A Data-Driven Population-Based Targeted Intervention for Diabetes Prevention and Management: A Multidimensional Approach to Enhance Community Health Outcomes

KRUSHI PATEL, SHRINIDHI RAJESH

Introduction

- Diabetes is a widespread metabolic disorder affecting millions worldwide.
- Rising diabetes cases necessitate proactive measures.
- Complications include heart disease, kidney failure, and blindness.
- Understanding lifestyle's role is crucial for effective prevention.
- **Project Goal: Develop targeted, data-driven interventions using advanced analysis to prevent diabetes in at-risk populations.**

Factors

Demographics

Life Style Habits

Healthcare Access

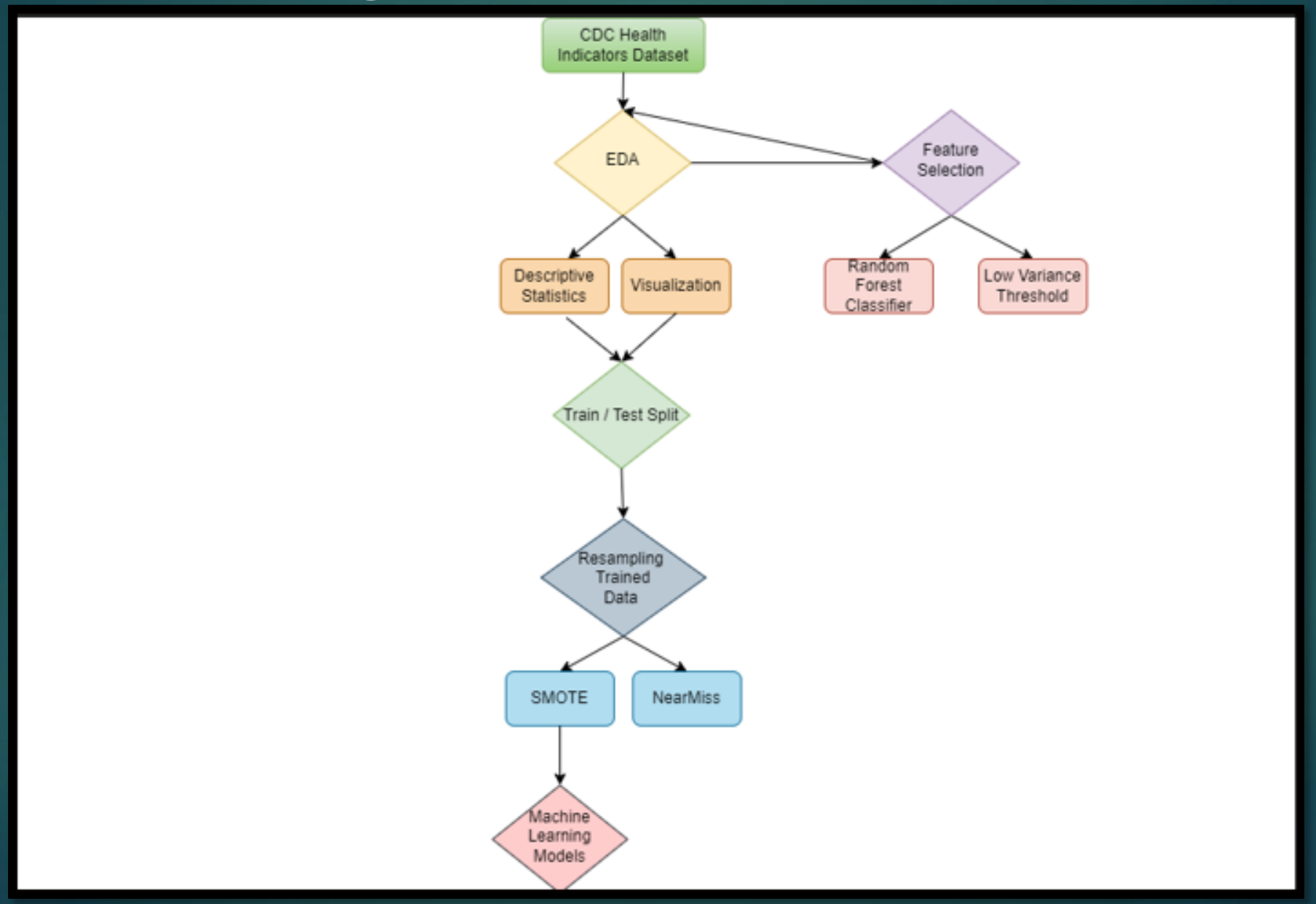
Clinical Factors

Data Set

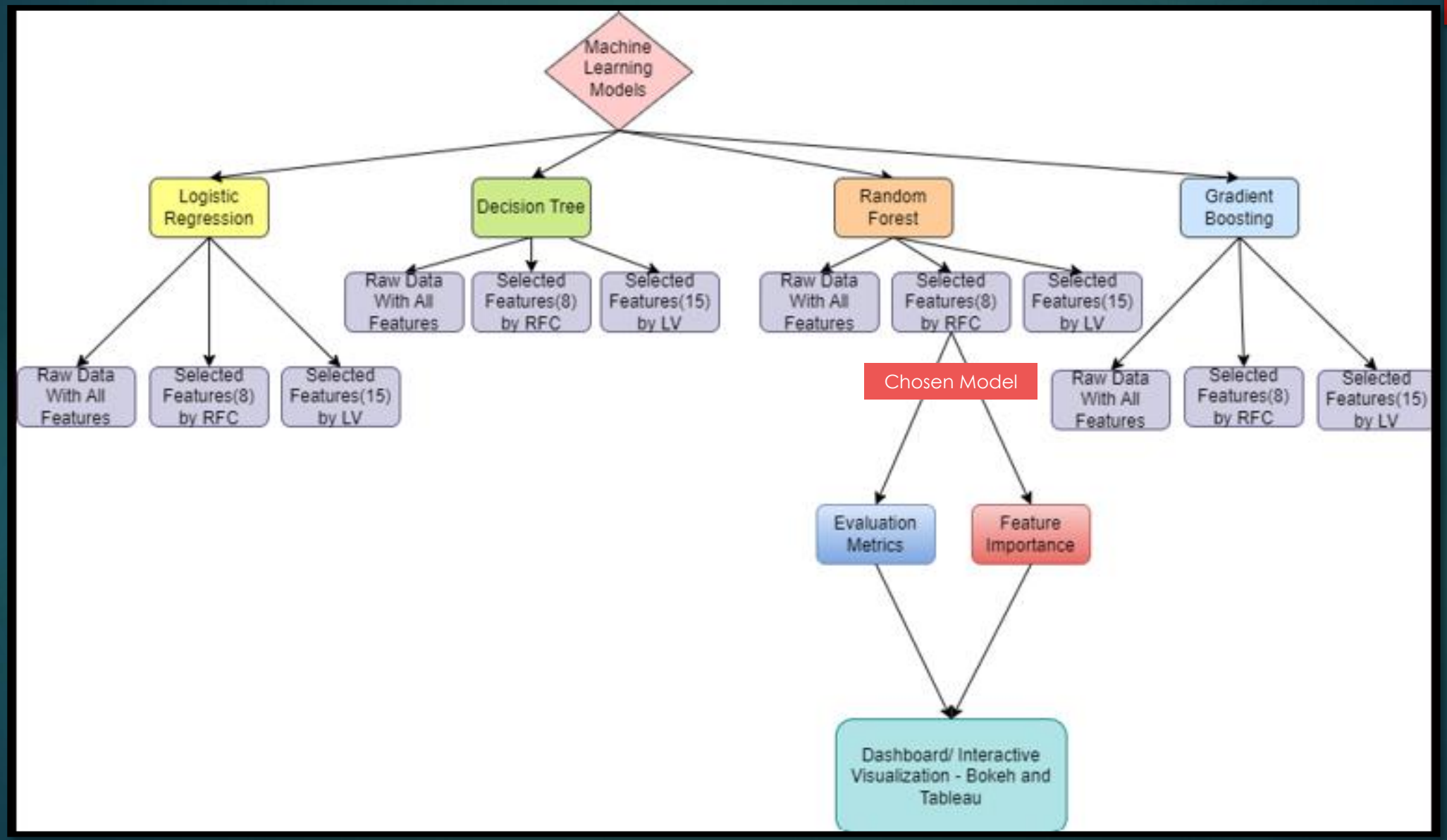
- Data Collection: UCI Machine Learning Repository
- Instances: 253,680
- Features: 21
- Target: 1 (Having/ Not Having Diabetes)

