Metody techniki systemów w medycynie 2 - Komputerowe wspomaganie diagnozowania zawałów z wykorzystaniem algorytmu k-NN.

Leszek Błażewski, 241264
Karol Noga, 241259

Semestr zimowy 2020/2021

Spis treści

1 Opis problemu medycznego jako zadania klasyfikacji wraz z identyfikacją klas oraz cech

Zadanie klasyfikacji w projekcie polegało na wspomożeniu rozpoznawania stanów zawałowych wśród pacjentów na podstawie danych zebranych podczas badań na osobach u których potwierdzono jedną z poniższych diagnoz:

- Ból nie pochodzący z organu serca
- Dusznica bolesna (dławica piersiowa)
- Dusznica Prinzmetala (dławica naczynioskurczowa)
- Pełnościenny zawał serca
- Podwsierdziowy zawał serca

Wynikiem zadania klasyfikacji było przydzielenie każdego z pacjentów do jednej z powyższych klas. W projekcie zbadano jakość klasyfikacji z wykorzystaniem algorytmu k najbliższych sąsiadów, w zależności od ilości cech uwzględnionych podczas uczenia oraz zastosowanej metryki odległości. Cechy wykorzystane podczas uczenia wyznaczone zostały z pomocą rankingu, który opisano w dalszej części pracy.

1.1 Opis danych dostępnych w zbiorze

Zestaw danych wykorzystanych podczas uczenia oraz testowania klasyfikatorów składał się z pięciu plików, które zawierały przypadki pacjentów z daną chorobą. Ilość wszystkich próbek wynosiła 901 i dla każdej z nich zawarto zestaw 59 cech określających stan zdiagnozowanego pacjenta. W zbiorze danych nie znaleziono próbek o błędnych lub brakujących wartościach.

1.1.1 Opis cech

W tabeli 1 przedstawiono zgrupowane cechy wraz z ich charakterystyką. Dostarczony opis danych, posiadał już opisane w tabeli grupowanie. Wszystkie cechy posiadały wartości numeryczne.

Tab. 1: Badane cechy

L.p.	Cecha	Charakterystyka	Wartości
Ogólne			
1.	wiek	dyskretna	liczby naturalne
2.	płeć	kategoryczna	0 - kobieta, 1 - mężczyzna
Ból			
3.	miejsce bólu	kategoryczna	tabela 2
4.	promieniowanie bólu w klatce piersiowej	kategoryczna	tabela 3
5.	charakter bólu	kategoryczna	tabela 4
6.	początek występowania bólu	kategoryczna	tabela 5
7.	liczba godzin od rozpoczęcia bólu	dyskretna	liczby naturalne
8.	długość trwania ostatniego wystąpienia	kategoryczna	tabela 6
Powiązane objawy			
9.	nudności	kategoryczna	0 - brak, 1 - obecność
10.	nadmierna potliwość	kategoryczna	0 - brak, 1 - obecność
11.	kołatanie serca	kategoryczna	0 - brak, 1 - obecność
12.	duszności	kategoryczna	0 - brak, 1 - obecność
13.	zawroty głowy/omdlenia	kategoryczna	0 - brak, 1 - obecność
14.	bekanie	kategoryczna	0 - brak, 1 - obecność
Czynniki paliatywne			
15.	czynniki paliatywne	kategoryczna	tabela 7
Historia podobnego bólu			
16.	wcześniejszy ból tego rodzaju w klatce piersiowej	kategoryczna	0 - brak, 1 - obecność
17.	konsultacja lekarska przy wcześniejszym bólu	kategoryczna	0 - brak, 1 - obecność
18.	wcześniejszy ból powiązany z sercem	kategoryczna	0 - brak, 1 - obecność
19.	wcześniejszy ból spowodowany zawałem	kategoryczna	0 - brak, 1 - obecność
20.	wcześniejszy ból spowodowany chorobą niedokrwienną serca	kategoryczna	0 - brak, 1 - obecność
Historia medyczna			
21.	wcześniejszy zawał serca	kategoryczna	0 - brak, 1 - obecność
22.	wcześniejsza choroba niedokrwienna serca	kategoryczna	0 - brak, 1 - obecność
23.	wcześniejszy nietypowy ból w klatce piersiowej	kategoryczna	0 - brak, 1 - obecność
24.	niewydolność serca	kategoryczna	0 - brak, 1 - obecność
25.	choroba naczyń obwodowych	kategoryczna	0 - brak, 1 - obecność
26.	przepuklina rozworu przełykowego	kategoryczna	0 - brak, 1 - obecność
27.	nadciśnienie tętnicze	kategoryczna	0 - brak, 1 - obecność

