Final Presentation

- Liam Owens
- Xuan Liu
- Chensheng Wen

[Website Link]

ADELPHI UNIVERSITY
NEW YORK

Agenda:

- Our Goals
- Top Model -- Ridge Regression
- Top Model -- Multiple Linear Regression
- Top Model -- Support Vector Regression
- Comparison of Prediction
- Dash Visualization
- Conclusion

Our Goals

1. What are the important features to predict target variable(“Quantity”)?
2. What are the excellent models to evaluate Corona’s data?
3. What is the difference between data with 2020 and data without 2020?

Top Model 1-- Ridge Regression

Equation:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi} + \varepsilon_i$$

Quantity = -0.02247693 + (0.32916839 * Housing_total_sales_NO_SI) + (0.07424731 * Gray_cement_dispatch_contractor)+ (-0.13035303 * RADAR_Plumbinglag_3) + (0.05395233 * Housing_total_launch_NO_SI) + (0.19237967 * Housing_total_sales) + (0.14497477 * Gray_cement_dispatch_other) + (-0.15374323 * RADAR_Tools) + (0.08755184 * Housing_total_launch_NO_SIlag_4) + (0.160274 * Seasonal) + (-0.01935941 * Offshore_microcredit_loans) + Error

Training & Testing Metrics:

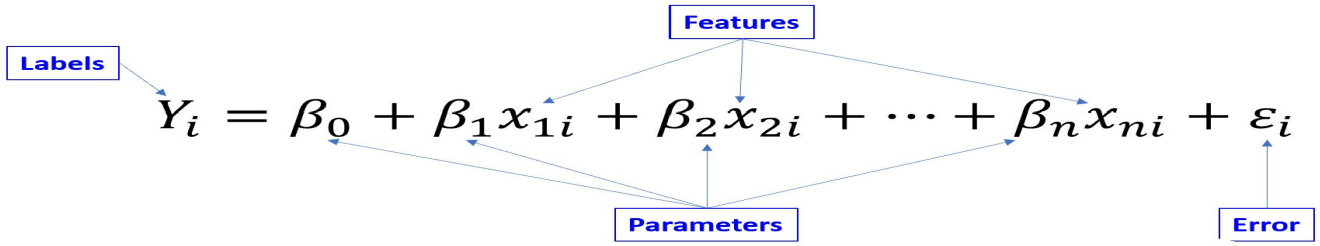
-Training Metrics:
R squared: 0.8297569444513031
Mean Absolute Error: 0.3018142569168214
Mean Squared Error: 0.1300680362503906
Root Mean Squared Error:
0.3606494645086702 5

-Testing Metrics:
R squared: 0.9021347876241389
Mean Absolute Error: 0.32789154500662643
Mean Squared Error: 0.1599865167572618
Root Mean Squared Error:
0.3999831455914884

Ridge Regression - Time Series

Top Model 2 -- Multiple Linear Regression

Equation:



Quantity = -0.02250867 + (0.33720736 * Housing_total_sales_NO_SI) + (0.07147608 * Gray_cement_dispatch_contractor)+(-0.12998208 * RADAR_Plumbinglag_3) + (0.05093813 * Housing_total_launch_NO_SI) +(0.19213823 * Housing_total_sales) + (0.14664222 * Gray_cement_dispatch_other) + (-0.15276257 * RADAR_Tools) + (0.08666063 * Housing_total_launch_NO_SIlag_4) + (0.16080055 * Seasonal) + (-0.0191351 * Offshore_microcredit_loans) + Error

