PREDICTING STOCK MARKET WITH MACHINE LEARNING & PYTHON

A Case Study on Tata Motors

Abhishek Anand Katare
PRN – 23050209103
BATCH: 2023 - 2025

Dr. D. Y. PATIL VIDYAPEETH, PUNE
CENTRE FOR ONLINE LEARNING

EXECUTIVE SUMMARY

- MARKET PREDICTION IS COMPLEX DUE TO NON-LINEARITY AND EXTERNAL INFLUENCES
- BUILT A HYBRID MODEL USING ML + SENTIMENT ANALYSIS
- CASE STUDY: TATA MOTORS (NSE: TATAMOTORS)
- TOOLS: PYTHON, YFINANCE, SCIKIT-LEARN, NLTK, TRANSFORMERS
- MODEL ARCHITECTURE: RANDOM FOREST CLASSIFIER WITH ENGINEERED FEATURES (RATIOS & TRENDS)
- INCLUDES BACKTESTING FRAMEWORK AND REAL-TIME SENTIMENT SCRAPING
- FINAL MODEL PRECISION: ~0.66 (↑ FROM 0.57)

PROBLEM STATEMENT & OBJECTIVES

🔶 **PROBLEM**:

❖ TRADITIONAL FORECASTING STRUGGLES TO PROCESS VAST VOLUMES OF UNSTRUCTURED DATA (E.G., NEWS, SOCIAL SENTIMENT).IT OFTEN LACKS TIMELINESS, OBJECTIVITY, AND ADAPTABILITY TO MARKET PSYCHOLOGY.

✅ **OBJECTIVES**:

✓ PREDICT NEXT-DAY STOCK DIRECTION (UP/DOWN) FOR TATA MOTORS.

✓ ENGINEER TECHNICAL INDICATORS (RATIOS, TREND COUNTS) TO IMPROVE MODEL PERFORMANCE.

✓ INTEGRATE SENTIMENT ANALYSIS FROM NEWS HEADLINES USING VADER & ROBERTA.

✓ EVALUATE MODEL WITH PRECISION SCORE AND VISUAL VALIDATION (ACTUAL VS PREDICTED).

✓ BUILD A BACKTESTING FRAMEWORK FOR REAL-WORLD TRADING SIMULATION.

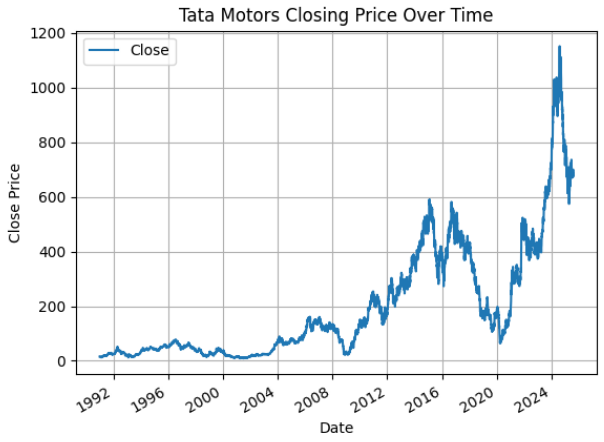
METHODOLOGY OVERVIEW

- 📌 **Data Sources**: Yahoo Finance, Economic Times, & NewsAPI.
- 🤖 **ML Model**: Random Forest Classifier, Baseline, & Enhanced
- 💬 **Sentiment Engines**: VADER (rule-based), RoBERTa (transformer)
- 📊 **Evaluation**: Precision Score, Backtesting, & Visual Plots.

DATA COLLECTION & PREPROCESSING

📈 Quantitative

- ~5000 OHLCV records from Yahoo Finance



✅ Initial Data:

Date	Open	High	Low	Close	Volume
1991-01-02 00:00:00+05:30	14.963198	15.604478	14.963198	15.604478	0
1991-01-03 00:00:00+05:30	14.963198	15.604478	14.963198	15.604478	0
1991-01-04 00:00:00+05:30	15.604478	15.604478	15.604478	15.604478	0
1991-01-07 00:00:00+05:30	14.535675	15.176956	14.321917	15.070076	0
1991-01-08 00:00:00+05:30	15.070076	15.070076	15.070076	15.070076	0

📰 Qualitative

- 100+ financial headlines scraped from Economic Times
- Supplemented with NewsAPI for recent and historical sentiment analysis
- Headlines mapped to dates to align with stock movement

⚙️ Transformations

- Feature engineering (ratios, trends)
- Target Creation: Binary label indicating next-day movement (1 = up, 0 = down)

✅ Data with Target:

Date	Close	Target
2025-07-21 00:00:00+05:30	687.450012	0
2025-07-22 00:00:00+05:30	673.400024	1
2025-07-23 00:00:00+05:30	690.099976	1
2025-07-24 00:00:00+05:30	700.500000	0
2025-07-25 00:00:00+05:30	687.400024	0

- Cleaning: Dropped missing values, removed non-informative columns

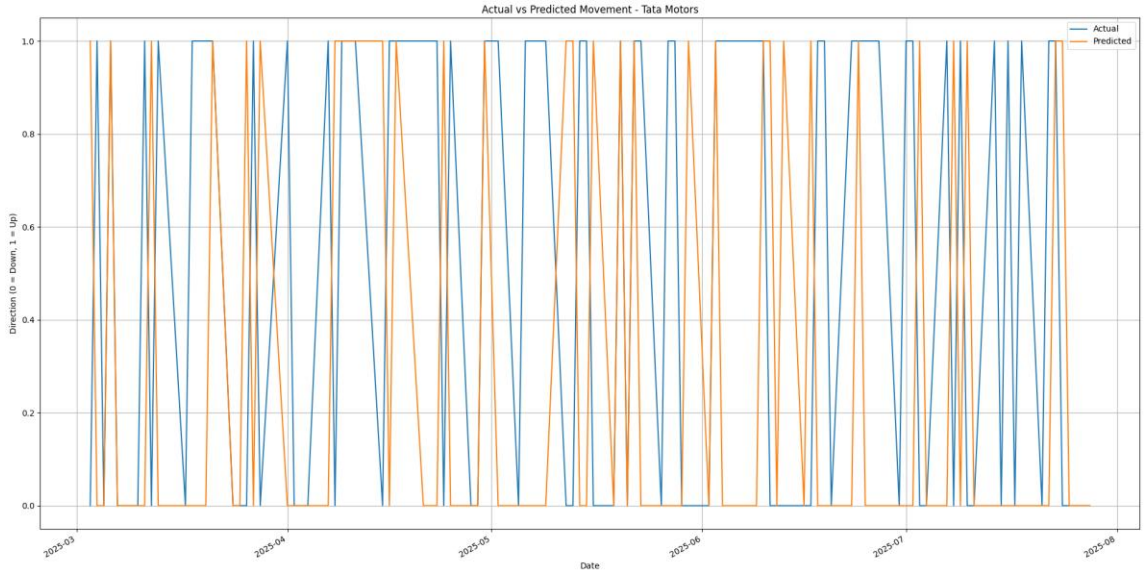
FEATURE ENGINEERING

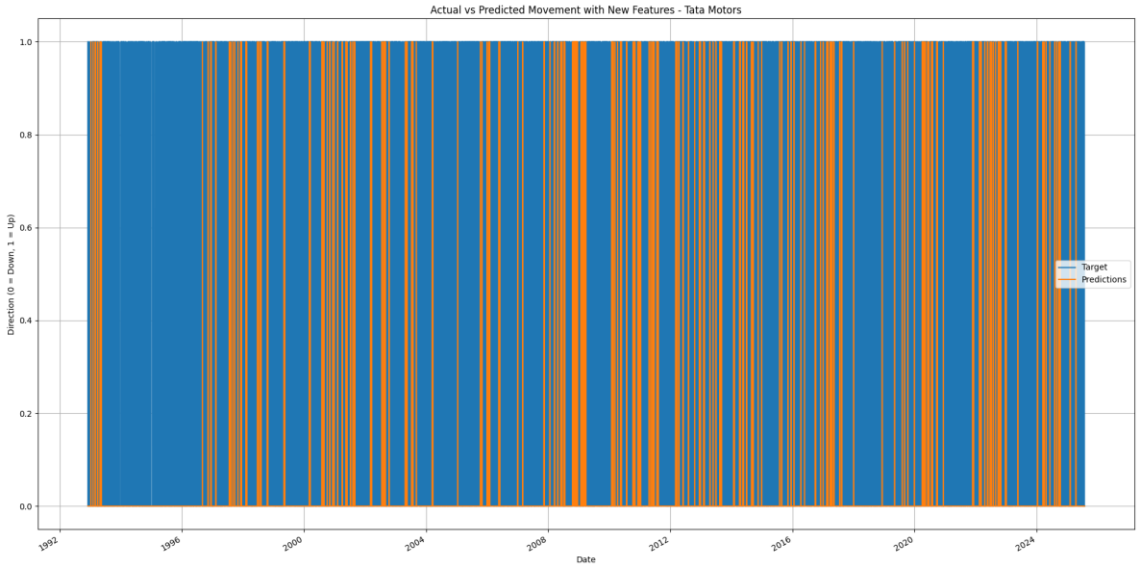
⚙️ Rolling Averages

- Rolling averages (2, 5, 20, 60, 250-day windows)
- Close Ratios & Trend indicators.

