PREDICTING BIKE RENTALS

By Linda Lacsivy

Agenda

1. data
2. identify ml problem
3. train models
4. evaluate models
5. optimize models
6. conclusion

1. DATA

Features

	instant	dteday	season	yr	mnth	hr	holiday	weekday	workingday	weathersit	temp	atemp	hum	windspeed	casual	registered	cnt
0	1	2011-01-01	1	0	1	0	0	6	0	1	0.24	0.2879	0.81	0.0	3	13	16
1	2	2011-01-01	1	0	1	1	0	6	0	1	0.22	0.2727	0.80	0.0	8	32	40
17376	17377	2012-12-31	1	1	12	21	0	1	1	1	0.26	0.2576	0.60	0.1642	7	83	90
17377	17378	2012-12-31	1	1	12	22	0	1	1	1	0.26	0.2727	0.56	0.1343	13	48	61
17378	17379	2012-12-31	1	1	12	23	0	1	1	1	0.26	0.2727	0.65	0.1343	12	37	49

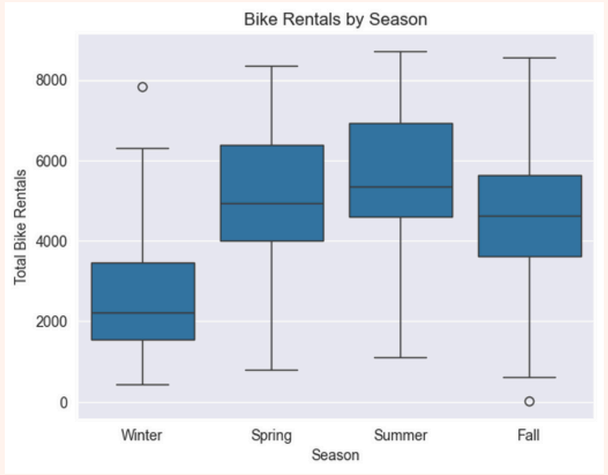
Target: bike rentals ‘cnt’

```
count    17379.000000
mean       189.463088
std        181.387599
min          1.000000
25%         40.000000
50%        142.000000
75%        281.000000
max        977.000000
Name: cnt, dtype: float64
```

