

Всеукраїнській конкурс студентських наукових робіт
з природничих, технічних та гуманітарних наук

Спеціальність: Комп'ютерні науки

Конкурсна робота

Шифр роботи «Дом»

**Кластеризація об'єктів нерухомості
за цінovими й територіальними ознаками**

2019-2020 навчальний рік

Анотація

Актуальність: прогнозування цінних показників тих чи інших об'єктів є однією з найбільш жваво впроваджуваних та досліджуваних проблем нашого сьогодення. Зокрема, однією з таких ознак є вартість одиниці площі житла в тих чи інших населених пунктах, районах і забудовах. Здатність автоматизовано розраховувати ціну помешкання і здійснювати цінове районування нерухомості допомогла б ріелторським агентствам ефективно та без рутини проводити оцінку апартаментів, посприяла би продавцям житла в точній оцінці власного майна, а також була би зручним індикатором сприятливих і не дуже зон майбутніх житлових комплексів для будівельних компаній.

Недоліки існуючих рішень: попри активну забудову українських міст, на сьогоднішній день ми все ще не маємо інтегрованого сервісу аналітики, котрий би був здатний надати всеохопну інформацію щодо цінової ситуації у великих населених пунктах. Наявні розробки або є недостатньо об'єктивними та глибинними, або ж мають недостатньо широке коло охоплених ознак і територій. Наше ж дослідження виправляє ці прорахунки, розглядаючи найбільш значущі міста, ігноруючи умовні обмеження у вигляді адміністративного поділу міст, беручи до уваги широкий спектр потенційно важливих ознак та факторів житла й візуалізуючи результати в зручній для сприйняття формі.

Мета роботи: покращення якості територіального групування об'єктів нерухомості з ціллю підвищення точності розрахунку зон спільної цінової політики.

Об'єкт досліджень: первинне та вторинне житло в українських містах.

Предмет досліджень: цінове районування помешкань у великих містах України.

Наукова новизна: запропоновано новий метод прогнозування вартості помешкань, що ґрунтується на поєднанні алгоритмів кластеризації та регресії, за рахунок розбиття зони міста на значну кількість невеликих за територіальним охопленням сегментів, що дозволяють локалізовано розраховувати ціну для заданої житлової забудови.

Практична цінність роботи: результуючі картограми районування міст можуть прискорити / спростити процес ріелторської оцінки нерухомості, бути застосовані урбаністами, забудовниками й банківськими організаціями для пошуку сприятливих територій майбутніх житлових районів.

Застосовані підходи: системний підхід, аналіз, синтез, емпіричний підхід.

Стислий опис результатів дослідження: було розроблено набір методик щодо цінового районування житлової нерухомості українських міст на основі поєднання технік кластеризації та регресії, виділено істотні ознаки, що формують результуючу вартість помешкання, а також спроектовано, проаналізовано та імплементовано додаток, котрий декомпонує територію обраного населеного пункту на самодостатні зони з подібною ціною квадратного метра апартаментів з метою подальшого прогнозування вартості житлової площі.

Ключові слова: кластеризація, групування, нерухомість, житло, квартири, розташування, аналіз даних, kmeans, fcm, python, scikit-learn.

Склад роботи: 26 аркушів, 5 малюнків, 3 таблиці, 10 джерел посилань.

Зміст

<u>Вступ</u>	4
<u>1. Аналіз конкурентних рішень</u>	5
<u>1.1. Наявні розробки</u>	5
<u>1.2. Порівняльна характеристика сервісів</u>	8
<u>1.3. Підсумки аналізу наявних додатків</u>	9
<u>2. Математичний апарат розв'язання</u>	10
<u>2.1. Наукова новизна</u>	10
<u>2.2. Характеристика вибірки</u>	12
<u>2.3. Істотні ознаки об'єктів</u>	13
<u>3. Алгоритм розв'язання</u>	14
<u>3.1. Покрокові розрахунок зон і прогнозування</u>	14
<u>3.2. Алгоритми кластеризації</u>	16
<u>4. Апробація удосконалених методів</u>	18
<u>4.1. Програмна реалізація</u>	18
<u>4.2. Візуалізація здобутків</u>	20
<u>Висновки</u>	23
<u>Джерела</u>	25

Вступ

Ринок нерухомості - одна з найбільш значущих галузей національної економіки, адже кожна комірка суспільства з часом прагне здобути власний дах над головою. Тож після економічної кризи 2014-2015 років ринок житла потроху стабілізується, переймаючи на себе частину капіталу, який втрачають банківські установи внаслідок падіння кредиту довіри громадян.

Більше того: у 2017 році Україна за кількістю новозбудованих квартир на 1000 осіб населення вперше в новітній історії обігнала ЄС (2.9 проти 2.8). Особливо відзначилася столиця, де у 2017 було побудовано 28 200 нових квартир: на Київ і область припало 43% прийнятого до експлуатації житла у багатоквартирних будинках [\[1\]](#).

