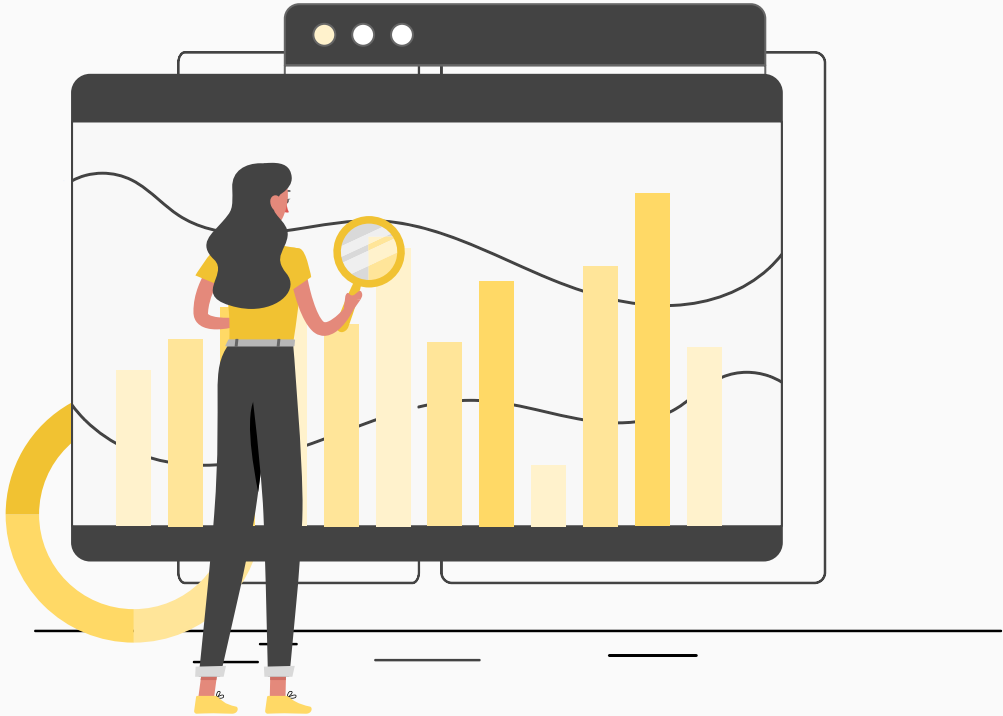
Options Pricing Group Project Team 25

En-Ning Chiang
ChinKai Huang
Sih-Yu Huang
Shih-Ting Liu



Agenda

01 Project Overview

02 Data Exploration

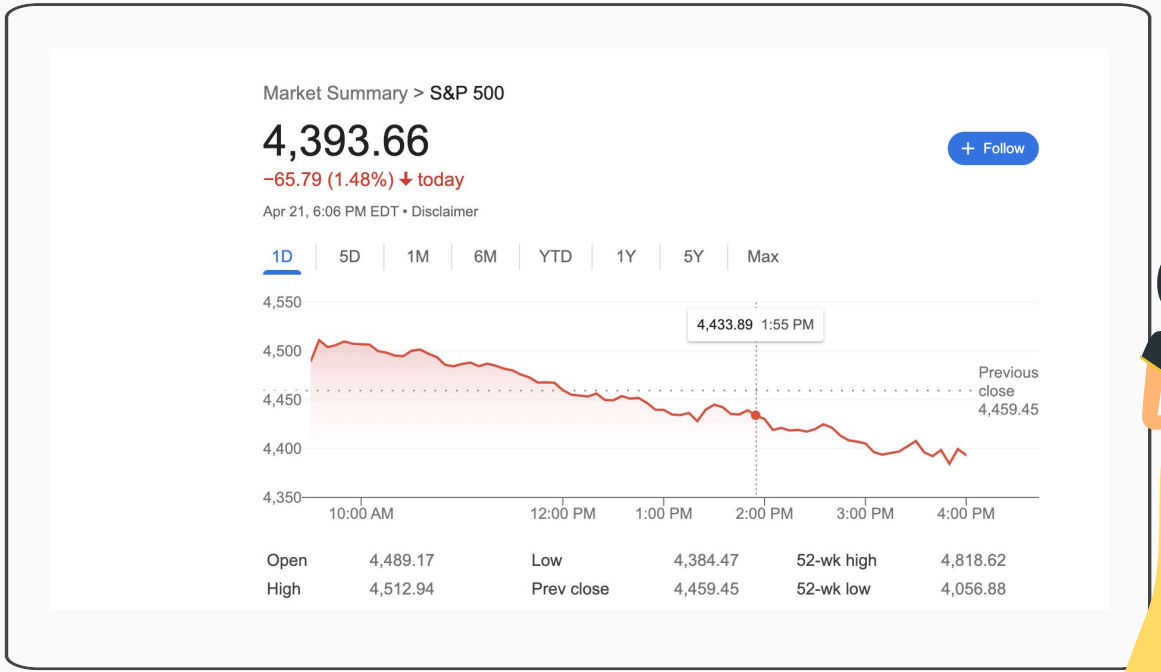
03 Data Preprocessing

04 Model Algorithm

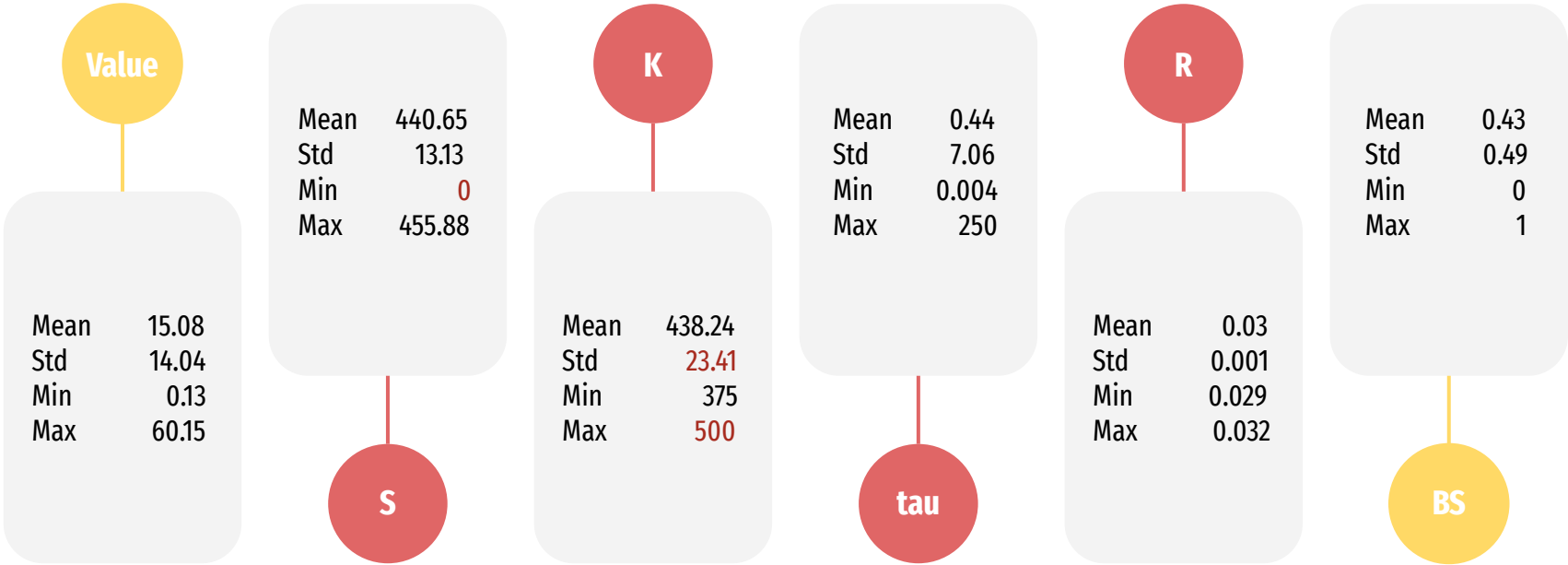
05 Performance Evaluation

06 Conclusion

Overview


Market Summary > S&P 500
4,393.66
−65.79 (1.48%) today
Apr 21, 6:06 PM EDT • Disclaimer
Follow
1D
5D
1M
6M
YTD
1Y
5Y
Max
4,550
4,500
4,450
4,400
4,350
4,433.89 1:55 PM
Previous close
4,459.45
10:00 AM
12:00 PM
1:00 PM
2:00 PM
3:00 PM
4:00 PM
Open 4,489.17
High 4,512.94
Low 4,384.47
Prev close 4,459.45
52-wk high 4,818.62
52-wk low 4,056.88