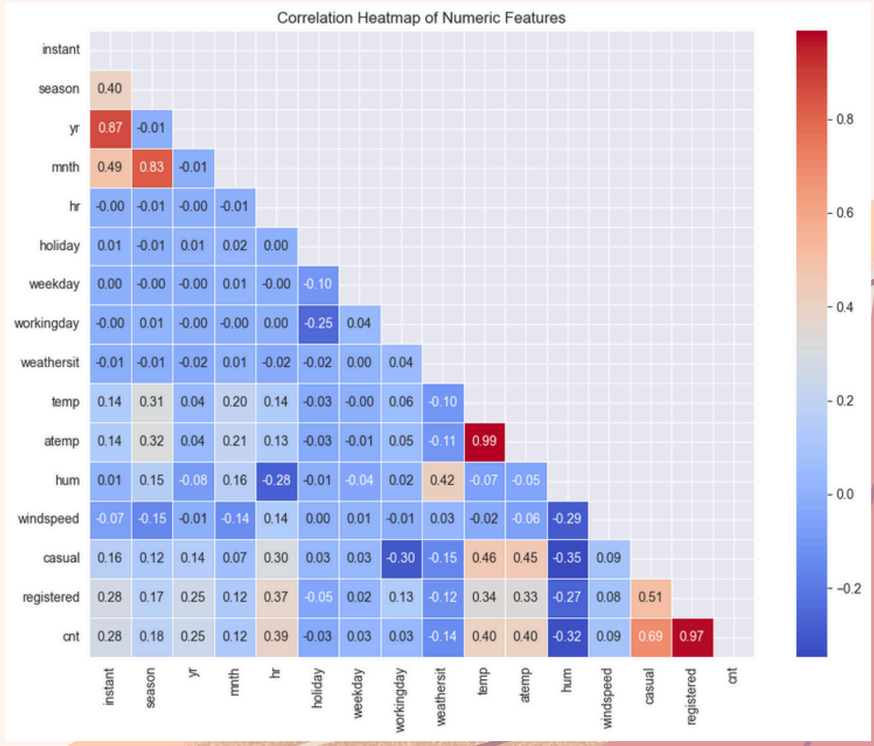
Outliers

explore, clean, adjust

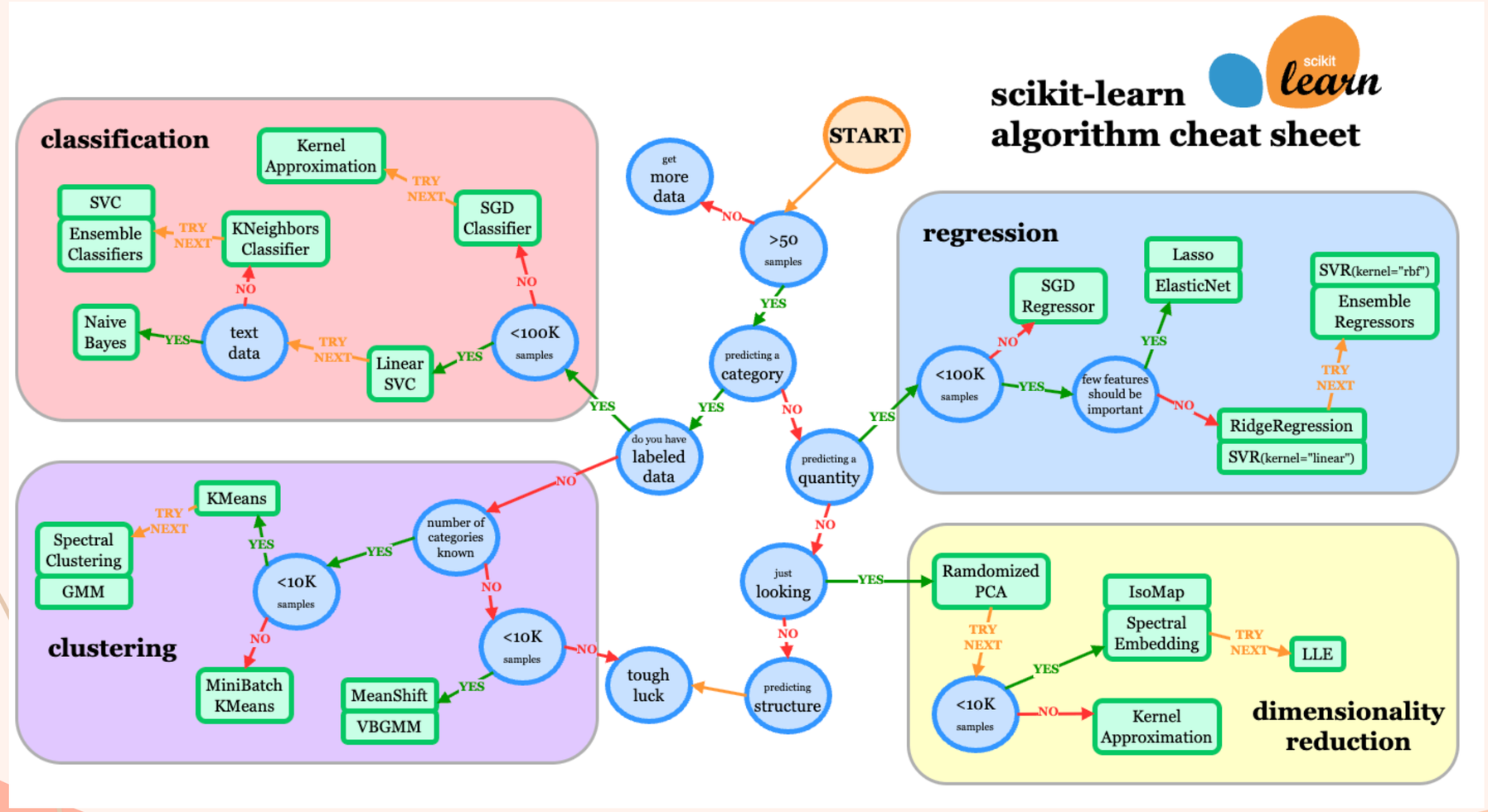
Plots



Correlation map



2. IDENTIFY ML PROBLEM

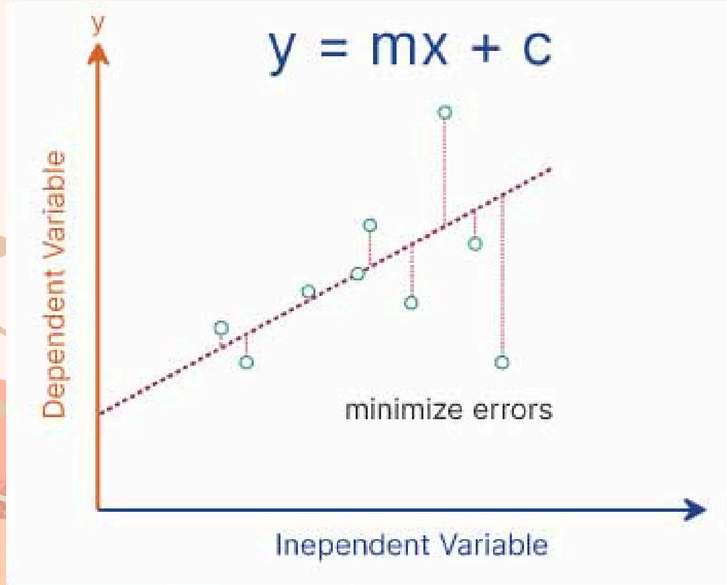


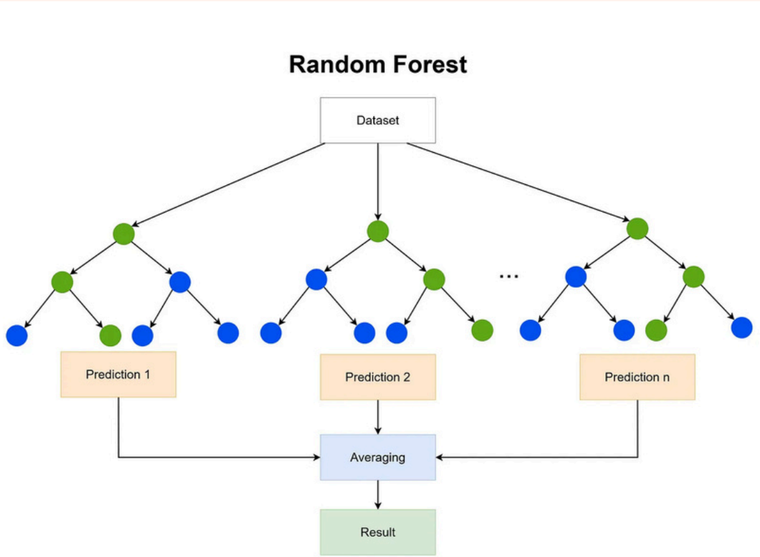
