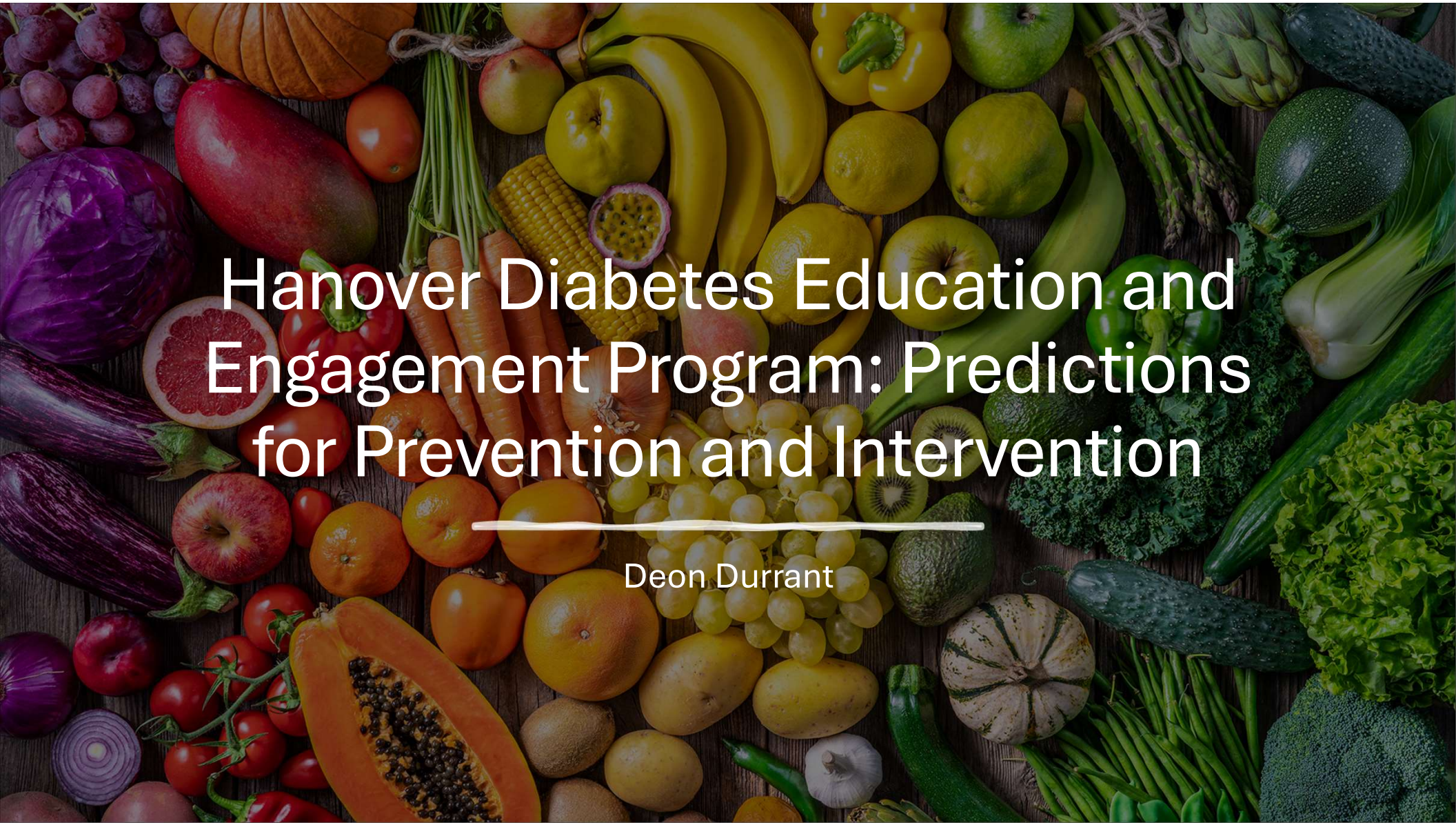


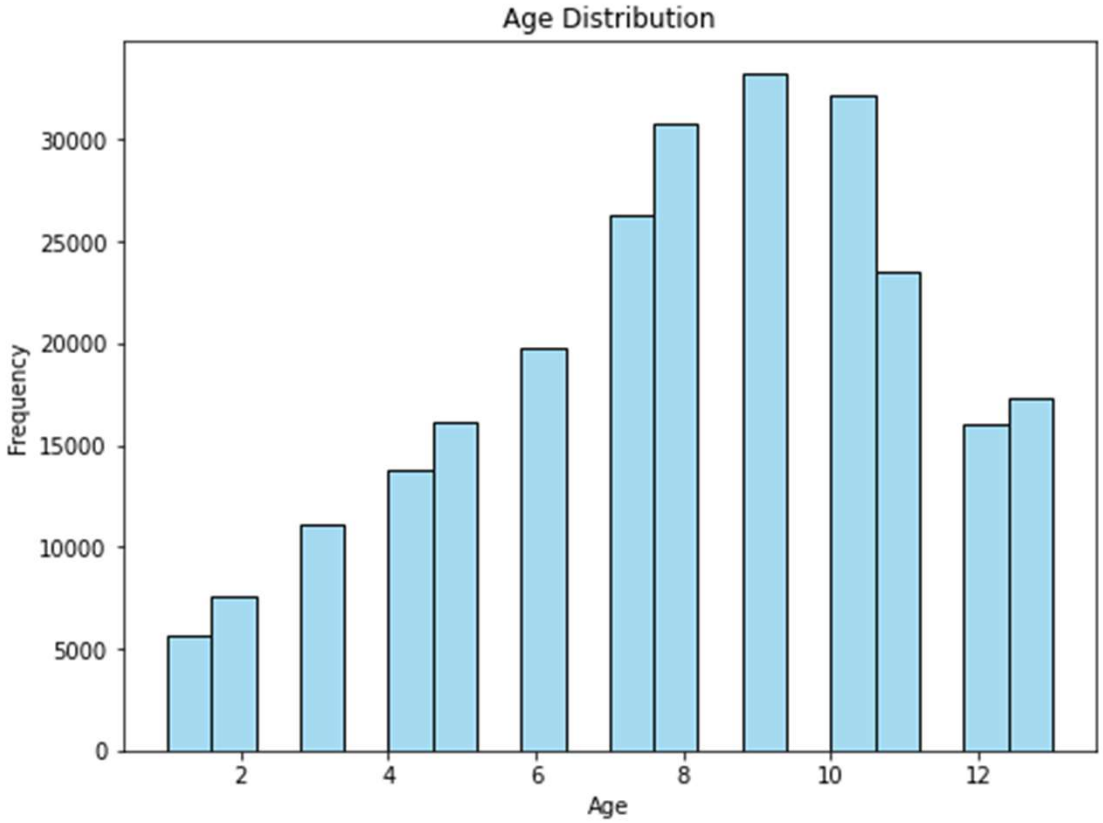
Hanover Diabetes Education and Engagement Program: Predictions for Prevention and Intervention

