Stock Analysis PC - User Documentation

**(Version 1.0 - April 9, 2025)**

# 1. Introduction & Important Disclaimer

Welcome to the Stock Analysis PC application! This tool is designed to provide historical stock price visualization, technical indicator analysis, and an *experimental* future price trend prediction using an LSTM model incorporating news sentiment analysis.

**IMPORTANT DISCLAIMER:**

- **NOT FINANCIAL ADVICE:** This application is provided for informational and educational purposes ONLY. It does NOT constitute financial, investment, trading, or any other form of advice or recommendation.

- **HIGH UNCERTAINTY:** Predicting stock market movements is extremely difficult and inherently uncertain. Past performance is not indicative of future results.

- **REFERENCE ONLY:** Use the information and predictions generated by this tool solely as a reference point for your own research and analysis. **DO NOT** base any financial decisions solely on the output of this application.

- **RISK:** All investments involve risk, including the potential loss of principal. You are solely responsible for your own investment decisions. Consult with a qualified financial advisor before making any investment choices.

# 2. Getting Started: How to Use the Application

1. **Launch:** Run the StockAnalysisPC.exe file.

2. **Enter Stock Symbol:** You will be prompted to enter a stock symbol (e.g., AAPL for Apple, MSFT for Microsoft, 1155.KL for specific KLSE stocks). The application will attempt to normalize the input (e.g., convert 1155 KL to 1155.KL).

3. **Select Date Range:** Choose the historical period you want to *display* on the plot.

   o **Relative:** Select a unit (Days, Months, Years) and enter a number to view data relative to today.

   o **Detailed:** Select specific Start and End dates using the calendars.

4.  **Select Training Period:** Enter the number of years of historical data (ending today) that the LSTM prediction model should be trained on (e.g., 3 years). More data provides more context but takes longer to process.

5.  **Processing:** The application will fetch historical price data, fetch and analyze news sentiment (if enabled), compute technical indicators, train the LSTM model, and generate predictions. Loading indicators will appear during these steps.

6.  **View Plot:** The main plot window will appear, showing:

    o  Historical closing prices (blue line).

    o  Predicted future closing prices (red dashed line).

    o  An **initial popup** will display key information and disclaimers (shown only once per application run).

7.  **Interact with Plot:**

    o  **Hover:** Move your mouse over the plot lines to see the specific Date and Price in a tooltip.

    o  **Zoom:** Use your mouse scroll wheel while the cursor is over the plot to zoom in or out.

    o  **Toggle Lines:** Click on items in the legend (top-right) to hide or show the corresponding plot line (Historical/Predicted).

8.  **Use Control Buttons (Bottom):**

    o  **Reselect Symbol:** Start the process over with a new stock symbol.

    o  **Reset View:** Return the plot zoom/pan to its original state.

    o  **Save Plot:** Save the current view of the plot as an image file (JPG).

    o  **Show News URLs:** (If sentiment analysis ran) Opens a window listing the URLs of the news articles that were successfully processed for sentiment scoring.

    o  **Documentation:** Opens this documentation file (documentation.pdf).

    o  **Feedback:** Opens a Google Form in your web browser for providing feedback about the application.

# 3. Technical Details: Model & Data

## 3.1 Data Sources

- **Stock Prices:** Historical stock data (Open, High, Low, Close, Volume) is sourced from **Yahoo Finance** (yfinance library). Data availability is dependent on Yahoo Finance.

- **News URLs:** Potential news article URLs related to the company are sourced via the **GDELT Project** database (gdeltdoc library).

## 3.2 Technical Indicators Used as Features

The LSTM model is trained using the following technical indicators, calculated from the historical stock data:

- **Close Price** $(Close_t)$

  - **Purpose:** The primary value being predicted and a key input feature.

  - **Formula:** The closing price of the stock on day t.

- **20-Day Exponential Moving Average** $(EM\ A20_t)$

  - **Purpose:** Identifies short-to-medium term trend direction by smoothing price data, giving more weight to recent prices.

  - **Formula:**
    $$EM\ A20_t = Close_t * k + EM\ A20_{t-1} * (1 - k)$$
    where $k = \dfrac{2}{span + 1}$ and $span = 20$.

