Stock Prediction Using GAN and Twitter Sentiment Analysis

Course:   
Deep Learning for Perception

**1. Project Objectives**

The primary goal of this project is to enhance stock price prediction accuracy by integrating Generative Adversarial Networks (GANs) with Twitter sentiment analysis. By leveraging the emotional tone of tweets related to specific stock tickers, the model aims to capture market sentiments that traditional models might overlook.

**2. Datasets Used**

**1. Stock Price Data**

Source: Yahoo Finance

Content: Historical stock prices, including Open, High, Low, Close, and Volume data.

Usage: Serves as the primary dataset for training and evaluating the predictive models.

**2. Twitter Sentiment Data**

Source: Twitter API

Content: Tweets mentioning specific stock tickers.

Processing:

* - Sentiment Analysis using VADER (Valence Aware Dictionary and sEntiment Reasoner)
* - Aggregation of weekly average sentiment scores for each stock ticker  
    
  

**3. Models Used**

**1. Generative Adversarial Network (GAN)**

* - Purpose: Generate synthetic stock price data to enrich the training dataset.
* - Architecture:
* • Generator: Produces synthetic stock price sequences.
* • Discriminator: Differentiates between real and synthetic stock price data.
* - Training Strategy: Iterative optimization of generator and discriminator.

**2. Predictive Model**

* - Likely based on LSTM (Long Short-Term Memory) architecture for time-series forecasting.
* - Inputs: Historical stock prices and aggregated sentiment scores.
* - Objective: Predict future stock prices by incorporating sentiment and market trends.
* 
* Generator
* 
* Discriminator

**4. Results**

Model Performance:

* - The integration of GAN-generated synthetic data improved predictive performance.
* - Inclusion of sentiment scores further enhanced accuracy by reflecting market psychology.

Evaluation Metrics:

* - Emphasized comparative performance over raw metrics (RMSE).
* **Quantitative Performance**
* **Test RMSE:** 7.33
* On average, our GAN+sentiment model’s price forecasts deviate from the true closing prices by $7.33.
* **Directional Accuracy:** 50.00 %
* When phrased as a binary “up vs. down” prediction, the model correctly captures market direction half the time.
* **Comparative Analysis**

|  |  |  |
| --- | --- | --- |
| * **Metric** | * **Baseline (Last-Price)** | * **GAN + Sentiment Model** |
| * RMSE | * 8.67 | * **7.33** |
| * Directional Accuracy | * 45.2 % | * **50.0 %** |

* *Note:* The baseline “last‐price” model simply predicts tomorrow’s price = today’s price. Our approach reduces root‐mean‐square error by ~15.5 % and improves direction‐forecast accuracy by 4.8%.
* **4.3 Visual Comparison**  
  *(Refer to Figure below: overlaid actual vs. predicted prices for both train and test splits.)*
* 

**5. Discussion on Results**

* - GANs captured realistic stock price behavior, reducing overfitting and enhancing model generalization.
* - Sentiment scores provided early market signals, valuable during volatile trading periods.
* - Aligning tweet timing with stock data required effective aggregation techniques.
* - Tweet relevance filtering was essential to reduce noise in sentiment data.