2024 /9 /30

21:00 - 23:00

11210CS460200 Group 23 Meeting Minutes

Topic	Discussion on Machine Learning Final Project Ideas
Place	Discord Voice Chat
Agenda	1. Decide project topic and implementation details 2. Assign a role(task) for each member 3. Create a project timeline
In attendance	All present
Task Assigned	游松澤, 曾柏勳: collecting datasets 楊立慈, 賴允中: data visualization 游松澤: researching related works 蕭以勝: designing model architecture
Next meeting	Date: 10/7 Time: 9 p.m Objective: the predictive to be used in the project Location: Discord Voice Chat

Final Project Selection Summary:

Baseball (MLB) Pitch Type Prediction Using Format Attention

After comprehensive discussions, we have provisionally chosen "Pitch Type Prediction Using Format Attention" as our final project topic. This decision was based on the availability of a **robust dataset from the MLB Statcast system**, which provides extensive details including each pitch's release point, type, and trajectory.

What's next? For the first week, our primary focus will be to explore the Statcast data and any accessible APIs that might be beneficial for our project. We plan to refine the scope of our data analysis by determining **(1) which pitchers' data to include** and **(2) identifying the advanced metrics** that will be most effective for our training models.

These initial steps will set the foundation for our project, ensuring we utilize the most relevant and comprehensive data available for predicting pitch types effectively.

Complete meeting records:

The meeting initiated with an open brainstorming session, exploring two main directions for our machine learning final project: Baseball Analytics and Quantitative Finance.

I. Baseball Analytics

The discussion on baseball was subdivided into two specific topics within Major League Baseball (MLB):

1) **Pitch Type Prediction Using Format Attention**

   Utilize supervised learning to predict the next pitch type based on historical pitch data.

   Employ player's release angle and other relevant metrics to predict the type of pitch. The correct pitch type will serve as the target variable y in our supervised learning model.

2) **MLB Player Value Prediction Based on Performance Metrics**

   Predict a player's value using comprehensive performance data, including advanced defensive and offensive metrics.

3) **CPBL pitching data collection using raw video of game**

   Since CPBL doesn't collect as much data as MLB, we still want to make a dataset by image recognition for our own country baseball for more advanced data.

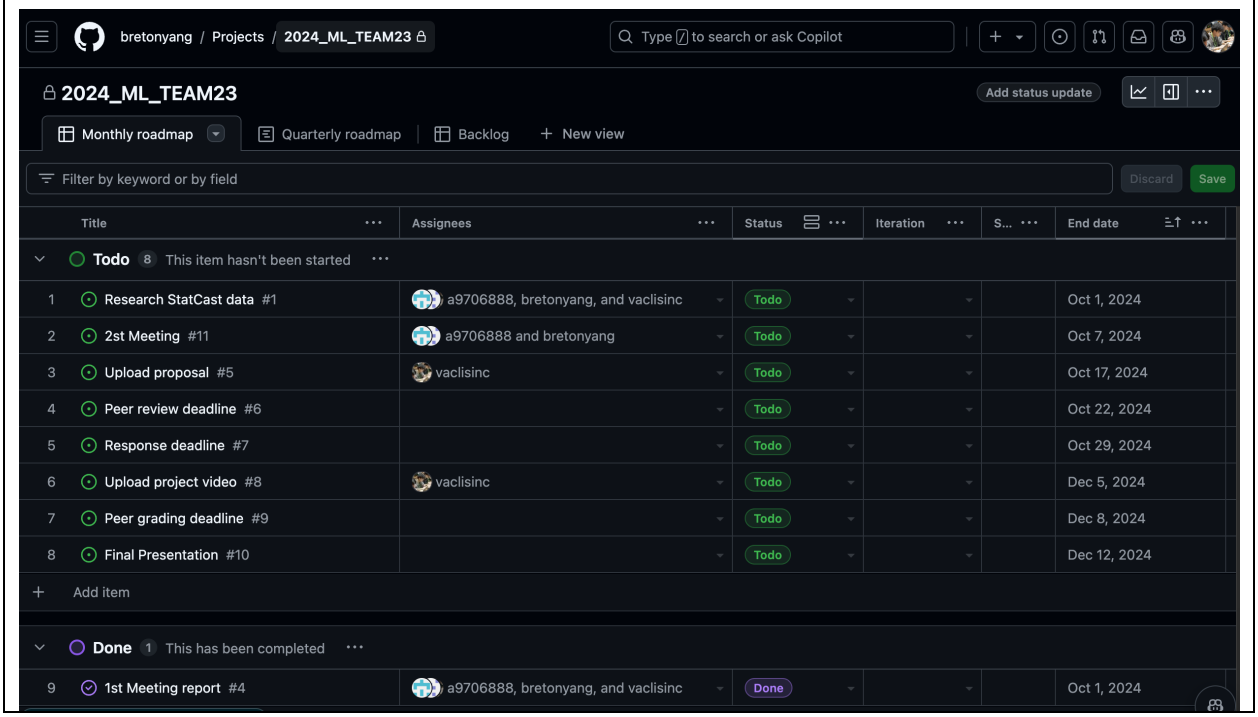
II. Quantitative Finance

Utilize two decades of financial data provided by WorldQuant to predict and identify alpha while avoiding overfitting.

Finally, we all agreed to choose “Pitch type prediction using format attention” as our final project topic since it is interesting and have comprehensive datasets.
We didn’t choose to the quantitative finance topic because most of us are not familiar with finance. Not only the reason above, we predicted that there will be too many teams do something like this.
We didn’t choose to do the CPBL raw game video pitching data collection due to the raw dataset are too complicated and really not an easy work to collect them.

Timeline planning records:

We created a repository in GitHub and using the project function to append the todo list and deadline on it.



A group photo of the discussion session: