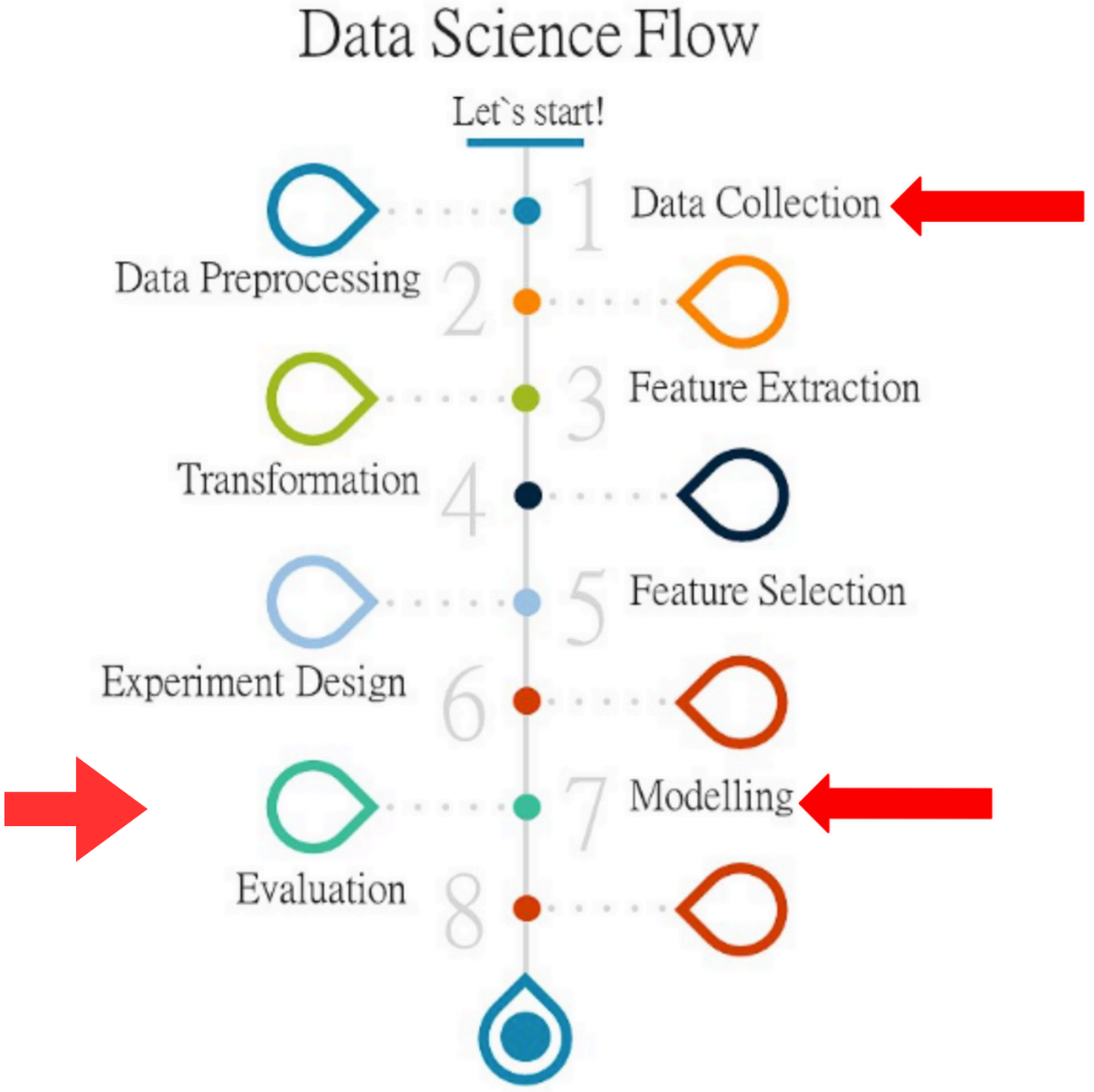
Applying Data Science to Predict Stock Prices

Group 24

Agenda

1. Data Science Flow
2. Problem Definition
3. What you can learn
4. Implementation
5. Conclusion


Data Science Flow
Let`s start!
1
Data Collection
Data Preprocessing
2
3
Feature Extraction
Transformation
4
5
Feature Selection
Experiment Design
6
7
Modelling
Evaluation
8

