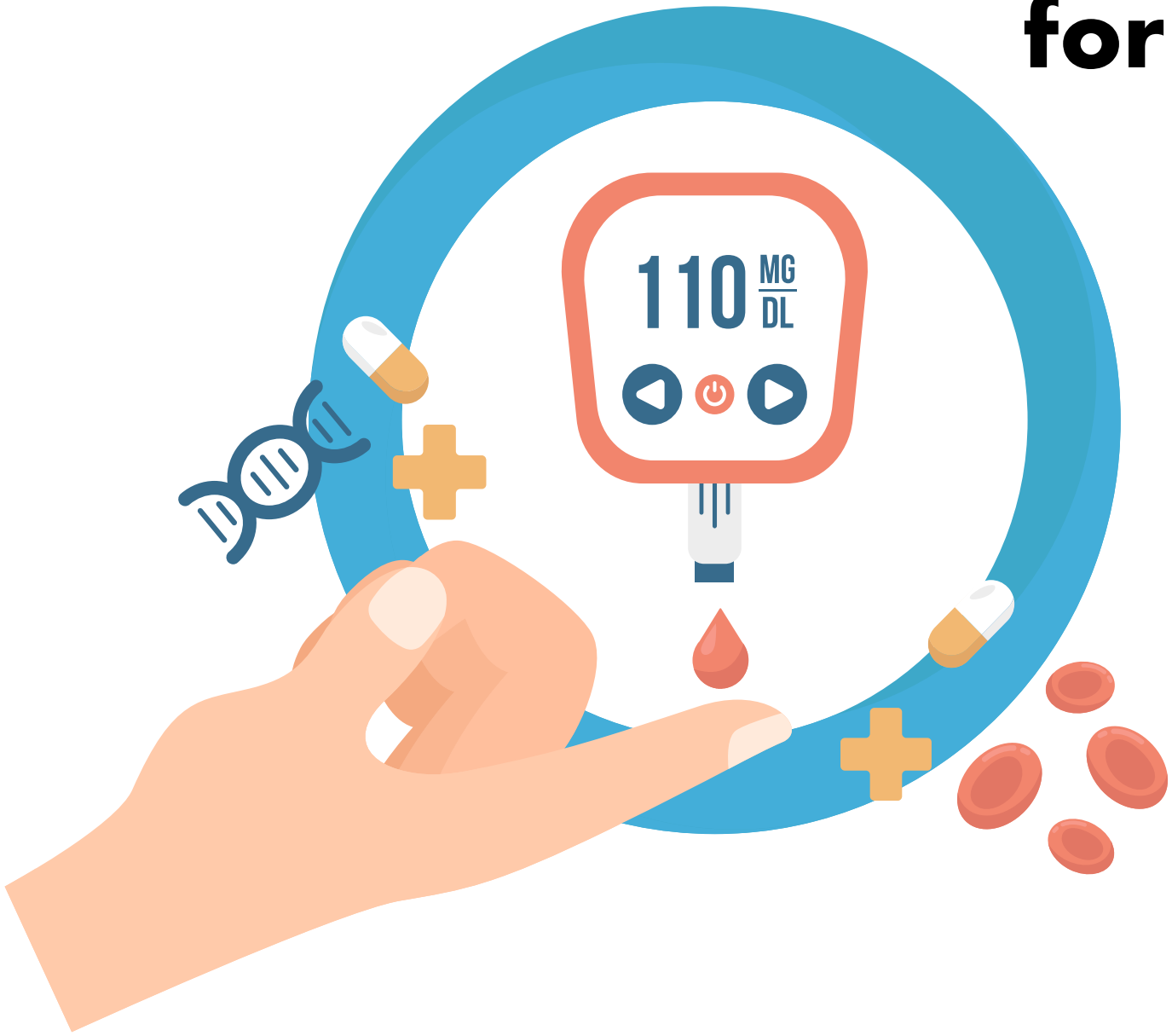
Predictive Analytics for Diabetes Detection



