

Kacper Dondziak

Inteligencja obliczeniowa: Projekt 2

Baza danych: Deszcz w Australii

link: <https://www.kaggle.com/jsphyg/weather-dataset-rattle-package?select=weatherAUS.csv>

repo: [https://github.com/Maniakrzela/weather\\_prediction](https://github.com/Maniakrzela/weather_prediction)

Baza zawiera zbierane przez 10 lat dane na temat pogody w różnych lokalizacjach Australii. Podstawowe pytanie brzmi czy jutro będzie padać? Baza zawiera **145 461** wierszy w formacie csv.

### 1. Obróbka bazy danych:

Baza zawiera puste dane w niektórych komórkach. Po odfiltrowaniu brakujących wartości pozostało **112 938** wierszy. Jest to moim zdaniem liczba do zaakceptowania do dalszej pracy. Do klasyfikatorów kategoriycznych wartości liczbowe zostały podzielone na 3 części: Low, Medium, High. Kolumna z datą została pominięta.

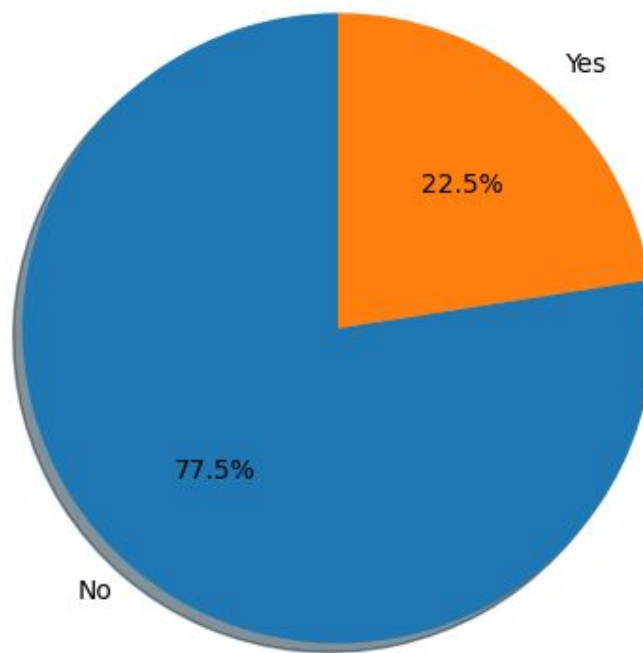
Baza została podzielona na zbiór testowy i treningowy w proporcji **3:1**

```
x_train, x_test, y_train, y_test = train_test_split(features, rain_tomorrow, test_size=0.33, random_state=42)
```

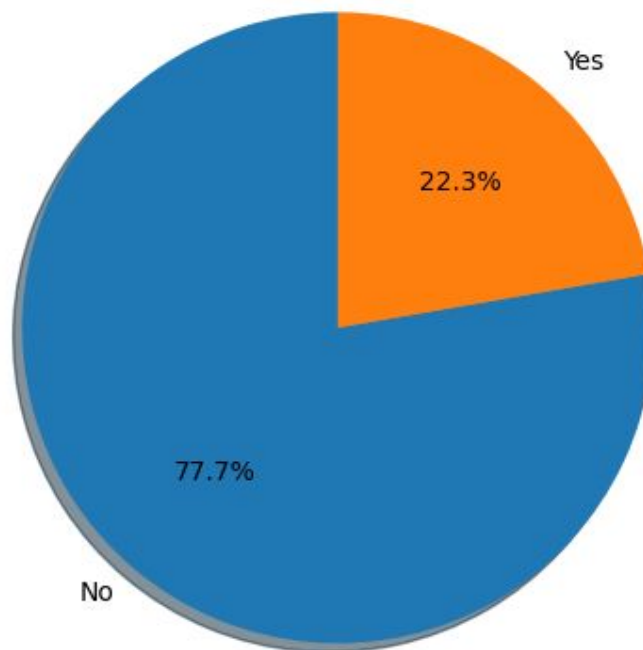
Kolumna	średnia	Max	Min
'MinTemp'	12.6	33.9	-8.2
'MaxTemp'	23.6	48.1	2.6
'WindGustSpeed'	40.7	87.0	2.0
'WindSpeed9am'	15.1	87.0	2.0
'WindSpeed3pm'	19.5	87.0	2.0
'Humidity9am'	67.4	100.0	0.0
'Humidity3pm'	50.6	100.0	0.0
'Pressure9am'	1017.4	1041.0	980.5
'Pressure3pm'	1015.0	1039.6	977.1
'Temp9am'	17.46	40.2	-3.1

