TCS Stock Data — Analysis & Prediction

Internship Project Report

Unified Mentor Internship Program

Abstract

This concise project report summarizes exploratory analysis and a simple predictive model built for Tata Consultancy Services (TCS) stock (historical price data provided). We perform data cleaning, compute summary statistics and technical indicators, and train a baseline Linear Regression model using lag features to predict the next day's closing price. Key performance metrics and visualizations are reported; recommendations for future work are provided.

1. Data Sources & Description

- TCS_stock_info.csv :- 150 rows, contains metadata and company info (fields: zip, 400001...)
- TCS_stock_history.csv :- 4463 rows, primary historical OHLCV time series (columns: Date, Open, High, Low, Close, Volume...)
- TCS_stock_action.csv :- 70 rows, corporate actions/events (columns: Date, Dividends, Stock Splits...)

2. Summary Statistics (Close price)

```
count   4463.0000
mean     866.5374
std      829.6113
min       26.3776
25%      188.5946
50%      529.7133
75%     1154.7849
max     3954.5500
```

3. Modeling — Baseline Linear Regression

A simple Linear Regression model was trained using the previous 5 days' Close prices and a 5-day moving average.
Performance on the last 20% of the dataset:
- MAE: 26.2845

- RMSE: 36.8426
- R^2: 0.9962

4. Conclusions & Recommendations

Conclusions:
- A simple Linear Regression baseline using 5 lagged closes + 5-day moving average yields MAE = 26.2845, RMSE = 36.8426, R^2 = 0.9962 on the test set.
- The model captures short-term structure but is limited (linear assumptions, few features).

Recommendations / Future Work:
1. Use richer features: technical indicators (RSI, MACD), calendar features, volume-based features, and corporate action indicators.
2. Try time-series-specific models: ARIMA/SARIMA, Prophet, or state-space models for better temporal dynamics.
3. Use machine learning models that capture non-linearity: Random Forests, Gradient Boosting, or LSTM/Transformer-based models for sequence learning.
4. Incorporate event/corporate-action data from TCS_stock_action.csv as exogenous variables.
5. Perform walk-forward validation to respect time ordering and avoid lookahead bias.