В умовах інформаційної епохи, коли цінною сировиною стають цифрові дані, а невід'ємним атрибутом - засоби їхньої автоматизованої обробки, великі гравці на ринку житла, такі як ріелорські агентства, забудовники, урбаністи та банківські організації, потребують інструменти аналітики, що здатні швидко та якісно розбивати величезні шматки інформації на групи, з'ясовувати, встановлювати, спростовувати гіпотези щодо залежностей та тенденцій на ринку житла.

Тож за мету роботи нами було встановлено покращення якості територіального групування об'єктів нерухомості з ціллю підвищення точності розрахунку зон спільної цінової політики.

1. Аналіз конкурентних рішень

1.1. Наявні розробки

- blagovist.ua - обраховує щомісячну статистику цін на житло в Києві з 2000 року з використанням власної БД [2].

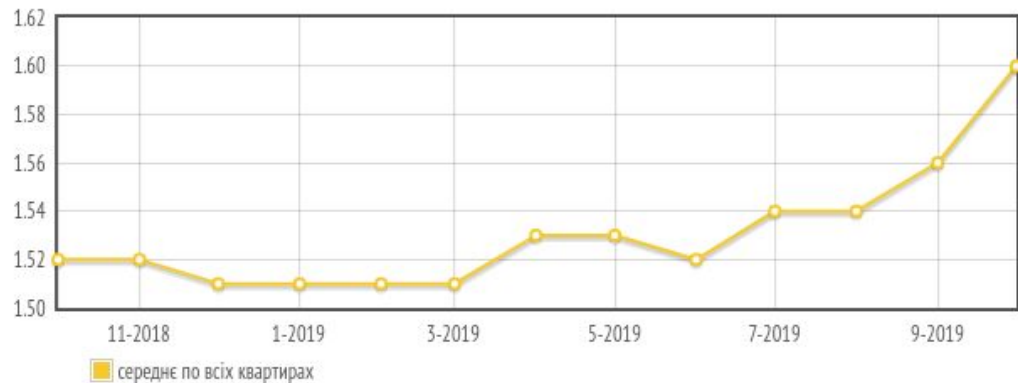


Рис. 1 - динаміка зміни середньої ціни за 1 кв. м загальної площі (у тис. дол. США) за жовтень 2018 - жовтень 2019. Джерело - blagovist.ua [2]

Даний ресурс має ту саму проблему, що й 100realty.ua - надто узагальнене значення середньої вартості позбавляє широкого охоплення різноманітних зон і місцевостей міста, де ціни апартаментів підпорядковуються локальним закономірностям [3]; але саме ці закономірності можуть бути вирішальними під час ріелторської оцінки невідомого помешкання. Відносно даного сайту наш застосунок має на **87%** більше територіальне охоплення, адже охоплює більше 4 найбільших міст України за винятком столиці, й у **2** рази більшу якість і обсяг оброблених даних, бо надає цілісну картину цінового районування для вказаного населеного пункту.

- olimp.net.ua - опрацьовує усю територію України, підраховуючи середню вартість одиниці площі житла, і відображає цінові зони на картограмі [4] .

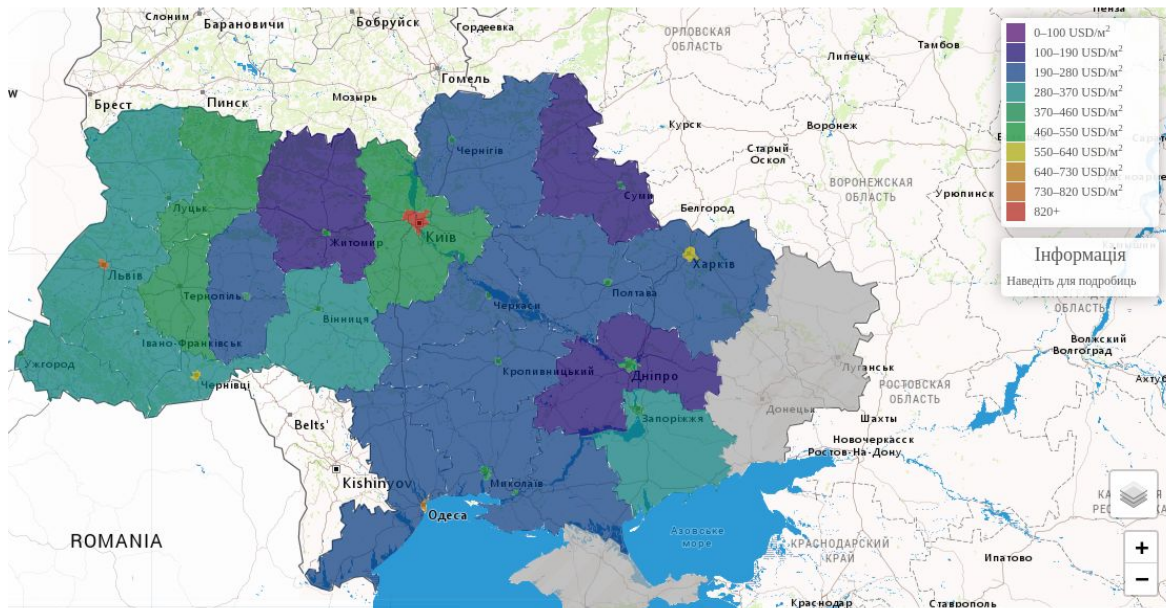


Рис. 2 - середня вартість житла в Україні за вересень 2019. Джерело - map.olimp.net.ua [4]

