Predicting Hourly Electrical Energy Output of a Combined-Cycle Power Plant

Andy Alvarez

6140523

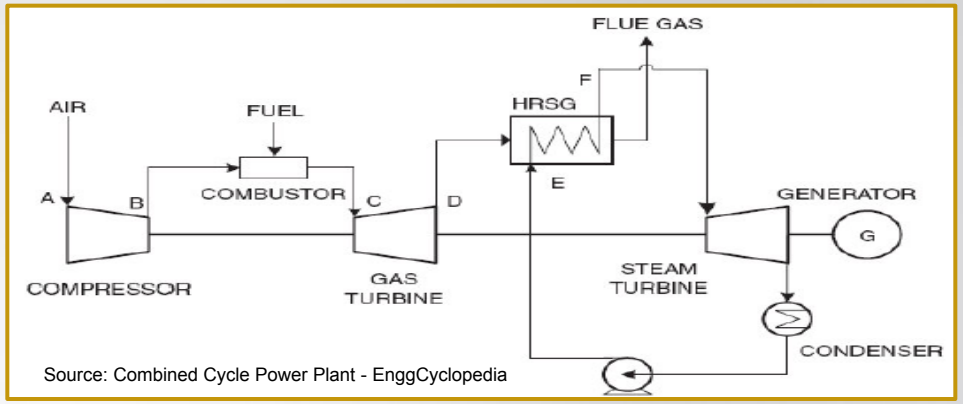
CNT 6154

Introduction

Goal: Predict net hourly electrical energy output of a combined cycle power plant.

Combined Cycle Power Plant:

- A power plant that uses both gas and steam powered turbines to generate electricity.



Source: Combined Cycle Power Plant - EnggCyclopedia

Why make this prediction?

- ❑ Estimate costs and forecast profit
- ❑ Price service accordingly
- ❑ Optimize feature values for improved efficiency and higher output production

FIU

Dataset

Train Data:
Rows, Columns (9568, 5)

Features | Target

	AT	V	AP	RH	PE
0	29.07	72.51	1009.24	58.38	449.371855
1	32.38	67.17	1006.97	60.04	450.861043
2	19.29	67.71	1007.94	63.71	460.894029
3	16.45	41.48	1016.64	45.21	467.577314
4	21.43	46.97	1013.94	61.25	469.805723
...	...	...	...	...	...
9563	4.44	38.44	1016.14	75.35	499.615488
9564	18.18	67.71	1004.50	87.26	461.130122
9565	10.96	45.01	1017.97	95.82	481.245635
9566	8.70	36.24	1013.34	89.50	490.928341
9567	27.97	58.84	1002.25	57.88	457.630310

Test Data:
Rows, Columns (38272, 4)

	AT	V	AP	RH
0	16.98	53.16	1013.95	82.80
1	9.60	41.03	1021.01	69.03
2	6.11	38.68	1017.53	79.23
3	12.34	43.22	1009.28	78.23
4	27.67	59.14	1016.51	61.20
...	...	...	...	...
38267	14.60	53.82	1016.28	64.83
38268	29.67	66.51	1015.60	34.10
38269	20.13	47.03	1012.59	83.03
38270	27.14	70.32	1007.08	73.08
38271	18.82	61.27	1019.50	72.23

Features:

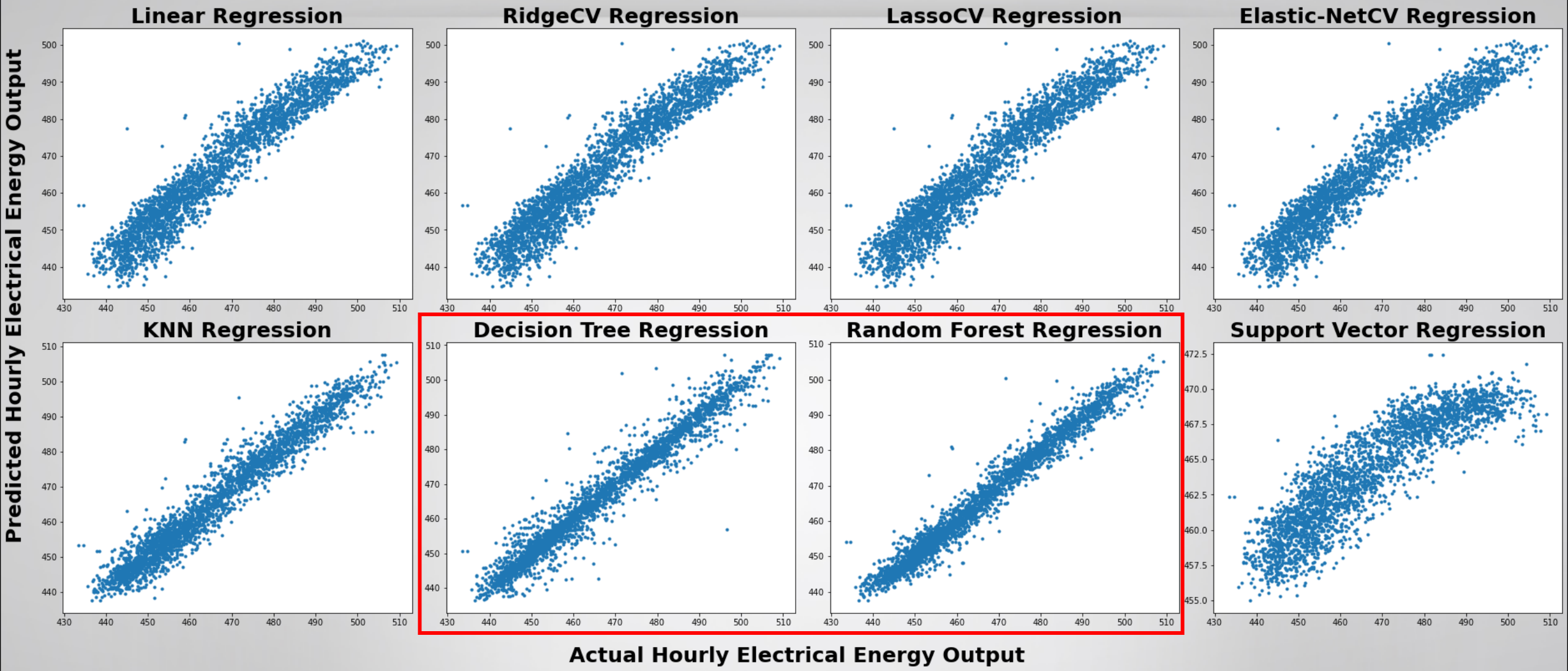
- Ambient Temperature (AT)
- Ambient Pressure (AP)
- Relative Humidity (RH)
- Exhaust Vacuum (V)

- ‘Train Data’ is treated as collected data since it already has the target column
- ‘Train Data’ uses 70/30 Train-Test Split
- ‘Test Data’ is treated as future data that needs to be predicted

