Bachelor’s Thesis Computer Science
Probabilistic Short-term Wind Power Forecasting

Johannes Gaus

18.07.2022

Contents

01 Introduction

Why bother with wind power forecasting?

- to plan and operate a wind power farm
- big shift to renewable energy sources
- cope with the shortage of traditional fossil energy and the environmental pollution
- successful integration of large amounts of wind power into the electricity supply system

02 Research Question

Research Question

- How to make a wind power forecasting model?

- What is a good approach?
  - Multiple Linear Regression
  - Support Vector Regression

- Which model works better with the dataset?

Approach Breakdown

- Artificial intelligence model with a supervised machine learning approach
  - Multiple Linear Regression
  - Support Vector Regression

- Tawn, R. and J. Browell (2022). “A review of very short-term wind and solar power forecasting”
- Tian, Zhongda (2021). “A state-of-the-art review on wind power deterministic prediction”

03 Approach

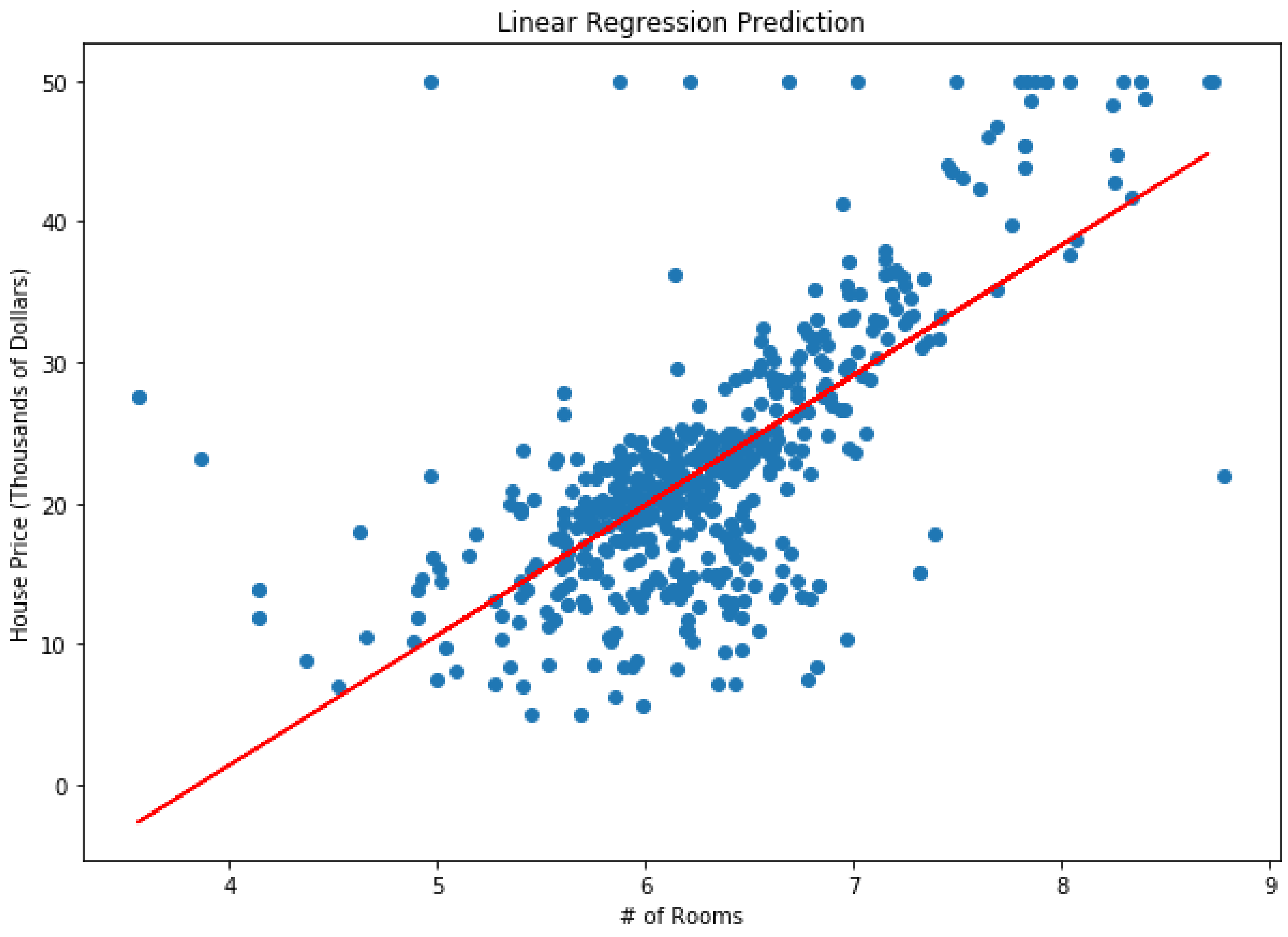
03 Theoretical Approach

Multiple Linear Regression

Support Vector Regression

Evaluation Score

Multiple Linear Regression



OLS Prediction of Boston Housing Prices

Figure 1 https://towardsdatascience.com/an-introduction-to-support-vector-regression-svr-a3ebc1672c2

SVR

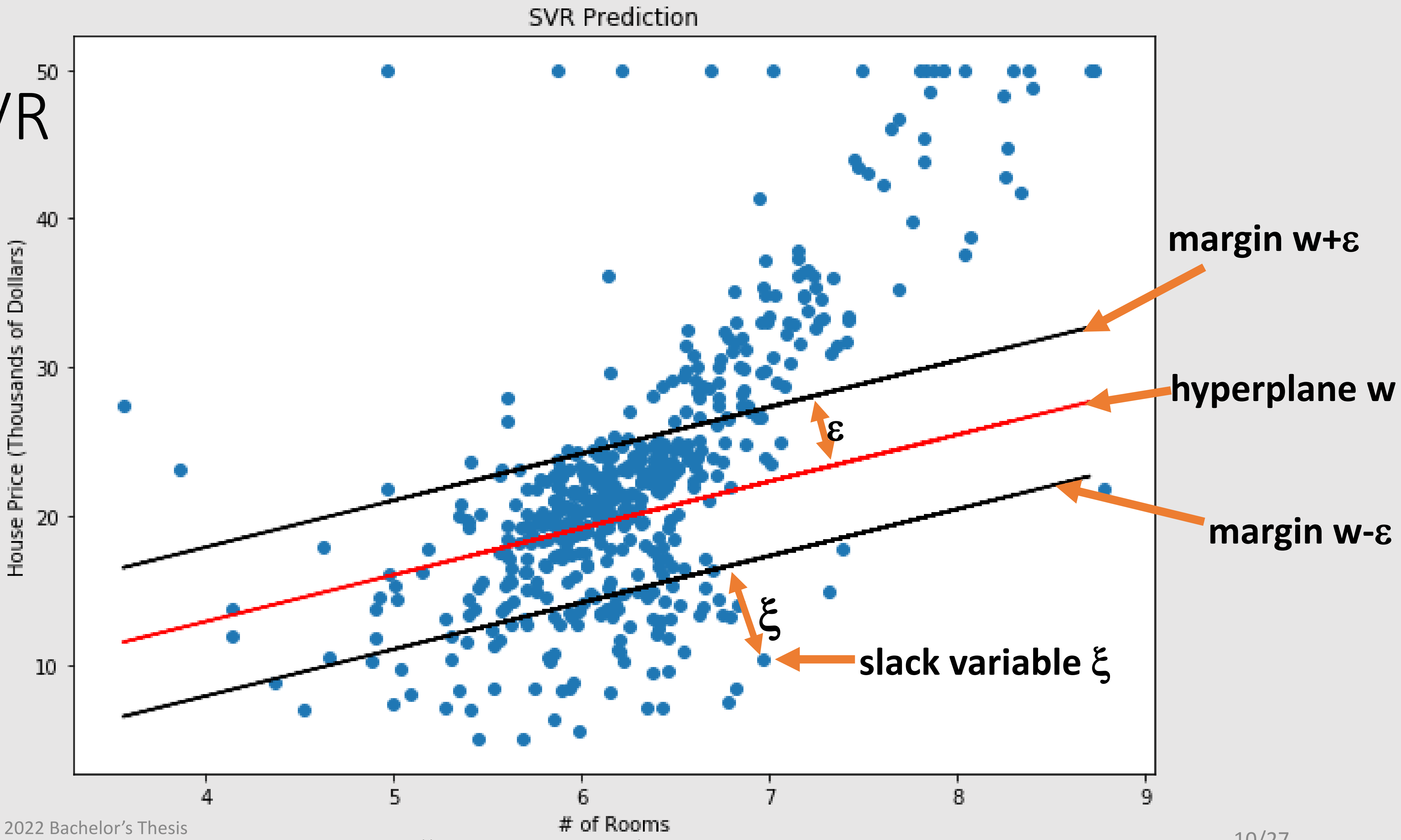
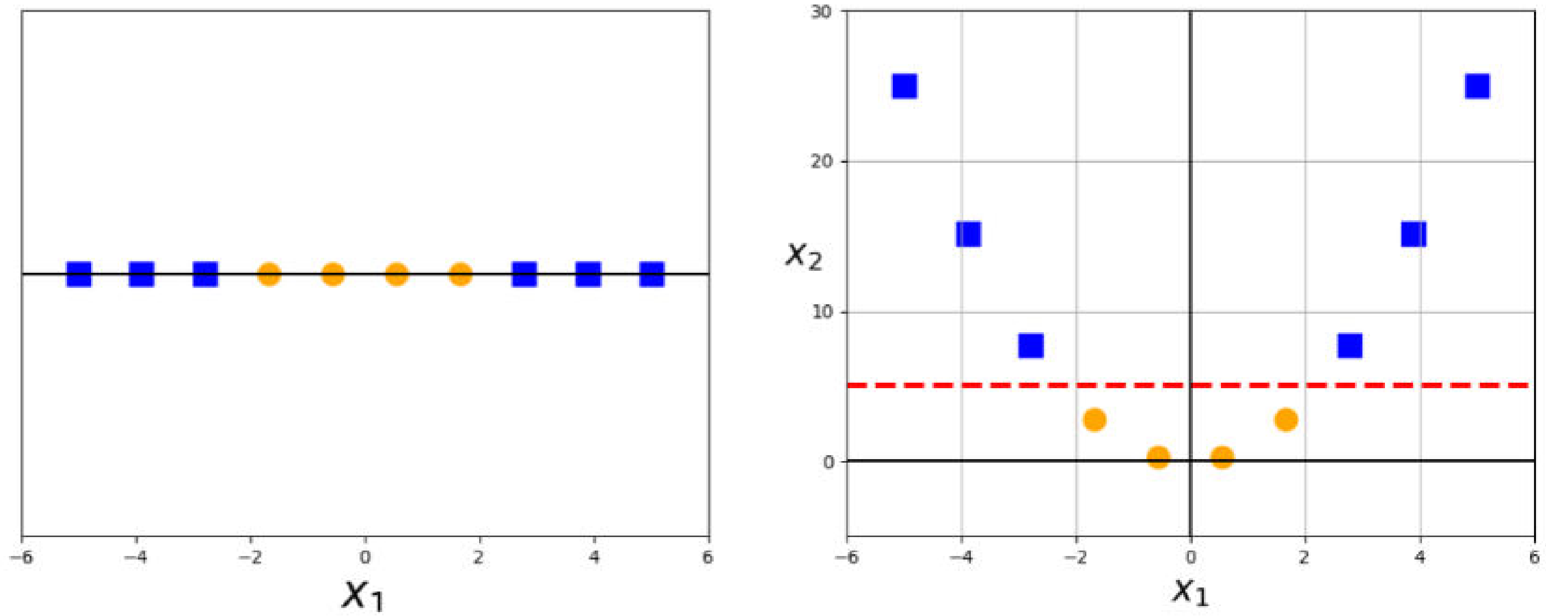


