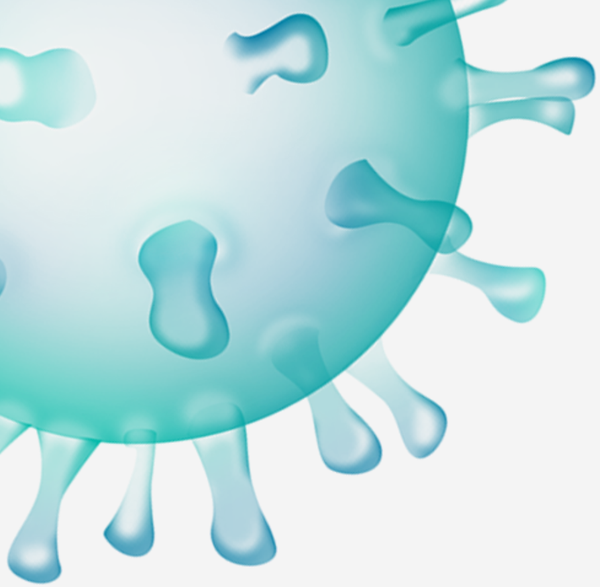
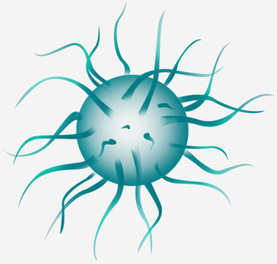
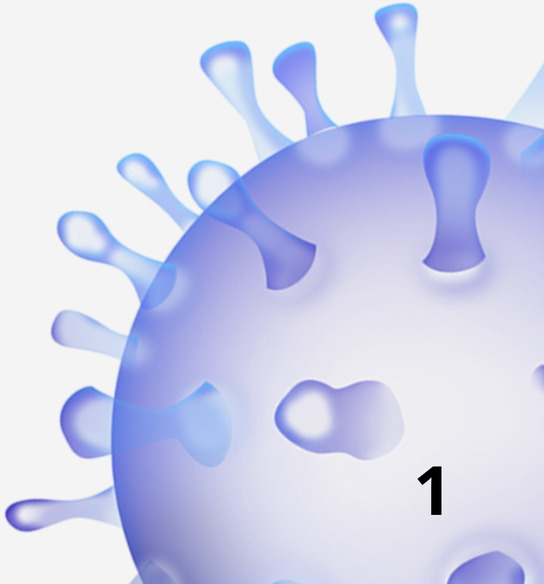


CANCER PREDICTIVE MODEL

By Haya Hadaya





Problem Statement Overview and Proposed Solution

Overview : Cancer diagnosis often arrives too late, impacting treatment success and patient outcomes. Early detection remains a challenge, hindering timely interventions and increasing healthcare costs.

Proposed Solution:

- Utilize Advanced ML
- Develop Predictive Model
- Enable Early Detection
- Improve Patient Outcomes

Early Diagnosis Impact: Machine Learning's Role in Healthcare Transformation

Early Intervention Reduces Costs

Shifts resources from late-stage treatments to cost-effective preventive measures.

Machine Learning for Early Detection

Slows cancer progression, improving treatment outcomes.

EMPOWERMENT THROUGH RISK AWARENESS:

Informs choices for healthier lifestyles.

Personalized Treatment Opportunities

Early detection enables tailored, effective therapies

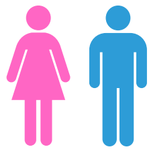
ANALYSIS

Dataset Overview

Features and Target Variable

1- Lifestyle Factors :

- Health Conditions 
- Healthcare Practices 
- General Lifestyle Choices 

2-Demographic Factors 

Target Variable : Cancer (Cancer Risk) 

