weCAN

A CANCER SURVIVABILITY PREDICTOR

BrainStation®

DATA SCIENCE CAPSTONE SPRINT 3

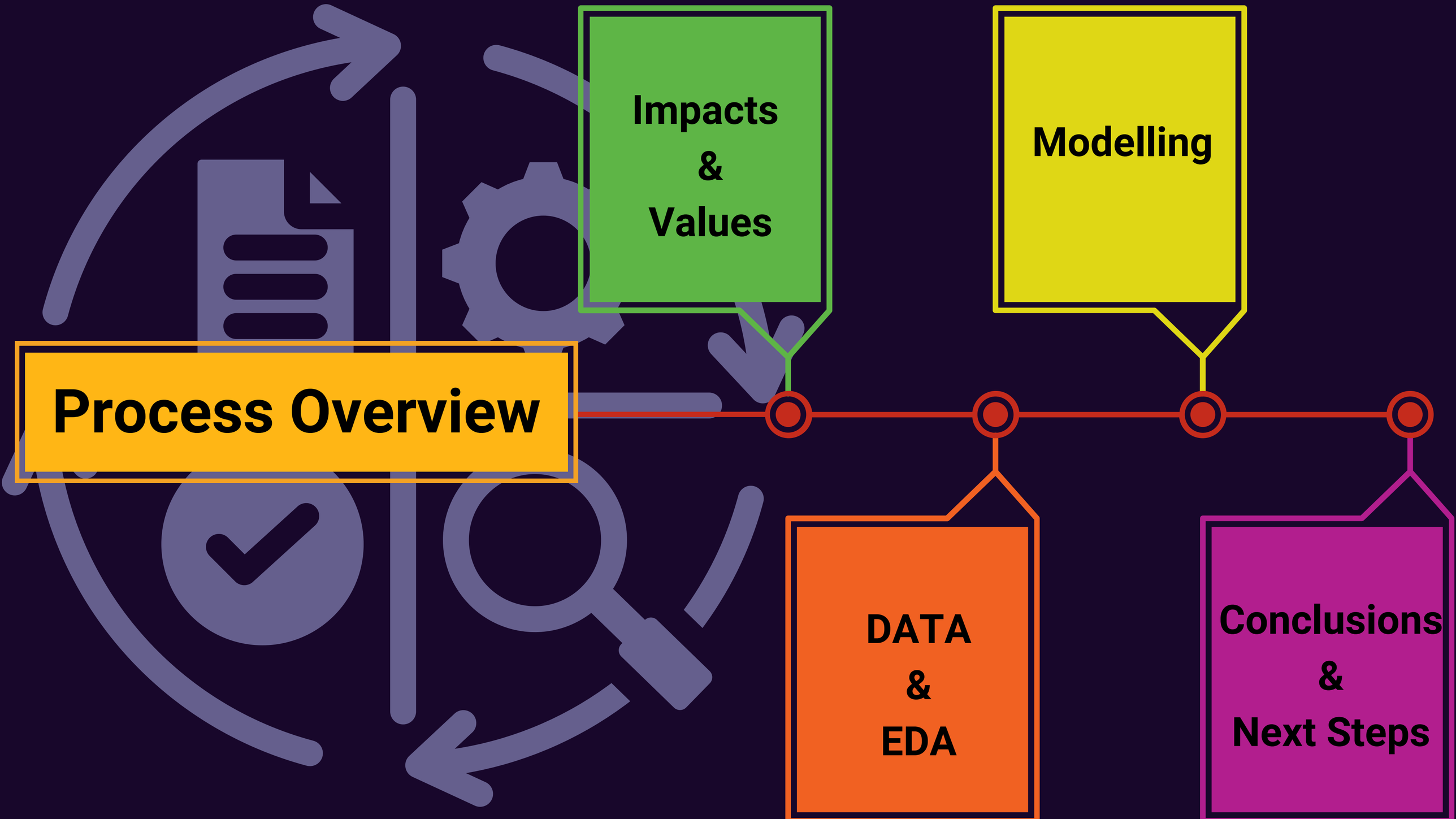
RAHIM KABA

24 NOVEMBER 2023

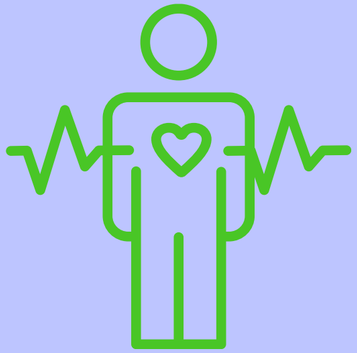
https://github.com/r-kaba/Cancer_Survivability_Predictor

The Problem:

How can we use Machine Leaning to increase the precision of prognostic estimates for cancer patients?


Process Overview
Impacts
&
Values
Modelling
DATA
&
EDA
Conclusions
&
Next Steps

Society:
Cancer prognosis affects everyone



Patients:
Accurate prognosis = better quality of life

Impacts &Values

Health Care:

- **Prognosis difficult to estimate**
- **Important for treatment and life planning**

Value:

- **Better allocation of resources**
- **Increased efficiency and accuracy of prognosis estimating**

The Data

MSK met2021 study

25,775 Patients, 55 Features

What we are tying to predict?

