Semester Project - Phase 1

Peter Martin

I. Frame the Problem and Look at the Big Picture

1.Define the objective of the project in business terms.

AIMed (pronounced EyeMed) is a startup developing an AI diagnosing tool for medical professionals. AIMed has tasked you with developing a model to predict fetal health using Cardiotocogram (CTG) data.

The business model is to sell API keys to the model. Medical device companies can integrate calls to the API into their embedded software, and each call to the API incurs a charge.

I. Frame the Problem and Look at the Big Picture

2.How will your solution be used?

Doctors will review the prediction generated from the API, as well as the data that was fed into the API to help inform their diagnosis.

3.How should performance be measured?

We should use goodness-of-prediction as the primary performance metric as the dataset's assumed target variable is categorical. We'll be able to visualize this with a confusion matrix.

I. Frame the Problem and Look at the Big Picture

4.Is the performance measure aligned with the business objective?

I believe it is, as the assumed success of this model is contingent on its ability to supply medical professionals with accurate predictions.

5.What would be the minimum performance needed to reach the business objective?

Your boss expects an accuracy of at least 97% for this model to be a minimum viable product. They cite an academic paper that they merely skimmed as the source for this goal.[1]

[1] Chen M, Yin Z. Classification of Cardiotocography Based on the Apriori Algorithm and Multi-Model Ensemble Classifier. Front Cell Dev Biol. 2022 May 11;10:888859. doi: 10.3389/fcell.2022.888859. PMID: 35646917; PMCID: PMC9130474.

I. Frame the Problem and Look at the Big Picture

6.List the assumptions you have made so far.

- The single biggest assumption is that a rival company hasn't beaten us to the market with a similar model. Given that an academic paper exists on this exact topic and dataset, it's likely that this startup is very late getting this product to the market.

I. Frame the Problem and Look at the Big Picture

7.Verify assumptions if possible.

- Interestingly, it appears that AI diagnostic tools are not widely accepted in CTG analysis. A 2022 study revealed that “none of [the approaches for AI in CTG] has gained big acceptance in clinical practice.” [2]
- This reveals a new assumption I was making that I had not realized; that doctors would be quick to adopt the use of highly accurate models in CTG analysis. This is clearly not the case and proves an additional hurdle for getting this product to market.

[2] Aeberhard JL, Radan AP, Delgado-Gonzalo R, Strahm KM, Sigurthorsdottir HB, Schneider S, Surbek D. Artificial intelligence and machine learning in cardiotocography: A scoping review. Eur J Obstet Gynecol Reprod Biol. 2023 Feb;281:54-62. doi: 10.1016/j.ejogrb.2022.12.008. Epub 2022 Dec 9. PMID: 36535071.

II. Get the Data

1.Document where you got the data by providing a URL.

https://www.kaggle.com/datasets/andrewmvd/fetal-health-classification

2.Get the data.

3. Convert the data to a format you can easily manipulate (without changing the data itself).

4. Perform a train/test split. Save the test set to a file.

Data and all other project materials available on github:
https://github.com/PetoriousBIG/CTG-Model-Project
Data locations and size:

- Original data - data/fetal_health.csv, (2126, 22)
- Train split - train_df.csv, (1700, 22)
- Test split - test_df.csv, (426, 22)