Project 2: Stock Portfolio Forecasting and Optimization Using Machine Learning

Applying Predictive Modeling to the S&P 500 Index

Moisés Santos and Hélder Oliveira

Overview of the Project

- Main goals:
  - Predict S&P 500 stock prices
  - Optimize a portfolio for daily investments

- Outline the approach using machine learning and optimization techniques

Context and Motivation

- Challenges in predicting and managing portfolios
  - **Market Volatility**: Financial markets are highly sensitive to both internal factors (like earnings reports) and external factors (such as geopolitical events).
  - **Overfitting**: Especially when using historical data, there is a risk of overfitting, where a model performs well on past data but poorly on new, unseen data.
  - **Non-Stationarity**: Financial time series data is often non-stationary, meaning statistical properties like mean and variance change over time.
  - **Unexpected Events (Black Swan Events)**: Events like the 2008 financial crisis or the COVID-19 pandemic are highly unpredictable but can have dramatic impacts on portfolios.

Project Objectives

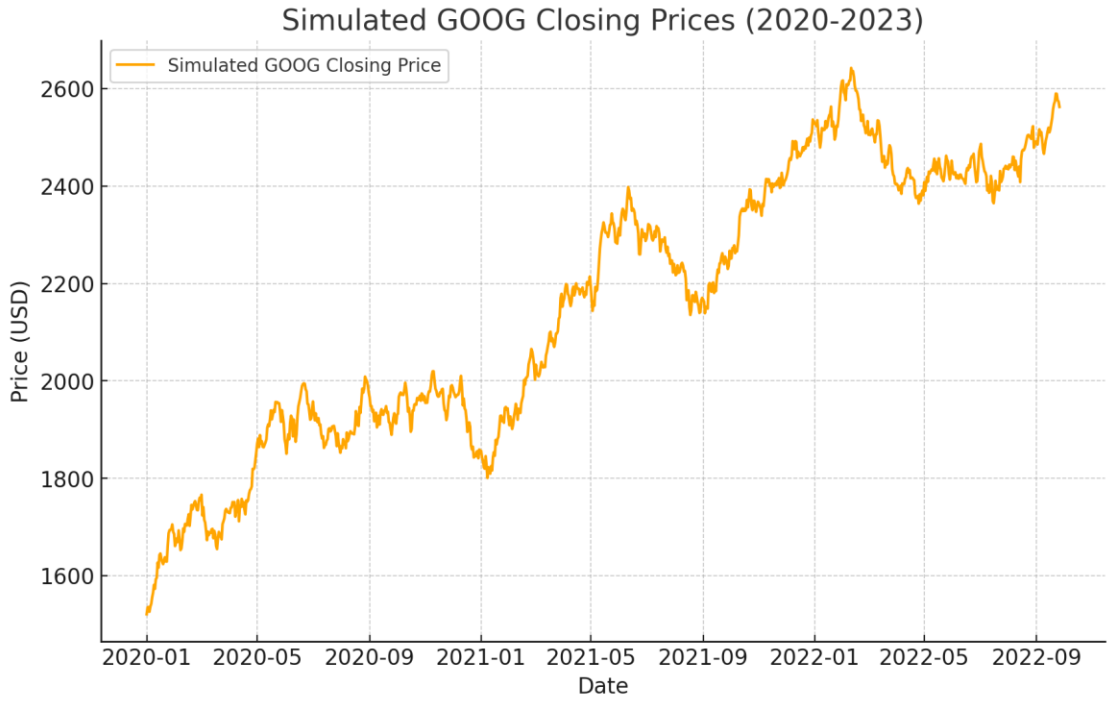
- **Forecasting**: Use machine learning to predict daily stock prices
- **Optimization**: Select the best stocks for a portfolio to maximize returns or minimize risks
- **Implementation**: Apply models from 2010-2023 data and evaluate in January 2024

Related Works

- Some key studies on machine learning in finance:
  - *Patel et al. (2015)* - Random Forest & SVM for market predictions
  - *Bao et al. (2017)* - LSTM for financial time series
  - *Glasserman (2003)* - Monte Carlo for portfolio optimization

Dataset - S&P 500

- Data source: Yahoo Finance API, etc.
- Overview of data structure (Open, High, Low, Close, Volume)
  - Close to Close: Open position on close price today and close position on close price tomorrow
  - Day ahead: Open position on close price today and close position on open price tomorrow
  - Intraday: Open position on open price tomorrow and close position on close price tomorrow
- **Time Frame**: Train on 2010-2023, simulate January 2024



Machine Learning Models for Forecasting

- **Algorithms**:
  - Classical: Random Forest, Decision Trees, SVM
  - Deep Learning: LSTM, MLP

Forecasting Workflow

Data Preprocessing: Cleaning, normalization, feature engineering

Model Training: Cross-validation, tuning hyperparameters

Evaluation Metrics: Use MAE, RMSE, and MAPE

Portfolio Optimization

Optimization Goals:

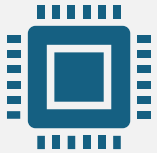
- MinMax: Minimize maximum risk
- MaxMin: Maximize minimum return

Techniques:

- Monte Carlo Simulation
- Genetic Algorithms
- Etc.

Portfolio Optimization Process

Define Objective: Maximize return, minimize risk



Simulate Scenarios: Apply Monte Carlo, genetic algorithms, etc

Evaluate Performance: Check the cummulative return

Negotiation Constraints

- Initial Amount: 10000 USD
- Maximum volume to buy: 100 stocks per day
- Operation cost: 1 USD

Project Workflow

- Data Collection
- Exploratory Data Analysis
- Forecasting or Prediction Models
- Portfolio Optimization
- Evaluation and Reporting

Submission Requirements

Notebook with well-documented code

Video (up to 5 minutes) summarizing problem, solution, and results

One-page document on ethical considerations

Evaluation Criteria

Product Fit: 15%

Business Impact: 20%

Technical Proficiency: 40%

Soft Skills: 15%

Ethics and Legal Awareness: 10%

Challenges & Solutions

Handling Large Datasets

Model Selection and Tuning

Risk Management

Practical Tips

- **Start Simple**: Test with classical models first
- **Visualize Results**: Use EDA and charts to understand data trends
- **Collaborate Effectively**: Assign roles within the team

Questions?