

# ІНТЕЛЕКТУАЛЬНИЙ МОДУЛЬ ВИЗНАЧЕННЯ ЦІНИ НА НЕ- РУХОМІСТЬ

Вінницький національний технічний університет;

## *Анотація*

*Запропоновано інтелектуальний модуль визначення ціни на нерухомість створений за допомогою методів машинного навчання, який допоможе автоматизувати процес визначення ціни на нерухомість та описані основні етапи задачі.*

**Ключові слова:** інтелектуальний модуль, визначення ціни, аналіз даних.

## *Abstract*

*The intelligent module for determining the price for houses created using machine learning methods is proposed to help automate the process of determining the price of real estate and the main stages of the task described.*

**Keywords:** intelligent module, determining the price, data analysis.

## **Вступ**

Сучасний ринок нерухомості є складним та динамічним середовищем, де ціни на нерухомість постійно коливаються під впливом численних факторів, таких як економічні умови, соціокультурні тенденції, регулююча політика та інші чинники. Прийняття ефективних та обґрунтованих рішень щодо цін на нерухомість стає важливим завданням для професіоналів у цій галузі, таких як ріелтори, інвестори, розробники та власники нерухомості.

Щоб допомогти вирішити цю складну проблему розглядається створення інтелектуального модуля прийняття рішень, який базується на машинному навчанні - це потужний інструмент інтелектуального аналізу даних, який дозволяє робити висновки та приймати рішення на основі великої кількості вхідних даних і правил.

Модуль буде заснований на методах машинного навчання та структурованих даних, що дозволить ефективно враховувати різноманітні змінні та взаємозв'язки між ними.

## **Результати дослідження**

Розробка інтелектуального модуля, який визначає ціни на нерухомість на основі аналізу різних факторів, що впливають на ринок нерухомості. Отже, вхідними даними програмного модуля є характеристики об'єкта нерухомості. Як вхідні дані було взято дані з датасету "House Prices - Advanced Regression Techniques"[1] за сайту Kaggle[2]. Всі дані що подаються на вхід моделі повинні пройти попередню обробку, очищення від пустих значень.

Прогнози цін на житло повинні бути точними. Це пов'язано з тим, що такі прогнози залежать від різних важливих характеристик, а їх точність визначається прогнозистами. Найважливішим фактором тут є використання різних регресійних методів.

Обираючи найкращий підхід для цієї роботи, ми враховуємо потребу в моделях, які можуть ефективно обробляти комбіновані дані та надавати точні та надійні прогнози цін на нерухомість. Тому надамо перевагу ансамблевим моделям, оскільки вони комбінують кілька моделей для отримання більш точного та стабільного прогнозу, зменшуючи вплив перенавчання та підвищуючи здатність до узагальнення на нові дані.

Серед найпоширеніших варіацій ансамблевих моделей можна виділити:

- Випадковий ліс[3]: Випадкові ліси ефективно працюють з числовими даними та категоріальними характеристиками, що робить їх ідеальними для прогнозування цін на нерухомість на основі характеристик житла.
- Градієнтний бустинг[4]: Цей підхід є ефективним при визначенні ціни на нерухомість, оскільки

дозволяє моделювати складні взаємозв'язки між об'єктами та забезпечує високу точність прогнозування.

- CatBoost: CatBoost - Алгоритм корисний у проектах з ціноутворення на нерухомість, оскільки він ефективно обробляє категоріальні дані та автоматично визначає їхню важливість для моделювання цінових тенденцій.

Поєднання моделей випадкового лісу, градієнтного бустингу та CatBoost є потужним підходом до вирішення проблеми визначення ціни на нерухомість.

Підхід злиття передбачає створення ансамблю, де кожна модель (Random Forest, Gradient Boost, CatBoost) додає свій унікальний внесок у прогноз. Наприклад, можуть використовуватися методи стекування, коли прогнози кожної моделі вводяться в метамодель, яка робить остаточний прогноз. Такий підхід дозволяє підсумовувати прогнози декількох моделей і підвищує точність прогнозів. Приклад стекування моделей зображено на рисунку 1.

