

РОЗВІДУВАЛЬНИЙ АНАЛІЗ ДАНИХ ДЛЯ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ПЕРЕДБАЧЕННЯ ВАРТОСТІ МОБІЛЬНИХ ТЕЛЕФОНІВ

Вінницький національний технічний університет

Анотація

Дана робота присвячена підготовці та розвідувальному аналізу даних для подальшого використання у інформаційній технології передбачення вартості мобільних телефонів методами машинного навчання.

Ключові слова: мобільний телефон, інформаційні технології, машинне навчання, аналіз даних, передбачення, ознаки, передбачення вартості мобільних телефонів.

Abstract

This work is dedicated to the preparation and exploratory data analysis for further use in information technology for predicting the cost of mobile phones using machine learning methods.

Keywords: mobile phone, information technology, machine learning, data analysis, predictions, signs, mobile phone cost prediction.

Вступ

Сучасний ринок мобільних телефонів характеризується динамічними змінами, що зумовлені швидким розвитком технологій, оновленням модельного ряду виробників та змінами в попиті споживачів. Великий обсяг інформації про технічні характеристики пристроїв, їх вартість на різних ринках та вплив зовнішніх факторів створює передумови для використання аналітичних методів з метою прогнозування цін.

Розвідувальний аналіз даних (EDA) відіграє ключову роль у розробці інформаційних технологій для передбачення вартості мобільних телефонів. Він дозволяє виявити закономірності, оцінити фактори, що впливають на зміну цін, та підготувати дані для побудови ефективних моделей прогнозування. Використання методів машинного навчання дає змогу покращити точність оцінки та допомогти як споживачам, так і бізнесу приймати обґрунтовані рішення щодо купівлі, продажу та ціноутворення пристроїв.

У цьому дослідженні проведено розвідувальний аналіз даних на основі відкритого набору даних «Mobiles Dataset (2025)», що доступний на платформі Kaggle [1]. Основна мета – визначити ключові фактори, що впливають на вартість пристроїв, та підготувати базу для побудови ефективної інформаційної технології прогнозування.

Розвідувальний аналіз

Датасет «Mobiles Dataset (2025)» містить 930 записів та 10 основних атрибутів, що є потенційними факторами визначення вартості мобільного телефона (рис. 1):

- а) Company Name – бренд або виробник мобільного телефона;
- б) Model Name – конкретна модель мобільного телефона;
- в) Mobile Weight – маса мобільного телефона у грамах;
- г) RAM – кількість оперативної пам'яті в телефоні в ГБ;
- д) Front Camera – роздільна здатність фронтальної камери в МП;
- е) Back Camera – роздільна здатність основної задньої камери в МП;
- ж) Processor – чіпсет або процесор, який використовується в пристрої;
- з) Battery Capacity – об'єм акумулятора в мАг;
- и) Screen Size – розмір дисплея в дюймах;
- к) Launched Year – рік офіційного випуску мобільного телефону.

Цільова змінна Launched Price визначає офіційну початкову ціну мобільного телефону після його випуску. У даному датасеті вона подана у валютах таких країн, як Пакистан, Індія, Китай, США та ОАЕ, що, дозволяє дослідити різні закономірності вартостей залежно від місця продажу.

	Company Name	Model Name	Mobile Weight	RAM	Front Camera	Back Camera	Processor	Battery Capacity	Screen Size	Launched Price (Pakistan)	Launched Price (India)	Launched Price (China)	Launched Price (USA)	Launched Price (Dubai)	Launched Year
0	Apple	iPhone 16 128GB	174g	6GB	12MP	48MP	A17 Bionic	3,600mAh	6.1 inches	PKR 224,999	INR 79,999	CNY 5,799	USD 799	AED 2,799	2024
1	Apple	iPhone 16 256GB	174g	6GB	12MP	48MP	A17 Bionic	3,600mAh	6.1 inches	PKR 234,999	INR 84,999	CNY 6,099	USD 849	AED 2,999	2024
2	Apple	iPhone 16 512GB	174g	6GB	12MP	48MP	A17 Bionic	3,600mAh	6.1 inches	PKR 244,999	INR 89,999	CNY 6,499	USD 899	AED 3,199	2024
3	Apple	iPhone 16 Plus 128GB	203g	6GB	12MP	48MP	A17 Bionic	4,200mAh	6.7 inches	PKR 249,999	INR 89,999	CNY 6,199	USD 899	AED 3,199	2024
4	Apple	iPhone 16 Plus 256GB	203g	6GB	12MP	48MP	A17 Bionic	4,200mAh	6.7 inches	PKR 259,999	INR 94,999	CNY 6,499	USD 949	AED 3,399	2024