Tab. 1: Badane cechy

L.p.	Cecha	Charakterystyka	Wartości
28.	cukrzyca	kategoryczna	0 - brak, 1 - obecność
29.	palacz	kategoryczna	0 - brak, 1 - obecność
Aktualne użycie leków			
30.	diuretyki	kategoryczna	0 - brak, 1 - obecność
31.	azotany	kategoryczna	0 - brak, 1 - obecność
32.	beta-blokery	kategoryczna	0 - brak, 1 - obecność
33.	naparstnica	kategoryczna	0 - brak, 1 - obecność
34.	niesteroidowe leki przeciwzapalne	kategoryczna	0 - brak, 1 - obecność
35.	leki zobojętniające kwas żołądkowy / blokery H2	kategoryczna	0 - brak, 1 - obecność
Badania fizyczne			
36.	skurczowe ciśnienie tętnicze	dyskretna	liczby naturalne
37.	rozkurczowe ciśnienie tętnicze	dyskretna	liczby naturalne
38.	tętno	dyskretna	liczby naturalne
39.	szybkość oddychania	dyskretna	liczby naturalne
40.	rzężenia	kategoryczna	0 - brak, 1 - obecność
41.	sinica	kategoryczna	0 - brak, 1 - obecność
42.	bladość	kategoryczna	0 - brak, 1 - obecność
43.	szmery skurczowe	kategoryczna	0 - brak, 1 - obecność
44.	szmery rozkurczowe	kategoryczna	0 - brak, 1 - obecność
45.	obrzęk	kategoryczna	0 - brak, 1 - obecność
46.	trzeci ton serca	kategoryczna	0 - brak, 1 - obecność
47.	czwarty ton serca	kategoryczna	0 - brak, 1 - obecność
48.	tkliwość ściany klatki piersiowej	kategoryczna	0 - brak, 1 - obecność
49.	obfite pocenie	kategoryczna	0 - brak, 1 - obecność
Badania EKG			
50.	nowy załamek Q	kategoryczna	0 - brak, 1 - obecność
51.	jakikolwiek załamek Q	kategoryczna	0 - brak, 1 - obecność
52.	nowe uniesienie odcinka ST	kategoryczna	0 - brak, 1 - obecność
53.	jakiekolwiek uniesienie odcinka ST	kategoryczna	0 - brak, 1 - obecność
54.	nowe obniżenie odcinka ST	kategoryczna	0 - brak, 1 - obecność
55.	jakiekolwiek obniżenie odcinka ST	kategoryczna	0 - brak, 1 - obecność
56.	nowy odwrócony załamek T	kategoryczna	0 - brak, 1 - obecność
57.	jakikolwiek odwrócony załamek T	kategoryczna	0 - brak, 1 - obecność
58.	nowe zaburzenie przewodnictwa śródkomorowego	kategoryczna	0 - brak, 1 - obecność
59.	jakiekolwiek zaburzenie przewodnictwa śródkomorowego	kategoryczna	0 - brak, 1 - obecność

Tab. 2: Wartości cechy miejsce bólu

Wartość	Znaczenie
1	zamostkowy
2	lewa strona, w okol. serca
3	prawa strona na wys. serca
4	lewy bok klatki piersiowej
5	prawy bok klatki piersiowej
6	brzuch
7	plecy
8	inna

Tab. 3: Wartości cechy promieniowanie bólu w klatce piersiowej

Wartość	Znaczenie
1	szyja
2	szczęka
3	lewe ramię
4	lewa ręka
5	prawe ramię
6	plecy
7	brzuch
8	inne

Tab. 4: Wartości cechy charakter bólu

Wartość	Znaczenie
1	ciągły
2	epizodyczny
3	częściej epizodyczny niż ciągły
4	częściej ciągły niż epizodyczny
5	tępy / nacisk
6	ostry
7	palący
8	opłucnowy

Tab. 5: Wartości cechy początek występowania bólu

Wartość	Znaczenie
1	podczas wysiłku
2	w spoczynku
3	podczas snu

Tab. 6: Wartości cechy długość trwania ostatniego wystąpienia

Wartość	Znaczenie
1	poniżej 5 min
2	5 - 30 min
3	30 - 60 min
4	1 - 6 godz.
5	powyżej 12 godz.

