Prototype Development and Business/Financial Modelling of Stock Price Prediction

Project submitted by

Rahul Kumar

In partial fulfilment of two months internship

At FeyNN Labs: AI for small Businesses

11th August 2023

1.Prototype Selection:

1.1 Data Preprocessing: Before training a stock price prediction model, it's important to preprocess the data. This involves cleaning, normalizing, and transforming the data to remove noise and inconsistencies. Selecting a subset of the data reduces the computational load and helps in focusing on relevant patterns.

1.2 Diverse Time Periods: By selecting prototypes from various time periods, you capture the changing market dynamics over time. Different market cycles, such as bullish and bearish phases, can have distinct patterns that your model needs to learn.

1.3 Volatility Levels: Stocks with varying levels of price volatility can teach the model how to handle different levels of risk and uncertainty in price movements.

1.4 Sector Diversity: Different sectors of the economy respond differently to market conditions. Including stocks from various sectors provides a holistic view of market behaviour and helps the model account for sector-specific factors.

1.5 Market Cap Range: The market capitalization of a company can influence how its stock price behaves. Including prototypes with different market caps allows the model to learn the impact of company size on price movements.

1.6 Liquidity: Liquidity, represented by trading volume, affects how easily a stock can be bought or sold without significantly impacting its price. Prototypes with varying levels of liquidity teach the model about the relationship between trading volume and price changes.

1.7 News Impact: Stocks that have been influenced by significant news events provide insights into how external factors can lead to rapid price fluctuations. This helps the model learn the impact of news sentiment on stock prices.

1.8 Cyclical vs. Non-Cyclical: Different sectors are influenced by economic cycles differently. Including prototypes from both cyclical (e.g., automotive) and non-cyclical (e.g., healthcare) sectors enables the model to understand varying sensitivities to economic conditions.

1.9 Technical Indicators: Technical indicators like moving averages, RSI, and MACD provide insights into price trends and momentum. Prototypes with diverse technical indicator patterns help the model recognize a wide range of signals.

1.10 Sentiment Analysis: Incorporating sentiment analysis data allows the model to learn the correlation between positive/negative sentiment and stock price movements. This is particularly useful in understanding the impact of public perception on stock prices.

1.11 Machine Learning Models: Trying different machine learning algorithms for stock price prediction, such as neural networks, random forests, or support vector machines, helps you identify which models perform best for different prototypes and market conditions.

1.12 Feature Importance: Analyzing feature importance for each prototype helps you understand which variables or factors have the most influence on stock price movements. This insight can guide feature selection and engineering efforts.

1.13 Outlier Detection: Prototypes that have experienced extreme outlier events (e.g., sudden price jumps or crashes) help the model learn to handle unexpected and rare market occurrences.

1.14 Market Regimes: Partitioning the data into different market regimes (e.g., bull markets, bear markets, sideways markets) helps the model adapt to changing conditions and tailor its predictions accordingly.

1.15 Validation and Testing: Evaluating the selected prototypes on validation and testing datasets ensures that the model's predictive performance holds across different time periods, thus validating the representative nature of the prototype selection.

2.Prototype Development:

2.1. Data Collection: Acquire historical stock price data from reputable financial data providers or APIs like Yahoo Finance or Alpha Vantage. Include attributes such as opening and closing prices, trading volume, and potentially relevant financial metrics.

2.2. Feature Selection: Identify features that might impact stock prices, such as technical indicators (moving averages, relative strength index), fundamental metrics (earnings per share, price-to-earnings ratio), and external factors (interest rates, market sentiment).

2.3. Data Preprocessing: Clean the data by addressing missing values, outliers, and anomalies. Normalize numerical features to a consistent scale (e.g., between 0 and 1) to prevent model bias.

2.4. Time-Series Split: Split the dataset chronologically, usually allocating around 70-80% for training, 10-15% for validation, and the remaining portion for testing. This ensures the model is evaluated on unseen data.

2.5. Model Selection: Depending on the complexity of the problem and the available data, consider models like ARIMA for linear trends, RNNs/LSTMs for capturing temporal dependencies, or hybrid models that combine various techniques.

2.**6. Hyperparameter Tuning:** Adjust parameters such as learning rate, batch size, and network architecture (for deep learning models) using techniques like grid search or more advanced methods like Bayesian optimization.

2.7. Feature Engineering: Create additional features like rate of change, Bollinger Bands, or sentiment scores from market news using natural language processing (NLP) techniques to provide the model with richer input.

2.8. Model Training: Use the training data to teach the model to recognize patterns and relationships between input features and target prices. Employ appropriate loss functions (e.g., mean squared error) and optimization algorithms (e.g., Adam) to fine-tune model weights.

2.9. Validation and Testing: Assess the model's performance on the validation set, monitoring metrics like mean absolute error or root mean squared error. Refine the model iteratively based on validation results before testing on the untouched test set.

2.10. Ensemble Techniques: Combine predictions from multiple models using techniques like bagging (bootstrap aggregating) or boosting to improve prediction accuracy and reduce overfitting.

2.11. Backtesting: Simulate trading scenarios by applying the model's predictions to historical data and measuring its hypothetical performance over time. This helps understand potential profitability and risk.

2.12. Real-time Data Integration: Implement mechanisms to periodically update the model with new data and retrain it to adapt to evolving market conditions and avoid model drift.

2.**13. Model Interpretability:** Utilize techniques like SHAP (SHapley Additive exPlanations) values or LIME (Local Interpretable Model-agnostic Explanations) to explain why the model makes specific predictions, enhancing trust and usability.

2.14. Risk Management: Develop strategies to manage risk, such as setting stop-loss orders based on predicted price changes, diversifying investments, or dynamically adjusting position sizes based on model confidence.

2.15. User Interface (UI): Create a user-friendly web or mobile interface that displays predictions, historical performance, and relevant market news. Enable users to interact with the model and make informed decisions.

3.Business Modelling:

Developing a comprehensive business model for stock price prediction involves a strategic approach to harness the potential of predictive analytics in the financial sector. Here's an elaborate breakdown of the key components and considerations for building a business model around stock price prediction:

3.1. Problem Definition and Value Proposition:

- Clearly define the problem you aim to solve, such as providing accurate short-term or long-term stock price predictions.

- Articulate the value proposition, emphasizing how accurate predictions can empower investors to make informed trading decisions and potentially enhance their returns.

3.2. Market Research and Target Audience:

- Conduct thorough market research to understand the demand for stock price prediction services, existing competitors, and gaps in the market.

- Identify your target audience, which could include individual investors, hedge funds, financial institutions, or even retail traders.

3.3. Data Acquisition and Partnerships:

- Establish partnerships or subscriptions with reliable financial data providers to access historical and real-time stock market data.

- Ensure data quality and integrity to build a robust prediction model.

4.4. Feature Engineering and Model Development:

- Engage data scientists and domain experts to engineer relevant features, combining technical indicators, fundamental metrics, and sentiment analysis.

- Develop and fine-tune predictive models such as machine learning algorithms (e.g., regression, random forests) or deep learning architectures (e.g., LSTM, CNN) based on your dataset.

3.5. Technology Infrastructure:

- Build a scalable and secure technology infrastructure to handle data storage, preprocessing, model training, and deployment.

- Consider cloud-based solutions for flexibility and scalability.

3.6. Model Interpretability and Transparency:

- Implement model interpretability techniques to explain predictions to users, enhancing trust and user engagement.

- Provide users with insights into how specific features and factors contribute to each prediction.

3.7. User Interface and Experience:

- Design an intuitive and user-friendly interface that displays stock predictions, historical performance, and other relevant information.

- Allow users to customize settings, set alerts, and interact with the predictions.

3.8. Subscription Plans and Monetization:

- Offer different subscription tiers with varying features, such as access to real-time predictions, advanced analytics, and personalized insights.

- Consider freemium models, one-time purchases, or monthly/annual subscriptions.

3.9. Risk Management and Disclaimers:

- Clearly communicate that predictions are not guaranteed outcomes and involve inherent risks.

- Provide educational resources on risk management and responsible trading practices.

3.10. Marketing and Customer Acquisition:

- Develop a marketing strategy to promote your platform through content marketing, social media, partnerships with financial influencers, and targeted advertising.

- Showcase case studies or backtesting results to demonstrate the effectiveness of your predictions.

