Maternal Health Risk Predictor

Lennon Au-Yeung, Chenyang Wang, Shirley Zhang (Team 14)

2022-12-04

Contents

Summary **1**

Introduction **1**

Data and EDA **2**
Data . . . 2
Exploratory Data Analysis (EDA) . . . 2

Model Building **4**
Exploratory Model Building . . . 4
Hyperparameter Optimization . . . 5
Evaluating on Test Data . . . 5

Assumptions and Limitations **6**

Future Directions **6**

References **7**

Summary

This data analysis project was created in fulfillment of the team project requirements for DSCI 522 (Data Science Workflows), for the Master of Data Science program at the University of British Columbia.

Introduction

Maternal mortality is a large risk in lower and lower middle-income countries, with about 810 women dying from preventable pregnancy-related causes each day (WHO, 2019). Often, there is a lack of information about the woman's health during pregnancy, making it difficult to monitor their status and determine whether they may be at risk of complications (Ahmed and Kashem, 2020). A potential solution to this issue is through using the 'Internet of Things (IoT),' or physical sensors which can monitor and report different health metrics of a patient to their health care provider. Medical professionals can then analyze this information to determine whether a patient may be at risk.

For this project, we aim to answer the question:

> **"Can we use data analysis methods to predict the risk level of a patient during pregnancy (low, mid, or high) given a number of metrics describing their health profile?"**

This is an important question to explore given that human resources are low in lower income countries, and non-human dependent classification methods can help provide this information to more individuals. Furthermore, classifying a patient's risk level through data-driven methods may be advantageous over traditional methods which may involve levels of subjectivity.

IoT sensors can collect a diverse range of health metrics, however not all of them may be useful in predicting whether a patient is at risk of adverse health outcomes (Sutton et al. 2020). Thus, time permitting, we also hope to use data analysis methods to infer (sub-question) whether some metrics may be more important in determining maternal health risk levels than others.

Data and EDA

The R programming language (R Core Team 2019) and the following R packages were used to perform the analysis: knitr (Xie 2014). The code used to perform the analysis and create this report can be found here: https://github.com/UBC-MDS/maternal_health_risk_predictor/blob/main/doc/final_report.md.

Data

Data used in this study was collected between 2018 and 2020, through six hospitals and maternity clinics in rural areas of Bangladesh (Ahmed and Kashem, 2020). Patients wore sensing devices which collected health data such as temperature and heart rate. The risk factor of each patient was determined through following a guideline based on previous research and consultation with medical professionals.

The full data set was sourced from the UCI Machine Learning Repository (Asuncion and Newman 2007), and can be found here. A .csv format of the data can be directly downloaded using this link. The data can be attributed to Marzia Ahmed (Daffodil International University, Dhaka, Bangladesh) and Mohammod Kashem (Dhaka University of Science and Technology, Gazipur, Bangladesh) (Ahmed and Kashem, 2020).

The data set contains six features describing a patient's health profile, including `age`, `SystolicBP` (systolic blood pressure in mmHG), `DiastolicBP` (diastolic blood pressure in mmHG), `BS` (blood glucose levels in mmol/L), `BodyTemp` (body temperature in Fahrenheit), and `HeartRate` (heart rate in beats per minute). There are 1014 instances in total, with each row corresponding to one patient. Finally, the data contains the attribute `RiskLevel`, corresponding to a medical expert's determination of whether the patient is at low, mid, or high risk (Ahmed et al., 2020).

Exploratory Data Analysis (EDA)

We first wanted to examine the distribution of counts of our three target classes in the training set, in order to see whether we would need to deal with class imbalance. Figure 1. shows this distribution as a bar graph. Upon visualization, there does not appear to be a drastic class imbalance.

