

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО”

ФАКУЛЬТЕТ БІОМЕДИЧНОЇ ІНЖЕНЕРІЇ
Кафедра біомедичної кібернетики
другого (магістерського) рівня вищої освіти ступеня «магістр»

ЗВІТ
з переддипломної практики

спеціальності 122 Комп'ютерні науки
(шифр назва)

спеціалізації Інформаційні технології в біології та медицині

За темою стартап-проекту магістерської дисертації
Ідентифікація осіб із зниженими регуляторними резервами

(тема індивідуального завдання)

Студент 6-го курсу гр. БС-71мп **ВОЙНИК БОГДАН ОЛЕКСІЙОВИЧ**

(підпис)

Керівник практики ст.викл. каф БМК Аверьянова О.А. _____
(вчені ступінь та звання, прізвище, ініціали) (підпис)
ас. каф. БМК Корнієнко Г.А. _____
(вчені ступінь та звання, прізвище, ініціали) (підпис)

Науковий керівник МД зав каф БМК, д.б.н., к.т.н., с.н.с, Настенко Є.А.
(вчені ступінь та звання, прізвище, ініціали)

(підпис)

Київ – 2018 р.

ЗМІСТ

ВСТУП.....	3
СПИСОК СКОРОЧЕНЬ	
АНОТАЦІЯ	5
АННОТАЦІЯ.....	6
ANNOTATION.....	7
РОЗДІЛ 1. ОГЛЯД ЛІТЕРАТУРНИХ ДЖЕРЕЛ З ТЕМИ ІНДИВІДУАЛЬНОГО ЗАВДАННЯ	8
1.2. МЕТОД ДИСКРИМІНАНТНОГО АНАЛІЗУ	8
1.3. БІНАРНА ЛОГІСТИЧНА РЕГРЕСІЯ	9
РОЗДІЛ 2. ОСНОВНА ЧАСТИНА ПЕРЕДДИПЛОМНОЇ ПРАКТИКИ	10
2.1. ТЕОРЕТИЧНА ЧАСТИНА	10
2.2. АНАЛІТИЧНА ЧАСТИНА	11
2.3. ПРОГРАМНА ЧАСТИНА.....	17
2.4. АНАЛІТИЧНА ЧАСТИНА ЗА ВИСНОВКАМИ ПРОГРАМНОЇ ЧАСТИНИ	20
АНАЛІЗ СТАРТАП-ПРОЕКТУ ТА МОЖЛИВОСТІ ЙОГО РИНКОВОГО ВПРОВАДЖЕННЯ ЗА ТЕМОЮ МД.....	24
ВИСНОВОК.....	27
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	29
ПРЕЗЕНТАЦІЯ ДО ЗАХИСТУ ПЕРЕДДИПЛОМНОЇ ПРАКТИКИ	30

ВСТУП

Актуальність:

Фізичні навантаження є основою здорового способу життя, але завжди існує певна небезпека вибору неправильної фізичної активності. В цьому випадку людина не покращить фізичні параметри, він чи вона навіть може мати негативні наслідки для здоров'я. Моніторинг фізіологічного стану протягом усього періоду спортивних тренувань може допомогти уникнути небажаних ситуацій. Крім того, це дає можливість контролювати та змінювати спортивну програму, щоб поліпшити фізичні результати. Оскільки стан фізичного здоров'я може змінюватися незалежно від фізичних навантажень, то регулярне проведення проби Мартіне дасть змогу спостерігати динаміку змін в організмі за певний період [1,2]. Це потребує вдосконалення програмного продукту та розробки нових модулів, які б реєстрували дані та автоматично проводили б дослідження стану фізичного здоров'я.

Об'єкт дослідження: база даних студентів Національного технічного університету України “Київський політехнічний інститут ім. Ігоря Сікорського”

Предмет дослідження: алгоритм ідентифікація осіб із зниженими регуляторними резервами.

Мета роботи: вдосконалення системи реєстрації змін функціонального стану системи кровообігу шляхом порівняння параметрів тесту з заданими функціональними схемами кровообігу на основі критерію мінімальної евклідової відстані та формування основних та додаткових висновків про стан організму.

У відповідності з метою ставлять такі завдання:

Дослідження існуючої проблеми; дослідження і порівняння алгоритмів та методів розрахунку; проведення глобальної кластеризації; поєднання та автоматизація обраних методів; реалізація програмного додатку; зменшення кількості кластерів та побудова таблиць; удосконалення, тестування та корекція розробленого продукту.

СПИСОК СКОРОЧЕНЬ

МНК – метод найменших квадратів

АТС – артеріальний систолічний тиск

АТД – артеріальний діастолічний тиск

АТР – різниця артеріального тиску

ЧСС – частота серцевих скорочень

ЛР – логістична регресія

SPSS – Statistical Package for the Social Sciences

АНОТАЦІЯ

1. Практика реалізувалась **Войником Богданом Олексійовичем** студентом 6-го курсу, гр. БС-71мп кафедри Біомедичної кібернетики факультету Біомедичної інженерії НТУУ «КПІ ім. Ігоря Сікорського».

2. Тема практики: «Ідентифікація осіб із зниженими регуляторними резервами».

3. Ціль та задачі практики.

Ціль – вдосконалення системи реєстрації змін функціонального стану системи кровообігу

Задачі:

1. порівняння алгоритмів та методів розрахунку;
2. проведення глобальної кластеризації;
3. поєднання та автоматизація обраних методів;
4. реалізація програмного додатку;
5. зменшення кількості кластерів та побудова таблиць;
6. удосконалення, тестування та корекція розробленого продукту.

4. Отримано навички роботи в середовищі розробки Microsoft Visual Studio 2017 на мові C#;

5. Реалізовано практичне застосування набутих знань шляхом створення програмного модулю для побудови ліній регресій, пошуку студента в базі даних та підключення додатку до основного продукту, попередньо розширивши його функціонал.

6. Зміст звіту по практиці:

Титульний аркуш, зміст, вступ, анотація, огляд літературних джерел з теми індивідуального завдання, основна частина, висновки, план публікацій по темі дипломної роботи, список використаних джерел, презентація до захисту з практики.

7. За практикою надані документи контролю проходження практики:

- щоденник практики;
- індивідуальне завдання;
- звіт на 28 листах, додаток до звіту на _____ листах;
- презентація на 18 слайдах;
- відгук керівника ДР;
- план публікацій чи впровадження;

8. Сформована тема дипломної роботи до наказу: Комп'ютерна система для визначення функціонального стану кровообігу студентів .

9. Ключові слова: кластер, кластеризація, регресія, мінімальна відстань.

АННОТАЦИЯ

1. Практика реализовывалась **Войником Богданом Алексеевичем** студентом 6-го курса, гр. БС-71мп кафедры Биомедицинской кибернетики факультета Биомедицинской инженерии НТУУ «КПИ им. Игоря Сикорского».

2. Тема практики: «Идентификация персон с пониженными регуляторными резервами».

3. Цель и задачи практики.

Цель – совершенствование системы регистрации изменений функционального состояния системы кровообращения.

Задачи:

1. сравнения алгоритмов и методов расчета;
2. проведение глобальной кластеризации;
3. совмещения и автоматизация выбранных методов;
4. реализация программного приложения;
5. уменьшение количества кластеров и построение таблиц;
6. усовершенствования, тестирование и коррекция разработанного продукта.

4. Получены навыки работы в среде разработки Microsoft Visual Studio 2017 на языке C#;

5. Реализовано практическое применение приобретенных знаний путем создания программного модуля для построения линий регрессий, поиска студента в базе данных и подключения приложения к основному продукту, предварительно расширив его функционал.

6. Содержание отчета по практике:

Титульный лист, содержание, введение, аннотация, обзор литературных источников по теме индивидуального задания, основная часть, выводы, план публикаций по теме дипломной работы, список использованных источников, презентация к защите на практике.

