

## ЗАСТОСУВАННЯ РЕКУРЕНТНОЇ НЕЙРОННОЇ МЕРЕЖІ ДЛЯ ОЦІНКИ ВАРТОСТІ ОБ'ЄКТІВ НЕРУХОМОСТІ

*О. С. Тимчук, кандидат технічних наук, доцент*

*E-mail: [oleh.tymchuk@knu.ua](mailto:oleh.tymchuk@knu.ua)*

*А. І. Пилипенко, кандидат технічних наук, доцент*

*E-mail: [anna.pylypenko@knu.ua](mailto:anna.pylypenko@knu.ua)*

*А. В. Кича, магістрант кафедри програмних систем і технологій*

*E-mail: [a.kicha@knu.ua](mailto:a.kicha@knu.ua)*

*Київський національний університет імені Тараса Шевченка*

**Анотація.** Розглянуто проблему оцінки вартості об'єктів нерухомості. Автоматизація вирішення зазначеної проблеми є одним із способів надання об'єктивної оцінки, в якому виключені такі суб'єктивні фактори як помилка арифметичного розрахунку, оцінка під впливом емоцій, оцінка під впливом переслідування власних цілей. Найчастіше для вирішення задачі оцінки вартості об'єктів нерухомості використовується регресійний аналіз (гедонічна регресія) та методи машинного навчання. Метою цього дослідження є побудова моделі для оцінки вартості об'єктів нерухомості не лише на звичних кількісних показниках (наприклад, площа, кількість кімнат, поверх тощо), а й на базі текстового опису об'єкта нерухомості. У статті як об'єкти нерухомості розглядаються квартири в новозбудованих житлових комплексах м. Києва. Для досягнення мети розроблено математичну модель для класифікації класу житла новобудови на основі текстового опису за допомогою рекурентної нейронної мережі. Модель розроблена як багатошарова нейронна мережа прямого зв'язку, яка приймає текстові дані опису житлового комплексу (новобудови) та пропускає їх через серію прихованих шарів, де кожен шар складається з нейронів. Змодельована категоріальна змінна 'predictedClass' була використана як незалежна змінна лінійної регресії для розрахунку вартості квартири в новобудові. У порівнянні з класичною лінійною регресією з кількісними регресорами, нова модель забезпечила високий показник  $R^2$  при мінімальній кількості змінних.

**Ключові слова:** *об'єкт нерухомості, категоріальна змінна, класифікація, рекурентна нейронна мережа, лінійна регресія*

**Актуальність.** Оцінка вартості об'єктів нерухомості – важлива задача для суспільства, яка щодня вирішується не тільки власниками та інвесторами нерухомості, але й банками, страховими компаніями, аналітичними центрами тощо. Зростання ринку нерухомості сигналізує про зростання попиту у цьому секторі як

від бізнесу, який шукає нові будівлі для розміщення офісів, так і від фізичних осіб, які шукають житло для себе та для своїх сімей. Так, наприклад, за даними сайту Grand View Research ринок нерухомості стрімко розвивається впродовж останніх десятиліть та вартує більш трьох трильйонів доларів США, що становить близько п'яти відсотків від вартості загальної світової економіки.

Підвищення точності оцінки об'єктів нерухомості – гаряча тема для дослідників останнього десятиліття, оскільки своєчасна ефективна оцінка вартості дає значну економічну вигоду. Методи об'єктивної оцінки об'єктів нерухомості набувають більшої актуальності зі зростанням попиту на нерухомість, оскільки допомагають зростаючій кількості інвесторів швидко та об'єктивно оцінити вартість об'єкту. Автоматизація оцінки вартості об'єктів нерухомості виключає можливість арифметичних помилок, спотворення оцінки через емоційний вплив та неналежне врахування власних інтересів при оцінці.

**Аналіз останніх досліджень та публікацій.** Оцінка вартості об'єктів нерухомості базується на двох підходах:

- прогнозування вартості на базі аналізу часових рядів;
- визначення вартості на базі аналізу поточного стану ринку нерухомості.

Наявні моделі та методи оцінки вартості об'єктів нерухомості умовно можна розділити на дві категорії:

- моделі і методи кількісного аналізу;
- моделі і методи якісного аналізу.

