

## ПРОГРАМНІ ЗАСОБИ ВИЯВЛЕННЯ КЛАСТЕРІВ АБІТУРІЄНТІВ ДЛЯ ВІДКРИТИХ ДАНИХ

*В статті представлено методичний підхід до оцінки ефективності стратегічного моніторингу основних кластерів абітурієнтів що вступають до закладів вищої освіти України. Реалізовано за допомогою спеціального пакету прикладних програм. Дані для кластеризації були взяті з відкритого доступу за допомогою WEB-парсингу.*

*Ключові слова: самоорганізаційні карти, кластеризація, інтелектуальний аналіз даних.*

*The article presents a methodical approach to assessing the effectiveness of strategic monitoring of the main clusters of applicants entering higher education institutions of Ukraine. Implemented using a special software package. Data for clustering were taken from open access using WEB-parsing.*

*Keywords: data mining, self-organizing maps, clustering.*

Одним з головних чиників розвитку закладів вищої освіти (далі ЗВО) стало прогнозування ефективності їх функціонування, аналіз показників за якими ЗВО оцінюється в масштабах світового прогресу. Відбувається ж цей аналіз, за допомогою різних рейтингів та моніторингу ефективності діяльності ЗВО що оцінюється Міністерством освіти і науки України. Важливим чиником для будь якого закладу є оцінювання поточної ситуації і вміння прогнозувати значення показників ефективності діючи на перспективу.

Наразі з'являється все більше різних даних з прогнозуванням – погоди, фінансовими чинниками, відсотками продажів, і купую інших. Інколи, виникають завдання для аналізу, які можливо представити лише в математичній та числовій формі. Як зразок уявіть ситуацію, вам необхідно обробити дані з не ясно вираженими принципами відбору: перспективний товар чи ні, пошук нових партнерів чи надійний обраний банк а клієнт платоспроможній і т.д. Саме для того, щоб вирішення цих завдань призводило до отримання максимально якісних та чітких результатів, слід користуватись різними методами аналізу даних. Одним з найбільш перспективних підходів, на мій погляд, є використання SOM (Self-organizing map) для візуалізації та кластеризації даних.

Першою чергою постало питання відбору даних, для цього було розглянуто декілька ресурсів з відкритими даними абітурієнтів. В таблиці 1 показано за якими критеріями йшов вибір.

Таблиця 1

### Порівняння сервісів з відкритими даними абітурієнтів

Критерії \ Назва ресурсу	abit-poisk.org.ua	vstup.osvit a.ua	https://vstup.edbo.gov.ua
Можливість парсингу	+	+	+
Обмеження на кількість запитів	До ~25 запитів	До ~25 запитів	До ~50 запитів
Зручність відбору даних	Немає зручного сортування	Доволі складно	Зручно
Є всі можливі дані	ні	ні	так
Беруть дані з інших ресурсів	+	+	-

Таким чином, після аналізу, для збору даних було обрано систему ЄДЕБО (ЄДИНА ДЕРЖАВНА ЕЛЕКТРОННА БАЗА з питань ОСВІТИ).

**Поля що були взяті для аналізу:** позиція заяви в ЄДЕБО, код університету ЄДЕБО, шифр спеціальності, пріоритет, СВО, обсяг держзамовлення, обсяг на контракт, вартість навчання за рік, коефіцієнт балу з укр. мови, коефіцієнт балу з історії, коефіцієнт балу з математики, коефіцієнт балу з іноземної мови, коефіцієнт балу з фізики, коефіцієнт балу з географії, коефіцієнт балу з біології, коефіцієнт середнього балу добументу про освіту, фахове випробування ( $k=$ ), коефіцієнт балу іноземної мови (ЄВІ), коефіцієнт інших показників конкурсного відбору ( $k=$ ), бал ЗНО з укр. мови, бал ЗНО з історії, бал ЗНО з математики, бал ЗНО з іноземної мови, бал ЗНО з фізики, бал ЗНО з географії, бал ЗНО з біології, середній бал добументу про освіту, фахове випробування, бал з іноземної мови (ЄВІ), інші показники конкурсного відбору, сільський коефіцієнт, квоти, конкурсний бал, до наказу (бюджет), до наказу (контракт), оригінали документів (табл. 2).

Таблиця 2

**Приклад відбору даних для абітурієнтів на другий (магістерський) рівень**

№	Код університету	Спеціальність	Пріоритет	Обсяг держзамовлення	Фахове випробування	ЄВІ (іноземна мова)	Оригінали документів
1	48	121	К	0	200	196	1
2	48	121	К	0	174	169	1
3	48	121	К	0	195	141	1
4	48	121	К	0	179	144	1

**Самоорганізаційна карта Кохонена** – «нейронна мережа з некерованим навчанням, яка використовується для конструювання багатовимірного простору в простір з нижчою розмірністю. Створює дискретне представлення вхідних просторів навчальних вибірок, які називаються картою», і тому використання цього типу нейронної мережі є методом для зниження розмірності частіше в двувимірне [1]. Застосовується також для вирішення задач моделювання, прогнозування та ін. Основа ідеї полягає в ітеративному налаштуванні вектору вагових коефіцієнтів  $w_j$  кожного нейрона,  $j = 1, 2, \dots, p$ , для цього застосовується модифікований алгоритм змагального навчання Хебба(2), який враховує не тільки внесок нейрона-переможця, але і найближчих його сусідів, розташованих в  $R$  - оточенні:

1. На стадії ініціалізації усім ваговим коефіцієнтам привласнюються невеликі випадкові значення  $w_{ij}^0$ ,  $i = 1, 2, \dots, m$ .

