Intraday Trading Buy/Sell Prediction

Team -20

Ananya Sarkar (ananya1@iisc.ac.in)

Nikita Dange (nikitadange@iisc.ac.in)

Raj Shekhar (rajshekhar1@iisc.ac.in)

S Vasanthakumar (vasanthakum1@iisc.ac.in)

Introduction

- Intraday trading prediction focuses on forecasting short-term price movements within the same trading day.
- It is characterized by extreme noise, non-stationarity, and what we call a 'Random Walk.'
- Essentially, predicting the next minute's price is mathematically incredibly difficult.
- Intraday trading prediction aims to use data-driven techniques to forecast short-term market direction within the same trading day.
- Traditional indicators such as EMA, RSI, and MACD capture trend, momentum, and reversal signals, while machine learning models attempt to learn complex patterns that are not visible through simple rule-based methods.
- The objective is to convert high-frequency market information into accurate buy/sell recommendations that improve decision-making for traders.

About Dataset

- ICICIPRULI - 816504 records (2017-09-27 09:44:00 2025-07-25 15:29:00)
- ICICIGI - 724352 records (2016-09-29 09:44:00 2025-07-25 15:29:00)
- HDFCAMC - 644432 records (2018-08-06 09:44:00 - 2025-07-25 15:29:00)

- **Source**: National Stock Exchange (NSE) of India **2015 to 2025**(Dataset)
- **Frequency**: 1-Minute intervals (High-frequency trading data)
- **Size**: Millions of rows representing minute-by-minute trading activity
- **Format**: Multi-file CSV file (chosen ICICIPRULI, ICICIGI, HDFCAMC bank sector)

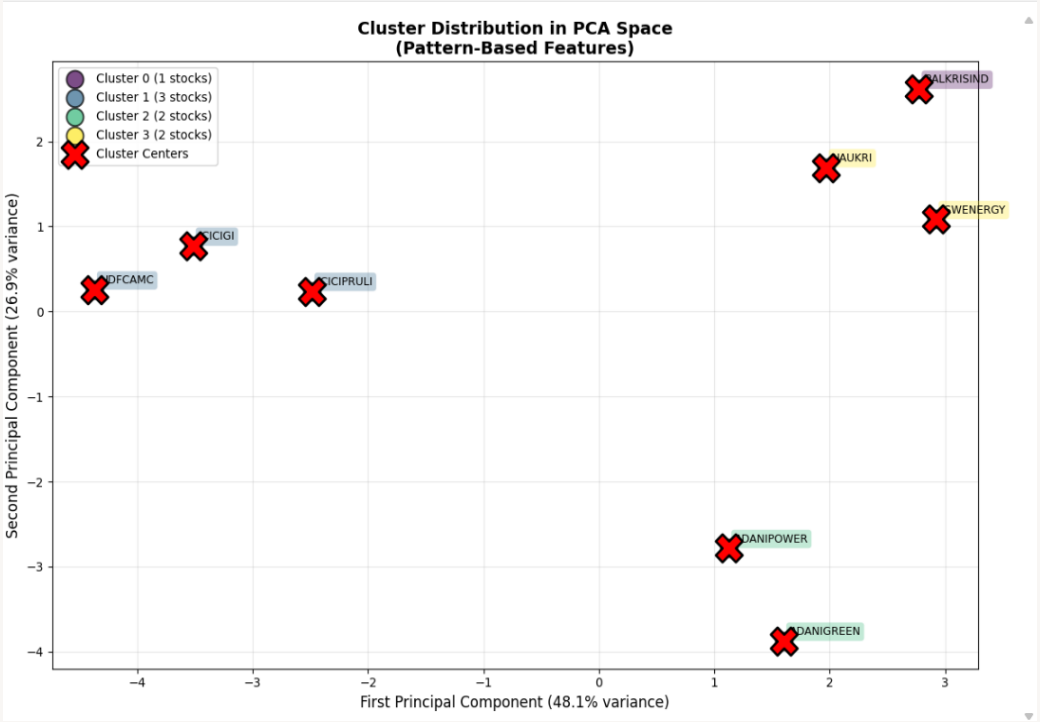
	date	open	high	low	close	volume
2015-02-02 09:15:00	52.50	52.70	52.40	52.60	142198
2015-02-02 09:16:00	52.50	52.65	52.35	52.45	106566
2015-02-02 09:17:00	52.40	52.50	52.25	52.40	140405
2015-02-02 09:18:00	52.35	52.40	52.05	52.20	103328
2015-02-02 09:19:00	52.20	52.25	52.15	52.20	75008
...	...	...	...	...	...	...
2025-07-25 15:25:00	569.75	569.95	569.60	569.95	12577
2025-07-25 15:26:00	569.70	570.35	569.65	570.05	21887
2025-07-25 15:27:00	569.55	569.90	569.50	569.70	12977
2025-07-25 15:28:00	569.70	570.95	569.60	570.95	15581

