TESLA STOCK DATA

TESLA

Dr. Nawana Coyle

TESLA STOCK DATA

Problem

Looking at previous data in a given timeframe, can we predict the adjusted closing price for TELSA stocks?

WHY ADJUSTED CLOSING PRICE IS IMPORTANT?

- Adjusted closing price looks at - dividends
  stock splits
  new stock offerings
- It begins where the closing price ends – Therefore, it's a more accurate measure of stock's value
- Helps investors evaluate the value of the stock quickly.
- Help compare stocks, especially in long term investments

WHY IS PREDICTING ADJUSTED CLOSING VALUE IMPORTANT?

- Cars with clean energy is the focus of the future.
- The popularity of electric cars have risen significantly which indicate larger sales in future.
- More sales generally increases the stock values, which is shown in Tesla stock prices.
- Predicting better stock values gets more investors interested in investing in the product.

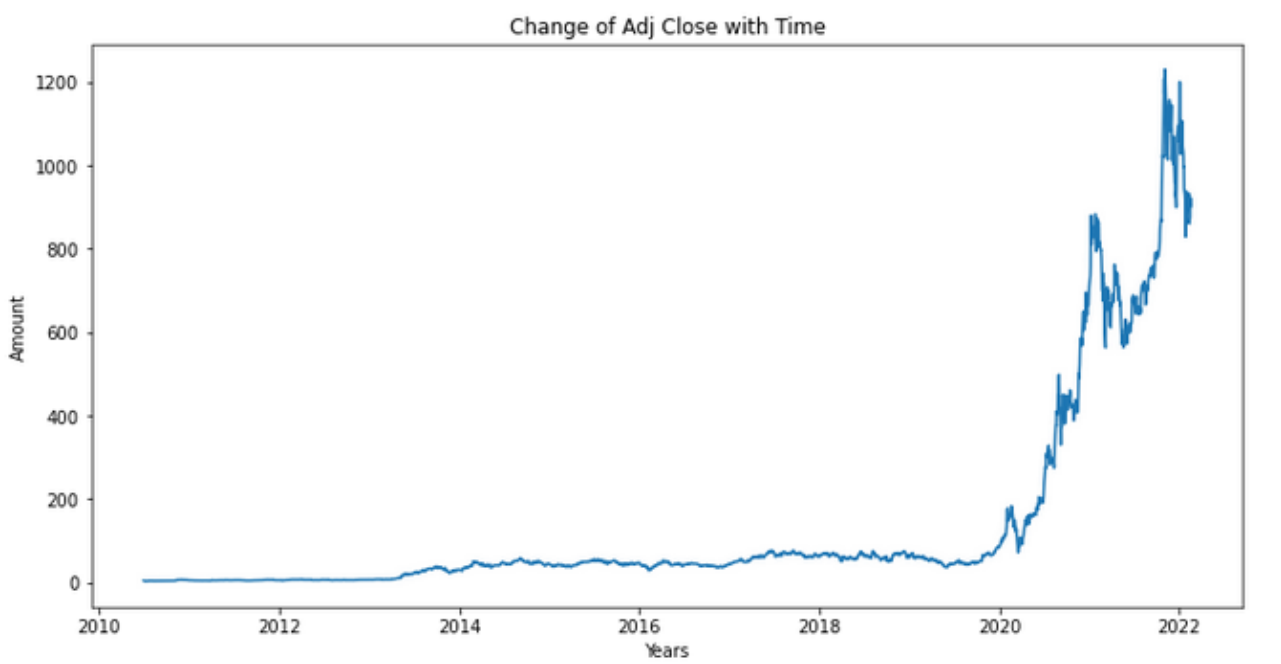
DATA FOR THE PROJECT

Original data for this project came from Kaggle.com at https://www.kaggle.com/datasets/varpit94/tesla-stock-data-updated-till-28jun2021

