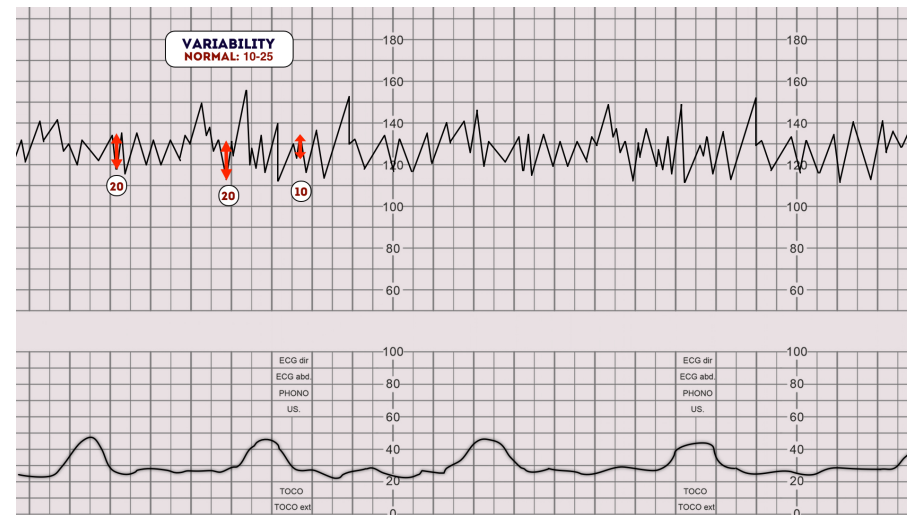




FETAL HEALTH CLASS MODELING

Overview

- Prevent child mortality
- Not everyone has access to technology
- CTGs scans are cost effective and widely spread
- Eliminate erroneous surgical intervention




Business Problem

CTG scans are currently interpreted via visual analysis by the physician and reading errors may increase fetal health risk





Business Questions



What feature
coefficients have the
greatest influence on
the model?



What, if any, new
features affect
the model?



Which model
makes the best
predictions of fetal
health class?





Data

- From Kaggle
- It was used in scientific researches
- It has 2,126 rows and 22 columns
- Target variables were healthy and distressed
- CTG Metrics

Methods



Exploratory Data Analysis



Vanilla and Experimental Modeling Process



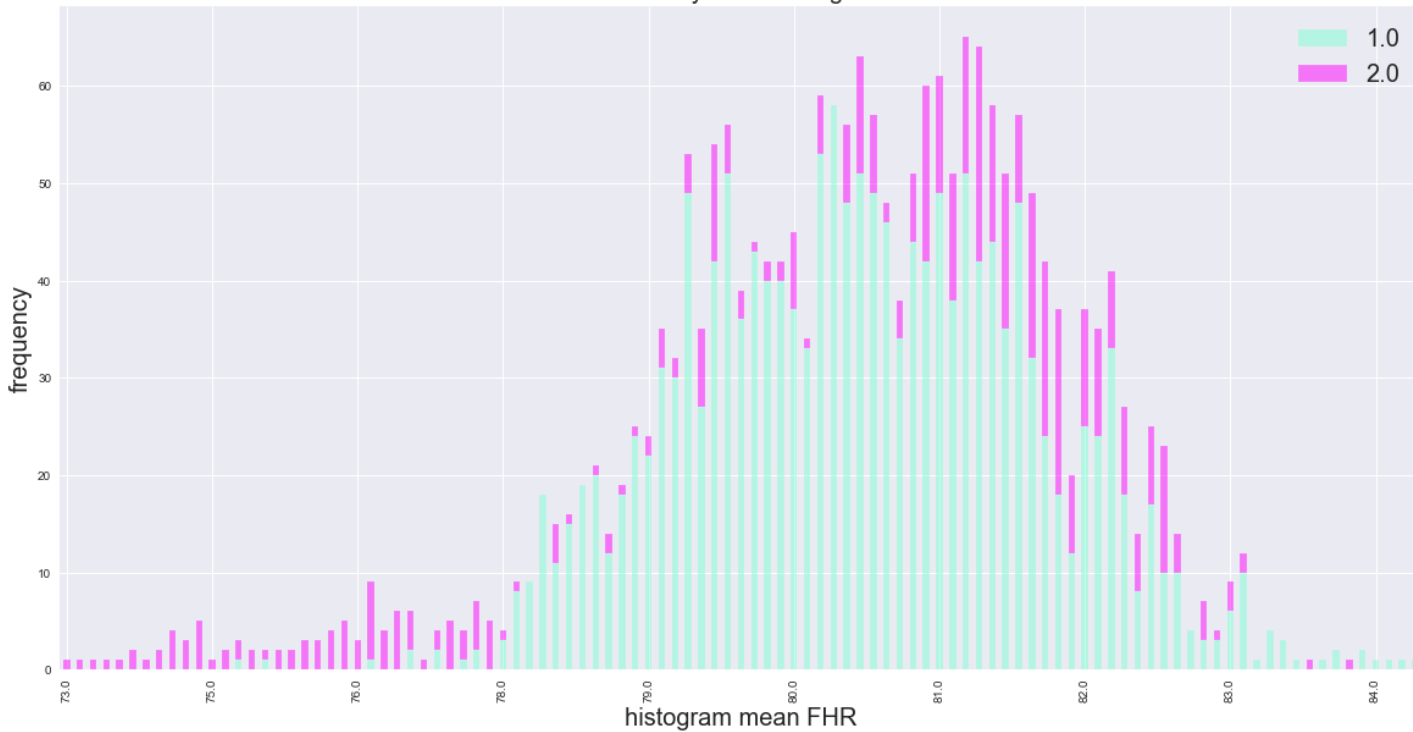
Logistic Regression, KNN, Decision Tree, Random Forest, Grid Search, and XGBoost