Tab. 7: Wartości cechy czynniki paliatywne

Wartość	Znaczenie
1	brak
2	nitrogliceryna w ciągu 5 min
3	nitrogliceryna po upływie 5 min
4	leki zobojętniające kwas żołądkowy
5	znieczulenie poza morfiną
6	morfina

2 Ranking cech

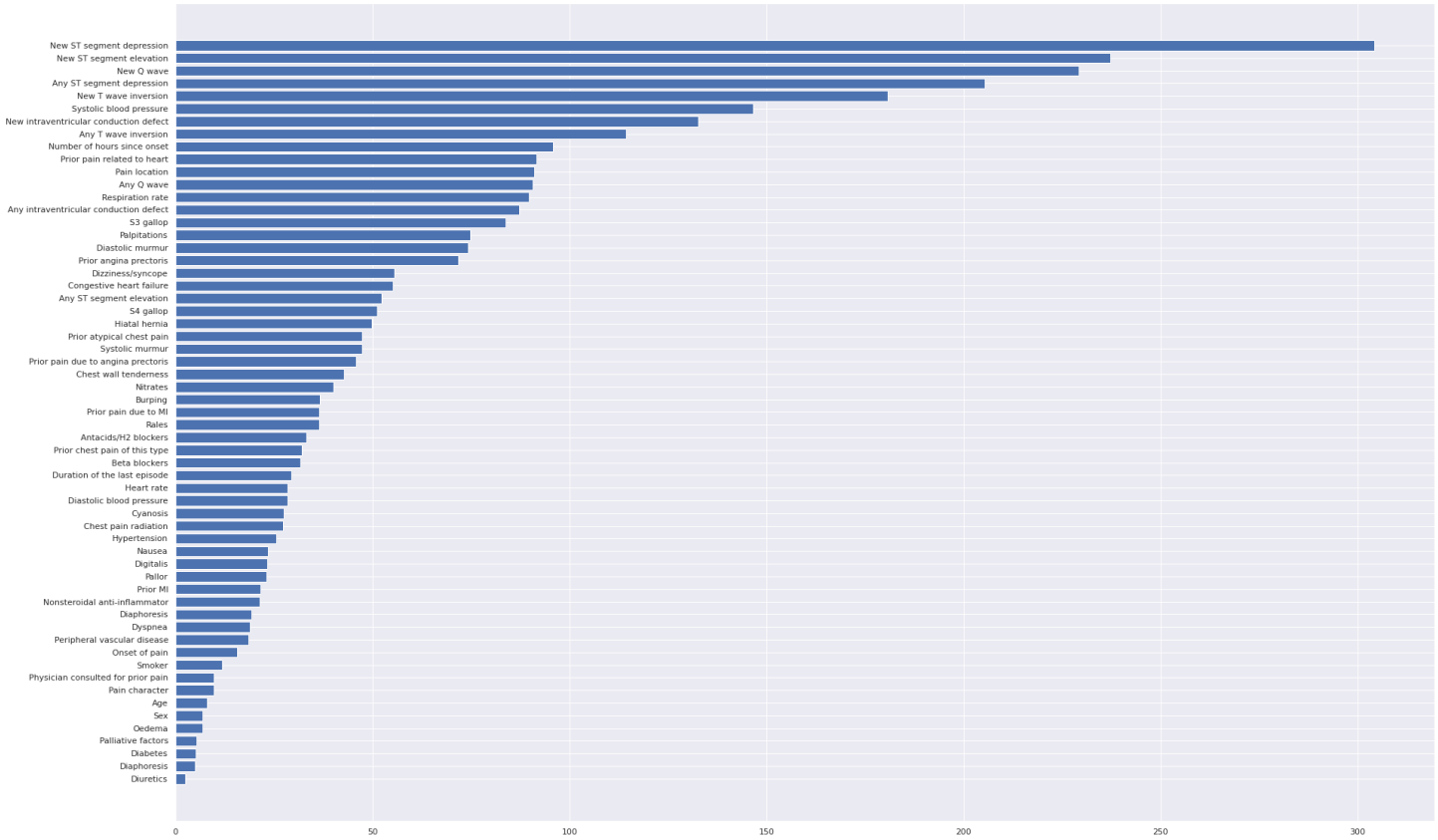
Duża ilość cech wykorzystywanych podczas trenowania algorytmu nie zawsze musi przynosić najlepszy rezultat. Podczas przeprowadzania badań istniała możliwość, iż część cech, silniej wpływała na jakość otrzymane klasyfikacji, ponieważ wybrane symptomy mogły stanowić szum w zbiorze uczącym, obniżając procent zidentyfikowanych poprawnie chorób. Proces selekcji cech miał za zadanie zbudowanie rankingu, który wyróżnił najistotniejsze symptomy wpływające na jakość klasyfikacji i wybranie z nich najlepszych cech.