Наведена хороплетова карта є чудовим засобом візуалізації, бо надає інформацію в більш розлогому і доступному вигляді, ніж часовий ряд середньої вартості, але все ж не бере до уваги специфіку кожного міста. Представлений розрахунок охоплює цілі області й міста у вигляді суцільних зон, ігноруючи регіональні особливості. Це позбавляє користувача конкретики й наочності, адже міста за своєю природою є доволі різношаровими утвореннями, у той час як більшість територій областей не містить житлової забудови взагалі. У нашій розробці відсутні ці недоліки, надаючи на **50%** більш значущі й на **60%** істотніші дані, бо не зосереджується на цілих регіонах країни, а точно фокусується на певних, найбільш заселених точках держави.

- misto.lun.ua - здійснює аналіз цін новобудов у Києві, Одесі та Львові [5].

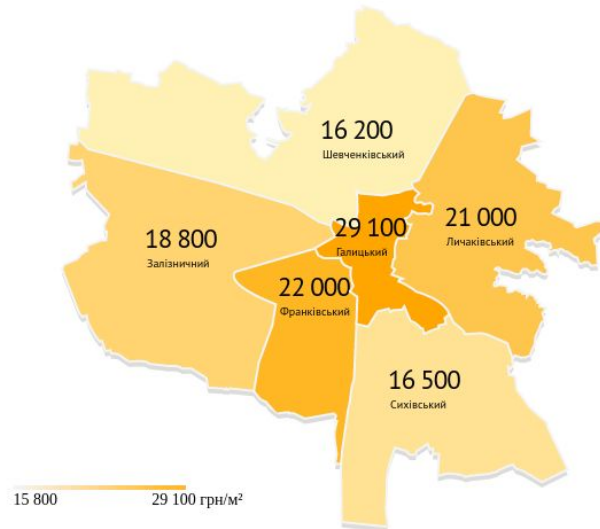


Рис. 3 - середня вартість 1 кв. м. житла у новобудовах Львова за жовтень 2019.

Джерело - misto.lun.ua [5]

Даний сервіс є найбільш комфортним з боку споглядання й представлення даних, його експертна оцінка візуалізації на **90-110%** вища за решту наявних ресурсів. до того ж, автори сайту будують карти комфорту, аналізуючи відкриті дані [6], що є надзвичайно наочною й корисною формою районування міста за певними ознаками. Тим не менш, цей ресурс не надає подібної цінової декомпозиції помешкань, адже інформація про середню вартість у кожному з адміністративних районів міста (аналогічно до попереднього додатку) нівелює специфічні зони вартості тих чи інших територіальних утворень.

1.2. Порівняльна характеристика сервісів

Для порівняння характеристик сервісів була застосована наведена нижче формула зваженого середнього, де w_i - вага фактору, x_i - значення фактору (в межах від 0 до 10). Всі величини були отримані в результаті колаборації з експертною групою у складі 3 чоловік, співробітників агентства veritexgroup.com.ua.

$$Q_j = \sum_{i=1}^n w_i x_{ij}$$

Таблиця 1. Аналіз наявних розробок за істотними факторами

	100realty.ua	blagovist.ua	olimp.net.ua	misto.lun.ua	наявне	план
Значимість отриманих результатів	5	5	6	7	9	10
Територіальне охоплення	4	4	9	6	7.5	8
Інформативність й обсяг результатів	4	4	5	7	8	10
Якість візуалізації	4	3	5.5	9	7	10
Можливість інтеграції з іншими сервісами	0	0	2	2	0	7
Q	3.85	3.65	5.65	6.75	7.175	9.34

Таблиця 2. Вагові коефіцієнти факторів

Значимість результатів, $w1$	Територіальне охоплення, $w2$	Обсяг результатів, $w3$	Якість візуалізації, $w4$	Можливості для інтеграції, $w5$
0.25	0.15	0.3	0.2	0.1

1.3. Підсумки аналізу наявних додатків

Найбільшими недоліками описаних ресурсів ми вважаємо нестачу ознак при аналізі й некоректну модель обрахунку вартості житла. За нашими спостереженнями, вони беруть до уваги не більше **2-4** ознак - *зазначену вартість* та *географічні координати*. В цей же час ми рахуємося із **7-9** параметрами й плануємо розширити це число **13-14** особливостей.

Окреме слово про безпосередньо нашу розробку: в порівняльній таблиці наведено додаток на 2 етапах: *наявний* та *план*. Перший з них - це окреслена кодова база, моделі, техніки, вибірка. Другий же відноситься до рівня комерційного продукту з налагодженими процесами, вдосконаленими підбірками даних й *можливістю інтеграції із зовнішніми ресурсами* - це дозволить колабораціонувати із клієнтськими сервісами для підвищення рівня автоматизації й здешевлення експлуатації.

