Unravelling the Diabetes Triggers Among Adults in the United States of America

Presenter: Felix Agyei-Sasu

Data Analyst

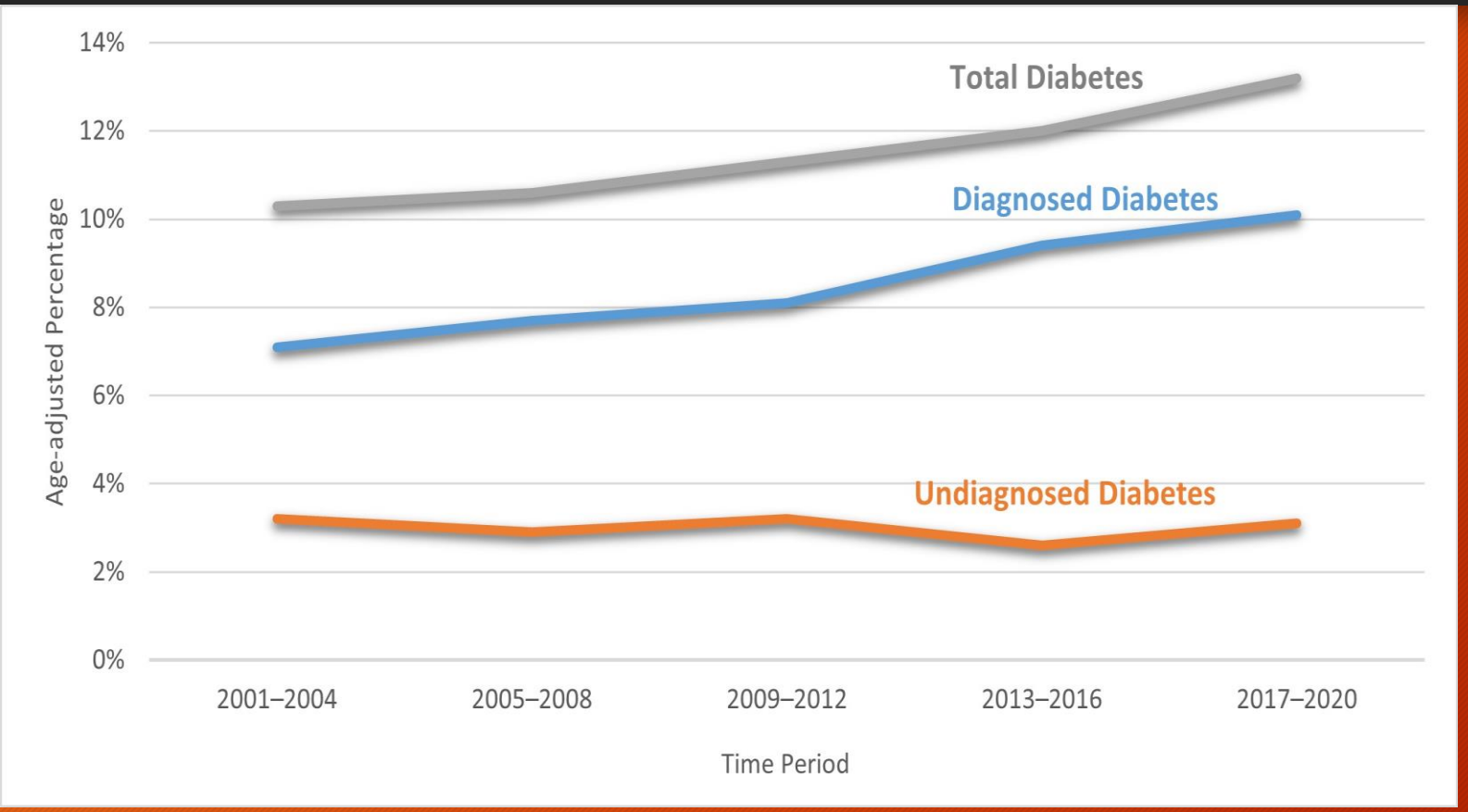
Overview

- Introduction
- Data and methods
- Main insight
- Recomendations

Introduction

- Diabetes, a metabolic disorder affects the body's ability to manage blood sugar.
  - Shortened lifespan
  - Diminished daily living quality
  - Economic toll –about $327 billion (diagnosed) and about $400 billion (undiagnosed) anually

