Introduction

This project is designed to assist new users in navigating the complexities of the stock market by providing them with intuitive tools and data-driven insights. With features like real-time stock tracking, detailed company information, and advanced prediction models, it simplifies the decision-making process for novice investors. By integrating machine learning and deep learning techniques the platform offers accurate forecasts for stock prices, helping users make informed investment decisions. The learning section and interactive chatbot also ensure that beginners can quickly grasp essential market concepts, making the stock market more accessible and less intimidating.

Problem Statement

The stock market is a cornerstone of the global economy. However, its unpredictability often intimidates novice investors. How can we leverage technology and mathematics to make smarter, data-driven investment decisions that enhance predictability and promote financial sustainability?

Solution

PredictHub is a comprehensive stock market analysis platform designed to simplify investing by leveraging advanced predictive models. It addresses the challenges of market unpredictability and investor inexperience by integrating key features such as:

- **Real-Time Insights**: A live stock ticker tape, detailed company metrics (Open, Close, High prices), and candlestick charts for in-depth analysis.
- **AI-Powered Predictions**: Advanced Deep learning and Machine learning models (LSTM, Bi-LSTM, ARIMA, and Random Forest) provide next-day stock price forecasts, enhancing decision-making with actionable insights.
- **User-Friendly Tools**: Search suggestions, a customizable watchlist, and a pre-prompted chatbot make navigation seamless, even for novice investors.
- **Learning and Support**: A dedicated learning section simplifies financial concepts, while personalized profiles, a support page, and feedback forms ensure user satisfaction.

Webapp URL: https://predicthub.vercel.app

GitHub Repository URL: https://github.com/deepraj21/PredictHub

Webapp QRCode

Scan it to go to Webapp

Mathematical computational Models

In PredictHub, we employ a combination of advanced machine learning and statistical models—LSTM (Long Short-Term Memory), Bi-LSTM (Bidirectional LSTM), ARIMA (AutoRegressive Integrated Moving Average), and Random Forest—to predict future stock prices with high accuracy.

1. **Long Short-Term Memory (LSTM)**
   LSTM is designed to capture temporal dependencies in sequential data, making it well-suited for stock price forecasting. Given a sequence of historical stock prices (x_1, x_2, ..., x_t), LSTM processes the sequence using its gated structure:

   - <u>Forget Gate</u>: $f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$, determining which information to discard.
   - <u>Input Gate</u>: $i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$, determining what new information to add.
   - <u>Cell State Update</u>: $\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$, where $C_t = f_t \cdot C_{t-1} + i_t \cdot C_t$
   - <u>Output Gate</u>: $o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$, producing the final output: $h_t = o_t \cdot \tanh(C_t)$.

   By training the LSTM on historical data, we obtain accurate predictions for the next time steps, incorporating real-time updates for dynamic adjustment.

2. **Bidirectional LSTM (Bi-LSTM)**
   Bi-LSTM extends LSTM by processing the sequence in both forward and backward directions, enabling the model to leverage future context along with past data:
   - <u>Forward pass</u>: Processes the input as $h_t^{forward}$.
   - <u>Backward pass</u>: Processes the reversed sequence as $h_t^{backward}$.

   The final output is $h_t = \text{Concat}(h_t^{forward}, h_t^{backward})$. This bidirectional structure improves pattern recognition in volatile markets by considering both preceding and succeeding trends.

3. **ARIMA (AutoRegressive Integrated Moving Average)**
   ARIMA is a statistical model suited for time series forecasting. It consists of three components:
   - <u>AutoRegression</u> (AR): Uses lagged values to predict the current value, $AR(p): y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t$.
   - <u>Integration</u> (I): Differencing to make the series stationary, $y_t' = y_t - y_{t-1}$.
   - <u>Moving Average</u> (MA): Models the error terms as linear combinations of past forecast errors, $MA(q): \epsilon_t = \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$.

   The combined ARIMA model is $y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q}$, where p, d, q are optimized using AIC (Akaike Information Criterion). ARIMA is particularly effective for capturing linear trends.