В цілому, *misto.lun.ua* є найбільш оптимальним з наявних ресурсів аналітики, чия комплексна оцінка лише на **6%** поступається здобуткам нашої розробки. Однак функціонал даного й решти додатків у ціновому районуванні на **30-60%** відстають від ефективності наших *наявних* методів, а при подальшій розробці безпосередньо наш застосунок зможе мати й **50-80%** перевагу над аналогами.

2. Математичний апарат розв'язання

2.1. Наукова новизна

Наріжним каменем, що дозволяє нам ефективніше (порівняно з наявними розробками) групувати апартаменти й у подальшому прогнозувати вартість помешкання, є техніка поєднання алгоритмів кластеризації та регресії, котра за рахунок розбиття зони міста на значну кількість невеликих за територіальним охопленням сегментів дозволяє локалізовано розраховувати ціну для заданої житлової забудови. Індикатором якості даного способу є показник r^2 , котрий для 65% угруповань сягає значень **0.7-0.85** і вище.

Головне завдання кластеризації - розбити територію міста на одиниці, споріднені за достатком інфраструктури й ціновою політикою. Було встановлено, що найбільш оптимальним є групування, за якого у кожен утворену зону потрапляє від **5** до **30** першопочаткових інформаційних центрів. Саме за таких обставин локальні кластери оточуватимуть території, на яких зручно виводити числові закономірності.

Після формування кластерів кожна зона надає міні-датасет, на основі якого можна побудувати поліноміальну багатофакторну регресійну модель прогнозування. Для обрахунку беруться ті з ознак, чий коефіцієнт кореляції Пірсона перевищує **0.5**.

Опісля ціна невідомої квартири прогнозується так: за геокоординатами житла визначається кластер, що містить помешкання, після чого застосовуються конкретні вагові коефіцієнти до ознак цільових апартаментів.

2.2. Характеристика вибірки

Цільовими об'єктами нашого дослідження виступили помешкання у багатоквартирних будинках 8 нижченаведених міст України. Вибір саме цих населених пунктів зумовлений тим, що вони є достатньо велелюдними, а також надають достатньо репрезентативну вибірку апартаментів.

Таблиця 3. Кількісний опис датасету

Місто	Первинне житло	Вторинне житло	Сумарно
Київ	5000	23600	28600
Одеса	6900	30500	37400
Харків	1500	3900	5400
Дніпро	900	2200	3100
Львів	800	1000	1800
Запоріжжя	-	900	900
Вінниця	2000	2200	4200
Миколаїв	-	1500	1500

Увесь масив інформації (об'ємом понад 80 тис. квартир) був зібраний за період листопада - грудня 2019 року. У якості джерел інформації слугували дані ресурсу з продажу нерухомості dom.ria.com, котрий надав напів-відкритий API.

Найбільш масовими є вибірки великих міст - Києва та Одеси, чия сумарна кількість квартир у новобудовах аж в **3** рази перевищила аналогічний показник в решті міст. У той же час, через нестачу даних, Запоріжжя й Миколаїв не беруться до уваги при аналізі первинного житла.

2.3. Істотні ознаки об'єктів

Групування квартир потребує набір певних числових характеристик досліджуваних сутностей. На наш погляд, найбільш суттєвими факторами, що й формують вартість квадратного метра, є не внутрішнє оформлення апартаментів чи суб'єктивна оцінка продавця, а саме довколишня інфраструктура - зупинки швидкісного транспорту, вузли територіального сполучення, історико-культурні й рекреаційні зони й пам'ятки, водойми, відсутність промислових забудов і установ, потужні торгові комплекси і тд. На даний момент до розрахунку беруться наступні ознаки помешкань:

- ***Longitude*** - географічна довгота (тут і надалі вважаємо квартири матеріальними точками);
- ***Latitude*** - географічна широта;
- ***City center distance*** - відстань до історичного центру міста, як-от Монумент Незалежності в Києві чи міська ратуша у Львові (Тут і надалі всі відстані між об'єктами рахуються, як географічна відстань [\[7\]](#));
- ***Subway station distance*** - відстань до найближчої станції метро;
- ***Mall distance*** - відстань до найближчого масивного торгового комплексу чи центру, наприклад: "Ocean Plaza", "Sky Mall", "Forum Lviv", і тд.;
- ***Railway station distance*** - відстань до найближчого великого залізничного вокзалу; у Києві, наприклад, це "Київ-Пасажирський" та "Дарниця";
- ***Airport distance*** - відстань до найближчого міжнародного аеропорту;
- ***Culture place distance*** - відстань до найближчого культурного осередку, як-от комплекс "Менора" в Дніпрі чи собор Св. Юра у Львові;
- ***Business center distance*** - відстань до найближчого бізнес-центру; наприклад, комплекс Gulliver у столиці.