При кількісному аналізі, який базується на аналізі числових ознак (наприклад, площа, поверх, кількість поверхів, часовий ряд продажів об'єктів нерухомості тощо), найчастіше використовується регресійний аналіз (гедонічна регресія, просторова регресія), рідше – методи теорії ігор, Байєсовські методи. Базова гедонічна регресійна модель [1] має ряд недоліків, які можуть бути успішно усунуті такими методами машинного навчання:

- моделі часових рядів (ARIMA, SARIMAX та інші) [2]. Вартість об'єктів нерухомості постійно змінюється в часі, тому часовий фактор має бути врахований під час моделювання. Також часові ряди є окремим випадком панельних даних.

Панельні дані за одними й тими одиницями нерухомості за різні проміжки часу дозволяють покращити висновок про причинність купівлі нерухомості порівняно з випадком, якби був доступний лише один зріз даних;

- штучні нейронні мережі (ANN) [3]. Ці моделі добре себе зарекомендували і дають більш точний прогноз порівняно з гедонічною регресією. Однак методами ANN складно оцінити такий рівень інформації, як вплив зовнішнього вигляду нерухомості на його ціну. Для усунення недоліків ANN можуть бути використані згорткові нейронні мережі (CNN) та рекурентні нейронні мережі (RNN). Мережі довготривалої пам'яті (LSTM) добре підходять для обробки часових рядів, таких як зміни цін на нерухомість;

- дерева рішень (Decision Tree Algorithm) [4]. На основі алгоритму дерева рішень можна побудувати алгоритм випадкового лісу. Алгоритм випадкового лісу також є типом ансамблевого алгоритму. Оскільки вибірки алгоритму випадкового лісу є випадковими, він має меншу ймовірність перенавчання, ніж традиційні алгоритми дерева рішень, а також більш точний, коли розмір вибірки великий. Це добре підходить для оцінки вартості нерухомості, де існує велика кількість вибірок та характеристик, і тому цей підхід може бути використаний для прогнозування продажів на основі даних про продаж нерухомості за попередні періоди;

- механізм уваги (Attention Mechanism) [5]. Спостерігаючи за об'єктами, люди схильні концентрувати свою увагу на тому, що становить більший інтерес. З розвитком досліджень було запропоновано додати до машинного навчання механізм «Attention», який можна розглядати як механізм перерозподілу ресурсів на основі важливості активації. Attention аналізує вхідний вміст, щоб визначити кореляцію між елементами, фіксує та посилює менш помітні, але важливіші характеристики, щоб збільшити їхню вагу впливу на навчання, і розподіляє більше обчислювальних ресурсів для більш значущих обчислювальних одиниць. Механізм Attention можна успішно застосовувати і до оцінки нерухомості.

- Нечіткі моделі та моделі, на базі обчислень зі словами (Fuzzy Logic, Computing with words) [6]. Під час оцінки інвестицій у нерухомість важливу роль відіграють соціальні та політичні чинники, які описуються словами. Відомо, що

одній й ті самі слова для різних людей означають різні речі, звідси виникає лінгвістична невизначеність у таких ознаках при описі об'єкта нерухомості. При цьому залучення більшої кількості експертів для опису таких параметрів об'єкта нерухомості не призводить до зниження рівня невизначеності. Також слід зазначити, що фактори мінливі у часі, оскільки світ активно розвивається і змінюється. Відсутність урахування невизначеності у вихідних даних призводитиме відповідно до появи невизначеності при прийнятті рішень. Тому інтеграція методів теорії нечітких множин, обчислень зі словами з методами машинного навчання для побудови адаптивних систем оцінки вартості нерухомості часто застосовується до оцінки нерухомості.

При оцінці вартості нерухомості або інвестицій у нерухомість, значну роль (у деяких випадках – основну) відіграють не кількісні ознаки, а категоріальні (наприклад, стан об'єкта нерухомості: «поганий», «задовільний», «добрий», клас об'єкта нерухомості). Як наслідок, використання кількісного аналізу для оцінки вартості об'єктів нерухомості стає дуже обмеженим. Моделі та методи якісного аналізу активно розвиваються останніми роками. Найбільш популярний напрямок серед дослідників – глибинне навчання.

**Мета дослідження** – розробити модель для оцінки вартості об'єктів нерухомості для м. Києва із застосуванням методів машинного навчання, а саме нейронної класифікації нерухомості за класом житла на базі текстового опису та лінійної регресії.

