

Raport dla projektu "Przewidywanie cen samochodów"

Oleksandr Bondarenko

Kod grupy: K01-21a

1 Wprowadzenie

Celem tego projektu jest zbadanie i prognozowanie cen samochodów w oparciu o bazę danych zebraną z popularnego serwisu OtoMoto. W dzisiejszym świecie rynek motoryzacyjny jest jednym z najbardziej dynamicznych i konkurencyjnych sektorów gospodarki. Znajomość i zrozumienie czynników wpływających na ceny samochodów może mieć duże znaczenie dla kupujących i sprzedających.

OtoMoto to platforma internetowa, na której użytkownicy mogą zamieszczać ogłoszenia kupna i sprzedaży samochodów. Baza danych OtoMoto zawiera liczne rekordy pojazdów, w tym markę, model, rok, przebieg i oczywiście cenę. Korzystając z tej bazy danych, możemy przeprowadzić analizę i zbudować model, który będzie w stanie przewidzieć ceny samochodów.

Przewidywanie cen samochodów może być przydatne zarówno dla kupujących, jak i sprzedających. Kupujący mogą wykorzystać te prognozy do oszacowania godziwej wartości samochodu i podejmowania świadomych decyzji zakupowych. Z drugiej strony, sprzedający mogą wykorzystać prognozy do określenia najlepszej ceny, jakiej mogą zażądać za swoje samochody, biorąc pod uwagę specyficzne cechy każdego pojazdu.

W ramach niniejszego raportu zbierzemy dane z bazy danych OtoMoto, przeanalizujemy uzyskane dane i zbudujemy model przewidywania cen samochodów. Wyniki tej pracy mogą być przydatne dla szerokiego grona interesariuszy i mogą przyczynić się do lepszego funkcjonowania rynku samochodowego.

2 Do czego służy ten program

Celem tego programu jest wykorzystanie bazy danych z serwisu OtoMoto do analizy i przewidywania cen określonych marek i modeli samochodów. Program zapewnia możliwość gromadzenia danych o sprzedaży samochodów z tego popularnego zasobu internetowego i przeprowadzania analizy analitycznej w celu zidentyfikowania wzorców i czynników wpływających na ceny.

Taki program może zaoszczędzić czas i wysiłek podczas badania rynku motoryzacyjnego, dając użytkownikowi dostęp do obszernej bazy danych i narzędzi analitycznych. Może również pomóc w podejmowaniu bardziej świadomych decyzji przy sprzedaży lub zakupie samochodów, biorąc pod uwagę przewidywane ceny i czynniki, które na nie wpływają.

W rezultacie oprogramowanie to pomaga poprawić wydajność i dokładność prognozowania cen samochodów, czyniąc proces kupna i sprzedaży bardziej świadomym i opartym na danych.

3 Opis każdej biblioteki

Program korzysta z następujących bibliotek:

- **csv** - do pracy z plikami CSV.
- **datetime** - do pracy z datą i godziną.
- **os** - dla systemu operacyjnego.
- **random** - do generowania liczb losowych.
- **time** - do pracy z czasem.
- **urllib.parse** - do analizowania adresów URL.
- **numpy** - do obsługi tablic i obliczeń.
- **pandas** - do pracy z danymi w formacie tabeli.
- **requests** - do wysyłania żądań HTTP i pobierania kodu HTML dla stron internetowych.
- **bs4** (BeautifulSoup) - do parsowania kodu HTML.
- **sklearn.ensemble** - do używania algorytmu Random Forest.
- **sklearn.model_selection** - do dzielenia danych na próbki treningowe i testowe.
- **matplotlib.pyplot** - do wizualizacji danych.

4 Opis każdej funkcji

4.1 Funkcja `get_urls(url)`

Ta funkcja pobiera adres URL i zwraca listę adresów URL stron z ogłoszeniami samochodów.

4.2 Funkcja `get_number_of_pages(url)`

Ta funkcja pobiera adres URL i zwraca liczbę stron z ogłoszeniami samochodów.

4.3 Function `parse_data()`

Ta funkcja analizuje dane z ogłoszeń samochodów i zapisuje wynikowe dane w pliku `output.csv`.

4.4 Funkcja `scrap_page(url)`

Ta funkcja pobiera adres URL strony z ogłoszeniami samochodów i analizuje dane z tej strony.

4.5 Function `del_dupl()`

Ta funkcja usuwa zduplikowane dane z pliku `new_data.csv` i zapisuje nowe dane w pliku `new_data.csv`.

4.6 Funkcja `get_cars()`

Ta funkcja zwraca słownik zawierający informacje o markach i modelach samochodów reprezentowanych w danych.