DATA CLEANING

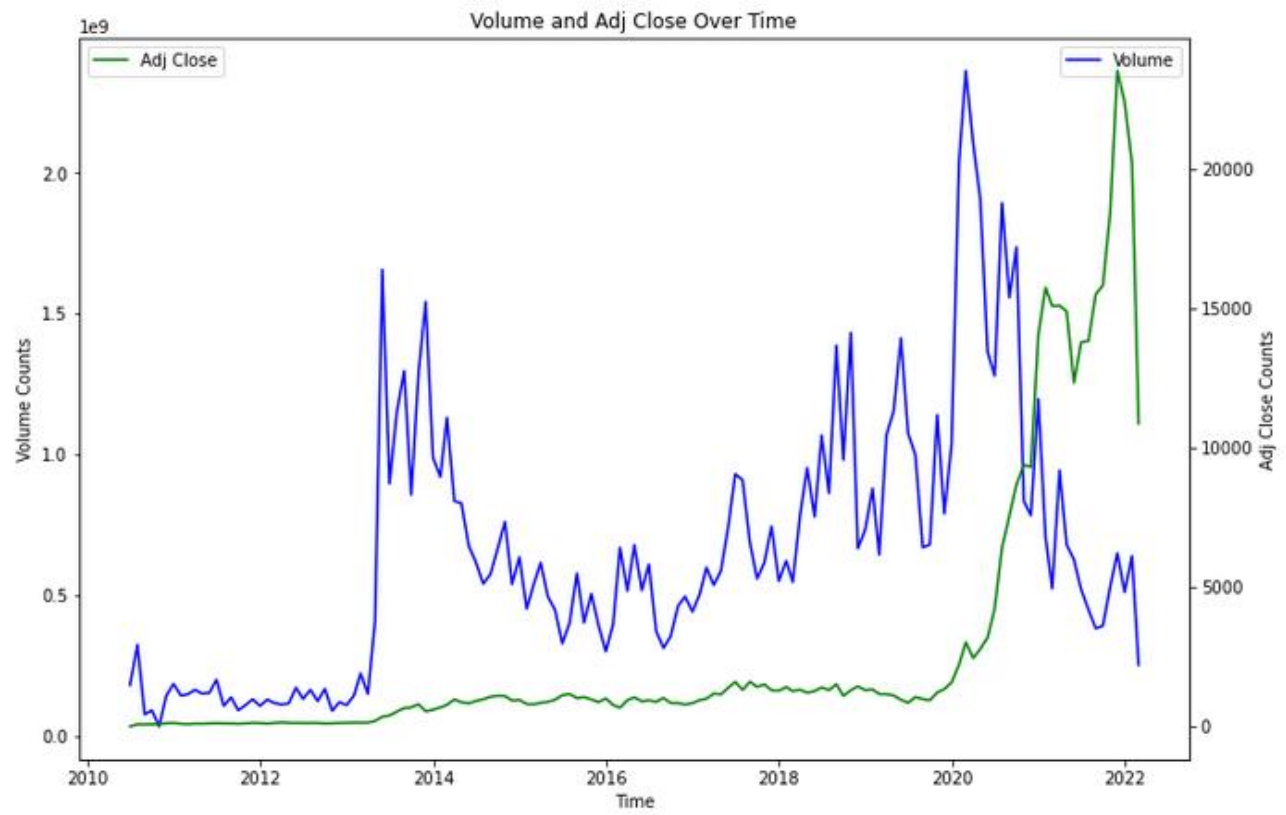
- Use pase_dates → convert the custom formatted date string to a standard formatted data string
- No missing values → No need to fill
- Apply toordinal() function to change the datetime dataset into an integer for fitting models.

EXPLORATORY DATA ANALYSIS (EDA)



- .info() → All data are of numeric data types
- No missing values
- There's a significant degree of increase in the adjusted closing price since 2020 → increase in interest in investing in Tesla ???

EDA CONTINUED…



- The volume of sales have an increased significantly at the beginning of 2020.
- The adjusted closing price gets a significant increase in January in 2021 and January 2022.

EDA CONTINUED…

- First quarter of 2020 seems to have the highest volume of sales.
- Adjusted closing price in the 1st quarter of 2020 was comparatively low.

