# Discussion

## Performance of Prediction Methods in High Volatility and Extreme Events

### Ensemble forest methods (e.g., random forests, gradient-boosted trees) outperform simpler decision tree methods in handling high volatility, missing data, and outliers.

### Overfitting with imputed data can decrease predictive ability, emphasizing the need for proper model validation and regularization techniques.

### Daily resampling proved to categorically provide better predictions on training and test set compared to hourly suggesting that more significant down sampling (24hrs vs 2 mins) may be a suitable approach for minimising the influence of noise in the context of single step predictions

## Addressing Heteroskedasticity and Extreme Events

### Residual analysis shows that imputation and outlier features did not improve model robustness for heteroskedasticity.

### Consider alternative approaches to address heteroskedasticity in future research.

## Handling Missing Data, Outliers, and Prediction Challenges

### Removing missing values typically yields better performance than imputing missing data.

### Including outliers as a feature does not noticeably improve results, suggesting that alternative outlier handling methods may be more effective.

## Impact of Pre-processing Techniques on Model Performance

### Interpolation can lead to increased error in train and test sets due to overfitting and bias.

### Using unimputed data without outliers results in better generalization to the test set.

# Discussion

## Leveraging Relationships Between Price Series for Prediction

### Predicting a stock using its own history is more accurate than using correlations between two stocks.

### Further research is needed to explore the potential benefits and limitations of cross-series prediction.

## Benchmarks and Limitations

### Most models exhibited larger errors than the benchmark random walk, indicating the need for further investigation into alternative models or feature engineering techniques.

### Consider visualizing model performance compared to the benchmark for better illustration.

## Comparing Data Imputation Techniques

### Artificially synthesizing data (e.g., interpolation) can reduce predictive ability and propagate bias in outlier features.

### Consider evaluating alternative imputation techniques, such as rolling mean.

## Integrating Outlier Detection Techniques for Enhanced Robustness

### Future research should explore other outlier detection methods (e.g., statistical tests, clustering, machine learning-based techniques) for integration into the stock price prediction process.

### Investigate the potential for cross-series prediction and its associated benefits and challenges to improve stock price predictions.

# Further Research and Limitations

## Slide 12: Conclusion and Future Research

### The Random Forest model demonstrated the best performance in terms of RMSE, but it is crucial to consider other factors such as model complexity, interpretability, and generalization ability.

### Model performance should be validated using additional test datasets or cross-validation techniques to ensure robustness and reliability.

### More testing is needed to identify models and features that can account for correlation, volatility, and provide good forecasting ability with robustness to extreme events.

### Investigate the potential of other models, feature engineering techniques, and evaluation metrics in the context of stock price prediction with high volatility and extreme events.