

Analiza możliwości wykorzystania modeli regresji do przewidywania liczby ludności w Polsce

Krzysztof Kulka
272667@student.pwr.edu.pl
MSiD Lab Wtorek 9.15 NP

May 15, 2024

Spis treści

1	Wstęp	3
2	Zbiór danych i jego analiza	3

1 Wstęp

Problemem projektu jest analiza możliwości modeli regresji liniowej do przewidywania liczby ludności w Polsce. W tym celu wykorzystane zostaną dane historyczne dotyczące demografii, oraz innych czynników wpływających na liczebność populacji. Przedstawiona analiza ma na celu rozstrzygnięcie czy model regresji liniowej jest odpowiedni do przewidywania liczby ludności w Polsce, oraz jakie modele sprawdzają się do tego najlepiej. Analizie zostaną poddane następujące czynniki:

- Historyczna liczba ludności
- Imigracja do kraju
- Wskaźnik dzietności
- Oczekiwana długość życia
- Urbanizacja
- Wskaźnik zmiany populacji na przestrzeni ostatnich 5 lat

Zbadane zaś zostaną następujące modele regresji:

- Regresja liniowa
- Regresja typu Ridge
- Regresja Decisions Trees
- Regresja Random Forest
- Regresja Lasso

2 Zbiór danych i jego analiza

Opis zbioru danych

Zbiór danych zawiera informacje na temat liczby ludności, imigracji do kraju, wskaźnika dzietności, oczekiwanej długości życia w momencie urodzenia na przestrzeni lat 1960-2023. Dane zostały pobrane z serwisu internetowego World Bank[1]. Dodatkowo informacje na temat urbanizacji zostały pobrane z serwisu internetowego Zintegrowana Platforma Edukacyjna Ministerstwa Edukacji Narodowej[2], a wskaźnik zmiany populacji na przestrzeni ostatnich 5 lat został obliczony na podstawie danych historycznych.

References

- [1] World Bank Data poland. <https://www.worldbank.org/pl/country/poland>.
- [2] Zintegrowana Platforma Edukacyjna Ministerstwa Edukacji Narodowej urbanizacja w polsce. <https://zpe.gov.pl/a/zroznicowanie-poziomu-urbanizacji-w-polsce/D19MUchJD>.