PREDICTEX

UNIVERSITY OF MANNHEIM
School of Business Informatics and Mathematics

Data Mining Group Project

6th of December 2023

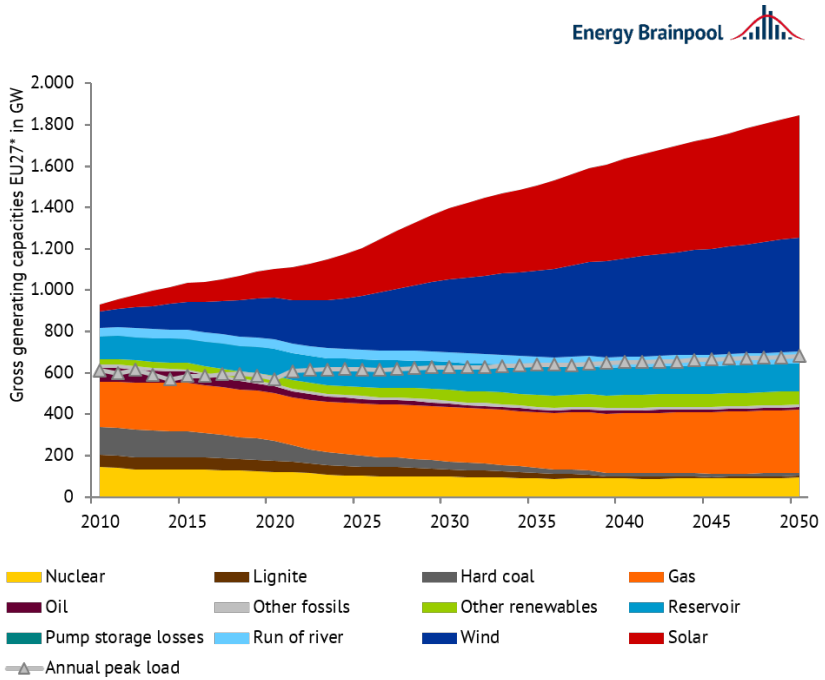
Max Lautenbach (1980683), Niklas Weidenfeller (1977441),
Lara-Aida Jopp (1978974), Babett Müller (Babett Müller (1979887),
Gregor Münker (1980671), Maximilian Heilmann (1979887)

UNIVERSITY OF MANNHEIM
School of Business Informatics and Mathematics

What is residual load?

Why predict the residual load?

UNIVERSITY OF MANNHEIM
School of Business Informatics and Mathematics



Handelsblatt

Energy industry sees security of supply at risk

The energy industry is calling on the German government to come up with a coherent concept for the construction of new power plants. Under the current conditions, security of supply is at risk.

McKinsey

Energy transition index from McKinsey: Security of supply under tension

Impending electricity shortfall: Peak load may exceed available capacity by 4 GW in 2025 and 30 GW in 2030 - expansion of renewables alone is not enough...

Data Selection

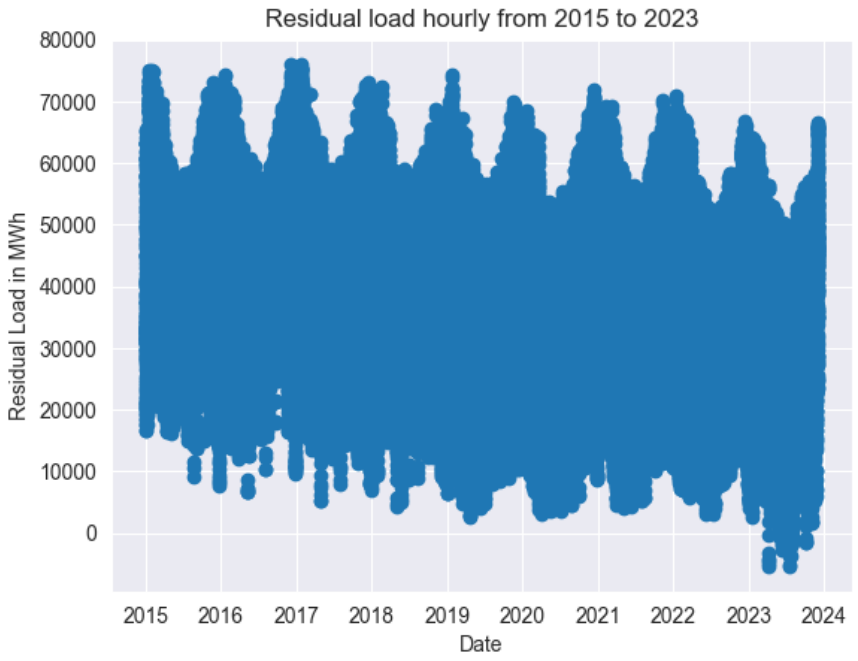
UNIVERSITY OF MANNHEIM
School of Business Informatics and Mathematics

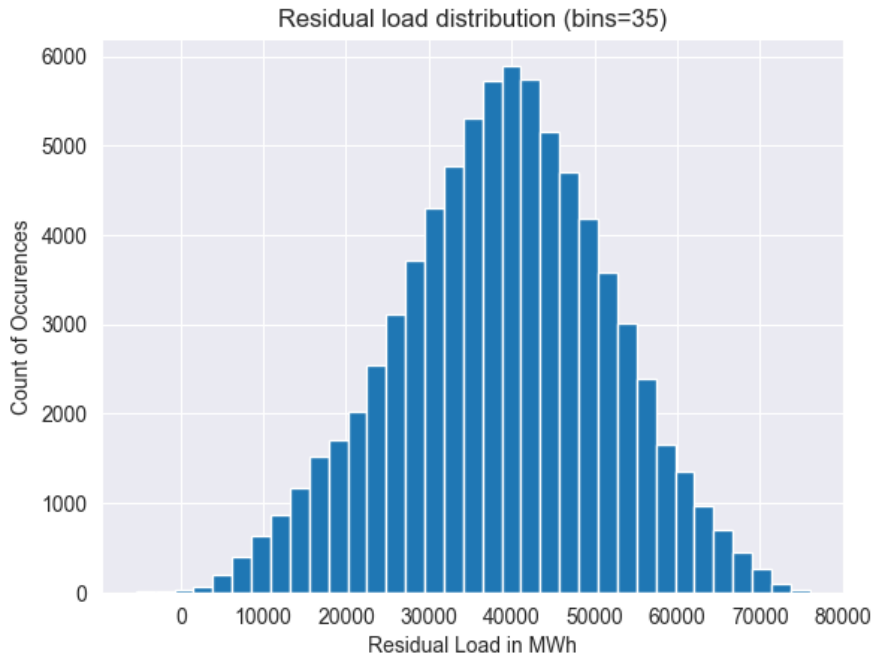
- Weather Data
  - Sun Duration
  - Precipitation Amount
  - Wind Velocity
  - Air Temperature
- Residual Load Data