- **Moving Average Convergence Divergence** $(MACD_t)$

  - **Purpose:** Shows the relationship between two EMAs of price, indicating momentum changes.

  - **Formula:**
    $$EMA12_t = EMA(Close, span = 12)$$
    $$EMA26_t = EMA(Close, span = 26)$$
    $$MACD_t = EMA12_t - EMA26_t$$
    *(Note: The MACD Signal Line (EMA of MACD) is not directly used as a feature in this version).*

- **Relative Strength Index (RSI) and related features** $(RSI_{MA_t},\ RSI_{Slope})$

- **Purpose:** Momentum oscillator measuring the speed and change of price movements (typically indicates overbought/oversold conditions, though used here as a momentum input).

- **Formula (RSI):**

    1. $Change_t = Close_t - Close_{t-1}$

    2. $Gain_t = Change_t \; if \; Change_t > 0 \; else \; 0$

    3. $Loss_t = -Change_t \; if \; Change\_t < 0 \; else \; 0$

    4. $AvgGain_t = EMA(Gain, period = 14) \; (Using \; Wilder's \; smoothing \; / \; EMA \; with \; com = period - 1)$

    5. $AvgLoss_t = EMA(Loss, period = 14)$

    6. $RS_t = \frac{AvgGain_t}{AvgLoss\_t} \quad (Handle \; AvgLoss\_t = 0 \; case)$

    7. $RSI_t = 100 - \frac{100}{1+RS_t}$

- **RSI Moving Average $(RSI_{MA_t})$:**
    $$RSI\_MA_t = SMA(RSI, period = 9) \; (Simple \; Moving \; Average \; of \; RSI \; over \; 9 \; days).$$

- **RSI Slope $(RSI_{Slope_t})$:**
    $$RSI\_Slope_t = RSI_{MA_t} - RSI_{MA_{t-1}}$$

- **On Balance Volume $(OBV_t)$**

  - **Purpose:** Relates price and volume, indicating buying/selling pressure.

  - **Formula (Conceptual):**

    - $If \; Close_t > Close_{t-1} : OBV_t = OBV_{t-1} + Volume_t$

    - $If \; Close_t < Close_{t-1} : OBV_t = OBV_{t-1} - Volume_t$

    - $If \; Close_t = Close_{t-1} : OBV_t = OBV_{t-1}$
      (Note: The calculated OBV is then scaled using MinMaxScaler before being fed to the model).

- **Average True Range $(ATR_t)$**

  - **Purpose:** Measures market volatility by considering the range between high/low prices and the previous close.

- o **Formula:**

  1. $TR_t = max[(High_t - Low_t), abs(High_t - Close_{t-1}), abs(Low\_t - Close_{t-1})]$ $(True\ Range)$

  2. $ATR_t = EMA(TR, span = 14)$

- **Standard Deviation** $(StdDev_t)$

  - o **Purpose:** Measures price volatility based on the standard deviation of closing prices over a period.

  - o **Formula:** $StdDev_t = StandardDeviation(Close, period = 20)$

- **Bollinger Band Width** $(BBWidth_t)$

  - o **Purpose:** Measures volatility relative to the current price level, based on Bollinger Bands.

  - o **Formula:**

    1. $UpperBB_t = EMA20_t + (2 * StdDev_t)$

    2. $LowerBB_t = EMA20_t - (2 * StdDev_t)$

    3. $BBWidth_t = \frac{(UpperBB_t - LowerBB_t)}{EMA20_t}$ $(Normalized\ width)$

- **Weighting Applied**

  - o Each computed indicator is multiplied by a corresponding weight to adjust its influence:

    1. **Close: 0.30**

    2. **EMA20: 0.20**

    3. **MACD: 0.15**

    4. **RSI_MA: 0.25**

    5. **OBV: 0.05**

    6. **ATR: 0.05**

  - o This weighted set forms the feature vector that the LSTM model uses as input.

## 3.3 Sentiment Analysis Process

1. **News Discovery (GDELT):** Potential relevant news articles URLs are sourced from the GDELT Project database using the company name as a keyword (searching roughly the last 60 days). GDELT aggregates news from global sources.