Training & Testing Metrics

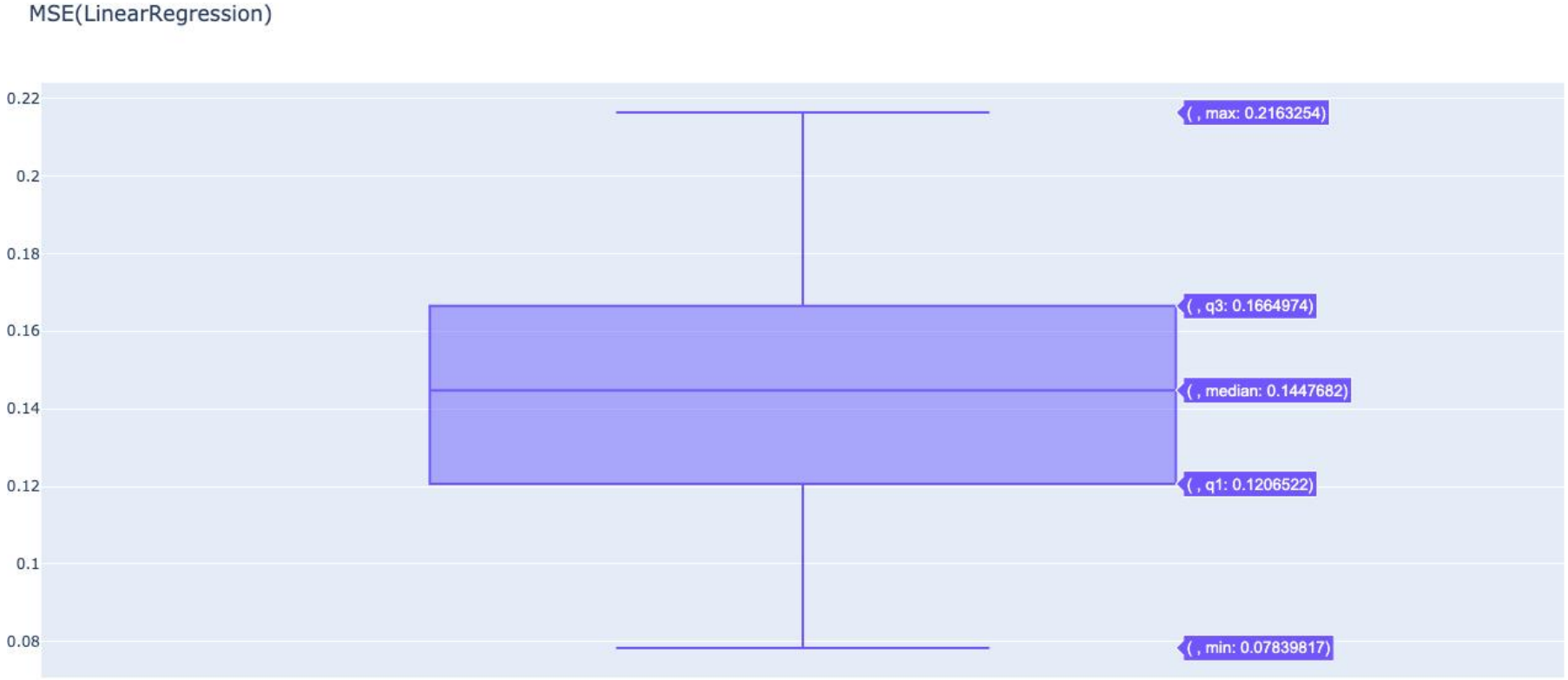
Training Metrics:
R squared: 0.829783817328369
Mean Absolute Error: 0.3018150550687958 6
Mean Squared Error: 0.1300475050026572 2
Root Mean Squared Error: 0.360620999114939 5

Testing Metrics:
R squared: 0.902157512973009 1
Mean Absolute Error: 0.327833111498284 2
Mean Squared Error: 0.159949366177198 07
Root Mean Squared Error: 0.399936702713314 44

