**1. Wprowadzenie do klasyfikacji osiągnięć studentów**

* **Kontekst:** Klasyfikacja osiągnięć studentów ma kluczowe znaczenie dla nauczycieli, umożliwiając monitorowanie postępów, identyfikację mocnych i słabych stron uczniów oraz ocenę skuteczności nauczania.
* **Problem badawczy:** Istnieje niewiele badań koncentrujących się na poprawie jakości modeli klasyfikacji osiągnięć studentów.
* **Cel:** Opracowanie modelu klasyfikacji osiągnięć studentów o wysokiej skuteczności poprzez zastosowanie różnych **technik normalizacji cech** i metody **k-NN**.

**2. Wykorzystane metody**

* **Dane:** Zebrano informacje o **66 studentach** uczestniczących w kursie statystyki online. Dane zawierają 7 cech:
  + **Obecność**, aktywność na zajęciach, oceny zadań domowych, egzaminu śródsemestralnego i końcowego, finalna ocena numeryczna oraz literowa.
* **Metody normalizacji cech:**
  + **Standaryzacja (z-score normalization)** – transformacja wartości na wspólną skalę poprzez odjęcie średniej i podzielenie przez odchylenie standardowe.
  + **Skalowanie min-max** – przekształcanie wartości do zakresu [0,1], co redukuje wpływ różnych zakresów danych.
  + **Centrowanie cech** – przesunięcie wartości tak, aby średnia wynosiła 0.
* **Klasyfikator:**
  + **k-NN (K-Nearest Neighbors)**
  + Metryka odległości: **odległość euklidesowa**
  + Wartości **K = 2, 4, 6**
* **Metody oceny modelu:**
  + **Cross-validation** (2, 3, 5, 10, 20 folds).
  + Metryki: **Accuracy, Precision, Recall, F1-score, AUC**.

**3. Wyniki eksperymentalne**

* **Najlepsza skuteczność (Accuracy = 84,52%)** osiągnięta przez **Standard\_K-NN i Scale\_K-NN** dla **K=2**.
* **Center\_K-NN osiągnął najgorsze wyniki (~80,28%)**, co sugeruje, że centrowanie cech nie jest optymalne dla tego zbioru danych.
* **ROC i AUC:**
  + Najlepsza wartość AUC dla Center\_K-NN (0.945 przy K=2 i fold=5),
  + Jednak **Standard\_K-NN i Scale\_K-NN przewyższyły Center\_K-NN pod względem Accuracy, Precision i Recall**.
* **Zastosowanie walidacji krzyżowej pozwoliło na uzyskanie bardziej stabilnych wyników.**

| **Metoda** | **Accuracy (%) (K=2)** | **AUC** |
| --- | --- | --- |
| **Standard\_K-NN** | **84.52** | 0.924 |
| **Scale\_K-NN** | **84.52** | 0.924 |
| **Center\_K-NN** | 80.28 | **0.945** |

📌 **Wniosek:** Normalizacja **standardowa i min-max** poprawia skuteczność k-NN, podczas gdy centrowanie cech daje gorsze wyniki.

**4. Wnioski końcowe**

* **Wpływ normalizacji na klasyfikację:** Wybór techniki normalizacji cech znacząco wpływa na skuteczność klasyfikacji.
* **Najlepsze metody:** Standaryzacja i min-max poprawiają działanie k-NN w klasyfikacji osiągnięć studentów.
* **Znaczenie walidacji krzyżowej:** Technika ta zwiększa wiarygodność oceny modelu.
* **Zastosowanie:** Wyniki mogą być wykorzystane w systemach wspomagających nauczycieli w ocenie studentów.