### Hybrid ModelS for Dow Jones Index Forecasting and Downward Trend Alert: Data-Driven Time Series Analysis with Risk Management Applications

*Juntian Xiao, Yuhong Yuan, Tianhe Zhao, Guangzheng Dong*

The University of Nottingham

#### Abstract

This study explores the use of advanced machine-learning techniques

**Index Terms—** Machine Learning, Dow Jones, Data Preprocessing, XGboot, Logistic Regression, MLP, LSTM, Prophet, Random Forest.

#### 1. Introduction

B is the most. And through the training of the basic dataset, it prepares for

#### 2. LITERATURE REVIEW

B

#### 3. data analysis and preprocessing

Each patient in this dataset contains 11 clinical features (Age, ER, PgG, HER2, TrippleNegative Status, Chemotherapy Grade, Tumour Proliferation, Histology Type, Lymph node Status, Tumour Stage and Gene) and 107 MRI-based features. Missing values were marked as “999”.

##### 3.1 Missing value handling

The data stability is improved by traversing the dataset to detect the position of 999 and replacing 999 with median filling or removing.

##### 3.2 Data Information

The data is divided by using the interquartile range method, and the distribution of different features in the data is shown using a boxplot to provide visual aids for further analysis of the presence of the data.

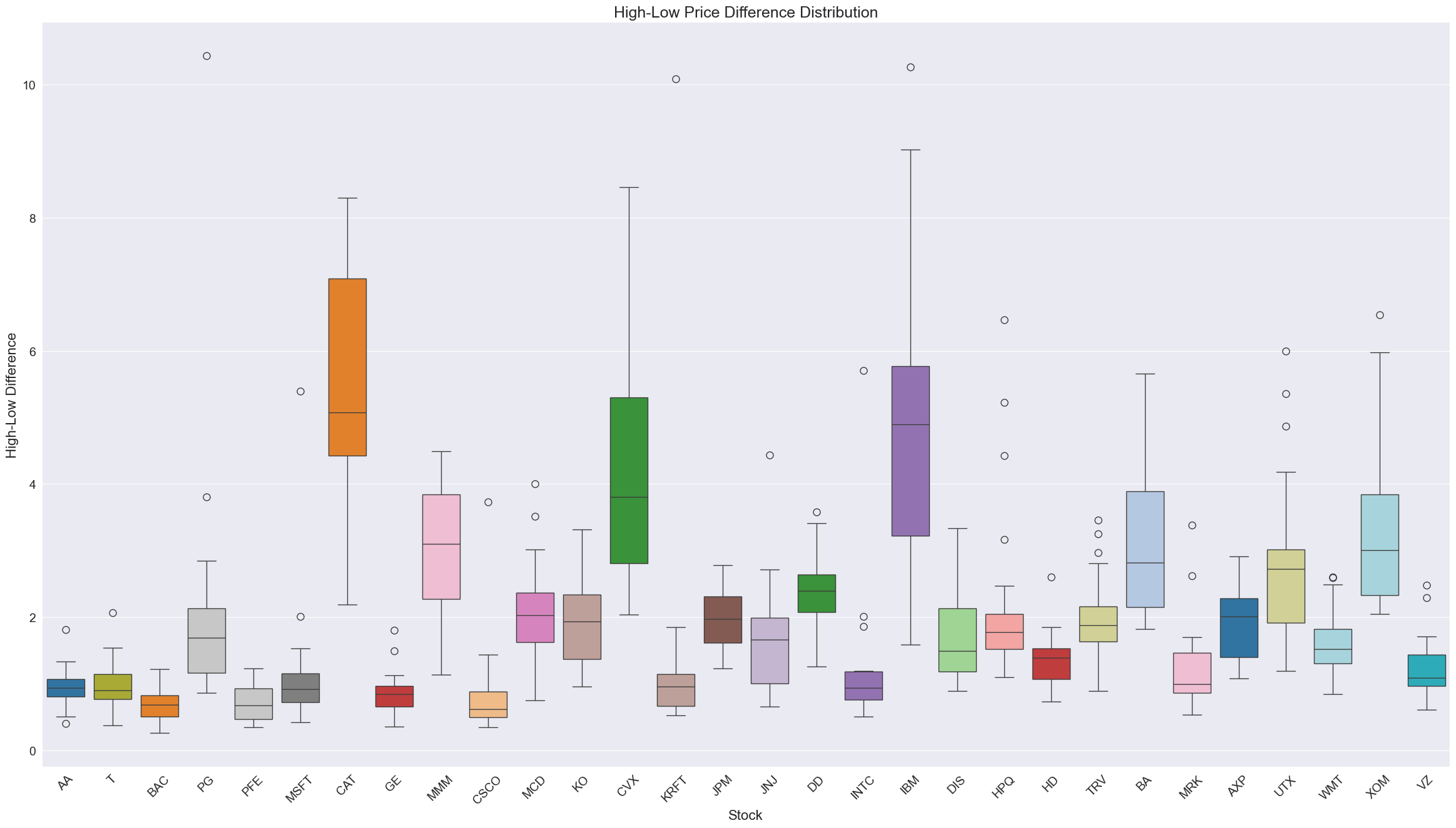


Fig. 1. Boxplot of data features

Use Seaborn to map the thermal mapping of your data to see the correlations between individual data features and use that as a basis to select efficient strategies to process your data.

