AIStock Pro — Technical Documentation

- DONE BY **K HARI PRASAD , IIT PATNA**

1) Product Overview

AIStock Pro is an AI-powered trading workspace that delivers:

- **Past Analysis** (historical intelligence, technical studies, pattern discovery)
- **Future Prediction** (ML-based forecasts, confidence, risk metrics)

It consists of a **React + Tailwind** frontend, a **Python processing layer** for data engineering & indicators, a **Node/Express API** (or similar) for orchestration, and **MongoDB** for secure, encrypted authentication and session management.

---

2) Screens & User Flows

2.1 Role Selection

- Two premium options: **Past Analysis** and **Future Prediction**.
- One-click navigation to `/dashboard` (Past) or `/prediction` (Future).
- Professional, large “choice” buttons; no emojis; gradient + elevation.

2.2 Trading Dashboard — Past Analysis

- **RADAPT Reprocessing System**: Recognition → Assimilation → Decision → Action → **PAST** → Transfer.
- **Past Data Analysis**: Historical patterns, seasonal trends, cyclical analysis, performance metrics, volatility cycles.
- **Stock Data Processing (Python)**: Cleans raw OHLCV, standardizes time index, computes indicators.
- **Stock Market Analysis (JavaScript)**: Interactive charts (price + SMA/EMA/BB), RSI panel, tooltips, legends.

- **System Status**: pipeline state, cache counts, processed symbols, success rate, Python integration status.

- **Analysis Modal**: Company card, overall stance (BUY/HOLD/SELL), **Sharpe** & **Max Drawdown**, chart overlays, **CSV/JSON export** buttons.

2.3 AI Stock Prediction Center — Future Prediction

- **Multi-Company Selection** (e.g., up to 5 symbols) with chips.

- **Target Date** picker and **Generate Predictions** action.

- **Prediction Summary** & **Detailed Analysis** (Predicted Price, % Change, Risk, Sentiment, RSI, confidence score).

- **Export as CSV** for predicted sets.

---

3) Security & Authentication (MongoDB, Encrypted)

- **User Store**: `users` collection in MongoDB with **hashed passwords** (e.g., bcrypt/Argon2).

- **Transport Security**: HTTPS; **JWT** access tokens + optional refresh tokens; short TTL for access, longer for refresh.

- **Data at Rest**:

    - Hashed credentials (never stored in plaintext).

    - Optional **MongoDB Client-Side Field Level Encryption (CSFLE)** for sensitive fields (e.g., email).

- **RBAC**: roles such as `user`, `analyst`, `admin` (optional).

- **Brute Force Protection**: login throttling and IP-based rate limits.

- **Secrets**: `.env` for API keys (Alpha Vantage/Yahoo/Polygon), JWT secrets, and Mongo connection strings.

---

4) Data Acquisition & Automation

4.1 Financial API Integration

- Providers: Alpha Vantage / Yahoo Finance / Polygon.io (pluggable).
- Endpoints cover **OHLCV** (daily/weekly/intraday) and optional fundamentals.
- **Batch requests** by **symbol list** and **date range**; interval selectable (Daily default).

4.2 Automated Collection

- API client supports **symbol arrays**: `["AAPL","MSFT","GOOGL","AMZN", ...]`.
- **Job Orchestrator** handles:
    1. schedule or user-triggered fetch
    2. per-symbol pull with rate-limit sleep/backoff
    3. normalization → cache → indicator computation
    4. persistence + export (if requested)

4.3 Caching

- **Raw cache**: provider JSON cached by `{provider}/{symbol}/{interval}/{date-range}` with TTL & versioning.
- **Processed cache**: parquet/CSV per symbol with computed indicators for speed.
- Cache keys include **indicator version** so changes invalidate stale files.

---

5) Data Quality & Preprocessing

5.1 Missing Values

- Align to **business-day index**; reindex series to fill non-trading gaps as `NaN`.

- **Forward-fill/Backward-fill** for non-critical fields; never fabricate OHLC—close is primary; volume kept `NaN` unless verified.
- Drop leading rows until the first valid window for indicators (e.g., first 20 rows before SMA-20).

5.2 Outliers

- Compute daily **returns**; mark outliers via |z-score| > 4 or **rolling MAD**; **winsorize** tails instead of hard deletion.
- Log any capped values for audit.

5.3 Corporate Actions / Holidays

- Use provider-adjusted prices where possible (split/dividend adjusted).
- Holiday calendar via **business day frequency**; never synthesize trades on closed days.

---

6) Feature Engineering (Derived Indicators)

- **6.1 Simple Moving Average (SMA)**

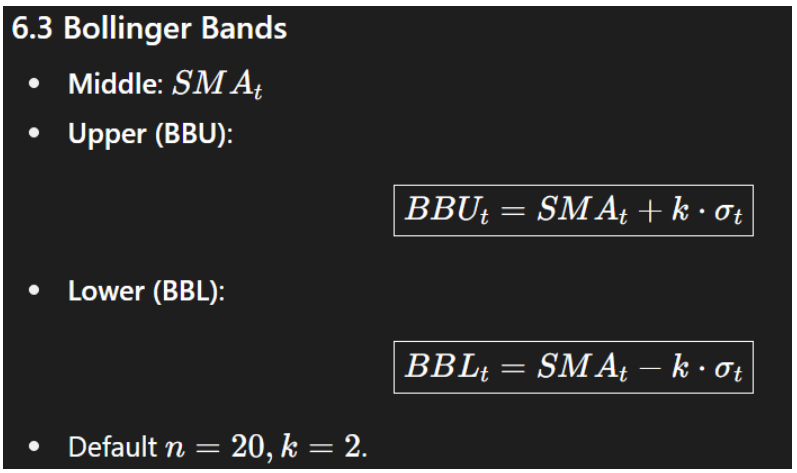
$$SMA_t = \frac{1}{n}\sum_{i=0}^{n-1} P_{t-i}$$