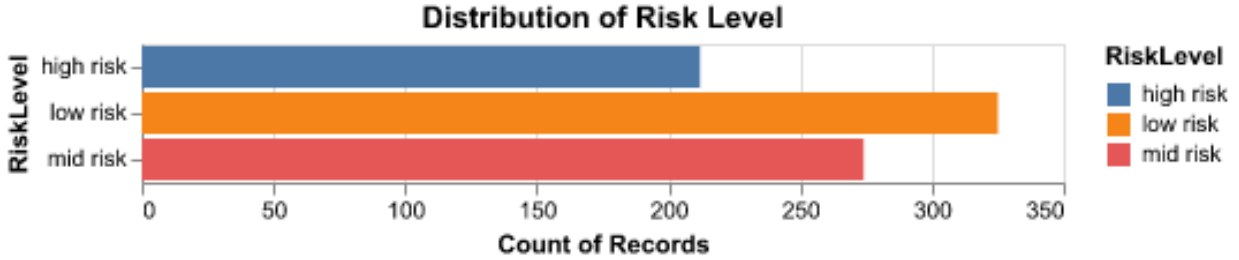

Figure 1. Counts of Observations for the Target Class RiskLevel in the Training Set

As all of our features were continuous, we plotted the density distribution of each feature colored by different target classes (high risk = blue, low risk = orange, mid risk = red). This is shown in Figure 2. below. These graphs provided us with insights on whether the distribution of some features are different for different target classes. For example, the distribution of `age` for the high risk class is more symmetrical than the low risk

and mid risk classes (which are both right-skewed). This could indicate that `age` could be a useful predictor in determining whether someone is at high risk. On the other hand, it appears that the feature `BodyTemp` has similar distributions for all classes, suggesting that it may not be not helpful in predicting the target.

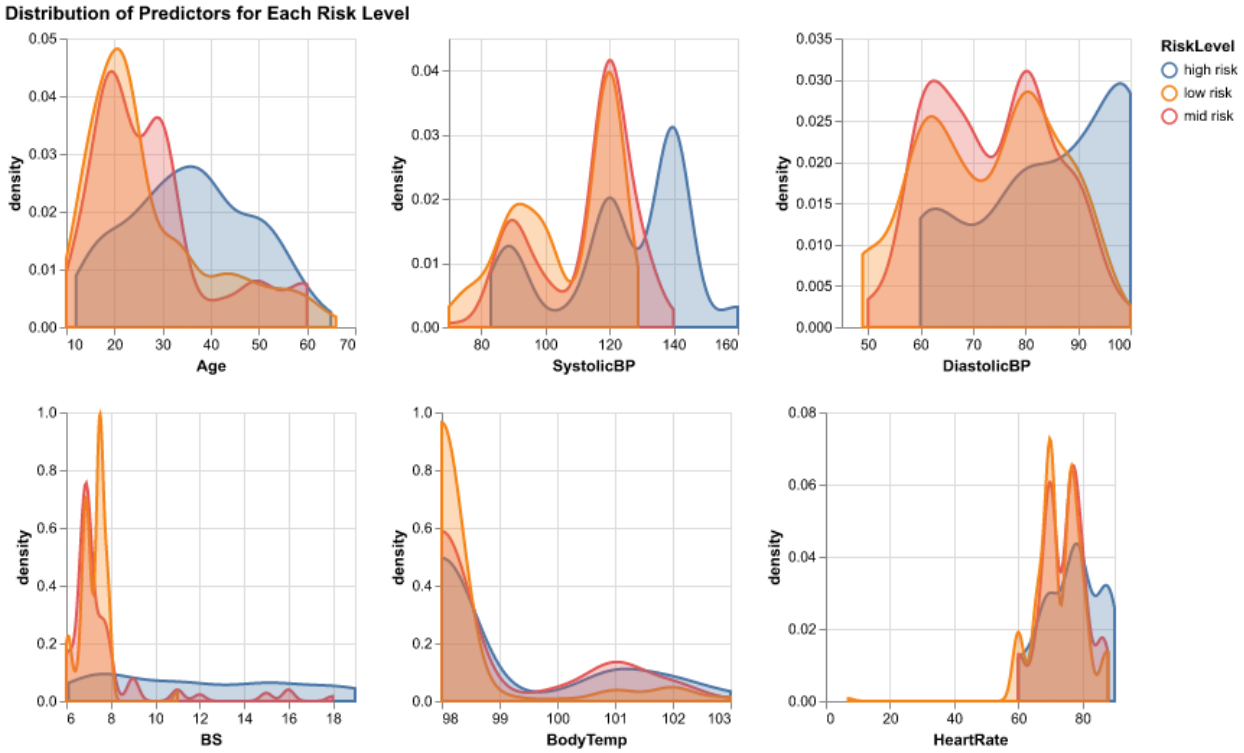

Figure 2. Distribution of Features for the Target Class RiskLevel in the Training Set

Finally, we created pairwise scatter plots for all features to examine whether some features may be correlated with one another. Figure 3. shows that the features `SystolicBP` and `DiastolicBP` have high correlation compared to other pairs of predictors, followed by the correlation between the two blood pressure levels and age. For other pairs of predictors, there are no significant correlations found.

