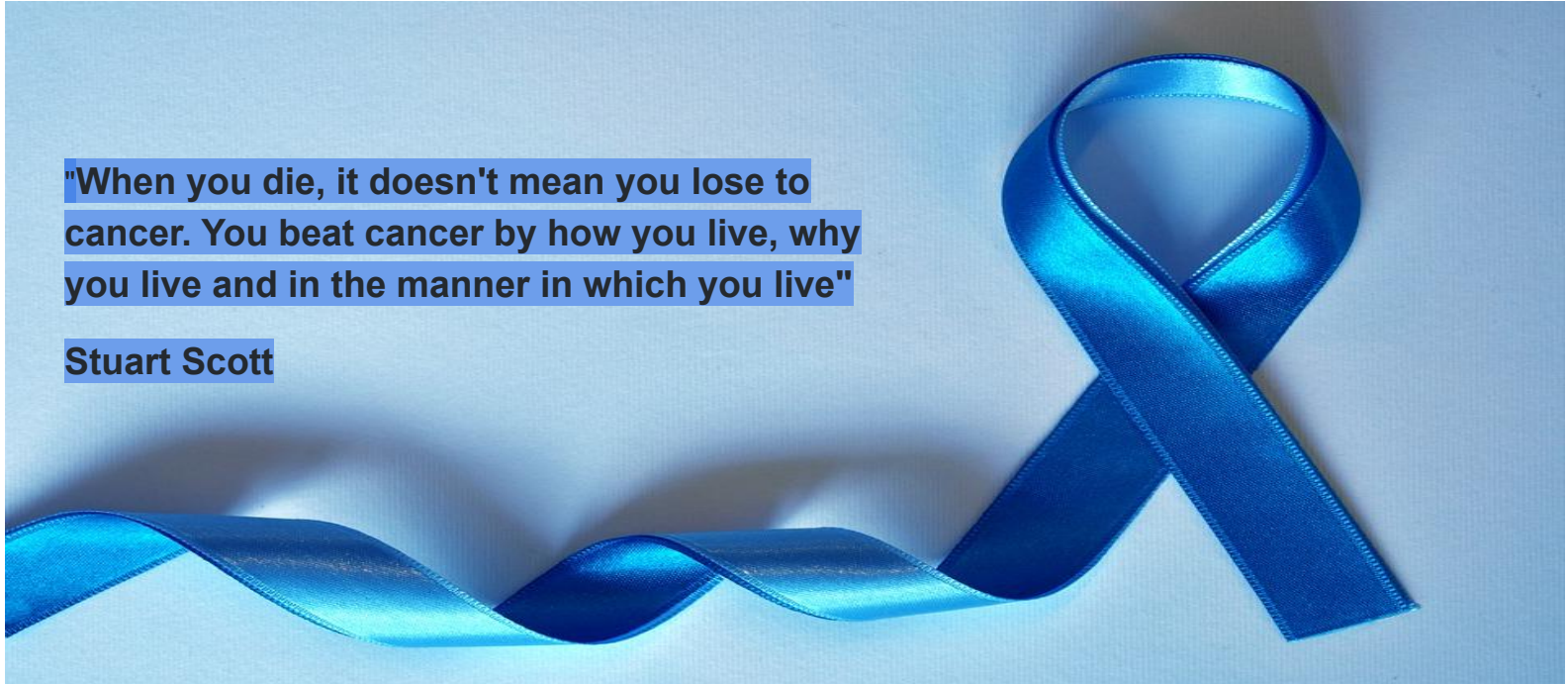

Predicting Cancer Patient Survival

A guide by Chip Heath & Dan Heath

Quality or Quantity

"When you die, it doesn't mean you lose to cancer. You beat cancer by how you live, why you live and in the manner in which you live"