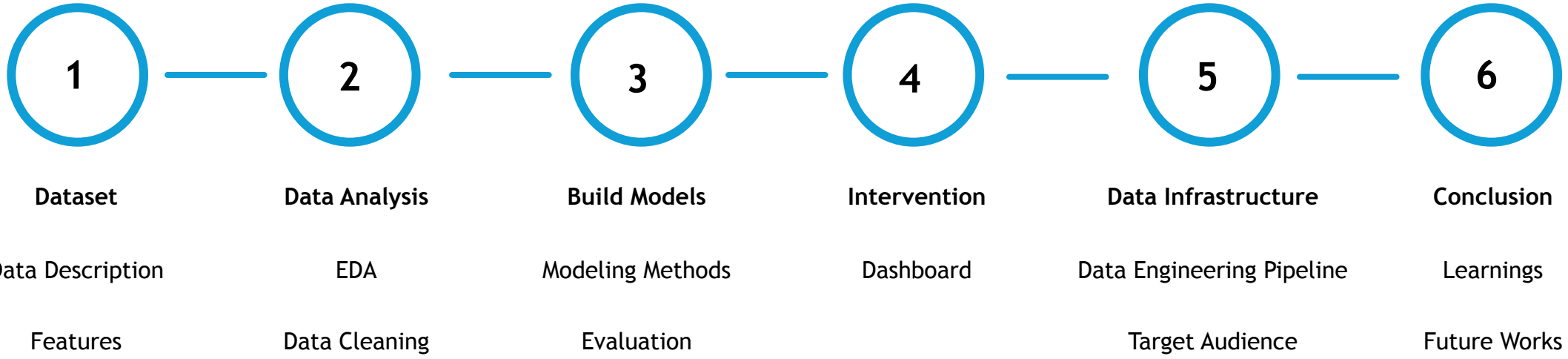
By Group 10

Suraj Nihal, Pushyap
Navdeep, Suthar

Problem Statement

Developing a predictive model to accurately identify individuals at high risk of diabetes within the Pima Indian community, using advanced machine learning techniques to facilitate early intervention and support CDC's HI-5 health initiative.

Project Overview

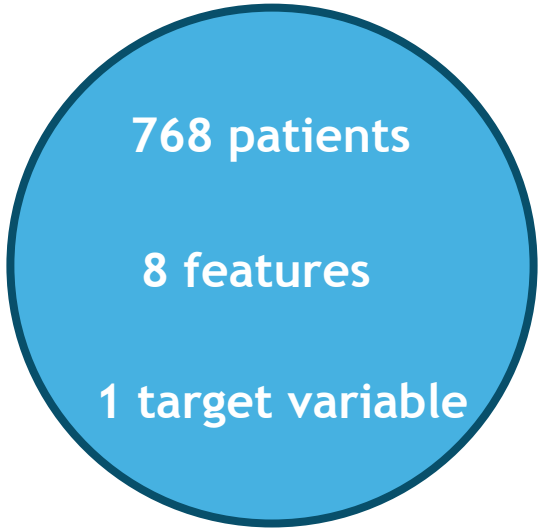


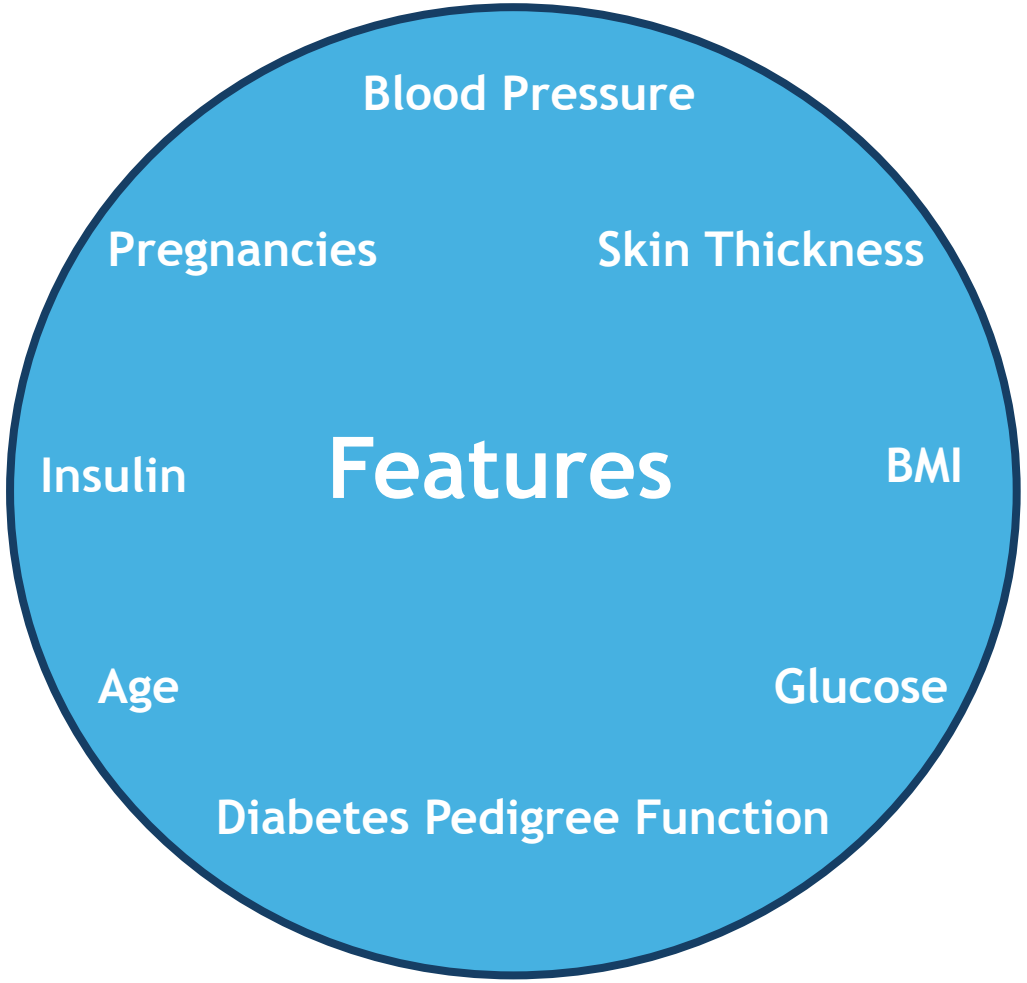
Pima Diabetes Dataset

Hosted on Kaggle from the National Institute of Diabetes and Digestive and Kidney Diseases.

