

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА»
ФАКУЛЬТЕТ ІНФОРМАЦІЙНО-КОМП'ЮТЕРНИХ ТЕХНОЛОГІЙ
КАФЕДРА ІНЖЕНЕРІЇ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

ПОЯСНЮВАЛЬНА ЗАПИСКА

до кваліфікаційної роботи освітнього ступеня «магістр»
за спеціальністю 121 «Інженерія програмного забезпечення»
(освітня програма «Інженерія програмного забезпечення»)

на тему:

«Система статистичного аналізу та ETL-трансформації даних електронних
декларацій»

Виконав студент групи ІПЗм-24-1
ПЕТРУК Олександр Олександрович

Керівник роботи:
САВІЦЬКИЙ Роман Святославович

Рецензент:
ЧИЖМОТРЯ Олексій Володимирович

Житомир – 2025

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
 ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА»
 ФАКУЛЬТЕТ ІНФОРМАЦІЙНО-КОМП'ЮТЕРНИХ ТЕХНОЛОГІЙ
 КАФЕДРА ІНЖЕНЕРІЇ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

«ЗАТВЕРДЖУЮ»

зав. кафедри інженерії
 програмного забезпечення

Тетяна ВАКАЛЮК

«18» вересня 2025 р.

ЗАВДАННЯ

на кваліфікаційну роботу

Здобувач вищої освіти: **ПЕТРУК Олександр Олександрович**

Керівник роботи: **САВІЦЬКИЙ Роман Святославович**

Тема роботи: **« Система статистичного аналізу та ETL-трансформації даних електронних декларації»,**

затверджена Наказом закладу вищої освіти від **«18» вересня 2025 р. № 491/с**

Вихідні дані для роботи: **зазначається предмет і об'єкт дослідження**

Консультанти з магістерської кваліфікаційної роботи із зазначенням розділів, що їх стосуються:

Розділ	Консультант	Завдання видав	Завдання прийняв
1	Савіцький Роман	22.09.2025	22.09.2025
2	Савіцький Роман	22.09.2025	22.09.2025
3	Савіцький Роман	22.09.2025	22.09.2025

РЕФЕРАТ

Випускна кваліфікаційна магістерська робота включає аналіз та вибір інструментарію розробки сервісів систематичної аналітики системи декларацій.

Мета роботи полягає в розробці сервісів систематичного аналізу системи декларацій. Результати дослідження можуть використовуватись в подальших наукових розробках систем медичних декларацій.

Пояснювальна записка складається зі вступу, трьох розділів, переліку умовних позначень та списку використаної літератури. Записка дипломної роботи містить 60 ілюстрацій, 25 формул та 41 посилання на джерела бібліографії. Повний обсяг роботи становить 97 сторінок, з них 95 сторінок основного тексту.

Перший розділ присвячений аналізу предметної області, зокрема проблематиці систематичного аналізу в медичних інформаційних системах. Досліджено специфіку організації медичних даних у сервісі декларацій, виявлено ключові виклики: прогнозування навантаження, виявлення похибок, сегментація користувачів та забезпечення їх утримання. Проведено огляд існуючих методів аналізу даних, включаючи прогнозування часових рядів, статистичні методи виявлення аномалій, RFM-аналіз, когортний аналіз, аналіз воронки конверсії та мережевий аналіз. Обґрунтовано вибір технологічного стеку для реалізації системи.

Другий розділ описує проектування та розробку елементів статистичного аналізу. Розроблено гібридний алгоритм прогнозування навантаження, що комбінує експоненціальне згладжування та лінійну регресію. Створено систему автоматичного виявлення похибок на основі паралельного застосування Z-Score та IQR методів. Адаптовано RFM-аналіз для сегментації лікарів за активністю та ефективністю роботи. Імплементовано когортний аналіз для відстеження утримання користувачів, аналіз воронки конверсії для виявлення вузьких місць у процесах, кореляційний та мережевий аналіз для виявлення взаємозв'язків між

метриками. Детально описано реалізацію ETL-трансформації даних з підтримкою експорту в Excel та CSV формати, а також імпорту даних назад у систему.

Третій розділ містить опис програмного продукту та оцінку його ефективності. Представлено комплексний веб-інтерфейс системи аналітики з інтерактивними панелями візуалізації для кожного типу аналізу. Детально описано функціональні можливості трьох ролей користувачів: адміністраторів з повним спектром управлінських інструментів, лікарів зі спеціалізованою панеллю обробки декларацій та пацієнтів з можливістю самостійного управління медичними даними. Проведено тестування системи на наборі даних з однією лікарнею, 15 лікарями, 250 пацієнтами та понад тисячею декларацій, що підтвердило ефективність розроблених алгоритмів та їх відповідність критеріям якості медичних інформаційних систем.

КЛЮЧОВІ СЛОВА: АНАЛІЗ, ПОХИБКА, ETL, ЛІКАР, ПАЦІЄНТ, АДМІНІСТРАТОР, ЛІКАРНЯ, СТАТИСТИКА, ДЕКЛАРАЦІЯ.

ABSTRACT

The master's graduation thesis includes analysis and selection of tools for developing systematic analytics services for a declaration system.

The aim of the work is to develop services for systematic analysis of the declaration system. The research results can be used in further scientific development of medical declaration systems.

The explanatory note consists of an introduction, three chapters, a list of abbreviations and a bibliography. The thesis note contains 60 illustrations, 25 formulas and 41 references to bibliographic sources. The total volume of the work is 97 pages, of which 95 pages are main text.

The first chapter is devoted to the analysis of the subject domain, particularly the issues of systematic analysis in medical information systems. The specifics of organizing medical data in the declaration service were studied, and key challenges were identified: workload forecasting, error detection, user segmentation and ensuring their retention. A review of existing data analysis methods was conducted, including time series forecasting, statistical anomaly detection methods, RFM analysis, cohort analysis, conversion funnel analysis and network analysis. The choice of technology stack for system implementation was justified.

The second chapter describes the design and development of statistical analysis elements. A hybrid workload forecasting algorithm was developed that combines exponential smoothing and linear regression. An automatic error detection system was created based on parallel application of Z-Score and IQR methods. RFM analysis was adapted for segmenting doctors by activity and work efficiency. Cohort analysis was implemented for tracking user retention, conversion funnel analysis for identifying bottlenecks in processes, correlation and network analysis for identifying relationships between metrics. The implementation of ETL data transformation with support for

export to Excel and CSV formats, as well as data import back into the system, was described in detail.

The third chapter contains a description of the software product and evaluation of its effectiveness. A comprehensive web interface of the analytics system with interactive visualization panels for each type of analysis was presented. The functionality of three user roles was described in detail: administrators with a full range of management tools, doctors with a specialized panel for processing declarations, and patients with the ability to independently manage medical data. The system was tested on a dataset with one hospital, 15 doctors, 250 patients and over a thousand declarations, which confirmed the effectiveness of the developed algorithms and their compliance with quality criteria for medical information systems.

KEY WORDS: ANALYSIS, ERROR, ETL, DOCTOR, PATIENT, ADMINISTRATOR, HOSPITAL, STATISTICS, DECLARATION.

Зміст

РЕФЕРАТ	3
ABSTRACT	5
ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ	8
ВСТУП.....	9
РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ	11
1.1 Аналіз проблем і особливостей процесу організації медичних даних у сервісі декларацій.....	11
1.2 Аналіз наявних методів організації та аналізу медичних даних	13
1.3 Вибір засобів реалізації.....	19
Висновки до розділу 1	22
РОЗДІЛ 2 РОЗРОБКА ТА ПРОЄКТУВАННЯ ЕЛЕМЕНТІВ СТАТИСТИЧНОГО АНАЛІЗУ	23
2.1 Розробка алгоритмів систематичного аналізу	23
2.2 Реалізація алгоритму ETL-трансформації даних.....	39
Висновки до розділу 2	43
РОЗДІЛ 3 ОПИС ПРОГРАМНОГО ПРОДУКТУ	46
3.1 Огляд статистичної аналітики	46
3.2 Огляд функціональні можливостей користувачів.....	70
3.3 Оцінка ефективності розробленої системи	83
Висновки до 3 розділу	87
ВИСНОВКИ	89
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	92

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

API – програмний інтерфейс застосунку

CSV – значення, розділені комами

EMA – експоненціальна ковзна середня

ETL – витягування, трансформація, завантаження

IQR – міжквартильний розмах

PDF – переносний формат документів

RFM – Recency, Frequency, Monetary

SQL – Structured Query Language

XLSX – Microsoft Excel Open XML Spreadsheet

ЕСОЗ – Електронна система охорони здоров'я

ЄДРПОУ – Єдиний державний реєстр підприємств та організацій України

МІС – медична інформаційна система

СУБД – система управління базами даних

α – згладжувальний коефіцієнт

μ – середнє арифметичне значення

σ – стандартне відхилення

Σ – символ суми

ВСТУП

Актуальність теми. Цифрова трансформація системи охорони здоров'я України потребує застосування методів статистичного аналізу та ETL-трансформації даних медичних декларацій для прогнозування навантаження на медичні заклади, виявлення похибок, сегментації користувачів та оптимізації процесів обробки.

Застосування методів інтелектуального аналізу даних, зокрема прогнозування часових рядів, статистичного виявлення похибок, RFM-аналізу, когортного аналізу та аналізу воронки конверсії, дозволяє вирішити проблеми та створити ефективну систему підтримки прийняття рішень для керівників медичних закладів.

Мета і завдання дослідження. Дослідження спрямоване на створення та вивчення можливостей використання алгоритмів прогностичної аналітики у системі управління медичними деклараціями.

Основними завданнями є:

1. Дослідити та застосувати методи прогнозування навантаження на систему медичних декларацій з використанням комбінованих підходів на основі експоненціального згладжування та лінійної регресії.
2. Розробити систему автоматичного виявлення похибок в часових рядах медичних даних з використанням статистичних методів для раннього виявлення проблемних ситуацій.
3. Адаптувати та застосувати RFM-аналіз для сегментації лікарів за рівнем активності та ефективності роботи в системі медичних декларацій.
4. Розробити систему когортного аналізу для оцінки рівня утримання користувачів у часі та ідентифікації критичних періодів відтоку.
5. Застосувати аналіз воронки конверсії для виявлення вузьких місць у процесі обробки медичних декларацій.

6. Дослідити застосування кореляційного та мережевого аналізу для виявлення взаємозв'язків між метриками та оптимізації розподілу навантаження між лікарями.

Об'єктом дослідження є процеси обробки електронних медичних декларацій у системах охорони здоров'я України.

Предметом дослідження є методи та алгоритми інтелектуального аналізу даних для прогнозування та виявлення закономірностей і похибок в медичних інформаційних системах.

Методи дослідження. У роботі використано методи аналізу часових рядів, статистичні, методи сегментації, аналіз воронки конверсії, кореляційний аналіз Пірсона та методи мережевого аналізу.

Наукова новизна. Запропоновано комплексну систему статистичного аналізу та ETL-трансформації даних електронних медичних декларацій, що інтегрує гібридні методи прогнозування, автоматизоване виявлення похибок та багаторівневу сегментацію користувачів для підтримки прийняття управлінських рішень у системі охорони здоров'я.

Практичне значення одержаних результатів. Розроблена система забезпечує можливість прогнозування навантаження на медичні заклади з точністю достатньою для планування ресурсів, дозволяє своєчасно виявляти технічні та організаційні проблеми, сегментувати лікарів за ефективністю роботи та оптимізувати процеси обробки декларацій. Результати дослідження можуть бути використані адміністраторами ЕСОЗ, керівниками медичних закладів та органами управління охороною здоров'я для підвищення ефективності функціонування системи електронних медичних декларацій.

Апробація результатів. Основні результати роботи апробовано при розробці аналітичної системи для сервісу електронних медичних декларацій в рамках ЕСОЗ України було написано дві тези [1-2].

РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Аналіз проблем і особливостей процесу організації медичних даних у сервісі декларацій

Цифрова трансформація систем охорони здоров'я супроводжується накопиченням величезних обсягів даних про пацієнтів, лікарів, медичні заклади та їх взаємодію. Загальна кількість електронних медичних записів в центральній базі даних ЕСОЗ станом на жовтень 2024 року сягає майже 4 мільярди записів [3].

Декларація – це документ, який підтверджує вибір лікаря первинної ланки [4]. Електронна медична декларація є ключовим документом, що формалізує відносини між пацієнтом та лікарем первинної медичної допомоги.

Сучасні медичні інформаційні системи в Україні стикаються з рядом аналітичних викликів при організації медичних даних таких як:

1. Проблема прогнозування навантаження.
2. Проблема своєчасного виявлення похибок.
3. Проблема сегментації та диференційованого управління.
4. Проблема утримання користувачів.

Розглянемо та проаналізуємо кожен з цих наведених пунктів детальніше.

Планування ресурсів медичних закладів ускладняється відсутністю достовірних прогнозів обсягу звернень пацієнтів. За результатами досліджень вітчизняних медичних установ, нераціональний розподіл робочого навантаження зумовлений порушенням пацієнтами встановленого маршруту між рівнями медичної допомоги та надмірними витратами робочого часу медичного персоналу на ведення документації. Традиційні методи планування, що базуються на середніх показниках попередніх періодів, не враховують сезонної компоненти, довгострокових трендів та короткострокових флуктуацій у структурі звернень. Методи аналізу та прогнозування часових рядів забезпечують вищу точність передбачення, проте потребують адаптації до специфічних умов функціонування медичної галузі [5].

Системні помилки, технічні несправності, організаційні дефекти та випадки недобросовісного використання системи можуть тривалий час залишатися невиявленими [6].

Користувачі системи характеризуються різними поведінковими патернами взаємодії з платформою. Оптимізація управління передбачає ідентифікацію таких сегментів та застосування диференційованих підходів до роботи з ними. Наукові праці демонструють можливість адаптації методів кластеризації для медичних інформаційних систем, що дозволяє індивідуалізувати стратегії взаємодії з різними категоріями користувачів [5].

Зміна пацієнтами закріпленого лікаря або зниження рівня активності медичних працівників у системі негативно позначається на якості надання медичних послуг та ефективності функціонування закладу. Ідентифікація користувачів із високою ймовірністю відтоку на ранніх етапах створює можливості для своєчасного впровадження превентивних заходів [7].

Специфіка процесу організації медичних даних у сервісі декларацій полягає в тому, що процеси характеризуються високою варіативністю поведінки учасників. Різні медичні заклади мають різну пропускну здатність, лікарі демонструють різну активність обробки декларацій, а пацієнти проявляють різні патерни взаємодії з системою [8]. Крім того, існує часова залежність процесів, оскільки навантаження на систему не є рівномірним – спостерігаються сезонні коливання, піки активності в певні дні тижня та години доби [9].

Практична значущість застосування методів аналізу даних та статистики в медичних інформаційних системах має безпосередній вплив на ефективність системи охорони здоров'я. Перехід від інтуїтивних рішень до рішень, базованих на даних, дозволяє керівникам медичних закладів отримувати об'єктивну інформацію про завантаженість лікарів, якість обробки декларацій, проблемні етапи процесів [5]. Прогнозування майбутнього навантаження на систему дозволить медичним закладам заздалегідь готуватися до пікових періодів,

розподіляти навантаження між лікарями, планувати графіки роботи персоналу. Розуміння патернів взаємодії пацієнтів та лікарів, виявлення вузьких місць у процесах, персоналізований підхід до різних сегментів користувачів сприятиме покращенню досвіду користувачів системи.

1.2 Аналіз наявних методів організації та аналізу медичних даних

Для вирішення визначених проблем організації медичних даних у сервісі декларацій застосовується комплекс аналітичних методів та алгоритмів. Вибір конкретних методів обумовлений характеристиками предметної області та практичними потребами медичних закладів [8].

Аналіз часових рядів є одним з найважливіших інструментів для прогнозування навантаження в системах (Рис. 1.1). Згідно з теорією аналізу часових рядів, прогнозування здійснюється на основі виявлення закономірностей у минулих даних з припущенням, що ці закономірності збережуться в майбутньому [10].

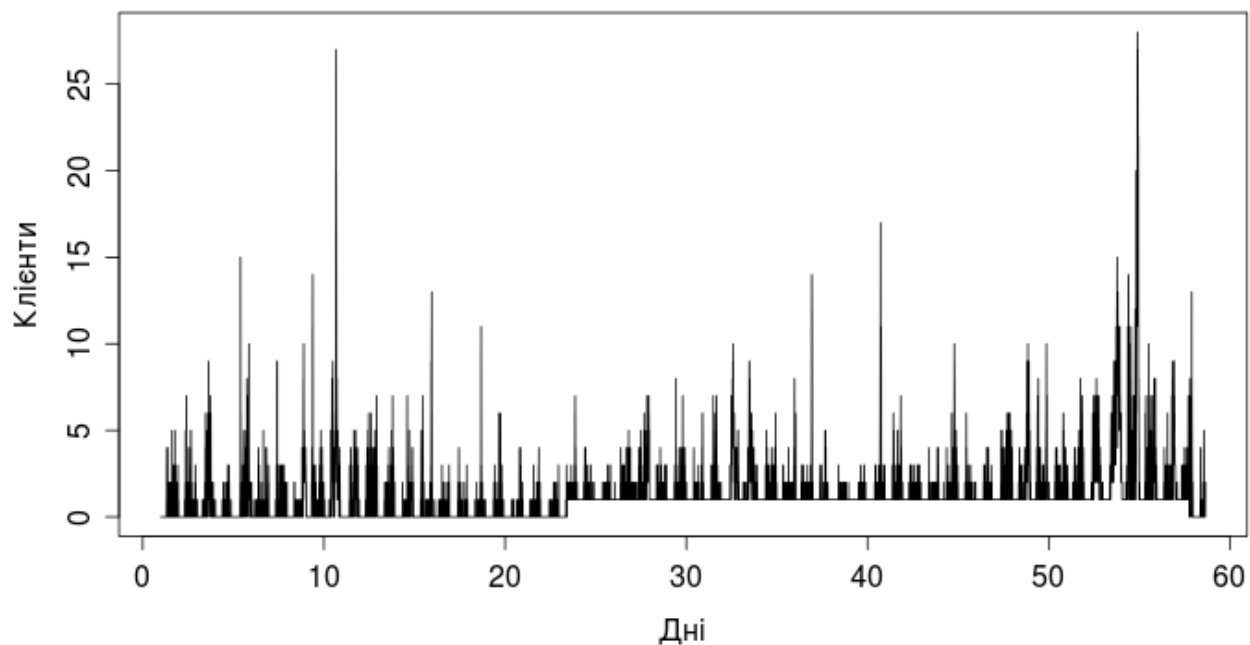


Рисунок 1.1. Приклад аналізу часових рядів

Exponential Moving Average (EMA) – метод експоненціального згладжування, який використовується для прогнозування часових рядів. При розрахунку прогнозу методом експоненціального згладжування враховується відхилення попереднього прогнозу від реального показника. Це популярний метод прогнозування, який був незалежно відкритий Броуном і Холтом [11]. ЕМА являє собою середньозважене значення минулих спостережень, де ваги зменшуються експоненціально. Чим новіше спостереження, тим більшу вагу воно має у розрахунку (Рис. 1.2).

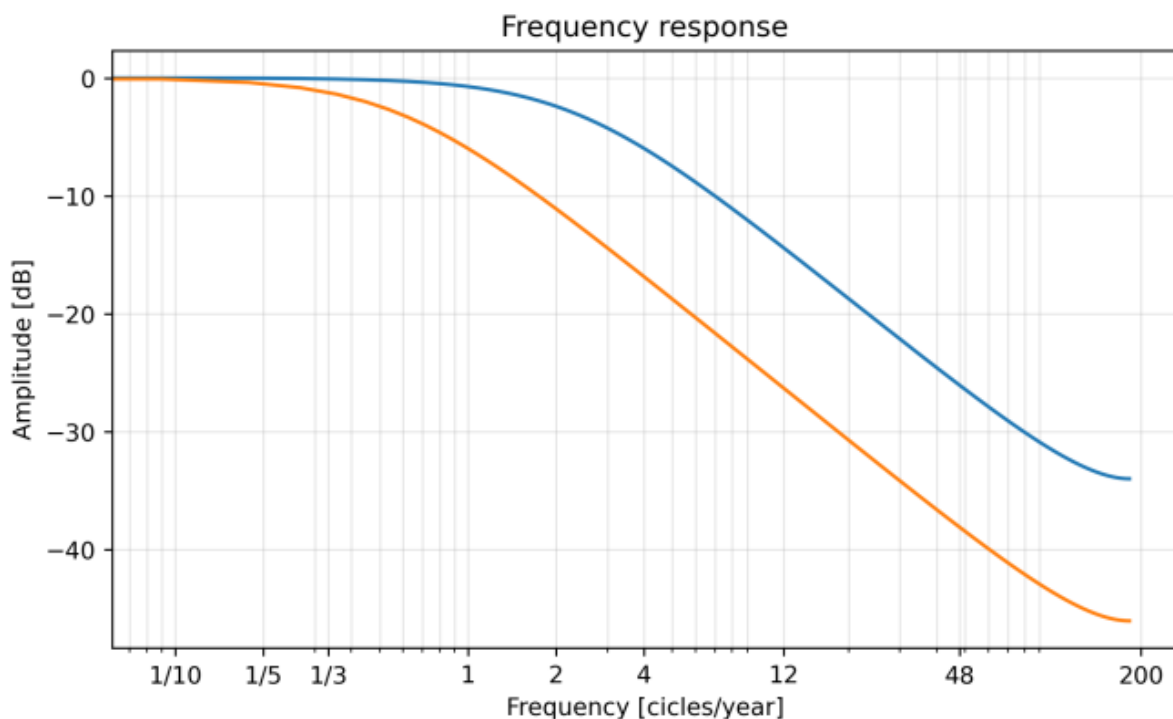


Рисунок 1.2. Приклад аналізу за допомогою методу Exponential Moving Average

Експоненціальне згладжування розраховується за формулою (1.1) [12]:

$$\text{EMA}(\text{today}) = \text{Price}(\text{today}) * \alpha + \text{EMA}(\text{yesterday}) * (1 - \alpha), \quad (1.1)$$

де $\alpha = 2 / (n + 1)$, ЕМА – експоненціальна ковзна середня на сьогодні, Price (today) – актуальна ціна на сьогодні (поточне значення показника), α (альфа) – згладжувальний коефіцієнт, розраховується як $\alpha = 2 / (n+1)$, де n – кількість періодів, ЕМА - експоненціальна ковзна середня за попередній період, $(1 - \alpha)$ –

вага для попереднього значення ЕМА. Дана формула представляє собою рекурсивне співвідношення, в якому поточне значення експоненціальної ковзної середньої обчислюється як зважена комбінація фактичного спостереження та попереднього згладженого значення [12].

Результатом застосування формули ЕМА є адаптивно згладжений часовий ряд, який зберігає інформацію про базовий тренд показника, водночас фільтруючи короткострокові випадкові коливання.

Лінійна регресія (Рис. 1.3) застосовується для виявлення довгострокових трендів у даних [13]. Цей метод дозволяє встановити математичну залежність між незалежними та залежними змінними, що є особливо корисним для аналізу часових рядів та побудови прогностичних моделей у різних галузях, включаючи економіку, фінанси та природничі науки.

