



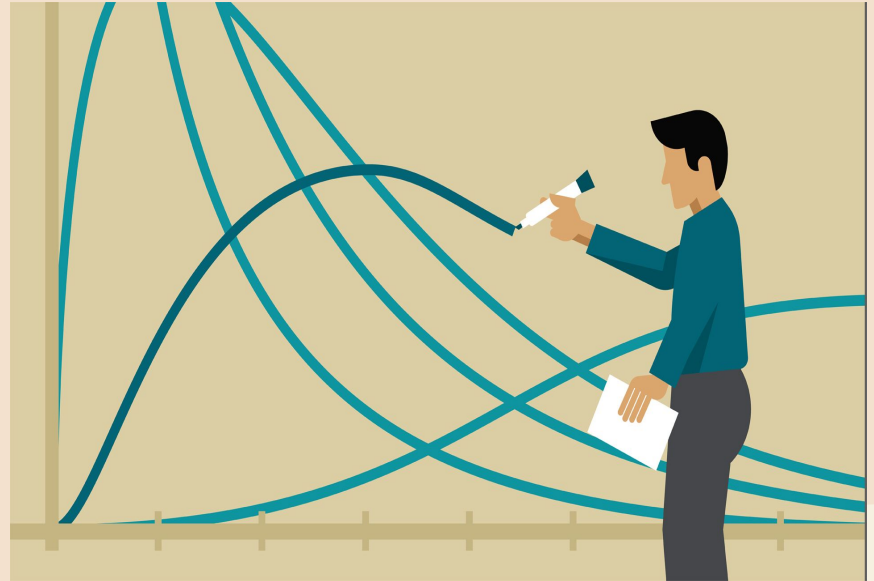
Prezentacja 6

24.05.2023r.



MODEL OLS

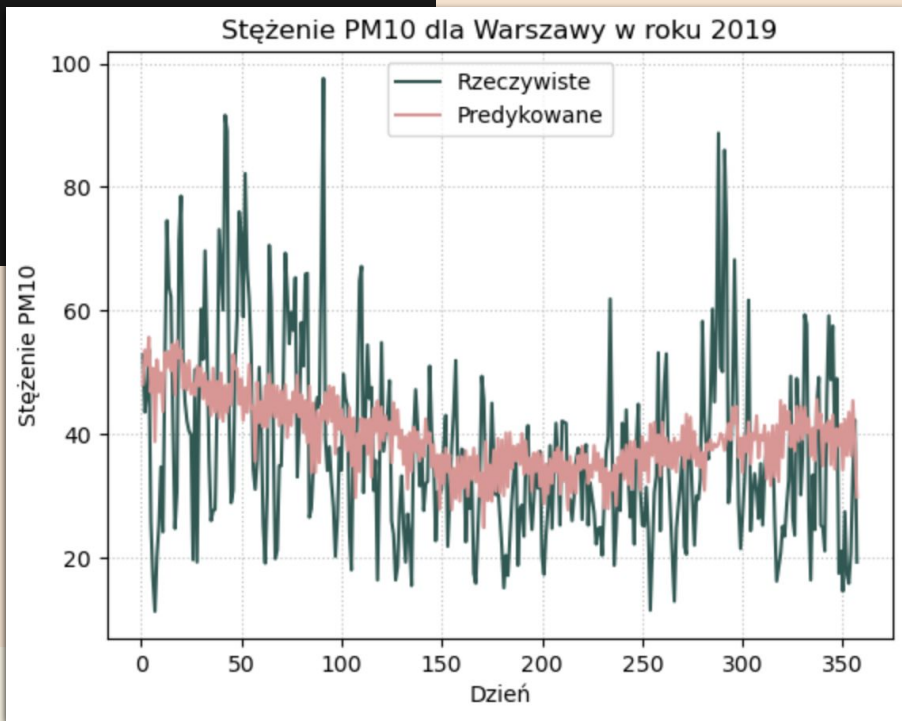
Model OLS (Ordinary Least Squares) - jeden z podstawowych modeli regresji w analizie statystycznej. Jest używany do estymacji parametrów i prognozowania wartości zmiennej zależnej na podstawie zmiennych niezależnych.



