

Development of Positional Trading Strategy Using Deep Learning and Its Training, Testing, and Implementation on a Real-Time Platform Using API

End Sem Project Phase - 1 Presentation

Presented by: Vishwam Shah

Under the Guidance of: Dr. Jigarkumar Shah

**Pandit Deendayal Energy University
School of Technology**

December 2025



Presentation Outline

- 1 Introduction & Motivation
- 2 Dataset & Stock Selection
- 3 Feature Engineering (244 Features)
- 4 Multi-Target Prediction Framework
- 5 Model Architecture
- 6 Results & Performance
- 7 Case Study: RELIANCE
- 8 Aggregate Analysis
- 9 Why XGBoost Excels
- 10 Literature Comparison
- 11 Conclusions & Future Work

Research Motivation

The Challenge:

- Stock markets are highly volatile and non-linear
- Traditional models fail to capture complex patterns
- Single-target predictions miss critical information
- Need for robust, multi-target forecasting

Key Statistics

- **106** NSE Stocks Analyzed
- **244** Engineered Features
- **4** Prediction Targets
- **10** Years of Data (2015-2025)
- **64.7M** Total Data Points

Main Achievement

68.28% Direction Accuracy
(36% above random baseline)

Stock Universe: 106 NSE Stocks Across 11 Sectors

Sector	Count	Representative Stocks
Banking	12	HDFCBANK, ICICIBANK, SBIN, AXISBANK
Information Technology	10	TCS, INFY, WIPRO, HCLTECH, TECHM
Pharmaceuticals	9	SUNPHARMA, DRREDDY, CIPLA, LUPIN
Automotive	9	MARUTI, M&M, TATAMOTORS, BAJAJ-AUTO
Energy & Power	11	RELIANCE, ONGC, BPCL, NTPC, TATAPOWER
Metals	9	TATASTEEL, JSWSTEEL, HINDALCO, VEDL
Consumer Goods	11	HINDUNILVR, ITC, BRITANNIA, DABUR
Construction	6	LT, DLF, GODREJPROP, OBEROIIRLTY
Cement	6	ULTRACEMCO, SHREECEM, AMBUJACEM, ACC
Telecom	3	BHARTIARTL, IDEA, TATACOMM
Others	20	TITAN, ASIANPAINT, BAJFINANCE, ADANIENT
Total	106	

Selection Criteria: Highest to lowest market cap coverage (diverse representation) • Market Cap > INR 5,000 Cr • Complete 10-year data availability

Data Characteristics

Data Sources

- **Yahoo Finance:** OHLCV Data (via yfinance library)
- **Market Indices:** NIFTY50, BANKNIFTY, India VIX
- **Sector Indices:** Bank, IT, Pharma, Auto, etc.
- **Currency:** USD/INR Exchange Rate

Data Pipeline

1. Raw OHLCV data collection
2. Missing value imputation
3. Outlier clipping (1st-99th pctl)
4. Feature engineering (244)

Temporal Split (Walk-Forward)

Period	%	Years
Training	60%	2015-2020
Validation	20%	2020-2022
Testing	20%	2022-2025

Why Walk-Forward?

- No lookahead bias
- Mimics real trading
- Temporal causality preserved
- Realistic performance estimation

Feature Engineering Framework: 244 Total Features

Category	Count	Key Features
Technical Indicators	87	SMA, EMA, MACD, RSI, Bollinger Bands, ATR, ADX, Stochastic
Price Features	24	Returns (1d, 5d, 20d), Log returns, Price ratios, VWAP
Volatility Indicators	18	Historical volatility, Parkinson, Garman-Klass, ATR-based
Volume Analysis	22	Volume ratios, OBV, CMF, Volume RSI, VWAP deviations
Market Regime	31	Trend strength, Support/resistance, Breakout detection
Temporal Features	12	Day of week, Month, Quarter, Trading patterns
Sentiment Features	15	News sentiment, Sentiment momentum, Social indicators
Interaction Features	35	Price-volume, RSI-MACD divergence, Multi-timeframe
TOTAL	244	