7. По практике предоставлены документы контроля прохождения практики:

- дневник практики;
- индивидуальное задание
- отчет на 28 листах, приложение к отчету на листах.
- презентация на 17 слайдах;
- отзыв руководителя ДР;
- план публикаций или внедрения;
- гарантийное письмо.

8. Сформированная тема дипломной работы в приказ: Компьютерная система для определения функционального состояния кровообращения студентов.

9. Ключевые слова: кластер, кластеризация, регрессия, минимальное расстояние.

ANNOTATION

1. The practice was implemented by **Voynyk Bohdan Oleksiyovich** 6th course student, group BS-71mp, *Biomedical cybernetics* department of *Biomedical engineering faculty*, NTUU «Igor Sikorsky KPI».

2. Practice theme: «Identification of persons with reduced regulatory reserves».

3. Goal and tasks:

Goal – improvement of the registration system of changes in the functional state of the circulatory system.

Tasks:

1. comparison of algorithms and calculation methods;
2. global clustering;
3. combining and automating selected methods;
4. implementation of software applications;
5. reducing the number of clusters and the construction of tables;
6. improvements, testing and correction of the developed product.

4. Acquired skills in development environment Microsoft Visual Studio 2017 Express language C#;

5. Implemented the practical application of acquired knowledge by creating a software module for building regression lines, searching for a student in the database and connecting the application to the main product, having previously expanded its functionality.

6. Contents of the practice report:

The title page, contents, introduction, annotation, review of the literature on the subject of individual assignment, the main part, the conclusions, the plan of publications on the topic of the theses, a list of sources used, the presentations.

7. The control documents in practice:

- practice diary;
- individual task;
- report on _28_ pages, attachment to report on _____ pages;
- presentation on ___17___ slides;
- scientific director's recall;
- plan of publishing or implementation;
- letter of indemnity.

8. Formed theme of the graduate work in order: Computer system for determining the functional state of blood circulation.

9. Key words: cluster, clustering, regression, minimum distance.

РОЗДІЛ 1. ОГЛЯД ЛІТЕРАТУРНИХ ДЖЕРЕЛ З ТЕМИ ІНДИВІДУАЛЬНОГО ЗАВДАННЯ

1.2. МЕТОД ДИСКРИМІНАНТНОГО АНАЛІЗУ

Дискримінаційний чи дискримінантний аналіз використовується в тому випадку, якщо є дані, класифіковані на кілька груп, і необхідно знайти одну або більше функцій кількісних вимірів, які допоможуть віднести спостереження до однієї з цих груп. В дискримінантному аналізі розрізняють дві мети:

- інтерпретація;
- класифікація.

В дискримінаційному аналізі, на відміну від кластерного, є навчальна вибірка, в якій відомо до яких класів відносяться об'єкти. За навчальною вибіркою необхідно отримати правила, які в подальшому дозволять визначити, до якого класу відносяться нові об'єкти.

Найбільш загальне застосування дискримінантного аналізу є включення у дослідження багатьох змінних з метою визначення тих з них, які найкращим чином поділяють сукупності між собою. Наприклад, дослідник в галузі освіти, який цікавиться прогнозом вибору, який зроблять випускники середньої школи щодо своєї подальшої освіти, зробить з метою одержання найбільш точних прогнозів реєстрацію більшої кількості параметрів учнів.

У ролі дискримінантного аналізу найчастіше береться лінійна функція записана у вигляді формули (1.1):

$$Z = C_1X_1 + C_2X_2 + \dots + C_mX_m, \quad (1.1)$$

де X_1, X_2, \dots, X_m – значення ознак у даного об'єкта;

C_1, C_2, \dots, C_m – дискримінанті множники.

За допомогою дискримінантних множників виконуємо перехід від m -мірного простору первинних показників до одновимірного простору.

Лінійну функцію можна розглядати як проекцію даного об'єкта на деяку (одновимірну) дискримінантну вісь.

Дискримінантний аналіз потрібно проводити з використанням мінімальної кількості функцій. Їхня кількість залежить від конфігурації класів в багатовимірному просторі дискримінантних змінних.

1.3. БІНАРНА ЛОГІСТИЧНА РЕГРЕСІЯ

Регресійний аналіз один із розділів математичної статистики та являє собою метод моделювання вимірюваних даних і дослідження їх властивостей. Дані складаються з пар значень залежної змінної (змінної відгуку) і незалежної змінної (пояснюватиме змінної).

За допомогою методу бінарної логістичної регресії можна дослідити залежність дихотомічних змінних від незалежних змінних, які мають будь-який вид шкали.

Як правило, у випадку з дихотомічними змінними йдеться про деяку подію, яка може відбутися або не відбутися; бінарна логістична регресія в такому випадку розраховує ймовірність настання події в залежності від значень незалежних змінних.

Ймовірність настання події для деякого випадку розраховується за формулою (1.2) та формулою (1.3):

$$p = \frac{1}{1 + e^{-z}}, \quad (1.2)$$

де

$$z = b_1 * X_1 + b_2 * X_2 + \dots + b_n * X_n + a, \quad (1.3)$$

де X_i - значення незалежних змінних;

b_i - коефіцієнти, розрахунок яких є завданням бінарної логістичної регресії;

a - деяка константа.

Якщо для p вийде значення менше 0,5, то можна припустити, що подія не настане; в іншому випадку передбачається настання події. Розрахована ймовірність p завжди вказує на виконання пророкування, яке відповідає більшій з двох кодувань залежних змінних.

РОЗДІЛ 2. ОСНОВНА ЧАСТИНА ПЕРЕДДИПЛОМНОЇ ПРАКТИКИ

2.1. ТЕОРЕТИЧНА ЧАСТИНА

Метод найменших квадратів

Задача МНК розв'язується шляхом параметричної оцінки функції регресії, що описує залежність однієї величини Y , значення якої (y_i) спостерігають з випадковими похибками (θ_i), від групи не випадкових величин X_1, X_2, \dots, X_k . Нехай відомо, що вихідний параметр процесу, який вивчається, позначимо його y , лінійно залежить від вхідного параметра x (суцільна пряма лінія на рисунку 2.1).

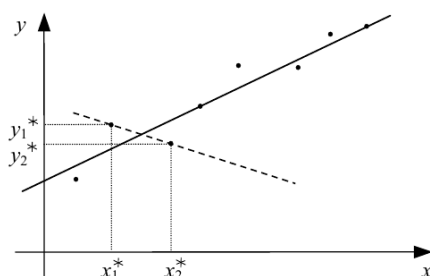


Рисунок 2.1 — Графічна інтерпретація причин, які обумовлюють необхідність використання МНК

Тобто припустимо, що статична характеристика цього процесу може бути подана у вигляді

$$y = ax + b \quad 2.1)$$

де a і b — коефіцієнти, для визначення числових значень яких необхідно, як мінімум, задати два значення x_1, x_2 вхідній величині x і заміряти відповідні їм значення y_1, y_2 вихідної величини y , оскільки лише під час виконання цих умов для моделі (2.1) можна скласти систему двох алгебраїчних рівнянь із двома невідомими a і b

$$\begin{cases} y_1 = ax_1 + b, \\ y_2 = ax_2 + b. \end{cases} \quad 2.2)$$

У свій час Гаусс запропонував інший спосіб визначення коефіцієнтів a, b моделі (2.1). Він запропонував сформувати суму квадратів різниць \sum^N між теоретично заданими за допомогою рівняння (2.1) значеннями вихідної координати y при значеннях аргументу $x_i, i = \overline{1, N}$ та її експериментальними значеннями y_i :

$$\sum^N = \sum_{i=1}^N (y(x_i) - (y_i))^2 \quad 2.3)$$

а потім знайти такі значення коефіцієнтів a, b рівняння (2.1), котрі мінімізують вираз (2.3). Від цієї процедури і назва методу — метод найменших квадратів.