Dataset source and Description

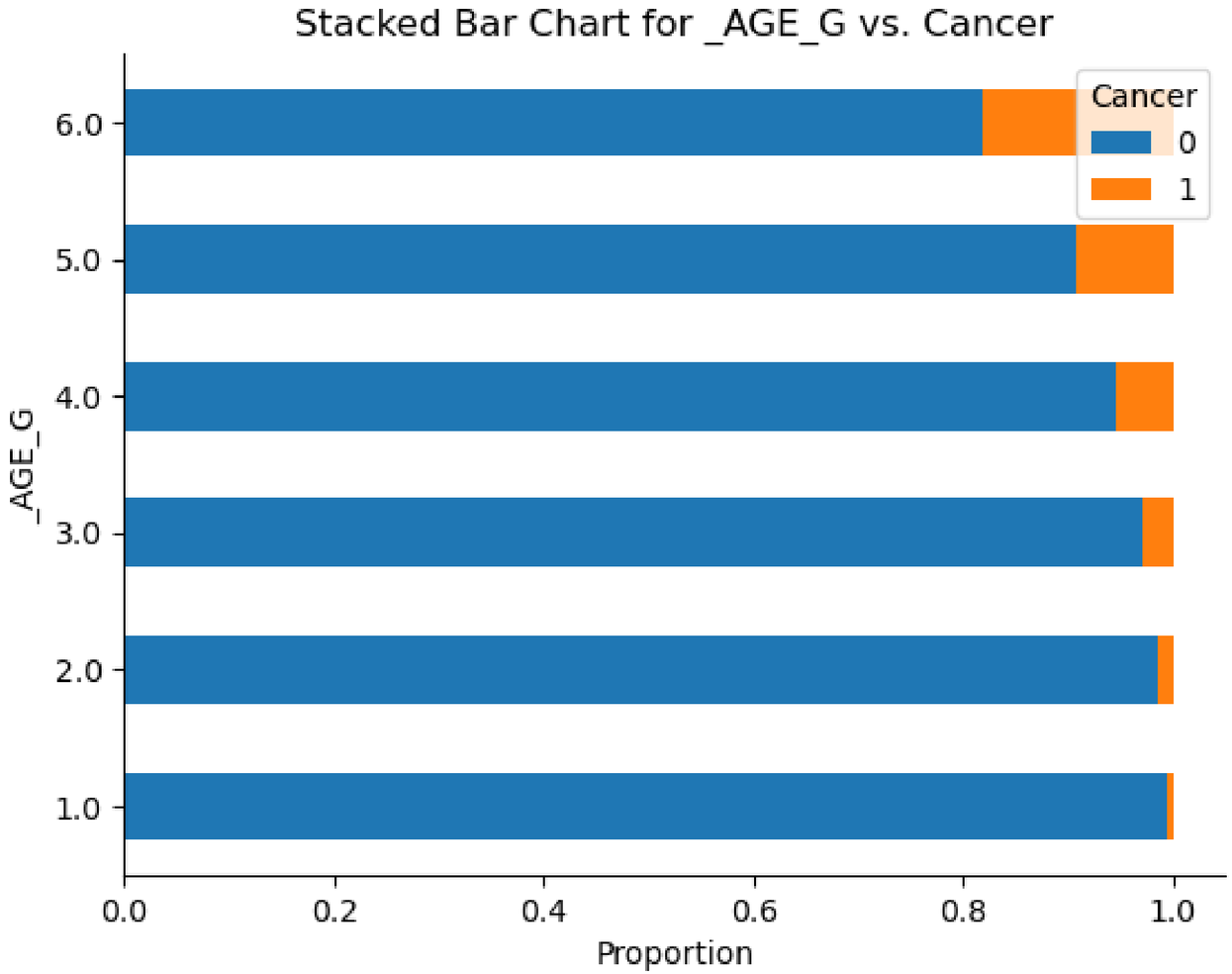
- The Centers for Disease Control and Prevention (CDC) Behavioral Risk Factor Surveillance System (BRFSS) for the year 2021.