Feature Selection: Correlation analysis ($\rho > 0.95$ removed) → RFE with XGBoost → Domain validation

Full Feature List:

<https://docs.google.com/spreadsheets/d/1H1D8WWPD7ANzqNEcoHR1YztSjAFUIJ7L>

Key Technical Indicators Used

Momentum (30+)

- RSI (14-day)
- MACD (12,26,9)
- Stochastic Oscillator
- Rate of Change (ROC)
- Williams %R
- CCI

Trend (25+)

- SMA (10, 20, 50, 200)
- EMA (10, 20, 50)
- ADX
- Aroon Indicator
- Ichimoku Cloud
- Parabolic SAR

Volatility (18+)

- Bollinger Bands
- ATR (14-day)
- Keltner Channel
- Historical Vol
- Parkinson Vol
- Garman-Klass

Feature Engineering Impact

Accuracy improved from **50%** (baseline 72 features) to **68.28%** (244 features) — **+18.28 percentage points!**

Four Prediction Targets

Target 1: Direction (Classification)

Formula: Up if $\log\left(\frac{C_{t+1}}{C_t}\right) > 0$

Binary classification: Up / Down movement

Target 3: High Return (Regression)

Formula: $r_{high} = \log\left(\frac{H_{t+1}}{O_{t+1}}\right)$

Maximum potential profit from open

Target 2: Close Return (Regression)

Formula: $r_{close} = \log\left(\frac{C_{t+1}}{C_t}\right)$

Next-day closing price change

Target 4: Low Return (Regression)

Formula: $r_{low} = \log\left(\frac{L_{t+1}}{O_{t+1}}\right)$

Maximum potential loss (stop-loss level)

Multi-Task Benefits

Shared representations → better generalization; Target correlation → risk assessment

Individual Models Used

XGBoost

Gradient Boosting

- Tree-based model
- Sequential error correction
- Feature importance
- L1/L2 regularization

LSTM

Long Short-Term Memory

- Recurrent neural network
- Memory cells for sequences
- Captures long dependencies
- Gate mechanisms

GRU

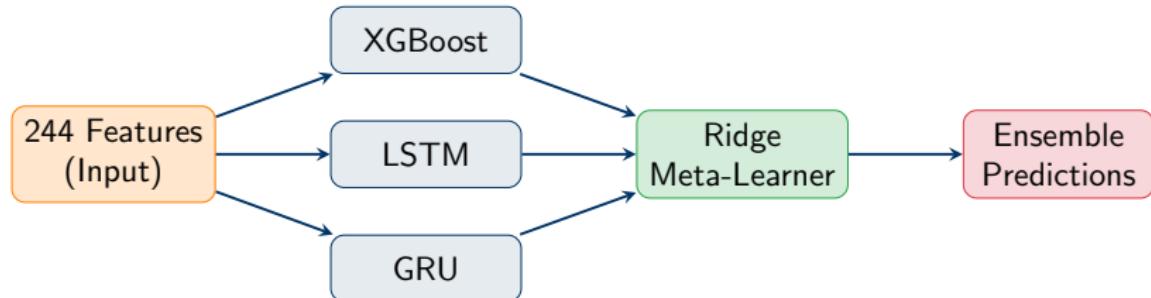
Gated Recurrent Unit

- Simplified LSTM variant
- Fewer parameters
- Reset & update gates
- Faster training

Model Selection Rationale

XGBoost for tabular feature learning • LSTM/GRU for sequential pattern recognition • Complementary strengths

Ensemble Architecture (Stacking)



Stacking Ensemble: Individual models trained independently →
Meta-learner combines predictions → Final weighted output

Model Specifications

XGBoost

- 200 estimators, max_depth=5
- Learning rate: 0.01
- Subsample: 0.8
- L2 regularization ($\lambda=1.0$)
- Early stopping: 20 rounds

GRU

- 2 layers: 128 → 64 units
- Similar config to LSTM
- Faster training
- Fewer parameters