📈 Trend Indicators

- Count of upward moves in past n days.





SENTIMENT ANALYSIS

🧠 VADER (Rule-Based)

- Classifies sentiment using compound scores.
- VADER Sentiment Scores (Top 3 Headlines)

Headline	Compound Score	Sentiment
Tata Motors reports record profits for Q1	+0.78	👍 Positive
Tata Motors stock drops after global chip shortage	-0.62	👎 Negative
New EV lineup unveiled by Tata Motors	+0.65	👍 Positive

🤖 RoBERTa (Transformer-Based)

- Deeper NLP for aspect-level sentiment (EVs, profits, etc.).
- RoBERTa Aspect-Level Sentiment

Headline	Aspect	Label	Score
Tata Motors reports record profits for Q1	profit	POSITIVE	0.99
Strong sales boost Tata Motors' revenue	sales	POSITIVE	0.96
Tata Motors faces chip shortage	shortage	NEGATIVE	0.94

🖥️ OUTPUT

```
📰 Recent News from Economic Times:

📰 Fetching latest Tata Motors headlines...
No news via yfinance—scraping Economic Times instead.

🧠 Sentiment Scores:
News sentiment exported to tatamotors_news_sentiment.csv
Fetching news from 2025-07-15 to 2025-07-29
Device set to use cpu

🧠 Aspect-Level Sentiment Analysis Results:
                                        headline           aspect    label     score
0         Tata Motors reports record profits for Q1           profit  LABEL_2  0.731115
1  New electric vehicle lineup unveiled by Tata M...  electric vehicle  LABEL_1  0.728203
2           Strong sales boost Tata Motors' revenue            sales  LABEL_2  0.626371
3           Strong sales boost Tata Motors' revenue               EV  LABEL_2  0.626371
4           Strong sales boost Tata Motors' revenue          revenue  LABEL_2  0.626371
5      Tata Motors invests in EV battery production               EV  LABEL_1  0.753659
📄 Exported to aspect_sentiment_results.csv and .xlsx
```