ZASTOSOWANIE sm.OLS

```
x_train = PM10_train_War[['Temperature', 'Wind', 'YEAR', 'MONTH']]
x_test = PM10_test_War[['Temperature', 'Wind', 'YEAR', 'MONTH']]
y_train = PM10_train_War[['Value']]
y_test = PM10_test_War[['Value']]
model = sm.OLS(y_train, x_train)
results = model.fit()

y_pred=results.predict(x_test)
```



```
print(results.summary())
```



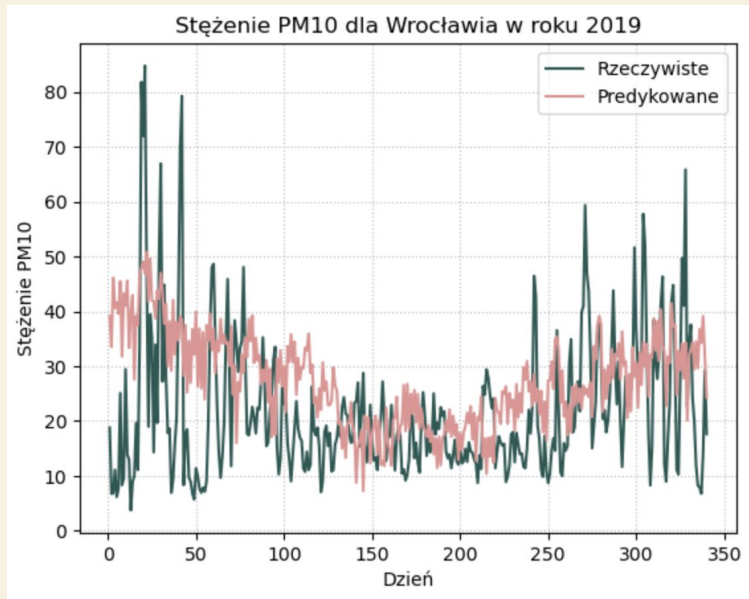
OLS Regression Results

```
=====
Dep. Variable:          Value    R-squared (uncentered):      0.809
Model:                  OLS      Adj. R-squared (uncentered):    0.808
Method:                 Least Squares    F-statistic:          1469.
Date:                  Mon, 22 May 2023    Prob (F-statistic):      0.00
Time:                  23:44:30    Log-Likelihood:        -6140.8
No. Observations:      1396    AIC:                   1.229e+04
Df Residuals:          1392    BIC:                   1.231e+04
Df Model:               4
Covariance Type:       nonrobust
=====
```

```
=====
              coef      std err          t      P>|t|      [0.025      0.975]
-----
Temperature    -0.5217      0.064     -8.194      0.000     -0.647     -0.397
Wind            -1.7562      0.260     -6.754      0.000     -2.266     -1.246
YEAR             0.0261      0.001     40.613      0.000      0.025      0.027
MONTH          -0.6731      0.158     -4.261      0.000     -0.983     -0.363
=====
```

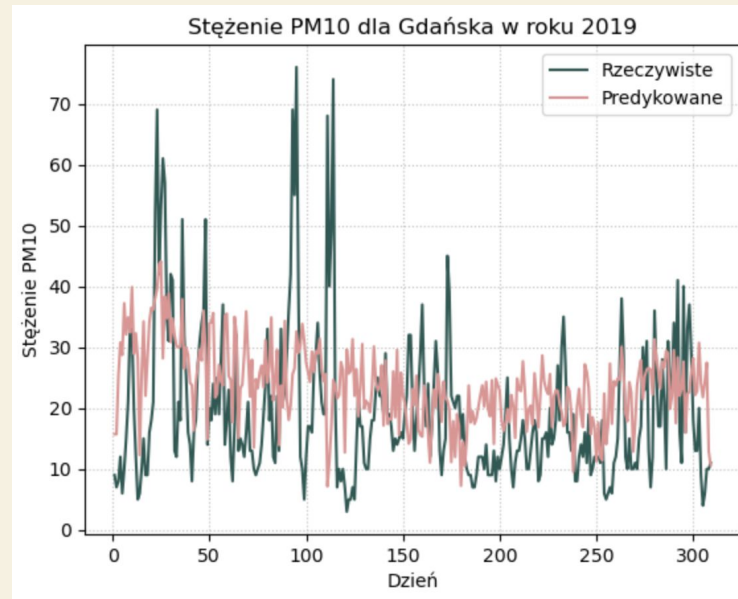
```
=====
Omnibus:          346.472    Durbin-Watson:           0.890
Prob(Omnibus):    0.000    Jarque-Bera (JB):        967.681
Skew:             1.271    Prob(JB):                7.43e-211
Kurtosis:         6.189    Cond. No.                994.
=====
```

Notes:



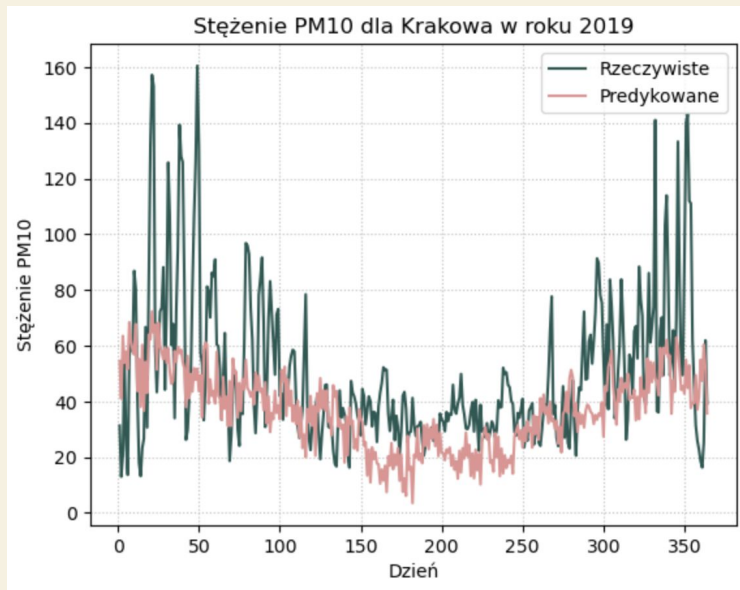
$$\sqrt{MSE} = 14.52548$$

	Współczynnik:
Wiatr	-1.5684
Temperatura	-1.0244
Rok	0.0225
miesiąc	-0.4766



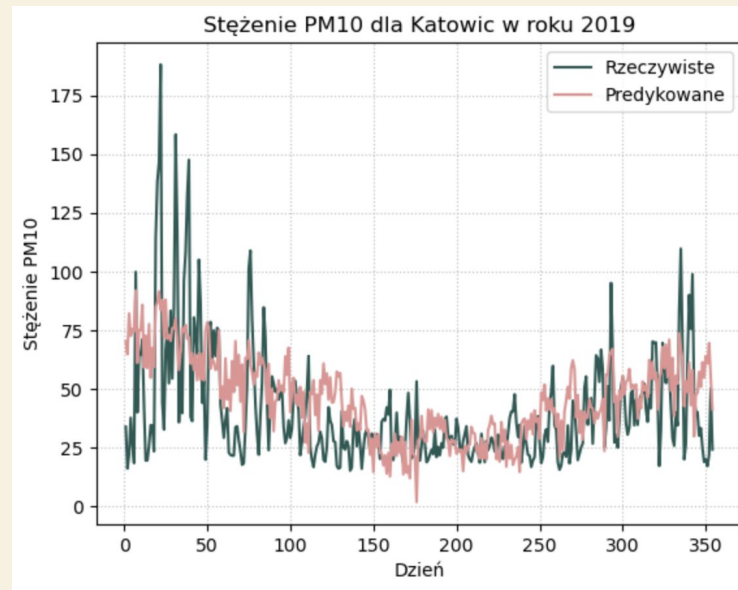
$$\sqrt{MSE} = 12.97182$$

	Współczynnik:
Wiatr	-2.9518
Temperatura	-0.5011
Rok	0.0223
miesiąc	-0.7300



$$\sqrt{MSE} = 27.25248$$

	Współczynnik:
