

# WSI Laboratorium 7 Raport

## Modele bayesowskie

Miłosz Andryszczuk 331355

### 1. Wykorzystane biblioteki

Do przeprowadzenia eksperymentu zostały wykorzystane następujące biblioteki:

- matplotlib  
<https://matplotlib.org/stable/index.html>
- numpy  
<https://numpy.org/doc/>
- pandas  
<https://pandas.pydata.org/docs/>
- scikit-learn  
<https://scikit-learn.org/stable/>

### 2. Opis eksperymentu

W ramach tego eksperymentu zbadano działanie algorytmu naiwnego klasyfikatora Bayesa, który wykorzystuje probabilistyczne podejście do klasyfikacji danych. W przypadku cech ciągłych zastosowano modelowanie ich rozkładu jako normalnego, natomiast cechy dyskretne traktowano jako zmienne kategoryczne.

Celem eksperymentu było zbadanie jakości klasyfikacji wykonywanej przez naiwny klasyfikator Bayesa dla różnych metod oceny, takich jak podział na zbiór treningowy i walidacyjny oraz k-krotna walidacja krzyżowa. Przeanalizowano wpływ różnych proporcji podziału zbioru na dane treningowe i walidacyjne, a także liczby foldów w walidacji krzyżowej na dokładność klasyfikacji i odchylenie standardowe wyników.

Na końcu przeprowadzono ostateczną ocenę jakości modelu na dedykowanym zbiorze testowym, aby uzyskać miarodajny wynik efektywności klasyfikatora.

### 3. Przebieg Eksperymentu

#### 3.1. Wpływ proporcji podziału na zbiór treningowy i walidacyjny na dokładność

W pierwszej części eksperymentu przeanalizowano wpływ proporcji danych przydzielanych do zbiorów treningowego i walidacyjnego na dokładność klasyfikacji. Proporcje testowano w przedziale  $[0.05; 0.6]$ , gdzie wartość oznacza odsetek danych przeznaczonych na zbiór walidacyjny. Dla każdej proporcji klasyfikator był trenowany na danych treningowych i oceniany na danych walidacyjnych.

W celu oceny stabilności wyników, każdy eksperyment powtórzono dziesięciokrotnie, obliczając średnią dokładność na zbiorze walidacyjnym. Dodatkowo porównano dokładności na zbiorach treningowym i walidacyjnym, aby zbadać różnice między nimi. Wyniki przedstawiono na wykresie z dokładnością i odchyleniem standardowym w zależności od proporcji podziału.

#### 3.2. Wpływ liczby foldów w walidacji krzyżowej na dokładność

W drugiej części eksperymentu przeanalizowano wpływ liczby foldów ( $k$ ) używanych w walidacji krzyżowej na jakość klasyfikacji. Liczba foldów była testowana w zakresie od 2 do 20, a dla każdego  $k$  algorytm trenowano na zbiorach treningowych i oceniano na zbiorach walidacyjnych wyznaczanych w procesie walidacji krzyżowej.

Dla każdego  $k$  obliczono średnią dokładność oraz odchylenie standardowe, co pozwoliło zbadać stabilność wyników dla różnych podziałów danych. Wyniki przedstawiono na wykresie ukazującym zależność dokładności i odchylenia standardowego od liczby foldów.

#### 3.3. Porównanie metod oceny i analiza ostatecznej jakości klasyfikacji

W ostatniej części eksperymentu porównano metody oceny klasyfikatora: podział na zbiór treningowy i walidacyjny oraz walidację krzyżową. Średnie dokładności i odchylenia standardowe obu metod zostały przedstawione na zrzucie ekranu. Na zakończenie, klasyfikator został oceniony na dedykowanym zbiorze testowym, aby wyznaczyć ostateczną dokładność modelu.