Pattern based clustering

Total 8 companies:

- ADANIGREEN
- ADANIPOWER
- BALKRISIND
- HDFCAMC
- ICICIGI
- ICICIPRULI
- JSWENERGY
- NAUKRI

- K-mean clustering
  - Fetch optimum clusters: Elbow method
- Chosen k=4



FEATURES

Trend Characteristics (3 features)
- trend_direction - Percentage of time in uptrend (SMA20 > SMA50)
- avg_price_vs_sma20 - Average deviation from 20-period moving average
- avg_price_vs_sma50 - Average deviation from 50-period moving average

Volatility Patterns (3 features)
- avg_rolling_volatility - Average 20-period rolling volatility
- volatility_of_volatility - Standard deviation of rolling volatility
- avg_intraday_range - Average daily high-low range relative to open

Momentum & Direction (2 features)
- momentum_roc - Rate of change over 10 periods
- positive_days_ratio - Percentage of positive price change days

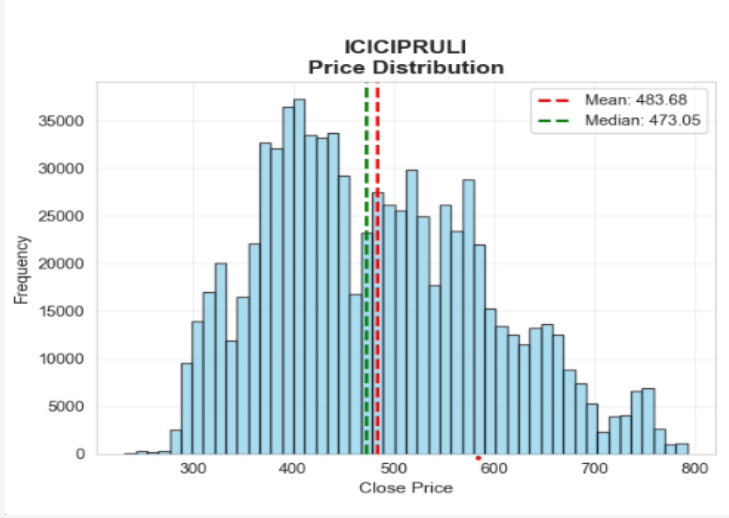
Volume Patterns (2 features)
- volume_trend - Overall volume trend direction
- avg_volume_ratio - Average volume relative to 20-period moving average

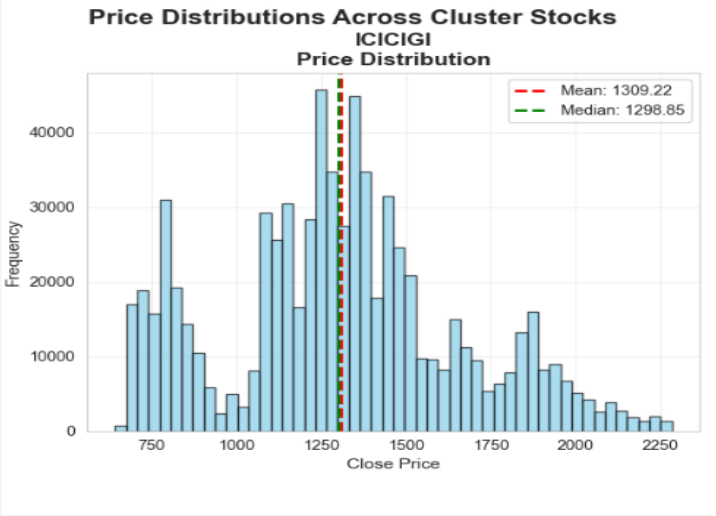
Persistence & Memory (2 features)
- autocorr_lag1 - 1-period autocorrelation (trend persistence)
- autocorr_lag5 - 5-period autocorrelation (longer-term memory)