2.1 Zastosowany algorytm

Do budowy rankingu oraz selekcji cech wykorzystano algorytm *SelectKBest* z biblioteki scikit-learn[4]. Do oceny przydatności cech w algorytmie, zastosowano funkcję *f_ classif*, która wykorzystywała metodę analizy wariancji do wyznaczenia wartości zależności przydzielonej klasy od danej cechy.

2.2 Uzyskany ranking

Wyniki uzyskane z zastosowania metody *SelectKBest* posortowane zostały malejąco, tak aby uwidocznić najistotniejsze cechy wpływające na klasyfikację. Na wykresie 1 przedstawiono uzyskany ranking, gdzie wartości poszczególnych cech równe były wynikom testu *ANOVA* przeprowadzanego na statystyce F z wykorzystaniem funkcji *f_ classif*. Dane na wykresie zilustrowane zostały z wykorzystaniem biblioteki *matplotlib*[2].



Rys. 1: Ranking cech utworzony z wykorzystaniem funkcji f_classif

3 Zastosowany algorytm klasyfikacji

W projekcie badano skuteczność klasyfikacji z wykorzystaniem algorytmu minimalno-odległościowego - k najbliższych sąsiadów. Sposób działania algorytmu opisany został na schemacie 1.

Metoda k najbliższych sąsiadów jest rozszerzoną wersją zachłannego algorytmu najbliższego sąsiada, która rozbudowuje go o dodatkową weryfikację uwzględniającą liczność

Algorithm 1 K Nearest Neighbors

1: **Input:**
   X = zestaw uczący
   L = etykiety klas zestawu
   x_q = niesklasyfikowana próbka
   k = liczba sąsiadów
2: **for** $(x', l') \in X$ **do**
3: Oblicz odległość $d(x', x_q)$
4: **end for**
5: Posortuj rosnąco obliczone odległości elementów zestawu uczącego X od x_q
6: Policz wystąpienia każdej z klas w L pośród najbliższych k sąsiadów x_q
7: Przydziel x_q do najczęściej występującej klasy

najbliższych elementów. Klasyfikowany obiekt przydzielany jest do klasy na podstawie głosowania większościowego przeprowadzonego wśród k instancji znajdujących się w najbliższej odległości. Parametrami algorytmu k-nn była ilość sąsiadów uwzględniana podczas głosowania oraz metryka definiująca miarę odległości.

3.1 Miary odległości

Odległości w algorytmie k najbliższych sąsiadów obliczane były z wykorzystaniem dwóch metryk, euklidesowej oraz manhattan. Metrykę euklidesową w przestrzeni R^n, gdzie n = liczba cech, definiuje się wzorem

$$d_e(x, y) = \sqrt{(y_1 - x_1)^2 + \cdots + (y_n - x_n)^2} \tag{1}$$

natomiast metrykę manhattan w przestrzeni R^n oblicza się przy użyciu wzoru

$$d_m(x, y) = \sum_{k=1}^{n} |x_k - y_k| \tag{2}$$

4 Opis środowiska programistycznego

W projekcie wykorzystano środowisko *Google Colab*[1], które zbudowane zostało w oparciu o interaktywne narzędzie *Jupter Notebook*[2]. Środowisko pozwalało na interaktywne wykonywanie komend języka Python oraz korzystanie z plików znajdujących się bezpośrednio w usłudze *Google drive*. Narzędzie posiadało wbudowany system uruchomieniowy dostarczy przy pomocy usługi *Google Cloud*, które zapewniło wystarczającą ilość zasobów potrzebnych do przeprowadzenia badań zawartych w projekcie. Narzędzie zawierało również zainstalowane wszystkie popularne biblioteki języki Python wykorzystywane w analizie danych.

