**Discussion questions**

Error Analysis

1. What are the key takeaways you’ve learned from using Error Analysis?

Answers:

* Gain a deep understanding of how model failures are distributed across a dataset.
* Discover erroneous cohorts to investigate.
* Traditional performance metric do not show where there are model failures in the data distribution

Model Overview

1. What were lessons on model reliability from disparities between Accuracy, Precision and Recall?

Answers:

* Having high a wholistic performance understanding of a model is important. Only relying on one model performance metric such as Accuracy score can be problematic.
* The Accuracy score for the cohorts was very high. However, as we started evaluating whether the predictions the model made were correct, we saw a decline in the precision score. Finally, validating if the model got the correct prediction for a given class, we saw a huge drop in Recall, which is alarming. Meaning the model predictions were not reliable.

1. What is an example of finding a Fairness issue in the performance metrics?

Answer:

If cohorts containing sensitive features are performing worse compared to cohorts with non-sensitive features. This will be a red flag for potential fairness issues. Meaning the model’s performance is favorable towards the non-sensitives features.

1. What issues did you observe from the confusion matrix? What hypothesis can you form based on the chart?

Answers:

* Data imbalance: The are more cases for Not Readmitted patients than Readmitted patients.
* The model is not learning well for "Readmitted" cases. This could be due to not having enough data sets for readmitted patients.

1. What did you learn about the Feature-based analysis for debugging?

Answers:

* auto cohort generation
* The ability to isolate the analysis into a specific feature to help pinpoint where model errors are coming from. One feature from a larger cohort may look fine, but another feature could have a lot of model errors. This helps in the process of understanding where the model issues are.

Data Analysis

1. What were examples of the data being imbalanced?

Answer:

* Gap in the number of patients among races
* Gap in the number of patients across different age groups

1. What are the Fairness and Reliability issues of an imbalance data?

Answer:

* The model predictions were skewed to favor patients that fall in groups such as race and age. This caused a fairness issue for the minority groups (e.g. race = Hispanic or age < 30) not to be represented in the “Readmitted” case.
* The model predictions were mostly “Not Readmitted” patients. The model did not learn well for “Readmitted” patients due to not enough data. Meaning the model is not reliable.

Feature Importance

1. Which features made a bigger impact on a diabetic patient's readmission classification?

Answer:

* Prior\_inpatient (past number of hospitalizations), Age and Number\_diagnoses (number of other diagnoses are on file for the patient). Which makes sense why a patient will be prone to be Readmitted vs Not Readmitted.

1. Did the aggregated Feature Importance have any Fairness issues? And why?

Answer:

* The answer is yes and no. Yes, because Age is a sensitive feature that is one of the top features driving the model’s predication. During the data analysis lab, we also saw disparities between the age groups. However, the answer can be No, because this could be a real-life representation of age in a diabetes case. That’s why it’s important for ML professionals to work with decision-makers such as clinicians to validate their behavior and expectation.

1. What are examples of how Feature importance addresses Accountability? (There is no one answer. Allow groups to come up with a scenario and provide justification).

* For example, a hospital’s AI system failed to flag a diabetic patient that is a high risk of being readmitted back to the hospital in 30 days which resulted in the patient having a severe diabetic medical emergency when sent back home. If the patient's case is audited, the hospital could use Feature Importance in trying to understand and explain what drove the model’s failure to predict the patient as likely to be Readmitted < 30 days.

-