

Klasyfikacja cen telefonów komórkowych

Weronika Kłujso (223599), Michał Korzeniewski (223399),
Miłosz Malinowski (223391), Piotr Misiejuk (223302)

24 maja 2025

Spis treści

1	Streszczenie	2
2	Wprowadzenie	2
3	Cel i zakres badania	2
4	Przegląd literatury	3
5	Słowa kluczowe	3
6	Zmienne wybrane do analizy oraz ich wizualizacja	3
7	Wstępna analiza danych	4
7.1	Statystyki opisowe	4
7.2	Brakujące oraz odstające dane	8
7.3	Skalowanie danych	8
8	Omówienie metod klasyfikacji	9
9	Rezultaty oraz omówienie wyników	9
10	Podsumowanie	9
11	Bibliografia	9

1 Streszczenie

Projekt dotyczy problemu klasyfikacji telefonów komórkowych do jednej z czterech klas cenowych na podstawie ich cech technicznych. Celem było zbudowanie modeli, które potrafią skutecznie przewidzieć przybliżoną cenę urządzenia bez znajomości jego rzeczywistej wartości rynkowej. W tym celu wykorzystano publiczny zbiór danych zawierający informacje o około 2000 telefonów, obejmujący zarówno zmienne ilościowe (np. pamięć RAM, rozdzielczość ekranu, liczba rdzeni procesora), jak i kategoryczne (np. obsługa 4G, ekran dotykowy).

W ramach projektu przeprowadzono analizę danych, ich wstępne przetworzenie i wizualizację, a następnie oceniono skuteczność wybranych algorytmów klasyfikacyjnych. Wyniki pokazują, że na podstawie odpowiednio dobranych parametrów technicznych możliwe jest z dużą dokładnością przewidywanie przynależności telefonu do określonego przedziału cenowego.

2 Wprowadzenie

Telefony komórkowe stanowią jedno z najpopularniejszych i najszybciej rozwijających się urządzeń codziennego użytku. Współczesny konsument ma do wyboru dziesiątki modeli różniących się parametrami technicznymi, jakością wykonania i – co istotne – ceną. Analiza tych różnic może pomóc w przewidywaniu, do jakiego przedziału cenowego należy konkretny model, co ma zastosowanie m.in. w marketingu, handlu internetowym czy planowaniu produkcji.

W niniejszym projekcie dokonano klasyfikacji telefonów komórkowych na cztery grupy cenowe (od najtańszych do najdroższych) na podstawie ich parametrów technicznych. Wykorzystano w tym celu publiczny zbiór danych Mobile Price Classification dostępny na platformie Kaggle, który zawiera dane o ponad 2000 modeli telefonów. Przeprowadzono analizę zmiennych, przygotowanie danych (w tym transformacje), a następnie ocenę skuteczności zastosowanych modeli klasyfikacyjnych.

3 Cel i zakres badania

Celem projektu jest podzielenie telefonów na cztery klasy, od najtańszych do najdroższych, na podstawie ich parametrów technicznych takich jak posia-

kanie ekranu dotykowego, ilość pamięci wewnętrznej czy długość i szerokość, za pomocą metody [TODO]. Zakres badania obejmuje:

- wczytanie i przygotowanie danych,
- skalowanie danych przy użyciu [TODO]
- [TODO]
- podsumowanie wyników i zweryfikowanie ich względem rzeczywistych cen telefonów.

4 Przegląd literatury

5 Słowa kluczowe

POL: klasyfikacja cen telefonów, analiza danych, parametry techniczne, predykcja klasy, dokładność modelu

ENG: mobile price classification, data analysis, technical parameters, class prediction, model accuracy

6 Zmienne wybrane do analizy oraz ich wizualizacja

Zbiór danych zawiera 20 zmiennych, będących cechami telefonu komórkowego. Zmienne zostały podzielone na ilościowe oraz kategoryczne.

- **Ilościowe:**
 - **battery_power** – pojemność baterii w mAh,
 - **clock_speed** – częstotliwość zegara procesora w GHz,
 - **fc** – rozdzielczość przedniego aparatu w MP,
 - **int_memory** – pamięć wewnętrzna w GB,
 - **m_dep** – głębokość telefonu w cm,
 - **mobile_wt** – masa telefonu w gramach,

- **n_cores** – liczba rdzeni procesora,
- **pc** – rozdzielczość tylnego aparatu w MP,
- **px_height** – wysokość ekranu w pikselach,
- **px_width** – szerokość ekranu w pikselach,
- **ram** – pamięć RAM w MB,
- **sc_h** – wysokość ekranu w cm,
- **sc_w** – szerokość ekranu w cm,
- **talk_time** – czas rozmów w godzinach.