Deon Durrant

Business and Data Understanding

Problem Statement

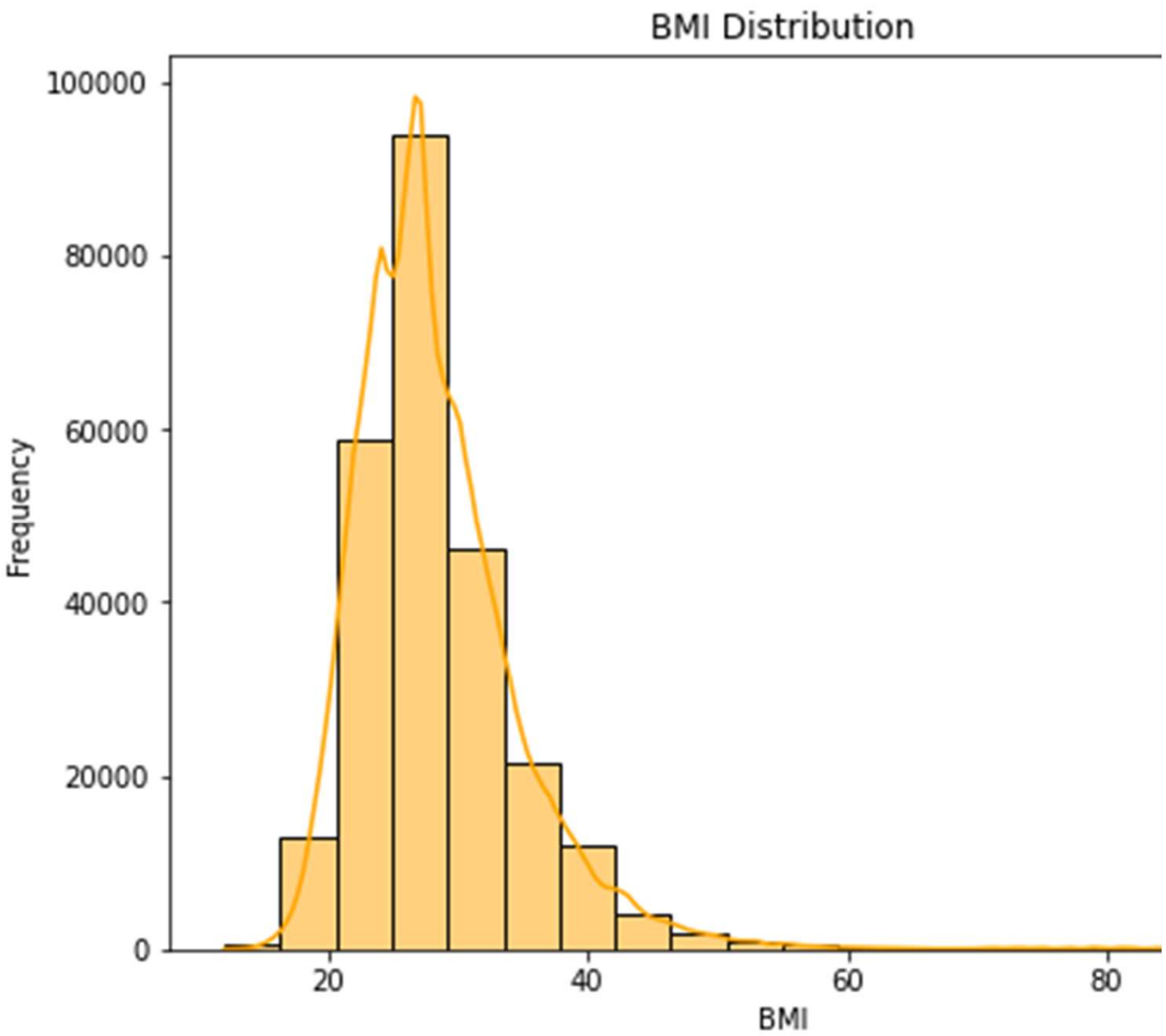
- In the U.S., 1 in 3 adults are prediabetic, yet 80% remain unaware of their condition. Without timely intervention, most will develop type 2 diabetes within 5 years, leading to severe health complications and significant financial burdens—$413 billion was spent on diabetes care in 2022 alone.
- To address this challenge, Hanover is launching the **Diabetes Education and Engagement Program (DEEP)**. This initiative leverages data analysis to identify at-risk individuals for early intervention, improve detection rates, and empowers the community through education to prevent and manage diabetes effectively



