

DIABETES RISK PREDICTION

Leveraging Advance Supervised Machine Learning Models to
Empower Early Detection For Diabetes Risk.



Table of Contents

- 01. Introduction
- 02. Problem Statement
- 03. Project Objective
- 04. Methodology Overview
- 05. Key Insights from DA
- 06. Model Performance
- 07. Recommendation
- 08. Conclusion
- 09. Q & A

INTRODUCTION

Key Points:

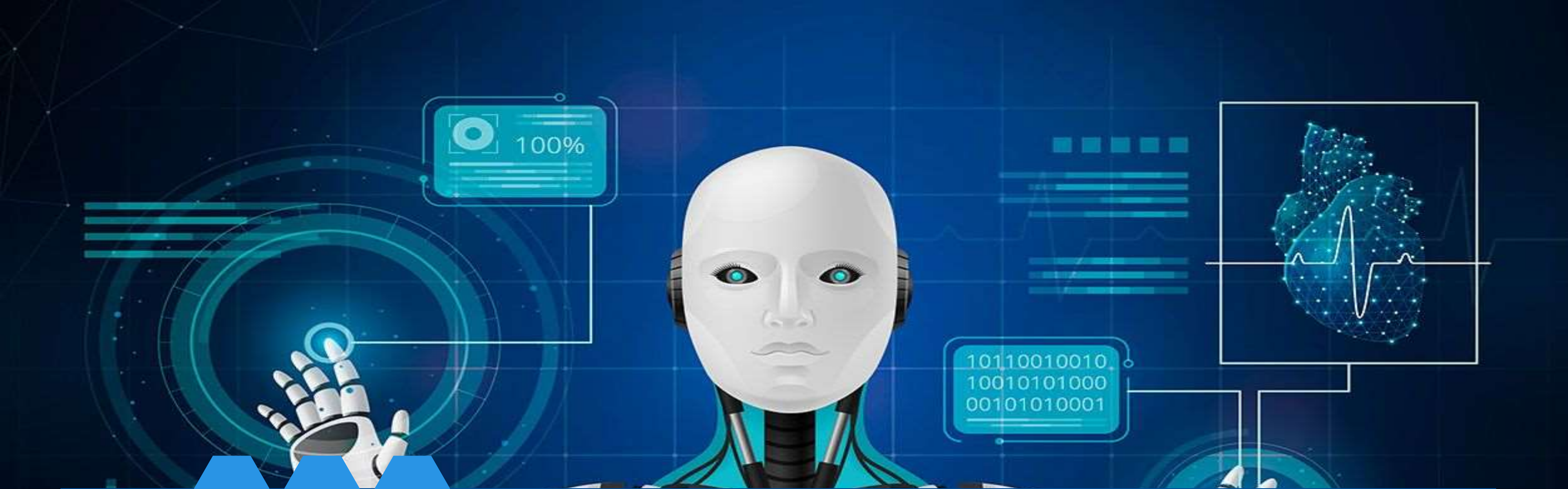
- Stark Health Clinic is committed to improving patient outcomes using technology and predictive modelling.
- Diabetes poses significant health risks and financial challenges.
- This project aims to develop a robust machine learning model to predict diabetes risk and enable early interventions.



PROBLEM STATEMENT

Challenges:

- Lack of precision in current diabetes detection methods.
- Missed opportunities for timely interventions.
- High healthcare costs due to late-stage diabetes management.