2. На виході мережі подаються послідовно у випадковому порядку образи у об'єктів вхідного шару, і для кожного з них обирається "нейрон-переможець" ВМУ (Best Matching Unit) з мінімальною відстанню  $\sum_{i=1}^m (y_i - w_{ij}^t)$ .

3. Визначається підмножина "найближчого оточення" ВМУ, радіус якого  $R$  зменшується з кожною наступною ітерацією  $t$ .

4. Перераховуються ваги  $w_j^t$  виділених вузлів з урахуванням їх відстаней до ВМУ і близькості до вектору  $u$ .

Кроки з другого по четвертий повторюються поки вихідні значення мережі не будуть стабілізовані з заданою точністю.

Приклад реалізації алгоритму Кохонена з даними абітурієнтів Херсонського державного університету (табл. 3):

Таблиця 3

**Фрагмент даних абітурієнтів ХДУ на магістерські освітні програми**

№	Код						Наказ бюд.	Наказ конт.	Оригінали док.	Вузли
	Шифр спец.	Обсяг	Ціна	ЄВФІ	ЄВІ	Інші				

1	48	121	4	24560	200	196	8,5	1	0	1	1
2	48	121	4	24560	174	169	6	1	0	1	2
3	48	121	4	24560	195	141	9	1	0	1	29
4	48	121	4	24560	179	144	17,5	1	0	1	11
5	48	121	4	24560	166	169	5,5	1	0	1	2
6	48	121	4	24560	134	164	5,5	1	0	1	2
7	48	121	4	24560	147	189	11,5	0	0	0	5

У результаті запровадження алгоритму карт Кохонена, що самоорганізуються (Self-organizing map, SOM), середня відстань до найближчих центрів знижується з 0,045 до 0,02 (рис. 1).

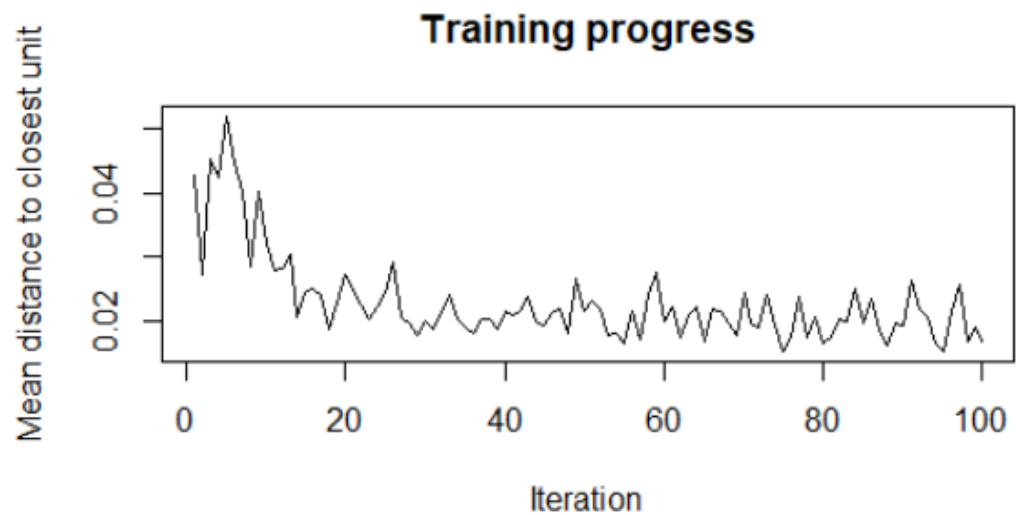


Рис. 1. Зменшення середньої відстані до найближчих нейронів після 100 ітерацій навчання мережі SOM із заданими значеннями параметра альфа

Серед 140 магістрантів, які вступали на освітні програми з галузі 12 Інформаційні технології було виявлено 6 кластерів (рис. 2).

