**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ**

**“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ**

**імені ІГОРЯ СІКОРСЬКОГО”**

**Факультет біомедичної інженерії**

**Кафедра біомедичної кібернетики**

ЗВІТ

**з переддипломної практики**

освітньо-кваліфікаційного рівня “**бакалавр**”

з напряму підготовки – 6.050101 «Комп’ютерні науки»

|  |  |
| --- | --- |
| **На тему** | **Реалізація програмного продукту для віднесення** |
| **студентів 1-2 курсу за функціональними реакціями на тестове** | |
| **навантаження до найближчого кластеру** | |
|  | |

(тема індивідуального завдання на початок практики)

**Студент 4 –го курсу гр. БС-31 Войник Богдан Олексійович \_\_\_\_\_\_\_\_\_\_\_**

(підпис)

**Керівник практики від кафедри БМК**

**\_\_\_\_\_*ст.викл. Аверьянова О.А*.\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_**

(вчені ступінь та звання, прізвище,ініціали)(підпис)

**\_\_\_\_\_\_*ас. Корнієнко Г.А.*\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_**

(вчені ступінь та звання, прізвище,ініціали)(підпис

**Керівник ДР \_\_\_\_\_*зав. каф. Настенко Є.А*.\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_**

(вчені ступінь та звання, прізвище,ініціали)(підпис)

**Консультант *кафедри*** *«Охорони праці, промислової та цивільної безпеки»*

**з розділу ДР «Безпеки життєдіяльності та охорони праці»**

**\_\_\_\_\_*доц. Демчук Г.В.\_\_\_\_\_\_\_\_\_* \_\_\_\_\_\_\_\_\_\_\_**

(вчені ступінь та звання, прізвище,ініціали)(підпис)

#### Київ – 2017 р.

Зміст

[ВСТУП 3](#_Toc481710203)

[АНОТАЦІЯ 4](#_Toc481710204)

[АННОТАЦИЯ 5](#_Toc481710205)

[ANNOTATION 6](#_Toc481710206)

[1 ОГЛЯД ЛІТЕРАТУРНИХ ДЖЕРЕЛ З ТЕМИ індивідуального завдання 7](#_Toc481710207)

[1.1. Кластерний аналіз 7](#_Toc481710208)

[1.2. Метод кластеризації k-середніх 8](#_Toc481710209)

[1.3. Міри відстані 8](#_Toc481710210)

[2 ОСНОВНА ЧАСТИНА ПЕРЕДДИПЛОМНОЇ ПРАКТИКИ 10](#_Toc481710211)

[2.1. Теоретична частина 10](#_Toc481710212)

[2.2. Аналітична частина 12](#_Toc481710213)

[2.3. Програмна частина 20](#_Toc481710214)

[ВИСНОВОК 23](#_Toc481710215)

[СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ 24](#_Toc481710216)

[ПРЕЗЕНТАЦІЯ ДО ЗАХИСТУ ПЕРЕДДИПЛОМНОЇ ПРАКТИКИ 25](#_Toc481710217)

# ВСТУП

**Актуальність:** При проходженні тестів пов’язаних з реакцією на фізичне навантаження студент хоче дізнатись свої результати та проблеми зі здоров’ям, зазвичай в навчальних корпусах не має програмного забезпечення для визначення цих даних. Наша програма дає змогу протягом 5 хвилин після проходження тесту на навантаження дізнатись більшість ризиків пов’язаних зі здоров’ям. Ця тема досить актуальна, оскільки програма може застосовуватись в різного роду клініках та центрах реабілітації. Результатом використання програми є пришвидшення роботи лікарів та визначення діагнозу без довготривалої затримки.

**Об’єкт дослідження:** студенти 1-2 курсу Національного Університету України “Київського технічного університету ім. Ігоря Сікорського”.

**Предмет дослідження:** визначення мінімальної відстані до кластеру.

**Мета роботи:** класифікація студентів за допомогою методу кластеризації k-середніх.

У відповідності з метою ставлять такі завдання:

1. Проаналізувати базу даних та визначити, який з методів буде найбільш підходячим для кластеризації.
2. Ознайомитись з методомами кластеризації. Проаналізувавши методи отримати дані про найоптимальніший з них.
3. Проаналізувати міри відстані, та визначити найоптимальнішу.
4. Провести дисперсійний аналіз даних.
5. Згрупувати результати в компактну таблицю.
6. Розробити програмний продукт для визначення мінімальної відстані до кластеру.

**Методи розробки проекту:** Для реалізації програмного продукту буде застосовано середовище програмування Microsoft Visual Studio 2013 Express, зокрема використано мова програмування C# та Windows Form Application.

# АНОТАЦІЯ

1 «Переддипломна практика» (практика) є частиною циклу нормативних дисциплін ООП *бакалавра* за напрямом підготовки *6.050101 «Комп'ютерні науки».*

2. Загальна трудомісткість практики становить 4,5 кредитів (ЕКТС), 135 годин. Період проведення практики с 18.04.2017р. по 12.05.2017р. (12.05.2017р.).

3. Практика реалізувалась **Войником Богданом Олексійовичем** студентом *4-го курсу, гр. БС-31* кафедри *Біомедичної кібернетики* факультету *Біомедичної інженерії НТУУ «КПІ»* .

4. Тема практики: «Реалізація програмного продукту для віднесення студентів 1-2 курсу за функціональними реакціями на тестове навантаження до найближчого кластеру».

5. Ціль та задачі практики.

Ціль – розробити програмний продукт для знаходження мінімальної відстані до кластеру.

Задачі:

1. Аналіз бази даних студентів;

2. Аналіз методів кластеризації та визначення оптимального;

3. Розробка програмного продукту;

4. Перевірка роботи алгоритму на тестовій вибірці.

6. Результати по темі практики:

Отримано навички роботи в середовищі розробки Microsoft Visual Studio 2013 Express на мові C#;

Реалізовано практичне застосування набутих знань шляхом створення програмного продукту для знаходження мінімальної відстані до кластеру.

7. Зміст звіту по практиці:

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

8. За практикою надані документи контролю проходження практики:

- щоденник практики;

- індивідуальне завдання;

- звіт на \_29\_\_ листах, додаток до звіту на \_\_\_\_\_\_ листах;

- презентація на \_\_10\_\_\_ слайдах;

- відгук керівника ДР;

- план публікацій чи впровадження;

- гарантійний лист.