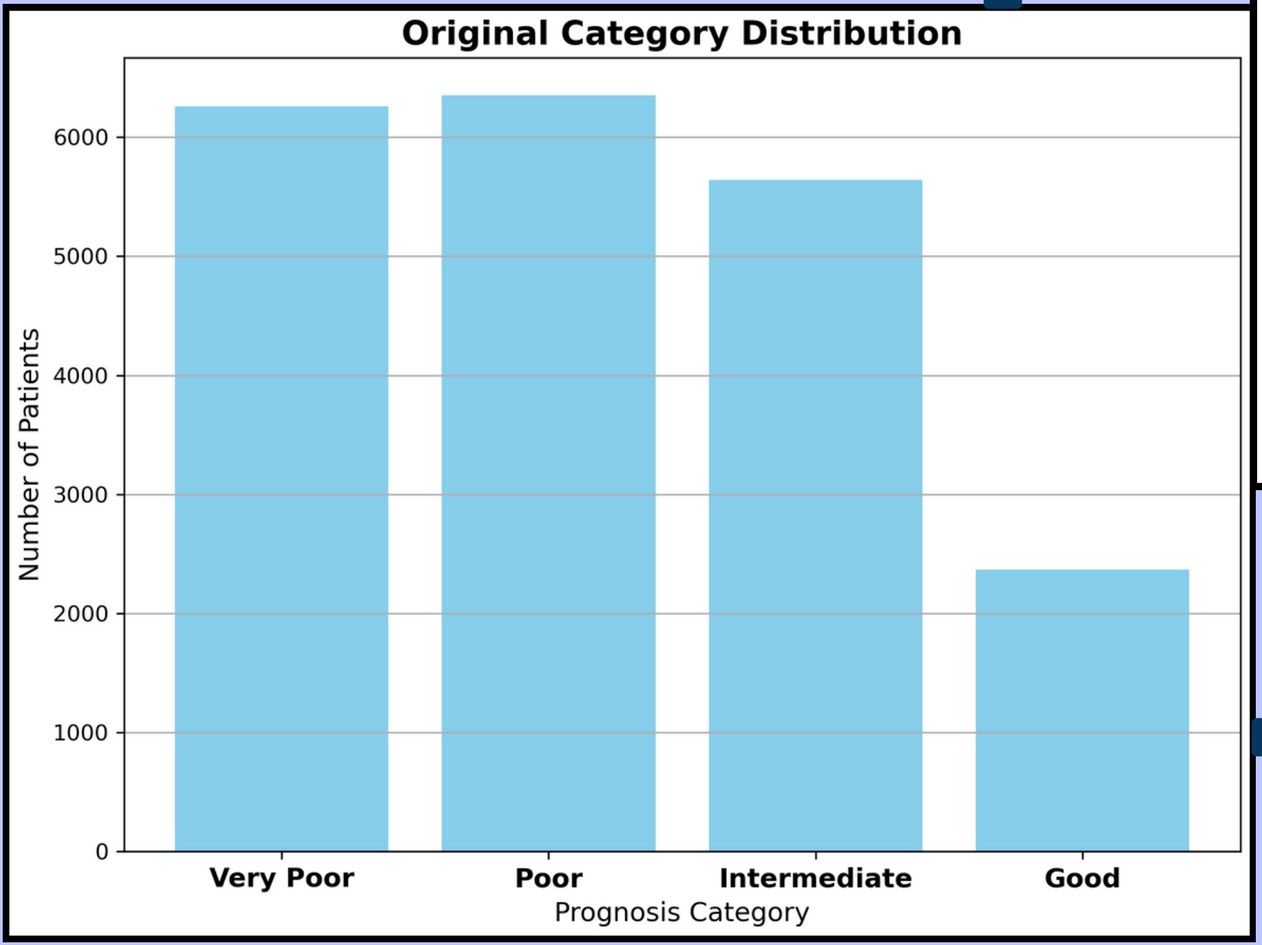
Data Processing

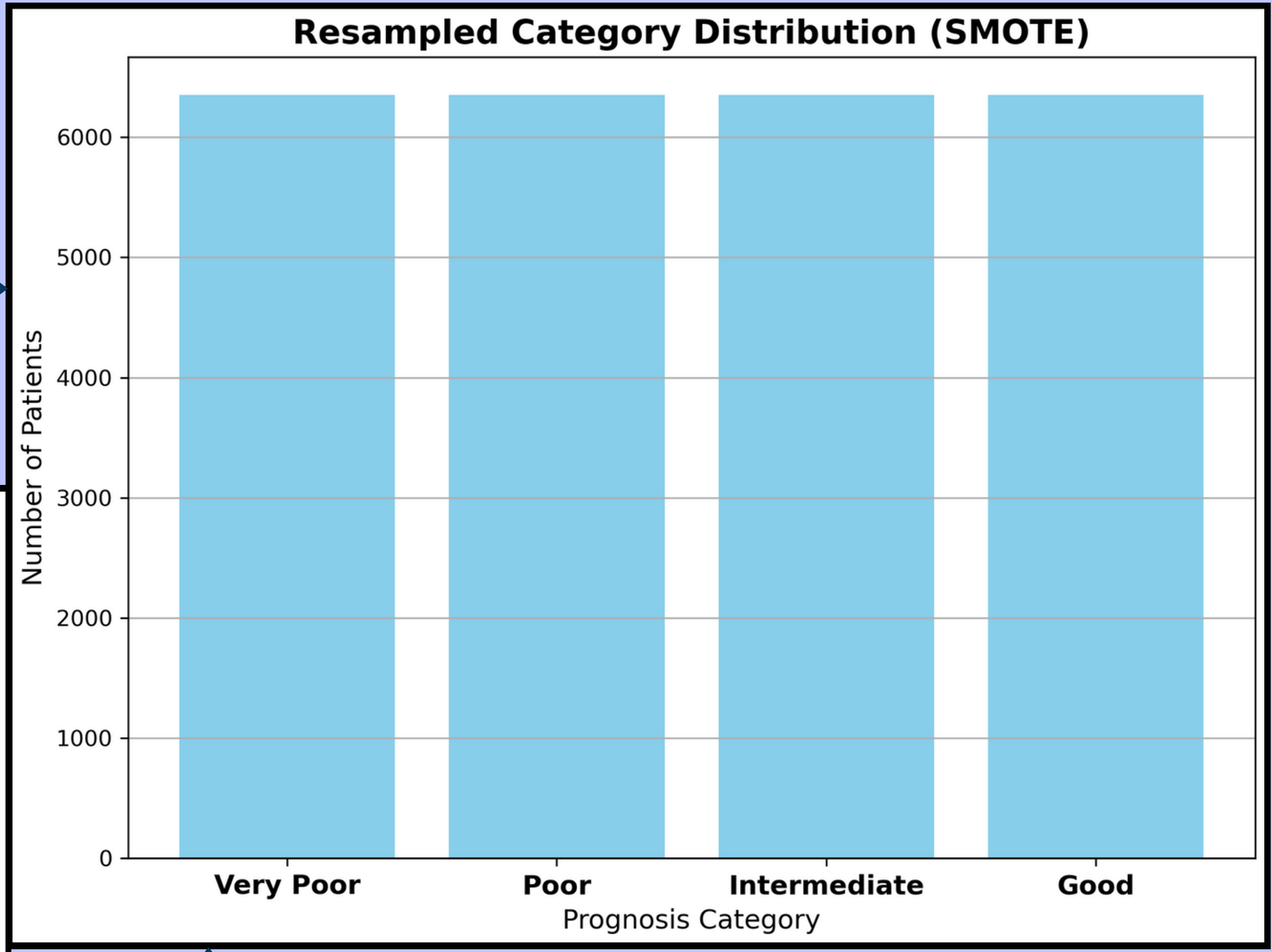
Data Cleaning

Feature Selection

Transform

