

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**  
**ХЕРСОНСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ**  
**Факультет комп'ютерних наук фізики та математики**  
**Кафедра інформатики, програмної інженерії та економічної**  
**кібернетики**

**ІНСТИТУЦІЙНІ АСПЕКТИ ІНТЕГРОВАНОГО ВИЯВЛЕННЯ**  
**КЛАСТЕРІВ АБИТУРІЄНТІВ ЗАСОБАМИ ІКТ**  
**Кваліфікаційна робота (проєкт)**

на здобуття ступеня вищої освіти “магістр”

Виконав: здобувач 2 курсу 261М групи  
Спеціальності 126 Інформаційні  
системи та технології  
Освітньо-професійної програми  
«Інформаційні системи та технології»  
другого (магістерського) рівня освіти  
Леденчук Станіслав Вадимович  
Керівник: доктор економічних наук,  
професор Кобець Віталій Миколайович  
Рецензент: доктор технічних наук,  
професор, професор кафедри  
програмних засобів і технологій  
Херсонського національного  
технічного університету Жарікова М.В.

## ЗМІСТ

<b>ВСТУП.....</b>	<b>3</b>
<b>РОЗДІЛ 1 ІНФОРМАЦІЙНІ СИСТЕМИ ДЛЯ ВСТУПУ</b>	
<b>АБІТУРІЄНТІВ В УКРАЇНІ .....</b>	<b>5</b>
1.1. Умови вступу до закладів вищої освіти .....	5
1.2. Перелік інформаційних систем з відкритими даними абітурієнтів	7
1.3. Аналіз пошукових запитів ЄДЕБО .....	9
1.4. Відбір критеріїв для аналізу.....	11
1.5. Порівняння та вибір ресурсу для парсингу даних абітурієнтів ...	13
<b>РОЗДІЛ 2 ПРОЄКТУВАННЯ ТА ПОБУДОВА АЛГОРИТМУ ДЛЯ</b>	
<b>КЛАСТЕРИЗАЦІЇ ДАНИХ .....</b>	<b>17</b>
2.1. Кластерний аналіз даних .....	17
2.2. Вибір алгоритму для аналізу даних .....	19
2.3. Застосування SOM за допомогою мови R та RStudio .....	25
<b>РОЗДІЛ 3 АНАЛІЗ ОТРИМАНИХ ДАНИХ ТА ВПРОВАДЖЕННЯ</b>	
<b>ПРОПОЗИЦІЙ.....</b>	<b>30</b>
3.1. Аналіз результатів отриманих після кластеризації даних абітурієнтів бакалаврського рівня.....	30
3.2. Аналіз даних вступників на магістерський рівень шляхом кластеризації заяв на ІТ-спеціальності до Херсонських ЗВО .....	37
3.3. Розробка та впровадження пропозицій щодо покращення умов вступу абітурієнтів до ЗВО .....	41
<b>ВИСНОВКИ .....</b>	<b>44</b>
<b>СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ .....</b>	<b>46</b>
<b>ДОДАТКИ.....</b>	<b>49</b>
Додаток А.....	49
Додаток Б .....	51

## ВСТУП

**Актуальність теми.** Різний рівень підготовки здобувачів, перелік сертифікатів ЗНО, пріоритетність, місце проживання створює передумови для забезпечення різних інституційних умов та індивідуальної траєкторії їх навчання. Цільова аудиторія абітурієнтів, розподілена на кластери, дозволить підготувати відповідні заходи з визначення цільових груп абітурієнтів для ІТ спеціальностей і перегляду освітніх програм для врахування специфіки попиту абітурієнтів на освітні послуги ІТ спеціальностей.

**Об'єктом дослідження** є абітурієнти ІТ спеціальностей, які планують вступ до закладів вищої освіти.

**Предметом дослідження** є визначення інституційних аспектів кластерів абітурієнтів інструментальними засобами на освітні програми ІТ спеціальностей ЗВО.

**Мета дослідження** – визначення кластерів абітурієнтів засобами ІКТ для адаптації освітніх програм ЗВО до попиту потенційних здобувачів вищої освіти.

### **Завдання дослідження:**

1. Проаналізувати існуючі інформаційні системи з відкритими даними учасників освітнього процесу.
2. Створити парсер для збору релевантних даних учасників освітнього процесу.
3. Розробити за допомогою алгоритмів самоорганізації кластери учасників освітнього процесу на основі їх релевантних характеристик для бакалаврського і магістерського рівнів.
4. Підготувати пропозиції для врахування характеристик абітурієнтів для підвищення частки вступників на ІТ спеціальності Херсонського державного університету.

### **Методи дослідження:**

- 1) Ієрархічна кластеризація (розподіл на кластери абітурієнтів бакалаврських і магістерських освітніх програм ІТ спеціальностей)
- 2) Аналіз множинної регресії (перевірка статистичної значимості факторів впливу на рівень вступу абітурієнтів)

***Технології дослідження:***

- 1) Мова програмування R у фреймворку RStudio
- 2) Бібліотека Selenium

***Практичне значення:*** розроблена методика дослідження даних абітурієнтів дозволяє визначити основні характеристики абітурієнтів, які вступають на ІТ спеціальності, встановити, яка цільова аудиторія для вступу на бакалаврські і магістерські освітні програми ІТ спеціальностей і врахувати ці характеристики при визначенні умов вступу на новий навчальний рік.

***Апробація результатів дослідження:***

Апробація роботи полягає у підготовці і публікації статті «Програмні засоби виявлення кластерів абітурієнтів для відкритих даних» в альманасу «Магістерські студії» (2021 р.).

***Структура роботи*** складається зі вступу, трьох розділів, висновків, списку використаних джерел, додатків. У розділі 1 розглядаються інформаційні системи для вступу до ЗВО України та основні принципи вступу. У розділі 2 досліджуються методи відбору і кластеризації даних абітурієнтів. У розділі 3 підготовлено звіт щодо виконаної кластеризації, та запропоновані міри для покращення умов вступу на освітні програми ІТ спеціальностей ЗВО.

## РОЗДІЛ 1

# ІНФОРМАЦІЙНІ СИСТЕМИ ДЛЯ ВСТУПУ АБІТУРІЄНТІВ В УКРАЇНІ

### 1.1. Умови вступу до закладів вищої освіти

Цьогоріч дійсними є сертифікати 2018, 2019, 2020, 2021 років. Вступники можуть подавати сертифікати ЗНО, комбінуючи їх за власним вибором.

Від вступників до магістратури прийматимуться результати єдиного вступного іспиту 2020 та 2021 років, результати єдиного фахового вступного випробування – тільки 2021 року. У вступній кампанії 2021 року відомі коригувальні фактори залишаються незмінними: сільські, регіональні та галузеві.

Регіональний коефіцієнт поширюється на студентів, які вступають до регіональних закладів вищої освіти (ЗВО): Донецьк, Житомир, Кіровоград, Луганськ, Миколаїв, Рівне, Суми, Херсон, Хмельницький, Чернігів, Черкаси. Рівень вступу до ЗВО в цих регіонах становить 1,04, а для інших регіонів - 1,02. Конкурсна пропозиція щодо ЗВО міста Києва (незалежний структурний підрозділ ЗВО, структурний підрозділ ЗВО розташованих у населених пунктах вищих навчальних закладів) – 1,00.

Для спеціальностей (дисципліни, спеціальності), зазначених у переліку спеціальностей, які отримують підтримку, коефіцієнт відомства для заявок з пріоритетами 1 та 2 дорівнює 1,02; 1,00 - в інших випадках. Регіональні та галузеві коефіцієнти застосовуються автоматично і можуть бути накопичені без будь-яких дій абітурієнтами.

Коефіцієнт села збільшує бал конкурсу на 2% і може накопичуватися для тих, хто зареєструвався у селі та закінчив сільську школу в рік зарахування. Сільський фактор не поширюється на внутрішньо переміщених осіб.

В більшості випадків для вступу на ІТ–спеціальності заклади вищої освіти встановлюють конкретні предмети ЗНО які необхідно здати для вступу, такі як: математика, українська мова, фізика/англійська мова для ступеня вищої освіти (далі СВО) бакалавр та єдиний вступний іспит (ЄВІ) з іноземної мови, єдине фахове вступне випробування (ЕФВВ) для СВО магістр. Кожен університет також додатково формує коефіцієнти за якими буде розраховано середній бал з усіх предметів. Для прикладу візьмемо коефіцієнти за якими відбувався вступ до Херсонського Державного Університету в 2020 році на спеціальність 121 Інженерія програмного забезпечення для бакалаврського (табл. 1.1) та магістерського рівнів відповідно (табл. 1.2).

Таблиця 1.1.

## Коефіцієнти бакалаврського рівня вищої освіти

1
<b>ЗНО</b>
Українська мова і література ( $k=0.35$ , $\text{min}=100$ )
2
<b>ЗНО</b>
Математика ( $k=0.35$ , $\text{min}=100$ )
3
<b>ЗНО</b>
Іноземна мова ( $k=0.2$ , $\text{min}=100$ )
<b>ЗНО</b>
Фізика ( $k=0.2$ , $\text{min}=100$ )
Середній бал документа про освіту ( $k=0.1$ )
Бал за мотиваційний лист ( $k=0$ )

Таблиця 1.2.

## Коефіцієнти магістерського рівня вищої освіти

1
<b>ІСПИТ</b>
Іноземна мова (англійська, німецька, французька або іспанська) ( $k=1$ , $\text{min}=100$ )
2
<b>ФАХ</b>
Фахове випробування ( $k=1$ , $\text{min}=100$ )
<b>ЗАЛІК</b>

Додаткове вступне випробування  
(к=0)  
Інші показники конкурсного відбору  
(к=1)

На рисунку 1.1, показано процедуру та терміни вступу бакалаврів до ЗВО у 2021 р.



Рисунок 1.1 - Процедура вступу бакалаврів до ЗВО

## 1.2. Перелік інформаційних систем з відкритими даними абітурієнтів

Насамперед для отримання офіційних даних про абітурієнтів ЗВО необхідний доступ до Єдиної Державної Електронної Бази з питань Освіти (ЄДЕБО), що доступна за покликанням <https://vstup.edbo.gov.ua/>.

Від 13.07.2011 постановою Кабінету Міністрів України №752 створено Єдину державну електронну базу з питань освіти, в якій

визначено, що ЄДЕБО є автоматизованою системою збирання, реєстрації, оброблення, зберігання та захисту відомостей та даних з питань освіти. Власником ЄДЕБО є держава, розпорядником – Міністерство освіти і науки України, а технічним адміністратором – державне підприємство «Інфоресурс» [1].

Окрім ЄДЕБО існують допоміжні сторонні ресурси, такі як:

- Інформаційна система "Vstup.OSVITA.UA" - розташована за постійною адресою [vstup.osvita.ua](http://vstup.osvita.ua) та управляється редакцією освітнього ресурсу [Osvita.ua](http://Osvita.ua). Інформація, опублікована на веб-сайті інформаційної системи [Vstup.OSVITA.UA](http://Vstup.OSVITA.UA), отримана з ЄДЕБО відповідно до угоди, підписаної з державним підприємством [Inforesource](http://Inforesource) як менеджером ЄДЕБО. Інформаційна система [Vstup.OSVITA.UA](http://Vstup.OSVITA.UA) інформує громадськість про процес подання заяв про вступ, рекомендації щодо вступу та вступу до ЗВО під час вступної діяльності для забезпечення відкритості та прозорості вступу до ЗВО. Веб-сайт має на меті поінформувати абітурієнтів, їхніх батьків, експертів в галузі освіти та інших зацікавлених сторін про вимоги до вступу до ЗВО України та надати статистичну інформацію про результати вступу до ЗВО за попередні роки. Веб-сайт містить інформацію про всі існуючі в Україні ЗВО, які приймають абітурієнтів на молодших спеціалістів, бакалаврський і магістерський рівень вищої освіти. ЄДЕБО надає контактну інформацію, професійну інформацію та освітні програми для підготовки молодших експертів, бакалаврський і магістерський рівень вищої освіти для кожного ЗВО. Ця інформація містить відомість про конкурсні пропозиції, кількість ліцензій, кількість національних замовлень тощо [2].

