The Maternal Health Risk Analysis Project

Zymek Olena Sba 23334
CCT College, Dublin

1. Abstract

One innovative project to improve prenatal care is creating a machine learning model to forecast maternal health risks based on patient data. With the use of cutting-edge computational methoACs, this novel approach analyses a variety of patient data to identify high-risk pregnancies and personalise interventions for better outcomes of both the expecting mother as well as the developing baby.

2. Introduction

This capstone project investigates the intricate interplay of health metrics during pregnancy, focusing on key factors such as diastolic blood pressure (DiastolicBP), blood glucose levels (BS), and heart rate. Through predictive modeling utilizing these vital attributes, the project aims to uncover complex patterns within health metrics and contribute valuable insights into pregnancy risk factors. The goal is to enhance our understanding of maternal health during pregnancy, underscore the importance of navigating health data complexities, and pave the way for future research opportunities in maternal healthcare improvement.

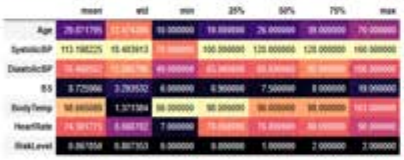

Figure 1. Descriptive Statistics

Figure 2 . Risk Level Pie Chart.

3. Methods & Materials

MLachine algorithms: KNN, LDA, Decision Tree, Random Forest, XGBoost, Naive Bayes, SVM, MLPClassifier, and Logistic Regression are utilised for predictive modeling of MHR.

Hyperparameter tuning and cross-validation.

User interface development.

Python libraries: Pandas, Numpy, Seaborn, Matplotlib, and Scipy ect.

Figure 3. ML Algorithms Chart

4. Results

The evaluation of machine learning algorithms for Maternal Health Risk analysis found XGBoost and Random Forest as top performers. Fine-tuning hyperparameters was crucial for generalizing models to unseten data and reducing overfitting or underfitting. Cross-validation further enhanced performance estimation. Despite Decision Tree unexpectedly outperforming Random Forest and XGBoost in one scenario, it's vital to consider various classifiers for Maternal Health Risk analysis.

Figure 4. Happy Pregnancy.

5. Discussion

Duplicate data can skew analysis results and threaten the integrity of the data set. However, in this instance, the patients may have the same value for certain health factors.

Insights: Preeclampsia and Hypertension are pregnancy complications that were predicted as a high risk at maternity risk levels based on symptoms and evidence found in our dataset. (HSE.ie, n.d.)

Model Performance: The unexpected performance of the Decision Tree algorithm compared to Random Forest in mean cross-validation accuracy, indicating a deeper understanding of the dataset.

Challenges: Balancing model complexity with the need for interpretability through 9 algorithms. Fine-tuning models with cross validation to provide transparent and actionable results.

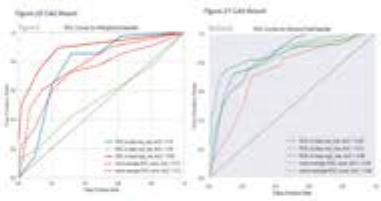

Figure 5. RocCurve CA2 and CA3

6. Conclusions

The Maternal Health Risk Analysis project successfully progressed through various stages, achieving significant milestones and gaining valuable insights.

In terms of unseen test data, XGBoost and Decision Tree models maintained their high accuracy, while Logistic Regression showed a lower accuracy of 54.19%. Overall, XGBoost and Decision Tree models emerged as top performers, showcasing robustness in accuracy across both training and testing datasets.

Positive correlation observed between blood sugar (BS) and age, as well as between systolic and diastolic blood pressure, leading to hypertension (pre-eclampsia and eclampsia).

It is essential to implement a system for continuous monitoring of the model's performance in real-world scenarios and to update the model as needed.

Contact

Zymek Olena
CCT College Dublin
Email: sba23334@student.cct.ie
Phone: +350876003575

References

1. https://www.genigraphics.com/
2. https://coolors.co/