Assignment 4: Predicting Maternal Health Risk Using Classification

By Victoria Hernandez

## Introduction

The dataset that will be analyzed in this report is titled “Maternal Health Risk” and it was created by Marzia Ahmed. Maternal Health is a highly important health issue that affects women worldwide. According to the World Health Organization (WHO), in 2020 around 287,00 women died both during and after childbirth. (World Health Organization & Shimizu, n.d.) Some of the leading causes include infection, high blood pressure, and excessive blood loss. (World Health Organization & Shimizu, n.d.) Predictive models can help analyze different indicators such as heart rate, blood pressure, blood glucose levels, age, and body temperature.

To analyze this data set exploratory statistics and classification models will be utilized. The exploratory methods used include analysis of the mean, minimum, maximum, standard deviation, and quartiles. The preprocessing methods used to prepare for classification are the holdout method and partitioning of the dataset. The machine learning methods used include a neural network, random forest, and an ensemble model with XGBoosting. In the upcoming sections the report will explain the data, methods used, explain the classification, provide a conclusion of the key findings, and discuss potential future analysis.

## Data

The dataset used for this report was sourced from the UC Irvine Machine Learning Repository. The creator of the dataset Marzia Ahmed sourced the data from different hospitals, maternal health care, and community clinics in rural areas of Bangladesh using an IoT based risk monitoring system. (Ahmed, 2020) An IoT-based risk monitoring system uses devices that are interconnected to collect and then analyze data in real time. (Aadil, Khan, Yu, Ali, & Kumar, 2024) The dataset contains 1,013 patient records and each record includes the patient’s age, blood glucose level (BS), body temperature, systolic and systolic blood pressure, heart rate, and risk level. The target variable for this analysis is the patient’s Risk Level.

**Table 1.** Attributes from dataset.

| Attributes | Description | Type | Range |
|---|---|---|---|
| Age | Age in years during pregnancy. | Numeric | 10-70 years |
| SystolicBP | Upper value of blood pressure in mmHg. | Numeric | 70-160 mmHg |
| DiastolicBP | Lower value of blood pressure in mmHg. | Numeric | 49-100 mmHg |
| BS | Blood glucose levels in terms of molar concentration. | Numeric | 6-19 mmol/L |
| BodyTemp | Body temperature in Fahrenheit. | Numeric | 98-103 Fahrenheit |

| | | | |
|---|---|---|---|
| HeartRate | Normal resting heart rate in bpm. | Numeric | 7-90 bpm |
| RiskLevel | Predicated risk intensity during pregnancy. | Categorical | Low Risk<br>Mid Risk<br>High Risk |

## Methods

Exploratory Data Analysis:

- The descriptive statistics used in this analysis are metrics such as the mean, minimum, maximum, standard deviation, and quartiles.

Data preprocessing:

- The RiskLevel variable was converted into a factor for classification in R.
- The dataset was partitioned by 70% for training and 30% for testing using the holdout method. The training set is used for model construction and the test set is used for accuracy estimation. (McGuire, 2025)
- The outliers were removed from all the variables using the IQR method. A custom R code was used to scan and remove rows containing outliers from numeric variables. These variables include HeartRate, SystolicBP, DiastolicBP, and BS, and age. The purpose of this process was to perform the analysis with consistent data that would allow the classification model to perform accurately.

Classification Models:

- Neural Network: Using the nnet package in R neural network was implemented. The model has an input layer, a hidden layer, and an output layer with three neurons. (McGuire, 2025) Cross validation was used to train the model in R.
- Random Forest: This is an ensemble method that builds decision trees and combines each prediction using a majority voting method. (McGuire, 2025) Each tree is trained using a different subset of the data which reduces overfitting and increases accuracy. (McGuire, 2025) Using R the model was trained using 10-fold cross validation with RiskLevel used as the outcome variable. The results of this model were evaluated using the sensitivity, F1 score, accuracy, specificity, precision, error rate, recall, and F-score.
- Boosting using XGBoost: This is in ensemble model initializes initial prediction and then goes through a few iterations, then uses those previous iterations too improve its accuracy. (McGuire, 2025) Through the caret package the xgbtree method was implemented and 10-fold cross validation to train the model.

**Use of generative AI:** The tool used was OpenAI ChatGPT. Generative AI was used to assist in creating an R code function to remove all outliers from numeric variables in the dataset. Specifically, it was used to create an IQR based function to scan through the dataset and remove rows with extreme values. The prompt used was "Write a function to remove all outliers from numeric variables using IQR method in R."