Predict Residual Load

Data Exploration

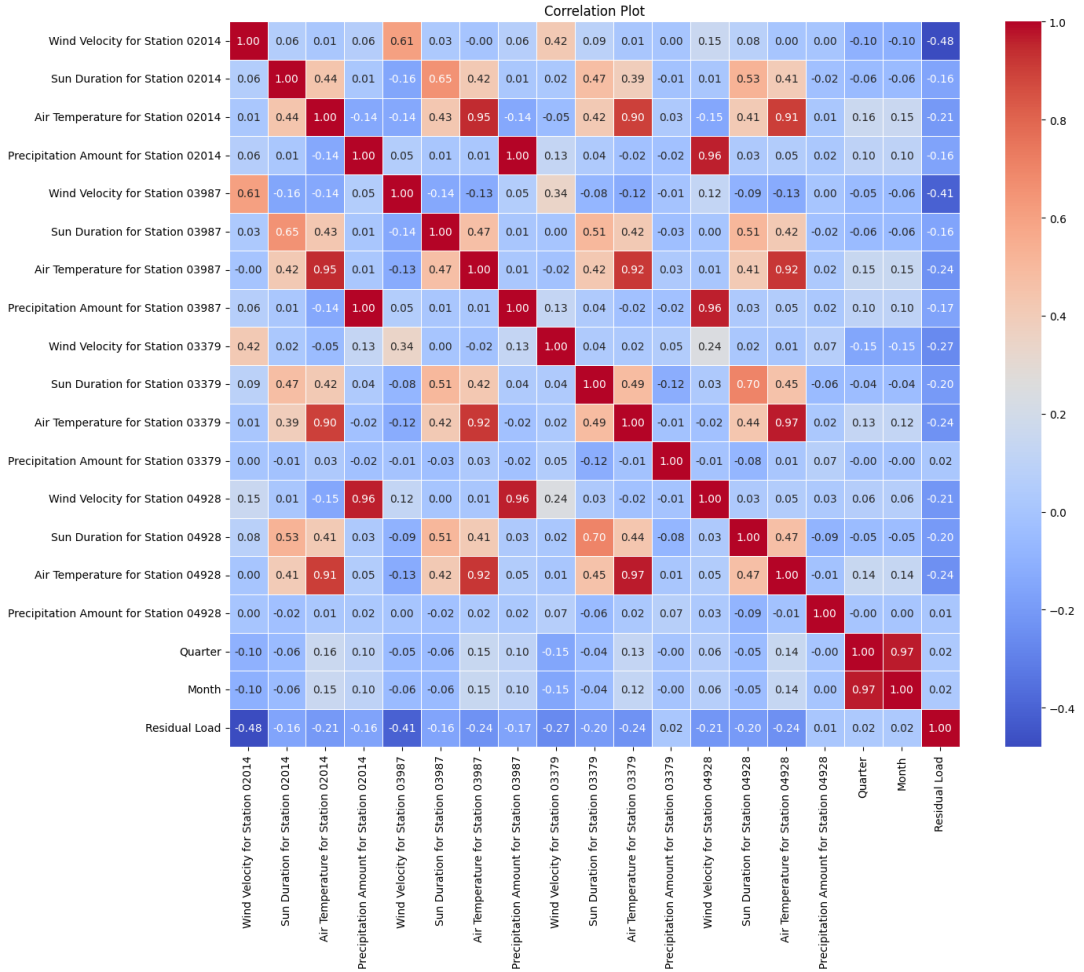
UNIVERSITY OF MANNHEIM
School of Business Informatics and Mathematics





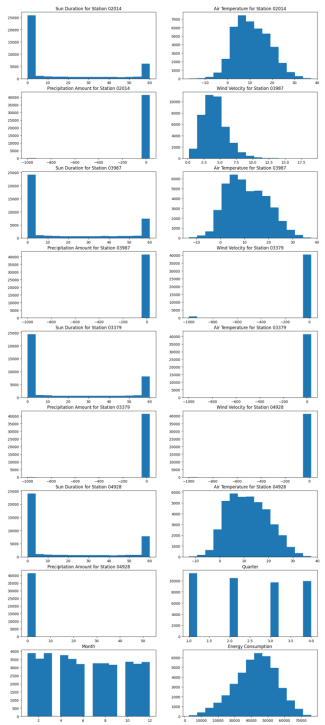
Data Exploration

UNIVERSITY OF MANNHEIM
School of Business Informatics and Mathematics

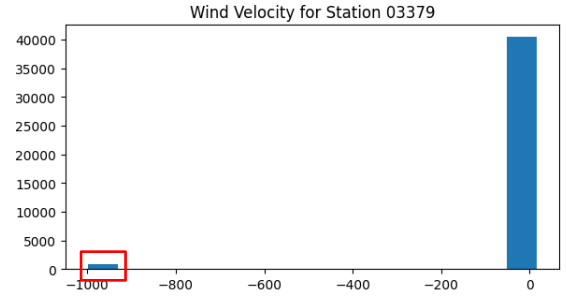


Preprocessing

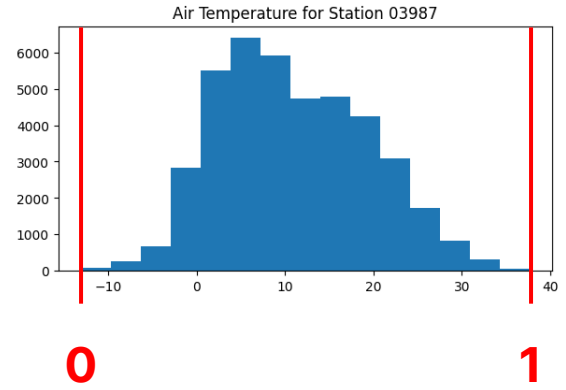
UNIVERSITY OF MANNHEIM
School of Business Informatics and Mathematics



