Stock Price Movement Prediction Using Financial News Sentiment Analysis From Transformers

Introduction

Stock price prediction is a complex task influenced by numerous factors, including market trends, economic indicators, and news sentiment. This project utilizes **financial news headlines** to predict stock price movements. By fine-tuning a **BERT model** for sentiment analysis and designing a machine learning model, we aim to predict whether stock prices will increase or decrease.

Workflow

1.Data Collection

- **News Headlines**: Scrape headlines from the archives of the *Economic Times* website for the past year.
- **Stock Data**: Gather historical stock price data and other financial indicators to complement the sentiment data.

2.Fine-tuning BERT for Sentiment Analysis

- Use a pre-trained BERT model, fine-tuned on a dataset of financial news articles labeled with sentiment (positive, negative, neutral).
- This enables the model to interpret and assign sentiment scores to financial news headlines effectively.

3.Sentiment Analysis

- Apply the fine-tuned BERT model to infer sentiment scores for each scraped news headline.
- Use these scores as input features for predicting stock price movements.

4. Sentiment probability scores

- Apply the tiny BERT model to infer sentiment scores for each scraped news headline and corresponding probability scores for each sentiment.

Use these scores as input features for predicting stock price movements.

5. ML modelinng for prediction

We fitted the data obtained into many model

Model Implementation

Scraping Financial News

- Library: BeautifulSoup/Selenium for web scraping.
- Data Source: *Economic Times* archives.
- Output: A structured dataset of headlines with corresponding dates.

Fine-Tuning BERT

- Framework: PyTorch/Transformers (Hugging Face library).
- Dataset: Financial sentiment datasets (e.g., FiQA or a custom-labeled dataset).
- Output: A fine-tuned BERT model capable of classifying news sentiment into positive, negative, or neutral categories.

Sentiment Scoring

- Input: Scraped headlines.
- Model: Fine-tuned BERT.
- Output: Sentiment scores (e.g., -1 for negative, 0 for neutral, +1 for positive).

Sentiment probabilites

- Input: Scraped headlines.
- Model: tiny BERT.
- Output: Sentiment scores (e.g.0.9 model is 90 percent sure the sentiment is positive(1)).

Stock Price Prediction

- Features:
    - Sentiment scores.
    - Historical stock price data.
    - Sentiment probability
- Algorithms:random forest,light gbm
- Evaluation Metrics: Accuracy, F1-Score,Confusion matrix.

Technologies Used

- **Python Libraries**: Transformers, Scikit-learn, Pandas, Matplotlib, Seaborn.
- **Scraping Tools**: BeautifulSoup, Selenium.
- **Machine Learning Frameworks**: PyTorch, TensorFlow.
- **Visualization**: Plotly, Seaborn.

Project Repository

Structure

```
.
├── data/
│   ├── news_data.csv
│   ├── stock_data.csv
├── models/
│   ├── fine_tuned_bert/
│   ├── ml_model.pkl
├── notebooks/
│   ├── 1_webscrapping & merging.ipynb
│   ├── 2_bert_model_training.ipynb
│   ├── 3_bert_inference_.ipynb
    ├── 4_predictions_using_ml_model.ipynb
     ├── 5_predictions_using_tinybert_sentiment_inferencel.ipynb
```