

ВИКОРИСТАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ ОЦІНКИ ВАРТОСТІ ЖИТЛА

Вступ. Сучасні системи прийняття рішень у більшості випадків пов'язані з обробкою нечіткої інформації. Завдяки зростанню потужності обчислювальної техніки, з'явилися програмні засоби та бібліотеки функцій, які реалізують операції обробки над нечіткими множинами. Потужні бібліотеки методів машинного навчання реалізовані в мові програмування Python, що дозволяє розробляти та впроваджувати алгоритми нечіткого виводу.

Метою даної роботи – застосування методів машинного навчання для оцінки вартості житла у місті Києві. На сьогоднішній день для оцінки вартості житла використовується порівняльний підхід [1, 2]. Але складність підбору об'єктів аналогів для парної оцінки обмежує область застосування методу. Альтернативними методами є статистичні методи, зокрема, економетричні методи [3, 4], методи дисперсійного аналізу [5], методи нечіткої математики [6], а також нейромережеві методи [7].

У даній роботі розглянуто застосування методів нечіткої математики для оцінки вартості житла.

Методи нечітких множин особливо корисні за відсутності точної математичної моделі функціонування системи. Теорія нечітких множин дає можливість застосувати для прийняття рішень неточні та суб'єктивні експертні знання про предметну область без формалізації їх у вигляді традиційних математичних моделей. З використанням теорії нечітких множин вирішуються питання узгодження суперечливих критеріїв прийняття рішень.

Система нечіткого виводу – це нелінійне відображення, яке перетворює дані з чітких входів у чіткі результати виведення. Різні алгоритми нечіткого виводу (Мамдані, Цукамото, Ларсена, Сугено) відрізняються способом отримання чіткого виводу. Незалежно від виду алгоритму всі вони містять такі етапи (рис. 1):

а) фазифікації. В контексті нечіткої логіки під фазифікацією розуміють не тільки окремий етап виконання нечіткого виведення, а й власне процес або процедуру знаходження значень функцій належності;

У даній роботі розглянуто застосування методів нечіткої математики для оцінки вартості житла у місті Києві. Для системи нечіткого виводу застосовано алгоритм Мамдані. Виконана програмна реалізація нечіткої моделі на мові програмування Python з використанням бібліотеки sklearn. Обчислена середня відносна похибка розробленої моделі.

Ключові слова: нечітка логіка, машинне навчання, програмування на Пайтоні, лінгвістичні змінні, прогнозна модель.

б) нечітких множин (термів) на основі звичайних (не нечітких) вихідних даних. Фазифікацію ще називають введенням нечіткості. Мета етапу фазифікації – встановлення відповідності між конкретним (зазвичай – чисельним) значенням окремої вхідної змінної системи нечіткого виведення і значенням функції належності відповідного їй терма вхідної лінгвістичної змінної. Після завершення цього етапу для всіх вхідних змінних мають бути визначені конкретні значення функцій належності по кожному з лінгвістичних термів, які використовуються в передумовах бази правил системи нечіткого виведення;

в) агрегація – це процедура визначення ступеня істинності умов кожного з правил системи нечіткого виведення;

г) активізація в системах нечіткого виведення – процедура або процес знаходження ступеня істинності кожного з правил нечіткого виводу;

д) акумуляція в системах нечіткого виведення – процедура або процес знаходження функції належності для кожної з вихідних лінгвістичних змінних. Мета акумуляції полягає у тому, щоб об'єднати або акумулювати всі ступені істинності висновків для отримання функції належності кожної з вихідних змінних. Причина необхідності виконання цього етапу полягає у тому, що висновки, які відносяться до однієї і тієї ж вихідної лінгвістичної змінної, належать різним правилам системи нечіткого виводу.

