Self-Assessment

Over the course of completing the culminating project, multiple roles were undertaken. The first was establishing possible project scopes and finding the necessary data. In exploring potential data sources, it was essential to go through an exploration to evaluate their relevancy and adequacy to the project. The second role involved assistance in building the Machine Learning (ML) model. This involved starting a decision tree, regression, and XGboost model. Additionally, during regression analysis, I implemented a Random Search CV analysis to find optimal hyper-parameters for fine-tuning. Although these models were not the final model chosen, they all assisted in the pathway toward the group's last model. Furthermore, during the ML model process, I determined feature importance and correlational values that were implemented into the model and later used for dashboard visuals. For the initial dataset, I utilized Tableau to create the visualizations for feature engineering. Lastly, I made the presentation and its content in google slides for deliverables 2,3 and 4. During the 4th deliverable, the group determined that some slides were to be removed due to time restrictions.

The greatest personal challenge manifested during the ML model phase, where I struggled with frustration and technical issues. Before the new dataset was integrated, it seemed that it didn't matter which type of model we chose, as the error rates were always sub-optimal. During this time, I found ways to optimize the model, hyper-tuning, and implement strategies to decrease my frustration. This involved taking short breaks and placing value on the overall learning process. Talking to my group also helped as they ran into the same issues.

Team Assessment

Overall, the group's communication was adequate. We frequently communicated through slack and had additional meetings outside of class hours. However, I do wish that we were more organized with our roles. We often stepped into different positions that confused us, and some individuals put in more work than others. I believe implementing a short Gantt chart may help combat these issues. That way, those small weekly deliverables leading to the main deliverable are checked off, and it is clear to see which person is responsible for what. Additionally, it's a path to see which deliverables are on schedule, behind, or ahead of the project timeline. One of the strengths of our team was everyone's ability to step up to the plate and help one another when needed. No one shied away from additional work and wanted to contribute to the team.

For future cohorts, I recommend first looking at the project in a top-down approach. After developing a general idea of each project phase, then you can transfer to a bottom-up approach. Our group immediately went into the project with a detailed, oriented approach that produced stagnation during the early phases of the project.

Summary of the Project

Using Covid-19 as a reference, develop a model to prepare for future pandemic outbreaks by directing resources to poor-performing counties based on medical, political, and socio-economic factors. Multiple ML models were developed; however, the final model was a Random Forest Classification.

Results:

|  |  |  |
| --- | --- | --- |
| Avg/Total | **Cases** | **Deaths** |
| Precision | 0.99 | 0.98 |
| Recall | 0.99 | 0.95 |
| Specificity | 1.00 | 1.00 |
| F1 Score | 0.99 | 0.96 |
| Accuracy | 0.99 | 0.94 |
| Matthews Correlation Coefficient | 0.80 | 0.64 |