- Веб-сайт ["abit-poisk.org.ua"](http://abit-poisk.org.ua) використовує інформацію, отриману з офіційних джерел ([Vstup.edbo.gov.ua](http://Vstup.edbo.gov.ua), [Vstup.osvita.ua](http://Vstup.osvita.ua), веб-сайти університетів тощо) [4].



В табл.1.3. наведено коротка порівняльна характеристика описаних систем, їх недоліки та переваги а також обмеження з точки зору парсингу даних.

Таблиця 1.3.

Порівняльна характеристика інформаційних систем

Назва ресурсу	Офіційна сторінка ЄДЕБО	Інформаційна система "Vstup.OSVITA.UA"	Веб-сайт "abit-poisk.org.ua"
Наявність вичерпуючої інформації	Є вся необхідна інформація	Інформація не повна	Інформація не повна
Використовують інформацію зі сторонніх ресурсів	Є офіційним ресурсом країни	Є договір з ЄДЕБО	Дані беруться з офіційних сайтів
Наявність зручного сортування	Є розширене сортування	Пошук за регіонами та освітнім ступенем	Пошук за регіонами
Обмеження на кількість запитів до сторінки	До ~50 запитів	До ~25 запитів	До ~25 запитів

Аналіз онлайн ресурсів показав, що більшість з них використовують дані, взяті з офіційного сайту ЄДЕБО. Отже, раціональним є використовувати саме цей ресурс для отримання даних.

### 1.3. Аналіз пошукових запитів ЄДЕБО

Основною метою ЄДЕБО є надання фізичним та юридичним особам інформації у сфері освіти. Наразі EDEBO пропонує:

(i) Електронне ліцензування освітньої діяльності у сферах вищої освіти, професійної вищої освіти та професійної (професійної) освіти (реєстрація, подання документів до ліцензійного агентства та отримання відповідних рішень в електронній формі);

(ii) Атестація освітніх програм та спеціальностей у порядку, встановленому законодавством;

(iii) Підтримує вступ до професійної освіти та вищої освіти, надаючи такий функціонал:

- Абітурієнти можуть подавати в електронному вигляді заявки на конкурсний відбір до закладів освіти через Інтернет (електронна реєстрація);
- Розробити список рейтингів вступників та список вступників, рекомендованих для вступу до закладів освіти;
- Забезпечити цільові державні та регіональні замовлення;
- Інформувати громадськість про процес зарахування до ЗВО;
- Формування, реєстрація та облік інформації, виданої суб'єктами освіти освітні документи (документи про освіту), а також видача ліцензій на освітню діяльність, освітню проекти та сертифікати, необхідні закладам освіти відповідно до законодавства;
- Можливість перевірки достовірності документів про освіту, кваліфікації (з 2014 року), вищу освіти (з 2012 р.);
- Інформувати громадськість про видачу та анулювання ліцензій на освітню діяльність закладів освіти, розширення або скорочення їх освітньої діяльності, атестації ЗВО;
- Формування статистичних звітів у сфері вищої освіти;
- Інші потреби у сфері освіти, передбачені законодавством.

На рисунку 1.2. показані критерії за якими можливо відсортувати пропозиції від університетів. Для пошуку необхідно вказати регіон закладу і спеціальності (спеціалізації), або вказати назву (код) закладу.

Після пошуку за спеціальністю та регіоном, відкриваються всі доступні пропозиції від ЗВО в обраному регіоні.

**ПОШУК КОНКУРСНИХ ПРОПОЗИЦІЙ**

**Пошук**

Введіть ключові слова для пошуку конкурсної пропозиції (освітній ступінь, спеціальність, спеціалізацію, регіон, назву закладу)  
 Наприклад: «бакалавр психологія київ», «магістр стоматологія Богомольця», «бакалавр китайська одеса»  
 Також можна вибрати відповідні значення з випадуючого списку

**РОЗШИРЕНИЙ ПОШУК**

Освітній рівень: Бакалавр

База вступу: Повна загальна середня освіта

Спеціальність / спеціалізація: Спеціальність

Освітня програма:

Форма навчання:

Курс зарахування:

Регіон: УКРАЇНА

Назва закладу / код в ЄДЕБО:

ОЦ «Крим-Україна», ОЦ «Донбас-Україна»

Рисунок 1.2 - Вікно вибору критеріїв для пошуку спеціальностей

#### 1.4. Відбір критеріїв для аналізу

Криза в країні, економічний стан конкретного регіону, нові правила вступу до ЗВО та інші чинники впливають на вибір абітурієнтами спеціальностей і ЗВО для вступу.

Сервіс ЄДЕБО містить дані, необхідні для подальшого аналізу. На рисунку 1.3. показано вікно доступної для парсингу інформації з основними критеріями для аналізу вступу абітурієнтів на перший (бакалаврський -Б) і другий (магістерський - М) рівень:

- Назва університету;
- Шифр спеціальності;

- Пріоритет;
- Ступінь вищої освіти (СВО);
- Макс. обсяг держзамовлення;
- Обсяг на контракт;
- Вартість навчання за рік;
- Коефіцієнт балу з української мови (Б);
- Коефіцієнт балу з історії (Б);
- Коефіцієнт балу з математики (Б);
- Коефіцієнт балу з іноземної мови (Б);
- Коефіцієнт балу з фізики (Б);
- Коефіцієнт балу з географії (Б);
- Коефіцієнт балу з біології (Б);
- Коефіцієнт середнього балу документу про освіту (Б);
- Фахове випробування (к=) (М);
- Коефіцієнт балу іноземної мови (ЄВІ) (М);
- Інші показники конкурсного відбору (к=) (М);
- Бал ЗНО з української мови (Б);
- Бал ЗНО з історії (Б);
- Бал ЗНО з математики (Б);
- Бал ЗНО з іноземної мови (Б);
- Бал ЗНО з фізики (Б);
- Бал ЗНО з географії (Б);
- Бал ЗНО з біології (Б);
- Середній бал документа про освіту (Б);
- Фахове випробування (М);
- Іноземна мова ЄВІ (М);
- Інші показники конкурсного відбору (М);
- Сільський коефіцієнт;

- Квота;
- конкурсний бал ДО НАКАЗУ (бюджет);
- ДО НАКАЗУ (контракт);
- Оригінали документів.

Освітня програма: Інженерія програмного забезпечення  
 Факультет: Комп'ютерних наук, фізики та математики  
 Форма навчання: Денна  
 Курс зарахування: 1 курс  
 Вид пропозиції: Відкрита, із вказанням пріоритетності

Термін навчання: 05.09.2020 – 30.06.2024 (Зр 10м)  
 Термін подачі заяв: 13.08.2020 – 22.08.2020  
 Ліцензійний обсяг: 50  
 Макс. обсяг держзамовлення: 9  
 Обсяг на контракт: 41  
 Регіональний коефіцієнт: 1.04  
 Вартість навчання за рік (контракт): 19320  
 Кількість заяв: 149, на бюджет – 112  
 Конкурсний бал: СЕР=158; МІН=118.2; МАКС=192

1 ЗНО Українська мова і література (k=0.35, min=100)  
 2 ЗНО Математика (k=0.35, min=100)  
 3 ЗНО Іноземна мова (k=0.2, min=100)  
 ЗНО Фізика (k=0.2, min=100)  
 Середній бал документа про освіту (k=0.1)  
 Бал за мотиваційний лист (k=0)

Список вступників Розподіл балів

УСЬОГО ПОДАНО ЗАЯВ: 149  
 У випадку некоректного відображення інформації або виникненні помилок спробуйте оновити кешовані сторінки за допомогою комбінації Ctrl+F5 або очистити кеш браузера.

Прізвище

ЗАВАНТАЖЕНО ЗАЯВ: 100  
 Завантажити додаткові заявки

#	Статус	ПР	КБ	Предмети / показники / коефіцієнти	ОД
1	ДО НАКАЗУ (Бюджет)	1	156.000	150 Українська мова і література 150 Математика 150 Іноземна мова 7 Середній бал документа про освіту Квота 2	+

Рисунок 1.3 - Вікно перегляду поданих заяв за спеціальністю.

## 1.5. Порівняння та вибір ресурсу для парсингу даних абітурієнтів

Python є найпопулярнішою мовою програмування при розробці синтаксичного аналізатора завдяки своєму простому синтаксису, швидкості навчання та множині доступних бібліотек, що полегшує

роботу розробникам. Далі розглянемо деякі бібліотеки та фреймворки синтаксичного аналізу, доступні для парсингу даних за допомогою Python:

**Scrapy** - це платформа для сканування та аналізу веб-сайтів, написаних на Python, для розробників, які використовують цю мову програмування. Scrapy вважається зрілою структурою, тому за замовчуванням у ньому є все необхідне для синтаксичного аналізу веб-сайту, включаючи модулі для надсилання HTTP-запитів та отримання даних із завантажених HTML-сторінок [5].

Вони безкоштовні у використанні та з відкритим кодом. Крім того, дані можна зберігати під час аналізу. Однак Scrapy не виконує код JavaScript і потрібно звертатися за допомогою до іншої бібліотеки. Для цього можна використовувати популярний інструмент автоматизації браузера Splash або Selenium.

**PySpider** - ще один інструмент розбору веб-сайтів, який можна використовувати для розробки сценаріїв Python. На відміну від Scrapy, цей інструмент може виконувати код JavaScript, тому не потрібно використовувати Selenium. Однак PySpider не є настільки функціональним, як Scrapy, тому що Scrapy розвивається з 2008 р. з кращою документацією та більшою кількістю користувачів. Але ці факти не занижують переваги PySpider. На противагу цьому, PySpider містить кілька унікальних функцій, таких як веб-інтерфейс з редактором сценаріїв [6].

**BeautifulSoup** - спрощує процес отримання даних з веб-сторінок. Бібліотека використовує аналізатори коду HTML та XML, щоб надати спеціальні методи доступу до даних Python. BeautifulSoup забезпечує легкі для аналізу функції, тому він став одним з найважливіших інструментів для аналізу веб-сайтів на ринку [7].

Насправді, більшість навчальних ресурсів з розбору веб-сайтів використовують BeautifulSoup, щоб показати початківцям, як писати

синтаксичний аналізатор. У порівнянні з використанням Scrapy або PySpider, використання цієї бібліотеки та бібліотеки запитів одночасно для надсилання HTTP - запитів полегшує розробку синтаксичного аналізатора.

**Selenium** - якщо цільовий веб-сайт використовує метод AJAX, тобто якщо він спирається на асинхронні запити, призначені для завантаження певних частин веб-сторінки за допомогою JavaScript, Scrapy, Requests і BeautifulSoup, це не допоможе. Якщо відвідується така веб-сторінка, потрібно скористатися засобом автоматизації браузера Selenium. Його можна використовувати для автоматизації браузерів, які підтримують режим без заголовка, таких як Chrome і Firefox[8].

Порівняльна характеристика засобів для парсингу узагальнена в табл. 1.4.

*Таблиця 1.4.*

Порівняння засобів для парсингу

Засіб	Переваги	Недоліки
Selenium	Дозволяє окрім усіх інших засобів, парсити сайти що використовують метод AJAX	Складність в використанні та обмеженість літератури.
BeautifulSoup	Легкі для аналізу функції, один з найважливіших ресурсів для аналізу веб-сайтів.	Необхідна бібліотека для відправки запиту на сайт, для аналізу завантажених даних потрібен зовнішній аналізатор.
PySpider	Може виконувати код JavaScript, веб-інтерфейс з редактором сценаріїв	Є доволі молодим методом, мало документації та користувачів
Scrapy	Безкоштовний, є все необхідне для синтаксичного аналізу веб-сайту, відкритий код.	Не виконує код JavaScript, є доволі застарілим.