Business and Data Understanding

Demographic Distribution

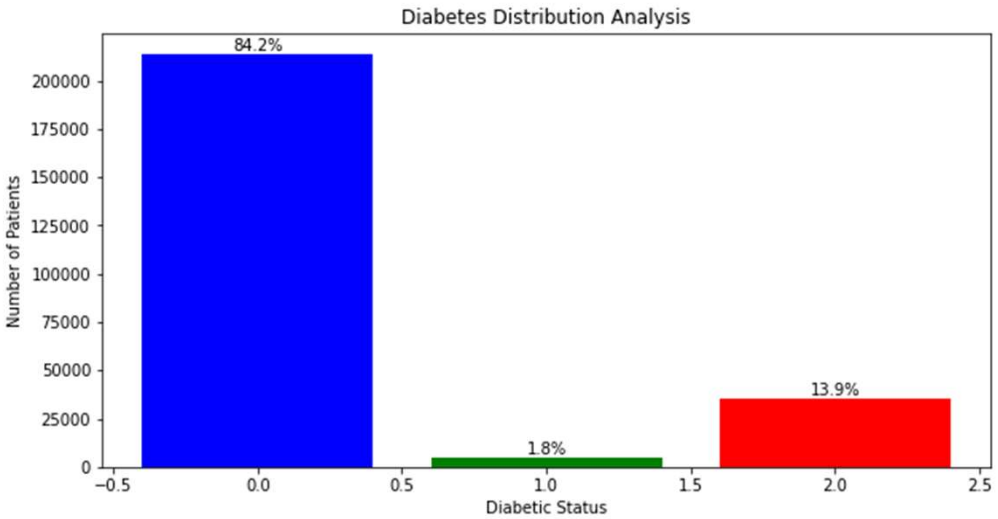
- Age ranges from 18-80 years, peak 55-69 years old.
- Sex: women outnumber men 56% to 44%
- Income: The majority income above $35000



Business and Data Understanding

- **Health indicators**
- Most BMI falls between 20-40.
- More Non-Smokers than Smokers in the dataset (140000: 105000)
- - Physical Activities: majority of individuals are "Active" - females are more physically active overall.

Business and Data Understanding



Diabetes Distribution

- **21,3703** labeled"0" (nondiabetic)
- **4631** labeled as "1" (pre-diabetes) **1.82%**
- **35,346** labeled as "2" (diabetes) **13.9%**

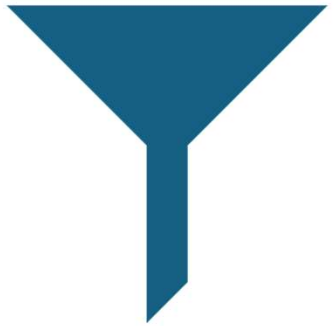
Classification System: Technique and Model selection

- XGBoost creates a predictive machine learning model of the data

- SHAP (SHapley Additive exPlanations)

Evaluation Metrics

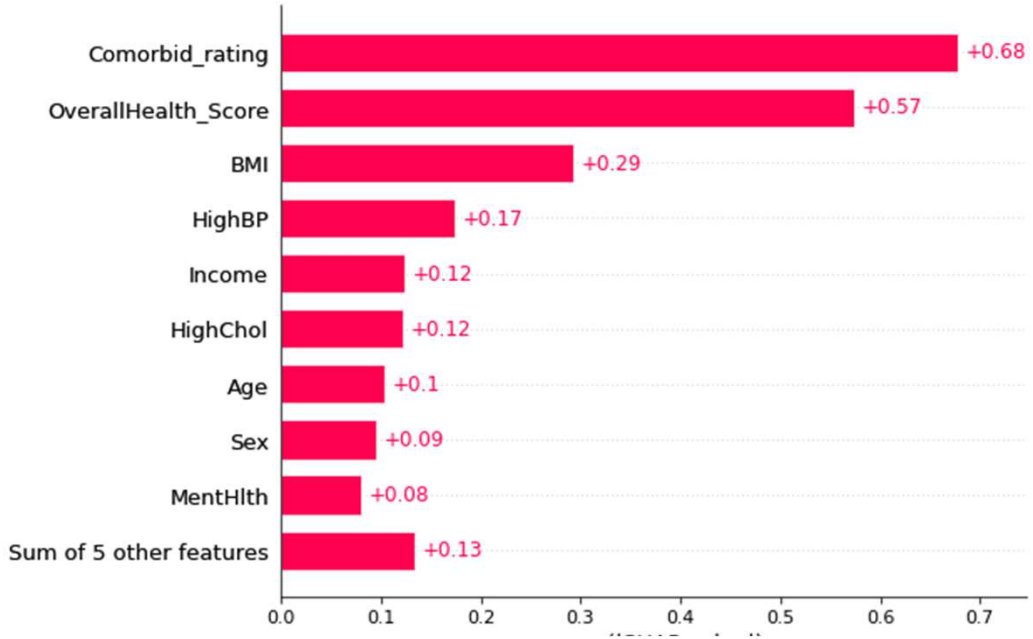
- Accuracy, precision, recall, and F1-score for balanced and relevant predictions.
- SHAP fidelity evaluation