PROJECT OBJECTIVE

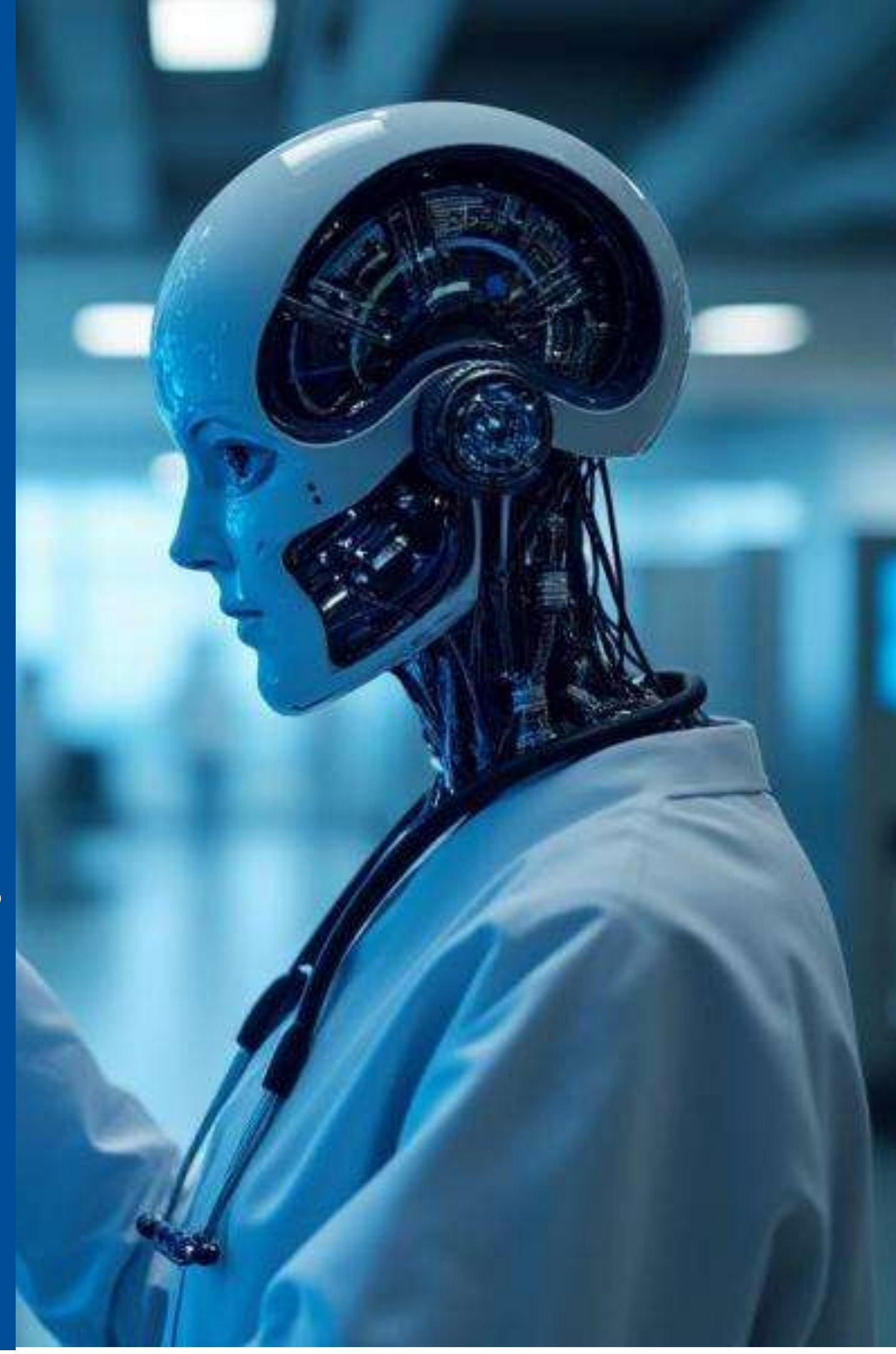
Goals:

- Develop an accurate model to predict diabetes onset.
- Enable timely intervention and targeted preventive measures.
- Reduce healthcare costs and enhance patient outcomes.

METHODOLOGY

Steps:

1. **Data Cleaning:** Address missing values, remove anomalies.
2. **EDA:** Analyze distributions, correlations, and trends.
3. **Feature Engineering:** Encode categorical variables and scale numerical features.
4. **Model Training:** Train and compare Logistic Regression, Decision Tree, SGD, and Random Forest models.
5. **Model Evaluation:** Use metrics like accuracy, precision, recall, and F1-score.
6. **Optimization:** Fine-tune the best model for improved performance.



