Alexandra Treml

Evolve DS Assessment Project Brief

12-19-2021

1. If this initial work formed the basis of a longer-term project, what additional steps would you take to increase performance and create a production-ready model?
    a. I would spend significantly more time understanding where the data came from, identifying trends over variables to determine what features may be more important, try other classification models such as XG Boost. While I scaled the model, I would also look into ways that the data could be scaled or represented more accurately.
2. Assume an existing churn prediction model is currently deployed. How would you determine when your model is ready to replace the existing model in production?
    a. I would create a baseline model with my data to test the existing churn prediction model and my new churn prediction model against each other. Then I would use the same test and training data to test each one again with the existing model and the one I've chosen. It would also be important to dig into the EDA to determine why columns may have been dropped, or what other considerations and assumptions may have been taken into consideration that differ from my model. When I believe that my model has taken the best from both, the assumptions are sound, and I'm getting an output that is reflecting a higher accuracy score, or a better representation of the data (if the previous model has been overfitted), I will replace the old one.
3. After putting your model in production, what potential issues could arise a few months down the road and how would you address them?
    a. It's common that extraneous data or outliers that were not considered or tested in the production of the model could arise, especially as a company, and thus the data, scales. There will likely be additional columns that affect the target variable differently, columns that don't make sense and need to be addressed with the data engineering team or clients, or simply there could be a lack of sample size for specific algorithms. It's also important to make sure that new variables do not contribute to overfitting, like time. It could also happen that the hypothesis or question that the model was originally made to help answer has changed, and thus the problem statement and statement of work needs to be revisited.
4. Any other limitations or considerations that you'd like to share.
    a. I appreciated how straight forward the dataset seemed. As I'm always growing my knowledge and toolset arsenal, I'm looking forward to feedback and other's ideas.