[1] https://colab.research.google.com/notebooks/intro.ipynb [dostęp 29.11.2020]
[2] https://jupyter.org [dostęp 29.11.2020]

W projekcie wykorzystano następujące biblioteki dostępne w języku Python:

1. numpy[1]
2. pandas[3]
3. scikit-learn[4]
4. matplotlib[2]
5. scipy[5]

Pierwsze dwie biblioteki wykorzystane zostały do odczytania oraz transformacji danych. Biblioteka *scikit-learn* posłużyła do klasyfikacji badanych danych. Czwarta pozycja wykorzystana została do zaprezentowania uzyskanych wyników na wykresach. Ostatnia biblioteka zawierała funkcje pozwalające na przeprowadzenia analizy statystycznej.

5 Przeprowadzone badania

W poniższym rozdziale opisano plan oraz wyniki przeprowadzonych eksperymentów z wykorzystaniem klasyfikatora k najbliższych sąsiadów.

5.1 Plan eksperymentu

Celem projektu było zbadanie jak ilość wykorzystanych cech, liczba sąsiadów oraz metryka odległości wpływają na jakość klasyfikacji. Badanie przeprowadzone zostało dla następujących kombinacji parametrów algorytmu:

- ilość sąsiadów - $1, 5, 10$
- miara odległości - euklidesowa, manhattan

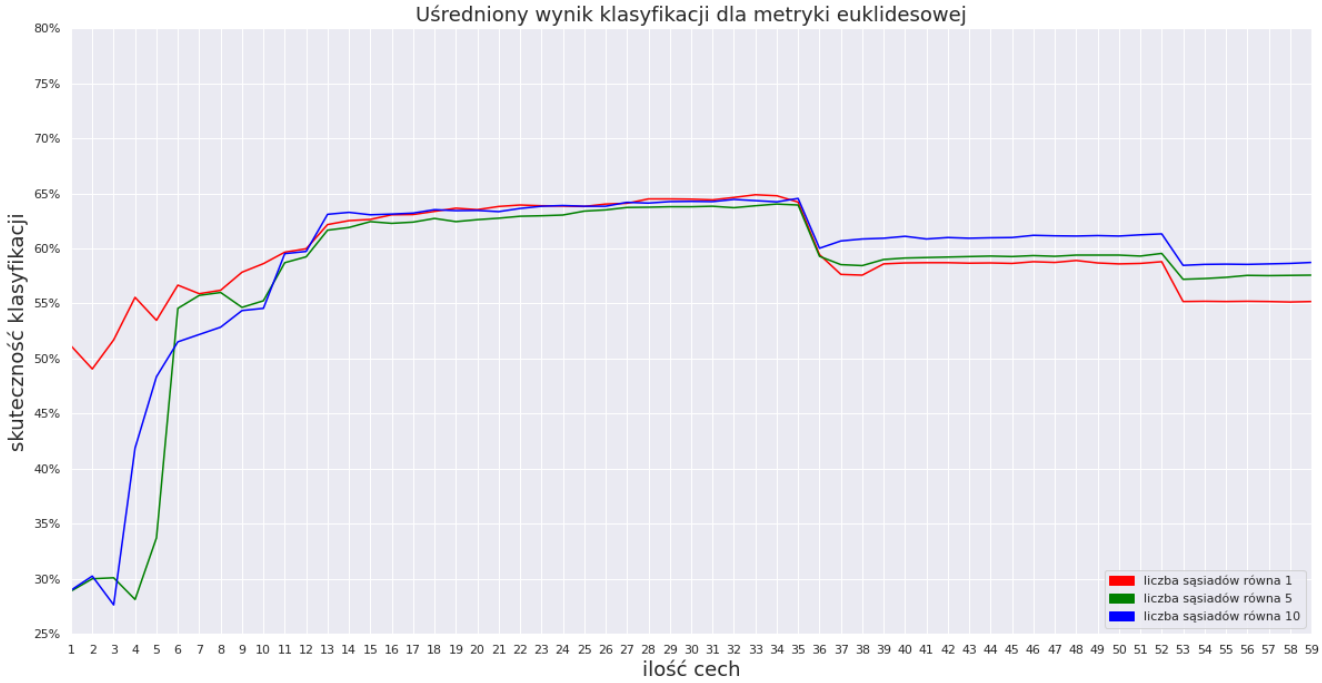
Dla każdego eksperymentu składającego się z kombinacji wymienionych parametrów i ilości cech, określano jakość klasyfikacji z wykorzystaniem pięć razy powtarzanej stratyfikowanej metody dwukrotnej walidacji krzyżowej (*z ang. repeated stratified cross validation*). Sprawdzian krzyżowy polegał na podzieleniu dostępnych danych na dwa segmenty, z których jeden wykorzystany został do uczenia modelu, natomiast drugi posłużył do walidacji uzyskanych wyników. Jakość klasyfikacji stanowiła ilość poprawnie zidentyfikowanych obiektów na zbiorze testowym. Cała procedura dzielenia danych i ewaluacji klasyfikatora, powtarzana była pięciokrotnie dla każdego zestawu parametrów i ilości cech, co pozwoliło na lepsze oszacowanie końcowej jakości klasyfikacji. Dodatkowo w badaniach zastosowano stratyfikowaną wersję algorytmu, która zapewniała zachowanie proporcji klas w dzielonym zbiorze. Metoda pozwoliła uniknąć sytuacji w której w zestawie treningowym nie znajdowały się obserwacje należące jedynie do jednej z klas lub ich ilość była znikoma.

