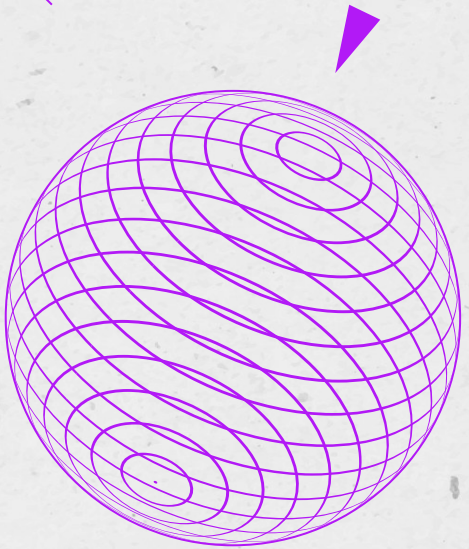
USC DSO 530

OPTION PRICING MODEL PROJECT



2023

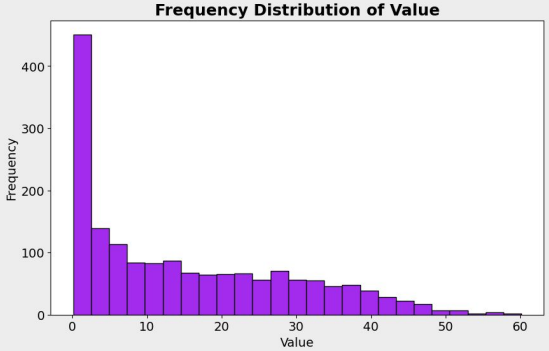
Tom Do, Jessie Xue, Connie Yau, Huiling Xiao, Jiayi Hu, Yanshuo Li

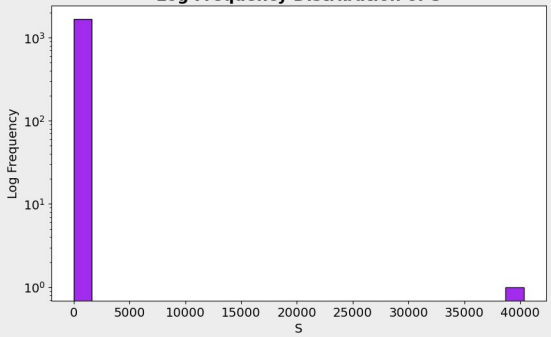
TABLE OF CONTENT

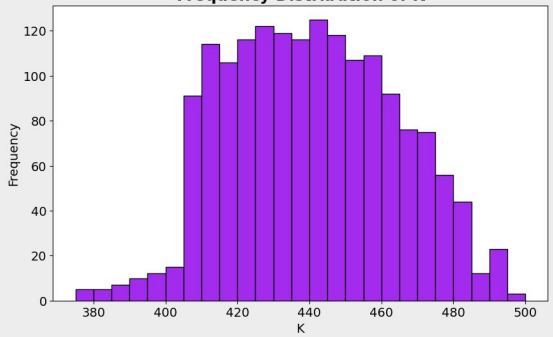
DATA EXPLORATIONS - Summary table

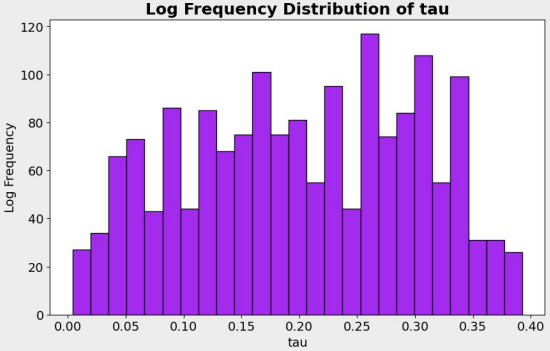
Field Name	% Populated	Min	Max	Mean	Stdev	# Zero
Value	99.88	0.125	60.149367	15.07	14.04	0
S	99.94	0	40333	464.40	973.65	1
K	99.88	375	500	438.24	23.41	0
tau	99.94	0.003968	250	0.44	7.06	0
r	100	0.029510	0.031880	0.03	0.000557	0