Variables Table			
Variable Name	Role	Type	Converted Type
target	Target	Binary	Binary
HighBP	Feature	Binary	Binary
HighChol	Feature	Binary	Binary
CholCheck	Feature	Binary	Binary
BMI	Feature	Integer	Integer
Smoker	Feature	Binary	Binary
Stroke	Feature	Binary	Binary
HeartDiseaseorAttack	Feature	Binary	Binary
PhysActivity	Feature	Binary	Binary
Fruits	Feature	Binary	Binary
Veggies	Feature	Binary	Binary
HvyAlcoholConsump	Feature	Binary	Binary
AnyHealthcare	Feature	Binary	Binary
NoDocbcCost	Feature	Binary	Binary
GenHlth	Feature	Integer	Category
MentHlth	Feature	Integer	Category
PhysHlth	Feature	Integer	Category
DiffWalk	Feature	Binary	Binary
Sex	Feature	Binary	Binary
Age	Feature	Integer	Category
Education	Feature	Integer	Category
Income	Feature	Integer	Category

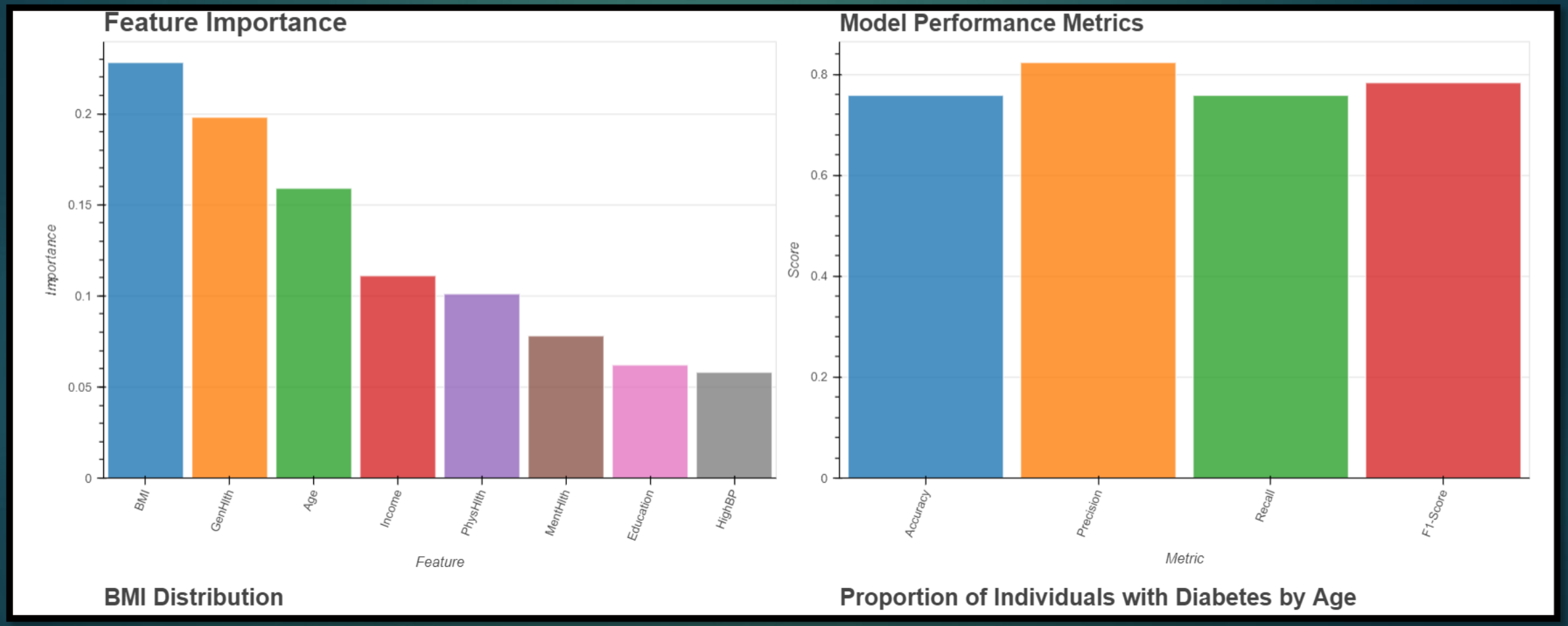
Methodology



Methodology(Contd.)