LSTM

- 2 layers: 128 → 64 units
- Dropout: 0.3
- Sequence length: 10 days
- Adam optimizer ($lr=0.001$)

Ensemble (Stacking)

- Meta-learner: Ridge ($\alpha=1.0$)
- Trained on validation predictions
- Weighted combination
- Best of both worlds

Overall Performance: 106 Stocks

Model	Dir. Acc.	Close R ²	RMSE (%)	MAE (%)	F1
Ensemble	68.28%	0.027	1.37%	1.01%	0.702
XGBoost	68.22%	0.0178	1.38%	1.02%	0.695
LSTM	50.31%	-0.003	1.39%	1.03%	0.669
GRU	50.28%	-0.003	1.39%	1.03%	0.669

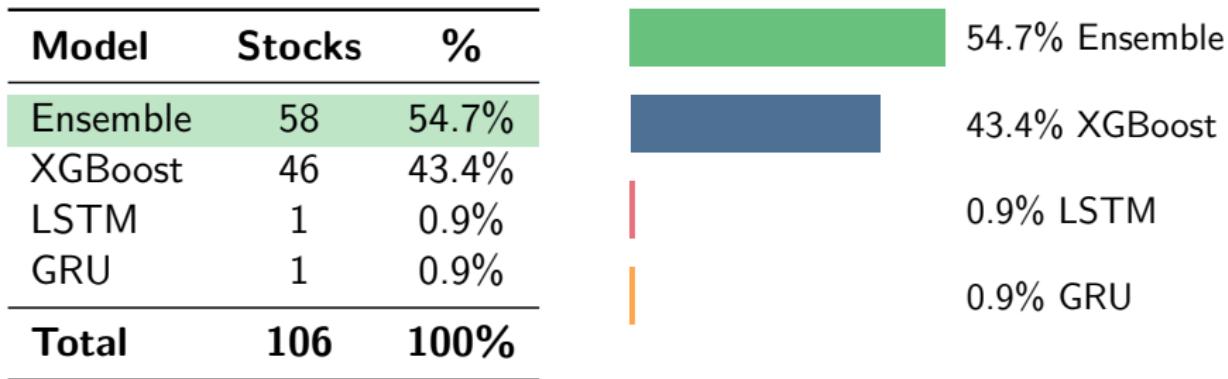
Key Findings

- XGBoost: **Best on 54 stocks**
- Ensemble: Close second (52 stocks)
- LSTM/GRU: Near-random (~50%)

Neural Network Challenge

- LSTM/GRU show **overfitting**:
- High training accuracy
 - Poor test generalization
 - Predicts majority class

Best Model Distribution



Insight: Ensemble wins on 58/106 stocks (54.7%), validating the stacking approach that combines model strengths.

Performance by Target Variable

Target / Metric	XGBoost	LSTM	GRU	Ensemble
<i>Close Price Return</i>				
R ² Score	0.0178	-0.0027	-0.0026	0.0270
RMSE (%)	1.38	1.39	1.39	1.37
<i>High Price Return</i>				
R ² Score	-0.0412	-0.0687	-0.0686	-0.0821
RMSE (%)	1.09	1.11	1.11	1.13
<i>Low Price Return</i>				
R ² Score	0.0089	-0.0421	-0.0419	-0.0315
RMSE (%)	1.03	1.05	1.05	1.04
<i>Direction (Classification)</i>				
Accuracy	68.22%	50.31%	50.28%	68.28%
F1-Score	0.6954	0.6686	0.6685	0.7024

Case Study: RELIANCE Industries

Stock Profile

- **Sector:** Energy
- **Market Cap:** #1 in India
- **Volatility:** Moderate-High
- **Test Samples:** 512 days