4. **Random Forest**
   Random Forest is an ensemble learning method that builds multiple decision trees from subsets of the data and averages their predictions. For stock price forecasting:
   - Input features include historical prices, technical indicators, and market sentiments.
   - Each decision tree makes a prediction, $y_i = f_i(x)$, where f_i is the ith tree.
   - Final prediction: $\hat{y} = 1/n \sum_{i=1} y_i$.

   Random Forest is robust against overfitting and can handle non-linear relationships in stock data.

Tech Stacks

Backend and Model

- Python
- tensorflow, keras, scikit-learn (python libraries)
- flask (python framework)

Frontend

- React.js
- Tailwind CSS

Database

- mongoDB

APIs

- yahoo finance
- Google Gemini

Development and Deployment

- VSCode
- Git & GitHub
- Vercel & Render

Key features of PredictHub

- Authentication System
- Realtime Stock ticker tape
- Search any ticker with suggestions
- Shows Open, Close, High prices, Company's Details of searched ticker
- Adding/Removing ticker functionality into Watchlist
- Pre-prompted chatbot for Stock Market
- Candle Stick Chart for searched ticker
- Description about used Stock Prediction Methods
- 4 Stock Prediction DL/ML Models for next day price prediction
  - LSTM (Long Short Term Memory)
  - BiLSTM (Bidirectional Long Short Term Memory)
  - Random Forest
  - ARIMA (AutoRegressive Integrated Moving Average)
- Learning Section for novice users
- Sharable and personalized Profile page
- Support page for resolving user query
- Contact and Feedback form connected to the mongoDB
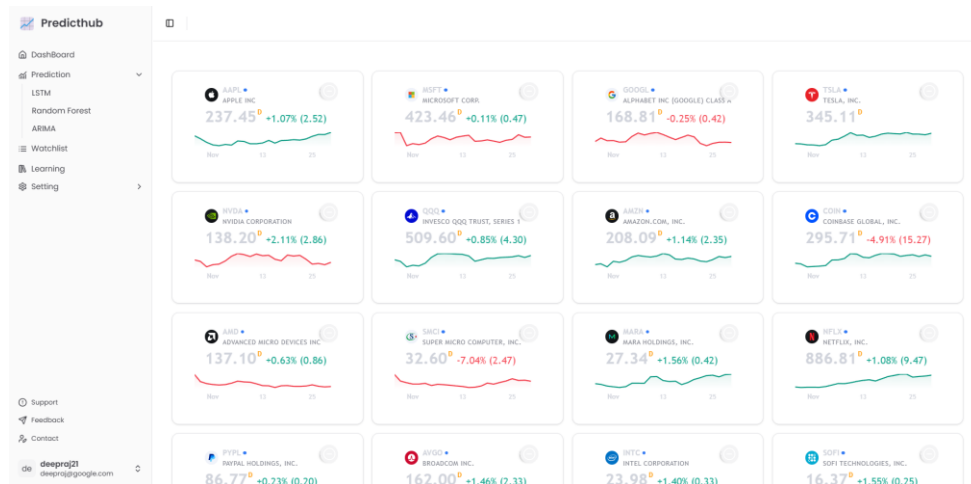- Supports Dark Mode

Applications

1. Personalized Investment Guidance
   Helps novice investors make informed decisions by providing real-time market insights, historical data, and predictive analytics tailored to their portfolios.

2. Portfolio Management
   Enables users to create and manage personalized watchlists, track their preferred stocks, and analyse market trends for better portfolio optimization.

3. Financial Advisory Support
   Assists financial advisors by offering data-driven forecasts and visualizations, allowing them to provide well-informed recommendations to their clients.