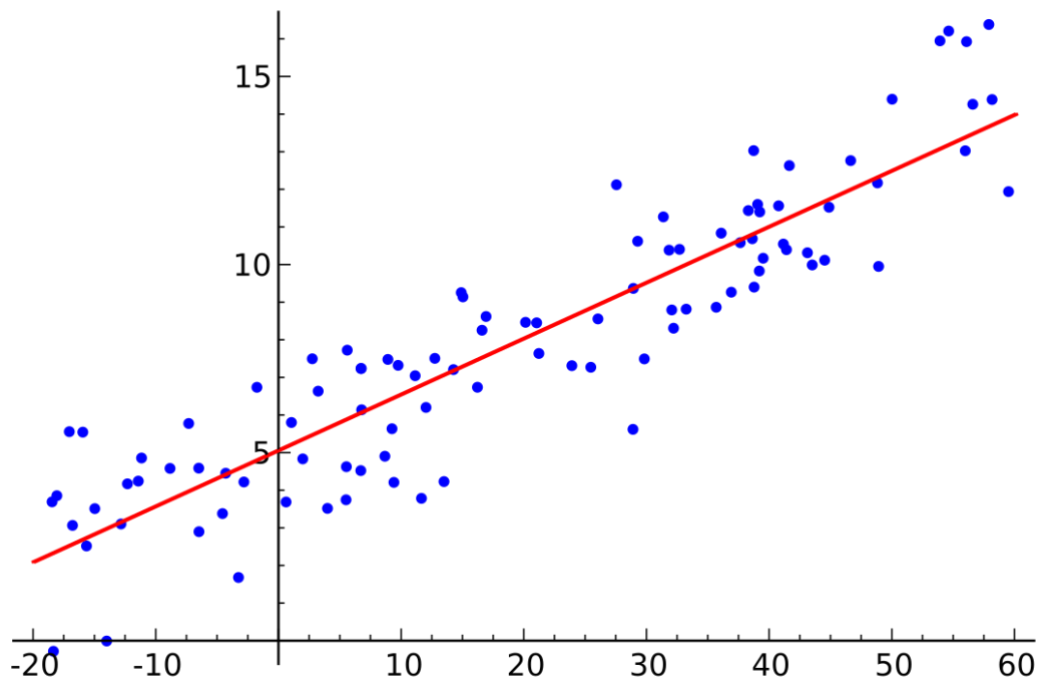


Рисунок 1.3. Приклад лінійної регресії

Метод найменших квадратів дозволяє побудувати лінію тренду, де коефіцієнти визначаються за формулами (1.2-1.3) [13-14].

$$m = \Sigma[(x_i - \bar{x})(y_i - \bar{y})] / \Sigma[(x_i - \bar{x})^2], \quad (1.2)$$

де m – коефіцієнт нахилу прямої регресії, x_i – i -те значення незалежної змінної x , y_i – значення залежної змінної y , $\bar{x}$ (x з ризиком) – середнє арифметичне всіх значень x : $\bar{x} = (x_1 + x_2 + \dots + x_n) / n$, $\bar{y}$ (y з ризиком) – середнє арифметичне всіх значень y : $\bar{y} = (y_1 + y_2 + \dots + y_n) / n$, Σ (сигма) – знак суми, означає підсумовування для всіх точок даних.

$$b = \bar{y} - m \cdot \bar{x}, \quad (1.3)$$

де b – вільний член, точка перетину прямої з віссю Y , $\bar{y}$ – середнє значення y , m – коефіцієнт нахилу (обчислений раніше), $\bar{x}$ – середнє значення x .

Рівняння лінійної регресії має класичну форму, де шукана залежність представляється у вигляді прямої лінії на координатній площині. У цьому рівнянні змінна y представляє залежну змінну або відгук. Незалежна змінна x представляє пояснювальну змінну або предиктор, на основі якого здійснюється прогнозування. Найчастіше в аналізі трендів роль незалежної змінної виконує час, виражений у відповідних одиницях вимірювання. Однак це може бути і будь-яка інша кількісна характеристика, що теоретично впливає на залежну змінну. Коефіцієнт m є кутовим коефіцієнтом або нахилом прямої регресії. Цей параметр визначає швидкість зміни залежної змінної при збільшенні незалежної змінної на одиницю. Позитивне значення коефіцієнта m вказує на зростаючий тренд, негативне – на спадний, а значення близьке до нуля свідчить про відсутність вираженої лінійної залежності. Параметр b є вільним членом або точкою перетину з віссю ординат. Він представляє теоретичне значення залежної змінної при нульовому значенні незалежної змінної. У чисельнику цієї формули знаходиться сума добутків відхилень кожного спостереження від відповідних середніх значень. Символ x_i позначає i -те спостереження незалежної змінної, тоді як $\bar{x}$ представляє середнє арифметичне всіх значень незалежної змінної, обчислене як сума всіх x_i , поділена на кількість спостережень. Таким чином гарантується, що побудована модель є збалансованою відносно фактичних спостережень, мінімізуючи систематичне зміщення прогнозів. Прогнозування на

основі лінійної регресії особливо ефективно для виявлення загальної тенденції зростання або спаду показників [13].

Виявлення похибок відіграє вирішальну роль в аналізі медичних даних для раннього виявлення проблемних ситуацій в системі [15].

Модель Альтмана (Altman model) – алгоритм інтегральної оцінки загрози банкрутства підприємства, заснований на комплексному обліку найважливіших показників, що діагностують кризовий фінансовий стан [16].

Z-Score метод вимірює, наскільки далеко точка даних відхиляється від середнього значення як кратне стандартного відхилення. Великі абсолютні значення Z-score вказують на похибку (Рис 1.4). Формула розрахунку Z-Score (1.4) [17]:

$$z = (x - \mu) / \sigma, \quad (1.4)$$

Де x – значення спостереження, μ – середнє, σ – стандартне відхилення.

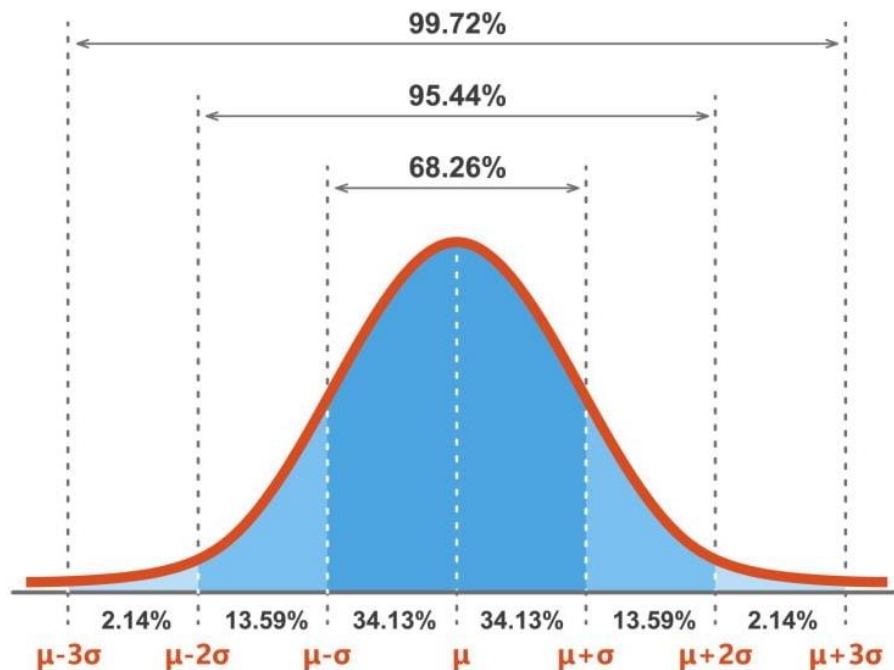


Рисунок 1.4. Приклад результату аналізу з використанням Z-score методу

Міжквартильний розмах (IQR) – це міра статистичного розсіювання, яка описує розкид середніх 50% набору даних. Він обчислюється як різниця між верхнім квартилем (Q3) та нижнім квартилем (Q1) [18]. Це показує діапазон, охоплений середніми 50% даних IQR метод забезпечує надійне виявлення викидів, особливо ефективний для даних зі скошеним розподілом. IQR розраховується як різниця між третім та першим квартилем (1.4) [19]:

$$IQR = Q3 - Q1 \quad (1.4)$$

Викидами вважаються значення, що знаходяться за межами інтервалу (1.5) [19]:

$$[Q1 - 1.5 \cdot IQR, Q3 + 1.5 \cdot IQR] \quad (1.5)$$

Множник 1.5 відповідає приблизно 3 стандартним відхиленням від медіани, що охоплює 99.7% даних при нормальному розподілі [20].

Recency-Frequency-Monetary (RFM) аналіз є технікою сегментації, яка класифікує користувачів згідно з тим, як нещодавно, як часто та з якою інтенсивністю вони взаємодіяли з системою. Дослідження показують ефективність модифікованої RFM моделі в медичній галузі, що враховує обмеження пропускної здатності сервісу [21].

Когортний аналіз дозволяє відстежувати поведінку користувачів шляхом групування їх у когорти – групи, що мають спільні характеристики або початкову точку. Коефіцієнт утримання є ключовою метрикою, яка розраховується як відсоток користувачів у когорті, які продовжують взаємодію після певного періоду [22].

Воронка конверсії є моделлю, що відображає шлях користувача через різні етапи процесу. У медичній галузі воронка зазвичай включає етапи від початкового контакту до довгострокового утримання [23].

Коефіцієнт кореляції Пірсона – це найпоширеніший спосіб вимірювання лінійної кореляції. Це число між -1 і 1 , яке вимірює силу та напрямок зв'язку між двома змінними. Коефіцієнт кореляції Пірсона є описовою статистикою, що

узагальнює характеристики набору даних. Зокрема, він описує силу та напрямок лінійного зв'язку між двома кількісними змінними [24].

Кореляція Пірсона використовується для виявлення лінійних взаємозв'язків між метриками. Коефіцієнт кореляції Пірсона розраховується за формулою (1.6) [25]:

$$R = \Sigma[(x_i - \bar{x})(y_i - \bar{y})] / \sqrt{[\Sigma(x_i - \bar{x})^2 \cdot \Sigma(y_i - \bar{y})^2]}, \quad (1.6)$$

де x_i – окреме значення змінної x , це кожне спостереження з першого набору даних. y_i – окреме значення змінної Y , це відповідне спостереження з другого набору даних. $\bar{x}$ (x з рискою) – середнє арифметичне всіх значень x , $\bar{y}$ (y з рискою) – середнє арифметичне всіх значень y , Σ (сигма) – символ суми, $(x_i - \bar{x})$ – відхилення кожного значення X від середнього X , $(y_i - \bar{y})$ – відхилення кожного значення Y від середнього Y , $(x_i - \bar{x})^2$ – квадрат відхилення X , $(y_i - \bar{y})^2$ – квадрат відхилення Y (завжди позитивне число).

Центральність ступеня є метрикою для аналізу впливовості вузлів у мережі зв'язків. У контексті медичних декларацій це дозволяє ідентифікувати найбільш завантажених та впливових лікарів [26]. Центральність розраховується за формулою (1.7) [27]:

$$C(v) = \deg(v) / (n - 1), \quad (1.7)$$

де $\deg(v)$ – кількість зв'язків вершини, n – загальна кількість вершин у графі. Побудова графа зв'язків лікар-пацієнт дозволяє візуалізувати розподіл навантаження та оптимізувати його між спеціалістами.

1.3 Вибір засобів реалізації

Для реалізації системи аналітики медичних декларацій обрано комплекс алгоритмів та технологій, що забезпечують ефективне вирішення поставлених задач.

Для реалізації системи обрано такий технологічний стек. Серверну частину вирішено реалізувати на NestJS. Це фреймворк для Node.js із вбудованою

підтримкою TypeScript, який забезпечує строгу типізацію та модульну архітектуру. Цей вибір дозволяє створювати масштабовані додатки з чіткою структурою, які легко підтримувати та розвивати.

В якості бази даних обрано PostgreSQL. Ця потужна реляційна СУБД здатна ефективно виконувати складні аналітичні запити. Вона підтримує віконні функції та агрегації, що є критично важливими для реалізації запропонованих алгоритмів аналітики.

Клієнтську частину вирішено побудувати на React. Ця сучасна бібліотека для створення користувацьких інтерфейсів забезпечує швидку розробку інтерактивних панелей для візуалізації аналітичних даних завдяки компонентному підходу.

Обраний технологічний стек забезпечує оптимальний баланс між продуктивністю, масштабованістю та зручністю розробки системи аналітики медичних декларацій. Для прогнозування навантаження на систему обрано гібридний підхід, що поєднує два додаткових методи. Exponential Moving Average (EMA) обраний для відстеження короткострокових коливань та адаптації до останніх змін у даних. Метод забезпечує швидку реакцію на зміни тренду завдяки експоненційним вагам, що надають більшу значущість свіжим даним. Ваговий коефіцієнт α обрано емпірично в діапазоні 0.2-0.3 для балансу між чутливістю та стабільністю прогнозу [8]. Паралельно використовується лінійна регресія для виявлення довгострокових трендів у даних про створення декларацій. Метод найменших квадратів забезпечує оптимальну апроксимацію лінійного тренду, що критично важливо для стратегічного планування ресурсів медичних закладів [28].

Комбінований підхід з динамічними вагами дозволяє одночасно враховувати як короткострокові коливання, так і загальну тенденцію розвитку. Додатково реалізовано розрахунок довірчого інтервалу прогнозу на основі

коефіцієнта варіації історичних даних, що дозволяє оцінювати надійність прогнозу.

Для виявлення похибок обрано паралельне застосування двох статистичних методів таких як Z-Score метод – обраний для виявлення похибок у даних з нормальним або близьким до нормального розподілом.

IQR метод – використовується як додатковий підхід. Метод не залежить від припущень про характер розподілу даних, що робить його універсальним для різних метрик. Комбінація обох методів підвищує надійність виявлення аномальних ситуацій. Система показує помилку лише у випадку, коли похибка виявлена обома методами одночасно, що знижує кількість хибних спрацювань.

В якості методу для сегментації на аналізу поведінки було вирішено обрати RFM-аналіз. RFM-аналіз обрано для сегментації лікарів через його доведену ефективність у класифікації великих обсягів даних. Квантільний розподіл на 5 категорій по кожній метриці забезпечує гранулярність сегментації, а агрегація в 7 основних сегментів робить результати практично застосовними для прийняття управлінських рішень.

Когортний аналіз обрано для відстеження утримання користувачів, оскільки він дозволяє виявити критичні періоди відтоку та оцінити ефективність заходів з утримання. Групування користувачів за місяцем першого контакту забезпечує достатню деталізацію при збереженні репрезентативності когорт.

Воронка конверсії реалізована для відстеження переходів між етапами обробки декларацій. Цей підхід дозволяє точно локалізувати вузькі місця в процесі та оцінювати ефективність кожного етапу через коефіцієнти конверсії.

В якості методів виявлення взаємозв'язків було вирішено обрати Кореляцію Пірсона та мережевий аналіз. Кореляція Пірсона обрана для виявлення лінійних взаємозв'язків між метриками як найпоширеніший та добре інтерпретований метод. Коефіцієнт кореляції дозволяє швидко оцінити силу та напрямок зв'язку між змінними. Мережевий аналіз з використанням метрики центральності

ступеня обрано для візуалізації та аналізу розподілу навантаження між лікарями. Побудова графа зв'язків лікар-пацієнт дозволяє ідентифікувати найбільш завантажених спеціалістів та оптимізувати розподіл пацієнтів.

Висновки до розділу 1

Проведено аналіз проблем організації медичних даних у сервісі електронних декларацій України. Виявлено ключові виклики: необхідність прогнозування навантаження, виявлення похибок, сегментації користувачів, оцінки утримання та оптимізації процесів обробки декларацій.

Проаналізовано наявні методи організації та аналізу медичних даних, включаючи методи прогнозування часових рядів, статистичні методи виявлення похибок, методи сегментації, когортний аналіз, аналіз воронки конверсії, кореляційний та мережевий аналіз.

Обґрунтовано вибір комбінованого підходу до прогнозування навантаження, що поєднує ЕМА та лінійну регресію з динамічними вагами для одночасного врахування короткострокових коливань та довгострокових трендів.

Визначено необхідність паралельного застосування Z-Score та IQR методів для підвищення надійності виявлення похибок у даних з різними характеристиками розподілу.

Обґрунтовано вибір технологічного стеку для реалізації системи аналітики, що забезпечує баланс між продуктивністю, масштабованістю та зручністю розробки.

РОЗДІЛ 2 РОЗРОБКА ТА ПРОЄКТУВАННЯ ЕЛЕМЕНТІВ СТАТИСТИЧНОГО АНАЛІЗУ

2.1 Розробка алгоритмів систематичного аналізу

Для прогнозування навантаження на систему медичних декларацій розроблено гібридний алгоритм, що поєднує переваги експоненціального згладжування та лінійної регресії. Такий підхід дозволяє одночасно враховувати короткострокові коливання та довгострокові тренди в даних про створення декларацій.

Архітектура модуля включає збирач даних, що відповідає за агрегацію історичних даних з бази даних за заданий період, модуль експоненціального згладжування для виявлення короткострокових закономірностей, модуль лінійної регресії для визначення довгострокового тренду, комбінатор прогнозів для об'єднання результатів та калькуляція надійності для оцінки якості прогнозу.

Процес прогнозування розпочинається з ідентифікації релевантного набору даних. Система визначає всі активні лікарні, до яких має доступ адміністратор, та формує список лікарів, що працюють у цих закладах. Історичні дані про створення декларацій агрегуються за добовими інтервалами за заданий період, що забезпечує достатню статистичну базу для аналізу.

Експоненціальне згладжування реалізовано як рекурсивний алгоритм. Початкове значення ЕМА встановлюється рівним першому спостереженню часового ряду (2.1) [29].

$$EMA_0 = y_0 \quad (2.1)$$

Параметр згладжування розраховується за формулою (2.2) [30]

$$\alpha = 2/(N + 1), \quad (2.2)$$

де N – фактична кількість наявних історичних точок даних.

Для кожної наступної точки виконується розрахунок що формує послідовність згладжених значень за формулою (2.3):

$$\text{EMA}_t = \alpha \times y_t + (1 - \alpha) \times \text{EMA}_{t-1}, \quad (2.3)$$

де EMA_t – експоненціальна ковзна середня в поточний момент часу t , α (альфа) – згладжувальний коефіцієнт, де $0 < \alpha \leq 1$. Цей параметр визначає, наскільки швидко ЕМА реагує на нові дані, більше α означає більшу вагу для нових даних, менше α означає більшу вагу для історичних даних, y_t – фактичне спостережуване значення в поточний момент часу t , $(1 - \alpha)$ – вага для попередньої ЕМА, EMA_{t-1} – експоненціальна ковзна середня з попереднього моменту часу.

Прогнозне значення для майбутніх періодів базується на останньому обчисленому значенні ЕМА, яке відображає актуальну тенденцію з урахуванням всіх попередніх спостережень (Рис. 2.1).

```
const alpha = 2 / (dataPoints.length + 1);
let ema = dataPoints[0] || 0;
const emaValues = [ema];

for (let i = 1; i < dataPoints.length; i++) {
    ema = dataPoints[i] * alpha + ema * (1 - alpha);
    emaValues.push(ema);
}
```

Рисунок 2.1. Фрагмент коду реалізації рекурсивного алгоритму

Лінійна регресія застосовується для виявлення довгострокового тренду в історичних даних. Часовий ряд перетворюється у формат, де незалежна змінна x представляє порядковий номер дня від початку спостережень, а залежна змінна y відображає кількість створених декларацій. Обчислюються середні арифметичні значення $\bar{x}$ та $\bar{y}$. Коефіцієнт нахилу визначається за формулою (2.4) [31]:

$$m = \Sigma[(x_i - \bar{x})(y_i - \bar{y})] / \Sigma[(x_i - \bar{x})^2], \quad (2.4)$$

де m – коефіцієнт нахилу регресійної прямої. Показує, на скільки одиниць змінюється y при зміні x на одиницю, Σ – символ суми, означає, що потрібно просумувати всі значення від $i = 1$ до n , x_i – i -те значення змінної x , y_i – i -те

значення змінної y , $\bar{x}$ (х з рискою) – середнє арифметичне всіх значень, $\bar{y}$ (у з рискою) – середнє арифметичне всіх значень.

Вільний член розраховується за формулою (2.5) [32]:

$$b = \bar{y} - m \times \bar{x}, \quad (2.5)$$

де b – вільний член або y -intercept Це значення y , коли $x = 0$. Показує базовий рівень залежної змінної, $\bar{y}$ (у з рискою) – середнє арифметичне всіх значень y , m – коефіцієнт нахилу, який ми розраховували у попередній формулі. Показує крутизну лінії регресії, $\bar{x}$ (х з рискою) – середнє арифметичне всіх значень x (Рис. 2.2).

```
const xMean = xValues.reduce((a, b) => a + b, 0) / n;
const yMean = yValues.reduce((a, b) => a + b, 0) / n;

let numerator = 0;
let denominator = 0;

for (let i = 0; i < n; i++) {
  numerator += (xValues[i] - xMean) * (yValues[i] - yMean);
  denominator += (xValues[i] - xMean) ** 2;
}

const slope = denominator !== 0 ? numerator / denominator : 0;
const intercept = yMean - slope * xMean;
```

Рисунок 2.2. Фрагмент коду реалізації лінійної регресії

Прогнозне значення для довільного майбутнього дня t обчислюється за рівнянням (2.6) [33]:

$$\hat{y}_t = m \times t + b, \quad (2.6)$$

де $\hat{y}_t$ (у з капелюшком, індекс t) – прогнозоване або передбачене значення залежної змінної в момент часу t , m – коефіцієнт нахилу, який показує швидкість зміни. У контексті часових рядів це тренд – наскільки швидко значення зростає або падає з часом. t – часова змінна або номер періоду, b – вільний член, початкове значення або базовий рівень. Це теоретичне значення y на початку відліку.

Система перевіряє знаменник перед діленням, щоб уникнути помилок у випадку нульової дисперсії незалежної змінної.

Фінальний прогноз формується як зважена комбінація методів (2.7) [34]:

$$\hat{y}_{comb} = 0.6 \times \hat{y}_{LR} + 0.4 \times EMA, \quad (2.7)$$

де $\hat{y}_{comb}$ – комбінований прогноз. Це фінальне передбачене значення, яке об'єднує переваги обох методів. $\hat{y}_{LR}$ – прогноз, отриманий методом лінійної регресії, ЕМА – експоненціальна ковзна середня, цей метод добре адаптується до останніх змін у даних, 0.6 – вага для лінійної регресії (60%). Показує, наскільки сильно прогноз LR впливає на фінальний результат, 0.4 – вага для ЕМА (40%). Показує, наскільки сильно ЕМА впливає на фінальний результат.

Більша вага надається лінійній регресії для врахування довгострокового тренду, а менша – експоненціальному згладжуванню для коригування на основі останніх спостережень. Такий розподіл ваг забезпечує баланс між стабільністю прогнозу та його чутливістю до останніх змін. Система генерує прогнози на задану кількість днів вперед, округлюючи результати до цілих чисел та гарантуючи невід'ємні значення (Рис. 2.3).

```
for (let i = 1; i <= daysAhead; i++) {
  const linearPrediction = slope * (n + i) + intercept;
  const combinedPrediction = linearPrediction * 0.6 + lastEma * 0.4;

  const predictionDate = new Date(now);
  predictionDate.setDate(predictionDate.getDate() + i);

  predictions.push({
    date: predictionDate.toISOString().split('T')[0],
    predicted: Math.max(0, Math.round(combinedPrediction)),
    confidence: this.calculateConfidence(dataPoints),
  });
}
```

Рисунок 2.3. Фрагмент коду розрахунку фінального значення

Для оцінки надійності прогнозу використовується аналіз варіабельності історичних даних. Розраховується коефіцієнт варіації за формулою (2.8) [35]:

$$CV = (\sigma / \mu) \times 100\%, \quad (2.8)$$

де σ – стандартне відхилення, μ – середнє значення.

За значенням коефіцієнта варіації прогнозу присвоюється рівень надійності: високий, середній або низький. Якщо кількість історичних точок менша за 7, надійність автоматично класифікується як низька через недостатність статистичної бази. Додатково система аналізує розподіл навантаження за днями тижня, підраховуючи загальну кількість декларацій для кожного дня тижня протягом історичного періоду. Визначаються три дні з найвищим навантаженням, що допомагає планувати розподіл ресурсів (Рис. 2.4).

```
private calculateConfidence(dataPoints: number[]): string {
  if (dataPoints.length < 7) return 'low';

  const mean = dataPoints.reduce((a, b) => a + b, 0) / dataPoints.length;
  const variance =
    dataPoints.reduce((sum, val) => sum + (val - mean) ** 2, 0) /
    dataPoints.length;
  const stdDev = Math.sqrt(variance);

  const cv = (stdDev / mean) * 100;

  if (cv < 20) return 'high';
  if (cv < 40) return 'medium';
  return 'low';
}
```

Рисунок 2.4. Фрагмент коду реалізації розрахунку коефіцієнта

Загальний тренд зміни кількості декларацій визначається на основі коефіцієнта нахилу: зростаючий тренд діагностується при $m > 1$, спадний – при $m < -1$, стабільний – у проміжному діапазоні. Система також розраховує середнє історичне навантаження для порівняння з прогнозованими значеннями.

