**1. Wprowadzenie do drzew decyzyjnych**

* Drzewa decyzyjne to popularna metoda klasyfikacji, która działa na zasadzie podziału danych na podstawie atrybutów.
* **Podział na uniwariacyjne i multiwariacyjne drzewa decyzyjne**:
  + **Uniwariacyjne drzewa** – wybierają jeden atrybut jako kryterium podziału.
  + **Multiwariacyjne drzewa** – używają kombinacji atrybutów do podziału, co może prowadzić do bardziej efektywnych klasyfikatorów.

**2. Zastosowanie teorii zbiorów przybliżonych (RST - Rough Set Theory)**

* **RST pomaga radzić sobie z niepewnością w danych**, co jest problemem dla klasycznych algorytmów drzew decyzyjnych.
* W RST do wyboru atrybutów używa się **regionu pozytywnego (positive region)**, który określa, które obiekty można jednoznacznie przypisać do klasy.
* W klasyfikacji RST kluczową rolę odgrywa **CORE** – minimalny zbiór istotnych cech.

**3. Eksperymenty i wyniki**

* **Testy przeprowadzono na 5 zbiorach danych z repozytorium UCI**, w tym m.in.:
  + **Breast Tissue** (9 atrybutów, 4 klasy)
  + **Liver Disorder** (26 atrybutów, 3 klasy)
  + **Diabetes** (8 atrybutów, 2 klasy)
* **Porównano skuteczność uniwariacyjnych i multiwariacyjnych drzew decyzyjnych**:
  + Multiwariacyjne drzewa miały wyższą dokładność w większości przypadków.
  + Wyjątek: dla zbioru **Diabetes i Primary Tumor** lepszą dokładność miały uniwariacyjne drzewa.

| **Zbiór danych** | **Uniwar. Drzewo (%)** | **Multiwar. Drzewo (%)** |
| --- | --- | --- |
| Breast Tissue | 44.67 | 50.95 |
| Liver Disorder | 91.00 | 93.20 |
| Thoraric Surgery | 73.26 | 78.00 |
| Diabetes | **65.71** | 63.20 |
| Primary Tumor | **40.12** | 38.00 |

* **Wnioski**:
  + Multiwariacyjne drzewa są bardziej efektywne w większości przypadków.
  + Uniwariacyjne drzewa mogą działać lepiej, jeśli dane są dobrze rozdzielone przy użyciu pojedynczych atrybutów.
  + **Złożoność drzewa**: Multiwariacyjne drzewa mają **mniej węzłów**, co oznacza mniejszą złożoność obliczeniową.

**4. Podsumowanie i możliwe zastosowania**

* **RST-based decision trees mogą być skuteczne w klasyfikacji medycznej**.
* **Multiwariacyjne drzewa generują mniejsze i bardziej ogólne modele**, co może zmniejszyć ryzyko przeuczenia.
* Możliwe **zastosowania w systemach eksperckich i diagnostycznych**.