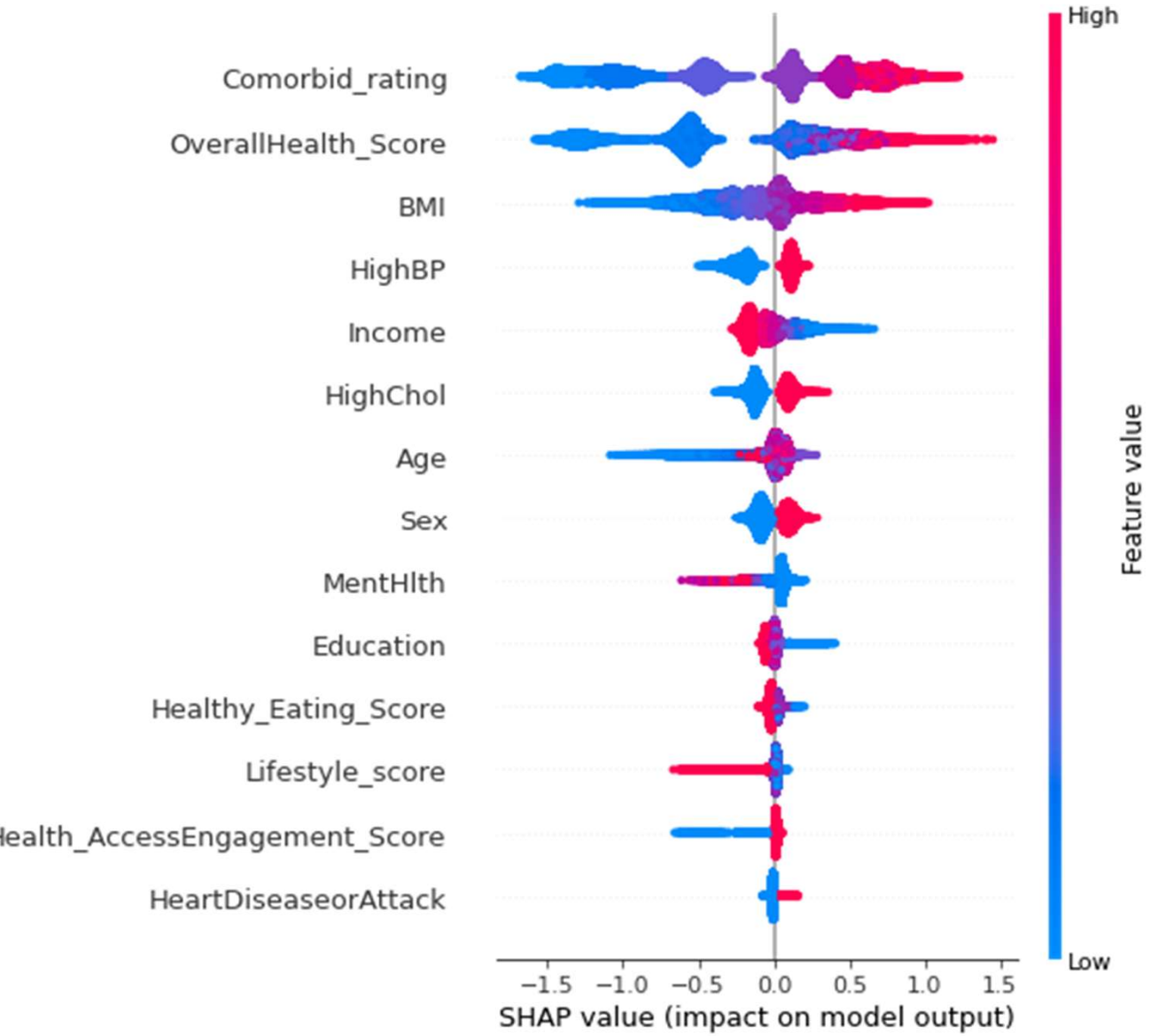
Classification System: Technique and Model selection