9. Сформована тема дипломної роботи до наказу: Оцінка функціональної реакції на тестове навантаження у студентів 1-2 курсу. Жінки. Чоловіки. Комплексний проект.

10. Проміжна атестація у формі *диференціального заліку*.

11. Ключові слова: кластер, кластеризація, k-середніх, мінімальна відстань.

# АННОТАЦИЯ

1 «Преддипломная практика» (практика) является частью цикла нормативных дисциплин ООП *бакалавра* по направлению подготовки *6.050101 «Компьютерные науки»*.

2. Общая трудоемкость практики составляет 4,5 кредитов (ЕКТС), 135 часов. Период проведения практики с 18.04.2017г по 12.05.2017г. (12.05.2017г.).

3. Практика реализовывалась **Войником Богданом Алексеевичем** студентом *4-го курса, гр. БС-31* кафедры *Биомедицинской кибернетики* факультета *Биомедицинской инженерии НТУУ «КПИ»* .

4. Тема практики: «Реализация программного продукта для отнесения студентов 1-2 курса за функциональными реакциями на тестовое нагрузки к ближайшему кластеру».

5. Цель и задачи практики.

Цель – разработать программный продукт для нахождения минимального расстояния к ближайшему кластеру.

Задачи:

1. Анализ базы данных студентов

2. Анализ методов кластеризации и определение самого оптимального

3. Разработка программного продукта

4. Проверка работы алгоритма на тестовой выборке.

6. Результаты по теме практики:

Получены навыки работы в среде разработки Microsoft Visual Studio 2013 Express на языке C#;

Реализовано практическое применение приобретенных знаний путем создания программного продукта для нахождения минимального пути к кластеру.

7. 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.

8. По практике предоставлены документы контроля прохождения практики:

- дневник практики;

- индивидуальное задание

- отчет на \_29\_\_ листах, приложение к отчету на \_\_\_\_\_\_ листах.

- презентация на \_\_10\_\_\_ слайдах;

- отзыв руководителя ДР;

- план публикаций или внедрения;

- гарантийное письмо.

9. Сформированная тема дипломной работы в приказ: Оценка функциональных реакций на тестовые нагрузки у студентов 1-2 курса. Женщины. Мужчины. Комплексный проект.

10. Промежуточная аттестация в форме *дифференциального зачета*.

11.Ключевые слова: кластер, кластерный анализ, k-средних, минимальное расстояние.

# ANNOTATION

1 «Undergraduate practice» (practice) is a part of cycle regulatory disciplines PLO undergraduate field of study *6.050101 «Computer Science»*.

2. Total labor practices is 4.5 credits (ECTS), 135 hours. Practice period from 18.04.2017 to 12.05.2017. (12.05.2017.).

3. The practice was implemented by **Voinyk Bohdan Oleksyiovich** 4th course student, group BS-31, *Biomedical cybernetics* department of *Biomedical engineering faculty*, NTUU «KPI».

4. Practice theme: « Implementation of the software to assign students 1-2 courses for functional responses to a test load to the closest cluster ».

5. Goal and tasks:

Goal – To develop software for finding minimum distance to the cluster.

Tasks:

1. Analysis of students database

2. Clustering analysis methods and finding the most optimal

3. Software product development

4. Checking the algorithm using a test sample.

6. Practice results:

Acquired skills in development environment Microsoft Visual Studio 2013 Express language C#;

Implemented practical usage of knowledge by creating a software for finding minimum distance to the cluster.

7. 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.

8. The control documents in practice:

- practice diary;

- individual task;

- report on \_29\_\_ pages, attachment to report on \_\_\_\_\_\_ pages;

- presentation on \_\_\_10\_\_ slides;

- scientific director’s recall;

- plan of publishing or implementation;

- letter of indemnity.

9. Formed theme of the graduate work in order: Assessment of functional response to stress test the students 1-2 course. Women. Men. Complex project.

10. Interim certification in the form of a differential offset.

11. Key words: k-means, clustering, minimum distance.

# 1 ОГЛЯД ЛІТЕРАТУРНИХ ДЖЕРЕЛ З ТЕМИ індивідуального завдання

# 1.1. Кластерний аналіз

Кластерний аналіз – багатовимірна статистична процедура, що виконує збір даних, котрі містять інформацію про вибірку об’єктів, і потім впорядковують об’єкти в порівняно однорідні групи. Завдання кластеризації відноситься до статистичної обробки, а також до широкого класу задач навчання без вчителя.

Суттєвою особливістю кластерного аналізу являється те, що він не є звичайним статистичним методом, оскільки до нього у більшості випадків незастосовні процеси перевірки статистичної значимості. Кластерний аналіз дає найбільш можливо-значиме рішення. Саме тому досить часто його використовують тоді, коли дослідник має набір даних, але не має жодної апріорної гіпотези про класи цих даних.

Метою кластерного аналізу є організація отриманих даних у наглядні структури. Фактично, кластерний аналіз є набором різноманітних алгоритмів класифікації. Згідно з авторами підручника Statsoft – кластеризувати можна більшість симптомів захворювань, а також види їх лікування, крім цього це дає змогу досить цікаво їх класифікувати.

При класифікації великих масивів даних на групи кластерний аналіз є незамінним інструментом, оскільки дає змогу зробити це без особливих проблем.

Особливості кластерного аналізу:

* Більшість методів кластерного аналізу є доволі простими з евристичної точки зору, а також, здебільшого, не мають статистичного обґрунтування
* Використання різноманітних методів кластеризації створює різні кластерні рішення для однакових даних

Етапи проведення кластерного аналізу:

1. Проведення дослідження.
2. Підготовка даних до кластерного аналізу.
3. Вибір методу кластерного аналізу.
4. Вибір міри відстані між об’єктами та її обчислення.
5. Вибір стратегії кластеризації.
6. Застосування обраної стратегії для утворення кластерів.
7. Перевірка результатів кластерного аналізу на осмисленість і їх інтерпретації.