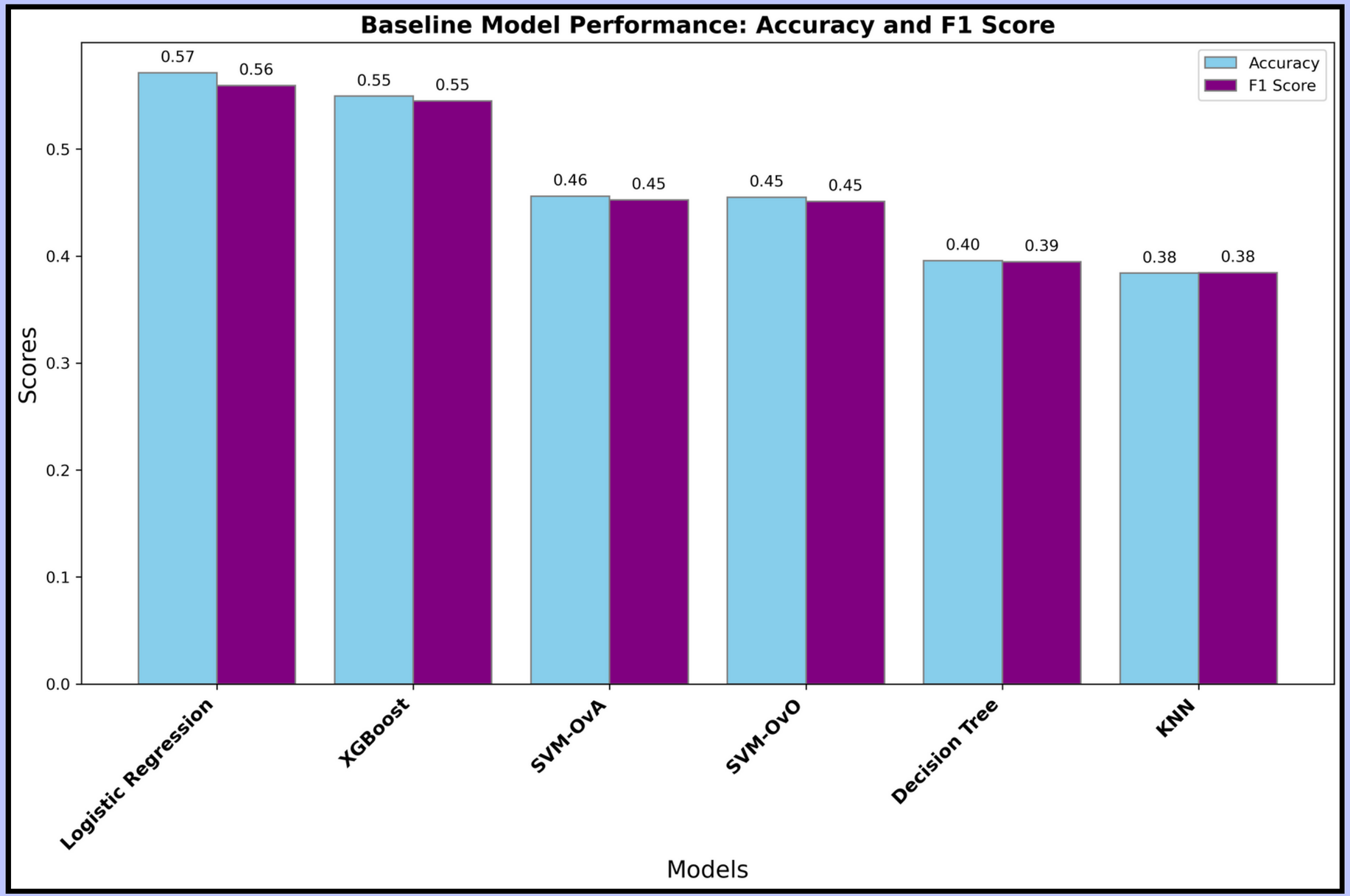
EDA


Original Category Distribution
Number of Patients
0
1000
2000
3000
4000
5000
6000
Very Poor
Poor
Intermediate
Good