Changed target class label into binary fetal health class

EDA

Fetal Health by Mean histogram FHR



Depth analysis of features and relations with other features and the target class

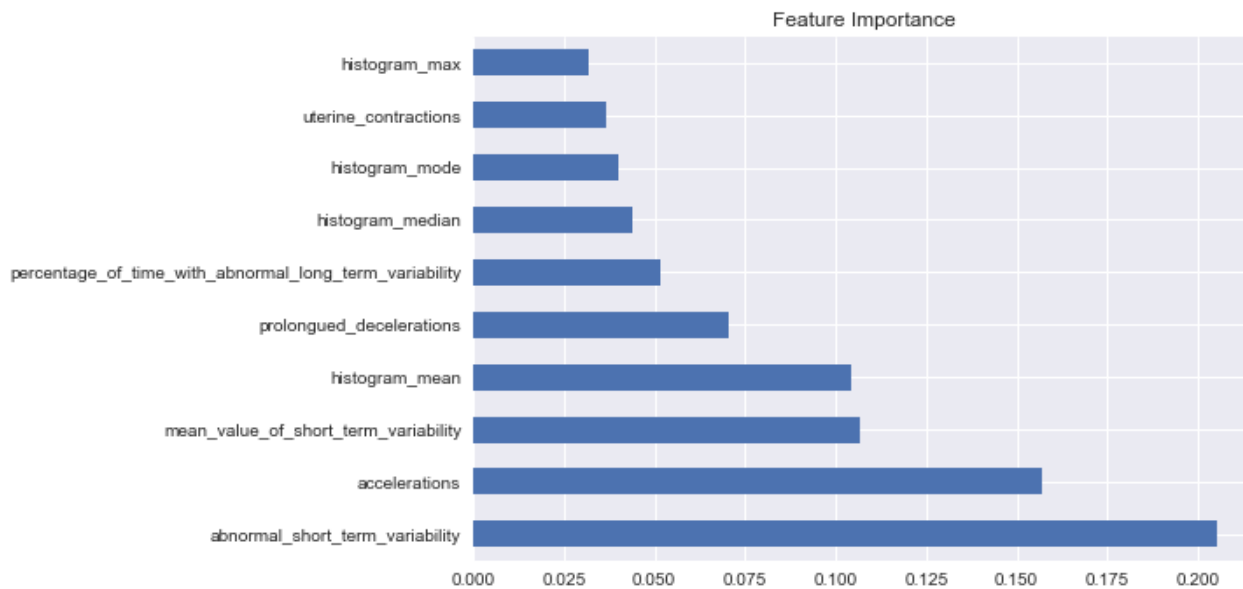


Our baseline models performed well even without feature engineering



Feature engineering for experimental model

Vanilla Modeling Process



Takeaways

- Unmodified dataset
- Our baseline models had a good performance
- Grid Search Random Forest had the best performance

