Healthcare Analytics Using Machine Learning

- Analyzing Patient Data for Better Healthcare Insights

Name: CH. ANANYA KUMAR

Date: 15th November, 2024

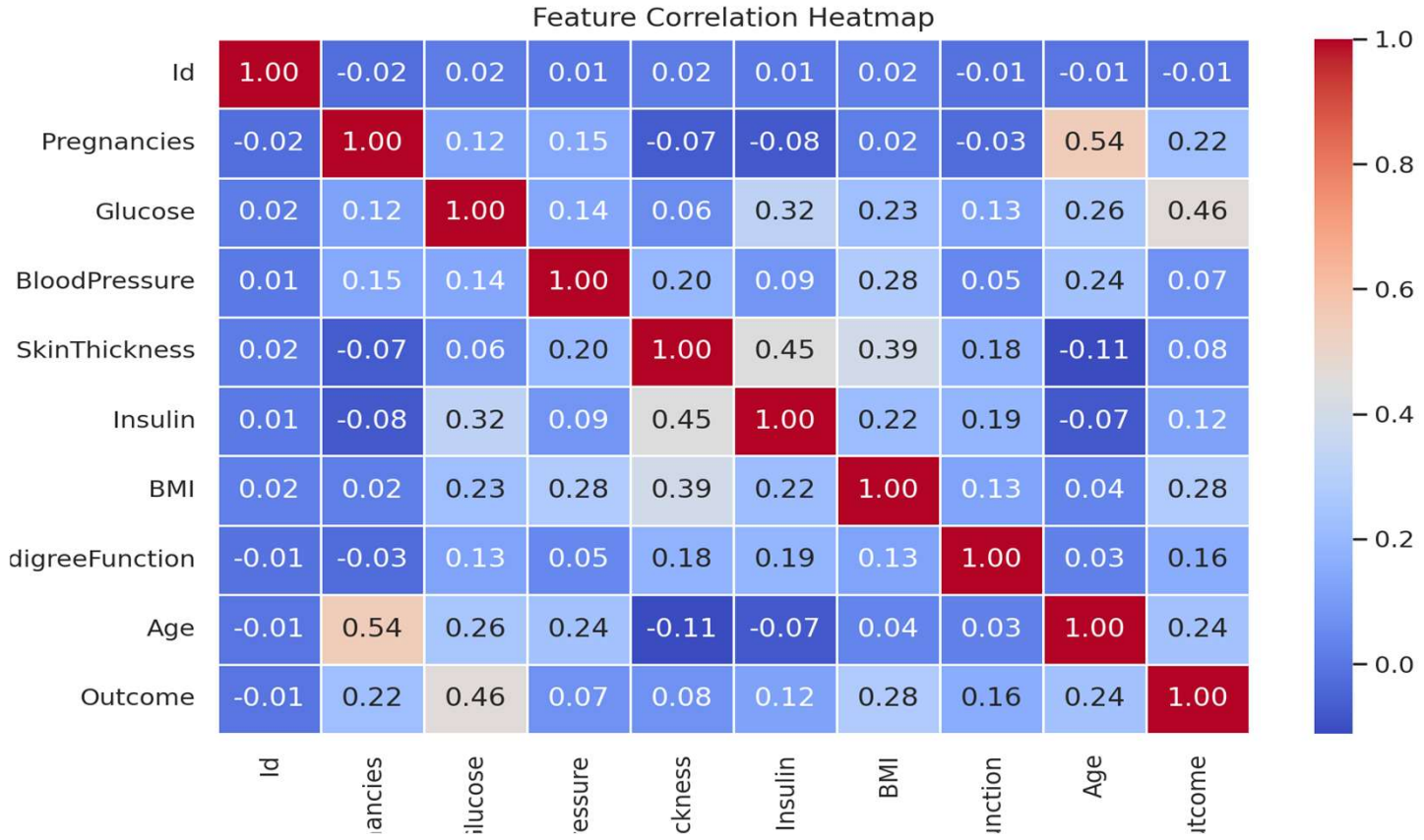
Introduction

- Importance of healthcare data analysis in predicting pregnancy outcomes.
- How machine learning enhances diagnosis and treatment.
- Overview of data-driven improvements in patient care.