Після аналізу можна заявити, що оптимальним для нашого типу завдання буде використовувати Selenium, так як він є найбільш сучасним та працює без підключення додаткових бібліотек.

Далі необхідно сформувати список посилань за якими буде робитись запит на сторінки для парсингу. Як приклад наведемо кілька посилань для Херсонських ЗВО:

<https://vstup2020.edbo.gov.ua/offer/722074/>

<https://vstup2020.edbo.gov.ua/offer/687149/>

<https://vstup2020.edbo.gov.ua/offer/715617/>

<https://vstup2020.edbo.gov.ua/offer/688749/>

<https://vstup2020.edbo.gov.ua/offer/803446/>

<https://vstup2020.edbo.gov.ua/offer/683328/>

На рис. 1.4 наведено приклад деяких отриманих даних.

Назва університету	Шифр спе	Пріоритет	СВО	Мабс. обс	Обсяг на €	Вартість	б коефіцієн	б коефіцієн	б коефіцієн	б коефіцієн
48	121	1	1	9	41	19320	0,35	0	0,35	0,2
48	121	1	1	9	41	19320	0,35	0	0,35	0,2
48	121	1	1	9	41	19320	0,35	0	0,35	0,2
48	121	1	1	9	41	19320	0,35	0	0,35	0,2
48	121	1	1	9	41	19320	0,35	0	0,35	0,2
48	121	1	1	9	41	19320	0,35	0	0,35	0,2
48	121	1	1	9	41	19320	0,35	0	0,35	0,2
48	121	2	1	9	41	19320	0,35	0	0,35	0,2
48	121	5	1	9	41	19320	0,35	0	0,35	0,2

Рисунок 1.4 - Результати парсингу

Усього було знайдено 839 абітурієнтів що подавали заявки до ЗВО на ІТ - спеціальності Херсонської області, серед яких 699 заяв на СВО бакалавр, та 140 заяв до СВО магістр відповідно (рис 1.4).



## РОЗДІЛ 2

### ПРОЄКТУВАННЯ ТА ПОБУДОВА АЛГОРИТМУ ДЛЯ КЛАСТЕРИЗАЦІЇ ДАНИХ

#### 2.1. Кластерний аналіз даних

Кластеризація полягає в автоматичному поділі елементів у наборі на групи на основі їх подібності. Елементами колекції можуть бути будь-які об'єкти, такі як дані або вектори. Саму групу також називають кластером.

Кластеризацію зазвичай порівнюють з класифікаціями. Відмінність між ними полягає в тому, що в класифікації дані чітко визначені багатьма групами результатів, тоді як при кластеризації вони визначаються алгоритмом під час його роботи. Кластеризація має багато практичних застосувань в різних сферах. Ось кілька прикладів кластерних програм:

- Аналіз даних;
- Спрощення інформаційної роботи;
- Візуалізація даних;
- Видобування та пошук інформації;
- Створити зручний класифікатор;
- Групування та розпізнавання об'єктів;
- Розпізнавання образів;
- Групування об'єктів.

Кластерний аналіз може бути виражений у такій послідовності операцій:

- 1) Виділення кількох об'єктів.
- 2) Визначення набору змінних для оцінки об'єктів та зображення векторів ознак.

3) Стандартизування власних векторів за допомогою доступного методу.

4) Визначення подібності між об'єктами за поданою метрикою.

5) Застосування вибраного методу аналізу кластерів, щоб розділити кілька об'єктів на кластери відповідно до ступеня подібності.

6) Презентування результатів аналізу.

Після попереднього аналізу результатів кластеризації відбувається налаштування вибраних параметрів, показників або методів кластеризації для покращення результатів. Ці вдосконалення спрямовані на покращення результатів кластеризації.

Класифікація алгоритмів кластеризації:

Плоский алгоритм розбиває заданий набір об'єктів на кластери і створює єдиний розділ. Щоб отримати інший розділ, потрібно повторити процес кластеризації з іншими параметрами.

У порівнянні з плоским алгоритмом, ієрархічний алгоритм створює не один розділ, а систему, яка ділить вкладені розділи на різні кластери. Як результат виконання цього алгоритму призведе до створення дерева розділів, коренем якого є кластер, що містить усі об'єкти, а листя - це менші кластери.

Існує також чітка або нечітка класифікація. Вона визначає, чи може зразок об'єкта належати до кількох кластерів одночасно або завжди належить до одного кластера [9].

У чіткому (або явному) алгоритмі кожен вибіркового об'єкт належить лише одному кластеру, тобто кожному об'єкту присвоюється унікальний номер кластера.

У нечіткому (або звичайному) алгоритмі кожен об'єкт порівнюється з набором істинних значень, які відображають ймовірність зв'язку між об'єктом і кожним кластером. Іншими словами, у нечіткому алгоритмі кожен об'єкт належить до всіх кластерів різного ступеня.

## 2.2. Вибір алгоритму для аналізу даних

**Алгоритм k-середніх (k-means).** Алгоритм k-mean конструює k кластерів, максимально віддалених один від одного. Основним типом проблеми, що вирішується алгоритмом k-mean, є наявність припущень (гіпотез) щодо кількості кластерів, і вони повинні бути максимально різними. Вибір числа k може ґрунтуватися на результатах попередніх досліджень, теоретичних міркуваннях або інтуїції. Загальна ідея алгоритму: полягає в пошуку таких центрів кластерів та наборів елементів кожного кластера при наявності деякої функції  $F(\circ)$ , що виражає якість поточного розбиття множини на k кластерів, коли сумарне квадратичне відхилення елементів кластерів від центрів цих кластерів буде найменшим [10]. Опис алгоритму:

Крок 1. Першочергове розподілення об'єктів за кластерами. Обирається число k. На першому кроці ці точки розглядаються як "центр" кластера. Кожен кластер відповідає центру. Вибір початкового центроїда може бути таким:

- вибір k спостережень, щоб збільшити початкову відстань;
- випадковий вибір k спостережень;
- вибір перших k спостережень.

Тож кожен об'єкт призначається до певного кластера.

Крок 2. Обчислюється центр кластера, а потім розглядається як середнє значення координат кластера. Об'єкт знову перепризначається новому кластеру з найближчим центром. Процес обчислення центру об'єкта та перерозподілу триває доти, доки не буде виконана одна з таких умов:

- центр кластера стабілізується, тобто усі спостереження належать до кластера, до якого вони належали при поточної ітерації;
- кількість ітерацій дорівнює максимальній кількості ітерацій.

Вибір кількості кластерів - складна задача. За відсутності припущень про кількість кластерів рекомендується спочатку створити 2 кластери, потім 3, 4, 5 тощо і порівняти результати, за яких умов відстань між кластерами збільшується.

**Самоорганізаційна карта Кохонена (SOM).** Опис алгоритму SOM як одношарової нейронної мережі, яка поширюється:

Крок 1. Підготовка даних для навчання. Навчальні вибірки мають бути репрезентативними, а не суперечливими, з трансформацією до чисельних значень та нормалізацією даних.

Крок 2. Ініціалізація початкової карти. На цьому етапі обирається кількість кластерів, та кількість нейронів у вихідному шарі відповідно. Перед вивченням карти потрібно ініціалізувати вагу мережевих нейронів. Існує два способи ініціалізації ваг:

- Ініціалізація за випадковими значеннями, коли всі ваги призначаються невеликим випадковим величинам.
- Через ініціалізацію екземпляра – випадкового значення з початкової вибірки, встановлюється на початкове значення.

Крок 3. Мережеве навчання. Карта Кохонена - це нейронна мережа, що складається з двох шарів: вхідного та вихідного (рис 2.1. зліва вхідні дані, справа вихідні).

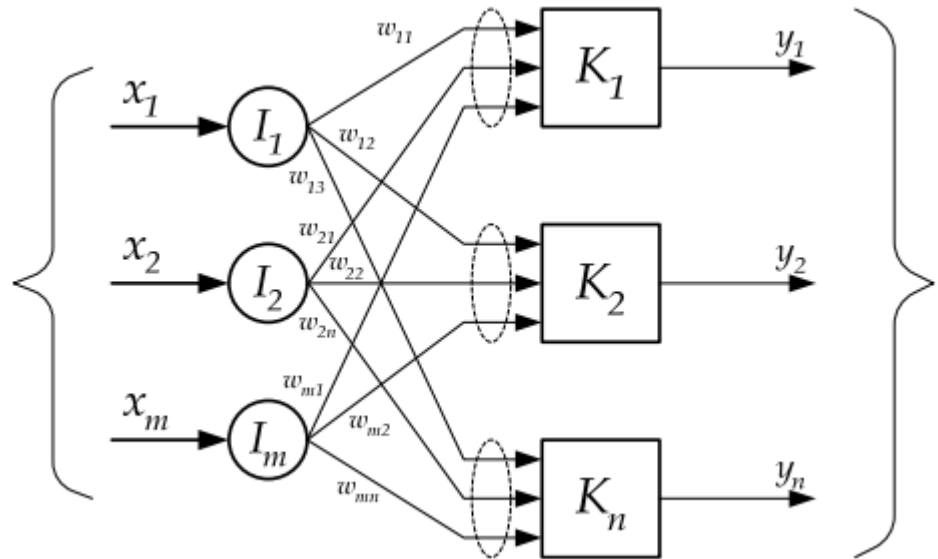


Рисунок 2.1 - Карта Кохонена

Навчання складається з серії нейронних векторних корекцій. На кожному етапі навчання вектор вибирається випадковим чином з вихідного набору даних, а потім здійснюється пошук найбільш подібного вектора коефіцієнта нейрона. У цьому випадку буде вибрано переможний нейрон, найбільш подібний до вхідного вектора. Схожістю цієї проблеми є відстань між вказівними величинами, яка зазвичай обчислюється в евклідовому просторі. Тому, якщо ми позначимо переможний нейрон як  $c$ , то отримаємо:  $\|x - w_c\| = \min_i \{\|x - w_i\|\}$ . Знайшовши переможний нейрон, йде відрегулювання ваги нейронної мережі. У цьому випадку вектор, що описує виграшний нейрон, і вектор, що описує його сусідів у сітці, рухаються у напрямку вхідного вектора. Щоб змінити вагу, застосовується наступна формула:

$$w_i(t + 1) = w_i(t) + h_{ci}(t) * [x(t) - w_i(t)],$$

де  $t$  - номер епохи, вектор  $x(t)$  випадковим чином вибирає ітерацію  $t$  з навчальної вибірки, а функція  $h(t)$  є функцією сусідства нейронів. Функція сусідства нейронів - це незростаюча функція часу і відстані між нейроном-переможцем і сусідніми нейронами в сітці. Функція ділиться на дві частини: функція відстані та швидкість навчання, що змінюються з часом.

$h(t) = h(\|r_c - r_i\|, t) * a(t)$ , де  $r$  - визначає розташування нейрона в сітці, а  $a(t)$  є функцією швидкості навчання мережі. Існує два типи функцій відстані:

Прості константи: 
$$h(d, t) = \begin{cases} const, d \leq q(t) \\ 0, d > q(t) \end{cases}$$

Гауссова функція: 
$$h(d, t) = e^{-\frac{d^2}{2 * q^2(t)}}$$

Функція швидкості навчання мережі:  $a(t) = \frac{A}{t+B}$ , де  $A$  і  $B$  – константи. Навчання складається з двох основних етапів: на початковому етапі обирається досить велика швидкість навчання та значення радіуса навчання, щоб дозволити нейронам розташовуватись відповідно до розподілу прикладів у зразку, а потім уточнюється шкала з коригуванням, коли параметр швидкості навчання значно менший. Якщо використовується лінійна ініціалізація, початковий етап регулювання можна пропустити [11].

Крок 4. Використання карти. При використанні карти вхідний вектор подається на вхід, а потім на виході активується нейрон або група нейронів, що відповідає кластерам, отриманим у процесі навчання мережі.