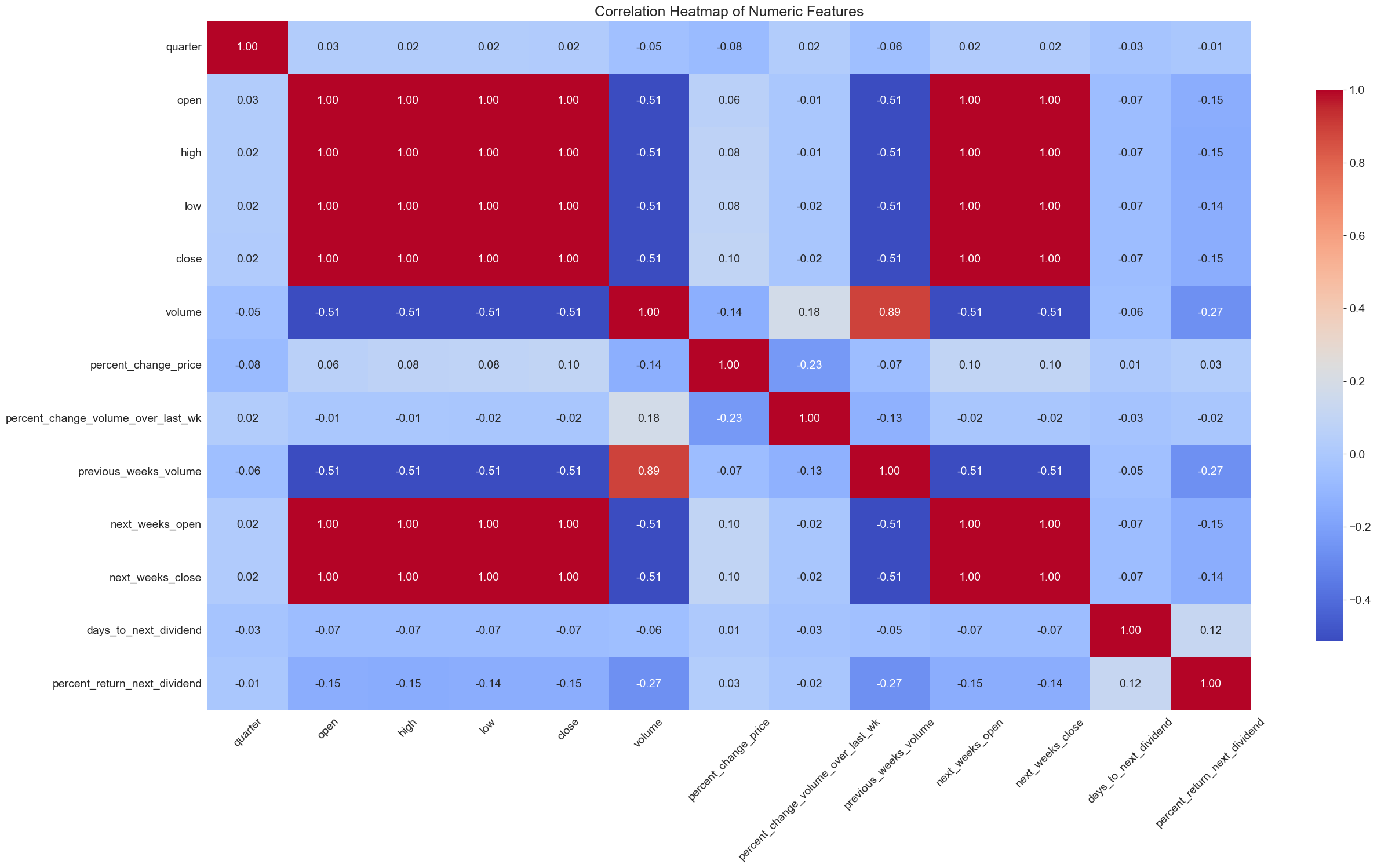


Fig. 2. Correlation heatmap

Finally, the histogram is used to understand the distribution pattern of the data, such as whether it conforms to a normal distribution, to determine whether the data is statistically significant or if there are problems in the data that need further processing.

