Financial Market Analysis and Prediction : Integrating Monte Carlo Simulations with Ensemble ML Models

TEAM MEMBERS :

Janapareddy Vidya Varshini - 230041013

Korubilli Vaishnavi - 230041016

Mullapudi Namaswi - 230041023

Project Statement :

Modern financial portfolio optimization faces significant hurdles due to the high volatility, non-linear dependencies, and inherent uncertainty of global markets. Traditional models often miss extreme market events and don't fully capture how stock prices change over time.

This project addresses these limitations by developing a hybrid intelligent framework that merges Deep Learning (LSTM) and Machine Learning (Random Forest) into an ***ensemble architecture***. By feeding these AI-driven price targets into a ***high-iteration Monte Carlo Simulation engine,*** the system provides a probabilistic range of future outcomes, enabling data-driven risk management (VaR/CVaR) and optimized capital allocation across a diversified banking portfolio.

Project Objectives :

- To develop a deep learning model (**LSTM**) to forecast future stock prices using historical market data
- To implement **Monte Carlo simulations** to generate probabilistic future price paths

- To Optimize portfolio allocation using **Sharpe ratio–based weighting**
- To Calculate key risk metrics including **Value at Risk (VaR)**
- To Build an interactive web application for portfolio visualization and decision support.

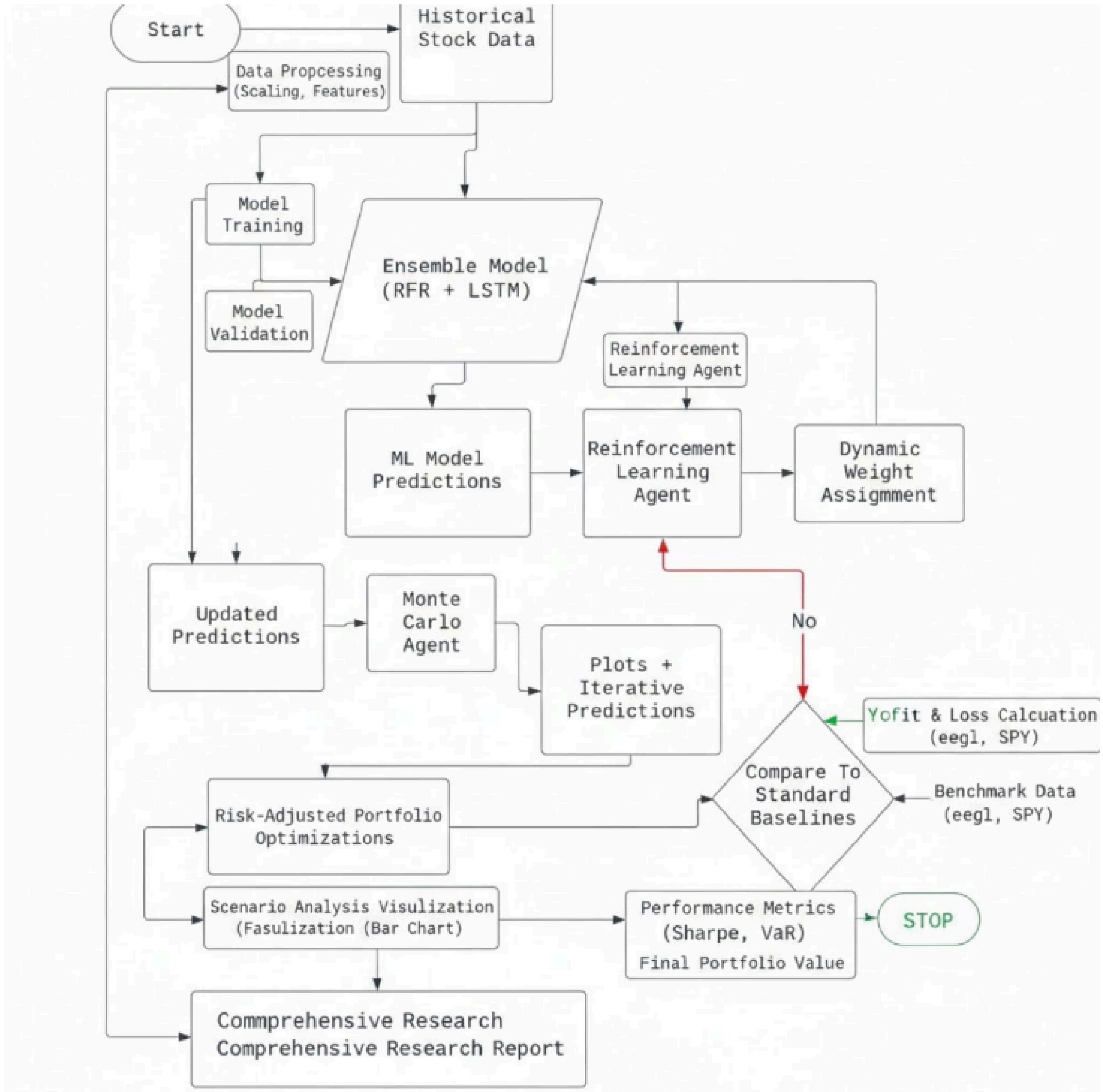
Role of Monte Carlo Simulation

In this framework, the Monte Carlo engine serves as the vital bridge between deterministic AI predictions and probabilistic reality. While the Ensemble models (LSTM and Random Forest) provide a singular target price, the Monte Carlo agent expands this point-prediction into 500 potential market paths to simulate the inherent "noise" and volatility of the banking sector.

The simulation is specifically utilized for two primary tasks:

- **Uncertainty Quantification**: The AI-derived target is converted into a **drift parameter (μ)**, while historical data provides the **volatility (σ)**. By running 500 iterations of Geometric Brownian Motion, the system captures a wide range of outcomes-from best-case growth to worst-case market corrections-ensuring the strategy is not over-fitted to a single "perfect" forecast.
- **Forward-Looking Stochastic Risk Assessment :** Rather than relying solely on historical averages or backward-looking volatility estimates, the framework conducts a forward-looking stochastic risk evaluation using Monte Carlo simulation. By analyzing the empirical distribution of 500 simulated terminal returns for each asset, the system estimates key downside risk measures, including Value at Risk (VaR) and Conditional Value at Risk (CVaR).

Operational Workflow



1. Data Ingestion : Historical and real-time financial data are collected for a selected basket of assets to form the modeling dataset.

2. AI-Based Ensemble Forecasting : Parallel models — **LSTM** (for temporal dependencies) and **Random Forest** (for nonlinear structure) — generate a 30-day price target.

Model outputs are dynamically weighted based on recent prediction performance.

3. Stochastic Simulation : The AI-predicted target acts as the drift input for a Monte Carlo engine, generating 500+ simulated price paths to capture market uncertainty.