Data Exploration



Data Preprocessing

01

Drop the data
that has empty records

Data Cleaning

02

Rename
‘Under’ as 0
‘Over’ as 1

Rename labels

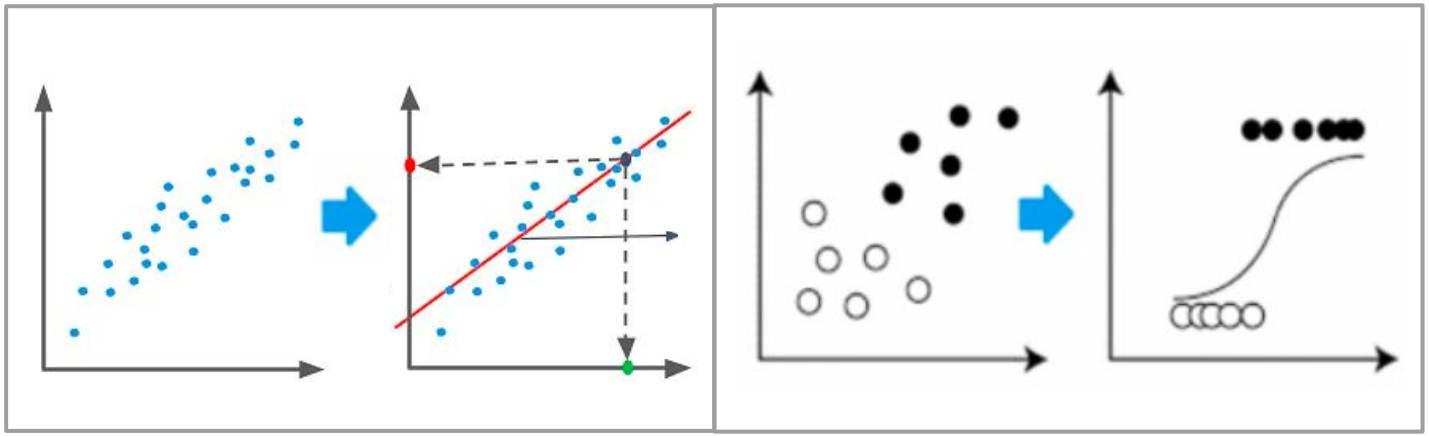
03

Deduct the mean value
then divide by stdv

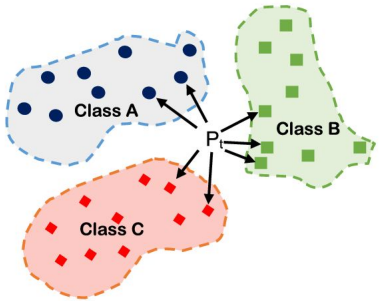
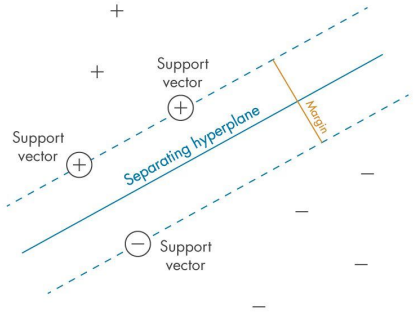
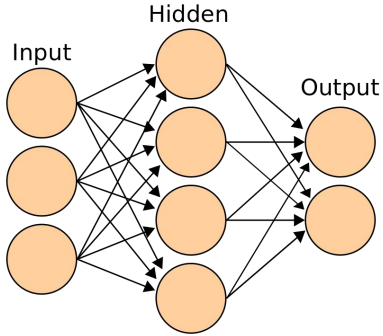
Standardization