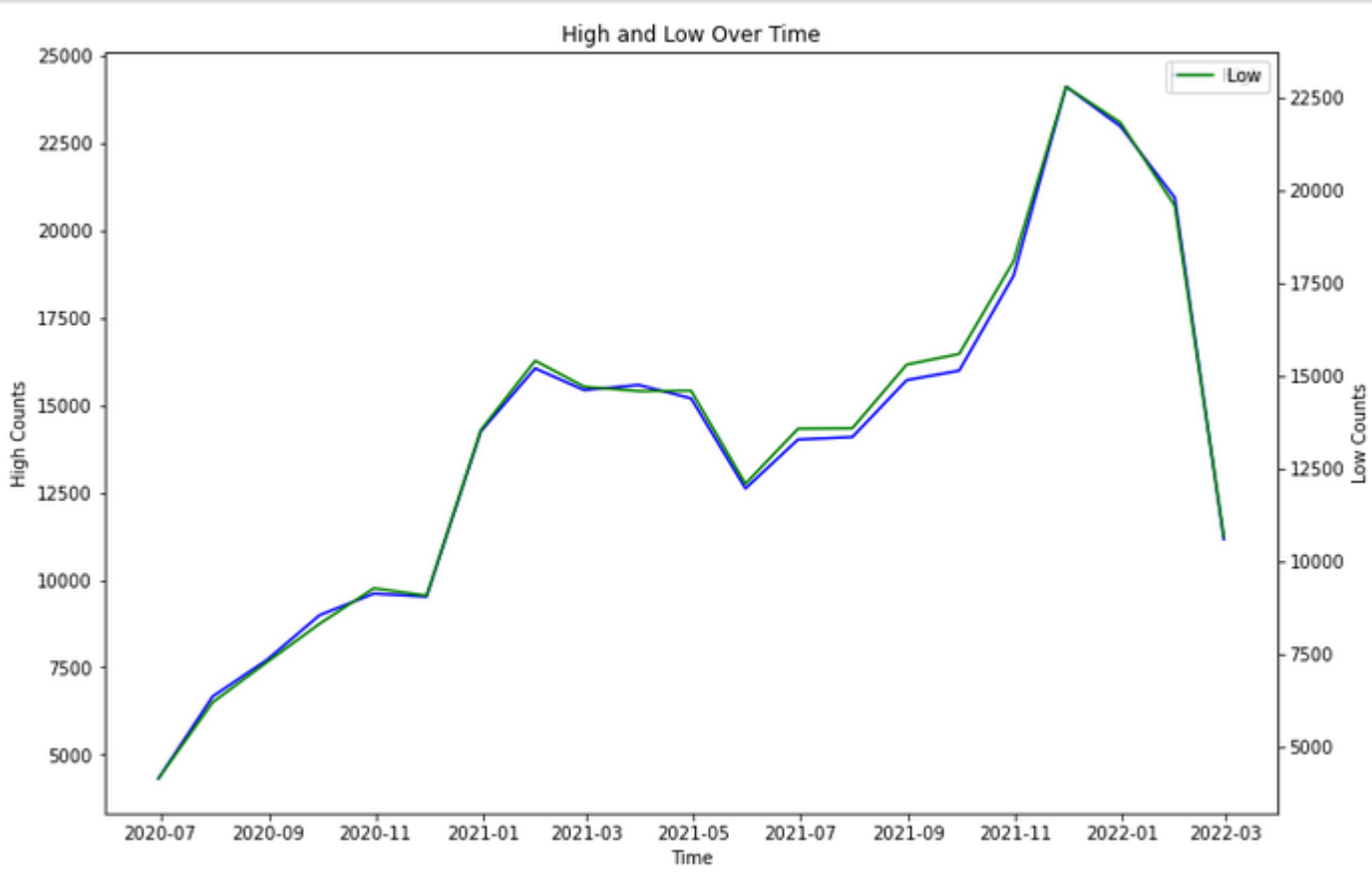
In 1st quarter of 2020→ Sales ⬆ → Investor interest ⬆ → Stock prices of Tesla ⬆ ???

EDA CONTINUED…

- Adjusted closing price is calculated taking closing price, stock splits, dividends and new stock offerings into considerations.
- Stock splits, dividends and new stock offerings brings the adjusted closing price down.
- According the graph, the adjusted closing price is the same as its closing price for Tesla stocks.

- Are there no negative factors which affects the Tesla stock values after closing, which makes it more attractive to investors which explains stock price going up in the recent 2 years ???