Кожна лінійна ознака кожної квартири була обрахована за наступною запропонованю нами формулою:

$$d(\lambda_{fi}, \phi_{fi}, \lambda_o, \phi_o) = \min \left(R \sqrt{(\phi_{fi} - \phi_{oi})^2 + \left(\cos \left(\frac{\phi_{fi} + \phi_{oi}}{2} \right) \cdot (\lambda_{fi} - \lambda_{oi}) \right)^2} \right)$$

$$\forall \lambda_{oi} \in \lambda_o, \phi_{oi} \in \phi_o$$

Де λ_{fi} , ϕ_{fi} - довгота й широта чергової квартири в радіанах, λ_o , ϕ_o - масиви довгот і широт об'єктів інфраструктури, до яких знаходиться відстань, наприклад, бізнес-центрів чи станцій метро, R - радіус Землі.

На додачу до вищенаведених ознак ми плануємо включити в список суттєвих особливостей ще й *близькість крупних водойм* (для оцифрування словесної характеристики “видової квартири”), *парків та скверів*, а також *віддаленість промислових зон*. Ці об'єкти за своєю природою займають доволі значні території, тож до нашого математичного апарату вони будуть додані у вигляді полігонів.

3. Алгоритм розв'язання

3.1. Покрокові розрахунок зон і прогнозування

Вище були наведені математичний апарат, опис датасету та суттєві ознаки досліджуваних об'єктів. Нижче наводиться більш детальна схема роботи

- Вибірка даних усіх квартир цільового населеного пункту заданого типу житла із персистентного сховища даних;
- Обрахунок відстаней між квартирами й цільовими об'єктами, такими, як ТРЦ, БЦ, станції метро і тд; кожна будова інфраструктури сприймається, як безрозмірна точка на площині (для спрощення обрахунку).
- Обрахунок центрів кластеризації. Кількість кластерів встановлюється емпірично для заданого міста й типу житла; для первинного житла в столиці, наприклад, це число близьке до **300**;
- Застосування діаграми Вороного й контурів міських меж для окреслення полігонів зон. Центри деяких кластерів можуть накладатися або бути близькими одне до одного, у такому разі їхні багатокутники зливаються ;
- Відсіювання невалідних зон - кожна з них повинна містити не менше **5** квартир для складання мінімально значущої вибірки для майбутнього прогнозування вартості нових апартаментів;
- Виявлення істотних ознак квартир, що потрапили до того чи іншого сегменту з допомогою кореляційних матриць. Цей відбір здійснюється за формулою:

$$F = \{p_i, \forall p_i \in P \mid r_i \geq 0.5\}$$

Де P - масив усіх чисельних ознак, r_i - коефіцієнт кореляції Пірсона p_i ознаки із вартістю 1 кв. м. житлової площі для усіх квартир чергового кластеру, F - масив значущих параметрів для квартир цільової територіальної одиниці.

Після цього для кожного угруповання обраховуються вагові коефіцієнти для майбутньої формули на основі поліноміальної багатofакторної регресії 2 порядку, котра виглядатиме таким чином:

$$R(f_1, \dots, f_{k_j}) = w_0 + w_1 f_1 + \dots + w_{k_j} f_{k_j} + w_{11} f_1^2 + w_{12} f_1 f_2 + \dots + w_{k_j k_j} f_{k_j}^2$$

- Збереження результуючих полігонів до БД з датою проведення підрахунку;
- Обрахунок ціни невідомої квартири: за геокоординатами в БД знаходиться зона, що містить задане помешкання, для квартири обраховуються істотні ознаки відстаней, опісля з допомогою збережених вагів і вищенаведеної формули прогнозується вартість 1 кв. м. житлової площі.

3.2. Алгоритми кластеризації

Кластеризація - одна з найбільш поширених задач *unsupervised learning*, метою якої є знаходження закономірностей в наборах нерозмічених даних, тобто процес організації елементів у групи подібних об'єктів за певним правилом. У той же час, формальний кластер - це одне з таких угруповань, в межах якого сутності є схожими, а поза межами - різними [8]. Кластеризація як техніка має багато реалізацій, серед усієї їхньої сукупності нами було випробувано та порівняно в дії дві алгоритма:

- *K-Means* - “алгоритм к-середніх“, котрий ділить задану множину об'єктів на k груп, кожен з учасників яких в середньому є ближчим саме до свого умовного центру, аніж до будь-якого іншого. Метод базується на мінімізації наступної цільової функції:

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2$$

Де x_i - черговий елемент вибірки, c_j - центр кластера, який на j -ітерації присвоєний елементу x_i , n - обсяг вибірки, k - число угруповань [9].

K-Means має ряд переваг, серед яких - простота в розумінні, відносно висока швидкість дії та ефективність при роботі з відносно невеликими кількостями даних (до 1000 квартир). У таких ситуаціях *K-Means* має навіть ліпшу продуктивність порівняно з аналогічними техніками, в середньому на **5-15%**. Даний алгоритм корисно застосовувати для групування житла в середніх за розмірами й населенням містах: Миколаєві, Вінниці, тощо.