3. TRAIN MODELS

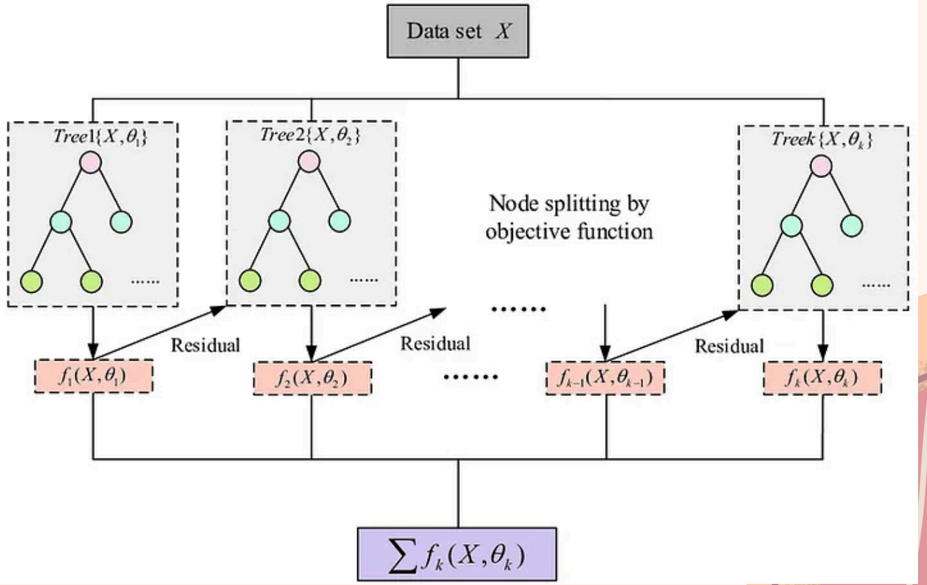
1 Linear Model

2 Random Forest Model

3 eXtrem Gradient Boost Model



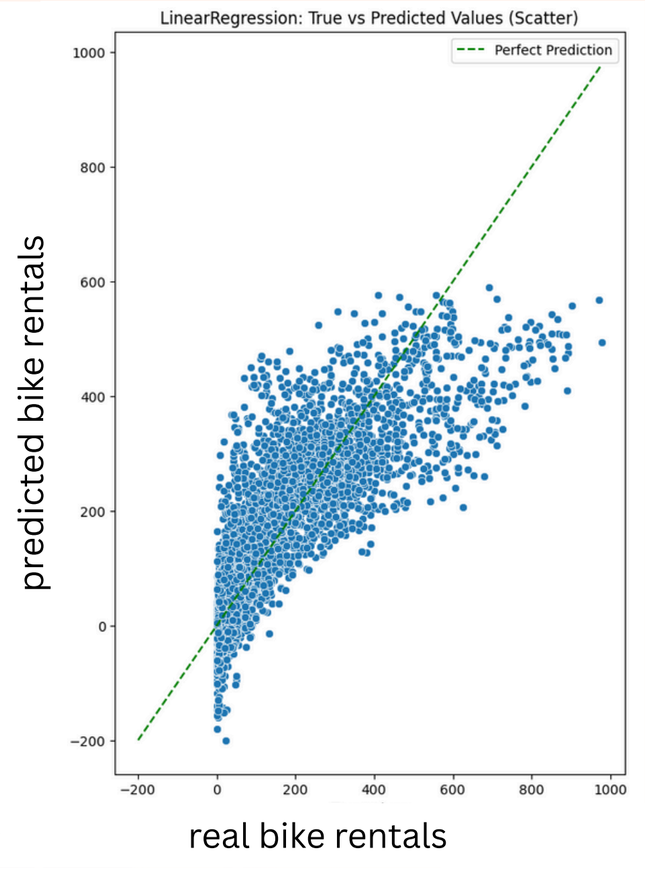




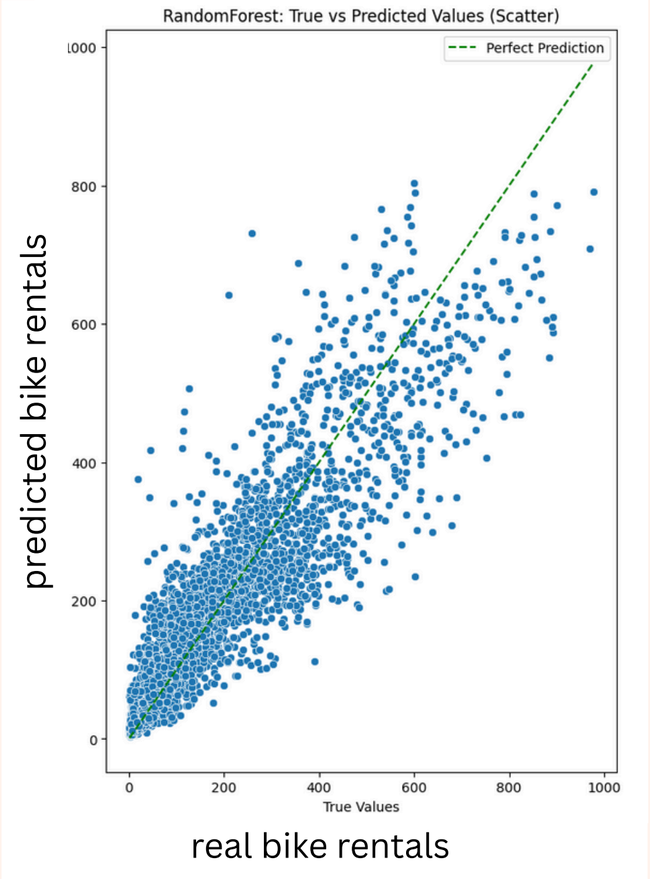
4. EVALUATE MODELS