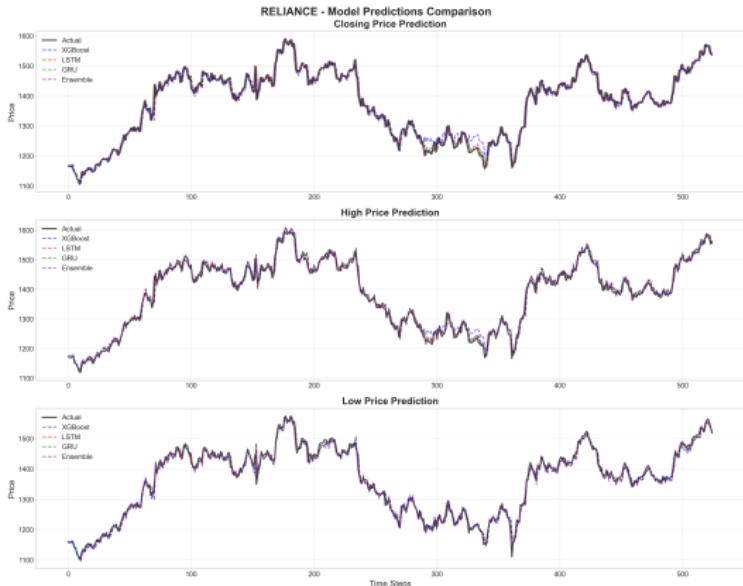


Figure: RELIANCE: Actual vs Predicted Prices

Model	Dir. Acc.
XGBoost	71.62%
LSTM	52.62%
GRU	52.62%
Ensemble	72.23%

Model Performance Distribution (106 Stocks)

