

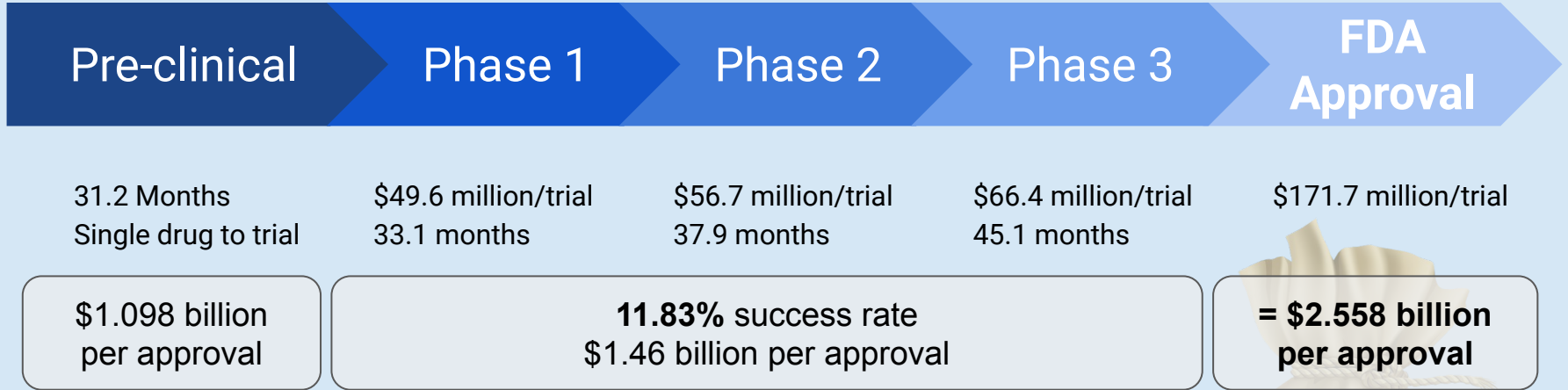


How will it end?

**Predicting whether or not cancer clinical trials
make it to completion**

**Team Pioneer: Aziz Burak Guelen, Kriti Sehgal, Asia Wyatt
Spring 2022 Erdős Institute Data Science Bootcamp**

The why



Project goal: Determine what are the most important features for a cancer intervention clinical trial to be completed or not and predict whether or not a test set of clinical trials will be completed

Clinical Trial Data and Feature Selection

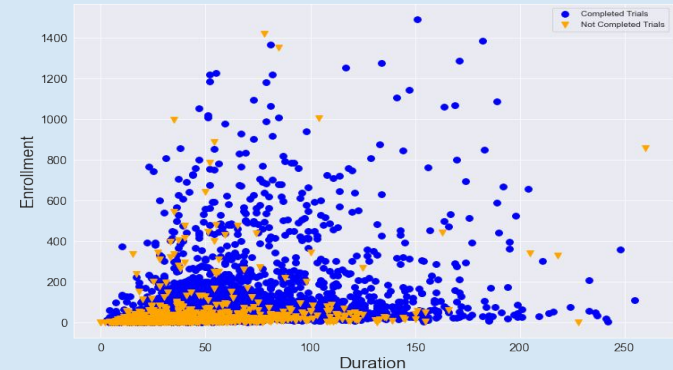
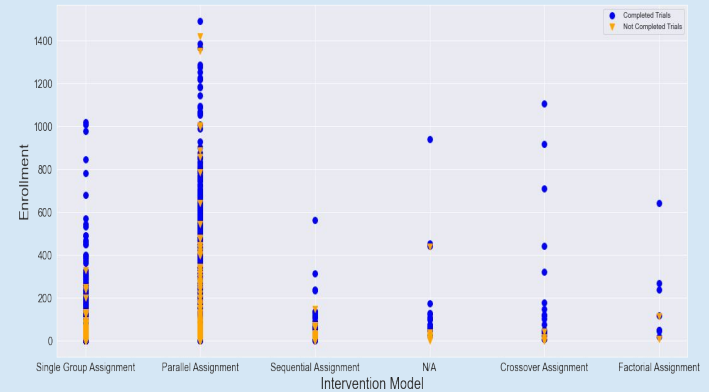
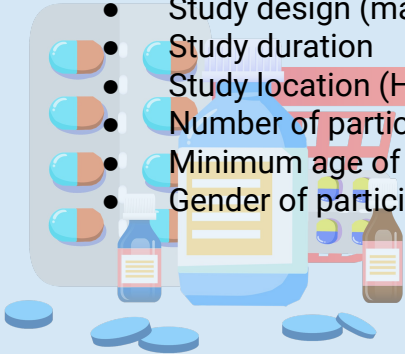
Data is taken from published <https://clinicaltrials.gov/> data

Criteria for data selection:

- Cancer
- Studies with results
- Interventional Studies

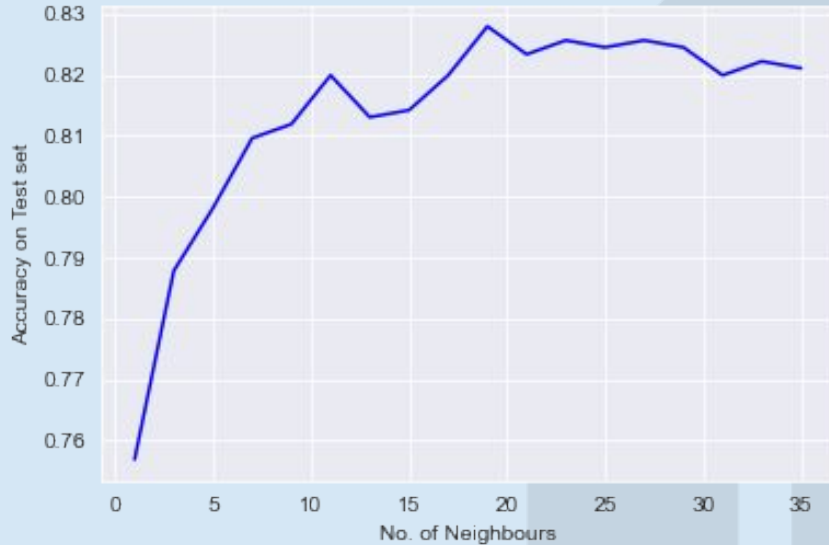
Features we chose to examine:

- Conditions (cancer types the therapy is tested for)
- Funder Type (NIH/Federal funding, Industry, or other)
- Study design (masking, etc)
- Study duration
- Study location (Hospital/University or not)
- Number of participants
- Minimum age of participants
- Gender of participants