Система виявлення похибок призначена для автоматичного моніторингу медичних даних та раннього виявлення проблемних ситуацій у процесі створення та обробки медичних декларацій. Основна мета полягає у забезпеченні

своєчасного виявлення відхилень від нормального функціонування для оперативного реагування та підвищення якості надання медичних послуг.

Реалізовано два статистичні методи – Z-Score та IQR, які працюють паралельно для максимального охоплення різних типів похибок. Z-Score ефективний для даних із близьким до нормального розподілом, тоді як IQR демонструє високу робастність до екстремальних значень і не потребує припущень про характер розподілу. Система складається з модуля збору та агрегації метрик для збору і попередньої обробки даних, Z-Score детектора для виявлення похибок на основі статистичного аналізу відхилень, IQR детектора для робастного виявлення викидів на основі міжквартильного розмаху, та генератора сповіщень для формування структурованих повідомлень про похибку. Система реалізована у вигляді двох основних методів класу сервісу. Перший метод призначений для виявлення похибок на рівні всієї системи з аналізом денної кількості декларацій та загальних показників відхилень. Другий метод фокусується на виявленні похибок у роботі окремих лікарів через аналіз їхньої продуктивності, якості роботи та активності. Обидва методи приймають ідентифікатор адміністратора та період аналізу в днях.

Процес починається зі збору історичних даних за визначений період. Система визначає лікарні, до яких має доступ адміністратор, та формує список активних лікарів. Далі відбувається агрегація денних показників декларацій з обчисленням загальної кількості створених декларацій, кількості відхилених декларацій та відсотка відхилень для кожного дня. На основі отриманих даних формуються масиви значень для статистичного аналізу.

Математична основа Z-score методу полягає в обчисленні базових статистичних показників. Для вибірки обчислюється середнє арифметичне значення за формулою (2.9) [36]:

$$\mu = (1/n) \times \sum x_i, \quad (2.9)$$

де μ – середнє арифметичне значення або математичне сподівання, n – кількість елементів у наборі даних, Σ (сигма) – символ суми, означає, що потрібно додати всі значення разом, x_i – i -тий елемент набору даних, де i може бути від 1 до n , $(1 / n)$ – множник, який ділить суму на кількість елементів.

Стандартне відхилення розраховується за формулою (2.10) [37]:

$$\sigma = \sqrt{[(1/n) \times \Sigma(x_i - \mu)^2]}, \quad (2.10)$$

де σ (сигма) – стандартне відхилення генеральної сукупності, $\sqrt{}$ (корінь квадратний) – витягує квадратний корінь з дисперсії, повертаючи результат до оригінальних одиниць виміру, n – кількість елементів у наборі даних.

Середнє арифметичне характеризує центральну тенденцію розподілу, тоді як стандартне відхилення описує ступінь розсіювання значень навколо середнього (Рис. 2.5).

```
const mean = counts.reduce((a, b) => a + b, 0) / counts.length;
const variance =
  | counts.reduce((sum, val) => sum + (val - mean) ** 2, 0) / counts.length;
const stdDev = Math.sqrt(variance);
```

Рисунок 2.5. Реалізація Z-score методу

Для кожного спостереження x розраховується Z-оцінка за формулою (1.4) яка показує, на скільки стандартних відхилень значення віддалене від середнього. Система встановлює два порогові значення для класифікації похибок. Діапазон $3 < |z| < 4$ відповідає попередженню, коли значення є незвичним, але не критичним, що вимагає уваги та моніторингу ситуації. Значення $|z| \geq 4$ вказує на критичну похибку, коли значення значно відхиляється від норми та вимагає негайного втручання. Вибір порогів базується на правилі трьох сигм, згідно з яким у нормальному розподілі приблизно 99.7% значень знаходяться в межах трьох стандартних відхилень від середнього.

IQR є робастним підходом до виявлення викидів, який не залежить від припущень про характер розподілу даних. На відміну від Z-Score методу, який чутливий до наявності викидів при обчисленні середнього та стандартного відхилення, IQR метод базується на квантилях, які є більш робастними характеристиками розподілу.

Метод базується на обчисленні квантилів, що ділять упорядкований набір даних на чверті. Перший квантиль Q1 відповідає 25-му перцентилю, третій квантиль Q3 відповідає 75-му перцентилю. Міжквантильний розмах обчислюється як різниця третього та першого квантиля і представляє діапазон, в якому знаходяться центральні 50% спостережень [38] (Рис. 2.6).

```
const sortedCounts = [...counts].sort((a, b) => a - b);
const q1Index = Math.floor(sortedCounts.length * 0.25);
const q3Index = Math.floor(sortedCounts.length * 0.75);
const q1 = sortedCounts[q1Index];
const q3 = sortedCounts[q3Index];
const iqr = q3 - q1;
```

Рисунок 2.6. Реалізація IQR методу

Для визначення меж нормальних значень використовується правило Тьюкі. Згідно з цим правилом, нижня межа обчислюється за формулою (2.11) [39]:

$$Q1 - 1.5 \times IQR, \quad (2.11)$$

де Q1 – значення, яке ділить дані так, що 25% даних знаходяться нижче нього, а 75% – вище. Також називається 25-м перцентилем, IQR – міжквантильний розмах, міра розкиду середніх 50% даних. 1.5 – емпірично визначена константа, яка встановлює межу для виявлення аномальних значень. Верхня межа розраховується за формулою (2.12):

$$Q3 + 1.5 \times IQR, \quad (2.12)$$

де Q3 – значення, яке ділить дані так, що 75% даних знаходяться нижче нього, а 25% – вище, IQR – міжквартильний розмах, 1.5 – стандартний коефіцієнт Тьюкі для визначення викидів.

Значення класифікується як викид, якщо воно знаходиться поза визначеними межами. Коефіцієнт 1.5 є стандартним у статистичному аналізі та забезпечує баланс між чутливістю до похибок та стійкістю до випадкових коливань. Система використовує комбінований підхід до виявлення похибок, при якому похибку вважається підтвердженою, якщо виявлена хоча б одним з методів. Така логіка забезпечує максимальну чутливість системи до різних типів відхилень, оскільки Z-Score метод краще виявляє поступові зміни та тренди у даних з близьким до нормального розподілом, тоді як IQR метод ефективніший для виявлення раптових викидів.

Система ідентифікує декілька типів похибок залежно від характеру відхилення. Стрибок кількості декларацій спостерігається, коли денна кількість декларацій значно перевищує очікуване значення, що може свідчити про масову реєстрацію пацієнтів, завершення реєстраційних кампаній або технічні проблеми з наступним накопиченням записів. Спад кількості декларацій фіксується, коли денна кількість значно нижча за очікуване значення, що може вказувати на проблеми з доступністю системи, технічні проблеми або низьку активність лікарів. Високий відсоток відхилень означає аномально високий відсоток відхилених декларацій, що може свідчити про систематичні проблеми з якістю даних, зміни у правилах валідації або недостатню кваліфікацію персоналу.

Для показника відсотка відхилених декларацій застосовується підхід, який поєднує статистичний аналіз з абсолютними порогам практичної значущості. Спочатку обчислюється середній відсоток відхилень за аналізований період. Похибка виявляється за двома обов'язковими умовами, які мають виконуватися одночасно, актуальне значення має перевищувати середнє більш ніж удвічі, абсолютне значення має перевищувати 20 відсотків. Ця подвійна умова гарантує,

що похибка буде зареєстрована лише у випадку значного та практично важливого відхилення.

Критичність визначається залежно від величини відсотка відхилень. Окремим напрямком роботи системи є моніторинг діяльності індивідуальних лікарів. Для кожного лікаря збирається детальна статистика за аналізований період, що включає ідентифікатор та повне ім'я лікаря, загальну кількість оброблених декларацій як показник обсягу роботи, кількість та відсоток відхилених декларацій як індикатор якості роботи, а також середній час обробки декларації в годинах.

Для виявлення лікарів з аномально високим відсотком відхилень спочатку обчислюється середній показник по всіх лікарях. Для кожного лікаря розраховується індивідуальний відсоток відхилень як співвідношення кількості відхилених декларацій до загальної кількості оброблених декларацій. Похибка виявляється за двома умовами: відсоток відхилень конкретного лікаря має перевищувати середній показник більш ніж удвічі, та абсолютне значення відсотка має перевищувати 20% (Рис. 2.7).

```
const rejectionRates = doctorStats.map((d) => {
  const total = parseInt(d.totalDeclarations);
  const rejected = parseInt(d.rejectedDeclarations);
  return total > 0 ? (rejected / total) * 100 : 0;
});

const processingTimes = doctorStats
  .map((d) => parseFloat(d.avgProcessingHours))
  .filter((t) => !isNaN(t));

const meanRejectionRate =
  rejectionRates.reduce((a, b) => a + b, 0) / rejectionRates.length;
const meanProcessingTime =
  processingTimes.reduce((a, b) => a + b, 0) / processingTimes.length;
```

Рисунок 2.7. Фрагмент коду розрахунку лікарів з найбільшою кількістю відхилень

Система аналізує середній час обробки декларацій кожним лікарем та порівнює його із загальним середнім показником. З масиву статистики витягуються значення середнього часу обробки з фільтрацією недійсних значень, після чого обчислюється загальний середній час. Похибка повільної обробки виявляється, якщо середній час обробки конкретного лікаря перевищує загальний середній показник більш ніж удвічі.

Система відстежує кількість оброблених декларацій кожним лікарем та порівнює її із середнім показником активності. Низька активність фіксується, якщо кількість декларацій менша за 30% від середньої кількості по всіх лікарях. Похибка реєструється лише для лікарів з ненульовою активністю для уникнення хибних спрацьовувань. Для похибок низької активності завжди встановлюється рівень попередження.

Система формує структуровану відповідь, яка містить повну інформацію про виявлені похибки та дозволяє адміністраторам швидко оцінити ситуацію. Відповідь складається з трьох основних секцій. Перша секція містить масив похибок, де кожна виявлена похибка описується набором полів:

1. Дата виявлення.
2. Тип похибки.
3. Рівень критичності.
4. Назва метрики.
5. Фактичне та очікуване значення.
6. Відсоток відхилення від норми.
7. Значення Z-оцінки або позначка.
8. Текстовий опис проблеми.

Друга секція містить підсумкову статистику з агрегованими даними про всі виявлені похибки:

1. Загальна кількість похибок.
2. Кількість критичних похибок та кількість попереджень.

Третя секція містить загальну статистику за аналізований період:

1. Середня кількість декларацій на день.
2. Стандартне відхилення.
3. Середній відсоток відхилень, та у деяких випадках діапазон IQR.

Система передбачає ретельну обробку граничних випадків для забезпечення стабільної роботи у будь-яких умовах. Особливої уваги потребує обробка ситуації з нульовим стандартним відхиленням, яка може виникнути, якщо всі значення у вибірці однакові. У такому випадку Z-оцінка встановлюється в нуль для уникнення ділення на нуль.

RFM-аналіз адаптовано для медичної галузі з метою сегментації лікарів за рівнем активності та ефективності роботи в системі електронних декларацій. Метод дозволяє класифікувати лікарів у значущі групи для застосування диференційованих підходів до управління та підтримки.

Класичні RFM-метрики трансформовано відповідно до специфіки медичних систем. Recency (R) визначається як кількість днів з моменту останньої обробки декларації лікарем. Менше значення R вказує на вищу активність лікаря. У системі обчислюється як різниця між поточною датою та датою останньої активності. Frequency (F) представляє загальну кількість декларацій, оброблених лікарем за весь період роботи в системі. Більше значення F свідчить про вищу активність. Метрика обчислюється як загальна кількість записів декларацій для кожного лікаря через SQL-агрегацію. Monetary (M) переосмислена як кількість активних декларацій лікаря на поточний момент. Більше значення M вказує на більше навантаження та цінність лікаря для системи. Розрахунок виконується через умовну агрегацію.

Для кожної метрики застосовується квантільний розподіл, що забезпечує рівномірну класифікацію лікарів. Обчислюються п'ять квінтилів, що розділяють вибірку на п'ять рівних груп. Система реалізує автоматичний розрахунок

квінтильних меж через сортування значень: масиви R сортуються за зростанням, а F та M за спаданням.

Функція визначає, до якого квінтиля потрапляє значення та порівнюючи значення з пороговими межами. Кожному лікарю присвоюється бал від 1 до 5 по кожній метриці. Для метрики R інвертується напрямок балів через параметр, бал 5 отримують лікарі з найменшою давністю активності. Для метрик F та M з параметром бал 5 отримують лікарі з найвищими значеннями. Для кожного лікаря формується RFM-профіль, що складається з трьох балів, які об'єднуються в комплексну оцінку для можливості ранжування лікарів.

Для практичної застосовності 125 можливих RFM-профілів агрегуються у сім основних сегментів на основі встановлених правил класифікації. Успішні лікарі включає найбільш активних та завантажених лікарів з $R \geq 4$ та $F \geq 4$ та $M \geq 4$, які регулярно працюють з деклараціями, мають високу кількість оброблених декларацій та значне поточне навантаження. Лояльні лікарі об'єднує лікарів з $R \geq 3$ та $F \geq 3$ та $M \geq 3$, що демонструють постійну якісну роботу та є стабільно активними фахівцями з хорошими показниками по всіх метриках. Потенційно лояльні лікарі включає лікарів з $R \geq 4$ та $F \leq 3$ та $M \leq 3$, тобто нещодавно активних лікарів з гарним потенціалом для зростання, які можуть збільшити свою продуктивність при відповідній підтримці. Лікарі в зоні ризику охоплює раніше активних лікарів з $R \leq 2$ та $F \geq 3$ та $M \geq 3$, що мають хорошу історію роботи, але тривалий час не проявляли активності. Критично важливі лікарі визначає критично важливих лікарів з ознаками відтоку: $R \leq 2$ та $F \geq 4$ та $M \geq 4$. Це найбільш цінні фахівці, які тривалий час не активні та потребують негайної уваги. Неактивні лікарі об'єднує лікарів з тривалою відсутністю активності $R \leq 2$ та $F \leq 2$, тобто неактивних фахівців з низькою історичною продуктивністю. Лікарі які потребують уваги включає всіх інших лікарів, що не потрапили в попередні категорії, зокрема лікарів з непостійною активністю, які потребують індивідуального аналізу. Система автоматично розраховує розподіл

лікарів за сегментами для надання адміністраторам загальної картини стану активності в лікарнях.

Аналіз воронки конверсії застосовується для виявлення вузьких місць у процесі обробки медичних декларацій. Метод дозволяє кількісно оцінити ефективність кожного етапу та ідентифікувати проблемні переходи між ними.

Реалізація збору статистики виконується через агреговані SQL-запити до бази даних системи. Кількість декларацій на кожному етапі визначається на основі їхніх статусів у системі. Для етапу розгляду враховуються декларації зі всіма статусами (Рис. 2.8).

```
const funnelStats = await this.declarationRepository
  .createQueryBuilder('declaration')
  .select('COUNT(*)', 'total')
  .addSelect(
    "SUM(CASE WHEN declaration.status IN ('pending_doctor_review', 'pending_doctor_sign', 'active', 'rejected', 'terminated') THEN 1 ELSE 0 END)",
    'reviewed',
  )
  .addSelect(
    "SUM(CASE WHEN declaration.status IN ('pending_doctor_sign', 'active', 'rejected', 'terminated') THEN 1 ELSE 0 END)",
    'signed',
  )
  .addSelect(
    "SUM(CASE WHEN declaration.status = 'active' THEN 1 ELSE 0 END)",
    'activated',
  )
  .where('declaration.inserted_at BETWEEN :startDate AND :endDate', {
    startDate,
    endDate,
  })
  .andWhere('declaration.doctor_email IN (:...doctorEmails)', {
    doctorEmails: doctorEmailList,
  })
  .getRawOne();
```

Рисунок 2.8. Запит на отримання інформації про декларації

Запит враховує часовий діапазон аналізу та фільтрує декларації за електронними адресами лікарів, що працюють у лікарнях, доступних для аналізу адміністратором. Такий підхід забезпечує коректність та релевантність отриманих статистичних даних для конкретного медичного закладу чи групи закладів.

Для кожного переходу між етапами розраховуються ключові метрики, що характеризують ефективність процесу обробки декларацій. Коефіцієнт конверсії

показує частку декларацій, що успішно перейшли на наступний етап, та обчислюється за формулою (2.13) [40]:

$$CR_i = \frac{N_{i+1}}{N_i} \times 100\%, \quad (2.13)$$

де N_i представляє кількість декларацій на етапі i .

Загальний коефіцієнт конверсії воронки визначається як відношення кількості активованих декларацій до загальної кількості створених (2.14) [40]:

$$CR_{total} = \frac{N_4}{N_1} \times 100, \quad (2.14)$$

Коефіцієнт відпаду характеризує втрати між суміжними етапами та розраховується за формулою (2.15) [40]:

$$DR_i = \frac{N_i - N_{i+1}}{N_i} \times 100\% = 100\% - CR_i, \quad (2.15)$$

Високий коефіцієнт відпаду вказує на проблемний перехід, що потребує детального аналізу та оптимізації.

Для початкового етапу створення декларацій коефіцієнт конверсії встановлюється на рівні 100%, оскільки це базова точка відліку для всієї воронки. Коефіцієнт відпаду на цьому етапі дорівнює нулю. Для наступних етапів коефіцієнти обчислюються відносно загальної кількості створених декларацій або кількості декларацій на попередньому етапі, що дозволяє оцінити як абсолютну, так і відносну ефективність кожного переходу.

Реалізована система автоматично виявляє вузькі місця воронки конверсії шляхом визначення етапу з максимальним коефіцієнтом відпаду. Алгоритм ідентифікації послідовно аналізує всі етапи воронки, порівнюючи значення коефіцієнтів відпаду та фіксуючи етап з найбільшим показником втрат (Рис. 2.9).

```

private identifyBottleneck(stages: any[]): string {
    let maxDropOff = 0;
    let bottleneckStage = '';

    stages.forEach((stage) => {
        const dropOff = parseFloat(stage.dropOff);
        if (dropOff > maxDropOff) {
            maxDropOff = dropOff;
            bottleneckStage = stage.stage;
        }
    });

    return bottleneckStage || 'Немає';
}

```

Рисунок 2.9. Алгоритм пошуку вузького місця в підписанні декларації

Етап визначається як вузьке місце, якщо для нього характерний найбільший коефіцієнт відпаду серед усіх переходів у воронці. Така методика дозволяє локалізувати проблемні ділянки процесу та визначити пріоритети для оптимізації системи обробки медичних декларацій.

Результати аналізу воронки представляються у структурованому форматі, що включає масив етапів з відповідними метриками, загальний коефіцієнт конверсії, загальну кількість декларацій та ідентифіковане вузьке місце. Такий формат забезпечує зручність інтерпретації результатів та прийняття управлінських рішень щодо оптимізації процесів медичного закладу.

Кореляційний аналіз застосовується для виявлення лінійних взаємозв'язків між метриками системи в розрізі часових параметрів. Метод реалізує аналіз активності створення декларацій за днями тижня та годинами доби з подальшим обчисленням кореляції між обсягом декларацій та кількістю відхилень. Для кожної пари метрик обчислюється коефіцієнт кореляції Пірсона за формулою (2.16) [41]:

$$r = \Sigma[(x_i - \bar{x})(y_i - \bar{y})] / \sqrt{[\Sigma(x_i - \bar{x})^2 \cdot \Sigma(y_i - \bar{y})^2]}, \quad (2.16)$$

де r – коефіцієнт кореляції Пірсона. Значення завжди знаходиться в діапазоні від -1 до +1, Σ (сигма) – символ суми всіх значень від $i = 1$ до n , x_i – i -те значення першої змінної, y_i – i -те значення другої змінної, $\bar{x}$ (х з рискою) – середнє арифметичне всіх значень, $\bar{y}$ (у з рискою) – середнє арифметичне всіх значень.

Система аналізує залежність між денним обсягом декларацій та кількістю відхилень, що дозволяє виявити перевантаження системи в пікові періоди. Для кожної виявленої значущої кореляції формується аналітичний висновок з поясненням можливих причинно-наслідкових зв'язків (Рис. 2.10).

```
private calculatePearsonCorrelation(x: number[], y: number[]): number {
  const n = x.length;
  if (n === 0) return 0;

  const xMean = x.reduce((a, b) => a + b, 0) / n;
  const yMean = y.reduce((a, b) => a + b, 0) / n;

  let numerator = 0;
  let xVariance = 0;
  let yVariance = 0;

  for (let i = 0; i < n; i++) {
    const xDiff = x[i] - xMean;
    const yDiff = y[i] - yMean;
    numerator += xDiff * yDiff;
    xVariance += xDiff ** 2;
    yVariance += yDiff ** 2;
  }

  const denominator = Math.sqrt(xVariance * yVariance);
  return denominator !== 0 ? numerator / denominator : 0;
}
```

Рисунок 2.10. Реалізація метода розрахунку коефіцієнта кореляції Пірсона

2.2 Реалізація алгоритму ETL-трансформації даних

Процес витягування, трансформації та завантаження даних є фундаментальним підходом до інтеграції даних між різними системами. У розробленому додатку реалізовано двонаправлений механізм обробки даних, який дозволяє як експортувати інформацію з PostgreSQL у формати Excel та CSV, так і імпортувати дані з файлів назад у базу даних. Архітектура системи

побудована бібліотеки TypeORM для роботи з базою даних та спеціалізованих інструментів XLSX і PapaParse для обробки структурованих файлів.

Фаза витягування даних реалізована двома способами залежно від напрямку потоку інформації. Для експорту з бази використовується компонент побудови запитів бібліотеки TypeORM, що дозволяє ефективно завантажувати дані разом зі зв'язаними сутностями через механізм лівого з'єднання таблиць. Система автоматично визначає необхідні зв'язки на основі аналізу конфігурації полів, які потрібно експортувати. Відповідний метод виконує розбір шляхів до полів та ідентифікує зв'язки першого рівня для подальшого завантаження. Для імпорту даних підтримуються два формати структурованих файлів.

Файли формату CSV обробляються спеціалізованою бібліотекою з автоматичним визначенням заголовків колонок та динамічним приведенням типів даних, що забезпечує коректну інтерпретацію числових та логічних значень. Файли формату Excel обробляються через відповідну бібліотеку, яка читає дані з робочих аркушів та конвертує їх у масив об'єктів для подальшої обробки.

Фаза трансформації даних є найбільш складною та критичною частиною всього процесу. Ключовим елементом трансформації для експорту виступає спеціалізований метод, який використовує функціональний підхід для навігації по глибоко вкладених структурах об'єктів.

Цей механізм дозволяє отримати значення з ієрархічних структур даних на кшталт вкладених властивостей об'єктів користувача та лікарні, що критично важливо для роботи з реляційними даними. Наступний метод виконує агрегацію інформації з декількох полів джерела, дозволяючи об'єднувати значення.

Для імпорту реалізовано два принципово різні підходи обробки інформації. Перший підхід використовує двофазну модель завантаження, де на першому етапі створюються основні сутності предметної області, а на другому етапі встановлюються зв'язки між цими сутностями через відповідні методи сервісу.

Другий підхід дозволяє створювати сутності та встановлювати зв'язки паралельно в рамках одного проходу по вхідним даним, що зручно для простих сценаріїв імпорту з денормалізованої структури. Особливої уваги при трансформації потребує обробка складних структурованих полів, таких як персональні дані, телефони та адреси, які можуть зберігатися як текстові рядки в табличних файлах і потребують десеріалізації у відповідні об'єкти мови програмування.

Фаза завантаження даних реалізована з урахуванням вимог до продуктивності системи при обробці великих обсягів інформації. При експорті в формат Excel дані спочатку проходять проміжну трансформацію, після чого формується структура робочого аркуша відповідним методом бібліотеки та генерується фінальний файл у вигляді бінарного буфера з явним зазначенням типу вихідного документа. При імпорті в реляційну базу даних застосовується стратегія пакетної обробки, коли дані зберігаються групами по сто записів за одну транзакцію, що суттєво знижує навантаження на з'єднання з базою даних та покращує загальну продуктивність операції. Така оптимізація особливо критична при імпорті великих масивів декларацій.

