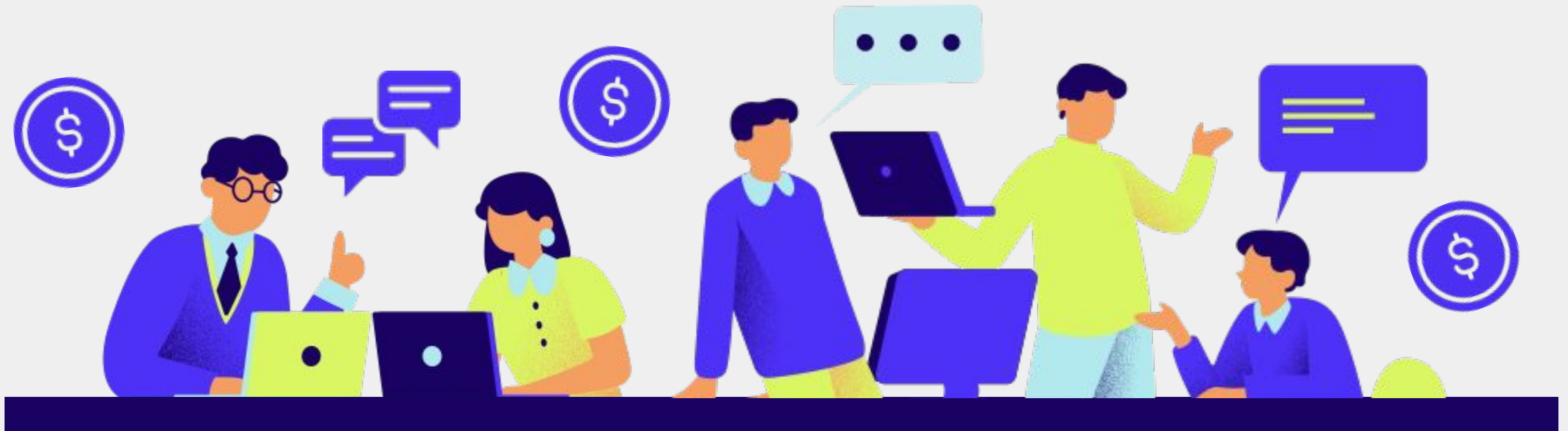


DSO 530 - 16275. Team 6: Linh Chu, Nhi Dang, Linh Pham, Cong Minh Tran

Introducing our Call Option Valuation & Pricing

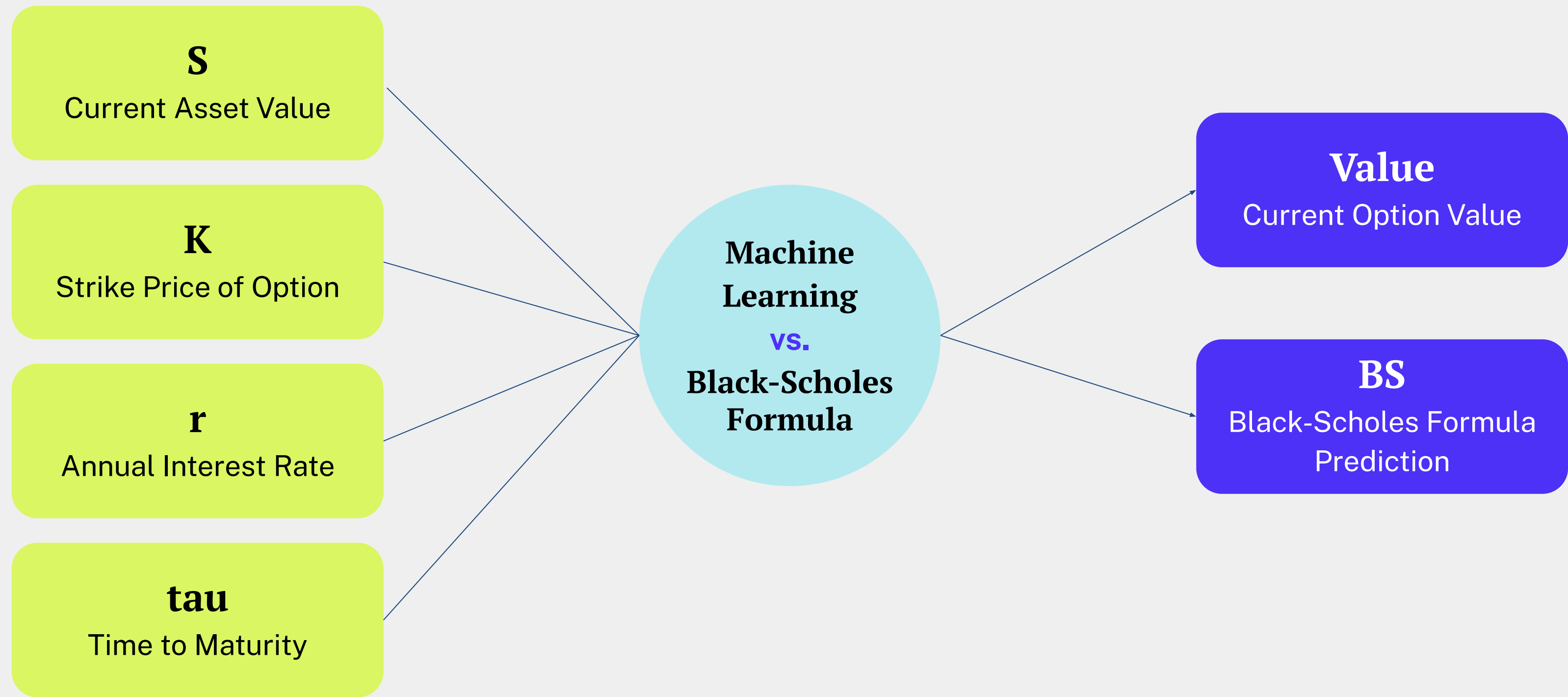




Agenda

- 1 Problem Statement
- 2 Exploratory Data Analysis
- 3 Feature Engineering
- 4 Value Prediction Model
- 5 BS Prediction Model
- 6 Considerations

Which one predicts better?