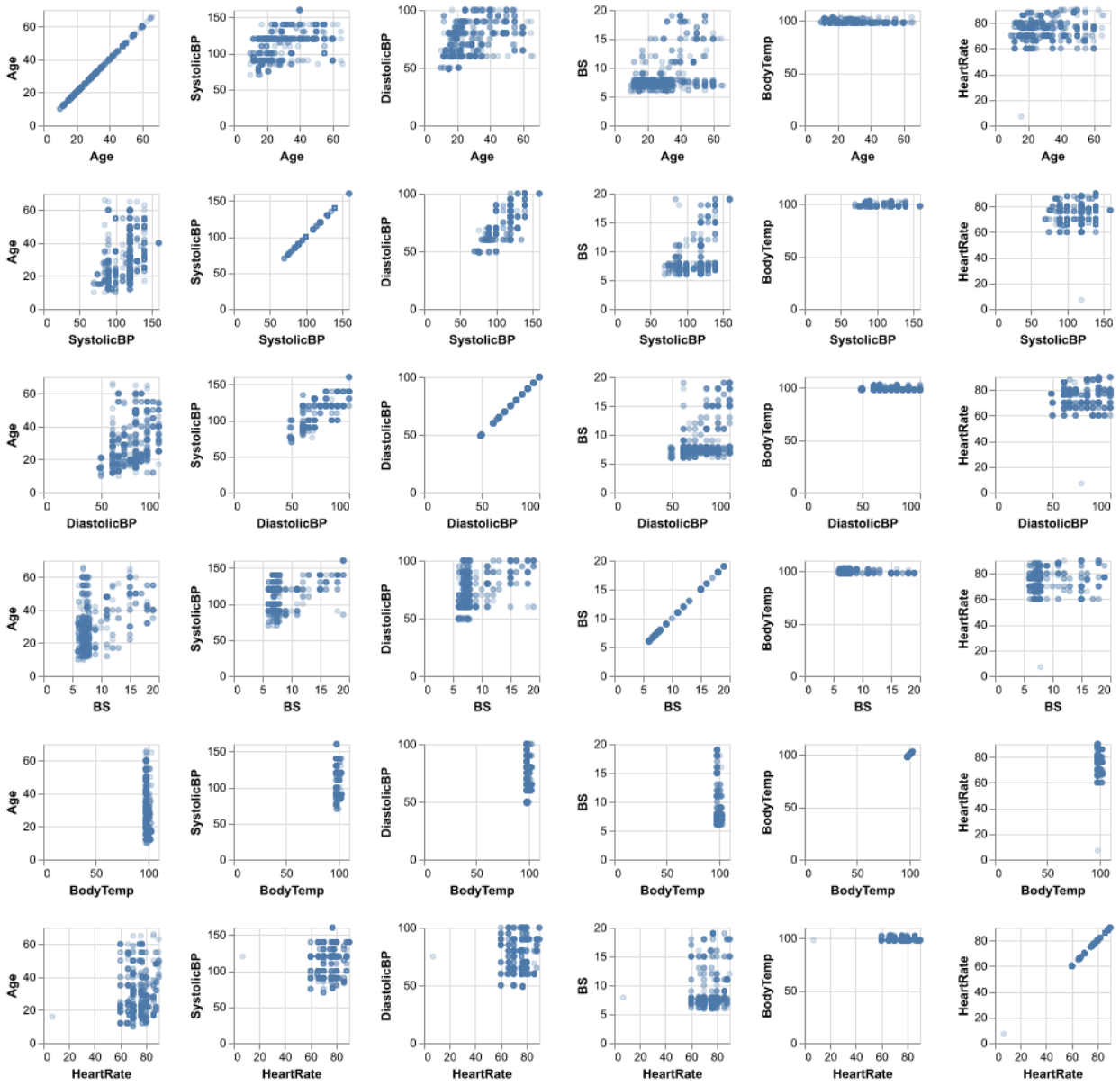

Figure 3. Pairwise Relationships Between Features

Model Building

Exploratory Model Building

Before applying our machine learning models, we separated our training set into 'X_train' and 'y_train.' As all of our features were numeric and continuous with different ranges, we performed standard scaling on 'X_train' with the Scikit-learn Standard Scaler package (Pedregosa et al. 2011).