4.1. PREDICTED VS TRUE BIKE RENTALS

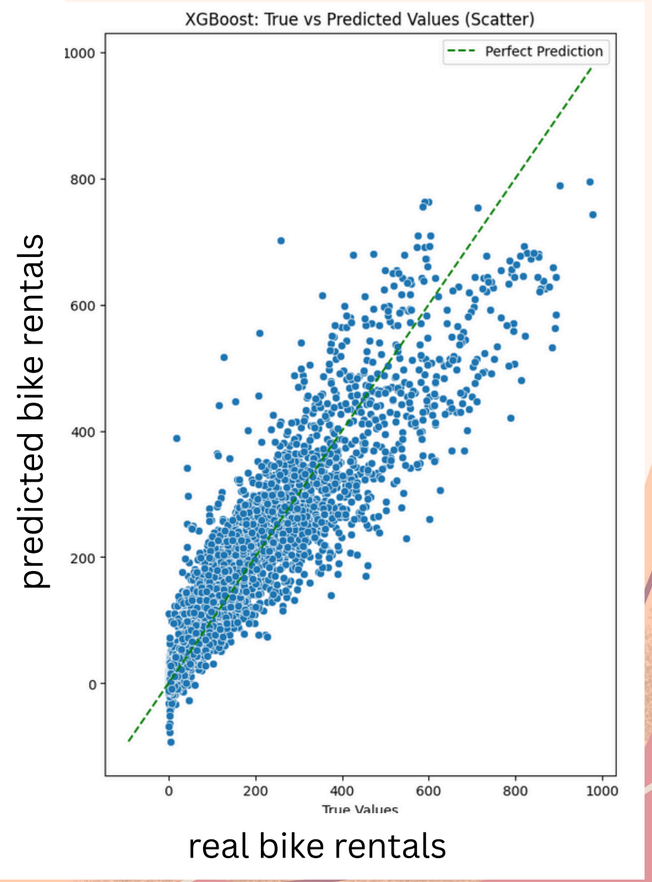
1 Linear Regression



2 Random Forest Regression



3 XGBoost Regression



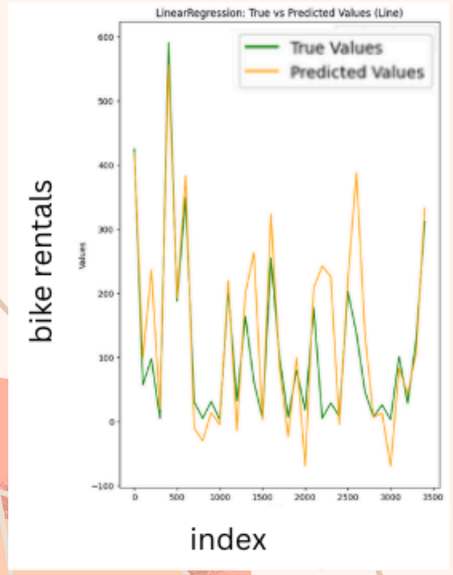
4. EVALUATE MODELS

