Stock Predict AI — Technical Documentation

PAMBALA SAHASRA –MALLA REDDY ENGINEERING COLLEGE FOR WOMEN

1) Product Overview

StockPredict AI is a machine learning–powered stock analysis and prediction platform that provides:

- **Financial Data Acquisition** (real-time & historical stock data via APIs like Yahoo Finance).
- **AI Predictions** (forecasting future stock prices, risk analysis, and insights).
- **Visualization** (interactive dashboards, charts, and analysis tools).

It consists of:

- **Frontend:** React + Tailwind for modern UI.
- **Backend:** Node.js/Express for API orchestration.
- **Data Processing:** Python layer for stock data cleaning & feature engineering.
- **Database:** MySQL for structured data storage and caching.

---

2) Screens & User Flows

2.1 Role Selection

- Users can access modules like **Data Acquisition**, **Data Processing**, **AI Prediction**, and **Visualization** directly.
- Simple sidebar navigation with intuitive icons and labels.

2.2 Data Acquisition Module

- Connects to APIs like Yahoo Finance.

- Fetches stock market data for multiple companies (AAPL, GOOGL, MSFT, AMZN, TSLA, NVDA, META, NFLX).
- Customizable **start & end date** filters.
- Results show **latest price, total data points, and success confirmation**.

2.3 AI Prediction Center

- Users select one or more stock symbols.
- Prediction outputs: **future price, % change, risk score, and confidence score**.
- Export predictions in CSV/JSON format.

---

3) Security & Authentication

MySQL database with encrypted user credentials

- All user login details (like passwords) are stored in **encrypted form** using hashing algorithms (bcrypt/Argon2).
- This ensures that even if the database is compromised, the actual passwords are not visible in plaintext.

JWT-based authentication for secure sessions

- **JWT (JSON Web Tokens)** are used to verify and manage user logins.
- Each user gets a token after login, which is used for every request, making sessions secure and reducing risk of hijacking.

HTTPS encryption for data transport

- All communication between the platform and the user is done over **HTTPS**.
- This encrypts the data during transmission, preventing hackers from reading sensitive information (like login details or financial data).

Role-based access (Premium User, Analyst, Admin)

- Users are divided into roles with different permissions:
    - **Premium User:** Can access predictions and insights.
    - **Analyst:** Can perform detailed data analysis and export.
    - **Admin:** Full control, including user management and system monitoring.
- This ensures **data security** and **controlled access**.

---

4) Data Acquisition & Automation

4.1 Financial API Integration

- Supports Yahoo Finance, extendable to other APIs.
- Fetches OHLCV data (Open, High, Low, Close, Volume).

4.2 Automated Collection

- Scheduled jobs to fetch and cache data.
- Rate-limit handling to prevent API overload.

4.3 Caching

- **Raw cache** for API JSON responses.
- **Processed cache** (CSV/Parquet) with computed indicators.

---

5) Data Quality & Preprocessing

5.1 Missing Values

- Aligns to trading days, fills gaps with forward/backward fill.

5.2 Outliers

- Detects anomalies using z-scores and rolling MAD.

5.3 Adjustments

- Uses adjusted prices for dividends & stock splits.
- Avoids synthesizing trades on holidays.

---

6) Feature Engineering

StockPredict AI generates advanced features to improve prediction accuracy:

- **Trend Indicators**:
    - SMA (Simple Moving Average)
    - EMA (Exponential Moving Average)
    - MACD (Moving Average Convergence Divergence)
- **Volatility Indicators**:
    - Bollinger Bands
    - ATR (Average True Range)
- **Momentum Indicators**:
    - RSI (Relative Strength Index)
    - Stochastic Oscillator
- **Volume-based Indicators**:
    - OBV (On-Balance Volume)
    - VWAP (Volume Weighted Average Price)
- **Custom Features**:
    - Rolling averages & variance
    - Lag features for time-series forecasting
    - Market sentiment integration (future scope)

7) Multi-Company Selection & Batch Processing

Parallel Processing: Analyzes multiple stock tickers simultaneously.

Batch Mode: Processes company arrays with a single query.

Unified Dashboard: Comparison charts for multiple companies.

Aggregated Output:

- JSON object (all symbols in one response).
- Zipped CSV bundle (individual company files).

Scalability: Handles up to 50+ companies in one request.

Custom Filters: Apply same date range & intervals across companies.

Error Handling: Skips failed symbols but continues with others.

8) Forecasting Models

- **ARIMA Model**: Captures trend, seasonality, and residual noise.
- **Machine Learning Models**: Random Forest, XGBoost (optional extensions).
- **Deep Learning Models** (Future Scope):
  - LSTM (Long Short-Term Memory Networks) for sequential data.
  - Prophet (by Facebook) for robust trend/seasonality modeling.
- **Hybrid Predictions**: Combines technical indicators + time-series data.
- **Evaluation Metrics**:
  - MAPE (Mean Absolute Percentage Error)
  - RMSE (Root Mean Square Error)
  - Confidence Interval ranges
- **Visualization of Forecasts**:
  - Future trend lines with shaded risk/confidence regions.

9) Export & Import (CSV / JSON)

Export Options:

- Historical Analysis Export: OHLCV data + indicators.
- Prediction Export: Forecasted prices + risk metrics.
- Multi-Company Export: Bundled in a **.zip** file.

Import Options:

- Accepts CSV or JSON user uploads.
- Validates schema (date, open, high, low, close, volume).
- Auto-detects missing indicators and recalculates.
- De-duplication of records for consistency.

System Checks:

- File size validation (non-empty files).
- Row count checks ($\geq$ indicator window).
- SHA-256 hash manifest for data integrity.
- Success notifications and error logs.

10) Deliverables Present in Build

Real-time Data Acquisition with API integration

- The platform connects to financial APIs (like Yahoo Finance) to fetch **real-time and historical stock data**.
- Supports multiple companies (AAPL, GOOGL, TSLA, etc.) and custom date ranges.
- Automated scheduling ensures fresh data without manual effort.

AI Predictions with price forecasting

- Uses **time-series models (ARIMA, ML algorithms)** to predict future stock movements.
- Provides confidence scores, risk metrics, and % change.
- Helps users make **data-driven investment decisions**.

Visualization dashboards for stock insights

- Interactive charts and graphs (candlesticks, SMA/EMA, RSI, Bollinger Bands).
- Side-by-side comparison of multiple companies.
- Easy-to-read panels for **trends, volatility, and predictions**.

Secure authentication (JWT + MySQL)

- MySQL database stores **encrypted user credentials**.
- **JWT tokens** ensure only verified users access features.
- Role-based access (User, Analyst, Admin) for controlled privileges.

Caching system for performance

- Stores previously fetched and processed stock data locally.
- Reduces repeated API calls → **faster response times**.
- Ensures smooth performance even with large datasets.

CSV/JSON Export functionality

- Users can export historical data and predictions in **CSV or JSON formats**.
- Multi-company exports bundled into a **zip file with manifest**.
- Enables offline analysis, backtesting, and report generation.