Рис. 1. Приклад даних із датасету «Mobiles Dataset (2025)»

Було виконано попереднє очищення даних для усунення спотворених числових ознак та перетворено категоріальні ознаки для подальшого аналізу. Побудовано гістограми розподілу значень числових ознак (рис. 2).

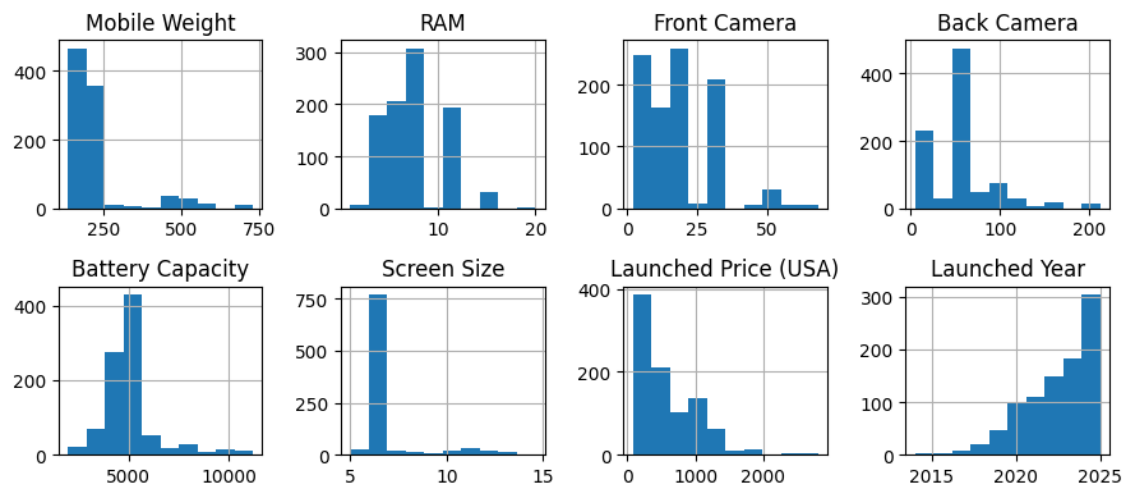


Рис. 2. Гістограми розподілу значень числових ознак

Побудовано коробковий графік вартості мобільного телефону в доларах США залежно від виробника його процесора (рис. 3).

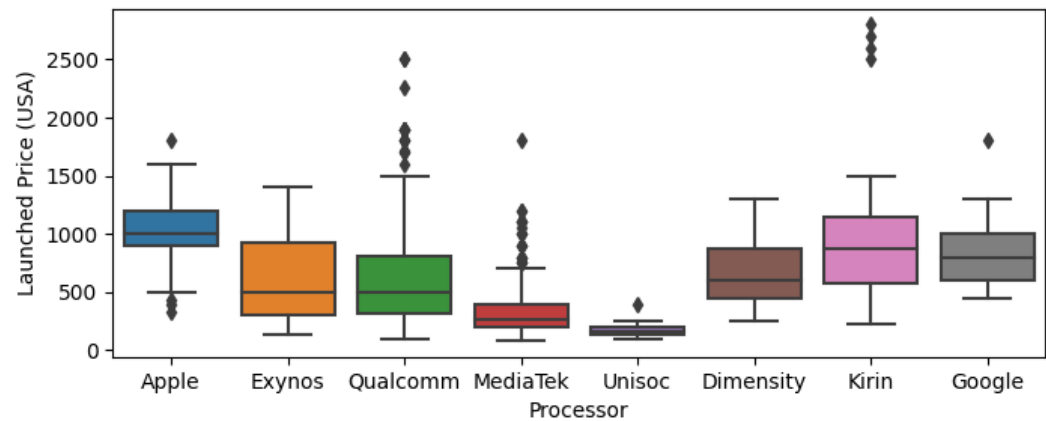


Рис. 3. Коробковий графік вартості мобільного телефона залежно від виробника його процесора

Також було виконано візуалізацію зведеної таблиці вартості мобільного телефона залежно від виробника його процесора та бренду (рис. 4).

