

## **Szkoleniowe Zadanie: Klasyfikacja Kwiatów Iris za pomocą Algorytmu Random Forest**

### **Cel Zadania:**

Celem zadania jest nauczanie uczestników, jak zbudować model klasyfikacyjny przy użyciu algorytmu Random Forest, bazując na zestawie danych „Iris”. Uczestnicy poznają proces wczytywania danych, dzielenia ich na zestawy treningowe i testowe, trenowania modelu oraz oceny jego skuteczności.

### **Wymagania wstępne:**

- Podstawowa znajomość Pythona.
- Zainstalowane biblioteki: pandas, numpy, matplotlib, scikit-learn, seaborn.

### **Treść Zadania:**

#### **Krok 1: Wprowadzenie do Zestawu Danych**

##### **1. Zrozumienie Zestawu Danych:**

- Zapoznaj się z zestawem danych „Iris”, który zawiera informacje o trzech gatunkach kwiatów (setosa, versicolor, virginica) oraz czterech cechach (długość i szerokość działki oraz długość i szerokość płatków).
- Zidentyfikuj cechy, które będą użyte do klasyfikacji, oraz odpowiadające im etykiety gatunków.

##### **2. Zadanie dla uczestników:**

- Przeczytaj dokumentację dotyczącą zestawu danych Iris, aby zrozumieć, co oznaczają poszczególne cechy.

#### **Krok 2: Przygotowanie Środowiska Pracy**

##### **1. Instalacja bibliotek:**

- Upewnij się, że wszystkie wymagane biblioteki są zainstalowane. Uczestnicy mogą użyć polecenia w terminalu do zainstalowania niezbędnych bibliotek.

##### **2. Zadanie dla uczestników:**

- Sprawdź, czy biblioteki są zainstalowane, importując je w swoim środowisku pracy.

#### **Krok 3: Wczytanie Zestawu Danych**

##### **1. Importowanie Bibliotek i Wczytanie Danych:**

- Uczestnicy powinni zaimportować potrzebne biblioteki i wczytać zestaw danych Iris, który będzie zawierał cechy i etykiety.

##### **2. Zadanie dla uczestników:**

- Uruchom proces wczytywania danych i sprawdź, jakie dane zawiera zestaw, zwracając uwagę na kształt macierzy cech i etykiet.

#### **Krok 4: Podział Danych na Zestaw Treningowy i Testowy**

### 1. Podział Danych:

- Uczestnicy będą musieli podzielić dane na zestaw treningowy (80%) i testowy (20%), aby mogli ocenić skuteczność modelu.

### 2. Zadanie dla uczestników:

- Uruchom proces podziału danych i sprawdź kształty nowych zbiorów danych.

## Krok 5: Tworzenie Modelu Random Forest

### 1. Inicjalizacja i Trening Modelu:

- Uczestnicy stworzą model Random Forest i wytrenują go na zestawie treningowym, wykorzystując odpowiednie parametry.

### 2. Zadanie dla uczestników:

- Uruchom proces treningu modelu i upewnij się, że model jest gotowy do dokonania predykcji.

## Krok 6: Predykcja

### 1. Dokonanie Predykcji:

- Uczestnicy użyją wytrenowanego modelu do przewidywania gatunków kwiatów w zestawie testowym.

### 2. Zadanie dla uczestników:

- Uruchom proces przewidywania i zapisz wyniki predykcji.

## Krok 7: Ocena Modelu

### 1. Ocena Skuteczności Modelu:

- Uczestnicy będą korzystać z macierzy pomyłek oraz raportu klasyfikacji, aby ocenić model, analizując dokładność, precyzję i recall.

### 2. Zadanie dla uczestników:

- Uruchom proces oceny modelu i przeanalizuj wyniki.

## Krok 8: Wizualizacja (Opcjonalne)

### 1. Wizualizacja Macierzy Pomyłek:

- Uczestnicy stworzą wizualizację macierzy pomyłek, aby lepiej zrozumieć wyniki modelu.

### 2. Zadanie dla uczestników:

- Uruchom proces wizualizacji i przeanalizuj graficznie wyniki macierzy pomyłek.

## Podsumowanie

- Po zakończeniu zadania uczestnicy będą mieli praktyczne doświadczenie w budowaniu modelu klasyfikacyjnego przy użyciu algorytmu Random Forest. Powinni zrozumieć

proces wczytywania danych, dzielenia ich na zbiory, trenowania modelu oraz oceny jego skuteczności.

- Zachęć uczestników do eksperymentowania z różnymi parametrami modelu oraz innymi zestawami danych w celu dalszego rozwoju umiejętności.

**Dodatkowe Pytania do Dyskusji:**

1. Jakie inne metody można by zastosować do klasyfikacji tych danych?
2. Jakie są zalety i wady użycia algorytmu Random Forest w porównaniu do innych algorytmów klasyfikacyjnych?
3. Jak można poprawić dokładność modelu?