Supervised Machine Learning Model

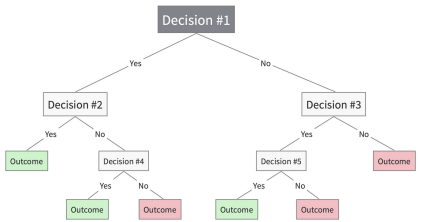
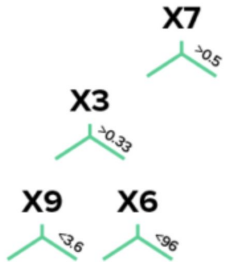
Multiple Linear Regression	Logistic Regression



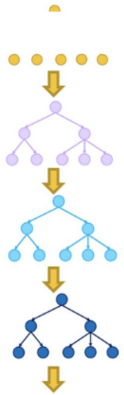
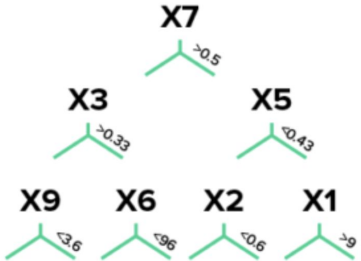
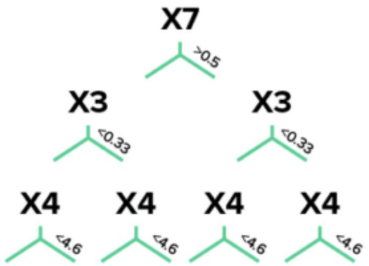
Supervised Machine Learning Model

KNN	Support Vector Machine	Neural Network
		

Supervised Machine Learning Model

Decision Tree	Boosted Tree	Random Forest
		

Supervised Machine Learning Model

Gradient Boosting Tree	XGBoost	CatBoost
		

Approach

01	Train/Test Split	• Set the test size as **20%** of the training dataset • Minimize the effects of data discrepancies and better understand the characteristics of the model
02	Cross-Validation	• Evaluating and comparing models by dividing data into two segments • Split **10** times with shuffle
03	Grid Search	• Tuning hyper parameters by building a model for every combination of hyperparameters specified and evaluating each model

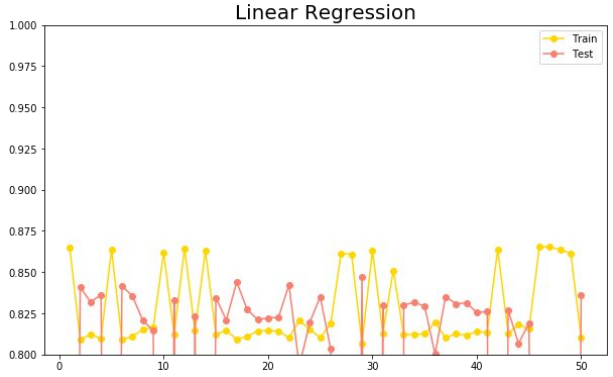
Approach

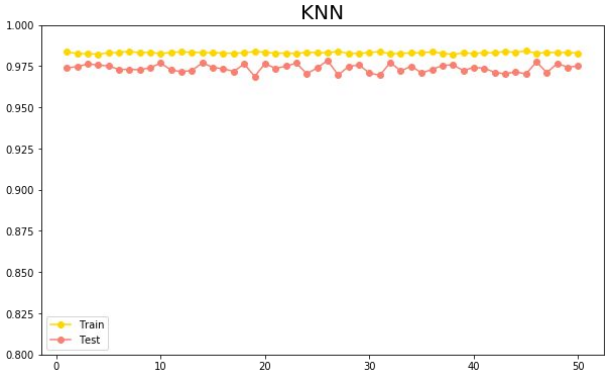
01	Train/Test Split	• Set the test size as **20%** of the training dataset • Minimize the effects of data discrepancies and better understand the characteristics of the model
02	Cross-Validation	• Evaluating and comparing models by dividing data into two segments • Split **10** times with shuffle
03	Grid Search	• Tuning hyper parameters by building a model for every combination of hyperparameters specified and evaluating each model