Stuart Scott

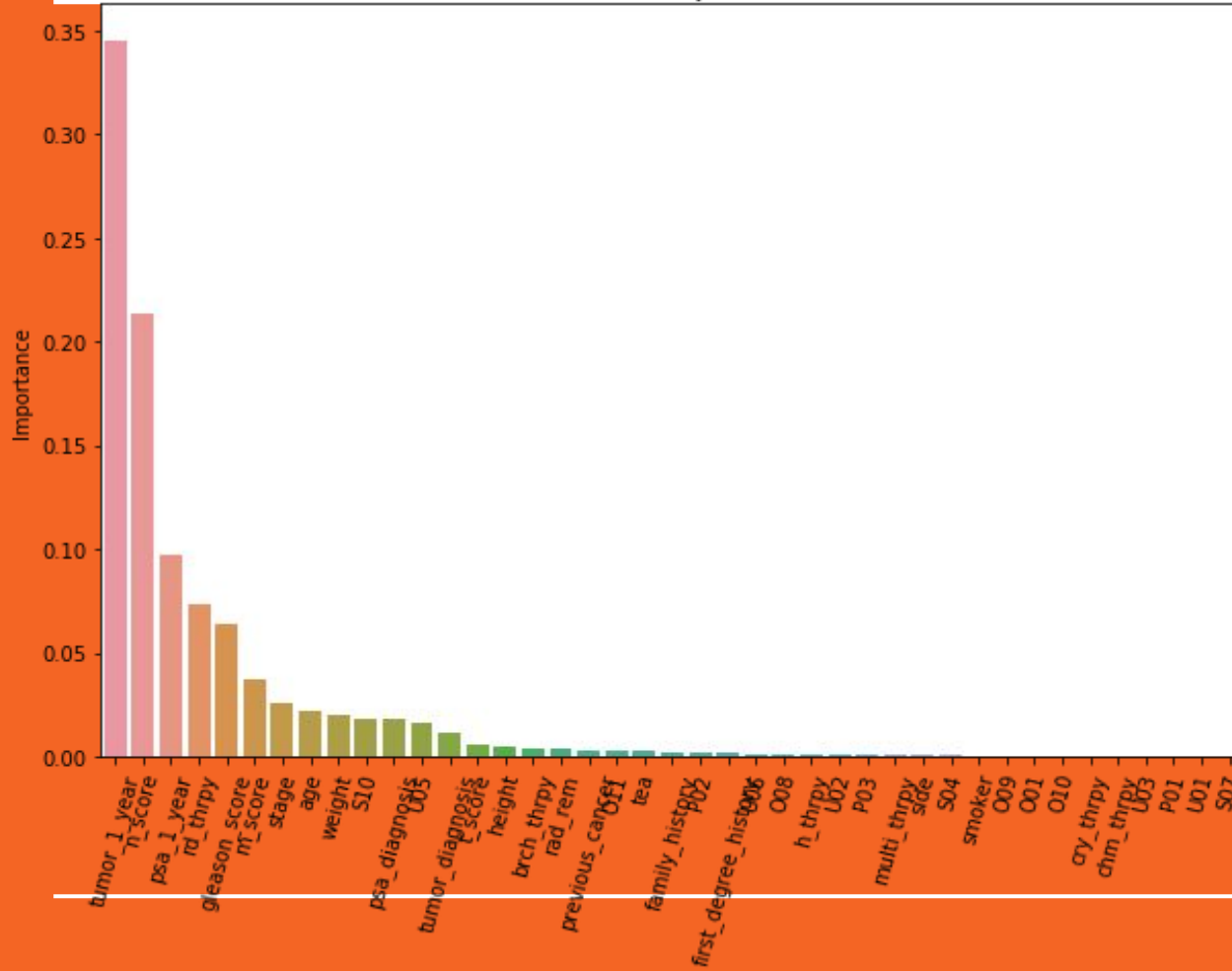




Objective

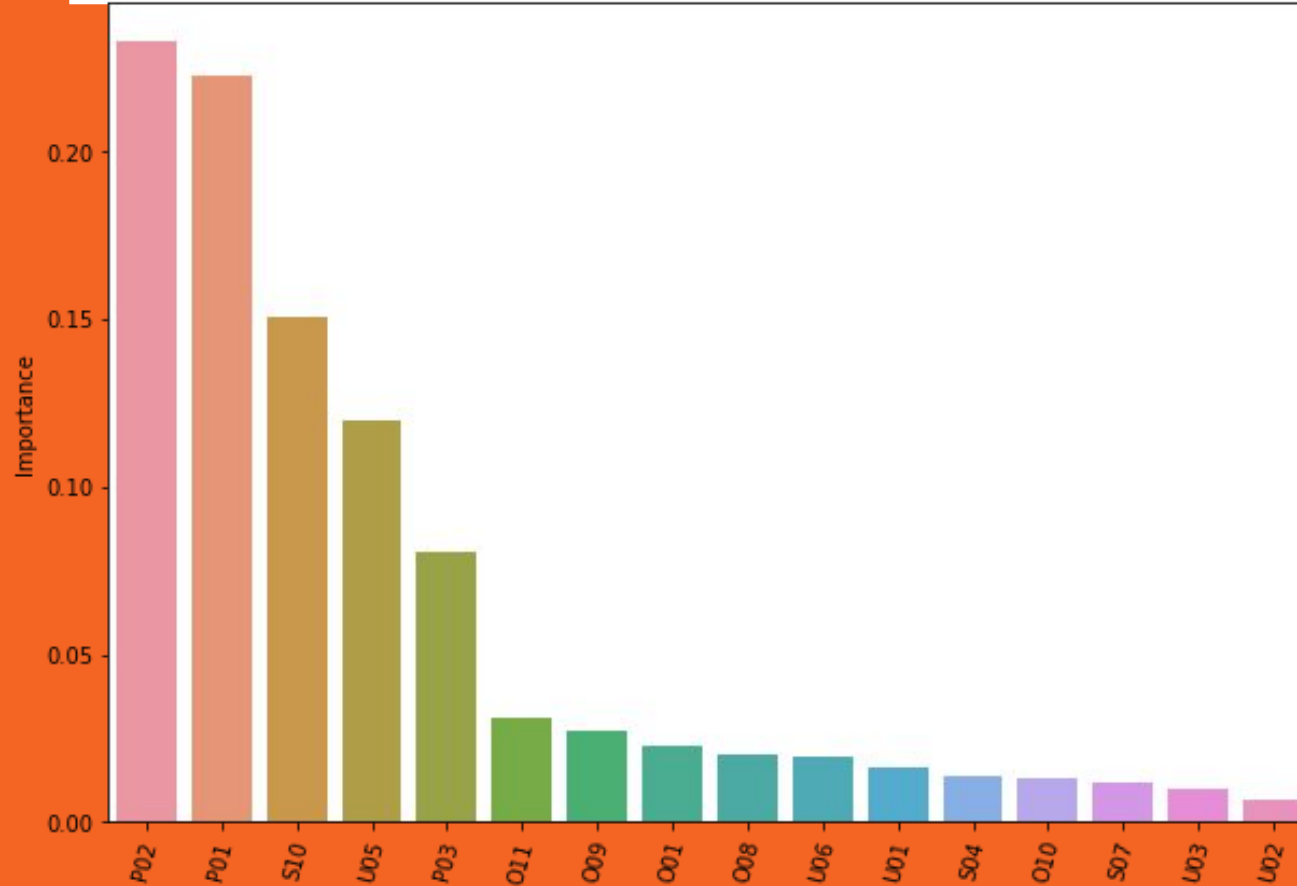
- Create a model to better predict the 7-year survival of prostate cancer patients using the data available to us at time of prognosis and 1yr check up.
- Provide some insight as to which feature has a drastic effect

Feature Importance



- 1) Tumor Size in 1year (mm)
- 2) Cancer has spread to the lymph nodes
- 3) Blood level of prostate-specific antigen after 1 year (ng/mL)
- 4) A measurement of how Cancer cell abnormality
- 5) Cancer has spread to distant parts of the body

