

Języki programowania w analizie danych

Zadanie 1.

Przeprowadzeniu analizy statystycznej dla wybranego
zbioru danych

Paweł Ciupka 234048

Bartosz Łuniewski 234086

Filip Woźniak 234131

Cel zadania

Celem zadania była implementacja aplikacji, która realizuje określone wymagania:

1. Dla poszczególnych atrybutów wyznaczyć medianę, minimum i maximum dla cech ilościowych i dominantę dla cech jakościowych.
2. Narysować histogramy dla dwóch cech ilościowych najbardziej ze sobą skorelowanych.
3. Zadbać o czytelność rezultatów oraz staranny i atrakcyjny wygląd histogramów.

Realizacja zadania

Aplikacja została napisana w języku Python w wersji 3.8.2. Do implementacji poszczególnych wymagań wykorzystano takie moduły i biblioteki jak: `statistic`, `collections`, `numpy`, `matplotlib`.

Do przeprowadzenia badań wykorzystano zbiór danych *fertility_diagnosis.data*, który posiada różnego rodzaju informację dotyczące ludzkiego życia.

Wyniki

Punkt 1.

Poniżej przedstawiono wyniki przeprowadzonej analizy danych dla poszczególnych atrybutów, dla których należało wyznaczyć medianę, minimum i maximum dla cech ilościowych i dominantę dla cech jakościowych.

Pora_roku

Minimalna wartość dla cech ilościowych wynosi = -1.0

Maksymalna wartość dla cech ilościowych wynosi = 1.0

Mediana dla cech ilościowych wynosi = -0.33

Wiek

Minimalna wartość dla cech ilościowych wynosi = 0.5

Maksymalna wartość dla cech ilościowych wynosi = 1.0

Mediana dla cech ilościowych wynosi = 0.67

Choroba_dziecieca

Minimalna wartość dla cech ilościowych wynosi = 0

Maksymalna wartość dla cech ilościowych wynosi = 1

Mediana dla cech ilościowych wynosi = 1.0

Wypadek__Uraz

Minimalna wartość dla cech ilościowych wynosi = 0

Maksymalna wartość dla cech ilościowych wynosi = 1

Mediana dla cech ilościowych wynosi = 0.0

Zabieg_chirurgiczny

Minimalna wartość dla cech ilościowych wynosi = 0

Maksymalna wartość dla cech ilościowych wynosi = 1

Mediana dla cech ilościowych wynosi = 1.0

Goraczka

Minimalna wartość dla cech ilościowych wynosi = -1

Maksymalna wartość dla cech ilościowych wynosi = 1

Mediana dla cech ilościowych wynosi = 0.0

Czestotliwosc_konsumpcji_alkoholu

Minimalna wartość dla cech ilościowych wynosi = 0.2

Maksymalna wartość dla cech ilościowych wynosi = 1.0

Mediana dla cech ilościowych wynosi = 0.8

Palenie

Minimalna wartość dla cech ilościowych wynosi = -1

Maksymalna wartość dla cech ilościowych wynosi = 1

Mediana dla cech ilościowych wynosi = -1.0

Liczba_godzin_spedzonych_na_siedzeniu_dziennie

Minimalna wartość dla cech ilościowych wynosi = 0.06

Maksymalna wartość dla cech ilościowych wynosi = 1.0

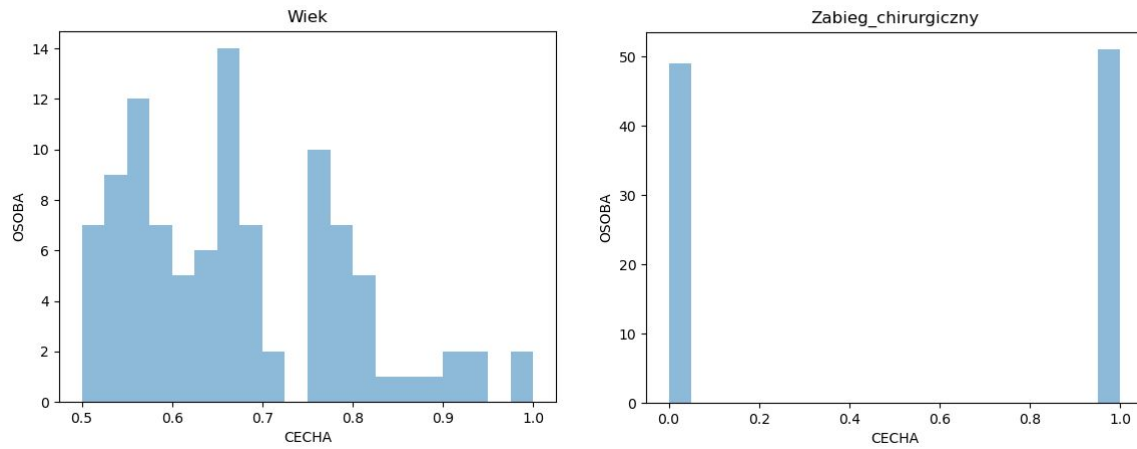
Mediana dla cech ilościowych wynosi = 0.38

Wynik

Dominanta dla cech jakościowych wynosi = [(b'N', 88)]

Punkt 2.

Poniżej przedstawiono histogramy wynikowe dla dwóch cech ilościowych najbardziej ze sobą skorelowanych.



Skorelowane cechy:

1. Wiek
2. Zabieg_chirurgiczny

Współczynnik korelacji:

0.27194480941572