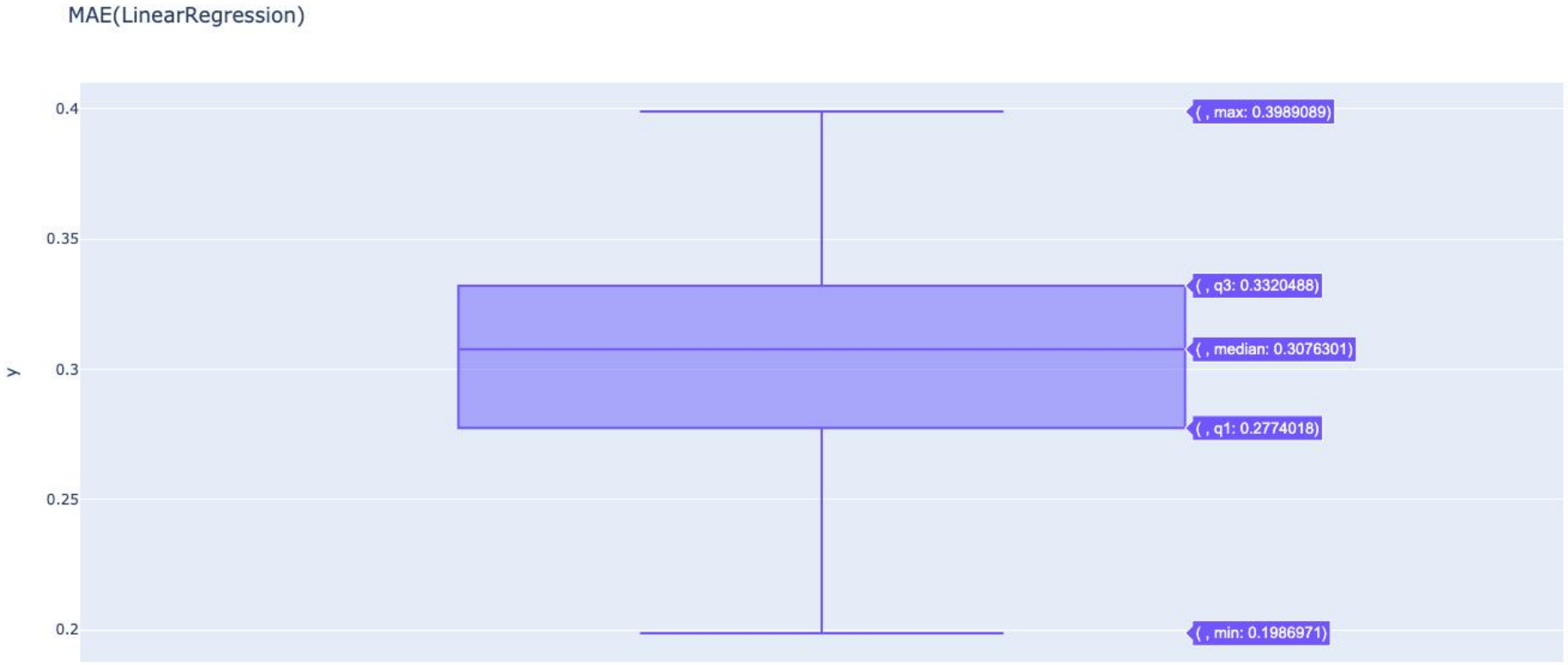
R2 Boxplot


R2(LinearRegression)
y
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
(, max: 0.9603982)
(, q3: 0.8997027)
(, median: 0.8105014)
(, q1: 0.7132471)
(, min: 0.5444621)