Symptoms Feature Importance



1) P02

2) P01

3) S10

4) U05

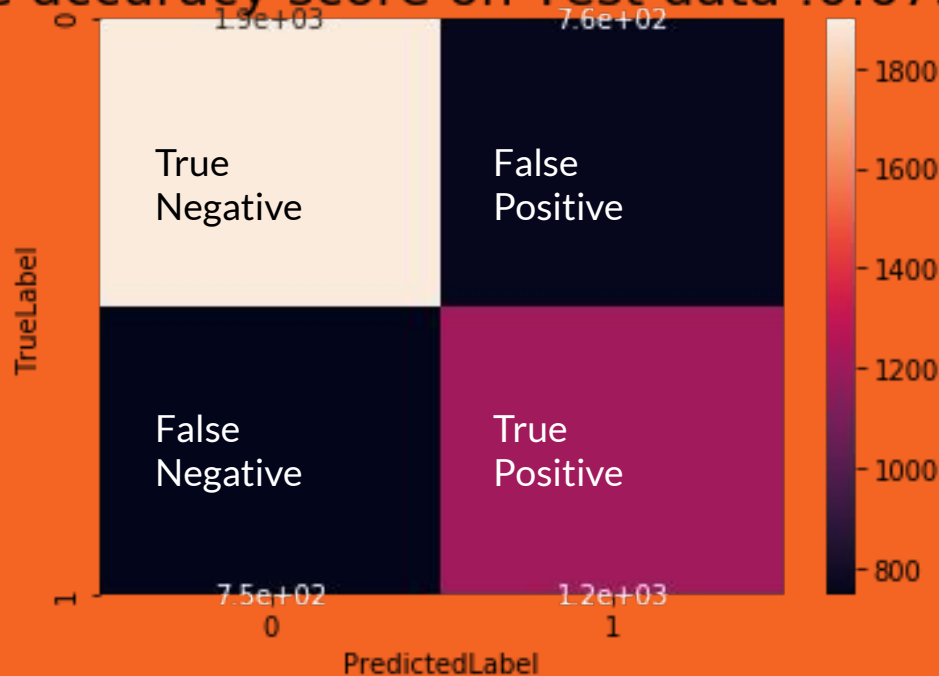
5) P03

6) O11

These codes had most effect on the prediction. Had a seooerate analysis using Symptoms

Final Model Accuracy

The accuracy score on Test data :0.673



GradientBoostingClassifier

Percent Accuracy on a test data was of 67.3%

Using Top 11 Important features + Top 6 Important Symptoms

This Model can be used to predict 7yr survival

Recomendations

If the Size of primary tumor 1 year after diagnosis is greater than 33.23 mm then chances are lower of survival.