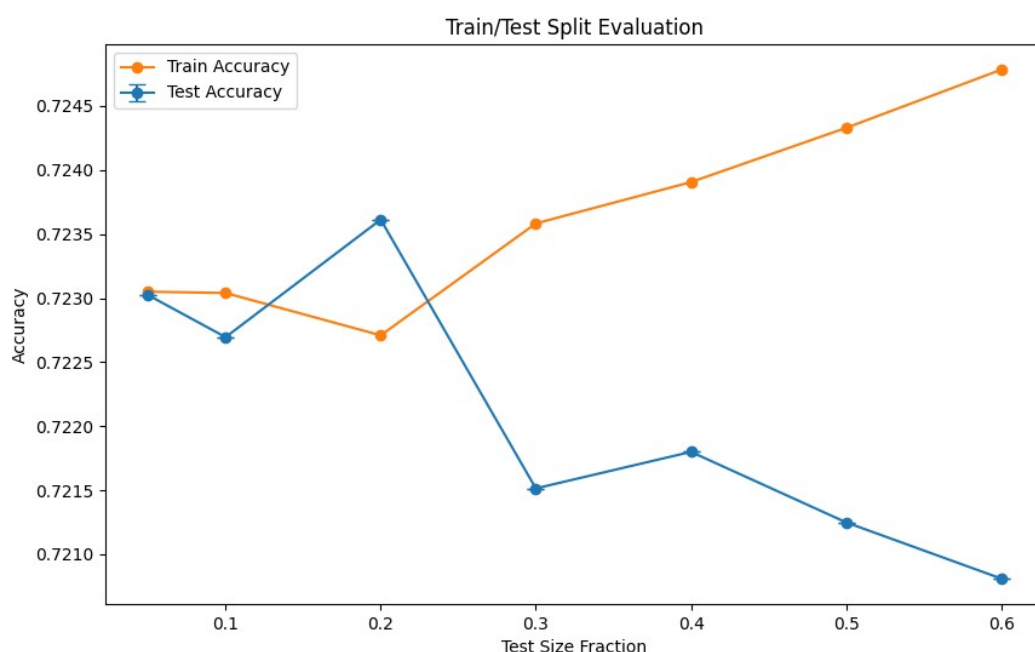
## 4. Wyniki

### 4.1. Wpływ proporcji podziału na zbiór treningowy i walidacyjny na dokładność

Na osi poziomej wykresu przedstawiono proporcję danych przeznaczonych na zbiór walidacyjny (Test Size Fraction), a na osi pionowej znajduje się dokładność klasyfikacji na zbiorach treningowym i walidacyjnym.

Wykres pokazuje, że zmiana proporcji podziału danych wpływa na dokładność modelu. Dokładność na zbiorze treningowym systematycznie rośnie wraz ze zwiększaniem rozmiaru zbioru walidacyjnego. Większy zbiór walidacyjny oznacza mniejszy zbiór treningowy, co prowadzi do bardziej ogólnych reguł klasyfikacji, umożliwiając lepsze dopasowanie do danych treningowych.

Z kolei dokładność na zbiorze walidacyjnym wykazuje tendencję spadkową wraz ze wzrostem proporcji przeznaczonych na zbiór walidacyjny. Maksymalna dokładność na zbiorze walidacyjnym jest osiągana dla mniejszego testowanego podziału, a następnie zaczyna spadać. Spadek ten może być wynikiem mniejszej ilości danych treningowych, co ogranicza zdolność modelu do efektywnego uchwycenia zależności w danych. W konsekwencji prowadzi to do pogorszenia generalizacji na danych walidacyjnych.