**Строгий алгоритм С середніх (Hard C-Means).** Обмеження: Обсяг даних невеликий. Переваги: простий у використанні; швидкість використання; чіткість і прозорість алгоритму. У порівнянні з k-mean алгоритм менш чутливий до викидів. Недоліки: Необхідно вказати кількість кластерів; повільно працює з великими базами даних. Опис алгоритму: Алгоритм бере як вхід множину  $S$  та кількість кластерів  $k$ . На виході алгоритм ділить множину  $S$  на  $k$  кластерів:  $S_1, S_2, \dots, S_k$ . Алгоритм подібний до алгоритму k-mean. Лише коли алгоритм спрацює, об'єкти будуть перерозподілені відносно медіани кластера замість його центру.

Строгий алгоритм C-Means (Hard C-Means; Jang, Sun і Mizutani, 1997) намагається визначити центри кластерів у багатовимірному просторі ознак. Мета - зіставити кожен точку простору об'єктів із відповідним кластером. Метою цього алгоритму є групування великої кількості числових даних, представлених описами об'єктів. Для цього встановлюється кількість точок і кластерів у скінченно-вимірному просторі ознак на вході алгоритму. Також можна встановити поріг, який використовується в критерії зупинки алгоритму: значення цільового функціоналу та швидкість зміни функціоналу під час виконання основного циклу алгоритму. Завданням цього алгоритму є усереднення відстані всередині кластера.

На першому кроці алгоритму ініціалізуються центри кластера. Це можна зробити, вибравши випадкові значення. Потім відбувається виконання основного циклу алгоритму: Розрахунок звичайної матриці, яка порівнює кожен точку простору ознак з найближчим центром кластера. В іншому випадку це можна розглядати як оптимізацію цільової функції на основі приналежності точки до кластера. У разі виконання циклу перевірте умови зупинки та виходу зі стандарту. Центр кластера перераховується як центроїд точки кластера. Вказується, що цільова функція оптимізована за координатами центру кластера. Цей алгоритм гарантує, що монотонність цільової функції не збільшується, тому вона сходиться до локального мінімуму. Цільова функція цього алгоритму повністю аналогічна функції K-Means [12].

**Нечіткий алгоритм C-середніх (Fuzzy C-means)** - Нечітка кластеризація, що дозволяє отримувати велику кількість числових даних, що в свою чергу дозволяє правильно визначати об'єкти на межах кластера. Однак реалізація цього алгоритму вимагає великої кількості обчислювальних ресурсів та початкового налаштування кількості кластерів. Крім того, об'єкти, розташовані далеко від центру всіх кластерів, можуть бути неоднозначними.

Алгоритм кластеризації нечітких С-засобів (FCM) був запропонований Дж. Данном у 1973 році та модифікований Дж. Бездеком у 1981 році. На відміну від більшості існуючих алгоритмів кластеризації, цей алгоритм є нечітким - кожен об'єкт не є однозначно включеним у будь-який кластер, а належить до всіх кластерів з різним ступенем належності. У випадку, коли кластери знаходяться близько один до одного і велика кількість точок розташована на їх межах, що надає перевагу для розподілу. Однак вартість цієї неоднозначності вище, ніж алгоритми розрахунків, такі як Hard C-Means та K-Means, при цьому зберігаються їх недоліки, такі як апріорне визначення кількості кластерів та відсутність гарантій глобальної оптимальності результатів [13].

Переваги та недоліки методів кластеризації зібрані в табл. 2.1.

Таблиця 2.1.

Переваги та недоліки описаних методів

Метод	Переваги	Недоліки
Алгоритм k-середніх (k-means).	Простота використання; швидкість використання; зрозумілість і прозорість алгоритму	Алгоритм занадто чутливий до викидів, які можуть спотворювати середнє; повільна робота на великих базах даних; необхідно ставити кількість кластерів; неможливість застосування алгоритму на даних, де є пересічні кластери
Самоорганізаційна карта Кохонена (SOM)	Використовується універсальний апроксиматор – нейронна мережа, навчання мережі без вчителя, самоорганізація мережі, простота	Робота тільки з числовими даними, мінімізація розмірів мережі, необхідно задавати кількість кластерів.



	реалізації, гарантоване отримання відповіді після проходження даних по шарах	
Строгий алгоритм С середніх (Hard С- Means)	Простий у використанні; швидкість використання; чіткість та прозорість алгоритму. У порівнянні з k-mean алгоритм менш чутливий до викидів.	Необхідно вказати кількість кластерів; повільно працювати над великими базами даних.
Нечіткий алгоритм С- середніх (Fuzzy С- means)	Нечіткість при визначенні об'єкта в кластер дозволяє визначати об'єкти, які знаходяться на кордоні, в кластери.	Обчислювальна складність, завдання кількості кластерів, виникає невизначеність з об'єктами, які вилучені від центрів всіх кластерів.

### 2.3. Застосування SOM за допомогою мови R та RStudio

Виберемо статистичну систему R для аналізу даних. Це вільно розповсюджуваний програмний продукт, який динамічно розробляється та використовується провідними науково - дослідними установами у різних наукових галузях, включаючи корпусне мовознавство. R - це система аналізу статистичних даних, створена Россом Іхакою та Робертом Джентлменом. Програма є одночасно мовою програмування та професійним програмним забезпеченням [14]. R вважається діалектом статистичної мови S, створеної експертом AT1T Bell Labs. Наразі підтримку та розробку нової версії надає спільна команда розробників з усього світу. R має ліцензію GPL GPL, тому будь -хто може використовувати цю програму у своїх дослідженнях. Єдине обмеження - це модифікація платформи. Будь -які зміни в коді продукту або

створення нових модулів на його основі також повинні поширюватися за тією ж ліцензією, що і сама програма.

Найважливіші переваги програмного пакета R включають:

- Ефективну обробку даних та прості методи зберігання результатів;

- Набір операторів для обробки масивів, матриць та векторів;

- Велику колекцію інтегрованих інструментів для статистичного аналізу;

- Використовуйте багатофункціональну бібліотеку для дослідження графічного дизайну даних;

- Ефективна інтерпретована мова програмування, що дозволяє самостійно розширювати функції програмного забезпечення та адаптувати його до конкретної ситуації завдання;

- відкритість та зручність використання програмного забезпечення та його модулів [14].

Серед недоліків, пов'язаних з програмою аналізу, слід зазначити наступні моменти:

- нестабільність нової версії R;

- всі запити потрібно виконувати в командному рядку;

- R-досить складний пакет програм для вивчення початківцями

Щоб спростити роботу R, зазвичай використовується додаткове спеціальне програмне забезпечення - R Studio. R Studio - це безкоштовне інтегроване середовище розробки, створене Джозефом Аллером [15], яке розповсюджується у двох версіях: R Studio Desktop (для персональних комп'ютерів та ноутбуків) та R Studio Server (серверна версія IDE). Ця програма полегшує написання коду в R, швидко встановлює та підключає додаткові модулі, проектує та зберігає статистичні графіки розподілу тощо. Поєднання функцій R та IDE R Studio допомагає спростити статистичний аналіз.

SOM в оболонці R:

Пакет `kohonen` реалізує різні форми самоорганізаційного відображення (SOM). Доступні алгоритми навчання онлайн та пакетно; пакетне навчання також можна проводити паралельно. До алгоритму навчання може бути представлено кілька шарів даних, і кожен рівень може мати різну метрику відстані. Загальна відстань - це середньозважене значення відстаней шарів. Шар можна вибрати за допомогою параметра `whatmap` або шляхом надання нульової ваги. Основна функція - `supersom`, просто використовує шар (класична форма) для обертання SOM. Ви можете використовувати функцію `map.kohonen` для відображення нових даних у навченому SOM. Функція `predict.kohonen` відображає дані в SOM і повертає прогнози тих шарів, які не знаходяться в новому об'єкті даних (тобто середнє значення одиниць - переможців). Кілька методів візуалізації представлені в графіку функції `Kohonen` [16]. В додатку А, наведено код який ми будемо використовувати для кластеризації даних абітурієнтів ЗВО Херсонського регіону.

Отже першим кроком згідно з алгоритмом роботи є зменшення середньої відстані до найближчих нейронів, для цього ми вкажимо приблизно 100 ітерацій для навчання SOM та задамо початкове значення параметру альфа. Результати роботи алгоритму можна побачити на рис. 2.2.

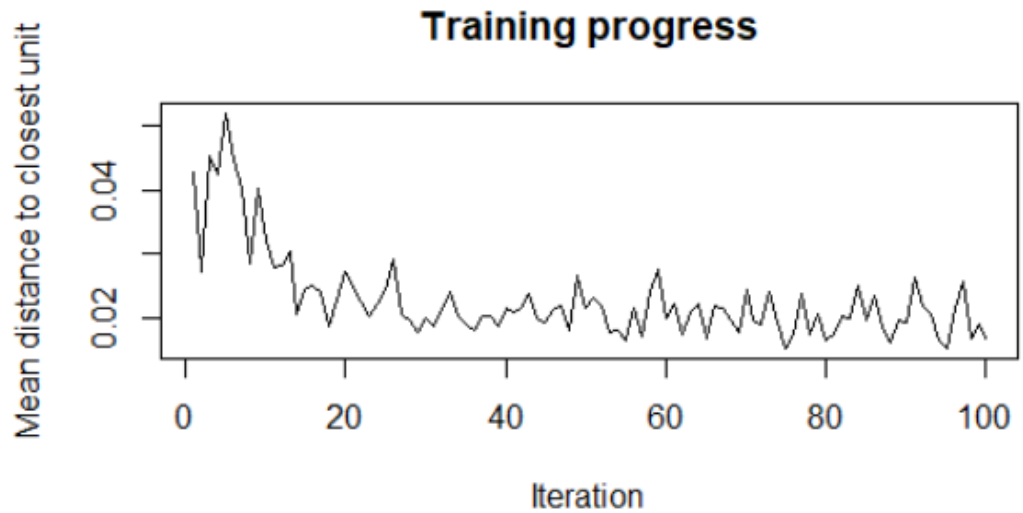


Рисунок 2.2 - Зменшення середньої відстані до найближчих нейронів

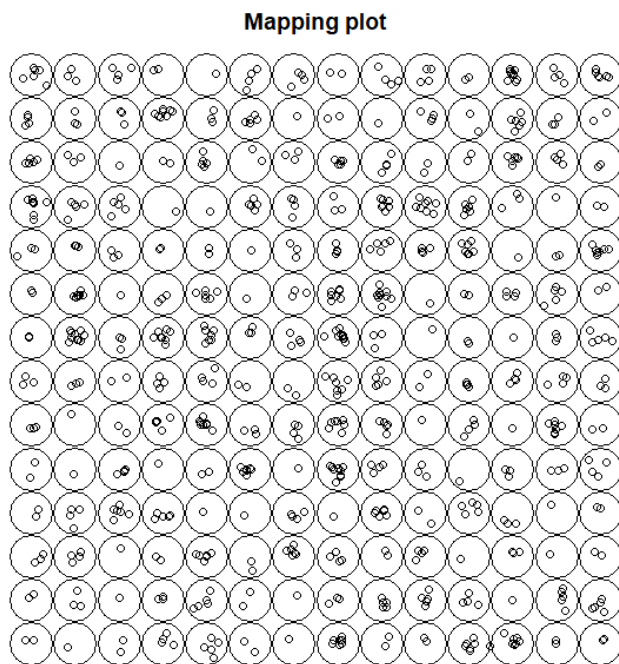


Рисунок 2.3 - Mapping plot

На рис. 2.3 представлено графік, на якому червоним кольором вказані кластери до яких потрапило менша кількість абітурієнтів, а білим і білим представлені кластери з найбільшою щільністю поданих заяв. Переконатись можна за допомогою графіку на рисунку 2.4, на якому в свою чергу представлено графік де чорними точками вказана

кількість абітурієнтів що потрапили в той чи інший кластер. Також необхідно звернути увагу наскільки близько до центру кластеру знаходиться точка. Чим ближче до центру тим більш точно можна сказати що заявка відноситься саме до цього кластеру і тим точніше будуть отримані результати з регресійного аналізу

## РОЗДІЛ 3

### АНАЛІЗ ОТРИМАНИХ ДАНИХ ТА ВПРОВАДЖЕННЯ ПРОПОЗИЦІЙ

#### 3.1. Аналіз результатів отриманих після кластеризації даних абітурієнтів бакалаврського рівня

Після проведення кластерного аналізу методом самоорганізаційних карт (SOM) було виявлено 6 (шість) основних кластерів абітурієнтів за першим (бакалаврським) рівнем. Ці кластери, в свою чергу, розподілені по кольорам (рис. 3.1), першому кластеру був присвоєний синій колір, другому кластеру помаранчевий колір, третьому – фіолетовий, четвертому - синій, п'ятому - червоний та шостому білий колір.

