David Dempsey

Final Exam

Dec 17, 2023

BA-64060

I am unsure of how to answer these questions as I was unable to perform an analysis of this data. I was able to clean the data properly and even filter out nonnumeric data. I will discuss more about my hardships attempting to complete this project in this document. I think its harder on an open ended final as this to find proper results when using something like machine learning which is already so difficult to work with. I was hoping to analyze the data by using the clustering method but for some reason I cannot use that on this data. I tried to filter out all non-numeric data to see if it was causing corruption in my code, but I could still not run it. I have spent 14 hours on this project and I am disappointed that I was unable to discover interesting forms in this data. I think working with such a huge dataset is what hurt me the most. The load time when running my script was a major hinderance and I am upset deeply that I could not solve this tricky problem. I also tried Knn method to solve this problem and was still met with error after error. Perhaps when you grade my assignment, you can tell me what I did wrong as I am desperate to have a firm grasp on machine learning. I am aware you will give me poor marks for not being able to meaningfully analyze this data but I do hope you understand the time and effort I poured into this project. I will willing accept any and all advice you can give me.

Some questions I would have liked to answer:

Are there any abnormal or anomalous flow conditions?

How stable are the identified clusters over time?

How well does clustering capture known operational states (normal, start-up, shutdown)?

Can clustering be used for predictive maintenance?

I cannot answer these questions one way or another with my missing data. I would have hoped to use my clustering method to answer these questions or at least find more meaning in there.