Problem definition

1. Problem Statement

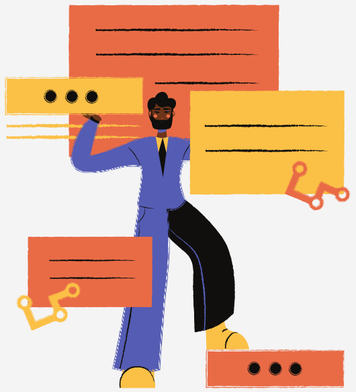
- Predicting stock prices accurately is crucial for investors and traders to make informed decisions. However, it's challenging due to the complex nature of financial markets and various influencing factors

2. Objective

- Develop a model to predict stock prices based on historical data, aiding investors in making better trading decisions

What you can learn

- **Data Science Application:** machine learning, for analysis and prediction
- **Model Development:** data collection, preprocessing, and evaluation.
- **Evaluation techniques**: Mean Squared Error, T-test, Wilcoxon-test, and Feature Importance
- **Real-world application:** how these techniques can be applied to real-world scenarios

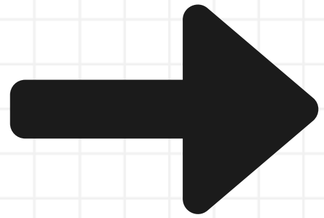


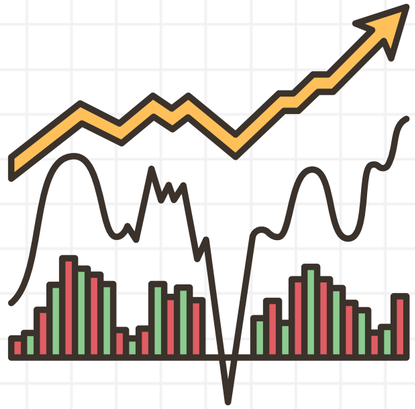
3. Implementation

Step 1: Data Collection

- Utilize Yahoo Finance API to fetch historical stock price data
- Take TSMC34.SA for example