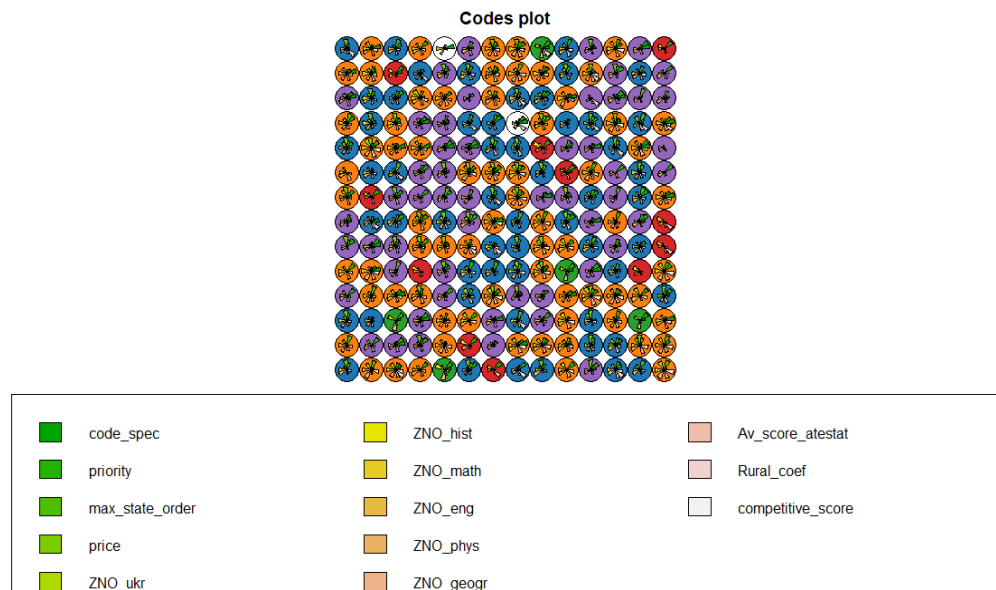


Рисунок 3.1 - Графік розбиття абітурієнтів за чотирнадцятьма критеріями на шість кластерів

Наступним завданням було віднести кожну заявку до кластеру та виділити їх відповідним кольором. Для цього ми скористались функцією «map\$unit.classif», яка показала нам, до якого саме кластеру потрапив

абітурієнт після виконання SOM алгоритму. Порахувавши та зробивши обчислення ми отримали, що:

- До першого кластеру потрапили 223 абітурієнти (31.9%)
- До другого 251 абітурієнт (35.9%));
- До третього 171 потенційний студент (24.4%);
- До четвертого 37 абітурієнта (5.3%);
- До п'ятого 14 абітурієнтів (2%);
- До шостого 4 абітурієнти (0.5%).

1	2	3	5	7	16	17	18	19	20	21	23	24	26	27	28	29	
edbo	code_spe	priority	max_state	price	ZNO_ukr	ZNO_hist	ZNO_matl	ZNO_eng	ZNO_phys	ZNO_geog	Av_score	Rural_coe	competiti	nakaz_bui	nakaz_cor	original_d	nodes
1	1	1	9	19320	150	0	150	150	0	0	7	0	156	1	0	1	1
1	1	1	9	19320	163	0	198	0	198	0	10,7	0	192,036	1	0	1	64
1	1	1	9	19320	188	0	180	185	0	0	10,7	0	191,88	1	0	1	27
1	1	1	9	19320	186	0	189	0	180	0	8,9	0	191,516	1	0	1	64
1	1	1	9	19320	173	0	179	185	0	0	10,4	1,02	189,459	1	0	1	162
1	1	1	9	19320	164	0	187	197	0	0	10,3	0	187,772	1	0	1	27
1	1	1	9	19320	170	0	180	0	180	0	7,1	0	180,544	1	0	1	64
1	1	2	9	19320	154	0	173	0	167	0	9,1	0	171,548	0	1	1	64
1	1	5	9	19320	138	0	186	0	164	0	10,6	0	171,392	0	1	1	190
1	1	2	9	19320	156	0	158	169	0	0	9,9	0	168,064	0	1	1	46

Рисунок 3.2 - Кольорове розбиття заяв бакалаврів згідно їх кластерам (фрагмент даних абітурієнтів)

Перший (синій) кластер (рис. 3.3):

		Av_score										priority					
b14	b13	b12	b11	b10	b9	b8	b7	b6	b5	b4	b3	b2	b1	const			
X14	X13	X12	X11	X10	X9	X8	X7	X6	X5	X4	X3	X2	X1				
0,02783	-0,07355	-0,05864	0	-0,0029	-0,00322	-0,0095	0	-0,00684	-5,9863E-05	-0,01142	-0,02435	0,005456	-0,03281	0,472197			
0,019244	0,067532	0,029327	0	0,004553	0,004571	0,007604	0	0,006236	0,00026319	0,007294	0,010709	0,043504	0,240516	4,860271			
0,212681	0,256498	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д			
4,013405	208	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д			
3,696642	13,68452	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д			
1,446108	1,08904	1,999662	#ДЕЛ/0!	0,637243	0,703616	1,249596	#ДЕЛ/0!	1,096613	0,227451945	1,565456	2,273545	0,125416	0,13643	0,097154			
tkr=	1,971435																

Рисунок 3.3 - Аналіз чинників за критерієм наказу до бюджету для першого кластеру бакалаврського рівня

- В кластері 1 зі зниженням абітурієнтом пріоритетності вступу на одиницю (X3) для бакалаврського рівня вищої освіти, знижується ймовірність вступу на бюджет на 2,4%
- В кластері 1 зі збільшенням середнього балу атестату абітурієнта (X12), ймовірність вступу на бюджет знижується на 5,9%

(оскільки збільшується ймовірність його/її вступу на інші бюджетні спеціальності з вищим пріоритетом)

nakaz_contract											max_statepriority				
											X4	X3			
0,005617	-0,01693	-0,02736	0	-0,00461	-0,00446	-0,00454	0	-0,00115	-0,00016864	0,019327	-0,08226	-0,03581	0,20966	4,208334	
0,025721	0,090259	0,039196	0	0,006085	0,006109	0,010163	0	0,008334	0,000351764	0,009748	0,014313	0,058145	0,32146	6,495953	
0,255288	0,34282	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	
5,09305	208	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	
8,379866	24,44525	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	
0,218387	0,18759	0,698045	#ДЕЛ/0!	0,757101	0,730164	0,446868	#ДЕЛ/0!	0,138023	0,479409467	1,982619	5,747593	0,615885	0,65221	0,647839	
tkr=	1,971435														

Рисунок 3.4 - Аналіз статистично значимих чинників за критерієм наказу до контракту для першого кластеру бакалаврського рівня

- В кластері 1 зі зниженням абітурієнтом пріоритетності вступу на одиницю (X3) для бакалаврського рівня вищої освіти, знижується ймовірність вступу на контракт на 8,2%
- В кластері 1 зі збільшенням середнього обсягу державного замовлення на одиницю (X4), ймовірність вступу на контракт зростає на 1,9% (абітурієнта приваблює можливість вступу на бюджет і у разі відсутності бюджетного міста, він охочіше вступає на контракт в повнокомплектну групу)

original_docs														
b14	b13	b12	b11	b10	b9	b8	b7	b6	b5	b4	b3	b2	b1	const
0,033447	-0,09048	-0,086	0	-0,00751	-0,00768	-0,01404	0	-0,00799	-0,0002285	0,007909	-0,10661	-0,03035	0,176846	4,680531
0,030223	0,106056	0,046056	0	0,007151	0,007178	0,011942	0	0,009793	0,000413327	0,011454	0,016818	0,068321	0,37772	7,632828
0,222163	0,402817	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
4,243442	208	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
9,639679	33,75046	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
1,106682	0,853105	1,867377	#ДЕЛ/0!	1,050105	1,069443	1,176	#ДЕЛ/0!	0,815743	0,55283565	0,690499	6,339215	0,444292	0,468193	0,613211
tkr=	1,971435													

Рисунок 3.5 - Аналіз статистично значимих чинників за критерієм подачі оригіналу документів для першого кластеру бакалаврського рівня

- В кластері 1 зі зниженням абітурієнтом пріоритетності вступу на одиницю (X3) для бакалаврського рівня вищої освіти, знижується ймовірність вступу на 10,7%

Другий (помаранчевий) кластер (рис. 3.6).



nakaz_budg															
Av_score												atestat		priority	
b14	b13	b12	b11	b10	b9	b8	b7	b6	b5	b4	b3	b2	b1	const	
X14	X13	X12	X11	X10	X9	X8	X7	X6	X5	X4	X3	X2	X1		
0,012138	0,014241	0,023346	0	-0,0016	-0,00161	-0,00433	0	-0,00385	-2,99677E-05	-0,00585	-0,08406	0,031195	-0,0255	0,295882	
0,016326	0,115026	0,02841	0	0,004163	0,004138	0,006318	0	0,005116	0,000110041	0,006884	0,011629	0,03403	0,097346	2,071817	
0,22968	0,294173	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	
5,004872	235	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	
6,063565	20,33644	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	
0,743469	0,123809	0,821759	#ДЕЛ/0!	0,384593	0,389239	0,685088	#ДЕЛ/0!	0,75303	0,27233166	0,850387	7,228743	0,91669	0,261942	0,142813	
tkr=	1,97011														

Рисунок 3.6 - Аналіз статистично значимих чинників за критерієм наказу до бюджету для другого кластеру бакалаврського рівня

- В кластері 2 зі зниженням абітурієнтом пріоритетності вступу на одиницю (X3) для бакалаврського рівня вищої освіти, знижується ймовірність вступу на бюджет на 8,4%

nakaz_contract															
max_state												priority		code_spec	
X4	X3	X2													
-0,01333	0,006942	0,021424	0	0,003919	0,003772	0,003568	0	0,001963	-3,32786E-05	-0,00889	-0,02708	-0,04072	0,040208	1,43567	
0,009325	0,065703	0,016228	0	0,002378	0,002364	0,003609	0	0,002922	6,28552E-05	0,003932	0,006642	0,019438	0,055604	1,183416	
0,143195	0,168031	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	
2,805341	235	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	
1,108902	6,635098	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	
1,429874	0,105653	1,320241	#ДЕЛ/0!	1,64814	1,595711	0,988651	#ДЕЛ/0!	0,671645	0,529449384	2,261538	4,077125	2,095018	0,72312	1,213157	
tkr=	1,97011														

Рисунок 3.7 - Аналіз статистично значимих чинників за критерієм наказу до контракту для другого кластеру бакалаврського рівня

- В кластері 2 для абітурієнтів, які подали заявки на контракт одночасно на декілька ІТ спеціальностей, при їх переході зі спеціальності 121 на 122 або зі 122 на 126 (X2) для бакалаврського рівня вищої освіти, їх ймовірність вступу на контракт знижується на 4,1%
- В кластері 2 зі зниженням абітурієнтом пріоритетності вступу на одиницю (X3) для бакалаврського рівня вищої освіти, знижується ймовірність вступу на контракт на 2,7%
- В кластері 2 зі збільшенням середнього обсягу державного замовлення на одиницю (X4), ймовірність вступу на контракт