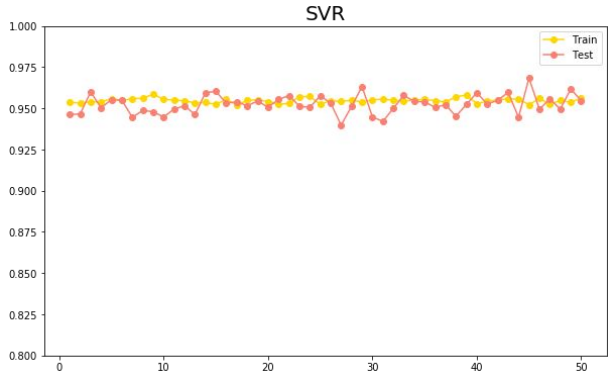
Regression - Model Performance Comparison

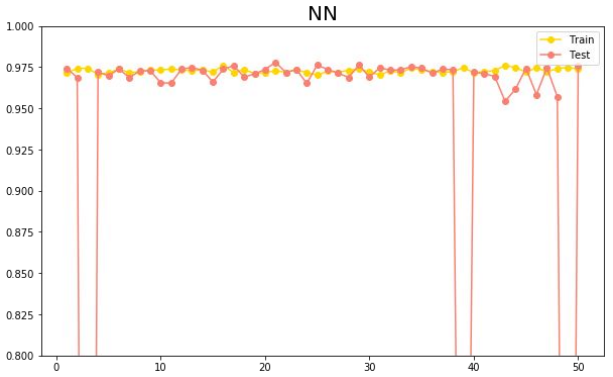
Model	Train / Test Split	Cross Validation
Linear Regression	81.7%	65.1%
Decision Tree	99.2%	99.2%
Boosted Tree	99.7%	99.6%
XGBoost	99.4%	99.5%
Random Forest	99.7%	99.8%
KNN	97.2%	97.4%
SVR	94.7%	95.5%
Neural Network	97.4%	97.3%

Regression - Model Performance Comparison

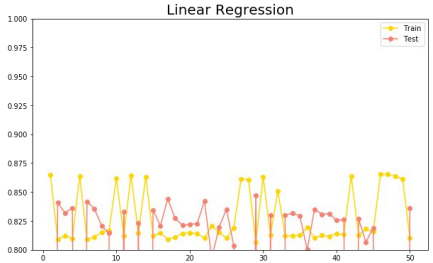


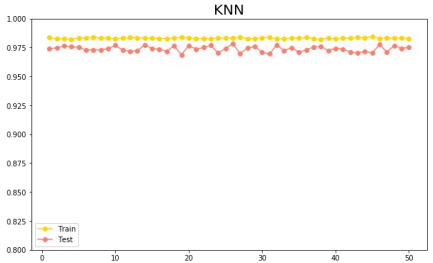


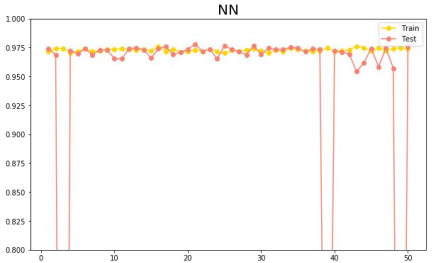


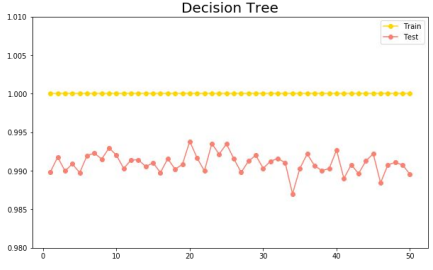


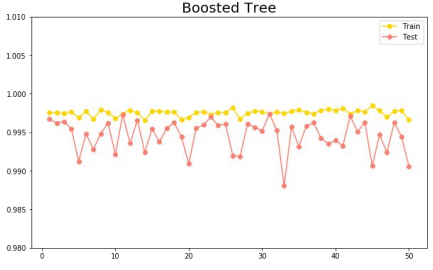
Regression - Model Performance Comparison