Power plant power prediction Machine Hack - Kaggle

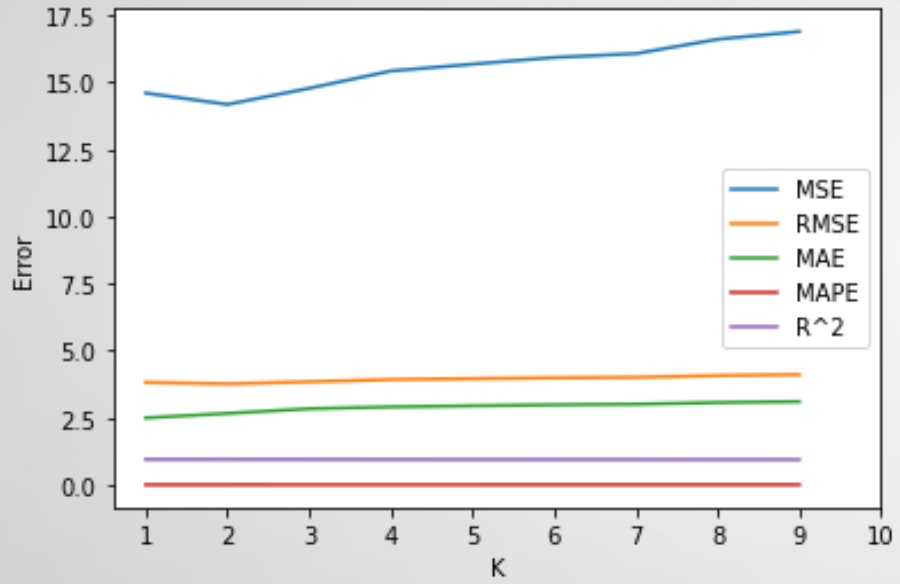
FIU

Model Performance Analysis



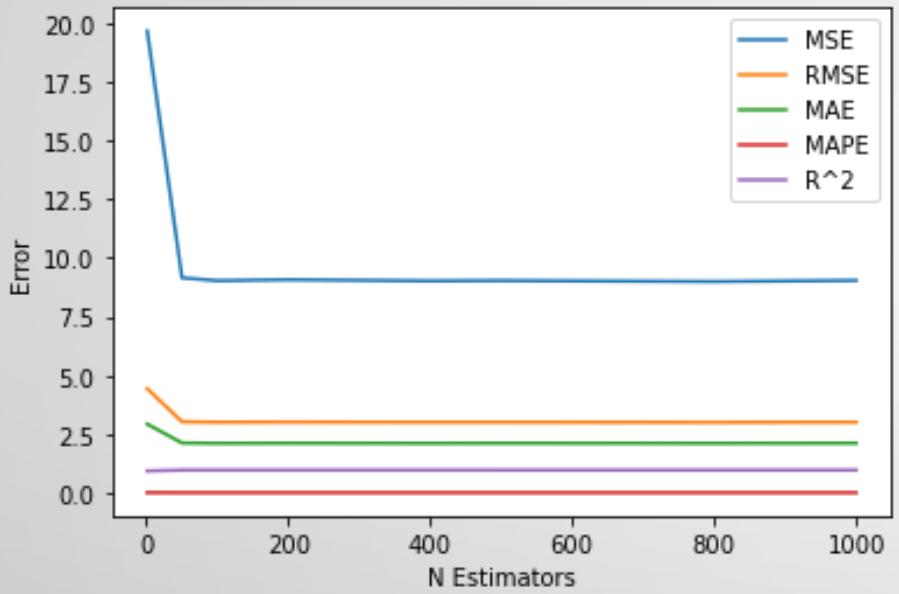
FIU | FLORIDA INTERNATIONAL UNIVERSITY

KNN Optimization



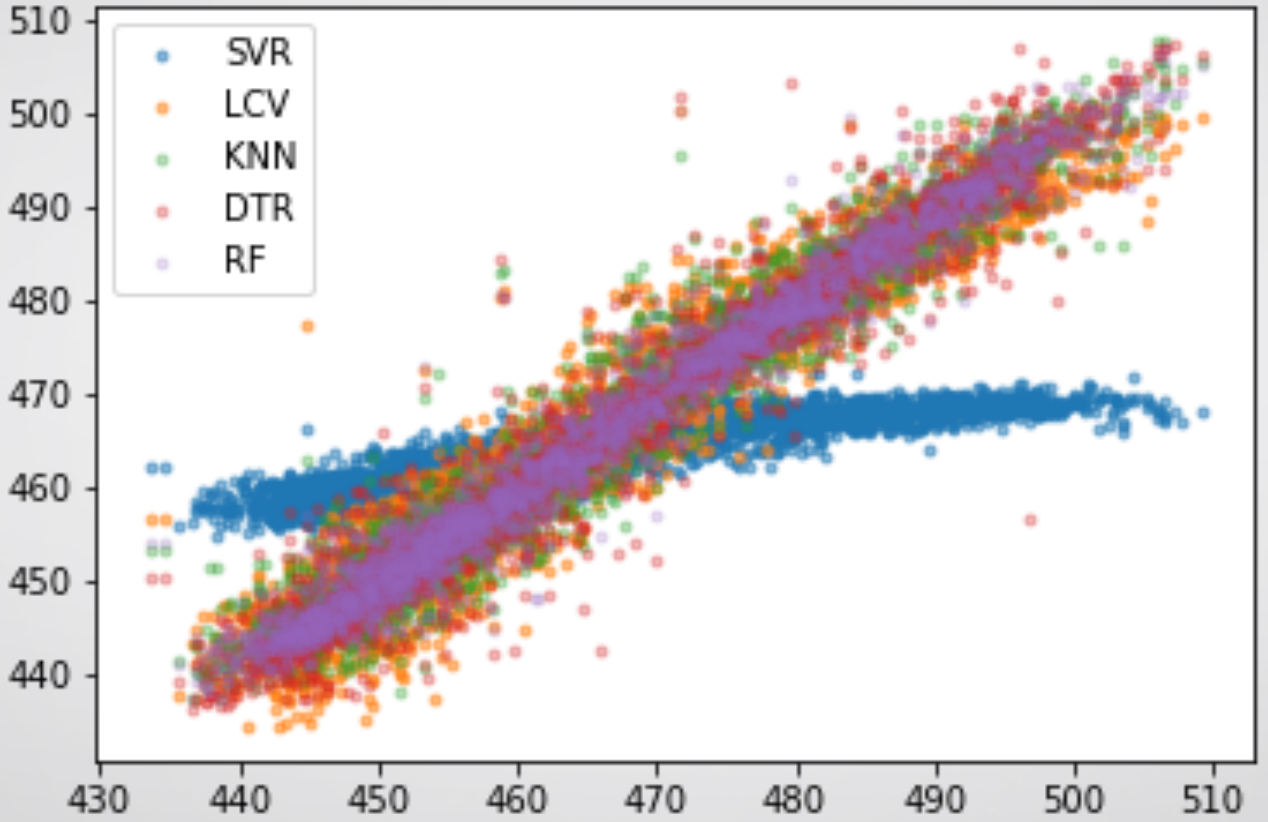
	K	MSE	RMSE	MAE	MAPE	R^2
0	1	14.604987	3.821647	2.498038	0.005366	0.948646
1	2	14.181611	3.765848	2.663358	0.005721	0.950135
2	3	14.782327	3.844779	2.838465	0.006092	0.948023
3	4	15.426672	3.927680	2.907300	0.006238	0.945757
4	5	15.681565	3.959996	2.948056	0.006325	0.944861
5	6	15.932094	3.991503	2.984852	0.006402	0.943980
6	7	16.075200	4.009389	3.007458	0.006450	0.943477
7	8	16.608688	4.075376	3.070116	0.006582	0.941601
8	9	16.898264	4.110750	3.104823	0.006656	0.940583