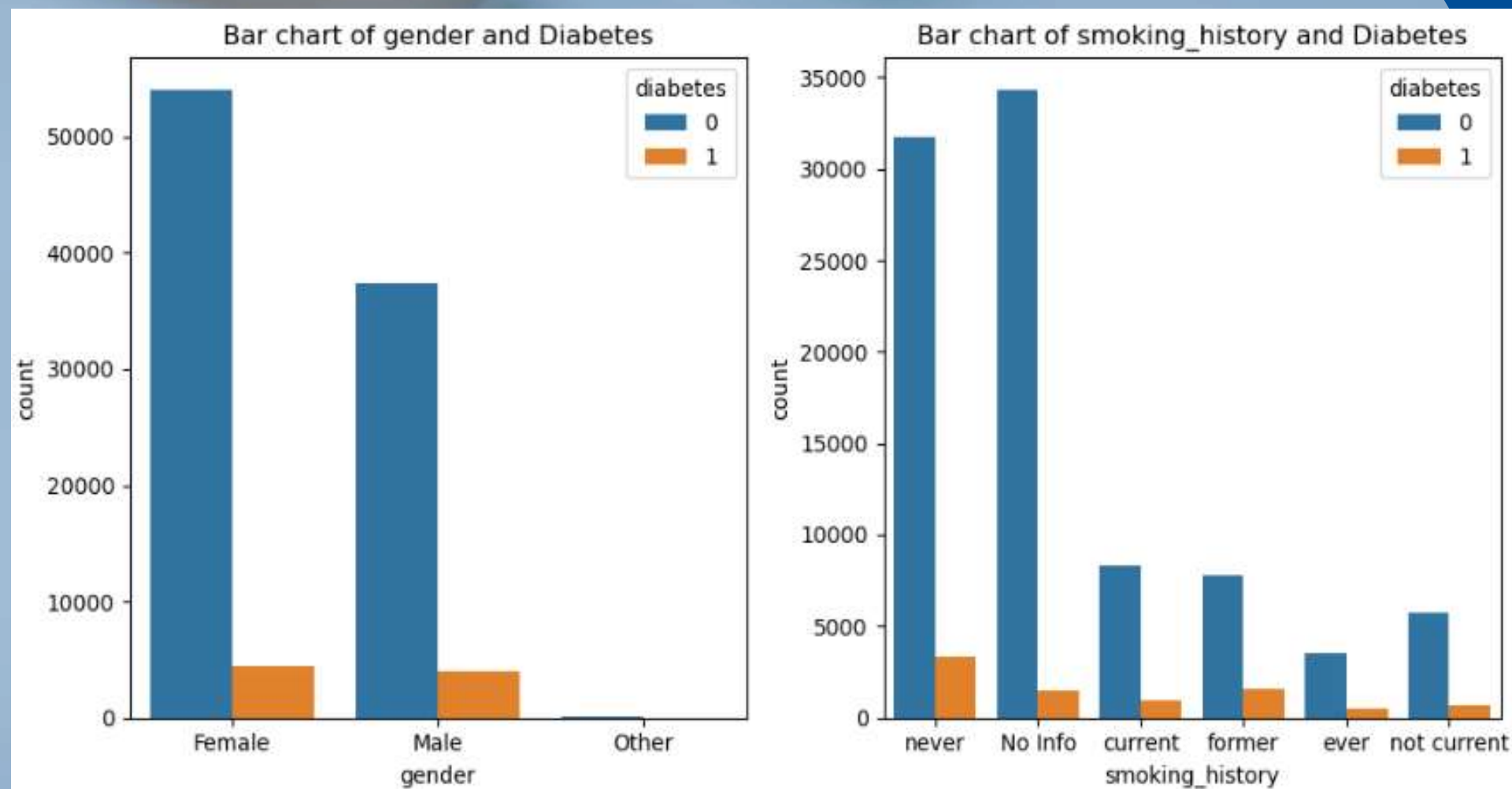
KEY INSIGHTS

Numerical Data

- Age distribution: Standard with no outliers.
- BMI: Right-skewed distribution without outliers.
- Strongest correlation: Blood Glucose Level vs. Diabetes (0.42).

Categorical Data

- Higher prevalence of diabetes among males.
- Former smokers have the highest diabetes risk.
- Patients over 50 are predominantly former smokers.



MODEL PERFORMANCE

Evaluation Metrics:

- Average performance across all 4 models trained: 96% (accuracy, precision, recall, F1-score).

