

Milestone 3: Preliminary Analysis

Holly Figueroa and Karlie Schwartzwald

College of Science and Technology, Bellevue University

DSC 630: Predictive Analytics

Professor Andrew Hua

January 22, 2023

Preliminary Analysis

Will I be able to answer the questions I want to answer with the data I have?

Our preliminary analysis suggests no reason our data should not perform suitably for modeling with some adjustments for scaling and normalization. No missing data was found. Histograms and descriptive statistics were used to conduct univariate analysis on all variables. Multivariate analysis was conducted using correlations matrices and box plots to illustrate variable distributions across our 3 target classes for fetal health outcome: Normal(1), Suspect(2), and Pathological(3).

Our preliminary analysis with correlations showed that variation in many of our features are associated with variations in our target classes. The relationships are not necessarily strong, but combined, they will likely offer appropriate training for our predictive models. The strongest relationships to our target classes were related to CTG features involving deceleration at .50 Tau and abnormal variability, with those Tau's ranging from .25 to .43

What visualizations are especially useful for explaining my data?

Histograms, while simple, were in many ways most effective in conveying data issues to consider. Knowing data is skewed, differently scaled, imbalanced, or lacking variation in measures, lends a lot for our preprocessing needs as mentioned above. Histograms related to heart decelerations per second were found to be highly imbalanced. But upon further inspection using value counts, the rare values represented cases with the most serious health outcome, 'pathological'. Box were effective in allowing us to understand how our features distribute and relate to our target classes. Short-term abnormal variability appeared to have a positive linear relationship across our target classes as they ranged from normal health to pathological.

Do I need to adjust the data and/or driving questions?

The analysis shows that some steps, such as scaling, and normalization for some variables will have to be taken prior to being used as input for our models. Strong correlations between variables were also shown between different central measures related to ETC histograms (mean, median, and mode), which led to the dropping of variables for median and mode. We also know that our target data will require us to use sampling techniques to balance our target classes when model training. So, while changes will have to be made, the data itself and questions remain appropriate.

Do I need to adjust my model/evaluation choices?

At this time, our model choices appear reasonable. We have chosen both a neural network classifier and a naive Bayes classifier for modeling and comparison. It is possible that the independent treatment of features within a naive Bayes may not be as well suited. The features may share probabilistic qualities with each other given the specialized topic. Box plots also illustrate what appear to be nuanced difference in feature distributions across target classes. While we do not have the domain expertise to objectively know, our current model choices can be compared to offer some insights. The confusions matrices we have chosen to evaluate results will help illustrate how well our model discerns between classes.

Are my original expectations still reasonable?

Given the steps we can take to prepare data for modeling, our plans for modeling seem reasonable. We will have a neural network classifier and a naïve Bayes classifier planned to compare. Our preliminary analysis shows no indication an effective model cannot result from this dataset. The target classes may be imbalanced, but we can utilize over sampling techniques

to give our model more proportionate data to work with. The data has no missing values, and there is no reason to believe any of the observations were made in error. We also already know that many of our features have some established relevance to predicting fetal outcomes as they represent output features from existing diagnostic techniques.

Updated Milestone 2:

A Machine Learning Approach to Cardiotocography Interpretation

Despite modern advances, childbirth remains a risky medical endeavor both for birthing people and the unborn. There exists no replacement for the expertise and direct involvement of a medical professional to manage care. Analytics, however, can offer significant aid in ways it could not before through advanced modeling. Cardiotocography (CTG) is widely used over the course of pregnancy and during labor (*Fetal Heart Monitoring*, 2019). CTG allows monitoring of fetal heart rate and uterine contractions to prevent fetal hypoxia, or oxygen loss in body tissues (Boudet et al., 2020). Creating effective modeling for CTG exam data can provide additional support to those dedicated preventing harm and loss of life. The goal of this project will be to build an effective model to classify CTG data using supervised machine learning.

Data Selection

The dataset chosen for model training contains 2126 instances of heart-health-related features taken from cardiotocogram exams of fetuses. These were then classified by three expert obstetricians into 3 classes: 1 – Normal, 2 – Suspect, or 3- Pathological, making the target a multi-class categorical feature. There are a total of 22 features, both numerical and categorical. We found this dataset on Kaggle, but it was originally published in the Journal of Maternal-Fetal & Neonatal Medicine in the year 2000. This dataset should be useful for creating a model that

can predict suspect and pathological health outcomes, with the assumption that some of the features provided will be correlated with fetal heart health.