Створення зв'язків типу багато-до-багатьох між сутностями користувачів та лікарень здійснюється через прямі SQL-запити з використанням спеціальної конструкції, що ігнорує конфлікти унікальності, дозволяючи уникнути помилок при спробі повторного створення існуючих зв'язків у проміжних таблицях асоціацій. Підсистема валідації забезпечує цілісність та коректність даних на кожному етапі процесу обробки. При імпорті структурованих файлів спочатку виконується перевірка наявності всіх обов'язкових полів у вхідному наборі даних шляхом порівняння конфігурації з реальною структурою файлу.

Додатково реалізовано механізм валідації на рівні метаданих об'єктно-реляційного відображення, який перевіряє фактичне існування полів та зв'язків у

схемі бази даних перед початком виконання операцій, що дозволяє виявляти помилки конфігурації на ранніх стадіях життєвого циклу обробки запиту.

Інтеграція системи обробки даних з аналітичним модулем демонструє практичне застосування концепції трансформації інформації в контексті підготовки тестових наборів даних. Алгоритм формує шість часових когорт користувачів за останні півроку з різними рівнями активності та характеристиками використання системи. При генерації записів про декларації застосовується механізм зваженого випадкового розподілу для днів тижня та годин доби, що імітує реальні паттерни поведінки користувачів медичної інформаційної системи. Статистичний аналіз показує, що активність у робочі дні становить приблизно 15-20 відсотків від загального обсягу, тоді як активність у вихідні дні знижується до 6-8 відсотків, а пікові години користування системою припадають на проміжок з десятої до дванадцятої години з імовірністю близько 13 відсотків. Крім типових патернів, система також генерує аномальні періоди з різкими сплесками активності, коли кількість створених декларацій перевищує середнє значення майже вдвічі, а також провалами активності з мінімальною кількістю операцій.

Для забезпечення високої продуктивності системи при обробці великих масивів даних застосовано комплекс технічної оптимізації на різних рівнях архітектури. Пакетна обробка запитів суттєво знижує загальну кількість звернень до бази даних та мінімізує накладні витрати на встановлення з'єднань. Використання параметризованих запитів не лише запобігає вразливостям типу впровадження SQL-коду, але й дозволяє базі даних ефективно кешувати плани виконання запитів. Компонент побудови запитів бібліотеки TypeORM генерує оптимізовані SQL-конструкції з явними операціями з'єднання таблиць замість виконання множинних окремих запитів, що є класичною проблемою продуктивності об'єктно-реляційних систем. Механізм пулу з'єднань автоматично керує життєвим циклом підключень до сервера PostgreSQL,

переповикористовуючи активні з'єднання та розподіляючи навантаження, що усуває накладні витрати на встановлення нових з'єднань для кожної операції. При роботі зі зв'язками типу багато-до-багатьох використовується спеціальна SQL-конструкція в прямих запитах, що дозволяє виконувати масові операції вставки без явної попередньої перевірки на існування записів, покладаючись на механізм унікальних індексів PostgreSQL для автоматичного запобігання створенню дублікатів.

Реалізована система обробки та трансформації даних забезпечує необхідну гнучкість у налаштуванні набору полів для операцій експорту та імпорту через спеціалізовані структури конфігурації, підтримує роботу зі складними реляційними структурами та можливість навігації по глибоко вкладених об'єктах через механізм точкової нотації у визначенні шляхів доступу до полів, гарантує валідацію даних на кожному етапі конвеєра обробки та демонструє високу продуктивність завдяки застосуванню пакетної обробки та оптимізованих стратегій формування запитів.

Висновки до розділу 2

У другому розділі магістерської роботи детально описано методологію дослідження та застосування аналітичних алгоритмів для системи управління медичними деклараціями.

Використано комплекс методів включаючи системний аналіз, математичну статистику, прогнозування часових рядів, виявлення похибок, сегментацію, когортний аналіз, аналіз воронки конверсії та мережевий аналіз.

Застосовано гібридний алгоритм прогнозування навантаження, що поєднує експоненціальне згладжування та лінійну регресію з динамічними вагами. Розроблено модульну архітектуру з компонентами збору даних, розрахунку прогнозів, комбінування результатів та обчислення довірчих інтервалів.

Реалізовано систему виявлення похибок на основі паралельного застосування Z-Score та IQR методів. Розроблено консенсусну стратегію прийняття рішень, що вимагає підтвердження похибки обома методами одночасно для зменшення кількості хибних спрацювань. Встановлено порогові значення для класифікації похибок за рівнями критичності.

Адаптовано RFM-аналіз для медичної галузі з трансформацією класичних метрик під специфіку системи декларацій. Розроблено алгоритм квантільної категоризації з присвоєнням балів від 1 до 5 по кожній метриці. Розроблено специфічні рекомендації для роботи з кожним сегментом.

Реалізовано когортний аналіз для відстеження утримання користувачів у часі. Розроблено методику формування місячних та тижневих когорт та розрахунку коефіцієнтів утримання і відтоку. Створено візуалізацію результатів у вигляді теплової матриці з колірним кодуванням рівнів утримання.

Застосовано аналіз воронки конверсії для виявлення вузьких місць у чотириетапному процесі обробки декларацій. Розроблено алгоритми розрахунку коефіцієнтів конверсії та відпаду між етапами. Створено систему автоматичного виявлення вузьких місць на основі трьох критеріїв: абсолютного відпаду, відносного відпаду та динаміки змін.

Реалізовано кореляційний аналіз для виявлення взаємозв'язків між метриками системи з використанням коефіцієнта Пірсона. Застосовано мережевий аналіз для дослідження розподілу навантаження між лікарями з розрахунком центральності ступеня та класифікацією за рівнем впливу.

Для кожного алгоритму розроблено модульну архітектуру з чітким розділенням відповідальності компонентів, що забезпечує масштабованість, підтримкою та можливість незалежного тестування. Використання технологічного стеку Nest.js, PostgreSQL, React забезпечує ефективну реалізацію всіх аналітичних алгоритмів.

Представлена методологія та застосування алгоритмів забезпечують комплексне вирішення задач аналізу медичних даних у системі електронних декларацій, що дозволяє здійснювати прогнозування навантаження, виявляти похибки, сегментувати користувачів, оцінювати утримання, ідентифікувати вузькі місця процесів, аналізувати взаємозв'язки метрик та оптимізувати розподіл навантаження між лікарями.

РОЗДІЛ 3 ОПИС ПРОГРАМНОГО ПРОДУКТУ

3.1 Огляд статистичної аналітики

Після завантаження ресурсу користувач бачить стартову сторінку з коротким описом продукту та кнопкою логіну або реєстрації, які направить користувача на сторінку аутентифікації.

Якщо користувач не зареєстрований в системі, існує два варіанти реєстрації. Перший за допомогою самостійної реєстрації де є можливість ввести ім'я, електронну адресу та пароль користувача. Користувачу дається змога обрати свою роль такі як: пацієнт, лікар, адміністратор лікарні (Рис. 3.1).

Інший метод за допомогою інструмент імпорту, при використанні даного інструмента користувачу дається змога імпортувати всі типи даних які є в системі: користувачів, декларації, лікарні, прив'язку адміністраторів до лікарень або прив'язка пацієнта до лікаря. Якщо користувач який буде імпортувати інформацію має вже існуючу базу даних в нього є можливість трансформувати дані зі старої бази даних в файл типу XLSX та імпортувати данні в систему.

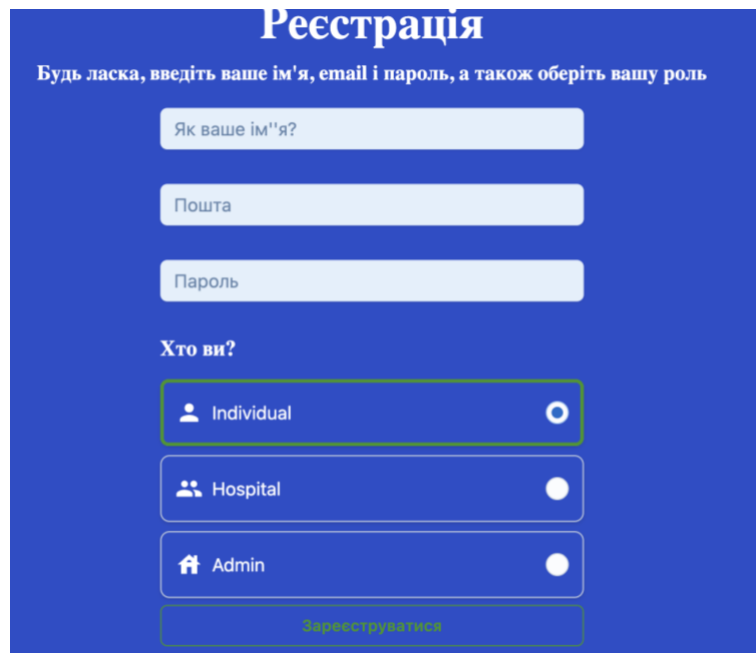


Рисунок 3.1. Сторінка реєстрації користувача

Якщо користувач вже має обліковий запис у системі, він може авторизуватися через спеціально призначену сторінку логіну. Для входу в систему користувачу необхідно ввести свої облікові дані, а саме адресу електронної пошти та пароль, які були вказані під час реєстрації. Після успішної перевірки введених даних система надає доступ до персонального кабінету та всіх функціональних можливостей системи (Рис. 3.2).

Рисунок 3.2. Сторінка логіну

В системі існує три ролі користувачів, які мають різні задачі та функціонал який в них існує, розглянемо кожну з них детальніше.

Роль адміністратора має найширші повноваження у системі та забезпечує управління всіма ключовими бізнес-процесами. Адміністратор має виключний доступ до створення, редагування та видалення юридичних осіб медичних закладів.

Система управління підрозділами надає можливість створення ієрархічної структури медичних закладів з повним спектром операцій.

Система управління запрошеннями дозволяє адміністратору обробляти заявки на приєднання адміністраторів до закладу через механізми підтвердження або відхилення запитів, організації процесу запрошення користувачів до лікарні та можливість припинення членства у медичному закладі.

Функціонал експорту забезпечує конвертацію даних з формату CSV у Microsoft Excel з гнучким налаштуванням. Адміністратор визначає тип об'єктів експорту, конфігурує перелік полів для вивантаження, встановлює обов'язковість окремих атрибутів та ініціює генерацію файлів у форматі Excel.

На сторінці аналітики користувач бачить статистику та аналітику в системі в діапазоні бачення своїх лікарень. Загальний фільтр по статистиці відбувається за допомогою вибору періоду фільтрації: день, тиждень, місяць, квартал, весь час.

В першому блокові статистики користувач може бачити загальну кількість користувачів які працюють в його лікарнях. Та активних користувачів що мається на увазі користувачів в лікарнях який мають активний статус профілю туди входять адміністратори, пацієнти та лікарі. Також представлена інформація про загальну кількість декларацій та кількість активних лікарень користувача.

В блокові статистики користувач може отримати інформацію про детальні статуси декларацій такі як завершенні, відхилені, активні та ті які очікують підпису. В блокові активності користувачів описується кількість користувачів в діапазоні лікарень та їх активність в системі (Рис. 3.3).

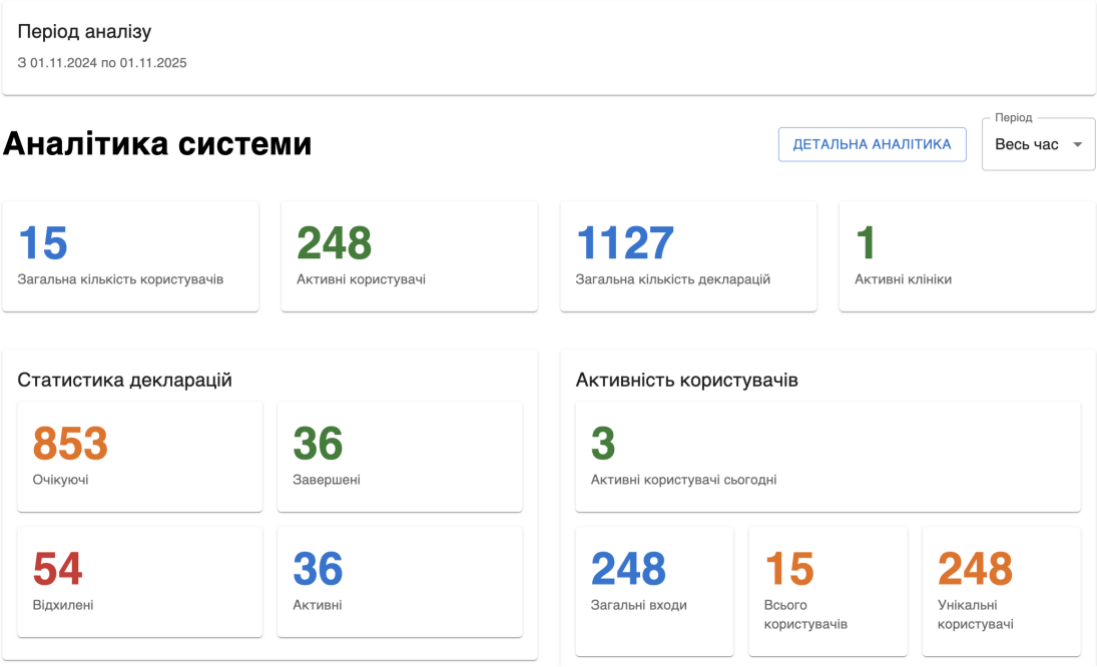


Рисунок 3.3. Сторінка аналітики системи, розділ з загальною статистикою

В блокові продуктивності лікарів користувач може побачити відсоток прочитаних нотифікацій та інформацію про загальну кількість лікарів в активних лікарнях та їх загальну кількість активних декларацій. В блокові взаємодії з повідомленнями відображено кількість відправлених нотифікацій в клініці та кількість переглянутих повідомлень (Рис. 3.4).

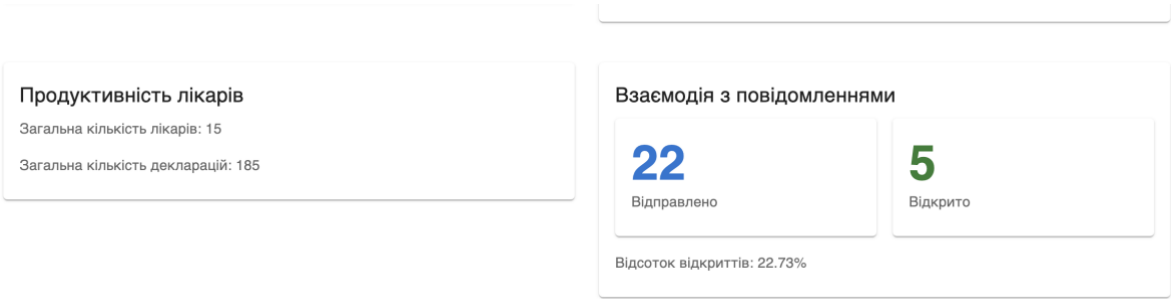


Рисунок 3.4. Сторінка аналітики системи, розділ статистики продуктивності лікарів та взаємодії з повідомленнями

На аналітичному графіку виявлення похибок в системі, адміністратор може переглянути кількість та деталі похибок в системі за періоди: 7, 14, 30, 60, 90 днів.

В короткій інформації по аналітиці користувач може побачити загальну кількість похибок, скільки з них критичних та скільки з них попереджають про можливу похибки. Також в інформації представлено середня значення похибок, стандартне відхилення та середній відсоток відхилень. На графіку зображено: фактичне значення похибок, очікуване значення, критичність похибки та попередження (Рис.3.5).

При наведенні на вісь графіка користувач може побачити детальну інформацію про випадок похибки, яка включає дату та час похибки, тип похибки, метрику похибки, фактичне значення, очікуване значення, Z-оцінку похибки та опис похибки.

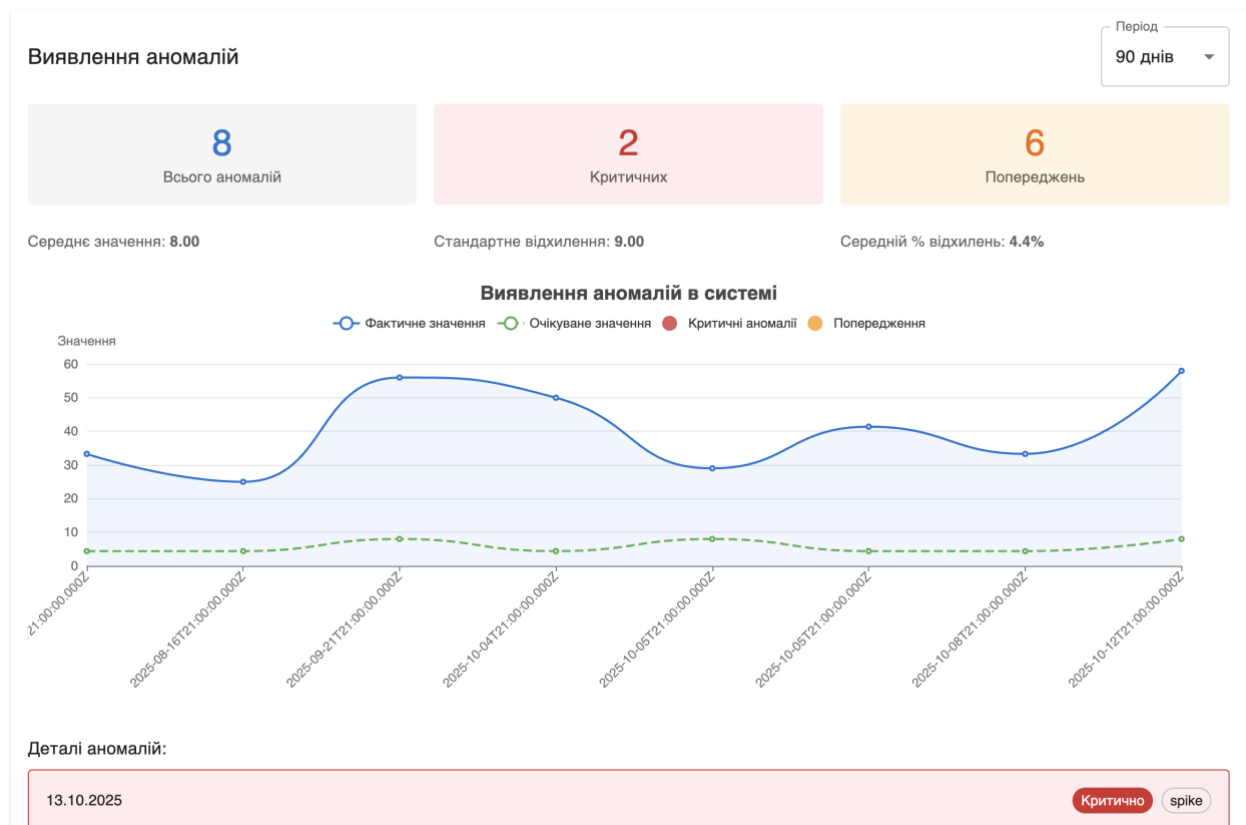


Рисунок 3.5. Сторінка аналітики графік виявлення похибок

Якщо під час аналізу даних похибок не було виявлено, замість графіку користувач буде бачити інформативне повідомлення на його місці, яке підтверджує відсутність відхилень у досліджуваному періоді. Це повідомлення

дозволяє користувачу зрозуміти, що всі метрики працюють у межах допустимих значень і не потребують додаткової уваги. У випадках, коли система виявляє похибки або помилку в даних, користувач отримує можливість детальніше розглянути специфіку кожної проблеми через спеціально організований список відхилень. Цей список структурований таким чином, щоб забезпечити зручність аналізу та швидку ідентифікацію критичних ситуацій (Рис. 3.6).

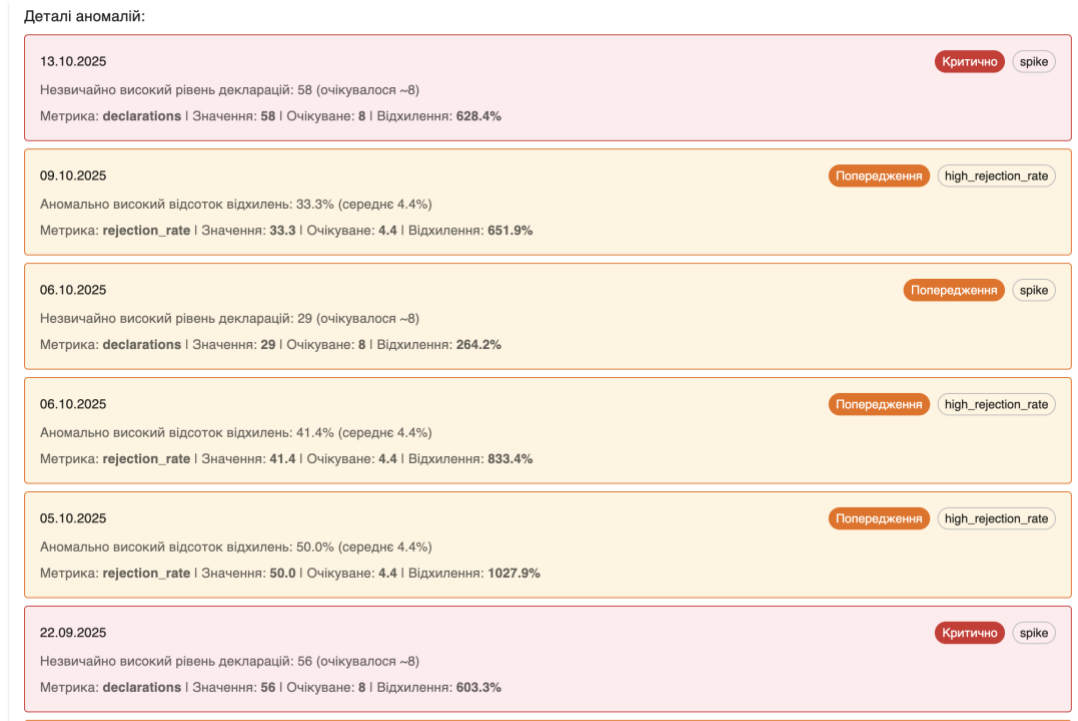


Рисунок 3.6. Сторінка аналітики, деталі графіку виявлення похибок

На аналітичному графіку похибок у лікарів адміністратор може переглянути похибки які виникли в певних лікарів та порівнює їх за періоди: 7, 14, 30, 60, 90 днів. В блоках з інформацією написано про загальну кількість лікарів, кількість критичних похибок, і кількість попереджень, в додатковій інформації написано відсоток відхилень та середній час обробки декларації лікарями. В графіку порівняння лікарів з похибками зображено лікарів та їх фактичні значення в порівнянні очікуваного значення. При наведенні на елемент

графіку користувач може побачити ім'я лікаря, тип та критичність проблеми, фактичне та очікуване значення (Рис. 3.7).