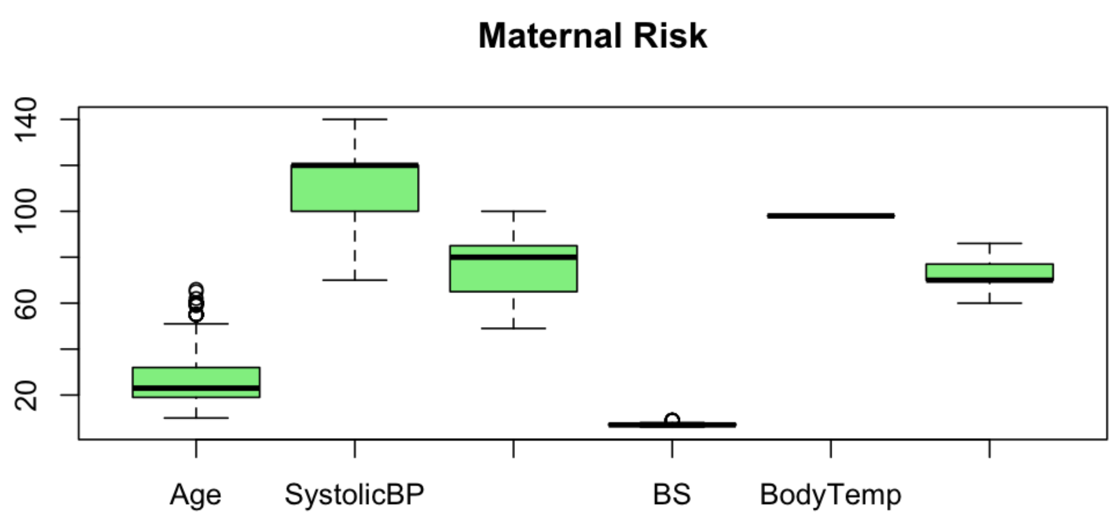
## Results

**Exploratory statistics:**

**Figure 1.** Exploratory statistics of the cleaned maternal health risk data.

```
> summary(Maternal_Risk_Clean)
      Age          SystolicBP     DiastolicBP           BS           BodyTemp     HeartRate
 Min.   :10.0   Min.   : 70.0   Min.   : 49.00   Min.   :6.000   Min.   :98   Min.   :60.00
 1st Qu.:19.0   1st Qu.:100.0   1st Qu.: 65.00   1st Qu.:6.800   1st Qu.:98   1st Qu.:70.00
 Median :23.0   Median :120.0   Median : 80.00   Median :7.000   Median :98   Median :70.00
 Mean   :27.9   Mean   :111.7   Mean   : 75.07   Mean   :7.159   Mean   :98   Mean   :72.51
 3rd Qu.:32.0   3rd Qu.:120.0   3rd Qu.: 85.00   3rd Qu.:7.500   3rd Qu.:98   3rd Qu.:77.00
 Max.   :66.0   Max.   :140.0   Max.   :100.00   Max.   :9.000   Max.   :98   Max.   :86.00
     RiskLevel
 high risk: 50
 low risk :336
 mid risk :214
```

**Figure 2.** Boxplots for variable of cleaned maternal risk data.




Age: Based on the 1st and 3rd quartiles most of the recorded ages are from 19 to 39 years.

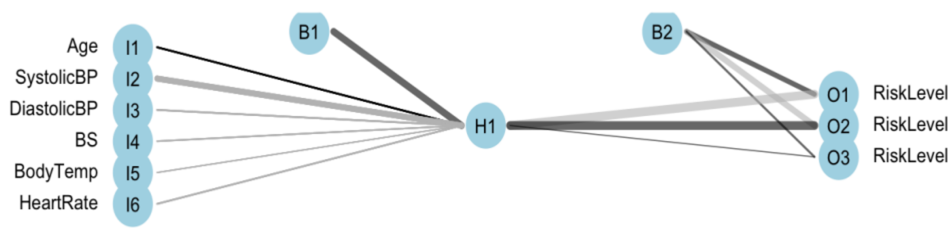
Heart Rate: The average heart rate was 72.51 bpm based on the mean.

Body Temperature: The average body temperature is 98 degrees F, which is in the normal range for body temperature.

Blood Glucose(BS): The mean is 7.5 mmol/L and the median is 7.0 mmol/L meaning it is skewed right.

**Neural Network**

**Figure 3. Neural Network plot.**




I1- I6 are input neurons representing each variable. H1 is 1 hidden layer with 1 neuron. There are 3 output neurons, O1 = High Risk, O2 = Low Risk, O3 = Mid Risk. The diagram shows that SystoicBP and Age have the strongest effect on the model's predictions. BodyTemp, BS, and HeartRate have lighter lines, showing that these variables have less impact on the model predictions.