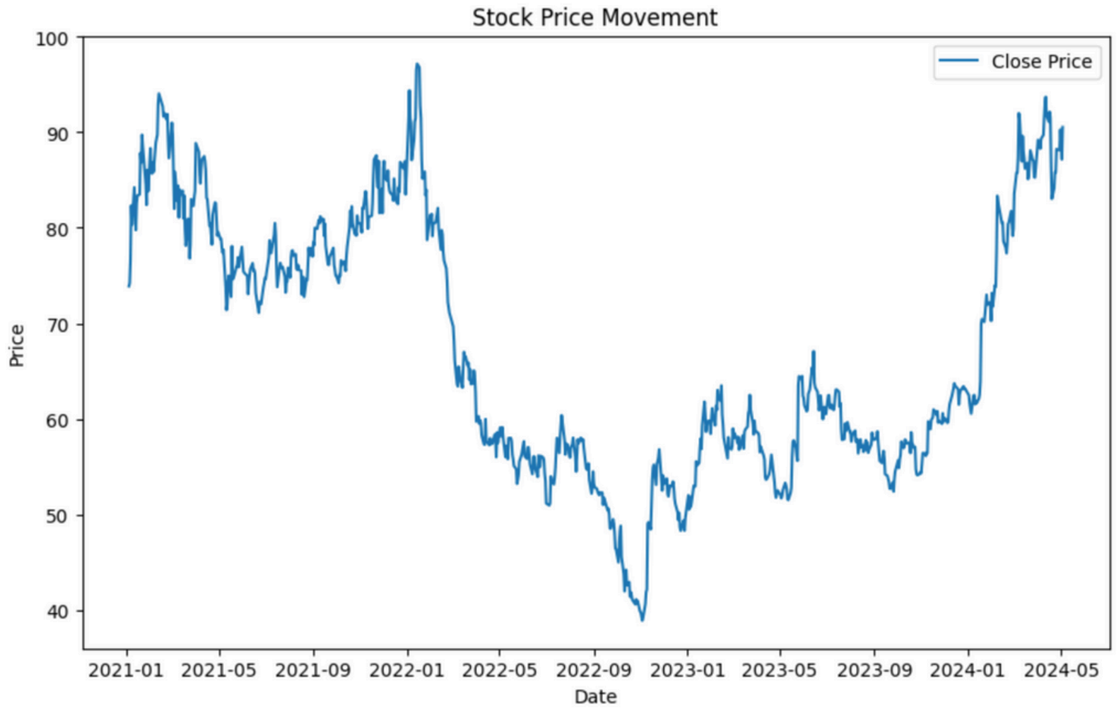
YAHOO! FINANCE

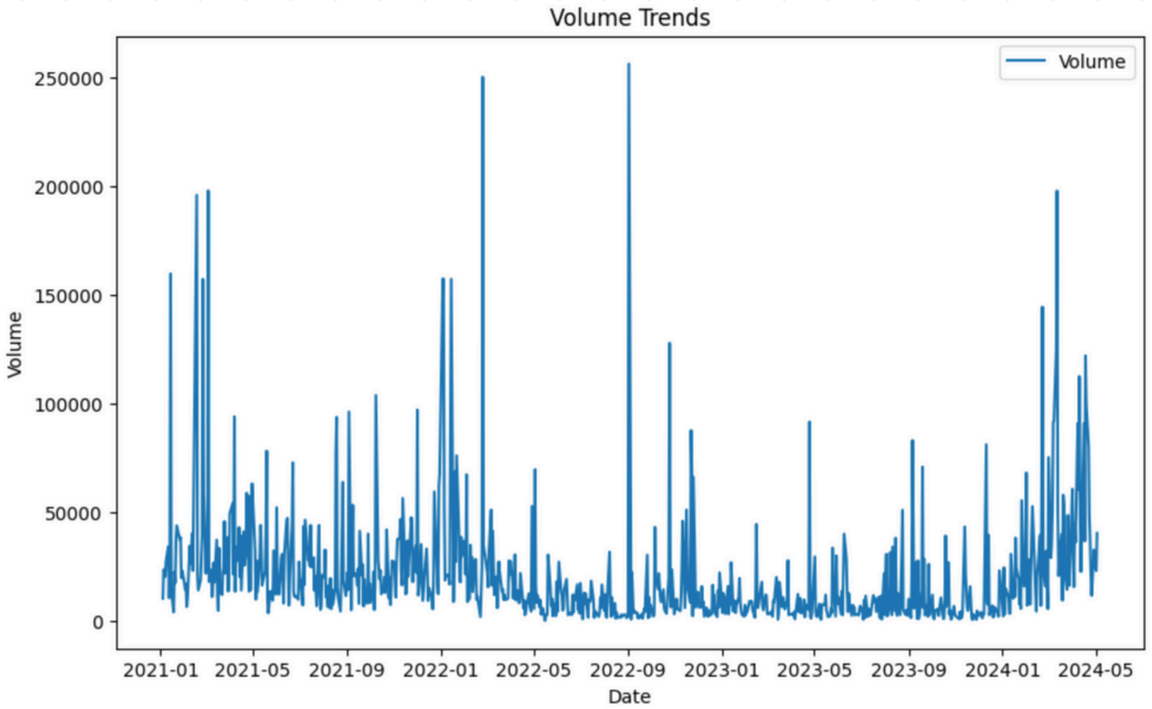




Step2: Data Analysis

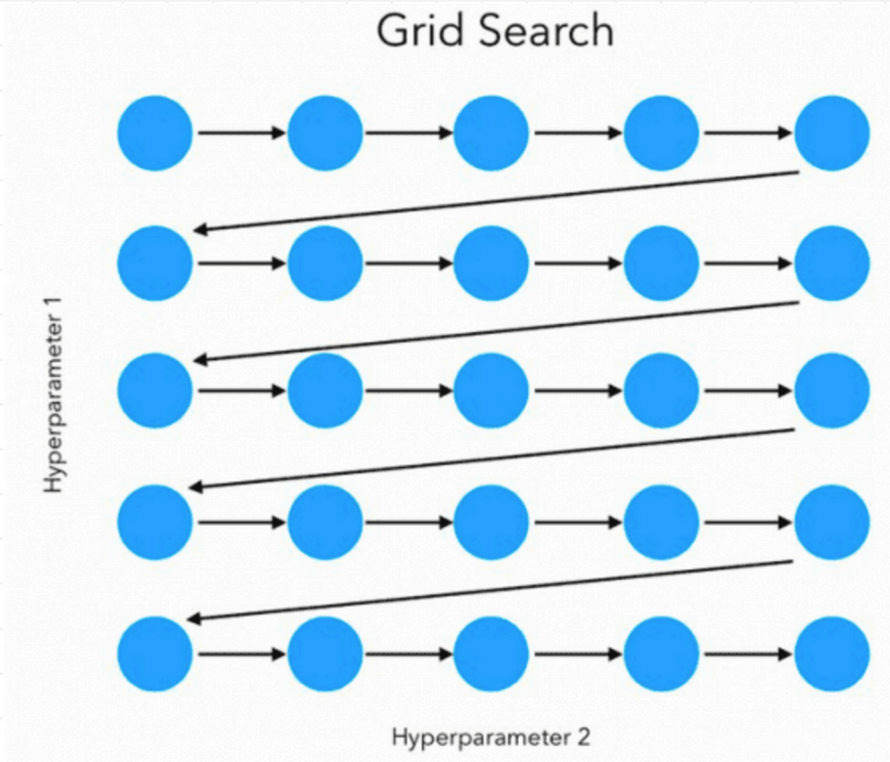
- Visualize stock price movements and volume trends to understand patterns and correlations



