

Group 4

# Bankruptcy Prediction

## Machine Learning Implementation on Financial Distress

Jin, Ran

Qin, Rui

Wang, Mengqing

Zhang, Xingyue

# Agenda

- Introduction
- Data Description
- What is Machine Learning?
- Project Pipeline
- Detailed Process
- Conclusion

# Introduction

- Motivation: Better Understand Potential Risks When Making Investments
- Goal: Maximize Prediction Sensitivity

# Data Description

## Column Variables



Company



Time



Financial Distress

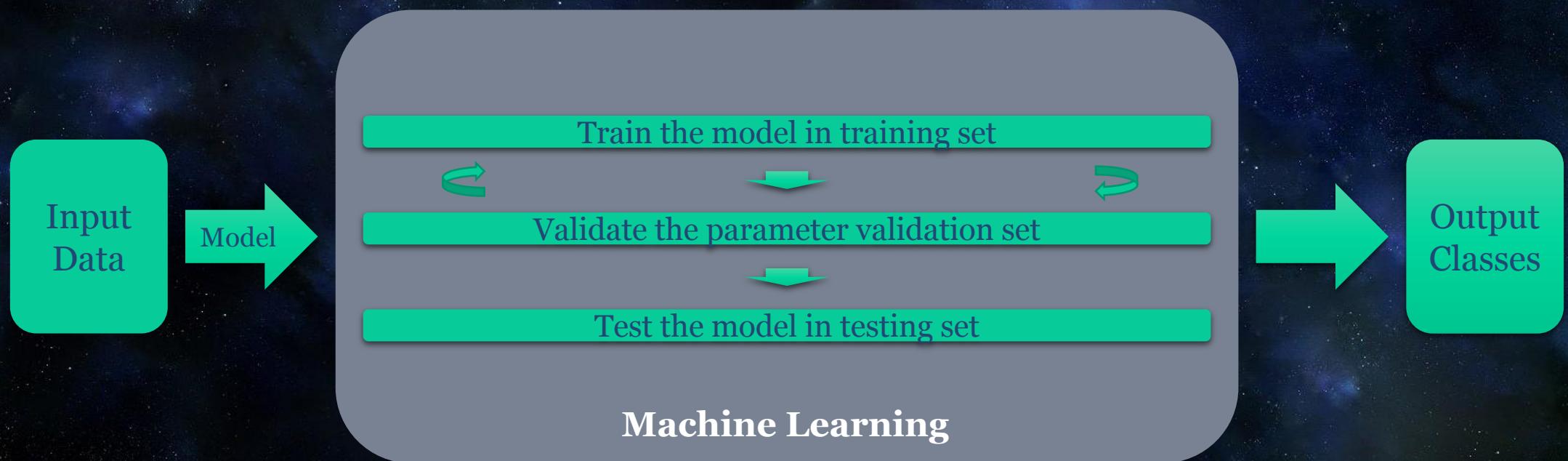


Other (80)

## Count

- 83 financial and non-financial characteristic variables
- 422 companies in 14 different time periods
- 136 out of 422 companies went bankrupt, 3.70% of data points
- Dimension  $3673 * 83$

# What is Machine Learning?



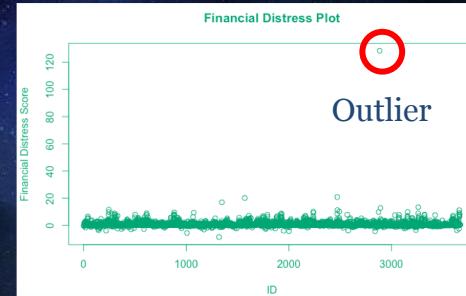
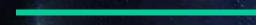
# Project Pipeline



# Data Preprocessing

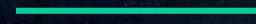


Identify Outlier



Financial Distress $\leq -0.5$

Identify Bankruptcy



Scale & Centering



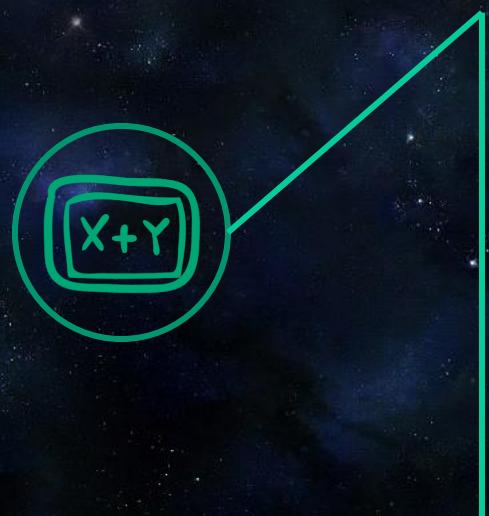
Data Splitting



```
scale(data, center = TRUE,  
      scale = TRUE)
```

Train: 70% (Cross validation, 5-fold)  
Validation: 15%  
Test: 15%

# Model Fitting



Train and test the data to find the model with best sensitivity

Choose SMOTE to balance the training data

Flexible discriminant analysis:  
validation sensitivity 33.78%  $\rightarrow$  56.25%

# Conclusion & Next Phase

- Extreme Gradient Boosting is the Best Model
  - Sensitivity: 56.25%
- Future Exploration:
  - Data Background for Every Variable X
  - Feature Construction

A dark blue space background filled with numerous small white stars of varying sizes. In the center, there is a faint, glowing nebula with a slight purple tint. The overall atmosphere is deep and mysterious.

Thank You!