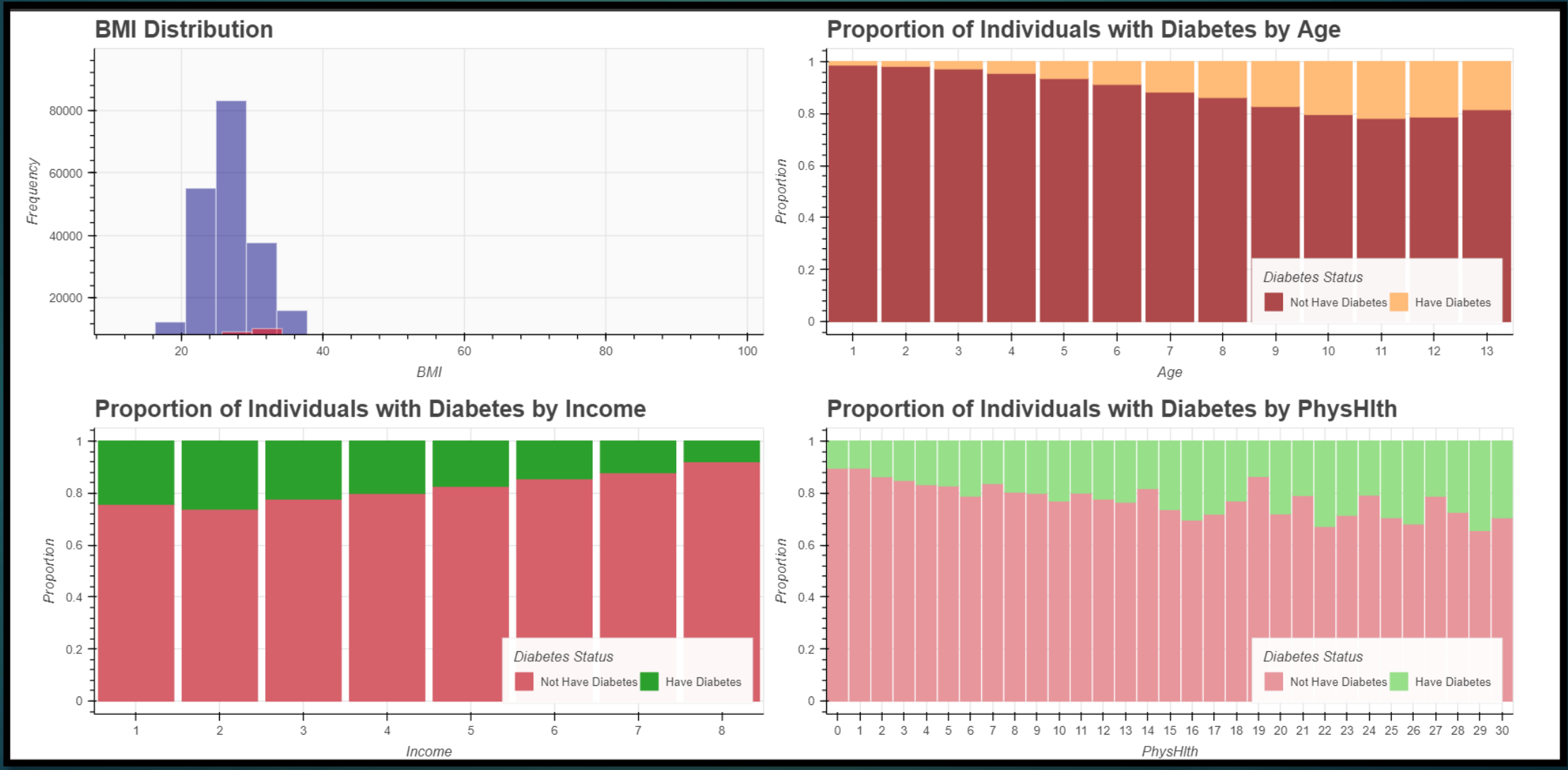
Results and Discussions



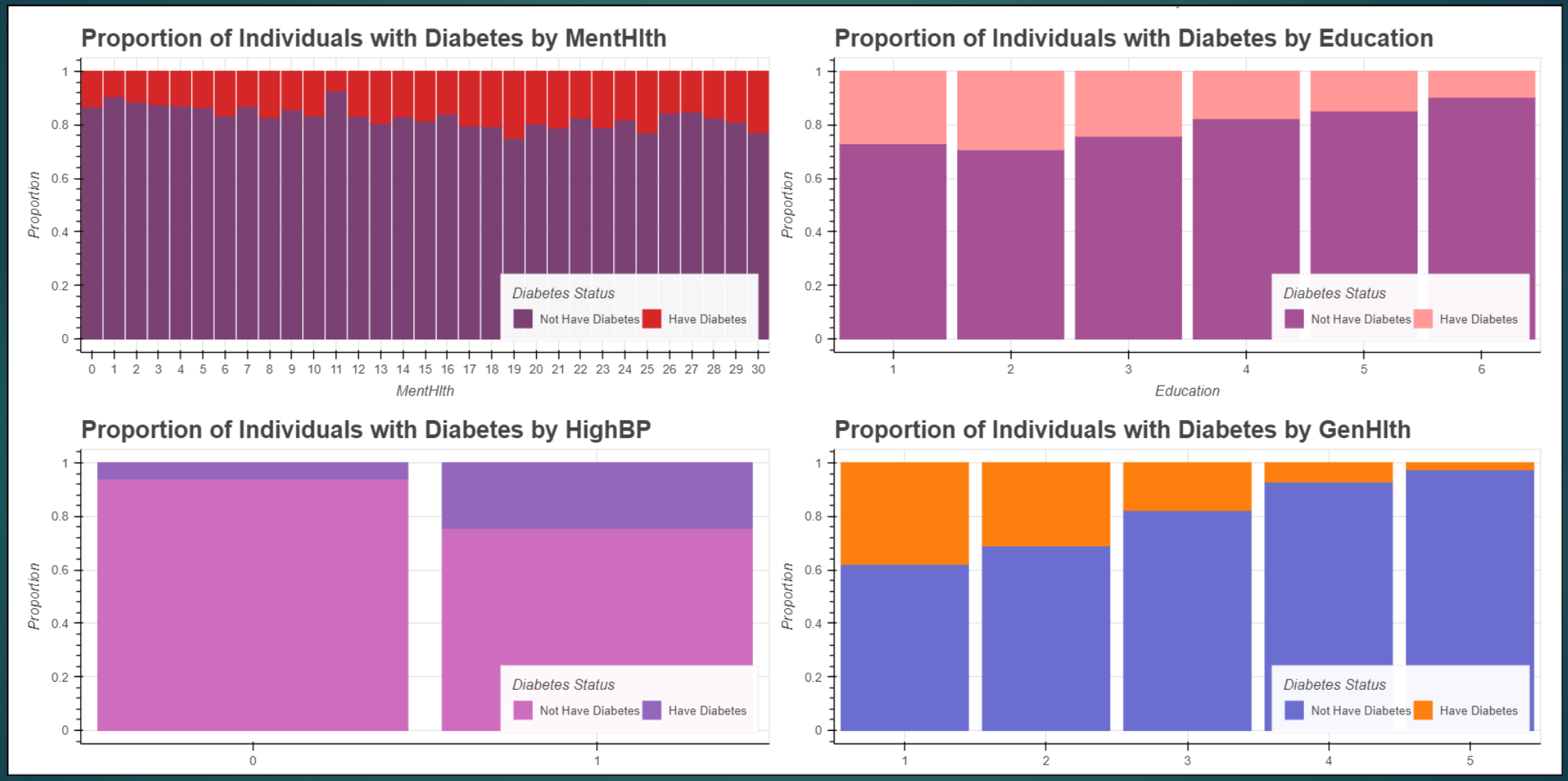
Model Comparison

Model	Weighted Sum	Accuracy	Precision	Recall	F1 Score
Logistic Regression					
all_features_resampled	3.11	0.74	0.84	0.74	0.77
selected_features_lv_resampled	3.1	0.73	0.84	0.73	0.77
selected_features_rf_resampled	3.1	0.73	0.85	0.73	0.77
Decision Tree					
all_features_dt	3.09	0.75	0.81	0.75	0.78
selected_features_lv_dt	3.08	0.75	0.81	0.75	0.77
selected_features_rf_dt	3.07	0.74	0.81	0.74	0.77
Random Forest					
all_features_rfc	3.23	0.8	0.83	0.8	0.81
selected_features_lv_rfc	3.22	0.79	0.83	0.79	0.81
selected_features_rf_rfc	3.12	0.76	0.82	0.76	0.78
Gradient Boosting					
all_features_xgb	3.11	0.75	0.84	0.75	0.78
selected_features_lv_xgb	3.1	0.74	0.84	0.74	0.77
selected_features_rf_xgb	3.06	0.72	0.85	0.72	0.76

Distribution of Features



Distribution of Features(Contd.)



Target Audience

US Based Population

BMI

Age

Income Level

Education Level

Blood Pressure

Proposed Interventions