**Матеріали і методи дослідження.** Для оцінки вартості об'єктів нерухомості було використано pipeline, що складається з трьох етапів: попередня обробка, витягування ознак та моделювання.

Етап 1 – попередня обробка. Для побудови набору даних було використано дані про об'єкти нерухомості (квартири в новобудовах) міста Києва, що надає популярна платформа lun.ua. Опис змінних в сформованому наборі даних представлено у таблиці 1.

## 1. Опис набору даних для моделювання

<i>Назва змінної</i>	<i>Опис змісту змінної</i>
<i>Residence</i>	Назва житлового комплексу, в якому розташована квартира
<i>Developer</i>	Забудовник житлового комплексу
<i>Description</i>	Опис житлового комплексу від забудовника
<i>Address</i>	Адреса житлового комплексу
<i>District</i>	Район, в якому знаходиться житловий комплекс
<i>CeilingHeight</i>	Висота стель в квартирі
<i>NumApartmentsTotal</i>	Кількість квартир в житловому комплексі
<i>Year</i>	Рік закінчення будівництва житлового комплексу та початку вводу в експлуатацію
<i>PricePerSquareMeter</i>	Ціна за квадратний метр
<i>price</i>	Загальна ціна квартири
<i>area</i>	Площа квартири в квадратних метрах
<i>rooms</i>	Кількість кімнат у квартирі (на lun.ua кількість кімнат часто відповідає саме кількості спальних кімнат)
<i>floor</i>	Поверх квартири
<i>minutesToCityCenter</i>	Кількість хвилин, що займе піша прогулянка до центру Києва
<i>assess</i>	Середня вартість за квадратний метр нерухомості в тому самому районі Києва, де знаходиться дана квартира
<i>crimeRateInDistrict</i>	Показник кримінальної активності в тому самому районі Києва, де знаходиться дана квартира

Загалом, сформований набір даних містить 169 унікальних записів і 17 характеристик (змінних) різних типів. Розглянуті дані стосуються об'єктів нерухомості, в яких рік закінчення будівництва та початку вводу в експлуатацію задається періодом з 2022 по 2025 роки. Середня ціна за квадратний метр становить більше 100 тисяч гривень ( $1.007e+05$ ), розмір квартир в середньому становить більше 80 квадратних метрів ( $8.289e+01$ ), до центру міста можна доїхати в середньому менш ніж за 100 хвилин ( $9.141e+01$ ). Кількість унікальних значень для категоріальних даних різна. Наприклад, для змінної 'district' маємо лише 5 різних значень, тоді як для змінної 'description' – 11. Описова статистика кількісних та категоріальних даних надана у таблицях 2(a) і 2(b).

Етап 2 – витягування ознак. Ключовим фактором, що впливає на вартість квадратного метра новобудови, є клас житла. Щоб об'єктивно визначити клас житла, необхідно спиратися на різні ознаки житлового комплексу. Однак, незрозумілість критеріїв для віднесення об'єкта нерухомості до певного класу може призвести до завищення або недооцінки його реальної вартості. Для вирішення цієї проблеми пропонується використати рекурентну нейронну мережу (RNN).

**2(a). Описова статистика кількісних змінних**

	mean	Std	Min	25%	50%	75%	max
ceilingHeight	2.443e+00	4.983e-01	2.0	2.00	2.0	3.0	3.0
numApartmentsTotal	1.731e+03	1.625e+03	125.0	328.00	1709.0	2857.0	5183.0
Year	2.023e+03	7.178e-01	2022.0	2023.00	2023.0	2023.5	2025.0
pricePerSquareMeter	1.007e+05	6.927e+04	34700.0	58987.50	69725.0	116825.0	317600.0
Price	9.366e+06	1.257e+07	1232500.0	3264581.25	4888850.0	9024400.0	79400000.0
Area	8.289e+01	5.303e+01	8.0	53.00	76.0	98.0	556.0
Rooms	1.852e+00	8.907e-01	1.0	1.00	2.0	2.0	5.0
Floor	1.179e+01	6.920e+00	2.0	7.00	11.0	14.0	25.5
minutesToCityCenter	9.141e+01	5.841e+01	5.0	40.00	109.0	156.0	169.0
Assess	6.915e+04	3.892e+04	34392.0	47299.00	52536.0	53927.0	136750.0
crimeRateInDistrict	2.720e+03	6.002e+02	2106.0	2317.00	2537.0	2884.0	3939.0