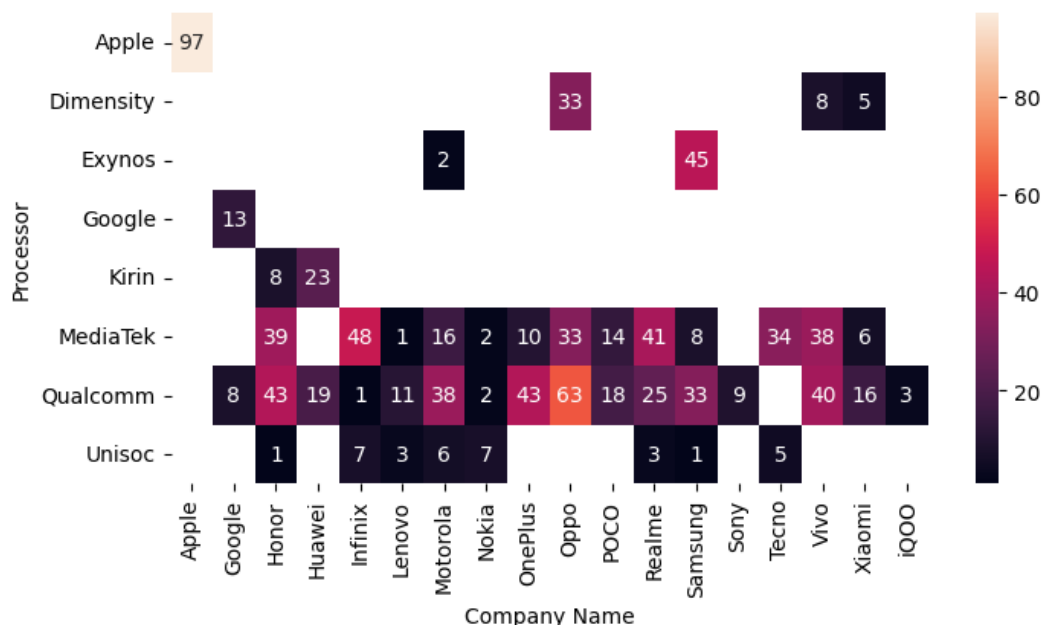


Рис. 4. Візуалізація зведеної таблиці вартості мобільного телефона залежно від виробника його процесора та бренду

Було побудовано теплову карту попарної рангової кореляції Спірмана для повного набору даних для визначення ознак, який мають високий зв'язок із цільовою ознакою (рис. 5).