- The dataset has 438,693 rows and 303 columns , and it was condensed down to just 18 columns

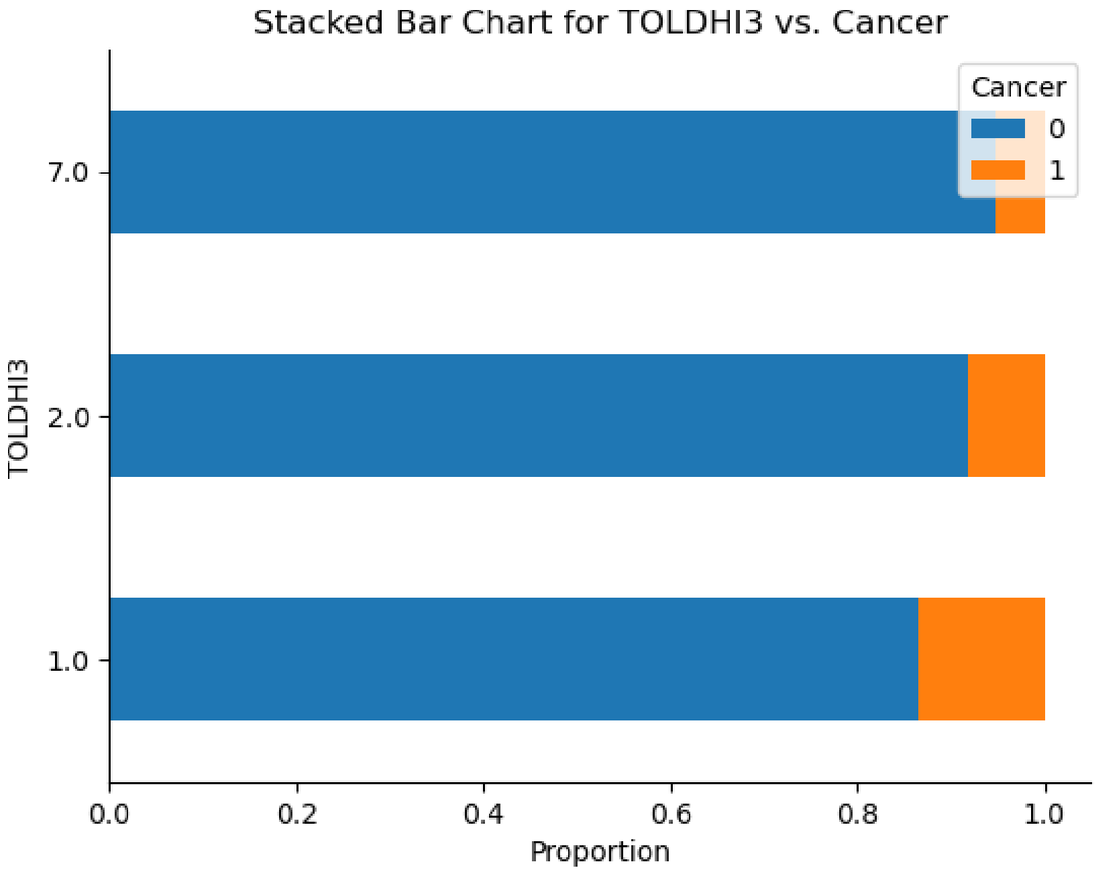
Preprocessing Procedures

1. **MISSING VALUES** — Imputation and Removal
2. **ONE-HOT ENCODING** — Dummy Variables
3. **FEATURE ENGINEERING** — Grouping Infrequent Values
4. **PRINCIPAL COMPONENT ANALYSIS (PCA)**