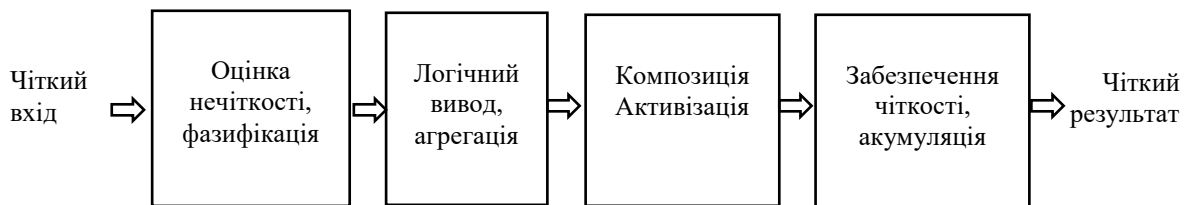


РИС. 1. Загальна схема нечіткого виводу

У даній роботі для системи нечіткого виводу застосовано алгоритм Мамдані. Даний алгоритм описано, зокрема, в [1].

Дані для моделювання взяті за період липень – жовтень 2020 року з єдиної бази звітів про оцінку майна [8]. Вибірка містить 2133 записи, яка поділена на навчальну та тестову в пропорції 85:15, тобто навчальна вибірка містить 1813 записів, тренувальна – 320. Далі показана частина таблиці з даними.

ТАБЛИЦЯ

Вид об'єкта оцінки	Дата оцінки	Дата реєстрації звіту	Оціночна вартість об'єкта оцінки, грн.
Об'єкт житлової нерухомості	25.09.2020	25.09.2020	77500
Об'єкт житлової нерухомості	25.09.2020	25.09.2020	413275
Об'єкт житлової нерухомості	25.09.2020	25.09.2020	224000
Об'єкт житлової нерухомості	25.09.2020	25.09.2020	22000

Таблиця містить такі колонки: Оціночна вартість об'єкта оцінки, грн; Статус звіту; Регіон об'єкта оцінювання; Населений пункт об'єкта оцінювання; Тип вулиці; Вулиця; Зона населеного пункту; Вид об'єкта нерухомості; Тип будинку; Клас будинку; Група капітальності; Рік введення в експлуатацію; Загальна площа, кв. м; Площа житлових приміщень, кв. м; Площа допоміжних/підсобних приміщень, кв. м; Поверх у будівлі (для квартири)/Поверховість будівлі; Інженерне обладнання; Фізичний стан; Наявність вбудованого паркінгу.

Оскільки робота присвячена оцінці вартості житла в Києві, то дані відфільтровані за регіоном об'єкта. Також до уваги були взяті тільки звіти зі статусом «Перевірено». Результуючою змінною буде Оціночна вартість об'єкта оцінки.

Програмна реалізація моделі була виконана на мові програмування Python з використанням бібліотеки sklearn (безкоштовна програмна бібліотека машинного навчання для мови програмування Python, яка надає функціональність для створення та тренування різноманітних алгоритмів класифікації, регресії та кластеризації, таких як: лінійна регресія, random forest, градієнтний бустинг). Для визначення лінгвістичних змінних використано клас Antecedent, для визначення логічного висновку – клас Consequent. Для визначення терм-множини та функцій належності використано функцію trapezmf для побудови функції трапеції. Для побудови правил використано клас Rule. Для побудови логічного виводу використано клас ControlSystem.

Реалізована система містить наступні параметри: S – площа житла; Z – зона населеного пункту; Y – рік здачі в експлуатацію; T – вид житла; P – вартість житла.

Створено 50 нечітких правил у термінах змінних виду:

П1: якщо S дуже мала, T гуртожиток, то P дуже низька;

П2: якщо S дуже мала, T гуртожиток, Z периферійна або серединна, то P дуже низька;

П3: якщо Y дуже старий будинок, S мала, то P низька;

П4: якщо Z центральна, S мала або середня, P середня;

Інші правила описують рік здачі в експлуатацію, вид житла, вартість житла тощо.

Побудовані функції належності $\mu(x)$ терм-множин вхідних та вихідних змінних (рис. 2 – 5).

$\mu(S)$: S – площа житла, кв. м, $S \in [10, 400]$.

Терм-множина: {дуже мала, мала, середня, велика, дуже велика}.

