Create a presentation based on this plan. Put together the presentation in dot pint form and organised accordin to slides. Where certain results of facts would be included indicate it with a {insert}:

Presentation

**Introduction**:

This section would introduce the topic of stock price prediction, its importance in the financial industry, and the motivation behind the research. It would also provide an overview of the paper's structure.

* Introduce the challenges in stock price prediction, particularly in the presence of high volatility, missing data, and extreme events
* Explain the significance of addressing these challenges for investors and financial institutions
* State the research questions, emphasizing the need to find effective techniques for handling missing data, outliers, and high volatility

c. Provide an overview of the presentation structure

1. Identify the main research objective: In your case, it is to investigate suitable approaches for stock price prediction in the presence of high volatility, missing data, and outliers.
2. Break down the main objective into sub-objectives: For example, you may want to explore the impact of volatility, missing data, and outliers on prediction accuracy, as well as identify methods to handle these challenges.
3. Formulate specific research questions based on the sub-objectives: These questions should be clear, focused, and answerable through your research methods.
4. How does high volatility in the stock price data impact the accuracy of traditional time series prediction models?
5. What are the most effective methods for handling missing data in the context of stock price prediction?
6. How can outliers be detected and managed in the stock price data to improve prediction accuracy?
7. Which machine learning or statistical models are most robust to high volatility, missing data, and outliers in stock price prediction?
8. Can alternative data sources, such as market sentiment or macroeconomic indicators, improve the prediction accuracy in the presence of high volatility, missing data, and outliers?
9. How do different preprocessing techniques, such as data imputation or outlier removal, affect the performance of prediction models in volatile stock price data?
10. What are the trade-offs between model complexity and prediction accuracy when dealing with high volatility, missing data, and outliers in stock price prediction?

**Background**

~~What are the main challenges in predicting stock prices for the given dataset, considering the presence of high volatility, missing data, and outliers due to extreme events?~~

* ~~Difficulty in accurately estimating the impact of external factors such as news, economic events, and market sentiment on stock prices~~
* ~~High-frequency data may be affected by noise, leading to challenging predictions and potentially impacting model performance~~
* ~~Heteroskedasticity in return distributions complicates modeling and requires appropriate techniques to account for it~~

Can these challenges be addressed using existing prediction methods, or is there a need to develop new methods specifically for this context?

* Ensemble methods, which combine the strengths of multiple prediction models, can potentially improve predictions in challenging situations
* Feature engineering and advanced statistical techniques may help improve the performance of existing models
* The development of domain-specific models, such as those that incorporate news sentiment analysis, can better capture the nuances of stock price prediction under volatile conditions

Which techniques are most effective for handling missing data and outliers in the context of stock price prediction with high volatility and extreme events?

* Advanced imputation techniques, such as matrix factorization or autoencoders, can help preserve the underlying structure of the data when filling in missing values
* Techniques that incorporate the temporal nature of the data, such as time series decomposition or dynamic time warping, can help identify and handle outliers in a time-sensitive manner

How do different prediction methods perform in the presence of high volatility and extreme events, and what factors contribute to their performance?

* Model interpretability and generalizability play a role in their performance, as more interpretable models can help identify the factors contributing to stock price changes and more generalizable models can adapt to new situations
* The choice of features, such as incorporating technical indicators, macroeconomic variables, or sentiment analysis, can impact model performance in the presence of high volatility and extreme events

How do various data imputation techniques, such as linear interpolation and rolling mean, impact the accuracy of stock price predictions when dealing with missing data?

* The choice of imputation technique can lead to different levels of preservation of the original data structure and noise levels, which can subsequently impact model performance
* The success of imputation techniques may depend on the specific characteristics of the data, such as the extent of missing data and the nature of the underlying patterns

**Data Description:**

A concise description of the data used in the study.

This section would describe the dataset used for the research, including the source of the data, the time covered, the frequency of the data (e.g., daily, 2-minute intervals), and any pre-processing steps taken to clean or transform the data.