The dataset comprises data from female patients of Pima Indian heritage, aged 21 and above.

Features Overview

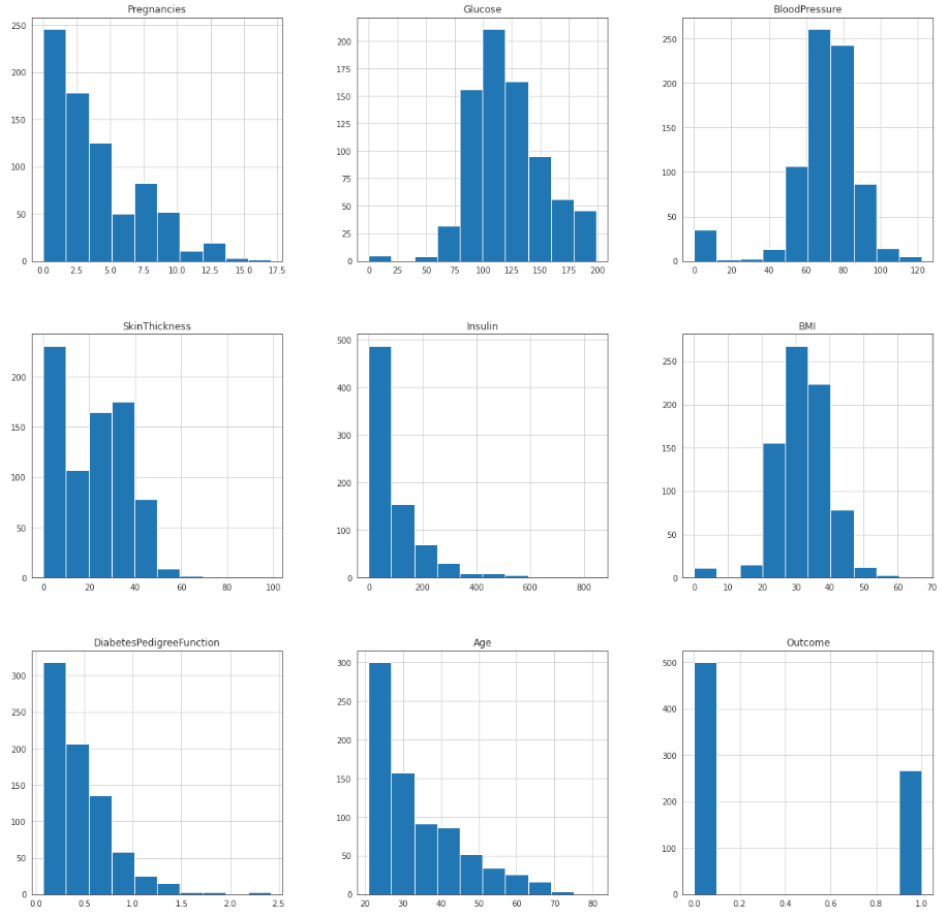




Exploratory Data Analysis

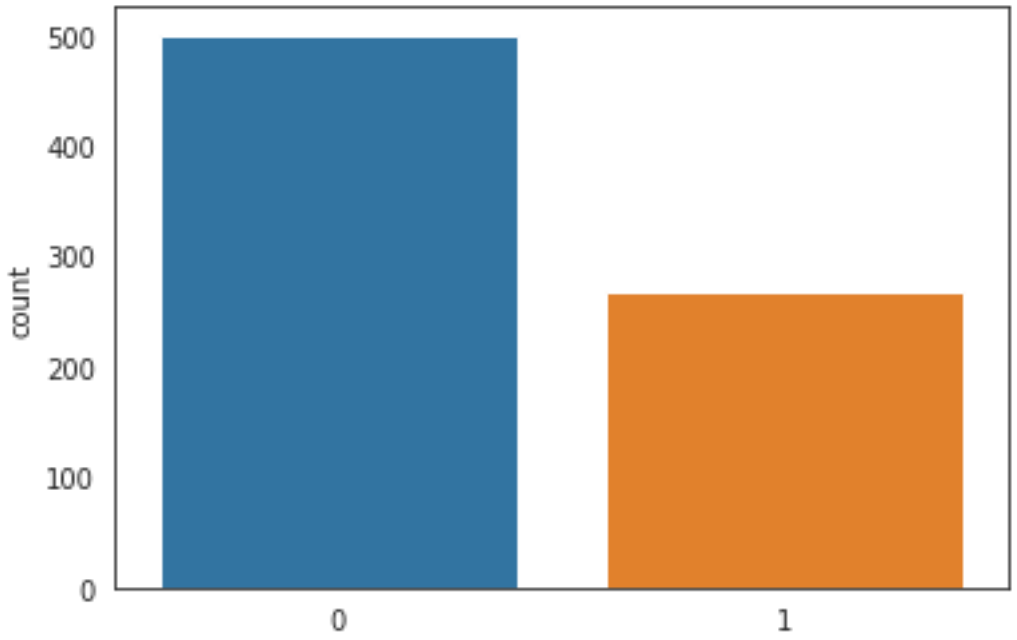
Feature	Missing Value (%)
Glucose	0.65
BloodPressure	4.56
SkinThickness	29.56
Insulin	48.70
BMI	1.43