3.11. Regulatory and Legal Compliance:

- Understand and adhere to financial regulations and securities laws that apply to providing financial insights or advice.

- Consult legal experts to ensure compliance with relevant regulations.

3.12. Continuous Improvement and Innovation:

- Implement feedback loops to gather user insights and enhance your models and platform based on user needs and preferences.

- Stay updated with advancements in predictive analytics and incorporate new techniques to improve prediction accuracy.

3.13. Performance Tracking and Reporting:

- Provide users with performance reports, comparing predicted prices against actual market movements.

- Use transparency in reporting to build credibility and maintain user trust.

3.14. Customer Support and Engagement

- Offer responsive customer support to address user inquiries, technical issues, and provide assistance with interpreting predictions.

3.15. Scaling and Future Expansion:

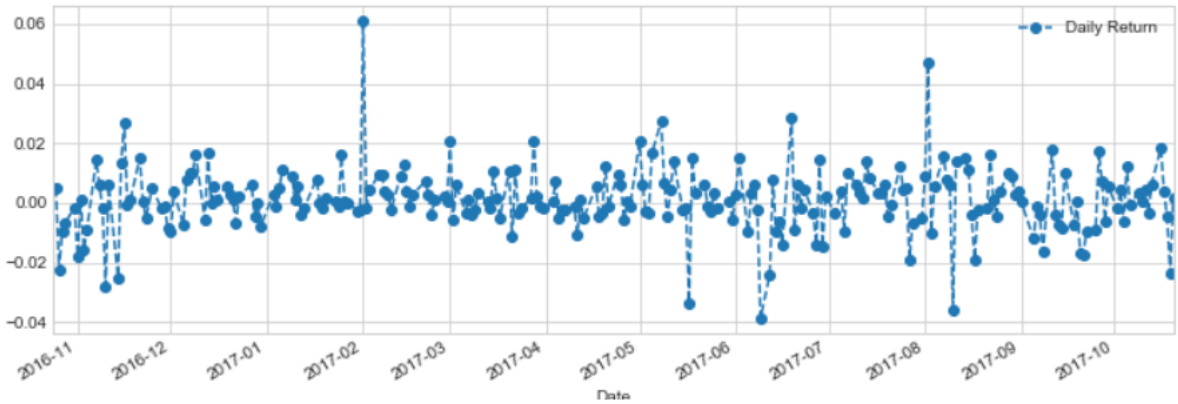
- Plan for scalability by optimizing your infrastructure to handle increased data volume and user traffic.

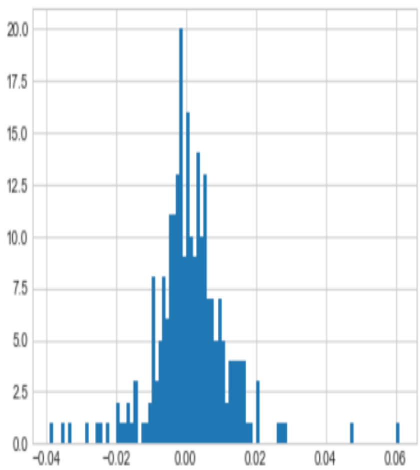
- Explore opportunities for expansion into related areas, such as other financial markets or investment products.

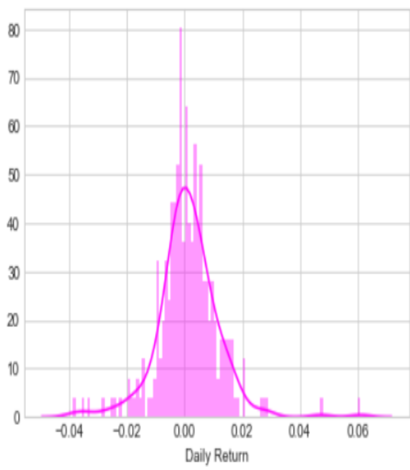
A successful business model for stock price prediction combines advanced data analytics, technology, and user-centric design to deliver actionable insights that empower investors to navigate the complexities of the financial markets more effectively. It requires continuous innovation, adaptability, and a commitment to providing real value to users.