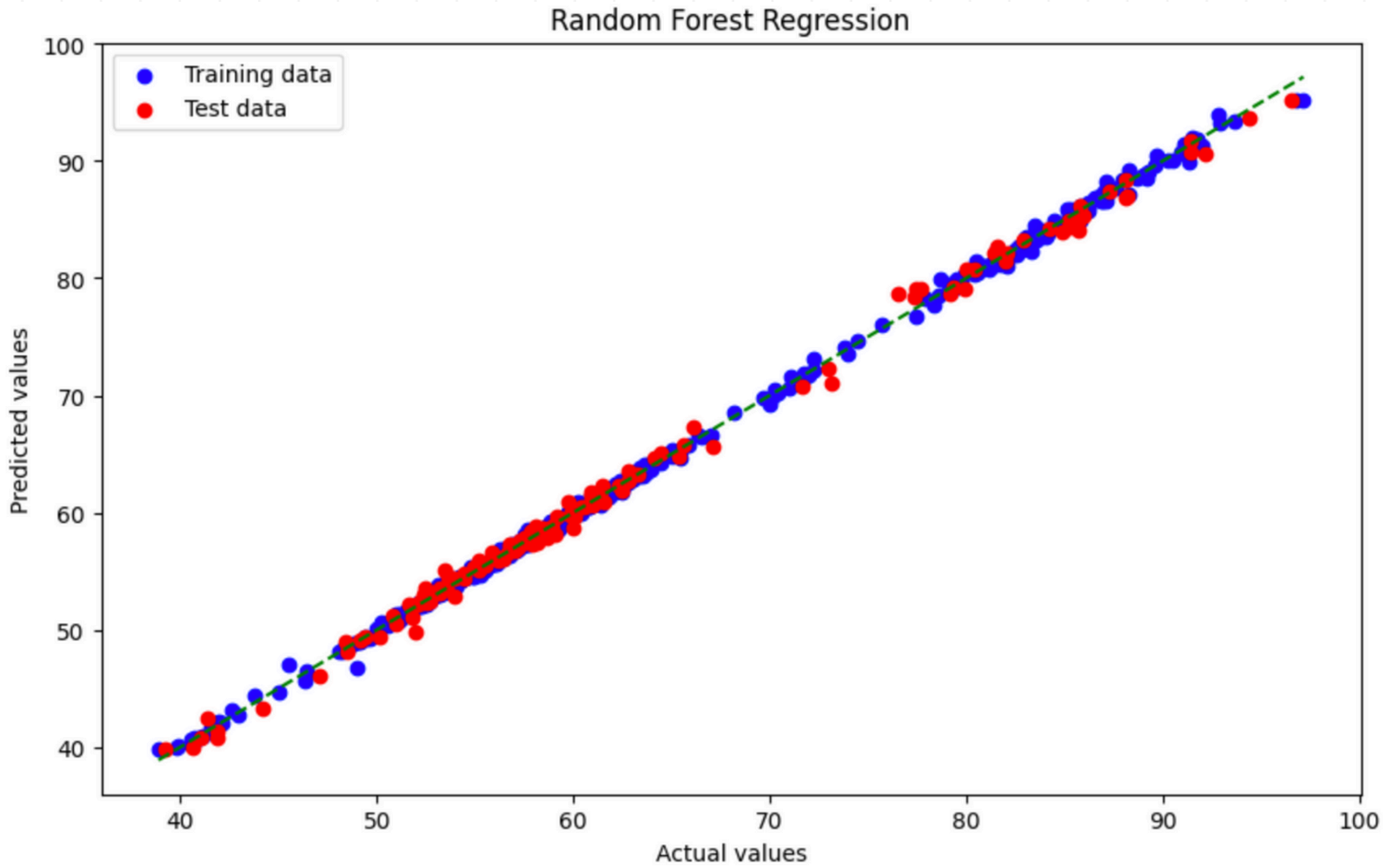


Step3: Model Building

- Split the data into training and testing sets, build a Random Forest Regressor model, and optimize its hyperparameters using GridSearchCV



Step3: Model Building



Step4: Model Evaluation (1)