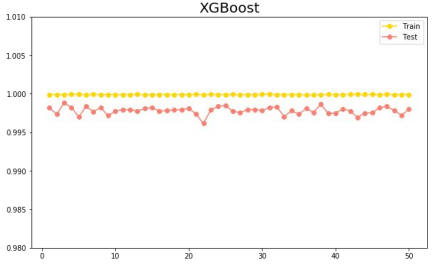


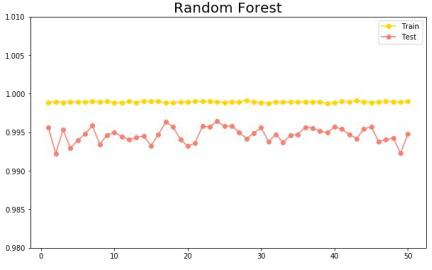




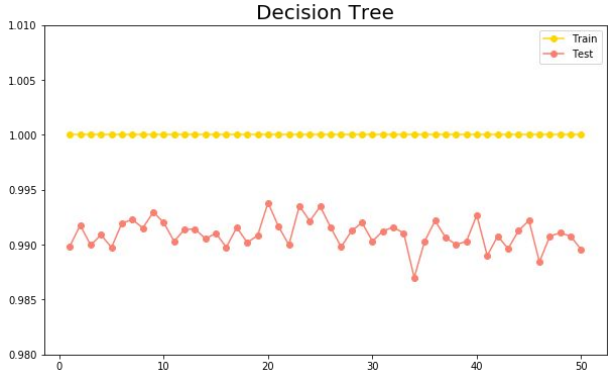


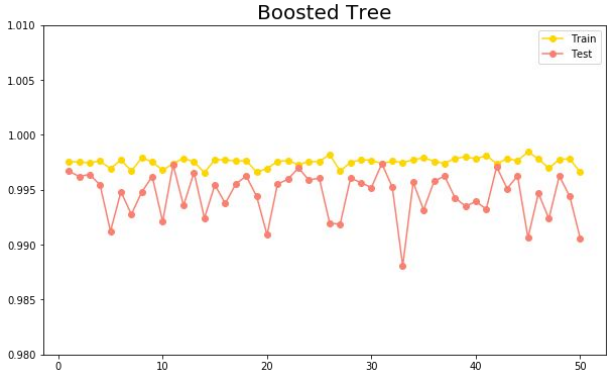


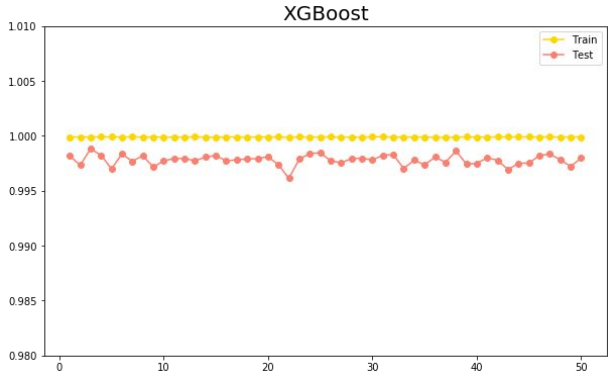


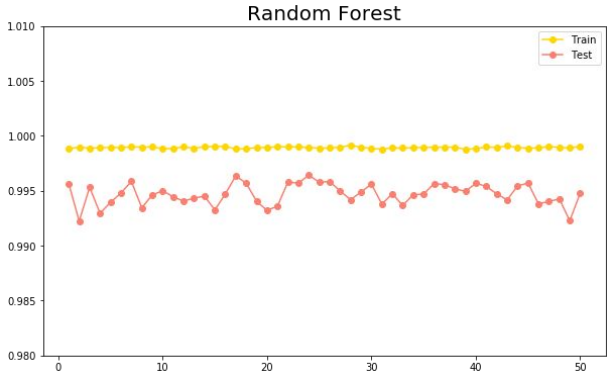


Regression - Model Performance Comparison