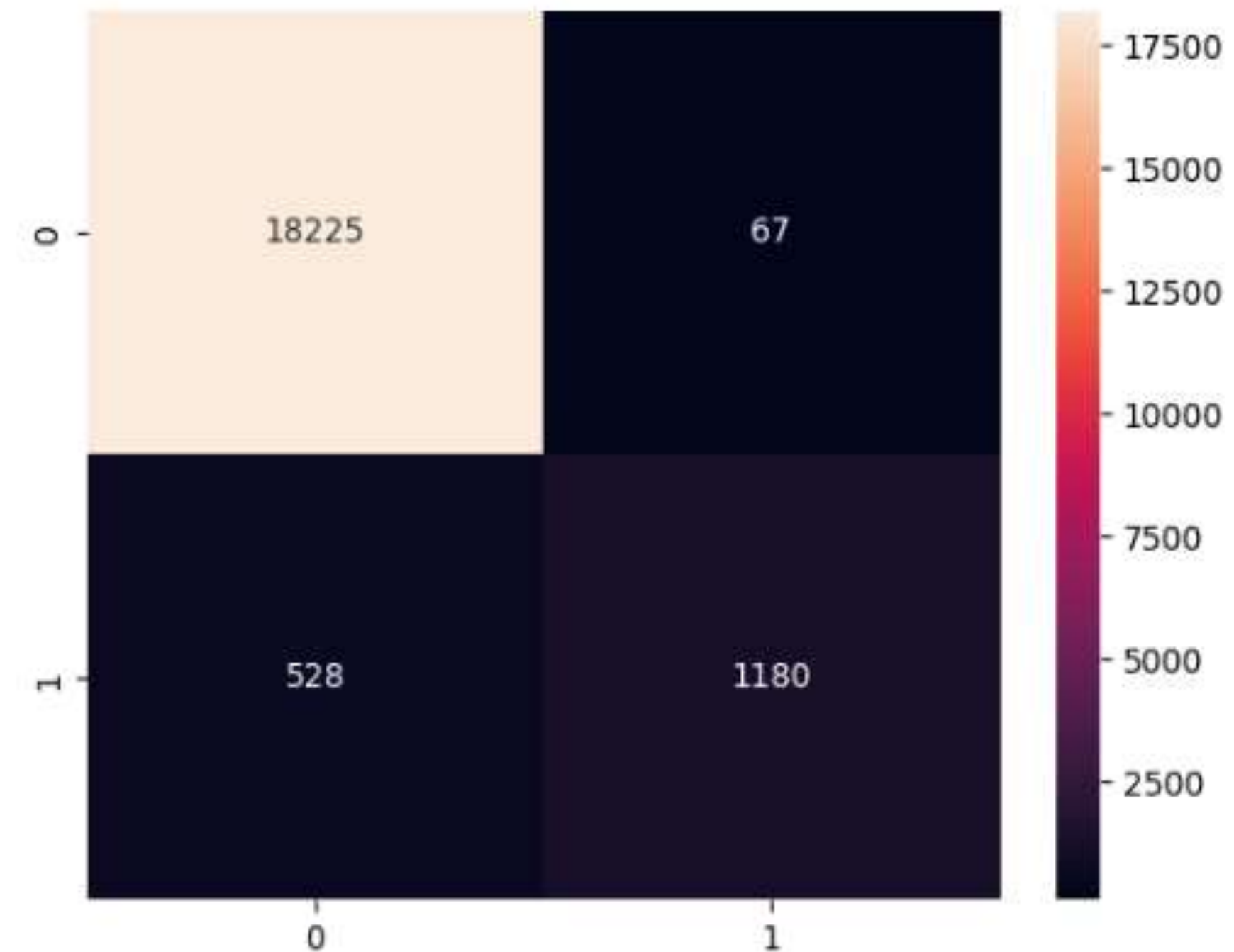
Best Model - Random Forest:

- Recall: 100%
- Precision: 97%
- F1 Score: 98%
- Accuracy: 97%

```
<-- Random Forest -->
Classification Report:

```

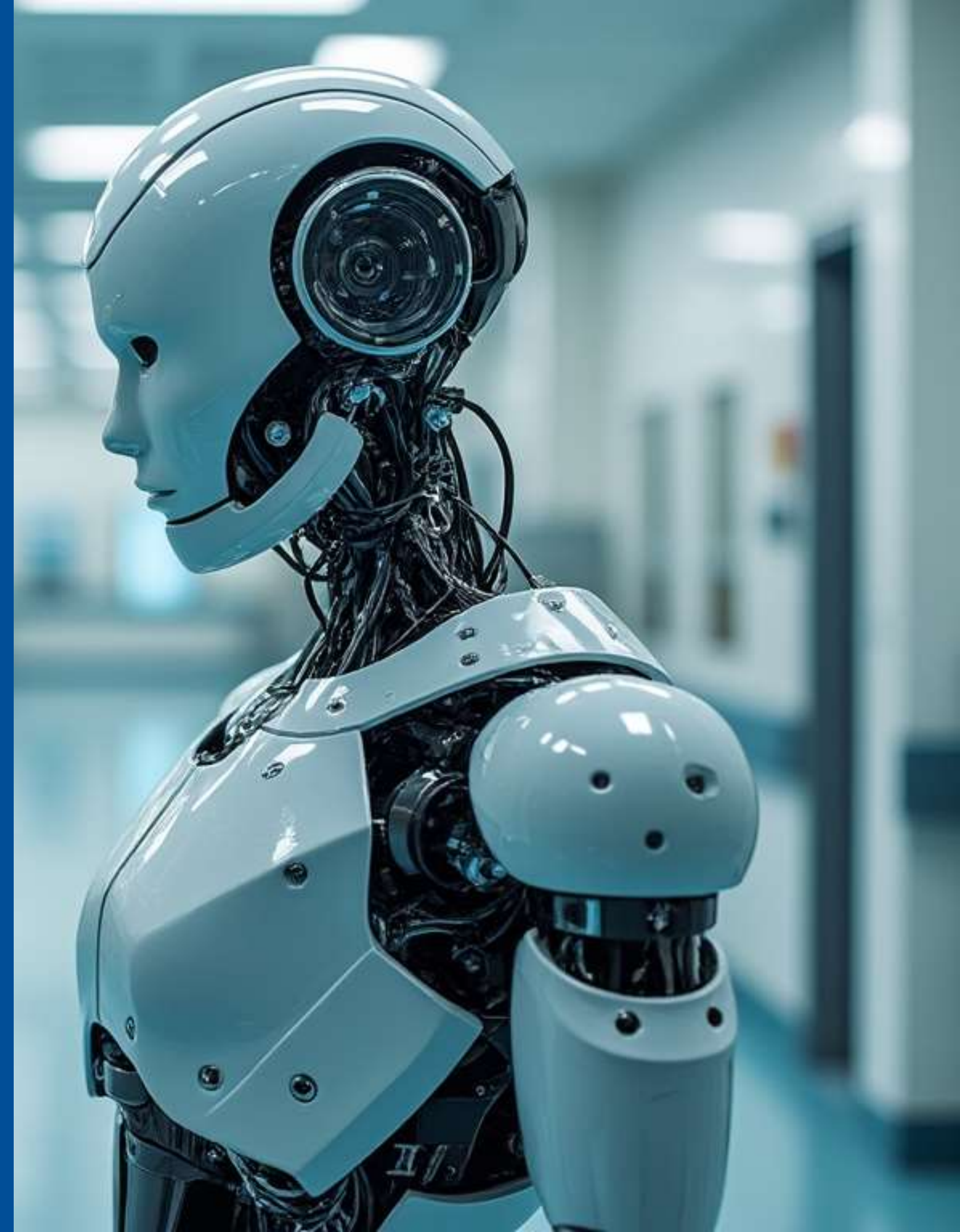
	precision	recall	f1-score	support
0	0.97	1.00	0.98	18292
1	0.95	0.69	0.80	1708
accuracy			0.97	20000
macro avg	0.96	0.84	0.89	20000
weighted avg	0.97	0.97	0.97	20000



RECOMMENDATION

For Stark Health:

- Monitor key predictors: Blood Glucose Level and HbA1c Level.
- Implement targeted interventions for high-risk groups.
- Integrate the predictive model into EHR systems for proactive care.
- Launch educational campaigns on smoking and obesity.
- Continuously update and refine the model with new data.





CONCLUSION

- The project successfully developed a robust diabetes prediction model.
- Random Forest demonstrated exceptional performance in predicting diabetes risk.
- Stark Health Clinic can leverage these insights to enhance patient care, reduce costs, and take a proactive approach to combating diabetes.



THANK YOU!

Question & Answer?