# 1.2. Метод кластеризації k-середніх

Алгоритм являє собою версію ЕМ-алгоритму, котрий застосовується також для розділення суміші гаусової функції. Він розбиває множину елементів векторного простору на завчасно відоме число кластерів k. Робота алгоритму зводиться до мінімізації середньоквадратичного відхилення на точках кожного кластеру. Основна ідея – це те, що на кожній ітерації перераховується центр кластерів для кожного кластеру, котрий був отриманий на попередньому кроці.

Після цього вектори розбиваються на кластери знову згідно з тим, який з обраних центрів виявився ближче до метрики.

# 1.3. Міри відстані

Евклідова відстань – найбільш популярна метрика, являється геометричною відстанню в багатомірному просторі. Дана метрика, як і більшість інших, чутлива до змін одиниць вимірювання осей. Тому, при використанні більшості метрик кластерного аналізу передбачає попередню стандартизацію.

Квадрат евклідової відстані – використовують, при необхідності надання більшої ваги віддаленішим одним від одного об’єктам.

Манхетенська відстань – зменшує вплив окремих великих різниць між однойменними координатами точок, так при обчисленні відстаней ці різниці не приводяться до квадрату.

Відстань Чебишева – застосовують, коли потрібно визначити два об’єкти, як відмінності, якщо вони відрізняються по будь-якій одній координаті.

Відстань Мінського – застосовуються, коли потрібно збільшити чи зменшити вагу, що відноситься до розмірності, для котрої відповідні об’єкти сильно відрізняються.

Процент незгоди – застосовується в тих випадках, коли данні являються категоричними.

# 2 ОСНОВНА ЧАСТИНА ПЕРЕДДИПЛОМНОЇ ПРАКТИКИ

# 2.1. Теоретична частина

Ця процедура намагається виявити відносно однорідні групи спостережень на основі обраних характеристик, використовуючи алгоритм, що дозволяє обробити велику кількість спостережень. Однак цей алгоритм вимагає вказівки числа кластерів. Ви можете задати початкові центри кластерів, якщо така інформація вам доступна. Ви можете вибрати один з двох методів класифікації спостережень, або ітеративно оновлюючи центри кластерів, або обмежуючись лише класифікацією. Ви можете зберегти приналежність до кластерів, інформацію про відстані і остаточні центри кластерів. Додатково ви можете задати змінну, значення якої будуть використовуватися в якості міток спостережень при виведенні результатів. Ви можете також запитати висновок F -Статистика дисперсійного аналізу. Відносні величини цих статистик дають інформацію про внесок кожної змінної в розділення груп.

*Приклад.* Чи можна розбити телевізійні шоу на групи, так щоб в кожній групі глядачі, яких вони залучають, були схожі? За допомогою кластерного аналізу методом k-середній ви можете розділити (кластеризувати) телевізійні шоу (спостереження) на k однорідних груп, виходячи з характеристик їх глядачів. Це можна використовувати при сегментації ринку. Або ви можете розбити міста (спостереження) на однорідні групи, що дозволить відбирати порівнянні міста для перевірки різних маркетингових стратегій.

*Статистики.* Повне рішення: початкові центри кластерів, таблиця дисперсійного аналізу. Для кожного спостереження: інформація про кластери, відстань від центру кластера.

*Дані для кластерного аналізу методом k-середніх*

*Дані.* Змінні повинні бути кількісними і вимірюваними в інтервального шкалою або шкалою відносин. Якщо змінні є двійковими або кількостями, скористайтеся процедурою Ієрархічний кластерний аналіз.

Порядок спостережень та початкових центрів кластерів. Алгоритм, який використовується за умовчанням для вибору початкових центрів кластерів, не є інваріантним щодо порядку спостережень. Параметр Використовувати ковзаючі середні в діалоговому вікні Ітерації робить виходить в результаті рішення потенційно залежним від порядку спостережень, незалежно від того, як вибираються початкові центри кластерів. При використанні будь-якого з цих методів, ви, можливо, захочете отримати кілька різних рішень з спостереженнями, розташованими у випадковому порядку, щоб упевнитися в стабільності даного рішення. Завдання початкових центрів кластерів і невикористання параметра Використовувати ковзаючі середні дозволить уникнути проблем, пов'язаних з порядком спостережень. Однак впорядкування початкових центрів кластерів може вплинути на рішення, якщо є збігаються відстані від спостережень до центрів кластерів. Щоб оцінити стабільність даного рішення, можна порівняти результати аналізу з різними перестановками значень початкових центрів.

*Припущення.* Для обчислення відстаней використовується просте евклідова відстань. Якщо необхідно задати інший тип відстані або міри схожості, зверніться до процедури Ієрархічний кластерний аналіз. Масштабування змінних грає важливу роль. Якщо ваші змінні мають різний масштаб вимірювань (наприклад, одна змінна виміряна в доларах, а друга - в роках), то результати можуть бути некоректними. У цій ситуації необхідно подумати про стандартизацію ваших змінних до виконання кластерного аналізу методом k-середній (це можна зробити за допомогою процедури Описові статистики). Передбачається, що вибрано відповідне число кластерів, а в аналіз включені всі істотні змінні. Якщо ви неправильно вибрали число кластерів або не включили важливі змінні, то отримані результати також можуть ввести вас в оману.

*Як запустити Кластерний аналіз методом k-середніх*

Для цієї процедури потрібно модуль Statistics Base.

1. Виберіть в меню:

Аналіз> Класифікація> Кластерний методом K середніх ...