Distribution of Diabetes Predictors

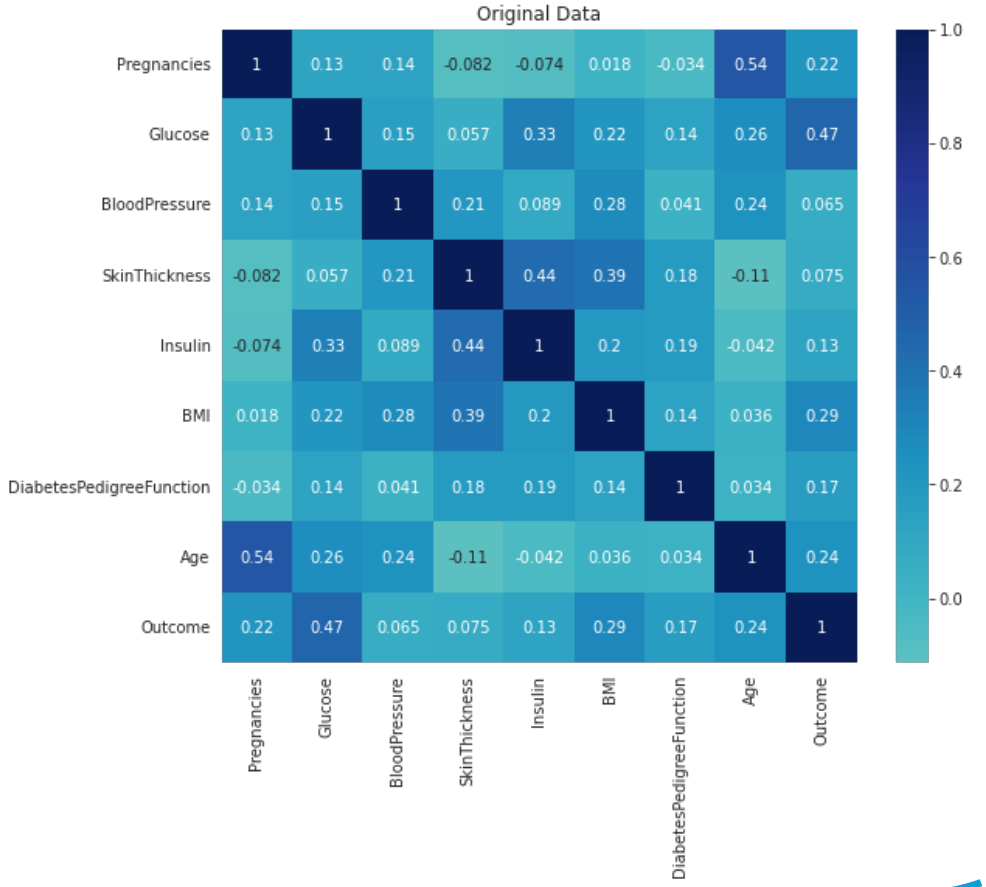


Exploratory Data Analysis

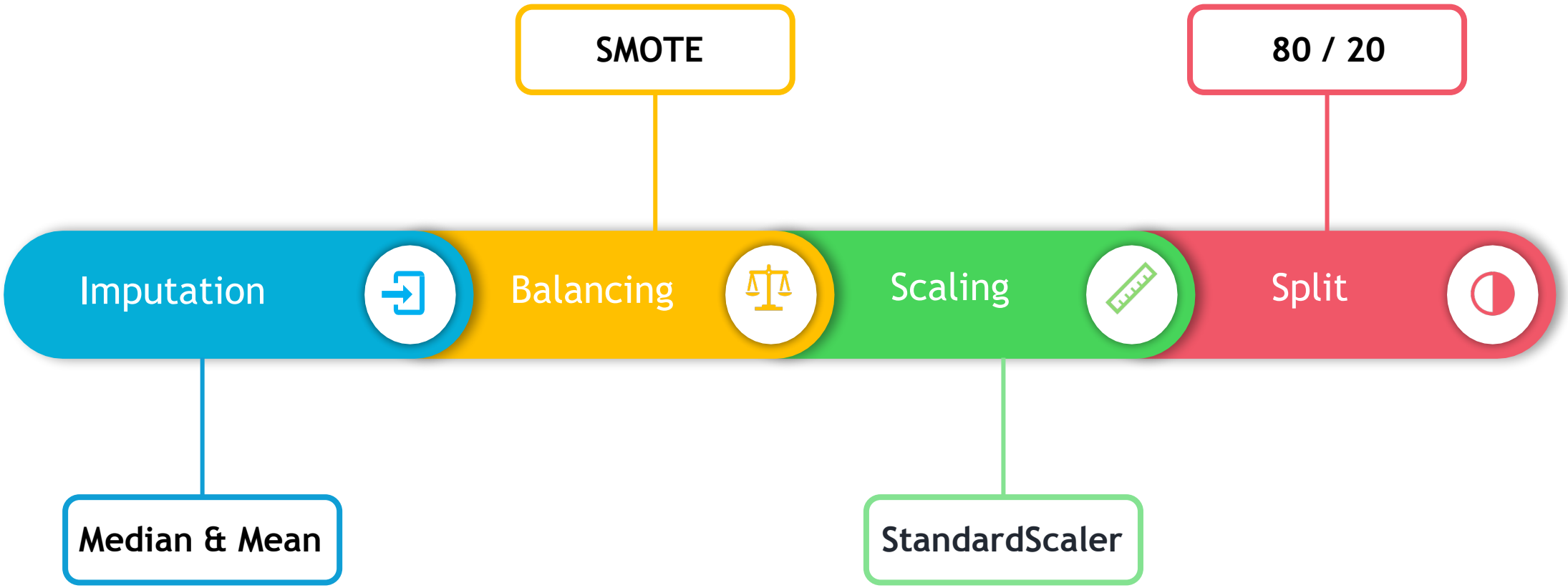
Diabetic vs. Non-Diabetic Cases



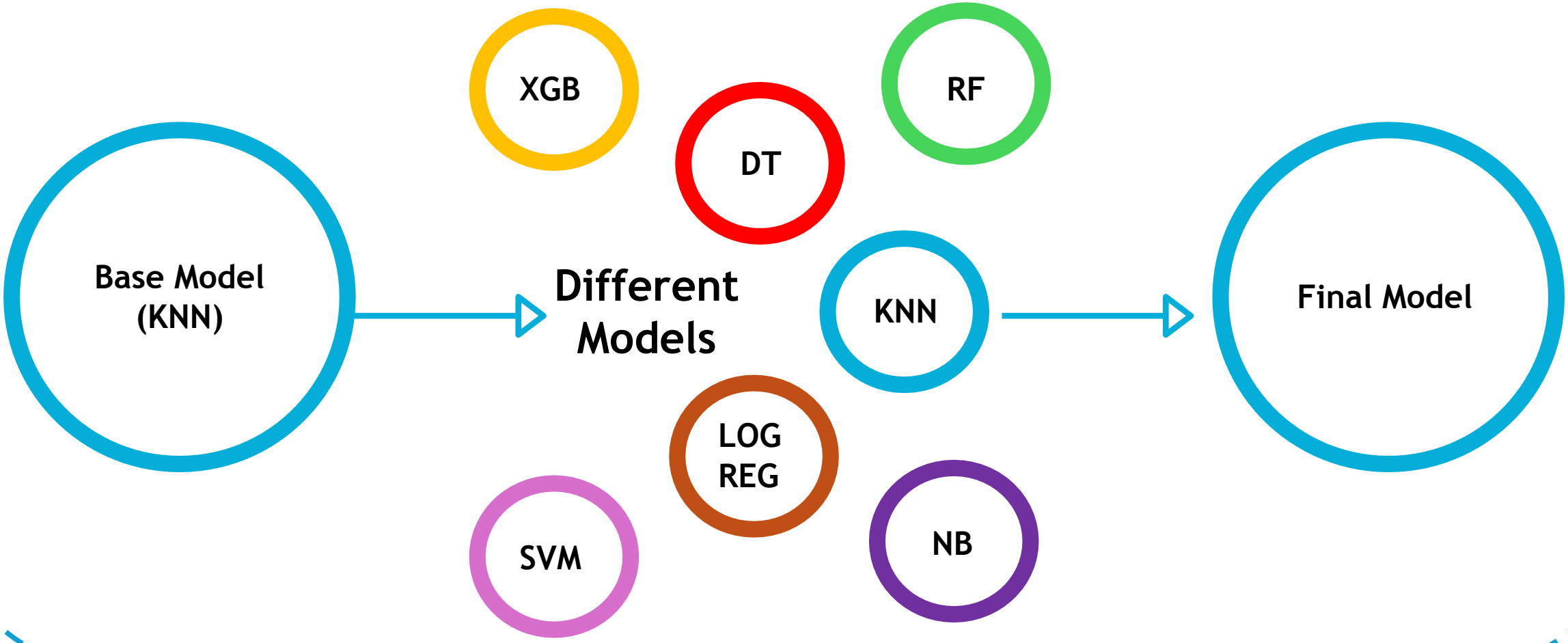
Heatmap of Variable Correlations



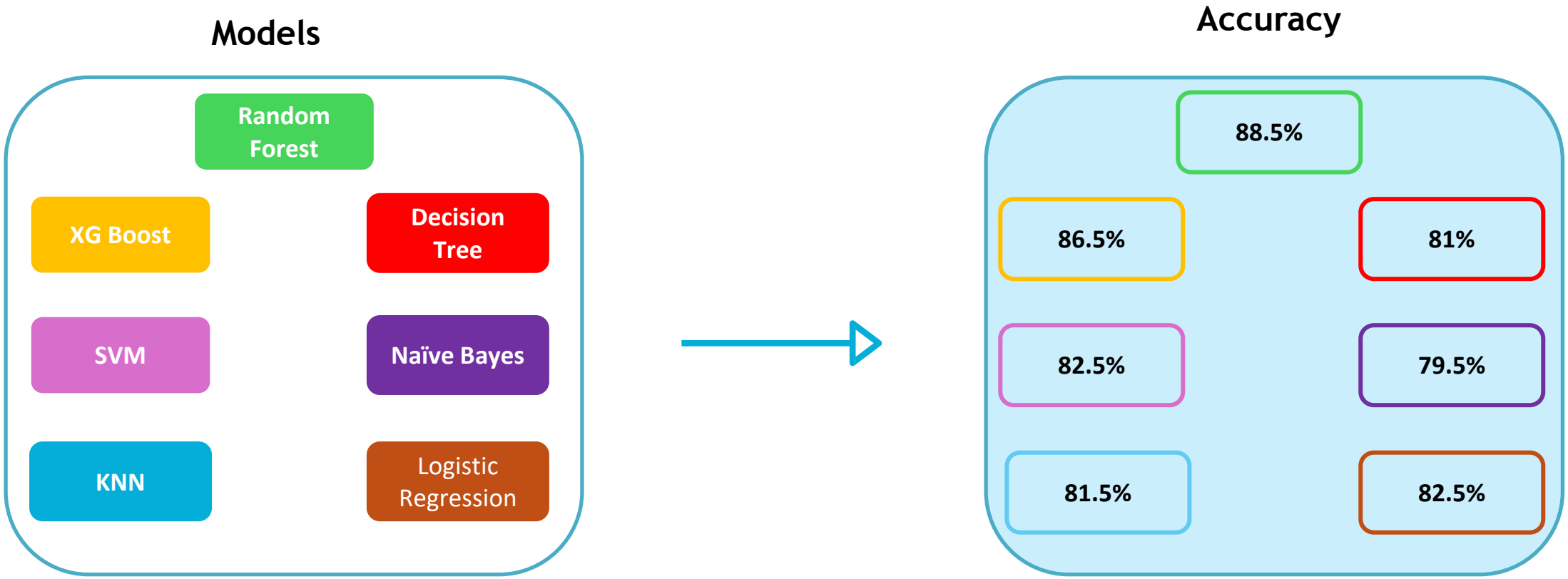
Data Cleaning