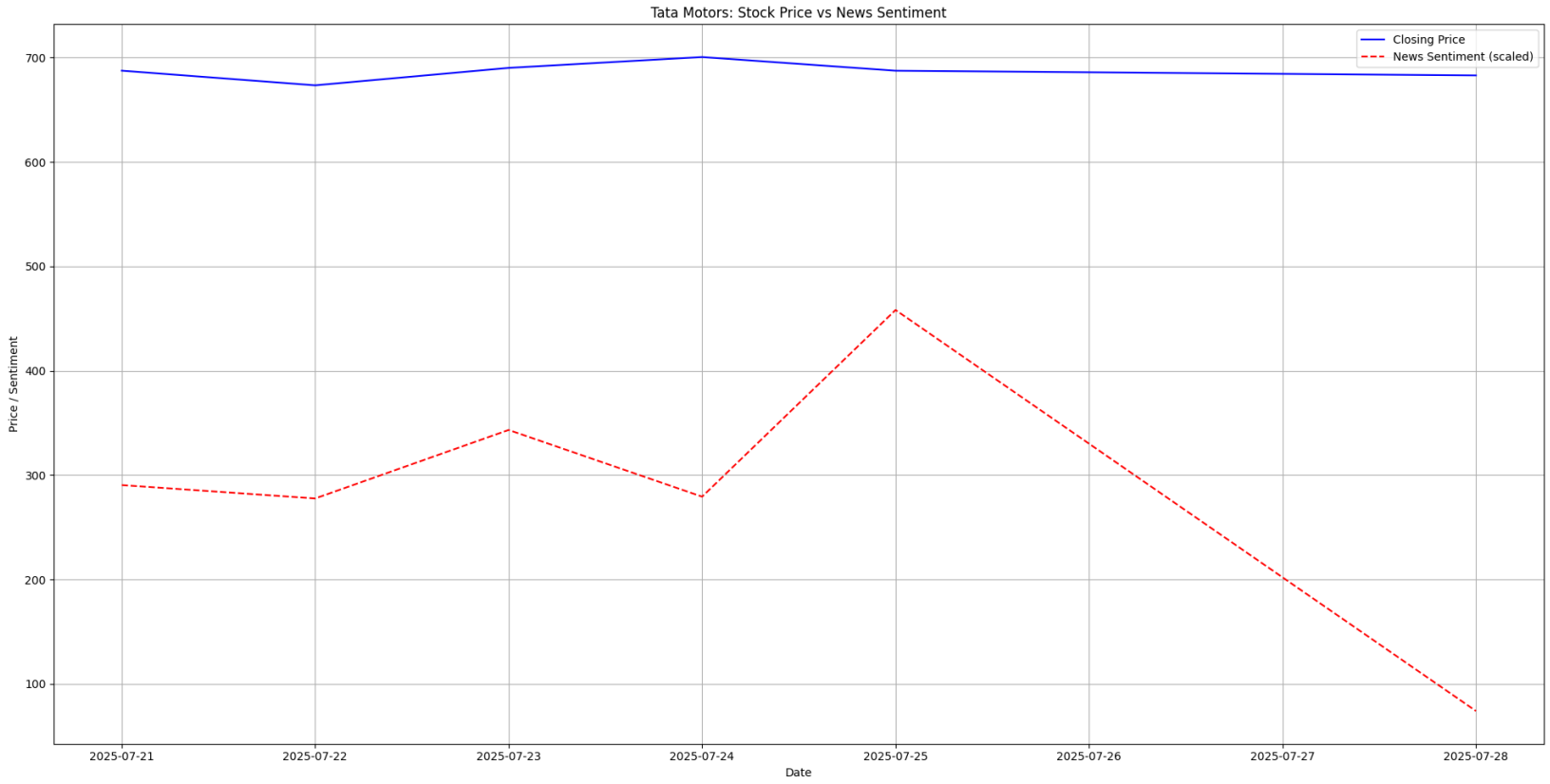
SENTIMENT ANALYSIS

📊 Sentiment-Price Correlation

- Merged **sentiment scores** with **stock prices**
- Identify patterns between **news tone and price movement**
- Sentiment used as **additional predictive feature**

✅ Final Predictions:

Date	Target	Predictions
2025-07-16 00:00:00+05:30	1	0
2025-07-17 00:00:00+05:30	0	0
2025-07-18 00:00:00+05:30	1	0
2025-07-21 00:00:00+05:30	0	0
2025-07-22 00:00:00+05:30	1	0
2025-07-23 00:00:00+05:30	1	0
2025-07-24 00:00:00+05:30	0	0
2025-07-25 00:00:00+05:30	0	0
2025-07-28 00:00:00+05:30	1	0
2025-07-29 00:00:00+05:30	0	0



MODEL RESULTS

Final model achieved **~0.66 precision**, outperforming baseline by ~9%

MODEL	FEATURES USED	SENTIMENT	PRECISION
🎯 Baseline	OHLCV only	❌	~0.57
🎯 Enhanced	OHLCV + Trends	❌	~0.63
🎯 Final	OHCVL + Trends + Sentiment	✅	~0.66

⚙️ **Threshold Tuning**:

- Custom threshold (≥ 0.6) reduced false positives
- Improved prediction **confidence and precision**

💬 **Sentiment Integration**:

- News-based sentiment features enhanced model's understanding of market psychology
- Strong correlation with price movement patterns

LIMITATIONS & FUTURE SCOPE

⚠️ **Limitations**:

- Focused on a single stock (Tata Motors)
- Daily-level prediction (no intraday insights)
- Not deployed for real-time forecasting

🚀 **Future Enhancements**:

- Use LSTM / Transformer models for time-series learning
- Extend to **multiple stocks** and **intraday predictions**
- Build a live dashboard with real-time alerts & sentiment tracking

CONCLUSION

📌 Summary

- Built a hybrid ML model combining **technical indicators** and **news sentiment**
- Demonstrated effectiveness using **Tata Motors** case study
- Achieved **~0.66 precision**, up from ~0.57 (baseline)

📈 Key Learnings

- **Feature engineering** boosts model performance
- **Sentiment signals** add real-world context to predictions
- **Backtesting** ensures realistic model validation

🔮 What's Next?

- 📊 Add more stocks & sectors
- ⏱️ Use intraday/tick-level data
- 🤖 Explore deep learning (LSTM, Transformer forecasting)
- 📑 Integrate macroeconomic indicators
- 📡 Deploy real-time dashboard + alert system

🧠 Impact

- Provides a blueprint for **algorithmic trading**, **decision support**, and **financial analytics**

THANK YOU

Abhishek Anand Katare

PRN – 23050209103

BATCH: 2023 - 2025

Dr. D. Y. PATIL VIDYAPEETH, PUNE

CENTRE FOR ONLINE LEARNING