

**Київський національний університет імені Тараса Шевченка**  
Факультет інформаційних технологій  
Кафедра програмних систем і технологій

УДК 004.891.3

*На правах рукопису*

**ВИПУСКНА КВАЛІФІКАЦІЙНА БАКАЛАВРСЬКА РОБОТА**

Тема: «Розробка додатку для аналізу й прогнозування стану здоров'я за результатами лабораторних досліджень»

Спеціальність – 121 «Інженерія програмного забезпечення»

**ПОЯСНЮВАЛЬНА ЗАПИСКА**

БР.ІПЗ – 22.00.00.000

**Студент**

ІПЗ-42 Яна КУШАКЕВИЧ  
(шифр групи)(підпис)(дата)(розшифровка підпису)

**Науковий керівник**

к.т.н., ас. Денис БЕРЕСТОВ  
(посада)(підпис)(дата)(розшифровка підпису)

**Консультант з питань  
нормоконтролю**

фахівець Тамара ЧАПОВСЬКА  
(посада)(підпис)(дата)(розшифровка підпису)

**Допускається до захисту  
з питань нормоконтролю  
Завідувач кафедри**

д.т.н., доц. Олексій БИЧКОВ  
(посада)(підпис)(дата)(розшифровка підпису)

**Київський національний університет імені Тараса Шевченка**  
**Факультет інформаційних технологій**  
**Кафедра програмних систем і технологій**  
**Спеціальність 121 “Інженерія програмного забезпечення”**

ЗАТВЕРДЖУЮ:

Завідувач кафедри програмних систем і технологій

\_\_\_\_\_ (Олексій БИЧКОВ)

„\_\_\_\_\_” \_\_\_\_\_ 20\_\_ р.

**ЗАВДАННЯ**  
**НА ВИПУСКНУ КВАЛІФІКАЦІЙНУ БАКАЛАВРСЬКУ РОБОТУ СТУДЕНТУ**

\_\_\_\_\_ Кушакевич Яні Титівні

(прізвище, ім'я, по батькові)

- 1. Тема випускної кваліфікаційної бакалаврської роботи** «Розробка додатку для аналізу й прогнозування стану здоров'я за результатами лабораторних досліджень» керівник проекту (роботи) Берестов Денис Сергійович к.т.н., асистент затверджені наказом вищого навчального закладу від „11” листопада 2020 р. № 6
- 2. Строк здачі студентом закінченої роботи** 31 травня 2021 року
- 3. Вихідні дані до роботи** Систематизовані дані інтервалів допустимих показників лабораторних досліджень, на основі яких можна робити висновки щодо наявності чи відсутності дефіцитних станів
- 4. Зміст пояснювальної записки (перелік питань, що їх належить розробити)**
  1. Аналіз основних понять та завдань аналізу даних
  2. Аналіз загальних методологій досліджень, що застосовуються при аналізі даних
  3. Розгляд технологій та алгоритмів інтелектуального аналізу даних
  4. Розробка моделі аналізу даних лабораторних досліджень та розробка на її основі онлайн-додатку для аналізу стану здоров'я людини
- 5. Перелік графічного матеріалу (з точним забезпеченням обов'язкових креслень)**
  1. Схема алгоритму виявлення прихованих закономірностей (рис. 1.1, ст. 29)
  2. Загальна схема дерева рішень (рис. 1.2, ст. 39)
  3. Життєвий цикл SPA-додатку (рис. 1.3, ст. 41)
  4. Схема роботи онлайн-додатку (рис. 1.4, ст. 41)

## 6. Консультанти з роботи із зазначенням розділів роботи, що їх стосуються

Розділ	Консультант	Підпис, дата	
		Завдання видав	Завдання прийняв
4	Берестов Д. С.		

7. Дата видачі завдання 8 листопада 2020 р.

Керівник Денис БЕРЕСТОВ  
(підпис) (розшифровка підпису)

Завдання прийняв до виконання Яна КУШАКЕВИЧ  
(підпис) (розшифровка підпису)

## КАЛЕНДАРНИЙ ПЛАН

Номер і назва етапів бакалаврської роботи	Термін виконання етапів роботи	Примітка
1. Підбір та вивчення літератури	15.01.2021	виконано
2. Аналіз концепцій та методологій аналізу даних	18.01.2021	виконано
3. Вивчення технологій та алгоритмів аналізу даних	25.01.2021	виконано
4. Побудова та опис дерева рішень	09.02.2021	виконано
5. Визначення переліку та інтервальних норм лабораторних аналізів	14.02.2021	виконано
6. Програмна реалізація онлайн-додатку	25.04.2021	виконано
7. Затвердження роботи науковим керівником	01.06.2021	виконано
8. Затвердження роботи завідувачем кафедри	11.06.2021	виконано

Студент – бакалавр Яна КУШАКЕВИЧ  
(підпис) (розшифровка підпису)

Керівник роботи Денис БЕРЕСТОВ  
(підпис) (розшифровка підпису)

## АНОТАЦІЯ

**Випускна кваліфікаційна бакалаврська робота:** 56 с., 4 рис., 3 додат., 18 джерел.

**Тема:** Розробка додатку для аналізу й прогнозування стану здоров'я за результатами лабораторних досліджень.

**Об'єкт дослідження:** технології та алгоритми обробки даних для вирішення завдань онлайн-додатку.

**Мета роботи:** розробка методу аналізу стану здоров'я та його програмна реалізація.

**Предмет дослідження:** дані лабораторних аналізів, представлені у вигляді таблиці.

**Результати дослідження:** Досліджено можливі шляхи вирішення проблем інтелектуального аналізу даних, оглянуто алгоритми їх реалізації. Створено додаток для аналізу та прогнозування стану здоров'я за допомогою показників лабораторних аналізів.

**Висновок:** У результаті досліджень було отримано дерево рішень для аналізу даних лабораторних аналізів, на основі якого побудовано онлайн-додаток для графічного відображення результатів досліджень.

АНАЛІЗ ДАНИХ, ДЕРЕВО ПРИЙНЯТТЯ РІШЕНЬ, ЛАБОРАТОРНІ АНАЛІЗИ, ДЕФІЦИТНИЙ СТАН.

## АННОТАЦИЯ

**Выпускная квалификационная бакалаврская работа:** 56 с., 4 рис., 3 доп., 18 источников.

**Тема:** Разработка приложения для анализа и прогнозирования состояния здоровья по результатам лабораторных исследований.

**Объект исследования:** технологии и алгоритмы обработки данных для решения задач онлайн-приложения.

**Цель работы:** разработка метода анализа состояния здоровья и его программная реализация.

**Предмет исследования:** данные лабораторных анализов, представленные в виде таблицы.

**Результаты исследования:** Исследованы возможные пути решения проблем интеллектуального анализа данных, рассмотрены алгоритмы их реализации. Создано приложение для анализа и прогнозирования состояния здоровья с помощью показателей лабораторных анализов.

**Вывод:** В результате исследований было получено дерево решений для анализа данных лабораторных анализов, на основании которого построено онлайн-приложение для графического отображения результатов исследований.

АНАЛИЗ ДАННЫХ, ДЕРЕВО ПРИНЯТИЯ РЕШЕНИЙ, ЛАБОРАТОРНЫЕ АНАЛИЗЫ, ДЕФИЦИТНЫЕ СОСТОЯНИЯ.

## ANNOTATION

**Final qualifying bachelor's paper:** 56 p., 4 fig., 3 annex, 18 sources.

**Topic:** Development of application for analyzing and predicting health status based on the results of laboratory tests.

**Object of research:** technologies and algorithms of data processing for solving problems of online application.

**Objective:** to develop a method of health analysis and its software implementation.

**Subject of research:** laboratory analysis data presented in the form of a table.

**Results of the study:** possible ways to solve problems of data mining are investigated, algorithms for their implementation are considered. An application for analyzing and predicting health status using indicators of laboratory tests has been created.

**Conclusion:** as a result of the research, a decision tree was obtained for analyzing laboratory analysis data, on the basis of which an online application was built for graphical displaying research results.

DATA ANALYSIS, DECISION TREE, LABORATORY TESTS, DEFICIENT CONDITION.

## ЗМІСТ

ПЕРЕЛІК ОСНОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, СКОРОЧЕНЬ .....	8
ВСТУП.....	9
<b>РОЗДІЛ 1</b>	
<b>ОСНОВНІ ПОНЯТТЯ Й ЗАВДАННЯ АНАЛІЗУ ДАНИХ</b>	
1.1. Класифікація ознак за шкалами вимірювання .....	12
1.2. Описова статистика .....	13
1.3. Дисперсійний аналіз .....	19
1.4. Кореляційний аналіз .....	20
<b>РОЗДІЛ 2</b>	
<b>ЗАГАЛЬНА МЕТОДОЛОГІЯ ДОСЛІДЖЕННЯ</b>	
2.1. Проблеми аналізу даних.....	23
2.2. Методи та алгоритми інтелектуального аналізу даних .....	24
2.3. Сфера застосування ІАД .....	25
<b>РОЗДІЛ 3</b>	
<b>ВИБІР ТЕХНОЛОГІЙ ТА АЛГОРИТМІВ ВИРІШЕННЯ ПРОБЛЕМИ ДОДАТКУ</b>	
3.1. Statistica Data Miner.....	29
3.2. RapidMiner .....	30
3.3. KNIME.....	31
3.4. Дерево прийняття рішень.....	31
<b>РОЗДІЛ 4</b>	
<b>РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ</b>	
4.1. Вибір критерію для аналізу стану здоров'я .....	35
4.2. Побудова дерева прийняття рішень .....	36
4.3. Програмна реалізація додатку .....	41
ВИСНОВКИ .....	43
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ .....	44
ДОДАТКИ.....	46
Додаток А. Скріншоти онлайн-додатку.....	46
Додаток Б. Таблиці даних меж визначення дефіцитних станів.....	48
Додаток В. Software Architecture Document.....	52

**ПЕРЕЛІК ОСНОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, СКОРОЧЕНЬ**

ІАД – інтелектуальний аналіз даних

Data mining – обробка даних

DNS – Domain Name System

SAD – Software Architecture Document

## ВСТУП

**Актуальність обраної теми** обумовлена тим, що результатом розвитку інформаційних технологій є колосальний обсяг даних, накопичених в електронному вигляді, який зростає кожного дня. При цьому дані, як правило, відрізняються своєю структурою (текст, зображення, аудіо, відео, гіпертекстові документи, бази даних). Накопичені за тривалий термін дані можуть містити в собі закономірності, тенденції та взаємозв'язок, що є цінною інформацією при плануванні, прогнозуванні, прийнятті рішень, контролю за процесами. Однак людина фізично не здатна ефективно аналізувати такі обсяги неоднорідних даних.

Технології інтелектуального аналізу даних, технології і методи оцінки якості даних розглядаються як одні з перспективних та найважливіших тем для досліджень і застосування в галузі інформаційних технологій.

**Напрямок досліджень** – аналіз наявності дефіцитних станів у людини – є одним із ключових у сучасній медицині, оскільки раціон сучасної людини в середньому є далеким від корисного і більшість людей має ті чи інші дефіцити (в Україні поширеними дефіцитами є дефіцит йоду та заліза). Наразі лікарі розробляють методичні матеріали щодо аналізу зовнішності людини на предмет пошуку візуальних маркерів наявності дефіцитних станів чи їх зв'язку з поширеними хворобами, дана ж робота є спробою систематизувати пошук дефіцитних станів на основі показників лабораторних досліджень.

**Метою** роботи є систематизація знань з технологій і методів оцінки якості, очищення та попередньої обробки даних, та практичне їх застосування на прикладі аналізу стану здоров'я людини.

Для досягнення мети в роботі вирішувалися наступні основні **задачі**:

- систематизація сутності і основних технологій (методів) оцінки даних, що аналізуються;
- практичний приклад аналізу даних для аналізу стану здоров'я за допомогою показників лабораторних аналізів.

**Об'єкт дослідження:** технології та алгоритми обробки даних для вирішення завдань онлайн-додатку.

**Предмет дослідження:** дані лабораторних аналізів, представлені у вигляді таблиці.

**Методи дослідження:** узагальнення та аналіз інформації щодо понять, завдань і методологій аналізу даних; дерево рішень як метод аналізу показників лабораторних досліджень.

**Наукова новизна** роботи зводиться до того, що вперше запропоновано узагальнений алгоритм пошуку дефіцитних станів людини на основі даних результатів лабораторних досліджень і вперше ж реалізовано у формі програмного забезпечення.

**Практичне значення одержаних результатів** полягає у реалізації готового до використання програмного продукту, який може знайти застосування як для лікарів (у першу чергу, терапевтів), так і для пацієнтів.

**Апробація результатів випускної кваліфікаційної бакалаврської роботи:** Результати бакалаврської роботи були представлені та схвалені на 8-мій Східно-Європейській конференції Математичних та програмних технологій Internet of Everything, що відбувалася 14 травня 2021 р.

**Публікації:** за результатами бакалаврської роботи підготовлено тези для доповіді у 8-мій Східно-Європейській конференції Математичних та програмних технологій Internet of Everything, що прийняті до друку в збірнику конференції MSTIoE-8.

## РОЗДІЛ 1

### ОСНОВНІ ПОНЯТТЯ Й ЗАВДАННЯ АНАЛІЗУ ДАНИХ

Математичні методи аналізу даних широко використовують при дослідженні різного роду систем і процесів – природних, екологічних, соціальних, технічних, економічних тощо.

Вперше про застосування статистичних методів аналізу даних згадано у Книзі чисел. Основи сучасних методів аналізу даних були закладені:

- Андрієм Миколайовичем Колмогоровим – статистичні методи контролю за якістю;
- Адрієном Марі Лежандром – метод найменших квадратів;
- Вільфредо Парето – розподіл Парето, діаграма Парето;
- Даніїлом Бернуллі – перші таблиці нормального розподілу, застосування нормального розподілу в теорії похибок, поділ похибок спостережень на випадкові й систематичні тощо;
- Карлом Гаусом – метод найменших квадратів;
- Карлом Пірсоном – критерії згоди, теорія кореляції, метод головних компонент;
- Томасом Байєсом – байєсівський підхід, байєсівські оцінки;
- Френсисом Гальтоном – теорія кореляції;
- Чарльзом Спірменом – техніка факторного аналізу, рангова кореляція).

Сучасні методи аналізу даних знайшли розвиток у працях А.Г. Івахненка, А.М. Дуброва, А. Стьюарта, А. Хьютсона, Б.В. Гнеденко, Б.Ю. Лемешка, В.В. Налімова, Г. Куллдорфа, Г. Сміта, Г. Хоттелінга, Г.В. Мартинова, Дж. Кіфера, Дж. Тьюкі, Е. Пітмена, Е. Пятецького-Шапіро, І.А. Ібрагімова, І.М. Парасюка, К.Х. Крамера, Л.М. Большева, М. Кендалла, М.С. Нікуліна, Н. Дрейпера, О.І. Орлова, О.О. Чупрова, П. Хьюбера, С.А. Айвазяна, С.Р. Рао, Т. Андерсона, Ю.В. Лінника, Ю.В. Прохорова, Ю.П. Адлера та багатьох інших дослідників.