2.2. АНАЛІТИЧНА ЧАСТИНА

Дослідження передбачають проведення глобальної кластеризації на базі даних студентів та викладачів, що пройшли пробу Мартіне декілька разів. Для кластеризації студентів і визначення їх субгрупи ризику ми використали програмний продукт «ClusterBox», попередньо модифікувавши його, додавши функцію глобального режиму, здатну розставити кластери та мітки (субгрупи ризику) не для одного студента, а для всіх одразу, що присутні в базі даних. На даному етапі дослідження дані зберігаються до текстових файлів.

Наступним кроком у дослідженні є аналіз графіків кластерів, побудованих на базі результуючих таблиць. Слід зазначити, що перед дослідженнями база даних була розщеплена по статі, оскільки попередні дослідження показують відмінність чоловічої групи від жіночої. Всі дослідження показані на прикладі чоловічої групи

Результуючі таблиці були побудовані на базі студентів НТУУ «КПІ ім. І. Сікорського» 1-2 курсу, за допомогою проведення дисперсійного аналізу. Приклад результуючої таблиці показано на рисунку 2.2.

Також дослідження показали, що кластеризація початкової бази даних по чоловікам збігається з методом кластеризації k-середніх на 79,49%. Отже алгоритм показує хороші результати.

Пок-ник	CL1		CL2		CL3		CL4		CL5		CL6		CL7		Ст. св.	Знач.	Сума квадратів
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD			
АТС0	135,84	9,647	119,08	9,518	124,02	9,499	131,24	9,641	137,68	12,649	142,39	13,807	117,49	9,740	793,00	0,000	134700,49
АТС1	154,70	18,842	125,13	15,299	131,29	14,327	144,25	14,831	153,73	22,810	156,46	21,457	123,42	13,779	793,00	0,000	338079,02
АТС2	148,49	13,435	125,05	11,198	131,41	11,468	139,28	9,200	145,89	15,579	151,91	16,134	124,54	9,263	793,00	0,000	180130,52
АТС3	140,12	11,333	120,78	9,397	127,61	9,891	132,22	8,740	138,95	11,068	144,48	14,077	120,43	8,749	793,00	0,000	131790,79
АТС4	135,68	10,231	118,25	8,410	123,07	9,255	130,85	9,756	136,43	12,141	140,98	11,769	117,41	8,045	793,00	0,000	122752,87
АТС5	132,20	10,482	116,83	8,083	121,17	8,975	126,88	7,644	134,30	12,324	138,04	13,623	115,91	8,091	764,00	0,000	105827,82
АТД0	80,23	6,291	70,73	6,193	76,43	6,243	74,44	5,686	83,52	6,048	86,59	8,676	67,72	6,237	793,00	0,000	56550,17
АТД1	84,60	9,502	68,51	9,489	75,14	9,024	76,02	7,244	82,77	9,009	85,67	10,418	65,57	8,020	793,00	0,000	100688,71
АТД2	81,62	7,814	69,12	7,176	76,95	7,641	73,11	5,992	81,07	7,726	86,41	9,446	65,63	7,232	793,00	0,000	75864,25
АТД3	77,93	7,459	67,62	7,353	73,46	5,682	70,19	4,971	80,43	7,896	84,48	9,227	63,73	7,179	793,00	0,000	67588,51
АТД4	76,69	6,992	66,62	6,842	71,52	6,886	69,91	6,284	79,61	8,255	82,59	7,095	62,34	6,902	793,00	0,000	66716,62
АТД5	75,30	6,565	66,02	6,396	71,44	6,649	68,11	5,131	78,26	7,384	83,17	5,405	61,59	7,068	764,00	0,000	60617,13
ЧСС0	84,44	6,610	88,80	8,443	98,94	11,080	69,69	9,037	71,75	9,684	99,46	10,899	72,82	8,493	793,00	0,000	159238,17
ЧСС1	107,34	9,108	111,55	9,173	122,70	12,034	89,75	11,557	87,30	12,702	120,46	11,051	96,73	10,892	793,00	0,000	2036660,07
ЧСС2	91,84	8,128	97,46	8,524	110,14	11,676	74,76	10,368	73,16	11,108	110,30	10,407	79,43	10,021	793,00	0,000	217153,74
ЧСС3	88,96	6,820	92,68	7,704	104,86	11,919	70,94	10,500	72,73	13,644	106,50	9,531	72,48	8,676	793,00	0,000	212435,31
ЧСС4	88,06	7,154	90,99	8,106	102,77	12,179	70,63	10,643	70,61	8,224	105,98	7,943	72,28	8,732	793,00	0,000	198600,74
ЧСС5	86,60	7,452	90,53	8,109	102,43	11,536	70,57	10,870	70,35	8,141	104,09	7,871	72,92	9,377	764,00	0,000	183981,60

Рисунок 2.2 – Згрупована таблиця результатів

За даними середніх значень із результуючих таблиць ми побудували графіки середніх значень АТД і АТС по кластерам

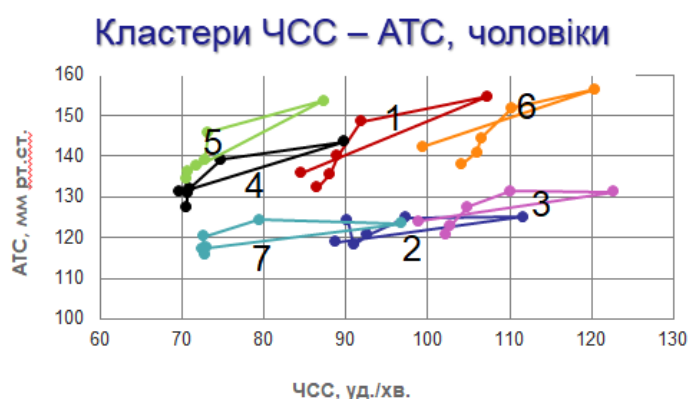


Рисунок 2.3 – Графіки ЧСС, АТС по кластерам

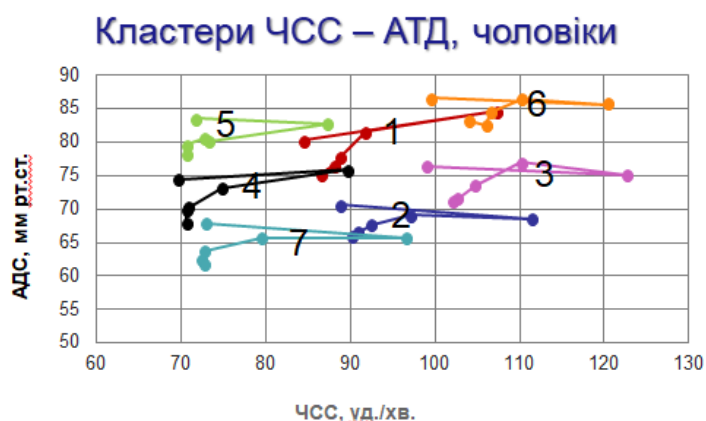


Рисунок 2.4 – Графік ЧСС, АТД по кластерам

Дані графіки показують динаміку артеріального тиску та пульсу при виконанні проби Мартіне у кожному кластері. Як ми бачимо, за АТС кластери 2 і 3, 7 і 2 перетинаються, а за АТД перетинаються кластери під номерами 1 і 4, 2 і 7, 1 і 6. Таким чином висуваємо гіпотезу про те, що вони мають спільну підгрупу. Також ми можемо побачити, що кластери 4 і 5 знаходяться приблизно

в одному діапазоні як за значеннями тиску, так і за значеннями пульсу, тому ми можемо зменшити кількість кластерів, щоб покращити наші результати.