2. **Article Fetching & Parsing (newspaper3k):** The application attempts to download the HTML content for each URL found and parse out the main article text using the newspaper3k library.

   o **Potential Failures:** This step frequently encounters issues. Websites may block automated access (scraping), require JavaScript, have unusual layouts, or timeouts may occur. The library attempts to handle these, but many articles might fail to download or parse correctly, resulting in them being skipped.

3. **Sentiment Scoring (VADER):** For articles where text is successfully extracted and meets a minimum length requirement (100 characters), the VADER (Valence Aware Dictionary and Sentiment Reasoner) tool analyzes the text.

   o **Output:** VADER provides a 'compound' score ranging from -1.0 (most negative) to +1.0 (most positive), representing the overall sentiment intensity of the text. A score near 0 is considered neutral.

4. **Aggregation:** Sentiment scores are grouped by their publication date (if available from GDELT), and the average compound score for each day is calculated.

5. **Limitation:** The final sentiment data used by the model may be based on a small number of successfully processed articles due to the challenges in step 2.

## 3.4 LSTM Prediction Model

- **Type:** A Long Short-Term Memory (LSTM) recurrent neural network, suitable for learning patterns in time-series data.

- **Implementation:** Built using TensorFlow and Keras libraries.

- **Input:** Takes sequences of the past 60 days' worth of the calculated technical indicators (listed in section 3.2) as input. All input features are scaled between 0 and 1 using MinMaxScaler.

- **Output:** Predicts the scaled closing price for the *next* day.

- **Training:** The model is trained from scratch each time you run an analysis, using the historical data length specified (e.g., 3 years).

- **Prediction Process:** To predict multiple days into the future, the model predicts one day, then uses that prediction as part of the input to predict the next day, and so on (autoregressive prediction).

## 3.5 Prediction Adjustments

After the LSTM model generates raw price predictions, two adjustments are applied:

1. **Level Shift:** The entire predicted sequence is shifted up or down so that the first predicted price aligns smoothly with the last *actual* historical closing price, avoiding an unnatural jump.

2. **Sentiment Adjustment (Optional):** If sentiment data was successfully gathered, a simple linear regression forecast of recent sentiment trends is made. This forecasted sentiment is used to slightly nudge the price prediction up (for positive sentiment) or down (for negative sentiment) using a small weighting factor. The impact of this adjustment is deliberately kept minor.

# 4. Accuracy, Considerations & Limitations

- **Directional Accuracy:** Based on internal testing against historical data, the model correctly predicts the general *direction* (up vs. down trend) over the next ~10 days about **60%** of the time.

- **Confidence Interval:** The 95% confidence interval for this directional accuracy is wide, roughly **23.1% to 88.2%**. This statistical range highlights the significant uncertainty involved.

- **Price Prediction UNRELIABILITY:** While the model outputs specific price points, **predicting the exact future price of a stock with high accuracy is extremely difficult, if not impossible.** The predicted red line should be viewed only as a potential *trend indication* based on the model's inputs, **NOT as a precise price target.** Do not make trading decisions based on these specific price values.

- **Data Dependency:**

  o The application is entirely dependent on **Yahoo Finance** for stock price data and the **GDELT Project / newspaper3k** for news sentiment. Service outages, API changes, website blocking, or data inaccuracies from these sources will affect the results.

- Sentiment analysis quality is highly dependent on the number and relevance of articles successfully downloaded and parsed, which can be very low for some symbols or time periods.

- **Model Scope Limitations:** The model only considers historical price/volume patterns (via technical indicators) and recent news sentiment trends. It **DOES NOT** account for:

  - Fundamental company analysis (revenue, earnings, debt, management).

  - Future macroeconomic factors (interest rates, inflation, GDP growth, recessions).

  - Unforeseen company news or events (product launches, mergers, scandals, earnings surprises).

  - Market shocks or geopolitical events.

  - Analyst rating changes.

  - Changes in overall market sentiment or structure.

# 5. Feedback

We appreciate your feedback to help improve this tool! Please use the **"Feedback"** button located on the plot window controls. This will open a Google Form in your web browser where you can share your thoughts, report any bugs you encounter, or suggest new features. Thank you!