Describe the dataset used for the research, including the source, time period, and frequency b. Explain any preprocessing steps taken to clean or transform the data

My source of data is unknown, i gave two unnamed price series. They may possibly be futures data. 2. My time period is 5 years 2008 to 2013 in two minute interval. Making predictions despite high volatility and outliers due to extreme events. Primary goal is investigate suitable approach for prediction when there is high volatility missing data and outliers due to extreme events

Series plots with returns and volatiltity

Missing streaks plots

**Research questions:**

1. Main challenges in predicting stock prices for the given dataset:

* High volatility: This may result in prediction models struggling to capture rapid fluctuations in the stock price, leading to decreased accuracy and higher prediction errors.
* Missing data: The presence of missing data may cause the models to be less reliable, as they might not have sufficient information to make accurate predictions.
* Outliers due to extreme events: These events can lead to sudden and unexpected changes in stock prices, making it difficult for prediction models to anticipate and account for them.

1. What are the main challenges in predicting stock prices for the given dataset, considering the presence of high volatility, missing data, and outliers due to extreme events?
2. Can these challenges be addressed using existing prediction methods, or is there a need to develop new methods specifically for this context?
3. Which techniques are most effective for handling missing data and outliers in the context of stock price prediction with high volatility and extreme events?
4. How do different prediction methods perform in the presence of high volatility and extreme events, and what factors contribute to their performance?
5. How do various data imputation techniques, such as linear interpolation and rolling mean, impact the accuracy of stock price predictions when dealing with missing data?
6. In what ways can outlier detection techniques be integrated into the stock price prediction process to enhance the model's robustness against extreme events and improve overall predictive performance?

* What is the impact of different data imputation techniques, such as linear interpolation and rolling mean, on the accuracy of stock price predictions in the presence of missing data?
* How can outlier detection techniques be incorporated into the stock price prediction process to improve the model's robustness to extreme events?
* Can we identify any patterns or features in the data that are particularly relevant for predicting stock prices during periods of high volatility and extreme events?
* How can the relationships between the two unnamed price series be leveraged to improve stock price predictions in the presence of high volatility, missing data, and extreme events?
* Investigate the potential for cross-series prediction, such as using one price series to predict the other or exploiting common trends and patterns between the two series

**Data prepraration :**

Explain any preprocessing steps taken to clean or transform the data

* 1. Prepare four versions of your dataset for each time series:
     1. i. Original data with missing values removed and outliers not identified.
     2. iii. Imputed data with outliers not identified.
     3. ii. Original data with missing values removed and outliers identified.
     4. iv. Imputed data with outliers identified.

1. Data resampling: a. Resample the time series to daily frequency for computational efficiency and to capture potential daily and weekly periodicities.
2. Impact of data imputation techniques on stock price prediction accuracy:

* Data imputation techniques can help improve the accuracy of stock price predictions by filling in missing data points and providing a more complete dataset for the prediction models.
* However, the choice of imputation method can significantly impact the accuracy of predictions, with more sophisticated methods potentially leading to better results.

1. Integrating outlier detection techniques into the stock price prediction process:

* Outlier detection techniques can be used to identify extreme events and remove or adjust them before feeding the data into the prediction models.
* This can help improve the model's robustness against extreme events and reduce the impact of outliers on the overall predictive performance.
* Integrating outlier detection techniques may involve preprocessing the data, modifying the prediction models to account for detected outliers, or using ensemble methods that combine the predictions of multiple models, some of which may be more robust to outliers.

**Methodology**

Present the specific approach or approaches being investigated b. Explain the rationale behind the chosen approach and any assumptions made c. Briefly discuss the model development process

**Model Development and Validation .**

a. Explain how the data was split into training and testing sets

b. Describe the process of model selection and tuning

c. Introduce the evaluation metrics used to assess the performance of the models This section would describe the process of developing and validating the prediction models. It would include details on how the data was split into training and testing sets, the process of model selection and tuning, and the evaluation metrics used to assess the performance of the models (e.g., mean squared error, mean absolute error).

a. Evaluate the performance of the models using appropriate metrics like Mean Squared Error (MSE), Mean Absolute Error (MAE), or R-squared.

b. Compare the performance of different models on different dataset versions (original, imputed, with/without outlier identification) for each hypothesis.

c. Investigate the impact of data imputation techniques, such as linear interpolation and rolling mean, on the accuracy of stock price predictions.

d. Assess the effectiveness of different outlier detection techniques in improving the model's robustness to extreme events.

e. Analyze how different prediction methods perform in the presence of high volatility and extreme events.