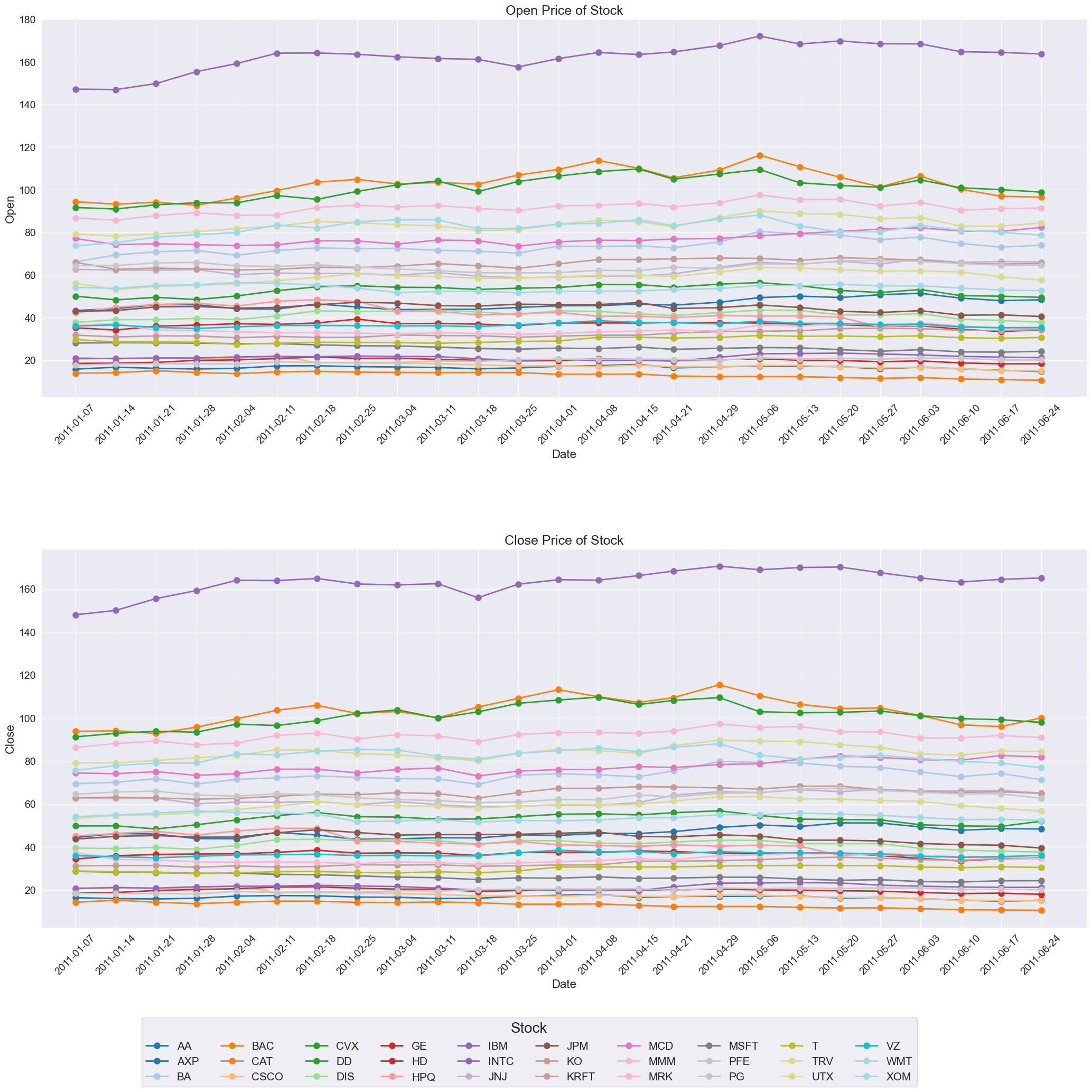


Fig. 3. Open and Close Subplot

Then, through data processing operations

##### 3.3 Data

According to the analysis of ,

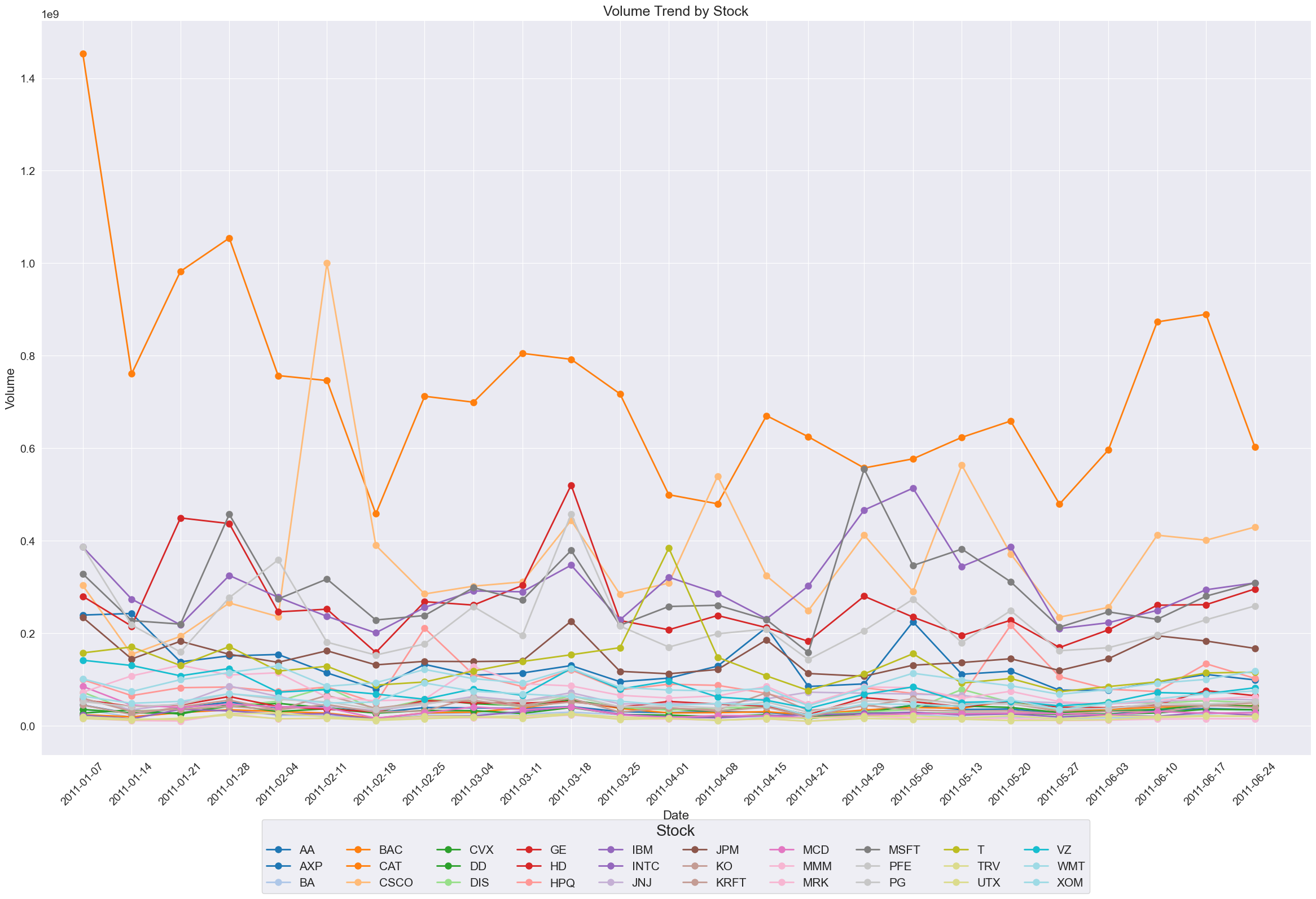


Fig. 4. S

##### 3.4 P

PCA can reduce the

#### 4. MODEL SELECTION AND Comparison

##### 4.1 PCR (classification task)

Logistic Regression:

* Based on its efficiency and ease of interpretation for binary classification tasks.
* Hyperparameters: Regularization strength (C=1.2), solver (liblinear), and maximum iterations (500). Threshold: 0.52.
* Achieved balanced accuracy of 70.26%, precision of 42.31%, and ROC-AUC of 70.26%.

Random Forest:

* Processes high-dimensional data and provides feature importance assessment.
* Hyperparameters: 75 estimators. Threshold: 0.48.