Exploratory Data Analysis (EDA)

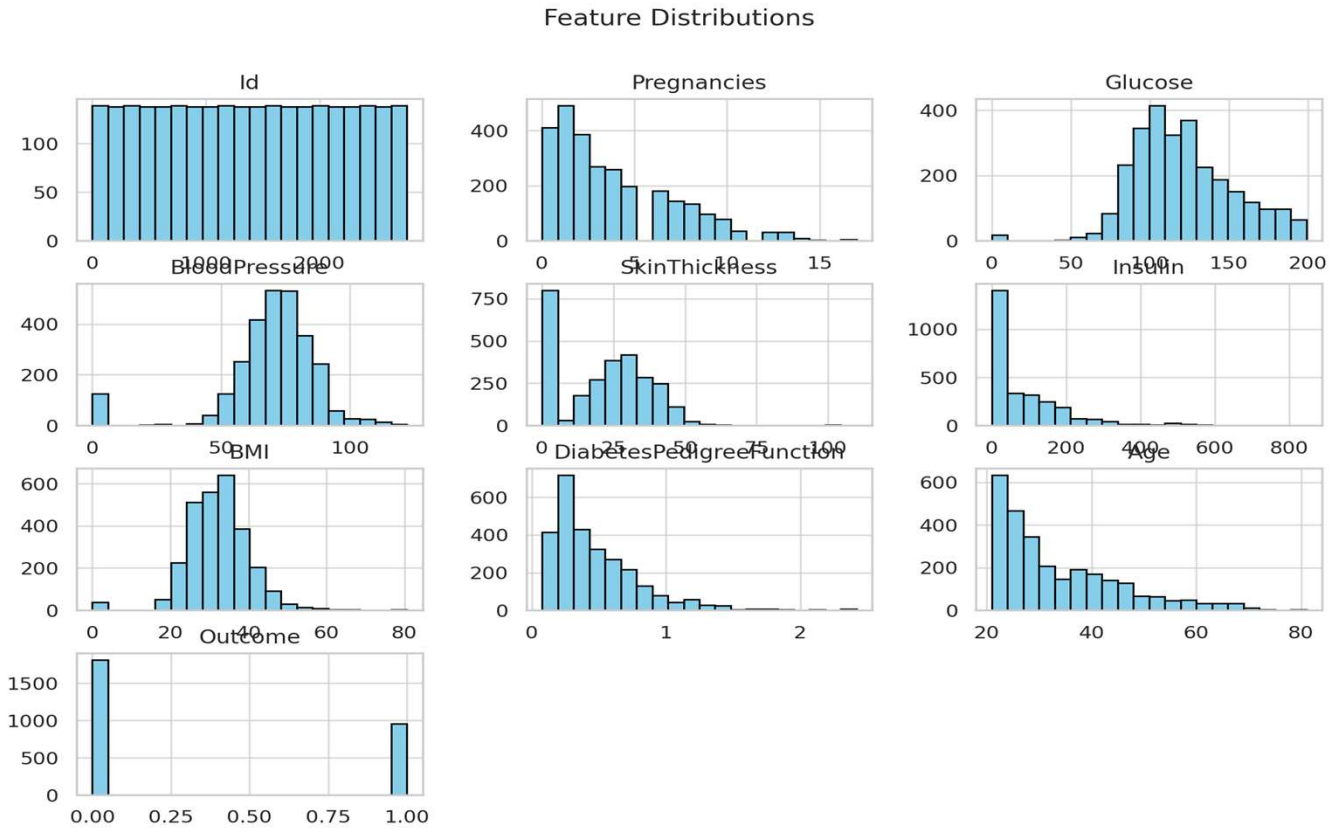
- EDA helps understand data distribution, patterns, and relationships.
- Various visualizations like heatmaps, histograms, and scatter plots help identify trends.

