

Badanie ogłoszeń używanych samochodów - Streszczenie raportu

401039, Michał Szczurek, czwartek 14⁴⁰
AGH, Wydział Informatyki Elektroniki i Telekomunikacji
Rachunek prawdopodobieństwa i statystyka 2020/2021

Kraków, 24 stycznia 2021

Raport powstał w oparciu o analizę danych dotyczących ogłoszeń sprzedaży używanych samochodów na Białorusi. Głównym celem analizy było sprawdzenie jakie cechy mają wpływ na cenę sprzedawanych samochodów. Analizy dokonano przy pomocy języka R i programu RStudio. W badaniu wykorzystano następujące cechy:

- manufacturer name - nazwa marki,
- color - kolor,
- odometer value - liczba przejechanych kilometrów,
- year produced - data produkcji samochodu,
- engine capacity - pojemność silnika (w litrach),
- body type - typ samochodu (np. suv, sedan),
- price usd - cena w dolarach,
- number of photos - liczba zdjęć w ogłoszeniu,
- days listed - liczba dni od wystawienia ogłoszenia

Analizując dane szczególnie wagę przywiązano do zależności między ceną a liczbą przejechanych kilometrów i rokiem produkcji. Wpływ marek i kolorów samochodów nie został jednak pominięty (jak i innych mniej istotnych cech). Stworzony podczas badania model wskazał, że dla wszystkich cech ilościowych za wyjątkiem liczby przejechanych kilometrów mają pozytywny (wraz ze wzrostem badanych wartości cena rośnie) wpływ na cenę samochodu. Podczas badań odkryto również, że wbrew intuicji większość cech nie najprawdopodobniej nie pochodzi z rozkładu normalnego. Prawdopodobnie jest to spowodowane, tym że na dane takie jak ilość zdjęć, cena ludzie mają bezpośredni wpływ.