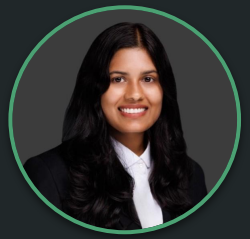
Applying Machine Learning for Options Pricing

Meet Our Team

Dipti Agrawal

Firas Rasyid



Jnana K P

Pranav Asoori

Sneha Venkatesh

Introduction

Objectives and Goals

Aim

Integrating machine learning into option pricing

Objective

Using Machine learning in predicting option value beyond the Black - Scholes formula

Goal

Building a model to predict current option value (Value) and classification for estimate of option value (BS)

Dataset

Columns	
Value	Current option value
S	Current asset value
K	Strike Price of option
Tau	Time to maturity (in years)
r	Annual interest rate
BS	Prediction estimate (Over/Under)

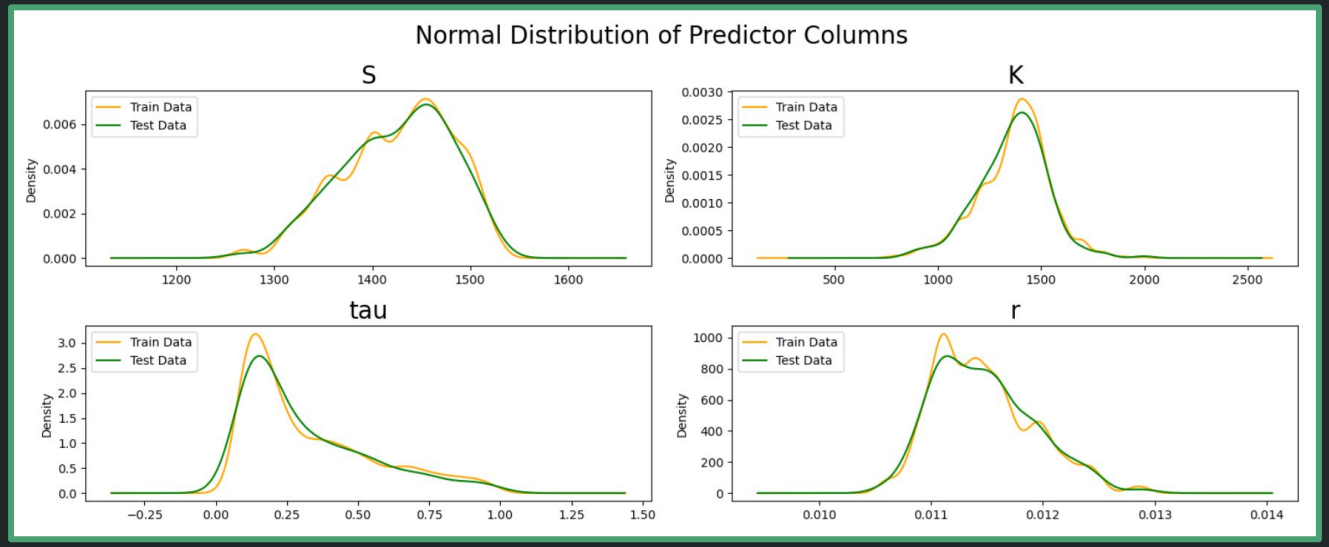
Train Data

	count	mean	std	min	25%	50%	75%	max
Value	5000.0	140.316869	125.155000	0.281250	45.750000	105.125000	200.406250	685.500000
S	5000.0	1426.643916	56.051523	1264.740000	1387.670000	1434.320000	1469.440000	1527.460000
K	5000.0	1370.244000	172.679107	750.000000	1275.000000	1400.000000	1475.000000	1995.000000
tau	5000.0	0.327615	0.231184	0.084932	0.142466	0.238356	0.463014	0.989041
r	5000.0	0.011468	0.000448	0.010600	0.011100	0.011400	0.011700	0.012900
BS_binary	5000.0	0.226400	0.418543	0.000000	0.000000	0.000000	0.000000	1.000000

Test Data

	count	mean	std	min	25%	50%	75%	max
S	500.0	1425.534580	55.269994	1264.740000	1386.170000	1432.220000	1467.285000	1527.460000
K	500.0	1360.390000	167.710021	850.000000	1250.000000	1375.000000	1470.000000	1995.000000
tau	500.0	0.316022	0.221916	0.084932	0.141781	0.235616	0.437671	0.986301
r	500.0	0.011476	0.000445	0.010600	0.011100	0.011400	0.011800	0.012900

Training and Test Data Distribution



Generally, both training and test dataset is normally distributed with similar trends. Tau and r show right skewness, while S and K shows left skewness.