EDA CONTINUED…

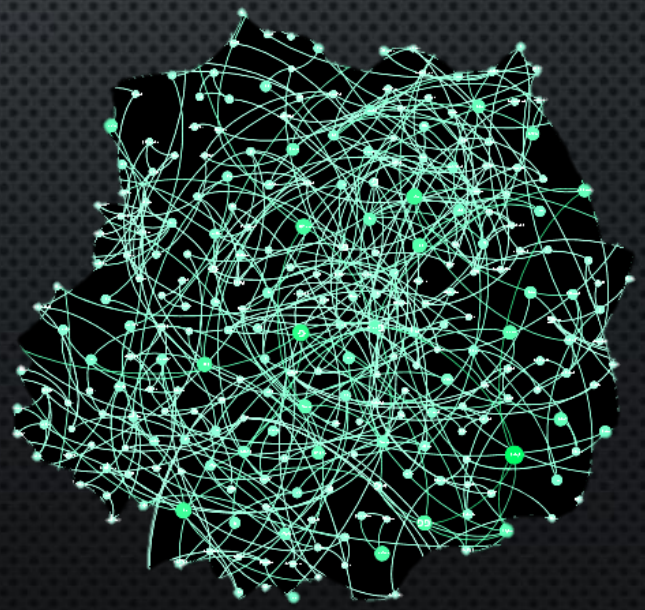


- High: maximum price in a trading day
- Low : minimum price in a trading day
- The shape of sales of each line is very similar → The number of items sold with highest and lowest prices for each month are similar.

DATA MODELING

Models used for the project:

- RandomForestRegressor
- Tuned RandomForestRegressor
- TWeedieRegressor
- Ridge Model

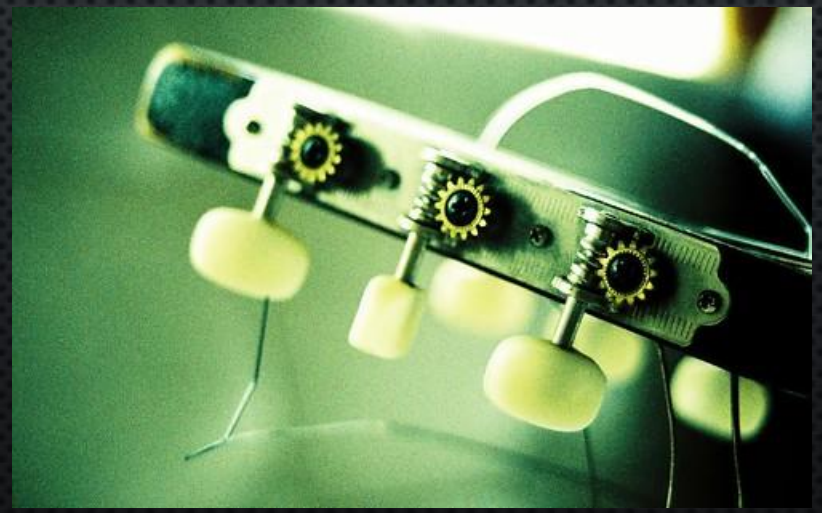


WHY CHOOSE RANDOMFORESTREGRESSOR?

- Simple, highly accurate and easy to use to predict values.
- Uses ensemble learning method which combines predictions from multiple machine learning algorithms to make predictions more accurate.
- Easier to tune hyperparameters during experimentations, and scales well if new features are added to the dataset.

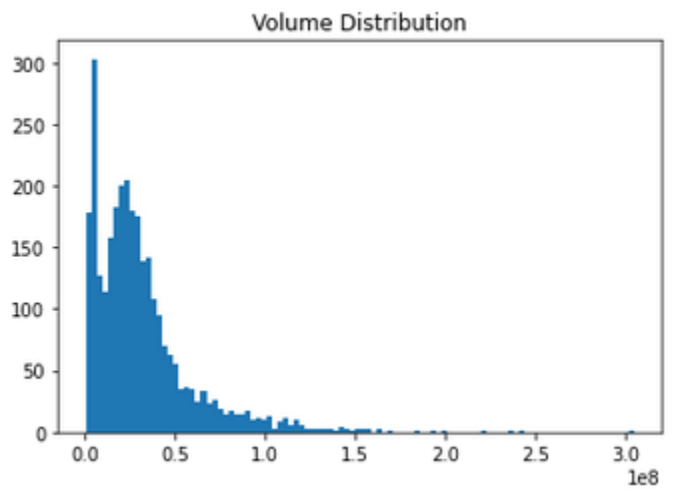
WHY TUNE PARAMATERS?

