

Project 3 Reflection

Nathan Young

How accurate was the time assessment?

After the first hour of work, I hoped to have a cleaned dataset, and have key features identified and plotted. There are at most 40 features that can be used, but a number of those are not helpful to predicting the target. I was able to start visualizing data, but when I attempted a `pairplot()`, I ran into an issue with the data, letting me know that I had not completed the processing stage successfully. More time would need to be spent on preprocessing and digging deeper to identify columns that are not helpful, like `Orbiting Body`. All the asteroids are orbiting Earth. That doesn't tell me if an asteroid is hazardous or not.

Could you hand this off?

There is not much information, other than single line comments in each cell briefly describing what is happening in that cell. I tend to work and then describe what I did later, which could be problematic if I was to leave this project for someone else. My thought process cannot be found, which could lead to the project getting completely redone.

3 Things Learned

1. I learned that adding commentary as you work is important. More than a single line comment, and probably in markdown, explaining what is going to happen and why. My thought process needs to be on the page, no matter what stage I am in. The first part of machine learning is to identify a question to answer and then plan how to answer that question. I did both, but without the plan, you would not know the goal I was trying to reach.
2. Machine learning projects take time, and that's okay. As nice as it would be to snag a data set and a few hours later, or at the end of a workday, have a model that makes sense of the data, that will likely not be the case. Machine learning requires an agile plan that changes as new insights are extrapolated.
3. Data processing is an annoying but necessary process. From what I have heard, this (data processing) takes up the majority of the work data scientists do and often this results in burn out, sometimes even resulting in people changing careers. The quick feedback is what draws many people to technology, but it can be a downfall too, if it is not as quick as you wanted or if you cannot resolve an issue. This is why having a plan and time is important. My planning will get more accurate as I learn how to better extrapolate information, but for now, time lines should be generous to account for extra time necessary to gain proper insights.

Is this project useful?

I think this project is useful. If done properly, it puts you in a career-like situation, allowing you to experience a simulation of what a day in data science could be like. That is beneficial because you can identify areas you need to work on (for me, writing a plan) and thus be prepared if an opportunity like this presents itself in the future.

I think this project should be included in future courses. It might even be beneficial if this project effected the final project. I see this like, the plan you developed and explored for an hour, you now get 7 hours to work on the final, following your plan to see how well your plan actually worked. The result could be a reflection, like this one, reviewing the plan created and discussing if it worked or not.

I personally plan on using this data source for Project 4 and will be using knowledge I gained with this project in that project.