Lachlan Sinclair

OSU CS 419

Assignment 1 Reflections

A)

IBM SPSS requires having all the data in memory before processing it, this means the size of the dataset that can be processed is limited by the machine's memory. So, from a practical perspective this application doesn't scale well making it unsuited for arbitrarily large datasets. More modern applications have their speed bound to the machine size not their ability to function.

B)

Benefits:

I think the main advantages of using this tool is its accessibility and rate of development both of which are useful in certain real data science projects.

The UI is intuitive and provides the user with a lot a prebuilt features they can easily use in their project. The work of developing all these features is taken care of and after adding them to a model the application guides you through the remaining configurations steps. Once a user gets familiar with this environment, they could produce a proof of concept very quickly. Also, there is no usage fee, users can spend as much time modeling as they like without fear of running up a large bill.

Challenges:

Once a project gets significantly complex the visual format would become problematic. Having to many lines drawn all over the place would get confusing very fast. This is notable since it doesn't provide project support like other applications.

The size of the data that can be processed using this application is also very limited which makes it impractical for many real data science projects.

Like many other similar applications users are limited to the features provided by the application which in turn limits its uses in real scenarios.

The lectures mentioned that it cannot produce a deployment solution in a straight forward manner, this can make applying it to a real-world project difficult.