• **Kategoryczne:**

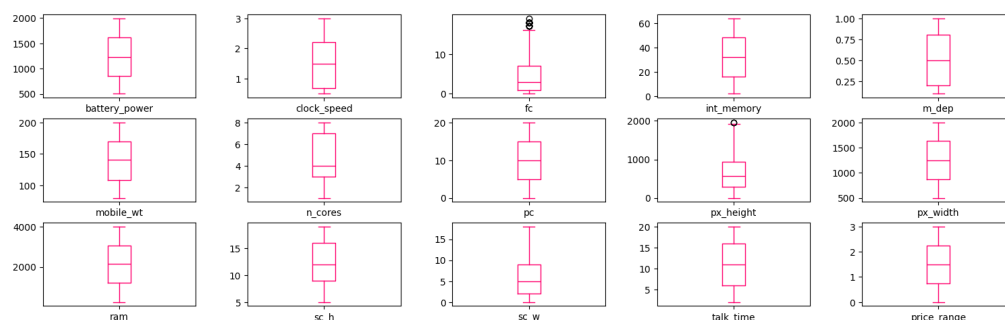
- **three_g** – obsługa 3G,
- **four_g** – obsługa 4G,
- **blue** – obsługa Bluetooth,
- **dual_sim** – obsługa dwóch kart SIM,
- **touch_screen** – ekran dotykowy,
- **wifi** – obsługa WiFi.

Na potrzeby projektu wprowadzimy własną zmienną **price_range**, pełniącą rolę zmiennej objaśnianej w procesie klasyfikacji. Będzie to nasza zmienna docelowa, kategoryczna wieloklasowa, podzielona na cztery klasy cenowe (0 – najtańsze, 3 – najdroższe).

7 Wstępna analiza danych

7.1 Statystyki opisowe

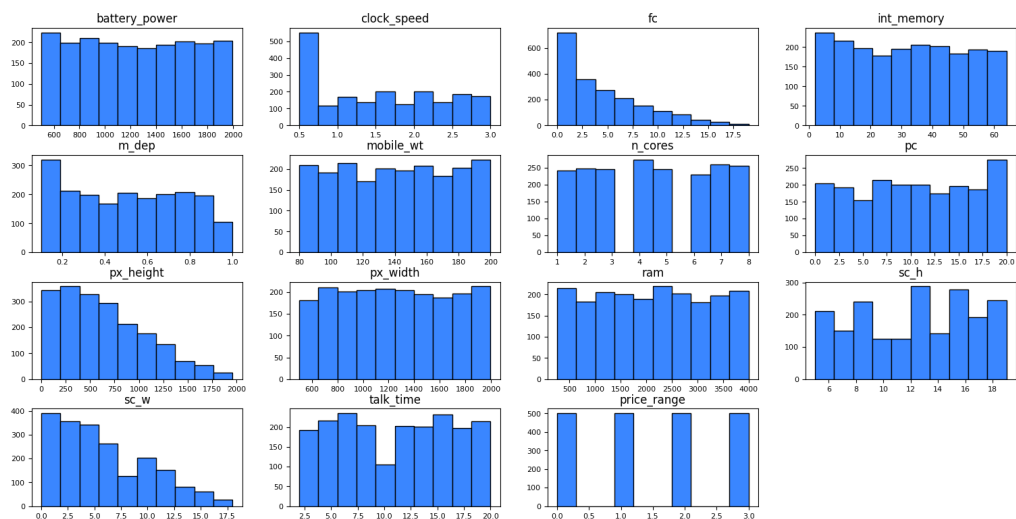
W celu lepszego zrozumienia zbioru danych przeprowadzono analizę podstawowych statystyk opisowych. Dane obejmują zarówno zmienne ilościowe, jak i kategoryczne. Poniżej zaprezentowano rozkłady i wykresy pudełkowe dla zmiennych ilościowych.



