Akademia Ekonomiczno-Humanistyczna w Warszawie

SPRAWOZDANIE

INTELIGENTNA ANALIZA DANYCH

**LAB2**

**WALIDACJA KRZYŻOWA i STRATYFIKACJA**

07.11.2021

JOANNA PRAJZENDANC

36358

MIŁOSZ SAKOWSKI

36381

**Spis treści**

[1. Cel i przebieg ćwiczenia 3](#_Toc21908)

[2. Definicje i założenia 3](#_Toc28075)

[2.1. Wybrana baza danych 3](#_Toc3490)

[2.2. Wyjaśnienie pojęć 3](#_Toc5115)

[3. Walidacja krzyżowa i stratyfikacja 3](#_Toc12644)

[3.1. Zadanie #1 3](#_Toc24127)

[i. Treść polecenia 3](#_Toc17689)

[ii. Rozwiązanie 3](#_Toc4669)

[4. Klasyfikator drzewa losowego 3](#_Toc2025)

[4.1. Zadanie #2 3](#_Toc10509)

[i. Omówienie kodu 3](#_Toc13013)

[ii. Pytania 3](#_Toc13010)

[4.2. Zadanie #3 3](#_Toc27667)

[i. Treść polecenia 3](#_Toc8469)

[ii. Rozwiązanie 4](#_Toc31340)

[4.3. Wnioski 4](#_Toc30022)

[5. Naiwny klasyfikator bayesa 4](#_Toc24652)

[5.1. Zadanie #4 4](#_Toc31722)

[i. Treść polecenia 4](#_Toc1198)

[ii. Rozwiązanie 4](#_Toc17016)

[5.2. Wnioski 4](#_Toc8)

[6. Porównanie klasyfikatorów 4](#_Toc9469)

[7. Wnioski końcowe 4](#_Toc5595)

# Cel i przebieg ćwiczenia

Celem ćwiczenia było

# Definicje i założenia

## Wybrana baza danych

## Wyjaśnienie pojęć

# Walidacja krzyżowa i stratyfikacja

## Zadanie #1

### Treść polecenia

Proszę pobrać inny zbiór danych, zmienić klasy w taki sposób żeby zbiór był niezbalansowany oraz przetestować drzewo decyzyjne stosując 5-krotną cross walidację ze stratyfikacją

### Rozwiązanie

# Klasyfikator drzewa losowego

## Zadanie #2

### Omówienie kodu

### Pytania

#### Czy uzyskane wyniki są zadowalające?

#### Czy model został przetestowany w prawidłowy sposób?

Nie, ponieważ istnieje ryzyko, że w zbiorze trenującym są same 0 i żadnych 1 albo w zbiorze testującym nie ma żadnych 1 więc nie zbadamy skuteczności modelu w dopasowywaniu 1

## Zadanie #3

### Treść polecenia

Proszę pobrać dowolny zbiór danych ze strony <https://archive.ics.uci.edu/ml/index.php.> Następnie proszę podzielić zbiór na dane trenujące i testujące, wytrenować 5 modeli lasów losowych z różną maksymalną głębokością i liczbą drzew w lesie, porównać wyniki. Proszę o sporządzenie sprawozdania z wnioskami.

### Rozwiązanie

#### 2 drzewa o głębokości równej 2

#### 10 drzew o głębokości równej 2

#### 30 drzew o głębokości równej 2

#### 50 drzew o głębokości równiej 2

#### 70 drzew o głębokości równej 2

#### 10 drzew o głębokości równiej 1

#### 10 drzew o głębokości równiej 3

#### Porównanie wyników

## Wnioski

# Naiwny klasyfikator bayesa

## Zadanie #4

### Treść polecenia

Proszę pobrać dowolny zbiór danych ze strony <https://archive.ics.uci.edu/ml/index.php> Następnie proszę podzielić zbiór na dane trenujące i testujące, wytrenować i przetestować klasyfikator GaussianNB. Proszę o sporządzenie sprawozdania z wnioskami.

### Rozwiązanie

## Wnioski

# Porównanie klasyfikatorów

# Wnioski końcowe