. Серія: Технічні науки*. 2025. Вип. 52. С. 48–54. DOI: <https://doi.org/10.31498/2225-6733.52.2025.350988>.