Prognosis Category


Resampled Category Distribution (SMOTE)
Number of Patients
0
1000
2000
3000
4000
5000
6000
Very Poor
Poor
Intermediate
Good
Prognosis Category

Modelling Overview

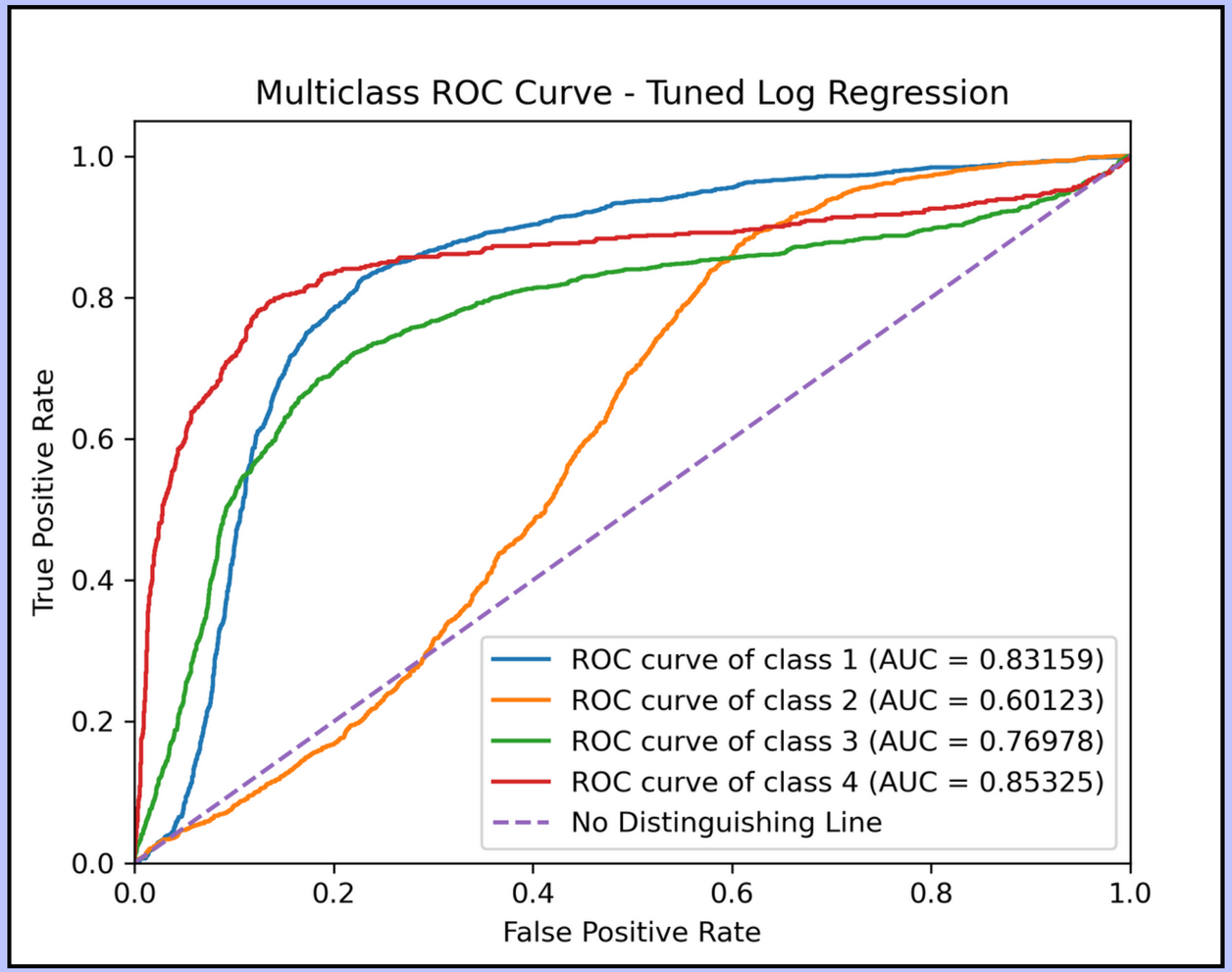
- 6 Models Tested
- Best Model by F1 and Accuracy
- Optimized Best Model