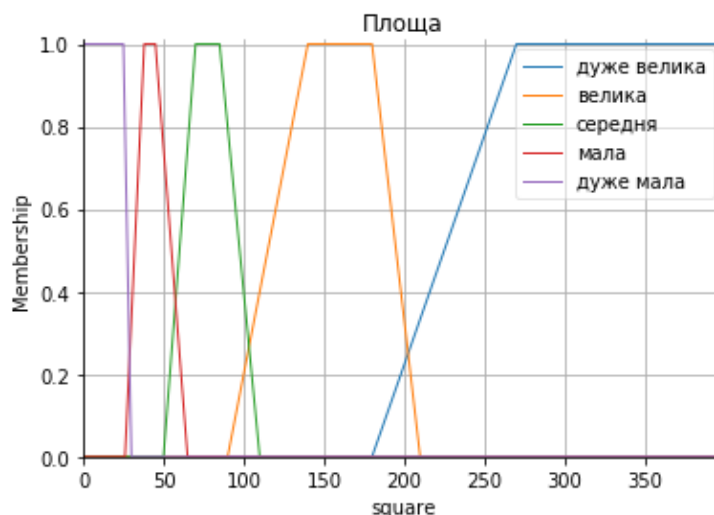


РИС. 2. Функція належності площі до терм-множин

$\mu(Z)$: Z – зона населеного пункту, $Z \in [0, 5]$.

Терм-множина: {поза містом (1), приміська (2), периферійна (3), серединна (4), центральна (5)}.

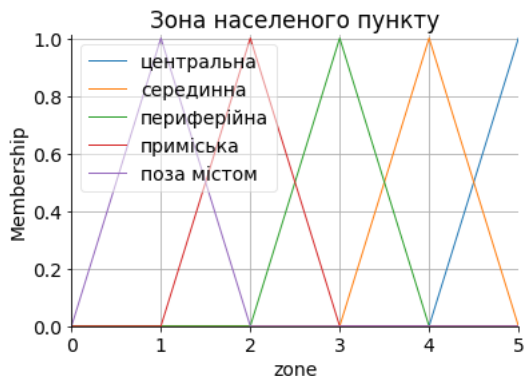


РИС. 3. Функція належності зони населеного пункту до терм-множин

$$\mu(Y): Y - \text{рік здачі в експлуатацію}, Y \in [1865, 2020].$$

Терм-множина: {дуже старий будинок, старий будинок, нормальний будинок, новий будинок, дуже новий будинок}.

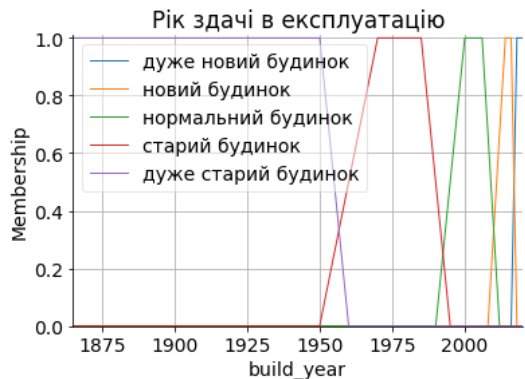


РИС. 4. Функція належності року здачі в експлуатацію до терм-множин

$$\mu(T): T - \text{вид житла}, T \in [0, 5].$$

Терм-множина: {гуртожиток (1), дачний будинок (2), квартира в багатоповерховому будинку (3), квартира в багатоповерховому будинку (4), індивідуальний будинок (5)}.

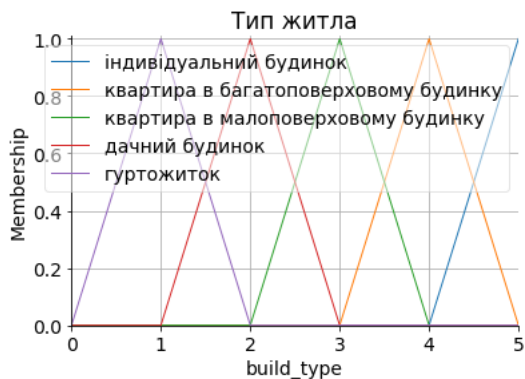


РИС. 5. Функція належності виду житла до терм-множин

$\mu(P)$: P – вартість житла. Одиниця виміру – тисячі гривень, $P \in [100, 10000]$.

