Data Science Capstone Project Proposal

**Title**: Stock Prediction and Portfolio Optimization App  
**Student Name**: David Heller  
**Instructor**: Professor Ananda Mondal  
**Project Mentor**: Professor Florence George

**Problem Statement**

Investing in stock markets is inherently complex, requiring accurate forecasting and effective risk management. Many individual investors lack access to tools that integrate robust stock predictions with optimized portfolio allocations, leaving them at a disadvantage in making informed financial decisions. Existing solutions are often either overly simplistic or prohibitively complex and expensive. This project aims to bridge this gap by creating an accessible, data-driven tool to empower individual investors.

**Dataset**

The dataset includes around 15 years of daily stock price data (from January 1, 2009, to October 1, 2024) for 8 prominent stocks: Netflix (NFLX), Google (GOOGL), IBM, Johnson & Johnson (JNJ), Coca-Cola (KO), Microsoft (MSFT), Nike (NKE), and Apple (AAPL). The data was obtained programmatically using the *yfinance* Python library (Yahoo Finance’s API), comprising 3,962 rows per stock and a total of 31,696 data points (3,962 x 8 stocks). The dataset contains six columns with the following data types:

* Adj Close, Close, High, Low, Open: float64
* Volume: int64

Additionally, several features will be engineered, including technical and momentum indicators (e.g., moving averages, Bollinger Bands, RSI) and date-based features (e.g., day of the week, month, quarter-end).

**Methodology**

1. **Data Preprocessing**: Feature engineering and data cleaning to prepare the dataset for analysis.
2. **Machine Learning Models**: Develop and evaluate ARIMA, XGBoost, and LSTM models, selecting the best-performing model based on prediction accuracy and robustness.
3. **Portfolio Optimization**: Implement Monte Carlo simulations to recommend portfolio allocations that balance risk and return.
4. **App Development**: Build an interactive Streamlit-based app that integrates prediction models and portfolio optimization tools for user accessibility.

**Expected Outcomes**

1. **Comprehensive and User-Friendly Tool**: Integrated stock data exploration, price prediction, and portfolio optimization.
2. **Reliable Stock Predictions**: Use of advanced ML models for precise stock price forecasting. Users will be able to select the time horizon for the predictions.
3. **Optimized Portfolio Allocation**: Recommendations based on Monte Carlo simulations, offering metrics such as expected returns, volatility, and Sharpe Ratio.
4. **Empowerment of Individual Investors**: Accessible and actionable insights for investors with minimal technical expertise. Provide financial insights through data-driven decision-making.
5. **Academic and Professional Growth**: Hands-on experience with advanced techniques, including deep learning, portfolio optimization, and app development. This project will showcase my ability to tackle real-world financial challenges.

**Impact**

1. **Practical Application**: Empowers individual investors by democratizing advanced financial analytics.
2. **Societal Benefits**: Reduces barriers to effective investing by simplifying complex tools and eliminating high costs.
3. **Future Opportunities**: Lays the foundation for expanding functionality, integrating more stocks, and improving predictive capabilities with hybrid machine learning models.