We decided to test a few different classification models to decide which we should choose in our final analysis:

1. Dummy Classifier
2. Decision Tree (Myles et al. 2004)
3. Support Vector Machines (SVMs) (Hearst et al. 1998)
4. Logistic Regression

Table 1: Table 1. Comparison of Classification Models

	Dummy	Decision Tree	SVM	Logistic Regression	K-Nearest Neighbors
Training Score	0.401	0.928	0.714	0.607	0.797
Mean Cross Validation Score	0.401	0.808	0.695	0.610	0.668

5. K-Nearest Neighbors (KNN)

For all above models, we used the default parameters and did not include hyperparameter optimization. Table 1. shows the comparison of the models with their respective training scores and mean cross validation scores (default 5-fold cross validation). As the Decision Tree model had the highest mean cross validation score, we decided to choose this one for our final analysis.

Hyperparameter Optimization

For hyperparameter optimization, we used a random search method for max depths ranging between 1 to 50. We used 50-fold cross-validation, and computed the average accuracy between folds to compare between different max depths. We plotted these averaged scores against the max depths as shown in Figure 4. The best depth found by the randomized search was 29.

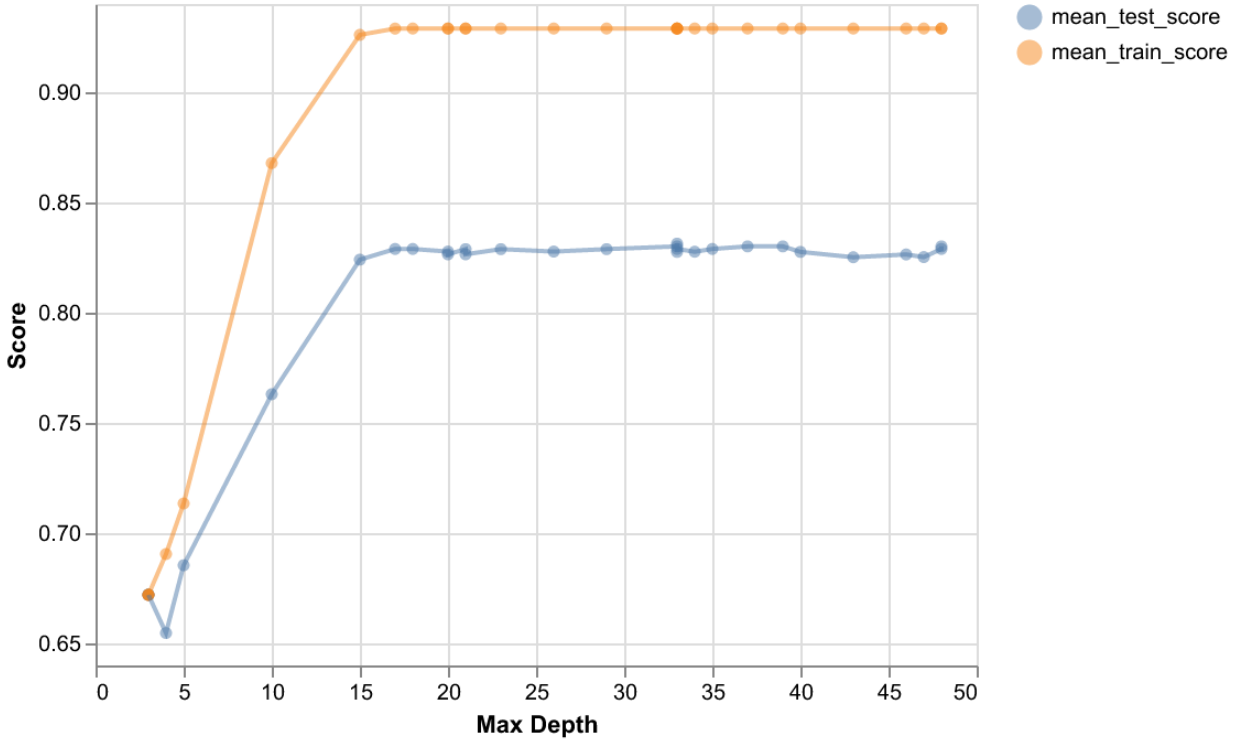

Figure 4. Hyperparameter Optimization for Max Depth with Mean Train and Test Scores

Evaluating on Test Data

With a max depth 29, we scored our best Decision Tree model on our test data set and obtained an accuracy score of **0.823**.

