Project Summary: Identifying the Best Option Price Model for FAANG Companies

1. Project Goal

The objective of this project is to determine the most accurate machine learning model for predicting option prices of FAANG companies (Meta, Apple, Amazon, Netflix, Google). The study evaluates different models and feature combinations to improve upon the traditional Black-Scholes model, incorporating real-world market behaviors such as:

- Bid-ask spreads
- Time to maturity
- Log-moneyness

2. Methodology

Data Collection & Preprocessing

- Data Source: Historical option price data fetched using the `yfinance` API.
- Key Features Extracted:
    - Strike price, last price, bid/ask spread, time to maturity, historical volatility.
    - Spot price: Latest traded stock price
    - Potential features:
        - Moneyness (`spot/strike`)
        - Log-moneyness
        - Ask-bid Spread
        - Black-Scholes price (`bs_price`): derived from historical volatility, not form implied volatility to avoid data leakage
        - Implied Volatility
- Cleaning & Filtering:
    - Removed missing/unrealistic values (e.g., negative prices, zero volatility).

- Kept options with:
  - Moneyness between 0.8 and 1.2
  - Time to maturity >1 day
  - Remove illiquid quotes: spread / midPrice < 0.5

Feature Engineering & Model Selection

- Candidate Features:
  1. `bs_price` (Black-Scholes price)
  2. `T` (time to maturity)
  3. `log_moneyness`
  4. `ask_bid_spread`
- Feature Subsets: All possible combinations (power sets) tested.
- Models Evaluated with groups of potential hyperparameters:
  1. Baseline: Black-Scholes (BS) model.
  2. Linear Models:
     - Multiple Linear Regression (MLR)
     - Ridge Regression
     - Lasso Regression
  3. Non-linear Models:
     - Polynomial Regression (Poly)
     - K-Nearest Neighbors (KNN)
  4. Ensemble Methods:
     - Bagging
     - Decision Trees
     - Random Forest
     - Gradient Boosting
     - XGBoost
  5. Neural Network: Keras-based regressor.

Training & Validation

- Cross-Validation: Using GridSearchCV to determine best hyperparameters for each model with `GroupKFold` (5 splits) to reduce data leakage (options of the same expiration date should not be separated differently).
- Performance Metric: Mean Squared Error (MSE) for model comparison.

3. Key Findings

Best Model & Performance

- Best Model: Polynomial Regression (degree=3) with various features for different companies.
- Improvement Over Black-Scholes:
    - 80% relative reduction in MSE for Meta.
    - Similar improvements for other FAANG stocks.
- Feature Importance:
    - `log_moneyness` and `T` had the strongest impact.
    - `ask_bid_spread` had a minor effect.

Model Comparisons

Model	**Performance (vs. BS)**	**Notes**
Polynomial Regression	Best (80%+ MSE ↓)	Captures non-linear trends
Random Forest	Strong (but slower)	Good for robustness
XGBoost	Competitive	Requires tuning
Neural Network	Moderate	Needs more data

4. Technologies Used

Category	**Tools/Libraries**
Data Handling	`yfinance`, `pandas`, `numpy`
Machine Learning	`scikit-learn`, `XGBoost`, `TensorFlow/Keras`

Validation	`GroupKFold`, `mean_squared_error`,`GridSearchCV`
Visualization	`Matplotlib, seaborn`

5. Result

We analysed 2220 call-option trades (after cleaning) for the big five tech names. With machine-learning tools, we improved the classic Black-Scholes price with a short, smooth curve fit. That change slashed pricing error by roughly one order of magnitude -- about 80 percent -- across every stock, even the most volatile. So with a handful of familiar inputs we get better prices estimates in real time, making this a drop-in upgrade for trading screens and risk dashboards.

5. Applications & Future Work

Practical Use Cases

- Algorithmic trading systems
- Improved hedging strategies
- Risk management in derivatives

Future Enhancements

- Collecting more historical data to make better predictions.
- Integrating advanced volatility models to the Black-Sholes model.
- Enhancing the preprocessing step.
- Testing different volatility models to better capture company-specific volatility patterns.
- Relating option price prediction to stock price prediction using advanced deep learning tools.