4. Risk Quantification : Statistical analysis of simulated paths computes:

- ***Value at Risk (VaR)*** : The maximum expected loss over a specified time horizon at a given confidence level (e.g., 5%).
- ***Sharpe Ratio*** : A measure of risk-adjusted return, calculated as excess return per unit of volatility.
- ***Rachev Ratio*** : A tail-risk performance measure comparing expected extreme gains to expected extreme losses.
- ***Conditional Value at Risk (CVaR) :*** measures the average loss occurring in the worst-case scenarios beyond the VaR threshold

This quantifies both downside exposure and reward potential.

5. *Capital Allocation* : Investment capital (e.g., $10,000) is distributed across assets based on risk-adjusted performance metrics to generate an optimized portfolio strategy.

6. *Web-Based Visualization & Deployment* : An interactive web application enables

- Real-time portfolio visualization
- AI forecast generation on demand
- Risk metric dashboards
- Downloadable portfolio reports

Data & Technology Stack

Datasets:

Primary Source : Yahoo Finance API (yfinance)
Scope : Five years of daily adjusted closing prices were collected for 10 major financial institutions (JPM, GS, MS, BAC, C, WFC, HSBC, RY, TD, USB) along with the SPY S&P 500 ETF as the market benchmark. From this data, daily returns, annualized volatility, and ***Beta*** were computed.

ML Ensemble Layer

The core predictive engine uses an **Ensemble Logic** to mitigate the bias of any single model:

- **LSTM (Long Short-Term Memory):** Captures sequential dependencies and long-term trends within the 60-day sliding window of price data.
- **Random Forest:** Provides a robust, non-linear regression that excels at handling outliers and preventing the LSTM from over-fitting.
- **Reinforcement Weighting:** The system calculates the Mean Absolute Error (MAE) for both models. It assigns a higher weight to the model that is currently "winning" (lower error), ensuring the final 30-day price target is the most reliable possible synthesis.

Monte Carlo Simulation Framework

To account for market uncertainty, a Monte Carlo simulation was implemented using **Geometric Brownian Motion (GBM)**. For each stock, 500 simulated 30-day price paths were generated. The framework begins by extracting historical **volatility** and price distributions to establish a statistical baseline. Using this volatile information as a foundation for randomness, the system projects hundreds of potential future scenarios, effectively transforming a singular AI prediction into a probabilistic "cloud" of market possibilities.

Geometric Brownian Motion Model :

$$\boxed{S_t = S_{t-1} \cdot \exp\left(\left(\mu - \tfrac{1}{2}\sigma^2\right) + \sigma Z\right)}$$

S_{t-1} = Previous stock price

μ = Drift (expected return)

σ = Historical daily volatility

$Z \sim N(0,1)$ = Standard normal random variable

Drift Estimation : The drift is derived from the ensemble AI prediction to project the most likely future trajectory based on current model intelligence..

$$\mu = \frac{P_{ensemble} - P_{current}}{P_{current}}$$

Volatility Estimation: $\sigma = \text{Std Dev of historical daily returns}$

Risk Quantization: Volatility is calculated as the standard deviation of historical daily returns to measure the inherent "noise" and stability of the asset.