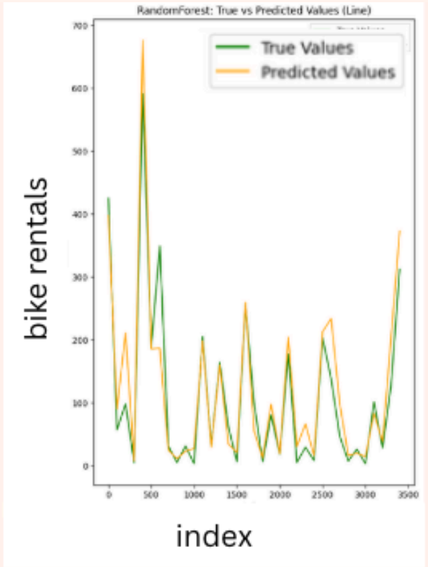
4.2. METRICS

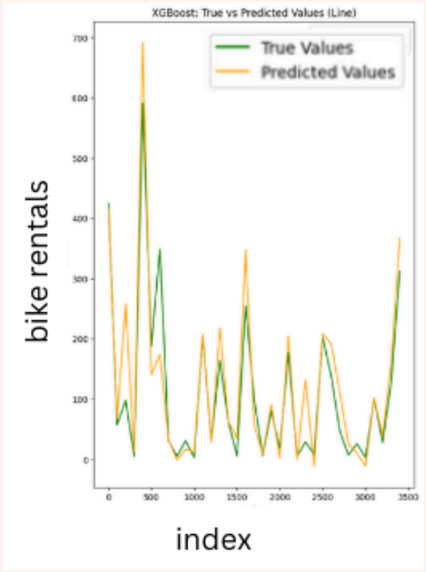
1 Linear Regression

2 Random Forest Regression

3 XGBoost Regression





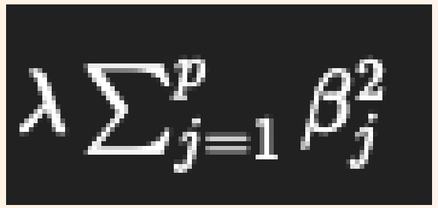
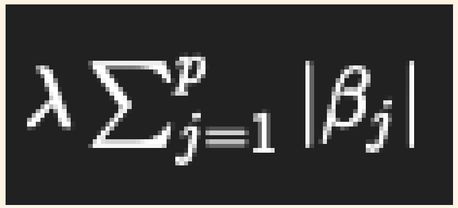