Figure 2 https://towardsdatascience.com/an-introduction-to-support-vector-regression-svr-a3ebc1672c2

The Kernel trick

- Smola, Alex J. and Bernhard Schölkopf (2004). “A tutorial on support vector regression”. In: Statistics and Computing 14.3, doi: 10.1023/B:STCO.0000035301.49549.88



This data becomes linearly separable after a quadratic transformation to 2-dimensions.

Figure 3 https://towardsdatascience.com/the-kernel-trick-c98cdbcaeb3f

Support Vector Regression

- choose an acceptable error margin(ϵ)
- tuning our tolerance of falling outside that acceptable error margin
- map linear data into a higher feature space to get a better regression line for the SVR algorithm
  - RBF Kernel
  - Linear Kernel
  - Poly Kernel
  - …

Evaluation Scores

- MAE $\Rightarrow$ mean absolute error
- RMSE $\Rightarrow$ root mean squared error
- R2 $\Rightarrow$ how well the data fit the regression model (the goodness of fit)
- CRPS $\Rightarrow$ comparison of the predicted cumulative density function and the true cumulative density function
- Shapley Values $\Rightarrow$ attempt to explain why an ML model reports the outputs that it does on an input

03 Practical Approach

Cleaning the dataset

Feature selection

SVR parameter and kernel selection

dataset

- La haute borne wind farm in France
- about 150 km west of Strasbourg
- 4 turbines

- https://opendata-renewables.engie.com/explore/dataset/01c55756-5cd6-4f60-9f63-2d771bb25a1a/information

Figure 4 https://www.thewindpower.net/windfarm_en_3354_la-haute-borne.php

Feature Selection

- tried some algorithms from sklearn
- build some models with different feature sets by hand
- evaluate the different features using
  - a Multiple Linear Regression model
  - MAE score, RMSE score, CRPS, and R2 score

Feature Selection

	model 1	model 2	model 3	model 4	model 5	model 6	model 7	model 8	model 9	model 10	model 11	model 12
Ba_avg							x					
Cosphi_avg					x							
Db2t_avg						x						
Gb1t_avg		x	x	x								
Gost_avg	x											
hard coded time stamp												x
month											x	
Nf_avg						x						
Ot_avg	x					x	x	x	x	x	x	x
Rbt_avg	x											
Rs_avg	x											
Rt_avg						x	x		x			
Va_avg						x						
Ws_avg	x				x		x	x				
Ws1_avg			x	x	x				x	x	x	x
Ws2_avg			x	x	x				x	x	x	x
year											x	
Yt_avg		x	x				x		x			
MAE	148.22	218.64	146.02	161.90	156.52	334.84	134.84	160.02	142.99	157.67	157.38	157.11
RMSE	200.16	288.30	192.54	209.55	202.67	453.81	190.18	204.02	191.99	202.56	202.34	202.30
R2	0.80	0.60	0.82	0.79	0.80	0.07	0.83	0.80	0.82	0.80	0.80	0.80
CRPS	142.84	132.55	261.38	275.20	270.80	108.21	238.69	135.35	256.58	273.19	272.57	272.42

Table 3.1: Evaluation scores of different Multiple Linear Regression models using different features to choose the best set of features

Figure 5

Feature Selection

	model 1	model 2	model 3	model 4	model 5	model 6	model 7	model 8	model 9	model 10	model 11	model 12
Ba_avg							x					
Cosphi_avg					x							
Db2t_avg						x						
Gb1t_avg		x	x	x								
Gost_avg	x											
hard coded time stamp												x
month											x	
Nf_avg						x						
Ot_avg	x					x	x	x	x	x	x	x
Rbt_avg	x											
Rs_avg	x											
Rt_avg						x	x		x			
Va_avg						x						
Ws_avg	x				x		x	x				
Ws1_avg			x	x	x				x	x	x	x
Ws2_avg			x	x	x				x	x	x	x
year											x	
Yt_avg		x	x				x		x			
MAE	148.22	218.64	146.02	161.90	156.52	334.84	134.84	160.02	142.99	157.67	157.38	157.11
RMSE	200.16	288.30	192.54	209.55	202.67	453.81	190.18	204.02	191.99	202.56	202.34	202.30
R2	0.80	0.60	0.82	0.79	0.80	0.07	0.83	0.80	0.82	0.80	0.80	0.80
CRPS	142.84	132.55	261.38	275.20	270.80	108.21	238.69	135.35	256.58	273.19	272.57	272.42

Table 3.1: Evaluation scores of different Multiple Linear Regression models using different features to choose the best set of features

Figure 5

- the pitch angle (BA)
- the hub temperature (Rt)
- the nacelle temperature (Yt)
- the wind speed (Ws)
- the outdoor temperature (Ot)

Parameter selection and Kernel selection

kernel	C	epsilon	MAE	RMSE	R2	CRPS
rbf	1.0	0.1	77.70	125.46	0.92	127.60
rbf	10.0	1.0	49.73	81.99	0.96	124.08
rbf	100.0	20.0	41.18	64.01	0.98	122.62
rbf	150.0	25.0	40.34	62.02	0.98	122.44
rbf	160.0	40.0	41.55	62.51	0.98	123.01
rbf	170.0	30.0	40.51	61.67	0.98	122.69
rbf	175.0	25.0	39.90	61.32	0.98	122.29
rbf	175.0	30.0	40.44	61.55	0.98	122.55
rbf	180.0	35.0	40.85	61.70	0.98	122.83
rbf	185.0	40.0	41.11	61.85	0.98	122.97
rbf	25.0	5.0	45.92	74.60	0.97	123.41
rbf	250.0	50.0	41.32	61.33	0.98	123.26
rbf	350.0	100.0	48.43	65.70	0.98	131.38
rbf	50.0	5.0	42.75	68.80	0.97	122.83