Stochastic Scaling: This parameter determines the spread of the simulated paths, ensuring the model respects the historical price fluctuations unique to each banking ticker.

Risk Quantification

The final phase of the framework involves translating the thousands of simulated price paths into actionable risk metrics. This process quantifies the potential downside and the efficiency of the expected returns for each of the 10 banking assets.Final simulated returns [R] are calculated by comparing the starting price (S_0) against the simulated price at the end of the 30-day forecast horizon (S_{30}):

$$R = \frac{S_{30} - S_0}{S_0}$$

From the distribution of simulated returns:

Value at Risk (VaR): $VaR_{5\%} = \text{Percentile}_{5\%}(R)$

Conditional Value at Risk (CVaR): $CVaR_{5\%} = E(R \mid R \leq VaR_{5\%})$

Sharpe Ratio: $\text{Sharpe} = \frac{\bar{R}}{\sigma}\sqrt{252}$

Rachev Ratio: $\text{Rachev} = \frac{E(R \mid R \geq P_{95})}{|E(R \mid R \leq P_5)|}$

Beta (Systematic Risk)

Using SPY as the market benchmark:

- ***Beta formula:***

$$\beta_i = \text{Cov}(R_i , R\_SPY) / \text{Var}(R\_SPY)$$

- Where: R_i = return of stock i
  R_SPY = return of SPY (market benchmark)
- ***Interpretation:***
- $\beta > 1 \rightarrow$ More volatile than the market
  $\beta < 1 \rightarrow$ Less volatile than the market
  $\beta = 1 \rightarrow$ Moves in line with the market

Ensemble Model Integration

To ensure high predictive accuracy, the system employs a hybrid intelligence layer that combines diverse algorithmic strengths. The **LSTM** model is utilized to capture complex temporal dependencies and long-term price trends, while the **Random Forest** regressor provides structural stability and handles non-linear relationships without overfitting. These models are not simply averaged; instead, the system acts as a reinforcement agent, calculating the **Mean Absolute Error (MAE)** for both and assigning higher weights to the model demonstrating superior recent performance. This optimized ensemble prediction serves as the critical "drift" parameter, guiding the Monte Carlo engine toward the most statistically likely future price target.

$$w_L = \frac{MAE_{RF}}{MAE_{LSTM} + MAE_{RF}}$$

$$P_{ensemble} = w_L P_{LSTM} + w_R P_{RF}$$

The ensemble prediction determines the drift used in Monte Carlo simulation.

Portfolio Allocation

The final stage of the project translates predictive data into a concrete investment strategy for a **user-defined principal** (e.g., $10,000). Rather than equal distribution, capital is allocated dynamically across the 10 banking tickers based on their individual **Sharpe Ratios**, which represent the return earned per unit of volatility. By favoring assets with higher risk-adjusted performance, the system maximizes potential gains while minimizing exposure to high-variance stocks. The resulting breakdown provides a clear

Weights are assigned based on Sharpe ratios:

$$w_i = \frac{\text{Sharpe}_i}{\sum \text{Sharpe}_j}$$

Portfolio volatility:

$$\sigma_p = \sqrt{w^T \Sigma w}$$

Portfolio Sharpe:

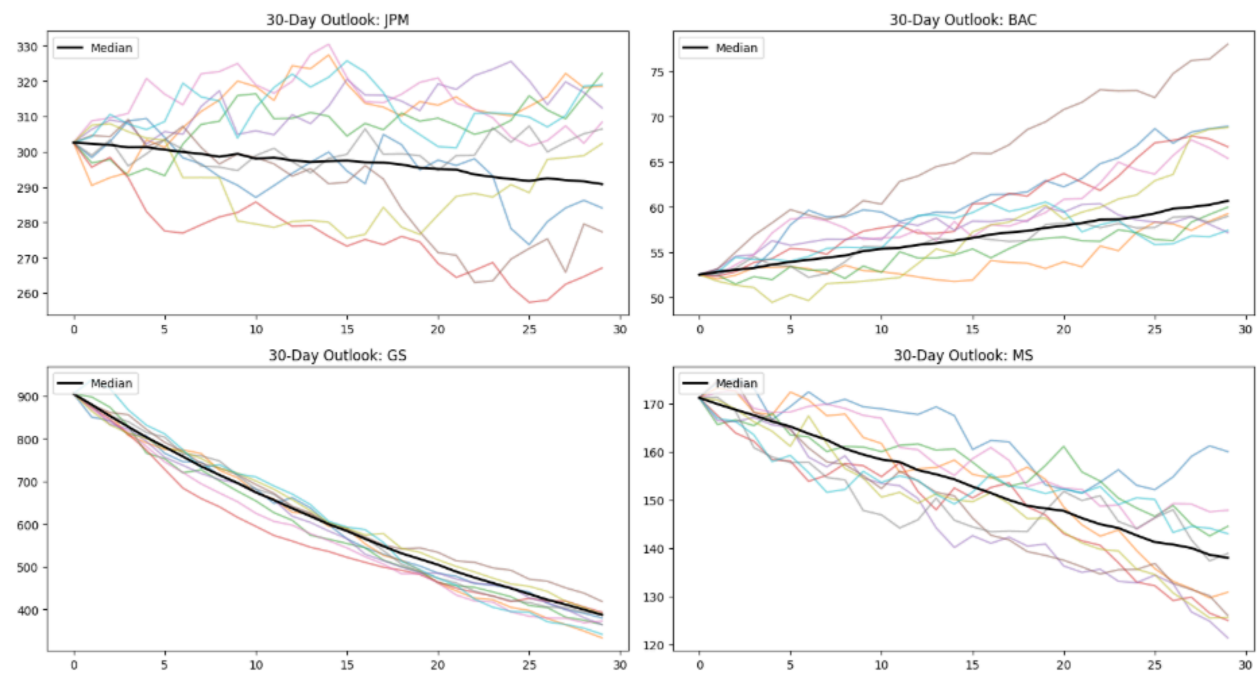
$$\text{Sharpe}_p = \frac{E[R_p]}{\sigma_p}$$

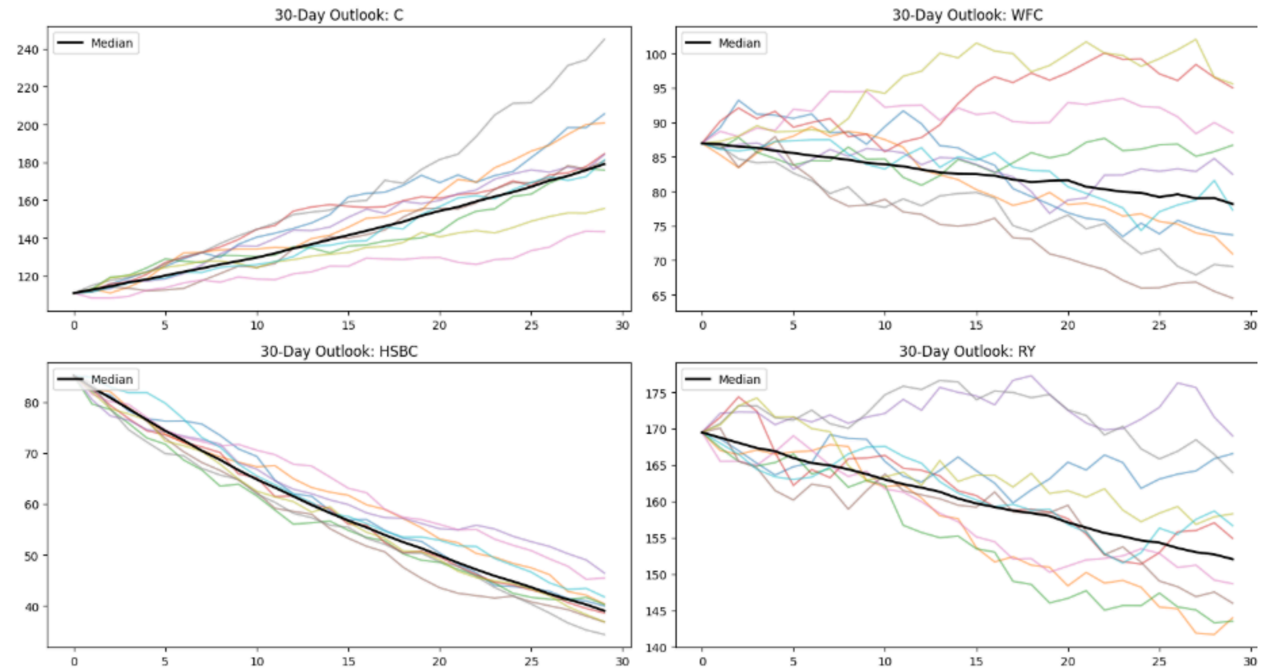
dollar-amount investment for each ticker, ensuring the total portfolio is optimized for both growth and safety based on the specific investment amount entered by the user.