HI-5 Bucket 1: Positive Health Impacts

Health Education Workshops

Support Groups

Financial Assistance

Affordable Fitness Classes

Health Screening Camps

HI-5 Bucket 2: Achieving Results Within Five Years

Measured as the reduction in Incidence of Diabetes Cases

- Collaboration with Local Community Centers
- Continuous Monitoring and Evaluation Programs
- Practical and Easy to Integrate
- Data Centric Approach
- Merging of Technology and Human Insights

HI-5 Bucket 3: Cost-Effectiveness and Cost Savings:

- Detect Patient Diabetes Treatment Cost - Assessment with Insights Derived From Our Intervention Program.
- Advocacy Efforts – Affordable Healthcare Policy
- Alliance With Health Insurers - premium discounts or wellness rewards

Data Engineering Plan

- Data Collection
- Transmission and Storage
- Pre-processing and Integration
- Analysis and Modelling

Data Engineering Plan (Contd.)

- Testing
- Validation
- Monitoring and Maintenance
- Compliance and Security

Data Engineering Plan(Contd.)

- Network Infrastructure
- Collaboration and Communication
- Scalability and Flexibility
- Utilization of Analyzed Data for Intervention Enhancement

Limitations

- Self reported data – Bias
- Class Imbalance (86:14)
- External Factors Influence on the Intervention
- Choice of Feature Selection and Model
- False Negatives
- Correlation Does Not Imply Causation

Future Project Extension

- Relationship between the features themselves and the target.
- Comprehensive Statistical Analysis – Compare groups.
- Cluster Identification
- Feature Engineering

Ethical Consideration

- Inclusion, Fairness and Integrity.
- Obtaining Explicit Concern
- Regular Audits for Data Quality and Eliminating Bias
- Avoiding Data Leaks and Breaches
- Managing Threats and Handling Digital Footprint
- Avoid Control Creep

Conclusion

- Targeted Specific Risk Factors
- Data Driven Community based Interventions
- Empower individuals to combat diabetes
- Commitment to continuous refinement
- Promote health equity and enhance population health

THANK YOU!