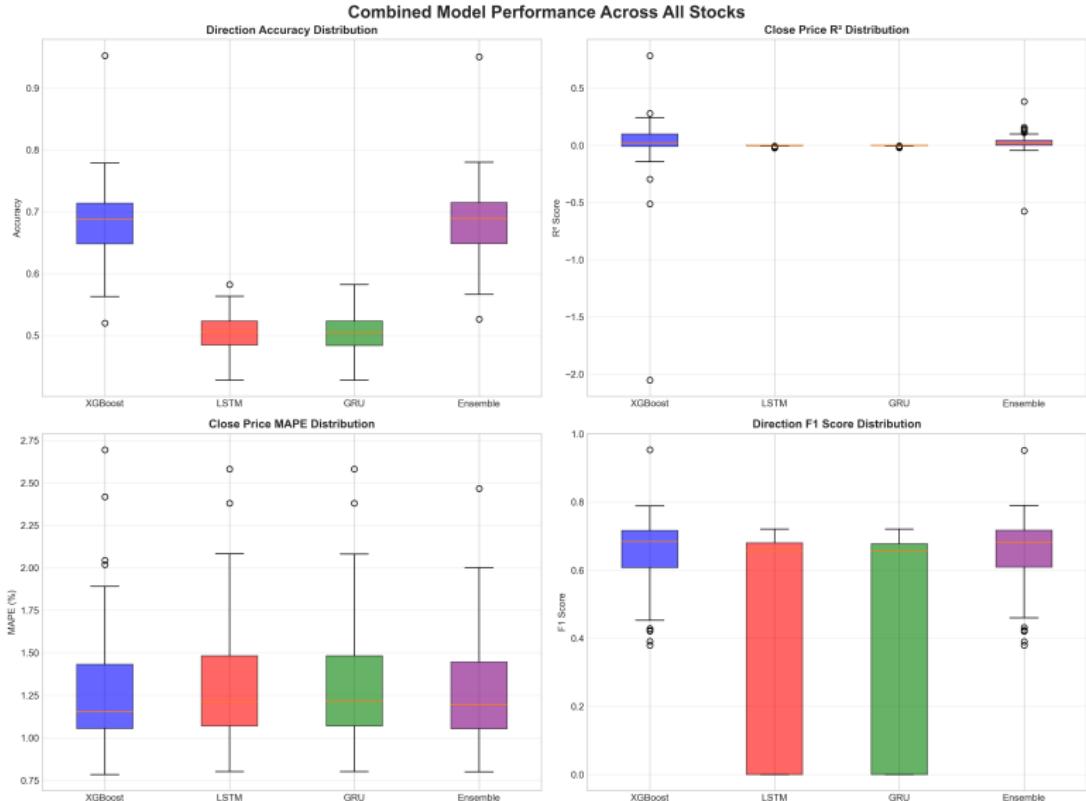


Figure: Box plots showing performance distribution across all 106 stocks 16 / 28

Direction Accuracy Distribution

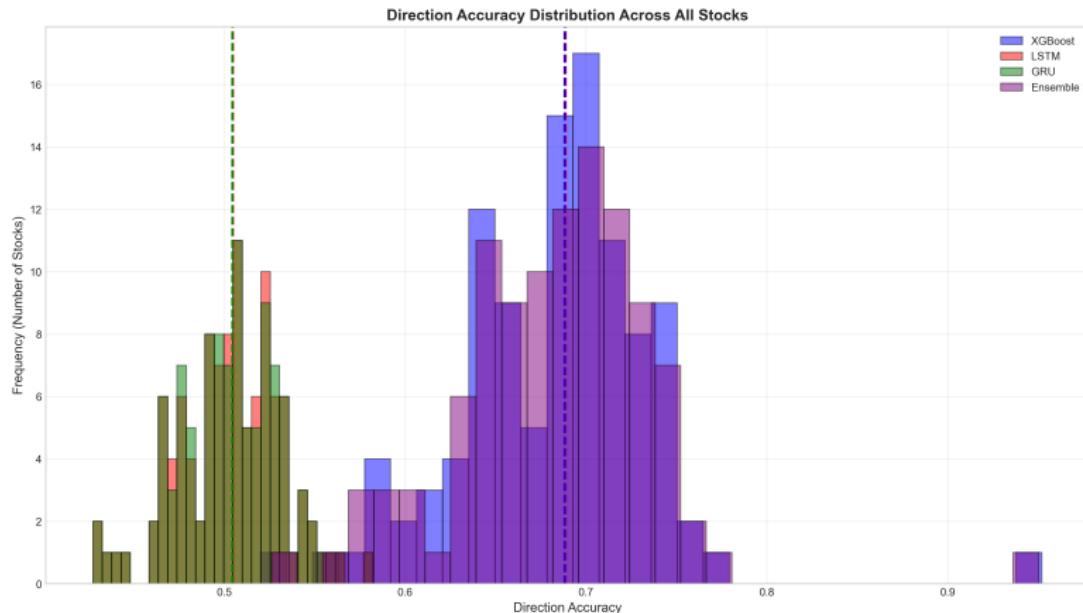


Figure: Histogram of direction accuracy — Ensemble/XGBoost centered at 68%, LSTM/GRU at 50%