## 1.1. Класифікація ознак за шкалами вимірювання

Визначення типу даних є першим етапом аналізу будь-яких даних. Класифікація даних за шкалами їх вимірювання є основною, і згідно з нею розрізняють наступні типи ознак.

Номінальні ознаки (класифікаційні ознаки, ознаки з неупорядкованими станами) – це дані, що вимірюють в номінальній шкалі (класифікаційній, шкалі найменувань). Найменування класів можуть бути виражені за допомогою чисел, але ці числа використовуються лише для відповіді на питання: належать два об'єкти до одного класу чи ні. Прикладами номінальних ознак є кольори, назви біологічних видів, назви навчальних дисциплін тощо. Із точки зору застосування стандартних алгоритмів і автоматизації аналізу даних доцільно обирати такі позначення класів: 0, 1, 2, 3, ... Проте з цими числами неможливо виконувати будь-які дії, окрім перевірки їх на рівність чи нерівність.

Порядкові ознаки (ординальні ознаки, ознаки з упорядкованими станами) – це дані, що вимірюють в порядкових шкалах. Такі дані можна порівнювати між собою у певному відношенні: “легше – важче”, “менше – більше”, “лівіше – правіше” тощо. Прикладами порядкових ознак є військові звання, сила землетрусу, оцінки школярів тощо. Якщо значеннями порядкової ознаки є числа, вони можуть також використовуватися для порівняння ступеня вияву класифікаційної ознаки, проте при цьому не будуть визначені відстані між класами.

Кількісні (варіаційні, числові) ознаки – це ознаки, що вимірюють у кількісних (відносних, циклічних, інтервальних та абсолютних) шкалах вимірювань. Залежно від шкали вимірювань з числовими характеристиками даних можуть виконуватися наведені нижче дії.

Отримані у шкалах вищих рангів дані можна приводити до шкал нижчих рангів. До прикладу, виміряні у шкалі відношень дані можна привести до інтервальної шкали. Такого роду перетворення називають зниженням шкали. Зазвичай необхідність у них виникає при обробці даних, що були виміряні у шкалах

різного типу. Зворотна операція – перетворення даних, що виміряні у нижчих шкалах, до вищих – вважається помилковою, оскільки зниження шкали призводить до втрати частини наявної інформації про досліджувані ознаки.

Точкові та інтервальні оцінки можуть використовуватися як характеристики вибірки. Точкові оцінки параметрів вибірки – це оцінки, що визначаються одним числом. Такими оцінками є медіани вибірок і середні арифметичні. Точкові оцінки можуть суттєво відхилятися від істинних значень оцінюваних параметрів за умови малих обсягів вибірок та при їх істотному відхиленні від нормального закону розподілу. Тому поряд з ними використовують інтервальні оцінки параметрів, що визначаються двома числами – межами інтервалу, до якого із заданою ймовірністю потрапляє оцінюваний параметр. [1]

## 1.2. Описова статистика

Описова статистика – це набір основних статистичних показників емпіричної вибірки значень кількісної ознаки. Здебільшого стандартні методи їх розрахунку розроблені з припущенням, що розподіл є нормальним. Підставою для цього є наявність зручних математичних моделей обробки відповідних даних. Не меншу роль у занадто широкому застосуванні методів аналізу нормально розподілених даних відіграє безгрунтовне припущення, що чи не всі випадкові дані підпорядковуються нормальному закону розподілу, однак це припущення часто виявляється хибним. За умови значного відхилення розподілу даних від нормального потрібно використовувати інші методи та формули. У зв'язку з цим процедуру аналізу емпіричної вибірки завжди необхідно починати із перевірки закону розподілу на нормальність. Відповідні методи будуть розглянуті нижче.

На практиці найчастіше користуються вибірковими характеристиками, які розраховуються за обмеженою кількістю значень досліджуваного показника, що становлять певну вибірку з генеральної сукупності. Вони є оцінками відповідних генеральних статистичних характеристик (параметрів розподілу). Разом з тим,

вибіркові характеристики є випадковими величинами і змінюються від вибірки до вибірки.

Оцінку називають спроможною (конзистентною), якщо із збільшенням обсягу вибірки вона за ймовірністю наближається до оцінюваного параметра. Для перевірки спроможності використовується її достатня умова: оцінка є спроможною, якщо при прямуванні обсягу вибірки до нескінченності її математичне сподівання наближається до істинного значення досліджуваного параметра, а її дисперсія – до нуля.

Серед основних завдань описової статистики виділяють визначення центру, ширини, протяжності й симетрії розподілу. Середнє значення, математичне сподівання, мода та медіана характеризують центр статистичного розподілу. Іноді його визначають як середнє арифметичне мінімального й максимального значень вибірки, тобто центру розмаху, але зазвичай похибка такого методу є недопустимо високою.

Математичне сподівання неперервної кількісної ознаки визначають як центр тяжіння розподілу й визначають за формулою:

$$\bar{X} = \int_{-\infty}^{+\infty} x f(x) dx ; \quad (1.1)$$

де  $f(x)$  – щільність розподілу.

Для випадкових дискретних величин, виміряних у кількісних шкалах, замість математичного сподівання використовують середнє арифметичне значення, яке обчислюють за формулою:

$$\mu = \sum_{i=0}^{\infty} x_i p_i ; \quad (1.2)$$

де  $x_i$  – значення, яких випадкова величина, розподілена на інтервалі  $(-\infty; +\infty)$ , набуває зі ймовірностями  $p_i$ .

Вибіркове середнє арифметичне та його середнє квадратичне відхилення обраховують за формулами:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, s_x = \frac{\sigma}{\sqrt{n}} ; \quad (1.3)$$

де  $x_i$  ( $i = 1, 2, \dots, n$ ) – значення результатів спостережень,  $n$  – обсяг вибірки,  $\sigma$  – вибіркове середнє квадратичне відхилення.

Якщо вихідні дані подано як частоти розподілу випадкової величини за інтервалами, то вибіркове середнє обчислюють за формулою:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k b_i v_i ; \quad (1.4)$$

де  $b_i$  ( $i = 1, 2, \dots, k$ ) – середини інтервалів,  $v_i$  – емпіричні частоти,  $k$  – кількість інтервалів.

Вибіркове середнє є спроможною незміщеною оцінкою математичного сподівання у випадку, якщо останнє існує. Широке застосування вибіркового середнього як оцінки центру розподілу пояснюється тим, що воно є єдиною оцінкою, для якої існує аналітичний вираз, що може бути використаний в інших співвідношеннях і формулах.

Значний вплив екстремальних значень, що можуть виявитися хибними, на значення середнього арифметичного як оцінки центру розподілу є його значним недоліком. Також застосування середнього арифметичного може виявитися некомфортним, якщо різні емпіричні точки мають різну важливість, або якщо діапазон зміни значень ознаки є занадто широким.

Виходячи з вищенаведеного, окрім середнього арифметичного використовують інші види середніх величин:

– середнє гармонічне:

$$h = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}} ; \quad (1.5)$$

– середнє геометричне (середнє пропорційне):

$$g = \sqrt[n]{\prod_{i=1}^n x_i} = \prod_{i=1}^n \sqrt[n]{x_i} ; \quad (1.6)$$

– степеневе середнє:

$$\omega_\alpha = \sqrt[\alpha]{\frac{1}{n} \sum_{i=1}^n x_i^\alpha}, \alpha > 0 ; \quad (1.7)$$

– зважені степеневі середні:

$$\Omega_\alpha = \sqrt[\alpha]{\frac{\sum_{i=1}^n w_i x_i^\alpha}{\sum_{i=1}^n w_i}}, \alpha > 0, \quad (1.8)$$

де  $w_i$  ( $i = 1, 2, \dots, n$ ) – вагові коефіцієнти; степеневі середні є окремим випадком зважених степеневих середніх.

Для розглянутих величин виконується нерівність  $h \leq g \leq \bar{x} \leq \omega_\alpha$  (остання нерівність виконується, якщо  $\alpha > 1$ ).

Середнє геометричне часто використовують для визначення середнього темпу зростання досліджуваного показника, а також у випадках, коли досліджувані значення утворюють геометричну прогресію або змінюються за експоненціальним законом.

Степеневі середні використовують з метою збільшення або зменшення внеску певних значень у результат.

Середнє гармонічне часто використовують, коли шуканим показником є величина, обернена значенню усереднюваної ознаки.

Вибіркова медіана є числовою характеристикою неперервно розподіленої випадкової величини, визначеної з умовою, що із показником ймовірності 0,5 випадкова величина може набувати значення, більші та менші ніж медіана, тобто:

$$\int_{-\infty}^m f(x)dx = \int_m^{+\infty} f(x)dx = \frac{1}{2} \quad (1.9)$$

Медіана є найбільш загальною й фундаментальною характеристикою центру розподілу, оскільки її базою є центр симетрії. Для дискретно розподіленої випадкової величини медіаною вважають таке ціле число  $m$ , що:

$$\sum_{i=0}^{m-1} p_i \leq \frac{1}{2}; \sum_{i=0}^m p_i \geq \frac{1}{2} \quad (1.10)$$

Вона може бути визначена як розв'язок рівняння:

$$F_n(x) = \frac{1}{2}, \quad (1.11)$$

де  $F_n(x)$  – емпірична функція розподілу випадкової величини. Похибка медіани  $s_{me} = \sigma \sqrt{\frac{\pi}{2n}}$ .

Модою  $m_x$  називають точку максимуму емпіричної функції щільності розподілу. Моду як характеристику центру можна використовувати лише для розподілів із симетричною кривою щільності розподілу.

Співвідношення між медіаною, модою та середнім арифметичним розподілу залежить від знака коефіцієнта асиметрії розподілу. Якщо він додатній, то  $m_x <$

$me_x < \bar{x}$ , у протилежному випадку –  $\bar{x} < me_x < m_x$ . У випадку, коли коефіцієнт асиметрії дорівнює нулю, ці показники центру розподілу є рівними.

Дисперсія характеризує ступінь відхилення елементів сукупності від середнього в одиницях вимірювання відповідної ознаки. Для ознак, що визначаються у кількісних шкалах, дисперсія розраховується за формулами:

$$- \text{для неперервного випадку} - \sigma^2 = \int_{-\infty}^{+\infty} x^2 f(x) dx - \mu^2; \quad (1.12)$$

$$- \text{для дискретного випадку} - \sigma^2 = \sum_{i=0}^{\infty} x_i^2 p_i - \mu^2; \quad (1.13)$$

Якщо середнє значення сукупності відоме, то:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \quad (1.14)$$

Якщо середнє значення оцінюється за самою вибіркою, для розрахунку дисперсії використовують скореговану формулу:

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{\sum_{i=1}^n x_i^2 - n(\sum_{i=1}^n x_i)^2}{n(n-1)} \quad (1.15)$$

Похибку (стандартне відхилення) дисперсії визначають за формулою:

$$s_{\sigma^2} = \sigma^2 \sqrt{\frac{2}{n}} \quad (1.16)$$

Основною перевагою дисперсії як характеристики вибірки є те, що дисперсія суми статистично незалежних вибірок є сумою їх дисперсій незалежно від законів розподілу складових вибірок:

$$\sigma_{\Sigma}^2 = \sum_{i=1}^n \sigma_i^2 \quad (1.17)$$

У загальному випадку дисперсія суми двох вибірок дорівнює:

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2 \quad (1.18)$$

де  $\rho$  – коефіцієнт кореляції.

Деякі розподіли не мають скінченної дисперсії. Для її існування необхідно, щоб при  $x \rightarrow \infty$  крива щільності розподілу спадала швидше за  $1/x^3$ . В іншому випадку дисперсія рівна нескінченності. Якщо порівняти ці умови з умовою існування математичного сподівання, можна дійти до висновку, що для певних розподілів математичне сподівання може існувати, а скінченна дисперсія – ні. У цьому разі неможливо правильно визначити значення математичного сподівання за вибіркою скінченного обсягу. [2]

### 1.3. Дисперсійний аналіз

Дисперсійний аналіз – це статистичний метод вивчення взаємозв'язку, що застосовується для дослідження впливу однієї або декількох якісних змінних на одну залежну кількісну змінну. Основою дисперсійного аналізу є припущення, що деякі змінні можна розглядати як причини (незалежні змінні), а інші – як наслідки (залежні змінні чи відгуки). Незалежні ж змінні називають факторами, оскільки в ході експерименту можна змінювати їх значення і аналізувати отриманий результат залежною кількісною змінною.

Основною метою дисперсійного аналізу є дослідження значимості відмінності між середніми значеннями залежної кількісної змінної по групах фактору. Досягають цього за допомогою розкладання загальної дисперсії залежної змінної на складові: дисперсію за рахунок розбиття на групи (міжгрупова дисперсія) і дисперсію за рахунок інших факторів (внутрішньогрупова дисперсія). Аналізуючи ці компоненти дисперсії можна оцінити частку впливу кожного фактора на залежну змінну. Окремим завданням дисперсійного аналізу є виявлення, за рахунок яких конкретно груп йде відмінність середніх значень залежної змінної.

Існують різні моделі дисперсійного аналізу. В залежності від числа факторів існують однофакторні та багатфакторні дисперсійні моделі. Залежно від числа спостережень у досліджуваних групах розрізняють збалансовані і незбалансовані моделі дисперсійного аналізу. Одномірні та багатовимірні моделі розрізняють залежно від числа залежних змінних. За природою факторів розрізняють

детерміновані, випадкові і змішані моделі. Також виділяють дисперсійні моделі з незалежними вимірами та із повторними вимірами залежно від того, залежні чи незалежні вибірки утворені категоріями фактора.

Окрім того, виділяють параметричні методи дисперсійного аналізу, а також непараметричний дисперсійний аналіз: критерій Крускала-Уолліса і критерій Джонкхіера-Терпстра для незалежних вибірок, критерій Фрідмана і критерій Пейджа для повторних вимірів.

Області застосування дисперсійного аналізу досить велика. Дисперсійний аналіз застосовується в плануванні експерименту в управлінні виробництвом і в ряді областей економічних, соціологічних, медико-біологічних досліджень. [3]

#### **1.4. Кореляційний аналіз**

Кореляційний аналіз – розділ математичної статистики, що досліджує залежність між двома або більше випадковими величинами. Термін «Кореляція» означає взаємодію, взаємозв'язок.

Щоб визначити, наскільки сильний взаємозв'язок між змінними (ознаками) використовується кореляційний аналіз.