Modelling

	precision	recall	f1-score	support
1	0.60	0.62	0.61	1565
2	0.51	0.53	0.52	1588
3	0.57	0.60	0.58	1411
4	0.62	0.37	0.47	591
accuracy			0.56	5155
macro avg	0.57	0.53	0.54	5155
weighted avg	0.56	0.56	0.56	5155

	precision	recall	f1-score	support
1	0.57	0.71	0.63	1270
2	0.49	0.41	0.44	1270
3	0.54	0.38	0.45	1270
4	0.65	0.79	0.71	1270
accuracy			0.57	5080
macro avg	0.56	0.57	0.56	5080
weighted avg	0.56	0.57	0.56	5080



Next Steps

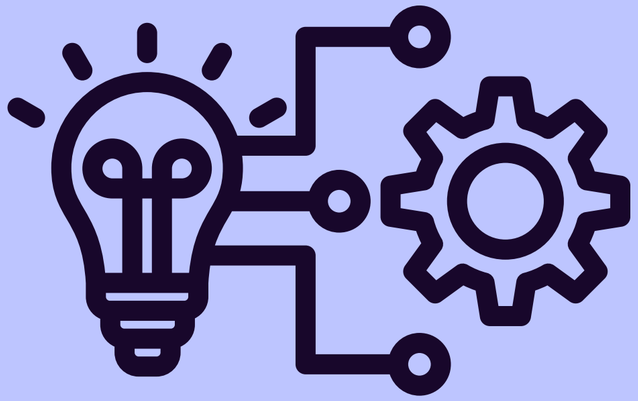
Improve and maintain the model:

- **Collect more data**
- **Add more features**

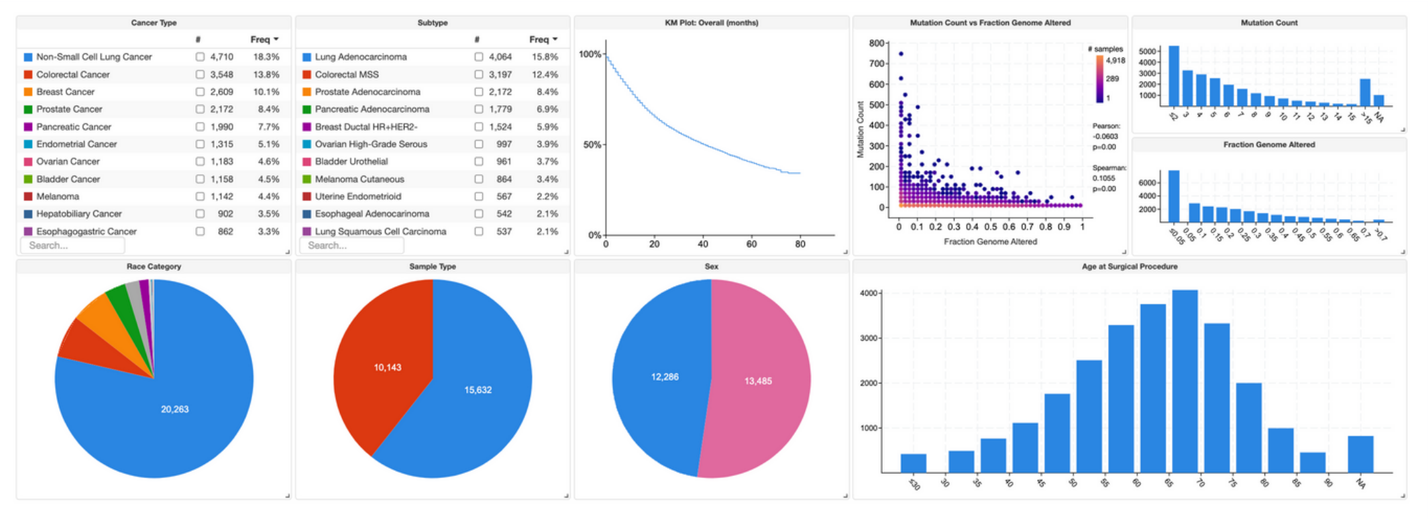
Design and Train Neural Network for more predictive power

Demo the predictor:

- **Streamlit App**
- **Tableau Dashboard**



Demo Day


Cancer Type
Non-Small Cell Lung Cancer 4,710 18.3%
Colorectal Cancer 3,548 13.8%
Breast Cancer 2,609 10.1%
Prostate Cancer 2,172 8.4%
Pancreatic Cancer 1,990 7.7%
Endometrial Cancer 1,315 5.1%
Ovarian Cancer 1,183 4.6%
Bladder Cancer 1,158 4.5%
Melanoma 1,142 4.4%
Hepatobiliary Cancer 902 3.5%
Esophagogastric Cancer 862 3.3%
Subtype
Lung Adenocarcinoma 4,064 15.8%
Colorectal MSS 3,197 12.4%
Prostate Adenocarcinoma 2,172 8.4%
Pancreatic Adenocarcinoma 1,779 6.9%
Breast Ductal HR+HER2- 1,524 5.9%
Ovarian High-Grade Serous 997 3.9%
Bladder Urothelial 961 3.7%
Melanoma Cutaneous 864 3.4%
Uterine Endometrioid 567 2.2%
Esophageal Adenocarcinoma 542 2.1%
Lung Squamous Cell Carcinoma 537 2.1%
KM Plot: Overall (months)
Mutation Count vs Fraction Genome Altered
Mutation Count
Fraction Genome Altered
Race Category
20,263
Sample Type
10,143
15,632
Sex
12,286
13,485
Age at Surgical Procedure


Cancer Survivability Predictor
Predect Survivability
SMB
Age
Met Count
Met Site Count
MGB
TMB
FTB
FMB
TGT
Style

THANK YOU

"weCAN make a difference"

weCAN

A CANCER SURVIVABILITY PREDICTOR