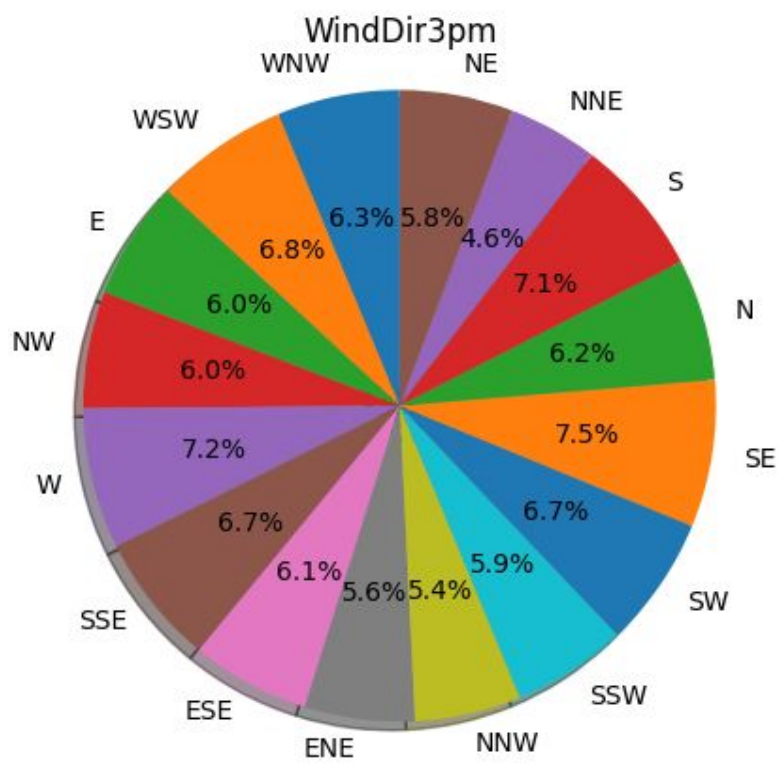
'Temp3pm'	22.1	46.7	1.7
-----------	------	------	-----