1. Establish benchmark model: a. Perform a random walk simulation on the original data with missing values as a benchmark for comparison.
2. Experiment with different hypotheses:
   1. a. H1: ts1 can predict ts1
   2. b. H2: ts2 can predict ts2
   3. c. H3: ts1 can predict ts2
   4. d. H4: ts2 can predict ts1

we use k-fold cross-validation for model tuning, that is, finding the optimal

hyperparameter values that yields a satisfying generalization performance.

Once we have found satisfactory hyperparameter values, we can retrain the model

on the complete training set and obtain a final performance estimate using the

independent test set. The rationale behind fitting a model to the whole training

dataset after k-fold cross-validation is that providing more training samples to a

learning algorithm usually results in a more accurate and robust model.

**Results**

a. Present the results of the prediction models, including their performance on the test dataset

b. Compare the performance of the proposed approach with existing methods

c. Highlight any insights gained from the analysis

a. Summarize the findings, highlighting the most effective techniques for handling missing data, outliers, and the challenges of stock price prediction in the presence of high volatility and extreme events.

b. Provide insights into the relationships between the two time series and their predictability.

c. Discuss the limitations of the study and potential avenues for future research.

This section would present the results of the prediction models, including their performance on the test dataset and any insights gained from the analysis. It would also compare the performance of the proposed approach with existing methods discussed in the literature review.

**Discussion**

a. Discuss the implications of the results, including any limitations of the study b. Suggest potential areas for future research c. Provide recommendations for practitioners, such as how the findings could be used to improve trading strategies or risk management

section would discuss the implications of the results, including any limitations of the study and potential areas for future research. It would also provide recommendations for practitioners, such as how the findings could be used to improve trading strategies or risk management.

1. Addressing challenges using existing prediction methods or developing new methods:

* Existing prediction methods, such as ARIMA, GARCH, and LSTM, can be adapted to address these challenges by incorporating additional features or modifying model parameters.
* However, it may be necessary to develop new methods specifically designed for this context, especially when dealing with high-frequency data, extreme events, and non-normal return distributions.

1. Techniques for handling missing data and outliers:

* Data imputation techniques like linear interpolation, rolling mean, or more advanced methods such as K-nearest neighbors imputation can be used to fill in missing data points.
* Outliers can be handled by using robust prediction models, such as robust regression, or incorporating outlier detection methods, like Z-score or IQR-based techniques, into the prediction process.

1. Performance of different prediction methods in the presence of high volatility and extreme events:

* Factors contributing to the performance of prediction methods include the model's ability to capture non-linear relationships, adapt to changing market conditions, and account for heteroskedasticity.
* Machine learning models, such as LSTMs or recurrent neural networks (RNNs), might perform better due to their ability to capture complex patterns and relationships in the data.

1. What role does the granularity of the data play in the effectiveness of stock price prediction methods and the applicability of one-step forecasting?

* Critically examine the impact of data granularity on the performance of various prediction methods and their ability to capture short-term and long-term market dynamics
* Investigate the potential limitations of one-step forecasting when applied to high-frequency or low-frequency data, considering the influence of market microstructure and macroeconomic factors
* Assess the role of data aggregation and resampling techniques in mitigating the challenges posed by different levels of data granularity and improving the performance of stock price prediction methods

Future:

* Investigate whether ensemble methods, combining multiple models with different outlier handling techniques, can improve prediction performance
* Examine the relationship between outlier detection sensitivity and model robustness, considering the risk of overfitting or underfitting due to outlier handling
* Dynamic outlier detection methods, which adapt to the evolving nature of stock prices, can help identify and handle extreme events more effectively
* Techniques that consider the underlying structure of the data, such as subspace or manifold learning, can help improve the model's robustness by identifying outliers in a more context-specific manner
* Using outlier detection techniques in conjunction with other preprocessing steps, such as feature scaling or transformation, can further enhance the model's ability to handle extreme events and improve overall performance

**Conclusion**

a. Summarize the main findings of the research b. Reiterate the contributions of your research to the field of stock price prediction This section would summarize the main findings of the paper and reiterate its contributions to the field of stock price prediction.