Wiatr	-2.5295
Temperatura	-1.6943
Rok	0.0292
miesiąc	-0.0675



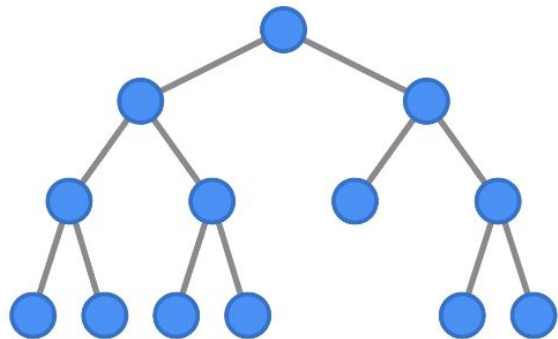
$$\sqrt{MSE} = 22.68565$$

	Współczynnik:
Wiatr	-3.3856
Temperatura	-2.0401
Rok	0.0380
miesiąc	-0.7741

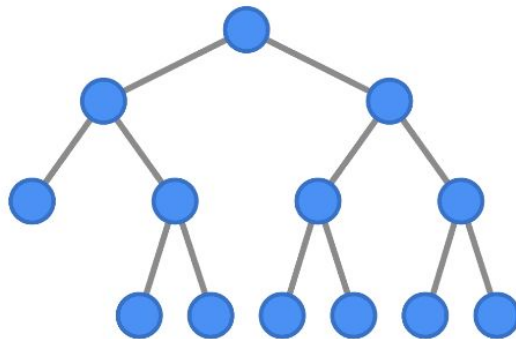
DZIAŁANIE LASU LOSOWEGO

EXAMPLES

Tree-1

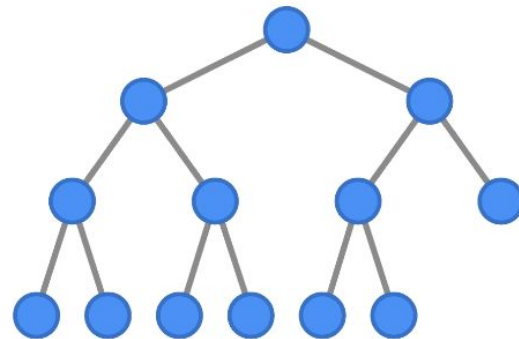


Tree-2