Heatmap



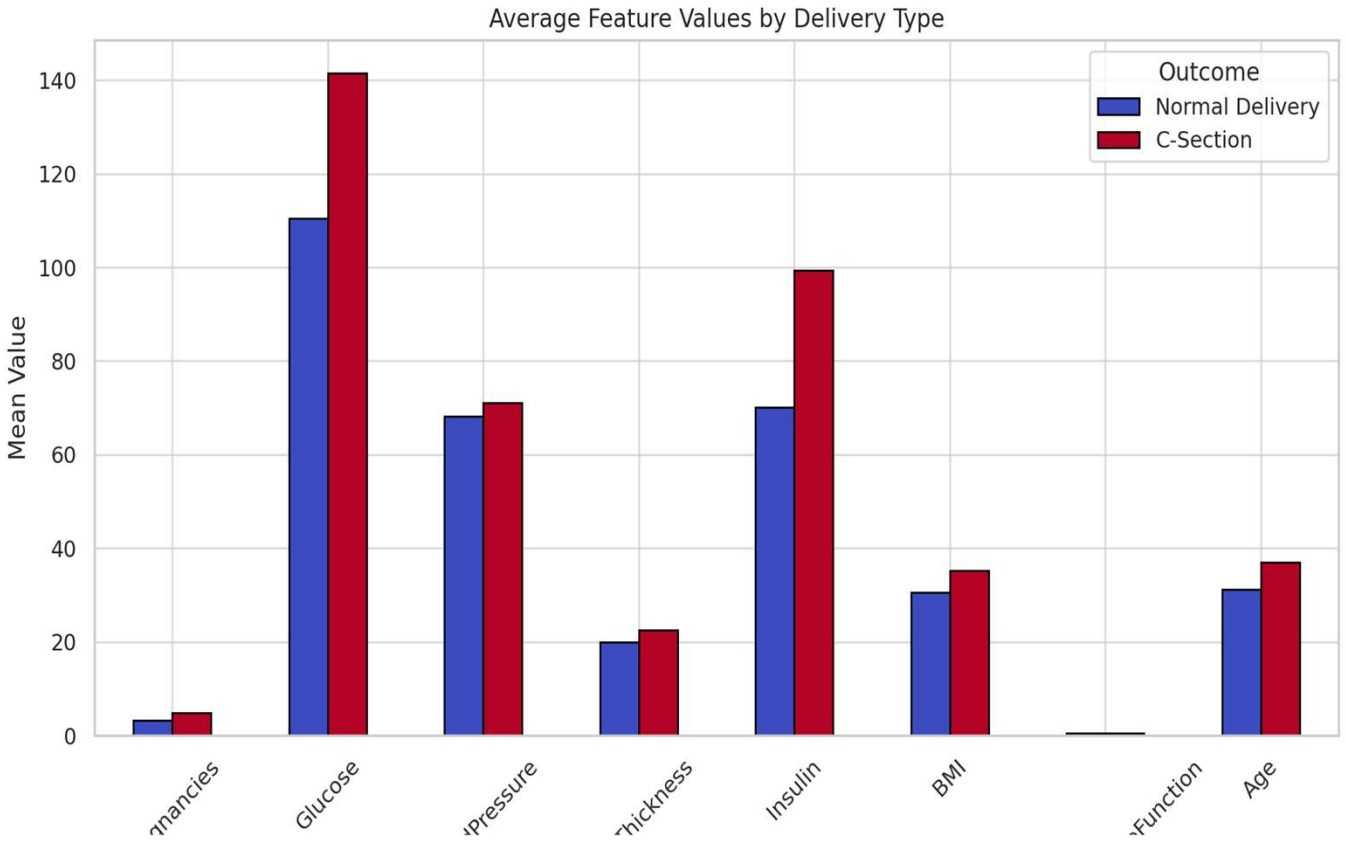
The heatmap visualizes the correlation between medical features. Strong correlations can help identify redundant or important variables for prediction.