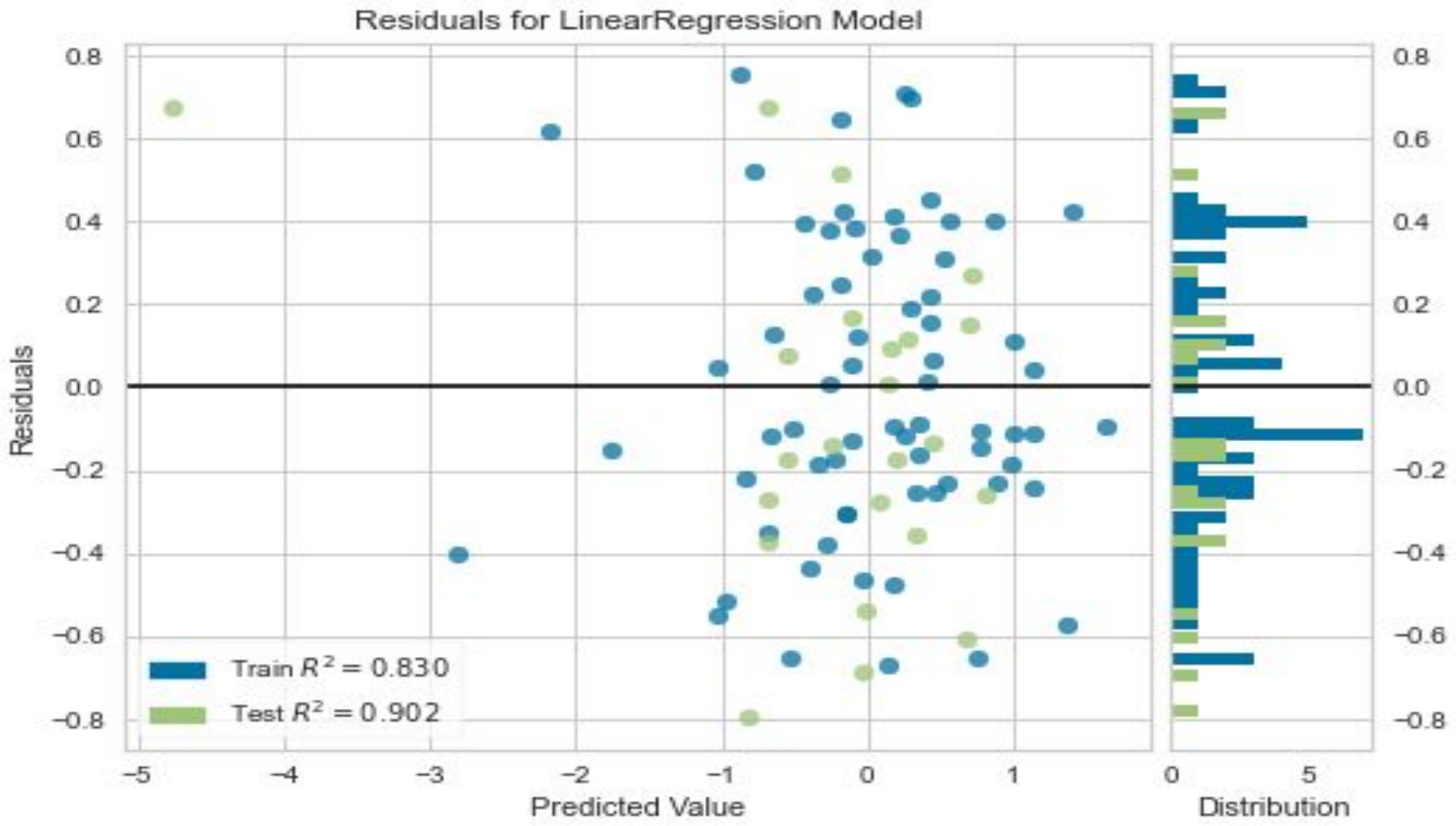
MSE Boxplot


MSE(LinearRegression)
0.22
0.2
0.18
0.16
0.14
0.12
0.1
0.08
(, max: 0.2163254)
(, q3: 0.1664974)
(, median: 0.1447682)
(, q1: 0.1206522)
(, min: 0.07839817)

MAE Boxplot


MAE(LinearRegression)
y
0.4
0.35
0.3
0.25
0.2
(, max: 0.3989089)
(, q3: 0.3320488)
(, median: 0.3076301)
(, q1: 0.2774018)
(, min: 0.1986971)

