Podstawy Sztucznej Inteligencji

Projekt 2 - Sprawozdanie

**Zespół:**

Jarosław Zabuski

Jakub Strawa 300266

**Temat:**

Przewidywanie czy grzyb jest jadalny przy użyciu zmodyfikowanej implementacji algorytmu ID3. Do wyboru testu w drzewie stosujemy zasadę koła ruletki - większe prawdopodobieństwo wyboru mają lepsze testy. Porównanie wyników z wynikami wersji klasycznej algorytmu.

**Doprecyzowanie tematu i założenia:**

Założyliśmy, że w podstawowej wersji algorytmu ID3 dodajemy nowe gałęzie do drzewa decyzyjnego wybierając atrybut z największym zyskiem informacyjnym aż cały zbiór testowy nie zostanie w pełni opisany. Dla zmodyfikowanej wersji ID3 nowe gałęzie są dodawane do momentu aż żaden z pozostałych atrybutów nie niesie ze sobą żadnego zysku informacyjnego. Prawdopodobieństwo wyboru atrybutu jest równe zyskowi informacyjnemu podzielonego przez sumę zysków informacyjnych wszystkich atrybutów które są możliwe do wyboru na tym etapie.

**Podział prac:**

Jarosław Zabuski:

- implementacja klasycznej wersji ID3

- opisanie eksperymentów

Jakub Strawa:

- implementacja zmodyfikowanej wersji ID3

- przeprowadzenie eksperymentów

**Wykorzystane narzędzia i biblioteki:**

Program został napisany w języku Python. Korzystaliśmy wyłącznie z modułów biblioteki standardowej: math, random, sys, gc, time(w celach testowych) oraz klasy Counter z collections.

Do przeprowadzenia testów i eksperymentów wykorzystaliśmy plik z danymi ze strony: <https://archive.ics.uci.edu/ml/datasets/mushroom>

**Przeprowadzone eksperymenty:**

1. Zmiana skuteczności wraz ze wzrostem zbioru uczącego
   1. Założenia:
   2. Eksperyment:
   3. Wniosek:
2. Średni czas wykonywania
   1. Założenia:
   2. Eksperyment:
   3. Wniosek:
3. Średnia głębokość drzewa
   1. Założenia:
   2. Eksperyment:
   3. Wniosek:
4. Średnia ilość wykorzystanych atrybutów
   1. Założenia:
   2. Eksperyment:
   3. Wniosek:

**Wnioski:**

**Instrukcja obsługi i sposób odtworzenia eksperymentów:**

Program uruchamiany jest poleceniem:

*python main.py iterations percentage if\_print\_tree*

Gdzie za *iterations* wpisujemy liczbę iteracji algorytmów, za *percentage* wpisujemy liczbę określającą jaki procent zbioru danych będzie zbiorem przykładów uczących, a za *if\_print\_tree* wpisujemy 1 jeśli chcemy wyświetlić drzewo decyzyjne lub 0 jeśli nie chcemy (1 zalecana jest tylko dla pojedynczego uruchomienia algorytmów).

Przykład:

*python main.py 100 5 0*

Program wykona 100 iteracji dla obu algorytmów, gdzie zbiorem uczącym będzie 5% zbioru danych, drzewo decyzyjne nie zostanie wypisane.

**Aby odtworzyć eksperyment:**

1.

2.

3.

4.