

2024 /10 /08

21:00 - 23:00

11210CS460200 Group 23 Meeting Minutes

Topic	Model and Data Feature Choosing for our final project
Place	Discord Voice Chat
Agenda	<ol style="list-style-type: none">1. Decide model we use2. Discuss what features we use3. Discuss the target pitcher we want to predict
In attendance	All present
Task Assigned	<ul style="list-style-type: none">- Watch Hung-Yi-Lee's lecture about RNN, which is the model we will use in the final project.- Data exploration and feature selection.
Next meeting	<p><u>Date:</u> 10/14</p> <p><u>Time:</u> 8 p.m</p> <p><u>Objective:</u> Record proposal video and discuss data preprocessing</p> <p><u>Location:</u> Discord Voice Chat</p>
<p>Meeting Summary:</p> <p><u>Which model we decided</u></p> <p>RNN is designed specifically to handle sequential data, where the order and context of data points are crucial. RNN has the ability to maintain information across different time steps, which allows them to model dependencies between elements in a sequence. Because pitchers will choose pitch type based on the previous pitch, so we choose Recurrent Neural Network as the model of pitch prediction.</p>	

Which pitcher to use as our prediction target

With so many pitchers in the league, each throwing a different mix of pitches, we figured it would be really tough to start by designing an ML model that works for all pitchers and pitch types. Take Lance Lynn, for example—he's famous for leaning heavily on his fastball, throwing it about 97% of the time. While that's interesting, it doesn't give us a good range to see how our model works when predicting pitch types. On the other hand, Yu Darvish, with his crazy variety of more than 8 different pitch types, makes the categorization process tricky and probably a bit too complicated for our first attempt at a model.

So, we decided to look for a pitcher who has a lot of data but sticks to a reasonable number of pitch types (around 4-5). After some thought, we landed on Max Scherzer, a well-known pitcher, as the perfect target for our first naive implementation.

A group photo of the discussion session:

