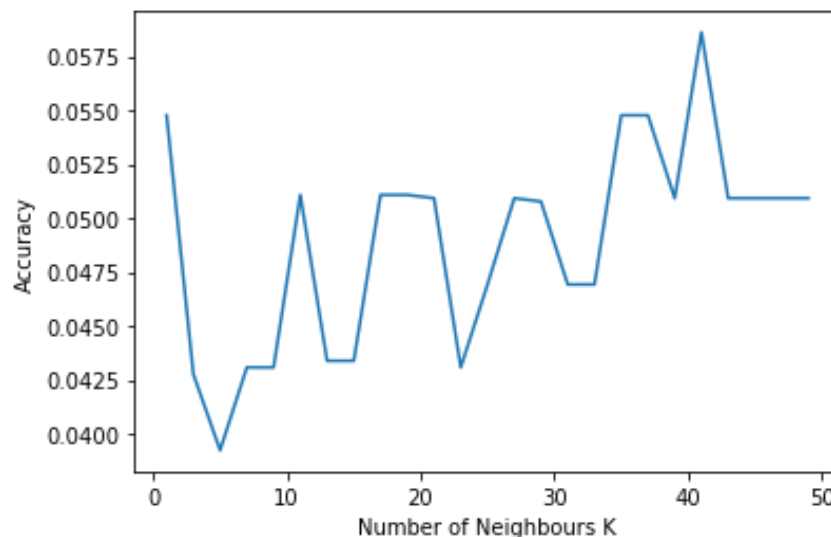


Jednym z nieparametrycznych podejść regresyjnych stosowanych w statystyce do prognozowania wartości pewnej zmiennej losowej jest algorytm k najbliższych sąsiadów. Za jego pomocą można również dokonać klasyfikacji. Gdy związek między zmiennymi objaśniającymi i objaśnianymi jest złożony lub nietypowy (np. niemonotoniczny), co sprawia, że modelowanie w konwencjonalny sposób stanowi wyzwanie, podejście k najbliższych sąsiadów jest bardzo pomocne. Klasyczne techniki, takie jak regresja liniowa, będą zazwyczaj dawały dokładniejsze wyniki, gdy związek jest prosty do interpretacji (np. liniowy), a zbiór nie zawiera obserwacji odstających.

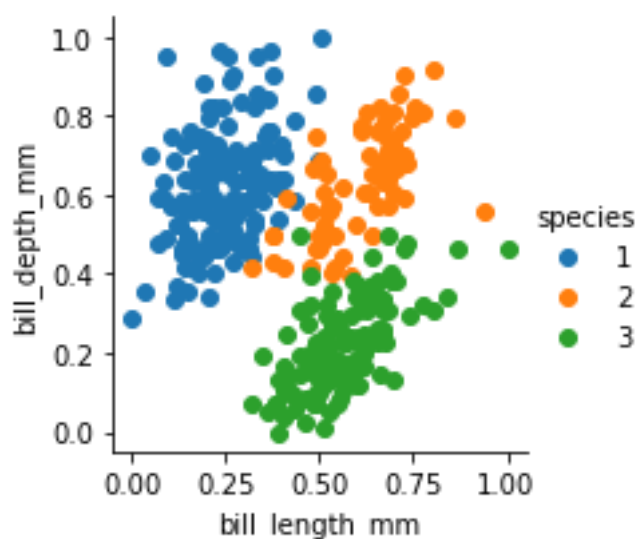
Do stworzenia modelu wykorzystaliśmy dane z biblioteki seaborn-data. Dane zawierały informacje dot. pingwinów.

- Model klasyfikacyjny
 - Wybór k najbliższych sąsiadów dla modelu klasyfikacyjnego



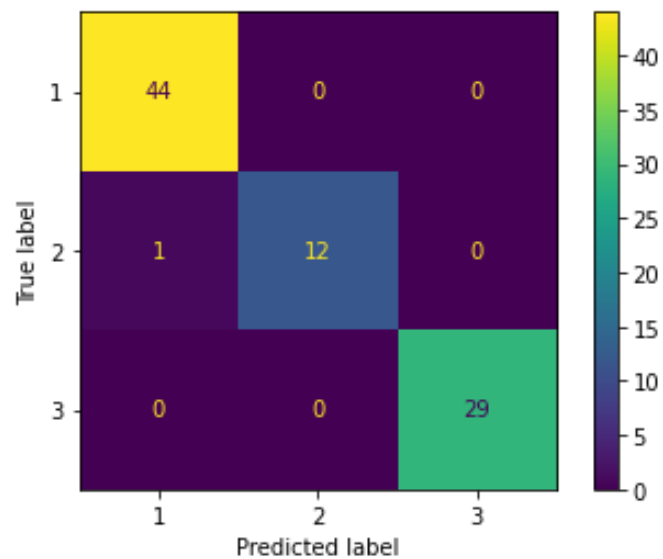
Rysunek 1 Wykres pokazujący optymalną ilość sąsiadów (optymalna liczba to 5).

- Zbuduj model KNN (klasyfikacyjny)



Rysunek 2 Wykres KNN

- Stwórz macierz pomyłek



Rysunek 3 Macierz pomyłek

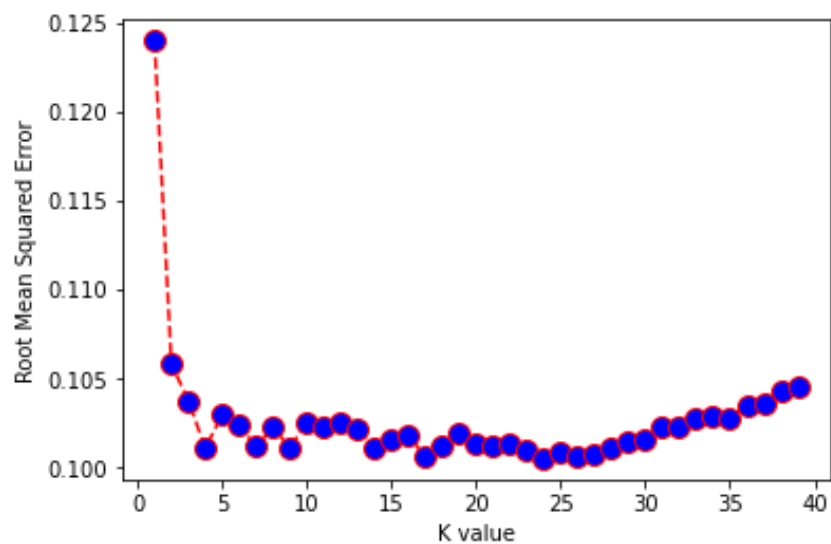
- Oblicz statystyki dla modelu

	precision	recall	f1-score	support
1	0.98	1.00	0.99	44
2	1.00	0.92	0.96	13
3	1.00	1.00	1.00	29
accuracy			0.99	86
macro avg	0.99	0.97	0.98	86
weighted avg	0.99	0.99	0.99	86

Rysunek 4 Wartości statystyczne dla modelu

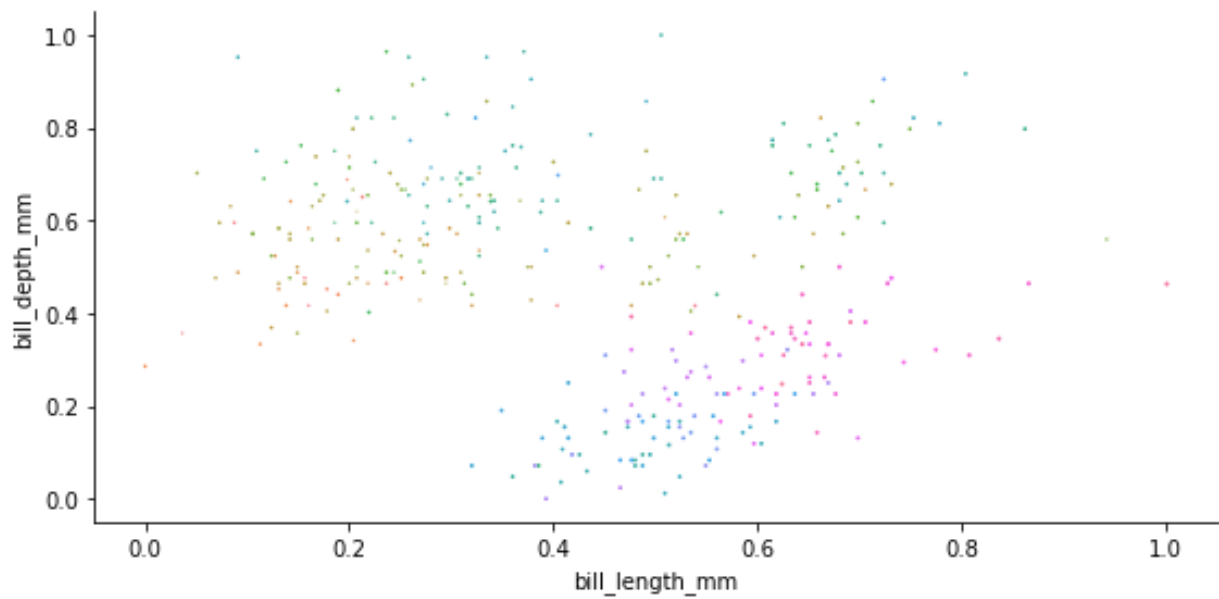
- Model regresyjny

- Wybór k najbliższych sąsiadów dla modelu regresyjnego



Rysunek 5 Wykres pokazujący optymalną ilość sąsiadów (optymalna liczba to 4)

- Zbuduj model KNN (regresyjny)



- Oblicz statystyki modelu

MIARA STATYSTYCZNA	WARTOŚĆ
Współczynnik determinacji (R^2)	0.85
Średni kwadratowy błąd kalibracji ($RMSE_c$)	0.0881