Early Detection and controlling the rate of growth

If Level of prostate-specific antigen in blood 1 year after diagnosis, is greater than 8.48 ng/mL then chances of survival is less.

Control and monitor the prostate-specific antigen

While analyzing symptoms, I found that death percentage increases if people are having "S01,P01,P02 and P03". (Where as chances of survival is more if the person is having O11 as a symptom.) still have to verify

Future Work

WORK IN PROGRESS

- Using rate of change in the tumor size in analysis
- Understanding the symptom codes
- Understanding current research and how prediction can be better



Thank You!

Q&A

Simranjit Nepaul

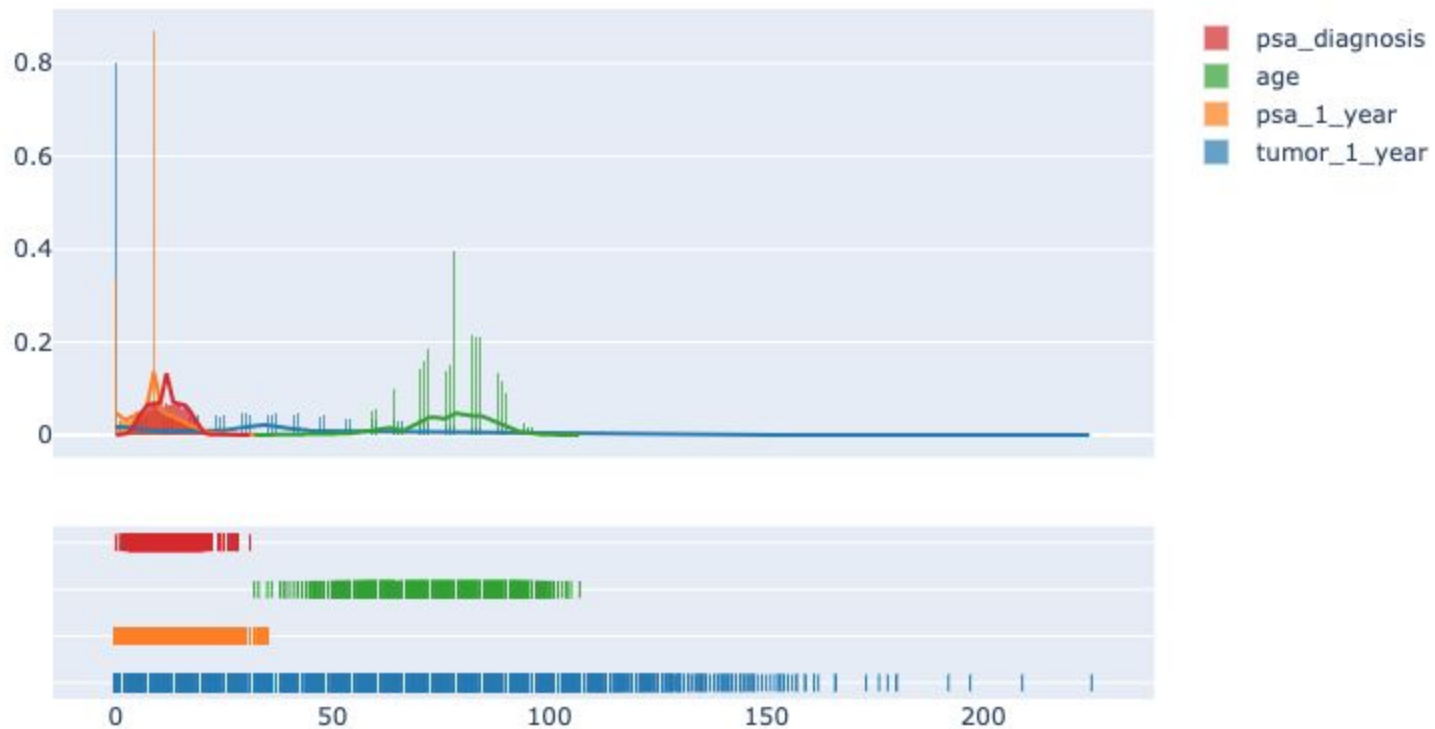
Module 3 Presentation

[GitHub](#)

[linkedin.com/in/ssnepaul](https://www.linkedin.com/in/ssnepaul)

Appendix

Data Distribution



Aggregations Plots

use dropdown to change aggregation

avg