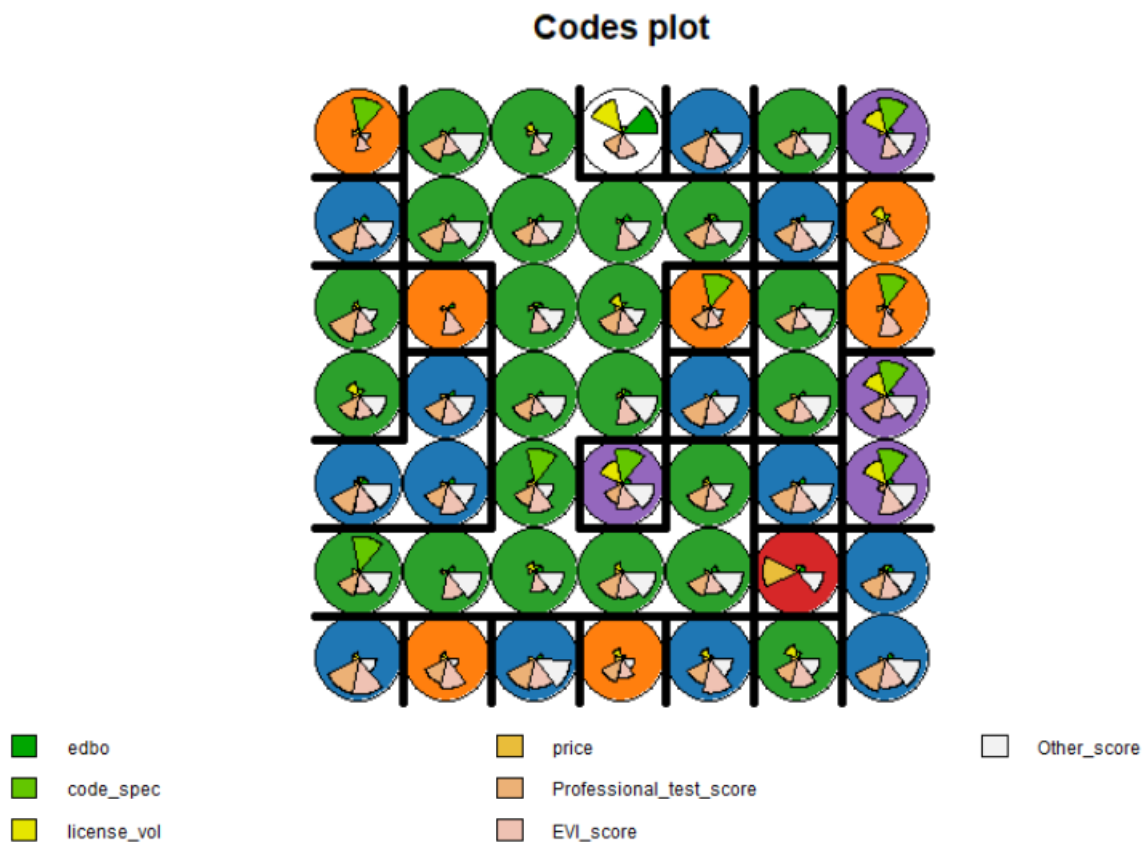


Рис. 2. Кластеризація вузлів карти Кохонена (SOM)

Перший і найчисельніший кластер абітурієнтів (зелений) відображає абітурієнтів із середніми балами за ЄВІ і фаховий іспит.

Другий кластер абітурієнтів (синій) відображає абітурієнтів із вищими, ніж у інших, балами за ЄВІ і фаховий іспит.

Третій кластер (помаранчевий) включає абітурієнтів, у яких фаховий іспит і ЄВІ здані на бали, нижче за середні, або високий бал ЄВІ і низький бал за фаховий іспит.

Четвертий (фіолетовий) кластер демонструє студентів, у яких зданий фаховий іспит і ЄВІ і є додаткові бали для вступу.

По одному вузлу п'ятий кластер представляє абітурієнтів, які з мінімальними балами прагнуть вступити на контрактну форму навчання і шостий кластер представляє вузол для абітурієнтів, які віддають перевагу ІТ спеціальностям з великим ліцензійним обсягом.

**Висновки:** перші два кластери представляють абітурієнти із високими вступними балами, які планують вступ на бюджетні місця, абітурієнти третього і четвертого кластеру планують вступати на бюджет і відмовляються від вступу в магістратуру, якщо проходять на контракт, абітурієнти п'ятого і шостого кластеру прагнуть вступити на контракт, якщо наберуть мінімальні бали для вступу.

#### ЛІТЕРАТУРА:

1. Самоорганізаційна Карта Кохонена URL: <https://artsandculture.google.com/entity/m0k6rk?hl=uk>
2. The Organization of Behavior – New York: Psychology Press, 2002. – 378 с. – (Behavioral Sciences; кн. 2147483647).
3. ЄДЕБО. – 2020. – URL: <https://vstup2020.edbo.gov.ua/>
4. Самоорганізующиеся карты Кохонена URL: <https://ranalytics.github.io/data-mining/105-Cohonen-Maps.html>.

5. Анисимова Э.С. Самоорганизующиеся карты Кохонена в задачах кластеризации / Анисимова Э.С. – URL: <https://cyberleninka.ru/article/n/samoorganizuyuschiesya-karty-kohonena-v-zadachah-klasterizatsii/viewer>.
6. Kobets, V., Yatsenko, V., Voynarenko, M. (2019). Cluster analysis of countries inequality due to IT development. CEUR-WS. 2393, 406–421.
7. 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.

**Рекомендує до друку науковий керівник професор Кобець В.М.**