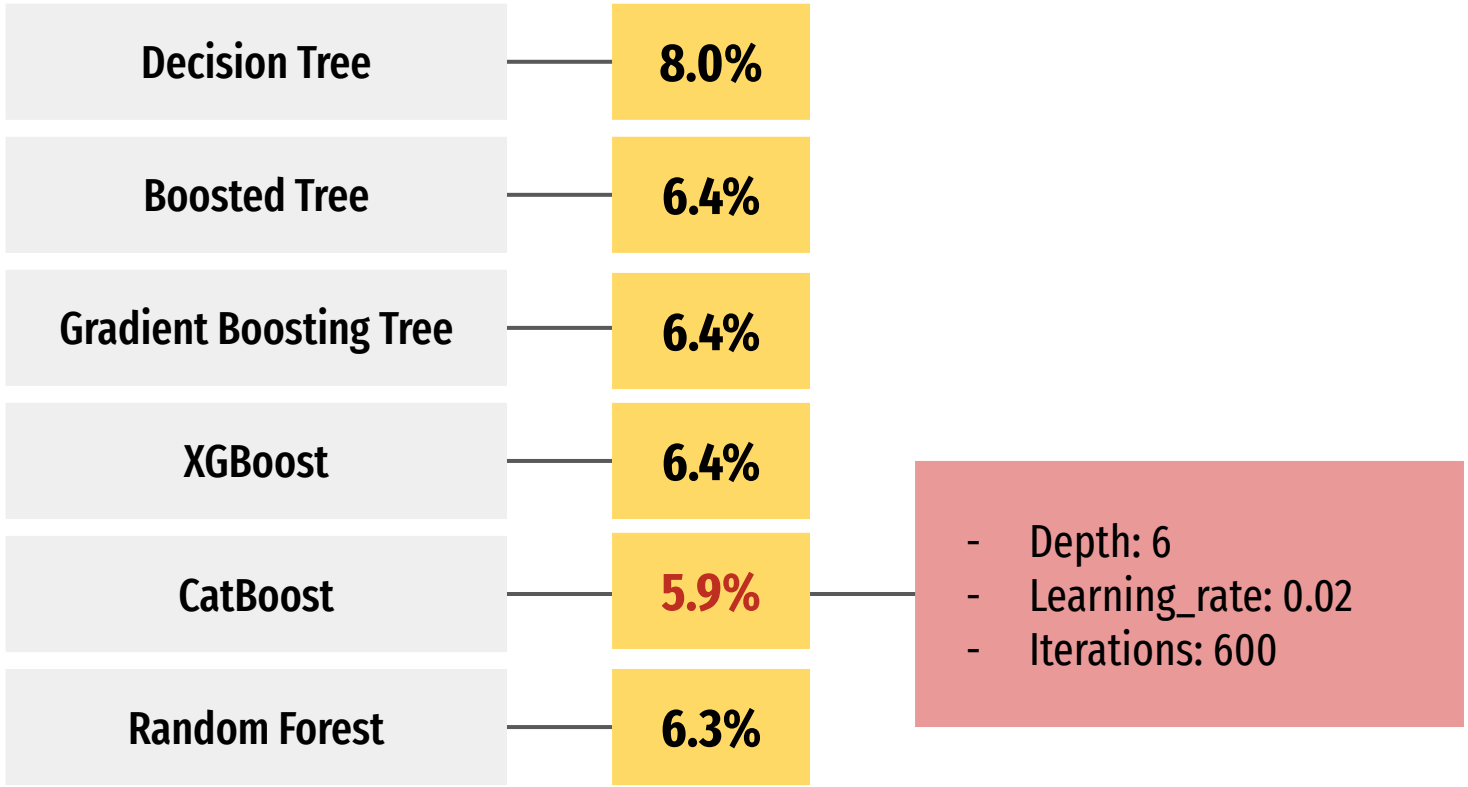
Regression - Model Performance Comparison

Model	Train / Test Split	Cross Validation
Linear Regression	81.7%	65.1%
Decision Tree	99.2%	99.2%
Boosted Tree	99.7%	99.6%
XGBoost	99.4%	99.5%
Random Forest	99.7%	99.8%
KNN	97.2%	97.4%
SVR	94.7%	95.5%
Neural Network	97.4%	97.3%