Increase Quality
Delete missing values

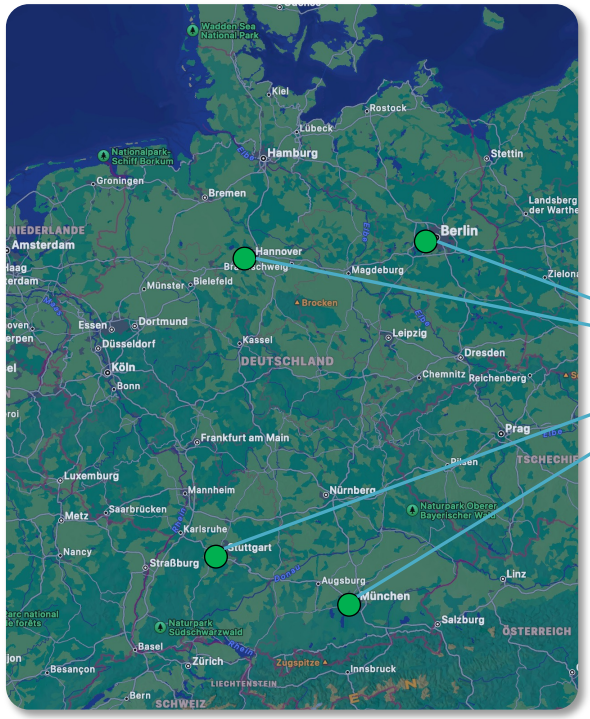


Data Normalizing
Min-Max-Scaler



Data Transformation

UNIVERSITY OF MANNHEIM
School of Business Informatics and Mathematics



Join on Date

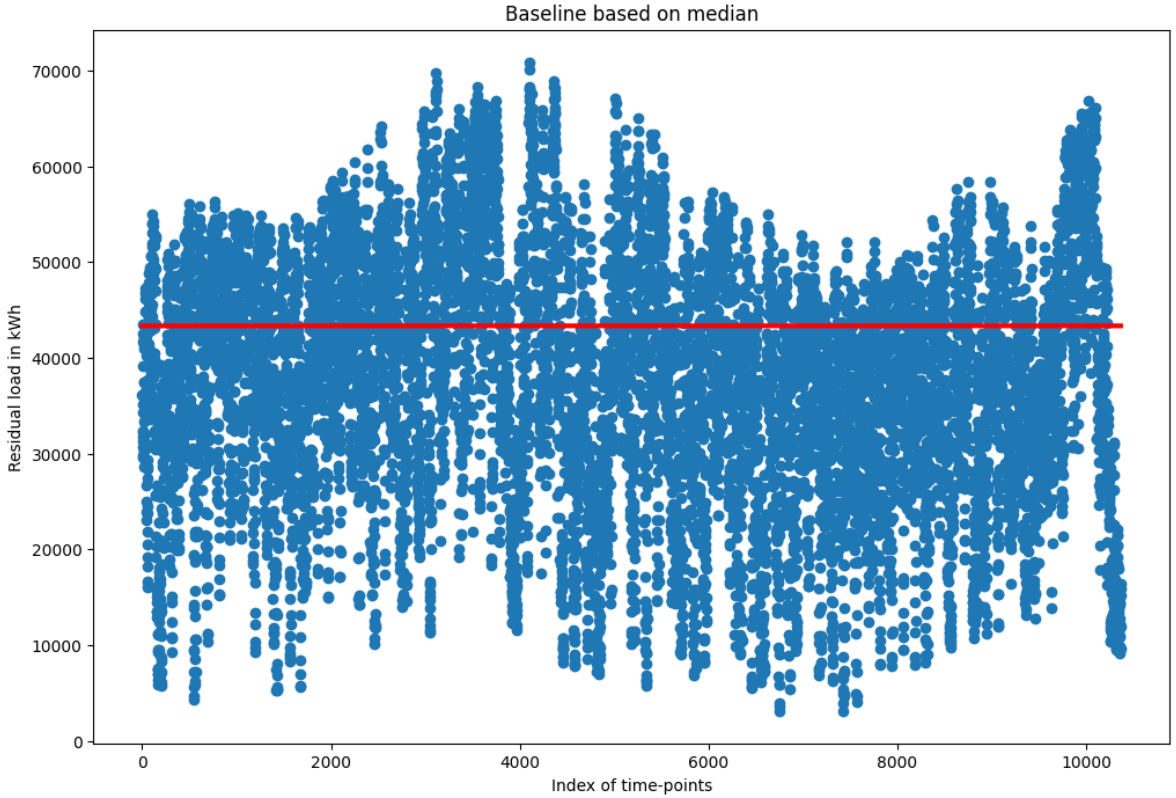
+

Extend Date by

- Quarter
- Month

Baseline

UNIVERSITY OF MANNHEIM
School of Business Informatics and Mathematics

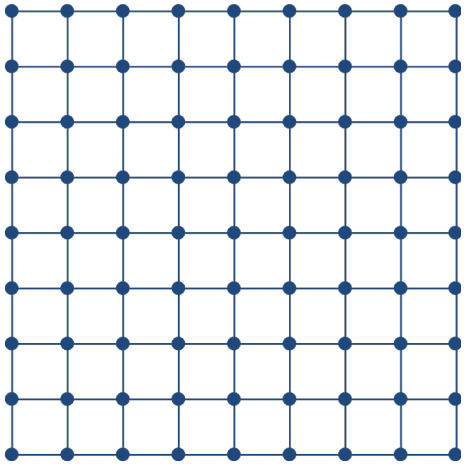


RMSE (Baseline):
14290.49 MWh

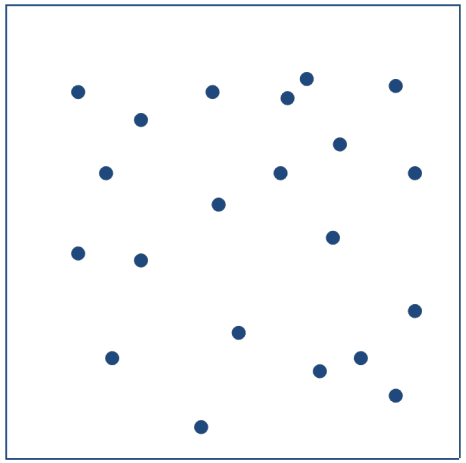
Hyperparameter Tuning

UNIVERSITY OF MANNHEIM