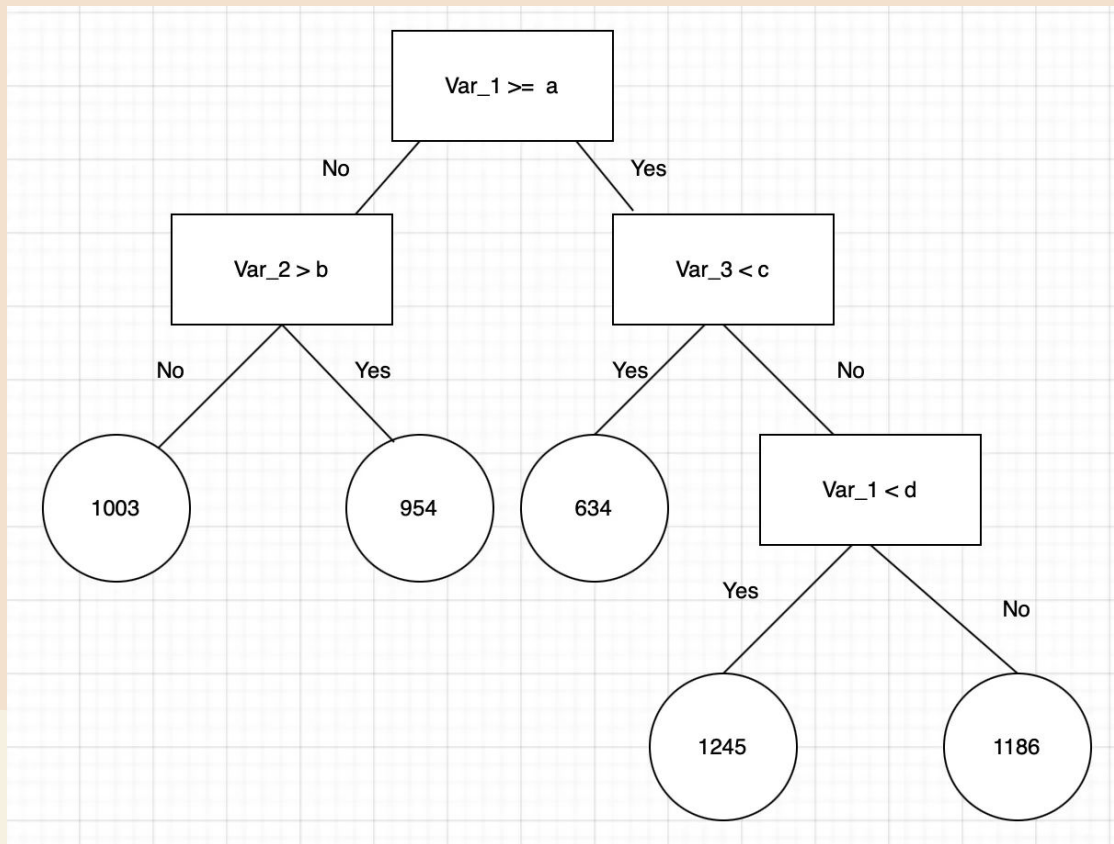


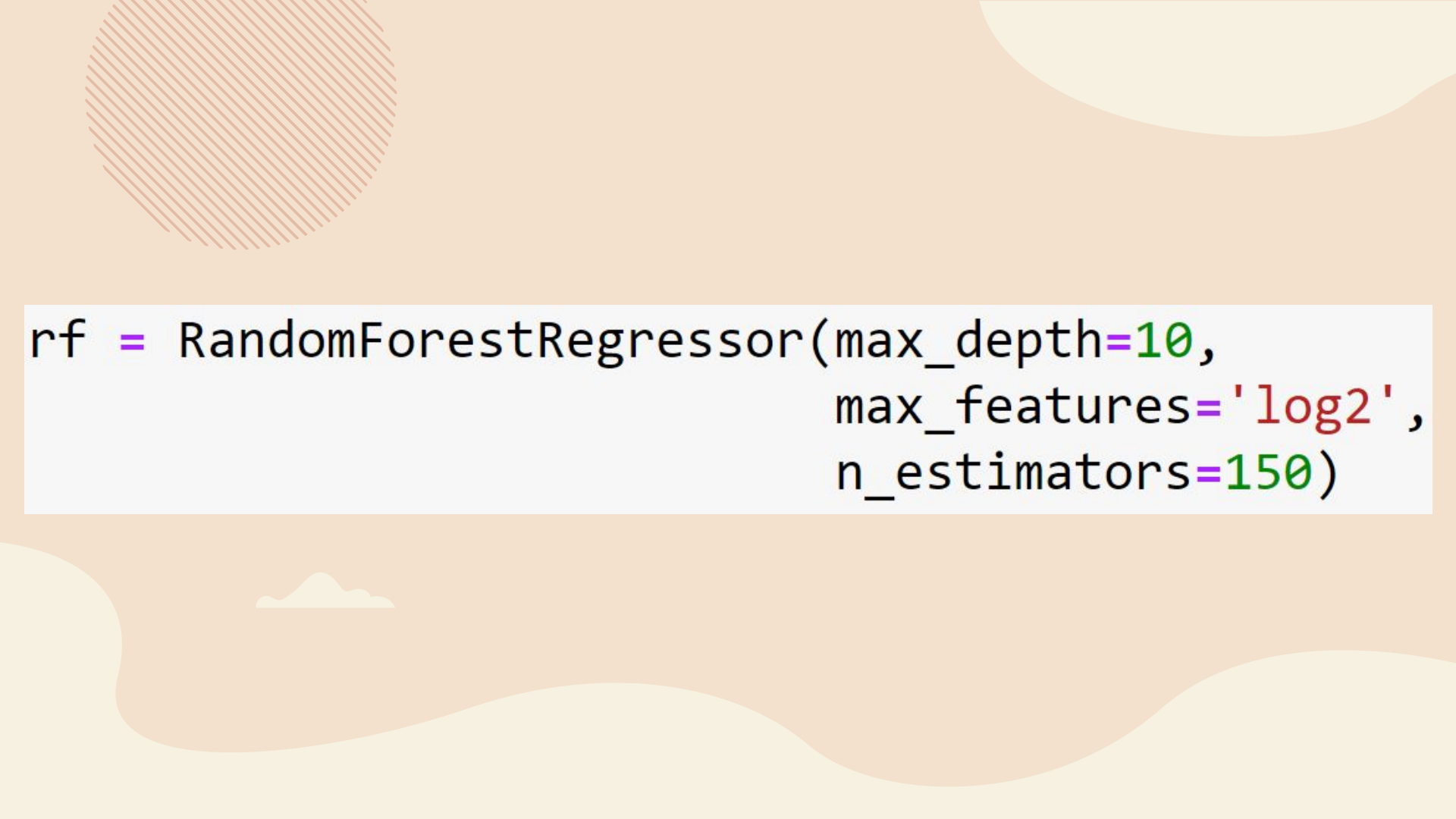
...

Tree-n



DZIAŁANIE LASU LOSOWEGO





```
rf = RandomForestRegressor(max_depth=10,  
                           max_features='log2',  
                           n_estimators=150)
```

WYNIKI PREDYKCJI



Dane treningowe

Stężenie PM10 w latach 2015-2018.