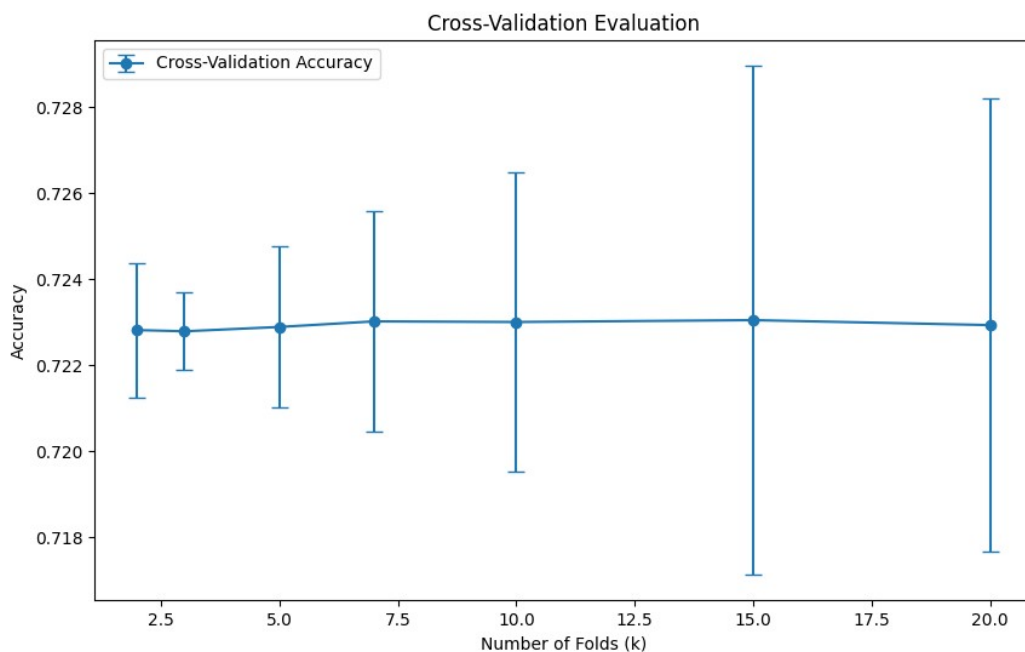


#### 4.2. Wpływ liczby foldów w walidacji krzyżowej na dokładność

Na osi poziomej wykresu przedstawiono liczbę foldów  $k$  używanych w  $k$ -krotnej walidacji krzyżowej, a na osi pionowej znajduje się średnia dokładność klasyfikacji wraz z odchyleniem standardowym dla poszczególnych wartości  $k$ .

Wyniki pokazują, że średnia dokładność klasyfikatora jest stosunkowo stabilna w całym zakresie badanych wartości  $k$ . Najlepsze wyniki uzyskano dla umiarkowanych wartości  $k$ , takich jak  $k=10$  lub  $k=15$ . Warto zauważyć, że odchylenie standardowe wzrasta znacząco przy większej liczbie foldów. Wynika to z faktu, że większa liczba foldów prowadzi do mniejszych zbiorów walidacyjnych, co zwiększa zmienność wyników.

Przy niskich wartościach  $k$ , takich jak  $k=2$ , wyniki są bardziej stabilne, ale mogą być mniej dokładne ze względu na mniej szczegółowy podział danych. Z kolei przy dużych wartościach  $k$ , takich jak  $k=15$  lub  $k=20$ , większe odchylenie standardowe wskazuje na trudności z uzyskaniem spójnych wyników między poszczególnymi foldami.



#### 4.3. Porównanie metod oceny i analiza ostatecznej jakości klasyfikacji

Porównanie metod oceny wykazało, że metoda podziału na zbiór treningowy i walidacyjny z proporcją 0.2 osiągnęła dokładność 0.7236, przy czym wyniki były w pełni stabilne. W przypadku walidacji krzyżowej najlepsze wyniki uzyskano dla  $k=15$ , osiągając dokładność 0.7230 z niewielkim odchyleniem standardowym 0.0059.

Ostateczna ocena modelu na zbiorze testowym wykazała dokładność 0.7212, co świadczy o dobrej zdolności klasyfikatora do generalizacji. Wyniki obu metod były zbliżone, co wskazuje, że zarówno metoda podziału, jak i walidacja krzyżowa są skuteczne w ocenie jakości modelu w tym zadaniu.

```
Comparison of Methods:  
Train/Test Split: Test Size = 0.2, Accuracy = 0.7236 ± 0.0000  
Cross-Validation: Folds = 15, Accuracy = 0.7230 ± 0.0059  
Final Test Set Accuracy: 0.7212
```

### 5. Wnioski

W przeprowadzonych eksperymentach oceniono skuteczność naiwnego klasyfikatora Bayesa oraz różne metody oceny jego jakości. Analiza proporcji podziału na zbiór treningowy i walidacyjny wykazała, że model osiągał najwyższą dokładność na zbiorze walidacyjnym przy mniejszych proporcjach przeznaczonych na walidację. Większe proporcje danych walidacyjnych prowadziły do spadku dokładności, co wynikało z ograniczonej liczby danych treningowych.

Badanie liczby foldów w walidacji krzyżowej pokazało, że najbardziej stabilne wyniki uzyskiwano dla umiarkowanej liczby foldów, takich jak 3. Przy większych wartościach foldów odchylenie standardowe rosło, co można tłumaczyć mniejszymi rozmiarami zbiorów walidacyjnych w podziałach.

Porównanie obu metod oceny – podziału na zbiór treningowy i walidacyjny oraz walidacji krzyżowej – wykazało zbliżoną dokładność, co świadczy o ich równoważnej skuteczności. Ostateczna ocena na zbiorze testowym wykazała dobrą zdolność klasyfikatora do generalizacji.