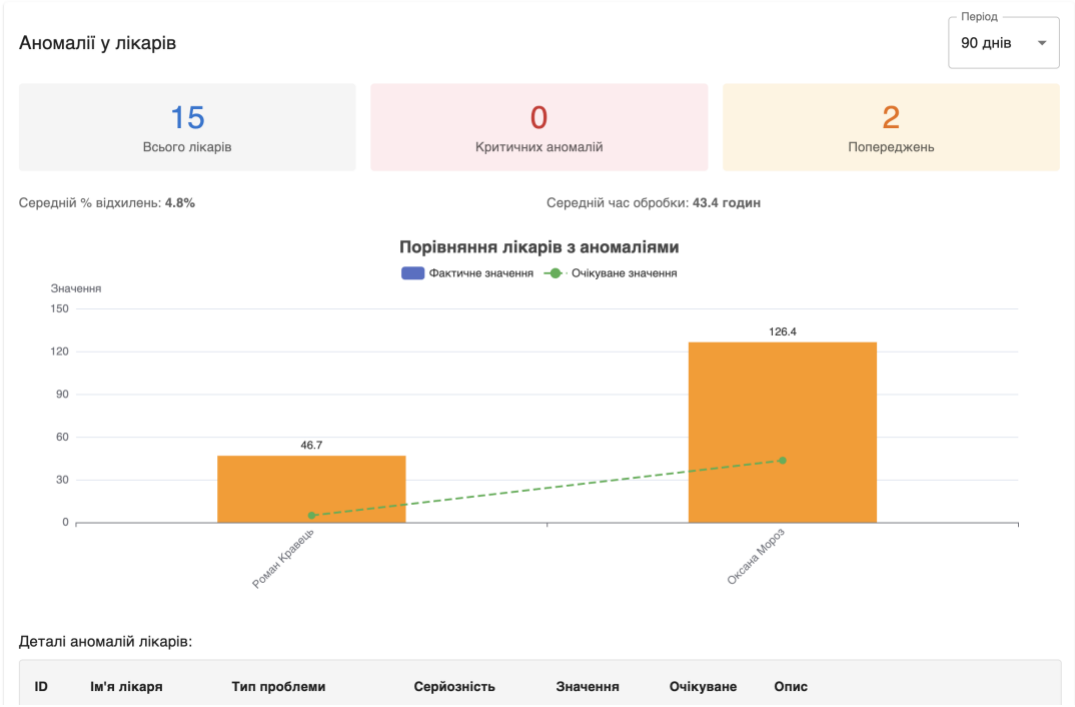


Рисунок 3.7. Сторінка аналітики, графік похибки у лікарів

Якщо похибок не було виявлено, замість графіку користувач буде бачити відповідне повідомлення на місці графіка. У разі наявності помилок у роботі лікарів, користувач має можливість ознайомитися з деталями цих помилок через відповідний перелік. В список включена така інформація як ідентифікатор лікаря, ім'я лікаря, тип проблеми, серйозність проблеми, фактичне значення, очікуване значення, опис проблеми (Рис. 3.8).

Деталі аномалій лікарів:

ID	Ім'я лікаря	Тип проблеми	Серйозність	Значення	Очікуване	Опис
644	Роман Кравець	Високий % відхилень	Попередження	46.7%	4.8%	Високий відсоток відхилень декларацій
645	Оксана Мороз	Повільна обробка	Попередження	126.4 годин	43.4 годин	Повільна обробка декларацій

Рисунок 3.8. Сторінка аналітики, деталі графіку похибок у лікарів

На аналітичному графіку когортний аналіз пацієнтів по тижням або місяцям. В короткій інформації користувач може побачити кількість когорт, та середній коефіцієнт утримання користувачів в системі. В графіку когортного аналізу у вигляді матриці, користувач може побачити коефіцієнт утримання пацієнтів та їх активність по когортам. При наведенні на певну когорту користувач може побачити назву когорти, відсоток утримання користувачів, кількість пацієнтів в системі на та кількість активних декларацій (Рис. 3.9).

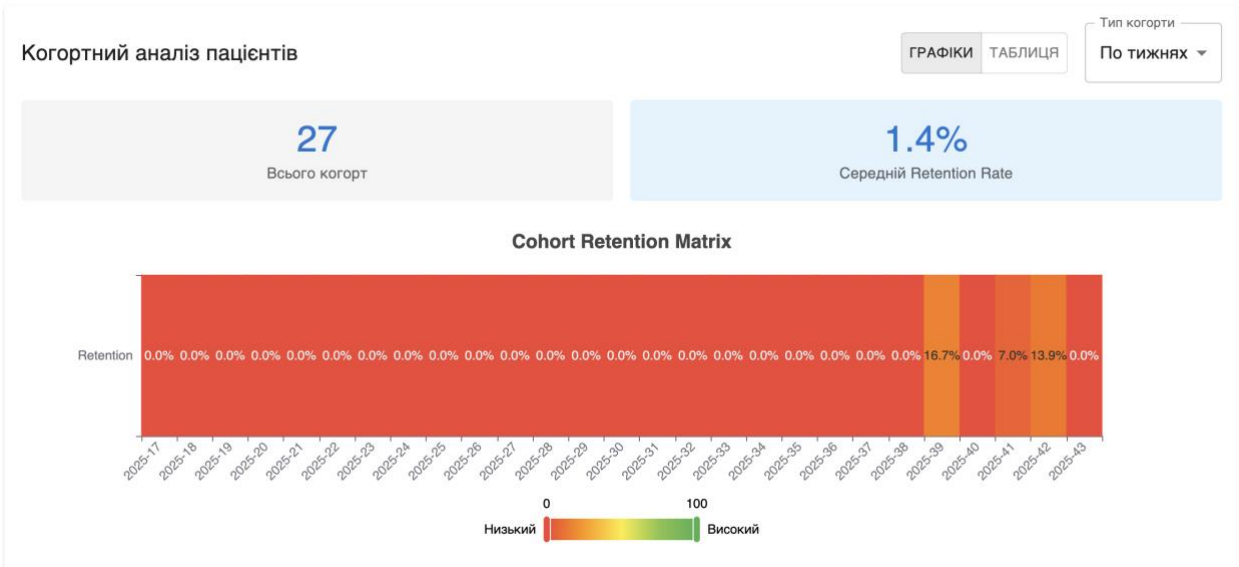


Рисунок 3.9. Сторінка аналітики, графік когортний аналіз пацієнтів

Також в цьому блокові знаходиться графік відвідування по когортам. Де користувач може переглянути відсоток відвідування користувачів системи відповідно до заданого періоду в когортах. При наведенні на один елемент графіку, можна побачити яка саме когорта, відсоток відвідування, кількість пацієнтів, декларацій та активних декларацій в системі за даний період (Рис. 3.10).

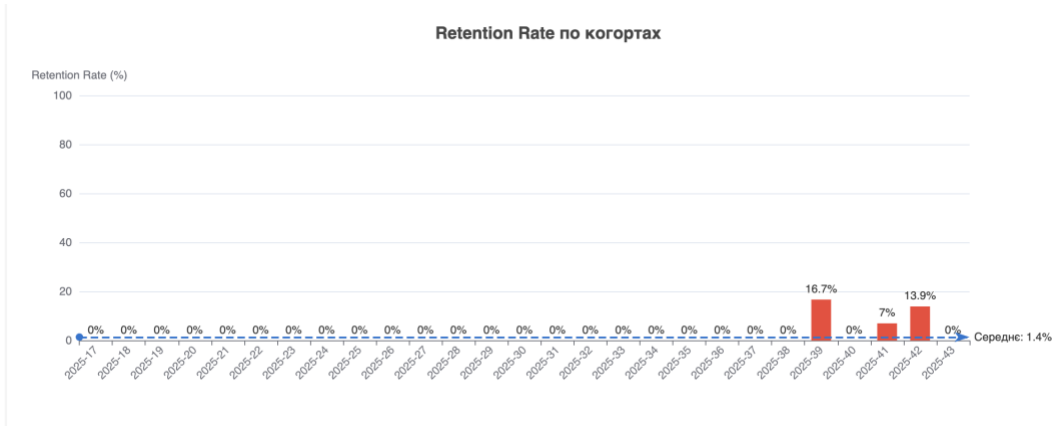


Рисунок 3.10. Графік активності користувачів по когортам

Якщо користувач хоче переглянути інформацію в табличному вигляді, він може переключитися у відповідний режим перегляду інформації по когортам. В таблиці представлена інформація про когорту, кількість пацієнтів, загальна кількість та кількість активних декларацій на відсоток активності користувачів (Рис. 3.11).

Когортний аналіз пацієнтів

ГРАФІКИ

ТАБЛИЦЯ

Тип когорти
По місяцях

7

Всього когорт

2.3%

Середній Retention Rate

Когорта	Кількість пацієнтів	Кількість декларацій	Активних	Retention Rate
2025-04	32	35	0	0.0%
2025-05	96	126	0	0.0%
2025-06	114	136	0	0.0%
2025-07	124	182	0	0.0%
2025-08	127	190	0	0.0%
2025-09	160	256	17	6.6%
2025-10	131	202	19	9.4%

Рисунок 3.11. Когортний аналіз пацієнтів у вигляді таблиці

На аналітичному графіку воронки декларації користувач може переглянути які є вузькі етапи при створенні декларації між лікарем та пацієнтом. Користувач може обрати період за яким він хоче переглянути інформацію. В блоках з

основною інформацією наведено про загальну кількість декларацій, загальну конверсію у відсотках та вузьке місце в етапі створення декларації.

На графіку воронки декларацій користувач може переглянути кожен етап за певний період. Процес створення декларації складається з чотирьох послідовних етапів спочатку пацієнт створює декларацію, потім лікар розглядає її та доповнює необхідною інформацією, після чого лікар підписує декларацію, і нарешті декларація стає активною.

При наведенні на елемент користувач може побачити кількість декларацій, їх зміну у відсотках, відсоток відпаду декларацій. Під графіком розписана основна інформація про воронку (Рис. 3.12).

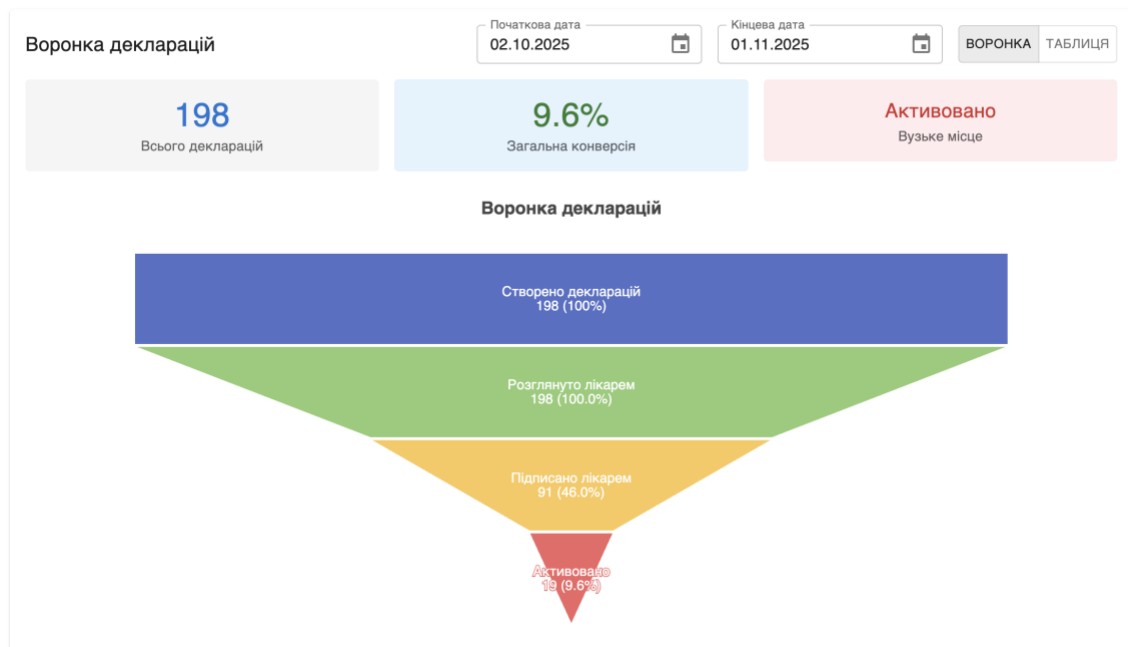


Рисунок 3.12. Графік воронки декларації

Користувач може переглянути інформацію про воронки декларацій у вигляді таблиці. Таблиця включає в себе поля: назва етапу декларації, кількість декларацій на етапові, відсоток конверсії декларації, відсоток відпаду декларації (Рис. 3.13).

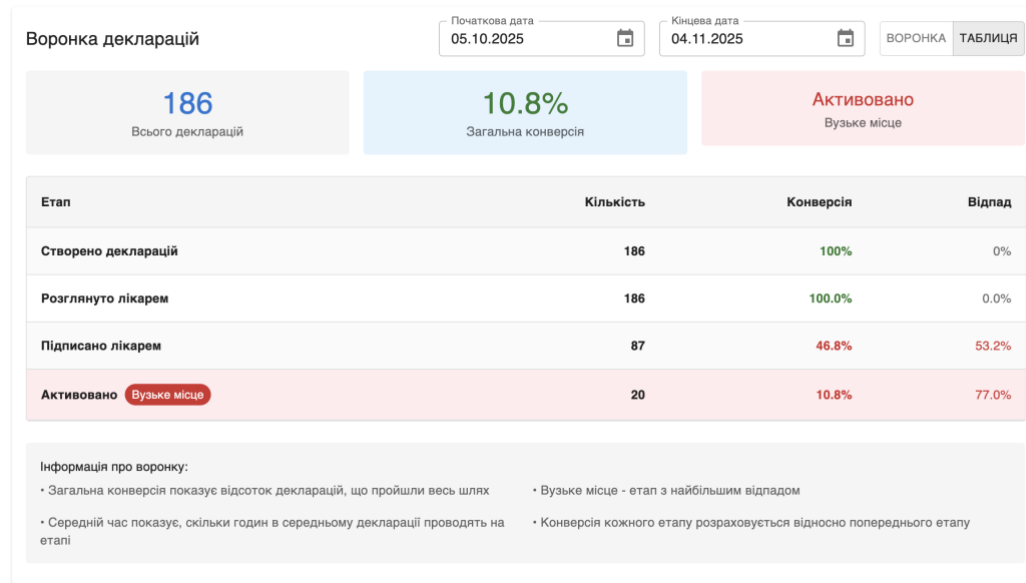


Рисунок 3.13. Таблиця воронки декларації

На аналітичному графіку RFM аналізу лікарів адміністратор може розглянути аналіз кожного лікаря та переглянути, до якої групи він належить. Аналіз відбувається по трьом основним параметрам, а саме R (Recency), що відповідає за активність лікаря (чим більше значення, тим більше часу лікар був неактивний), F (Frequency), яка показує, як часто лікар обробляє декларації (чим вище значення, тим краще), та M (Monetary), що відповідає за кількість активних декларацій у лікаря.

Враховуючи параметри цього методу, можна розділити лікарів на кілька груп, зокрема Hibernating (неактивні лікарі), Need attention (лікарі, які є в зоні ризику), Champions (лікарі з найбільшою активністю), Loyal (стабільно активні лікарі), Potential loyalist (лікарі, які можуть бути найбільш активними) та At risk (лікарі, які були раніше активні, але зараз мають меншу активність). В короткій інформації про аналітику можна побачити кількість користувачів до кожної групи.

Графік RFM діаграми розсіювання параметр R проти F, розмір точки по параметру M показує користувачів відповідно до двох параметрів. При наведенні

на точку можна побачити інформацію про ім'я користувача групи сегмента, та значення параметрів (Рис. 3.14).



Рисунок 3.14. Графік RFM діаграма розгортання

Графік розподілу лікарів по сегментах надає візуалізацію структури користувацької бази. У цьому графіку представлено інформацію про те, який відсоток користувачів належить до кожної категорії сегментації. Інтерактивні можливості графіка дозволяють отримати додаткову інформацію, при наведенні курсора миші на будь-який сегмент користувач може переглянути точну кількість лікарів, що входять до цієї групи, а також її повну назву (Рис. 3.15).

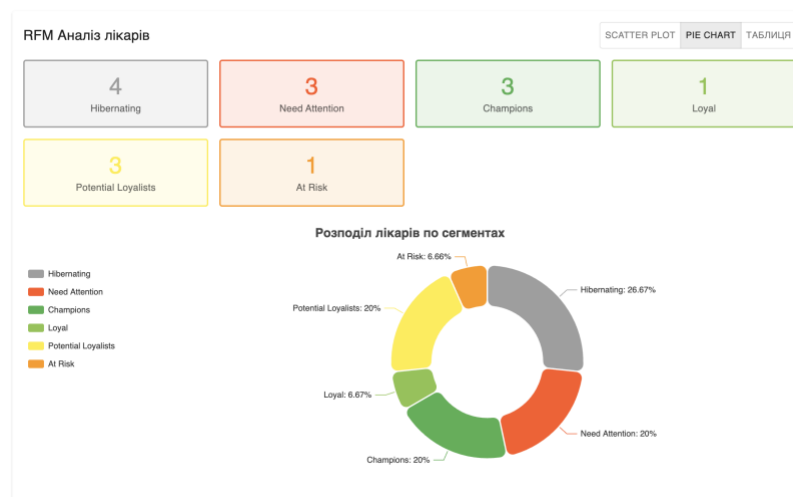


Рисунок 3.15. Графік розподіл лікарів по сегментам

Користувач може переглянути інформацію про лікарів у табличному вигляді, в таблиці представлена інформація про ідентифікатор лікаря, ім'я лікаря, остання активність в днях, кількість декларацій, монетарна що відповідає за кількість активних декларацій, параметри RFM, та RFM Score, з назвою сегмента користувача до якого лікар відноситься (Рис. 3.16).

ID	Ім'я лікаря	Остання активність	Частота	Монетарна	R	F	M	RFM Score	Сегмент
636	Андрій Мельник	15 днів	70	1	5	4	4	544	Champions
639	Юлія Коваль	15 днів	66	1	5	4	4	544	Champions
646	Ігор Данилюк	15 днів	72	3	5	3	2	532	Potential Loyalists
641	Тетяна Савченко	14 днів	62	1	4	5	4	454	Champions
644	Роман Кравець	14 днів	82	4	4	2	1	421	Potential Loyalists
638	Віктор Семенов	12 днів	83	3	4	1	2	412	Potential Loyalists
637	Наталія Поліщук	11 днів	65	1	3	5	4	354	Loyal
633	Марія Шевченко	11 днів	75	5	3	3	1	331	Need Attention
640	Сергій Захарчук	11 днів	81	1	3	2	4	324	Need Attention
645	Оксана Мороз	10 днів	62	4	2	5	1	251	Need Attention
643	Ірина Гончар	10 днів	70	2	2	4	3	243	At Risk
632	Іван Коваленко	10 днів	81	1	2	2	4	224	Hibernating
642	Дмитро Литвин	7 днів	88	2	1	1	3	113	Hibernating

Рисунок 3.16. RFM аналіз лікарів у табличному вигляді

На аналітичному графіку аналізу патернів активності адміністратор може переглянути активність підписання декларацій в періоди: 7, 14, 30, 60, 90 днів. В блокові інформації представлено кореляція декларацій. В графіку активності по дням тижня можна переглянути співвідношення кількості створених декларацій від відсотку відхилення декларацій. При наведенні на елемент графіку можна отримати інформацію про день тижня, загальну кількість декларацій за день та відсоток відхилення декларацій (Рис. 3.17).

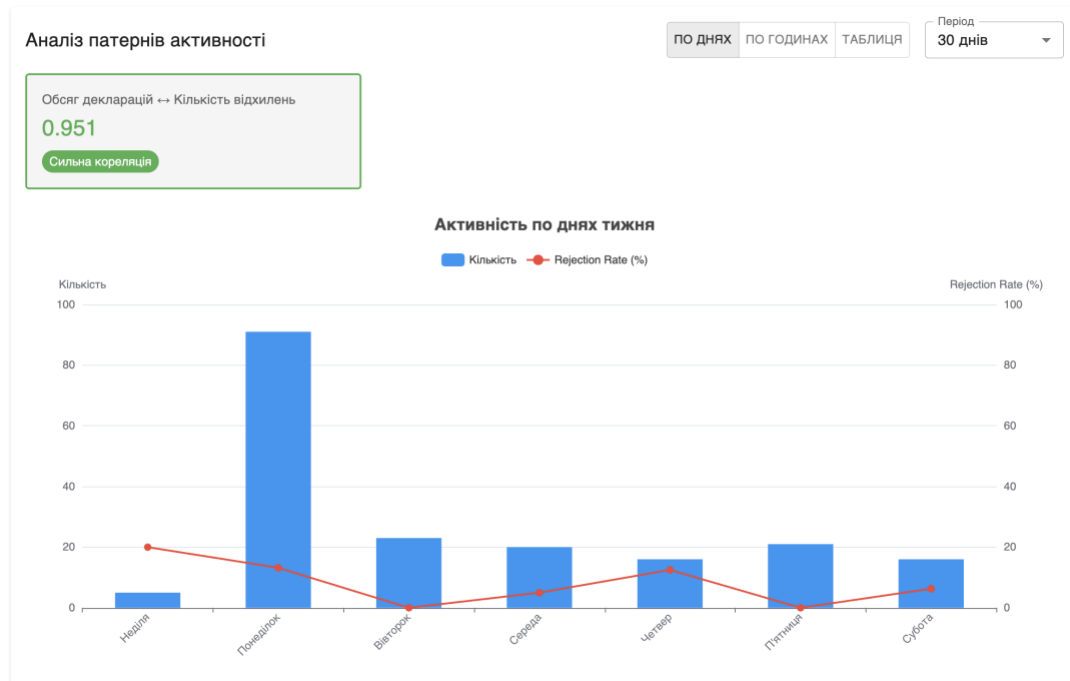


Рисунок 3.17. Графік активності по дням тижня

Графік активності по годинам надає детальну візуалізацію того, в які саме години протягом доби найчастіше здійснювалося створення декларацій у системі. Цей аналітичний інструмент дозволяє відстежити закономірності та піки активності користувачів. При наводженні на будь-який елемент графіку, система автоматично відображає спливаючу підказку з деталізованою інформацією. Ця підказка містить конкретну годину доби, яка відповідає обраному елементу, а також точну кількість декларацій, що були створені саме в цей часовий проміжок. Така інтерактивна функціональність забезпечує та зручний доступ до детальних даних без необхідності переходу до додаткових звітів чи таблиць (Рис. 3.18).

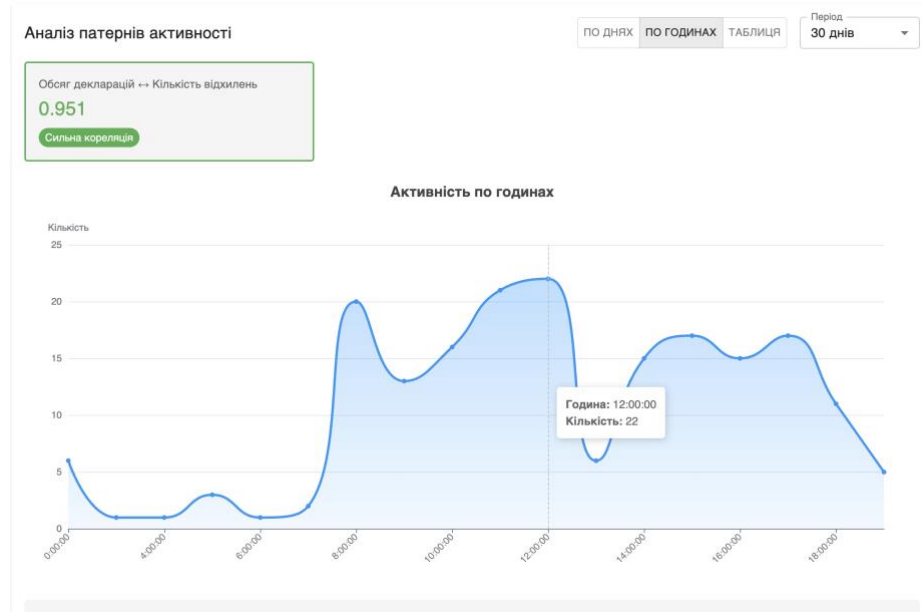


Рисунок 3.18. Графік активності створення декларацій по годинам

Для графіків існує альтернатива у вигляді таблиць з активністю створення декларації по дням тижня. В цих таблицях адміністратор може переглянути назву дня тижня, кількість створених декларацій та відсоток відхилення декларацій (Рис. 3.19).

Аналіз патернів активності

ПО ДНЯХ ПО ГОДИНАХ **ТАБЛИЦЯ** Період 30 днів

Обсяг декларацій ↔ Кількість відхилень
0.951
Сильна кореляція

Активність по днях тижня

День	Кількість	Rejection Rate
Неділя	5	20.0%
Понеділок	91	13.2%
Вівторок	23	0.0%
Середа	20	5.0%
Четвер	16	12.5%
П'ятниця	21	0.0%
Субота	16	6.3%

Активність по годинах

Година	Кількість
0:00:00	6
3:00:00	1
4:00:00	1
5:00:00	3
6:00:00	1
7:00:00	0

Рисунок 3.19. Аналіз патернів активності у вигляді таблиці

В аналітичному графіку аналізу настрою системи по статусам декларацій адміністратор може переглянути аналіз системи за період 7, 14, 30, 60 або 90 днів. У короткій інформації про графік адміністратор має можливість переглянути загальну оцінку тональності, кількість позитивних, нейтральних та негативних подій.