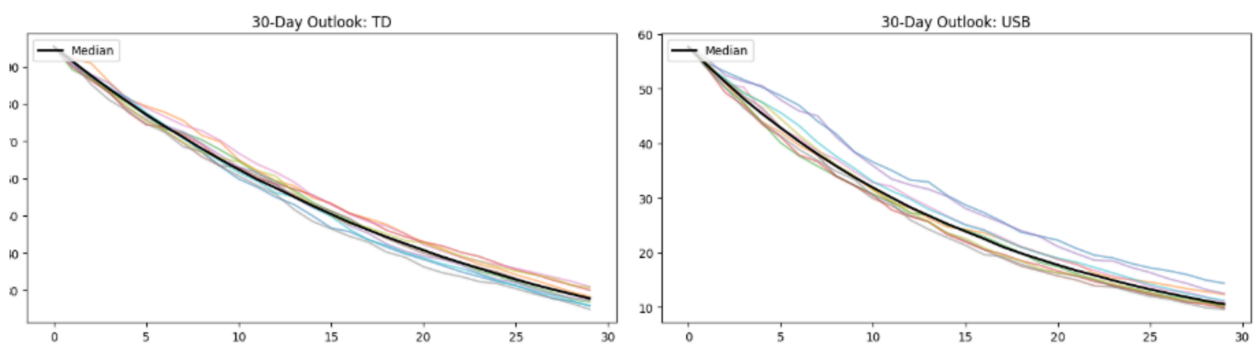
Simulation Results & Visual Inference :

Sector-Wide 30-Day Outlook :

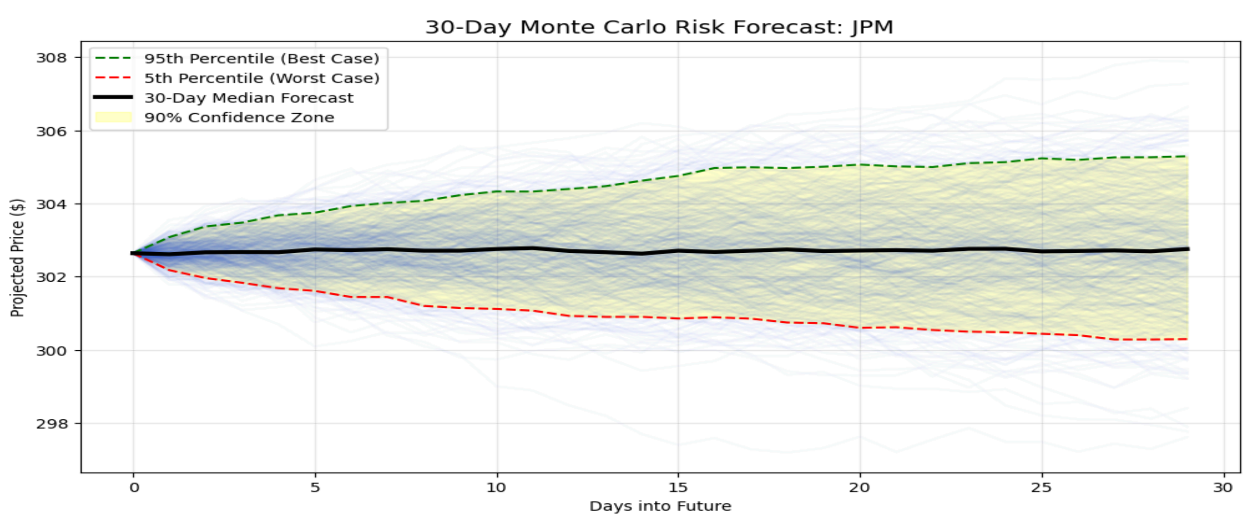
Visualization of Uncertainty : 500 simulations are executed per stock for statistical robustness, only the first 10 paths are plotted to maintain visual clarity.
Median Forecast Path : The black line represents the 50th percentile of simulated outcomes, indicating the most probable trajectory based on AI-estimated drift.
Axis Definition : The X-axis represents the 30-day forecast horizon, while the Y-axis shows projected stock price in USD.
Sentiment Inference : The slope of the median path indicates directional bias downward suggests bearish outlook; upward suggests bullish outlook.
Volatility Interpretation : The vertical dispersion of simulated paths reflects historical volatility, with wider spreads indicating higher uncertainty.