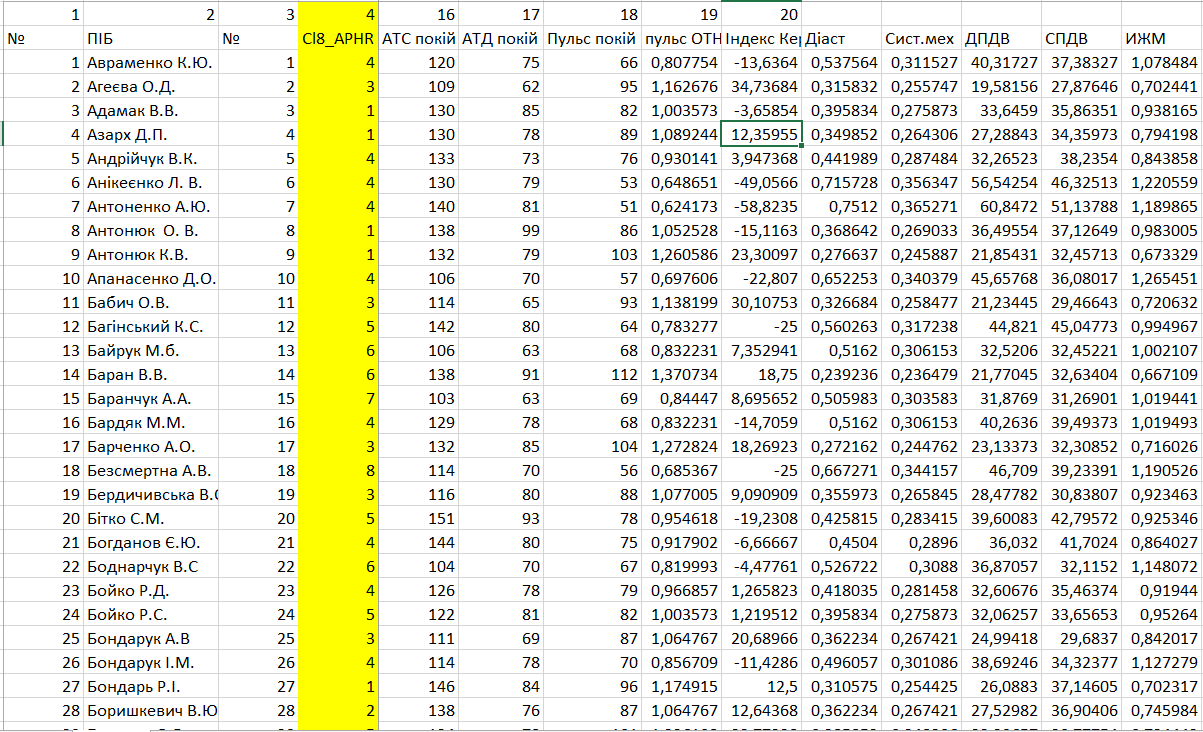
1. Виберіть змінні для використання в кластерному аналізі.
2. Задайте число кластерів. (Воно повинно бути не менше двох і не більше числа спостережень в файлі даних.)
3. Виберіть або метод Ітерації і класифікація, або метод Тільки класифікація.
4. Додатково можна вибрати таку нормативну змінну, щоб мітити спостереження.

# 2.2. Аналітична частина

Алгоритм включає в себе:

1. Приведення бази даних до прийнятного для SPSS виду
2. Виведення описових статистик
3. Дисперсійний аналіз для отримання центрів кластерів
4. Аналіз отриманих результатів
5. Зведення даних по ЧСС, АТД, АТС до окремої таблиці
6. Реалізація квадрата евклідової відстані для визначення мінімальної дистанції до кластеру

Алгоритм досліджувався на основі бази даних фізичних навантажень студентів 1-2 курсу.



*Рис. 1. База даних студентів*

База даних включає в себе 1500 студентів, що пройшли дослідження дихальної, кровоносної, нервової систем, а також дослідження системи фізичного стану. Кожна категорія містить певні параметри, що в сумі складають більше 300 зміних.

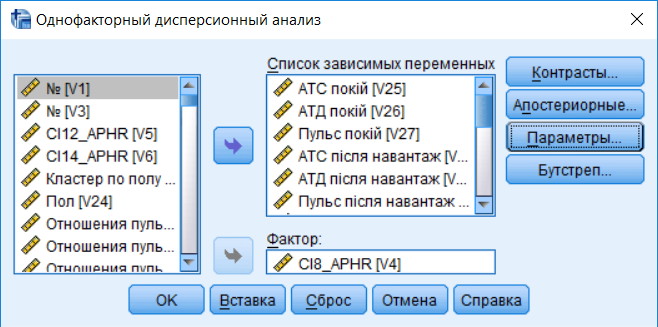
Після приведення бази даних Excel до прийнятного виду, ми можемо запустити її в SPSS та провести наступні операції перед початком аналізу БД:

* Встановлення імені змінних
* Встановлення мітки відповідності змінних
* Встановлення числовим змінним шкалу ”кількісну”

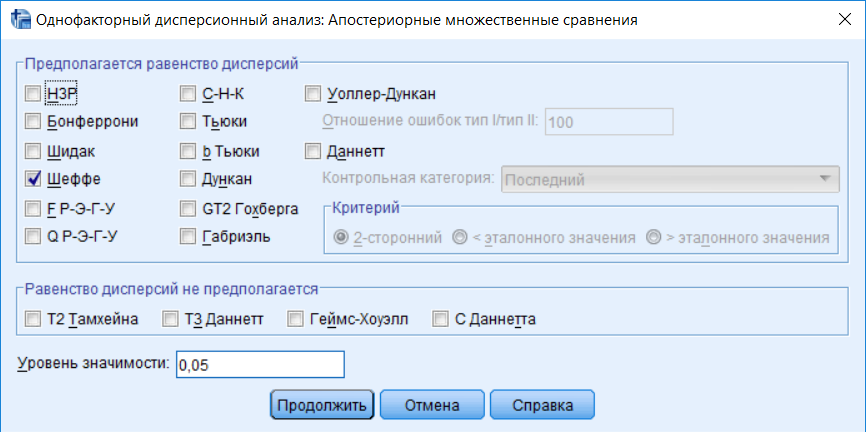


*Рис. 2. База даних готова до аналізу*

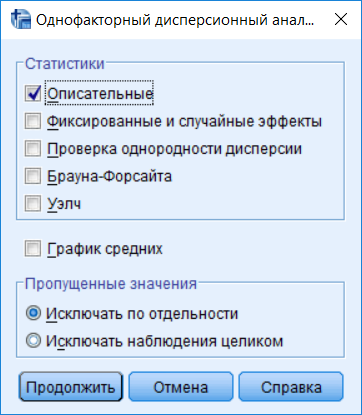
Наступним кроком є проведення дисперсійного аналізу для визначення стандартного відхилення в певному кластері, середнього значення, суми квадратів, степені свободи та значимості.