**Figure 4.** Neural Network Confusion Matrix and Overall Performance.

```
Confusion Matrix and Statistics

           Reference
Prediction  high risk low risk mid risk
  high risk        13        0        0
  low risk          1       99       49
  mid risk          1        1       15

Overall Statistics

               Accuracy : 0.7095
                 95% CI : (0.6371, 0.7748)
    No Information Rate : 0.5587
    P-Value [Acc > NIR] : 2.399e-05

                  Kappa : 0.413

 Mcnemar's Test P-Value : 2.048e-10
```

Statistics by Class:

| | Class: high risk | Class: low risk | Class: mid risk |
|---|---|---|---|
| Sensitivity | 0.86667 | 0.9900 | 0.23438 |
| Specificity | 1.00000 | 0.3671 | 0.98261 |
| Pos Pred Value | 1.00000 | 0.6644 | 0.88235 |
| Neg Pred Value | 0.98795 | 0.9667 | 0.69753 |
| Precision | 1.00000 | 0.6644 | 0.88235 |
| Recall | 0.86667 | 0.9900 | 0.23438 |
| F1 | 0.92857 | 0.7952 | 0.37037 |
| Prevalence | 0.08380 | 0.5587 | 0.35754 |
| Detection Rate | 0.07263 | 0.5531 | 0.08380 |
| Detection Prevalence | 0.07263 | 0.8324 | 0.09497 |
| Balanced Accuracy | 0.93333 | 0.6785 | 0.60849 |

Accuracy: The neural network model had 70.95% accuracy.

- High Risk: This class has a very strong performance with almost no missed cases. Its sensitivity shows 87% of actual high risk cases were accurate. Its precision is show 1.00 meaning it predicted high risk 100% of the time. Its F1-score is 93% meaning the balance between its sensitivity and precision is strong. Overall, these statistics show that the model is very good at detecting high-risk cases.
- Low Risk: The low risk class has a sensitivity of 99%, so it almost never misses these cases. Its precision is 66%, meaning about 34% were misclassified as either mid risk or high risk. Its F1 score is 80%, which is decent, but its precision is causing the imbalance. Overall, the model can adequately classify low risk but sometimes classifies low risk as other classes.
- Mid Risk: The mid risk class has a sensitivity of 23%, so it misses most of the low risk cases. Its precision is 88%, so when it does predict mid risk cases it predicts correctly. Its F1-Score is 37%, meaning the balance between precision and sensitivity is weak. Overall, the model is not good at detecting mid risk cases.

**Random Forest**

**Figure 5.** Random forest confusion matrix and statistics.

```
Confusion Matrix and Statistics

           Reference
Prediction  high risk low risk mid risk
  high risk        15        2        0
  low risk          0       81        9
  mid risk          0       17       55

Overall Statistics

               Accuracy : 0.8436
                 95% CI : (0.7819, 0.8935)
    No Information Rate : 0.5587
    P-Value [Acc > NIR] : 4.114e-16

                  Kappa : 0.7243

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: high risk Class: low risk Class: mid risk
Sensitivity                   1.00000          0.8100          0.8594
Specificity                   0.98780          0.8861          0.8522
Pos Pred Value                0.88235          0.9000          0.7639
Neg Pred Value                1.00000          0.7865          0.9159
Precision                     0.88235          0.9000          0.7639
Recall                        1.00000          0.8100          0.8594
F1                            0.93750          0.8526          0.8088
Prevalence                    0.08380          0.5587          0.3575
Detection Rate                0.08380          0.4525          0.3073
Detection Prevalence          0.09497          0.5028          0.4022
Balanced Accuracy             0.99390          0.8480          0.8558
```

Accuracy: The random forest model had 84.36% accuracy.

- High Risk: Its sensitivity is 100%, meaning it was able to predict all high risk cases. It also has a high precision of ~88% and a high F1-score of ~93%. Overall, it is highly accurate when classifying high risk cases.
- Low Risk: Its sensitivity is 81%, meaning it is very good at classifying low risk cases. Its precision is 90%, meaning about 10% were misclassified. Its F1-score was 85% meaning it was strong at classifying low risk cases overall.
- Mid Risk: The sensitivity is ~ 85% meaning it has a strong ability to classify mid risk cases. Its precision is 76%, meaning it is fairly good at predicting mid risk cases. Its F1-score is 80% meaning it has a good balance between precision and sensitivity.