Histogram



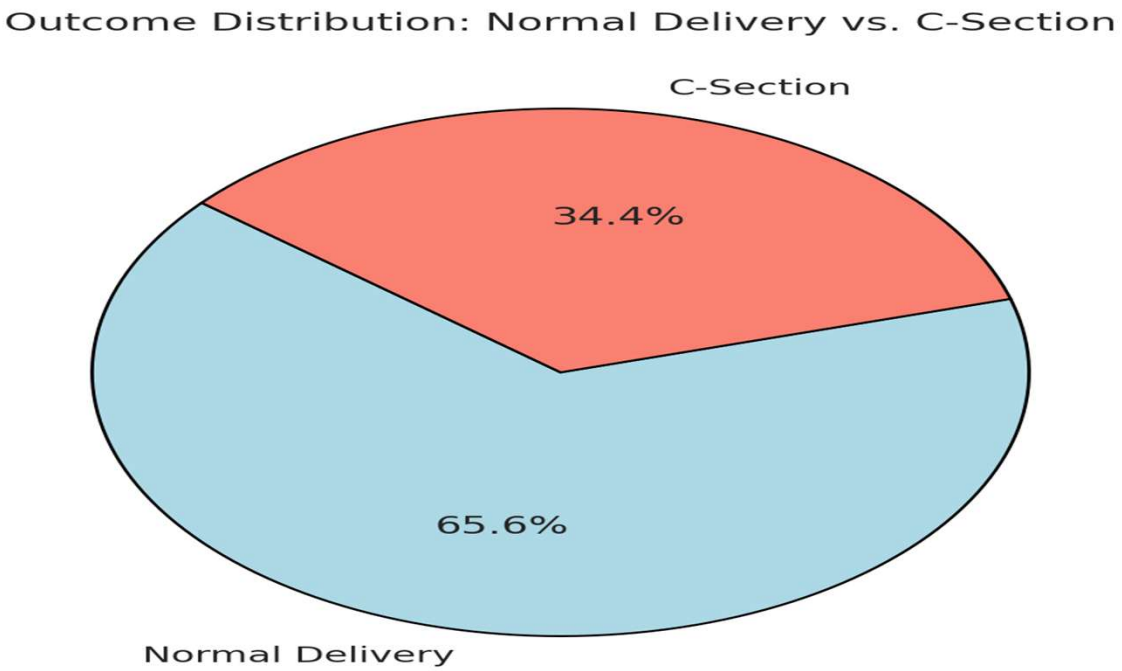
Histograms show the distribution of different numerical variables, revealing patterns such as skewness and outliers.