*Рис. 3. Вибір потрібних змінних для аналізу*



*Рис. 4. Вибір потрібного критерію*



*Рис. 5. Вибір потрібних статистик*

Результатом проведення дисперсійного аналізу з заданими вимогами є наступні таблиці:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Таблиця 1. Однофакторний дисперсійний аналізa.. Жінки. | | | | | | |
|  | | Сумма квадратов | ст.св. | Средний квадрат | F | Знч. |
| АТС покій | Между группами | 34450,461 | 7 | 4921,494 | 71,182 | ,000 |
| Внутри групп | 42451,476 | 614 | 69,139 |  |  |
| Итого | 76901,937 | 621 |  |  |  |
| АТД покій | Между группами | 20616,569 | 7 | 2945,224 | 62,801 | ,000 |
| Внутри групп | 28795,420 | 614 | 46,898 |  |  |
| Итого | 49411,989 | 621 |  |  |  |
| Пульс покій | Между группами | 54030,519 | 7 | 7718,646 | 114,324 | ,000 |
| Внутри групп | 41454,621 | 614 | 67,516 |  |  |
| Итого | 95485,140 | 621 |  |  |  |
| АТС після навантаж | Между группами | 55007,034 | 7 | 7858,148 | 73,117 | ,000 |
| Внутри групп | 65989,122 | 614 | 107,474 |  |  |
| Итого | 120996,156 | 621 |  |  |  |
| АТД після навантаж | Между группами | 25574,601 | 7 | 3653,514 | 71,602 | ,000 |
| Внутри групп | 31329,469 | 614 | 51,025 |  |  |
| Итого | 56904,071 | 621 |  |  |  |
| Пульс після навантаж | Между группами | 66685,796 | 7 | 9526,542 | 84,734 | ,000 |
| Внутри групп | 69031,670 | 614 | 112,429 |  |  |
| Итого | 135717,466 | 621 |  |  |  |
| АТС після 2 виміру | Между группами | 44431,952 | 7 | 6347,422 | 76,635 | ,000 |
| Внутри групп | 50855,876 | 614 | 82,827 |  |  |
| Итого | 95287,828 | 621 |  |  |  |
| АТД після 2 виміру | Между группами | 21895,578 | 7 | 3127,940 | 82,789 | ,000 |
| Внутри групп | 23198,250 | 614 | 37,782 |  |  |
| Итого | 45093,828 | 621 |  |  |  |
| Пульс після 2 виміру | Между группами | 105621,524 | 7 | 15088,789 | 207,314 | ,000 |
| Внутри групп | 44688,431 | 614 | 72,782 |  |  |
| Итого | 150309,955 | 621 |  |  |  |
| АТС після 3 виміру | Между группами | 35797,927 | 7 | 5113,990 | 74,789 | ,000 |
| Внутри групп | 41984,448 | 614 | 68,379 |  |  |
| Итого | 77782,375 | 621 |  |  |  |
| АТД після 3 виміру | Между группами | 20879,867 | 7 | 2982,838 | 82,496 | ,000 |
| Внутри групп | 22200,623 | 614 | 36,157 |  |  |
| Итого | 43080,490 | 621 |  |  |  |
| Пульс після 3 виміру | Между группами | 103975,365 | 7 | 14853,624 | 226,977 | ,000 |
| Внутри групп | 40180,854 | 614 | 65,441 |  |  |
| Итого | 144156,219 | 621 |  |  |  |
|  | | | | | | |

Таблиця 2. Описові статистикиa. Жінки.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | N | Среднее | Стд. Отклонение | Стд. Ошибка |
|
| АТС покій | 1 | 118 | 115,61 | 8,351 | ,769 |
| 2 | 43 | 130,40 | 11,776 | 1,796 |
| 3 | 127 | 116,63 | 8,584 | ,762 |
| 4 | 41 | 122,07 | 9,193 | 1,436 |
| 5 | 40 | 127,03 | 9,088 | 1,437 |
| 6 | 97 | 111,02 | 6,677 | ,678 |
| 7 | 76 | 103,32 | 8,503 | ,975 |
| 8 | 80 | 107,79 | 6,037 | ,675 |
| Итого | 622 | 114,78 | 11,128 | ,446 |
| АТД покій | 1 | 118 | 75,20 | 6,162 | ,567 |
| 2 | 43 | 85,84 | 8,574 | 1,308 |
| 3 | 127 | 74,21 | 6,108 | ,542 |
| 4 | 41 | 78,68 | 5,488 | ,857 |
| 5 | 40 | 84,48 | 13,001 | 2,056 |
| 6 | 97 | 68,96 | 5,422 | ,550 |
| 7 | 76 | 65,66 | 7,359 | ,844 |
| 8 | 80 | 69,79 | 4,883 | ,546 |
| Итого | 622 | 73,73 | 8,920 | ,358 |
| Пульс покій | 1 | 118 | 82,82 | 6,920 | ,637 |
| 2 | 43 | 83,84 | 8,171 | 1,246 |
| 3 | 127 | 92,99 | 9,536 | ,846 |
| 4 | 41 | 68,51 | 8,623 | 1,347 |
| 5 | 40 | 99,08 | 11,947 | 1,889 |
| 6 | 97 | 74,88 | 7,165 | ,728 |
| 7 | 76 | 82,18 | 8,296 | ,952 |
| 8 | 80 | 68,58 | 6,121 | ,684 |
| Итого | 622 | 81,92 | 12,400 | ,497 |