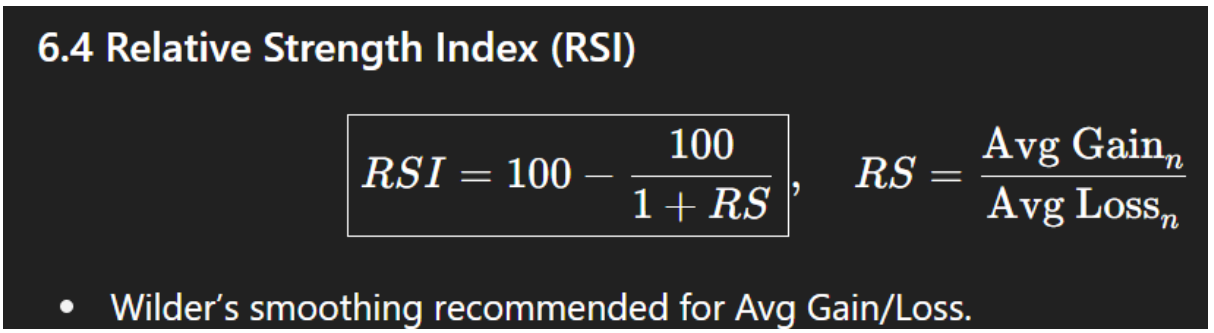
  - Trend smoother; e.g., **SMA-20, SMA-50.**

- **6.2 Exponential Moving Average (EMA)**

$$EMA_t = \alpha P_t + (1-\alpha)EMA_{t-1}, \quad \alpha = \frac{2}{n+1}$$

  - More reactive than SMA; e.g., **EMA-12, EMA-26.**

- 



---

7) Multi-Company Selection & Batch Processing

- UI supports selecting **multiple companies (e.g., up to 5)**.
- Backend accepts `symbols[]` and common `date_range`.
- Jobs run **sequential with rate-limit awareness** or **concurrently** with bounded pool size.
- Results aggregated per symbol and returned as:
    - **Combined JSON** (array of symbol objects), and/or
    - **Zipped CSVs** (one CSV per symbol) for download.

---

8) Forecasting with ARIMA Model

- Uses **ARIMA (AutoRegressive Integrated Moving Average)** for stock price forecasting.
- Captures **trend + seasonality + noise** in time series.
- Helps predict **future stock movements** based on historical data

ARIMA formula:

$$ARIMA(p,d,q) \quad \Rightarrow \quad \phi(B)(1-B)^d X_t = \theta(B)\epsilon_t$$

Where:

- p = AR order
- d = Differencing order
- q = MA order
- $\phi(B), \theta(B)$ = lag polynomials
- ϵ_t = white noise

8) Export & Import (CSV / JSON) + System Checks

8.1 Export

- **From Dashboard (Past)**: export historical + indicators per symbol.
- **From Prediction Center (Future)**: export predictions with confidence & risk.
- **Formats**:
  - **CSV** columns (example): `date, open, high, low, close, volume, sma20, ema12, bb_upper, bb_lower, rsi14, ...`
  - **JSON**: per-symbol object with `meta`, `series[]`.
- **Batch Export** for multiple symbols creates a **zip bundle**:
  - `AAPL_past_2024-01-01_2025-08-29.csv`
  - `MSFT_past_2024-01-01_2025-08-29.csv`
  - `manifest.json` (symbols, date range, indicator versions, row counts).

8.2 System Checks (Post-Export)

- Verify **file exists** & **non-zero size**.
- Validate **row count ≥ indicator window** (e.g., ≥ 20 for SMA-20).
- Write a **manifest** with SHA-256 hashes for integrity.
- Return a success toast and manifest summary to UI.

8.3 Import

- Accepts **CSV or JSON** generated by AIStock Pro or provider-compatible format.
- On import:
    1. Schema validation (required columns + date type).
    2. De-duplication on `(symbol, date)`.
    3. Optional re-compute of indicators if columns missing.
    4. Store in **processed cache**; mark as “user-imported”.
- Import is available for **backtesting**, **replays**, or **offline runs**.

---

9) RADAPT Analysis — Concept & Mapping

RADAPT is the platform’s operational framework for turning market data into decisions:

1. **Recognition** — Detect patterns (trend direction via SMA/EMA; anomalies; volume spikes).
2. **Assimilation** — Fuse signals (technical + sentiment, risk consolidation).
3. **Decision** — Score signals → **Signal Generation**, **Risk Assessment**, **Timing Optimization**.
4. **Action** — Translate decisions into actions (position sizing, stop-loss, take-profit).
5. **PAST** — Historical evaluations: **seasonality**, **cyclical analysis**, **volatility cycles**, **performance metrics**.

6. **Transfer** — Feed learnings back (strategy adaptation, model updates, knowledge reuse).

UI Mapping (from your cards):

- The six RADAPT boxes under the dashboard summarize which sub-modules completed (green ticks) and the **tags** underneath reflect the techniques enabled in each stage.

Deliverables Present in Your Build

- **Encrypted Authentication with MongoDB** (hashed credentials, JWT).
- **Multi-Company Selection** with batched processing.
- **Automated API fetch** by symbol & date range (rate-limit aware).
- **Caching** (raw + processed datasets).
- **CSV/JSON Export** (single or zip, with manifest).
- **CSV/JSON Import** (schema validation, de-dup, optional recompute).
- **System Checks** after export (file existence, size, row count, hashes).
- **Professional UI** for Past & Future modules, with indicator overlays.