School of Business Informatics and Mathematics

GridSearchCV

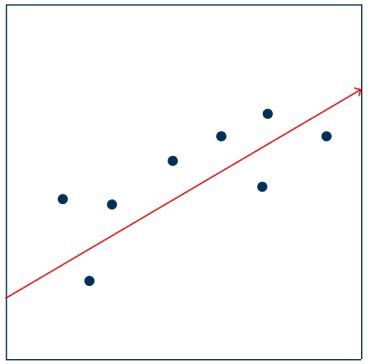


RandomizedSearchCV



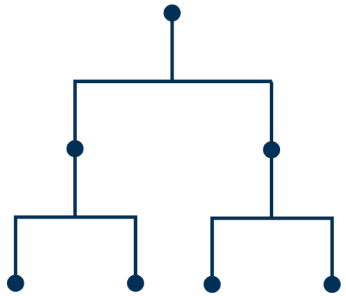
Used Algorithms and Evaluation Critera

UNIVERSITY OF MANNHEIM
School of Business Informatics and Mathematics



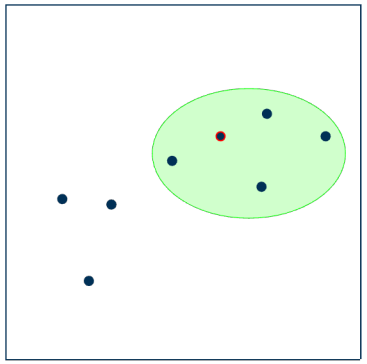
Simple Regression

(Linear, Ridge, Lasso)



Tree-Based-Methods

(Decision Tree, Random Forest)



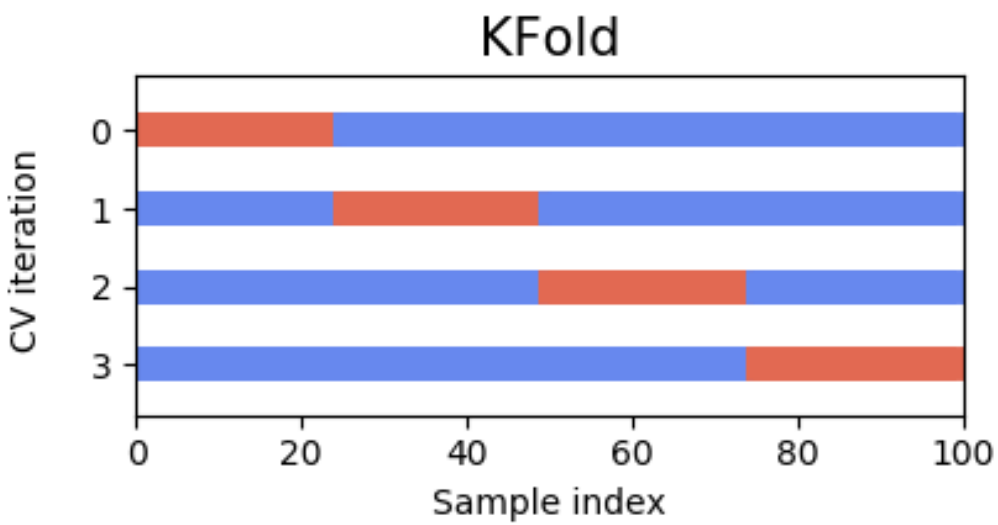
kNN-Algorithm

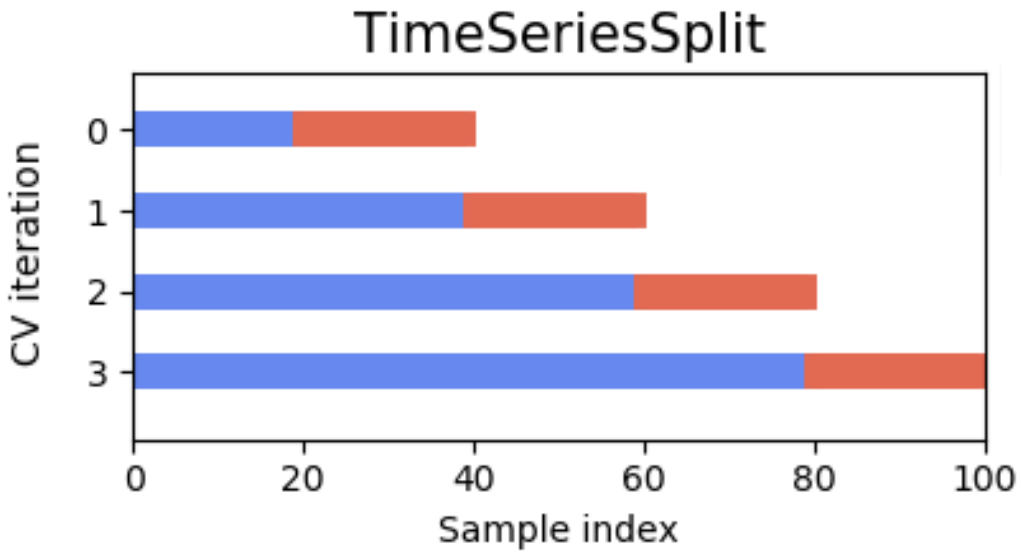
Evaluation Criterion:

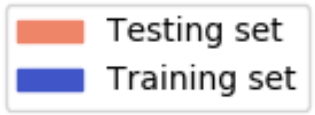
RMSE

Time Series data & Cross-Validation

UNIVERSITY OF MANNHEIM
School of Business Informatics and Mathematics



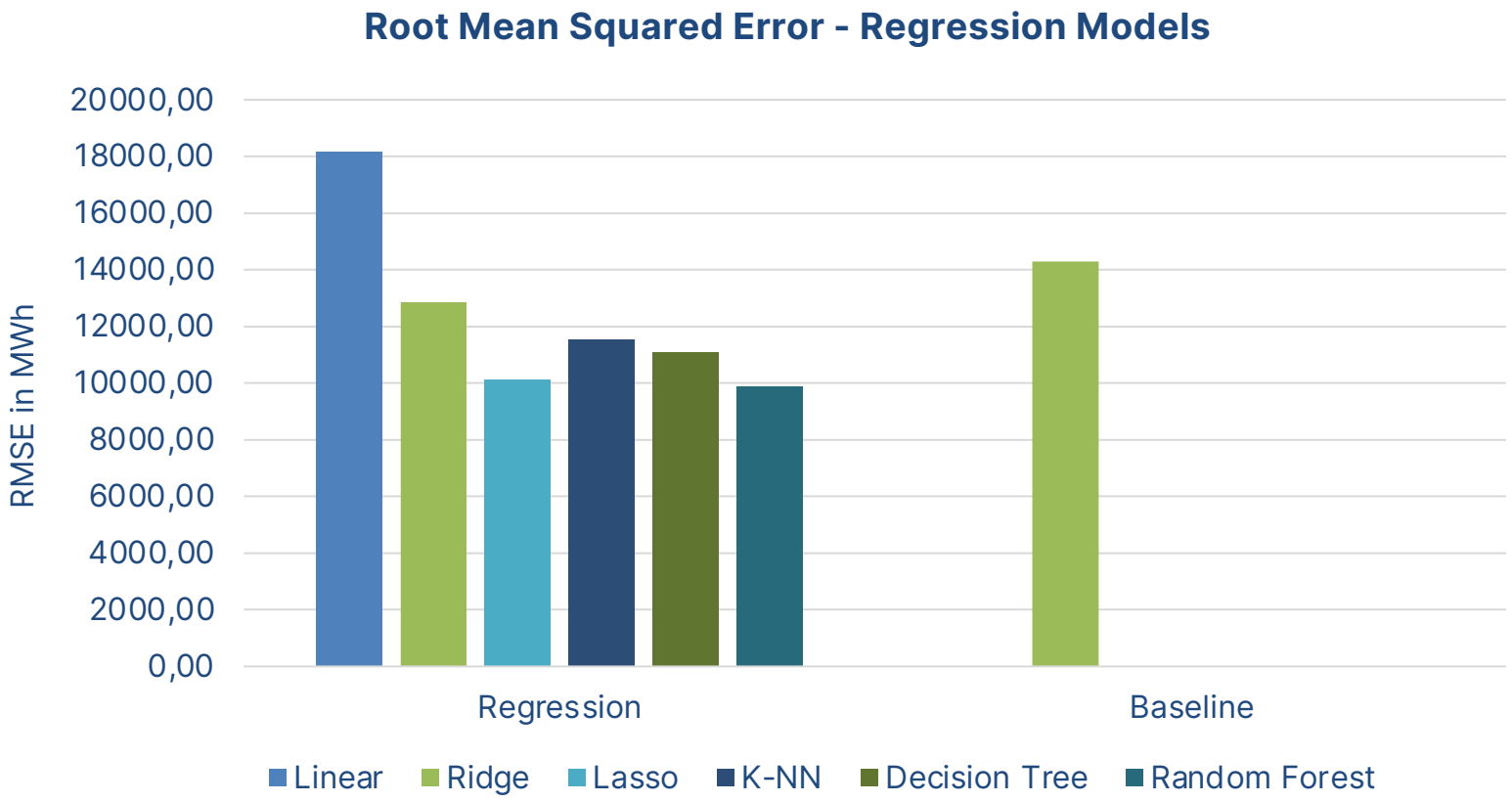




Source: https://medium.com/@soumyachess1496/cross-validation-in-time-series-566ae4981ce4

Data Mining – Application of Regression Models

UNIVERSITY OF MANNHEIM
School of Business Informatics and Mathematics



Result Discussion & Recommendations

UNIVERSITY OF MANNHEIM
School of Business Informatics and Mathematics

Results
Better than baseline

Error
Too high for reliable power-grid operation

Too few
influencers in dataset

Behavioral Aspect
not included in our regressions

Questions?

UNIVERSITY OF MANNHEIM
School of Business Informatics and Mathematics

Max Lautenbach (1980683), Niklas Weidenfeller (1977441),
Lara-Aida Jopp (1978974), Babett Müller (1979887),
Gregor Münker (1980671), Maximilian Heilmann (1979887)

Appendix

UNIVERSITY OF MANNHEIM
School of Business Informatics and Mathematics

Max Lautenbach (1980683), Niklas Weidenfeller (1977441),
Lara-Aida Jopp (1978974), Babett Müller (1979887),
Gregor Münker (1980671), Maximilian Heilmann (1979887)