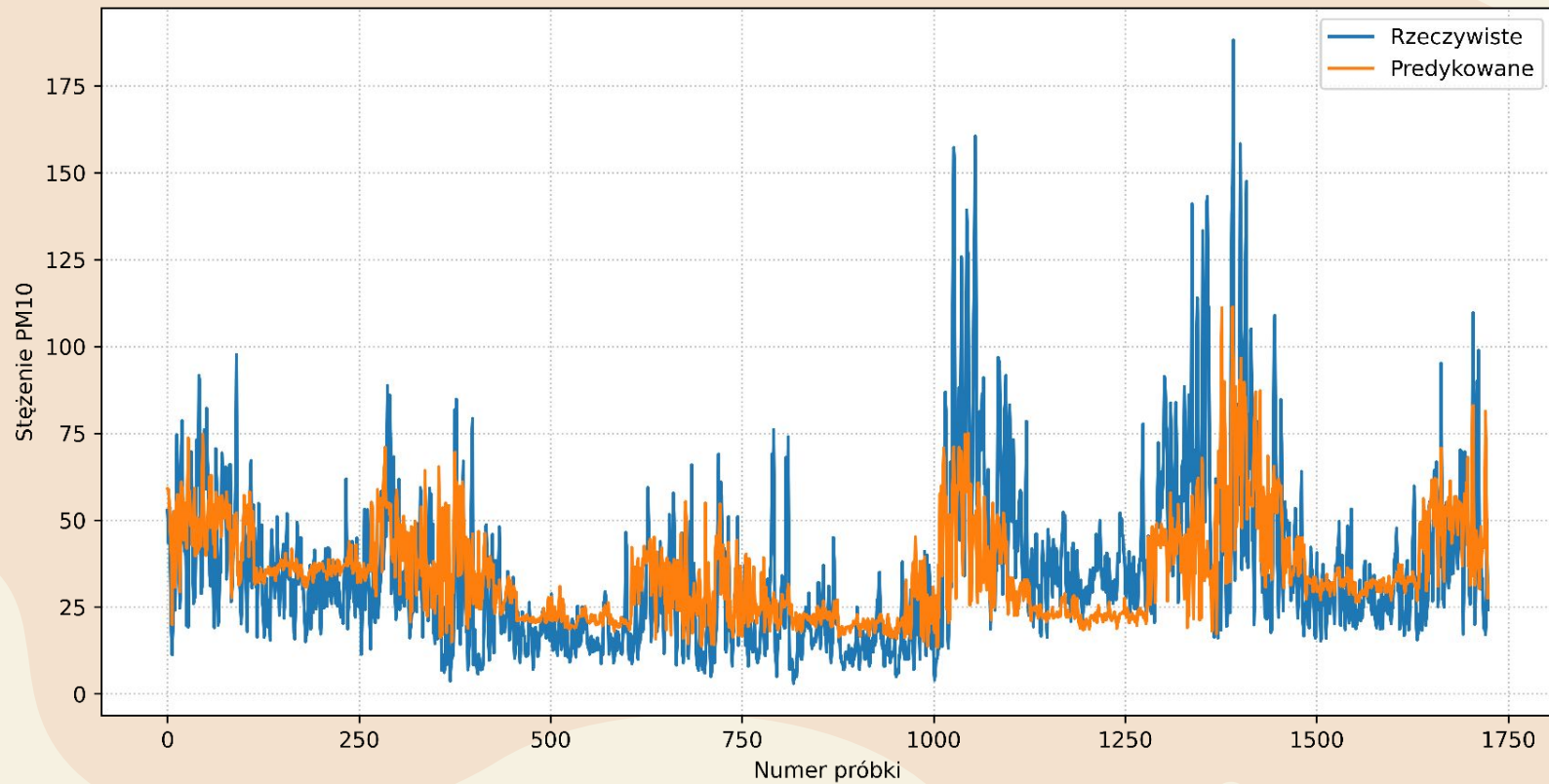
Dane testowe

Stężenie PM10 w roku 2019.

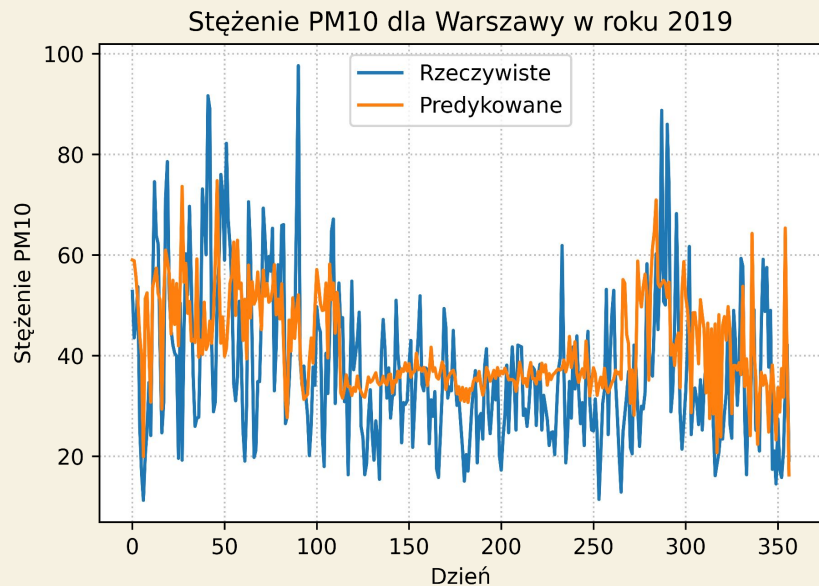
Zmienne wzięte pod uwagę:

- miasto,
- temperatura,
- wiatr,
- rok,
- miesiąc.