Також слід перевірити чи дійсно алгоритм квадрату евклідової відстані дає найоптимальніший результат. Для цього необхідним було провести логістичну регресію та дискримінантний аналіз для побудови моделей з їх подальшою оцінкою та порівнянням із алгоритмом знаходження мінімальної дистанції за квадратом евклідової відстані.

Для проведення даних аналізів було застосовано статистичний пакет для обробки даних «Statistical Package for the Social Sciences » або просто SPSS.

Для початку проведемо логістичну регресію. Оскільки класифікатор студентів є не бінарною змінною, тому було прийнято рішення розбити пацієнтів на групи методом «один проти всіх» та вирівняти дані в групах, де були отримані показники класифікації занадто асиметричні. Результати дослідження представлені для третього кластеру, але аналогічна процедура проводилася для кожного кластеру окремо. Результуючу таблицю для всіх кластерів зображено у таблиці 2.1. На рисунку 2.5 зображено встановлення параметрів, які вводилися для отримання класифікації даних.

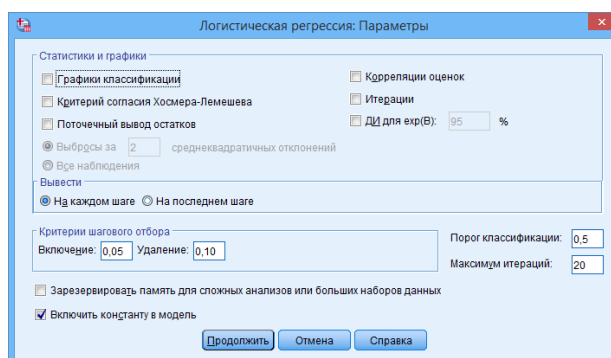


Рисунок 2.5 – Зображення параметрів

Розбиття на класи «один проти всіх»

На рисунку 2.6 наведені результати класифікації при об'єднанні у групи: перша група – студенти, що знаходяться в 3 кластері, друга – студенти в інших кластерах (3 проти 1, 2, 4, 5, 6, 7).

Для побудови логістичної регресії для великої кількості предикторів використано метод включення (условне).

Таблица классификации^а

	Наблюдаемые	Предсказанные		
		З		Процент корректных
		0	3	
Шаг 1	З, 0	238	39	85,9
	3	36	234	86,7
	Общий процент			86,3
Шаг 2	З, 0	243	34	87,7
	3	30	240	88,9
	Общий процент			88,3
Шаг 3	З, 0	244	33	88,1
	3	24	246	91,1
	Общий процент			89,6
Шаг 4	З, 0	245	32	88,4
	3	18	252	93,3
	Общий процент			90,9
Шаг 5	З, 0	245	32	88,4
	3	18	252	93,3
	Общий процент			90,9
Шаг 6	З, 0	244	33	88,1
	3	18	252	93,3
	Общий процент			90,7
Шаг 7	З, 0	245	32	88,4
	3	12	258	95,6
	Общий процент			92,0
Шаг 8	З, 0	245	32	88,4
	3	18	252	93,3
	Общий процент			90,9

^а. Разделяющее значение = .500

Рисунок 2.6 – Результаты класифікації ЛР

За результатами дослідження ми бачимо, що загальний відсоток коректно спрогнозованих даних складає 90,9%.

Таким чином, рівняння регресії набуває вигляду:

$$Y_1 = -31,370 \cdot \text{АТДО} - 0,095 \cdot \text{ЧСС0} + 0,096 \cdot \text{ЧСС2} + 0,128 \cdot \text{ЧСС3}$$

Для поліпшення якості класифікаторів було вирішено розширити матрицю змінних за допомогою нелінійних перетворень. Тому необхідним кроком було встановлення в параметрах аналізу покрокового режиму для відбору змінних

Таблица классификации^а

	Наблюдаемые	Предсказанные		
		З		Процент корректных
		0	3	
Шаг 1	З, 0	239	38	86,3
	3	36	234	86,7
	Общий процент			86,5
Шаг 20	З, 0	277	0	100,0
	3	0	270	100,0
	Общий процент			100,0

^а. Разделяющее значение = .500

Рисунок 2.7 – Результати класифікації ЛР з нелінійними перетвореннями

За результатами дослідження ми бачимо, що відсоток коректно спрогнозованих даних складає 100,0%. Але слід зазначити, що це при умові неповної моделі, оскільки при обробці даних SPSS видав наступне попередження:

Предупреждения

Оценивание не состоялось из-за числовых проблем.
Возможные причины: (1) по крайней мере один из критериев сходимости LCON и BCON равен нулю или слишком мал; (2) значение EPS слишком мало (если оно не задавалось, то, возможно, значение по умолчанию слишком мало для этих данных).

Рисунок 2.8 – Попередження

Таким чином, рівняння регресії має вигляд:

$$Y_1 = 172652,633 \cdot \text{АТС0} - 47,939 \cdot \text{ЧСС1} + 831,154 \cdot \text{ЧСС3} + 0,156 \cdot \text{КвАТС0} - 8,628 \cdot \text{КвЧСС3} - 0,003 \cdot \text{КубАТС0} + 0,000 \cdot \text{КубЧСС0} + 0,001 \cdot \text{КубЧСС1} + 0,030 \cdot \text{КубЧСС3} + 0,001 \cdot \text{КубЧСС5} - 731813,859 \cdot \text{ДелАТД2} - 268776,880 \cdot \text{ДелАТС4} - 3523,495 \cdot \text{ДелАТР3}$$

Порівнюючи дві моделі ми можемо дійти висновку, що модель, яка побудована на базі даних з додатковими змінними нелінійних перетворень є більш складною, хоча й не повною, але водночас дає вищий результат в порівнянні з моделлю, що включають лише істинні змінні.

Таблиця 2.1 – Порівняльна характеристика

Класифікатор	Позначення	ЛР на істинних даних	ЛР з нелінійними перетвореннями	Результат	
1-Всіх	Заг. Кор., %	90,4	100	2	
	Ск-ть	3/21	13/84		
	Ск, %	14,29	15,48		1
2-Всіх	Заг. Кор., %	77,4	91	2	
	Ск-ть	4/21	7/84		
	Ск, %	19,05	8,33		2
3-Всіх	Заг. Кор., %	90,9	100*	2	
	Ск-ть	4/21	13/84		
	Ск, %	19,05	15,48		2
4-Всіх	Заг. Кор., %	97,5	95,8	1	
	Ск-ть	5/21	13/84		
	Ск, %	23,81	15,48		2
5-Всіх	Заг. Кор., %	99,7*	100	1	
	Ск-ть	3/21	6/84		
	Ск, %	14,29	7,14		2
6-Всіх	Заг. Кор., %	100*	100	1	
	Ск-ть	4/21	5/84		
	Ск, %	19,05	5,95		2
7-Всіх	Заг. Кор., %	88	100	2	
	Ск-ть	4/21	15/84		
	Ск, %	19,05	17,86		2

Розшифрування міток в таблиці 2.1:

Заг.Кор,% - загальний відсоток коректно спрогнозованих даних

Ск-ть – складність моделі

Ск,% - відсоток складності моделі (чим нижче, тим краще)

21 – кількість істинних змінних

84 – кількість істинних змінних разом з нелінійними перетвореннями

1- виграш ЛР на істинних даних

2 – виграш ЛР з нелінійними перетвореннями

* - неповність моделі

З таблиці 2.1 видно, що аналіз, проведений на базі нелінійних перетворень дає складніші рівняння моделі, але їх складність, беручи до уваги всі змінні, в переважній кількості менша. Слід також зауважити, що у 4 тестах із 7 відсоток коректно спрогнозованих даних збільшився

Беручи до уваги таблицю 2.1 ми можемо зробити висновок, що проведення логістичної регресії, включаючи нелінійні перетворення, а саме: операція взяття квадрату та кубу, і операція взяття оберненої змінної, буде складніші рівняння регресії, але самі моделі є простішими за складністю і дають більш високий відсоток коректності.