Although this score appears high, we also wanted to further examine the predictions made by our model for each of the target classes. Table 2. shows the confusion matrix. We are mainly interested in seeing whether

Table 2: Table 2. Confusion Matrix with Test Data

	Predicted High Risk	Predicted Low Risk	Predicted Mid Risk
True High Risk	53	1	6
True Low Risk	1	67	13
True Mid Risk	4	10	48

our model can correctly predict whether an individual is at high maternal health risk, as it would be more detrimental to incorrectly label a high risk individual as mid or low risk.

Our model predicted 53 high risk observations correctly out of a total of 60 true high risk observations (recall = 0.88). Our model also predicted 58 high risk observations in total (5 of these predictions were incorrect, giving a precision of 0.91). The f1-score for the high risk class would be 0.90.

Assumptions and Limitations

Before our analysis, we made the following assumptions:

1. The maternal risk dataset that we used is representative of the population of patients.
2. The risk level classified in the data set is a good indicator of the patient's risk level.
3. The data collected is unbiased.

The dataset we used was collected from the rural areas of Bangladesh, and it might be possible that patients in different regions have different characteristics of health information that affects their maternal risk level. Thus, a limitation to our model is that it may not be as accurate when predicting patients from other regions.

During our analysis, we also assumed that:

1. Our data set did not contain any class imbalance
2. A Decision Tree model may be the best choice based on initial modeling looking at cross-validation scores of accuracy

However, it is possible that there is class imbalance that we did not account for. Furthermore, it may be that Decision Trees had the best cross-validation scores without hyperparameter tuning, but another model with optimal hyperparameters could outperform our best model. Furthermore, we only looked at accuracy during cross-validation when choosing the model, and it may be more beneficial to look at recall or f1-scores.

Future Directions

As mentioned in the introduction, we are trying to determine whether a patient is at risk, and identifying patients with a high risk level should be our priority. We propose the following next steps to improve our model:

1. We could combine low and mid risk level into singular class so that we would have a binary classifier. In this way, we could focus on improving the prediction of the high risk class. We would explore classification metrics such as recall instead of using accuracy, as we are trying to minimize the number of false negatives such that high risk patients are not being misclassified by the model.
2. As mentioned in the limitations, we could add `class_weight` in our hyperparameter optimization to see whether handling class imbalance would improve our model.
3. We could also try hyperparameter tuning with other classification models (instead of just Decision Trees) to see whether they may have better performance.
4. We could drop some features which do not seem meaningful, like `BodyTemp` to see if a simpler model is better.
5. Finally, we can examine Precision-Recall (PR) curves and adjust the threshold of our model to improve the recall for the high risk class.

References

Asuncion, Arthur, and David Newman. 2007. "UCI Machine Learning Repository." Irvine, CA, USA.

Hearst, Marti A., Susan T Dumais, Edgar Osuna, John Platt, and Bernhard Scholkopf. 1998. "Support Vector Machines." *IEEE Intelligent Systems and Their Applications* 13 (4): 18–28.

Myles, Anthony J, Robert N Feudale, Yang Liu, Nathaniel A Woody, and Steven D Brown. 2004. "An Introduction to Decision Tree Modeling." *Journal of Chemometrics: A Journal of the Chemometrics Society* 18 (6): 275–85.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12 (Oct): 2825–30.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Sutton, Elizabeth F, Sarah C Rogan, Samia Lopa, Danielle Sharbaugh, Matthew F Muldoon, and Janet M Catov. 2020. "Early Pregnancy Blood Pressure Elevations and Risk for Maternal and Neonatal Morbidity." *Obstetrics and Gynecology* 136 (1): 129.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.