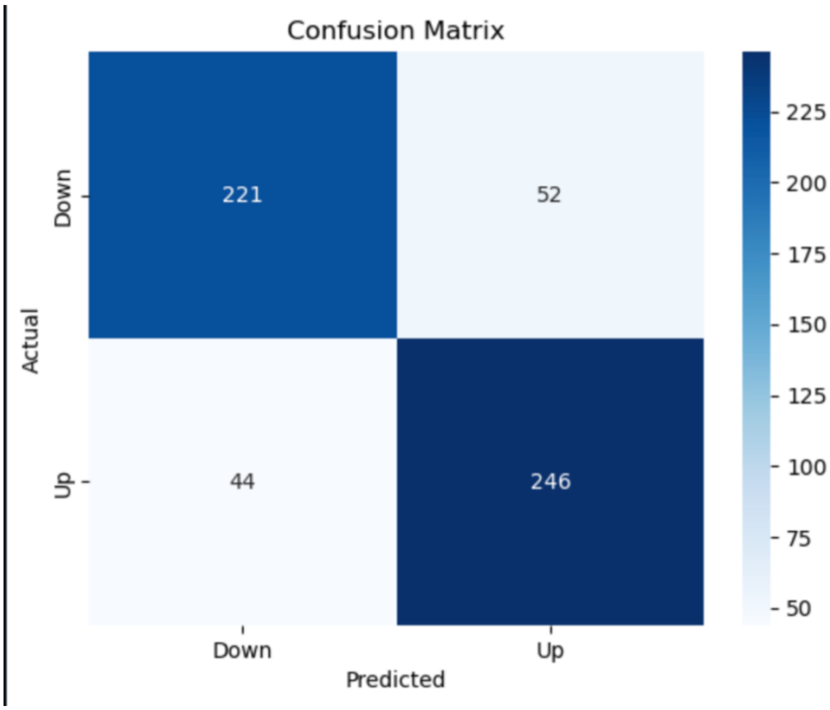
GA-XGBoost Model for Stock Price Movement Prediction

Steps in Model Implementation

1. **Data Preparation**:
    - Defined a prepare_data function to fetch historical stock data for TSLA.
    - Calculated and added various technical indicators as features, using the calculate_technical_indicators function. These indicators included momentum, volatility, volume, and cycle indicators, among others.
    - Defined the target variable as a binary label: 1 if the closing price went up compared to the previous day and 0 if it went down.
    - Result: Created a feature matrix X containing all technical indicators and a target vector y representing the stock price movement.
2. **Genetic Algorithm (GA) for Feature Selection**:
    - Implemented the GA to iteratively search for the optimal subset of features from the list of all technical indicators.
    - The GA used a fitness function, where it trained an XGBoost model on each feature subset and evaluated its accuracy as the fitness score.
    - Steps in GA included:
        - **Initialization**: Generated a random population of feature subsets.
        - **Fitness Evaluation**: Trained XGBoost models for each subset and recorded their accuracy.
        - **Selection**: Chose the top-performing individuals for reproduction.
        - **Crossover and Mutation**: Combined features from parents to create offspring, with occasional random mutations to introduce variability.
    - Result: GA identified an optimal feature subset that maximized the model's predictive accuracy.
3. **XGBoost Model Training**:
    - Defined a train_xgboost function to train an XGBoost model on the feature subset chosen by the GA.
    - Split the data into training and testing sets to evaluate model performance.
    - Used the optimal feature subset returned by the GA to train a final XGBoost model, measuring its performance on unseen data.
4. **Evaluation and Results**:
    - Generated a confusion matrix and classification report to evaluate the model.
    - Key metrics:
        - **Accuracy**: 0.83 (83%)
        - **Precision**: 0.83 for both classes ("Up" and "Down")
        - **Recall**: 0.81 for "Down" and 0.85 for "Up"
        - **F1-Score**: 0.82 for "Down" and 0.84 for "Up"
    - Confusion Matrix:



5. **Feature Subset Identified by GA**:
   - The GA identified the following optimal feature subset:

['STOCH_slowk', 'BB_middle', 'STOCH_slowd', 'BB_upper', 'HT_PHASOR_inphase', 'WMA', 'HT_TRENDMODE', 'MOM', 'MIDPOINT', 'CMO', 'HT_DCPERIOD', 'AROON_DOWN', 'WCLPRICE', 'MACD_hist', 'CCI', 'AROON_UP', 'BB_lower', 'TEMA', 'STOCH_fastk', 'ADXR', 'HT_DCPHASE', 'ADX', 'HT_PHASOR_quadrature', 'STOCH_fastd', 'MEDPRICE', 'MFI', 'MIDPRICE', 'TRANGE', 'PLUS_DI', 'SMA', 'T3']

Plan for Future Improvements

1. **Feature Importance Analysis**:
   - Use XGBoost's built-in feature importance metrics to analyze which features within the selected subset contribute most to the model's predictions.
   - This could help refine the feature subset further and potentially remove less important indicators.
2. **Hyperparameter Tuning**:
   - Experiment with hyperparameters for the XGBoost model (e.g., learning rate, max depth, n_estimators).
   - This tuning could be done through grid search or random search, aiming to improve the model's accuracy and robustness.
3. **Expand Dataset and Validate on Different Time Periods**:
   - Increase the dataset size to include more years or apply the model to different stocks for validation.
   - This would test the model's generalizability and robustness to market changes over time.
4. **Consider Alternative Targets**:

    - Experiment with different target variables, such as predicting the magnitude of price movement or multi-day trends, to make the model adaptable to different trading strategies.
5. **Explore Additional Evaluation Metrics**:
    - Use metrics like AUC-ROC for a more nuanced evaluation, especially if the dataset becomes imbalanced.
    - Track metrics such as mean squared error (MSE) if considering a regression approach for predicting actual price changes.
6. **Experiment with Other Models and Ensembles**:
    - Consider trying other machine learning models, such as Random Forest or Support Vector Machine (SVM), and compare their performance with XGBoost.
    - Use ensemble methods to combine predictions from different models to potentially increase accuracy and stability.

Summary

The GA-XGBoost model has demonstrated promising predictive capabilities for stock price movement, achieving an accuracy of 83% and balanced precision/recall. Moving forward, I plan to fine-tune the model, validate it on different data, and potentially explore alternative targets to enhance its effectiveness in a real-world trading environment.