Random Forest Configuration

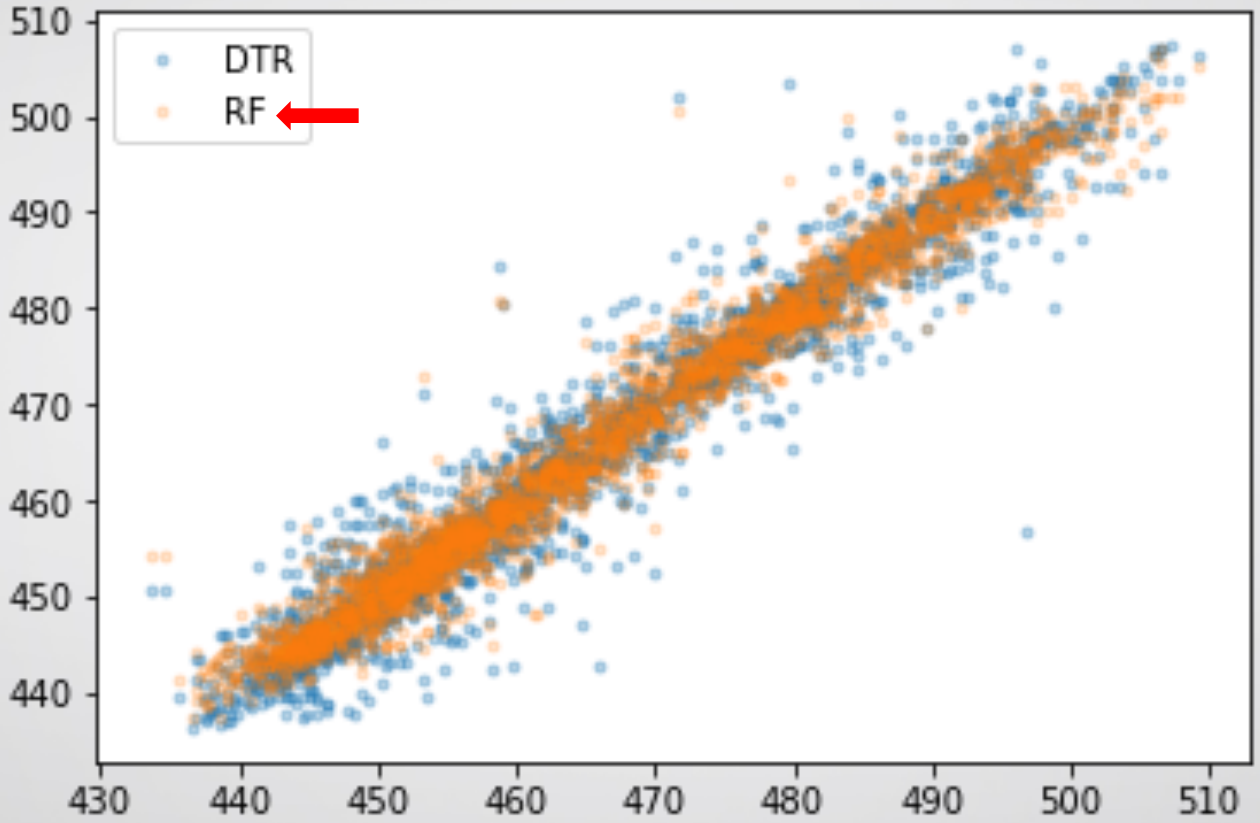


	n	MSE	RMSE	MAE	MAPE	R^2
0	1	19.677679	4.435953	2.929579	0.006278	0.930810
1	50	9.158551	3.026310	2.122525	0.004551	0.967797
2	100	9.038739	3.006450	2.108166	0.004519	0.968218
3	200	9.078228	3.013010	2.113968	0.004532	0.968079
4	400	9.029537	3.004919	2.108857	0.004520	0.968251
5	500	9.040920	3.006812	2.113495	0.004531	0.968210
6	800	9.001730	3.000288	2.108794	0.004521	0.968348
7	1000	9.047273	3.007869	2.114193	0.004532	0.968188

Overlayed Plots



Decision Tree v Random Forest



Error Metrics Analysis

Goal: $R^2 > 0.90$

	MSE	RMSE	MAE	MAPE	R^2
Linear	22.208463	4.712586	3.758167	0.008065	0.921911
Ridge	22.208463	4.712586	3.758167	0.008065	0.921911
Lasso	22.208459	4.712585	3.758166	0.008065	0.921911
ElasticNet	22.208460	4.712585	3.758168	0.008065	0.921911
KNN	14.782327	3.844779	2.838465	0.006092	0.948023
DTR	14.354885	3.788784	2.452637	0.005256	0.949526
RF	9.016432	3.002737	2.109750	0.004522	0.968297
SVR	190.444886	13.800177	11.369381	0.024092	0.330361