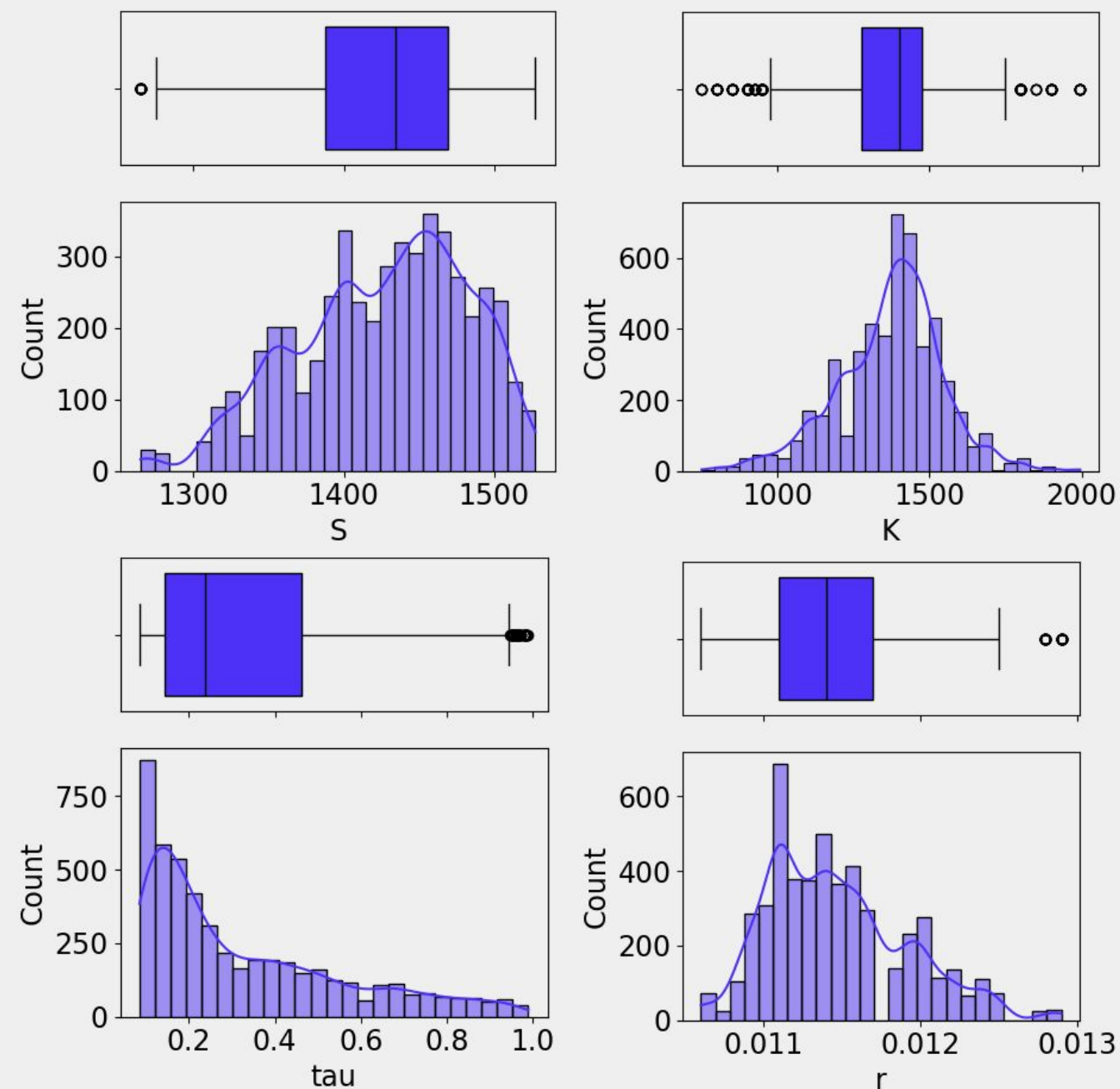
Exploratory Data Analysis

1. Descriptive Statistics

No missing value + no extreme outliers

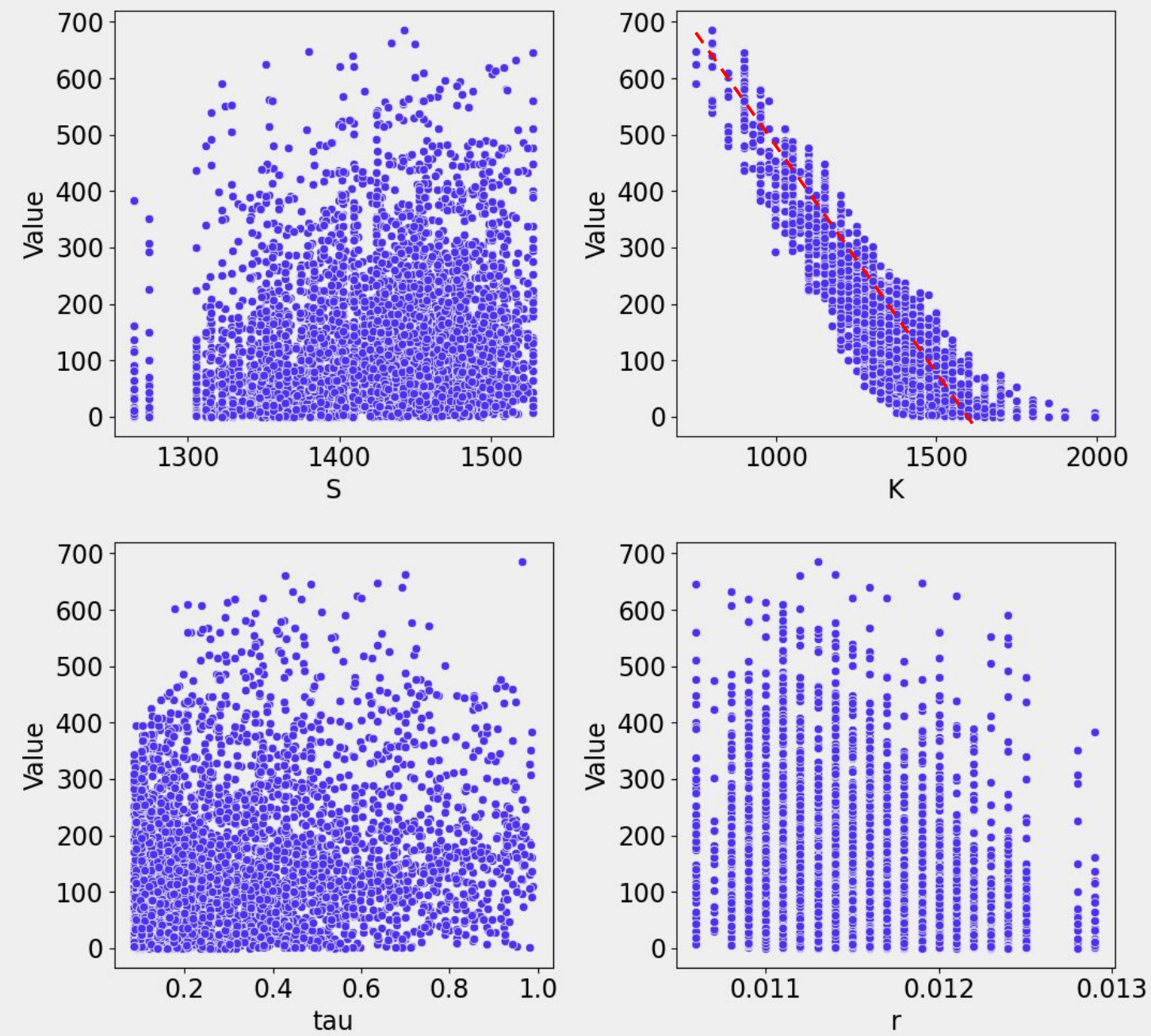
→ Keep all 5,000 rows