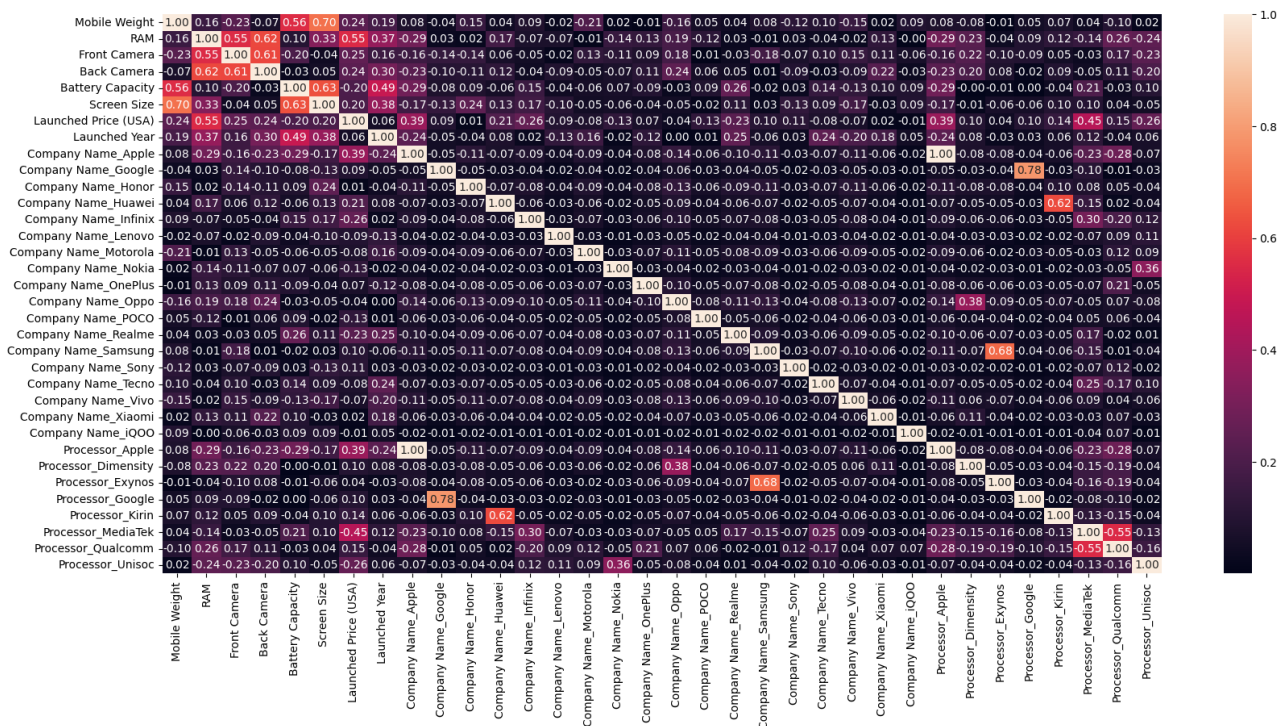


Рис. 5. Теплова карта попарної рангової кореляції Спірмана для повного набору даних

Висновки

Виконаний розвідувальний аналіз представлених даних дає можливість зробити деякі висновки. Більшість ознак у наборі даних мають розподіл, скошений вправо, окрім року випуску, який скошений вліво. Найпоширеніший розмір екрану – 6 дюймів, але є моделі з екранами понад 10 дюймів, що, ймовірно, є планшетами. Аналіз маси підтверджує цю гіпотезу, оскільки деякі пристрої важать понад 500 г. З кожним роком кількість нових моделей зростає, що, ймовірно, пов'язано з розвитком технологій.

Найдорожчі телефони зазвичай мають процесори Apple, хоча моделі з Kirin можуть бути винятково дорогими. Процесори Apple і Google використовуються лише в однойменних смартфонах, тоді як Exynos здебільшого зустрічається у пристроях Samsung. Найпоширенішими процесорами є Qualcomm і MediaTek, причому смартфони з MediaTek загалом дешевші, що може свідчити про нижчу продуктивність.

Аналіз кореляційної теплової карти показує, що вартість телефону найсильніше корелює з обсягом оперативної пам'яті, ємністю акумулятора та розміром екрана. Процесори Apple, Google та Exynos зазвичай асоціюються з вищими цінами, тоді як MediaTek – з нижчими. Деякі компанії, зокрема Apple і Google, мають жорстку прив'язку до власних процесорів, тоді як інші використовують різні варіанти.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Mobiles Dataset (2025). Mobile specs: Company, Model, RAM, Cameras, Battery, Prices (PK, IN, CN, US, UAE). URL: <https://www.kaggle.com/datasets/abdulmalik1518/mobiles-dataset-2025>
2. Йоганссон Р. Numerical Python: Scientific Computing and Data Science Applications with Numpy, SciPy and Matplotlib. – New York: Apress, 2019. – 709 с. – ISBN 978-1-4842-4245-2.
3. Seaborn Tutorial. URL: <https://seaborn.pydata.org/tutorial.html>

Рудзевич Олександр Володимирович – студент групи СА-21б, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, м. Вінниця, e-mail: sasha.rtrr.1@gmail.com

Жуков Сергій Олександрович – к.т.н., доцент кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, e-mail: sazhukov@gmail.com

Rudzevych Olexandr V. – student of group SA-21b, Faculty of Intellectual Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: sasha.rtrr.1@gmail.com

Zhukov Serhii O. – Ph.D., Assistant Professor of the Department of System Analysis and Information Technologies, Vinnytsia National Technical University, Vinnytsia, e-mail: sazhukov@gmail.com