Терм-множина: {дуже низька, низька, середня, висока, дуже висока, надзвичайно висока}.

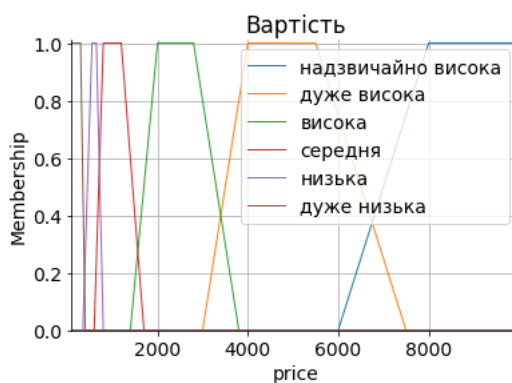


РИС. 6. Функція належності вартості житла до терм-множин

Програмна реалізація моделі на мові програмування Python розраховує коефіцієнти α_1 моделі, згідно алгоритму Мамдані, оцінює похибку моделі та дозволяє робити прогноз вартості житла на основі введених характеристик.

Середня відносна похибка розробленої моделі становить 4.22 %. На рис. 7 показано відносні похибки моделі для всіх записів тестової вибірки. Більшість похибок знаходяться в діапазоні від 0 до 50 %. Незначна кількість похибок перевищує 75 %.

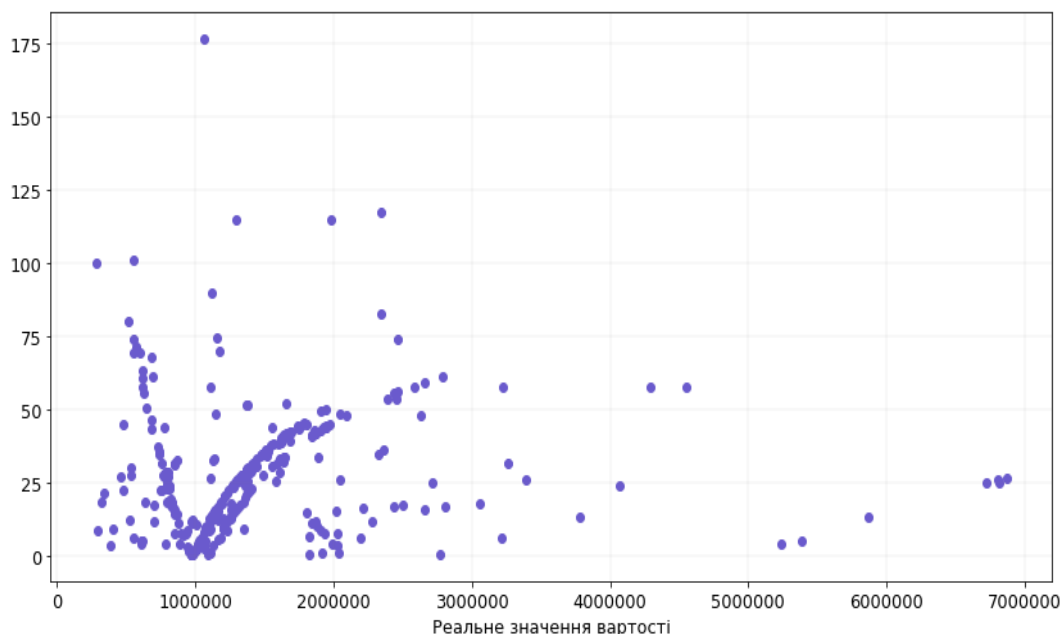


РИС. 7. Похибки апроксимації моделі, %

Висновок. У роботі побудована нечітка модель оцінки вартості житла у місті Києві, розроблена програма на мові Python, яка дозволяє прогнозувати вартість житла базуючись на основі аналізу бази даних попередніх продажів.