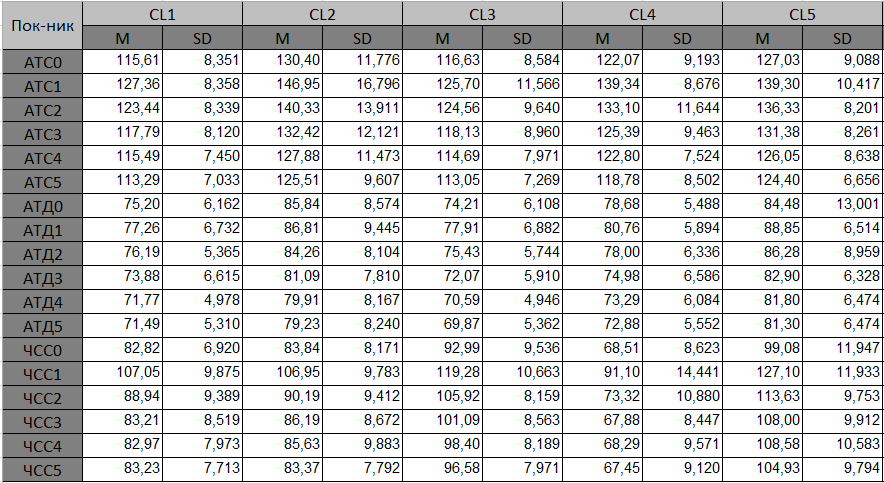
Таблиця 3. Описові статистикиa. Чоловіки.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | N | Среднее | Стд. отклонение | Стд. Ошибка |
|
| АТС покій | 1 | 113 | 135,84 | 9,647 | ,908 |
| 2 | 130 | 119,08 | 9,518 | ,835 |
| 3 | 106 | 124,02 | 9,499 | ,923 |
| 4 | 156 | 131,24 | 9,641 | ,772 |
| 5 | 44 | 137,68 | 12,649 | 1,907 |
| 6 | 46 | 142,39 | 13,807 | 2,036 |
| 7 | 199 | 117,49 | 9,740 | ,690 |
| Итого | 794 | 126,50 | 13,033 | ,463 |
| АТД покій | 1 | 113 | 80,23 | 6,291 | ,592 |
| 2 | 130 | 70,73 | 6,193 | ,543 |
| 3 | 106 | 76,43 | 6,243 | ,606 |
| 4 | 156 | 74,44 | 5,686 | ,455 |
| 5 | 44 | 83,52 | 6,048 | ,912 |
| 6 | 46 | 86,59 | 8,676 | 1,279 |
| 7 | 199 | 67,72 | 6,237 | ,442 |
| Итого | 794 | 74,45 | 8,445 | ,300 |
| Пульс покій | 1 | 113 | 84,44 | 6,610 | ,622 |
| 2 | 130 | 88,80 | 8,443 | ,740 |
| 3 | 106 | 98,94 | 11,080 | 1,076 |
| 4 | 156 | 69,69 | 9,037 | ,724 |
| 5 | 44 | 71,75 | 9,684 | 1,460 |
| 6 | 46 | 99,46 | 10,899 | 1,607 |
| 7 | 199 | 72,82 | 8,493 | ,602 |
| Итого | 794 | 81,45 | 14,171 | ,503 |

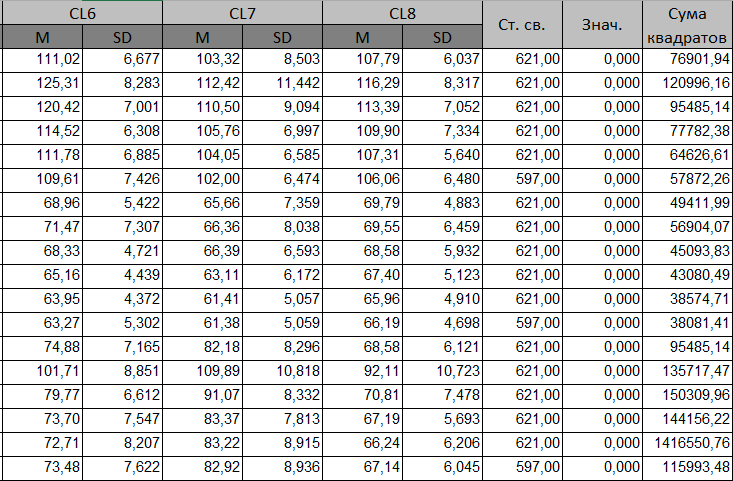
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Таблиця 4. Однофакторний дисперсійний аналізa. Чоловіки. | | | | | | |
|  | | Сумма квадратов | ст.св. | Средний квадрат | F | Знч. |
| АТС покій | Между группами | 54467,199 | 6 | 9077,866 | 89,044 | ,000 |
| Внутри групп | 80233,290 | 787 | 101,948 |  |  |
| Итого | 134700,489 | 793 |  |  |  |
| АТД покій | Между группами | 25405,688 | 6 | 4234,281 | 106,997 | ,000 |
| Внутри групп | 31144,484 | 787 | 39,574 |  |  |
| Итого | 56550,171 | 793 |  |  |  |
| Пульс покій | Между группами | 95941,075 | 6 | 15990,179 | 198,813 | ,000 |
| Внутри групп | 63297,096 | 787 | 80,428 |  |  |
| Итого | 159238,171 | 793 |  |  |  |
| АТС після навантаж | Между группами | 131794,771 | 6 | 21965,795 | 83,802 | ,000 |
| Внутри групп | 206284,253 | 787 | 262,115 |  |  |
| Итого | 338079,024 | 793 |  |  |  |
| АТД після навантаж | Между группами | 41166,650 | 6 | 6861,108 | 90,717 | ,000 |
| Внутри групп | 59522,062 | 787 | 75,632 |  |  |
| Итого | 100688,712 | 793 |  |  |  |
| Пульс після навантаж | Между группами | 112537,680 | 6 | 18756,280 | 161,993 | ,000 |
| Внутри групп | 91122,390 | 787 | 115,784 |  |  |
| Итого | 203660,071 | 793 |  |  |  |
| АТС після 2 виміру | Между группами | 77670,396 | 6 | 12945,066 | 99,432 | ,000 |
| Внутри групп | 102460,128 | 787 | 130,191 |  |  |
| Итого | 180130,524 | 793 |  |  |  |
| АТД після 2 виміру | Между группами | 33749,244 | 6 | 5624,874 | 105,112 | ,000 |
| Внутри групп | 42115,010 | 787 | 53,513 |  |  |
| Итого | 75864,253 | 793 |  |  |  |
| Пульс після 2 виміру | Между группами | 139342,218 | 6 | 23223,703 | 234,889 | ,000 |
| Внутри групп | 77811,521 | 787 | 98,871 |  |  |
| Итого | 217153,739 | 793 |  |  |  |
| АТС після 3 виміру | Между группами | 54558,799 | 6 | 9093,133 | 92,660 | ,000 |
| Внутри групп | 77231,988 | 787 | 98,135 |  |  |
| Итого | 131790,787 | 793 |  |  |  |
| АТД після 3 виміру | Между группами | 30446,111 | 6 | 5074,352 | 107,519 | ,000 |
| Внутри групп | 37142,403 | 787 | 47,195 |  |  |
| Итого | 67588,514 | 793 |  |  |  |
| Пульс після 3 виміру | Между группами | 140567,948 | 6 | 23427,991 | 256,554 | ,000 |
| Внутри групп | 71867,362 | 787 | 91,318 |  |  |
| Итого | 212435,310 | 793 |  |  |  |

Результуючі таблиці виводяться окремо для чоловіків та жінок, оскільки вони мають різну кількість кластерів.

