Predicting Adani Port Close Price With Respect To The Open Price

Ganga Prasanna Mogaparthi

School of Computer Science and Information Systems, Northwest Missouri State University

Dataset Introduction and Goals

- Dataset obtained from Kaggle.

- 3323 Rows and 15 Columns.

- Data from 2007 to 2021

- Numeric Features include Prev Close, Open, High, Low, Last, Close, VWAP, Volume, Turnover, Trades, Deliverable Volume and %Deliverable.

- Replaced null values in Trades column with 0.

- Goal: Predicting the Close price based on the Open price.

Models Used

- Supervised Learning: The dataset was split into 75% training and 25% test set.
  - Linear Regression Model
  - Decision Trees
  - SVM
  - Random Forest
  - Clustering
  - Neural Nets

Model Outputs

```
LinearRegression()
value of R2
0.9956225580219896
The value of Mean Squared Error is:  162.1680920810809
The root mean squared error is  12.734523629923537
```
Linear Regression

```
Accuracy = 0.5618892508143323
Precision = 0.5561292382291975
Sensitivity = 0.5618892508143323
F1 = 0.4775368206817104
```
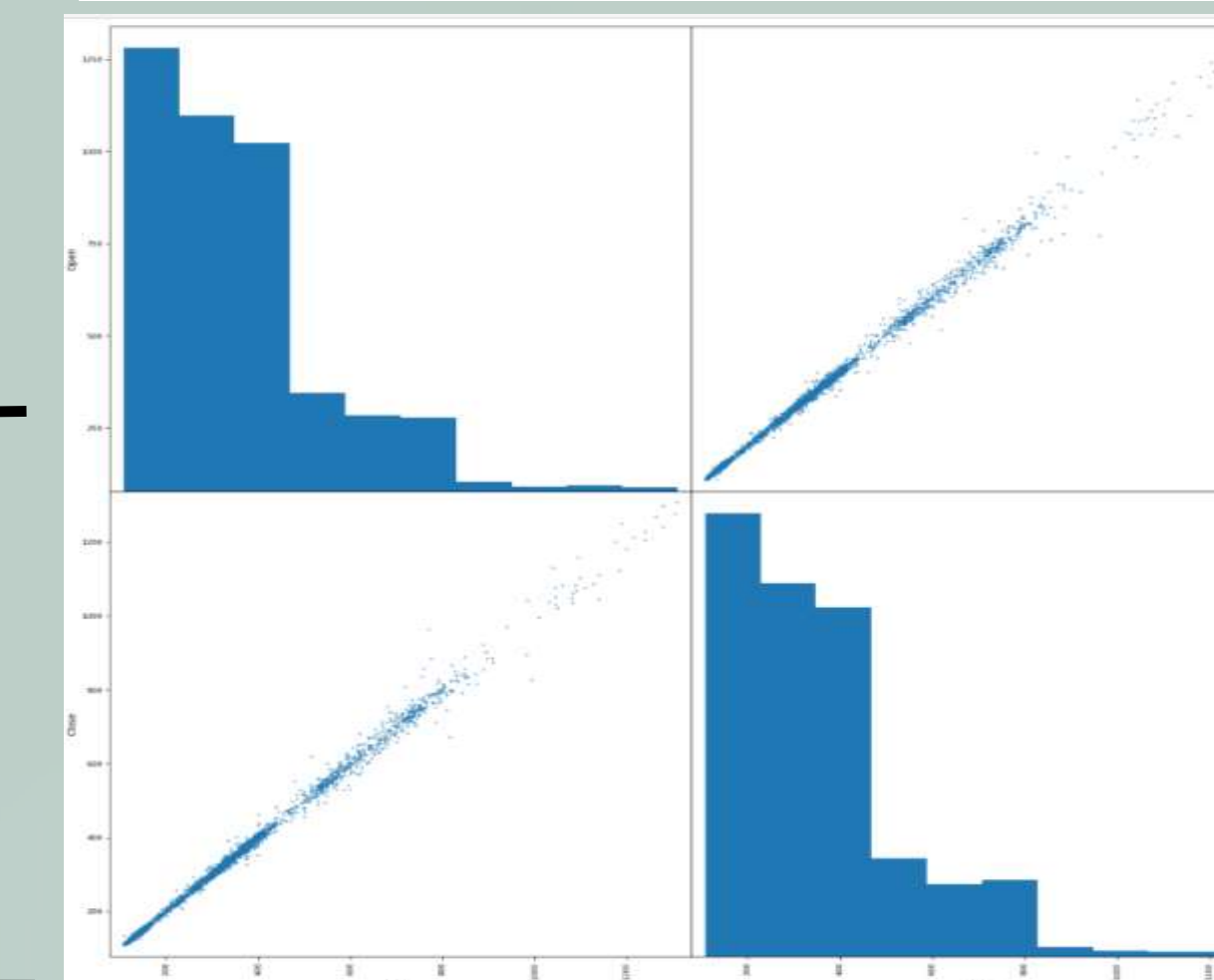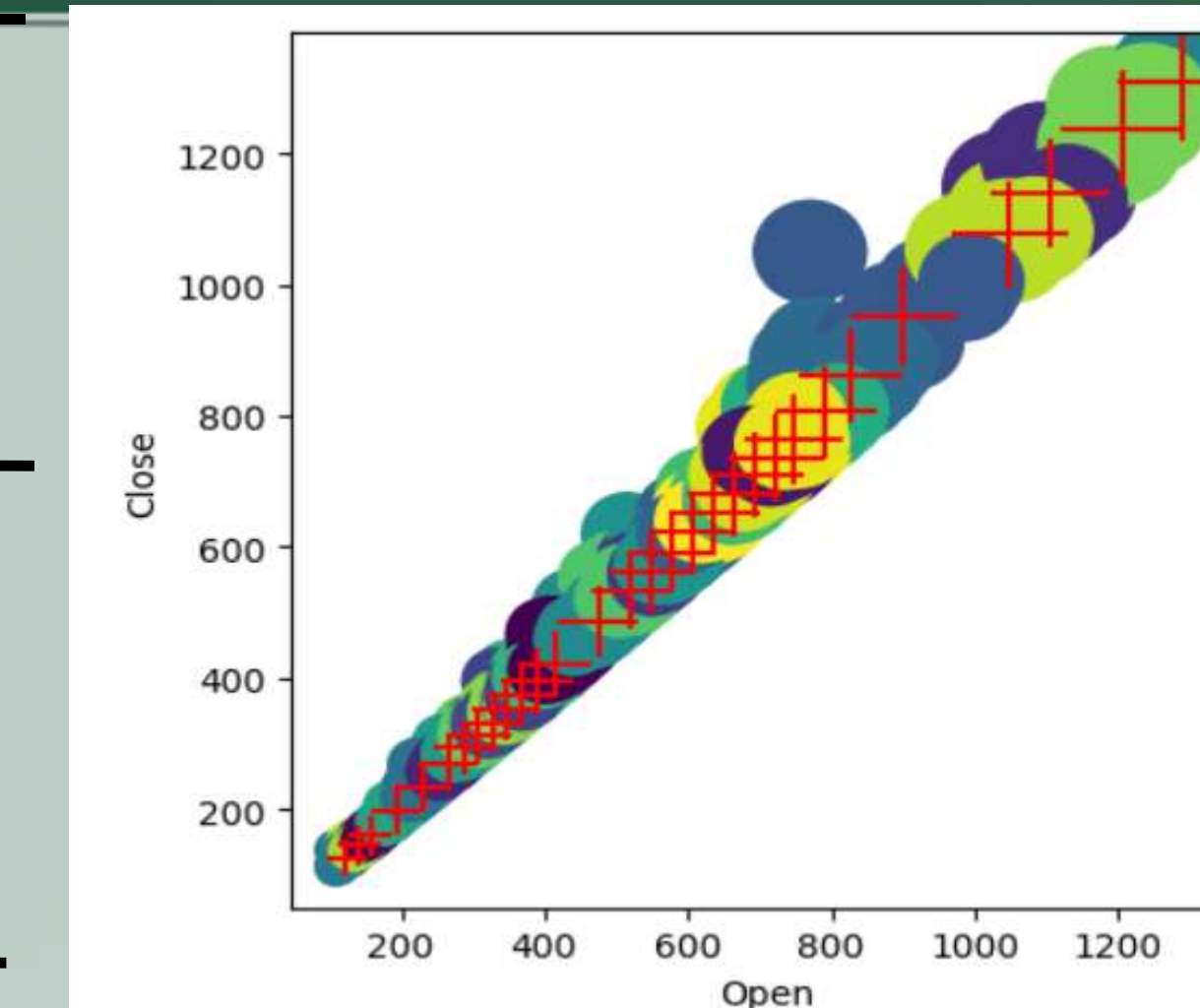**Decision Tree Test Data**

```
Accuracy = 0.5266015200868621
Precision = 0.27730916095779384
Sensitivity = 0.5266015200868621
F1 = 0.3633026134311922
```
SVM Test Set

```
Accuracy (training): 0.9998275340194038
Accuracy (testing): 0.9989571423141078
R2 score : 1.00
Mean squared error: 36.36
Root Mean squared error: 6.03
```
Random Forest
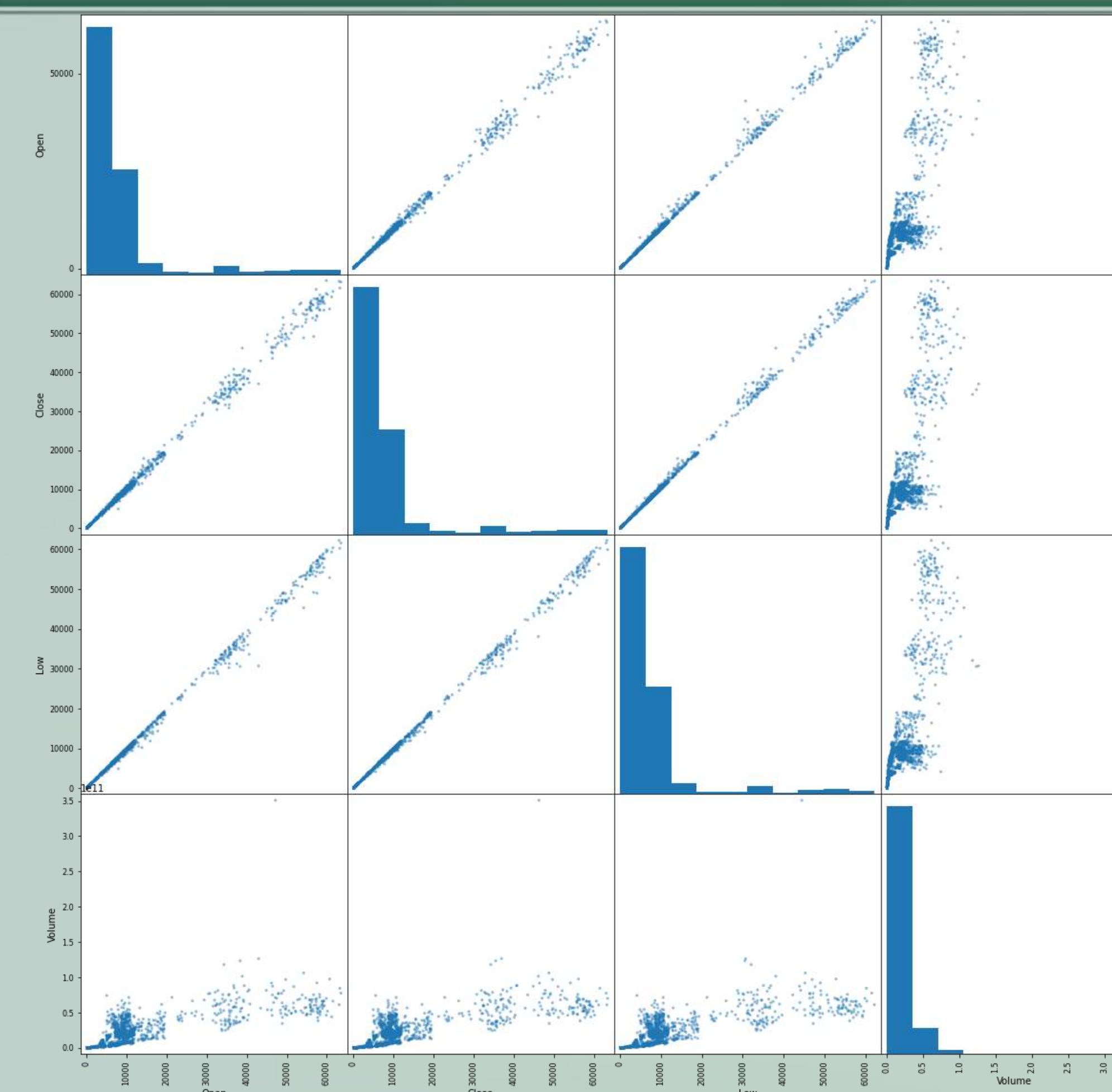

Discussion and Results

Linear Regression Results:
- Linear Regression R2 value is 0.99 shows a strong relationship.

Decision Tree
- Decision Tree has slight increase in performance on the test set with accuracy from 52% to 56%.
- F1 score dropped to 47% from 55% on the test set.

SVM
- SVM accuracies are in the range of 52% which performs similar to decision trees.

Random Forest
- Metrics are very good with a R2 score of 1.0

Interpretation and Conclusion

- A high R2 value of 0.99 for linear regression suggests an excellent fit and shows strong relationship between Open and Close price.

- Performance of Decision tree is low with only 52% of accuracy.

- SVM metrics also seem to be poor. It could be because the dataset is imbalanced, or the kernel choice was not the best.

- Random Forest metrics appeared to be good when compared to other models.

- SVM model with different kernel can be employed in the **future** in order to verify if the metrics are improved.

Additional Resources

- https://www.kaggle.com/datasets/rohanrao/nifty50-stock-market-data?select=ADANIPORTS.csv
- https://github.com/44-566-Machine-Learning-S24/ml-s24-project-MogaparthiGanga

Acknowledgements

Repository

GitHub

https://github.com/44-566-Machine-Learning-S24/ml-s24-project-MogaparthiGanga