Список літератури

1. Горбаченко В.И., Ахметов Б.С., Кузнецова О.Ю. Интеллектуальные системы: нечеткие системы и сети: учебное пособие для вузов. 2-е изд. М.: Юрайт, 2019. 105 с.
2. Снитюк В.Е., Мирошник О.Н. Моделирование и прогнозирование процессов на рынке недвижимости. Черкассы. 2014. 332 с.
3. Васильева Г.Ю., Лютіков А.А., Маляр В.А. Методи і моделі оцінювання нерухомого майна. *Містобудування та територіальне планування: наук.-техн. зб.* Київ: нац. ун-т буд-ва і архіт.; відп. ред. М.М. Осетрін. Київ: КНУБА, 2017. Вип. 63. С. 52–55.
4. Сихимбаев М.Р., Кумисбекова Ж.А. Математические методы оценки стоимости недвижимого имущества. *Международный журнал прикладных и фундаментальных исследований.* 2015. **6-1**. С. 119–123.
5. Глухов Ю., Яцура А. Методика оцінки впливу якісних характеристик нерухомого майна на його вартість. *Галицький економічний вісник.* 2015. **2** (49). С. 160–169.
6. Renigier-Biżozor M., Janowski A., d'Amato M. Automated valuation model based on fuzzy and rough set theory for real estate market with insufficient source data. *Land Use Policy.* 2019. **87**. P. 147–154. <https://doi.org/10.1016/j.landusepol.2019.104021>
7. Yalpir S., Ozkan G. Knowledge-based FIS and ANFIS models development and comparison for residential real estate valuation. *International Journal of Strategic Property Management.* 2018. **22**. P. 110–118. <https://doi.org/10.3846/ijspm.2018.442>
8. Єдина база даних звітів про оцінку [Електронний ресурс] – Режим доступу до ресурсу: <http://www.spfu.gov.ua/ua/content/spf-estimate-basereport.html>

Одержано 02.02.2021

Третиник Віолета Вікентіївна,

кандидат фізико-математичних наук, доцент кафедри прикладної математики

НТУ України «КПІ ім. Ігоря Сікорського»,

<https://orcid.org/0000-0002-3538-8207>

viola.tret@gmail.com

Возняк Анастасія Тарасівна,

магістрант кафедри прикладної математики НТУ України «КПІ ім. Ігоря Сікорського»,

Домрачев Володимир Миколайович,

кандидат фізико-математичних наук, доцент кафедри прикладних інформаційних систем Київського національного університету імені Тараса Шевченка.

<https://orcid.org/0000-0002-0624-460X>

mipt@ukr.net

UDC 519.67

V. Tretynyk^{1*}, A. Voznyak¹, V. Domrachev²

Using Machine Learning Methods to Estimate the Cost of Housing

¹ The National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

² Taras Shevchenko National University of Kyiv

* Correspondence: viola.tret@gmail.com

Introduction. Nowadays, the state has enshrined at the legislative level the definition of appraised value for tax purposes in sales of real estate as mandatory. The comparative approach most often used by appraisers

has disadvantages such as the inability to find analogues in some cases and the need to make corrections, which affects the reliability of the results. The module of electronic determination of appraisal value (Module) similar to the object of property appraisal of the Unified database of appraisal reports works on the same approach and quite often overestimates appraisal value that leads to increase in the size of the tax during sales as the real estate cannot be sold for the price less than the estimated cost.

Today to determine the price of an automated system correctly, it is necessary to fill the Unified Valuation Database in the State Property Fund with large knowledge bases - a huge IT system. So far, the thoughtless machine still determines the price by the average value. Currently there are often situations when the appraised value of real estate, determined by the Module, exceeds its real market value. Given that the approach used by the Valuation Module does not always give the correct result, there is a need to find a better method to determine the value of housing that could be used by the Module.

The purpose of the paper. In this paper, an approach based on fuzzy logic was used to estimate the cost of housing in Kyiv. Fuzzy methods allow to apply a linguistic description of complex processes, to establish fuzzy relationships between concepts, to predict the behavior of the system, to create a set of alternative actions, to formally describe fuzzy decision-making rules.

Results. The software implementation of the model in Python programming language was performed. Data for modeling were taken for the period July – October 2020 from a single database of property valuation reports. The sample contained 2133 records, it was filtered, divided into training and testing in the proportion of 85 : 15. To assess the quality of the program, the average relative error of the developed model was calculated.

Keywords: fuzzy logic, machine learning, Python programming, linguistic variables, predictive model.