FIU

Electrical Energy Prediction for Future Data

Without Prediction

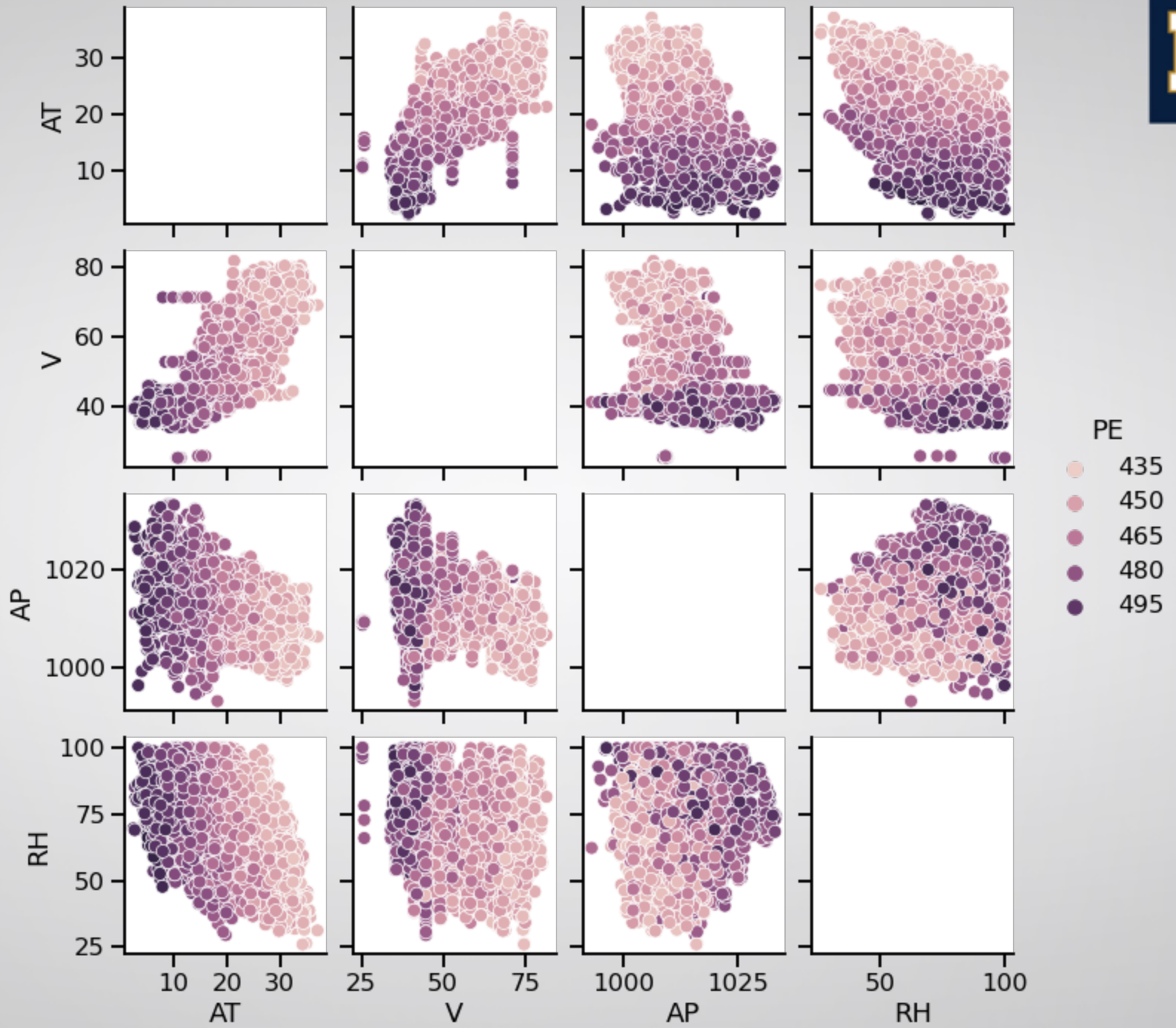
	AT	V	AP	RH
0	16.98	53.16	1013.95	82.80
1	9.60	41.03	1021.01	69.03
2	6.11	38.68	1017.53	79.23
3	12.34	43.22	1009.28	78.23
4	27.67	59.14	1016.51	61.20
...	...	...	...	...
38267	14.60	53.82	1016.28	64.83
38268	29.67	66.51	1015.60	34.10
38269	20.13	47.03	1012.59	83.03
38270	27.14	70.32	1007.08	73.08
38271	18.82	61.27	1019.50	72.23