Table 3.3: Evaluation of different parameters to compare different SVR models

Figure 6

kernel	C	epsilon	MAE	RMSE	R2	CRPS
linear	1.0	0.1	116.00	210.76	0.78	151.95
rbf	1.0	0.1	77.70	125.48	0.92	127.60
polynomial	1.0	0.1	132.67	248.66	0.70	117.77

Table 3.2: Evaluation scores for different SVR kernels (linear, rbf, poly)

Figure 7

04 RESULTS

Multiple Linear Regression model

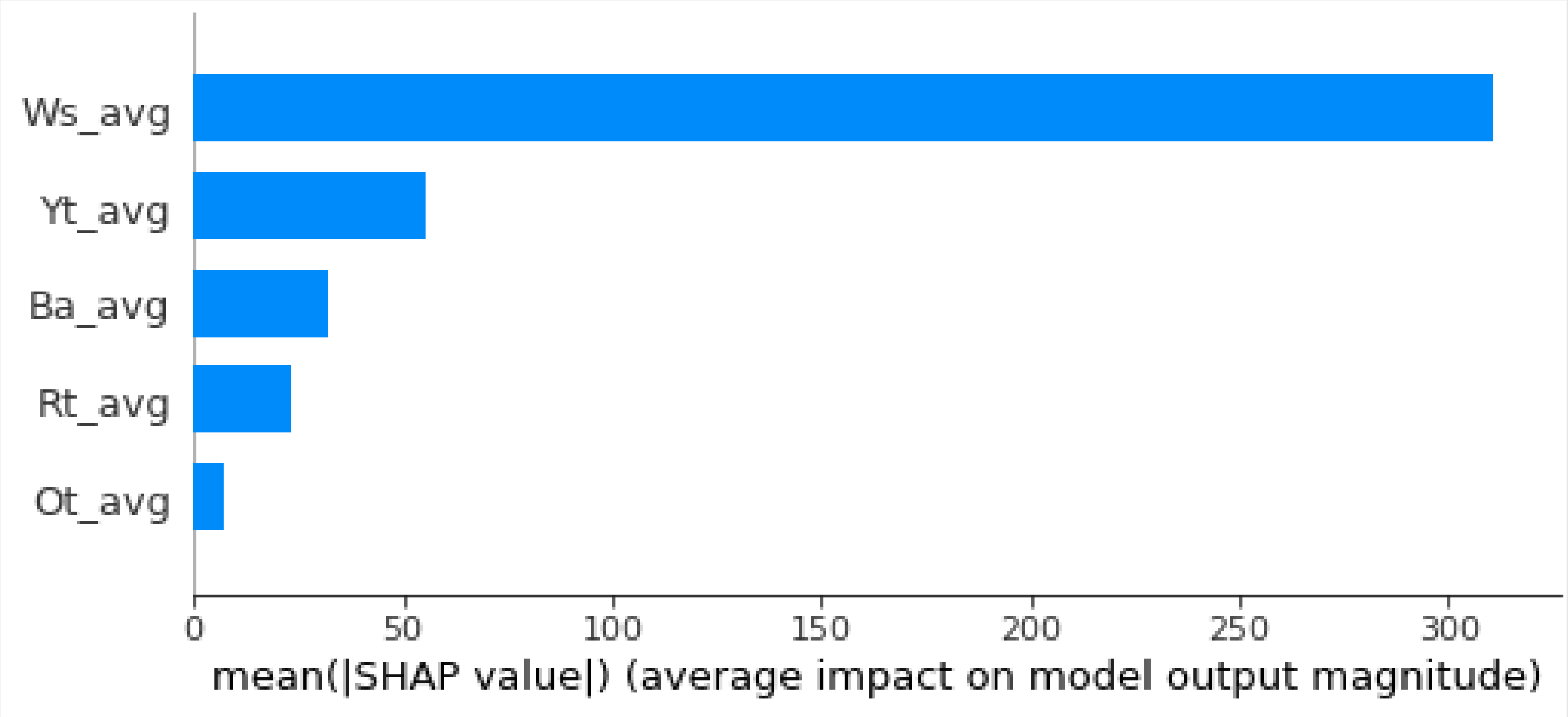


Figure 8

Support Vector Regression model

