

## Brief overview of our exploration for the Data Mining project on bankruptcy prediction

Presented By

# DARing ANALysts

Darshan Upadhyay  
Ananth Mohan



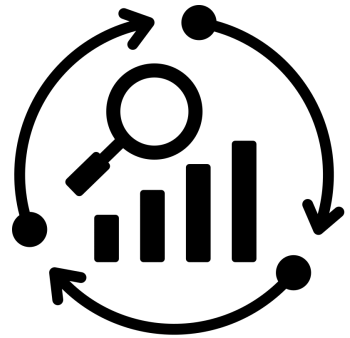
# Problem Statement

The goal is to develop a predictive model using different models in SAS EM to foresee a firm's financial condition and **probability of bankruptcy**, using 64 given financial metrics



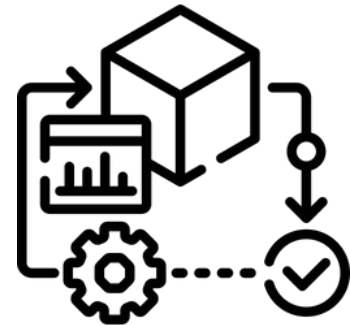
<b>Training Data</b>	10,000 samples
<b>Test Data</b>	8,000 samples
<b>Attributes</b>	64 financial ratios and metrics, e.g., Net Profit / Total Assets, Total Liabilities / Total Assets, etc.
<b>Response Variable</b>	Binary classification - Class (Y): 0 (Did not bankrupt: 9,789), 1 (Bankrupt: 211)
<b>Missing Values</b>	No missing values in the dataset

# Model Selection



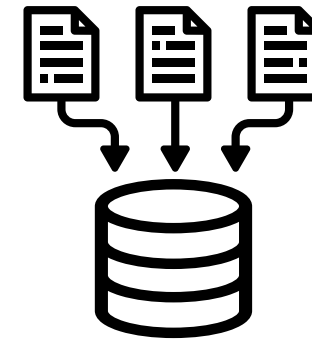
## Preprocessing

- Transformation
- Filtering
- Sampling
- **Replacement**
- Variable selection



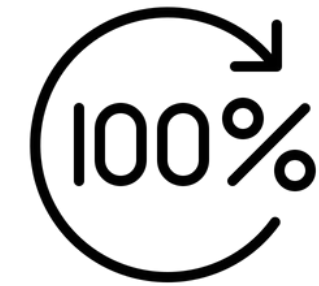
## Models

- LASSO, and Logistic **Regression** (Forward, **Backward**, Stepwise)
- Decision Trees
- **Neural Networks**
- **Gradient Boosting**
- HP Models



## Ensemble Models

Built ensemble models by merging predictions from the best individual models.



## Training with 100% Data

- Decided to train our best model on the full 100% data, which performed really well on the public leaderboard

Validation ROC Index	0.909
Test ROC Index - Public	0.94303

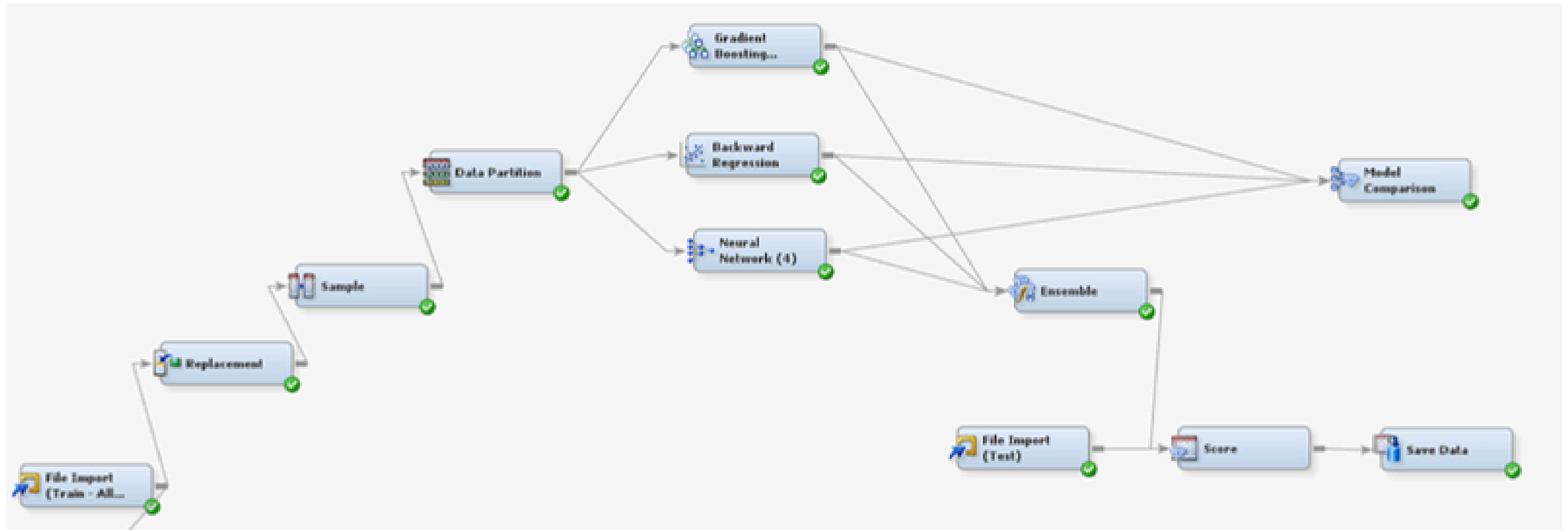
# Model I

Ensemble Model comprising of:

- Gradient Boosting
- Backward Regression
- Neural Network

With selection statistic as average

Test ROC Index - Public	0.94303
Test ROC Index - Private	0.94145



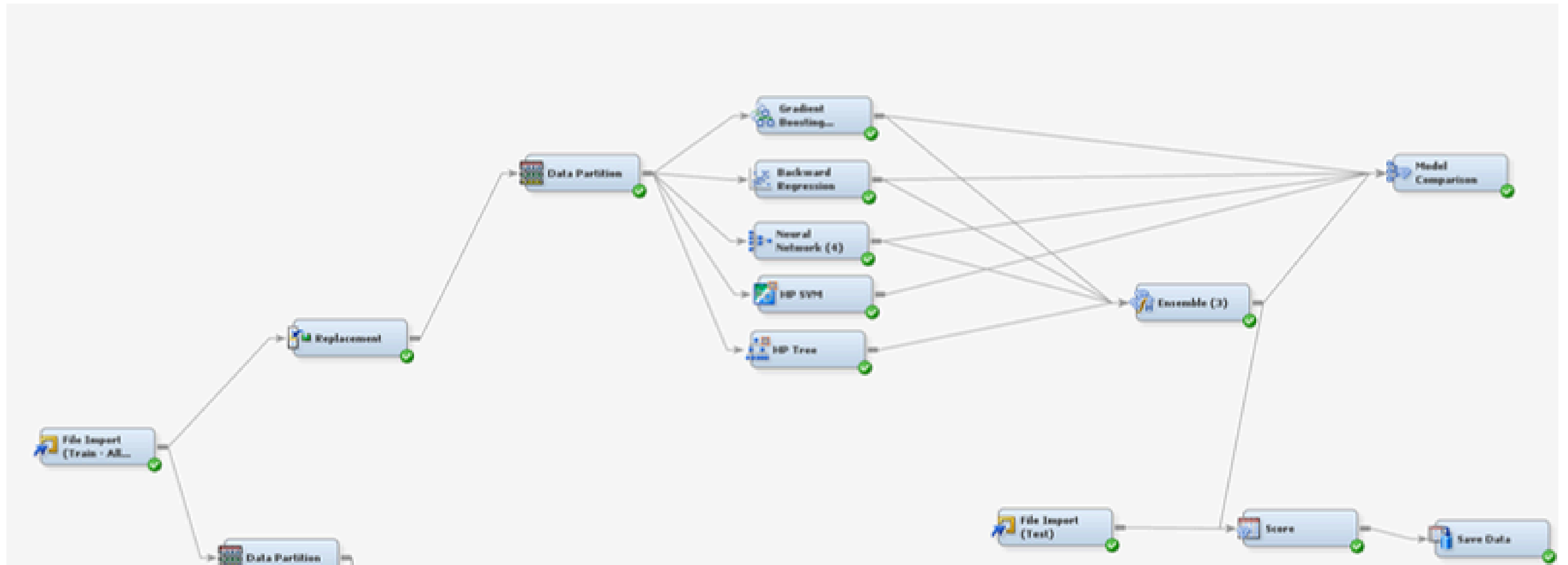
# Model II

Ensemble Model comprising of:

- Gradient Boosting
- Backward Regression
- Neural Network
- HP Tree

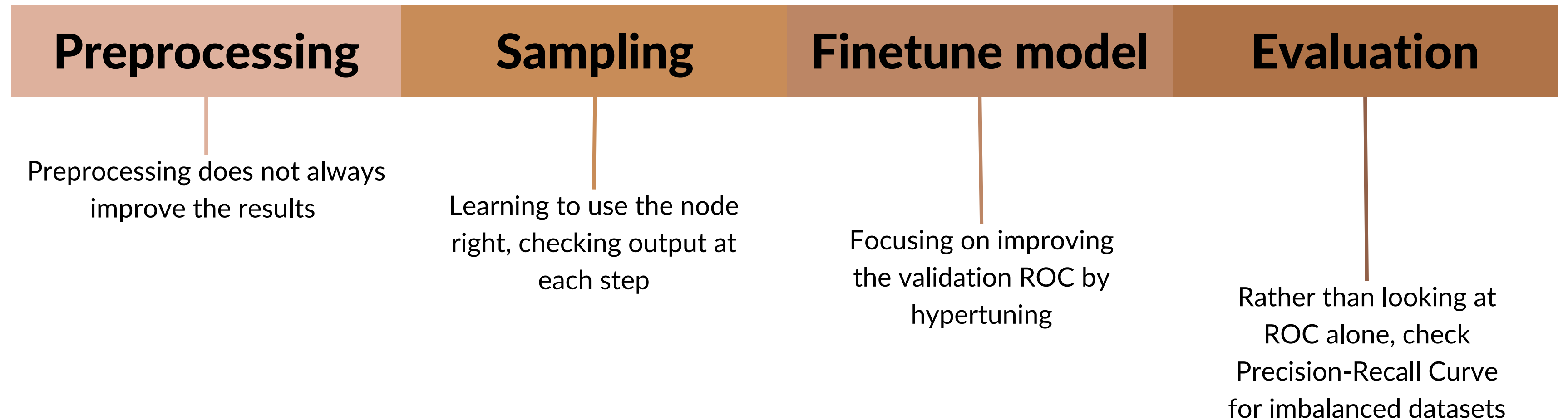
With selection statistic as average

Test ROC Index - Public	0.94472
Test ROC Index - Private	0.94013

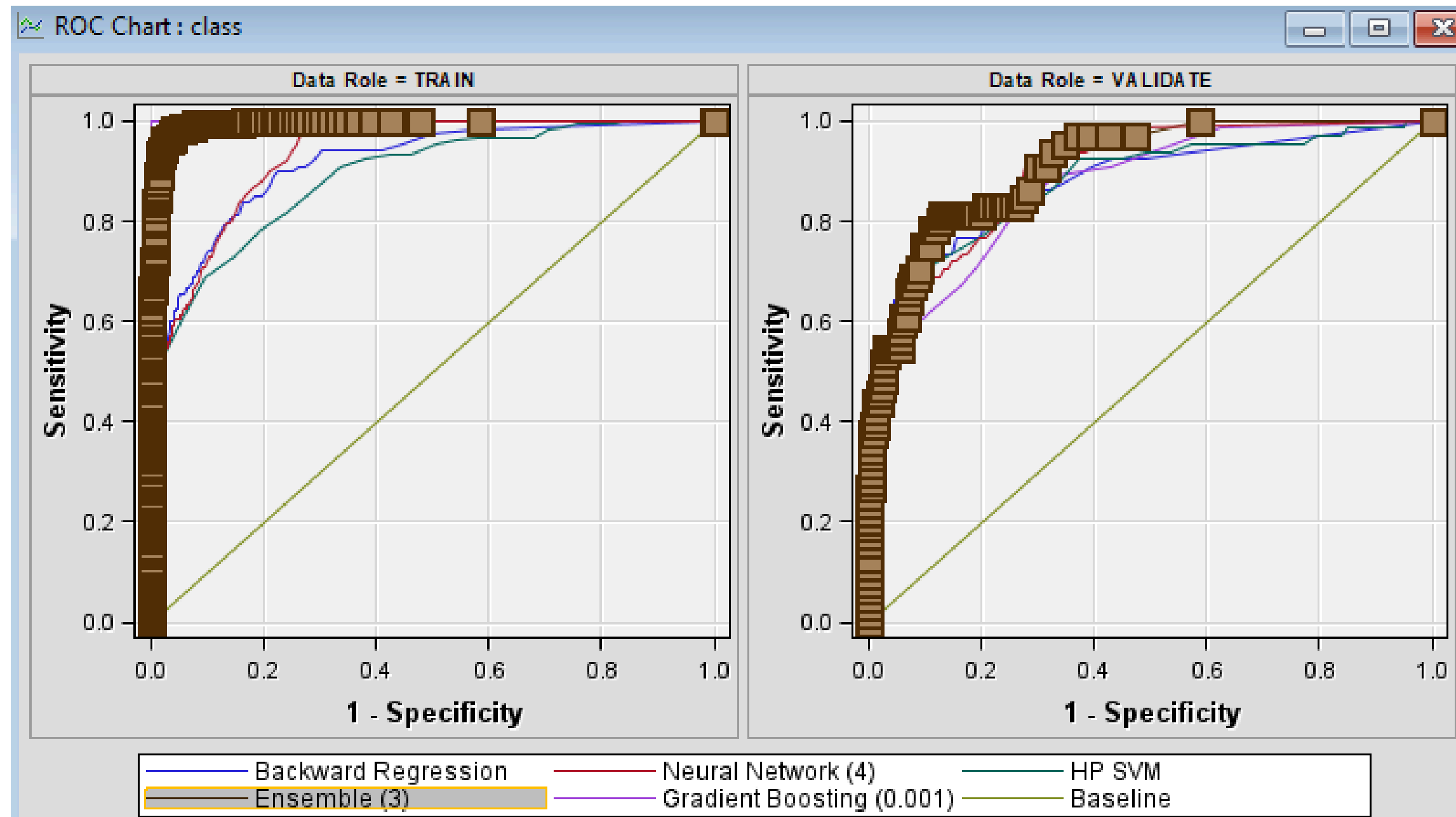


# Key Learnings

Through this project, we were able to apply the concepts we learned in class and learned a few more things to and not to do



# Overfitting



# Where we ended up





bankruptcy\_sample\_submission 4.csv

Complete · Darshan Upadhyay · 6d ago




0.94145

0.94303

## On Friday Night

3	DARing ANALysts	 	0.94010	29	12s
---	-----------------	---	---------	----	-----

## After Results

20	 8 DARing ANALysts	 	0.94145	46	6d
----	---	---	---------	----	----





Thank You!