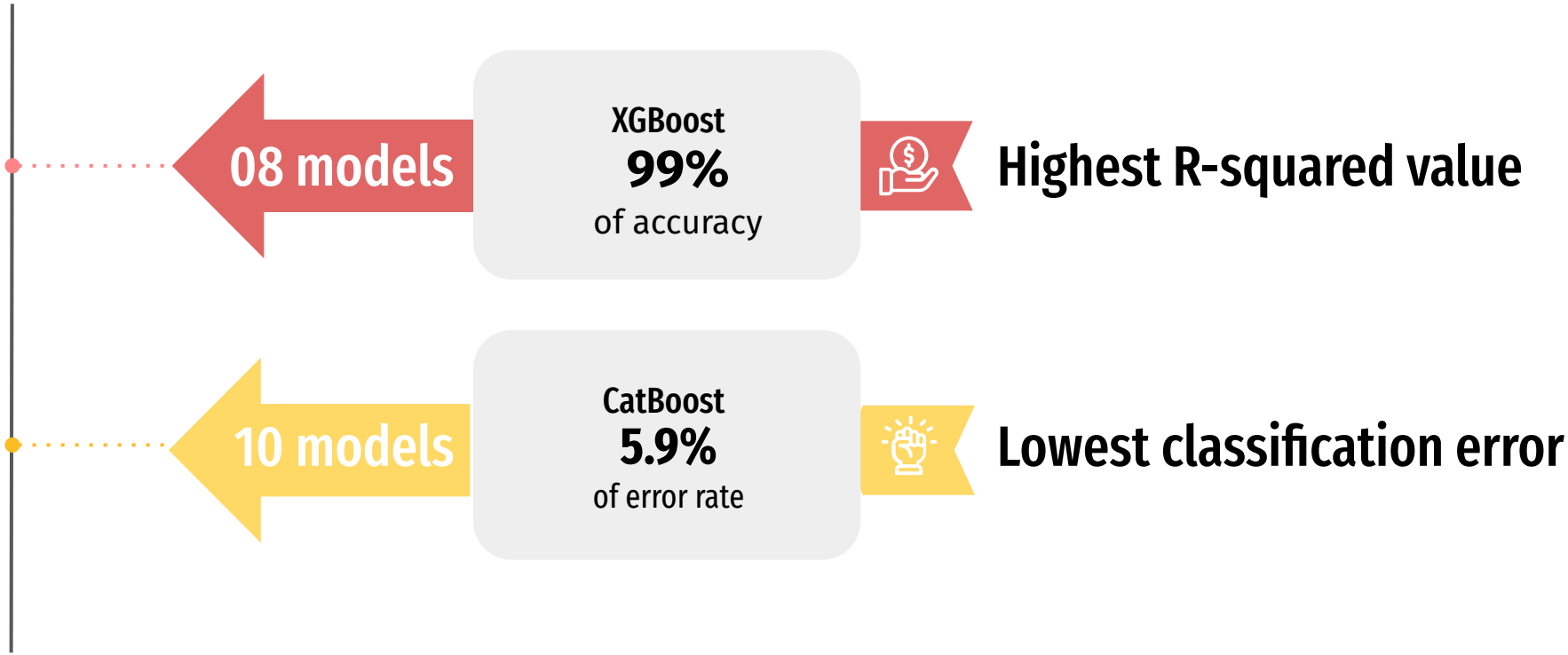
Classification - Model Performance Comparison

Model	Train / Test Split	Cross Validation
Logistic Regression	13.1%	12.3%
Decision Tree	8.6%	8.5%
Boosted Tree	6.3%	6.6%
Gradient Boosting Tree	6.5%	6.7%
XGBoost	6.3%	6.6%
CatBoost	6.8%	6.2%
Random Forest	7.4%	6.4%
KNN	10.4%	9.3%
SVM	11.0%	10.6%
Neural Network	12.2%	11.6%

Hyperparameters Tuning



Models Decision



4 Risk Indicators of Our Models

S	Current Asset Value	Evaluation of market trend
K	Strike Price of Option	Evaluation of reward payoff
tau	Time to Maturity	Evaluation of time value
R	Annual Interest Rate	Evaluation of ROI

Research Insights

Pros of ML Model

- Flexibility: capable of fitting a large number of function forms
- Fixed model vs. adjustable model

Limitation of BS Model

- Europe Options
- Dividends, volatility, and risk-free rates
- Taxes, commissions, or trading costs

Future Research Development

- Non-parametric ML models
- High frequency data

Thank you