Bar Plot



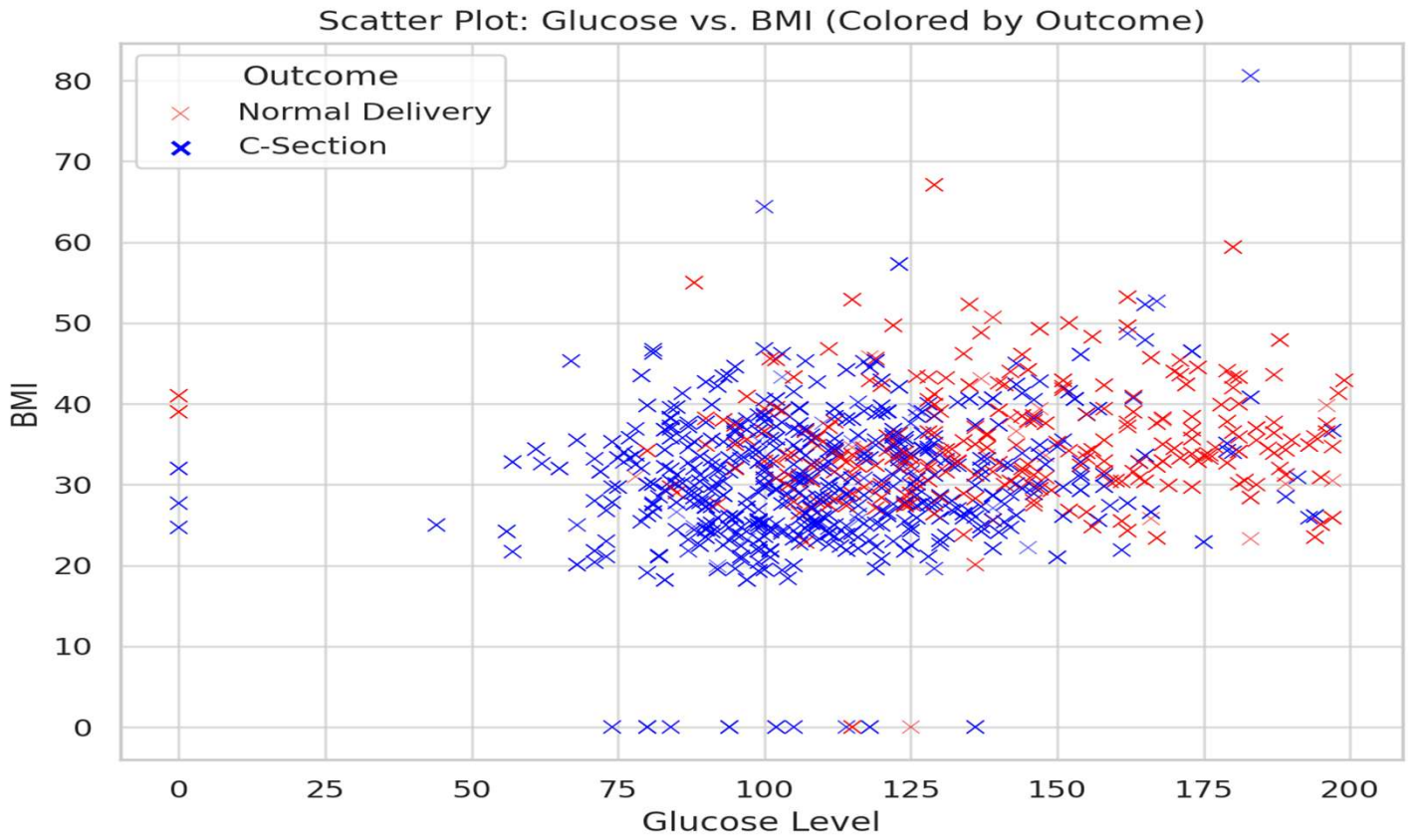
Bar plots display average feature values grouped by delivery type, highlighting important differences.