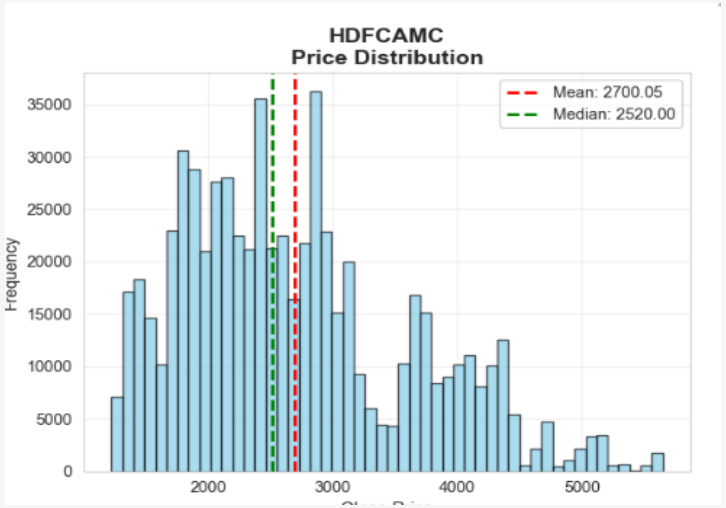
Price Action Patterns (2 features)
- avg_gap - Average opening gap from previous close
- reversal_tendency - Tendency for large moves to reverse

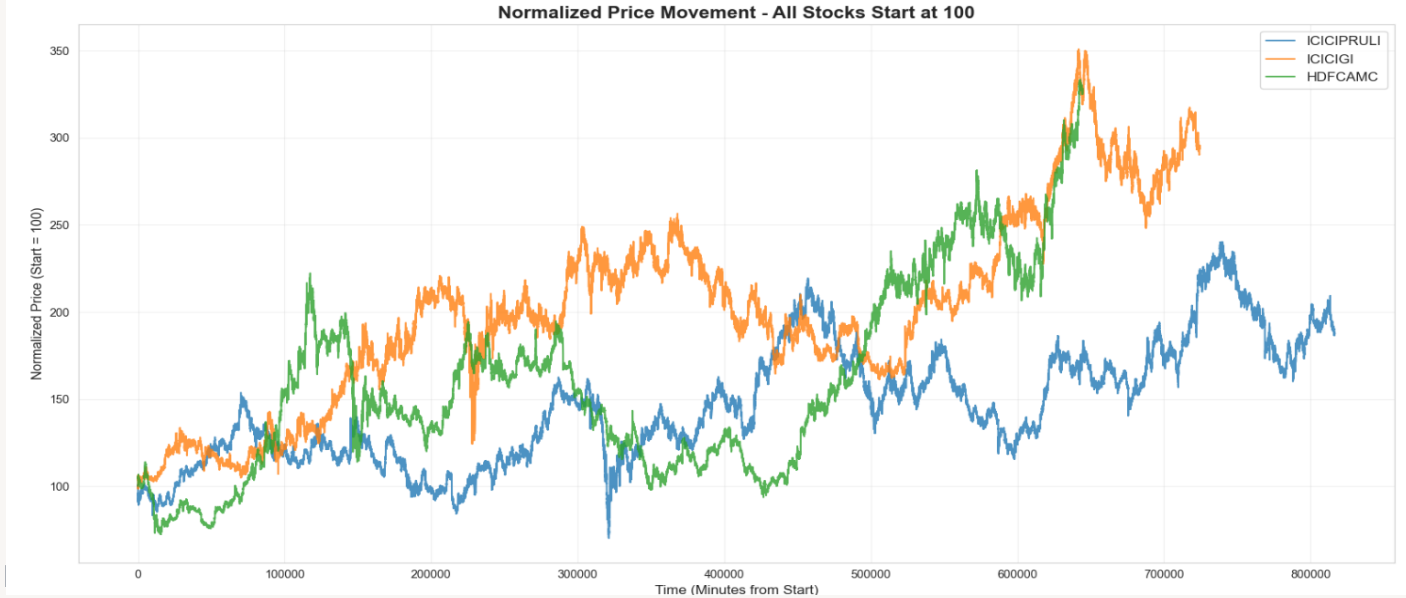
Distribution Shape (2 features)
- skewness_returns - Skewness of returns distribution
- kurtosis_returns - Kurtosis of returns (tail heaviness)