→ Standardize all features



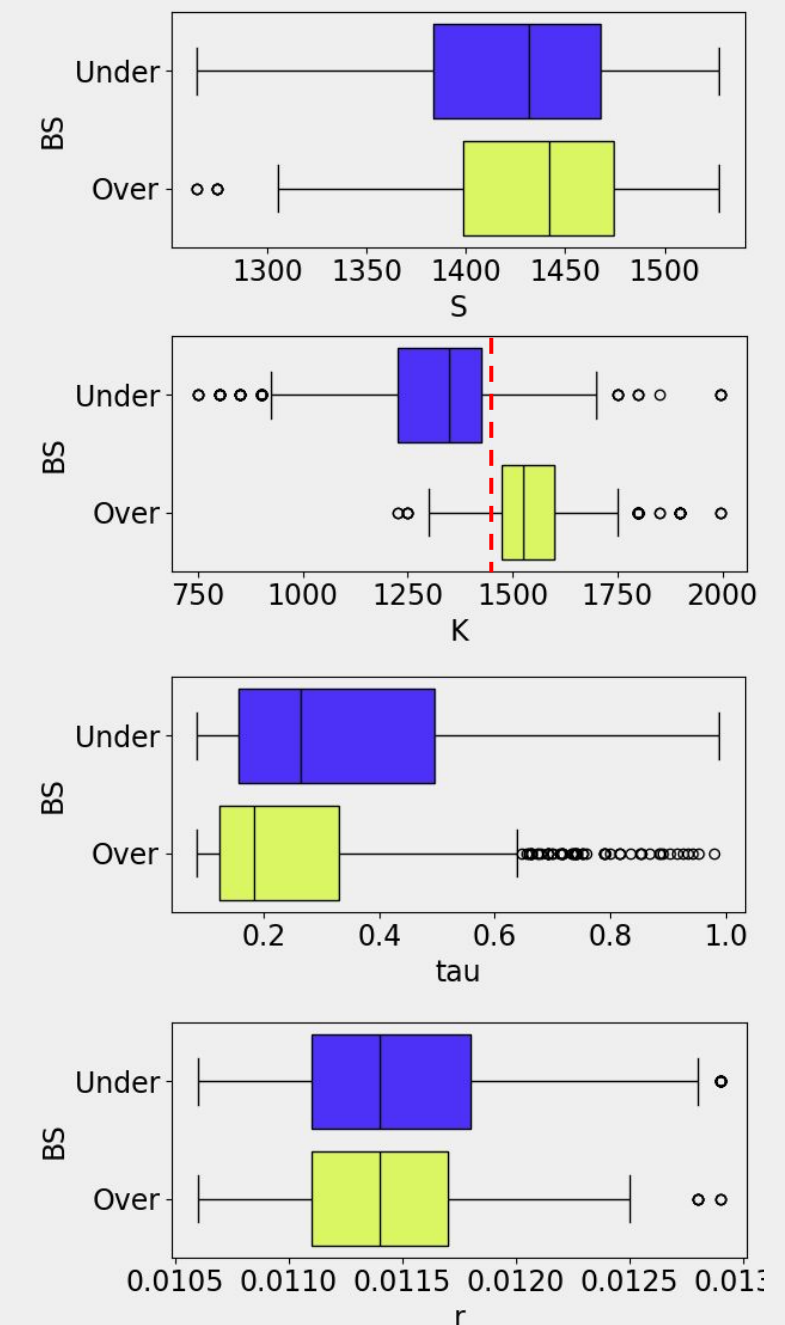
2. Correlation with 'Value'

- K seems negatively correlated
- No clear linear correlation with other features