- *FCM, Fuzzy C-Means* - “нечіткий” аналог попереднього методу, котрий передбачає включення кожного елементу в кожен кластер із певним

чисельним членством. Математично задача відшукування угруповань зводиться до знаходження оптимуму цільової функції:

$$J = \sum_{i=1}^n \sum_{j=1}^c u_{ij}^m \cdot \|x_i - c_j\|^2, \quad 1 \leq m \leq \infty$$

де n - розмірність вибірки, c - кількість кластерів, x - масив елементів, c - масив центрів кластерів, u - матриця членства усіх елементів в кожному кластері [10].

В контексті нашої задачі нечітке групування є більш пріоритетним тоді, коли якісна декомпозиція є особливо бажаною, а вибірка є доволі великою. В середньому, він є на **10-25%** швидшим (в реалізації мови *Python*) і на **8-20%** точнішим у таких ситуаціях, котрі є характерними для великих міст (Києва, Харкова, Одеси).

Порівнюючи вищенаведені алгоритми, хотілося б відзначити техніку нечіткої кластеризації, котра в загальному випадку краще підходить для використання в нашій задачі, адже її застосування дає **20%** вигреш у продуктивності й **15%** вигреш у коректності та влучності групування.

4. Апробація удосконалених методів

4.1. Програмна реалізація

Назва додатку: *halo*

GitLab репозиторій: <https://gitlab.com/hard-sunrise/sunrise>

Дане посилання вказує на сховище всіх висхідних кодів програмного забезпечення. Більш стислі фрагменти коду знаходяться в **Додатках**.

Мова програмування: *Python, v3.7*

На нашу думку, серед всіх сучасних мов програмування, причетних до написання штучного інтелекту, машинного навчання та аналізу даних, саме *Python* є найбільш універсальним та довершеним інструментом у вирішенні потенційних задач, і на те є декілька підтверджень:

- *Python* має простий та зрозумілий синтаксис, є легким у опануванні й терплячим до помилок;
- Має потужну підтримку комм'юніті програмістів та компаній; свого часу корпорація *Google* взяла *Python* під власну опіку, прискоривши популяризацію та розвиток даної мови;
- *Python* володіє широким стеком бібліотек та фреймворків для найбільш різноманітного спектру потреб і задач, котрі зручно оформлені у вигляді репозиторію пакетів під назвою *PyPI*.
- Розширення та прискорення у вигляді модулів мов *C/C++*, котрі покращують продуктивність *Python* в “тонких місцях”.
- Легкість у поєднанні із вебom, що часто дає рішення, котрі сполучають в собі моделі машинного навчання та високонавантажені веб-сервіси.

РСУБД: PostgreSQL

Дана система управління БД є однією з найбільш поширених у світі завдяки потужній підтримці спільноти та надійним системним рішенням. *PostgreSQL* має велику кількість вбудованих функцій та можливостей, що дозволяє з коробки отримати бойовий агрегат зберігання даних. Також надзвичайно приємним доповненням є розширення під назвою *PostGIS*, котре надає величезну кількість просторових та географічних об'єктів і структур даних до стандартного функціоналу *PostgreSQL*. Саме його ми й використали під час проектування БД.

Основні бібліотеки та фреймворки:

- *aiohttp* - бібліотека асинхронних HTTP-запитів, слугує HTTP-клієнтом скраперів веб-сайтів;
- *beautifulsoup4* - один з найліпших модулів *Python* для парсингу XML та HTML розмітки. Використовується у спайдерах під час збору даних;
- *sqlalchemy* - один з найбільш популярних *ORM* фреймворків, котрий є необхідним для доступу до БД й збору первинного датасету;
- *pandas* - одна з важливих бібліотек машинного навчання, котра привносить в код спеціалізовану структуру даних - *data frame* , аналог таблиць в базах даних;
- *scikit-learn* - базовий фреймворк машинного навчання, бібліотека наукових алгоритмів та структур даних, серед яких є базові реалізації *kmeans* та конструктора діаграми Вороного;
- *scikit-fuzzy* - набір інструментів для маніпуляцій з нечіткою логікою;
- *plotly* - універсальна бібліотека візуалізації та побудови різноманітних діаграм і чартів.