Дві випадкові величини  $X$  і  $Y$  можуть бути:

- пов'язані функціональною залежністю (жорстко, як залежність змінних в математичному аналізі);
- незалежними;
- пов'язані стохастичною (ймовірнісною залежністю) при якій зміна однієї величини тягне за собою зміну розподілу іншої.

У якості запобіжного зв'язку між випадковими величинами використовується коефіцієнт кореляції. Коефіцієнт кореляції для генеральної сукупності позначається  $\rho$ . Однак, як правило, він невідомий. Тому він оцінюється за експериментальними даними, які представляють вибірку обсягу  $n$ , отриману при спільному вимірі двох змінних (ознак)  $X$  і  $Y$ .

Коефіцієнт кореляції, який визначається за вибірковими даними, називається вибірковим коефіцієнтом кореляції (або просто коефіцієнтом кореляції). Його прийнято позначати символом  $r$ . Найбільш часто в якості оцінок генерального коефіцієнта кореляції використовується коефіцієнт кореляції Пірсона ( $r$ ) і коефіцієнт кореляції Спірмена ( $r_s$ ).

#### Коефіцієнт кореляції Пірсона ( $r$ )

Щоб правильно застосовувати кореляційний аналіз в наукових дослідженнях, потрібно враховувати умови застосування цього методу.

Умови, при яких можливий розрахунок коефіцієнта кореляції Пірсона:

- експериментальні дані повинні бути представлені тільки інтервальною шкалою або шкалою відносин;
- розподіл експериментальних даних підпорядковується нормальному закону;
- передбачається лінійна залежність між випадковими величинами  $X$  і  $Y$ .

#### Коефіцієнт кореляції Спірмена ( $r_s$ )

При розрахунку коефіцієнта кореляції Спірмена вимоги до вихідних даних менш суворі, а саме:

- дані можуть бути представлені у вигляді порядкової, інтервальної шкали, або шкали відносин;
- допускається будь-який закон розподілу випадкових величин  $X$  і  $Y$ ;
- між випадковими величинами  $X$  і  $Y$  повинна існувати монотонно-зростаюча або монотонно-спадна залежність.

#### Властивості оцінок коефіцієнтів кореляції

Розраховані коефіцієнти кореляції можуть набувати значень від  $-1$  до  $+1$ .

Якщо коефіцієнт кореляції дорівнює:  $r = +1$  і  $r = -1$ , це означає, що випадкові величини  $X$  і  $Y$  пов'язані жорсткою лінійною залежністю.

Якщо  $r \neq 0$ , то чим ближче  $|r|$  до одиниці, тим сильніша лінійна залежність випадкових величин  $X$  і  $Y$ .

Якщо коефіцієнт кореляції позитивний ( $r > 0$ ), то між випадковими величинами  $X$  і  $Y$  існує позитивна кореляція (або іншими словами позитивна кореляційна залежність). Прикладом позитивної кореляційної залежності є збільшення результату стрибка в довжину із збільшенням сили м'язів ніг.

Якщо коефіцієнт кореляції негативний ( $r < 0$ ), то між випадковими величинами  $X$  і  $Y$  існує негативна кореляція (або іншими словами негативна кореляційна залежність). Прикладом негативної кореляційної залежності є зменшення результату бігу 100 м зі збільшенням сили м'язів ніг.

Якщо коефіцієнт кореляції дорівнює нулю ( $r = 0$ ) – це означає, що кореляції немає; випадкові величини  $X$  і  $Y$  некорельовані. Іншими словами, між випадковими величинами  $X$  і  $Y$  немає взаємозв'язку.[4]

## РОЗДІЛ 2

### ЗАГАЛЬНА МЕТОДОЛОГІЯ ДОСЛІДЖЕННЯ

#### 2.1. Проблеми аналізу даних

У зв'язку зі швидким ростом технологій та способів зберігання даних людина отримала об'ємні потоки інформації в усіляких областях. Діяльність підприємств потребує реєстрації та запису всіх подробиць їх діяльності. Стає очевидним, що без обробки дані не мають цінності. Специфіка сучасних вимог до такої переробки є такою:

- необмежений обсяг даних;
- різноманітність даних (кількісні, якісні, текстові);
- коректність та зрозумілість результатів;
- прості у використанні інструменти для обробки сирих даних.

Термін «data mining» з'явився в 1990-х роках, але сама обробка даних виникла у 18 столітті, ґрунтуючись на теоремі Байєса, а трохи пізніше – на регресійному аналізі.

Із ростом кількості даних з'являлися нові технології в галузі інформатики (генетичні алгоритми, нейронні мережі, метод дерев рішень і т.д.), також з'явилася можливість за допомогою комп'ютерів зберігати велику кількість даних і швидко обробляти інформацію. Інтерес до data mining невпинно зростає і незабаром обробка даних почала вважатися окремою дисципліною. Зараз data mining включає в себе обробку не лише текстових даних (text data mining), але й графічних і мультимедійних (web mining).

Усталеного перекладу словосполучення «data mining» українською мовою немає, дослівно термін перекладається як «видобуток даних». Найчастіше ж використовується формулювання «інтелектуальна обробка даних».

Наразі data mining є частиною більшого поняття – Big data, що включає також збір і зберігання даних окрім їх обробки. [5]

## 2.2. Методи та алгоритми інтелектуального аналізу даних

Для того, щоб дані дійсно були цінними для організації, ви повинні мати можливість виявити закономірності та взаємозв'язки в межах цих даних. Цим займається пошук даних. Ці зв'язки та розуміння можуть забезпечити кращі бізнес-рішення. Обмін даними також може знизити ризик, допомагаючи виявити шахрайство, помилки та невідповідності, які можуть призвести до втрати прибутку та пошкодження репутації. У різних галузях промисловості використовується пошук даних у різних контекстах, але мета однакова: краще зрозуміти клієнтів та бізнес.

Фундаментально data mining ґрунтується на 3-х поняттях:

- математична статистика – є основою більшості технологій, що використовуються для data mining (регресійний аналіз, кластерний аналіз, дискримінуючий аналіз тощо);
- штучний інтелект – відтворення нейронної мережі мислення людини в цифровому вигляді;
- машинне навчання – сукупність статистики і штучного інтелекту, що сприяє розумінню комп'ютерами даних, які вони обробляють, для вибору найбільш відповідного методу або методів аналізу.

У data mining терміни вживаються в такому класі завдань:

- виявлення відхилень – виявлення даних, що відрізняються по будь-яким параметрам із загальної маси;
- навчання асоціаціям – пошук взаємозв'язків між подіями;
- кластеризація – групування наборів даних, без заздалегідь відомих шаблонів;
- класифікація – узагальнення відомого шаблону для застосування до нових даних;
- регресія – пошук функції, що відображає набір даних із найменшим відхиленням;
- підведення підсумків – відображення у стислому вигляді вихідної інформації, включно із наданням звітів і візуалізацією. [6]

## 2.3. Сфера застосування ІАД

Застосування інтелектуального аналізу даних наразі є всюди. Методи ІАД в основному цікавлять комерційні підприємства, що розгортають свої проекти на основі Data Warehousing (інформаційних сховищ даних). ІАД являють собою велику цінність для всіх, хто займається наукою та бізнесом.

### 2.3.1. Постачальники послуг

У першу чергу пошуком даних займаються постачальники послуг у галузі мобільних телефонів та комунальних послуг. Компанії, що займаються мобільними телефонами та комунальними службами, використовують майнінг даних, щоб передбачити «збій», умови, за яких клієнт покидає компанію, щоб отримати телекомунікаційні послуги / газ / широкосмуговий зв'язок від іншого провайдера. Вони порівнюють платіжну інформацію, взаємодію з обслуговування клієнтів, відвідування веб-сайтів та інші показники, щоб дати кожному клієнту оцінку ймовірності, а потім націлюють пропозиції та заохочення на клієнтів, яких, як вони вважають, піддають більшому ризику розбиття. [7]

### 2.3.2. Роздрібна торгівля

Інший приклад пошуку даних та бізнес-аналітики походить із сектору роздрібною торгівлі. Роздрібні торговці сегментують клієнтів у групи «Відсотність, частота, грошові кошти» (ВЧГ) і спрямовують маркетинг та акції на ці різні групи. Для клієнта, який витрачає мало коштів, але часто, і востаннє робив це недавно, маркетинг буде дещо інший, ніж для клієнта, який витратив великі гроші, але лише один раз, а також деякий час тому. Перші можуть отримати пропозиції щодо лояльності, продажу та перехресних продажів, тоді як останнім може бути запропоновано виграшна угода, наприклад.

### *2.3.3. Електронна комерція*

Можливо, деякі з найбільш відомих прикладів Data Mining та аналізу походять із сайтів електронної комерції. Багато компаній із електронної комерції використовують Data Mining, щоб пропонувати перехресні розпродажі та розпродажі через свої веб-сайти. Один із найвідоміших з них – це, звичайно, Amazon, який використовує складні методи майнінгу, щоб пропонувати: "Людам, які переглядали цей продукт, також сподобалися...".

### *2.3.4. Супермаркети*

Супермаркети – це ще один хороший приклад пошуку даних та бізнес-аналітики в дії. Відомо, що програми карт лояльності супермаркетів, як правило, керуються в основному, якщо не виключно, бажанням збирати вичерпні дані про клієнтів для використання в обробці даних. Один помітний останній приклад цього – це американський роздрібний торговий центр Target. У рамках програми Data Mining компанія розробила правила, щоб передбачити, чи буде ймовірно, що їх покупці будуть вагітні. Переглядаючи вміст кошиків для покупок своїх клієнтів, вони могли помітити клієнтів, які, на їхню думку, ймовірно очікують дитину, і почати орієнтуватися на рекламу підгузків (памперсів), вати тощо. Прогноз був настільки точним, що Target розсилав рекламні талони сім'ям, які ще не усвідомлювали (або ще не оголосили), що вони вагітні!

### *2.3.5. Правоохоронні органи*

Правоохоронні органи використовують пошук даних та бізнес-аналітику не лише для корпоративних програм. Поза ними, органи із запобігання злочинності використовують аналітику та обмін даними, щоб виявити тенденції в безлічі даних, що допомагають у всьому: де розмістити штат поліції (де скоріше за все злочин трапиться і коли?), кого шукати на прикордонній території (на основі про вік / тип транспортного засобу, кількість / вік пасажирів, історію перетину кордонів) і навіть

до яких даних розвідки слід серйозно поставитись для антитерористичної діяльності. [8]

### РОЗДІЛ 3

## ВИБІР ТЕХНОЛОГІЙ ТА АЛГОРИТМІВ ВИРІШЕННЯ ПРОБЛЕМИ ДОДАТКУ

Медицина являє собою одну з найбільш важливих галузей для розвитку із використанням ІАД. Відомо багато експертних систем для постановки медичних діагнозів. Вони побудовані переважно на основі патернів, які описують зрівняння симптомів із різними захворюваннями. Дані допомагають прогнозувати не тільки діагноз, але й можливі шляхи лікування. Правила допомагають обирати ліки, визначати їх вплив, створювати умови кращого лікування, передбачати результати призначеного курсу лікування тощо. Технології аналізу даних дозволяють виявляти ці шаблони, складові зазначених правил. [9]

Проблема, яку необхідно вирішити, є створення додатку для аналізу стану здоров'я за допомогою показників лабораторних аналізів. У наш час існує безліч рішень, які використовують принципи ІАД для вирішення медичних потреб. Є причини вважати, що цей напрям є найбільш перспективним та потребує розвитку.

Показовим прикладом потреби розвитку даної сфери є пандемія, яка наразі охопила майже весь світ. Китай та інші країни прогнозують швидкість поширення та масштаб розповсюдження вірусу.

За допомогою технології аналізу великих даних, система може швидко прогнозувати тенденцію розвитку епідемії в ключових районах відповідно до ситуації в реальному часі, і заздалегідь сповіщати людей про необхідність прийняття точкових протиепідемічних заходів.

Система дозволяє створювати різні експериментальні варіанти розвитку подій залежно від заходів щодо запобігання і контролю епідемії для проведення порівняльного аналізу.

### 3.1. Statistica Data Miner

Statistica Data Miner – досить популярний пакет Statistica, розробником якого є компанія StatSoft. Компанією StatSoft була розроблена ціла система Statistica Data Miner, яка реалізує технологію Data Mining . Дана система спроектована і реалізована як універсальний всебічний засіб аналізу даних (від взаємодії з різними базами даних до створення готових звітів), що реалізує так званий графічно-орієнтований підхід.

Для більш повного і зручного уявлення про Data Mining слід розглядати це поняття як процес аналізу даних, в якому на різних етапах застосовуються різні методи. Даний підхід полегшує початкове знайомство з можливостями і методами Data Mining. Data Mining є процесом аналізу даних, починаючи з вибору, зберігання даних, і закінчуючи аналізом результатів, отриманих в ході виконання даного процесу. Весь процес можна розбити на 5 основних частин – Вибірка, Розвідницький аналіз, Обробка і трансформація даних, Застосування моделі або виду аналізу, Аналіз результатів. Аналізуючи результати, отримані після кожної стадії процесу аналізу, ви можете переналаштувати модель або структуру наступної стадії. У разі отримання незадовільних результатів ви завжди можете повернутися назад і провести деякі стадії аналізу або весь аналіз заново.

Система Statistica Data Miner підходить для виявлення прихованих правил і закономірностей, проведення поглиблених досліджень, де «не працюють» класичні методи математичної статистики. Авторами розроблений алгоритм виявлення прихованих закономірностей, схема якого представлена на рис. 1.1 . Алгоритм був апробований при вирішенні подібних прикладних дослідницьких завдань.