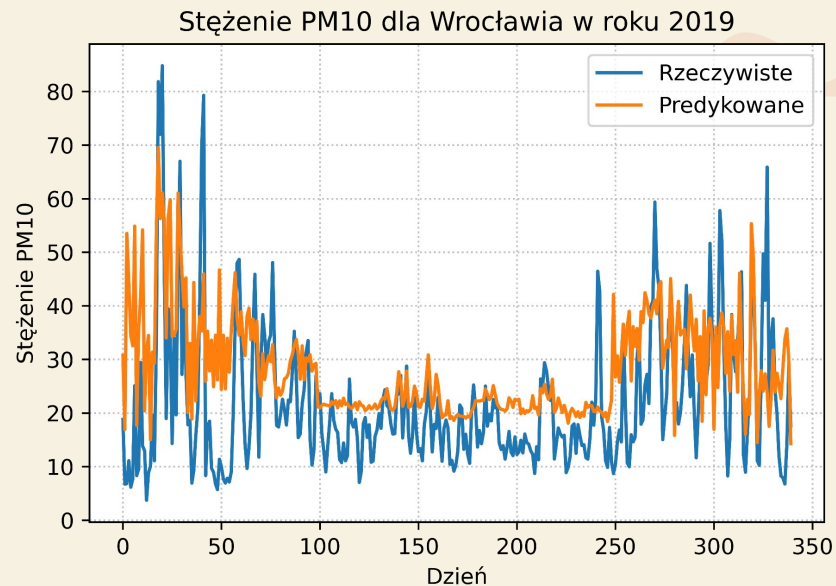
Stężenie PM10 w roku 2019



	Błąd średniokwadratowy (MSE)	Pierwiastek z błędu średniokwadratowego	Średni błąd bezwzględny (MAE)	Średni bezwzględny błąd procentowy (MAPE)	Współczynnik determinacji R ²
Wzór	$\frac{1}{n} \sum_i^n (y_i - \hat{y}_i)^2$	$\sqrt{\frac{1}{n} \sum_i^n (y_i - \hat{y}_i)^2}$	$\frac{1}{n} \sum_1^n y_i - \hat{y}_i $	$\frac{1}{n} \sum_{i=1}^n \left \frac{y_i - \hat{y}_i}{y_i} \right $	$\frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$
Wynik dla danych treningowych	185,25862 (0,95263)	13,61093 (0,03499)	9,24818 (0,01798)	0,33483 (0,00077)	0,71781 (0,00145)
Wynik dla danych testowych	330,25786 (1,66448)	18,17294 (0,04580)	12,81912 (0,03009)	0,50327 (0,00155)	0,36349 (0,00321)



