Predicting Bitcoin High Price With Respect To The Opening Price

Bikash Adhikari

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

NORTHWEST MISSOURI STATE UNIVERSITY

Dataset Introduction and Goals

- Dataset obtained from Kaggle.
- 2991 Rows and 9 Columns.
- Daily Data from 2013 to 2021
- Numeric Features include prices opening, closing, low, high, volume and market cap.
- Goal: Predicting the bitcoin high price based on the opening price.

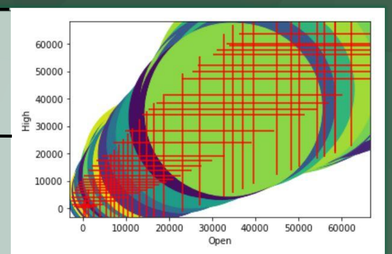
Models Used

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

Model Outputs

```
LinearRegression()
0.9994568683960177
Mean Squared Error:  75744.28108339632
Root Mean Squared Error:  275.21678924694316
```

Linear Regression

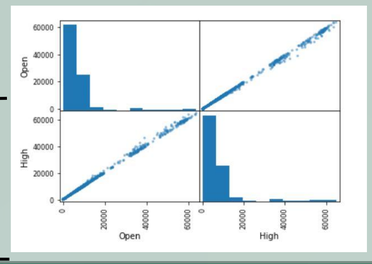



```
Accuracy:  0.982620320855615
Precision:  0.985942915466421
Sensitivity:  0.982620320855615
F1 Score:  0.9824600428406657
```

Decision Tree Test Data

```
Accuracy is  0.7647058823529411
Sensitivity is  0.7647058823529411
Precision is  0.7057921541660694
F1 is  0.7176769726075992
```

SVM Test Set




```
Accuracy (training): 0.9999296833312191
Accuracy (testing): 0.9994286956435832
R2 score : 1.00
Mean squared error: 79673.21
Root Mean squared error: 282.26
```

Random Forrest

Discussion and Results

Linear Regression Results:

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

Decision Tree

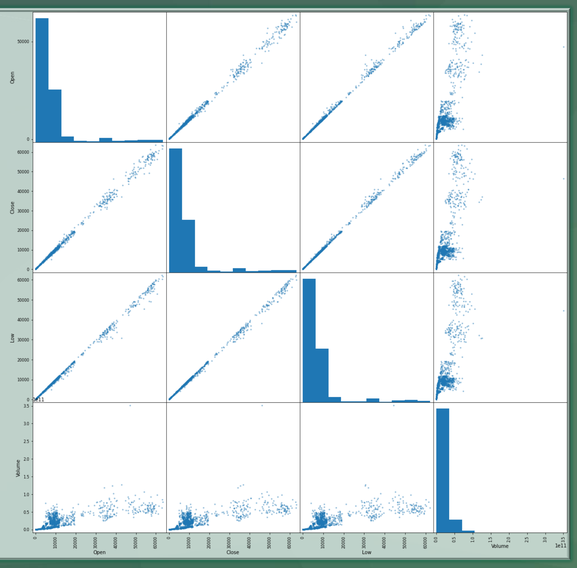
- Decision Tree was overfitting on the train set with all metrics exactly equal to 1.0
- Overfitting is classic for decision trees.
- However, analysis on the test set yields metrics close to 0.98 and interpreted in the conclusion.

SVM

- SVM accuracies are in the upper 0.70 and performs lesser than decision trees.

Random Forrest

- Metrics are very good with a R2 value 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 opening and high price.
- It was suspected that the decision tree was overfitting and while it could have, the metrics for test set are high showing good predictions.
- SVM metrics seem to be poor. It could be because the dataset is imbalanced, or the kernel choice was not the best.
- To address overfitting of the decision tree metrics Random Forrest with 1000 trees was employed and high metrics yielded suggest it might not be overfitting.
- A PCA model can be employed in the **future** in order to determine if the predictions can be based off a single variable.

Additional Resources

- https://www.kaggle.com/datasets/sudalairajkumar/cryptocurrencypricehistory
- https://www.kaggle.com/datasets/jessevent/all-crypto-currencies
- https://github.com/44-599-machine-learning-S22/project-machine-learning-s22-bikash30851

Acknowledgements

Kaggle.com

Dr. Charles Hoot

Repository

GitHub

https://github.com/44-599-machine-learning-S22/project-machine-learning-s22-bikash30851