Modeling Method

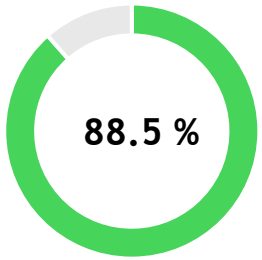


Model Evaluation

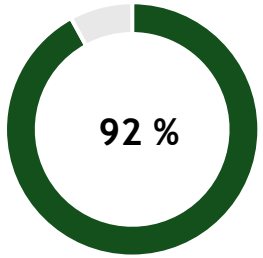


Models Selected

Random Forest

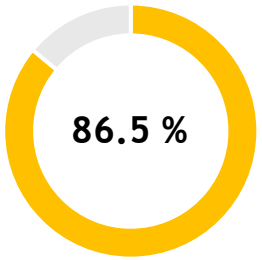


Accuracy

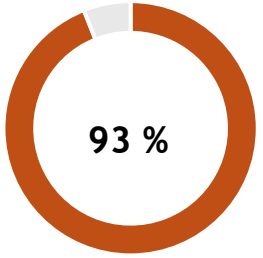


Recall

XG Boost

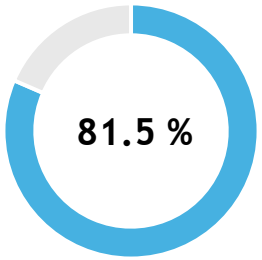