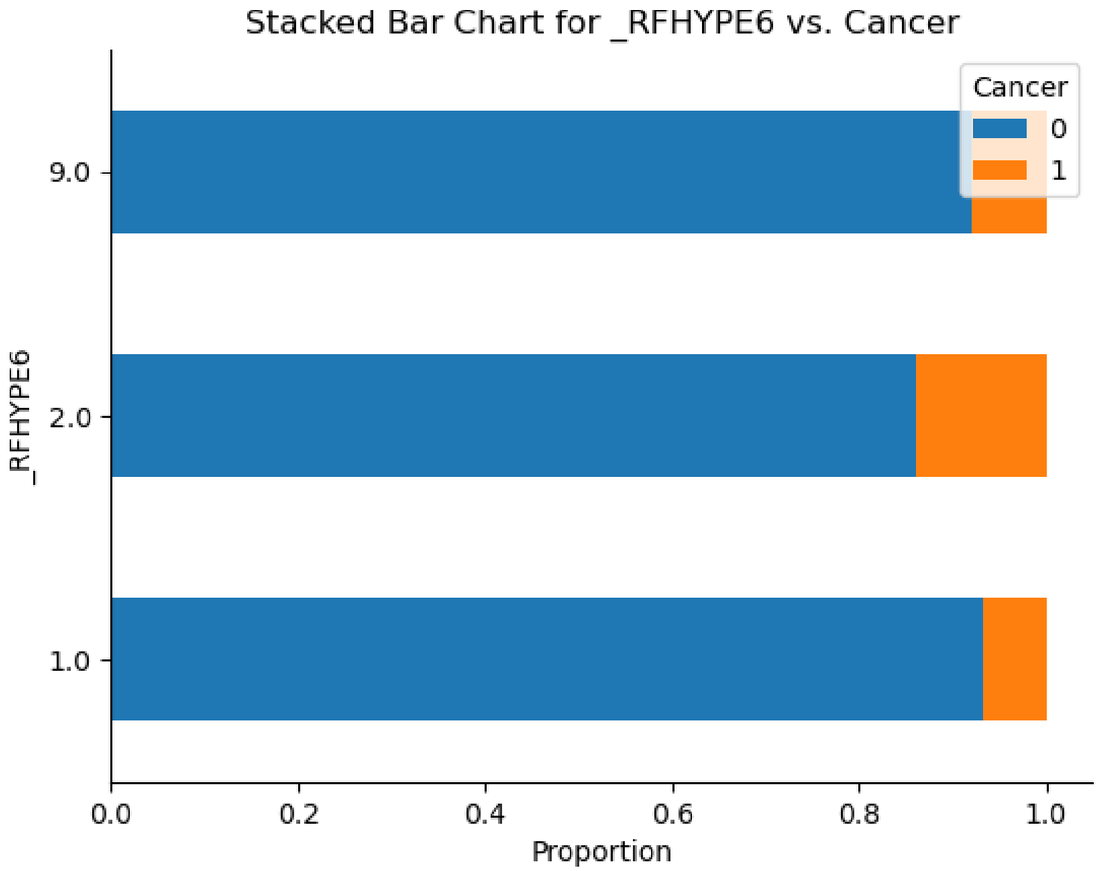
Top Key Findings from EDA



There is a higher association between older age groups and cancer presence.