Model Selection & Evaluation

We plan to implement two different models for this project, both of which will perform classification on the fetal heart health target feature. The first model will be a neural network. We think that this dataset provides a good opportunity to train a neural network because of the substantial number of instances in the data. Neural networks are also considered some of the most accurate models, and they are ideal in a medical setting like this where accuracy is important. We plan to use the percentage of accuracy to score and evaluate this model. In this context where we are discussing medical diagnostics, a false negative has more detrimental outcomes to the patient's life than a false positive. Thus, we will review the confusion matrices in addition to the accuracy percentage, so that we can calculate rates for false positives and false negatives.

The second model we plan to use is Naïve Bayes classification model. Although the dataset contains over 2,000 instances, it is possible that this amount of data is not sufficient to train a competent neural network. The Naïve Bayes model is recommended as an alternative in that circumstance. It also has a reputation as an effective classifier and has been proven useful in medical diagnostics (Simfukwe et al., 2015). Since we are still doing classification, we will be using accuracy and confusion matrices to evaluate the results of our model. We can compare the accuracy score of this model to the neural network model in this proposal.

To ensure quality and accuracy, we plan to implement a variety of techniques to improve both models. We will need to clean the data, do some feature scaling, and feature selection. If

there are any categorical features that are relevant to the model, we can create dummy variables with them before training the model. We plan to use a grid search to perform hyperparameter tuning on our models as well. Since the target feature has imbalanced classes, we will need to perform oversampling on the underrepresented classes.

Risks and Ethical Implications

Machine learning and artificial intelligence has already earned space in healthcare with the increased use of AI diagnostic systems. Some common examples involve the interpretation of x-rays, MRI scans, CT scans, and other image-related tools of measurement. While machine learning healthcare applications (ML-HCAs) and AI recommendation systems can offer powerful insights for the medical field, limits must be acknowledged. The most fundamental limit being correlation is not causation, therefore their role must be additive to care, not a replacement for expertise. A common hope is that the addition of ML-HCAs and AI diagnostics will circumvent bias within the healthcare system, however, these tools are just as susceptible (Naik, 2022). The data used to train a diagnostic model is critical to how accurately these systems perform for diverse populations. It is unknown how these concerns apply to CTG exams of the unborn. For example, we do not know if CTG exams, in practice, predict or protect all groups equally. We also do not know at this time if the anomalies associated with fetal heart issues present the same across groups. For this project, however, we will assume the data is sound towards its purpose.

Contingency Proposal

In the case that our original proposal is unsuccessful, the contingency plan will be to use an alternative dataset to train a regression model. We found a dataset that is intended to train

models to predict the insurance premium price of an individual given certain health history data points about them. This dataset has 10 predicting features about the health history of an individual and 986 instances with no missing values. Since our target is continuous and we want to include multiple features, we will use a multiple regression model.

References

- Boudet, S., Houzé l'Aulnoit, A., Demailly, R., Delgranche, A., Peyrodie, L., Beuscart, R., & Houzé de l'Aulnoit, D. (2020). A fetal heart rate morphological analysis toolbox for MATLAB. *SoftwareX*, 11, 100428. <https://doi.org/10.1016/j.softx.2020.100428>
- Fetal Heart Monitoring*. (2019, August 14). Johns Hopkins Medicine.
<https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/fetal-heart-monitoring>
- Naik, N. (2022, March 14). *Legal and Ethical Consideration in Artificial Intelligence in Healthcare: Who Takes Responsibility?* Frontiers.
<https://www.frontiersin.org/articles/10.3389/fsurg.2022.862322/full>
- Simfukwe, M., Kunda, D., & Chembe, C. (2015). *Comparing Naive Bayes Method and Artificial Neural Network for Semen Quality Categorization*. International Journal of Innovative Science, Engineering & Technology, 2(7)
https://ijiset.com/vol2/v2s7/IJSET_V2_I6_90.pdf