Accuracy

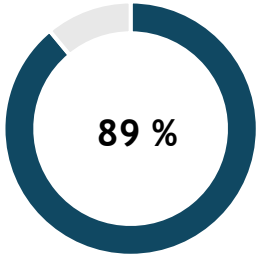


Recall

KNN



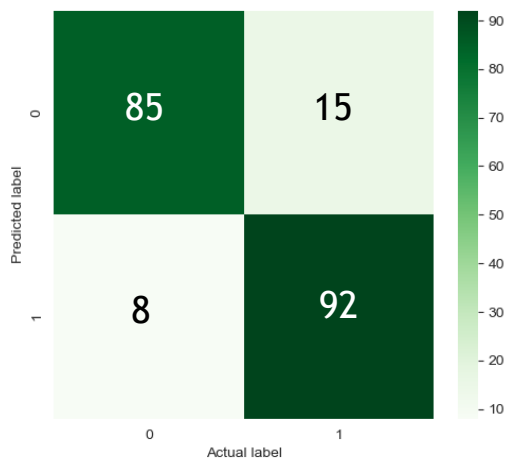
Accuracy



Recall

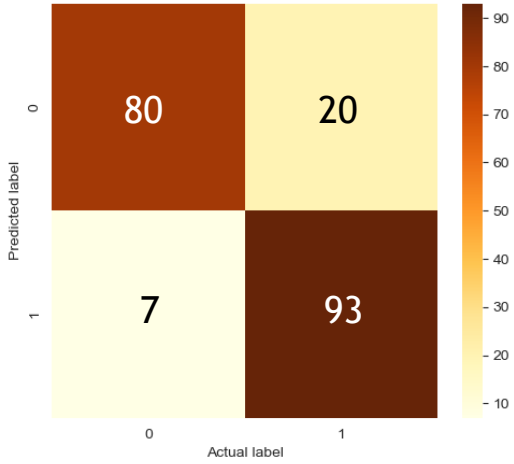
Evaluation Metrics

Random Forest



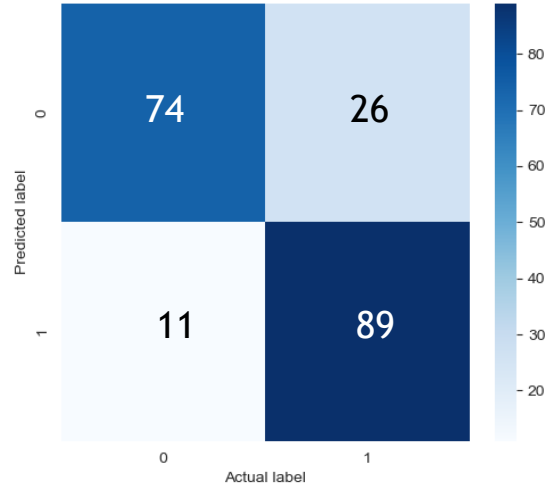
Metric	Score
Sensitivity	92%
Specificity	85%
AUC	94%

