weCAN

A CANCER SURVIVABILITY PREDICTOR

BrainStation®

DATA SCIENCE CAPSTONE SPRINT 2

RAHIM KABA

10 NOVEMBER 2023

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

The Problem:

Increase the precision of prognostic estimates for cancer patients using data and Machine Learning

Motivations/Impacts

Why do society care?

- Cancer Prognosis affects everyone
- 50% life risk being diagnosed and 20-25% life risk of dying
- Accurate prognosis = increased quality of life

Why do I care?

- Personal connection
- Working in cancer health care, I understand the impact of accurate prognosis

Why does industry care?

- Prognosis can be difficult to determine
- Important for determining treatment and life planning
- Better allocation of resources

The Data

MSK met2021 study

25,775 Patients, 55 Features

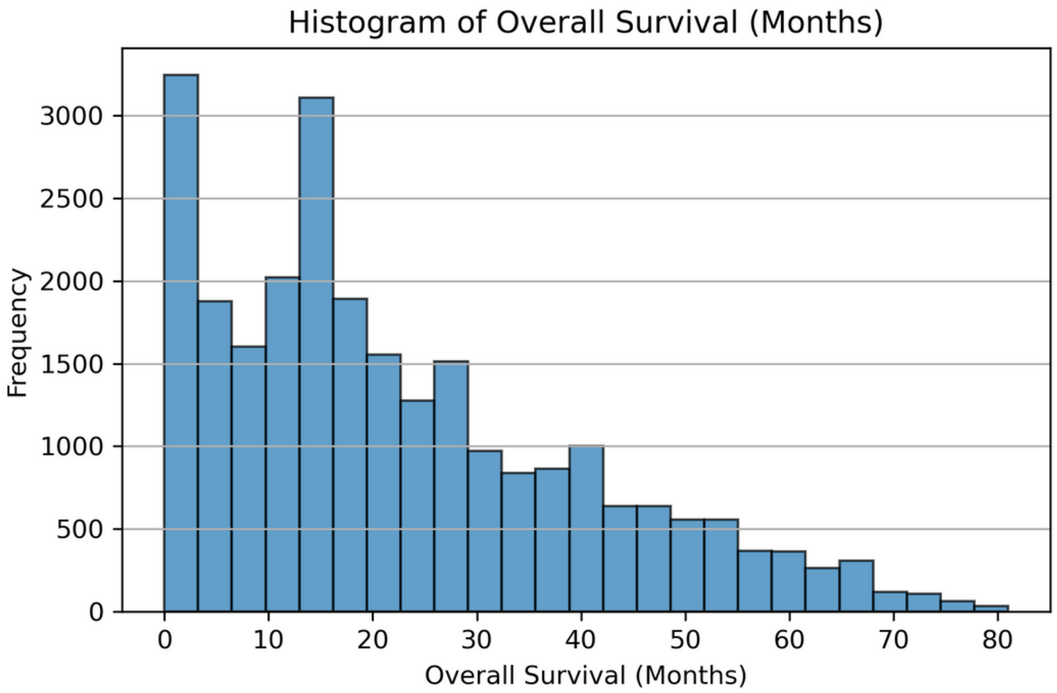
What we are tying to predict?