W kolejnych eksperymentach ilość cech wykorzystana podczas uczenia zwiększana była zgodnie z rankingiem uzyskanym na wykresie 1. Końcową jakość rozpoznawania stanowiła średnia uzyskanych klasyfikacji w każdym z przebiegów w walidacji krzyżowej.

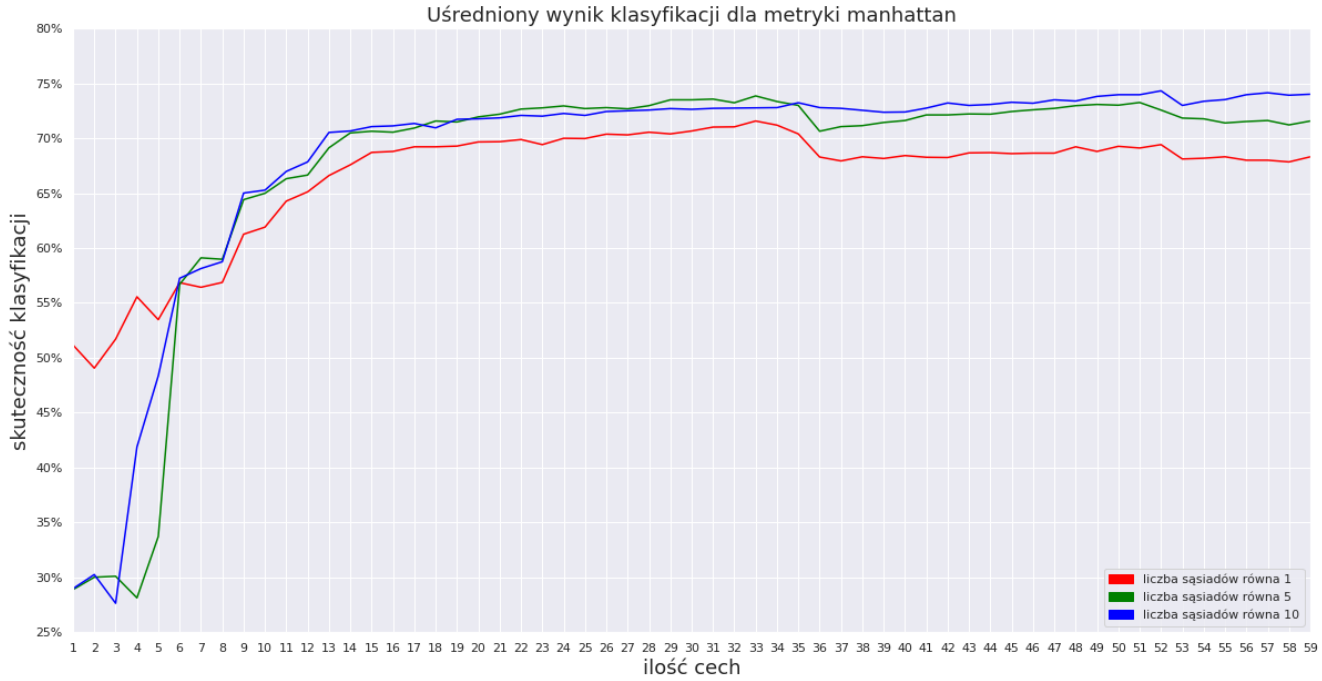
5.2 Uzyskane wyniki

Wyniki przeprowadzonych eksperymentów przedstawione zostały na dwóch wykresach podzielonych względem zastosowanych w badaniach metryk. Poszczególne serie na diagramach, reprezentują klasyfikatory o różnej liczbie sąsiadów. Na osi X umieszczono ilość cech zastosowanych w eksperymencie, natomiast oś Y reprezentowała średnią jakość klasyfikacji wyników uzyskanych w procesie powtarzanej walidacji krzyżowej.

Najlepszą jakość klasyfikacji wynoszącą 74% uzyskano dla klasyfikatora o liczbie sąsiadów równej 10, ilości cech wynoszącej 52 oraz wykorzystanej metryce manhattan. Odchylenie standardowe najlepszego zestawu danych wynosiło 0.01.



Rys. 2: Zależność jakości średniej klasyfikacji od ilości cech dla metryki euklidesowej



Rys. 3: Zależność jakości średniej klasyfikacji od ilości cech dla metryki manhattan

5.3 Dyskusja otrzymanych wyników

Podczas analizy otrzymanych wyników, zauważono, że dla dowolnych parametrów branie pod uwagę dwóch bądź trzech cech nie polepsza, a nawet pogarsza skuteczność klasyfikacji względem użycia wyłącznie jednej cechy. Jednocześnie zaobserwowano, że dla małej liczby cech, użycie liczby sąsiadów równej 1 zapewnia zdecydowanie lepsze wyniki niż dla 5, bądź 10. Dla każdej kombinacji miary odległości i liczby sąsiadów liczba cech wynosząca 13 jest granicą, po przekroczeniu