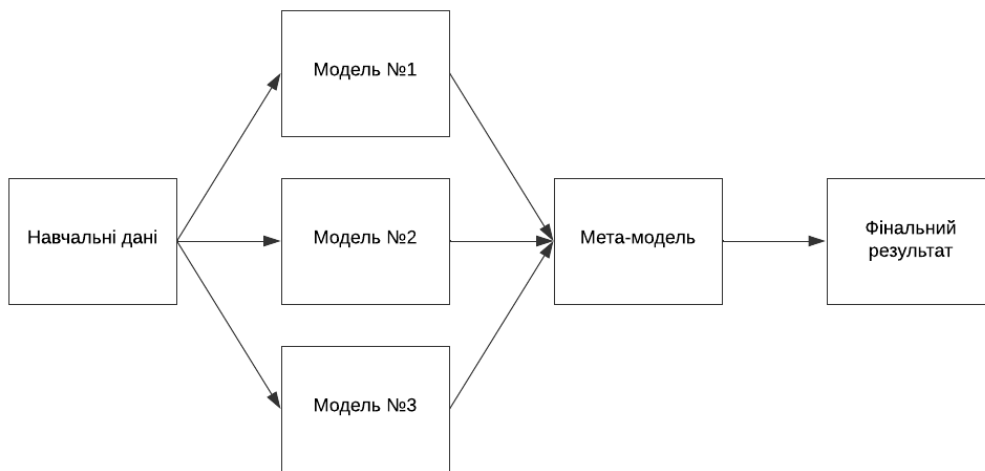


Рис. 1. Приклад стекування моделі

Було обрано середовище розробки Anaconda, мова програмування Python та бібліотеки scikit-learn, XGBoost для розробки модуля визначення ціни на нерухомість.

Результати навчання моделей зображені на рисунку 2.

	Model	R2_Score	Adjusted_R2_Score	MAE	MSE	RMSE
0	SVR	0.874823	0.753874	0.096051	0.018952	0.137665
1	RandomForestRegressor	0.866908	0.738312	0.098616	0.020150	0.141950
2	GradientBosstRegressor	0.879098	0.762280	0.097009	0.018304	0.135294
3	XGBosstRegressor	0.857981	0.720759	0.098525	0.021501	0.146634
4	CatBoostRegressor	0.900580	0.804520	0.082663	0.015052	0.122686

Рис. 1. Результати навчання моделей

З рис 2. можна побачити що найкращий результат мають моделі CatBoostRegressor і GradientBoostRegressor, вони і були використані для стекінгової моделі.

## Висновки

Дослідження показало що для досягнення високої точності прогнозів доцільно використовувати ансамблевий методи моделювання, зокрема Random Forest, Gradient Boosting та CatBoost. Ансамблевий підхід, що включає стекування моделей, дозволив комбінувати їхні переваги, зменшити ризик перенавчання та підвищити здатність до узагальнення.

Після аналізу результатів навчання моделей виявилось, що найкращі результати продемонстрували CatBoostRegressor та GradientBoostingRegressor. Вони й були використані для створення стекінгової моделі, яка інтегрує прогнози декількох моделей для отримання більш точного та надійного прогнозу.

Таким чином, використання ансамблевих методів у поєднанні з ефективними алгоритмами обробки категоріальних та числових даних дозволило створити потужний інструмент для прогнозування цін на нерухомість, що відповідає вимогам точності та надійності.

#### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. House Prices - Advanced Regression Techniques [Електронний ресурс]. Режим доступу: <https://www.kaggle.com/competitions/house-prices-advanced-regression-techniques/overview>
2. Kaggle [Електронний ресурс]. Режим доступу: <https://www.kaggle.com/>
3. Random Forest [Електронний ресурс]. Режим доступу: [https://uk.wikipedia.org/wiki/Random\\_forest](https://uk.wikipedia.org/wiki/Random_forest)
4. Gradient Boosting [Електронний ресурс]. Режим доступу: [https://en.wikipedia.org/wiki/Gradient\\_boosting](https://en.wikipedia.org/wiki/Gradient_boosting)

*Мельник Євгеній Олегович* — студент групи ЗКН-206, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, Вінниця, e-mail: euggenvschole@gmail.com

Науковий керівник: *Паночішин Юрій Миколайович* — к.т.н. доцент кафедри комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця

*Melnik Eugene O.* — Department of Intelligent Information Technology and Automation, Vinnytsia National Technical University, Vinnytsia, email : euggenvschole@gmail.com

Supervisor: *Panochyshyn Yurii M.* — Ph.D. (Eng.), docent of the Chair of Computer science, Vinnytsia National Technical University, Vinnytsia