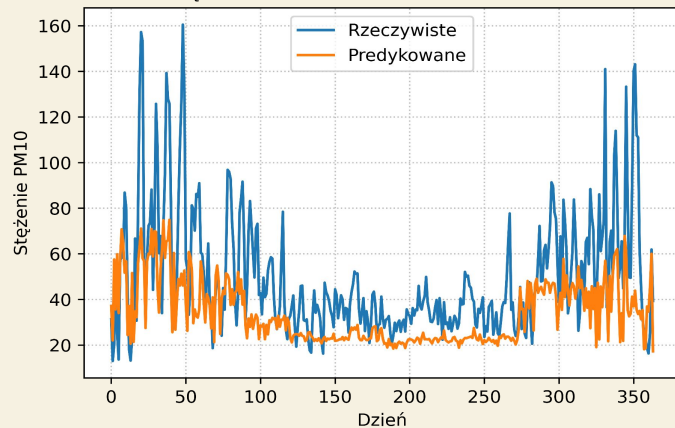
$$\sqrt{MSE} = 14.76868$$



$$\sqrt{MSE} = 12.83363$$

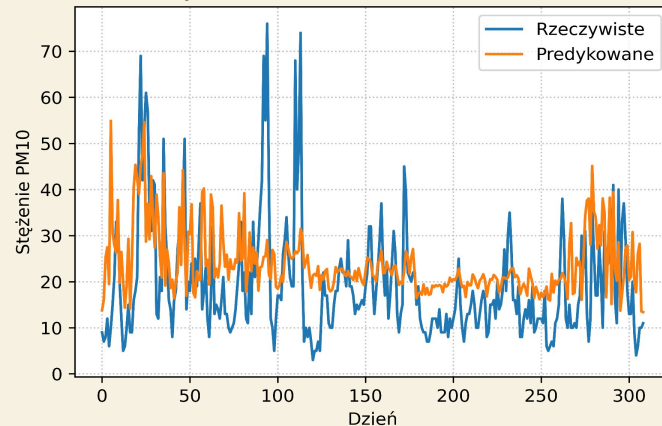


Stężenie PM10 dla Krakowa w roku 2019



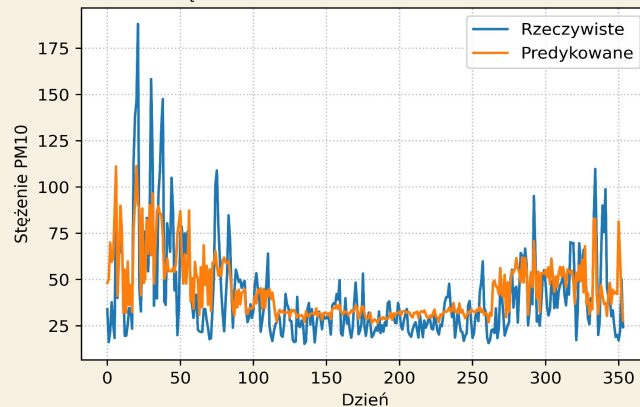
$$\sqrt{MSE} = 26.77637$$

Stężenie PM10 dla Gdańska w roku 2019



$$\sqrt{MSE} = 12.03381$$

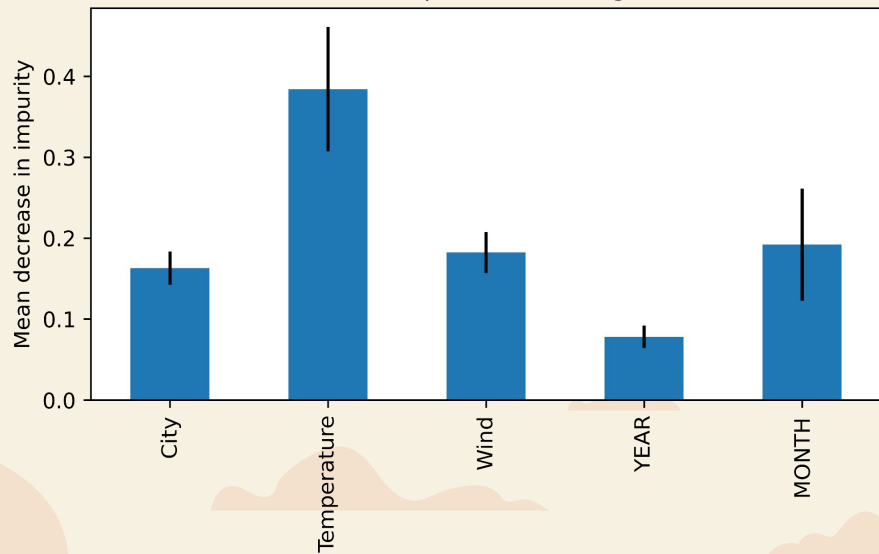
Stężenie PM10 dla Katowic w roku 2019



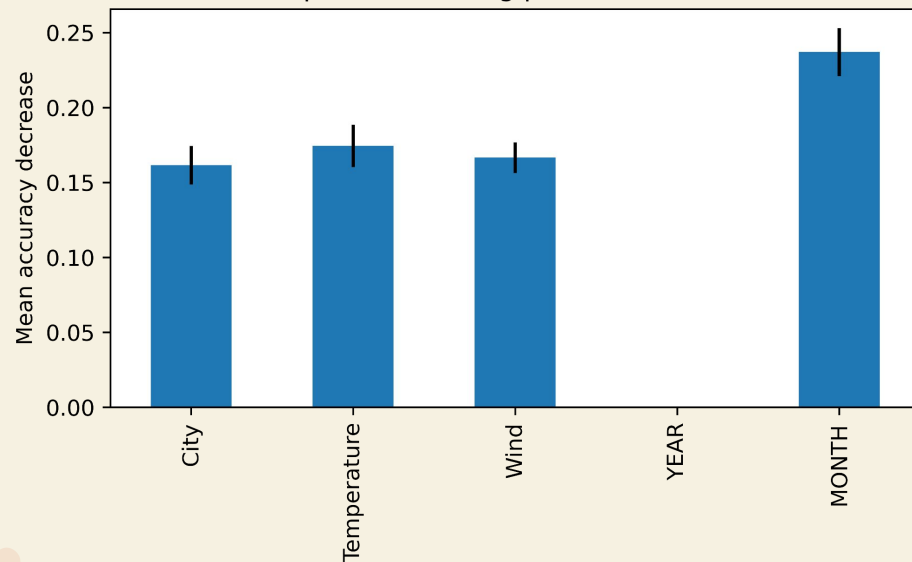
$$\sqrt{MSE} = 18.85465$$

WAŻNOŚĆ CECH

Feature importances using MDI



Feature importances using permutation on full model



DZIĘKUJEMY ZA UWAGĘ!

PREZENTACJĘ PRZYGOTOWALI:

Paulina Iwach

Julia Mazur

Ewa Trębacz

Małgorzata Kowalczyk

Kamil Kowalski