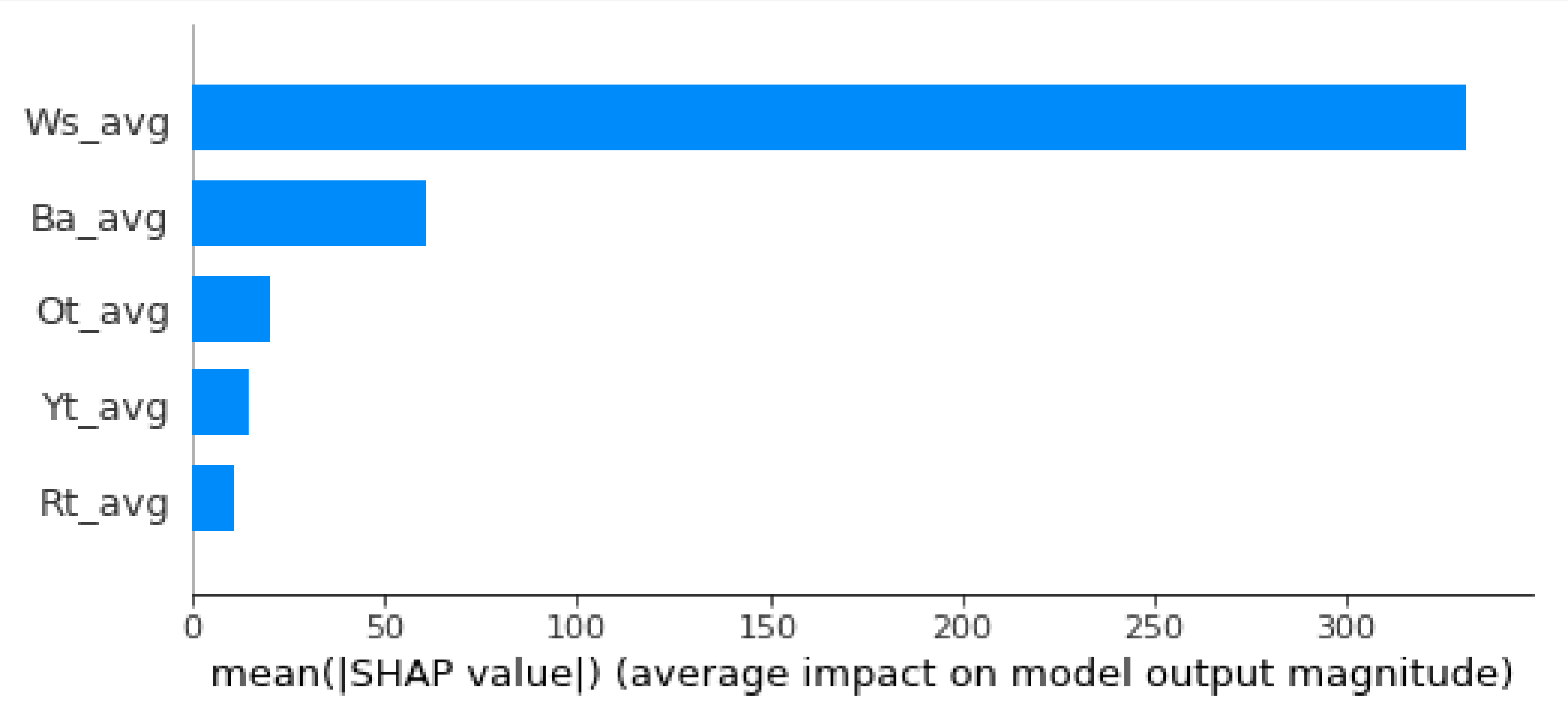


Figure 9

Comparison

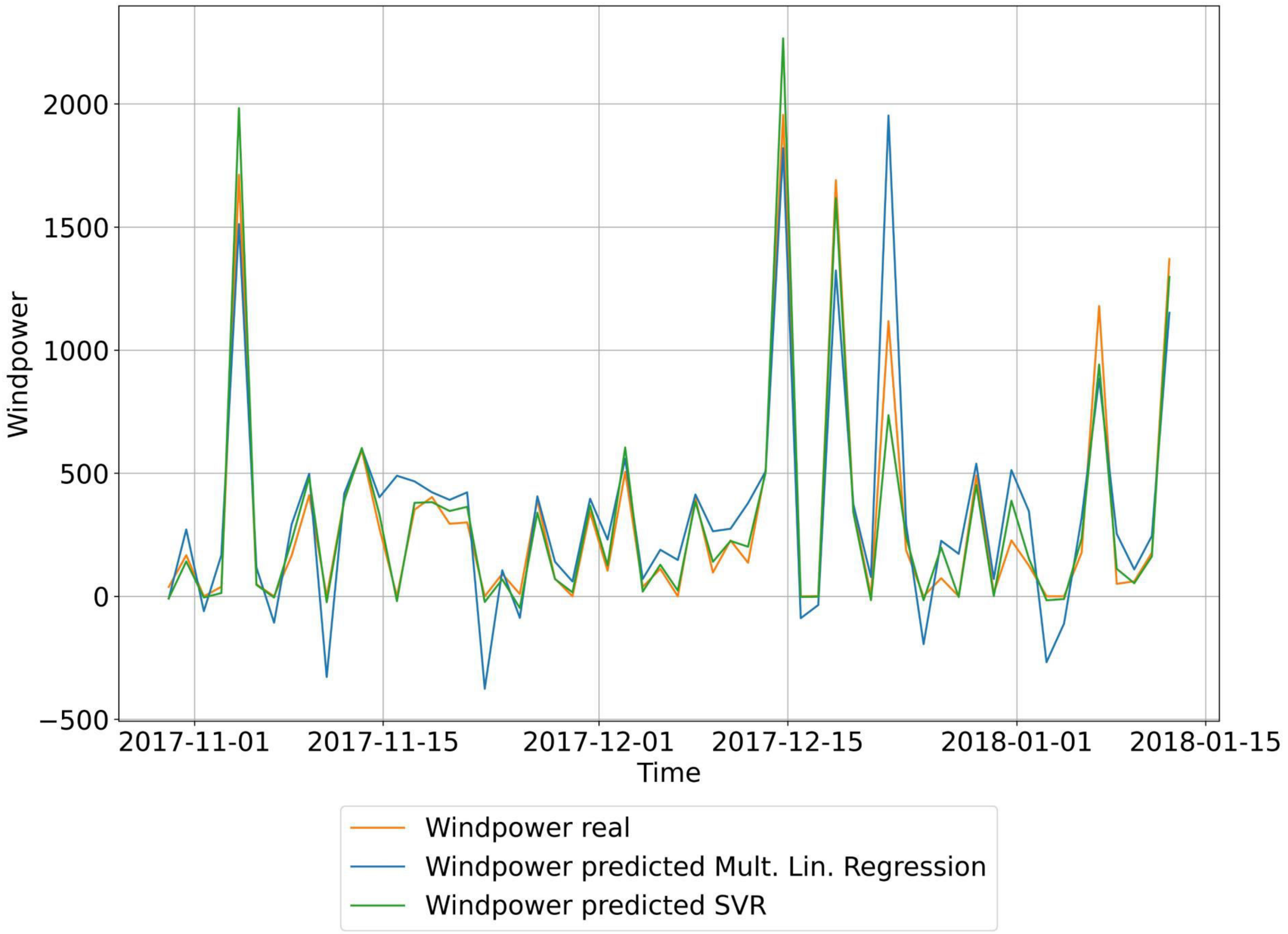


Figure 10

Comparison

	Multiple Linear Regression	Support Vector Regression
MAE	133.29	39.58
RMSE	188.86	60.87
R2	0.829	0.982
CRPS	174.04	122.19

Table 4.1: Evaluation scores to compare the Multiple Linear Regression model to the Support Vector Regression model

Figure 11

05 Conclusion

Conclusion

Discussion

Conclusion

- Main work
  - data processing
  - parameter selection
  - modeling a Multiple Linear Regression model
  - modeling a Support Vector Regression model
  - evaluating and comparing the different models
- SVR approach works better for this particular wind farm dataset
- We assume an SVR approach generally works better for wind power prediction than a Multiple Linear Regression approach

Discussion

- Thank you so much for your interest and attention
- Any questions?