**2(b). Описова статистика категоріальних змінних**

	count	unique	Top	Freq
residence	169	10	ЖК Метрополіс	31
developer	169	6	DIM	71
description	169	11	The new "Metropolis" residential complex in Ky...	31
address	169	10	вул. Академіка Заболотного, 1	31
district	169	5	Голосіївський р-н	70

На вхід цій мережі подається текстовий опис об'єкта нерухомості (змінна 'description'), перекладений на англійську мову за допомогою Google Cloud Translation API. Перекладений текстовий опис на англійську мову допомагає забезпечити однорідність та зрозумілість вхідних даних для моделі.

Композиція моделі складається з трьох шарів. Перший шар використовує nnlm-en-dim50/2 – словникове вбудування, створене за допомогою нейронної мережі, що відображає слова англійською мовою у векторний простір розмірністю 50. Другий та третій шари є прихованими шарами з 16 нейронами в кожному, що використовують функцію активації ReLU. Останній шар складається з 4 нейронів, що відповідають класам житла (економ, комфорт, бізнес, преміум), і використовує функцію активації softmax для мультикласової класифікації.

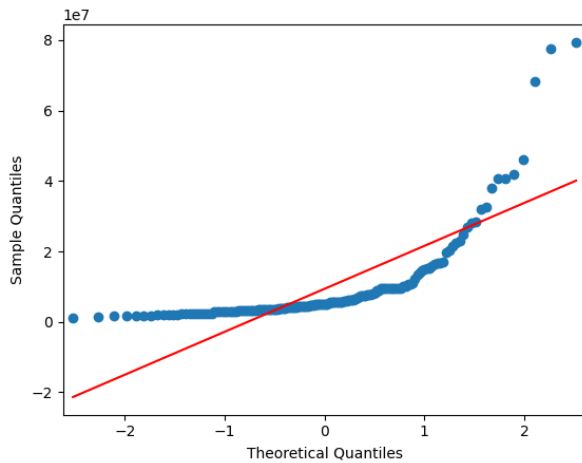
Таким чином, на основі текстового опису об'єкта нерухомості нейронною мережею створюється нова категоріальна змінна 'predictedClass' – клас житлового комплексу, в якому знаходиться квартира. Змінна 'predictedClass' може бути передана на вхід моделі лінійної регресії для розрахунку вартості квартири в новобудові.

Етап 3 – моделювання. У цій роботі для моделювання вартості об'єкта нерухомості використовується лінійна регресія, в якій регресором є нова категоріальна змінна 'predictedClass':

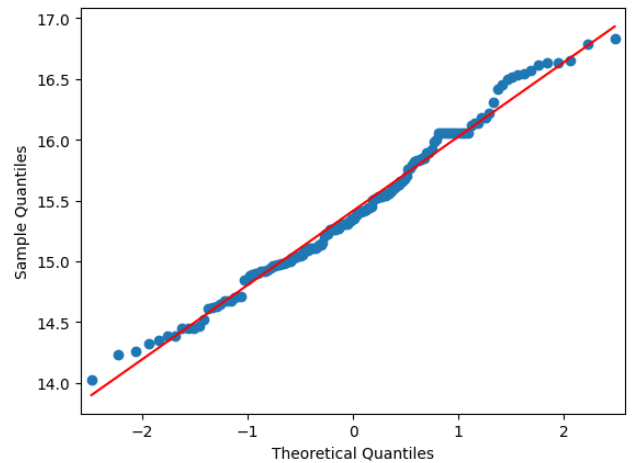
$$\log(\text{price}) = \beta_0 + \beta_1 \cdot \text{predictedClass}[T.\text{комфорт}] + \beta_2 \cdot \text{predictedClass}[T.\text{преміум}] + \beta_3 \cdot \log(\text{area}) + u \quad (1)$$

де  $\beta_0$  – постійний член регресії (intercept parameter);  $\beta_1, \beta_2$  – коефіцієнти регресії, які вказують, як змінюється логарифм ціни при зміні класу 'predictedClass' на «комфорт» та «преміум» відповідно (slope parameters). Значення «економ» є базовим рівнем, порівняння проводиться саме з ним, тому з метою уникнення мультиколінеарності немає необхідності включати його окремо в рівняння регресії;  $\beta_3$  – еластичність ціни відносно площі квартири;  $u$  – помилка регресії (disturbance), яка містить неспостережувані фактори, які впливають на ціну нерухомості.