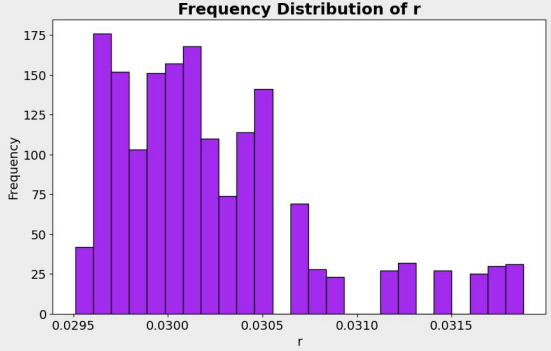
DATA EXPLORATIONS -
Visualization of fields

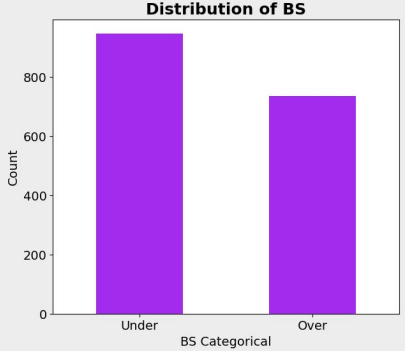












2023

DATA EXPLORATIONS - Clean data

- Drop Outliers which are 3 standard deviations from the mean
- Drop Records that has null values

	Value	S	K	tau	r	BS
12	2.31500144	448.688109	470	250	0.03013	Over
33	2.56500002	445.04224	455	146	0.03003	Over
47	11.4512725	40333	425	0.04365079	0.03147	Under
53	NAN	446.718974	430	0.16666667	0.02962	Under
205	59.8251691	430.638719	375	0.28968254	0.03178	Under
292	8.625	NAN	NAN	NAN	0.03003	Over
818	NAN	431.284616	NAN	0.23015873	0.02972	Over
879	4.125	0	455	0.17063492	0.03003	Over
1275	60.1493671	431.020394	375	0.28571429	0.03168	Under

FEATURES - Feature Engineering

All four predictor variables are essential:

- Current asset value (S):
  - Provide information about the value of the underlying asset.
  - Gain a better understanding of the current market conditions and the potential risks and rewards.
- Strike price of option (K):
  - Represents the price at which the option can be exercised.
  - Provide insight into the expectations of the market and the potential profits or losses.
- Annual interest rate (r):
  - It can impact the cost of borrowing and may influence the value of the underlying asset.
  - Help businesses understand the potential impact of interest rate changes on the option value and make informed decisions about whether to hold, buy, or sell the option.
- Time to maturity (tau):
  - Represents the amount of time until the option expires.
  - Provide information about the potential risks and rewards associated with holding the option over a particular period of time.

FEATURES - New features

Future_Mult = (1 + Annual Interest Rate)^Time to maturity

- Future value of an investment or the total amount of interest earned on an investment.
- For example, if you have an investment with an annual interest rate of 5% and a time to maturity of 10 years, the expression (1 + 0.05)^10 would give you the total amount of interest earned over the 10-year investment period.

S/K = Current Asset Value / Strike Price of Option

- Intrinsic value of a call option. If the result is greater than 1, the option is "in the money" because the option holder could buy the underlying asset at a lower price than the current market price. Vice Versa.
- For example, if a call option has a K = $100 and S = $120, the intrinsic value of the option would be $20. This means the option holder could exercise the option and buy the underlying asset for $100, then sell it on the market for $120, realizing a profit of $20.

MODEL EXPLORATIONS - Strategy

Establish a baseline

- Start with Linear Regression and Logistic Regression as the baseline models

Experiment with non-linear models

- Try nonlinear models: Decision Tree, Random Forest, XGBoost, LightGBM, CatBoost, Gradient Boost, tuning hyperparameters

Split train/test and cross validation

- Split 80% training and 20% testing
- 5-fold cross validation on randomly shuffled dataset
- Repeat the process 100 times to evaluate the model