której wzrost poprawności klasyfikacji jest znikomy. Dla odległości euklidesowej i liczby cech powyżej 10, każda z badanych liczb sąsiadów daje niemal identyczne wyniki, natomiast w metryce manhattan już od sześciu cech obserwujemy spadek dokładności klasyfikacji dla jednego sąsiada względem pięciu lub dziesięciu sąsiadów o około trzy do czterech punktów procentowych.

6 Analiza statystyczna otrzymanych wyników

Z uwagi na dużą liczbę symptomów dostarczonych w wykorzystanym zbiorze danych, analiza statystyczna przeprowadzona została wyłącznie dla siedmiu kolejnych najlepszych cech, wyznaczonych w rankingu zaprezentowanym na wykresie 1. Pojedyncza instancja dla której wyznaczano parametry statystyczne składała się z kombinacji ilości cech, ilości sąsiadów oraz zastosowanej metryki. Poniżej przedstawiono wyniki przeprowadzonej analizy z podziałem na ilość cech zastosowanych w procesie uczenia.

6.1 Uzyskane wyniki

Tab. 8: Charakterystyka statystycznie lepsza dla klasyfikatorów uczonych z wykorzystaniem jednej cechy

Statystycznie lepszy dla 1 cechy	k1euklides	k5euklides	k10euklides	k1manhattan	k5manhattan	k10manhattan
k1euklides	0	1	1	0	1	1
k5euklides	0	0	0	0	0	0
k10euklides	0	0	0	0	0	0
k1manhattan	0	1	1	0	1	1
k5manhattan	0	0	0	0	0	0
k10manhattan	0	0	0	0	0	0

Tab. 9: Charakterystyka statystycznie lepsza dla klasyfikatorów uczonych z wykorzystaniem dwóch cech

Statystycznie lepszy dla 2 cech	k1euklides	k5euklides	k10euklides	k1manhattan	k5manhattan	k10manhattan
k1euklides	0	1	1	0	1	1
k5euklides	0	0	0	0	0	0
k10euklides	0	0	0	0	0	0
k1manhattan	0	1	1	0	1	1
k5manhattan	0	0	0	0	0	0
k10manhattan	0	0	0	0	0	0