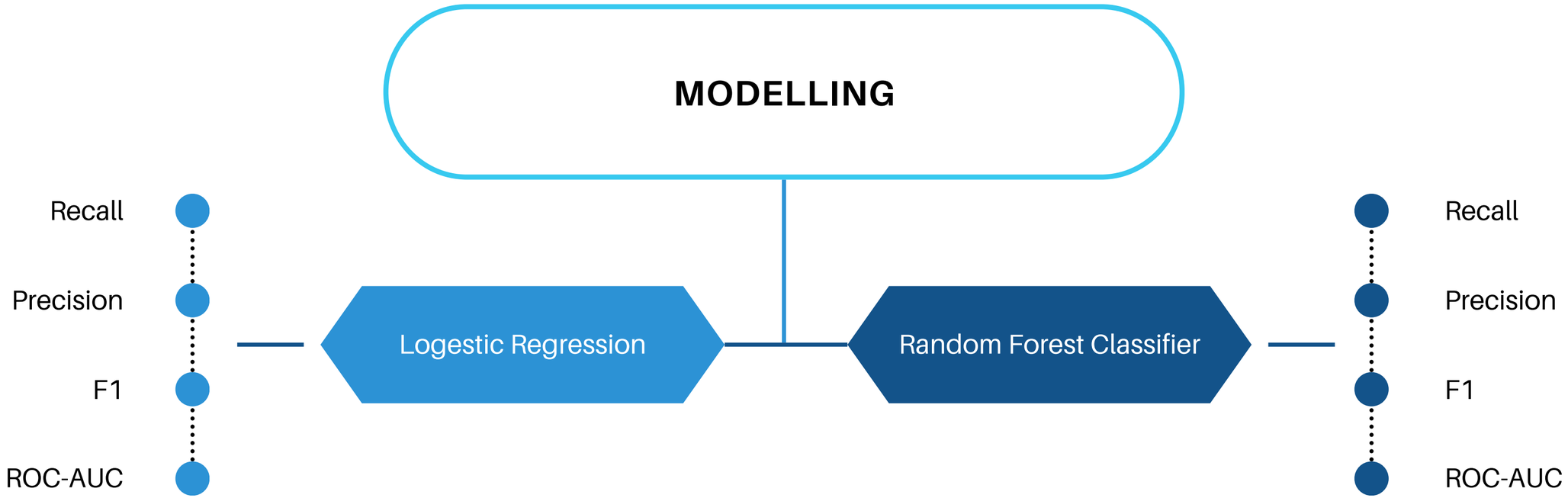


A notable proportion of cancer patients have reported high cholesterol levels

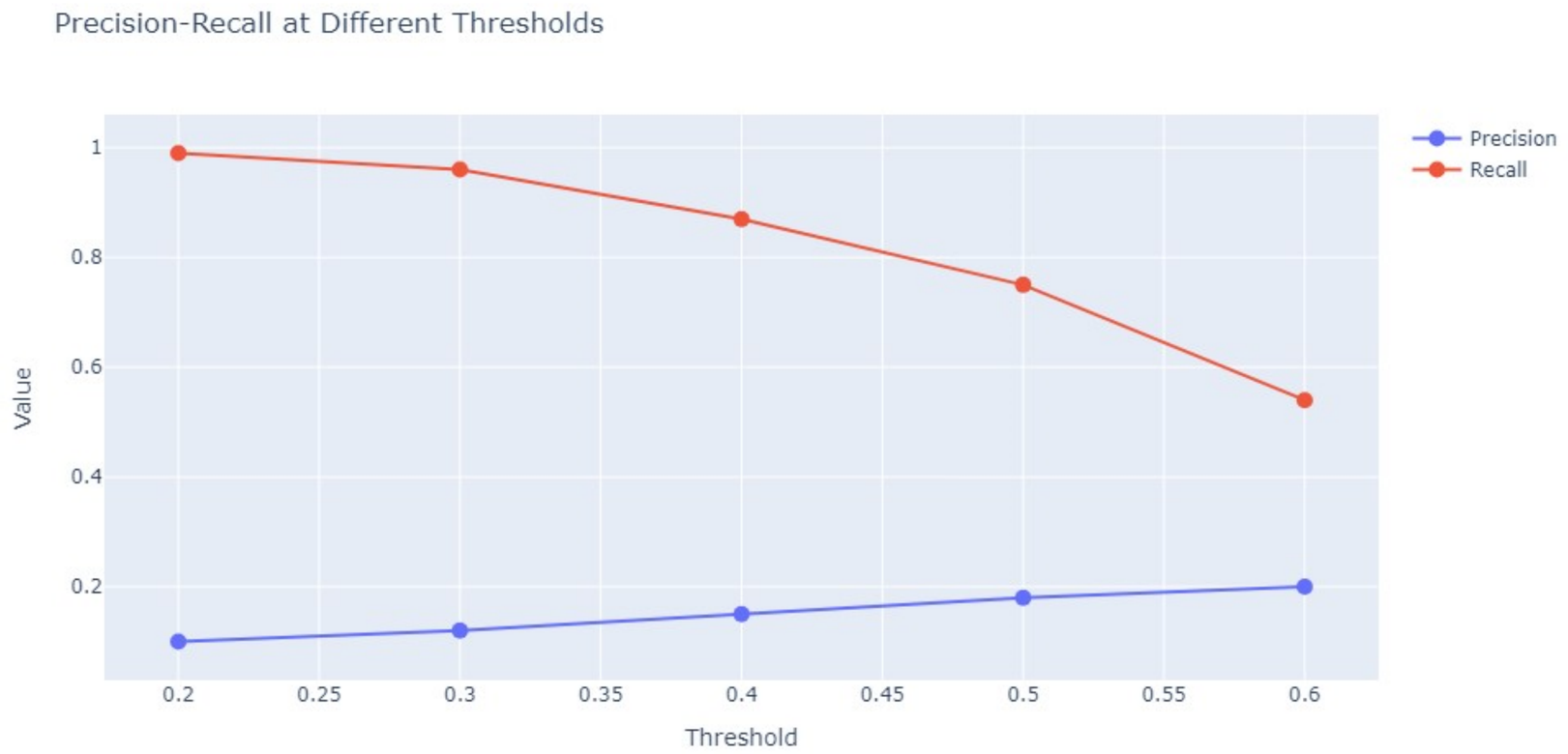


High blood pressure is prevalent among patients who have been diagnosed with cancer.

