



Prediction of Cervical Cancer Risk

JiEun Song
Feb 12, 2020

Introduction

- Fourth most common type of cancer among women
- More than 10,000 new cases in the US every year
- Most preventable type of cancer

Regular screening!

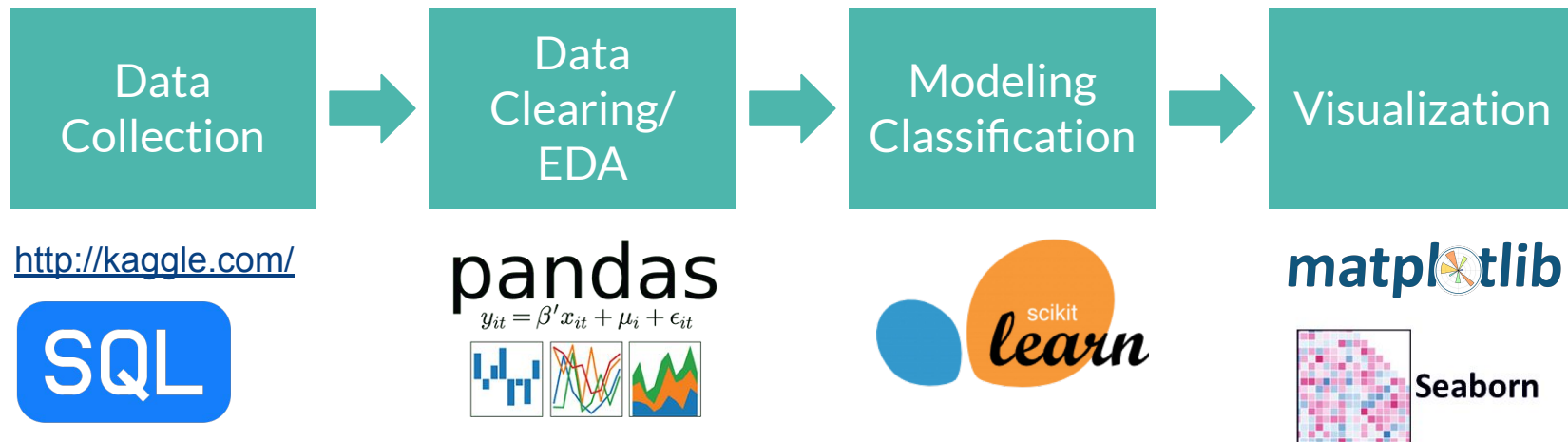


Goal

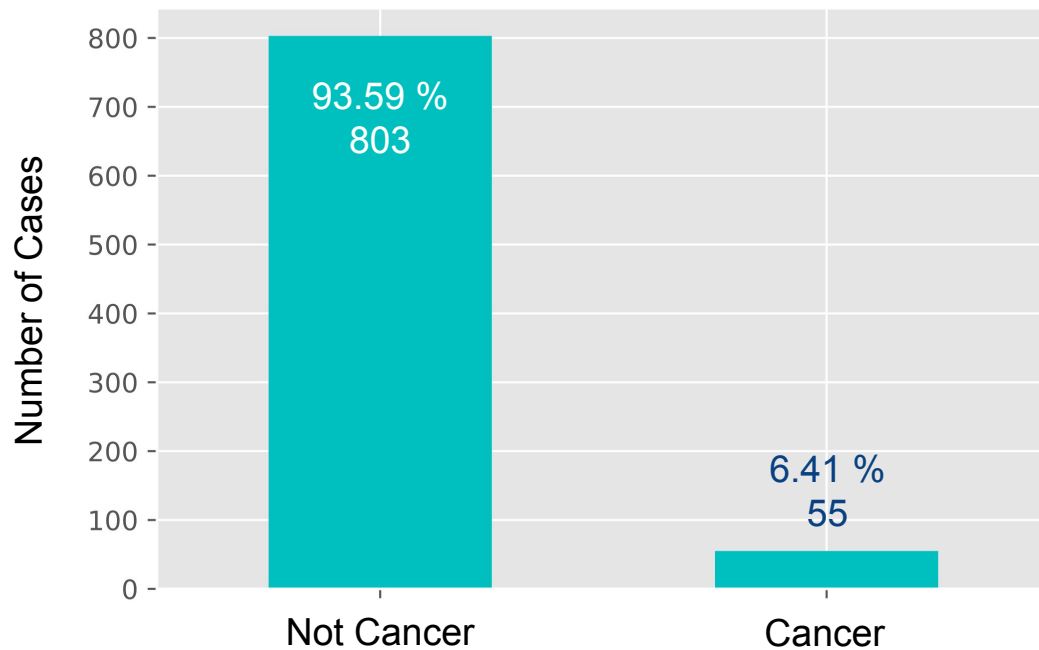
- Build a model to predict cancer positive case
- Recommend to potential patients more frequent screening



