**Raport Walidacyjny zespołu nr 4**

**Dermanet**

1. W Feature Engineering została użyta funkcja PCA wykonana dla 100 komponentów dla której wartość Cumulative Explained Variance wynosi 0.40-0.45. Czy nie jest to za mała wartość, aby redukować wymiary do 100 komponentów? Czy nie utraciły w ten sposób za dużo informacji o zbiorze?
2. Ręczne usuwanie outlierów, nie jest to optymalna metoda postępowania. Co gdyby zdjęć było kilka milionów?
3. Nie jesteśmy pewni odnośni użycia minmax scalera do standaryzacji danych. Czy nie wpłynie on negatywnie na tworzenie się klastrów? Warto by było sprawdzić jak poradzi sobie np.: Standard Scaler i porównać wyniki.
4. Brak pipeline w Feature Engineering. Znacznie utrudnia to sprawdzenie obserwacji zespołowi walidacyjnemu oraz powoduje pojawienie się błędów w kodzie.

Obraz zawierający tekst, zrzut ekranu, oprogramowanie

Opis wygenerowany automatycznie

1. Można spróbować dodać strides przy features enginnering, skoro tez eksperymentujecie z padding. Strides i padding razem umożliwiają dostosowanie wyjściowych wymiarów oraz poziomu szczegółowości ekstrakcji cech.
2. Sam dobór modeli wygląda bardzo dobrze i wyniki na zbiorze testowym prezentują się podobnie.
3. Finalny wskaźnik Silhouette wygląda dosyć kiepsko, bo tylko 0.141528. Oznacza to, że klastry nie są wyraźnie oddzielone. Sugeruje to, że klastery mogą się na siebie nakładać, a punkty w obrębie klastrów nie są wyraźnie rozdzielone. Może warto rozważyć lepszą normalizację danych, upewnienie się, że dane są odpowiednio przeskalowane oraz znormalizowane.
4. Liczba różnych chorób skórnych do przewidywania wynosi 23, natomiast ostateczna liczba klastrów wynosi 12. Czy nie mija się to z celem biznesowym rozpoznawania każdego rodzaju choroby? Oznacza to że wrzucamy średnio 2 różne choroby do tych samym klastrów, co nie jest chyba najlepszą wiadomością dla potencjalnego pacjenta.
5. Sam model, przynajmniej na zbiorze testowym, lepiej klasteryzuje części ciała niż same choroby. Oczywiście są choroby związane tylko z daną częścią ciała, wtedy klastry wyglądają dosyć dobrze, jednak w przypadku chorób, które mogą występować na różnych częściach ciała wynik klasteryzacji jest słaby. Czy nie warto byłoby stworzyć modelu który rozpoznaje części ciała, a następnie na danej części ciała klasteryzuje choroby?

Wykresy na zbiorze testowym:

Obraz zawierający tekst, zrzut ekranu, linia, diagram

Opis wygenerowany automatycznie

Obraz zawierający tekst, linia, Wykres, diagram

Opis wygenerowany automatycznieObraz zawierający tekst, diagram, Wykres, linia

Opis wygenerowany automatycznie