3. 'BS' differences

K can separate the majority of 2 classes

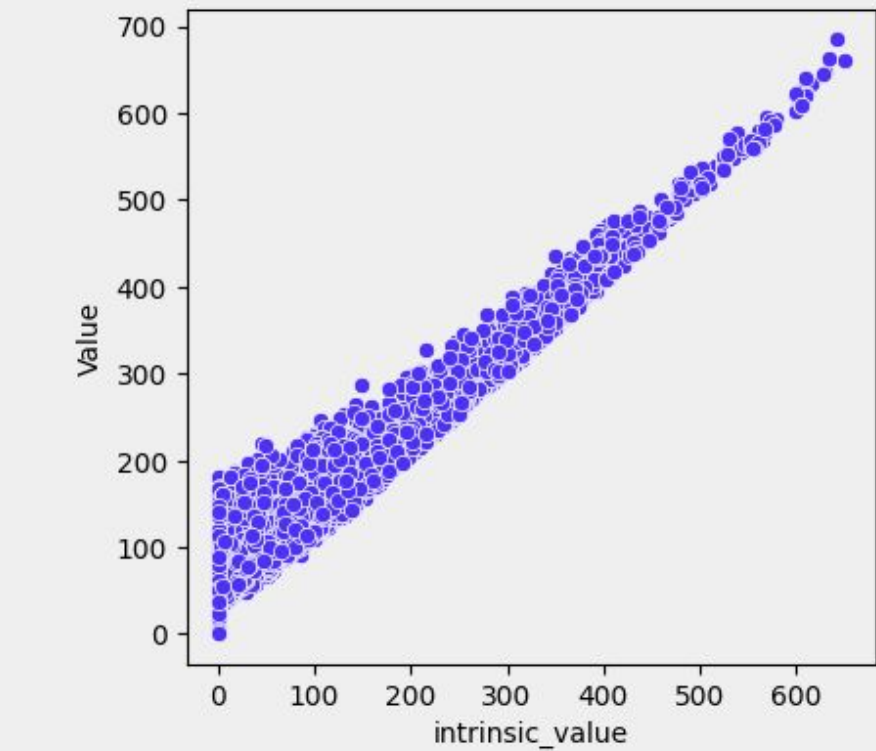


Feature Engineering

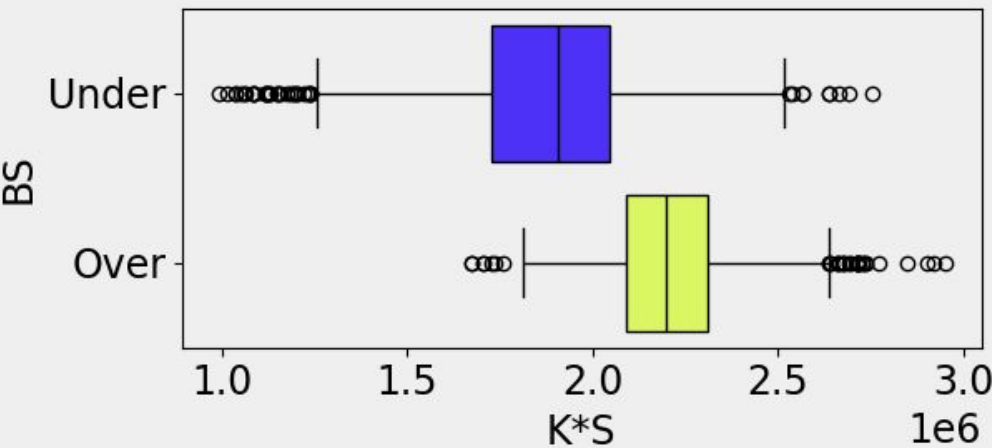
Creating new features based on options terminologies

Features	Definitions
S/K	Moneyness - Indicator for in-the-money (ITM), at-the-money (ATM), or out-of-the-money (OTM) options
$\max(S-K,0)$	Capture intrinsic value, enhancing the interpretability of the model by incorporating a clear and intuitive measure of option value
$\text{abs}(S-K)$	A direct measure of the distance between the current asset price S and the strike price K - assess the potential impact of asset price fluctuations on option values
tau_days	Offer additional granularity and capture more nuanced temporal information
$K*S$	Monetary value of the underlying asset relative to the strike price

→ $\max(S-K,0)$ has a strong linear relationship with option value



→ $K*S$ can partially separate BS value



Modeling Approach

Regression

(Criteria: R^2)

Linear Regression
Decision Tree
Gradient Boosting
XGBoost
Random Forest

Classification

(Criteria: Classification Error)