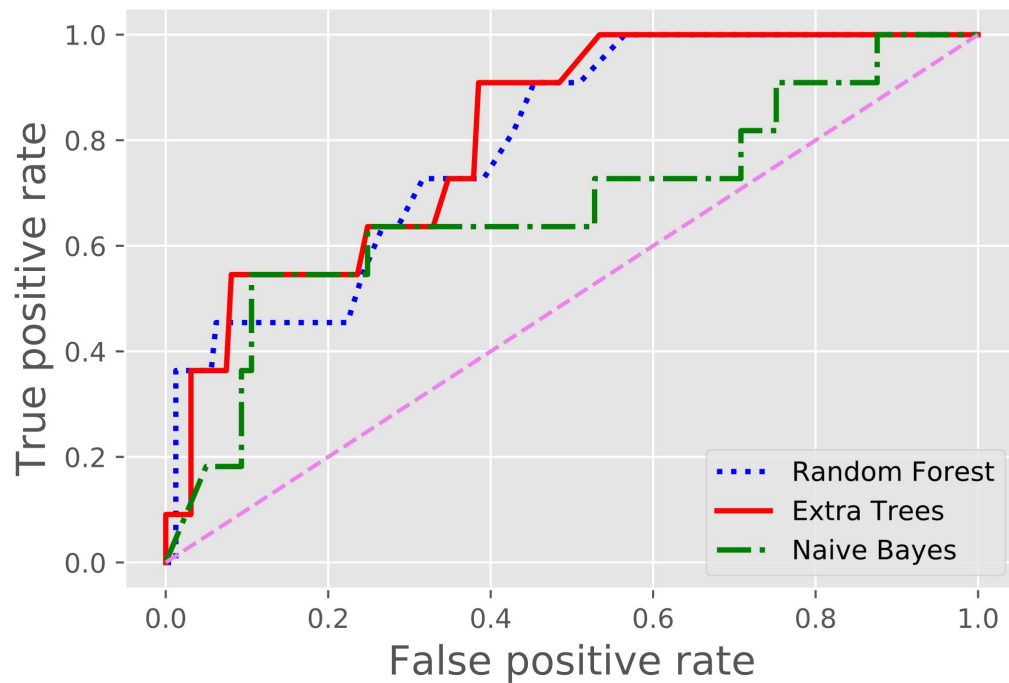
Methodology



Class Imbalance



ROC curve

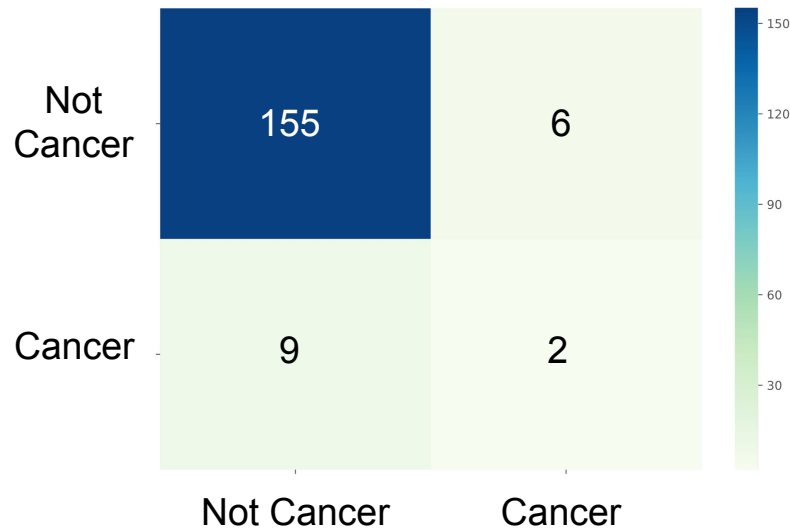


ROC-AUC score
RF: 0.79
ET: 0.81
NB: 0.68

Confusion Matrix

Random Forest

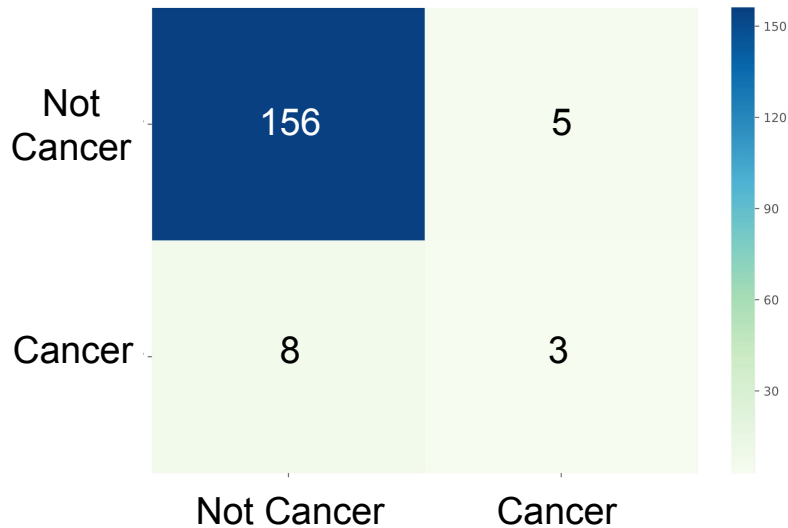
Ture Label



