R Project: Statistical Model for Neonatal Weight Predictions

Project Description:

The project, conducted for *Neonatal Health Solutions*, focuses on developing a statistical model capable of accurately predicting neonatal birth weight using clinical variables collected from 2,500 newborns across three hospitals. The model aims to support clinicians in early identification of at-risk pregnancies, optimize hospital resource allocation, and provide actionable insights for improving neonatal health outcomes. The work includes data collection, exploratory analysis, hypothesis testing, regression modelling, model selection, validation, and creation of visualizations to support clinical and strategic decision-making.

Key Responsibilities and Tasks

1. Data Collection and Preparation

- Gathered and organized clinical data from three hospitals, including maternal age, number of pregnancies, smoking habits, gestational duration, neonatal anthropometric measures, delivery type, newborn sex, and birth hospital.
- Cleaned and structured the dataset, identifying missing values, outliers, and inconsistencies.

2. Exploratory Data Analysis (EDA)

- Conducted descriptive statistics to understand variable distributions and detect anomalies.
- Analysed relationships among variables and inspected potential predictors of birth weight.

3. Hypothesis Testing

Performed statistical tests to verify key assumptions, including:

- Differences in caesarean delivery rates among hospitals.
- Comparison of sample neonatal weight and length means with known population values.
- Significant differences in anthropometric measurements between male and female newborns.

4. Predictive Modelling

- Built a multiple linear regression model including all clinically relevant variables.
- Interpreted coefficients to quantify the impact of factors such as maternal smoking and gestational weeks on birth weight.
- Investigated interaction effects and non-linear patterns.

5. Model Optimization and Validation

- Applied model selection criteria (AIC, BIC) to identify the most parsimonious model.
- Evaluated model accuracy using R^2, RMSE, residual diagnostics, and influence analysis.
- Ensured robustness by assessing influential observations and correcting model deviations when necessary.

6. Insights and Forecasting

- Generated predictive scenarios (e.g., estimating birth weight for specific maternal profiles).
- Identified key risk factors that negatively influence neonatal birth weight.

7. Visualization and Communication

- Created clear and intuitive visualizations to illustrate variable relationships, model performance, and predicted outcomes.
- Summarized results for clinical and strategic stakeholders.

Outcome

- Development of a validated predictive model capable of estimating neonatal birth weight with high accuracy.
- Identification of the strongest predictors, including gestational duration and maternal smoking status.
- Insight into hospital-specific differences in delivery practices and neonatal outcomes.
- Enhanced ability for clinicians to anticipate complications such as low birth weight or prematurity.
- Support for strategic planning to improve neonatal care protocols, allocate resources more efficiently, and reduce perinatal risks.
- Visual and analytical outputs usable for reporting, research, and decision-making.

Technologies and Tools Used

R Programming Language