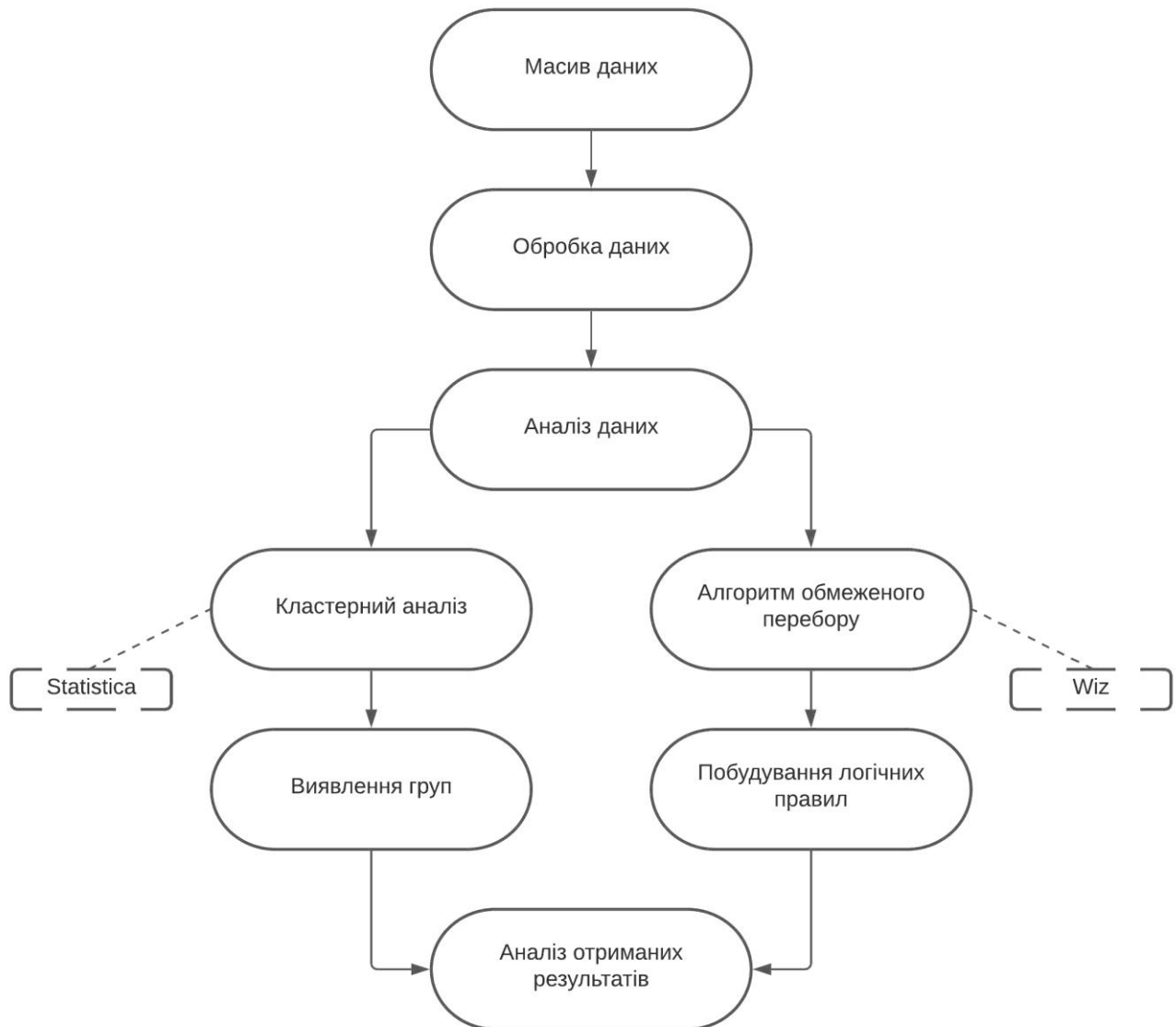


Рис. 1.1. Схема алгоритму виявлення прихованих закономірностей

### 3.2. RapidMiner

RapidMiner – це безкоштовне опенсорсне середовище для прогнозної аналітики.

Можливості RapidMiner можуть бути розширені за допомогою доповнень, окремі з яких також доступні безкоштовно. Система підтримує всі етапи глибокого аналізу даних, включаючи результуючу візуалізацію, перевірку та оптимізацію.

Величезна перевага RapidMiner полягає в тому, що для його використання не потрібно знати програмування. Тут реалізований принцип візуального програмування, тобто сам код писати не потрібно, як і не потрібно проводити складні математичні обчислення. Все відбувається наступним чином: користувач закидає дані на робоче поле, а потім просто перетягує оператори в GUI, формуючи процес обробки даних. Розбиратися у створеному кодї можна, але в більшості випадків не потрібно.

### 3.3. KNIME

KNIME – ще одна безкоштовна система для інтелектуального аналізу даних, яка навіть у базовій версії володіє якісним функціоналом. Подібно RapidMiner, KNIME пропонує інтуїтивно зрозуміле робоче середовище без необхідності програмувати. Тут також є ряд операторів, присутніх в RapidMiner (в KNIME вони називаються «вузлами»).

Якщо говорити про аналіз тексту, дана платформа дозволяє виконувати наступні завдання:

- перетин: згортання варіацій ключових термінів у вихідні форми;
- фільтрація стоп-слів: видалення незначних слів;
- розбиття на лексеми: розбиття рядків тексту на більш дрібні одиниці, наприклад, слова і фрази, через зазначені користувачем правила.

Крім цього, є deep-learning, веб-аналіз, обробка зображень, аналіз соціальних мереж та інше.

### 3.4. Дерево прийняття рішень

За загальним визначенням, дерево прийняття рішень – це засіб підтримки прийняття рішень при прогнозуванні, що широко застосовується в статистиці й аналізі даних.

Для вирішення завдань класифікації та регресії дерева прийняття рішень є одним із найбільш ефективних інструментів інтелектуального аналізу даних.

Вони являють собою ієрархічні деревоподібні структури, що складаються з рішучих правил виду «Якщо ..., то ...». У процесі навчання на навчальній множині правила генеруються автоматично і, оскільки вони формуються практично на природній мові, дерева рішень як аналітичні моделі у більшій мірі інтерпретованими, ніж, скажімо, нейронні мережі.

Оскільки правила в деревах рішень устанавлюються шляхом узагальнення безлічі навчальних прикладів, що описують предметну область, то за аналогією з відповідним методом логічного висновку їх називають індуктивними правилами, а сам процес навчання – індукцією дерев рішень.

У навчальній множині для прикладів обов'язково задається цільове значення, адже дерева рішень є моделями, що будуються на основі навчання (з учителем). Якщо цільова змінна є дискретною, модель називають деревом класифікації, а якщо безперервною – деревом регресії.

У 1950-х роках були закладені основні ідеї в області досліджень моделювання поведінки людини за допомогою комп'ютерних систем, що дали поштовх появі і розвитку дерев рішень. Серед них виділяють роботи К. Ховеленда «Комп'ютерне моделювання мислення» [10] і Е. Ханта «Експерименти по індукції» [11].

Подальший розвиток дерев рішень як самонавчальних моделей для аналізу даних пов'язано з іменами Джона Р. Квінлена [12], який розробив алгоритм ID3 і його вдосконалені модифікації C4.5 і C5.0, а також Лео Бреймана [13], який запропонував алгоритм CART і метод випадкового лісу.

### *3.4.1. Структура дерева рішень*

Дерево рішень – це спосіб подачі вирішальних правил в ієрархічній структурі, що складається з елементів двох типів – вузлів (node) і листя (leaf). У вузлах знаходяться вирішальні правила і проводиться перевірка відповідності прикладів цього правила з будь-якими атрибутами навчальної множини.

У найпростішому випадку, безліч заданих у вузлі прикладів в результаті перевірки розбиваються на дві підмножини, одна з яких містить приклади, що

задовільняють правило, а інша – що не задовільняють. Далі до кожної з підмножин знову застосовується зазначене правило і перевірка рекурсивно повторюється доти, доки алгоритм не зупиниться у зв'язку з досягненням умови зупинки. У результаті отримується вузол, не проводиться перевірка та розбиття, який оголошується листом. Для кожного прикладу, що потрапив у нього, лист визначає рішення. Для дерева регресії таким рішенням є відповідний листу модальний інтервал цільової змінної, а для дерева класифікації – клас, що асоціюється з вузлом.

Таким чином, лист містить не правило, а підмножину об'єктів, що відповідають усім правилам гілки, яка закінчується цим листом.

Підводячи підсумок, приклад повинен відповідати усім правилам, визначеним на шляху до листа, щоб потрапити у цей лист. Оскільки шлях до кожного листа у дереві єдиний, то в один лист може потрапити тільки один приклад, що забезпечує єдиний розв'язок. [14]

### 3.4.2. Задачі

Підтримка процесів прийняття управлінських рішень, яка використовується в статистиці, аналізі даних і машинному навчанні, є основною сферою застосування дерев рішень.

Даний апарат розв'язує наступні завдання:

- класифікація – розподілення кожного об'єкта до одного із зазначених класів (при цьому цільова змінна повинна мати дискретні значення);
- регресія – прогноз для кожного заданого вектора числового значення незалежної змінної;
- опис об'єктів – опис об'єктів і різноманітних складних структур через набір правил у дереві рішень.

Процесом побудови дерев рішень є послідовне, рекурсивне розбиття навчальної множини на підмножини із застосуванням вирішальних правил в вузлах. Розбиття проводиться доти, доки всі вузли на кінцях усіх гілок не будуть оголошені листям. Вузол може стати листом природним чином (коли буде містити лише один

об'єкт або ж об'єкти тільки одного класу), або після досягнення заданої умови зупинки.

Алгоритми побудови дерев рішень відносять до категорії так званих жадібних алгоритмів. Жадібними називаються алгоритми, які допускають, що локально-оптимальні рішення на кожному кроці (розбиття в вузлах), призводять до оптимального підсумкового рішення. У разі дерев рішень це означає, що якщо один раз був обраний атрибут, і по ньому було вироблено розбиття на підмножини, то алгоритм не може повернутися назад і вибрати інший атрибут, який дав би краще підсумкове розбиття. Тому на етапі побудови можна сказати, чи забезпечить обраний атрибут, в кінцевому підсумку, оптимальне розбиття. [15]

Наразі існує значна кількість алгоритмів навчання дерева рішень, наприклад: ID3, CART, C4.5, C5.0, NewId, ITrule, CHAID, CN2 і т.д. Найбільше поширення отримали наступні алгоритми:

- ID3 (Iterative Dichotomizer 3) – алгоритм побудови класифікуючих дерев рішень (дозволяє працювати тільки з дискретною цільовою змінною). Дерево може мати необмежене число нащадків у вузлах, але не може працювати з пропущеними даними;
- C4.5 – вдосконалена версія алгоритму ID3, що може працювати з пропущеними значеннями атрибутів (за версією видання Springer Science в 2008 році, алгоритм зайняв 1-е місце в топ-10 найбільш популярних алгоритмів Data Mining);
- CART (Classification and Regression Tree) – алгоритм навчання дерев рішень, що дозволяє вирішувати як завдання класифікації, так і регресії (наявна можливість використання як дискретної, так і безперервної цільової змінної). Дерева, побудовані за даним алгоритмом, у кожному вузлі можуть мати лише два нащадки.

## РОЗДІЛ 4

### РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

Проаналізувавши доступні та існуючі рішення поставленої задачі дипломного проекту – створення додатку для аналізу стану здоров'я за допомогою показників лабораторних аналізів – було прийнято рішення побудувати додаток із використанням алгоритму побудови дерева прийняття рішень.

#### 4.1. Вибір критерію для аналізу стану здоров'я

Лабораторні аналізи – це дослідження стану здоров'я організму пацієнта в лабораторних умовах. Лабораторні дослідження виконують для профілактики, за показанням лікаря чи за бажанням пацієнта.

За результатами лабораторних досліджень можна встановити або уточнити діагноз, визначити ступінь тяжкості хвороби, а також контролювати динаміку одужання.

Вітаміни та мінерали необхідні для підтримання нормальної роботи усіх систем та органів людського організму. Якщо людина не отримує їх разом із їжею у достатній кількості, або ж при наявності хвороб, що впливають на якість абсорбції вітамінів та мінералів із їжі, виникає дефіцит і людина починає хворіти. Саме із дефіцитом вітамінів та мінералів пов'язана велика частина як косметичних проблем людини (випадіння волосся, акне, ламкість нігтів, передчасна сивина тощо), так і проблем із функціонуванням організму (починаючи із постійної безпричинної втоми і закінчуючи депресією, судомою, хворобами нирок, галюцинаціями). Тому у даній роботі дослідження стану здоров'я і його прогнозування пов'язано із дослідженням наявності дефіцитів вітамінів та мінералів.

На можливий дефіцит вітамінів та мінералів можуть вказати не лише лабораторні дослідження щодо, власне, вітамінів та мінералів, а й загальний аналіз крові та біохімічний аналіз крові. Наприклад, низький рівень лейкоцитів може вказати на дефіцит вітамінів групи В, а низький трансферин – на дефіцит цинку. [16]

Лабораторії рідко надають у висновках щодо лабораторних досліджень межі норм того чи іншого аналізу, натомість завжди вказують референсні дані – середні показники лабораторних досліджень умовно здорових людей. При цьому встановлення норм є результатом рекомендацій медичних організацій, які досліджують, які показники тих чи інших аналізів необхідні для здорового функціонування людського організму.

Для визначення норм показників лабораторних аналізів у роботі використовуються дані збірника методичних рекомендацій «Анализы. Элементарно» Попової М.М.. Даний збірник засновується на низці джерел – починаючи з робіт окремих лікарів і закінчуючи даними клінічних рекомендацій всесвітніх медичних організацій чи організацій окремих країн світу.

Для прогнозування стану здоров'я, що для даної роботи відповідає визначенню розвитку дефіцитних станів, були використані референсні межі лабораторних аналізів збірника «Лабораторный справочник СИНЭВО», що випускається мережею медичних лабораторій Synevo.

Аналіз дефіцитних станів доступний для обмеженої кількості лабораторних аналізів, тож у перелік були включені лише ті аналізи, щодо яких є підтверджені дані залежності рівня їх результатів від рівня запасів тих чи інших вітамінів та мінералів (додаток Б).

## **4.2. Побудова дерева прийняття рішень**

У ході побудови дерева рішень вирішується кілька основних проблем, із кожною з яких пов'язаний відповідний крок процесу навчання:

- Вибір атрибута, за яким буде проводитися розбиття в даному вузлі;
- Вибір критерію зупинки навчання;
- Вибір методу відсікання гілок (спрощення);
- Оцінка точності побудованого дерева.

Розглянемо ці етапи нижче.

#### *4.2.1. Вибір атрибуту розбиття*

При формуванні правила розбиття у вузлі дерева обирається атрибут, за яким це буде зроблено. Загальне правило формулюється наступним чином: обраний атрибут розбиває множину даних у вузлі таким чином, що результуючі підмножини містять приклади з однаковими мітками класу або максимально наближені до цього, аби кількість об'єктів з інших класів в кожній підмножині була якомога меншою.

Вхідні дані для аналізу надаються у вигляді масивів об'єктів, де кожен масив – окремий вид аналізу. Суть розрахунків полягає у вирахуванні вірогідності наявності дефіцитів вітамінів та мінералів у людини.

Всього представлено чотири таблиці (масива) – «Клінічний аналіз крові», «Біохімічний аналіз крові» та «Вітаміни та мінерали» (додаток Б). Показник із таблиці «Вітаміни та мінерали» (а також «Феритин») має найвищий коефіцієнт для обрахування – в роботі було обрано цим показником число 1 – адже вони є прямими показниками, тобто показниками, які напряду вказують на запаси в організмі тих чи інших мікроелементів. Решті аналізів із таблиць «Клінічний аналіз крові» та «Біохімічний аналіз крові» був присвоєний коефіцієнт для обрахування 0.

