

Project Postmortem

BAH Machine Learning

Team 3 – Analytics Avengers:

Ahmed Sodeinde

Erick Orellana Morales

Rami Tello

What went well...

- Collaboration between the team members
- Trying multiple models (traditional and more modern techniques)
- Building and tuning models manually as well as using automated tools to verify results
- Applying techniques for feature selection and hyperparameter tuning
- Applying the lessons from the Technical Excellence course to the project
- Learning how to have difficult conversations when there isn't sufficient signal in a dataset to build a model
- Working as a team to get through issues that came up but asking for help as needed
- Staying within the scope of the project

Areas for improvement...

- Using more than one platform (we were set on SageMaker)
- Verifying the lineage and validity of a dataset early on
- Saving reference materials in a location where it can be retrieved quickly when needed

Lessons Learned...

- SageMaker comes with restrictions:
 - Size limits on batch processing
 - Optimized to run on SageMaker models and does not always handle others well
 - Difficult to package/register anything that was not made with a SageMaker library
- Use better naming convention for endpoints on SageMaker to make it easier to find them when looking back at a list of endpoints
- The best point at which to balance data when dealing with unbalanced datasets
- Importance of data lineage to validate results
- Using the official documentation for libraries and tools when working through issues or building new code
- Building models is building experiments... a bad model does not necessarily mean failure; it could mean there isn't enough signal in the data to make the type of inferences needed
- The importance of feature selection
- Having a subject matter expert can help especially during the data exploration stage and possibly when verifying the results of the model