With Prediction

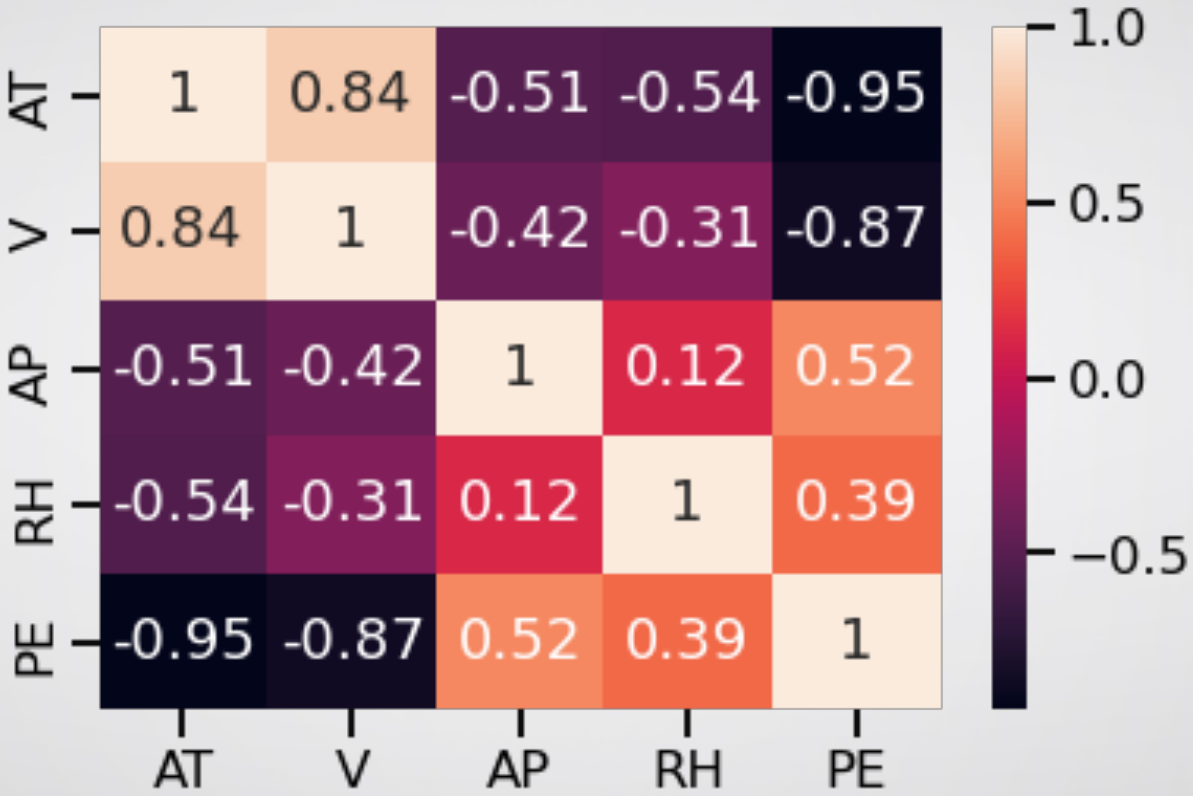
	AT	V	AP	RH	PE
0	16.98	53.16	1013.95	82.80	473.254274
1	9.60	41.03	1021.01	69.03	490.843863
2	6.11	38.68	1017.53	79.23	495.703278
3	12.34	43.22	1009.28	78.23	480.961193
4	27.67	59.14	1016.51	61.20	448.666474
...	...	...	...	...	...
38267	14.60	53.82	1016.28	64.83	474.870863
38268	29.67	66.51	1015.60	34.10	447.316109
38269	20.13	47.03	1012.59	83.03	468.323381
38270	27.14	70.32	1007.08	73.08	447.117325
38271	18.82	61.27	1019.50	72.23	465.664402

FIU

Pair Plot


AT
V
AP
RH
30
20
10
80
60
40
1020
1000
100
75
50
25
10 20 30
25 50 75
1000 1025
50 100
AT
V
AP
RH
PE
435
450
465
480
495

Correlation Matrix



Conclusion

- ✓ Random Forest Regression performed the best for this application.
- ✓ Net hourly electrical energy output was predicted for a future dataset.
- ✓ To improve production, scientists may want to explore methods to decrease **Ambient Temperature** and **Exhaust Vacuum**.