

Факультет комп'ютерних наук та кібернетики

Звіт
“Knowledge Discovery”

Виконав:
Студент 4 курсу
групи ТТП-41
Корнієнко Олександр Віталійович

Київ – 2025

ЗМІСТ

1. Класифікація категорій піци	3
2. Прогнозування денного доходу	4
3. Прогнозування кількості піц у замовленні	5
4. Висновки	7

1. Класифікація категорій піци

Мета та опис проєкту:

Було поставлено завдання розробити модель, яка на основі характеристик піци (розмір, ціна, список інгредієнтів) класифікує її за категоріями (Classic, Veggie, Supreme).

Методика:

- Попередня обробка даних: видалення пропусків, кодування розміру за допомогою One-Hot, векторизація інгредієнтів (TF-IDF).
- Побудова Pipeline із використанням Random Forest.
- Розділення даних: 80% – навчальна вибірка, 20% – тестова.

Результати:

- Точність (Accuracy) моделі становила приблизно 100%.
- Було отримано деталізований Classification Report із показниками Precision, Recall та F1-score для кожного класу.

Висновок:

Модель успішно класифікує категорії піци, причому найбільш коректно розпізнається категорія з 100% точністю, що, ймовірно, пов'язано з більшою кількістю прикладів у даних.

2. Прогнозування денного доходу

Мета та опис проєкту:

Було поставлено завдання розробити модель для прогнозування щоденного доходу піцерії на основі історичних даних замовлень.

Методика:

- Дані групувано за датою замовлення для отримання щоденного доходу.
- Дані перетворено у формат, необхідний для моделі Prophet (стовпці ds та y).
- Розділення даних: 80% – навчальна вибірка, 20% – тестова.
- Навчання моделі Prophet, що враховує сезонність та тренди.
- Оцінка точності 74.64%

Результати:



- Графік порівняння реального та прогнозованого доходу демонструє чітке розділення на навчальну та тестову вибірки.

Висновок:

Модель Prophet добре відтворює динаміку доходу та сезонні коливання, що робить її корисною для планування та аналізу бізнес-показників.

3. Прогнозування кількості піц у замовленні

Мета та опис проєкту:

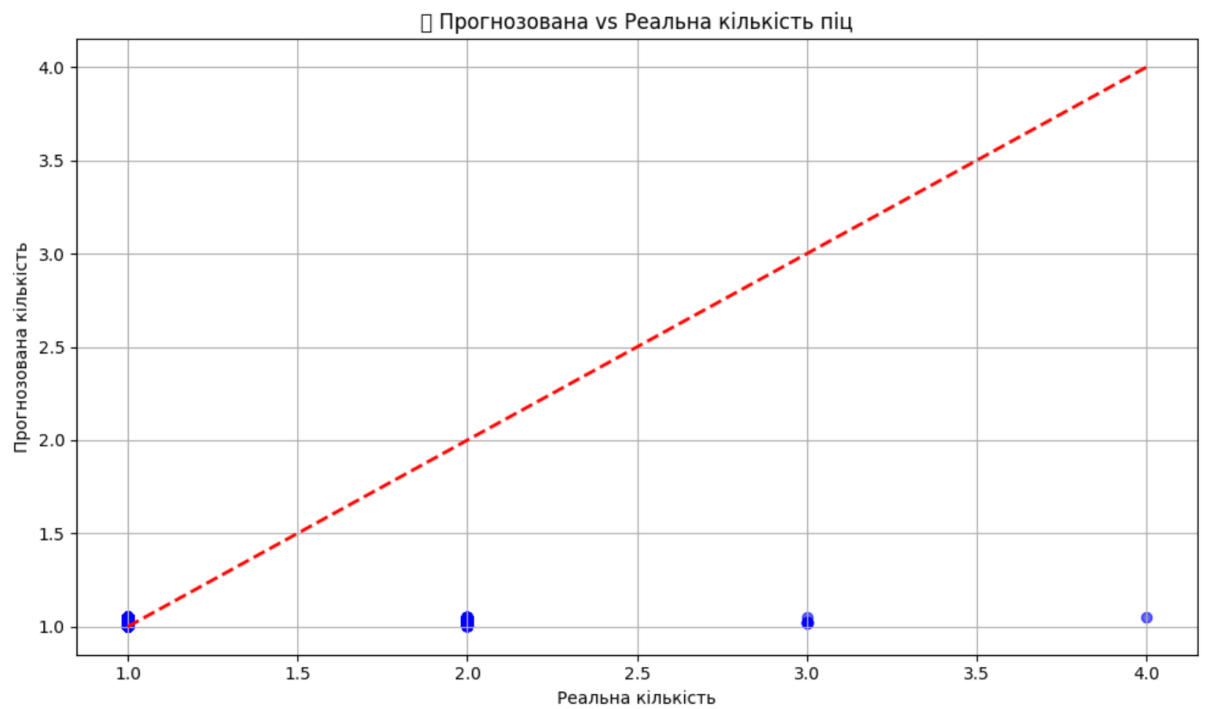
Було поставлено завдання розробити модель для прогнозування кількості піц у замовленні (quantity) за допомогою регресійної моделі.

Методика:

- Вхідні ознаки: розмір піци, ціна, категорія та список інгредієнтів.
- Попередня обробка даних: видалення пропусків, One-Hot кодування для категоріальних ознак, TF-IDF векторизація тексту.
- Побудова Pipeline із використанням моделі **GradientBoostingRegressor**.
- Розділення даних: 80% – тренувальна вибірка, 20% – тестова.
- Оцінка точності проводилася за метриками MAE, RMSE, а також MAPE. Для зручності обчислено "accuracy" як $(100 - \text{MAPE})$.
- Для візуалізації побудовано графік «Реальна vs Прогнозована кількість піц».

Результати:

- MAE: 0.04
- RMSE: 0.15
- MAPE: 2.86%
- Accuracy: 97.14%



Висновок:

Модель показала високу точність прогнозування – середня помилка становить менше однієї піци. Це свідчить про практичну застосовність моделі для оптимізації бізнес-процесів.

4. Висновки

У межах проєкту було реалізовано три задачі з використанням різних методів машинного навчання:

- **Класифікація категорій піци** – моделі Random Forest та XGBoost забезпечують хорошу точність класифікації за характеристиками піци.
- **Прогнозування денного доходу** – модель Prophet ефективно враховує сезонність та тренди, забезпечуючи точність прогнозу на рівні $\approx 74\%$
- **Прогнозування кількості піц** – застосування GradientBoostingRegressor дозволило отримати дуже низьку середню похибку та високу точність (більше 97%).