Позитивною подією вважається переведення статусу декларації в активний статус. До негативних подій відноситься розривання декларації зі сторони лікаря або пацієнта на будь-якому етапі підписання. Нейтральними подіями є створення декларації пацієнтом та доповнення інформації зі сторони лікаря без фактичного підписання декларації.

В графіку розподілу категорій по настрою системи адміністратор може побачити візуально стан системи на обраний період. Дані представлені в відсотковому форматі, при наведенні на сегмент графіку адміністратор може отримати детальну інформацію по категорії, таку як: назва категорії, фактична кількість та відсоток від загальної суми (Рис. 3.20).

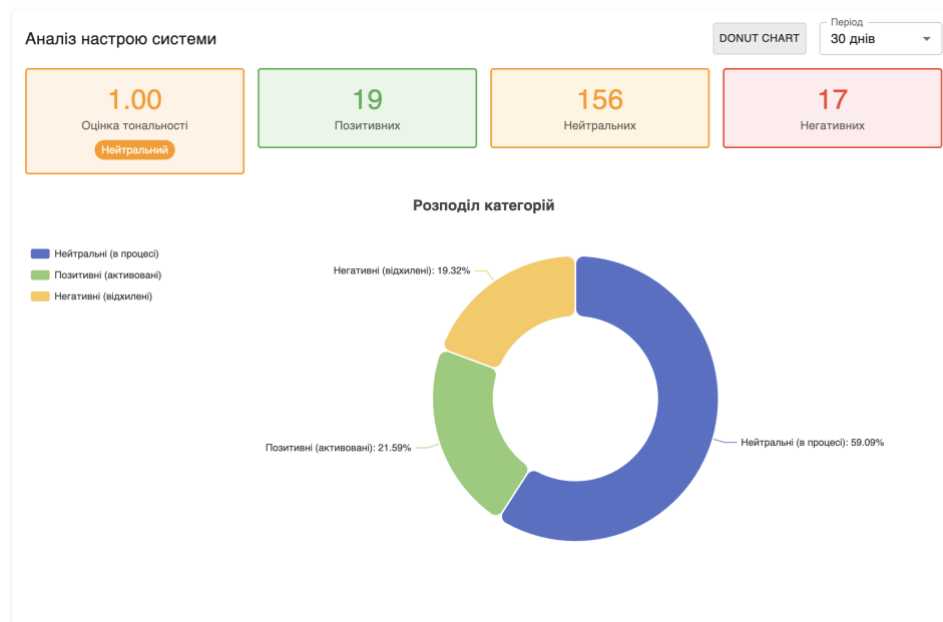


Рисунок 3.20. Графік аналіз настрою системи

Проблема відтоку користувачів від медичних систем є актуальною темою на сьогоднішній час, тому потрібно аналізувати існуючу ситуацію в системі, щоб розуміти, які ланки користувачів від'єднуються від системи та з яких причин. Основними факторами ризику є тривалість пасивності користувача, кількість відхилених декларацій та загальна кількість декларацій.

Користувачів було розбито на три категорії залежно від рівня ризику відтоку. Високий ризик характеризується відсотком від 70 до 100 і сигналізує про те, що пацієнт з високою ймовірністю може відтекти та потребує негайної уваги. Середній ризик варіюється від 40 до 70 відсотків і вказує на помірну ймовірність відтоку пацієнта, тому рекомендується приділити йому увагу. Низький ризик коливається в межах від 0 до 40 відсотків і свідчить про те, що пацієнт має низький ризик відтоку від системи.

На графіку розподілу пацієнтів по рівнях ризику відтоку можна побачити кожену групу ризику розділену у відсотковому форматі. При наведенні на сегмент показується інформація про назву ризику, кількість пацієнтів та відсоток від загальної кількості (Рис. 3.21).

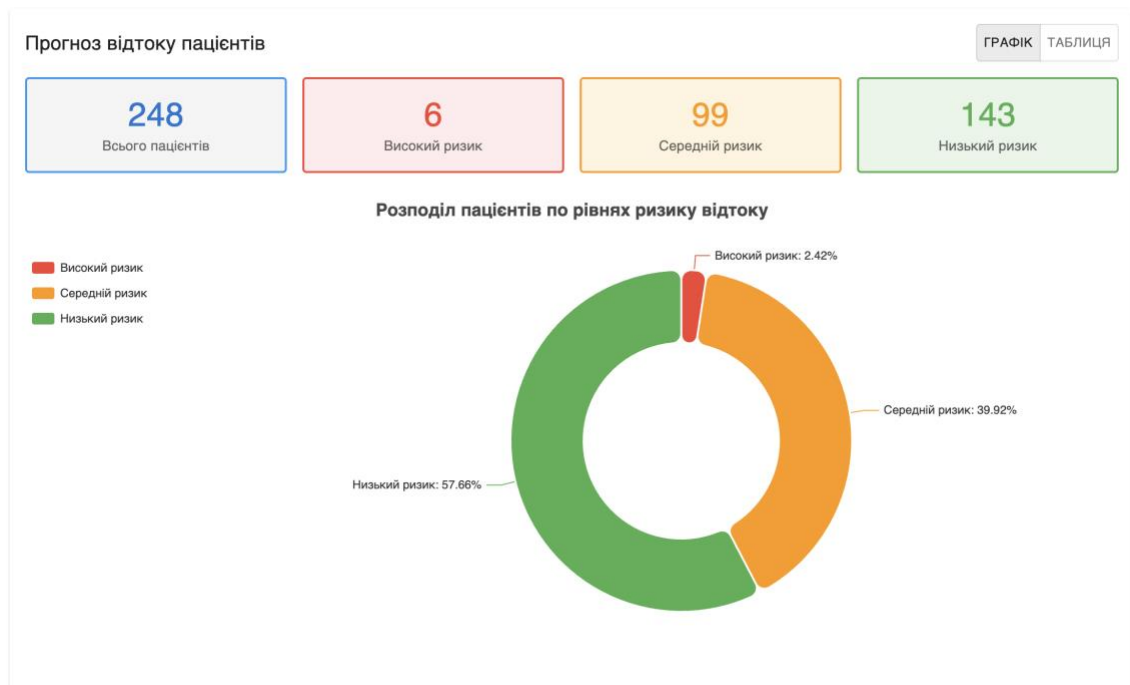


Рисунок 3.21. Графік відтоку пацієнтів

Для перевірки активності кожного з пацієнтів адміністратор може переглянути табличний варіант прогнозу відтоку пацієнтів. В таблиці представлені з самого початку найбільш вразливі користувачі. Поля таблиці включають такі інформацію: електронну адресу пацієнта, день останньої активності, активність включає в себе огляд нотифікацій від лікаря або взаємодії з деклараціями, загальна кількість декларацій, відсоток відхилення декларацій, відсоток ризику відтоку, рівень ризику відтоку (Рис. 3.22).



Рисунок 3.22. Таблиця прогнозу відтоку пацієнтів

В аналітичному графіку рекомендації лікарів адміністратор може переглянути лікарів з найвищою оцінкою в клініці. В графіку рекомендованих лікарів за оцінкою адміністратор може переглянути 10 найкращих лікарів активних лікарів по всім активним лікарням. Максимальна оцінка рекомендації 100 балів, оцінка базується на кількості декларацій, середній час обробки та відсоток успішних декларацій. При наведенні на сегмент графіку відображається інформація: ім'я лікаря, оцінка рекомендації, відсоток успішного підписання

декларацій лікарем, середній час обробки декларацій, загальна кількість декларацій, причина успішності лікаря (Рис. 3.23).

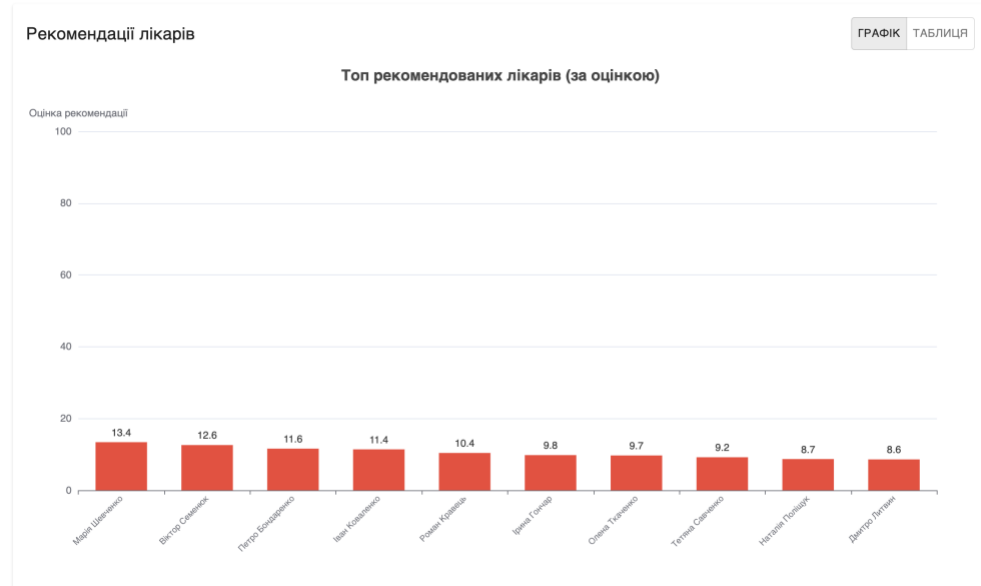


Рисунок 3.23. Графік найкращих лікарів за оцінкою

Для перегляду деталей по кожному лікареві адміністратор може переглянути рекомендації лікарів у вигляді таблиці. В таблиці представлені всі активні лікарі з активних лікарень адміністратора (Рис. 3.24).

В таблиці представлені поля, що характеризують продуктивність лікарів. Номер по списку лікаря відповідає за його місце за продуктивністю в порівнянні з іншими лікарями. Ім'я лікаря вказує на те, яке ім'я використовується в системі. Success Rate показує відсоток успішно оброблених декларацій, де вищий відсоток означає вищу оцінку. Середній час обробки відображає швидкість обробки декларації лікарем, при цьому менший час відповідає вищій оцінці. Поле "Всього декларацій" містить загальну кількість декларацій, які є в лікаря. Оцінка рекомендації виражається у відсотках і дорівнює співвідношенню загальної кількості декларацій до кількості активних декларацій. Причина рекомендації визначається на основі оцінки рекомендації.

Рекомендації лікарів

ГРАФІК ТАБЛИЦЯ

#	Ім'я лікаря	Success Rate	Середній час обробки (год)	Всього декларацій	Оцінка рекомендації	Причина рекомендації
1	Марія Шевченко	6.7%	35.5	75	13.4 / 100	Стабільна робота
2	Віктор Семенюк	3.6%	33.2	83	12.6 / 100	Стабільна робота
3	Петро Бондаренко	4.8%	36.3	83	11.6 / 100	Стабільна робота
4	Іван Коваленко	1.2%	32.4	81	11.4 / 100	Стабільна робота
5	Роман Кравець	4.9%	38.4	82	10.4 / 100	Стабільна робота
6	Ірина Гончар	2.9%	36.9	70	9.8 / 100	Стабільна робота
7	Олена Ткаченко	3.4%	37.8	87	9.7 / 100	Стабільна робота
8	Тетяна Савченко	1.6%	36.6	62	9.2 / 100	Стабільна робота
9	Наталія Поліщук	1.5%	37.2	65	8.7 / 100	Стабільна робота
10	Дмитро Литвин	2.3%	38.3	88	8.6 / 100	Стабільна робота

Рисунок 3.24. Таблиця рекомендації лікарів

В аналітичному графіку аналізу мережі взаємодії адміністратор може переглянути лікарів з найбільшим впливом на пацієнтів. У короткій інформації адміністратор має можливість ознайомитися із загальною кількістю активних лікарів в активних лікарнях, середньою кількістю зв'язків, а також рівнями впливу лікарів.

Високий вплив визначається, коли кількість пацієнтів у лікаря перевищує половину від загальної кількості пацієнтів в клініці. Середній вплив характеризується ситуацією, коли кількість пацієнтів у лікаря становить більше п'ятої частини від загальної кількості пацієнтів в клініці. Низький вплив спостерігається тоді, коли кількість пацієнтів у лікаря є меншою за п'яту частину від загальної кількості пацієнтів в лікарні. В графіку топ-20 лікарів за кількістю зав'язків по параметру кількості активних пацієнтів які належить одному активному лікарю. На графіку зображено кількість зав'язків та імена лікарів з їх значеннями. При наведенні на сегмент графіку відображається ім'я лікаря, кількість його пацієнтів, міра значення взаємодії та рівень впливу лікаря (Рис. 3.25).



Рисунок 3.25. Графік найкращих лікарів за кількістю зав’язків

Для перегляду всіх лікарів та детальнішої інформації про їх взаємодію адміністратор може перейти в режим перегляду таблицею. Таблиця має такі поля: унікальний ідентифікатор користувача, ім’я лікаря, кількість зав’язків, значення впливу лікаря, рівень впливу лікаря в текстовому форматі (Рис. 3.26).

Аналіз мережі взаємодій

15
Всього лікарів

64.7
Середня к-ть зв'язків

15
Високий вплив

0
Середній вплив

0
Низький вплив

ID	Ім'я лікаря	Кількість зв'язків	Degree Centrality	Рівень впливу
635	Олена Ткаченко	77	5.500	Високий вплив
642	Дмитро Литвин	74	5.286	Високий вплив
640	Сергій Захарчук	72	5.143	Високий вплив
638	Віктор Семенов	71	5.071	Високий вплив
632	Іван Коваленко	69	4.929	Високий вплив
644	Роман Кравець	68	4.857	Високий вплив
634	Петро Бондаренко	67	4.786	Високий вплив
633	Марія Шевченко	65	4.643	Високий вплив
646	Ігор Данилюк	64	4.571	Високий вплив
643	Ірина Гончар	61	4.357	Високий вплив

Рисунок 3.26. Таблиця аналізу мережі взаємодії лікарів

Адміністратор зі сторінки аналітики та статистики може перейти на сторінку детальної статистики, яка складається з кількох розділів. Розділ активності користувачів відображає інформацію про загальну кількість активних користувачів в лікарнях. Сторінка процесу декларації містить дані про загальну кількість декларацій та статистику по них. У розділі продуктивності лікарів представлено список лікарів з їх статистикою по деклараціям. Статистика по лікарнях показує кількість лікарень та моменти їх створення. Сторінка повідомлень відображає статистику по нотифікаціям.

На сторінці активності користувачів адміністратор може переглянути активність користувачів в активних лікарнях (Рис. 3.27).

Інформація включає в себе блоки з даними, де активність користувачів показує число користувачів, які заходили в систему в заданий період. Загальні входи відображають загальну кількість входів всіх типів користувачів. Показник "Всього користувачів" містить загальну кількість лікарів. Унікальні користувачі представляють загальне число активних користувачів, які існують в лікарнях.

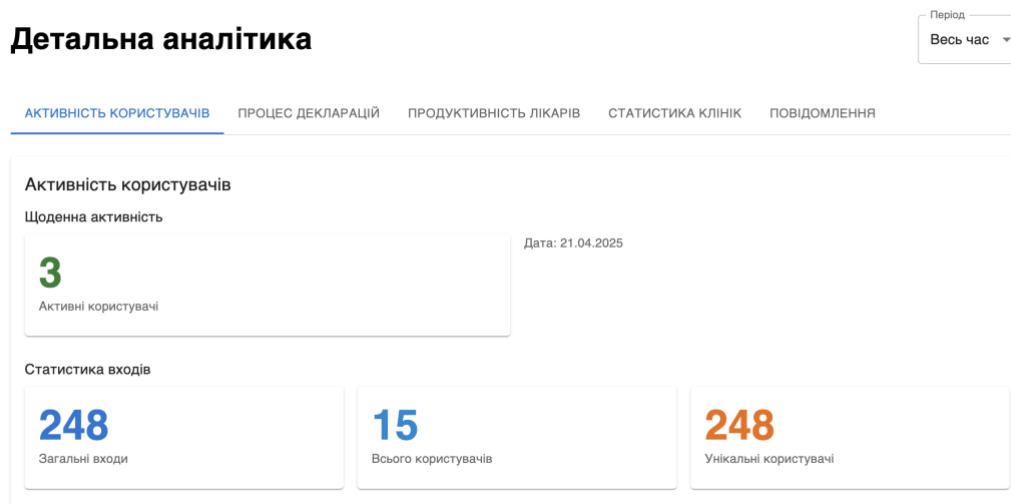


Рисунок 3.27. Статистика активних користувачів

На сторінці процесів декларацій адміністратор може переглянути все що пов'язано з деклараціями (Рис. 3.28-3.29).

Перший блок присвячений процесу обробки декларацій. У ньому розписана загальна кількість декларацій, розділених на три групи очікуючі, завершені та відхилені.

Другий блок демонструє розподіл статусів декларацій у відсотковому форматі. Тут кожен зі статусів декларацій зображено у вигляді індикаторів виконання, а значення відображаються у відсотках. Третій блок містить детальну статистику по статусам. У ньому у вигляді списку відображено назви системних статусів та кількість декларацій, які належать до кожного статусу. Четвертий блок показує час обробки декларації по статусам. У вигляді списку представлено назви системних статусів та середній час виконання декларацій, які належать до відповідного статусу.

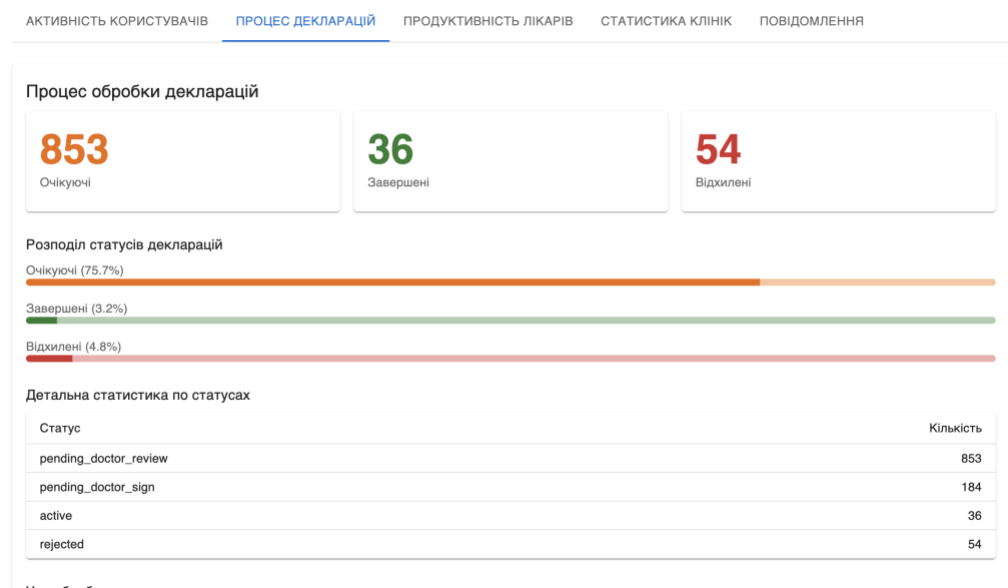


Рисунок 3.28. Статистика процесів обробки декларацій

Час обробки за статусами		
Статус	Кількість	Середній час (год)
pending_doctor_review	853	42.79
pending_doctor_sign	184	44.47
active	36	39.11
rejected	54	35.06

Рисунок 3.29. Статистика часу обробки по статусам декларацій

На сторінці статистики продуктивності лікарів адміністратор може переглянути таблицю з інформацією про кількість та результати підписання декларацій кожного лікаря. Таблиця включає поля, що відображають загальну кількість декларацій, загальну кількість активних декларацій, загальну кількість відхилених декларацій та середній час обробки декларацій в годинах (Рис. 3.30).

АКТИВНІСТЬ КОРИСТУВАЧІВ ПРОЦЕС ДЕКЛАРАЦІЙ **ПРОДУКТИВНІСТЬ ЛІКАРІВ** СТАТИСТИКА КЛІНІК ПОВІДОМЛЕННЯ

Продуктивність лікарів
Загальна кількість лікарів: 15

Ім'я лікаря	Всього декларацій	Активні декларації	Відхилені декларації	Середній час обробки (год)
Дмитро Литвин ID: 642	88	2	0	38.34
Олена Ткаченко ID: 635	87	3	0	37.83
Віктор Семенюк ID: 638	83	3	0	33.24
Петро Бондаренко ID: 634	83	4	1	36.29
Роман Кравець ID: 644	82	4	43	38.41
Іван Коваленко ID: 632	81	1	1	32.44
Сергій Захарчук ID: 640	81	1	3	37.67

Рисунок 3.30. Статистика продуктивності лікарів

На сторінці статистики повідомлень адміністратор може переглянути кількість відправлених повідомлень та кількість їх переглядів. Сторінка складається з блоків, що містять різну інформацію про повідомлення. Блок взаємодії з повідомленнями показує кількість відправлених та прочитаних нотифікацій за період. Показники ефективності представлені у вигляді індикатора виконання з відсотковим значенням. Статистика за період відображається у таблиці, яка містить інформацію про дату, кількість відправлених або відкритих повідомлень та відсоток відкритих нотифікацій. (Рис. 3.31).

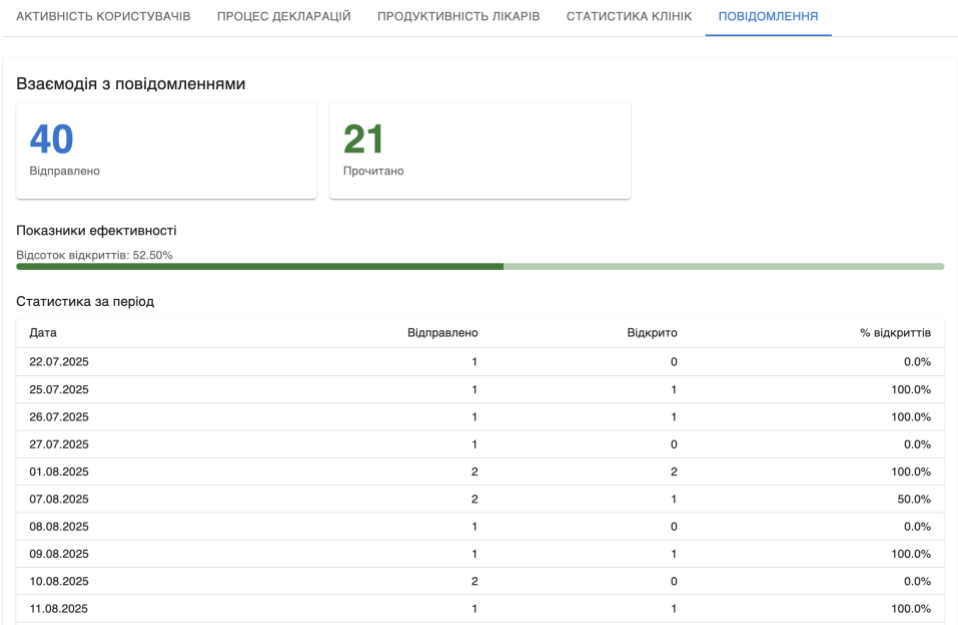
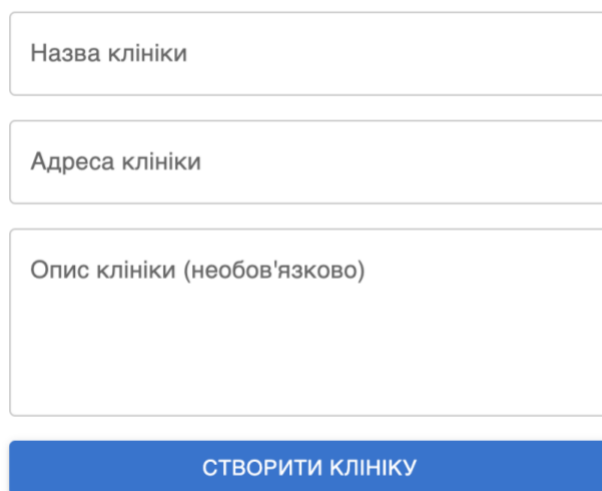


Рисунок 3.31. Статистика по повідомленням

Розроблена система статистичної аналітики забезпечує комплексний моніторинг та аналіз ключових показників ефективності медичного закладу. Реалізований функціонал дозволяє адміністраторам отримувати повну картину роботи системи через інтерактивні графіки та таблиці, які охоплюють різні аспекти діяльності від загальної статистики користувачів до детального аналізу продуктивності лікарів.

