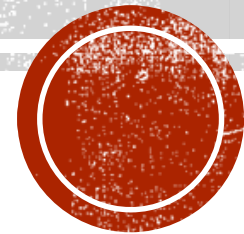


ПРОГНОЗНА ЦІНОУТВОРЮЮЧА МОДЕЛЬ ДЛЯ ТОВАРІВ З ХОЛОДНИМ СТАРТОМ

Величко Ростислав



АКТУАЛЬНІСТЬ ДОСЛІДЖЕННЯ

- У сучасному економічному середовищі зростає важливість розробки ефективних методів ціноутворення, зокрема для товарів з холодним стартом - продуктів, які не мають історії продажів. Прогнозна ціноутворююча модель, заснована на градієнтному бустингу, стає ключовою в умовах швидкозростаючого ринку та конкурентної боротьби.

Ключові виклики:

- Стара ціноутворююча модель
- Різний попит на різні товари та сегменти ринку



МЕТА ДОСЛІДЖЕННЯ

- Метою даного дослідження є розробка та дослідження прогностичної ціноутворюючої моделі для товарів з холодним стартом за допомогою моделей градієнтного бустингу.



МЕТОД ГРАДІЄНТНОГО БУСТИНГУ

- Модель, заснована на градієнтному бустингу, в контексті задачі прогнозування ціни, показує свою ефективність саме через те, що вона формує ціну на основі цін товарів, які обираються за рахунок найближчих сусідів до досліджуваного **SKU** (**stock keeping unit**).
- Найближчі сусіди до досліджуваного **SKU** визначаються за допомогою комбінації методу К-найближчих сусідів (**K-Nearest Neighbors, KNN**) та методів обробки природної мови (**NLP**)



МЕТОД ГРАДІЄНТНОГО БУСТІНГУ

- Ми маємо навчальний набір даних, який складається з n спостережень. Кожне спостереження містить інформацію про ціну товару y_i та його ознаки (назву та категорію товару). Також, нам дано певний **SKU**, для якого ми хочемо здійснити прогноз ціни.
- Навчальний набір позначається як (x_i, y_i) , де x_i - вектор ознак i -го товару, y_i - його ціна. Наша мета - побудувати модель $F(x)$, яка прогнозує ціну для нового **SKU** на основі ознак товару x .
- Модель градієнтного бустінгу буде апроксимувати функцію $F(x)$ шляхом додавання слабких моделей у вигляді ансамблю. Кожна наступна модель намагатиметься покращити попередні прогнози шляхом апроксимації залишкової функції.



МЕТОД ГРАДІЄНТНОГО БУСТИНГУ

- Апроксимація

$$F(x) \approx F_0(x) + \sum_{m=1}^M h_m(x),$$

$F_0(x)$ - це початкове базове значення, а $h_m(x)$ - вклад m -тої моделі.

При добавленні m -тої моделі, ми намагаємося мінімізувати наступну функцію втрат

$$L(y, F_{m-1}(x) + h(x)) = L(y - (F_{m-1}(x) + h(x)))^2,$$

де L - функція втрат (квадратична втрата), y - цільова змінна, $F_{m-1}(x)$ - прогноз, зроблений попередніми моделями, а $h(x)$ - нова модель, яку ми намагаємося навчити на цій ітерації.



МЕТОД ГРАДІЄНТНОГО БУСТИНГУ

- Після M ітерацій ми отримуємо модель у вигляді ансамблю

$$F_M(x) = F_0(x) + \sum_{m=1}^M h_m(x),$$

$F_0(x)$ - початкове базове значення (середнє значення ціни в навчальному наборі), а $h_m(x)$ - вклад m -тої моделі.

Отже, для прогнозу нової ціни для заданого SKU за допомогою цін його найближчих сусідів за допомогою GBDT, ми будемо використовувати дану модель $F_M(x)$, де x - ознаки нового товару.