Residual Plot



Top Model 3-- Support Vector Regression

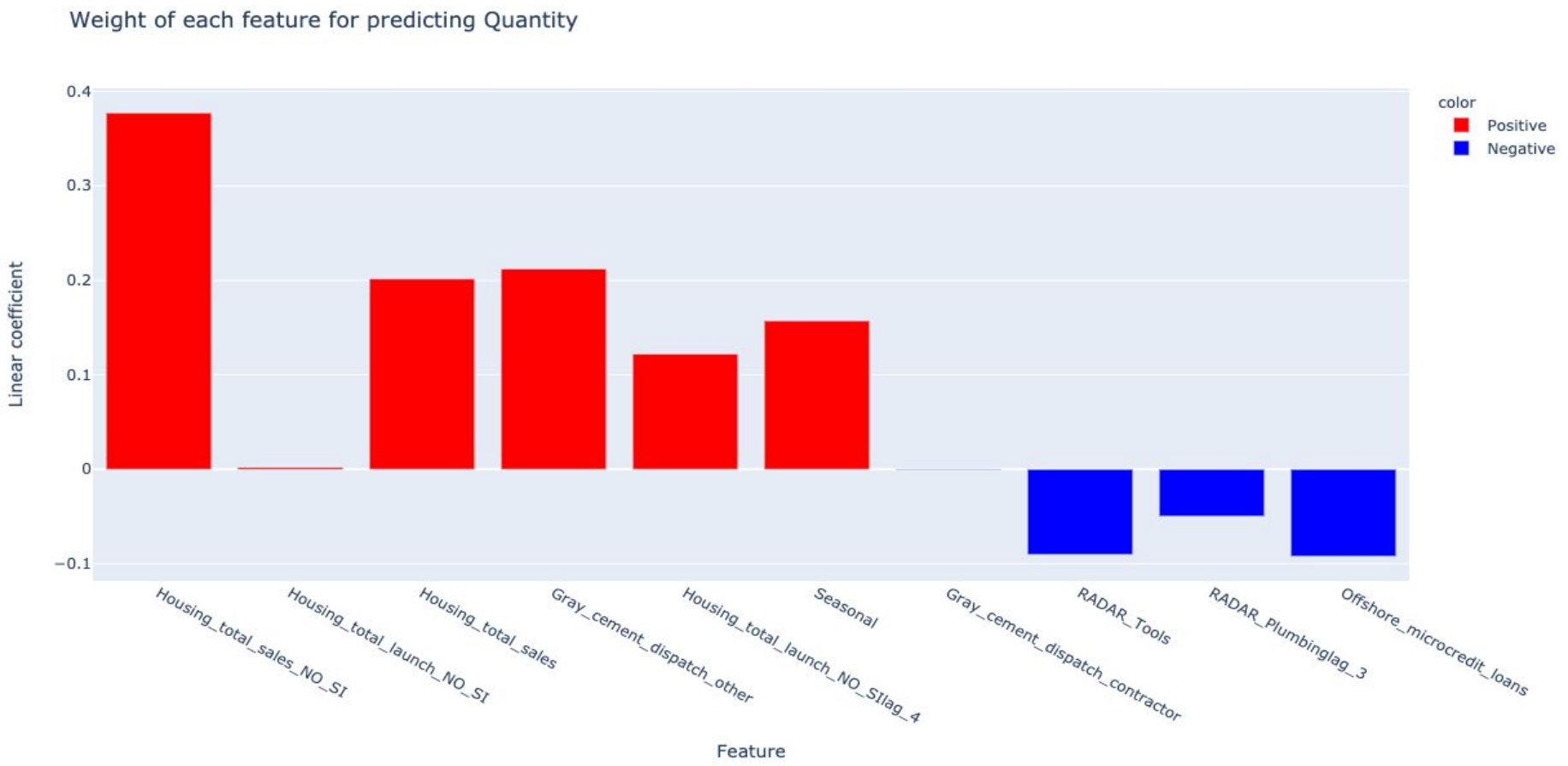
without 2020 data
Training Metrics:
R squared: 0.8229853456278702
Mean Absolute Error:
0.30454132150801483
Mean Squared Error:
0.1352416308995271 2
Root Mean Squared Error:
0.367752132420095

Testing Metrics:
R squared: 0.8987589015416149
Mean Absolute Error:
0.3220724685425985
Mean Squared Error:
0.165505293472710 5
Root Mean Squared Error:
0.4068234180485564 3

with 2020 data
Training Metrics:
R squared: 0.7008272989028212
Mean Absolute Error:
0.3363752347981861 5
Mean Squared Error:
0.1690571216228632 4
Root Mean Squared Error:
0.4111655647337982 3

Testing Metrics:
R squared: 0.9090677902662546
Mean Absolute Error:
0.3372428066811648 5
Mean Squared Error:
0.2003518037138827 3
Root Mean Squared Error:
0.447606751193369 2

Weight of Each Feature for Predicting Quantity



Comparison of Prediction

ADELPHI UNIVERSITY

Dash Visualization

http://127.0.0.1:8050/

Conclusion

- These are the most important features:
  - Housing_total_sales_NO_SI
  - Gray_cement_dispatch_contractor
  - RADAR_Tools,Housing_total_launch_NO_SI
  - Housing_total_sales
  - Gray_cement_dispatch_other
  - RADAR_Plumbinglag_3
  - Housing_total_launch_NO_SIlag_4
  - Seasonal
  - Offshore_microcredit_loans
- The best model is Support Vector Regression
- The trend of data is different.

All the plotly graphs we made are on this website: **Website Link**