Аналогічно до процедури проведення логістичної регресії дослідження були проведені методом дискримінантного аналізу. В результаті ми отримаємо аналогічну підсумкову таблицю.

Таблиця 2.2 – Порівняльна характеристика

Класифікатор	Позначення	ДА на істинних даних	ДА з нелінійними перетвореннями	Результат	
1-Всіх	Заг. Кор., %	87,2	94	2	
	Ск-ть	3/21	9/84		
	Ск,%	14,29	10,71		2
2-Всіх	Заг. Кор., %	77,2	84,6	2	
	Ск-ть	4/21	4/84		
	Ск,%	19,05	4,76		2
3-Всіх	Заг. Кор., %	91,4	92,1	2	
	Ск-ть	4/21	9/84		
	Ск,%	19,05	10,71		2
4-Всіх	Заг. Кор., %	85,4	88,9	2	
	Ск-ть	4/21	7/84		
	Ск,%	19,05	8,33		2
5-Всіх	Заг. Кор., %	94,3	97,6	2	
	Ск-ть	5/21	12/84		
	Ск,%	23,81	14,29		2
6-Всіх	Заг. Кор., %	93,9	97,1	2	
	Ск-ть	4/21	12/84		
	Ск,%	19,05	14,29		2
7-Всіх	Заг. Кор., %	88	88,5	2	
	Ск-ть	3/21	3/84		
	Ск,%	14,29	3,57		2

З таблиці 2.2 видно, що аналіз, проведений на базі нелінійних перетворень дає складніші рівняння моделі, але їх складність, беручи до уваги всі змінні, менша. Слід також зауважити, що у 7 тестах із 7 відсоток коректно спрогнозованих даних збільшився.

За допомогою дискримінантного аналізу було побудовано функції прогнозування для кожного кластера, при чому результати дослідження показують, що введення нелінійних змінних (для кожної вхідної змінної) до аналізу покращують результат класифікації.

Виходячи з результатів досліджень, ми можемо сказати, що при порівнянні алгоритмів, у 2 випадках із 7 (виключаючи неповну модель при логістичній регресії) найбільш точну і просту модель будує дискримінантний аналіз, ще в 2 із 7 – метод логістичної регресії. Ще у одному випадку логістична регресія та дискримінантний аналіз дають ідентичний результат. При цьому дискримінантний аналіз є більш простим та універсальним у використанні, оскільки при його застосуванні ми завжди маємо справу тільки з однією статистичною процедурою, в якій беруть участь одна категоріальна залежна змінна і кілька незалежних змінних з будь-яким типом шкали. Також з досліджень видно, що логістична регресія не завжди може побудувати повну модель. Тому ми ввели нелінійні змінні для методів логістичної регресії та дискримінантного аналізу.

2.3. ПРОГРАМНА ЧАСТИНА

Дослідження кластерів було вирішено проводити не на початковій базі даних, що містить 1500 досліджень і виступає в ролі бази для побудови результуючих таблиць, а на базі студентів та викладачів НТУУ «КПІ ім. І. Сікорського», що пройшли пробу Мартіне більше одного разу, оскільки на початковій базі даних результати відповідатимуть графікам кластерів, так як вони побудовані на основі даних із цієї бази даних.

Програмний продукт розроблено в середовищі Microsoft Visual Studio 2017 Community за допомогою мови програмування C# з використанням елементів WinForms та фреймворку .Net Framework. Для демонстрації

результатів кластеризації був розроблений багатофункціональний інтерфейс користувача.

Для того, щоб розпочати дослідження необхідно натиснути клавішу «Старт» і вибрати базу даних, що на одному листі містить артеріальний тиск та пульс, кластер, мітку, мінімальну та субмінімальну відстань. Слід зазначити, що для дослідження необхідно попередньо розрахувати АТР1, АТР2, АТР3, що показують різницю між систолічним та діастолічним тиском за першу, другу і третю хвилині після навантаження.

Для побудови графіків необхідно заповнити таблицю відповідності кластер-мітка (вона показує всі переходи між вибраним кластером і міткою, а також між міткою та кластером) або таблицю кластер-кластер (показує всі дані в обраних кластерах).

The screenshot displays the 'ClusterBox 2.0' software interface. It includes a top menu bar with 'Довідка' and 'Програма'. Below it, there are tabs for 'Дані' and 'Пошук та рахунок'. The main area contains several sections:

- Завантажена таблиця:** A table with columns: №, ПІБ, АТC0, АТC1, АТC2, АТC3, АТC4. It lists data for individuals like Artemeva I.S., Baturina K.Yu., and Bogdanov K.S.
- Таблиця відповідності кластер-мітка:** A table with columns: ПІБ, АТC1, АТC2, АТC3, АТD1, АТD2, АТD3, ЧСС1, ЧСС2, ЧСС3, АТР1, АТР2, АТР3, Кп., М, Р4. It shows data for Baturina K.Yu., Burdakov M.K., and Vadovchenko M.Yu.
- Таблиця відповідності кластер-кластер:** A table with columns: ПІБ, АТC1, АТC2, АТC3, АТD1, АТD2, АТD3, ЧСС1, ЧСС2, ЧСС3, АТР1, АТР2, АТР3, Кп., М. It shows data for Baturina K.Yu., Burdakov M.K., and Vadovchenko M.Yu.
- Таблиця кількості переходів кластер-кластер:** A small table with columns: Cluster, 1, 2, 3, 4, 5, 6, 7, 8. It shows transition counts between clusters.
- Таблиця кількості переходів кластер-мітка:** A small table with columns: Cluster, 1, 2, 3, 4, 5, 6, 7, 8. It shows transition counts between clusters and labels.

On the right side, there is a vertical label 'Графіки' (Graphs). At the bottom right, there are fields for 'Поранувати' (Sort by) and 'Загальна кількість: №' (Total count: #).

Рисунок 2.9 – Заповнені таблиці для дослідження

Якщо натиснути на клавішу «Графіки», то побачимо меню, в якому є таблиця коефіцієнтів регресії і клавіша «Графік АТС». При натисненні на «Графік АТС» побудуються лінії трендів «кластер 4» та «кластер 5», дані яких завантажені до таблиці відповідності кластер-кластер.

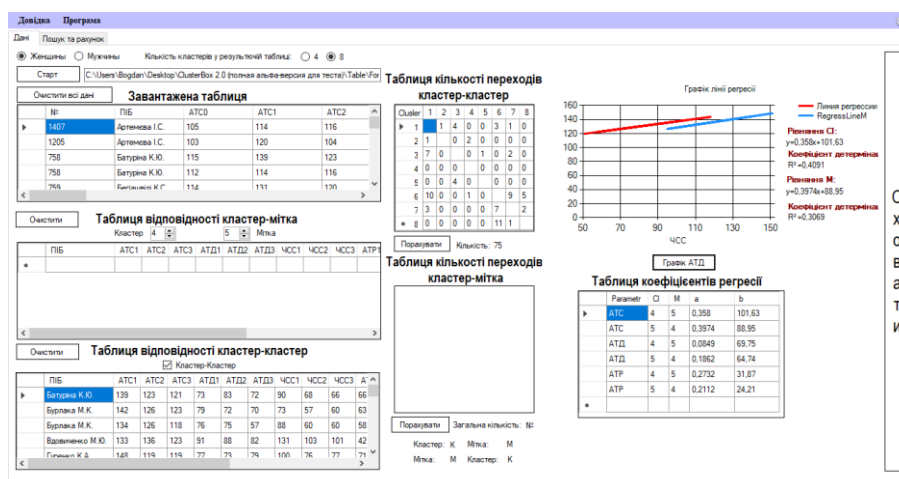


Рисунок 2.10 – виведення графіків і таблиці коефіцієнтів регресії

За проведеними дослідженнями за кластерами 4 і 5 ми бачимо, що за АТС коефіцієнт регресії a по кластеру 4 збігається з коефіцієнтом регресії a по кластеру 5 на 90,1%, а коефіцієнт b – на 87,5%; за АТД коефіцієнт регресії a збігається на 45,6%, а коефіцієнт b – на 92,8%; за АТР коефіцієнт регресії a збігається на 77,3%, а коефіцієнт b – на 76%;