Modelling and Evaluation Metrics

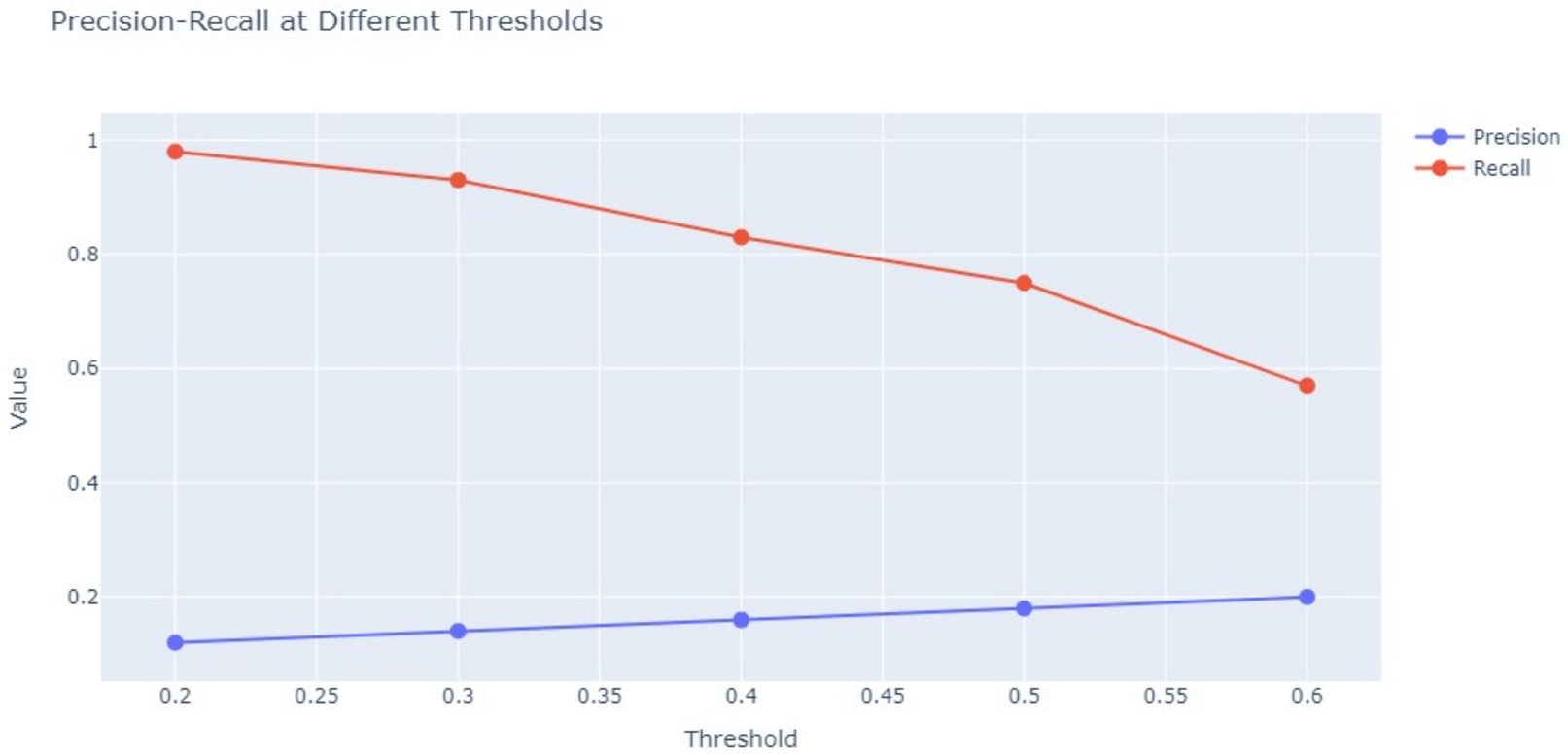


Precision-Recall Curve at Different Thresholds



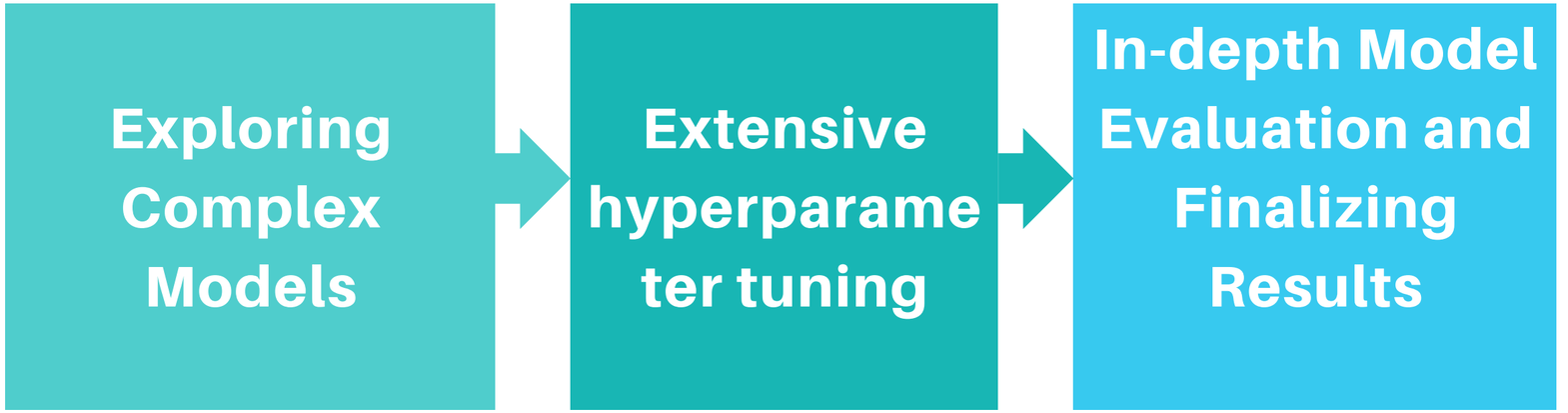
Random Forest

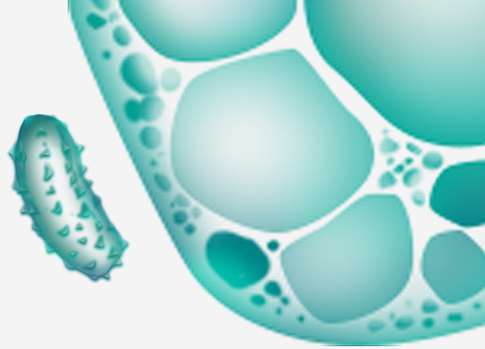
Precision-Recall Curve at Different Thresholds



Logestic Regression

NEXT STEPS


Exploring Complex Models
Extensive hyperparame ter tuning
In-depth Model Evaluation and Finalizing Results

Thank You