Evaluation Metrics:

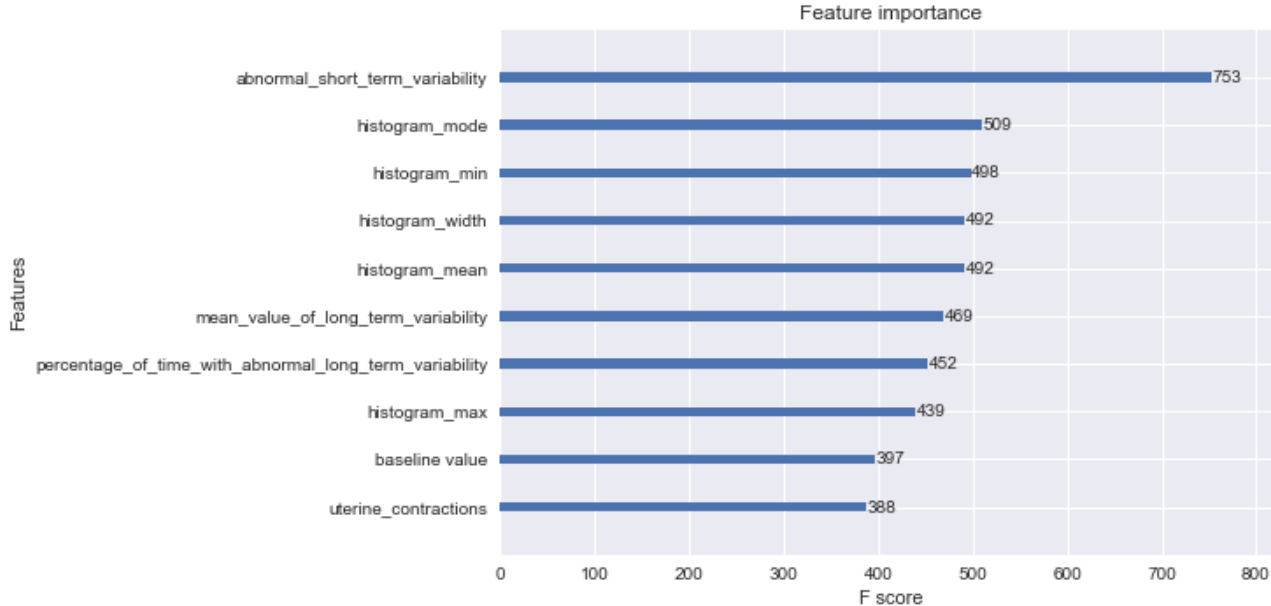
Accuracy: 0.9605

Recall: 0.9755

F1 Score: 0.9743

Precision: 0.9731

Experimental Modeling Process



Takeaways

- Feature engineering in the dataset
- We were able to improve our models even more
- XGBoost Random Forest had the best performance

Evaluation Metrics:

Accuracy: 0.9661

Recall: 0.9902

F1 Score: 0.9782

Precision: 0.9665

Final Model Evaluation



Vanilla and Experimental Model Evaluation

- XGBoost Random Forest was the overall best model
- Lowest recall and the a very high precision score
- Random Forest Grid Search performed well, but XGBoost Random Forest had a better Recall

Random Forest Grid Search Metrics:

Accuracy: 0.9605
Recall: 0.9755
F1 Score: 0.9743
Precision: 0.9731

XGBoost Random Forest Metrics:

Accuracy: 0.9661
Recall: 0.9902
F1 Score: 0.9782
Precision: 0.9665

Conclusion



Most Important CTG Metrics

- Abnormal short term variability
- Histogram mode
- Histogram mean
- Histogram width
- Histogram min



Final Conclusion

- Our identified key features are the same features that doctors look for during a visual analysis. Because our model is so precise, we can be confident that our model will predict the class at a better rate than the visual analysis, eliminating human error.



Next Steps

- Make a multiclass prediction for fetal health
- Find a larger dataset
- Consider maternal health and other diagnostic metrics into the model (heart rate, oxygen level, what anesthetics are used, etc.)