Logistic Regression
K-NN
Decision Tree
Random Forest
XGBoost
SVM

Cross validation 10-folds + Hyperparameter Tuning

Model selection based on **best mean CV score**

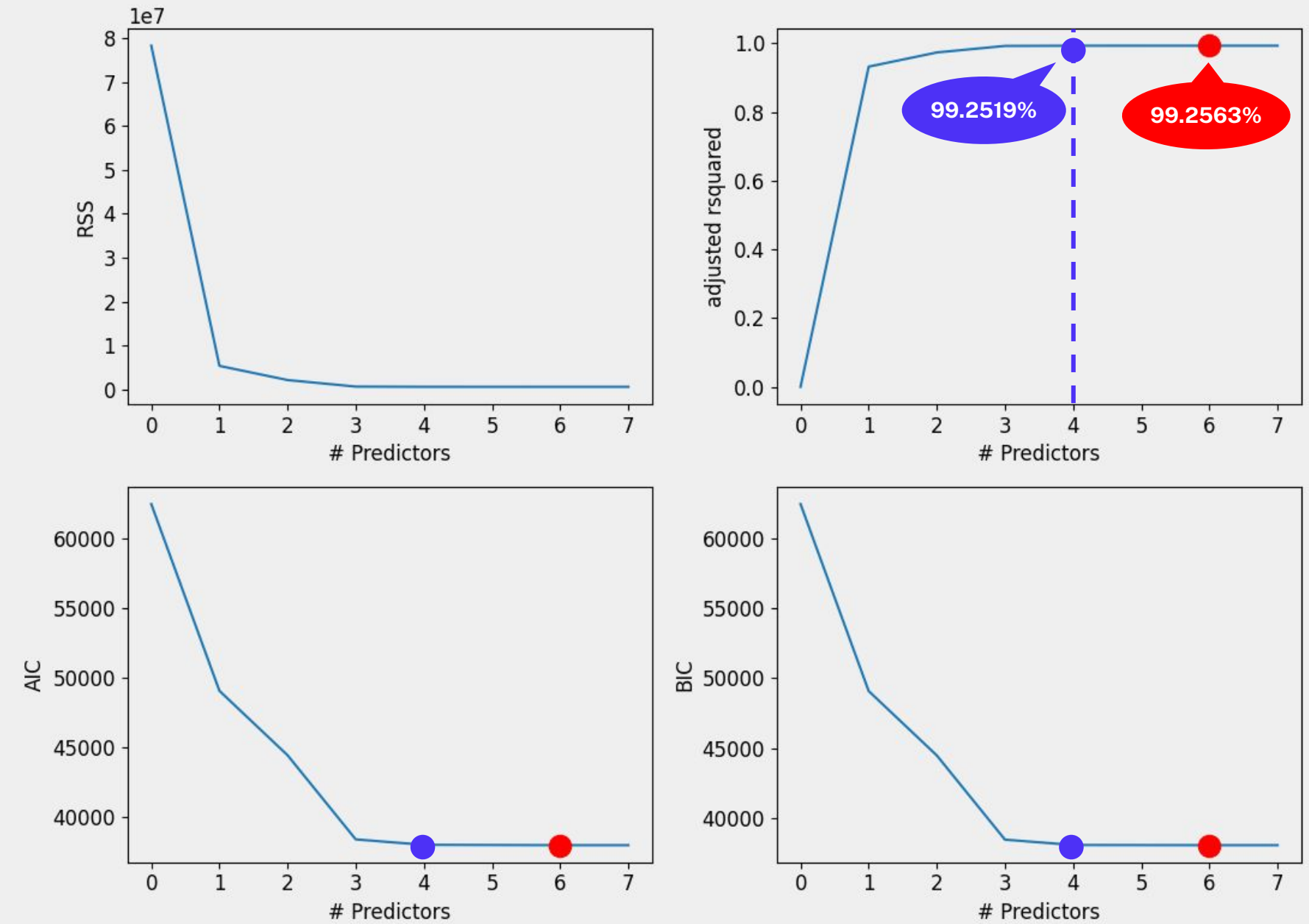
Regression

Model	Features	Mean CV R-squared
Linear Regression w/ Best Subset Selection	tau, S/K, S-K , intrinsic value	99.25

Linear Regression with Best Subset Selection

Adjusted R squared does not increase significantly for 5+ predictors

→ stop at 4 best predictors: **tau, S/K, |S-K|, intrinsic_value** for better model interpretability



	coef	std err	t	P> t	[0.025	0.975]
const	-130.5409	7.821	-16.691	0.000	-145.874	-115.208
tau	122.4174	0.678	180.551	0.000	121.088	123.747
S/K	160.3511	7.946	20.179	0.000	144.773	175.930
S-K_abs	-0.1724	0.005	-37.137	0.000	-0.182	-0.163
intrinsic_value	0.8795	0.012	72.230	0.000	0.856	0.903

Regression

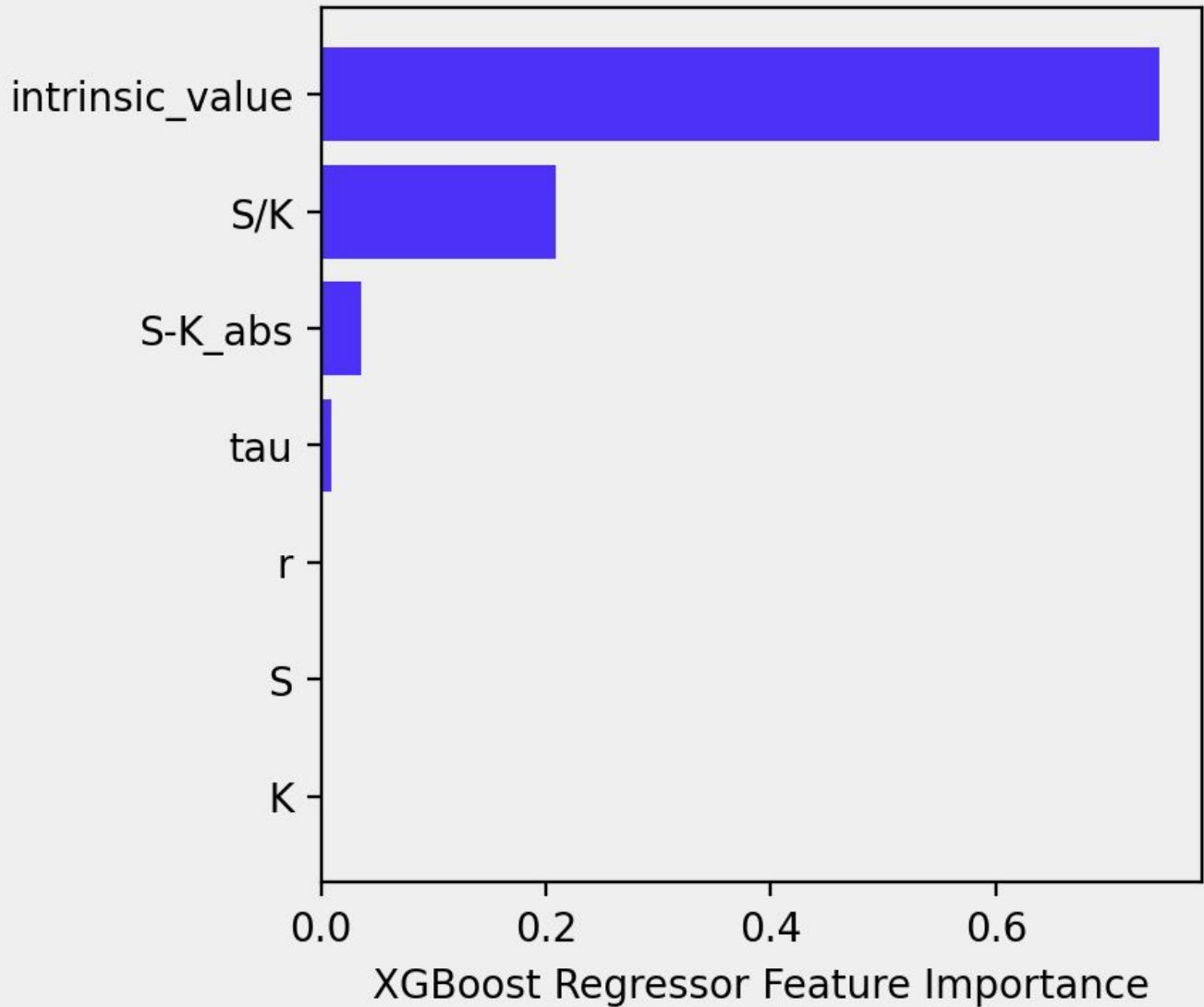
Model	Features	Mean CV R-squared
Linear Regression w/ Best Subset Selection	tau, S/K, S-K , intrinsic value	99.25
Decision Tree		99.58
Random Forest	S, K, tau, r, S/K, S-K , intrinsic value	99.77
Gradient Boosting		99.81
XGBoost		99.87