Target Variable Description
Diabetes & Pre-Diabetes = **Diabetes_status**
Non-Diabetes as the second category.
Early Detection Focus: flags anyone at risk of diabetes,.
Preventive Action: Treats both diabetic and pre-diabetic as high-risk groups, prompting similar preventive measures.

Analysis of the feature importances



- **Feature Importance Order:**
- Top contributors to the model's predictions
    - Comorbid_rating
    - OverallHealth_Score
    - BMI

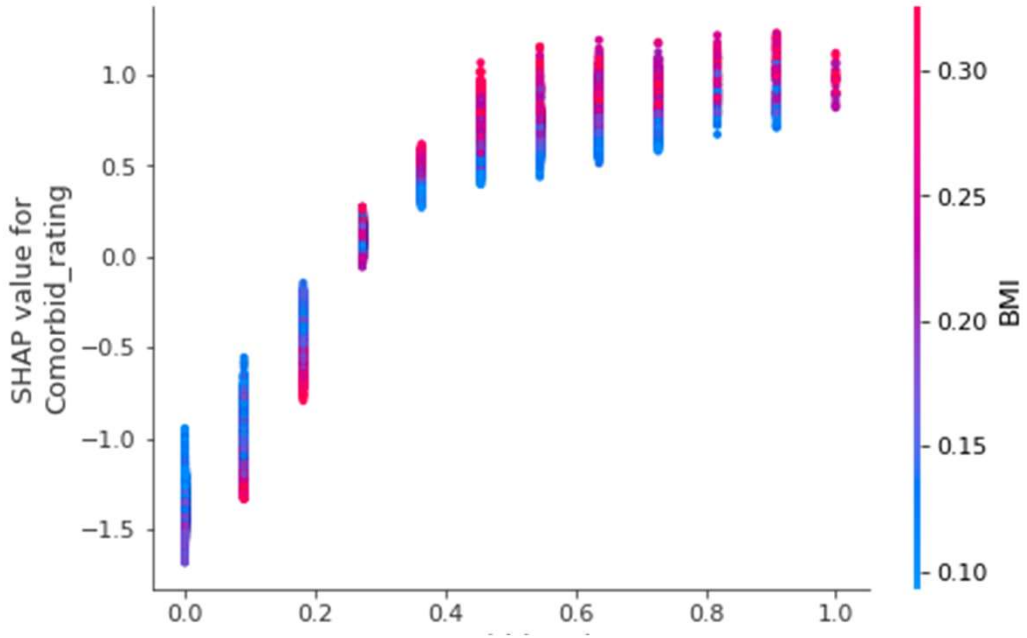


Analysis of the feature importances

Features impact on model's output

- High importance features
- Moderate importance features
- Low importance features