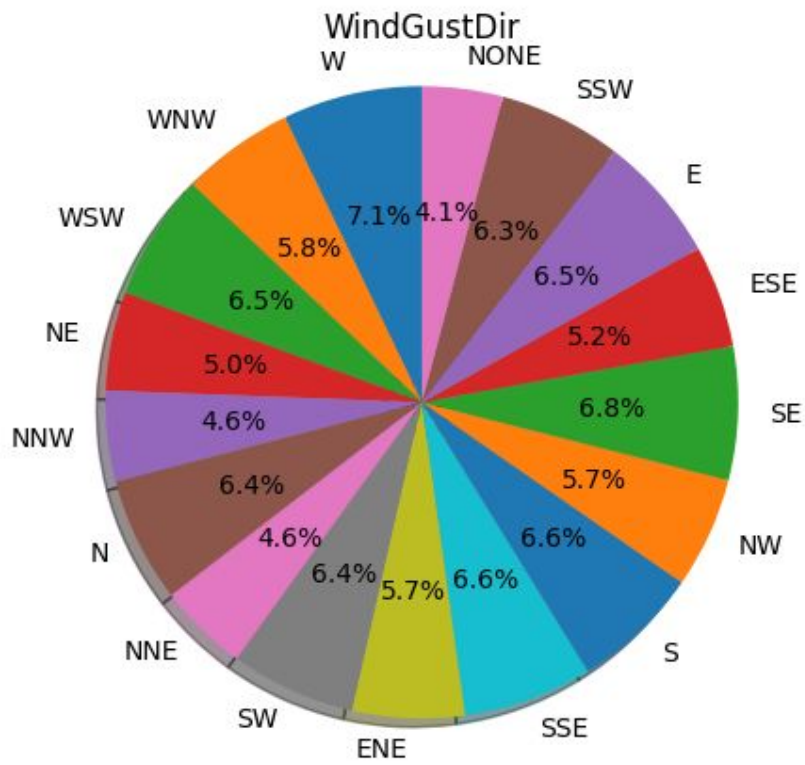
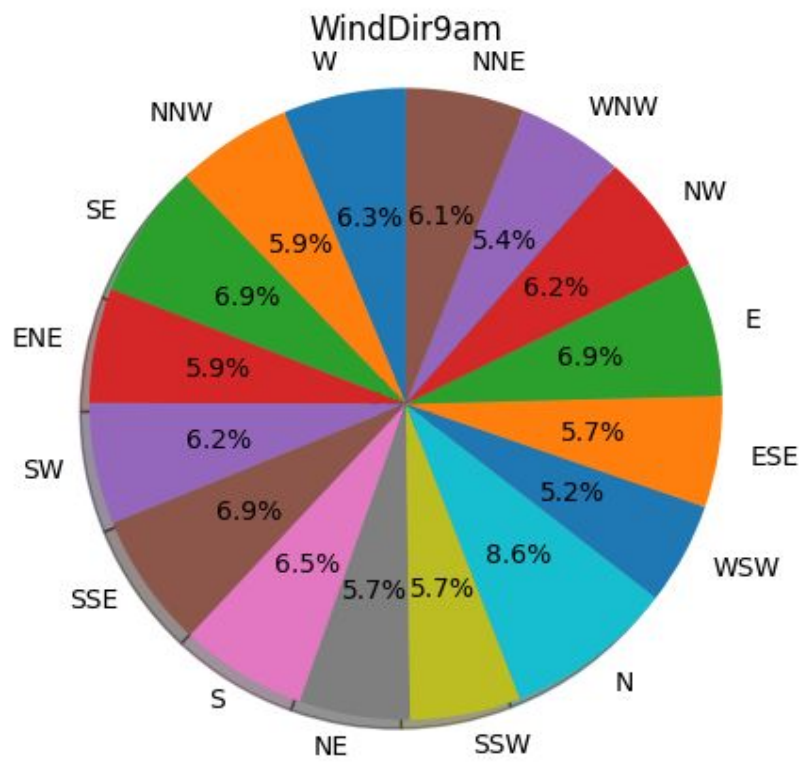
RainToday



RainTomorrow







## 2. Porównanie klasyfikatorów

Użyte klasyfikatory:

- Naive bayes
- KNN-3
- KNN-5
- KNN-11
- KMeans
- Decision Tree
- Random Forest

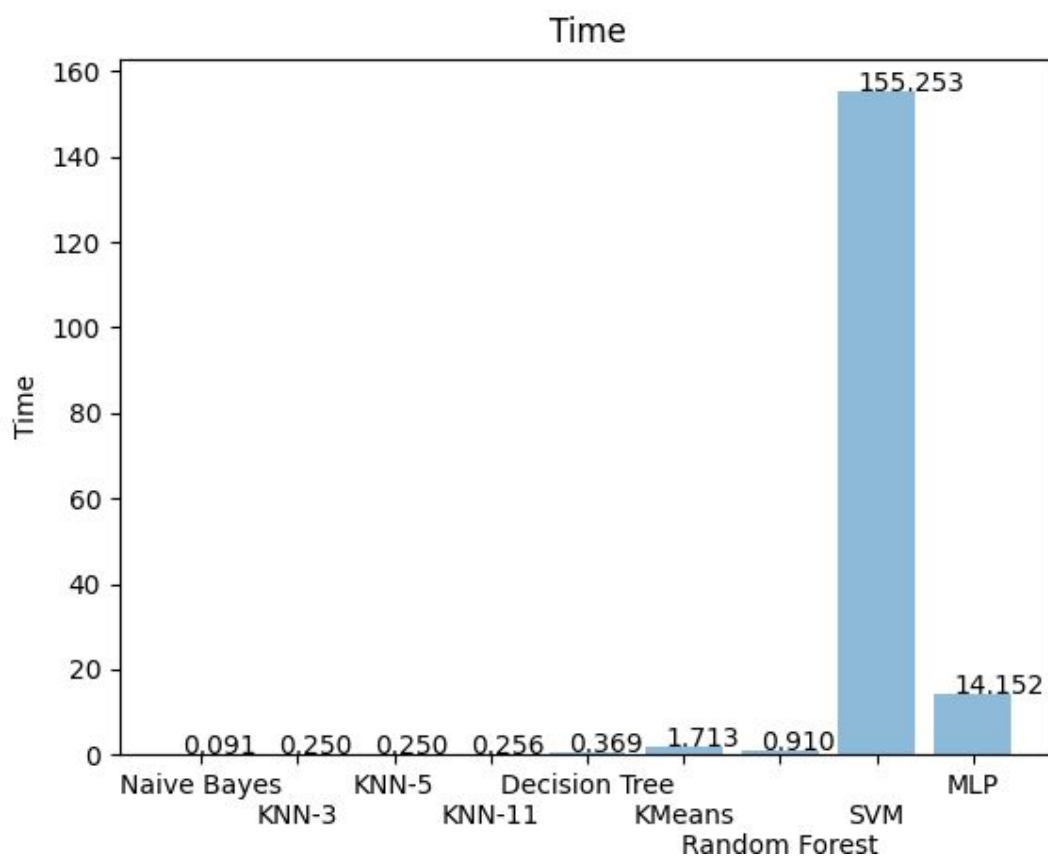
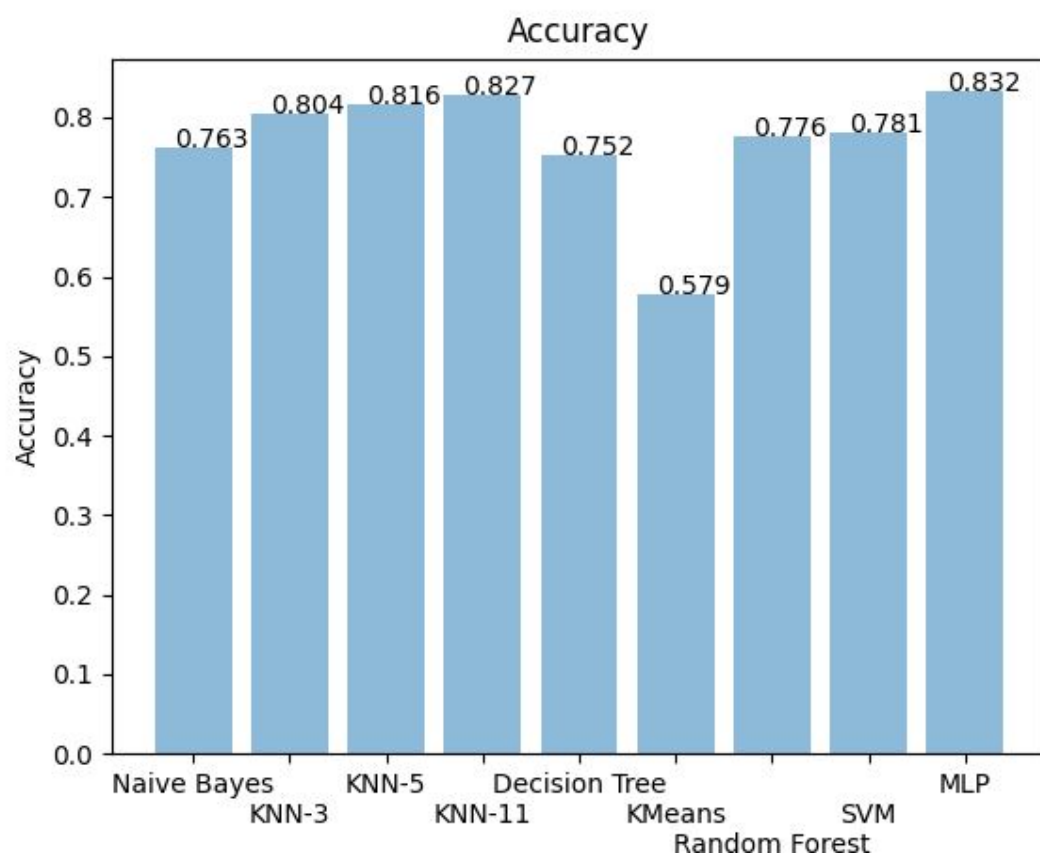
Jest to metoda bazująca na wielu drzewach decyzyjnych.