4.Financial Modelling:

4.1 Daily Return Analysis: Now, that we've done some baseline analysis, let's go ahead and dive a little deeper. We're now going to analyze the risk of the stock.In order to do so, we need to take a closer look at the daily changes of the stock, and not just its absolute value. Let's go ahead and use pandas to retrieve the daily returns for the APPL stock.

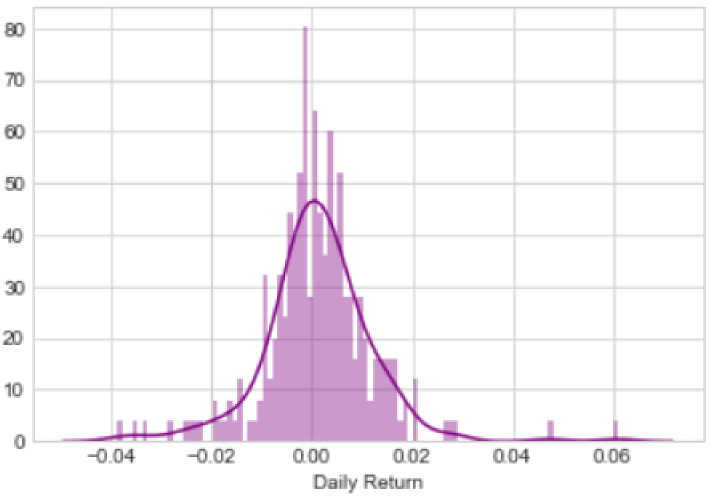






Risk Analysis:

Let's go ahead and define a value at risk parameter for our stocks. We can treat value at risk as the amount of money we could expect to lose (aka putting at risk) for a given confidence interval. There's several methods we can use for estimating a value at risk. Let's go ahead and see some of them in action.



Value at Risk using the Monte Carlo method:

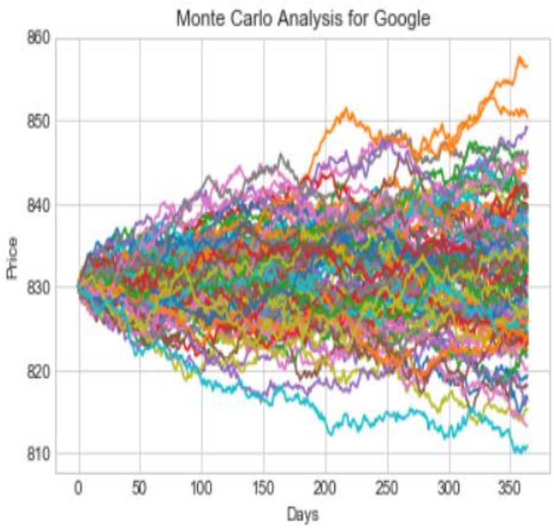
Using the Monte Carlo to run many trials with random market conditions, then we'll calculate portfolio losses for each trial. After this, we'll use the aggregation of all these simulations to establish how risky the stock is.

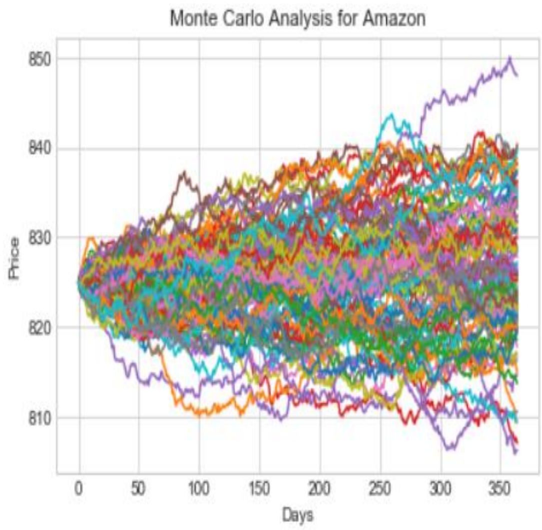
Let's start with a brief explanation of what we're going to do:

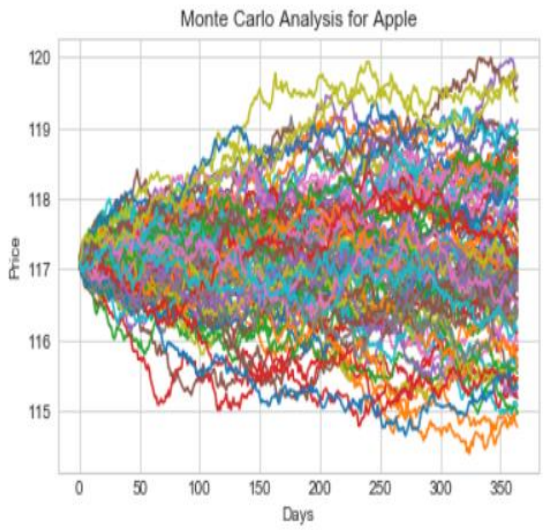
We will use the geometric Brownian motion (GBM), which is technically known as a Markov process. This means that the stock price follows a random walk and is consistent with (at the very least) the weak form of the efficient market hypothesis (EMH): past price information is already incorporated and the next price movement is "conditionally independent" of past price movements.

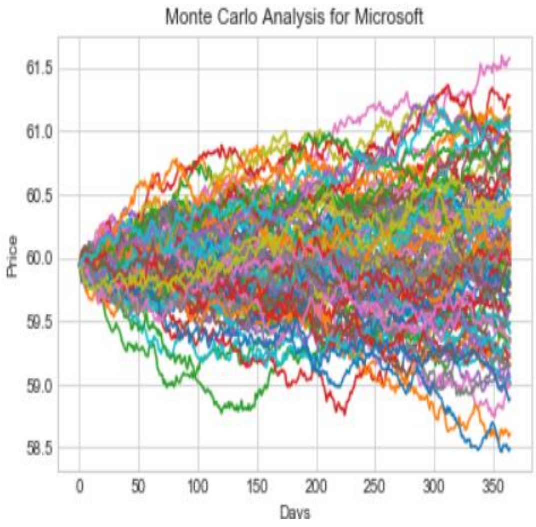
This means that the past information on the price of a stock is independent of where the stock price will be in the future, basically meaning, you can't perfectly predict the future solely based on the previous price of a stock.

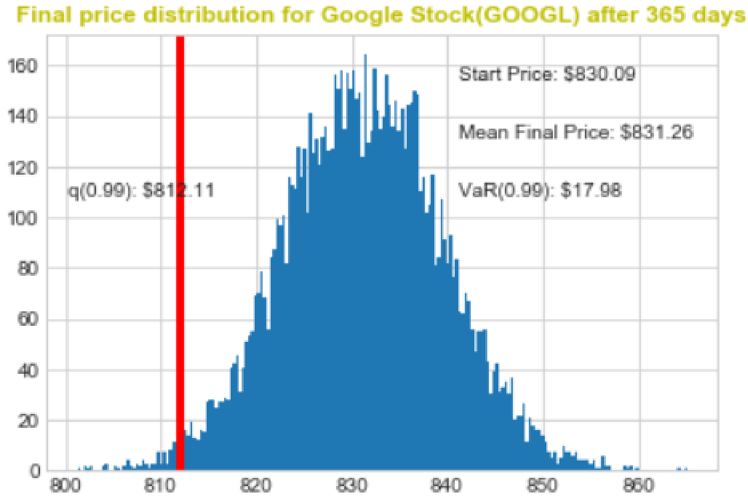
Now we see that the change in the stock price is the current stock price multiplied by two terms. The first term is known as "drift", which is the average daily return multiplied by the change of time. The second term is known as "shock", for each time period the stock will "drift" and then experience a "shock" which will randomly push the stock price up or down. By simulating this series of steps of drift and shock thousands of times, we can begin to do a simulation of where we might expect the stock price to be.











Awesome! Now we have looked at the 1% empirical quantile of the final price distribution to estimate the Value at Risk for the Google Stock(GOOGL), which looks to be $17.98 for every investment of 830.09 (The price of one initial Google Stock).

This basically means for every initial GOOGL stock you purchase you're putting about $17.98 at risk 99% of the time from our Monte Carlo Simulation.

This basically means for every initial AMZN stock you purchase you're putting about $18.13 at risk 99% of the time from our Monte Carlo Simulation.

Github link:Link