4.2. Візуалізація здобутків

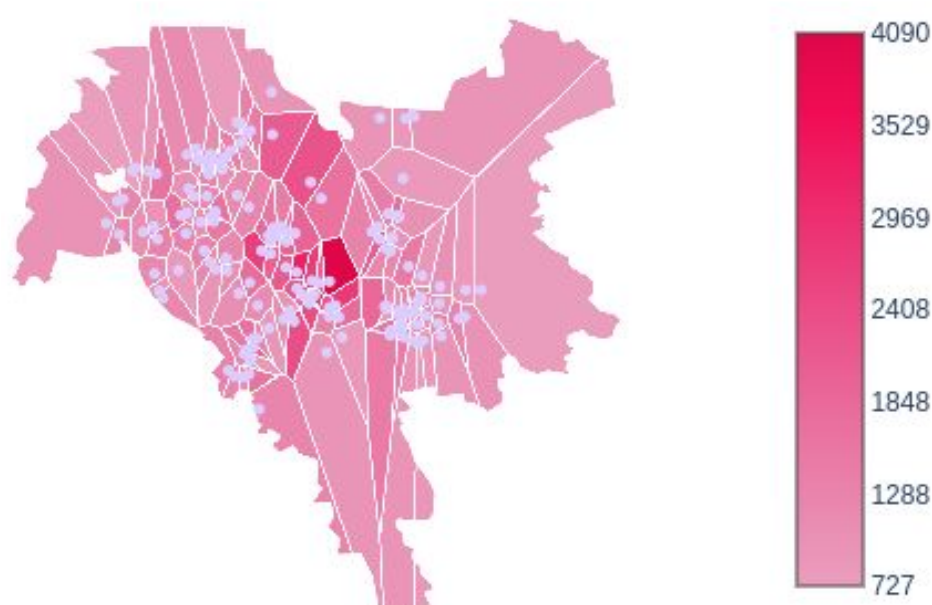


Рис. 4. Декомпозиція первинного житла Києва через FCM на 120 кластерів.

Наведена вище картограма ілюструє одне з можливих розбиттів території міста Київ. На рахунок неї хочеться зазначити наступне:

- Достатньо передбачувано, що найбільш дорогівартісними районами столиці є Клов, Липки, Черепанова й Чорна гори з середньою вартістю 1 м² у **3500-4000\$+**. Це пояснюється близькістю декількох станцій метрополітену, бізнес-центрів та офісних споруд, відносно сильним наближенням центру, розміщенням на території мальовничих схилів правого берега Дніпра.
- Тим не менш, не поступаються за рівнем ціни й північні райони міста, в т.ч. Оболонь, Мінський масив, Виноградар. Висока вартість помешкань у новобудовах (**2500-3000\$**) тут зумовлена наближеністю до Дніпра (Оболонська набережна), концентрація кількох потужних ТРЦ (Blockbuster,

Dream Town, Karavan), кілька станцій синьої гілки та цьогорічне відкриття 3 нових станцій зеленої гілки метро.

- Новобудови із середньою вартістю житла в **1000\$+** гуртуються переважно довкола вузлів сполучення, таких як станції метро та залізничні вокзали, видових зон (береги Дніпра, Батьківщина-Мати).

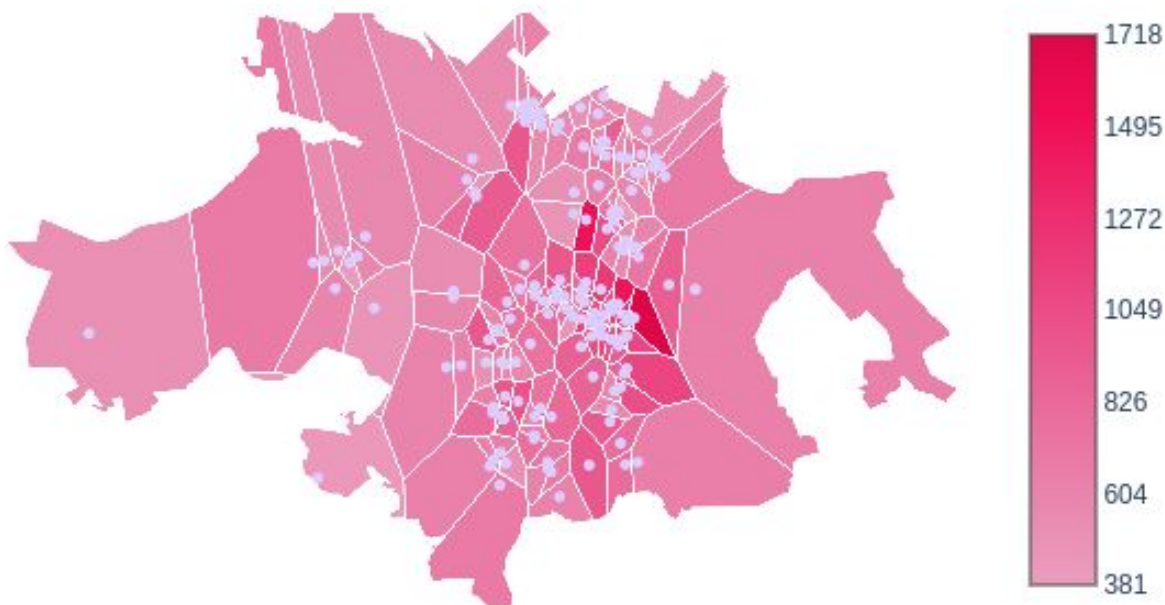


Рис. 5. декомпозиція вторинного житла Дніпра через FCM на 100 кластерів.

Кластеризацію вторинного житла візьмемо на прикладі міста Дніпро. Дана картограма підводить нас до наступних висновків:

- На відміну від столиці, через коротку протяжність гілки міського метрополітену, даний вид транспорту не є вирішальним формувальником вартості житла;
- Достатньо передбачувано, що район міста з його масивними ТРК (MOST-City, Passage та Grand Plaza), історико-культурною областю (Менора, Монастирський острів) є вирішальними ознаками при формуванні ціни **1200-1800\$** за 1 м².