Introduction



- Prevalence of diabetes is on the increase

https://archive.cdc.gov/www_cdc_gov/diabetes/library/reports/reportcard/national-state-diabetes-trends.html

Introduction

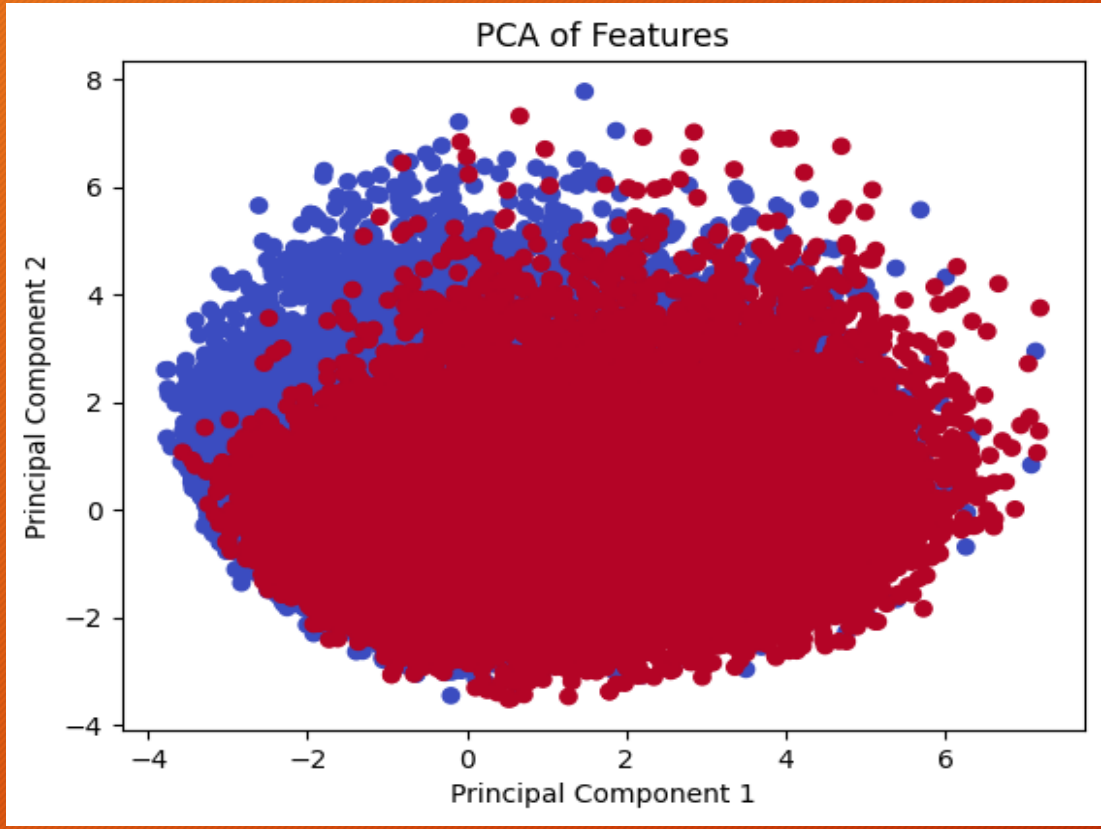
- **Objective**: to identify key factors associated with diabetes and prediabetes

Data and methods

- dataset of 70,692 survey responses
- Target variable is (Binary Diabetes) non-diabetic (0) and prediabetic or diabetic (1) – evenly split
- 21 explanatory variables that offer insights
- Initial EDA –Visualizations including principal component analysis (PCA) chat

Data and methods

Figure 1: Principal component analysis for features and target parameters



- Overlapping cluster
  - => limited linear separability
  - => more complex models to capture non-linearity (≠ Logistic model)

Data and methods

- Choice of an appropriate statistical modelling approach – Random Forest (RF) -80/20
- RF
  - ➢Mitigates overfitting
  - ➢Handles non-linear data
  - ➢Works well with both categorical and numerical features
  - ➢ -less interpretable than a single decision tree and can be intensive computationally