* Achieved balanced accuracy of 68.12%, precision of 41.67%, and ROC-AUC of 68.12%.

A:

* It makes good use of weak classifiers for cascade, and different classification algorithms can be used as weak classifiers, and it also has high accuracy.
* Hyperparameters: SAMME algorithm, learning rate (0.47), and 170 estimators. Threshold: 0.51.
* Achieved balanced accuracy of 75.33%, precision of 44.83%, and ROC-AUC of 75.33%.

Ada

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Precision (Class 0) | Recall (Class 0) | F1-Score (Class 0) |
| Random Forest (75estimators) |  |  |  |
| A |  |  |  |
| L |  |  |  |
|  |  | Recall (Class 1) | F1-Score (Class 1) |
| Random Forest |  |  |  |
| A |  |  |  |
| L |  |  |  |
|  | Accuracy | Balanced Accuracy | ROC-AUC |
| Random Forest |  |  |  |
| A |  |  |  |
| L |  |  |  |

Table. 1. Model accuracy comparison

##### 4.2 R

L

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | M | M | R | R |
| R |  |  |  |  |
| Random Forest |  |  |  |  |
| R |  |  |  |  |

Table. 2. R

#### 5. CONCLUSION

This study explored the use of machine learning techniques to predict

#### 6. References

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Task and Weighting** | **Data pre-processing (10%)** | **Feature**  **Selection**  **(25%)** | **ML method**  **development**  **(25%)** | **Method**  **Evaluation**  **(10%)** | **Report**  **Writing**  **(30%)** |
| **Juntian Xiao** | **0%** | **0%** | **0%** | **0%** | **50%** |
| **Yuhong Yuan** | **0%** | **0%** | **0%** | **0%** | **10%** |
| **Guangzheng Dong** | **0%** | **0%** | **0%** | **0%** | **5%** |
| **Tianhe Zhao** | **0%** | **0%** | **0%** | **0%** | **5%** |