Атрибутом для розбиття слугує коефіцієнт для обрахування.

#### *4.2.2. Вибір критерію зупинки навчання*

Гіпотетично, алгоритм навчання дерева рішень працюватиме доти, доки в результаті не будуть отримані «чисті» підмножини, в кожній із яких будуть приклади одного класу. Є ймовірність, що при цьому буде побудовано дерево, в якому кожному прикладу відповідатиме окремий лист. Таке дерево виявиться марним, оскільки воно буде перенавчаним – кожному прикладу відповідатиме один унікальний шлях в дереві, а отже набори правил будуть актуальними лише для заданого прикладу.

Наслідками перенавчання дерева рішень є точне розпізнавання заданих прикладів і при цьому повна неспроможність розпізнати нові дані. Окрім цього,

перенавчані дерева мають доволі складку структуру, через що їх складно інтерпретувати.

Очевидним рішенням проблеми є розробка примусової умови зупинки навчання. Існуючі підходи для примусової зупинки:

- Рання зупинка – алгоритм зупиняється, як тільки досягається задане значення певного критерію, наприклад, частки правильно розпізнаних прикладів. Перевагою підходу є зниження часу навчання. Основним недоліком є те, що рання зупинка шкодить точності дерева, тому рекомендується віддавати перевагу відсіканню гілок.
- Обмеження глибини дерева – задання максимальної допустимої кількості розбиття в гілках, після досягнення якої навчання дерева зупиняється. Цей метод також призводить до зниження точності дерева.
- Задання мінімально допустимої кількості прикладів у вузлі – заборона алгоритму створювати вузли з кількістю прикладів менше заданої (наприклад, 3). Це дозволяє запобігти тривіальному розбиттю і, відповідно, створенню малозначущих правил.

Всі перераховані підходи є евристичними, тобто не гарантують ідеального результату або ж працюють лише в окремих випадках, тому їх слід використовувати з обережністю. Обґрунтованих рекомендацій по тому, який метод працює найкраще, на даний час не існує, тому аналітики зазвичай використовують метод проб і помилок.

Як було зазначено вище, якщо «зростання» дерева обмежувати, результатом побудови стає складне дерево з великою кількістю вузлів і листя, яке важко інтерпретується. Разом із тим, вирішальні правила в таких деревах, вузли яких містять два-три приклади, виявляються малозначимими на практиці. Краще мати дерево, що має малу кількість вузлів, яким відповідає велика кількість прикладів із навчальної вибірки. Тому необхідно розробити таке дерево, яке при обмеженій

глибині забезпечує достатній рівень помилки розпізнавання, тобто знайти баланс між складністю і точністю дерева.

У роботі було використано змішані підходи для отримання кращого результату. Шляхом перебору було побудовано дерево, яке загальним чином описує процес прийняття рішення щодо наявності дефіциту певного елемента чи його розвитку.

#### *4.2.3. Опис утвореного дерева рішень*

Користувач вибирає перелік аналізів, які він хотів би перевірити. Потім в систему вводяться показники по кожному обраному критерію. На основі вхідних даних спочатку звіряються введені показники з нормами, зазначеними для них, із урахуванням статі та віку людини. Стать та вік обираються на першому етапі разом із переліком аналізів.

Далі формується масив із аналізами, що мають відхилення від норми, і передається для обрахування результатів. Дані сортуються по атрибуту для розбиття – в залежності від того, є прямі показники дефіцитів чи ні.

На рисунку 1.2 розглянуто прийняття рішень для поелементного розгляду наявності дефіцитних станів чи їх розвитку.

На початку аналізуються усі показники, що ввів користувач, для візуального відображення їх статусу в додатку – в нормі чи ні. Після цього для показників, що знаходяться поза нормою, формується масив елементів (вітамінів та мінералів) і показників, що до них відносяться (наприклад, згідно з додатком 2, для Фосфору такими можуть бути «Кальцій загальний», «Фосфор», «25-ОН вітамін D»).

Утворений масив аналізується на наявність прямих показників для того чи іншого елемента. Якщо такий показник є, аналіз на дефіцитний стан проводиться лише за ним для кожного окремого такого елемента, після чого результат аналізу виводиться на екран, а самі елементи прибираються з масиву для запобігання дублікації. Для елементів, які не мають у масиві прямих показників, визначається загальна кількість показників, по яких може бути проведений аналіз на наявність дефіцитних станів.

Якщо таких показників два чи більше, на екран виводиться відповідний результат; якщо ж один – така кількість не вважається достатньою для визначення дефіцитного стану, тож елемент ігнорується.

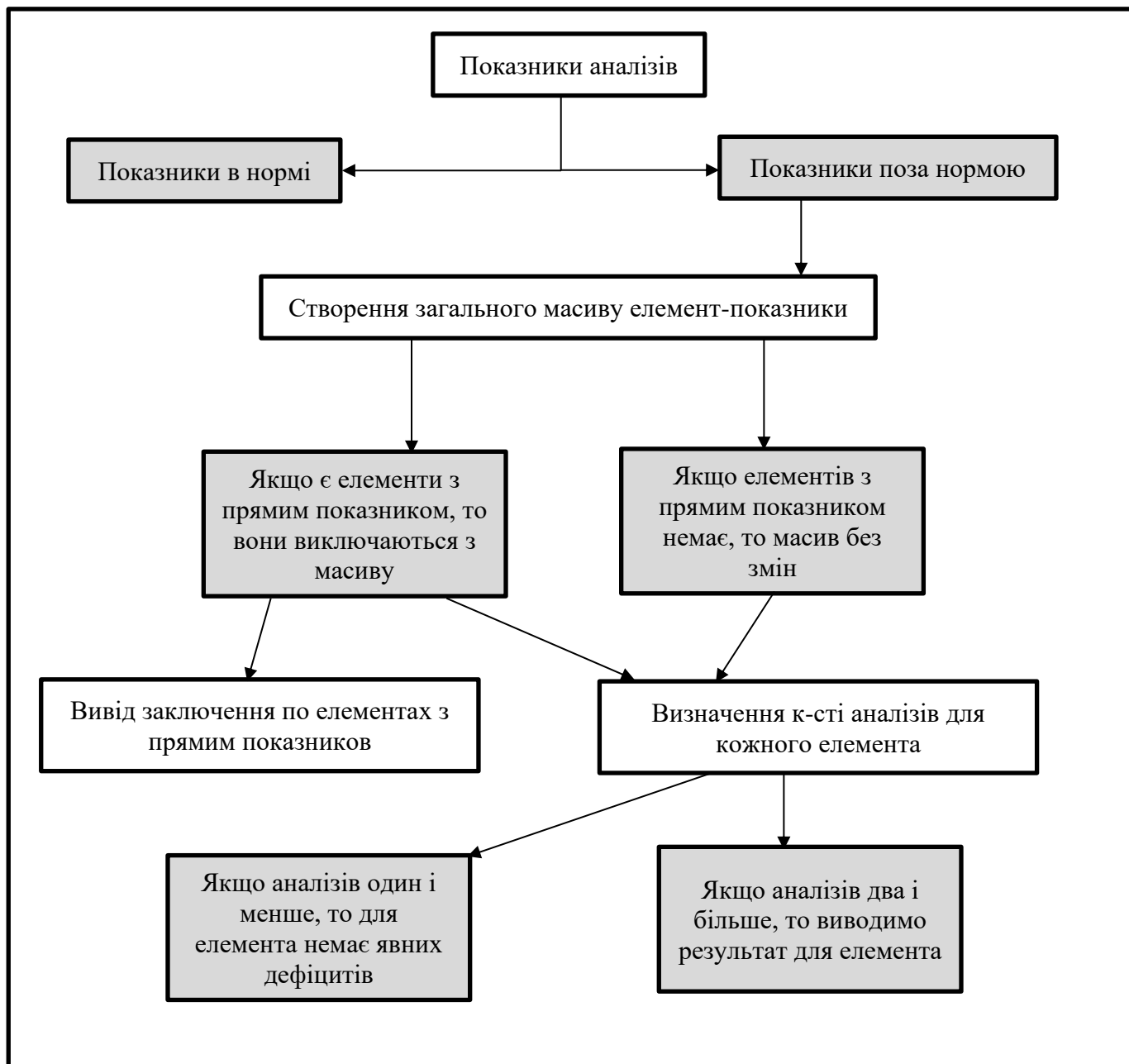


Рис. 1.2. Загальна схема дерева рішень

Для вилучення правил відстежено всі шляхи від кореневого вузла до листя дерева. Кожен шлях дав результуюче правило, яке складається з умов, що являють собою перевірку в кожному вузлі шляху.

Таким чином отримано проходження по всім можливим причинам відхилення показників лабораторних аналізів від норми та вибір результатів, які відповідають коефіцієнту обрахування чи мають достатню побічну базу для створення висновку.

### 4.3. Програмна реалізація додатку

Онлайн-додаток створений на мові програмування JavaScript, візуал виконаний за допомогою мов HTML і CSS3.

Додатково використовувалася JavaScript-бібліотека jQuery, а також для створення pdf-файлу підключена бібліотека pdfmake-0.1.js.

Дані представлені в форматі JSON.

Для створення онлайн-додатку було обрано життєвий цикл SPA (Single Page Application) (рис. 1.3) – він дозволяє динамічно оновлювати контент сторінки, при цьому роблячи лише один запит на сервер при вході на сайт, а решту часу всі дані знаходяться з боку користувача і скрипт працює лише з цими даними. Таким чином було розроблено так званий односторінковий сайт, що дозволяє зекономити час на завантаженні одних і тих же елементів. [18]

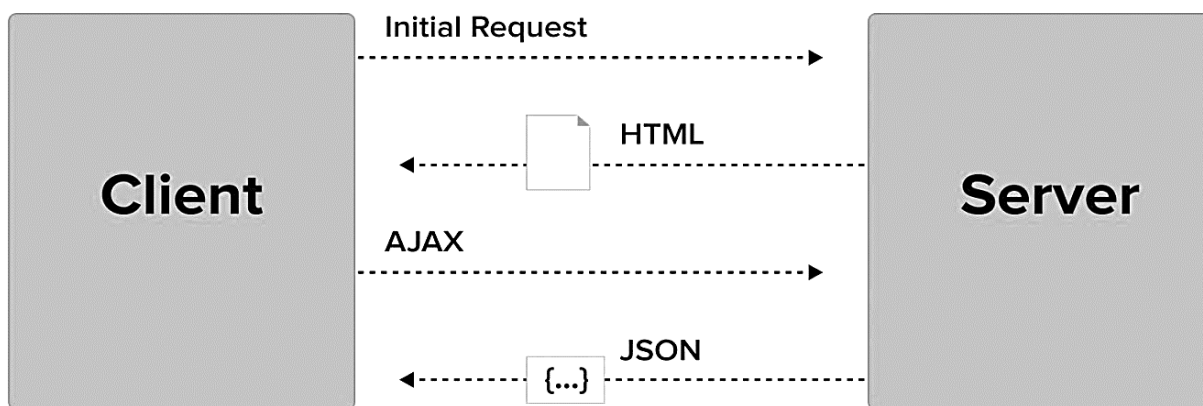


Рис. 1.3. Життєвий цикл SPA-додатку

Онлайн-додаток працює за схемою, наведеною на рисунку 1.4.

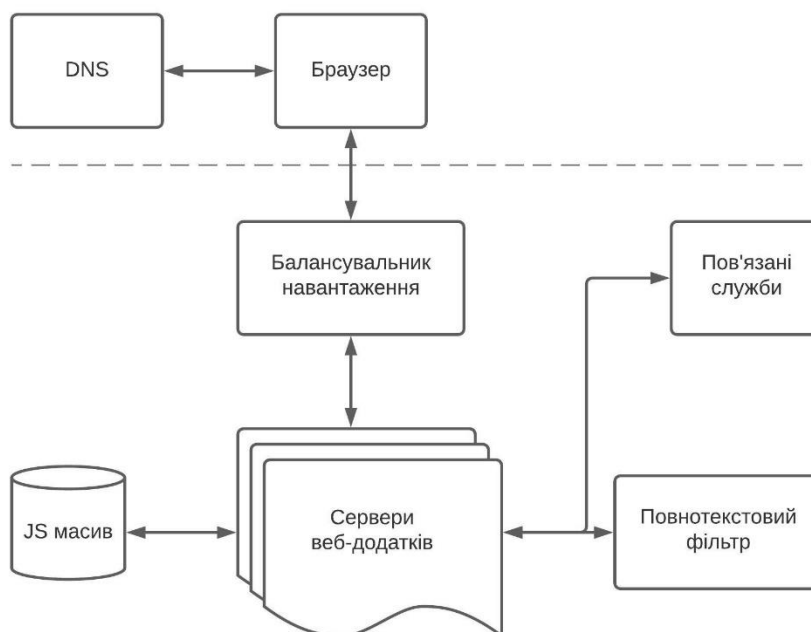


Рис. 1.4. Схема роботи онлайн-додатку

Онлайн-додаток реалізовано доволі простим, аби будь-який користувач без сторонньої допомоги міг ним користуватися. Сайт не вимагає авторизації – результати обробки аналізів можливо скачати у вигляді pdf-файлу. Сайт є односторінковим і має лише три варіанти відображення – «Про ресурс», «Ввід показників лабораторних аналізів», «Результат обробки» (додаток А). Виходячи з вищенаведеного, відповідно до стандартів Software Architecture Document, діаграми прецедентів (Use Case) та послідовності (Sequence) є доволі простими та не вимагають деталізації (додаток В).

Посилання для входу на сайт: <http://m85356.hostua5.fornex.host/>

## ВИСНОВКИ

Із розвитком глобалізації раціон людини, з одного боку, став набагато різноманітнішим із появою напівфабрикатів, широкого асортименту солодошів та їжі швидкого приготування. З іншого боку, така їжа є «бідною» на корисні вітаміни та мінерали, необхідні для здоров'я людини, і чим більша частка раціону віддана «смаколикам», тим більш розповсюдженими стають дефіцити специфічних вітамінів та мінералів. Сьогодні в Україні велика кількість людей мають дефіцит йоду та заліза, і все більше кількість людей, що мають інші дефіцити.

Лікарю дефіцитні стани легко діагностувати за аналізами, що прямо вказують на той чи інший вітамін/мінерал, і важче за аналізами інших категорій. Пацієнт же зовсім не має можливості в домашніх умовах розпізнати за наявними аналізами можливі проблеми зі здоров'ям, що може спричинити несвоєчасне звернення до лікаря.

Результатом роботи є доступний усім бажаючим онлайн-додаток, що за показниками лабораторних аналізів може підказати наявність у людини дефіцитних станів чи їх розвиток.