→ **XGBoost** is our final choice for the best **prediction accuracy**

→ **Linear Regression** can also be considered for **model interpretability**

XGBoost Regressor

- Out of tree-based models, XGBoost gave the highest R-squared
- 4 most important features in XGBoost match with 4 features selected in Linear Regression



Hyperparameter	Value
gamma	0.1
learning_rate	0.1
max_depth	7
n_estimators	300
subsample	0.8

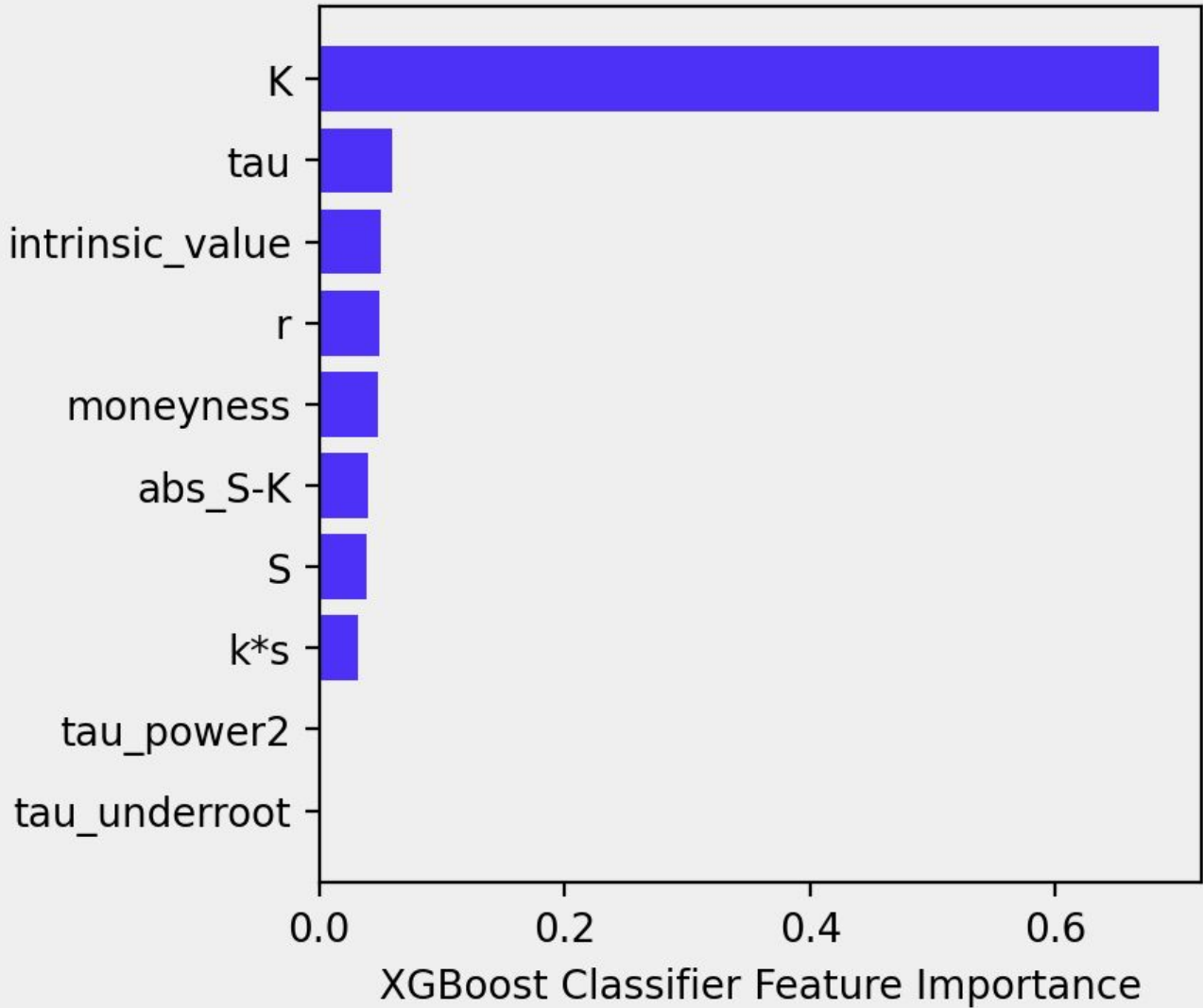
Classification

Model	Features	Mean CV Classification Error
Logistic Regression	K, tau, r, S/K, S-K , intrinsic value, tau ²	10.36
SVM	K, r, S-K , tau_underroot	9.16
KNN	S, K, r, S/K, intrinsic value, √tau, K*S	8.38
Decision Tree	S, K, tau, r, √tau, tau ²	9.28
Random Forest	all	6.46
XGBoost	all	5.58