Main insight

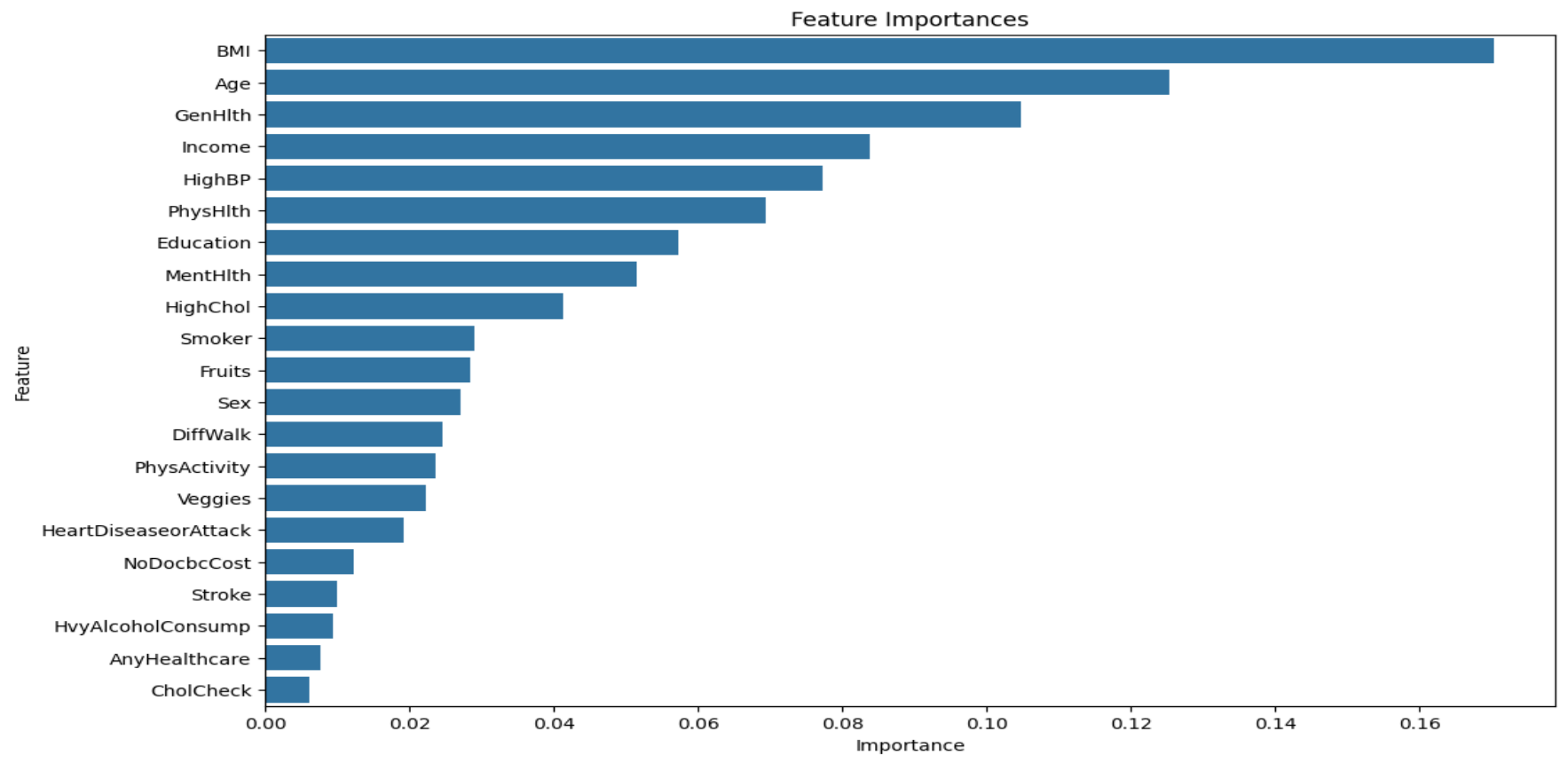
- Table 1: Model performance matrix

Matrix	Values				
Accuracy:	0.74				
Classification Report:					
	precision	Precision	Recall	f1-score	support
	0	0.76	0.71	0.73	8858
	1	0.72	0.77	0.75	8815
	accuracy			0.74	17673
	macro avg	0,74	0,74	0.74	17673
	weighted avg	0.74	0.74	0.74	17673

- 74% accuracy is a reasonable predictive power
- f1-score of 0.74 for both cases indicate consistent performance across both cases with no significant bias