У роботі були досягнені наступні цілі:

- досліджені методології інтелектуального аналізу даних, оглянуто алгоритми їх реалізації;
- створено онлайн-додаток для аналізу стану здоров'я за показниками лабораторних досліджень.

Розроблений додаток може використовуватися лікарями та їх пацієнтами для швидкого аналізу стану здоров'я. Результати обробки даних є достовірними, адже засновуються на роботах окремих лікарів, офіційно виданих методичних рекомендаціях та клінічних рекомендаціях всесвітніх медичних організацій і організацій окремих країн світу.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Бахрушин В.Є. Методи аналізу даних : навчальний посібник для студентів / В.Є. Бахрушин. – Запоріжжя : КПУ, 2011. – 268 с.
2. Василенко О. А. Математично-статистичні методи аналізу у прикладних дослідженнях: навч. посіб. / О. А. Василенко, І. А. Сенча. – Одеса: ОНАЗ ім. О. С. Попова, 2011. – 166 с.
3. Дослідження операцій: Навч. посіб. / М. Г. Медведєв, О. В. Колодінська. – [2-ге вид., перер. і доп.]. – К.: Вид-во Європ. ун-ту, 2006. – 158 с.
4. Химмельбау Дж. Анализ процессов статистическими методами / Дж. Химмельблау. – М. : Мир, 1973. – 957 с.
5. Knowledge Discovery Through Data Mining: What Is Knowledge Discovery? – Tandem Computers Inc., 1996 – 253 p.
6. Кречетов Н. Продукты для интеллектуального анализа данных // Рынок программных средств, N14-15\_97. – 1997. – С. 32–39.
7. Data Mining and Image Processing Toolkits. – [Електронний ресурс]. – Режим доступу <http://datamining.itsc.uah.edu/adam/>.
8. Барсегян Ф. Методы и модели анализа данных OLAP и DataMining / Ф. Барсегян, М. Куприянов, В. Степанеенко, И. Холод. – СПб.: БХВ-Петербург, 2008. – 354 с.
9. Чубукова И.А. Data Mining: учебн. пособ. – М.: Интернет-университет информационных технологий БИНОМ: Лаборатория знаний, 2006. – 382 с.
10. Novland, C. I. (1960). Computer simulation of thinking. American Psychologist, 15 (11), 687-693.
11. Hunt, Earl B.; Janet Marin; Philip J. Stone (1966). Experiments in Induction. New York: Academic Press. ISBN 978-0-12-362350-8.
12. Quinlan, J. R. (1986). Induction of decision trees. Machine Learning, 1(1):81-106.

13. Quinlan, J. Ross. C4.5: Programs for Machine learning. Morgan Kaufmann Publishers 1993.
14. Шеннон, К. Работы по теории информации и кибернетике. М. Иностранная литература, 1963.
15. Айвазян С. А., Мхитарян В. С Прикладная статистика и основы эконометрики, М. Юнити, 1998.
16. Анализы.Элементарно – [Электронный ресурс]. – Режим доступа: <https://ae.healthnorms.com/>
17. Лабораторный справочник СИНЭВО – [Электронный ресурс]. – Режим доступа: <https://spravochnik.synevo.ua/>
18. ASP.NET - Single-Page Applications: Build Modern, Responsive Web Apps with ASP.NET – [Электронный ресурс]. – Режим доступа: <https://docs.microsoft.com/en-us/archive/msdn-magazine/2013/november/asp-net-single-page-applications-build-modern-responsive-web-apps-with-asp-net>

# ДОДАТКИ

## Додаток А

### Скріншоти онлайн-додатку

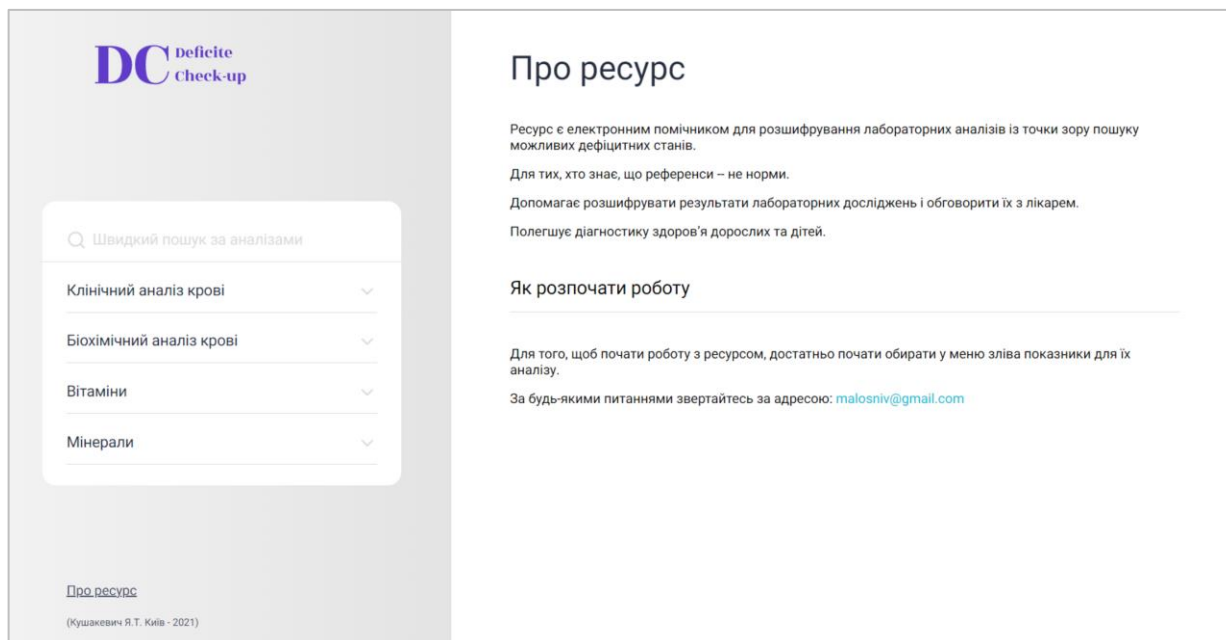


Рис. А.1 – Сторінка «Про ресурс»

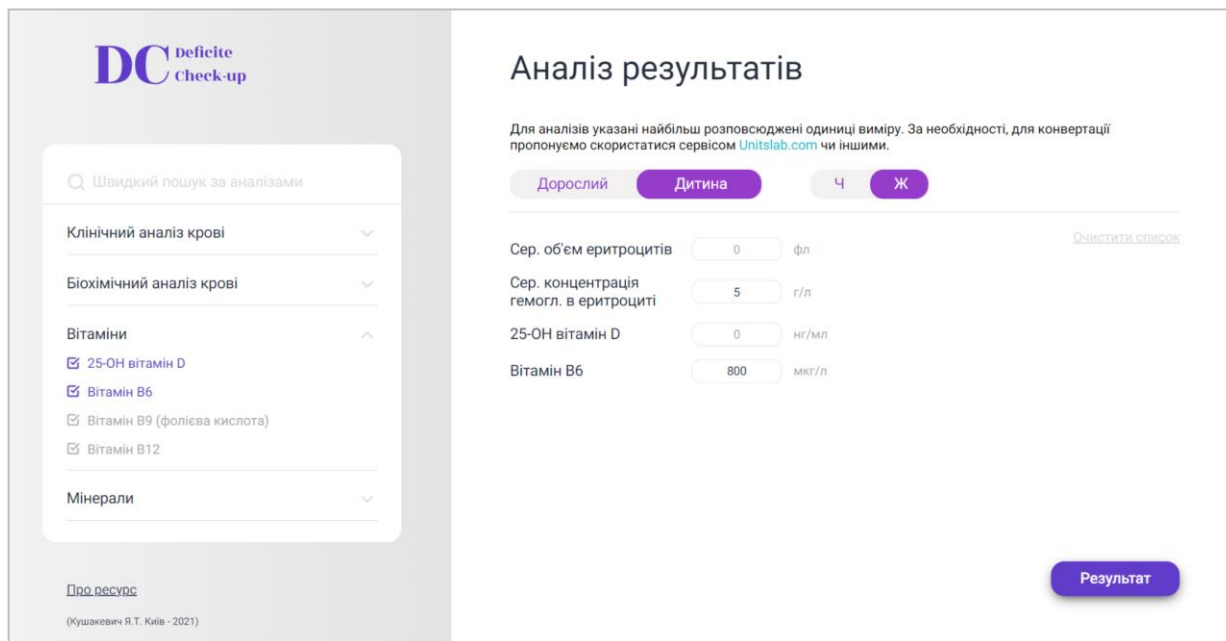


Рис. А.2 – Сторінка вибору аналізів та вводу даних

The screenshot displays the 'DC Deficite Check-up' interface. On the left, there is a navigation menu with categories: 'Клінічний аналіз крові', 'Біохімічний аналіз крові', 'Вітаміни', and 'Мінерали'. The main content area is titled 'Аналіз результатів' and includes a disclaimer: 'Дані засновуються виключно на довідкових посібниках та клінічних рекомендаціях щодо норм тих чи інших аналізів. Щодо розгляду можливих дефіцитів та лікування звертайтеся до лікаря!'. Below this is the 'Оцінка аналізів' section, which lists three parameters: 'Сер. об'єм еритроцитів' (12 фл) with a note 'Можливий дефіцит заліза.', 'Сер. концентрація гемогл. в еритроциті' (5 г/л) with notes 'Можливий дефіцит В6.', 'Можливий дефіцит заліза.', and 'Можливий дефіцит міді.', and 'Феритин' (8 мкг/л) with a note 'Можливий дефіцит заліза.'. The 'Результат обробки' section shows 'Дефіцит заліза.'. At the bottom right, there are two buttons: 'Повернутись до вибору' and 'Завантажити pdf файл'. The footer contains the text 'Про ресурс (Кушакевич Я.Т., Київ - 2021)'.

**DC Deficite Check-up**

Швидкий пошук за аналізами

Клінічний аналіз крові

Біохімічний аналіз крові

Вітаміни

Мінерали

Про ресурс  
(Кушакевич Я.Т., Київ - 2021)

## Аналіз результатів

Дані засновуються виключно на довідкових посібниках та клінічних рекомендаціях щодо норм тих чи інших аналізів.  
Щодо розгляду можливих дефіцитів та лікування звертайтеся до лікаря!

### Оцінка аналізів

Сер. об'єм еритроцитів 12 фл  
Можливий дефіцит заліза.

Сер. концентрація гемогл. в еритроциті 5 г/л  
Можливий дефіцит В6.  
Можливий дефіцит заліза.  
Можливий дефіцит міді.

Феритин 8 мкг/л  
Можливий дефіцит заліза.

### Результат обробки

Дефіцит заліза.

[Повернутись до вибору](#) [Завантажити pdf файл](#)

Рис. А.3 – Сторінка виводу результату обробки даних

## Додаток Б

## Таблиці даних меж визначення дефіцитних станів

Аналіз	одиниці	Норма	Розвиток (вище норми)	Коментар	Розвиток (нижче норми)	Коментар	Вище інтервалів	Нижче інтервалів
Клінічний аналіз крові								
Гемоглобін	г/л	135-170			130-135	Можливий розвиток чи наявність дефіциту заліза. Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.		Можливий дефіцит заліза. Можливий дефіцит В9. Можливий дефіцит В12.
Еритроцити	Т/л	4,5-5	5-5,5	Можливий розвиток чи наявність дефіциту заліза.	4-4,5	Можливий розвиток чи наявність дефіциту В6. Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.	Можливий дефіцит заліза.	Можливий дефіцит В6. Можливий дефіцит В9. Можливий дефіцит В12.
Сер. об'єм еритроцитів	фл	88-95	95-98	Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.	76-88	Можливий розвиток чи наявність дефіциту заліза.	Можливий дефіцит В9. Можливий дефіцит В12.	Можливий дефіцит заліза.
Сер. вміст гемогл. в еритроциті	пг	28-32	32-33	Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.	27-28	Можливий розвиток чи наявність дефіциту заліза. Можливий розвиток чи наявність дефіциту міді.	Можливий дефіцит В9. Можливий дефіцит В12.	Можливий дефіцит заліза. Можливий дефіцит міді.
Сер. концентрація гемогл. в еритроциті	г/л	320-360	360-370	Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.	310-320	Можливий розвиток чи наявність дефіциту заліза. Можливий розвиток чи наявність дефіциту міді. Можливий розвиток чи наявність дефіциту В6.	Можливий дефіцит В9. Можливий дефіцит В12.	Можливий дефіцит заліза. Можливий дефіцит міді. Можливий дефіцит В6.
Тромбоцити	*10 <sup>9</sup> /л	180-360	360-399	Можливий розвиток чи наявність дефіциту заліза.	170-180	Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.	Можливий дефіцит заліза.	Можливий дефіцит В9. Можливий дефіцит В12.
Сер. об'єм тромбоцитів	фл	7-10			6-7	Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.		Можливий дефіцит В9. Можливий дефіцит В12.
Лейкоцити	*10 <sup>9</sup> /л	4-9			3,5-4	Можливий розвиток чи наявність дефіциту В6. Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.		Можливий дефіцит В6. Можливий дефіцит В9. Можливий дефіцит В12.
Тромбоцит	%	0,12-0,4	0,4-0,45	Можливий розвиток чи наявність дефіциту заліза.	0,1-0,12	Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.	Можливий дефіцит заліза.	Можливий дефіцит В9. Можливий дефіцит В12.
Біохімічний аналіз крові								
Феритин	мкг/л	70-150			30-70	Можливий розвиток чи наявність дефіциту заліза.		Можливий дефіцит заліза.
Трансферин	г/л	2-3	3-3,6	Можливий розвиток чи наявність дефіциту заліза. Можливий розвиток чи наявність дефіциту В6. Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.	1,8-2	Можливий розвиток чи наявність дефіциту цинку.	Можливий дефіцит заліза.	Можливий дефіцит цинку.
Гомоцистеїн	мкмоль/л	5-7	7-15				Можливий дефіцит В6. Можливий дефіцит В9. Можливий дефіцит В12.	
Фібриноген	г/л	2-4			1,8-2	Можливий розвиток чи наявність дефіциту В12. Можливий розвиток чи наявність дефіциту цинку. Можливий розвиток чи наявність дефіциту магнію. Можливий розвиток чи наявність дефіциту В6. Можливий розвиток чи наявність дефіциту В12.		Можливий дефіцит В12. Можливий дефіцит цинку. Можливий дефіцит магнію. Можливий дефіцит В6. Можливий дефіцит В12.
Лужна фосфатаза	Од/л	40-130	130-140	Можливий розвиток чи наявність дефіциту кальцію. Можливий розвиток чи наявність дефіциту фосфору.	30-40		Можливий дефіцит кальцію. Можливий дефіцит фосфору.	Можливий дефіцит В6. Можливий дефіцит В12.
ТТГ	мМе/л	0,5-2	2-4,2	Можливий розвиток чи наявність дефіциту йоду. Можливий розвиток чи наявність дефіциту заліза.			Можливий дефіцит йоду. Можливий дефіцит заліза.	
T4 вільний	пмоль/л	12,5-17			9,3-12,5	Можливий розвиток чи наявність дефіциту йоду. Можливий розвиток чи наявність дефіциту цинку. Можливий розвиток чи наявність дефіциту заліза.		Можливий дефіцит йоду. Можливий дефіцит цинку. Можливий дефіцит заліза.
Вітаміни								
25-ОН вітамін D	нг/мл	30-100			10-30	Можливий розвиток чи наявність дефіциту вітаміну D.		Можливий дефіцит вітаміну D.
Вітамін В6	мкг/л	14,86-24,7			8,7-14,86	Можливий розвиток чи наявність дефіциту В6.		Можливий дефіцит В6.
Вітамін В9 (фолієва кислота)	нг/мл	6,6-20			4,6-6,6	Можливий розвиток чи наявність дефіциту В9.		Можливий дефіцит В9.
Вітамін В12	пг/мл	600-800			400-600	Можливий розвиток чи наявність дефіциту В12.		Можливий дефіцит В12.
Мінерали								
Міде сироваткова	мкмоль/л	11,01-22	22-23,5	Можливий розвиток чи наявність дефіциту В12. Можливий розвиток чи наявність дефіциту заліза.	10,9-11,01	Можливий розвиток чи наявність дефіциту міді.	Можливий дефіцит В12. Можливий дефіцит заліза.	Можливий дефіцит міді.
Цинк сироватковий	мкг/мл	0,99-1,5			0,7-0,99	Можливий розвиток чи наявність дефіциту цинку.		Можливий дефіцит цинку.
Магній	ммоль/л	0,9-1,1			0,66-0,9	Можливий розвиток чи наявність дефіциту магнію.		Можливий дефіцит магнію.
Селен сироватковий	мкг/л	101-135			74-101	Можливий розвиток чи наявність дефіциту селену.		Можливий дефіцит селену.
Йод у сечі	мкг/л	100-300			50-100	Можливий розвиток чи наявність дефіциту йоду.		Можливий дефіцит йоду.
Кальцій загальний	ммоль/л	2,2-2,6			2,15-2,2	Можливий розвиток чи наявність дефіциту кальцію. Можливий розвиток чи наявність дефіциту магнію. Можливий розвиток чи наявність дефіциту вітаміну D.		Можливий дефіцит кальцію. Можливий дефіцит магнію. Можливий дефіцит вітаміну D.
Фосфор	ммоль/л	0,8-1,45	1,45-1,5	Можливий розвиток чи наявність дефіциту кальцію.	0,7-0,8	Можливий розвиток чи наявність дефіциту фосфору. Можливий розвиток чи наявність дефіциту вітаміну D.	Можливий дефіцит кальцію.	Можливий дефіцит фосфору. Можливий дефіцит вітаміну D.