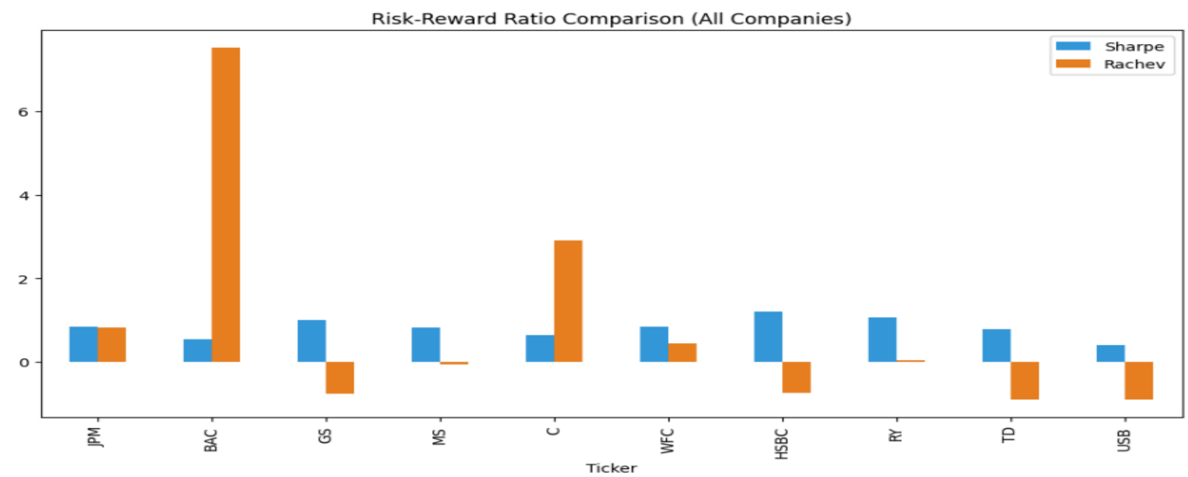


30-Day Price Risk Outlook for JPMorgan Chase & Co using 500 paths:



The Monte Carlo simulation projected JPMorgan Chase & Co. price over 30 days, showing the median forecast with best-case (95th percentile) and worst-case (5th percentile) scenarios.

Risk-Reward Ratio Comparison :



Risk-reward profiles vary significantly across companies, highlighting differences in performance efficiency and return consistency.

Daily Return & Risk Summary:

--- SECTION 1: DAILY RETURN STATISTICS ---

Ticker	Mean Daily Return	Daily Std Dev	Annual Volatility	Beta
JPM	0.000807	0.015318	0.243162	0.887908
BAC	0.000580	0.016914	0.268494	0.965683
GS	0.001092	0.017243	0.273719	1.101221
MS	0.000933	0.017849	0.283338	1.143490
C	0.000740	0.018110	0.287491	1.069555
WFC	0.001005	0.019089	0.303022	0.997057
HSBC	0.001201	0.015738	0.249840	0.676979
RY	0.000760	0.011241	0.178444	0.648840
TD	0.000610	0.012359	0.196199	0.624508
USB	0.000483	0.018767	0.297916	0.980714

Highest return : GS ; **Lowest return** : USB.
Highest volatility : WFC ; **Lowest volatility** : RY.
Most market-sensitive (High Beta) : MS, GS, C.
Defensive (Least Sensitive - Low Beta) : TD, RY, HSBC.

- GS & MS = high return–high risk;
- RY & TD = low risk–stable profile.

ML Performance Summary:

--- SECTION 2: ML PERFORMANCE METRICS ---

Ticker	MAE	MSE	RMSE	LSTM_Weight
JPM	0.036059	0.002109	0.045920	0.806988
BAC	0.040524	0.002530	0.050295	0.788505
GS	0.037087	0.002339	0.048361	0.875509
MS	0.036059	0.002100	0.045823	0.858513
C	0.039229	0.002449	0.049485	0.864492
WFC	0.046229	0.003312	0.057549	0.677702
HSBC	0.037547	0.002483	0.049832	0.849439
RY	0.023726	0.000984	0.031362	0.900500
TD	0.052865	0.004269	0.065340	0.833614
USB	0.037487	0.002819	0.053091	0.506252

- **Lowest error (Best performance):** RY (lowest MAE, MSE, RMSE).
- **Highest error (Worst performance):** TD and WFC.
- **Highest LSTM weight:** RY (~0.90) → strongest model influence.
- **Lowest LSTM weight:** USB (~0.51) → weaker model contribution.
- The model performs best on stable stocks (e.g., RY) and worse on more volatile ones (e.g., TD, WFC).

Risk & Ratio Summary:

--- SECTION 3: RISK & RATIO SUMMARY ---

Ticker	VaR (5%)	CVaR (5%)	Sharpe	Rachev
JPM	-0.164841	-0.180754	0.835946	0.829662
BAC	0.003446	-0.052130	0.544342	7.518473
GS	-0.633956	-0.646571	1.005241	-0.753066
MS	-0.312821	-0.336389	0.829757	-0.066329
C	0.386132	0.332396	0.648285	2.916502
WFC	-0.231298	-0.268455	0.836112	0.447217
HSBC	-0.602180	-0.611277	1.211711	-0.750148
RY	-0.190577	-0.210445	1.072973	0.044171
TD	-0.737961	-0.746591	0.783812	-0.894745
USB	-0.847394	-0.852993	0.408614	-0.910698

Sharpe Ratio: HSBC and RY have higher Sharpe → better return per unit of total risk; GS and USB have low Sharpe → poor risk-adjusted performance.
VaR (5%): USB and TD show more negative VaR → higher potential loss in worst 5% scenarios; JPM and RY have comparatively lower extreme loss risk.