4. Educational Tool
   Serves as a learning platform for students, educators, and novice traders, providing tutorials, resources, and interactive features to understand stock market fundamentals and advanced analytics.
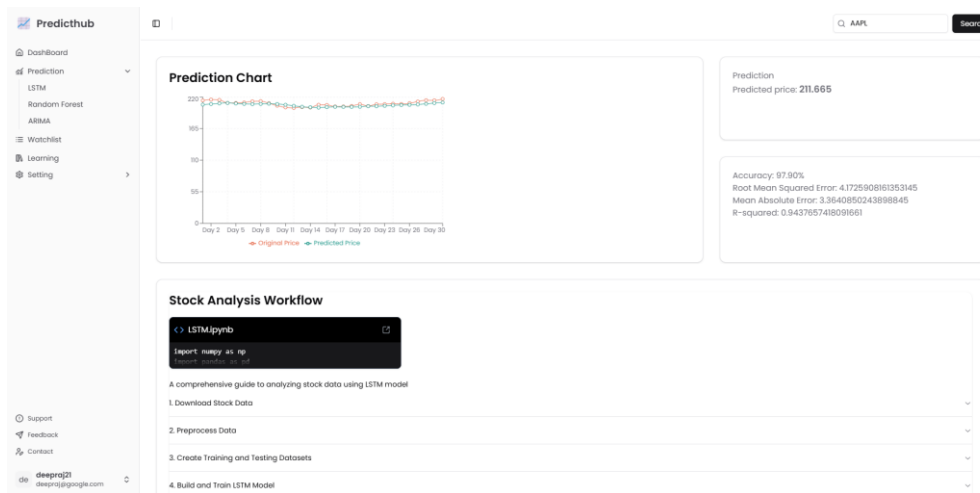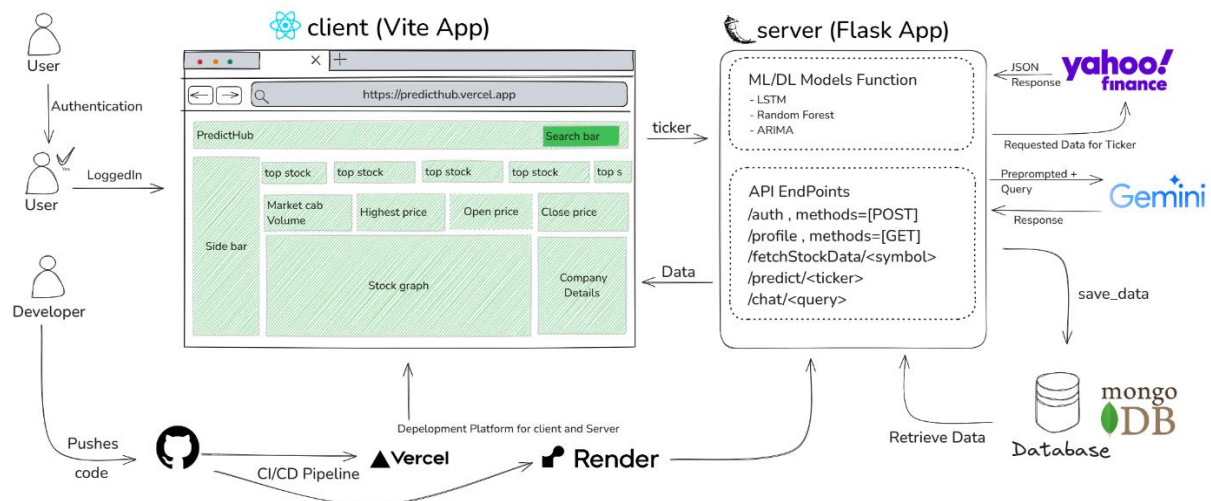
Webapp Images and Useful Links

Dashboard

Watchlist

LSTM Prediction

Architecture Diagram

References

https://colah.github.io/posts/2015-08-Understanding-LSTMs/

https://python-yahoofinance.readthedocs.io/en/latest/api.html

https://ui.shadcn.com/docs

https://developers.google.com/machine-learning/crash-course/neural-networks

https://www.investopedia.com/terms/r/regression.asp

https://www.investopedia.com/terms/t/technicalanalysis.asp

https://flask.palletsprojects.com/en/stable/