**Результати досліджень та їх обговорення.** Перед тим, як застосовувати лінійну регресію, необхідно перевірити передумови, щоб бути впевненим, що модель дійсно підходить для даних та її результати є надійними. Перш за все потрібно перевірити нормальність розподілу змінної відгуку ('price'). Для цього було побудовано Q-Q діаграму (квантиль-квантиль графік), а також застосовано критерій Шапіро-Вілка для тестування нормальності. Як видно на рисунку 1 (а) розподіл змінної 'price' не є нормальним, тому для досягнення нормальності застосовано логарифмування та видалення викидів. У результаті таких перетворень отримали нормальний розподіл (рис. 1(b)). Р-значення дорівнює 0.082, що означає, що нульова гіпотеза про нормальний розподіл не може бути відхилена на рівні значущості 0.05.



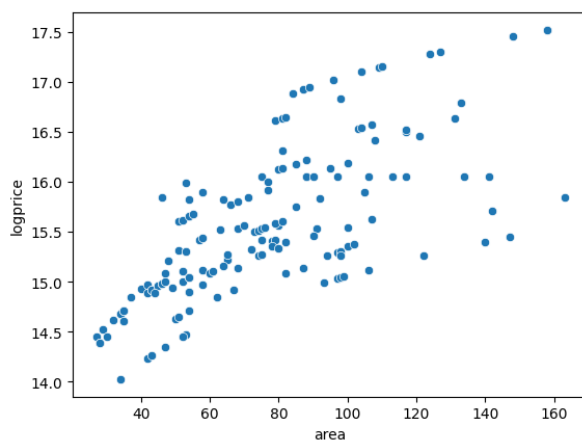
ShapiroResult (statistic=0.574,  
pvalue=3.121e-20)  
(a) початкові дані 'price'



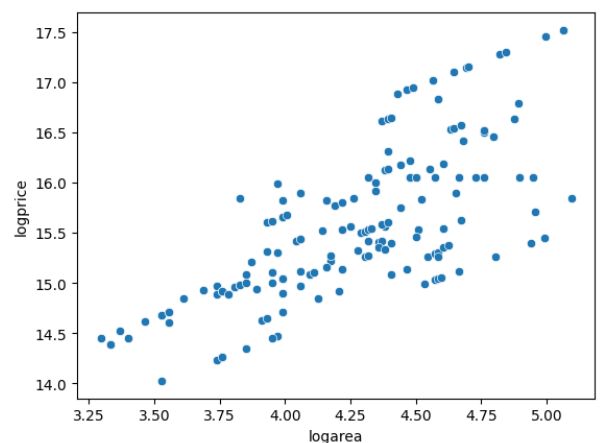
ShapiroResult (statistic=0.984,  
pvalue=0.082)  
(b) дані після перетворень 'logprice'

**Рис. 1. Q-Q діаграма змінної відгуку**

Далі необхідно перевірити чи є залежність між змінними 'logprice' та 'area' лінійною. Для цього використано діаграму розсіювання (рис. 2). Як видно на лівій частині рисунку 2(a), залежність не є лінійною, тому було застосовано логарифмування регресора 'area'. Логарифмування регресорів в лінійній регресії сприяє зменшенню нестабільності регресії, налагодженню лінійної залежності, поліпшенню нормальності розподілу. На рисунку 2(b) показано, що логарифмування регресора може допомогти побудувати лінійну залежність між залежною змінною та регресором.



(a) нелінійна залежність



(b) лінійна залежність

**Рис. 2. Діаграма розсіювання змінної відгуку та регресора**



З огляду обмеженого обсягу публікації подальші кроки перевірки передумов лінійної регресії та її побудови в цій статті не розкривається, розглянемо лише отриманий результат. Побудовано таке рівняння регресії:

$$\logprice = 647.4151 + 1.2723\logarea - 0.0806rooms - 0.0001crimeRateInDistrict - 0.0047/minutesToCityCenter - 0.3145 * year \quad (2)$$

	coef	std err	t	P> t	[0.025	0.975]
Intercept	647.4151	80.071	8.085	0.000	489.252	805.579
Logarea	1.2723	0.101	12.626	0.000	1.073	1.471
Rooms	-0.0806	0.041	-1.993	0.049	-0.161	8.21e-05
crimeRateInDistrict	-0.0001	5.02e-05	-2.441	0.016	-0.000	-2.34e-05
minutesToCityCenter	-0.0047	0.001	-9.319	0.000	-0.006	-0.004
Year	-0.3145	0.040	-7.940	0.000	-0.393	-0.236