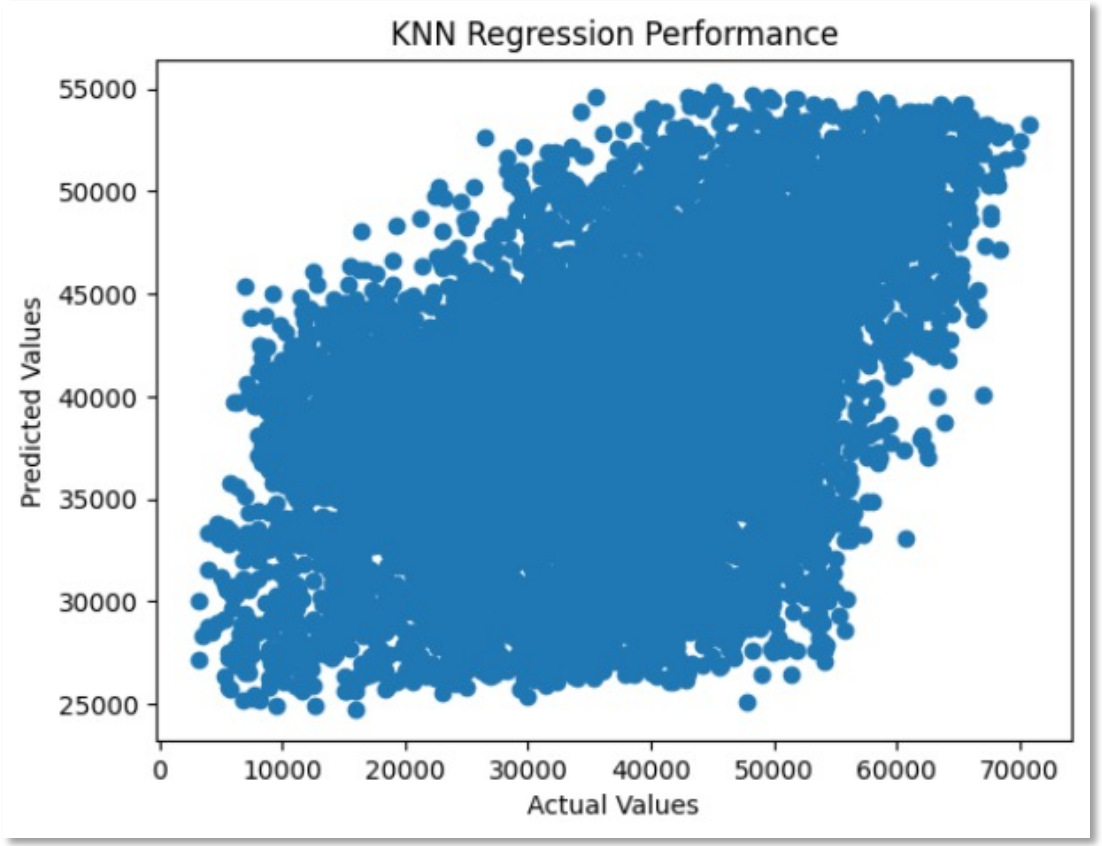
Data Exploration

UNIVERSITY OF MANNHEIM
School of Business Informatics and Mathematics

	Wind Velocity for Station 02014	Sun Duration for Station 02014	Air Temp. for Station 02014	Preicp. Amount for Station 02014	Wind Velocity for Station 03987	Sun Duration for Station 03987	Air Temp. for Station 03987	Precip. Amount for Station 03987	Wind Velocity for Station 03379	Sun Duration for Station 03379	Air Temp. for Station 03379	Precip. Amount for Station 03379	Wind Velocity for Station 04928	Sun Duration for Station 04928	Air Temp. for Station 04928	Precip. Amount for Station 04928	Quarter	Month	Year	Energy Consumption
count	40496,00	40496,00	40496,00	40496,00	40496,00	40496,00	40496,00	40496,00	40496,00	40496,00	40496,00	40496,00	40496,00	40496,00	40496,00	40496,00	40496,00	40496,00	40496,00	40496,00
mean	0,21	0,26	0,52	0,00	0,21	0,30	0,48	0,00	0,18	0,30	0,50	0,00	0,19	0,30	0,49	0,00	0,49	6,41	2017,78	42417,15
std	0,11	0,38	0,14	0,01	0,10	0,40	0,16	0,02	0,10	0,41	0,17	0,02	0,10	0,41	0,16	0,01	0,37	3,44	1,84	12562,49
min	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	2015,00	2550,00
25%	0,12	0,00	0,41	0,00	0,14	0,00	0,35	0,00	0,11	0,00	0,37	0,00	0,11	0,00	0,36	0,00	0,00	3,00	2016,00	34337,38
50%	0,20	0,00	0,51	0,00	0,20	0,00	0,46	0,00	0,15	0,00	0,49	0,00	0,17	0,00	0,48	0,00	0,33	6,00	2018,00	43305,50
75%	0,28	0,52	0,62	0,00	0,27	0,68	0,60	0,00	0,22	0,72	0,62	0,00	0,24	0,70	0,61	0,00	0,67	9,00	2019,00	51314,13
max	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	12,00	2021,00	76049,00

kNN-Regession

UNIVERSITY OF MANNHEIM
School of Business Informatics and Mathematics



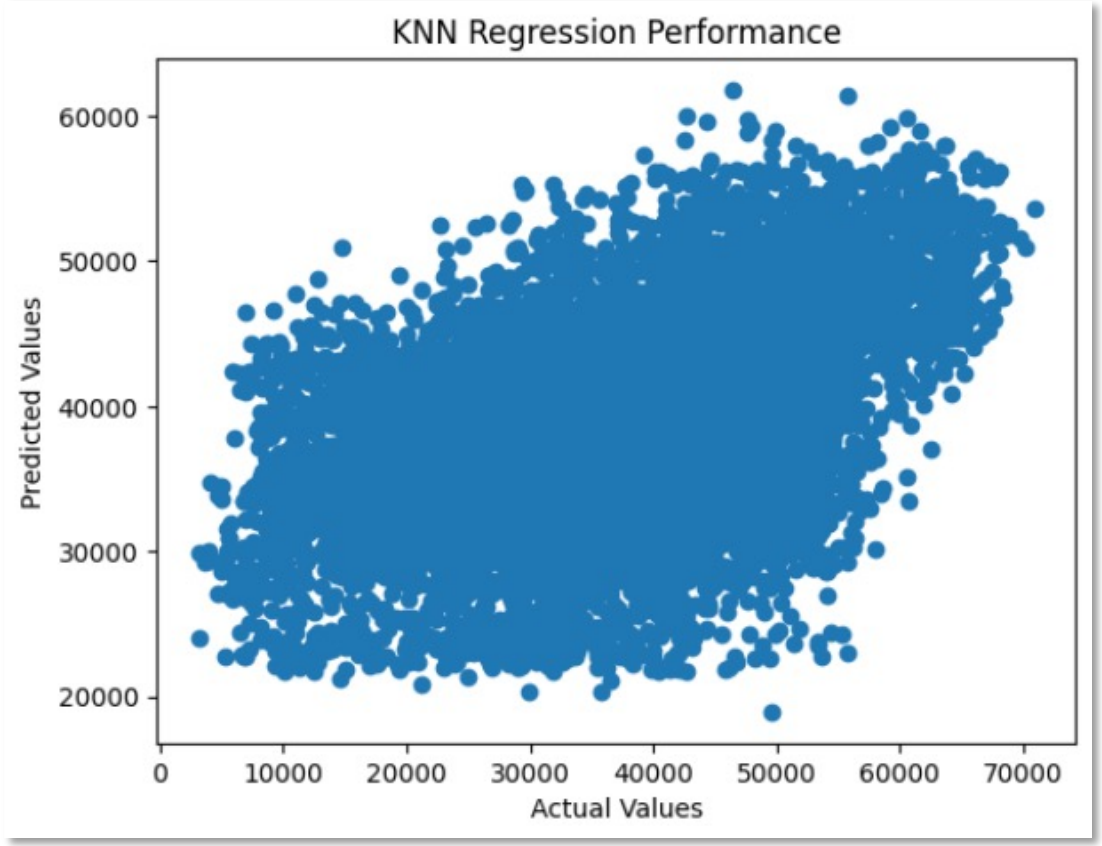
All Parameters k=100

MSE: 135790366.54084754
MAE: 9405.08743388908
RMSE: 11652.912363046738

CV-RMSE: 10796.02

kNN-Regession

UNIVERSITY OF MANNHEIM
School of Business Informatics and Mathematics



All Parameters k=38

MSE: 133188652.8830532
MAE: 9293.466625683628
RMSE: 11540.738836099412

CV-RMSE: 10831.98

kNN-Regession

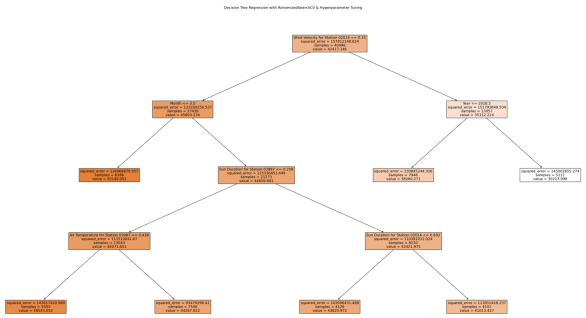
UNIVERSITY OF MANNHEIM
School of Business Informatics and Mathematics

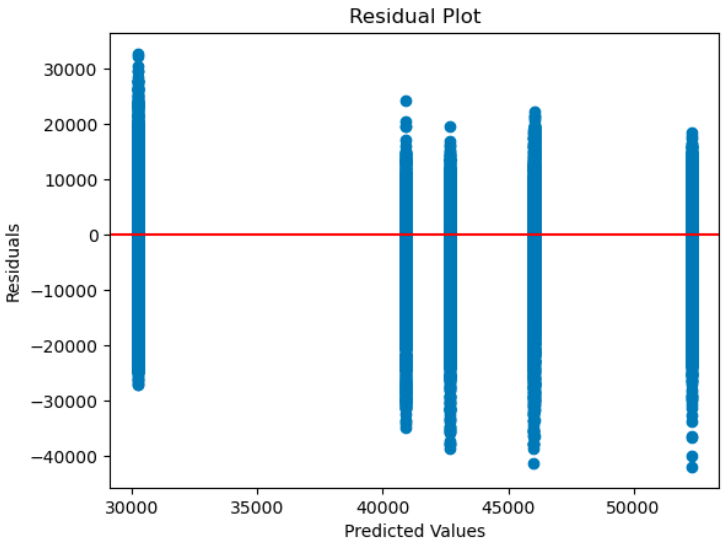
- Approach
  - By Station
  - By Weather feature

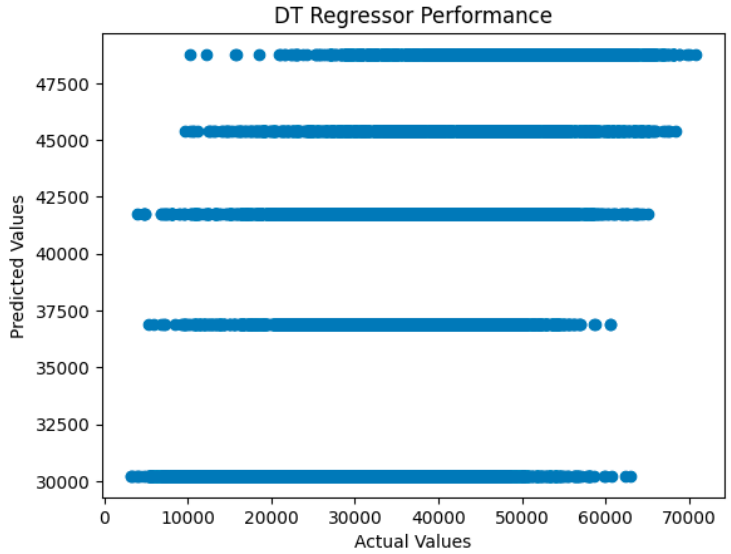
  - Random search
    - "Wind Velocity for Station 02014", "Wind Velocity for Station 03987“, "Wind Velocity for Station 03379", "Air Temperature for Station 04928“
      - K=400 CV-RMSE: 10446.23
    - "Wind Velocity for Station 02014",
    - "Wind Velocity for Station 03987",
    - "Wind Velocity for Station 03379",
    - “Precipitation Amount for Station 03379", “Sun Duratio for Station 03379",
    - "Air Temperature for Station 04928"
      - K=400 CV-RMSE:10514.81