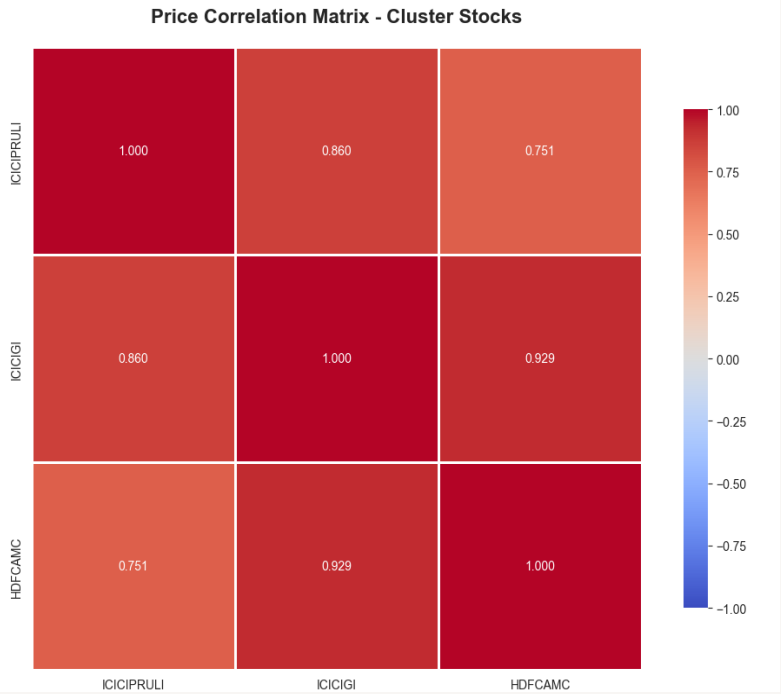
EDA - > Cluster 1 : HDFCAMC, ICICIGI, ICICIPRULI











Trend

- ICICIGI
  - consistent, smoother upward trend with moderate volatility
  - decent capital appreciation
- ICICIPRULI
  - lowest volatility among three
  - Capital Preservation Asset – Stable
- HDFCAMC
  - most aggressive growth trajectory
  - "high risk, high reward" asset in the portfolio

Feature engineering (28 feature)

- **1. Opening Period Statistics (8 features):**
- open_price, open_high, open_low, open_close, open_range, open_range_pct, open_change_pct, open_volatility
- **2. Time-Segmented Returns (4 features):**
- first_20min_return, second_20min_return, third_20min_return, open_returns_std
- **3. Volume Features (5 features):**
- avg_volume_per_min, volume_trend, total_volume_opening, volume_surge, volume_consistency
- **4. Position Features (3 features):**
- high_in_first_half, low_in_first_half, price_above_open
- **5. Technical Indicators (4 features):**
- opening_rsi, bb_position, price_momentum, price_acceleration
- **6. Additional Features (4 features):**
- hl_ratio, upper_shadow_avg, lower_shadow_avg, trend_strength
- **7. Target Variables (2):**
- day_high, day_low (full day's actual high and low prices)