→ **XGBoost** is our final choice for the best **prediction accuracy**

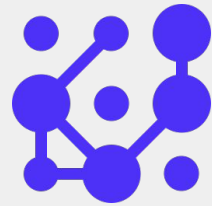
XGBoost Classifier

- K is the most important feature in XGBoost Classification



Hyperparameter	Value
colsample_bytree	1.0
gamma	0.1
learning_rate	0.1
max_depth	7
n_estimators	300
subsample	0.8

CONSIDERATIONS



Model complexity vs. prediction accuracy

- Hyperparameters & features lead to better performance
- Linear regression model has lower accuracy but higher interpretability



Inclusion of all four variables

- Not all variables are equally important
- Redundant predictors increase model complexity

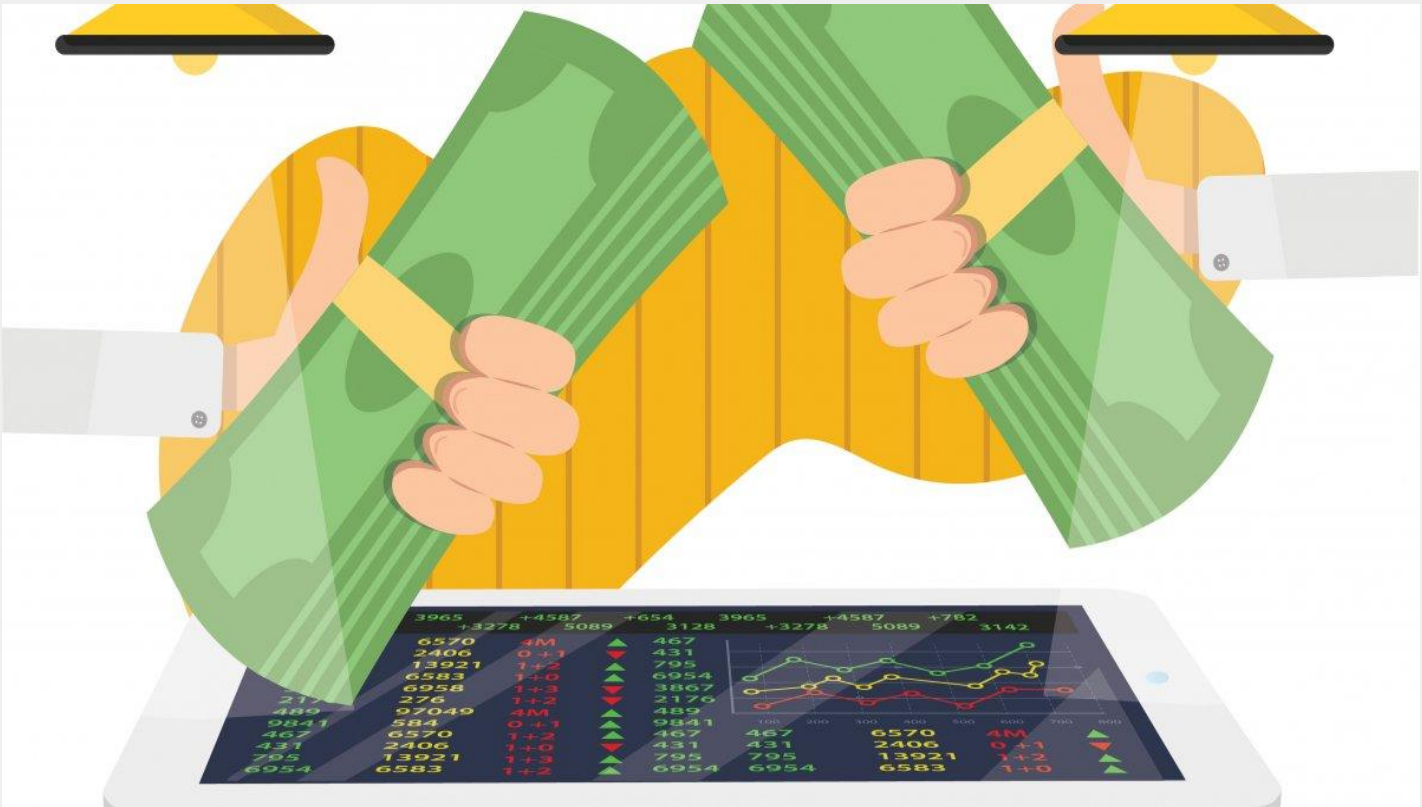


Application on Tesla stocks

- Can't directly be applied to Tesla stocks
- Application depends on diverse set of stocks/industries/time period & external factors

CONCLUSION

Best Model	Value Prediction (Regression) Mean CV R-squared	BS Prediction (Classification) Mean CV Classification Error
XGBoost	99.87	5.58



Thank you!