- Mean_absolute_error (MAE) is a good indication for evaluating RandomForestRegressor model.
- Tuning will help choose specific parameters which can make an impact on the predictions.
- The MAE could be a great metric which could indicate if RandomRorestRegressor or the tuned RandomRorestRegressor model did better in predicting the data.



WHY CHOOSE TWEEDIEREGRESSOR?

- Data show a tweedie distribution, and TweedieRegressor will do great in analyzing data.



(Number of units sold a day show a tweedie distribution)

- This model gives the flexibility to be customized depending on the distribution using the *power* parameter.

p = 0 : Normal distribution
p = 1: Poisson distribution
1 < p < 2: Compound Poisson/gamma distribution,
p = 2: gamma distribution etc.

WHY CHOOSE RIDGE MODEL?

- This model helps regularize coefficient estimates, and bringing the coefficients close to 0 which works better with new data.
- Solve the problems associated with overfitting – calculates the squared error differences between the predicted value and actual labels
- Many independent variables in the dataset has high intercorrelation - multicollinearity

  E.g.: closing price and adjusted closing price

  Highest and lowest sales prices of items on a given day

MODELING

1 Splitting data

- 2 methods of data splitting were used to understand which dataset would do better in making predictions:
    1. First 70% of data as training dataset and remainder as test dataset
    2. Randomly chosen 70% of data as training set with train_test_split () function
- Mean_absolute_error (MAE) metric was used to evaluate both datasets.

Observation:

- Randomly split dataset using train_test_split() function did much better than the other.
- MAE 0.6847 vs 73629.6424

Conclusion:

Since there is a significant degree of elevation in the graph later in years, compared to the first 8 years, randomly chosen dataset provided much better results.

MODELING CONTINUES…

2. Hyperparameter tuning

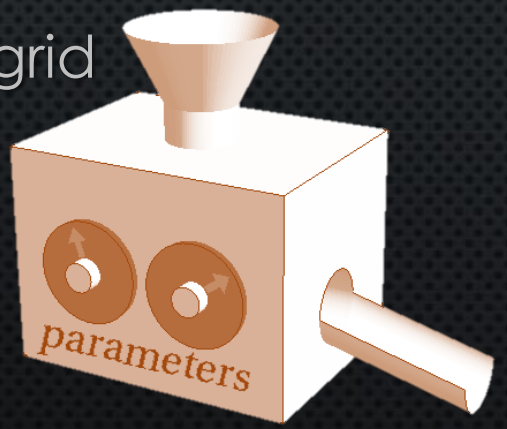
a) A grid was created with parameters that needed to be tuned.

```
grid = {'n_estimators': np.arange(10,100,10),
        'max_depth': [None, 3,5,10],
        'min_samples_split': np.arange(2,20,2),
        'min_samples_leaf': np.arange(1,20,2),
         'max_features': [0.5, 1, 'sqrt', 'auto']}
```

b) Initiated RandomizedSearchCV with param_distributions = grid

c) Get the best parameters with model.best_params_

d) Assign the best parameters to the model.



MODELING CONTINUES…

3. Calculating mean_absolute_error (MAE) for all models

a) Created a function to calculate all 4 models: score_models()
b) Created an empty dictionary to save the scores of all: model_scores = { }
c) Created an empty dictionary to save all predictions the models made: all_predictions = { }
d) Run the score_models() function

Results:

```
{'RandomForestRegressor MAE': 0.6847052478636401,
 'Tuned RandomForest MAE': 1.1584727843960612,
 'TweedieRegressor MAE': 0.3769953183456556,
 'Ridge Model MAE': 4.944438644064918e-05}
```