MODEL EXPLORATIONS -
Linear Regression

Original Dataset

R2: 0.798
Adj. R2: 0.797

Drop Outliers

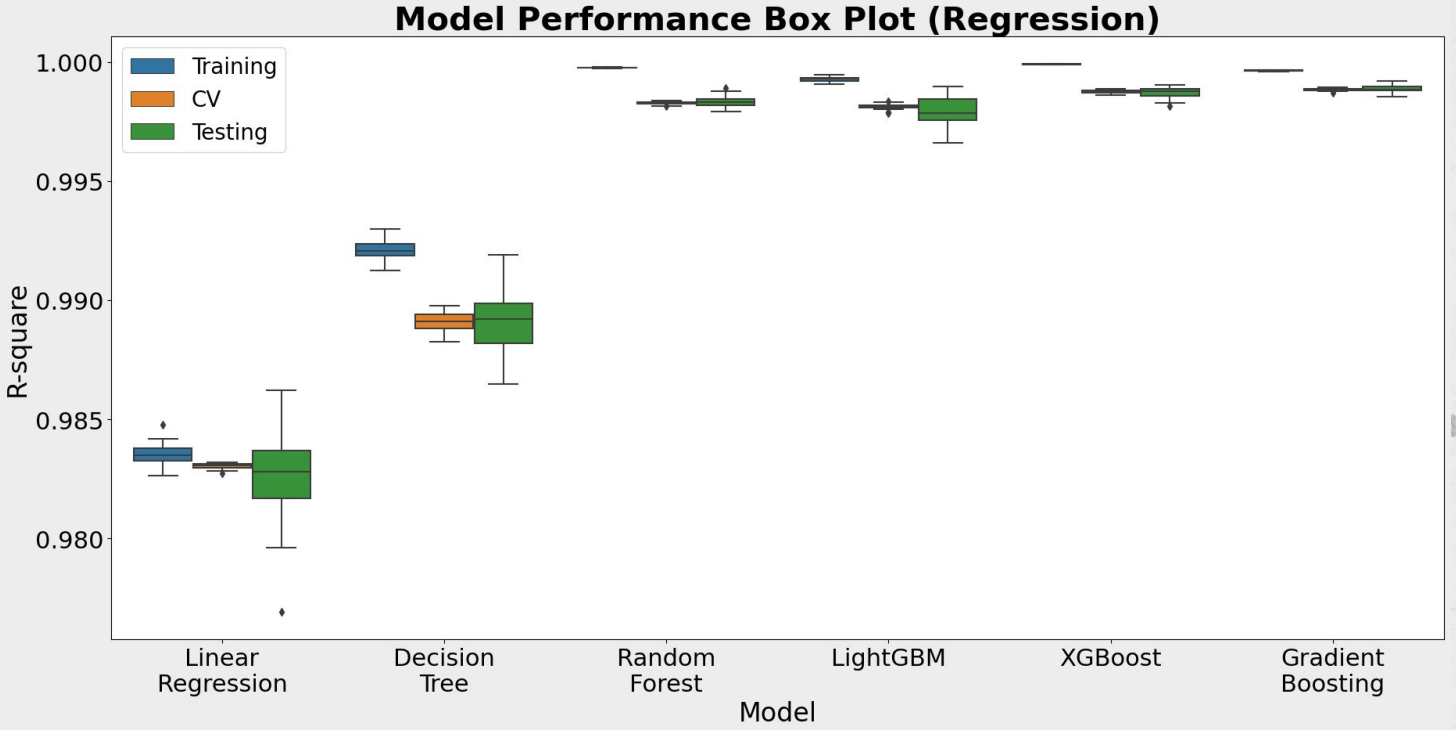
R2: 0.911
Adj. R2: 0.911

Add New Features

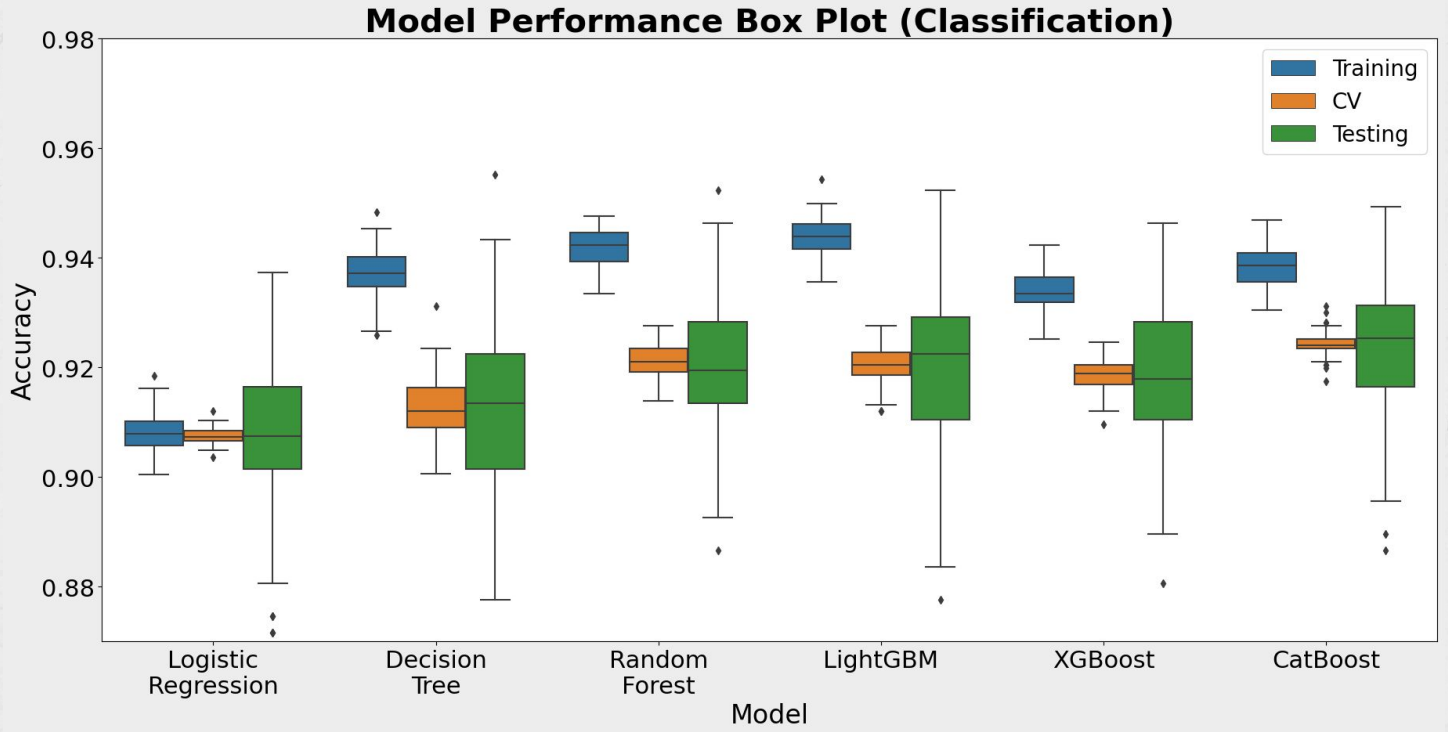
R2: 0.983
Adj. R2: 0.983

	coef	std err	t	P>|t|	[0.025	0.975]
const	-1424.4651	849.109	-1.678	0.094	-3089.900	240.970
S	-2.3056	0.035	-65.520	0.000	-2.375	-2.237
K	2.3491	0.035	67.664	0.000	2.281	2.417
tau	19.2775	25.600	0.753	0.452	-30.933	69.488
r	-95.0147	183.677	-0.517	0.605	-455.277	265.247
future_mult	133.4478	854.332	0.156	0.876	-1542.231	1809.127
S/K	1280.1535	15.103	84.759	0.000	1250.530	1309.777