- Evaluate the model's performance using Mean Squared Error(MSE) and interpret the results

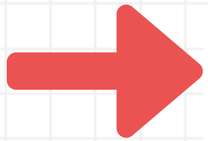
$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

MSE = mean squared error

n = number of data points

Y_i = observed values

$\hat{Y}_i$ = predicted values



```
Mean Squared Error: 0.5471133018891651
```

Step4: Model Evaluation (2)

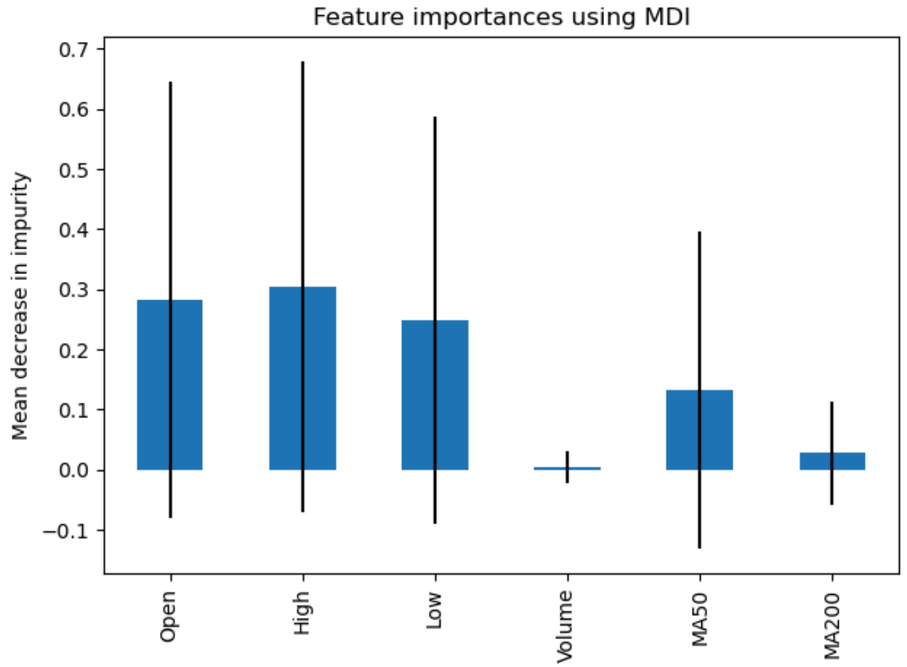
- Evaluate the model's performance using T-test and Wilcoxon Test

```
T-statistic: 0.7887427539651363
P-value: 0.4317439238842424 3
T-test result: There is no significant difference between actual and predicted prices.
```

```
Wilcoxon Statistic: 3821.0
P-value: 0.5587477653582841
Wilcoxon Test result: There is no significant difference between actual and predicted prices.
```

Step4: Model Evaluation (3)

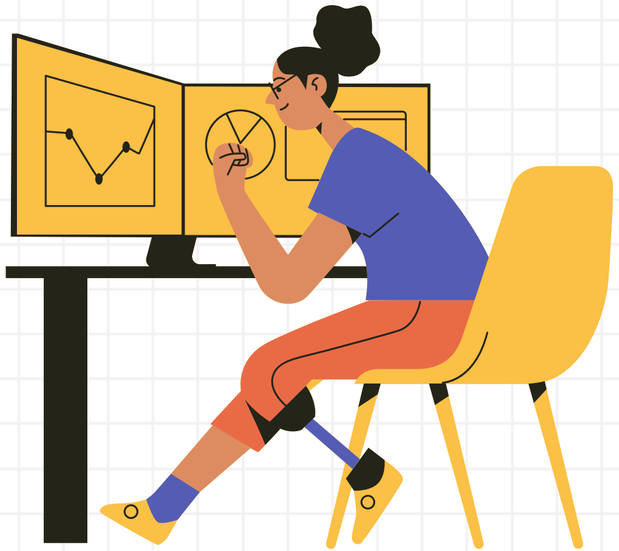
- Feature importances based on the mean decrease in impurity (MDI)
- The blue bars are the feature importances of the forest.
- Inter-tree variability represented by the error bars.



Step5: Prediction

- Demonstrate how the trained model can predict stock prices for the next day based on the latest data

```
Predicted Price for Next Day: 90.01515644155232
```



Conclusion

- Learned how machine learning techniques provide insights into financial markets
- Demonstrated the power of data science in empowering investors
- Validated models using MSE, T-test, Wilcoxon-test, and Feature Importance
- Showcased the practical value of data science in financial markets