Рис. Б.1 – Таблиця даних для групи користувачів «Дорослі чоловічої статі»

Аналіз	єдиниці	Норма	Розвиток дефіцита (вище норми)	Коментарій	Розвиток дефіцита (ниже норми)	Коментарій	Вище інтервалов	Нижче інтервалов
<b>Клінічний аналіз крові</b>								
Гемоглобін	г/л	128-160			122-128	Можливий розвиток чи наявність дефіциту заліза. Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.		Можливий дефіцит заліза. Можливий дефіцит В9. Можливий дефіцит В12.
Еритроцити	Т/л	4,3-4,7	4,7-4,9	Можливий розвиток чи наявність дефіциту заліза.	3,7-4,3	Можливий розвиток чи наявність дефіциту В6. Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.	Можливий дефіцит заліза.	Можливий дефіцит В6. Можливий дефіцит В9. Можливий дефіцит В12.
Сер. об'єм еритроцитів	фл	88-93	93-96	Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.	76-88	Можливий розвиток чи наявність дефіциту заліза.	Можливий дефіцит В9. Можливий дефіцит В12.	Можливий дефіцит заліза.
Сер. вміст гемогл. в еритроциті	пг	28-32	32-33	Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.	27-28	Можливий розвиток чи наявність дефіциту заліза. Можливий розвиток чи наявність дефіциту міді.	Можливий дефіцит В9. Можливий дефіцит В12.	Можливий дефіцит заліза. Можливий дефіцит міді.
Сер. концентрація гемогл. в еритроциті	г/л	320-360	360-370	Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.	310-320	Можливий розвиток чи наявність дефіциту заліза. Можливий розвиток чи наявність дефіциту міді. Можливий розвиток чи наявність дефіциту В6.	Можливий дефіцит В9. Можливий дефіцит В12.	Можливий дефіцит заліза. Можливий дефіцит міді. Можливий дефіцит В6.
Тромбоцити	*10 <sup>9</sup> /л	180-360	360-399	Можливий розвиток чи наявність дефіциту заліза.	170-180	Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.	Можливий дефіцит заліза.	Можливий дефіцит В9. Можливий дефіцит В12.
Сер. об'єм тромбоцитів	фл	7-10			6-7	Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.		Можливий дефіцит В9. Можливий дефіцит В12.
Лейкоцити	*10 <sup>9</sup> /л	4-10			3,5-4	Можливий розвиток чи наявність дефіциту В6. Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.		Можливий дефіцит В6. Можливий дефіцит В9. Можливий дефіцит В12.
Тромбоцит	%	0,12-0,4	0,4-0,45	Можливий розвиток чи наявність дефіциту заліза.	0,1-0,12	Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.	Можливий дефіцит заліза.	Можливий дефіцит В9. Можливий дефіцит В12.
<b>Біохімічний аналіз крові</b>								
Феритин	мкг/л	50-100			23-50	Можливий розвиток чи наявність дефіциту заліза.		Можливий дефіцит заліза.
Трансферин	г/л	2-3	3-3,6	Можливий розвиток чи наявність дефіциту заліза.	1,8-2	Можливий розвиток чи наявність дефіциту заліза. Можливий розвиток чи наявність дефіциту цинку.	Можливий дефіцит заліза.	Можливий дефіцит цинку.
Гомоцистеїн	мкмоль/л	5-7	7-15	Можливий розвиток чи наявність дефіциту В6. Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.			Можливий дефіцит В6. Можливий дефіцит В9. Можливий дефіцит В12.	
Обриноген	г/л	2-4			1,8-2	Можливий розвиток чи наявність дефіциту В12. Можливий розвиток чи наявність дефіциту цинку. Можливий розвиток чи наявність дефіциту магнію. Можливий розвиток чи наявність дефіциту В6. Можливий розвиток чи наявність дефіциту В12.		Можливий дефіцит В12. Можливий дефіцит цинку. Можливий дефіцит магнію. Можливий дефіцит В6. Можливий дефіцит В12.
Лужна фосфатаза	Од/л	35-105	105-110	Можливий розвиток чи наявність дефіциту кальцію. Можливий розвиток чи наявність дефіциту фосфору.	30-35		Можливий дефіцит кальцію. Можливий дефіцит фосфору.	
ТТГ	мМе/л	0,5-2	2-4,2	Можливий розвиток чи наявність дефіциту йоду. Можливий розвиток чи наявність дефіциту заліза.			Можливий дефіцит йоду. Можливий дефіцит заліза.	
T4 вільний	пмоль/л	12,5-17			9,3-12,5	Можливий розвиток чи наявність дефіциту йоду. Можливий розвиток чи наявність дефіциту цинку. Можливий розвиток чи наявність дефіциту заліза.		Можливий дефіцит йоду. Можливий дефіцит цинку. Можливий дефіцит заліза.
<b>Вітаміни</b>								
25-ОН вітамін D	нг/мл	30-100			10-30	Можливий розвиток чи наявність дефіциту вітаміну D.		Можливий дефіцит вітаміну D.
Вітамін В6	мкг/л	14,86-24,7			8,7-14,86	Можливий розвиток чи наявність дефіциту В6.		Можливий дефіцит В6.
Вітамін В9 (фолієва кислота)	нг/мл	6,6-20			4,6-6,6	Можливий розвиток чи наявність дефіциту В9.		Можливий дефіцит В9.
Вітамін В12	пг/мл	600-800			400-600	Можливий розвиток чи наявність дефіциту В12.		Можливий дефіцит В12.
<b>Мінерали</b>								
Мідь сироваткова	мкмоль/л	12,6-24,4	24,4-31,5	Можливий розвиток чи наявність дефіциту В12. Можливий розвиток чи наявність дефіциту заліза.	10,7-12,6	Можливий розвиток чи наявність дефіциту міді.	Можливий дефіцит В12. Можливий дефіцит заліза.	Можливий дефіцит міді.
Цинк сироватковий	мкг/мл	0,99-1,5			0,7-0,99	Можливий розвиток чи наявність дефіциту цинку.		Можливий дефіцит цинку.
Магній	ммоль/л	0,9-1,1			0,66-0,9	Можливий розвиток чи наявність дефіциту магнію.		Можливий дефіцит магнію.
Селен сироватковий	мкг/л	101-135			74-101	Можливий розвиток чи наявність дефіциту селену.		Можливий дефіцит селену.
Йод у сечі	мкг/л	100-300			50-100	Можливий розвиток чи наявність дефіциту йоду.		Можливий дефіцит йоду.
Кальцій загальний	ммоль/л	2,2-2,6			2,15-2,2	Можливий розвиток чи наявність дефіциту кальцію. Можливий розвиток чи наявність дефіциту магнію. Можливий розвиток чи наявність дефіциту вітаміну D.		Можливий дефіцит кальцію. Можливий дефіцит магнію. Можливий дефіцит вітаміну D.
Фосфор	ммоль/л	0,9-1,5	1,5-1,55	Можливий розвиток чи наявність дефіциту кальцію.	0,81-0,9	Можливий розвиток чи наявність дефіциту фосфору. Можливий розвиток чи наявність дефіциту вітаміну D.	Можливий дефіцит кальцію.	Можливий дефіцит фосфору. Можливий дефіцит вітаміну D.

Рис. Б.2 – Таблиця даних для групи користувачів «Дорослі жіночої статі»

Аналіз	єдиниці	Норма	Розвиток дефіцита (вище норми)	Коментарій	Розвиток дефіцита (ниже норми)	Коментарій	Вище інтервалов	Нижче інтервалов
<b>Клінічний аналіз крові</b>								
Гемоглобін	г/л	130-170			125-130	Можливий розвиток чи наявність дефіциту заліза. Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.		Можливий дефіцит заліза. Можливий дефіцит В9. Можливий дефіцит В12.
Еритроцити	Т/л	4,5-4,7	4,7-5,1	Можливий розвиток чи наявність дефіциту заліза.	3,5-4,5	Можливий розвиток чи наявність дефіциту В6. Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.	Можливий дефіцит заліза.	Можливий дефіцит В6. Можливий дефіцит В9. Можливий дефіцит В12.
Сер. об'єм еритроцитів	фл	85-90	90-98	Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.	78-85	Можливий розвиток чи наявність дефіциту заліза.	Можливий дефіцит В9. Можливий дефіцит В12.	Можливий дефіцит заліза.
Сер. вміст гемогл. в еритроциті	пг	28-32	32-33	Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.	27-28	Можливий розвиток чи наявність дефіциту заліза. Можливий розвиток чи наявність дефіциту міді.	Можливий дефіцит В9. Можливий дефіцит В12.	Можливий дефіцит заліза. Можливий дефіцит міді.
Сер. концентрація гемогл. в еритроциті	г/л	320-360	360-370	Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.	310-320	Можливий розвиток чи наявність дефіциту заліза. Можливий розвиток чи наявність дефіциту В6.	Можливий дефіцит В9. Можливий дефіцит В12.	Можливий дефіцит заліза. Можливий дефіцит міді.
Тромбоцити	*10 <sup>9</sup> /л	180-360	360-390	Можливий розвиток чи наявність дефіциту заліза.	160-180	Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.	Можливий дефіцит заліза.	Можливий дефіцит В9. Можливий дефіцит В12.
Сер. об'єм тромбоцитів	фл	7-10			6-7	Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.		Можливий дефіцит В9. Можливий дефіцит В12.
Лейкоцити	*10 <sup>9</sup> /л	5-9			4,3-5	Можливий розвиток чи наявність дефіциту В6. Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.		Можливий дефіцит В6. Можливий дефіцит В9. Можливий дефіцит В12.
Тромбоцит	%	0,12-0,4	0,4-0,45	Можливий розвиток чи наявність дефіциту заліза.	0,1-0,12	Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.	Можливий дефіцит заліза.	Можливий дефіцит В9. Можливий дефіцит В12.
<b>Біохімічний аналіз крові</b>								
Феритин	мкг/л	50-100			30-50	Можливий розвиток чи наявність дефіциту заліза.		Можливий дефіцит заліза.
Трансферин	г/л	2-3	3-3,6	Можливий розвиток чи наявність дефіциту заліза.	1,8-2	Можливий розвиток чи наявність дефіциту заліза.	Можливий дефіцит заліза.	Можливий дефіцит заліза.
Гомоцистеїн	мкмоль/л	5-7	7-10	Можливий розвиток чи наявність дефіциту В6. Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.				Можливий дефіцит В6. Можливий дефіцит В9. Можливий дефіцит В12.
Фібриноген	г/л	2,4			1,8-2	Можливий розвиток чи наявність дефіциту В12.		Можливий дефіцит В12.
Лужна фосфатаза	Од/л	89-300	300-517	Можливий розвиток чи наявність дефіциту кальцію. Можливий розвиток чи наявність дефіциту фосфору.	40-89	Можливий розвиток чи наявність дефіциту цинку. Можливий розвиток чи наявність дефіциту магнію. Можливий розвиток чи наявність дефіциту В6. Можливий розвиток чи наявність дефіциту В12.	Можливий дефіцит кальцію. Можливий дефіцит фосфору.	Можливий дефіцит цинку. Можливий дефіцит магнію. Можливий дефіцит В6. Можливий дефіцит В12.
ТТГ	мМЕ/л	0,4-2,5	2,5-4,3	Можливий розвиток чи наявність дефіциту йоду. Можливий розвиток чи наявність дефіциту заліза.			Можливий дефіцит йоду. Можливий дефіцит заліза.	
Т4 вільний	пмоль/л	12,5-17			9-12,5	Можливий розвиток чи наявність дефіциту йоду. Можливий розвиток чи наявність дефіциту цинку. Можливий розвиток чи наявність дефіциту заліза.		Можливий дефіцит йоду. Можливий дефіцит цинку. Можливий дефіцит заліза.
<b>Вітаміни</b>								
25-OH вітамін D	нг/мл	30-100			10-30	Можливий розвиток чи наявність дефіциту вітаміну D.		Можливий дефіцит вітаміну D.
Вітамін В6	мкг/л	14,86-24,7			8,7-14,86	Можливий розвиток чи наявність дефіциту В6.		Можливий дефіцит В6.
Вітамін В9 (фолієва кислота)	нг/мл	6,6-20			4,6-6,6	Можливий розвиток чи наявність дефіциту В9.		Можливий дефіцит В9.
Вітамін В12	пг/мл	500-800			400-500	Можливий розвиток чи наявність дефіциту В12.		Можливий дефіцит В12.
<b>Мінерали</b>								
Мідь сироваткова	мкмоль/л	12,6-18,4	18,4-25,2	Можливий розвиток чи наявність дефіциту В12. Можливий розвиток чи наявність дефіциту заліза.	10,1-12,6	Можливий розвиток чи наявність дефіциту міді.	Можливий дефіцит В12. Можливий дефіцит заліза.	Можливий дефіцит міді.
Цинк сироватковий	мкг/мл	0,99-1,5			0,7-0,99	Можливий розвиток чи наявність дефіциту цинку.		Можливий дефіцит цинку.
Магній	ммоль/л	0,9-1,1			0,7-0,9	Можливий розвиток чи наявність дефіциту магнію.		Можливий дефіцит магнію.
Селен сироватковий	мкг/л	101-135			69-101	Можливий розвиток чи наявність дефіциту селену.		Можливий дефіцит селену.
Йод у сечі	мкг/л	100-300			50-100	Можливий розвиток чи наявність дефіциту йоду.		Можливий дефіцит йоду.
Кальцій загальний	ммоль/л	2,2-2,6			2,1-2,2	Можливий розвиток чи наявність дефіциту кальцію. Можливий розвиток чи наявність дефіциту магнію. Можливий розвиток чи наявність дефіциту вітаміну D.		Можливий дефіцит кальцію. Можливий дефіцит магнію. Можливий дефіцит вітаміну D.
Фосфор	ммоль/л	0,95-1,78	1,78-1,85	Можливий розвиток чи наявність дефіциту кальцію.	0,87-0,95	Можливий розвиток чи наявність дефіциту фосфору. Можливий розвиток чи наявність дефіциту вітаміну D.	Можливий дефіцит кальцію.	Можливий дефіцит фосфору. Можливий дефіцит вітаміну D.