4.7 Funkcja `get_car()`

Ta funkcja umożliwia użytkownikowi wybranie marki i modelu samochodu w celu uzyskania informacji o cenie.

4.8 Funkcja `get_data_car(marka, model)`

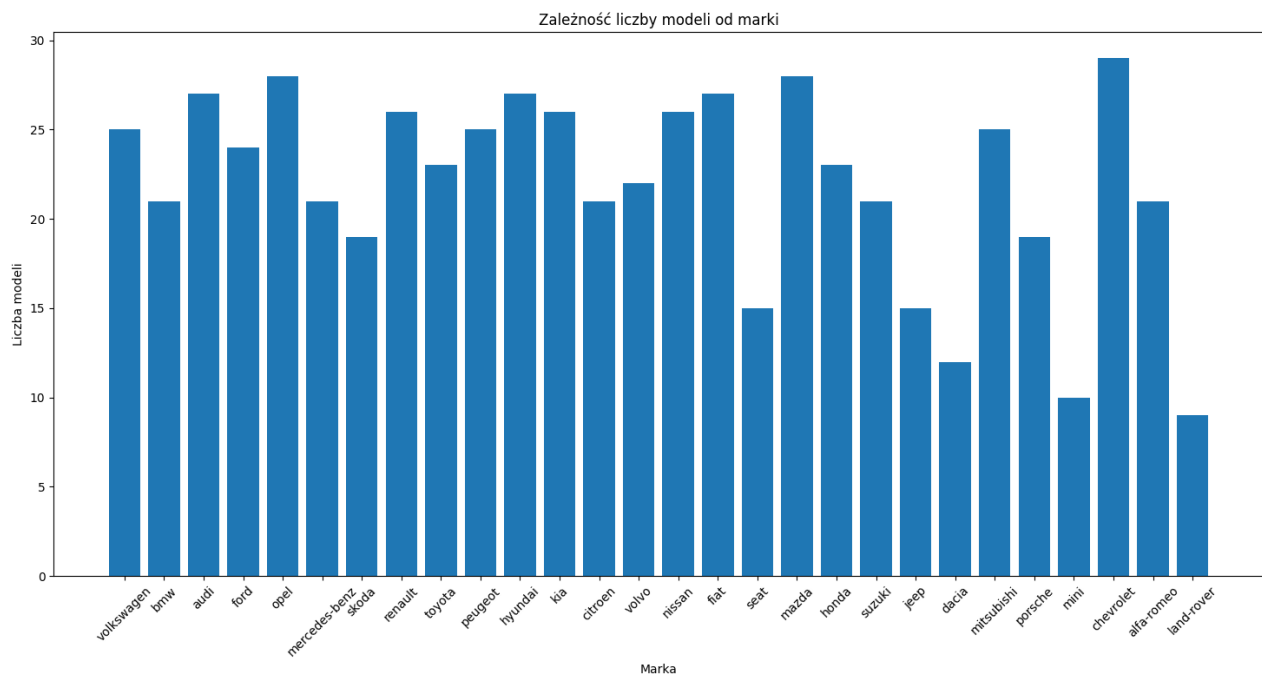
Ta funkcja zwraca dane dotyczące samochodów określonej marki i modelu.

4.9 function `get_price()`

Ta funkcja pyta użytkownika o rok produkcji, przebieg i pojemność silnika samochodu, a następnie przewiduje cenę tego samochodu.

5 Wyniki

Łącznie otrzymano 186157 unikalnych pojazdów. W analizie danych oceniono związek między liczbą modeli a markami samochodów. Wyniki analizy wykazały różnorodność i zróżnicowanie marek samochodów na rynku.



Rysunek 1: Zależność liczby modeli od marki

Wprowadziłem następujące dane dla przykładu:

Samochód: OPEL VECTRA

Rok produkcji: 2000

Przebieg: 227000 km

Pojemność silnika: 1600 cm3

Po przetworzeniu wprowadzonych danych zaproponowano nam cenę: 5432,3 zł.

```
Wybrany samochód: OPEL VECTRA
Wpisz rok produkcji: 2000
Wpisz przebieg: 227000
Wpisz pojemność skokowa silnika: 1600
-----
Dla samochodu: OPEL VECTRA
Rok produkcji: 2000
Przebieg: 227000 km
Pojemność skokową silnika: 1600 cm3
Oferujemy cenę: 5432.3 PLN
-----
```

Rysunek 2: Przybliżone działanie programu

6 Podsumowanie

Ten kod implementuje program do web scrapingu danych sprzedaży samochodów przy użyciu różnych bibliotek. Program umożliwia pobranie informacji o cenie samochodu określonej marki i modelu na podstawie dostępnych danych. Wyniki programu można zapisać w pliku `new_data.csv` do późniejszego wykorzystania.