XG Boost



Metric	Score
Sensitivity	93%
Specificity	80%
AUC	92%

KNN

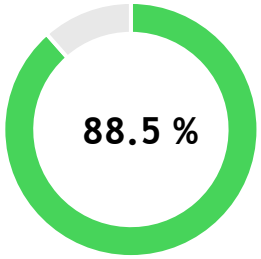


Metric	Score
Sensitivity	89%
Specificity	74%
AUC	89%

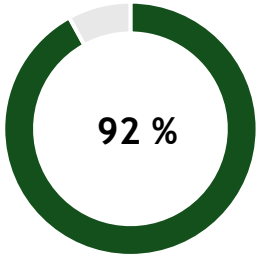
Random Forest

Best Model

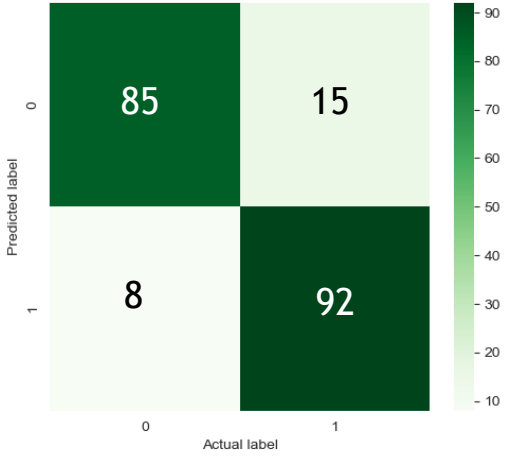
Accuracy



Recall

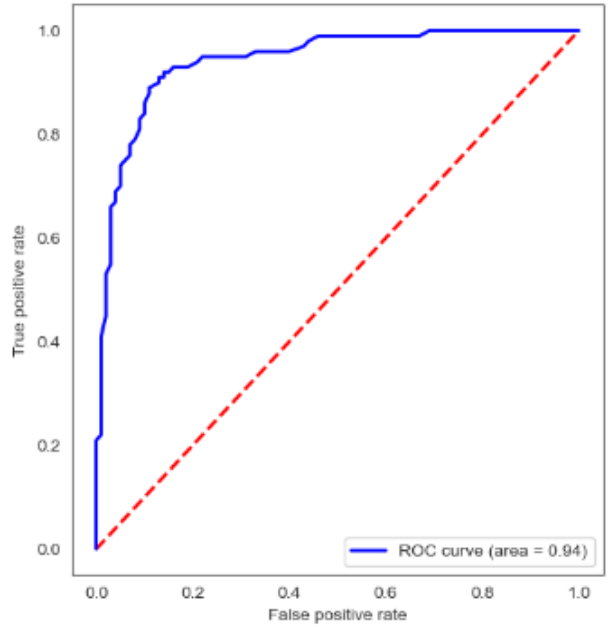


Confusion Matrix



Metric	Score
Sensitivity	92%
Specificity	85%
AUC	94%

ROC Curve



Metric	Score
AUC	94%

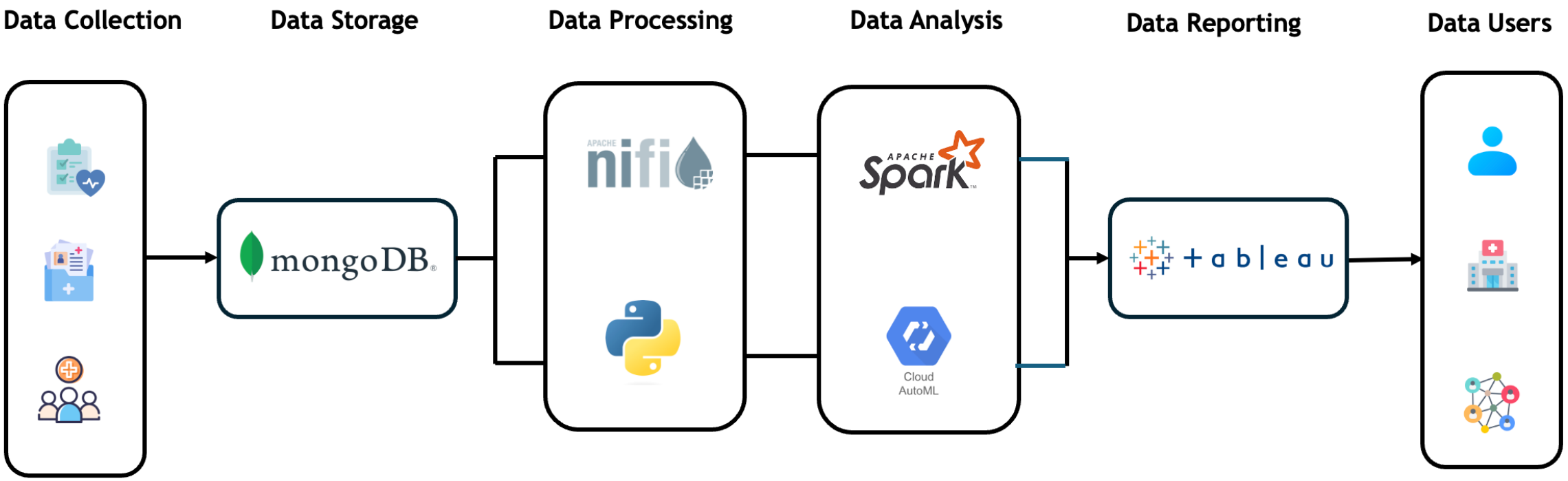
Intervention

- Implement personalized health programs targeting high-risk individuals identified through our model, incorporating lifestyle changes and preventive care.
- Utilize a Tableau dashboard to track intervention effectiveness, monitoring engagement rates and health outcomes, with interactive features for real-time data analysis.

Monitoring

- Displays trends and relationships through plots and a pie chart for quick insights.
- Highlights key metric relationships to guide predictive modeling.
- Allows customized exploration with dynamic feature selection.

Data Infrastructure



Target Audience

High-Risk Individuals

Healthcare Providers

Local Community Organizations

Data Science Learnings

Modeling Proficiency

Data Engineering

Real-World Application

Collaboration

Challenges

Ethics

Healthcare Implementation

- **Enhanced Patient Outcomes**: Utilize predictive analytics to enable healthcare providers to identify and intervene with high-risk patients earlier, improving patient management and outcomes.
- **Alignment with CDC's HI-5 Initiative**: Project supports the CDC's HI-5 model by advancing community-wide health interventions that focus on prevention and have proven health impacts and cost-effectiveness.
- **Dashboard for Real-Time Intervention Monitoring**: Deploy an interactive Tableau dashboard that allows healthcare professionals and community organizations to monitor, evaluate, and enhance the effectiveness of health interventions in real-time.

Future Works

- **Advanced Model Development**: Enhance accuracy with advanced algorithms and integrate more diverse datasets to improve robustness.
- **Healthcare System Integration**: Develop capabilities for seamless integration with electronic health records (EHR) to facilitate real-time patient risk identification and notifications.
- **Expansion to Other Conditions**: Extend the predictive model to cover additional chronic conditions, leveraging existing model infrastructure.
- **Global Application**: Adapt the model for use in low-resource settings, aligning with global health initiatives to broaden impact.

Together, we can beat diabetes.

Eat Healthy

Exercise Regularly

Maintain a Healthy Weight

Thank You