Рис. Б.3 – Таблиця даних для групи користувачів «Діти чоловічої статі»

Аналіз	єдиниці	Норма	Розвиток дефіцита (вище норми)	Коментарій	Розвиток дефіцита (ниже норми)	Коментарій	Вище інтервалов	Нижче інтервалов
<b>Клінічний аналіз крові</b>								
Гемоглобін	г/л	125-160			121-125	Можливий розвиток чи наявність дефіциту заліза. Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.		Можливий дефіцит заліза. Можливий дефіцит В9. Можливий дефіцит В12.
Еритроцити	Т/л	4,5-4,7	4,7-5,1	Можливий розвиток чи наявність дефіциту заліза.	3,5-4,5	Можливий розвиток чи наявність дефіциту В6. Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.	Можливий дефіцит заліза.	Можливий дефіцит В6. Можливий дефіцит В9. Можливий дефіцит В12.
Сер. об'єм еритроцитів	фл	85-90	90-98	Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.	78-85	Можливий розвиток чи наявність дефіциту заліза.	Можливий дефіцит В9. Можливий дефіцит В12.	Можливий дефіцит заліза.
Сер. вміст гемогл. в еритроциті	пг	28-32	32-33	Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.	27-28	Можливий розвиток чи наявність дефіциту заліза. Можливий розвиток чи наявність дефіциту міді.	Можливий дефіцит В9. Можливий дефіцит В12.	Можливий дефіцит заліза. Можливий дефіцит міді.
Сер. концентрація гемогл. в еритроциті	г/л	320-360	360-370	Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.	310-320	Можливий розвиток чи наявність дефіциту заліза. Можливий розвиток чи наявність дефіциту міді. Можливий розвиток чи наявність дефіциту В6.	Можливий дефіцит В9. Можливий дефіцит В12.	Можливий дефіцит заліза. Можливий дефіцит міді. Можливий дефіцит В6.
Тромбоцити	*10 <sup>9</sup> /л	180-360	360-390	Можливий розвиток чи наявність дефіциту заліза.	160-180	Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.	Можливий дефіцит заліза.	Можливий дефіцит В9. Можливий дефіцит В12.
Сер. об'єм тромбоцитів	фл	7-10			6-7	Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.		Можливий дефіцит В9. Можливий дефіцит В12.
Лейкоцити	*10 <sup>9</sup> /л	5-9			4,3-5	Можливий розвиток чи наявність дефіциту В6. Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.		Можливий дефіцит В6. Можливий дефіцит В9. Можливий дефіцит В12.
Тромбоцит	%	0,12-0,4	0,4-0,45	Можливий розвиток чи наявність дефіциту заліза.	0,1-0,12	Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.	Можливий дефіцит заліза.	Можливий дефіцит В9. Можливий дефіцит В12.
<b>Біохімічний аналіз крові</b>								
Феритин	мкг/л	50-100			30-50	Можливий розвиток чи наявність дефіциту заліза.		Можливий дефіцит заліза.
Трансферин	г/л	2-3	3-3,6	Можливий розвиток чи наявність дефіциту заліза.	1,8-2	Можливий розвиток чи наявність дефіциту цинку.	Можливий дефіцит заліза.	Можливий дефіцит цинку.
Гомоцистеїн	мкмоль/л	5-7	7-10	Можливий розвиток чи наявність дефіциту В6. Можливий розвиток чи наявність дефіциту В9. Можливий розвиток чи наявність дефіциту В12.			Можливий дефіцит В6. Можливий дефіцит В9. Можливий дефіцит В12.	
Фібриноген	г/л	2-4			1,8-2	Можливий розвиток чи наявність дефіциту В12.		Можливий дефіцит В12.
Лужна фосфатаза	Од/л	54-300	300-460	Можливий розвиток чи наявність дефіциту кальцію. Можливий розвиток чи наявність дефіциту фосфору.	35-54	Можливий розвиток чи наявність дефіциту цинку. Можливий розвиток чи наявність дефіциту магнію. Можливий розвиток чи наявність дефіциту В6. Можливий розвиток чи наявність дефіциту В12.	Можливий дефіцит кальцію. Можливий дефіцит фосфору.	Можливий дефіцит цинку. Можливий дефіцит магнію. Можливий дефіцит В6. Можливий дефіцит В12.
ТТГ	мМЕ/л	0,4-2,5	2,5-4,3	Можливий розвиток чи наявність дефіциту йоду. Можливий розвиток чи наявність дефіциту заліза.			Можливий дефіцит йоду. Можливий дефіцит заліза.	
T4 вільний	пмоль/л	12,5-17			9-12,5	Можливий розвиток чи наявність дефіциту йоду. Можливий розвиток чи наявність дефіциту цинку. Можливий розвиток чи наявність дефіциту заліза.		Можливий дефіцит йоду. Можливий дефіцит цинку. Можливий дефіцит заліза.
<b>Вітаміни</b>								
25-ОН вітамін D	нг/мл	30-100			10-30	Можливий розвиток чи наявність дефіциту вітаміну D.		Можливий дефіцит вітаміну D.
Вітамін В6	мкг/л	14,86-24,7			8,7-14,86	Можливий розвиток чи наявність дефіциту В6.		Можливий дефіцит В6.
Вітамін В9 (фолієва кислота)	нг/мл	6,6-20			4,6-6,6	Можливий розвиток чи наявність дефіциту В9.		Можливий дефіцит В9.
Вітамін В12	пг/мл	500-800			400-500	Можливий розвиток чи наявність дефіциту В12.		Можливий дефіцит В12.
<b>Мінерали</b>								
Мідь сироваткова	мкмоль/л	12,6-18,4	18,4-25,2	Можливий розвиток чи наявність дефіциту В12. Можливий розвиток чи наявність дефіциту заліза.	10,1-12,6	Можливий розвиток чи наявність дефіциту міді.	Можливий дефіцит В12. Можливий дефіцит заліза.	Можливий дефіцит міді.
Цинк сироватковий	мкг/мл	0,99-1,5			0,7-0,99	Можливий розвиток чи наявність дефіциту цинку.		Можливий дефіцит цинку.
Магній	ммоль/л	0,9-1,1			0,7-0,9	Можливий розвиток чи наявність дефіциту магнію.		Можливий дефіцит магнію.
Селен сироватковий	мкг/л	101-135			69-101	Можливий розвиток чи наявність дефіциту селену.		Можливий дефіцит селену.
Йод у сечі	мкг/л	100-300			50-100	Можливий розвиток чи наявність дефіциту йоду.		Можливий дефіцит йоду.
Кальцій загальний	ммоль/л	2,2-2,6			2,1-2,2	Можливий розвиток чи наявність дефіциту кальцію. Можливий розвиток чи наявність дефіциту магнію. Можливий розвиток чи наявність дефіциту вітаміну D.		Можливий дефіцит кальцію. Можливий дефіцит магнію. Можливий дефіцит вітаміну D.
Фосфор	ммоль/л	1,05-1,7	1,7-1,8	Можливий розвиток чи наявність дефіциту кальцію.	0,87-1,05	Можливий розвиток чи наявність дефіциту фосфору. Можливий розвиток чи наявність дефіциту вітаміну D.	Можливий дефіцит кальцію.	Можливий дефіцит фосфору. Можливий дефіцит вітаміну D.

Рис. Б.4 – Таблиця даних для групи користувачів «Діти жіночої статі»

## Додаток В

Development of application for analyzing and predicting health status based on the results of laboratory tests

Software Architecture Document (SAD)

Content owner: Kushakevych Yana

30.05.2021

## **Software Architecture Document**

### **1. Introduction**

This document provides a general overview of the architecture of the Application for Analyzing and Predicting Health Status Based on the Results of Laboratory Tests system.

The document defines the architecture goals and the use cases supported by the system. The document provides a rationale for the architecture and design solutions, from conceptual idea to its implementation.

#### **1.1. Purpose**

The Software Architecture Document (SAD) provides a exhaustive architectural overview of the Application for Analyzing and Predicting Health Status Based on the Results of Laboratory Tests system. It presents an architectural view to depict the different aspects of the system.

#### **1.2. Scope**

The scope is to explain the architecture of the Application for Analyzing and Predicting Health Status Based on the Results of Laboratory Tests system.

This document depicts the variant aspects of the system design that are considered to be architecturally important. These elements and behaviors are fundamental for understanding this project as a whole.

#### **1.3. Definitions, Acronyms, and Abbreviations**

SAD - Software Architecture Document

RDBMS – Relational Database Management System

UML – Unified Modeling Language

User – This is any user who is registered on the DMM website

## 1.4. Overview

To fully document all aspects of architecture, this SAD contains the following sections:

Section 2: describes the use of each view

Section 3: describes the architectural goals and limitations of the system

Section 4: describes the general use-case realization

## 2. Architectural Representation

Use Case view is the view used to document the Application for Analyzing and Predicting Health Status Based on the Results of Laboratory Tests system.

**Audience:** all are interested in understanding the architecture.

**Area:** describes the set of use cases that represent central functionality of the system. This view describes the actors and use cases for the system, and also presents the needs of the user and is elaborated to describe discrete flows and constraints in more detail.

## 3. Architectural Goals and Constraints

Key requirements and system constraints that have a great bearing on the architecture are:

1. The system in the future, with the expansion of the amount of data, can be expanded to work with artificial intelligence. Therefore, among the tasks of architectural design, there is a need to show the application so that it can be extended if possible.
2. The system will be written using JavaScript and will use a JSON for data persistence. In the future, with more data, it is possible to use a RDBMS.

Architectural design should further consider these specific deployment requirements.

## **4. Use-Case View**

To give additional context to explain the usage of the system is the purpose of the use-case view. In section 4.1 the current actors are listed and a brief description of each in the general use context of the system is given. Section 4.2 outlines and illustrates the use-cases using UML use-case diagrams and sequence diagrams to clarify the interactions between components.

### **4.1. Actors**

#### **User**

The user drives all operation of the software, and there is no distinction in regards to type of user. The user interacts with all available interfaces to initiate all possible application operations.

#### **Web Portal**

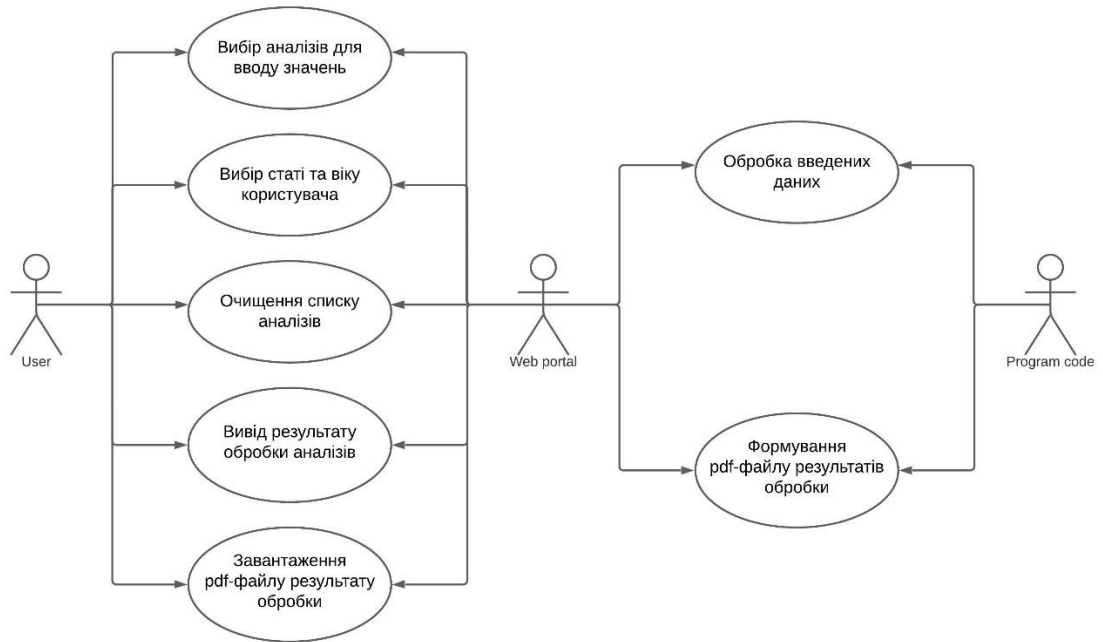
The web portal is the main user interface for the system.

#### **Program code**

Serves for data processing and communication with components.

## 4.2. Use-Case Realization

**Figure 1. General Use Case Diagram**



**Figure 2. General Sequence Diagram**