CVaR (5%): USB and TD have very negative CVaR → heavier tail risk (larger losses during extreme downturns).
Rachev Ratio: Higher values (e.g., BAC, C) indicate better upside-to-downside tail balance; negative or low values indicate unfavorable extreme risk profile.Overall, HSBC and RY appear more efficient on a risk-adjusted basis, while USB and TD carry higher tail risk.

Investment Strategy - (Ex : $10,000)Optimized Allocation:

```
INVESTMENT STRATEGY: $10,000 ALLOCATION
==================================================
Ticker  Allocation (%)  Amount ($)
   JPM       10.223399 1022.339932
   BAC        6.657159  665.715861
    GS       12.293827 1229.382698
    MS       10.147702 1014.770190
     C        7.928353  792.835301
   WFC       10.225424 1022.542411
  HSBC       14.818897 1481.889673
    RY       13.122179 1312.217949
    TD        9.585813  958.581347
   USB        4.997246  499.724638

-----------------------------------
Expected Annual Return: 21.97%
Risk (Volatility):      20.65%
Portfolio Sharpe Ratio: 1.06
```

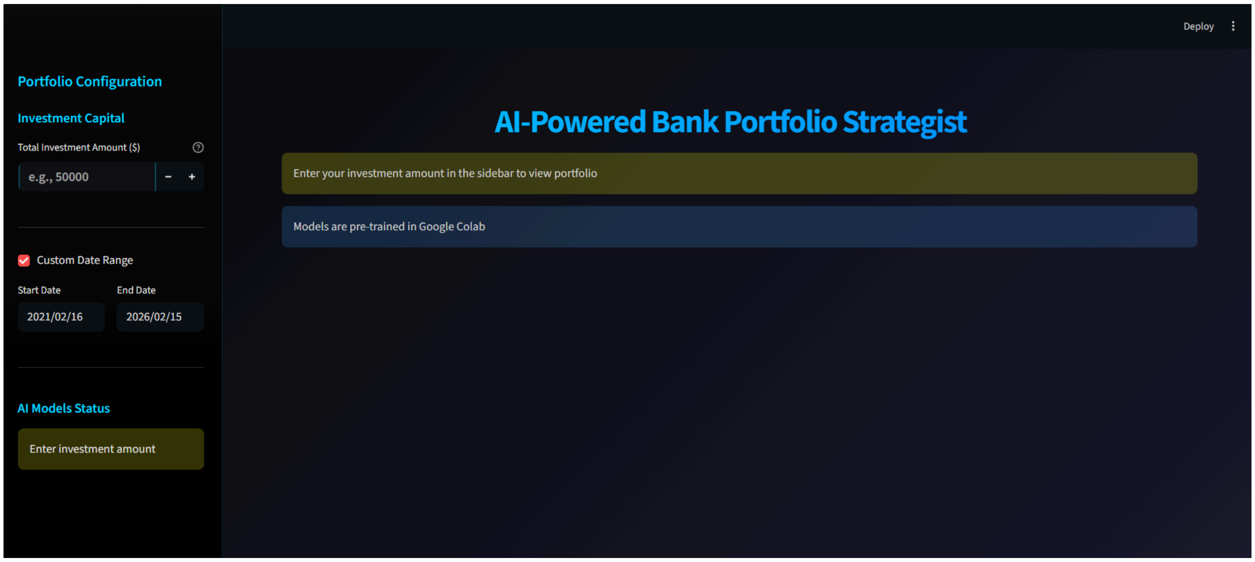
Highest allocations: HSBC (14.82%) and RY (13.12%) → They had the highest Sharpe ratios (HSBC ≈ 1.21, RY ≈ 1.07), so they provide the best return per unit of risk.
Moderate allocations : GS, JPM, MS, WFC → These stocks had mid-range Sharpe ratios (around 0.82–0.93), so they contribute to returns but are not the most efficient.
Lower allocations: BAC (6.66%), C (7.93%) → They had lower Sharpe ratios (around 0.54–0.65), so they contribute less efficiently to risk-adjusted return.
Lowest allocation: USB (4.99%) → It had the lowest Sharpe ratio (~0.41) and higher negative tail risk (VaR/CVaR), so the optimizer assigns minimal weight.

Financial Analysis Dashboard & Web Interface :

The project concludes with a user-centric web interface, allowing investors to personalize their strategy in real-time.

A. Portfolio Configuration

Users can input any investment amount and a custom date range to immediately see an optimized strategy.