ОЦІНКА МОДЕЛІ

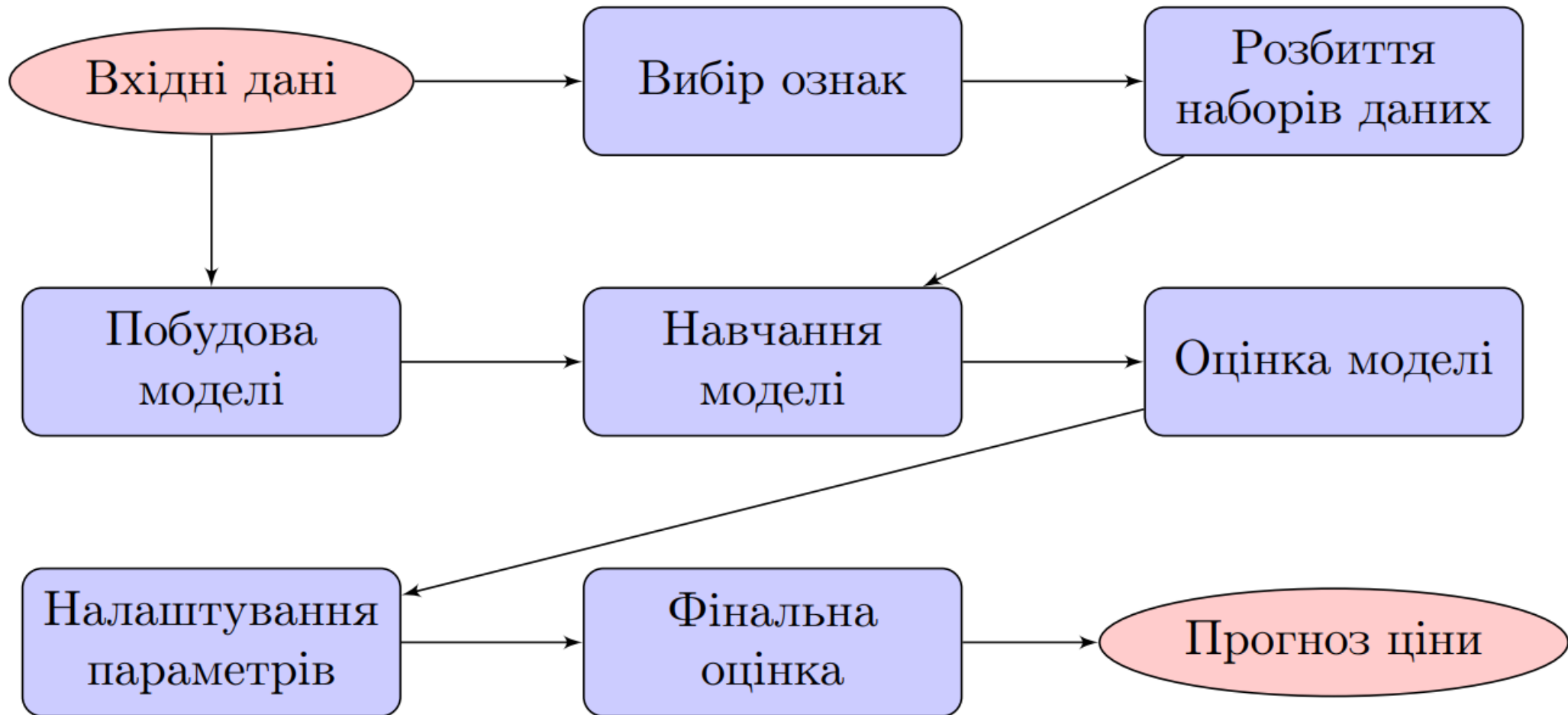
- Для оцінки нашої моделі використовуються **MSE** (середньоквадратична помилка) та **MAE** (середня абсолютна помилка).
- Середньоквадратична помилка (**MSE**) вимірює середньоквадратичну різницю між фактичними (y_i) та прогнозованими (\hat{y}_i) значеннями ціни товару для всіх n спостережень.
- Середня абсолютна помилка (**MAE**) вимірює середню абсолютну різницю між фактичними (y_i) та прогнозованими (\hat{y}_i) значеннями ціни товару для всіх n спостережень.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|.$$



БЛОК-СХЕМА МОДЕЛІ GBDT



ВИСНОВКИ

- Використання ансамблю слабких моделей дозволяє досягти високої точності прогнозів, а градієнтний спуск на кожній ітерації допомагає покращувати результати. Дана модель може бути корисним інструментом для вирішення задач ціноутворення та прогнозування попиту на товари.