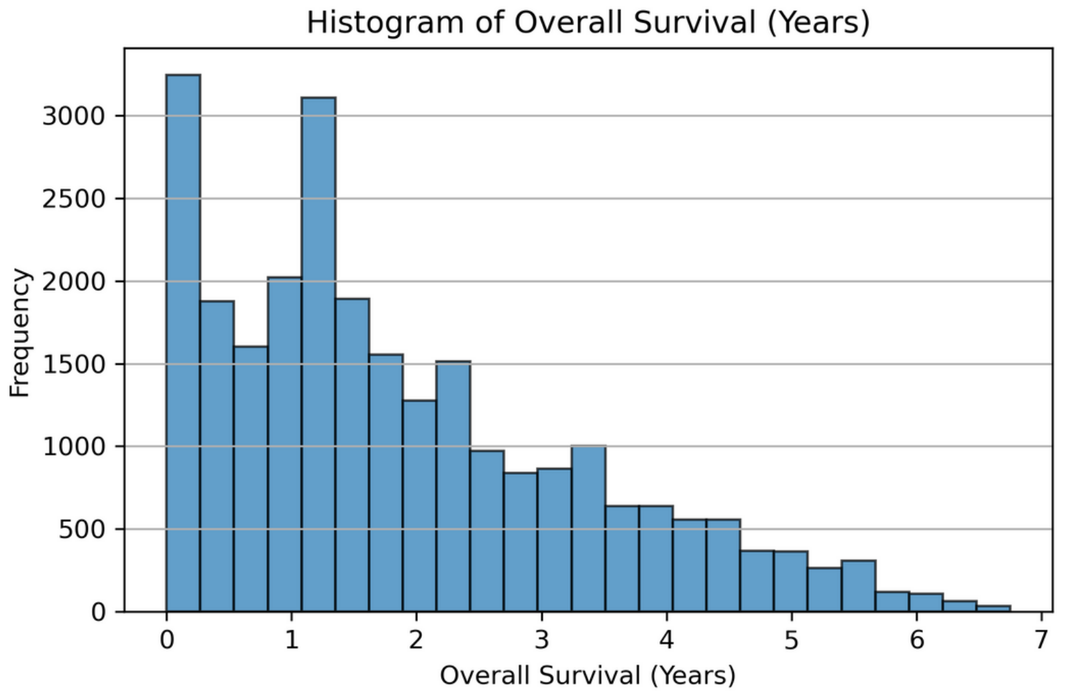
Data Processing

Data Cleaning

Feature Selection

Transform



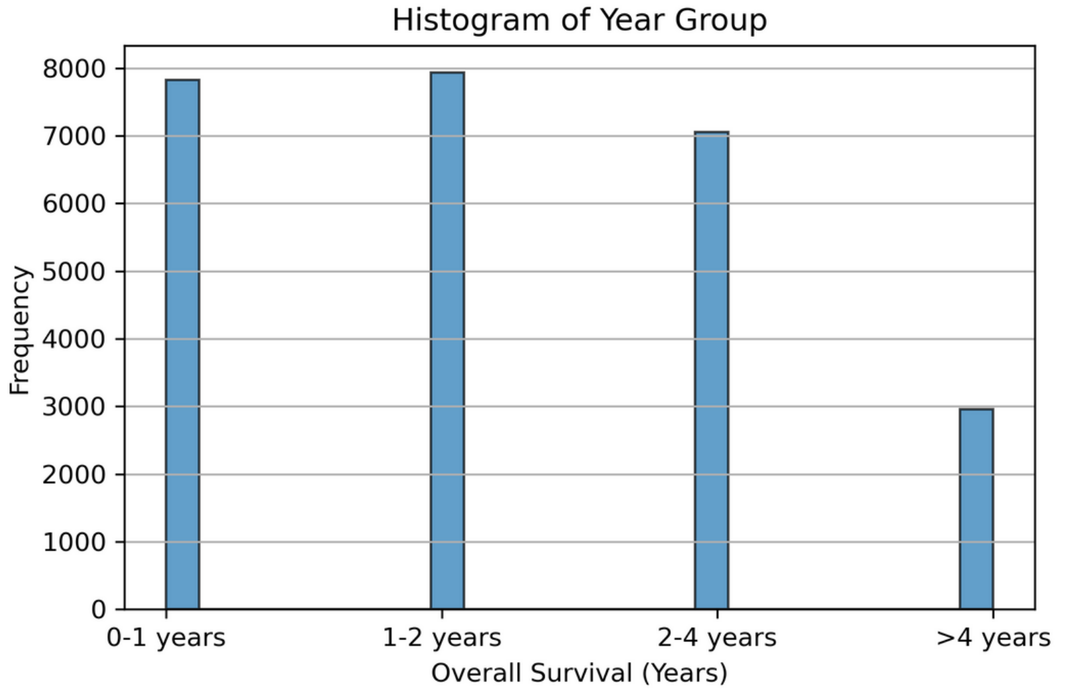


EDA Findings

- **Highly correlated features**
- **Missing values**
- **Some features not important for predicting**