Decision Tree Regression
K-Fold CrossValidation

UNIVERSITY OF MANNHEIM
School of Business Informatics and Mathematics

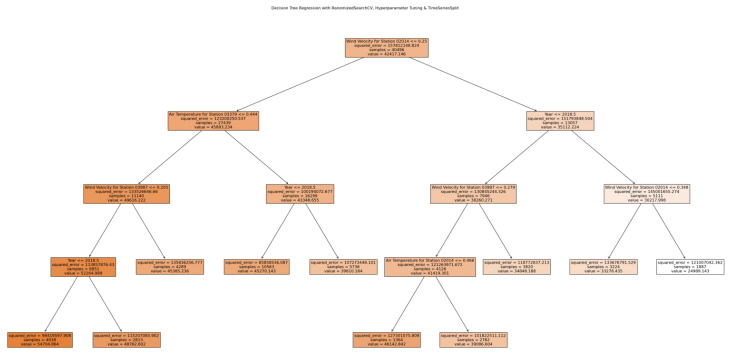


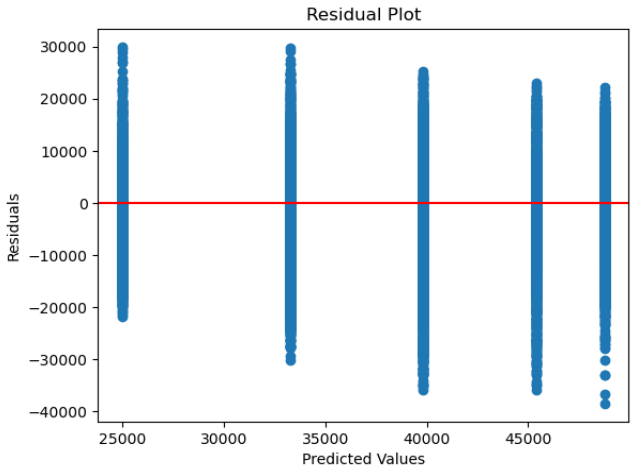


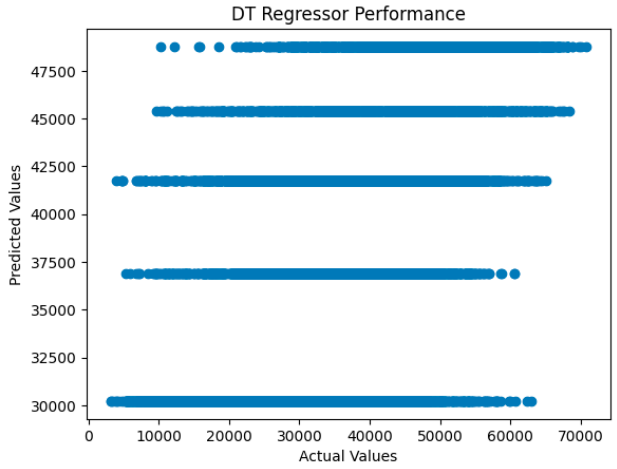



```
RMSE: 11533.000197356256
NRMSE: 0.17028910901804706
```

Decision Tree Regression
TimeSeriesSplit



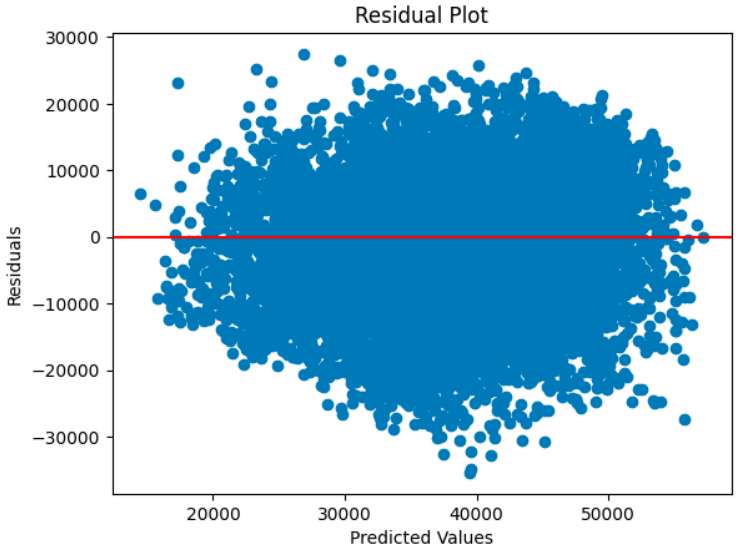


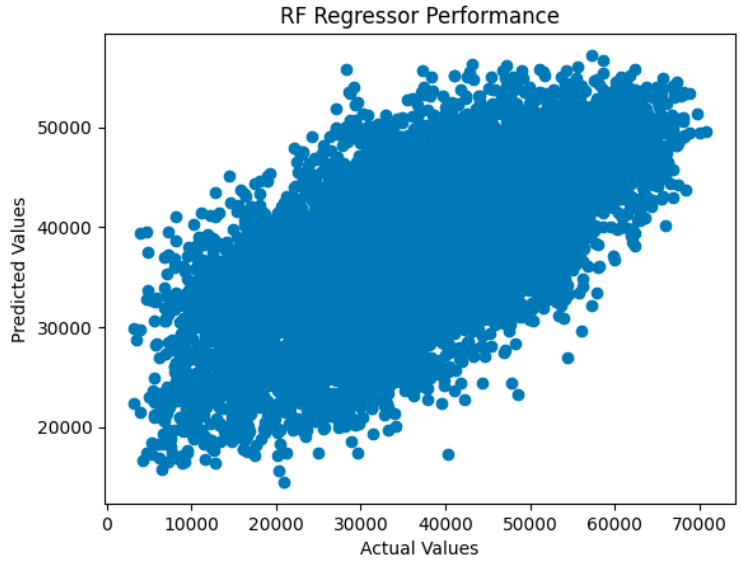


RMSE: 11093.479600694762
NRMSE: 0.1637994212074353

Best Random Forest
using TimeSeriesSplit CV

UNIVERSITY OF MANNHEIM
School of Business Informatics and Mathematics






```
best_params = {
    'n_estimators': 283,
    'min_samples_split': 7,
    'min_samples_leaf': 2,
    'max_samples': 0.7142857142857143,
    'max_leaf_nodes': None,
    'max_features': 'log2',
    'max_depth': 75
}
✓ 0.0s
```

```
# Train the model on the whole dataset and predict unseen data to assess it's real performance
rf = RandomForestRegressor(**best_params, random_state=42)
rf.fit(X_train, y_train)
y_pred = rf.predict(X_test)
rmse = sqrt(mean_squared_error(y_test, y_pred))
print("RMSE: ", rmse)
print("NRMSE: ", rmse / (y_test.max() - y_test.min()))

plt.scatter(y_pred, y_test - y_pred)
plt.axhline(y=0, color='r', linestyle='-')  # Add a horizontal line at y=0
plt.xlabel('Predicted Values')
plt.ylabel('Residuals')
plt.title('Residual Plot')
plt.show()
✓ 10.4s
```

```
RMSE:  9870.294855657023
NRMSE:  0.1457386359102416
```

RMSE: 9870.294855657023
NRMSE: 0.1457386359102416

Random Forest Regression

UNIVERSITY OF MANNHEIM
School of Business Informatics and Mathematics

```
tscv = TimeSeriesSplit(n_splits=5)

# Perform grid search
grid_search = RandomizedSearchCV(regressor, param_grid, cv=tscv,
scoring='neg_mean_squared_error', n_iter=500, n_jobs=-1)

Best Hyperparameters: {
        'n_estimators': 283,
        'min_samples_split': 7,
        'min_samples_leaf': 2,
        'max_samples': 0.7142857142857143,
        'max_leaf_nodes': None,
        'max_features': 'log2',
        'max_depth': 75
}
Best Score (MSE): 98751144.72589421
RMSE: 9870.294855657023
NRMSE: 0.1457386359102416
R-squared: 0.4163565581054164
MAE: 7964.750220017011
MAPE: 30.778891229779425
```