Predicted Label

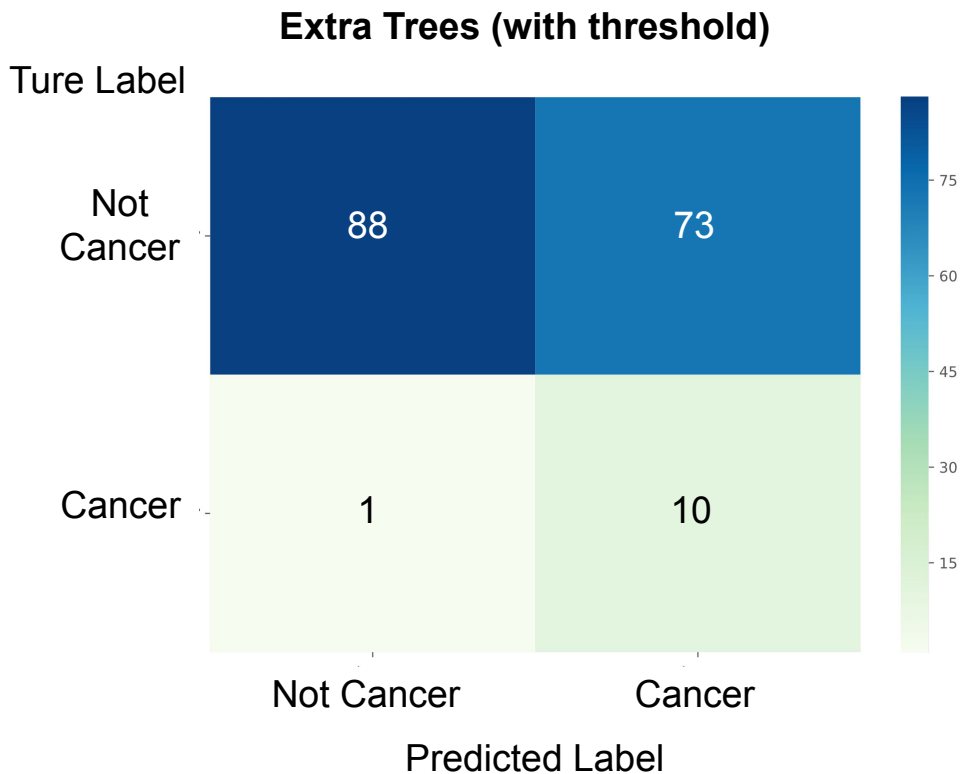
Extra Trees

Ture Label



Predicted Label

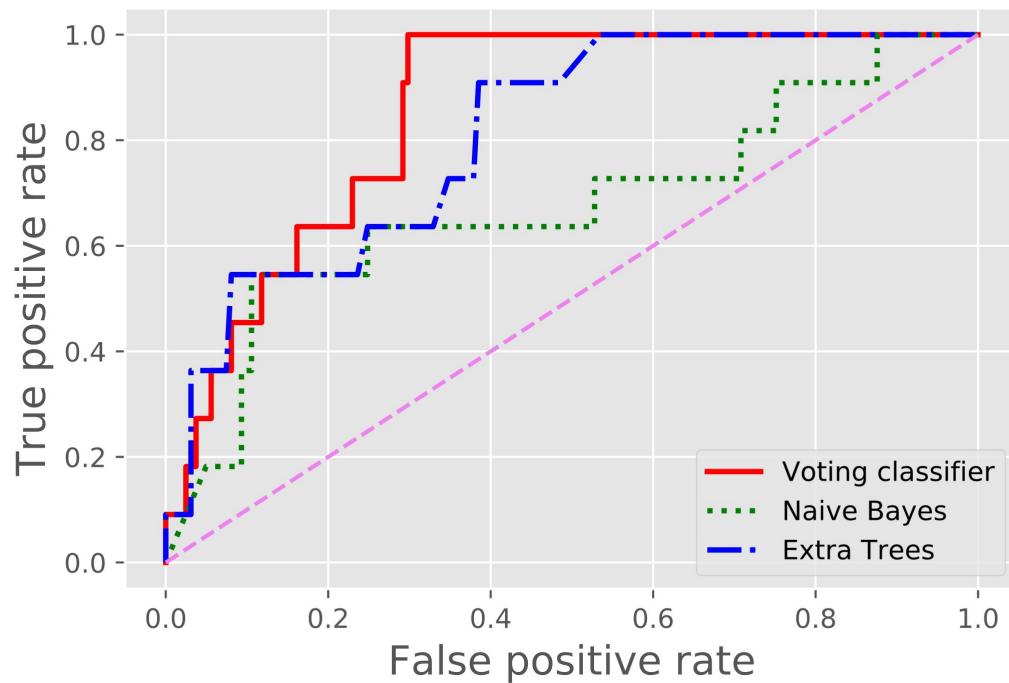
Confusion Matrix



Methodology

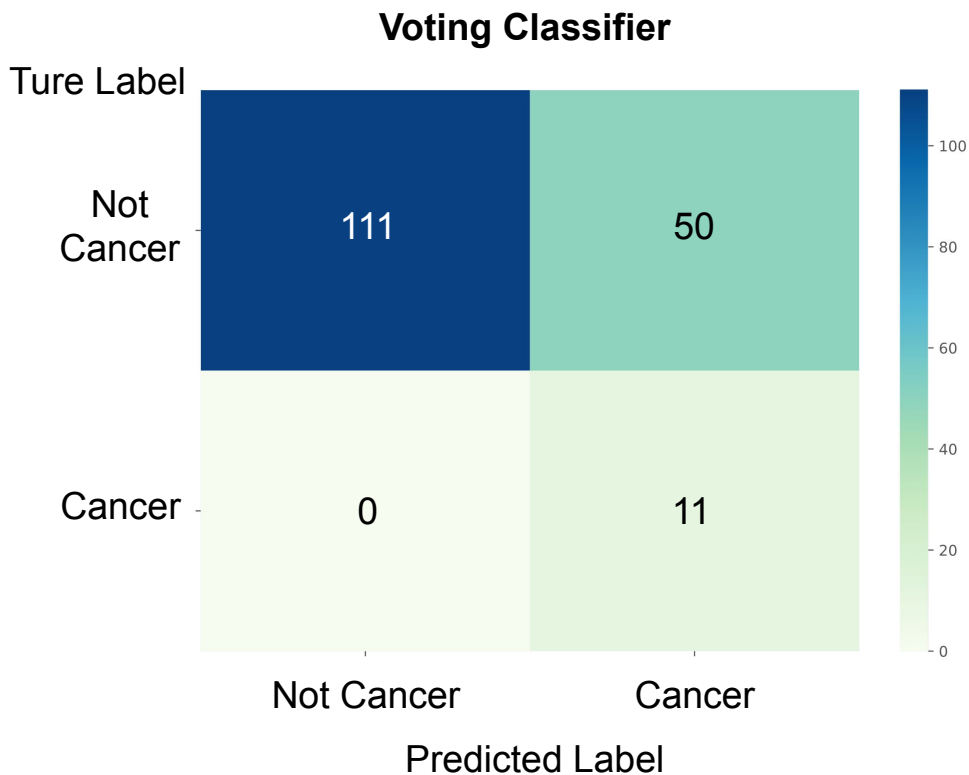


ROC curve



ROC-AUC score
VC: 0.84
ET: 0.81
NB: 0.68

Confusion Matrix



Conclusion

- Most important risk factors:
 - Number of pregnancy
 - Hormonal contraceptives
 - Number of partners
 - Age

Future Works

- Add more data to improve the model
- Feature engineering
- Train with different model



Questions?