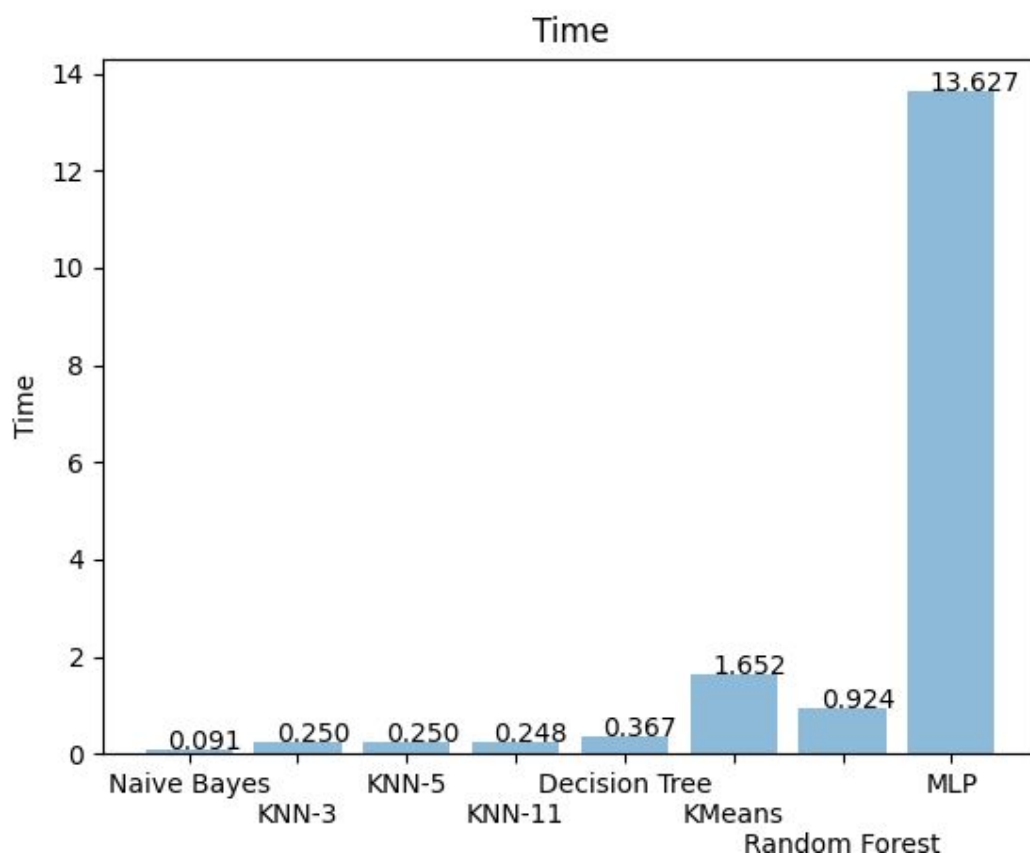
- Support vector machine

Metoda polega na wykorzystywaniu hiperpłaszczyzny do rozdzielania klas. Dla 2 wymiarowego modelu będzie to prosta, dla 3-wymiarowego płaszczyzna.

- Multilayer Perceptron (11, 5, 11)

Reguły asocjacyjne (min\_support = 0.1 min\_confidence = 0.9 )





**Najlepszy klasyfikator MLP**

Prawdziwy(Nie) && Przewidywany(Nie)	Prawdziwy(Nie) && Przewidywany(Tak)
Prawdziwy(Tak) && Przewidywany(Nie)	Prawdziwy(Tak) && Przewidywany(Tak)

Naive Bayes	24409	4520
	4326	4015
KNN-3	26260	2669
	4652	3689
KNN-5	26816	2113
	4761	3580
KNN-11	27362	1567



	4866	3475
Decision Tree	24065	4864
	4434	3907
KMeans	14412	14517
	1139	7202
Random Forest	28929	0
	8341	0
SVM	28248	681
	7463	878
MLP	28108	821
	5422	2919