3.2 Огляд функціональні можливостей користувачів

Адміністратор може створити лікарню, заповнивши поля назви, адреси та опису, після чого вона відразу стає активною та доступною в системі для нових лікарів та адміністраторів (Рис 3.32).



The form consists of three stacked input fields and a submit button. The first field is labeled 'Назва клініки' (Clinic Name). The second field is labeled 'Адреса клініки' (Clinic Address). The third field is labeled 'Опис клініки (необов'язково)' (Clinic Description (optional)). Below these fields is a blue button with the text 'СТВОРИТИ КЛІНІКУ' (Create Clinic).

Рисунок 3.32. Форма створення лікарні

Для перегляду та управлінням списком юридичних осіб адміністратор може увійти на відповідну сторінку (Рис. 3.33).

На сторінці юридичних осіб є можливість створення нових юридичних осіб в системі, пошуку юридичної особи за допомогою пошуку або фільтрів за такими полями, як назва організації, ЄДРПОУ, статус юридичної особи, електронна адреса юридичної особи та організаційно-правова форма юридичної особи.

Таблиця юридичних осіб включає поля з короткою інформацією про організацію (повна назва організації, скорочена назва організації, організаційно-правова форма), ЄДРПОУ, статус організаційної особи, контактну інформацію (електронна адреса, номер телефону), адресу юридичної особи, дату створення юридичної особи та меню дій, що містить можливості детального перегляду юридичної особи, зміни інформації юридичної особи та видалення юридичної особи з системи.

Юридичні особи

+ ДОДАТИ ОРГАНІЗАЦІЮ

Пошук та фільтрація

🔍 РОЗШИРЕНИЙ ПОШУК

Назва організації

ЄДРПОУ

Статус

🔍 ПОШУК

Організація	ЄДРПОУ	Статус	Контактна інформація	Адреса	Дата створення	Дії
ТОВ "Головна лікарня №10" Головна лікарня ТОВ	12345678	active	clinic-number-ten@gmail.com mobile: +380642575567	Кропивна, Шевченка 1/4	12.08.2025, 15:37	👁️ ✎️ 🗑️

Рисунок 3.33. Сторінка юридичних осіб

Сторінка створення та зміни інформації про юридичну особу мають схожі поля в формі (Рис. 3.34).

Форма розбита на декілька логічних блоків, що полегшують роботу з інформацією. Блок основної інформації містить повну назву організації, скорочену назву організації, організаційно-правову форму, публічну назву, ЄДРПОУ, статус та електронну адресу юридичної особи. Розділ телефонів включає контактну інформацію юридичної особи з можливістю додавання декількох мобільних номерів. Блок адрес містить адресну інформацію юридичної особи та надає можливість додавання декількох адрес.

← 📄 Нова організація

Основна інформація

Повна назва організації *

Скорочена назва *

Організаційно-правова форма *

Публічна назва

ЄДРПОУ *

Статус *

New

Email *

📞 Телефони

+ ДОДАТИ ТЕЛЕФОН

Тип телефону

Мобільний

Номер телефону

📍 Адреси

+ ДОДАТИ АДРЕСУ

Адреса 1 (Регистрація)

Рисунок 3.34. Форма створення юридичної особи

Для перегляду та зміни інформації про лікарню адміністратор може перейти на сторінку власних лікарні (Рис. 3.35).

В списку представлена коротка інформація про лікарню, її опис з датою створення та блоком взаємодії з лікарнею, куди входить редагування інформації про лікарню, перегляд запитів нових адміністраторів в лікарню, вихід з лікарні.

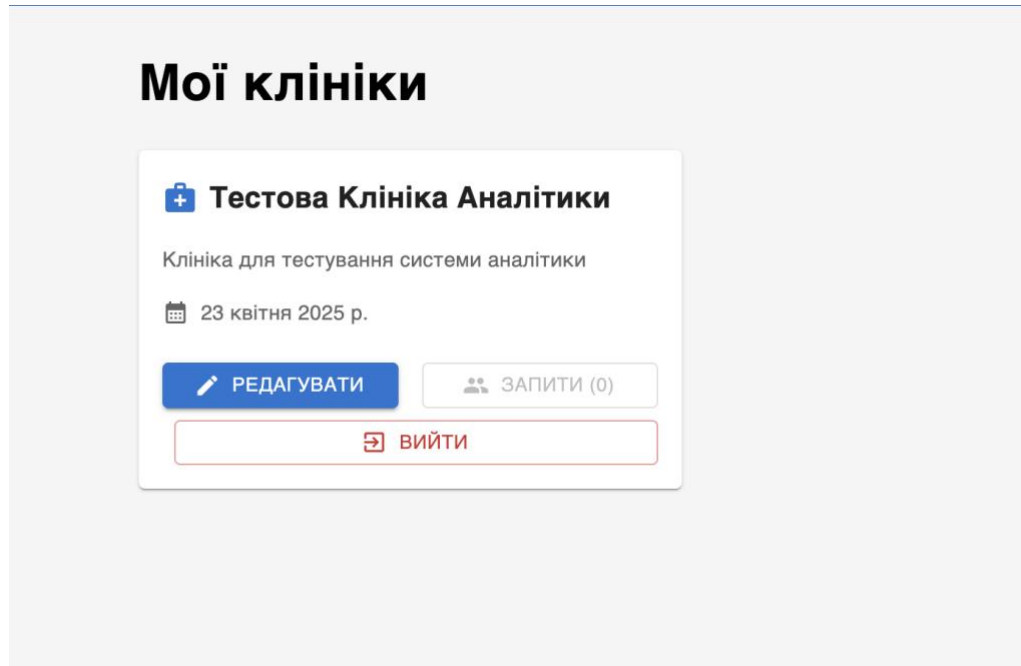


Рисунок 3.35. Сторінка списку лікарень адміністратора

Для комунікації між користувачами в межах лікарні існує сторінка системи нотифікацій. За допомогою системи нотифікацій адміністратор може інформувати адміністраторів або працівників лікарні. Якщо повідомлення відправляється на всю лікарню, кожне повідомлення індивідуально відправляється кожному користувачеві з індивідуальним статусом перегляду. Якщо користувач якому була відправлена нотифікація не прочитав її, даний статус буде бачити адміністратор (Рис.3.36).

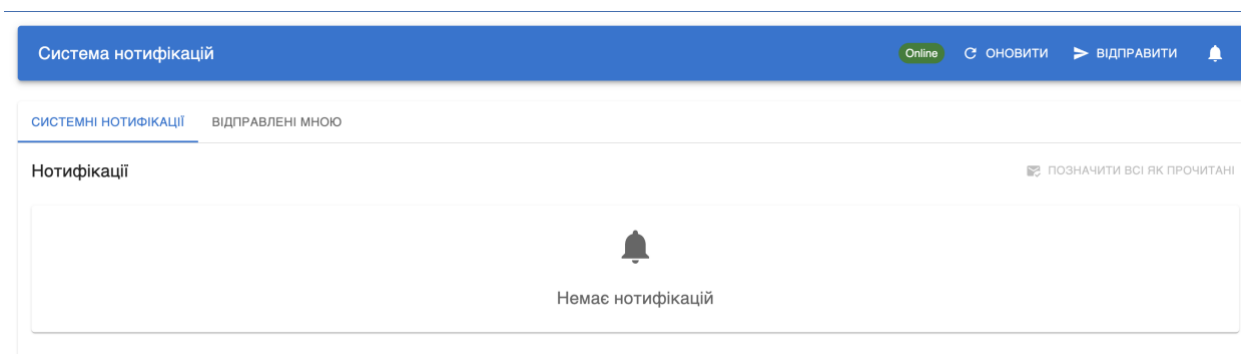


Рисунок 3.36. Сторінка системи нотифікацій

При відправленні нотифікації адміністратором з'являється форма з полями, де потрібно вказати лікарню (назву активної лікарні, до якої буде відправлена нотифікація), обрати відправників (це можуть бути тільки адміністратори лікарні, тільки працівники або адміністратори та працівники одночасно), ввести заголовок нотифікації та написати основний текст повідомлення (Рис.3.37).

Рисунок 3.37. Заповнена форма відправки нотифікації адміністратором

Після успішного відправлення нотифікації адміністратор у пункті сторінки «відправлено мною» може переглянути основну інформацію про нотифікацію, яка включає заголовок нотифікації, текст повідомлення, ім'я та електронну адресу відправника, ім'я та електронну адресу отримувача, назву лікарні в діапазоні якої була відправлена нотифікація, дату та час відправлення нотифікації, а також колір повідомлення (Рис. 3.38).

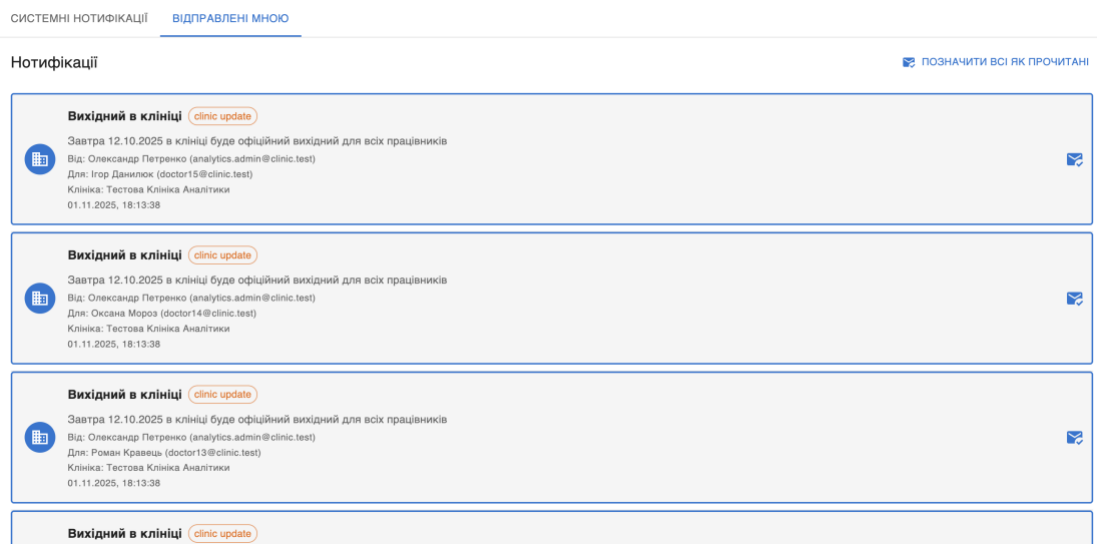


Рисунок 3.38. Сторінка відправлених нотифікацій

Роль лікаря в системі електронних медичних декларацій є центральною та багатоаспектною, оскільки саме медичні працівники становлять основну операційну ланку між пацієнтами та закладами охорони здоров'я. Лікар у контексті даної інформаційної системи виступає як кваліфікований спеціаліст, що надає первинну медичну допомогу та здійснює безперервне медичне обслуговування громадян на основі укладених електронних декларацій.

Функціональна роль лікаря в системі передбачає можливість автономної роботи з базою даних пацієнтів, що подали заявки на приєднання до його медичної практики. Лікар має повноваження переглядати вхідні запити від потенційних пацієнтів, аналізувати їхні базові медичні дані та приймати рішення щодо прийняття або відхилення декларації. Цей процес є критично важливим для

формування збалансованого навантаження на медичного працівника та забезпечення якості медичного обслуговування.

У рамках інформаційної взаємодії лікар отримує доступ до персоналізованої панелі управління, що агрегує всю релевантну інформацію про його базу пацієнтів. Система надає можливість перегляду списку активних декларацій, історії взаємодії з пацієнтами та актуального статусу кожної декларації. Це дозволяє медичному працівнику ефективно планувати свою робочу діяльність та своєчасно реагувати на зміни у складі обслуговуваного контингенту.

Важливим аспектом ролі лікаря є його інтеграція в організаційну структуру медичного закладу. Лікар може бути асоційований з однією лікарнею, що функціонують у системі. Процес приєднання до лікарні передбачає надсилання запиту адміністратору медичного закладу.

Система передбачає розширені можливості для управління професійним профілем лікаря. Медичний працівник може актуалізувати інформацію про свою спеціалізацію, кваліфікацію, контактні дані та регіон практики.

Роль лікаря також включає відповідальність за своєчасне оновлення статусу декларацій та комунікацію з пацієнтами через вбудовані механізми системи.

Таким чином, роль лікаря в системі є комплексною та багатофункціональною, поєднуючи клінічні, адміністративні та інформаційні аспекти медичної практики в умовах цифрової трансформації системи охорони здоров'я. На сторінці панелі лікаря з деклараціями лікар може переглядати існуючі та підписувати нові декларації (Рис.3.39).

Для спрощеного пошуку лікар може використовувати поле пошуку, в яке можна вписати ім'я або номер пацієнта. Для навігації по сторінці є три основних групи. Перша група є сторінкою декларацій на розгляді, в якій представлено декларації, надіслані пацієнтами, які потребують доповнення інформації зі

сторони лікаря. Друга група є сторінкою декларацій до підписання, куди декларація переходить після доповнення інформації про лікаря, отримуючи статус очікування підпису лікаря. Цей етап є фінальним для створення декларації, коли лікар може переглянути, чи всі дані заповнені та чи підходять вони для створення декларації між лікарем та пацієнтом. Третя група є сторінкою списку всіх декларацій, в якій представлений список декларацій з усіма існуючими статусами.

Панель лікаря - Декларації

НА РОЗГЛЯД ДО ПІДПИСАННЯ ВСІ ДЕКЛАРАЦІЇ

Пошук за ПІБ або телефоном оновити

ПІБ пацієнта	Телефон	Дата створення	Статус	Дії
Мельник Сергій Іванович	+380670000187	30.04.2025 16:54	Не розглядає	ЗАВАНТАЖИТИ ДОПОВНИТИ ВІДХИЛИТИ
Поліщук Дмитро Іванович	+380670000120	02.05.2025 08:17	Не розглядає	ЗАВАНТАЖИТИ ДОПОВНИТИ ВІДХИЛИТИ
Коваль Тетяна Іванович	+380670000109	03.05.2025 11:13	Не розглядає	ЗАВАНТАЖИТИ ДОПОВНИТИ ВІДХИЛИТИ
Мельник Юлія Іванович	+380670000231	13.05.2025 16:05	Не розглядає	ЗАВАНТАЖИТИ ДОПОВНИТИ ВІДХИЛИТИ
Мельник Марія Іванович	+380670000127	23.05.2025 11:48	Не розглядає	ЗАВАНТАЖИТИ ДОПОВНИТИ ВІДХИЛИТИ
Коваль Сергій Іванович	+380670000029	24.05.2025 11:30	Не розглядає	ЗАВАНТАЖИТИ ДОПОВНИТИ ВІДХИЛИТИ
Бондаренко Юлія Іванович	+380670000215	25.05.2025 10:54	Не розглядає	ЗАВАНТАЖИТИ ДОПОВНИТИ ВІДХИЛИТИ

Рисунок 3.39. Сторінка панелі лікаря з деклараціями

В будь який момент підписання та існування декларації лікар може відхилити її, вписавши причину відхилення. При відхиленні декларації, вона отримає окремий статус та її неможливо буде відкрити знову. Причина відхилення буде записана в документі декларації (Рис. 3.40).

Відхилення декларації

⚠ Ви відхиляєте декларацію для пацієнта **Поліщук Дмитро**

Причина відхилення*

Неправильно заповнено поле: номер телефону пацієнта

СКАСУВАТИ ВІДХИЛИТИ

Рисунок 3.40. Форма відхилення декларації

При доповненні інформації про декларацію лікаря відкривається модальне вікно з формою, в яку треба заповнити основну інформацію про лікаря та медичний заклад. Після успішного доповнення інформації в декларацію, з'явиться модальне вікно з інформацією про успішне доповнення інформації. Форма включає блоки, що охоплюють різні аспекти інформації. Блок основної інформації містить дату початку дії декларації, дату закінчення дії декларації, причини декларації та опис причини. Окремо представлені розділи підрозділу лікарні та юридичної особи. Блок даних про співробітника включає посаду, ПІБ лікаря, дату початку роботи та можливу дату закінчення роботи. Розділ освіти та кваліфікації лікаря містить інформацію про навчальний заклад, місто навчання, спеціальність, ступінь, номер диплому та дату видачі диплому (Рис. 3.41).

Доповнення декларації

Основна інформація

Дата початку дії 

Дата закінчення дії 

Причина

Опис причини

Підрозділ

Оберіть підрозділ 

Юридична особа

Оберіть юридичну особу 

Рисунок 3.41. Сторінка доповнення декларації

Після успішного доповнення інформації в декларацію, лікар може підписати декларацію. При натисканні відповідної кнопки, з'явиться модальне вікно з інформацією про те з ким підписується декларація. Після підпису декларації, декларація переходить в активний статус, лікар та пацієнт

прив'язуються в системі зав'язків. Дана функціональність надає змогу лікарю надсилати нотифікації особисто пацієнтові (Рис. 3.42).

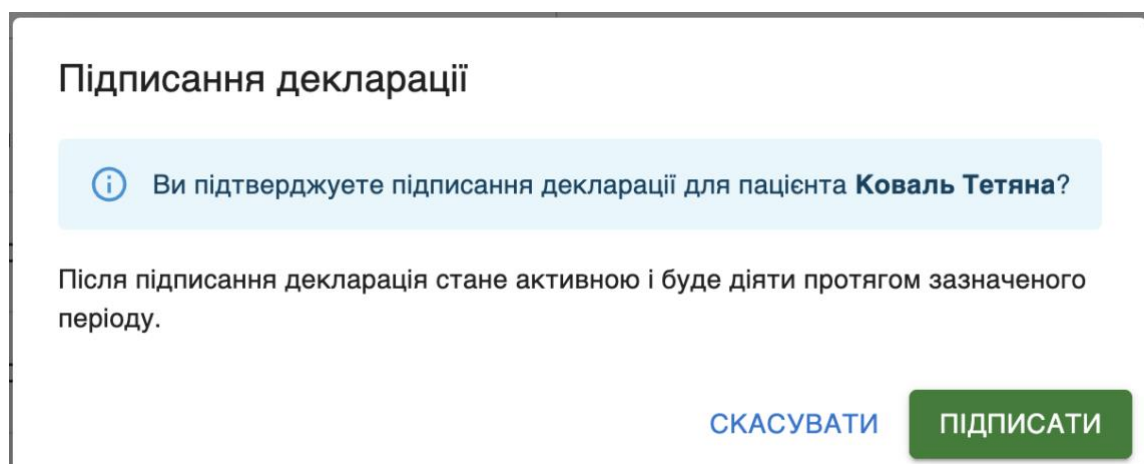


Рисунок 3.42. Модальне вікно підпису декларації

На будь-якому етапі процесу оформлення декларації як лікар, так і пацієнт мають можливість завантажити файл декларації у форматі PDF та детально ознайомитися з усім вмістом документа. Ця функція доступна протягом всього процесу підписання, що дозволяє обом сторонам у будь-який зручний момент отримати повну версію декларації для перегляду, збереження або роздрукування. Таким чином забезпечується прозорість процедури та можливість ретельної перевірки всієї інформації, що міститься в декларації, перед її остаточним підписанням (Рис. 3.43).

ДЕКЛАРАЦІЯ ПРО ВИБІР ЛІКАРЯ, ЯКИЙ НАДАЄ ПЕРВИННУ МЕДИЧНУ ДОПОМОГУ

№ DEC-000082

Дата створення: 03.05.2025

ЗАГАЛЬНА ІНФОРМАЦІЯ

Статус декларації:	АКТИВНА
Дата початку дії:	01.11.2025
Дата закінчення дії:	01.01.2026
Дата підписання:	01.11.2025
Причина:	Форма 1.22 П.3
Опис причини:	Підписано
Сфера дії:	family_doctor

Рисунок 3.43. Приклад файлу декларації

Кожен лікар належить до певної лікарні. Лікар в будь який момент може перейти з однієї лікарні в іншу, перенісши з собою пацієнтів та декларації. При переході аналітична та статистична інформація двох лікарень оновиться (Рис. 3.44).

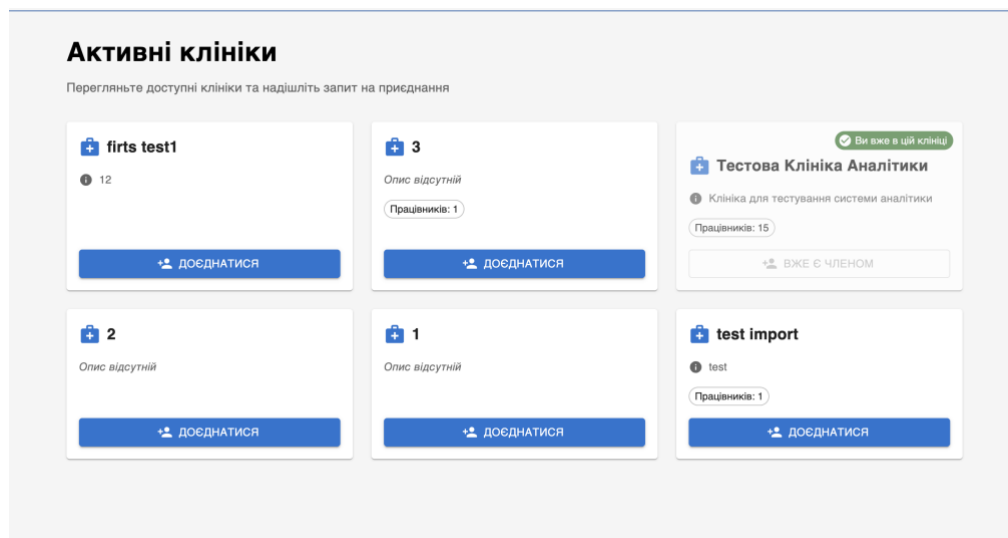


Рисунок 3.44. Сторінка вибору активних лікарень

Роль пацієнта в системі електронних медичних декларацій є фундаментальною та визначальною, оскільки саме пацієнти становлять кінцевих

бенефіціарів цифрової трансформації системи охорони здоров'я. Пацієнт виступає як активний учасник процесу медичного обслуговування, що має широкі можливості для самостійного управління своїми медичними даними та взаємодією з медичними закладами та лікарями.

Центральною функцією пацієнта в системі є можливість укладання електронної декларації про вибір лікаря, що надає первинну медичну допомогу. Цей процес реалізується через інтуїтивно зрозумілий інтерфейс користувача, де користувач може переглядати списки доступних медичних закладів та лікарів.

Пацієнт має постійний доступ до інформації про свою активну декларацію, історію взаємодії з лікарями та медичними закладами, а також статус обробки поданих запитів.

Важливим аспектом ролі пацієнта є можливість управління власним медичним профілем. Громадянин може актуалізувати персональні дані, контактну інформацію, відомості про місце проживання та інші релевантні відомості, що використовуються для забезпечення якості медичного обслуговування

Пацієнт наділяється правом зміни лікаря або медичного закладу в будь-який момент, що відповідає принципам мобільності та гнучкості в системі охорони здоров'я. Процес розірвання існуючої декларації та укладання нової є спрощеним та не вимагає складних адміністративних процедур. Така модель забезпечує конкуренцію між медичними працівниками та стимулює підвищення якості медичних послуг.

Система передбачає механізми комунікації між пацієнтом та лікарем через вбудовані інструменти сповіщень та повідомлень. Пацієнт отримує своєчасну інформацію про важливі події, що стосуються його медичного обслуговування, зміни в роботі лікаря або медичного закладу, а також інші релевантні повідомлення. Така комунікація сприяє підтриманню постійного зв'язку між пацієнтом та системою охорони здоров'я.

На сторінці особистих декларацій пацієнт може переглянути всі свої декларації. Пацієнт може відмінити активні декларації вписавши причину відхилення. Якщо в декларації закінчився термін дії, пацієнту буде відображена інформація (Рис. 3.45).

Рисунок 3.45. Сторінка особистих декларацій пацієнта

Пацієнт може перейти до форми декларації про вибір лікаря через меню системи або через список своїх декларацій. На сторінці створення декларації представлена форма, яка складається з шести етапів заповнення інформації (Рис. 3.46).