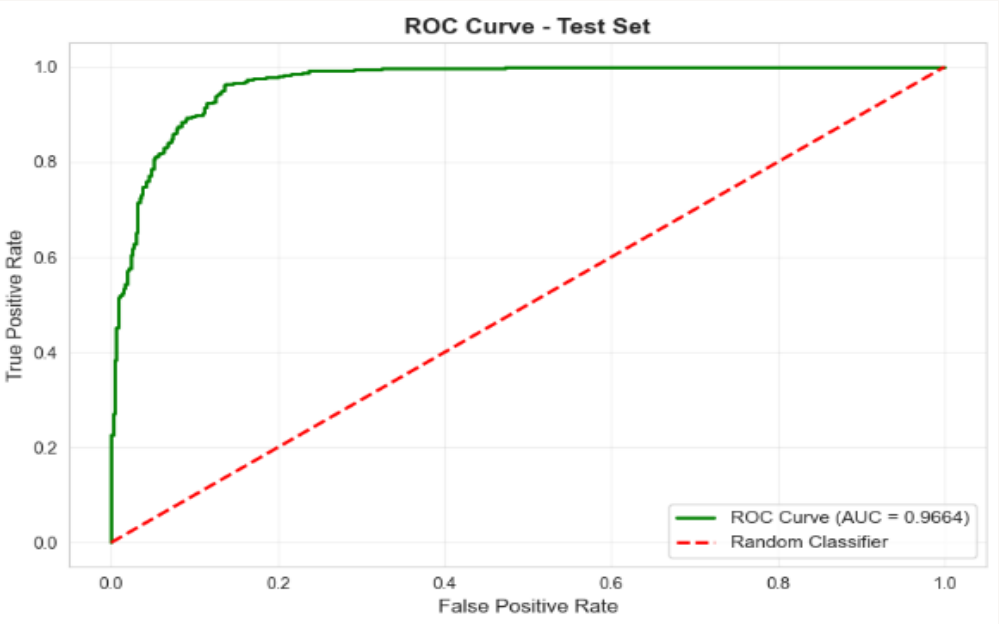
Training

- SPLIT: training: 70%, testing: 15%, validation:15%
- Train/prediction on first 60 minute.
- Training:

    - Classification Model -> ensemble of 5 models (XGBoost, LightGBM, CatBoost, Neural Network with Focal Loss, and Stacking Ensemble) combined using weighted averaging.
        - Predicts -> BUY/SELL signal
    - Regression
        - Model -> 2 LSTM model
            - **HIGH Price Model**: Predicts the daily maximum price from 15-day sequences of opening period features using a 3-layer LSTM architecture (128→64→32 units)
            - **LOW Price Model**: Predicts the daily minimum price from 15-day sequences of opening period features using a 3-layer LSTM architecture (128→64→32 units)

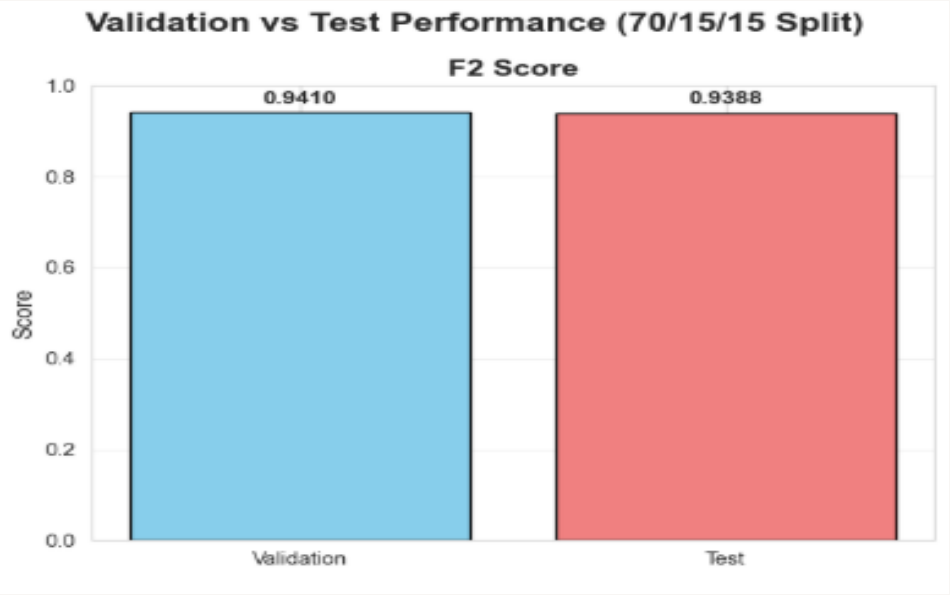
Classification model

- **Target Variable:**
- **BUY :** Daily LOW occurs before daily HIGH (upward price movement)
- **SELL :** Daily HIGH occurs before daily LOW (downward price movement)



Model Composition:
The ensemble combines **5 classification models** with optimized weights:

Model	Weight	SMOTE Variant	Training Details
XGBoost	0.3 (highest)	Standard (0.95)	700 estimators, lr=0.01, depth=5
LightGBM	0.2	Borderline (0.98)	700 estimators, lr=0.01, depth=5
CatBoost	0.2	Balanced (1.0)	700 iterations, lr=0.01, depth=5
Neural Network	0.2	Standard (0.95)	128→64→32, Focal Loss
Stacking Ensemble	0.1 (lowest)	Borderline (0.98)	Meta: LogisticRegression, CV=5