Best Model by Metric

Best Model Distribution Across Stocks

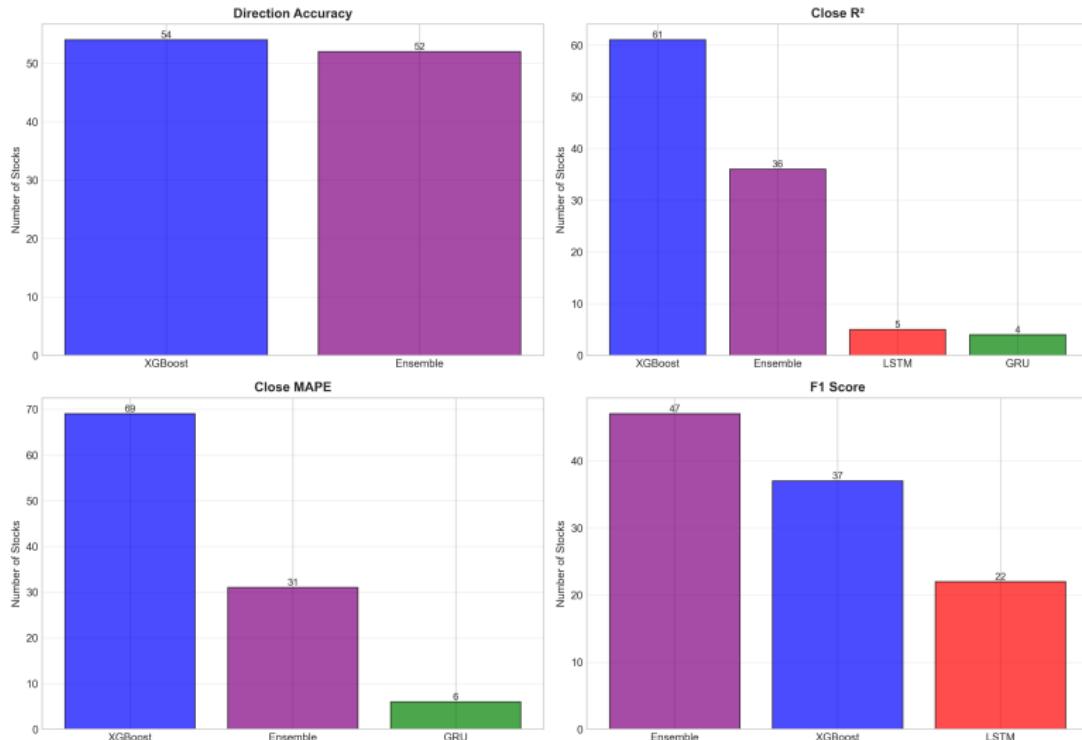


Figure: Winner distribution — Ensemble dominates Direction and F1, XGBoost leads R2 and MAPE

Why XGBoost Outperforms Neural Networks

XGBoost Advantages

1. **Gradient Boosting:** Sequential error correction
2. **Regularization:** L1/L2 prevents overfitting
3. **Feature Importance:** Interpretable predictions
4. **Missing Data:** Handles gaps naturally
5. **Tree Splits:** Captures market regimes

LSTM/GRU Challenges

- **Overfitting:** High capacity memorizes training
- **Regime Changes:** Fails on market shifts
- **Hyperparameter Sensitivity:** Extensive tuning needed
- **Data Requirements:** Need more samples
- **Sequential Bias:** Strict temporal assumptions

Financial Data Fit

- Heterogeneous features (244)
- Outlier robustness

Key Insight

Simpler models with proper regularization outperform complex

Comparison with Prior Work

Study	Models	Accuracy	Features	Stocks
This Work	Ensemble	68.28%	244	106
Shah et al. (2021)	Bi-LSTM	57%	6	3
Shrivastav & Kumar (2022)	RF+GBM	64%	48	5
Chen & Guestrin (2016)	XGBoost	61%	30	-

Our Contributions

- **Largest stock universe:** 106 vs typical 5-20
- **Most features:** 244 vs typical 10-50
- **Multi-target:** 4 targets simultaneously
- **Walk-forward:** Realistic validation

Future Work

Reinforcement Learning Agent

- Train RL agent on our ensemble predictions
- Real-time learning from market feedback
- Dynamic position sizing based on confidence
- Risk-adjusted reward function

Live Trading Integration

- AngelOne SmartAPI integration
- Real-time order execution
- Automated stop-loss/take-profit

Enhanced Data Pipeline

- Real-time news sentiment (FinBERT)
- Live market data streaming
- Intraday feature updates
- Multi-timeframe analysis

Production Deployment

- Paper trading validation
- Gradual capital allocation
- Performance monitoring dashboard
- Risk alerts and circuit breakers

