Twitter-Based Stock Movement Analysis

Introduction

This project leverages Twitter data to analyze stock market sentiment and movement. It consists of four main Python scripts: data scraping, data preprocessing, machine learning model creation, and model evaluation. The goal is to predict stock trends based on sentiment analysis of tweets.

1. data_scrapper.py

This script is responsible for collecting tweets using the Twitter API. It fetches tweets based on specific keywords related to stocks and saves them in a CSV file for further processing.

Key Functions:

o   `twitter_auth`: Authenticates with the Twitter API using a bearer token.
o   `fetch_tweets`: Fetches tweets based on a query and stores their metadata.
o   `save_to_csv`: Saves the fetched tweets into a CSV file.

Output: `tweets.csv` containing raw tweet data.

2. data_preprocess.py

This script preprocesses the raw tweet data to clean and analyze sentiment. It removes URLs, mentions, hashtags, and special characters from tweets and uses the TextBlob library to determine sentiment polarity.

Key Functions:

o   `clean_tweet`: Cleans raw tweets by removing unwanted elements.
o   `analyze_sentiment`: Analyzes the sentiment of cleaned tweets.
o   `preprocess_data`: Applies cleaning and sentiment analysis and saves the results.

Output: `processed_tweets.csv` containing cleaned tweets and sentiment labels.

3. svm_ga_model.py

This script builds a Support Vector Machine (SVM) model to classify sentiment and optimizes the SVM hyperparameters using a Genetic Algorithm (GA).

Key Functions:

o   `svm_model`: Trains an SVM model with specified kernel and regularization parameter C.
o   `evaluate`: Evaluates the model accuracy during GA optimization.
o   `run_genetic_algorithm`: Uses DEAP to optimize SVM parameters.
o   `final_model`: Trains and evaluates the SVM model using the best parameters.

Output: Classification report and accuracy score.

4. model_evaluation.py

This script evaluates the performance of the trained SVM model on the test dataset. It provides detailed metrics, including accuracy, precision, recall, and F1-score.

Key Components:

- o   Loads the test dataset.
- o   Uses the trained SVM model for predictions.
- o   Outputs performance metrics for analysis.

Conclusion

This project demonstrates how Twitter data can be used to analyze stock market sentiment and movement. By combining natural language processing, machine learning, and genetic algorithms, it provides insights into stock trends based on social media activity.