### 3. Reguły asocjacyjne

Jedyną regułą dla 'będzie padało jutro' jest taka, że Rano i po południu jest niskie ciśnienie atmosferyczne.

```
['Pressure3pm-Low', 'Pressure9am-Low'] -> RainTomorrowYes
```

Wybrane reguły

- Jutro nie będzie padało jeżeli cały dzień jest gorąco
- Jutro nie będzie padał jeżeli wilgotność jest niska
- Jutro nie będzie padał jeżeli dzisiaj nie padało i wilgotność jest niska

```
['Humidity3pm-Low'] -> RainTomorrowNo
['No', 'Humidity3pm-Low'] -> RainTomorrowNo
['Humidity9am-Low'] -> RainTomorrowNo
['No', 'Humidity9am-Low'] -> RainTomorrowNo
['Temp3pm-High', 'MaxTemp-High'] -> RainTomorrowNo
['No', 'Temp3pm-High'] -> RainTomorrowNo
['Pressure9am-High', 'Pressure3pm-High'] -> RainTomorrowNo
['No', 'MaxTemp-High'] -> RainTomorrowNo
['No', 'Pressure9am-High'] -> RainTomorrowNo
['No', 'Temp3pm-High', 'MaxTemp-High'] -> RainTomorrowNo
['No', 'Pressure3pm-High'] -> RainTomorrowNo
['No', 'Pressure9am-High', 'Pressure3pm-High'] -> RainTomorrowNo
['Temp3pm-Low', 'MaxTemp-Low'] -> RainTomorrowNo
['Humidity3pm-Low', 'Humidity9am-Low'] -> RainTomorrowNo
['Temp3pm-High', 'Temp9am-High', 'MaxTemp-High'] -> RainTomorrowNo
['No', 'Humidity3pm-Low', 'Humidity9am-Low'] -> RainTomorrowNo
['No', 'Temp9am-High', 'MaxTemp-High'] -> RainTomorrowNo
['No', 'Temp9am-High', 'Temp3pm-High'] -> RainTomorrowNo
['MaxTemp-High', 'No', 'Temp3pm-High', 'Temp9am-High'] -> RainTomorrowNo
['Temp9am-Low', 'Temp3pm-Low', 'MaxTemp-Low'] -> RainTomorrowNo
['Temp3pm-High', 'Humidity3pm-Low'] -> RainTomorrowNo
['Temp3pm-High', 'MinTemp-High', 'MaxTemp-High'] -> RainTomorrowNo
['No', 'Humidity3pm-Low', 'Temp3pm-High'] -> RainTomorrowNo
['Humidity3pm-Low', 'MaxTemp-High'] -> RainTomorrowNo
['Temp9am-High', 'MinTemp-High', 'MaxTemp-High'] -> RainTomorrowNo
['Temp3pm-High', 'Humidity3pm-Low', 'MaxTemp-High'] -> RainTomorrowNo
['Temp3pm-High', 'Temp9am-High', 'MinTemp-High'] -> RainTomorrowNo
['No', 'Humidity3pm-Low', 'MaxTemp-High'] -> RainTomorrowNo
['MinTemp-High', 'MaxTemp-High', 'Temp3pm-High', 'Temp9am-High'] -> RainTomorrowNo
['Pressure9am-High', 'MinTemp-Low'] -> RainTomorrowNo
['Temp9am-Low', 'MinTemp-Low', 'MaxTemp-Low'] -> RainTomorrowNo
['MaxTemp-High', 'No', 'Temp3pm-High', 'Humidity3pm-Low'] -> RainTomorrowNo
['MinTemp-Low', 'Pressure3pm-High'] -> RainTomorrowNo
['MinTemp-Low', 'Temp3pm-Low', 'MaxTemp-Low'] -> RainTomorrowNo
['Temp3pm-High', 'Humidity9am-Low'] -> RainTomorrowNo
['Humidity9am-Low', 'MaxTemp-High'] -> RainTomorrowNo
['Temp9am-Low', 'MinTemp-Low', 'Temp3pm-Low'] -> RainTomorrowNo
['No', 'Temp3pm-Low', 'MaxTemp-Low'] -> RainTomorrowNo
```