## Ensemble Model with XGBoosting

Figure 6. Boosted Classifier using XG-boost method confusion matrix and statistics

```
Confusion Matrix and Statistics

           Reference
Prediction  high risk low risk mid risk
  high risk        15        2        0
  low risk          0       84        9
  mid risk          0       14       55

Overall Statistics

               Accuracy : 0.8603
                 95% CI : (0.8008, 0.9075)
    No Information Rate : 0.5587
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.7524

 Mcnemar's Test P-Value : NA

Statistics by Class:

                     Class: high risk Class: low risk Class: mid risk
Sensitivity                   1.00000          0.8400          0.8594
Specificity                   0.98780          0.8861          0.8783
Pos Pred Value                0.88235          0.9032          0.7971
Neg Pred Value                1.00000          0.8140          0.9182
Precision                     0.88235          0.9032          0.7971
Recall                        1.00000          0.8400          0.8594
F1                            0.93750          0.8705          0.8271
Prevalence                    0.08380          0.5587          0.3575
Detection Rate                0.08380          0.4693          0.3073
Detection Prevalence          0.09497          0.5196          0.3855
Balanced Accuracy             0.99390          0.8630          0.8688
```

Accuracy: The Boosted Classifier using XGboost method model had 86.03% overall accuracy.

- High Risk: The model had a sensitivity and recall of 100% and didn't miss any high risk cases. Its precision was ~88% of the cases predicted correct. Its F1-Score was ~93, meaning it was almost perfectly balanced in classifying high risk cases. Overall, this model was highly accurate when classifying and predicting high risk cases.
- Low Risk: Looking at the models sensitivity and recall it was able to correctly classify 84% of low risk cases. With a precision of ~90%, the model misclassified very few low

risk cases. Its F1-score of ~86% shows that its sensitivity and precision is fairly balanced. Overall the model did a good job in classifying low risk cases.
- Mid Risk: The model had a sensitivity of ~86% for mid risk patients. It had a precision of ~79%, meaning it classified most of the cases correctly. Its F1-score of ~86% showed that it had good balance considering this is the most difficult class because mid risk overlaps with high and low features.

**Overall analysis:**

**Table 1.** Overall classification model analysis.

| **Model** | **Accuracy** | **High Risk Recall** | **Low Risk Recall** | **Mid Risk Recall** |
|---|---|---|---|---|
| **Neural Network** | **70.95%** | **100%** | **99%** | **87%** |
| **Random Forest** | **84.36%** | **100%** | **81%** | **86%** |
| **XGBoost** | **86.03%** | **87%** | **84%** | **86%** |

## Conclusion and Discussion

This report used classification models to predict maternal health risk levels. The data was cleaned by checking for missing values and removing outliers from each variable. Exploratory data analysis was performed to get a better understating of the dataset. The classification models utilized to perform this analysis were random forest, XGBoost, and a neural network. The data was split 70/30 using the holdout method and optimized using cross validation.

**Key Finding:**

- The XGBoost model had the best performance overall with an overall accuracy of 86%. It had a strong precision of and recall for all three classes. It had the most success with the mid risk case which were particularly challenging for the other models.
- The random forest model also performed well with an overall accuracy of 84.36%. It was able to predict all the high risk cases at 100%.
- The neural network model had the weakest overall performance of ~70.95%. While it had a 100% recall for high risk cases, it struggled with classifying mid risk case.

From this analysis, the model that would be most useful is the ensemble model with XGBoosting because of not only its high overall accuracy but its ability to handle the mid risk cases better than the other two models.

**Limitation:**

- In the data used in the analysis there were far more low risk cases that high risk or low risk cases. This may have affected the accuracy or balance of the some of the model.
- There are only 6 variables being considered in this dataset and they may not be a full representation of maternal risk.
- It seems like the data is only taken from one visit for each patient. Since a patient's condition can change over time, collecting data for multiple visits per patient would be useful.

**Potential future analysis:**

A possible future analysis could include more variables like the patient's medical history, diet, or financial status. These factors may give a better insight into key indicators for maternal risk. Another possible future analysis could expand the data collection outside of the original data collection location.

References

Ahmed, M. (2020). Maternal Health Risk. [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5DP5D.

World Health Organization. (n.d.). Maternal health. World Health Organization. Retrieved March 30, 2025, from https://www.who.int/health-topics/maternal-health#tab=tab_1

Aadil, M., Khan, S. A. R., Yu, Z., Ali, S. M., & Kumar, A. (2024). Analysis of Internet of Things implementation barriers in the cold supply chain: An integrated ISM-MICMAC and DEMATEL approach. ResearchGate. https://www.researchgate.net/publication/382209123

McGuire, M. (2025). Supervised Learning and Classification [Slides 1-52]. CIS 468: Data Science, Towson University.

McGuire, M. (2025). Bagging, Boosting, and Random Forests Bagging, Boosting, and Random Forests [Slides 1-13]. CIS 468: Data Science, Towson University.

McGuire, M. (2025). Neural Networks[Slides 1-50]. CIS 468: Data Science, Towson University.