- На північ й на південь від центральних районів міста смугами розташовуються переважно спальні райони, де ціна за одиницю площі помешкання тримається на рівні **300-800\$**, причиною чого є велика кількість індустріальних забудов, промислових зон та робітничих цехів. Екологічна ситуація поблизу підприємств навряд чи може бути названа задовільною, тож цілком резонним є те, що близькість заводів не сприяє, якщо не погіршує, привабливості апартаментів на вищезгаданих територіях.

Висновки

1. Було спроектовано й імплементовано інструмент декомпозиції територій міст на самодостатні зони подібної за розміщенням і ціновою політикою нерухомості, що дозволяє прогнозувати вартість житла в межах тих чи інших полігонів на карті населених пунктів. Самі по собі міста є дуже великими утвореннями, тож врахувати всі ціноутворюючі фактори надзвичайно важко, але якщо виокремити з-поміж всієї маси житлових масивів ті угруповання, котрі фізично й інфраструктурно є близькими, ми отримуємо ділянки, котрі містять в собі ціноформувальні закономірності й фактори, а це, у свою чергу, дозволить автоматизовано проводити ріелторську оцінку житла й аналіз територій для майбутніх новобудов.
2. З-поміж великої кількості особливостей, котрі варто враховувати при оцінці тієї чи іншої квартири, вирішальну роль грає розміщення - деяка абстрактна величина суб'єктивної якісної оцінки помешкання, котра є по суті дуже мінливим набором інших, більш матеріальних і вимірюваних ознак. Серед числа таких ми можемо виділити близькість зупинок швидкісного транспорту (метрополітен у Києві та Харкові), наближеність крупних водойм (береги Дніпра / Чорноморське узбережжя / великі озера, затоки і т.д.), досяжність бізнес-центрів, а також історико-культурних і паркових зон (особливо актуально для туристичних міст, таких як Львів) .
3. Дослідження охоплює достатньо активну й стрімку галузь національної економіки, тож резонною є ніша для подальших досліджень. Сюди ми відносимо такі напрямки, як:
 - Пошук нових істотних ознак, котрі кількісно формують якісне розміщення, оточення й інфраструктуру житлових масивів;

- Побудова часових рядів цінових зон, що дозволить прогнозувати “мутації” вже наявних і появу нових сегментів житла;
- Застосування *reinforced learning* / *neural networks* для більш точної оцінки й розрахунку вартості помешкань;
- Застосування наших досліджень не лише для квартир, а й для інших видів нерухомості - приватних будинків, офісів, земельних ділянок і т.д.

4. Реалізовано засіб групування житла за інфраструктурними ознаками й розміщенням на самодостатні зони, котрі дозволяють прогнозувати вартість невідомих квартир. Прикладом якості оцінки є показник r^2 для формули розрахунку вартості житлової площі, котрий для **65%** кластерів сягає значень **0.7-0.85** і вище, що є непоганою точністю для інструментів класичного навчання. Даний додаток може бути особливо цінним для ріелторських агентств, адже автоматизує сам процес пошуку й приблизної оцінки подібної до цільової нерухомості.

Джерела

1. Наталія Задерей, Олена Артеменко. Квартирне питання: в якому стані ринок нерухомості і чого очікувати від цін.
<https://voxukraine.org/uk/metr-zmin-v-yakomu-stani-rinok-neruhomosti-i-chogo-ochikuvati-vid-tsin/>
2. 100realty.ua. Аналіз цін на вторинному ринку житлової нерухомості Києва: жовтень 2019. <https://100realty.ua/uk/analytics/52716>
3. blagovist.ua. Статистика цін на квартири в Києві за 2019-10.
<https://blagovist.ua/uk/realtystat/2019-10>
4. map.olimp.net.ua. Статистика цін на житло. <https://map.olimp.net.ua/>
5. misto.lun.ua. Статистика ринку нерухомості у Львові за жовтень 2019.
<https://misto.lun.ua/price?city=113&date=10-2019>
6. misto.lun.ua. Карта комфорту Києва. <https://misto.lun.ua/map>
7. uk.wikipedia.org. Географічна відстань.
https://uk.wikipedia.org/wiki/%D0%93%D0%B5%D0%BE%D0%B3%D1%80%D0%B0%D1%84%D1%96%D1%87%D0%BD%D0%B0_%D0%B2%D1%96%D0%B4%D1%81%D1%82%D0%B0%D0%BD%D1%8C
8. home.deib.polimi.it. Clustering.
https://home.deib.polimi.it/matteucc/Clustering/tutorial_html/index.html
9. home.deib.polimi.it. Clustering - KMeans.
https://home.deib.polimi.it/matteucc/Clustering/tutorial_html/kmeans.html
10. home.deib.polimi.it. Clustering - Fuzzy C-means.
https://home.deib.polimi.it/matteucc/Clustering/tutorial_html/cmeans.html