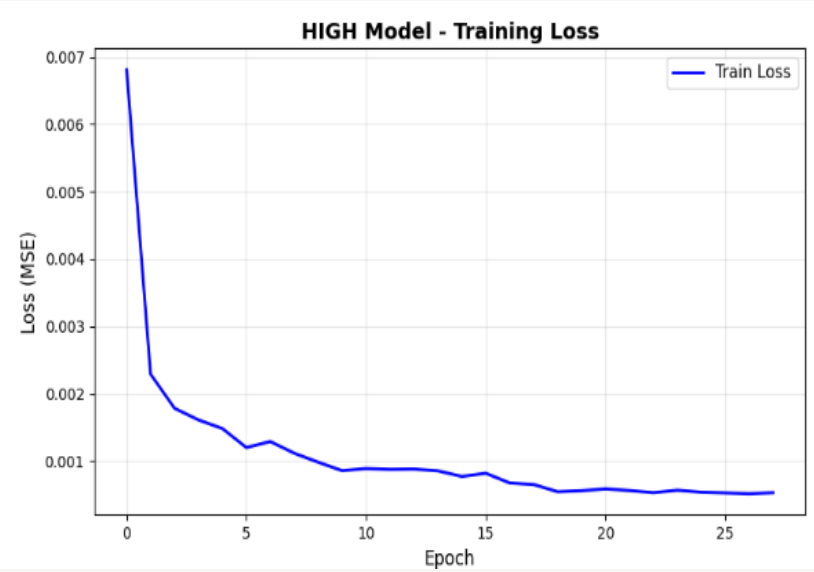
Regression Model

HIGH Price Model

- **Architecture**:
  - **Input**: 15-day sequences × 28 features (opening period characteristics)
  - **Layer 1**: LSTM with 128 units, returns sequences for next layer, **Dropout 1**: 30% dropout for regularization
  - **Layer 2**: LSTM with 64 units, returns sequences, **Dropout 2**: 30% dropout
  - **Layer 3**: LSTM with 32 units (final LSTM layer), **Dropout 3**: 20% dropout
  - **Dense Layer**: 16 neurons with ReLU activation
  - **Output**: Single value (predicted high price)
- **Training Configuration**:
  - **Optimizer**: Adam (learning rate: 0.001)
  - **Loss Function**: MSE (Mean Squared Error)
  - **Metrics**: MAE (Mean Absolute Error)
  - **Max Epochs**: 200
  - **Batch Size**: 32
  - **Early Stopping**: Patience of 30 epochs on validation loss
  - **Learning Rate Reduction**: Factor of 0.5 after 15 epochs of no improvement