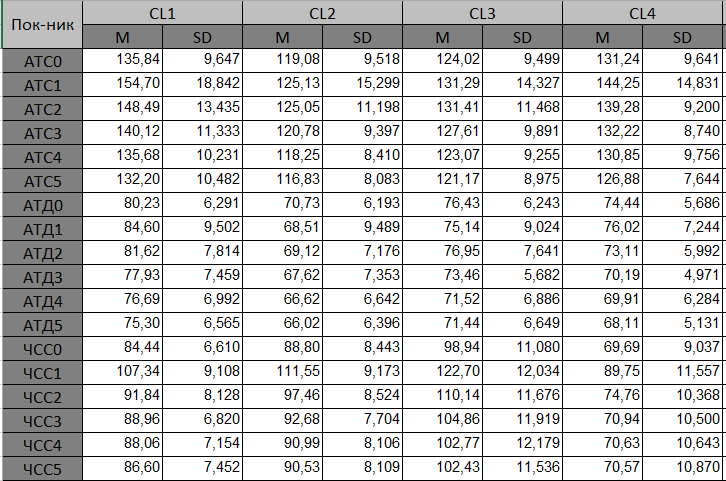
Після детального аналізу виведених результатів ми можемо згрупувати їх в компактну таблицю.



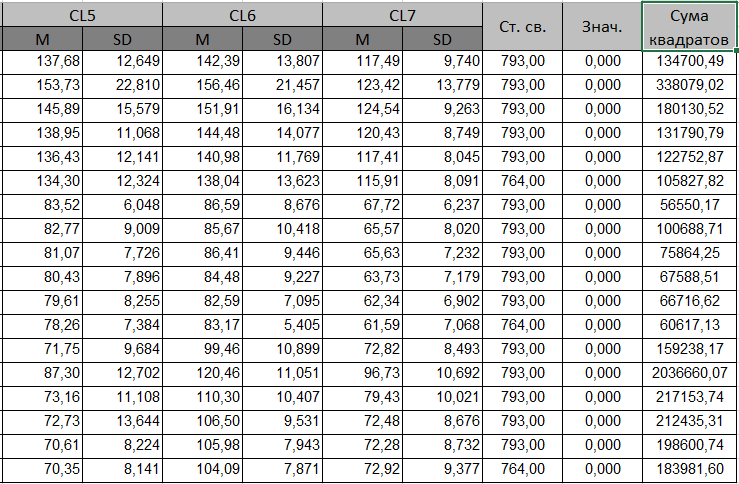
*Рис. 6. Згрупована таблиця результатів. Жінки.*



*Рис. 7. Продовження згрупованої таблиці результатів. Жінки.*



*Рис. 8. Згрупована таблиця результатів. Чоловіки.*

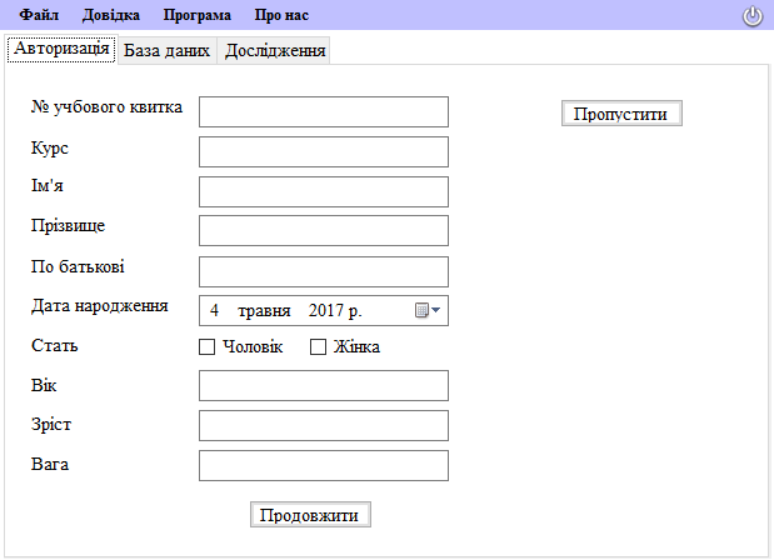


*Рис. 9. Продовження згрупованої таблиці результатів. Чоловіки.*

Як можна побачити на рисунках 6-9 дані стандартного відхилення, середнього значення, суми квадратів, значимості та степені свободи згруповані по змінним ЧСС, АТС, АТД, котрі відображають відповідні значення на кожній хвилині після навантаження та відносяться до певного кластеру.