зростає на 0,9% (абітурієнта приваблює можливість вступу на бюджет і у разі відсутності бюджетного міста, він охочіше вступає на контракт в повнокомплектну групу)

											max_state priority				
b14	b13	b12	b11	b10	b9	b8	b7	b6	b5	b4	b3	b2	b1	const	
-0,0012	0,021183	0,04477	0	0,002318	0,002161	-0,00076	0	-0,00189	-6,32463E-05	-0,01475	-0,11114	-0,00953	0,014709	1,731552	
0,017201	0,121186	0,029931	0	0,004385	0,00436	0,006656	0	0,00539	0,000115934	0,007252	0,012252	0,035853	0,102559	2,182771	
0,299499	0,309928	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	
7,176715	235	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	
9,651044	22,57296	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	
0,069546	0,174797	1,495773	#ДЕЛ/0!	0,528516	0,495681	0,114255	#ДЕЛ/0!	0,350611	0,54553615	2,033281	9,071758	0,265747	0,143421	0,793282	
tkr=	1,97011														

Рисунок 3.8 - Аналіз статистично значимих чинників за критерієм подачі оригіналів документів для другого кластеру бакалаврського рівня

- В кластері 2 зі зниженням абітурієнтом пріоритетності вступу на одиницю (X3) для бакалаврського рівня вищої освіти, знижується ймовірність вступу на 11,1%
- В кластері 2 зі збільшенням середнього обсягу державного замовлення на одиницю (X4), ймовірність вступу скорочується на 1,5% (при невеликих обсягах державного замовлення абітурієнт шукає альтернативи в інших ЗВО)

Наступним проаналізуємо третій (фіолетовий) кластер (рис. 3.9)

nakaz_budg															немає бюджету
b14	b13	b12	b11	b10	b9	b8	b7	b6	b5	b4	b3	b2	b1	const	
x14	x13	x12	x11	x10	x9	x8	x7	x6	x5	x4	x3	x2	x1		
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	0	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	
#ЧИСЛО!	156	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	
0	0	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	
#ДЕЛ/0!	#ДЕЛ/0!	#ДЕЛ/0!	#ДЕЛ/0!	#ДЕЛ/0!	#ДЕЛ/0!	#ДЕЛ/0!	#ДЕЛ/0!	#ДЕЛ/0!	#ДЕЛ/0!	#ДЕЛ/0!	#ДЕЛ/0!	#ДЕЛ/0!	#ДЕЛ/0!	#ДЕЛ/0!	
tkr=	1,975288														

Рисунок 3.9 - Аналіз статистично значимих чинників за критерієм наказу до бюджету для третього кластеру бакалаврського рівня

- В кластері 3 немає заявок на бюджетні місця (рис. 3.10)

	Rural_coef X13										max_statepriority	code_spec			
nakaz_contract											X4	X3	X2		
-0,01495	<b>0,519207</b>	-0,00847	0	0,004066	0,004849	0,006831	0	0,003221	2,79E-05	-0,01834	<b>-0,09729</b>	-0,02231	-0,02377	0,379627	
0,014093	0,227014	0,029258	0	0,004155	0,003961	0,00514	0	0,0044	0,000111	0,006096	0,010954	0,031286	0,089046	2,092324	
0,453872	0,263986	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	
9,260541	156	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	
9,034982	10,87145	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	
1,061048	<b>2,287112</b>	0,289453	#ДЕЛ/0!	0,978587	1,22408	1,329127	#ДЕЛ/0!	0,731864	0,252165	<b>3,007933</b>	<b>8,882128</b>	0,713055	0,266986	0,181438	
tkr=	1,975288														

Рисунок 3.10 - Аналіз статистично значимих чинників за критерієм наказу до контракту для третього кластеру бакалаврського рівня

- В кластері 3 зі зниженням абітурієнтом пріоритетності вступу на одиницю (X3) для бакалаврського рівня вищої освіти, знижується ймовірність вступу на контракт на 9,7%.
- В кластері 3 зі збільшенням середнього обсягу державного замовлення на одиницю (X4), ймовірність вступу на контракт скорочується на 1,8% (абітурієнтам важливий вступ насамперед на бюджет і якщо не на дану спеціальність, то на іншу).
- В кластері 3 для абітурієнтів, які подали заявки на контракт з сільським коефіцієнтом (X13) для бакалаврського рівня вищої освіти, їх ймовірність вступу на контракт збільшується на 51,9%.

											max_statepriority			
b14	<b>b13</b>	b12	b11	b10	b9	b8	b7	b6	b5	<b>b4</b>	<b>b3</b>	b2	b1	const
-0,01495	<b>0,519207</b>	-0,00847	0	0,004066	0,004849	0,006831	0	0,003221	2,79E-05	-0,01834	<b>-0,09729</b>	-0,02231	-0,02377	0,379627
0,014093	0,227014	0,029258	0	0,004155	0,003961	0,00514	0	0,0044	0,000111	0,006096	0,010954	0,031286	0,089046	2,092324
0,453872	0,263986	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
9,260541	156	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
9,034982	10,87145	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
1,061048	<b>2,287112</b>	0,289453	#ДЕЛ/0!	0,978587	1,22408	1,329127	#ДЕЛ/0!	0,731864	0,252165	<b>3,007933</b>	<b>8,882128</b>	0,713055	0,266986	0,181438
tkr=	1,975288													

Рисунок 3.11 - Аналіз статистично значимих чинників за критерієм подачі оригіналів документів для третього кластеру бакалаврського рівня

- Результати повністю співпадають із контрактними студентами через відсутність у даному кластері студентів, які вступили на бюджет.

Четвертий (червоний) кластер представлений на рис. 3.12.

nakaz_budg	немає бюджету														
b14	b13	b12	b11	b10	b9	b8	b7	b6	b5	b4	b3	b2	b1	const	
X14	X13	X12	X11	X10	X9	X8	X7	X6	X5	X4	X3	X2	X1		
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	0	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	
#ЧИСЛО!	22	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	
0	0	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	
#ДЕЛ/0!	#ДЕЛ/0!	#ДЕЛ/0!	#ДЕЛ/0!	#ДЕЛ/0!	#ДЕЛ/0!	#ДЕЛ/0!	#ДЕЛ/0!	#ДЕЛ/0!	#ДЕЛ/0!	#ДЕЛ/0!	#ДЕЛ/0!	#ДЕЛ/0!	#ДЕЛ/0!	#ДЕЛ/0!	
tkr=	2,073873068														

Рисунок 3.12 - Аналіз Статистично значимих чинників за критерієм наказу до бюджету для четвертого кластеру бакалаврського рівня

- В кластері 4 немає вступників на бюджет. Представлені студенти з низькими вступними балами, які не можуть претендувати на бюджет на ІТ спеціальності. Це абітурієнти, які вступили на 121 спеціальність до інших ЗВО (не ХДУ)
- В кластері 4 немає статистично значимих факторів, які впливають на вступ контрактників
- Дані для бюджету повністю співпадають із контрактними студентами через відсутність у даному кластері студентів, які вступили на бюджет

П'ятий (зелений) та шостий (білий) кластери включають в себе невелику кількість заяв, а саме 14 та 4 відповідно.

До п'ятого кластеру потрапили студенти без ЗНО з математики, які не можуть претендувати на вступ на ІТ спеціальності, але планували вступ на спеціальності 122 і 126 (тобто мають ЗНО з історії чи географії) в інші ЗВО (не ХДУ).

В шостий кластер потрапили абітурієнти, які здали ЗНО з фізики на низькі бали і мають в цілому низький конкурсний бал. Вступив в

ХДУ на бюджет лише абітурієнт із першим пріоритетом. Точно оцінити у кластері вплив факторів неможливо, оскільки кількість спостережень (4) менша, ніж кількість параметрів (14), тому регресійний аналіз не застосовується.

### **3.2. Аналіз даних вступників на магістерський рівень шляхом кластеризації заяв на ІТ-спеціальності до Херсонських ЗВО**

Після проведення кластерного аналізу методом самоорганізаційних карт було розподілено абітурієнтів на 6 (шість) основних кластерів, які розподілені за кольорами:

- Перший кластер - синій
- Другий – зелений
- Третій – помаранчевий
- Четвертий – фіолетовий
- П'ятий – червоний
- Шостий - білий

Обчисливши кількість абітурієнтів, що потрапили до кожного з кластерів (рис. 3.13), та провівши їх аналіз ми отримали що:

- До першого кластеру потрапили 27 абітурієнтів (19.2 %);
- До другого 72 абітурієнти (51.4%);
- До третього 18 абітурієнтів (12.9 %);
- До четвертого 18 абітурієнтів (12.9 %);
- До п'ятого 1 абітурієнт (0.7 %);
- До шостого 4 абітурієнти (2.9%).

## Codes plot

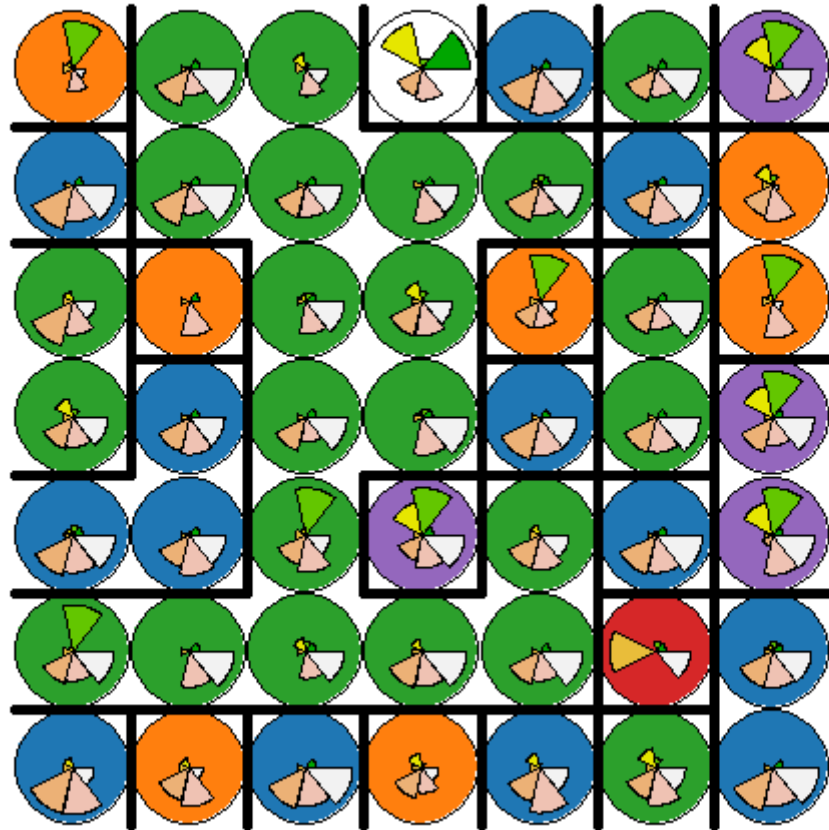


Рисунок 3.13 - Графік розбиття вступників до магістратури на 6 кластерів за сімома критеріями

edbo	code_spec	license_vol	price	Professional_test_score	EVI_score	Other_sco	nakaz_bu	nakaz_cor	original_d	Number	Nodes
48	121	4	24560	200	196	8,5	1	0	1	1	1
48	121	4	24560	174	169	6	1	0	1	2	2
48	121	4	24560	195	141	9	1	0	1	3	29
48	121	4	24560	179	144	17,5	1	0	1	4	11
48	121	4	24560	166	169	5,5	1	0	1	5	2
48	121	4	24560	134	164	5,5	1	0	1	6	2
48	121	4	24560	147	189	11,5	0	0	0	7	5
48	121	4	24560	150	147	12	0	0	0	8	19
48	121	4	24560	139	158	11,5	0	0	0	9	19
48	121	4	24560	155	141	11	0	0	0	10	19
48	121	4	24560	137	153	2,5	0	0	0	11	2
48	121	4	24560	110	158	17	0	0	0	12	11

