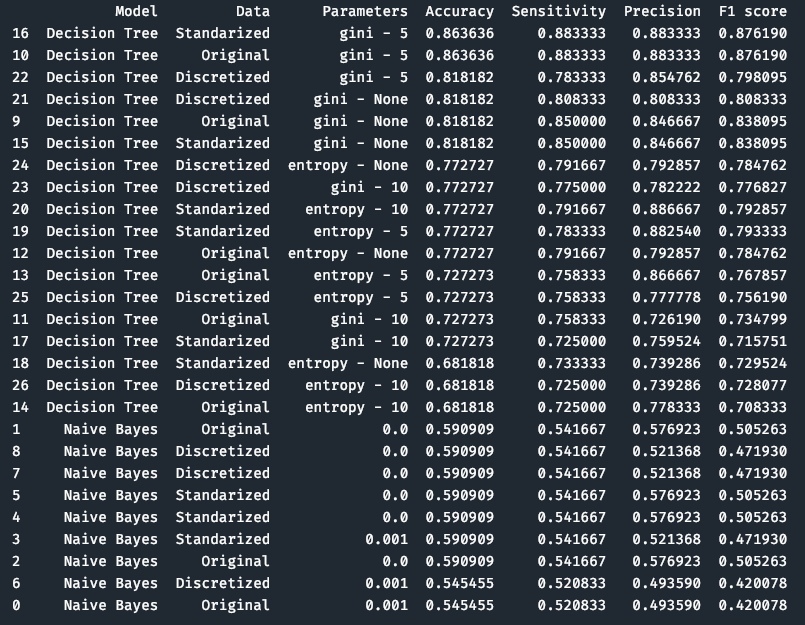
Sprawozdanie 4 – Jakub Radzik 260366

Wykonanie kroki:

* Załadowanie danych
* Eksploracyjna analiza danych
  + Wyświetlenie podstawowych danych statystycznych
  + Sumowanie wartości do 100%
  + Jakie typy występują w danych
  + Średnie wartości danych dla każdego typu
  + Korelacje między danymi
* Przygotowanie danych
  + Usunięcie 5% danych – wpisanie NaNów
  + Wstawienie średnich wartości w puste miejsca
  + Podział na zbiory weryfikacyjne i testowe
  + Przetworzenie standaryzacją, dyskretyzacją i selekcją cech
* Klasyfikacja
  + Klasyfikacja naiwnego Bayesa
  + Klasyfikacja drzewa decyzyjnego
* Ocena modeli
  + Dokladnosc, czulosc, precyzja, F1 (Verstappen > Hamilton)
  + Bayes < Drzewo decyzyjne

1. Biblioteki:
   1. Pandas
   2. NumPy
   3. Scikit-learn
2. Podsumowanie.



Obraz zawierający tekst, zrzut ekranu, Czcionka

Opis wygenerowany automatycznie

Model drzewa decyzyjnego osiągał zazwyczaj lepsze wyniki dla oryginalnych danych, zarówno pod względem dokładności, jak i innych miar jakości.

Standaryzacja i dyskretyzacja danych nieznacznie obniżyły wyniki dla modelu drzewa decyzyjnego. W niektórych przypadkach wyniki dla standaryzowanych i dyskretyzowanych danych były takie same.

Dla modelu Naive Bayes wyniki były niskie i zbliżone dla wszystkich zestawów danych i wartości hiperparametrów. Oznacza to, że Naive Bayes miał niską wydajność w tym przypadku.

klasyfikator bayesa - nie ma on dobrych wyników, ponieważ zakłada on niezależność cech

Wyniki precyzji, czułości i miary F1 były zróżnicowane w zależności od danych i hiperparametrów, ale ogólnie były podobne dla podobnych zestawów.