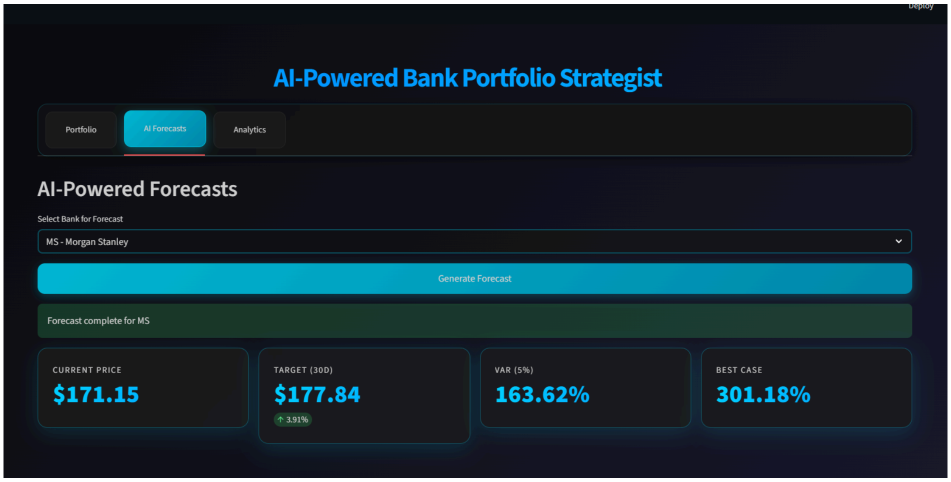
B. Optimized Allocation Strategy

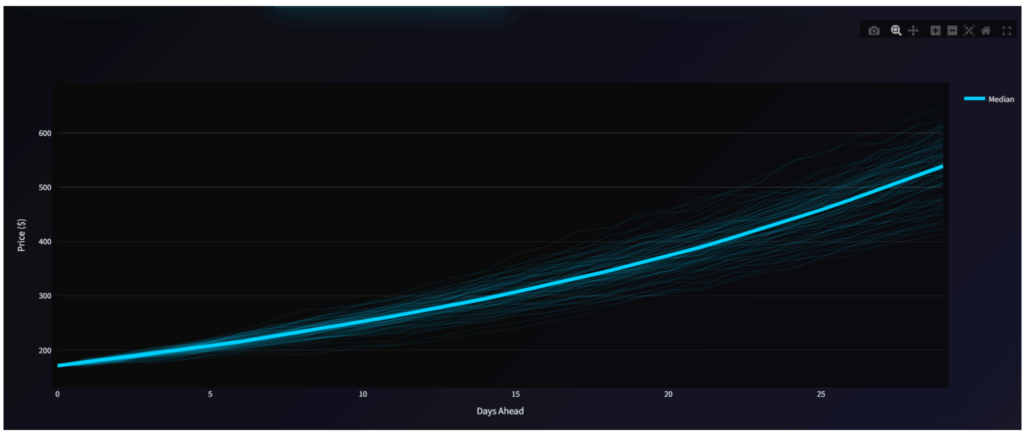
The dashboard provides a visual breakdown of the suggested investment.

- **Capital Distribution**: Displays exactly how much money to invest in each bank based on its **Sharpe Ratio**.
- **Portfolio Metrics**: Shows the total expected **Annual Return**, **Risk (Volatility)**, and the combined **Sharpe Ratio** for the entire portfolio.

C. Individual AI Forecasts

Users can select a specific bank to view detailed 30-day price targets, Value at Risk (VaR), and "Best Case" scenarios generated by the Monte Carlo agent.





D. Performance Analytics

A dedicated analytics view tracks the historical normalized performance of all 10 tickers.

Conclusion:

This project demonstrates the effectiveness of bridging Ensemble Machine Learning with Monte Carlo simulations to create a robust, data-driven framework for financial analysis. By utilizing a Reinforcement Learning agent to dynamically weight LSTM and Random Forest predictions, the system generates a high-fidelity "drift" parameter that accounts for both temporal trends and non-linear market behaviors. The transition from static price targets to a probabilistic cloud of 500 simulated paths allows for a precise quantification of Value at Risk (VaR) and Sharpe Ratios. Ultimately, this integrated pipeline successfully automates the path from raw historical data to an optimized, risk-adjusted portfolio, providing investors with a significant analytical advantage in navigating market uncertainty.

Future Improvements :

- **Sentiment Analysis Integration**: Incorporating a Natural Language Processing (NLP) layer to analyze financial news and social media sentiment would allow the model to adjust "drift" based on qualitative market expectations that historical data alone cannot capture.
- **Dynamic Asset Expansion**: Scaling the framework beyond 10 tickers to include hundreds of diverse assets across various sectors to minimize industry-specific risk.

Code Repository:

https://github.com/Namaswi24/MonteCarlo-Ensemble-Finance

Google Colab link:

MONTE_CARLO_PROJECT.ipynb

References:

Advanced financial market forecasting: integrating Monte Carlo simulations with ensemble Machine Learning models