	short form	formula	interpretation
r-squared	R^2	$\frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$	$R^2 = 1$ means perfect prediction
mean absolut error	MAE	$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$	average absolute prediction error
root mean squared error	RMSE	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$	average squared prediction error

METRICS	Linear Regression	Random Forest Regression	XGBoost Regression
R^2	0.62	0.81	0.82
mae	77	51	51
rmse	108	76	74

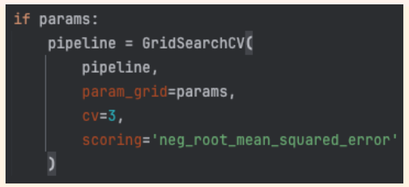
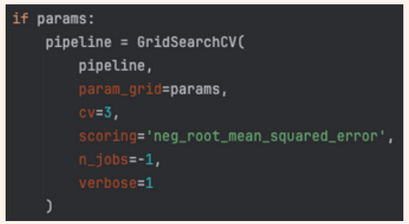
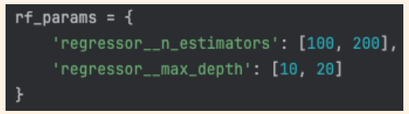
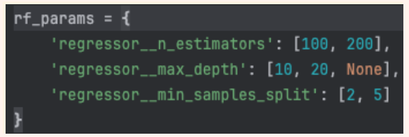
5. OPTIMIZE MODELS

5.1. Ridge and Lasso Regression

OPTIMIZATION	Ridge Regression	Lasso Regression
penalty term added	sum of squared coefficients	sum of absolute coefficients
formula		

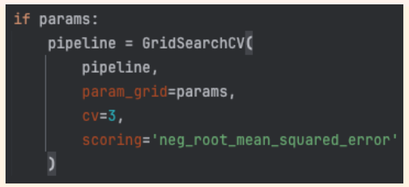
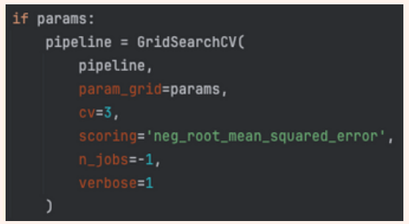
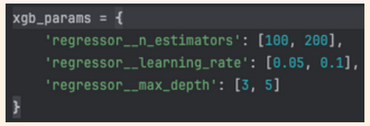
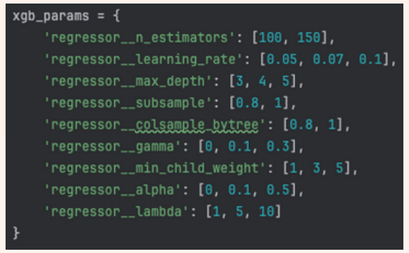
METRICS	Linear Regression	Ridge Regression	Lasso Regression
R^2	0.6266	0.6271	0.6276
mae	77.8781	77.8275	77.7407
rmse	108.7329	108.6538	108.590

5.2. tuned Random forest Regression

OPTIMIZATION	Random Forest Regression	Tuned Random Forest Regression
extended GridSearch		
tuned parameters		

METRICS	Random Forest Regression	Tuned Random Forest Regression
R^2	0.8133	0.8300
mae	51.8918	47.7931
rmse	76.8847	73.3524

5.3. TUNED XGBOOST REGRESSION

OPTIMIZATION	XGBoost Regression	Tuned XGBoost Regression
extended GridSearch		
tuned parameters		

METRICS	XGBoost Regression	Tuned XGBoost Regression
R^2	0.8251	0.82427
mae	51.0699	52.0300
rmse	74.40827	74.5939

6. Conclusion

best result

Tuned Random Forest Regression

Great Machine Learning course

We learned a lot of new tools and got plenty of hands-on experience in only 1 week.

only little optimizations

my optimizations did hardly improve the results

Thank You