Rys. 1. Wykresy pudełkowe dla zmiennych ilościowych

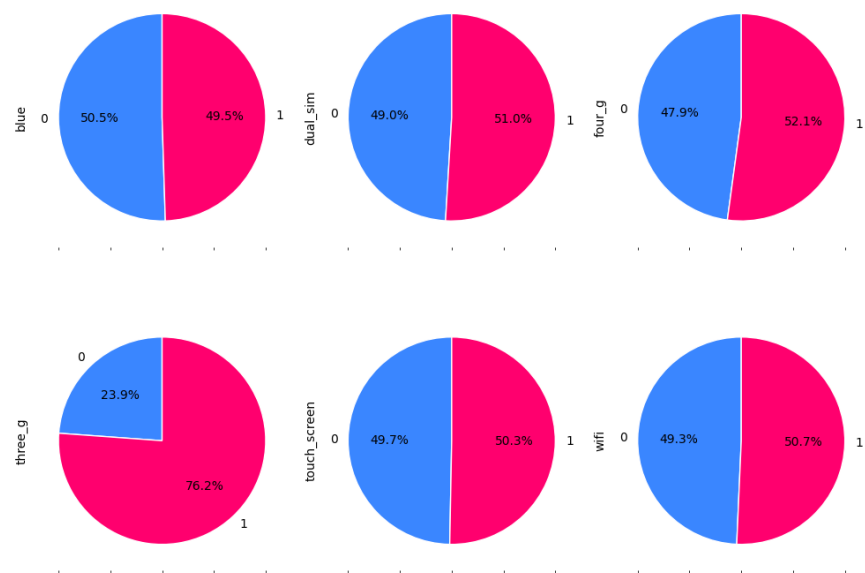
W większości zmiennych obserwujemy stosunkowo symetryczne rozkłady, choć dla niektórych, jak np. `fc` (przedni aparat), występują wartości odstające. Wartości takie mogą świadczyć o nielicznych modelach o bardzo wysokich parametrach w danej kategorii.

Na poniższych histogramach przedstawiono rozkłady zmiennych ilościowych. Widać, że wiele z nich (jak `px_height`, `fc`, `m_dep`) ma rozkłady skośne lub nieregularne, co może mieć wpływ na skuteczność wybranych algorytmów klasyfikacji.

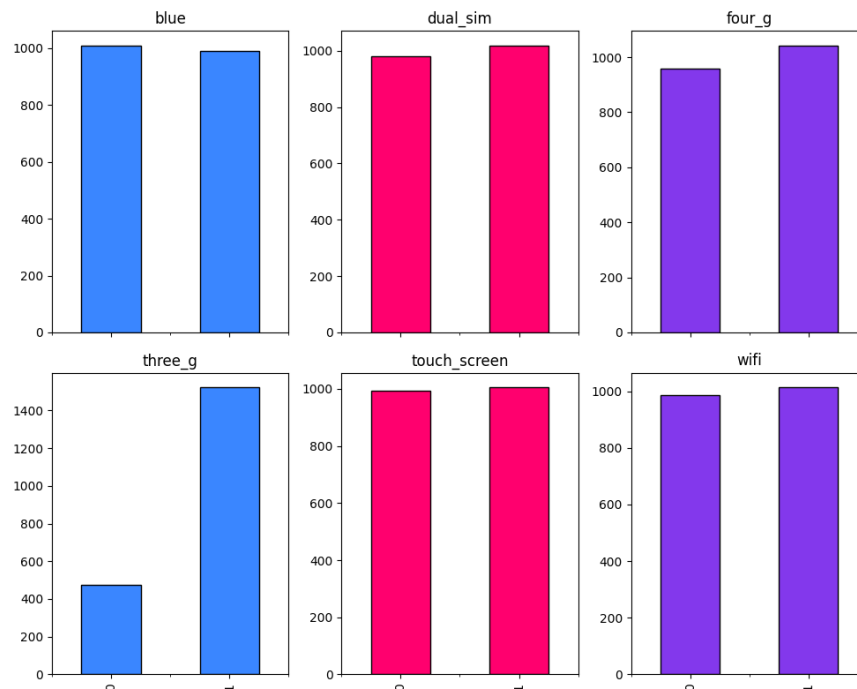


Rys. 2. Histogramy zmiennych ilościowych

W zbiorze występuje sześć zmiennych binarnych opisujących dostępność wybranych funkcji telefonu (np. Bluetooth, ekran dotykowy). Na poniższych wykresach kołowych i słupkowych przedstawiono ich proporcje.



Rys. 3. Udział klas dla zmiennych binarnych (wykresy kołowe)



Rys. 4. Liczność klas dla zmiennych binarnych (wykresy słupkowe)

Zmienna **three_g** dominuje pod względem udziału klasy pozytywnej (76% telefonów obsługuje 3G), co może być istotnym predyktorem w klasyfikacji cen. Pozostałe cechy binarne mają rozkłady bardziej zrównoważone.

7.2 Brakujące oraz odstające dane

7.3 Skalowanie danych

W celu zapewnienia spójności w zakresie wartości cech ilościowych przeprowadzono ich skalowanie, co umożliwia prawidłowe działanie wielu algorytmów klasyfikacyjnych. Przeskalowanie zmiennych redukuje wpływ różnych jednostek miar i ułatwia porównywalność.

8 Omówienie metod klasyfikacji

9 Rezultaty oraz omówienie wyników

10 Podsumowanie

11 Bibliografia

Literatura