MODEL EXPLORATIONS -
Regression models

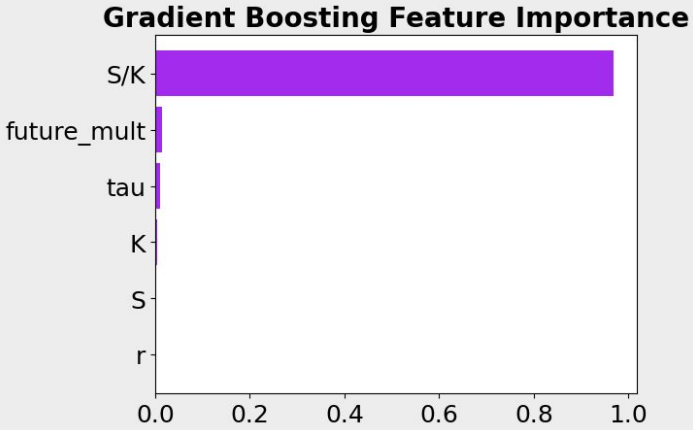


MODEL EXPLORATIONS -
Classification models



FINAL MODELS

Gradient boosting for regression



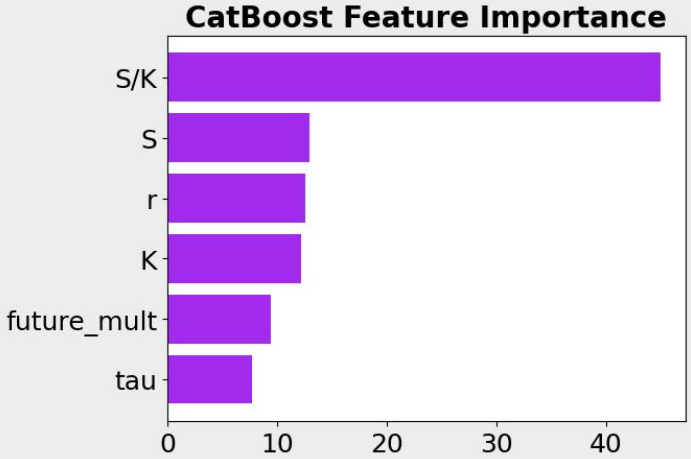
Train R2	CV R2	Test R2
0.999639	0.998847	0.998873

GradientBoostingRegressor(n_estimators=1000, learning_rate=0.1, max_depth=4)

- High in-sample and out-of sample R-square, low deviation
- High interpretability with feature importance scores and tree plot
- Speed and scalability for large dataset, fits the requirements for quantitative finance modeling

FINAL MODELS

Catboost for classification



Train	CV	Test
0.9379	0.9208	0.9268

CatBoostClassifier(iterations=75, depth = 10, learning_rate = 0.1, verbose = 0, l2_leaf_reg = 25, loss_function='Logloss')

- High in-sample and out-of sample accuracy (low classification error), relatively low variation in cross validation
- Achieved above 90% accuracy 96% of the time and above 92% accuracy 60% of the time
- High accuracy and faster training

Prediction Accuracy vs. Interpretation

Prediction accuracy is priority in the context of valuing options

- Ability to accurately value asset prices can inform profitable trading and risk management decisions
- Small errors in model prediction can result in significant losses for investors, which may have not only short-term financial consequences but also long-term repercussions
- In a competitive industry like finance, firms with an edge on accuracy will rise on top

ML Models vs. Black-Scholes

Machine learning models outperform Black-Scholes when predicting option values

- ML models allow for greater flexibility of inputs and assumptions, necessary for dynamic market conditions
  - Can account for other factors & engineer additional features as needed
- Black-Scholes makes unrealistic assumptions of implied volatility, lognormal stock price distribution, no transaction costs or taxes, etc
  - Can only be applied to valuing European options on non-dividend paying stocks

Our Confidence and Rationale

Changing Conditions

Stock market is always changing & these changes impact the option $$

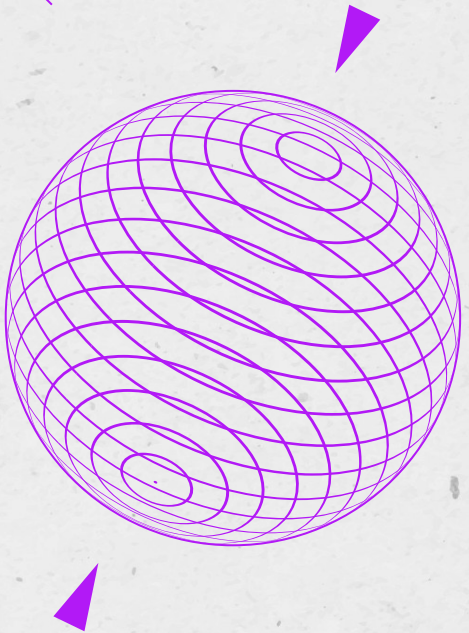
Training Data

Data may not be directly applicable

Errors or Biases

biases in the training data that could impact the accuracy

THANK YOU



Tom Do, Jessie Xue, Connie Yau, Huiling Xiao, Jiayi Hu, Yanshuo Li

2023

FINAL MODELS

Gradient boosting