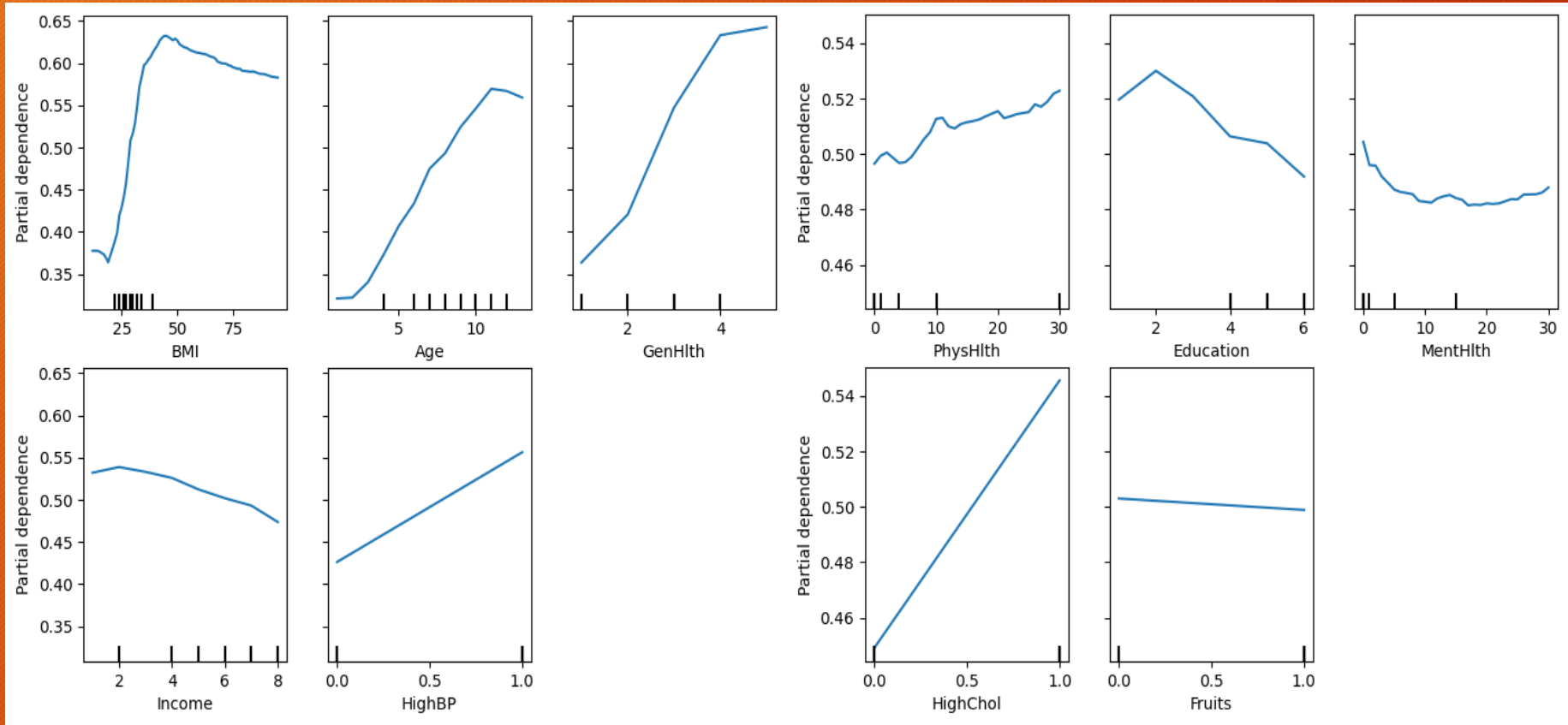
Main insight

Figure2: Features and their contribution to diabetes incidences



Main insight

Figure 4: Partial dependence plots (PDP) of the top 10 important features



- The complexity of the disease predictors due to non-linear relationships

Recommendations

- **Public Health Initiatives:**

✓Target public health initiatives on modifiable risk factors like weight management, diet, and physical activity to reduce the incidence of diabetes.

- **Personalized Interventions:**

✓Personalized interventions may be more effective due to the non-linear relationships shown in the PDPs. Healthcare providers should tailor prevention and treatment strategies to each patient's specific risk profile, rather than using a one-size-fits-all approach

.....the end.....

Thank you