EVALUATING MODELS

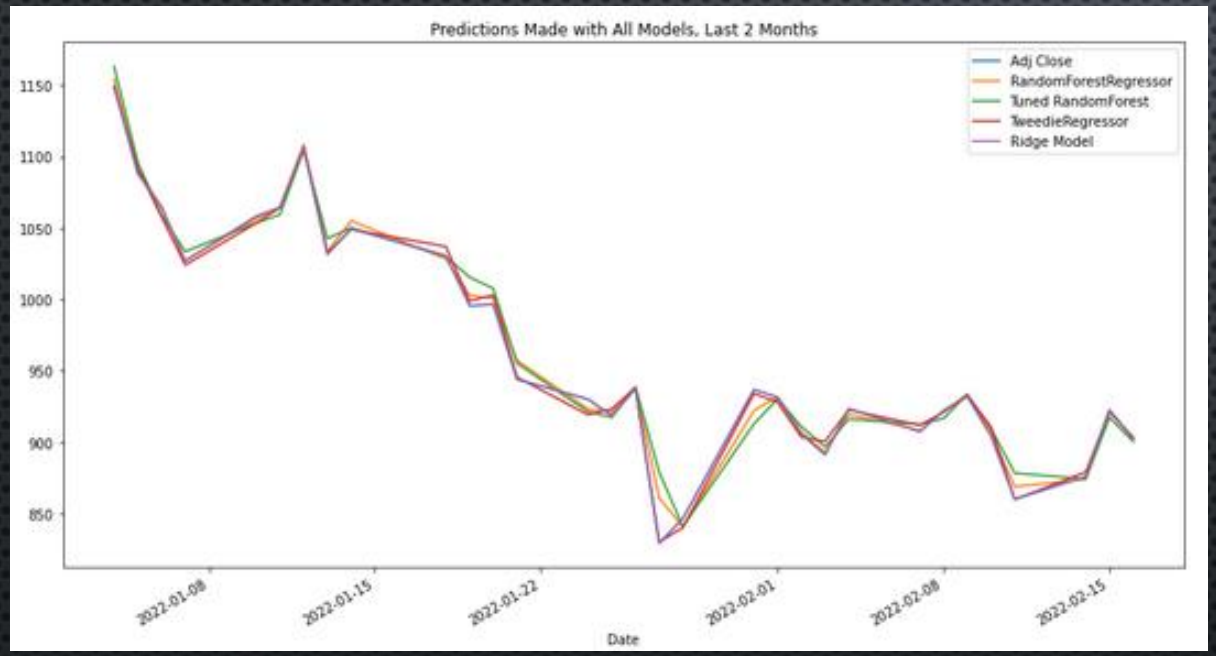
```
{'RandomForestRegressor MAE': 0.6847052478636401,
 'Tuned RandomForest MAE': 1.1584727843960612,
 'TweedieRegressor MAE': 0.3769953183456556,
 'Ridge Model MAE': 4.944438644064918e-05}
```

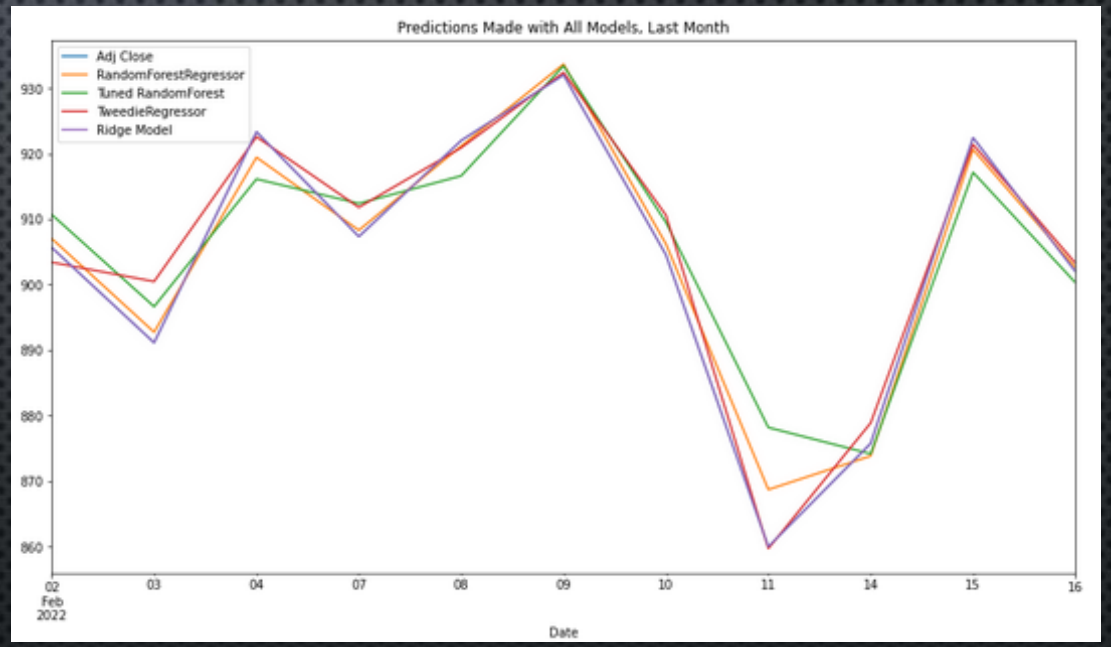
- Ridge MAE < TweedieRegressor MAE, RandomForest MAE, tuned RF MAE
- Ridge model has the lowest MAE, followed by RandomForestRegressor