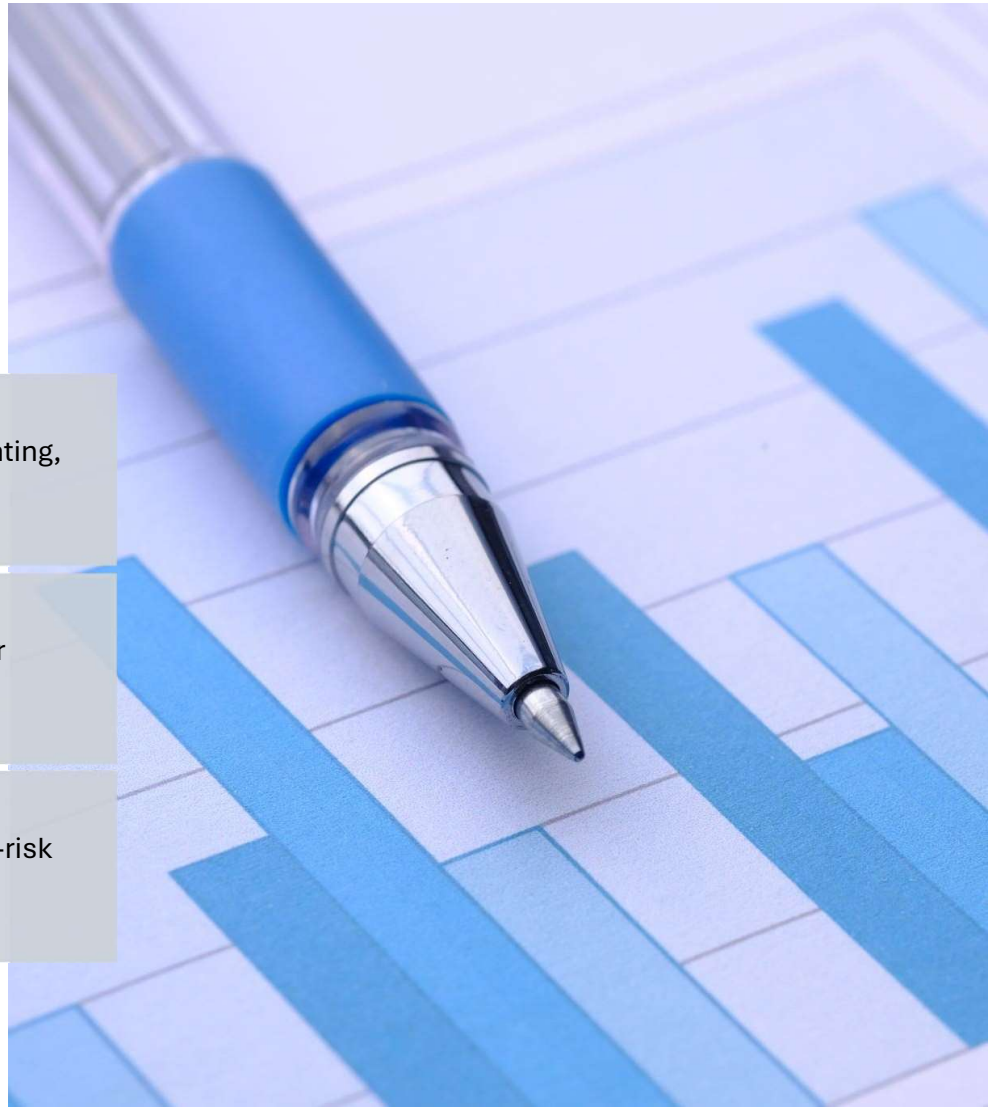
Analysis of the feature importances



- Feature Interactions:
- High Comorbid_rating & low interacting feature (blue) moderate SHAP values, (protective effect) offsets the severity of comorbidities.
- High Comorbid_rating & high interacting feature (pink) strongest SHAP values, maximum contribution to risk.

Recommended Interventions

Target	Target High-Risk Groups: Individuals with high Comorbid_rating, poor health, low income, and unhealthy lifestyles.
Leverage	Leverage Protective Factors: Promote healthy eating, higher education, better healthcare access, and income growth.
Address	Address Interactions: Target scenarios where multiple high-risk features compound effects to implement tailored



Next Steps

Improve model predictive performance and accuracy

- Validate Findings
- Feature Engineering
- Enhance Model Interpretability
- Cluster-Based Analysis

Questions

Thank You!

Email:DurrantDeon@gmail.com
LinkedIn: linkedin.com/in/dr-d-durrant