Methodology

Training Data Prep

- Create a dummy variable for classification
- Separate the training data using **K-Fold Cross Validation.**

Model Selection

- Regression
  - OLS Regression
  - Decision Tree
  - Random Forest
  - XGBoost
- Classification
  - Logistic Regression
  - Decision Tree
  - Random Forest

Metric: R-Squared

Prediction using Test Data

- Fit the selected model to test data

Regression Model Comparison & Evaluation

	R-squared	Mean Squared Error	Mean Absolute Error
Linear Regression	0.923956	1180.578350	23.671689
Decision Tree	0.992507	117.153131	6.672887
XGBoost	0.997152	44.519726	4.552015
Random Forest	0.996515	54.530377	5.003687

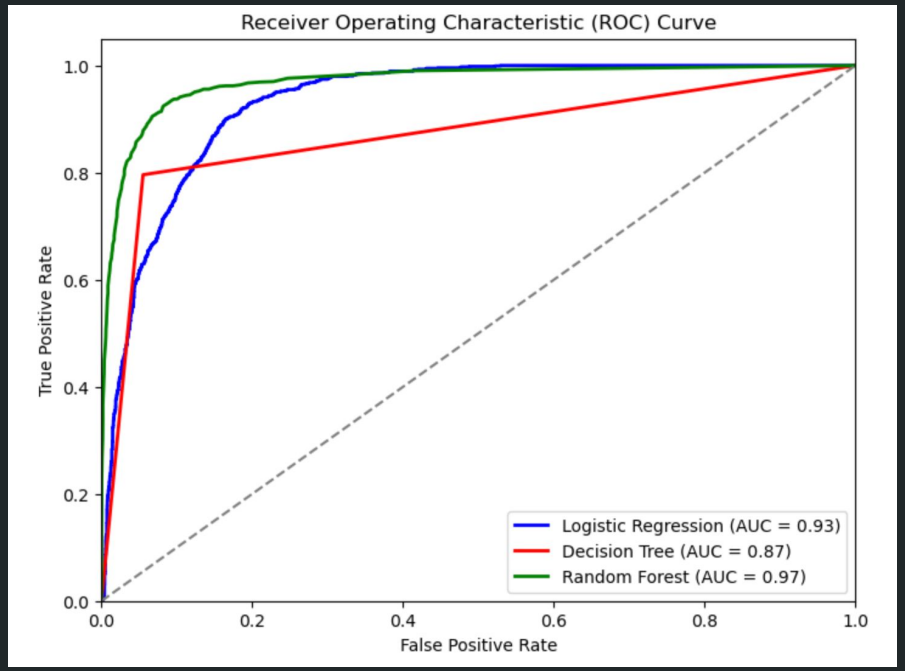
XG Boost performed the best with an Out of Sample R Squared of 0.997 and the lowest MSE & MAE

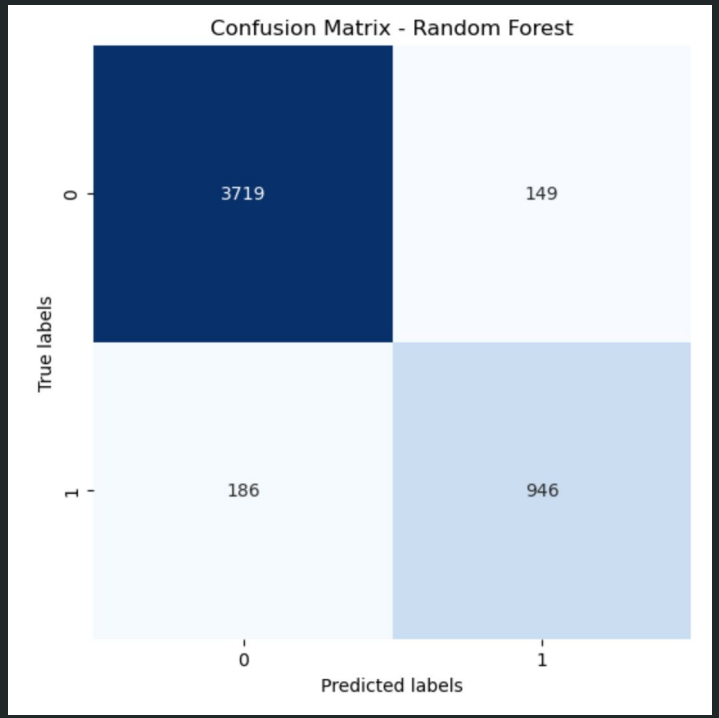
Classification Model Comparison & Evaluation

	Methods	Accuracy	F1 Score
0	Logistic Regression	0.8732	0.691648
1	Decision Tree	0.9110	0.805575
2	Random Forest	0.9330	0.850344

Random Forest performed the highest accuracy (0.933) and F1 (0.85)

Classification Model Comparison & Evaluation





Business Understandings

Prediction Accuracy

While both prediction accuracy and interpretation are important, in this context, as these decisions lead to significant consequences (high-frequency trading) prediction accuracy is paramount

Machine Learning Model

ML models are well suited to identify non-linear patterns while handling a large number of input variables. They can also adapt to changing market conditions

Variable Selection

From a business perspective, we need at least these many predictors to have a chance at predicting the dynamic nature of options price

Predicting Stock Options

We cannot be sure of using the training data as the data quality may be comprised or not of high-quality. There can be risks of overfitting too

Beyond Black-Scholes: Alternative Option Pricing Models

Binomial Option Model

Monte Carlo Simulation

Merton’s Jump Diffusion Model

Heston Model

Thank You!