Рисунок 3.14 - Кольоровий розподіл магістерських заяв поданих на ІТ спеціальності до Херсонських ЗВО

Перший (синій) кластер (табл. 3.1):



-0,01164	-0,00054	<b>0,001968</b>	-0,00016	-0,03352	0,149803	0	4,085618
0,023885	0,003435	0,000935	0,000319	0,047773	0,089163	0	7,665011
<b>0,127916</b>	0,472087	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
1,589012	65	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
2,124821	14,48629	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
0,4875	0,158595	<b>2,103756</b>	0,499739	0,701645	1,680097	#ДЕЛ/0!	0,533022
2,093024							

Третій (помаранчевий) кластер (табл. 3.3):

- В помаранчевому кластері 3 кожен додатковий бал, отриманий абітурієнтом за фазовий іспит (X5), підвищує на 0,7% ймовірність вступу на бюджет
- В помаранчевому кластері 3 кожен додатковий бал, отриманий абітурієнтом за фазовий іспит (X5), підвищує на 1% ймовірність вступу на бюджет і контракт
- В помаранчевому кластері 3 при переході з 121 на 122 або зі 122 на 126 (X2), ймовірність вступу на бюджет і контракт підвищується на 48,5%

Таблиця 3.3

Аналіз статистично значимих чинників за критерієм подачі оригіналів документів для третього кластеру магістерського рівня

original_docs							
-0,06458	0,003734	<b>0,01014</b>	-0,00084	-0,03067	<b>0,48503</b>	0	18,99306
0,062134	0,004116	0,002815	0,000813	0,064116	0,177774	0	19,25386
<b>0,584795</b>	0,401832	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
2,582154	11	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
2,501622	1,776156	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
1,039355	0,907053	<b>3,602462</b>	1,030782	0,478275	<b>2,728353</b>	#ДЕЛ/0!	0,986455
2,093024							

Четвертий (фіолетовий) кластер (табл. 3.4):

- В фіолетовому кластері 4 кожен додатковий бал, отриманий абітурієнтом за фазовий іспит (X5), підвищує на 0,5% ймовірність вступу на бюджет



- В фіолетовому кластері 4 кожен інші бали (середній бал диплому бакалавра), отримані абітурієнтом (X7), знижують на 8,7% ймовірність вступу на бюджет
- В фіолетовому кластері 4 кожен додатковий бал, отриманий абітурієнтом за фазовий іспит (X5), підвищує на 0,44% ймовірність вступу на бюджет і контракт
- В фіолетовому кластері 4 кожен інші бали (середній бал диплому бакалавра), отримані абітурієнтом (X7), знижують на 9% ймовірність вступу на бюджет і контракт

Таблиця 3.4

Аналіз статистично значимих чинників за критерієм подачі оригіналів документів для четвертого кластеру магістерського рівня

original_docs							
-0,09104	-0,00484	0,00444	0	0	0	0	2,11737
0,040112	0,003983	0,001202	0	0	0	0	0,869204
<b>0,785706</b>	0,302771	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
5,237831	10	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
3,361075	0,916703	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
<b>2,269583</b>	1,215928	<b>3,694839</b>	#ДЕЛ/0!	#ДЕЛ/0!	#ДЕЛ/0!	#ДЕЛ/0!	2,435987
2,093024							

### 3.3. Розробка та впровадження пропозицій щодо покращення умов вступу абітурієнтів до ЗВО

Побудовані кластери абітурієнтів показали ефективність в аналізі інформації по вступній кампанії закладів вищої освіти Херсонського регіону. Аналіз є дуже корисним насамперед для адміністрації ЗВО на рівні університету, факультету і кафедр, оскільки допомагає зробити висновки про поточний стан показників ефективності діяльності ЗВО, визначити, який кластер абітурієнтів віддає перевагу саме їх закладу та на які критерії необхідно звернути увагу для подальшого розвитку контингенту абітурієнтів. Представлені результати статистичної обробки показників можуть бути корисні завідувачам кафедр при

проведенні самоаналізу, включаючи оцінку поточного стану за цими показниками серед інших ЗВО в різних регіонах.

Таблиця 3.5

Звіт по значимим критеріям для чотирьох основних кластерів бакалаврського рівня

№ кластеру	Перший	Другий	Третій	Четвертий
Залежні змінні				
Рекомендовано до наказу на бюджет	Середній бал атестату та пріоритет	Пріоритет	Немає вступників на держ. замовлення	Немає держ. Замовлення
Рекомендовано до наказу на контракт	Пріоритет та кількість держ замовлення	Спеціальність, пріоритет та обсяг держ. замовлення	Сільський коефіцієнт, пріоритет та обсяг держ. замовлення	Немає
Оригінали документів	Пріоритет	Пріоритет та обсяг держ. замовлення	Результати ідентичні до наказу на контракт	Немає

В таблиці 3.5 наведено приклад основних значимих критеріїв, що впливали на кластери з найбільшою кількістю поданих заяв для вступу на бакалаврський рівень. Як бачимо, основним чинником, що впливав на рівень вступу абітурієнтів, стала пріоритетність заявки, можна зробити висновок що навіть при середніх балах за ЗНО можливо вступити до ЗВО, якщо правильно розподілити пріоритети заяв.

Таблиця 3.6

Звіт по значимим критеріям для чотирьох основних кластерів магістерського рівня

№ кластеру	Перший	Другий	Третій	Четвертий
Залежні змінні				

Рекомендовано до наказу на бюджет	Фаховий іспит	Немає	Фаховий іспит	Фаховий іспит, середній бал диплому
Рекомендовано до наказу на контракт	Мала кількість контрактних заяв	Немає	Мала кількість контрактних заяв	Мала кількість контрактних заяв
Оригінали документів	Фаховий іспит, зміна спеціальності з 121 на 122	Фаховий іспит	Фаховий іспит, зміна спеціальності з 121 на 122 та зі 122 на 126	Фаховий іспит, середній бал диплому

В таблиці 3.6 наведено приклад основних значимих критеріїв, що впливали на кластери з найбільшою кількістю поданих заяв на магістерському рівні. Як видно, основним чинником що впливав на вступ в основних трьох кластерах, стали результати здачі фахового іспиту, можна зробити висновки, що студентам слід більше приділяти увагу саме цьому параметру та сконцентруватися на успішній здачі іспиту за фахом. Університетам же, в свою чергу, для збільшення попиту на спеціальності необхідно створити додаткові курси для підготовки вступників до магістратури за конкретними спеціальностями. Також можна відмітити низький рівень вступників на контрактну форму навчання, що можна пояснити високою ціною та небажанням студентів продовжувати навчання

## ВИСНОВКИ

При виконанні дослідження було визначено та оброблено 839 заяв абітурієнтів, що подавалися на ІТ-спеціальності 121 Інженерія програмного забезпечення, 122 Комп'ютерні науки, 126 Інформаційні системи та технології в 2020 році до ЗВО Херсону. Виходячи з отриманих даних, які були нами розділили на шість кластерів за допомогою SOM кластеризації, було з'ясовано, що більшість студентів віддає перевагу університетам та спеціальностям з високим рівнем державного замовлення.

Дослідження за допомогою кластеризації для бакалаврського рівня показує, що навіть із середніми балами при правильному розподіленні пріоритетів у студента є високі шанси вступу на бюджетну форму навчання.

Для вступників до магістратури основним чинником, що впливав на успішний вступ, став результат єдиного фахового вступного випробування. Отже, можна зробити висновок, що студентам необхідно приділити більше уваги саме успішній здачі іспиту за спеціальністю. Для університетів, в свою чергу, необхідно створити спеціальні курси з підготовки абітурієнтів до вступу на конкретні спеціальності та забезпечити необхідні умови для отримання державного замовлення на ІТ спеціальності.

Результати цього дослідження сприятимуть прийняттю необхідних ефективних управлінських рішень керівництва ЗВО для визначення таких умов вступу, що підвищать конкурентоспроможність освітніх програм університетів та забезпечать проведення заходів, таких як можливість проходження абітурієнтами додаткових професійних курсів на базі ЗВО. Крім того, отримані результати є аналітикою для адміністративного персоналу факультету і кафедр для моніторингу

вступу абітурієнтів на освітні програми бакалаврського і магістерського рівнів на наступний навчальний рік.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Офіційний сайт ЄДЕБО [Електронний ресурс] – Режим доступу до ресурсу: <https://info.edbo.gov.ua/>.
2. Вступ.ОСВІТА.UA [Електронний ресурс] – Режим доступу до ресурсу: <https://vstup.osvita.ua/>
3. Wehrens R. Self- and Super-organizing Maps in R: The kohonen Package / R. Wehrens, L. Buysdens. // Journal of Statistical Software. – 2007. – №5.
4. Сайт абіт-пошук [Електронний ресурс] – Режим доступу до ресурсу: <https://abit-poisk.org.ua/terms>.
5. Scrapy [Електронний ресурс] – Режим доступу до ресурсу: <https://ru.wikipedia.org/wiki/Scrapy>.
6. Spyder [Електронний ресурс] – Режим доступу до ресурсу: [https://ru.wikipedia.org/wiki/Spyder\\_\(IDE\)](https://ru.wikipedia.org/wiki/Spyder_(IDE))
7. Beautiful Soup [Електронний ресурс] – Режим доступу до ресурсу: <https://www.crummy.com/software/BeautifulSoup/bs4/doc.ru/bs4ru.html>.
8. Selenium для Python [Електронний ресурс] – Режим доступу до ресурсу: <https://habr.com/ru/post/248559/>.
9. Бериков, В. С. Современные тенденции в кластерном анализе / В. С. Бериков, Г. С. Лбов. — Всероссийский конкурсный отбор обзорноаналитических статей по приоритетному направлению "Информационно-телекоммуникационные системы", —2008. —26 с.
10. Кластеризация методом k-средних [Електронний ресурс] – Режим доступу до ресурсу: [https://ru.wikipedia.org/wiki/%D0%9C%D0%B5%D1%82%D0%BE%D0%B4\\_k%D1%81%D1%80%D0%B5%D0%B4%D0%BD%D0%B8%D1%85](https://ru.wikipedia.org/wiki/%D0%9C%D0%B5%D1%82%D0%BE%D0%B4_k%D1%81%D1%80%D0%B5%D0%B4%D0%BD%D0%B8%D1%85)
11. Самоорганизующиеся карты Кохонена [Електронний ресурс] – Режим доступу до ресурсу: <https://ranalytics.github.io/data-mining/105-Cohonen-Maps.html>.
12. Шишков И. С. Алгоритм HCM (Hard C – Means) [Електронний ресурс] / Илья Сергеевич Шишков // AlgoWiki. – 1912. – Режим доступу до ресурсу: [https://algowiki-project.org/ru/%D0%A3%D1%87%D0%B0%D1%81%D1%82%D0%BD%D0%B8%D0%BA:%D0%A8%D0%B8%D1%88%D0%BA%D0%BE%D0%B2%D0%98%D0%BB%D1%8C%D1%8F%D0%A1%D0%B5%D1%80%D0%B3%D0%B5%D0%B2%D0%B8%D1%87/%D0%90%D0%BB%D0%B3%D0%BE%D1%80%D0%B8%D1%82%D0%BC\\_HCM\\_\(Hard\\_C\\_%E2%80%93\\_Means\)](https://algowiki-project.org/ru/%D0%A3%D1%87%D0%B0%D1%81%D1%82%D0%BD%D0%B8%D0%BA:%D0%A8%D0%B8%D1%88%D0%BA%D0%BE%D0%B2%D0%98%D0%BB%D1%8C%D1%8F%D0%A1%D0%B5%D1%80%D0%B3%D0%B5%D0%B2%D0%B8%D1%87/%D0%90%D0%BB%D0%B3%D0%BE%D1%80%D0%B8%D1%82%D0%BC_HCM_(Hard_C_%E2%80%93_Means)).
13. Муромцев Н. А. Нечеткий алгоритм C-средних (Fuzzy C-means) [Електронний ресурс] / Н. А. Муромцев // AlgoWiki – Режим доступу до ресурсу: <https://algowiki-project.org/ru/%D0%A3%D1%87%D0%B0%D1%81%D1%82%D0%BD>

