CAPSTONE PROJECT PROPOSAL 2

What is the problem you want to solve?

Predicting what should be paid out to medical practitioners based on what was paid out to them in the past.

Who are your clients and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn't have otherwise?

My clients are anyone healthy or unhealthy need to see a medical practitioner for either a routine yearly checkup or for those visiting an emergency room (ER) unit of a hospital, to know how much one needs to pay out of pocket that is not being covered by one's medical insurance deductible.

What data are you going to use for this? How will you acquire this data?

I would be using the imbalanced dataset from Centers for Medicare and Medicaid Services and the US Food and Drug Administration.

In brief, outline your approach to solving this problem (knowing that this might change later)?

- Run a hypothesis test on the dataset to check what factors are most significant in explaining variability in the amount being paid out to medical doctors at a statistically significant level of 0.05 or if such are barely by chance

- Comparing different algorithm in Python before settling on a suitable one after preprocessing the data. And, fine-tuning the model before evaluating it on an out of sample data to check if it generalizes well.

What are your deliverables? Typically, this would include code, along with a paper and a slide deck.

Deliverables will be code and paper explaining my analysis in details.