Таким чином ми бачимо, що лінії регресії достатньо схожі між собою і можемо припустити, що дані кластери теж схожі. Це й підтверджують попередні графіки, на яких видно, що кластери 4 та 5 знаходяться в одному діапазоні. А, отже, це є ще одним фактором зменшення кількості кластерів та перерахунку результуючих таблиць вже на новій базі даних

Також дослідження алгоритму передбачає пошук студента в завантаженій базі даних для того, щоб подивитись чи перевищує радіус студента радіус кластера. Дані дослідження показують динаміку роботи алгоритму та можуть бути використані для моделювання системи виводу характеристик кластера, до якого відноситься студент, та рекомендацій мітки, якщо радіус студента перевищує радіус кластера. Робота алгоритму наведена на рисунку 2.24.

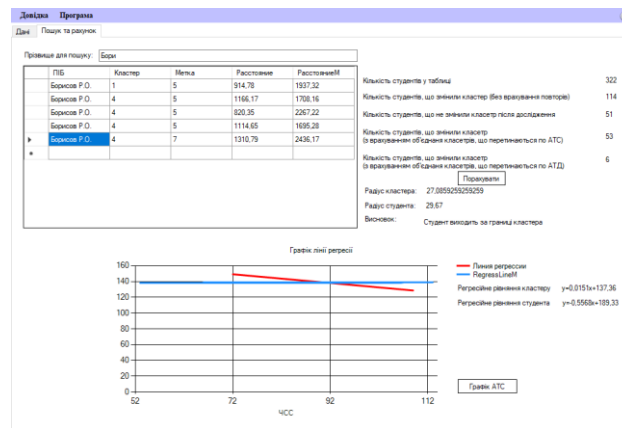


Рисунок 2.11 – Пошук студента в БД з виводом графіків

Якщо радіус студента перевищує радіус кластера, то автоматично розраховуються коефіцієнти ліній регресії і виводяться відповідні графіки.

2.4. АНАЛІТИЧНА ЧАСТИНА ЗА ВИСНОВКАМИ ПРОГРАМНОЇ ЧАСТИНИ

Наші дослідження показали, що зменшення кількості кластерів є необхідним кроком для поліпшення роботи алгоритму класифікації на основі квадрату евклідової відстані.

Таким чином ми повторно провели кластеризацію алгоритмом k-середніх. Спочатку було вирішено зменшити кількість кластерів на дві позиції, тому кількість кластерів становить 5.

Після кластеризації ми занесли результати до окремої таблиці Excel. Після цього завантажили її до SPSS та провели дисперсійний аналіз.

Описательные статистики								
		N	Среднее	Стд. отклонение	Стд. Ошибка	95% доверительный интервал для среднего		
						Нижняя граница	Верхняя граница	
Рs0	0	46	103,37	37,571	5,540	92,21	114,53	12
	1	54	111,44	14,127	1,922	107,59	115,30	90
	2	83	120,24	15,940	1,750	116,76	123,72	65
	3	75	120,24	13,195	1,524	117,20	123,28	74
	4	49	116,12	15,351	2,193	111,71	120,53	72
	5	48	130,48	15,739	2,272	125,91	135,05	81
	6	27	134,78	23,097	4,445	125,64	143,91	80
	19	2	19,00	,000	,000	19,00	19,00	19
	40	2	40,00	,000	,000	40,00	40,00	40
	41	2	41,00	,000	,000	41,00	41,00	41
	46	2	46,00	,000	,000	46,00	46,00	46
	67	2	67,00	,000	,000	67,00	67,00	67
	75	2	75,00	,000	,000	75,00	75,00	75
	Итого	394	116,60	24,452	1,232	114,17	119,02	12

Рисунок 2.12 – Описові статистики

За результатами дисперсійного аналізу було побудовано результуючу таблицю, яка містить показники стандартного відхилення та середніх значень за кожною змінною для кожного кластера.

Param	1		2		3		4		5	
	M	SD	M	SD	M	SD	M	SD	M	SD
Pa0	110,7391	12,01931	120,72	9,599493	121,1642	8,578686	117,1707	9,97723	131,675	8,558659
Pa0	64,45652	7,770904	69,54667	5,91118	75,29851	5,351338	71,82027	7,358337	80,175	6,511873
HR0	80,91304	11,378	69,45333	7,713262	82,92537	7,644197	101,7805	9,637199	91,525	8,249281
Pa1	110,7609	14,74552	131,4667	15,41001	126,7761	13,6035	123,7561	15,60253	135,8	12,83265
Pa1	57,86957	7,289593	68,32	10,09972	72,65672	8,557386	65,90244	9,087917	77,225	8,244618
HR1	105,2609	11,71596	92,24	10,86507	110,9552	10,66775	127,439	10,15886	117,35	12,65833
Pa2	118,8478	10,26097	131,6533	12,39932	132,1493	10,99484	126,6341	12,17324	140,85	9,434117
Pa2	60,47826	6,228034	67,56	8,109454	75,1194	7,56892	71,14634	6,440345	80,15	7,423283
HR2	87,80435	13,36766	73,44	9,091487	90	9,916317	111,439	11,14237	100,7	8,846845
Pa3	114,2826	10,26247	128,64	11,50868	127,6866	9,157228	125,2439	10,16804	134,65	9,509914
Pa3	58,69565	7,164324	66,09333	6,883536	73,52239	5,758678	70,60976	6,204345	78,8	6,536682
HR3	79,45652	12,92148	65,77333	8,554394	81,47761	7,634368	104,0244	8,341126	97,05	6,857337
Pa4	113,6522	8,353888	124,92	11,06707	123,9403	8,310053	121,7317	9,287153	130,075	9,965035
Pa4	58,58696	5,964995	65,12	6,959264	71,85075	6,631545	68,2439	5,351544	77,225	7,198602
HR4	77,84783	12,34318	66,68	8,226786	81,58209	7,744302	102,561	7,674792	96,325	6,627013
Pa5	110,5435	8,958439	122,6533	9,97769	122,4478	6,948476	118,9024	9,956919	128,325	9,32157
Pa5	59,1087	6,887116	63,96	7,010494	71,35821	6,285388	69,63415	9,093284	76,125	6,680195
HR5	77,86957	11,88577	67,98667	9,011244	81,62687	7,275604	102,1707	7,385467	95	5,193685

Рисунок 2.13 - Результуюча таблиця

Результуюча таблиця виступає базою даних для побудови нових графіків.

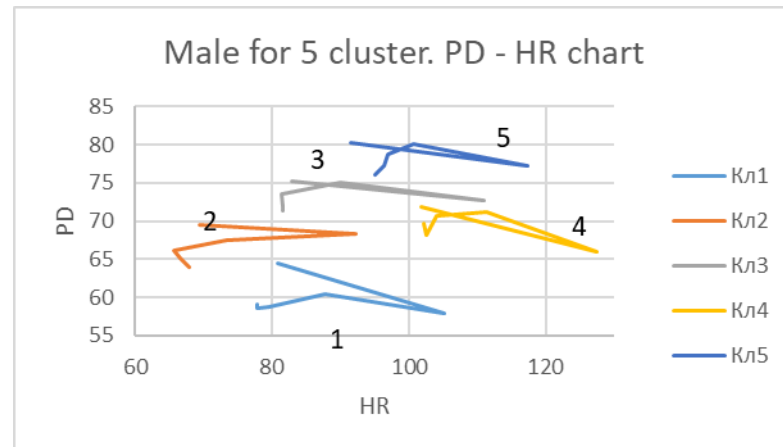


Рисунок 2.14 - Графіки АТД, ЧСС для 5 кластру

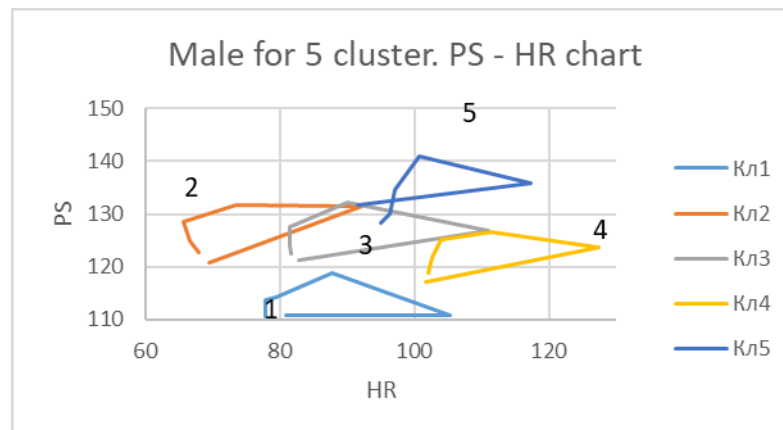


Рисунок 2.15 - Графіки АТС, ЧСС для 5 кластеру

Аналогічні аналізи при провели для чотирьох і трьох кластерів, отримали відповідні результуючі таблиці та побудували наступні графіки:

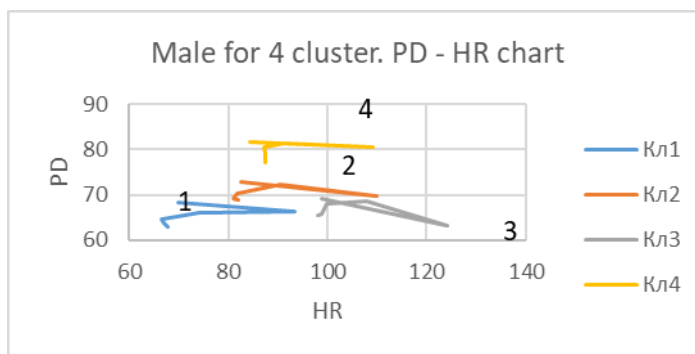


Рисунок 2.16 - Графіки АД, ЧСС для 4 кластру

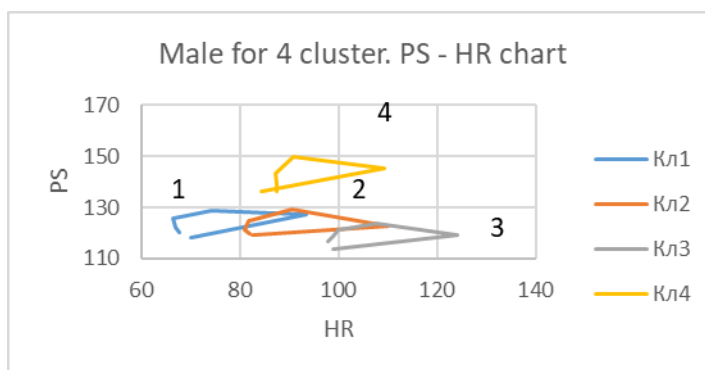


Рисунок 2.17 - Графіки АТС, ЧСС для 4 кластру

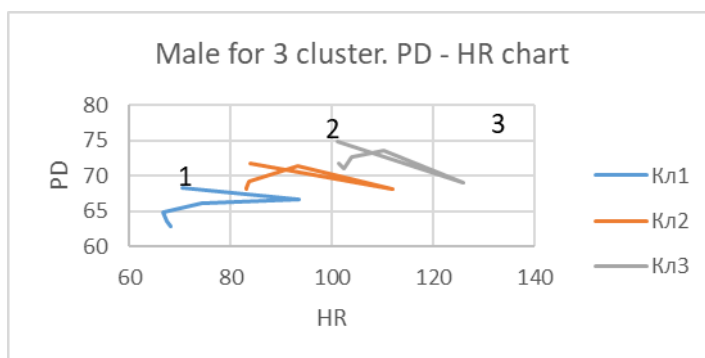


Рисунок 2.18 - Графіки АД, ЧСС для 3 кластеру

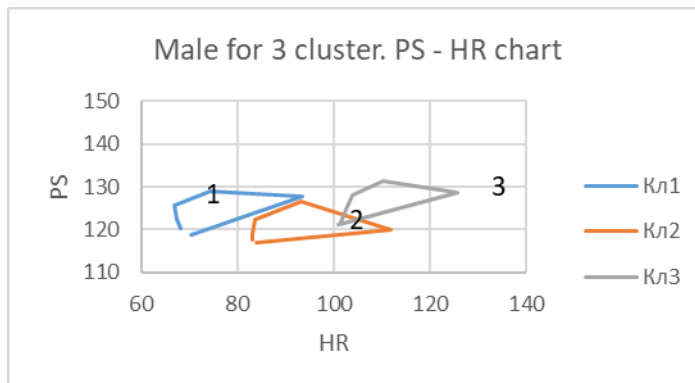


Рисунок 2.19 - Графіки АТС, ЧСС для 3 кластру

Дивлячись на графіки ми бачимо, що графіки на базі трьох кластерів є найменш інформативними, оскільки ми втрачаємо групу з даними, що показують високі показники систолічного тиску. Графіки на базі чотирьох кластерів показують, що група з низькими показниками систолічного тиску зникає, але в цілому інші графіки показують, що кластери досить різні. Тому виходячи з графіків ми можемо сказати, що найбільш інформативними є графіки на базі п'яти кластерів. Також ми можемо подивитися як змінюється дисперсія зі зменшенням кількості кластерів

Таблиця 2.3 – Зміна дисперсії

Параметр	1	2	3	4	5	6	7
Внутгруп.	80,00887	78,13669	79,88485	73,29032	41,71111	96,80202	81,71799
Міжгруп.	75,93598						
Внутгруп.	106,4002	92,65487	72,12717	89,51328	74,038		
Міжгруп.	86,9467						
Внутгруп..	110,0338	84,21743	93,54562	92,33333			
Міжгруп.	95,03256						
Внутгруп.	107,6799	85,97063	104,1654				
Міжгруп.	99,27197						

З порівняльної таблиці ми бачимо, що дисперсія досить сильно змінюється при зменшенні кластерів на 2 і 3 позиції, але при зменшенні на 4, дисперсія майже не змінилася, тому вибрати найоптимальнішу кількість виходячи з аналітичної частини треба серед таблиць на 5 і 4 кластери.

Виходячи з графіків та дисперсійного аналізу чітко видно, що таблиці на 5 кластерів дають найоптимальніший результат, тому нами було вирішено розширити функціонал програмного продукту Clusterbox і додати до нього можливість вибору результуючої таблиці на 5 кластерів. Таким чином ми надаємо вибір автоматичного застосування різних таблиць, просто вибираючи їх у меню програми.

АНАЛІЗ СТАРТАП-ПРОЕКТУ ТА МОЖЛИВОСТІ ЙОГО РИНКОВОГО ВПРОВАДЖЕННЯ ЗА ТЕМОЮ МД

I. Звіт про проект

1. ПІБ (команда)

ВОЙНИК БОГДАН ОЛЕКСІЙОВИЧ, БОРИСОВА ГАЛИНА ВІКТОРІВНА

2. Назва проекту

ІНФОРМАЦІЙНА СИСТЕМА «CLUSTERBOX» ДЛЯ ІДЕНТИФІКАЦІЇ
ОСІБ ІЗ ЗНИЖЕНИМИ РЕГУЛЯТОРНИМИ РЕЗЕРВАМИ

3. Короткий опис проекту

Проект виробництва та реалізації програмного продукту для визначення мінімальної відстані до кластеру з подальшим наданням оцінки функціонального стану організму людини.

Відмінними якостями продукту є:

- Надійність
- Точність
- Достовірність

Виконує вимоги:

- Перевірки коректності введених даних у всіх вкладках програми
- Розрахунку мінімальної відстані до кластеру
- Виведення результатів дослідження
- Розрахунку субмінімальної відстані
- Порівняння мінімальної та субмінімальної відстані
- Порівняння радіусу кластера з даними студента, що обстежується
- Виведення додаткових характеристик і рекомендацій, якщо це необхідно
- Визначення індексу маси тіла

4. Споживчі властивості товару

Програмний продукт для покупців має головну споживчу властивість: в результаті використання програми за призначенням споживач має можливість визначити регуляторні резерви організму, визначити свою групу ризику та необхідні рекомендації стосовно стану здоров'я.

Споживчий продукт: параметри вірогідності для встановлення приналежності особи до певної групи ризику:

№	Назва параметру	Розмірність	Норма	Факт
1	АТД (артеріальний діастолічний тиск)	<u>мм рт. ст.</u>	80	індивідуально
2	АТС (артеріальний систолічний тиск)	<u>мм рт. ст.</u>	120	індивідуально
3	ЧСС (частота серцевих скорочень)	уд./мин	60-80	індивідуально

5. Дослідження ринку

На світовому ринку програмних продуктів медичного спрямування зі спеціальними властивостями автоматизації процесу визначення функціональних реакцій на тестове навантаження шляхом реалізації розрахунку квадрату евклідової відстані аналоги відсутні (власне НОУ-ХАУ авторів проекту)

6. Дослідження конкурентного оточення

Конкурентне оточення продажу аналогів програмного продукту - відсутнє

7. Маркетингова стратегія просування

Новий товар на існуючий ринок програмних продуктів медичного спрямування

8. Елементи фінансового плану

1) Опис бізнес-проекту

Організація виробництва та реалізації інформаційної системи «clusterbox» для ідентифікації осіб із зниженими регуляторними резервами у формі стартап фірми у відповідності до заявленого проекту.

2) Опис товару/послуги/технології

Програмний продукт медичного спрямування зі спеціальними властивостями автоматизації процесу визначення функціональних реакцій на тестове навантаження шляхом реалізації розрахунку квадрату евклідової відстані

3) Маркетинг та продаж

Дослідження ринку програмних продуктів медичного спрямування зі спеціальними властивостями автоматизації процесу визначення функціональних реакцій на тестове навантаження показав: В Україні випуск та продаж подібних систем - відсутній

4) Фінансовий план (розробляється для конкретного інвестора)

5) Резюме

В сучасних умовах життя неможливо бути впевненим у гарному стані свого здоров'я. Відомі неодноразові випадки загибелі людини підчас помірних фізичних навантажень. Тобто ризик існує завжди: ранкова пробіжка, заняття з фізкультури у школі чи фізичне виховання в університеті, не кажучи вже про високі навантаження підчас тренувань у спортзалах. Данна система, на базі вже існуючих даних інших осіб та даних особи яка проходить обстеження допомагає визначити приналежність людини до тієї чи іншої групи ризику в залежності від реакції тіла на певне фізичне навантаження. Таким чином ми отримуємо засіб, що здатен попередити негативні наслідки від фізичних навантажень, допомагає скорегувати та підібрати необхідні навантаження та виключає негативний вплив того чи іншого фізичного навантаження на здоров'я людини.

ВИСНОВОК

В ході роботи було реалізовано та автоматизовано в програмному додатку наступний алгоритм:

1. Визначення кластера з мінімальною відстані до аналізованого об'єкта за допомогою межі близькості;
2. Визначення субмінімальної евклідової відстані до наступного кластера, тобто наступного мінімуму відстані до наступного кластера;
3. Визначення параметрів цього тесту в межах середнього радіусу кластеру ("так" або "ні");
4. Якщо параметри тесту знаходяться в межах середнього радіусу кластера, то в результаті буде відображатися лише інформація про функціональні та психофізичні особливості найближчого кластера.
5. В іншому випадку буде показано додаткову інформацію про функціональні та психофізичні особливості кластера з мінімальною відстані.

Також результати системної роботи порівнювалися з експертними оцінками. Точність співпадиння обох типів оцінок становила 87%. Програма передбачає можливість зміни. Розмір вхідного вектора та кількість функціональних патернів не фіксуються і можуть бути змінені в процесі вдосконалення системи.

Реалізація цього алгоритму дозволяє проаналізувати отримані дані, визначити умови, які призводять до змін в кровообігу, таких як зменшення регуляторних резервів та значне підвищення артеріального тиску. Це дає можливість контролювати індивідуальний рівень фізичної активності, змінювати навчальну програму та виявляти ситуації, що потребують додаткового медичного контролю. Крім того, система, заснована на цьому алгоритмі, може бути використана для перевірки контролю великих груп студентів. Це допомагає ідентифікувати людей з високим ризиком несподіваних розладів кровообігу та вчасно приймати обмежувальні та медичні

дії.

Майбутні дослідження будуть присвячені пошуку більших відмінностей між кластерами для розширення їх характеристик. Наступним завданням також буде пошук закономірностей між кластерами та мітками студентів, які пройшли пробу Мартіне більше одного разу для можливості прогнозування їх групи ризику.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Брехман И.И. Валеология – наука о здоровье. – М.: ФиС, 1990. – 208с.
2. Chekalova N. Functional reserves of childrens' and adolescents' organisms. Methods of research and evaluation. Nizhny Novgorod: NizhGMA; 2010.
3. Chekalova N. Methods of research and evaluation of functional reserves of childrens and adolescents organisms: Guidelines. Nizhny Novgorod: NizhGMA; 2009.
4. Mikhailova S, Kuzmichev J, Zhulin N. Evaluation of the functional state of students on the results of step test and MartineKushelevsky test. Vestnik Zdorov'e i Obrazovanie v XXI Veke. 2016;18(12):36-9.
5. Boyko A, Nastenکو I, Nosovets O, Voynyk B, Fedchishin M. Evaluation of coronary blood supply conditions for Martine's modified test for junior students. Visnyk Universytetu "Ukraine". 2017;1:51-62;
6. B.A. Voynyk, G.V. Borisova, V.S. Umanets¹, G.L. Boiko¹, A.V. Pavlov, Ie.A. Nastenکو. Automated assessment of a students circulatory system functional state using martine's test. Innov Biosyst Bioeng, 2018, vol. 2, no. 3, 144–148.
7. A.K. Jain, M.N. Murty, P.J. Flynn – “Data Clustering: A Review” [Електронний ресурс] – Режим доступу: <http://www.csee.umbc.edu/nicholas/clustering/p264-jain.pdf>
8. J. Kogan, C. Nicholas, M. Teboulle – “Clustering Large and High Dimensional data” [Електронний ресурс] – Режим доступу: <http://www.csee.umbc.edu/nicholas/clustering/tutorial.pdf>
9. Кластерный анализ [Електронний ресурс] – Режим доступу: http://statlab.kubsu.ru/sites/project_bank/claster.pdf
10. Кластерный анализ [Електронний ресурс] – Режим доступу: <http://statsoft.ru/home/textbook/modules/stcluan.html>
11. K-means [Електронний ресурс] – Режим доступу: <https://ru.wikipedia.org/wiki/K-means>

ПРЕЗЕНТАЦІЯ ДО ЗАХИСТУ ПЕРЕДДИПЛОМНОЇ ПРАКТИКИ