Pie Chart



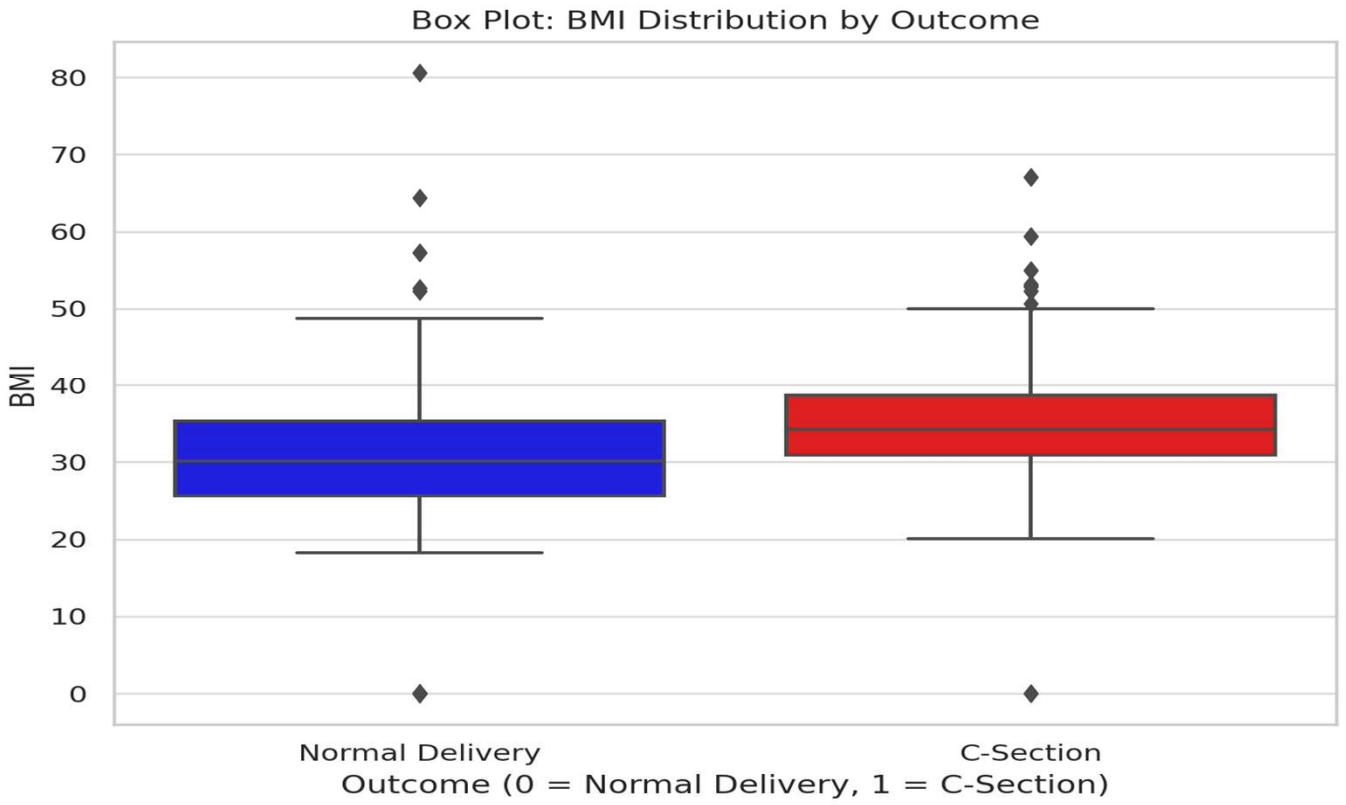
The pie chart illustrates the proportion of Normal Deliveries vs. C-Sections, indicating possible class imbalance.

Scatter Plot



Scatter plots reveal relationships between two numerical features. Here, we observe Glucose vs. BMI with different outcomes.

Box Plot



Box plots show the spread and median of BMI for different delivery types, highlighting possible outliers.

Data Preprocessing

- Handling missing values and inconsistencies.
- Encoding categorical variables.
- Feature scaling using StandardScaler/MinMaxScaler.
- Addressing class imbalance using SMOTE or undersampling.

Feature Selection

- Identifying relevant features using statistical methods.
- Removing redundant features to improve efficiency.
- Feature importance visualization (Decision Trees, SHAP values).

Model Selection

- Logistic Regression: Baseline model, interpretable.
- Decision Tree: Handles non-linearity.
- Random Forest: High accuracy, robust.
- SVM: Effective for complex data relationships.
- Preprocessing: Feature scaling, handling class imbalance.

Model Evaluation Metrics

- Logistic Regression: Accuracy=0.7708, Precision=0.7652, Recall=0.7708, F1-Score=0.7616
- Random Forest: Accuracy=0.9964, Precision=0.9964, Recall=0.9964, F1-Score=0.9964
- SVM: Accuracy=0.8231, Precision=0.8245, Recall=0.8231, F1-Score=0.8151
- Decision Tree: Accuracy=0.9765, Precision=0.9768, Recall=0.9765, F1-Score=0.9766
- Best Model: Random Forest with Score: 0.9964

Results & Insights

- Summary of trained model findings.
- Key patterns in pregnancy outcomes.
- Best Model: Random Forest due to highest accuracy and balanced performance.
- Limitations and improvements.
- Future work: Deep learning, ensemble methods.

Conclusion

- Machine Learning can effectively predict pregnancy outcomes.
- Random Forest emerged as the best-performing model.
- Future improvements can focus on deep learning and larger datasets.
- Continuous improvements in healthcare analytics.
- Thank you!