What we are trying to predict: Overall Survivability

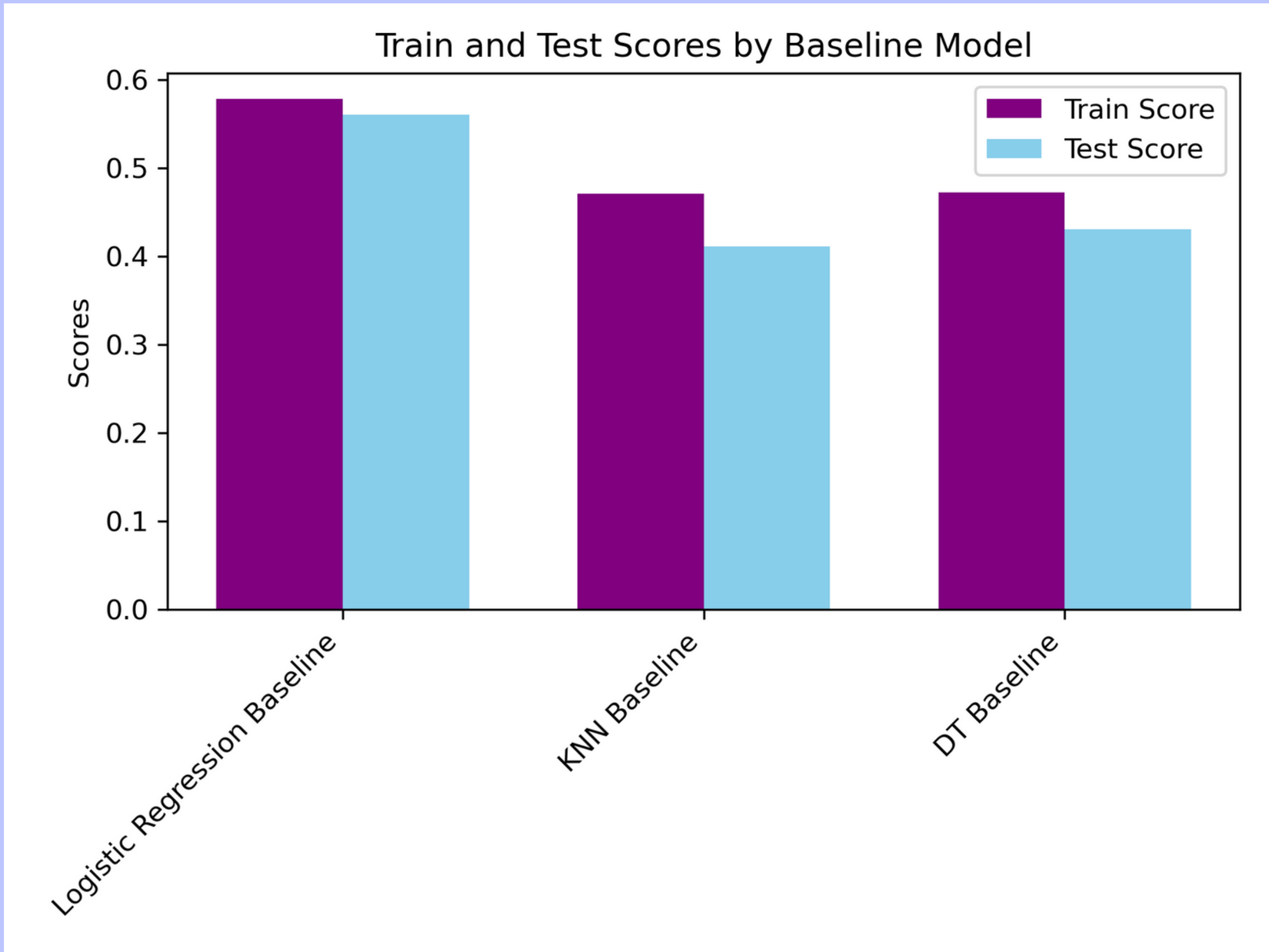
- **Month > Years > Groups**



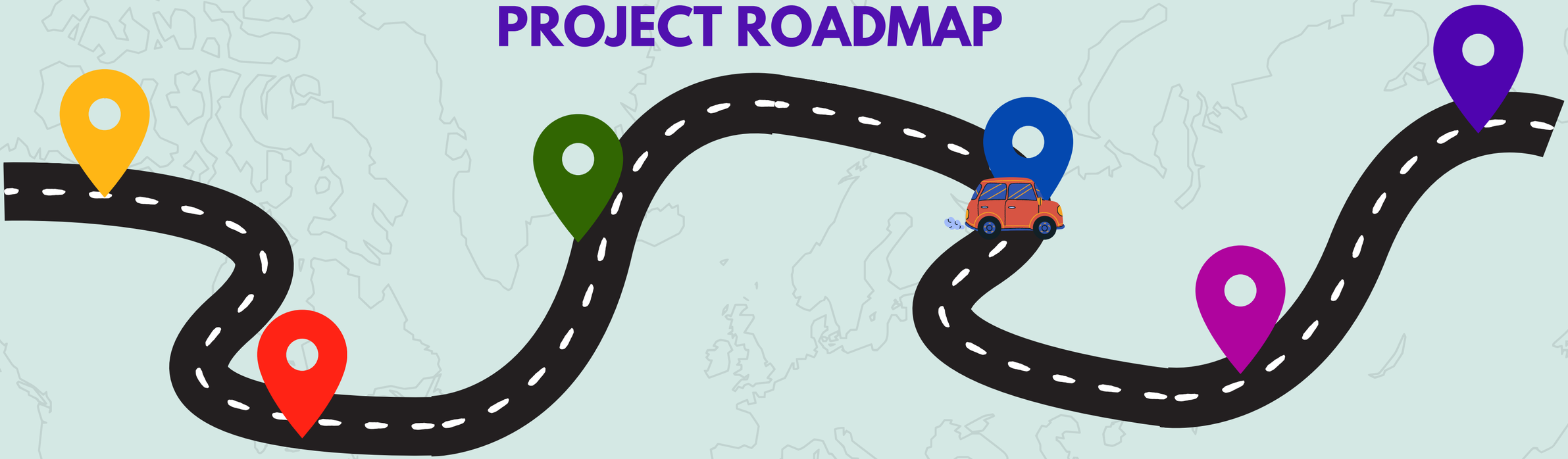
Baseline Modelling

3 baseline models:

- **Multi Class Logistic Regression**
- **K- Nearest Neighbors**
- **Decision Tree**



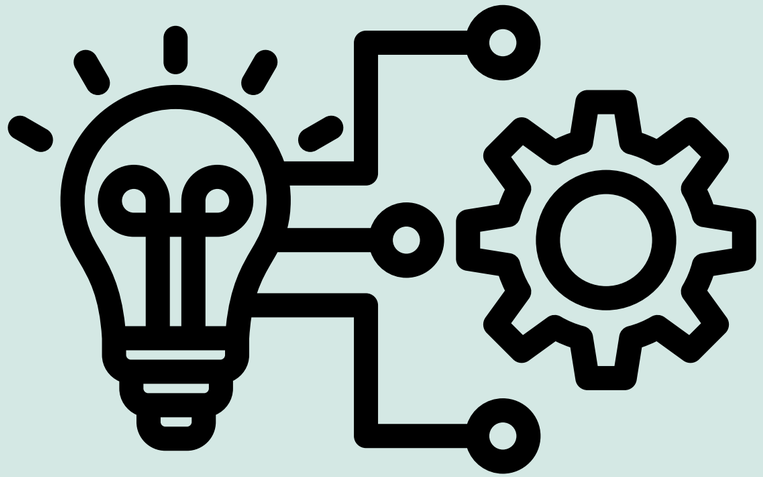
PROJECT ROADMAP



1. DATA COLLECTION
2. DATA CLEANING
3. EDA / FEATURE ENGINEERING
4. MODEL SELECTION/ PARAMETER TUNING
5. MODEL EVALUATION
6. MODEL DEPLOYMENT

Next Steps

MORE FEATURE ENGINEERING



OPTIMIZING BEST MODEL

MODEL EVALUATION AND DEPLOYMENT

THANK YOU

"weCAN make a difference"

weCAN

A CANCER SURVIVABILITY PREDICTOR