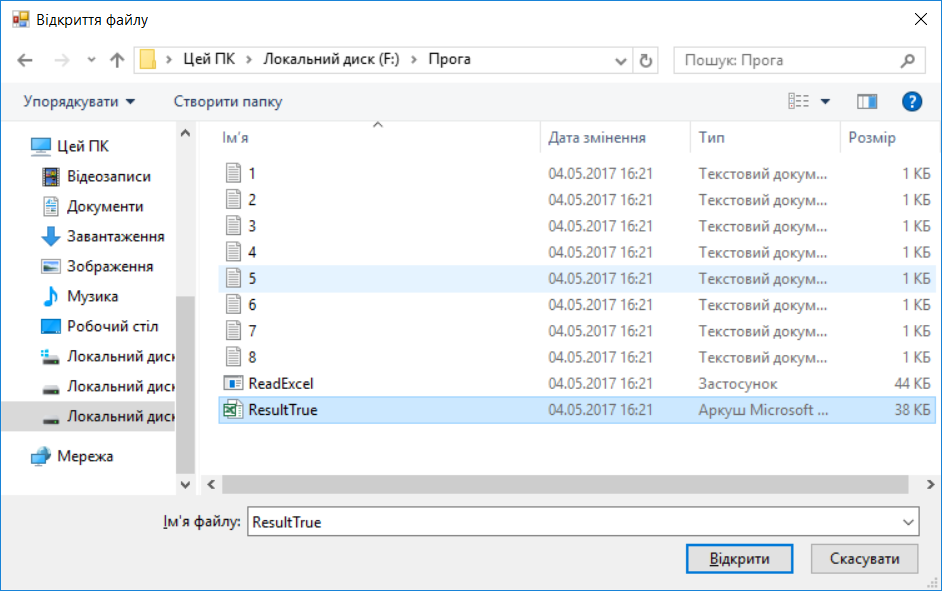
# 2.3. Програмна частина

Програма для визначення мінімальної відстані до центру кластера була розроблена за допомогою середовища програмування Microsoft Visual Studio 2013 Express та написана на мові програмування C#. Для демонстрації результатів кластеризації був розроблений багатофункціональний інтерфейс користувача.



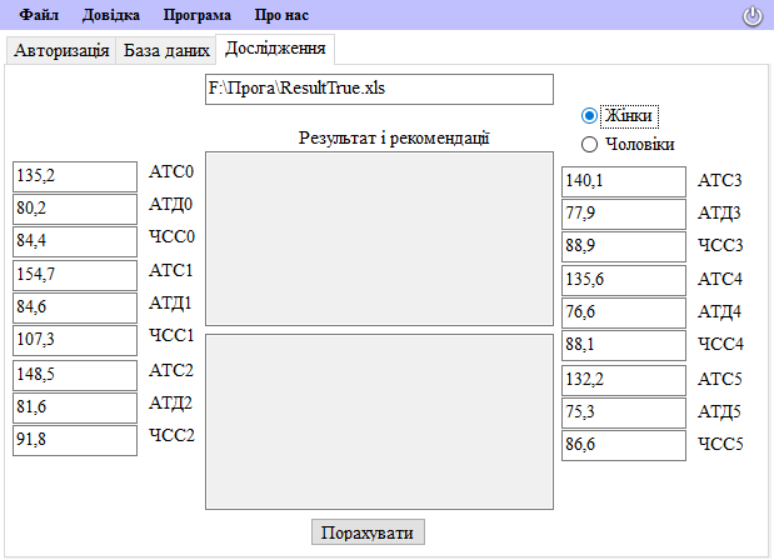
*Рис. 10. Інтерфейс програмного продукту*

Для того щоб розпочати процес визначення мінімальної відстані, необхідно вибрати таблицю формату Excel із вихідними даними(рис. 11.):



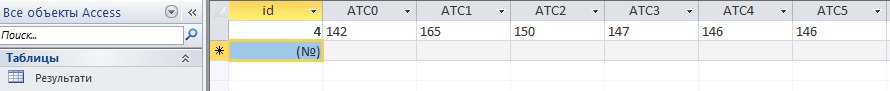
*Рис. 11. Вікно вибору файлу з даними*

Вибраний файл має містити вже відфільтровані дані без пропущених значень. Після вибору файлу можна приступити до заповнення полів тиску та пульсу студента, де АТС0, ЧСС0, АТД0 – дані спокою, а всі інші дані після навантаження, де індекс показує хвилину відпочинку.



*Рис. 12. Вікно проведення розрахунків*

Після проведення розрахунків, програма перевіряє чи існує база даних в котру можна зберегти результати, якщо її немає, то вона створюється і кожен наступний студент буде зберігатись в цю БД.



*Рис. 13. База даних Access з результатами розрахунків*

Програма також може проводити розрахунки без авторизації користувача, крім того в вкладці довідка можна дізнатись призначення програми та всі умовні позначення, а вкладка програма дає змогу отримати інструкцію користувача.

# ВИСНОВОК

У результаті роботи над переддипломною практикою було опрацьовано багато наукових джерел інформації стосовно теми практики, набуто професійних вмінь та навичок у роботі з сучасними інформаційними технологіями.

В ході виконання проекту було проаналізовано запропоновану базу даних студентів 1-2 курсів НТУУ “КПІ ім. Сікорського”. Після детального аналізу були розглянуті основні методи кластеризації даних, найоптимальнішим виявився метод k-середніх, оскільки він використовує квадрат евклідової відстані, що є невід`ємною частиною роботи програмного продукту.

Проведено дисперсійний аналіз, що дав змогу вирахувати середні показники значення, стандартне відхилення, степені свободи, значимість та суму квадратів.

Результати дисперсійного аналізу аналізу були згруповані в компактні таблиці для подальшої обробки програмним продуктом.

Результатом переддипломної практики став програмний продукт, що здатний визначати мінімальну відстань до кластеру спираючись на дані дисперсійного аналізу. Програмним продуктом також передбачене збереження результатів до бази даних формату Access.

Таким чином, переддипломна практика, що є джерелом накопичення практичного досвіду роботи по обраній спеціальності, закріпила наші теоретичні знання та показала нам, на що ми спроможні в роботі на підприємстві, в тому числі в колективі, зокрема і в команді. Вона відіграла важливу роль у підготовці нас, як спеціалістів високого рівня.

# СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. 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

1. 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

1. Кластеризация данных – Автор: Александр Котов, Николай Красильников, - 2 октября 2006 г.

[Електронний ресурс] – Режим доступу: http://yury.name/internet/02ia-seminar-note.pdf

1. Алгоритмы кластеризации на службе Data Mining

[Електронний ресурс] – Режим доступу: https://basegroup.ru/community/articles/datamining

1. Кластерный анализ

[Електронний ресурс] – Режим доступу: http://statlab.kubsu.ru/sites/project\_bank/claster.pdf

1. Кластерный анализ

[Електронний ресурс] – Режим доступу: http://statsoft.ru/home/textbook/modules/stcluan.html

1. K-means

[Електронний ресурс] – Режим доступу: https://ru.wikipedia.org/wiki/K-means

# ПРЕЗЕНТАЦІЯ ДО ЗАХИСТУ ПЕРЕДДИПЛОМНОЇ ПРАКТИКИ

