**Meeting Summary for Yousaf Abdul Khaliq's Personal Meeting Room**

Mar 31, 2025 08:09 PM Central Time (US and Canada) ID: 644 948 8621

Top of Form

Quick recap

The team discussed the challenges of replicating a study due to the large dataset and computational power required, and the importance of reproducing the study as a team of researchers to further the research. They also discussed their approach to replicating a paper's results using provided Python scripts from GitHub, with each team member taking responsibility for one or two scripts. The team agreed to document every change they need to make and focus on the ethical implications, and decided to work on different variations of the Lstm.

Next steps

Yousaf to set up and share a shared document for the team to collaborate on.

All team members to document changes made to the original code and add comments explaining the modifications.

All team members to work on their assigned scripts:

Seth to work on intraday Random Forest

John to work on next day Random Forest

Yousaf to work on next day LSTM

Clay to work on intraday LSTM

All team members to document challenges encountered while reproducing the paper's results.

All team members to consider and discuss the ethical implications of the project.

Clay to share his code with John after completing his attempt to replicate the paper step-by-step.

All team members to upload their modified code to GitHub with appropriate documentation and comments.

Summary

Reproducing Study With Large Dataset

The team discussed the challenges of replicating a study due to the large dataset and computational power required. They considered sampling a smaller subset of companies for testing purposes, but agreed that reproducing the study with all 500 companies was necessary for accuracy. The team also discussed the importance of reproducing the study as a team of researchers to further the research.

Data Accuracy and Feature Selection

The team discussed the accuracy of their sample data, with Seth suggesting a cherry-picked sample for quick results. However, John raised concerns about the potential for skewness in the results. Clay proposed a new approach, setting all null values equal to 0, but faced challenges with feature selection due to missing data. Yousaf suggested searching for more information on the concept of "study periods" mentioned in the paper.

Reproducing Paper Results With GitHub Scripts

The group discusses their approach to replicating a paper's results using provided Python scripts from GitHub. Clay has been trying to recreate the code from scratch but is encountering issues with feature generation for LSTM. The others (Yousaf, Seth, and John) are using the GitHub scripts directly, running them in environments like Google Colab or Jupyter hub, and addressing any errors or deprecation issues they encounter. They divide the six available scripts among the four team members, with each taking responsibility for one or two scripts. The team agrees that this approach gives them the best chance of reproducing the paper's results and provides good discussion points for their project.

Reproducing Paper Results: Approaches and Significance

In the meeting, Seth, Clay's, and John discussed their approaches to replicating a paper's results. Seth emphasized the importance of reproducing the code used to generate the paper and addressing any discrepancies. Clay's shared his experience of encountering issues when trying to replicate the paper manually. They also discussed the significance of reproducibility in their work, with Seth sharing his experience of testing machine learning algorithms for insurance companies. The team agreed that their approaches were valid, with Seth suggesting that Clay's method was the right way to go.

Project Next Steps and Code Uploads

In the meeting, Yousaf, Seth, Clay's, and John discussed the next steps for their project. They agreed to document every change they need to make and focus on the ethical implications. Seth raised a question about how to handle code uploads to their Github, and Yousaf suggested making comments on the areas they changed and adding a user guide at the top. Yousaf also mentioned he would set up a shared document and send an invite. The team decided to work on different variations of the Lstm, with Seth doing a random forest, Yousaf doing a regular Lstm, and John and Clay's working on other scripts. They agreed to work on one script each and clarified the tasks for each team member.

Bottom of Form