PREDICTIONS MADE BY THE MODELS

	Date	Adj Close	RandomForestRegressor	Tuned RandomForest	TweedieRegressor	Ridge Model
0	2010-06-29	4.778	4.74708	4.421653	5.380991	4.778000
1	2010-06-30	4.766	4.76518	4.766238	5.528464	4.766033
2	2010-07-01	4.392	4.40000	4.403660	4.710741	4.392005
3	2010-07-02	3.840	3.85242	3.885240	4.095452	3.840007
4	2010-07-06	3.222	3.33950	3.440041	3.528185	3.222008
5	2010-07-07	3.160	3.25608	3.381507	3.455507	3.160000
6	2010-07-08	3.492	3.43342	3.381507	3.808026	3.491998
7	2010-07-09	3.480	3.48224	3.422187	3.700299	3.479999
8	2010-07-12	3.410	3.45250	3.461547	3.607090	3.410003
9	2010-07-13	3.628	3.62594	3.573771	3.830443	3.628000
10	2010-07-14	3.968	3.92486	3.905588	4.212808	3.968000
11	2010-07-15	3.978	3.97124	3.974537	4.236579	3.978006
12	2010-07-16	4.128	4.12796	4.132491	4.332145	4.128001
13	2010-07-19	4.382	4.38448	4.349765	4.575859	4.382000
14	2010-07-20	4.060	4.06134	4.103525	4.231954	4.060002
15	2010-07-21	4.044	4.04114	4.042269	4.189517	4.043999
16	2010-07-22	4.200	4.19946	4.193821	4.362694	4.200000
17	2010-07-23	4.258	4.25290	4.281214	4.415118	4.258001
18	2010-07-26	4.190	4.19618	4.197229	4.319514	4.189998
19	2010-07-27	4.110	4.10798	4.115693	4.258118	4.110001

EVALUATING MODELS CONTINUES…





- Looking at data for the last month and 2 month time, ridge model comes very close to the actual adjusted closing price of Tesla stocks, compared to other models.
- RandomForestRegressor comes second.

SUMMARY OF THE PROJECT

Problem

Looking at previous data in a given timeframe, can we predict the adjusted closing price for TELSA stocks?

Findings

- In 1st quarter of 2020→ Sales ↑ → Investor interest ↑ → Stock prices of Tesla ↑ ???
- Are there no negative factors which affects the Tesla stock values after closing, which makes it more attractive to investors which explains stock price going up in the recent 2 years ???
- Number of items sold with highest and lowest prices for each month are similar.

Results

Ridge regression makes the best predictions, then RandomForestRegressor

AREAS FOR IMPROVEMENTS

- Try different parameters to tune in RandomForestRegressor
- Entertain other regression models
- Collect more data variables other than the 7 variables provided.
- Get clarity on how adjusted closing price for Tesla is calculated and determine if other factors play a role in determining the pricing.
- Conduct more experimentations with hyperparameter running for both Ridge regression and Random Forest Regression

Thank you!