Tab. 10: Charakterystyka statystycznie lepsza dla klasyfikatorów uczonych z wykorzystaniem trzech cech

Statystycznie lepszy dla 3 cech	k1euklides	k5euklides	k10euklides	k1manhattan	k5manhattan	k10manhattan
k1euklides	0	1	1	0	1	1
k5euklides	0	0	0	0	0	0
k10euklides	0	0	0	0	0	0
k1manhattan	0	1	1	0	1	1
k5manhattan	0	0	0	0	0	0
k10manhattan	0	0	0	0	0	0

Tab. 11: Charakterystyka statystycznie lepsza dla klasyfikatorów uczonych z wykorzystaniem czterech cech

Statystycznie lepszy dla 4 cech	k1euklides	k5euklides	k10euklides	k1manhattan	k5manhattan	k10manhattan
k1euklides	0	1	1	0	1	1
k5euklides	0	0	0	0	0	0
k10euklides	0	1	0	0	1	0
k1manhattan	0	1	1	0	1	1
k5manhattan	0	0	0	0	0	0
k10manhattan	0	1	0	0	1	0

Tab. 12: Charakterystyka statystycznie lepsza dla klasyfikatorów uczonych z wykorzystaniem pięciu cech

Statystycznie lepszy dla 5 cech	k1euklides	k5euklides	k10euklides	k1manhattan	k5manhattan	k10manhattan
k1euklides	0	1	0	0	1	0
k5euklides	0	0	0	0	0	0
k10euklides	0	1	0	0	1	0
k1manhattan	0	1	0	0	1	0
k5manhattan	0	0	0	0	0	0
k10manhattan	0	1	0	0	1	0

Tab. 13: Charakterystyka statystycznie lepsza dla klasyfikatorów uczonych z wykorzystaniem sześciu cech

Statystycznie lepszy dla 6 cech	k1euklides	k5euklides	k10euklides	k1manhattan	k5manhattan	k10manhattan
k1euklides	0	1	1	0	0	0
k5euklides	0	0	1	0	0	0
k10euklides	0	0	0	0	0	0
k1manhattan	0	1	1	0	0	0
k5manhattan	0	1	1	0	0	0
k10manhattan	0	1	1	0	0	0

Tab. 14: Charakterystyka statystycznie lepsza dla klasyfikatorów uczonych z wykorzystaniem siedmiu cech

Statystycznie lepszy dla 7 cech	k1euklides	k5euklides	k10euklides	k1manhattan	k5manhattan	k10manhattan
k1euklides	0	0	1	0	0	0
k5euklides	0	0	1	0	0	0
k10euklides	0	0	0	0	0	0
k1manhattan	0	0	1	0	0	0
k5manhattan	1	1	1	1	0	0
k10manhattan	1	1	1	1	0	0

6.2 Wnioski

Analiza statystyczna uzyskanych danych pozwoliła na potwierdzenie zależności wynikających z wykresów zawartych w sekcji 5.2. Na podstawie danych zawartych w tabeli 8 stwierdzić można, iż metryka euklidesowa była statystycznie lepsza od pozostałych modeli klasyfikatorów uczonych z wykorzystaniem jednej cechy. Porównując tabele z wynikami dla kolejnych ilości cech, zauważyć można, że wraz z wzrostem ilości uwzględnianych symptomów, statystycznie lepsze stają się klasyfikatory o większej liczbie sąsiadów. Wartości statystyki zawarte w tabeli 14 pozwalają zauważyć iż, dla siedmiu cech klasyfikator z metryką manhattan o większej liczbie sąsiadów, rozpoznawał poprawne wyniki z większą dokładnością niż modele korzystające z odległości euklidesowej. Analiza statystyczna sugeruje więc, że wraz z wzrostem liczby wykorzystanych cech oraz zwiększenia ilości sąsiadów uzyskujemy statystycznie lepsze klasyfikatory. Znaleziony najlepszy klasyfikator o liczbie sąsiadów równej 10, korzystający z metryki manhattan, który uczony był na zbiorze testowym składającym się z 52 cech również potwierdza tą tezę.

Literatura

[1] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del R'ıo, M. Wiebe, P. Peterson, P. G'erard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, T. E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, Wrze. 2020.

[2] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.

[3] T. pandas development team. pandas-dev/pandas: Pandas, Luty 2020.

[4] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[5] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.