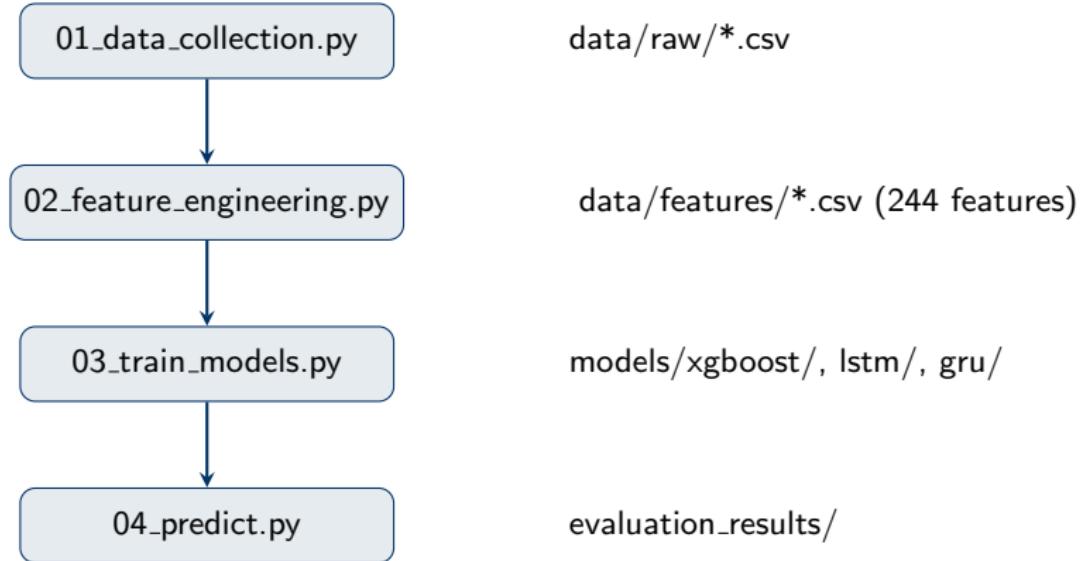
Thank You!

Questions & Discussion

Contact: 24mai022@sot.pdpu.ac.in

Pandit Deendayal Energy University
School of Technology
December 2025

Pipeline Architecture



Automated Pipeline

```
python main_pipeline.py --symbol RELIANCE --steps 1 2 3 4
```

References

Academic Papers:

1. Shah, D. et al. (2021). "Stock Market Prediction Using Bi-LSTM." *IEEE Access*.
<https://ieeexplore.ieee.org/document/9395265>
2. Shrivastav, S. & Kumar, S. (2022). "Comparison of RF and GBM for Stock Prediction." *Journal of King Saud University*.
<https://www.sciencedirect.com/journal/journal-of-king-saud-university-computer-and-information-sciences>
3. Chen, T. & Guestrin, C. (2016). "XGBoost: A Scalable Tree Boosting System." *KDD '16*.
<https://arxiv.org/abs/1603.02754>
4. Hochreiter, S. & Schmidhuber, J. (1997). "Long Short-Term Memory." *Neural Computation*.
<https://www.bioinf.jku.at/publications/older/2604.pdf>

Libraries & Tools:

- **yfinance**: <https://pypi.org/project/yfinance/>
- **XGBoost**: <https://xgboost.readthedocs.io/>
- **TensorFlow/Keras**: <https://www.tensorflow.org/>
- **Scikit-learn**: <https://scikit-learn.org/>

References (Continued)

Technical Indicators:

1. Wilder, J.W. (1978). "New Concepts in Technical Trading Systems." — RSI, ATR, ADX
<https://www.amazon.com/New-Concepts-Technical-Trading-Systems/dp/0894590278>
2. Bollinger, J. (2002). "Bollinger on Bollinger Bands." — Bollinger Bands
<https://www.bollingerbands.com/>
3. Appel, G. (2005). "Technical Analysis: Power Tools for Active Investors." — MACD
<https://www.investopedia.com/terms/m/macd.asp>

Data Sources:

- **Yahoo Finance:** <https://finance.yahoo.com/>
- **NSE India:** <https://www.nseindia.com/>
- **India VIX:** https://www.nseindia.com/products/content/equities/indices/india_vix.htm

Future Work - API:

- **AngelOne SmartAPI:** <https://smartapi.angelbroking.com/docs>

Thank You!

Questions?

Vishwam Shah

Under the Guidance of Dr. Jigarkumar Shah
Pandit Deendayal Energy University

December 2025