

Introduction:	15%	<p>14/15</p> <p>Overall, the intro is nicely written. Two points: (1) Throughout the paper, there are two kinds of wording about the authors' work: "our methods" and "our metric". What are they? Are they the same thing? If not, what is the relationship between them? They need to be explicitly and clearly defined. (2) The objective for prediction was described so long: "predicting baseball pitcher performance that can quantify the likelihood that a player who has done well in the Minor Leagues will continue to do well in the Major Leagues". Readers can hardly know the concrete thing/value to predict in the data. More examples will be helpful, like describing Netflix mechanism – users can give ratings to films, from 1 star to 5 stars, and the goal was to predict the unknown rating give a user and a film. We would have a sense about where the value comes from and what the range of the value is.</p>
Related Work:	10%	<p>9/10</p> <p>The related work section was supposed to broadly link to related topics. Instead of focusing on "intersection" (those that both did baseball data and used data science), it will be great to discuss the related topics broadly and separately: (1) Studies on baseball player data; (2) Data science methods for sports analytics. Through a broader survey, we will see the uniqueness of our work. Specifically for "data science in baseball data", if there has been a good number of work, we can survey on (1) the popular/effective features that have been used in existing work and (2) the popular/effective data science models that have been used. Then we will see the motivation of our model design.</p>
Solution/Method:	25%	<p>25/25</p> <p>This section was clearly written. I enjoyed reading the pipeline through the steps in the diagram of data science research.</p>
Data and Experiments:	10%	<p>8/10</p> <p>The final, cleaned, integrated, selected dataset for study needs a clear description. At the end of the paragraph of Data Cleaning, I can see eventually the authors use a dataset of 219 players and 13 attributes per pitch. However, what is the data object? Is it pitch? How many data objects/pitches? What are the attributes? Are they all "raw" attributes? Did you add some new attributes based on your domain knowledge (for example, in basketball, given raw attributes "pts", "rebounds", and "assists", we can add attributes like the number of triple doubles or the number of 20+10+5)? What is the label? Does it directly come from the dataset or need some calculation to generate?</p> <p>In the evaluation, it will be good to discuss the size of training and testing sets (how many objects). 3-fold cross validation is not 75%-25%. 4-fold cross validation is.</p> <p>What is <math>R^2</math> measure? Is it the coefficient of determination in regression model? What is the range of this measure? Is it the bigger the better?</p>
Evaluation and Results	25%	<p>23/25</p> <p>I have a few questions on Table 1: (1) Why <math>R^2</math> can be 1.18 or 28.29... bigger than 1? What does that mean? Why the smallest, 0.99, is the best? (2) Why was linear regression so bad? Why it was <math>10^{20}</math> not <math>10^3</math> or <math>10^5</math>? Without knowing the range of the label, I cannot have a good understanding of the physical meaning of the error.</p>
Writing Quality:	15%	<p>14/15</p> <p>Generally it was nicely written. A few points: (1) Use past tense to write related work section; (2) typos: "there work" ("their"), "what we plan to do is take..." (add "to"); (3) acknowledge Python packages/libraries if used for any model (AdaBoost, NN, SVM, etc.) – add references.</p>