Cancer Trial Completion Classification

KNN: K Selection

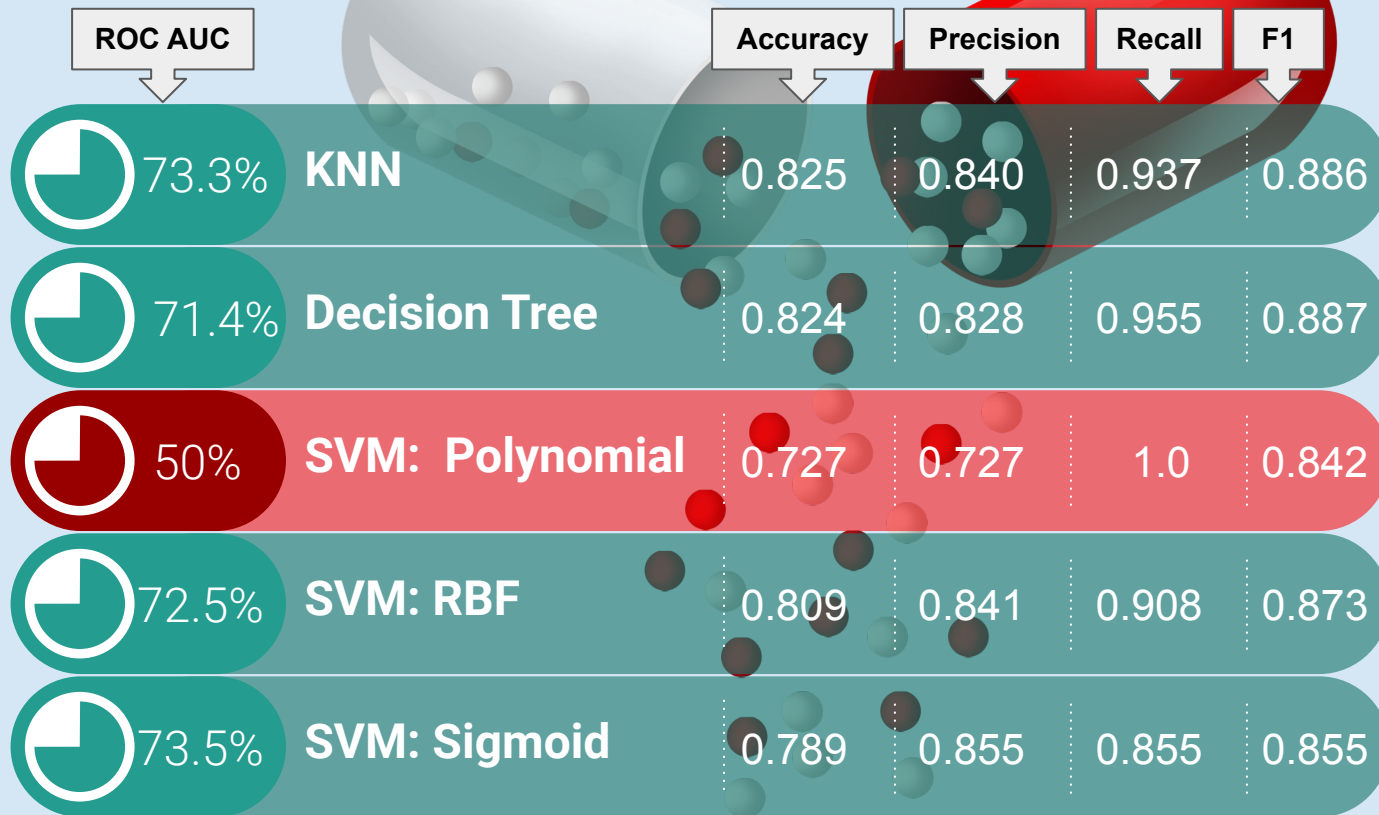


Decision Tree: Branch Selection



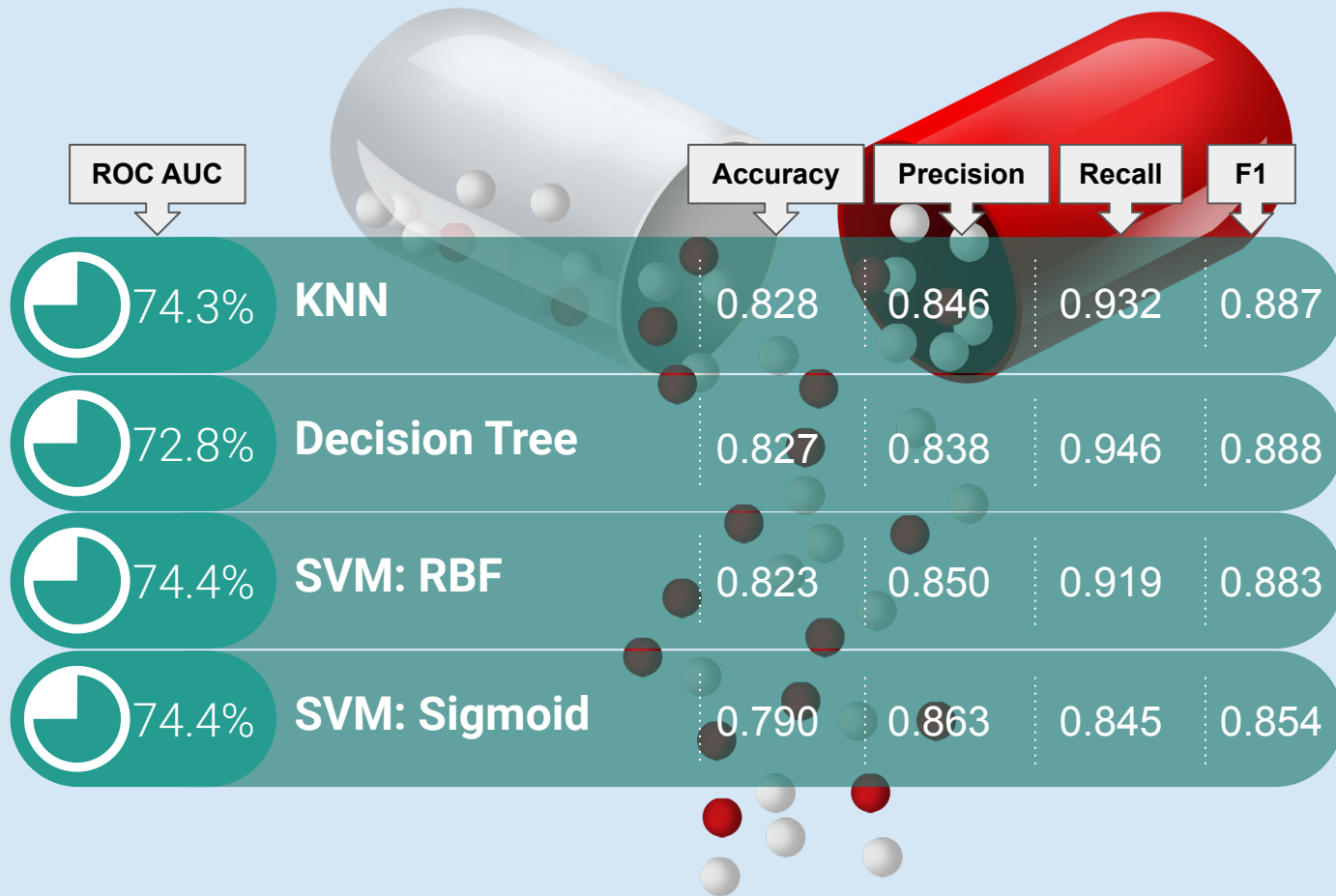
Training Results

The final metric used is the ROC AUC for the training set. This metric averages the true positive rate and the true negative rate allowing to see how good the model is at predicting the completed trials as completed and non-completed trials as non-completed.



Test Results

Removing the polynomial SVM, we see similar results across remaining model types





That's how it ends!

Team Pioneer:

Aziz Burak Guelen: guelen.1@osu.edu

Kriti Sehgal: sehgal.50@osu.edu

Asia Wyatt: awyatt314@gmail.com

<https://github.com/Asiaw Wyatt/Erdos-Project2022>