Рисунок 3.46. Сторінка створення декларації пацієнтом

Процес формування декларації включає етап персональних даних, де пацієнт вписує свої особисті дані, етап документів для заповнення інформації про документи, етап адреси проживання з інформацією про місця проживання та етап контактних даних, де вказуються особисті контактні дані й контактні дані осіб для екстрених випадків. Далі йде етап вибору лікаря, на якому пацієнт обирає лікаря зі списку з інформацією про ім'я лікаря та кількість вільних місць. Завершальним є етап підтвердження, що містить коротку інформацію про декларацію для перевірки

3.3 Оцінка ефективності розробленої системи

Для тестування системи було використано набір даних, що включає одну тестову лікарню з 15 лікарями, 250 пацієнтів та понад тисячею медичних декларацій, розподілених протягом шести місячних когорт. Дані охоплюють різноманітні сценарії роботи системи, включаючи типові робочі процеси та патерни користувацької активності.

Модуль прогнозування навантаження продемонстрував здатність оцінювати майбутні обсяги роботи системи на основі аналізу історичних даних за останні тридцять днів. На імпортованих протягом 180 днів система визначає середнє навантаження на рівні приблизно 6 декларацій на день, виявляє пікові дні тижня, зокрема понеділок та вівторок з показниками активності 20% та 18% від тижневого обсягу відповідно, що на 40-26% вище середнього рівня 14,3% на день, та встановлює тренд зміни навантаження на основі лінійної регресії з розрахунком нахилу та перетину прямої. Важливою характеристикою системи є автоматичне визначення рівня довіри до прогнозу на основі коефіцієнта варіації, високий рівень присвоюється при коефіцієнті варіації менше 20%, середній при значеннях від 20% до 40%, низький при коефіцієнті понад 40%. Система комбінує результати експоненціального згладжування з вагою 40% та лінійної регресії з вагою 60% для формування фінального прогнозу, що дозволяє враховувати як

короткострокові коливання, так і довгострокові тренди в роботі системи медичних декларацій.

Система автоматичного виявлення помилок успішно ідентифікує відхилення від типових патернів роботи системи. В тестових даних закладено 4 аномальні дні для перевірки коректності виявлення: 2 дні зі сплесками активності з кількістю 55 та 50 декларацій відповідно, що становить перевищення норми приблизно в 9-10 разів, один день з різким падінням активності до 2 декларацій, що становить зниження на 60-70% від очікуваного значення, та один день з аномально високим відсотком відхилення декларацій на рівні 60% при 20 деклараціях. Система автоматично визначає рівень серйозності кожної аномалії на основі Z-показника, класифікуючи відхилення з показником понад 4 як критичні, від 3 до 4 як попередження. Додатково система виявляє помилок на рівні окремих лікарів у одного спеціаліста закладено відсоток відхилених декларацій на рівні 50% що значно перевищує середній показник по системі 20%, в другого лікаря середній час обробки становить від 100 до 200 годин, що перевищує типову норму 1-72 години в 2-3 рази, та у третього лікаря зафіксовано низьку активність на рівні 4 декларацій.

Модуль RFM-аналізу забезпечив сегментацію лікарів за трьома ключовими параметрами давність останньої активності, частота роботи та кількість успішно активованих декларацій. На основі імпортованих даних система класифікувала лікарів по семи сегментах: до сегменту «Чемпіони» віднесено 3 лікарів (20%) з показниками активності від 60 до 80 декларацій та давністю останньої активності не більше 7 днів, сегмент «Лояльні» об'єднує 2 лікарів (13,3%) з кількістю декларацій від 40 до 45 та давністю активності від 10 до 14 днів, «Потенційно Лояльні» включає 2 нових лікарів (13,3%) з кількістю декларацій від 10 до 12 та давністю активності не більше 4 днів, до категорії «В Зоні Ризику» потрапили 2 лікарі (13,3%), які не проявляли активності від 65 до 70 днів при історичній кількості від 30 до 35 декларацій, сегмент «Не Можемо Їх Втратити» включає 1

лікаря (6,7%) з давністю активності 95 днів та історичним показником 55 декларацій, «Сплячі» включає 2 лікарів (13,3%) з давністю активності понад 125 днів та кількістю декларацій від 6 до 8, решта 3 лікарі (20%) з різними аномаліями в роботі віднесено до категорії «Потребують Уваги».

Система когортного аналізу успішно групує декларації за періодом їх створення та розраховує показники утримання для кожної когорти. Імпортовані дані включають шість місячних когорт декларацій з розподілом найстаріша когорта шість місяців тому містить 120 декларацій, п'ять місяців тому 140 декларацій, чотири місяці тому 160 декларацій, три місяці тому 180 декларацій, два місяці тому 200 декларацій, та найновіша когорта один місяць тому 200 декларацій. Розподіл статусів декларацій включає 500 активних (50%), 150 на розгляді (15%), 100 на підписанні (10%), 200 відхилених (20%) та 50 припинених (5%), що дозволяє системі когортного аналізу виявляти динаміку зміни утримання користувачів у часі.

Аналіз воронки конверсії відслідкував проходження декларацій через чотири ключові етапи обробки. На етапі створення зафіксовано 100% декларацій, всі декларації перейшли до етапу розгляду лікарем, оскільки система автоматично призначає декларації на обробку, 75,4% досягли етапу підписання зі статусами очікування підпису від лікаря, активні, відхилені, та 44,4% досягли фінального етапу активації. Найбільший відвал у 24,6% зафіксовано між розглядом та підписанням, що може вказувати на складнощі в процесі прийняття рішення лікарями або недостатність документів від пацієнтів. Відвал на етапі від підписання до активації склав 31%, що включає 200 відхилених (17,7%) та 50 припинених (4,4%) декларацій, а також 100 декларацій (8,9%), що очікують на підписання.

Отримані результати відповідають критеріям ефективності медичних інформаційних систем. Показник конверсії 44,4% перевищує типовий показник (35-40%), що свідчить про ефективність розробленого рішення. Відвал 24,6% між

етапами розгляду та підписання знаходиться в межах допустимих значень 20-30%, які є типовими для систем обробки медичної документації через причини необхідності додаткової перевірки, відмови пацієнтів або невідповідності документів. Коефіцієнт варіації менше 20% для модуля прогнозування відповідає критеріям високої статистичної надійності, а використання Z-показника з порогом 3 забезпечує оптимальний баланс між виявленням реальних аномалій та уникненням хибних спрацювань. Сегментація лікарів виявила 33,3% високоефективних користувачів у сегментах «Чемпіони» та «Лояльні», що перевищує типовий показник 25-30% для медичних систем, підтверджуючи успішність впровадження розробленого рішення.

Система розпізнавання патернів виявила закономірності в активності користувачів на основі імпортованих даних піковими годинами визначено період з 10:00 до 11:00 з показниками 12% та 13% відповідно, помірна активність спостерігається з 8:00 до 9:00 та з 15:00 до 16:00 по 8% кожен період, тоді як найменша активність припадає на ранкові години з 0:00 до 7:00 по 1% на годину та пізній вечір після 20:00 менше 2% сумарно. Розподіл по днях тижня демонструє найвищу активність у понеділок 20% та вівторок 18%, найнижчу в неділю 6% та суботу 8%.

Кореляційний аналіз розраховує коефіцієнт кореляції Пірсона між обсягом декларацій та кількістю відхилень по днях тижня з автоматичною класифікацією сили зв'язку сильна кореляція при абсолютному значенні понад 0,7, помірна від 0,3 до 0,7, слабка менше 0,3. Мережевий аналіз оцінює вплив кожного з 15 лікарів на основі кількості активних зв'язків з пацієнтами, розраховуючи показник центральності за ступенем як відношення унікальних пацієнтів до загальної кількості пацієнтів, з класифікацією впливу високий при показнику понад 0,5, середній від 0,2 до 0,5, низький менше 0,2. За результатами тестування можна констатувати успішне досягнення поставлених завдань дослідження.

Система прогнозування навантаження реалізована з використанням комбінованого підходу експоненціального згладжування та лінійної регресії з автоматичним визначенням рівня довіри до прогнозу на основі коефіцієнта варіації високий при значенні менше 20%, середній від 20% до 40%, низький понад 40%. Система виявлення помилок успішно ідентифікує відхилення за допомогою методів Z-показника з порогом 3 для виявлення помилки та міжквартильного розмаху для визначення викидів, автоматично класифікуючи їх за рівнем серйозності критичний при Z-показнику понад 4 та попередження від 3 до 4. RFM-аналіз забезпечив сегментацію 15 лікарів по семи категоріях на основі квінтільного розподілу показників давності, частоти та монетарного значення. Когортний аналіз дозволяє відслідковувати динаміку утримання користувачів через розрахунок відношення активних декларацій до загальної кількості для кожної з шести місячних когорт. Аналіз воронки конверсії виявляє етапи з найбільшим відвалом через порівняння кількості декларацій на чотирьох послідовних стадіях обробки.

Висновки до 3 розділу

У третьому розділі дипломної роботи було розглянуто систему аналітики, що включає десять спеціалізованих модулів для глибокого аналізу роботи системи.

Розглянуто комплексну рольову модель доступу з трьома типами користувачів, кожен з яких має чітко визначені функціональні можливості відповідно до своїх завдань. Адміністратори отримали повний спектр інструментів управління медичними закладами, включаючи створення та редагування лікарень, управління юридичними особами, обробку запитів на приєднання та систему внутрішніх нотифікацій. Лікарі забезпечені спеціалізованою панеллю для ефективної обробки декларацій з можливістю

доповнення інформації, підпису або відхилення документів з обов'язковим зазначенням причин.

Розроблена система вирішує поставлені завдання щодо автоматизації процесів управління медичними деклараціями, забезпечує аналітичний інструментарій для прийняття управлінських рішень та створює механізми взаємодії між усіма учасниками процесу медичного обслуговування.

Оцінено ефективність системи систематичного аналізу де було визначено практичну цінність розробленої системи що підтверджує можливість її використання адміністраторами Електронної системи охорони здоров'я, керівниками медичних закладів та органами управління охороною здоров'я для підвищення ефективності функціонування системи електронних медичних декларацій.

ВИСНОВКИ

У дипломній роботі вирішено науково-практичне завдання розробки системи статистичного аналізу та ETL-трансформації даних електронних медичних декларацій для підвищення ефективності управління медичними інформаційними системами.

Проведено комплексний аналіз проблем організації та обробки даних електронних медичних декларацій у системі охорони здоров'я України. Виявлено ключові виклики: необхідність прогнозування навантаження на систему, своєчасного виявлення похибок, сегментації користувачів та забезпечення їх утримання. Досліджено специфіку життєвого циклу електронної медичної декларації та особливості поведінки користувачів системи.

Проаналізовано наявні методи організації та аналізу медичних даних, включаючи методи прогнозування часових рядів, статистичні методи виявлення похибок, методи сегментації, аналіз воронки конверсії, кореляційний аналіз та мережевий аналіз. Обґрунтовано доцільність застосування комбінованих підходів для підвищення точності та надійності аналітичних висновків.

Розроблено та застосовано гібридний алгоритм прогнозування навантаження на систему медичних декларацій, що поєднує переваги експоненціального згладжування для короткострокових коливань та лінійної регресії для довгострокових трендів з динамічними вагами. Алгоритм забезпечує адаптивне прогнозування з урахуванням як поточних змін, так і загальних тенденцій у даних.

Створено систему автоматичного виявлення похибок в часових рядах медичних даних, що використовує паралельне застосування Z-Score та IQR методів. Система забезпечує надійне виявлення як статистичних викидів у даних з нормальним розподілом, так і похибок у даних зі скошеним розподілом, що підвищує універсальність та точність знаходження проблемних ситуацій.

Адаптовано та застосовано RFM-аналіз для медичного контексту, що дозволяє сегментувати лікарів за критеріями активності, частоти створення декларацій та загальної кількості активних декларацій. Розроблено систему класифікації лікарів на категорії, що забезпечує диференційоване управління різними групами користувачів.

Розроблено систему когортного аналізу для відстеження утримання користувачів у часі з групуванням за місяцем першої активності. Система дозволяє ідентифікувати критичні періоди відтоку користувачів та оцінювати ефективність заходів щодо їх утримання через розрахунок показників коефіцієнт утримання для кожної когорти.

Застосовано аналіз воронки конверсії для процесу обробки медичних декларацій з етапами: створення, огляд лікарем, підписання лікарем та активація. Система розраховує коефіцієнти конверсії між етапами та виявляє вузькі місця в процесі, що дозволяє цілеспрямовано оптимізувати проблемні ділянки.

Реалізовано кореляційний аналіз для виявлення взаємозв'язків між метриками системи з використанням коефіцієнта Пірсона та мережевий аналіз з метрикою центральності ступеня для візуалізації розподілу навантаження між лікарями. Ці інструменти забезпечують виявлення прихованих закономірностей та оптимізацію структури надання медичних послуг.

Розроблено ієрархічну архітектуру системи аналітики з контролем доступу на основі організаційної структури медичних закладів. Система забезпечує гранульований доступ до даних відповідно до ролі користувача, що гарантує конфіденційність інформації та відповідність нормативним вимогам.

Обрано та обґрунтовано технологічний стек для реалізації системи: NestJS для серверної частини з підтримкою TypeScript, PostgreSQL для зберігання та обробки даних з використанням оптимізованих SQL-запитів, React для клієнтської частини. Такий вибір забезпечує баланс між продуктивністю, масштабованістю та зручністю розробки.

Створено комплексний веб-інтерфейс системи аналітики з інтерактивними панелями візуалізації, що включають графіки для різних типів аналізу, графіки прогнозування, таблиці сегментації користувачів та діаграми воронки конверсії. Інтерфейс забезпечує інтуїтивно зрозумілий доступ до аналітичних інструментів для користувачів з різним рівнем технічної підготовки.

Практична цінність роботи полягає у створенні повнофункціональної системи аналітики, що може бути застосована адміністраторами ЕСОЗ та керівниками медичних закладів для прийняття обґрунтованих управлінських рішень щодо оптимізації роботи з електронними медичними деклараціями, планування ресурсів, виявлення проблемних ситуацій та підвищення якості медичного обслуговування населення.

Розроблена система демонструє ефективність застосування методів інтелектуального аналізу даних у медичній галузі та може слугувати основою для подальшого розвитку аналітичних можливостей електронних систем охорони здоров'я України.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Петрук О.О., Савіцький Р.С., ОСОБЛИВОСТІ МЕДИЧНИХ ІНФОРМАЦІЙНИХ СИСТЕМ (МІС) В УКРАЇНІ // Сучасні Інформаційні Системи і Технології Матеріали VII Всеукраїнської науково-практичної інтернет-конференції студентів, аспірантів та молодих вчених за тематикою «Сучасні комп'ютерні системи та мережі в управлінні», 29 листопада 2024 р Хмельницький. – С. 49-51
2. Петрук О.О., Савіцький Р.С., СТРАТЕГІЇ ОБРОБКИ ТА ТРАНСФОРМАЦІЇ МЕДИЧНИХ ДАНИХ У СИСТЕМАХ ОХОРОНИ ЗДОРОВ'Я // Львівський науковий форум матеріали XV Міжнародної Науково-практичної Конференції, практичні та теоретичні питання розвитку науки та освіти 29-30 серпня 2025 р. С. 107-108.
3. Електронні медичні записи в ЕСОЗ. Міністерство охорони здоров'я України. 16.04.2024. URL: <https://moz.gov.ua/uk/elektronni-medichni-zapisi-v-esoz> (дата звернення: 25.09.2025).
1. Пайтелемонов А. Як знайти свою декларацію з лікарем: все що потрібно знати. Doc.co.ua. 03.04.2025. URL: <https://doc.co.ua/medychna-reforma/yak-perevirytu-deklaratsiiu-z-likarem/> (дата звернення: 16.09.2025).
5. Кадрові ресурси системи охорони здоров'я в Україні / Д. Богдан та ін. ; ред.: Ю. Середа, С. Глушик. 2019.
6. Augustine M., Anita J. Real-time Anomaly Detection Systems. Chidiebere Joshua. 2025. С. 1–5. URL: https://www.researchgate.net/publication/391572570_Real-time_Anomaly_Detection_Systems (дата звернення: 08.10.2025).
7. Patient Churn Prediction And Retention Strategies In Healthcare. Byte IQ. 09.05.2023. URL: <https://www.byteiq.com.au/patient-churn-prediction-and-retention-strategies-in-healthcare/> (дата звернення: 08.10.2025).

8. Munoz-Gama J., Martin N. Process Mining for Healthcare: Characteristics and Challenges. *Journal of Biomedical Informatics*. 2022. С. 1–6. URL: <https://doi.org/10.1016/j.jbi.2022.103994>.

9. Santos A., Balancieri R. Process mining in healthcare: a tertiary study. *BMC Medical Informatics and Decision Making*. 2025. Т. 25. С. 2–11. URL: <https://doi.org/10.1186/s12911-025-02967-z>.

10. Методи прогнозування в розробці управлінських рішень. Learn.ztu.edu.ua. URL: https://learn.ztu.edu.ua/pluginfile.php/400380/mod_resource/content/0/%D0%9B%D0%B5%D0%BA%D1%86%D1%96%D1%8F%205.pdf (дата звернення: 14.10.2025).

11. Exponential Smoothing - General Introduction. TIBCO Statistica. URL: https://docs.tibco.com/pub/stat/14.0.1/doc/html/UsersGuide/user-guide/exponential-smoothing-general-introduction.htm?utm_source=chatgpt.com (дата звернення: 14.10.2025).

12. Chen J. Exponential Moving Average (EMA): Definition, Formula, and Usage. Investopedia. 09.09.2025. URL: <https://www.investopedia.com/terms/e/ema.asp> (дата звернення: 16.10.2025).

13. Ларкін О. МЕТОДИ РЕГРЕСІЙНОГО АНАЛІЗУ ТА ЇХ ЗНАЧЕННЯ В АНАЛІЗІ ДАНИХ КЛІНІЧНИХ ДОСЛІДЖЕНЬ. Науково-популярний альманах «Математика та інформатика навколо нас». 2024. В. 8. С. 52–55. URL: <https://dspace.vspu.edu.ua/items/315dbf1d-98ac-4236-bf07-69be4712f962> (дата звернення: 31.10.2025).

14. Simple Linear Regression. Online.stat.psu.edu. URL: <https://online.stat.psu.edu/stat200/book/export/html/244> (дата звернення: 15.10.2025).

15. Мартиненко О. Дослідження методів виявлення аномальних ситуацій у бізнес-процесах медичних закладів. Харківський національний університет радіоелектроніки. 2025. С. 20–59.

16. Кравченко В. Модель Альтмана. LivingFp. 10.05.2022. URL: <https://livingfo.com/model-altmana/> (дата звернення: 07.10.2025).

17. Liao L. An introduction of Normal distribution and Carl Fredrich Gauss. Medium. 21.04.2021. URL: <https://lamar-liao.medium.com/an-introduction-of-normal-distribution-and-carl-fredrich-gauss-373385dfb159> (дата звернення: 14.10.2025).

18. Sharma R. Interquartile Range (IQR): Meaning, Formula, Calculation & Examples. Vedantu. URL: <https://www.vedantu.com/maths/interquartile-range> (дата звернення: 14.10.2025).

19. Bhandari P. How to Find Outliers | 4 Ways with Examples & Explanation. Scibbr. 30.11.2021. URL: <https://www.scribbr.com/statistics/outliers/> (дата звернення: 14.10.2025).

20. Administering and Working with Enterprise Profitability and Cost Management. Docs oracle. URL: https://docs.oracle.com/en/cloud/saas/enterprise-profitability-cost-management-cloud/pcmpl/insights_metrics_IQR.html (дата звернення: 13.10.2025).

21. Organization, S. D. I. W. C. “Analyzing Patient Value by Modifying RFM Model With Consideration of The Limitation of Service Throughput: An Investigation of Dental Health Care.” URL: https://www.academia.edu/14371447/Analyzing_Patient_Value_by_Modifying_RFM_Model_With_Consideration_of_The_Limitation_of_Service_Throughput_An_Investigation_of_Dental_Health_Care

22. Shivkumar M. What is Cohort Analysis? Strategies to Boost Retention. CleverTap. 25.06.2025. URL: <https://clevertap.com/blog/cohort-analysis/> (дата звернення: 14.10.2025).

23. Funnel Marketing For Healthcare Clinics: Comprehensive Guide. Mediboost. 09.09.2025. URL: <https://www.mediboost.com.au/funnel-marketing/> (дата звернення: 25.10.2025).

24. Turnay S. Pearson Correlation Coefficient (r) | Guide & Examples. Scibbr. 13.05.2022. URL: <https://www.scribbr.com/statistics/pearson-correlation-coefficient/> (дата звернення: 29.10.2025).

25. Schober P., Boer C., Schwarte L. Correlation Coefficients: Appropriate Use and Interpretation. Anesth Analg.. 2018. URL: <https://doi.org/10.1213/ANE.0000000000002864>.

26. Соловійов В. Ранжування економічних суб'єктів за їх «важливістю» у системі. Теорія та методика навчання математики, фізики, інформатики. Кривий Ріг, 2015. Т. 13, вип. 3. С. 202–205. ISSN 2309-1479.

27. Coefficient of Variation – Definition, Formula, and Solved Examples. TestBook. URL: <https://testbook.com/maths/coefficient-of-variation> (дата звернення: 21.10.2025).

28. Pieter P. Exponential Moving Average. Tttapa. URL: <https://tttapa.github.io/Pages/Mathematics/Systems-and-Control-Theory/Digital-filters/Exponential%20Moving%20Average/Exponential-Moving-Average.html> (дата звернення: 21.10.2025).

29. Introduction to EMA. CodeSignal. URL: <https://codesignal.com/learn/courses/technical-indicators-in-financial-analysis-with-pandas/lessons/calculating-exponential-moving-average-ema-for-tesla-stock-using-pandas> (дата звернення: 13.10.2025).

30. Standard Deviation. Cuemath. URL: <https://www.cuemath.com/data/standard-deviation/> (дата звернення: 18.10.2025).

31. Jain A. 10 Assumptions of Linear Regression. Medium. 11.11.2021. URL: <https://medium.com/@amitjain2110/10-assumptions-of-linear-regression-c58f8703d657> (дата звернення: 24.10.2025).

32. Mckee A. Least Squares Method: How to Find the Best Fit Line. Datacamp. 12.07.2025. URL: <https://www.datacamp.com/tutorial/least-squares-method> (дата звернення: 23.10.2025).

33. What Is the $Y = mx + b$ Formula?. Vedantu. URL: <https://www.vedantu.com/maths/y-mx-b> (дата звернення: 15.10.2025).

34. A Step-by-Step Guide to Calculating the Least Squares Regression Line by Hand and Practice. DHgate. 23.10.2025. URL: <https://smart.dhgate.com/a-step-by-step-guide-to-calculating-the-least-squares-regression-line-by-hand-and-practice/> (дата звернення: 28.11.2025).

35. Hayes A. Coefficient of Variation: Definition and How to Use It. Investopedia. 17.06.2025. URL: <https://www.investopedia.com/terms/c/coefficientofvariation.asp> (дата звернення: 05.11.2025).

36. Nevil S. Z-Score: Meaning and Formula. Investopedia. 23.07.2025. URL: <https://www.investopedia.com/terms/z/zscore.asp> (дата звернення: 10.10.2025).

37. Rolitto E. Clustering Coefficient of the Tensor Product of Graphs. Asian Research Journal of Mathematics. 2022. С. 36–42. ISSN 2456-477. URL: <https://doi.org/10.9734/ARJOM/2022/v18i630382>.

38. Interquartile Range. ScienceDirect. URL: <https://www.sciencedirect.com/topics/computer-science/interquartile-range>, Book (дата звернення: 17.10.2025).

39. Identifying Outliers: IQR Method. Online.stat.psu.edu. URL: <https://online.stat.psu.edu/stat200/book/export/html/63> (дата звернення: 12.11.2025).

40. Correlation Coefficient Formula. Cuemath. URL: <https://www.cuemath.com/correlation-coefficient-formula/> (дата звернення: 12.11.2025).

41. Stewart K. Pearson's correlation coefficient. Britannica. 07.11.2025. URL: <https://www.britannica.com/topic/Pearsons-correlation-coefficient> (дата звернення: 19.11.2025).