Значення  $R^2$  (коефіцієнт детермінації) дорівнює 0.884, це доволі високий показник і свідчить про те, що модель (2) добре пояснює дані. Значення Adjusted  $R^2$  (скоригований коефіцієнт детермінації) дорівнює 0.880 і є трохи меншим за  $R^2$ . Це вказує на те, що додавання нових змінних до моделі може не принести значного покращення результатів. F-статистика дорівнює 237.0 (Prob (F-statistic) = 5.83e-71). Коефіцієнти перед усіма регресорами є статистично значущими на рівні значимості 0.05. Це свідчить про те, що модель (2) в цілому є статистично значущою.

Наступне рівняння регресії побудоване відповідно до моделі (1):

$$\logprice = 12.0690 - 0.6478predictedClass[T.комфорт] + 0.8667predictedClass[T.преміум] + 0.8859*logarea \quad (3)$$

	Coef	std err	T	P> t	[0.025	0.975]
Intercept	12.0690	0.228	53.019	0.000	11.619	12.519
predictedClass[T.комфорт]	-0.6478	0.045	-14.540	0.000	-0.736	-0.560
predictedClass[T.преміум]	0.8661	0.074	11.654	0.000	0.719	1.013
Logarea	0.8859	0.052	17.022	0.000	0.783	0.989

Отримане значення  $R^2$  вказує на те, що модель (3) виявилася добре придатною для опису даних, оскільки вона змогла пояснити близько 88.9 % дисперсії залежної змінної. Це можна підтвердити й за значенням F-статистики, яке дорівнює 420.2 (Prob (F-statistic) = 4.74e-75).

Загалом можна сказати, що обидві моделі (2) і (3) є значущими і добре пояснюють змінність залежної змінної. Однак, модель (3), яка використовує

змодельовану змінну 'predictedClass', має більш високий показник пояснювальної здатності. Крім того, вона має меншу кількість змінних та значимі коефіцієнти для кожної з них.

**Висновки і перспективи.** Отже, розроблено модель рекурентної нейронної мережі для класифікації класу житла новобудови на основі текстового опису. Модель розроблена як багатошарова нейронна мережа прямого зв'язку, яка приймає текстові дані опису житлового комплексу та пропускає їх через серію прихованих шарів, де кожен шар складається з нейронів. Змодельована категоріальна змінна 'predictedClass' була використана як незалежна змінна лінійної регресії для розрахунку вартості квартири в новобудові. Порівняно з класичною лінійною регресією з кількісними регресорами, нова модель забезпечила високий показник  $R^2$  при мінімальній кількості змінних. Подальші дослідження щодо оцінки вартості нерухомості за допомогою інтелектуального контролера на основі розробленої моделі слід проводити з урахуванням критеріїв, таких як кількість спостережень, вимоги до економічної адекватності, розподіл залишків, аналіз впливу викидів тощо.

#### Список використаних джерел

1. Sopranzetti, B.J. (2010). Hedonic Regression Analysis in Real Estate Markets: A Primer. In: Lee, CF., Lee, A.C., Lee, J. (eds) Handbook of Quantitative Finance and Risk Management. Springer, Boston, MA. [https://doi.org/10.1007/978-0-387-77117-5\\_78](https://doi.org/10.1007/978-0-387-77117-5_78)
2. Vishwakarma, Vijay Kumar, Forecasting Real Estate Business: Empirical Evidence from the Canadian Market (2013). Global Journal of Business Research, v. 7 (3) pp. 1-14, 2013, Available at SSRN: <https://ssrn.com/abstract=2148507>
3. Baur, Katharina & Rosenfelder, Markus & Lutz, Bernhard. (2023). Automated real estate valuation with machine learning models using property descriptions. Expert Systems with Applications. 213. 119147. 10.1016/j.eswa.2022.119147.
4. Fan, Gang-Zhi & Ong, Seow Eng & Koh, Hian. (2006). Determinants of House Price: A Decision Tree Approach. Urban Studies. 43. 2301-2316. 10.1080/00420980600990928.
5. Wanga, Pei-Ying & Chen, Chiao-Ting & Su, Jain-Wun & Ting-Yun, Wang & Huang, Szu-Hao. (2021). Deep Learning Model for House Price Prediction Using Heterogeneous Data Analysis Along With Joint Self-Attention Mechanism. IEEE Access. PP. 1-1. 10.1109/ACCESS.2021.3071306.
6. Tymchuk, O., Pylypenko, A., Iepik, M. Forecasting of Categorical Time Series Using Computing with Words Model (Conference Paper) 9th International Scientific Conference "Information Technology and Implementation" Workshop, IT and I-WS 2022; Kyiv; Ukraine; CEUR Workshop Proceedings. Volume 3384, 2022, Pages 151-159 URL:

<http://www.scopus.com/inward/record.url?eid=2-s2.0-85159621875&partnerID=MN8TOARS>

### References

1. Sopranzetti, B.J. (2010). Hedonic Regression Analysis in Real Estate Markets: A Primer. In: Lee, C.F., Lee, A.C., Lee, J. (eds) Handbook of Quantitative Finance and Risk Management. Springer, Boston, MA. [https://doi.org/10.1007/978-0-387-77117-5\\_78](https://doi.org/10.1007/978-0-387-77117-5_78)
2. Vishwakarma, Vijay Kumar, Forecasting Real Estate Business: Empirical Evidence from the Canadian Market (2013). Global Journal of Business Research, v. 7 (3) pp. 1-14, 2013, Available at SSRN: <https://ssrn.com/abstract=2148507>
3. Baur, Katharina & Rosenfelder, Markus & Lutz, Bernhard. (2023). Automated real estate valuation with machine learning models using property descriptions. Expert Systems with Applications. 213. 119147. 10.1016/j.eswa.2022.119147.
4. Fan, Gang-Zhi & Ong, Seow Eng & Koh, Hian. (2006). Determinants of House Price: A Decision Tree Approach. Urban Studies. 43. 2301-2316. 10.1080/00420980600990928.
5. Wang, Pei-Ying & Chen, Chiao-Ting & Su, Jain-Wun & Ting-Yun, Wang & Huang, Szu-Hao. (2021). Deep Learning Model for House Price Prediction Using Heterogeneous Data Analysis Along With Joint Self-Attention Mechanism. IEEE Access. PP. 1-1. 10.1109/ACCESS.2021.3071306.
6. Tymchuk, O., Pylypenko, A., Iepik, M. Forecasting of Categorical Time Series Using Computing with Words Model (Conference Paper) 9th International Scientific Conference "Information Technology and Implementation" Workshop, IT and I-WS 2022; Kyiv; Ukraine; CEUR Workshop Proceedings. Volume 3384, 2022, Pages 151-159 URL: <http://www.scopus.com/inward/record.url?eid=2-s2.0-85159621875&partnerID=MN8TOARS>

## A RECURRENT NEURAL NETWORK FOR FOR REAL ESTATE PRICE ESTIMATION

*O. Tymchuk, A. Pylypenko, A. Kicha*

**Abstract.** *The paper considers the problem of estimating the price of real estate. Automation of the solution to the specified problem is one of the methods of providing an objective assessment, which excludes such subjective factors as an arithmetic calculation error, assessment under the influence of emotions, assessment under the influence of the pursuit of one's own goals. Most often, regression analysis (hedonic regression) and machine learning methods are used to solve the problem of estimating the price of real estate. The purpose of this research is to build a model for estimating the price of real estate not only on the basis of the usual quantitative indicators (for example, area, number of rooms, floor, etc.), but also on the basis of the textual description of the real estate. In this paper we consider a real estate in new residential buildings in the Kyiv. To achieve the goal, a mathematical model was developed for the classification of real estate class based on a textual description using a recurrent neural network. The model is developed as a multi-layer feed-forward neural network that accepts textual data describing a new residential building and passes it through a series of hidden layers, where each layer consists of neurons. The simulated categorical variable 'predictedClass' was used as the*

*independent variable of the linear regression to calculate the price of an apartment in a new building. Compared to classical linear regression with quantitative regressors, the new model provided a high  $R^2$  with a minimal number of variables.*

**Key words:** *real estate, categorical data, classification, recurrent neural network, linear regression*