[%D0%B8%D0%BA:Nikmedoed/%D0%9D%D0%B5%D1%87%D0%B5%D1%82%D0%BA%D0%B8%D0%B9 %D0%B0%D0%BB%D0%B3%D0%BE%D1%80%D0%B8%D1%82%D0%BC %D0%A1-%D1%81%D1%80%D0%B5%D0%B4%D0%BD%D0%B8%D1%85\\_ \(Fuzzy C-means\).](https://www.r-project.org/about.html)

14. "What is R? " [Электронный ресурс]. Доступно: <https://www.r-project.org/about.html>.
15. "Why RStudio? " [Электронный ресурс]. Доступно: <https://www.rstudio.com/about/>
16. Анисимова Э.С. Самоорганизующиеся карты Кохонена в задачах кластеризации [Электронный ресурс] / Анисимова Э.С – Режим доступа до ресурсу: <https://cyberleninka.ru/article/n/samoorganizuyuschiesya-karty-kohonena-v-zadachah-klasterizatsii/viewer>.
17. Wehrens R. Package 'kohonen' [Электронный ресурс] / R. Wehrens, K. Johannes // Supervised and Unsupervised Self-Organising Maps. – 2019. – Режим доступа до ресурсу: <https://cran.r-project.org/web/packages/kohonen/kohonen.pdf>.
18. Самоорганизация Карта Кохонена [Электронный ресурс] – Режим доступа до ресурсу: <https://artsandculture.google.com/entity/m0k6rk?hl=uk>
19. Kobets, V., Yatsenko, V., Voynarenko, M. (2020). Cluster Analysis of Countries Inequality Due to IT Development Through Macros Application. Communications in Computer and Information Science. Springer. 1175, 415–439. doi: 10.1007/978-3-030-39459-2\_19.
20. The Organization of Behavior – New York: Psychology Press, 2002. – 378 с. – (Behavioral Sciences;).
21. Kobets, V., Yatsenko, V., Voynarenko, M. (2019). Cluster analysis of countries inequality due to IT development. CEUR-WS. 2393, 406–421.
22. Helsinki University of Technology – CIS Laboratory (2007). "Bibliography of SOM Papers." URL <http://www.cis.hut.fi/research/som-bibl/>.
23. Kohonen T (2001). Self-Organizing Maps. Number 30 in Springer Series in Information Sciences. Springer-Verlag, Berlin, 3 edition.
24. Методы и средства анализа данных [Электронный ресурс]: Режим доступа: <http://bourabai.kz/tpoi/analysis6.htm>
25. Хайкин С. Нейронные сети: полный курс. — М.: Вильямс, 2006. — 1104с.
26. Паклин Н. Б., Орешков В. И. Бизнес-аналитика: от данных к знаниям.— СПб.: Питер, 2013. — 704 с
27. Руководство по изучению языка R и его использование в Data Science [Электронный ресурс] – Режим доступа до ресурсу: <https://proglib.io/p/data-science-with-r>.

28. Ключников М. В. Технология кластерного анализа финансовых показателей банков / М. В. Ключников., 2006. – (Прикладная информатика. Научные статьи).
29. Duran B. S. Cluster Analysis / B. S. Duran, P. L. Odell., 1974.
30. Python: Scrapy, Selenium, BeautifulSoup что лучше для парсинга веб сайтов [Электронный ресурс] – Режим доступа до ресурсу: <https://dev-gang.ru/article/python-scrapy-selenium-beautiful-soup-czto-luczshe-dlja-parsinga-veb-saitov-rmv3n1q3us/>.
31. Кластерный анализ [Электронный ресурс] – Режим доступа до ресурсу: [https://ranalytics.github.io/data-mining/101-Partitioning-Algos.html#sec\\_10\\_1](https://ranalytics.github.io/data-mining/101-Partitioning-Algos.html#sec_10_1).
32. Лучшие инструменты для самостоятельного парсинга веб-сайтов — исчерпывающий список [Электронный ресурс] – Режим доступа до ресурсу: <https://vc.ru/services/249048-luchshie-instrumenty-dlya-samostoyatel'nogo-parsinga-veb-saytov-ischerpyvayushchiy-spisok>.
33. Парсинг веб-данных с помощью Python и Selenium [Электронный ресурс] – Режим доступа до ресурсу: <https://ichi.pro/ru/parsing-veb-dannyh-s-pomos-u-python-i-selenium-134271702427202>.
34. Нейский И. М. Классификация и сравнение методов кластеризации [Электронный ресурс] / И. М. Нейский – Режим доступа до ресурсу: [http://it-claim.ru/Persons/Neyskiy/Article2\\_Neiskiy.pdf](http://it-claim.ru/Persons/Neyskiy/Article2_Neiskiy.pdf).
35. Tian Zhang. «BIRCH: An Efficient Data Clustering Method for Very Large Databases / Tian Zhang..
36. Герман-Галкін С.Г. Комп'ютерне моделювання систем в MATLAB/ С.Г. Герман-Галкін. - СПб .: КОРОНА принт, 2001. - 320 с.
37. Обзор алгоритмов кластеризации данных [Электронный ресурс] – Режим доступа до ресурсу: <https://habr.com/ru/post/101338/>.
38. Методы кластерного анализа. Иерархические методы [Электронный ресурс] – Режим доступа до ресурсу: <https://intuit.ru/studies/courses/6/6/lecture/182?page=1>.
39. Ярушкіна Н.Г. Інтелектуальний аналіз часових рядів / Н.Г. Ярушкіна, Т.В. Афанасьєва. - Ульяновськ: УлГТУ, 2010. - 320 с.
40. Кластерний аналіз [Электронный ресурс] – Режим доступа до ресурсу: [https://pidru4niki.com/11800912/ekonomika/klasterniy\\_analiz](https://pidru4niki.com/11800912/ekonomika/klasterniy_analiz).



## ДОДАТКИ

### Додаток А

Кластеризація методом SOM для даних вступників до магістратури на прикладі Херсонських ЗВО

```

library("kohonen")
library("dplyr")

#Data
data <- read.csv("mag2.txt", header = TRUE, sep="\t", dec=",")

head(data)
tail(data)
str(data)
glimpse(data)

X<-data[,-1]
glimpse(X)
X <- scale(data[,-1])
summary(X)

#SOM
set.seed(135) # генеруємо випадкові числа для початкових ваг кожного
вузла
g<-somgrid(xdim = 7, ydim = 7, topo="rectangular") # вводим 7 вхідних і
вихідних вузлів для побудови SOM
help(somgrid)
'map <- som(X, grid=g, alpha=c(0.05, 0.01), radius=1) # пошук центральних
значень для мінімізації відстані між схожими елементами
plot(map, type='changes') # в ході застосування методу SOM відстань між
елементами зменшилась.
plot(map)
plot(map, type='codes')
plot(map, type='count')
plot(map, type='mapping')
plot(map, type='dist.neighbours')
plot(map, type = "quality")
help("type")

# Формуємо матрицю "Вузли - змінні"
mydata <- as.matrix(map$codes[[1]])

# Застосовуємо ієрархічну кластеризацію з вхідним коефіцієнтом при k=6
som_cluster <- cutree(hclust(dist(mydata)), 6)

```

```
help("cutree")  
# Реалізуємо палітру кольорів  
pretty_palette <- c("#1f77b4", '#ff7f0e', '#2ca02c', '#d62728', '#9467bd')  
  
# Показуємо різними кольорами кластери вузлів та змінні  
plot(map, type = "codes", bgcol = pretty_palette[som_cluster])  
  add.cluster.boundaries(map, som_cluster)
```

Додаток Б  
КОДЕКС АКАДЕМІЧНОЇ ДОБРОЧЕСНОСТІ  
ЗДОБУВАЧА ВИЩОЇ ОСВІТИ ХЕРСОНЬСЬКОГО  
ДЕРЖАВНОГО УНІВЕРСИТЕТУ

Я, Леденчук Станіслав Валдимович, учасник освітнього процесу Херсонського державного університету, **УСВІДОМЛЮЮ**, що академічна доброчесність – це фундаментальна етична цінність усієї академічної спільноти світу.

**ЗАЯВЛЯЮ**, що у своїй освітній і науковій діяльності **ЗОБОВ'ЯЗУЮСЯ**:

– дотримуватися:

- вимог законодавства України та внутрішніх нормативних документів університету, зокрема Статуту Університету;
- принципів та правил академічної доброчесності;
- нульової толерантності до академічного плагіату;
- моральних норм та правил етичної поведінки;
- толерантного ставлення до інших;
- дотримуватися високого рівня культури спілкування;

– надавати згоду на:

- безпосередню перевірку курсових, кваліфікаційних робіт тощо на ознаки наявності академічного плагіату за допомогою спеціалізованих програмних продуктів;
- оброблення, збереження й розміщення кваліфікаційних робіт у відкритому доступі в інституційному репозитарії;
- використання робіт для перевірки на ознаки наявності академічного плагіату в інших роботах виключно з метою виявлення можливих ознак академічного плагіату;
  - самостійно виконувати навчальні завдання, завдання поточного й підсумкового контролю результатів навчання;
  - надавати достовірну інформацію щодо результатів власної навчальної (наукової, творчої) діяльності, використаних методик досліджень та джерел інформації;
  - не використовувати результати досліджень інших авторів без використання покликань на їхню роботу;
  - своєю діяльністю сприяти збереженню та примноженню традицій університету, формуванню його позитивного іміджу;
  - не чинити правопорушень і не сприяти їхньому скоєнню іншими особами;
  - підтримувати атмосферу довіри, взаємної відповідальності та співпраці в освітньому середовищі;
  - поважати честь, гідність та особисту недоторканність особи, незважаючи на її стать, вік, матеріальний стан, соціальне становище, расову належність, релігійні й політичні переконання;
  - не дискримінувати людей на підставі академічного статусу, а також за національною, расовою, статевою чи іншою належністю;
  - відповідально ставитися до своїх обов'язків, вчасно та сумлінно виконувати необхідні навчальні та науково-дослідницькі завдання;
  - запобігати виникненню у своїй діяльності конфлікту інтересів, зокрема не використовувати службових і родинних зв'язків з метою отримання нечесної переваги в навчальній, науковій і трудовій діяльності;
  - не брати участі в будь-якій діяльності, пов'язаній із обманом, нечесністю, списуванням, фабрикацією;
  - не підроблювати документи;
  - не поширювати неправдиву та компрометуючу інформацію про інших здобувачів вищої освіти, викладачів і співробітників;
  - не отримувати і не пропонувати винагород за несправедливе отримання будь-яких переваг або здійснення впливу на зміну отриманої академічної оцінки;
  - не залякувати й не проявляти агресії та насильства проти інших, сексуальні домагання;
  - не завдавати шкоди матеріальним цінностям, матеріально-технічній базі університету та особистій власності інших студентів та/або працівників;
  - не використовувати без дозволу ректорату (деканату) символіки університету в заходах, не пов'язаних з діяльністю університету;
  - не здійснювати і не заохочувати будь-яких спроб, спрямованих на те, щоб за допомогою нечесних і негідних методів досягати власних корисних цілей;
  - не завдавати загрози власному здоров'ю або безпеці іншим студентам та/або працівникам.

**УСВІДОМЛЮЮ**, що відповідно до чинного законодавства у разі недотримання Кодексу академічної доброчесності буду нести академічну та/або інші види відповідальності й до мене можуть бути застосовані заходи дисциплінарного характеру за порушення принципів академічної доброчесності.

07.09.2020  
(дата)

\_\_\_\_\_ (підпис)

Леденчук Станіслав  
(ім'я, прізвище)