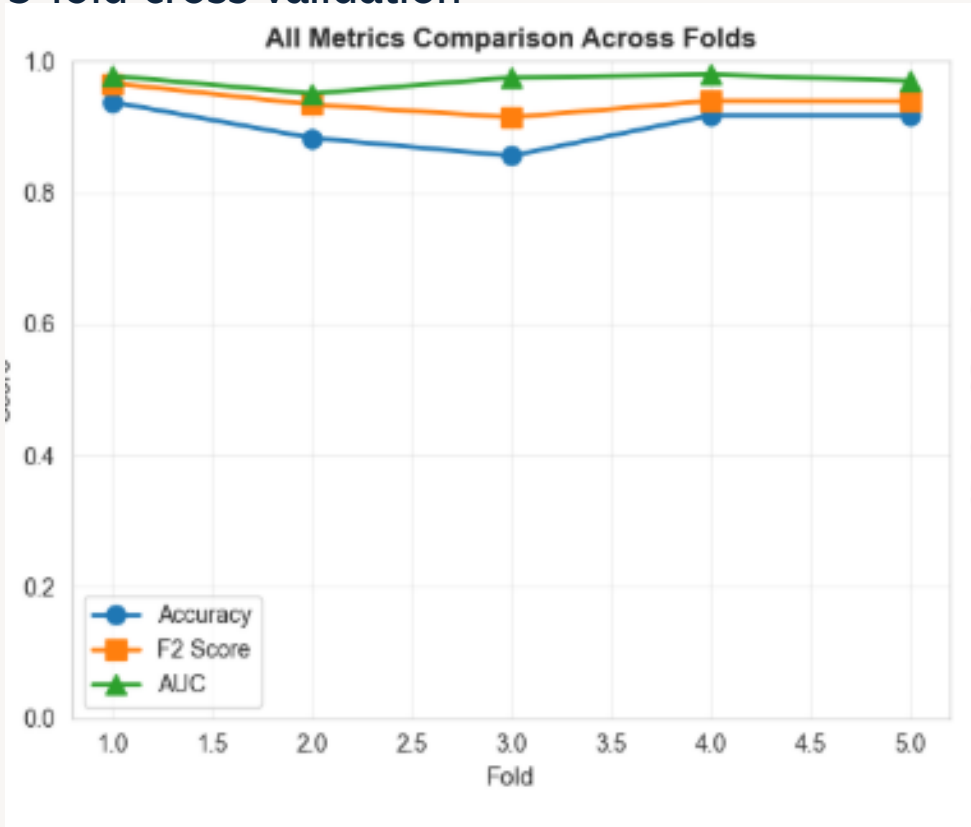
LOW Price Model

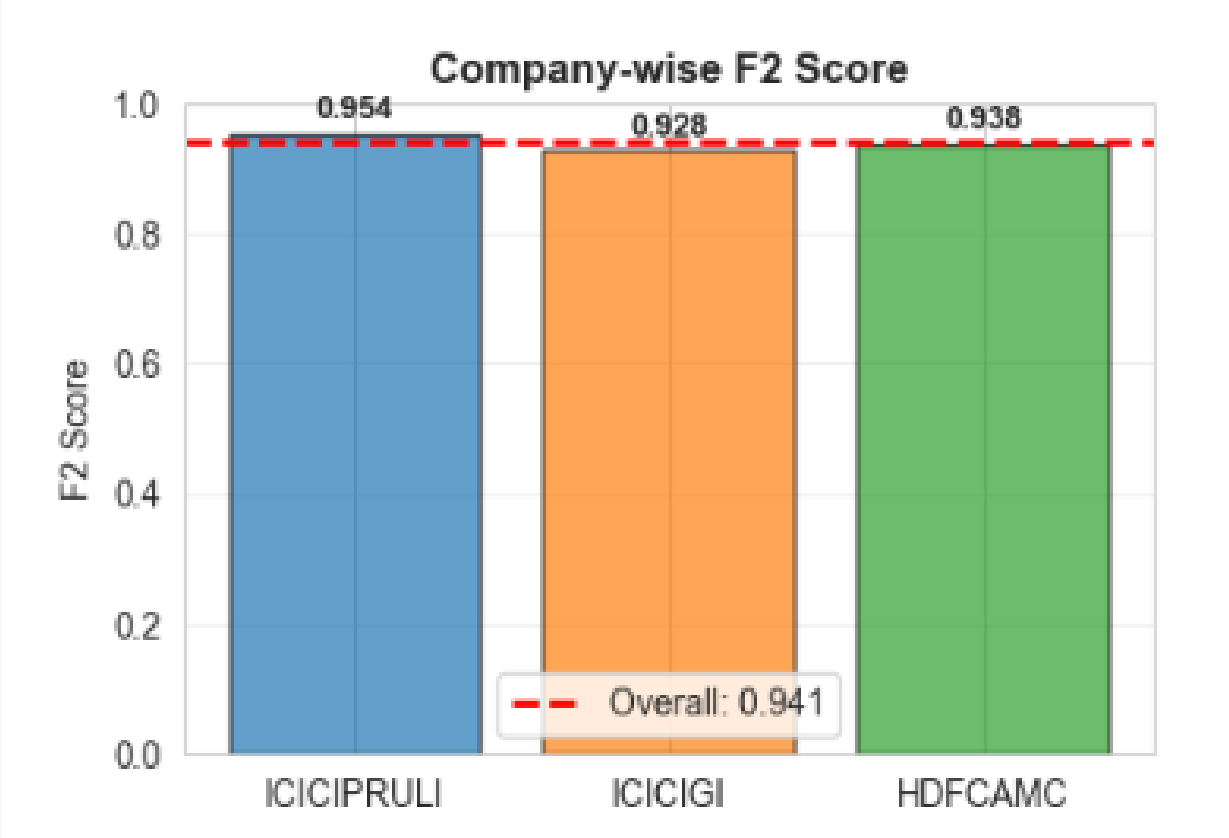
- **Architecture**:
  - Identical architecture to HIGH model
  - **Input**: 15-day sequences × 28 features
  - **Layers**: 128 → 64 → 32 LSTM units with dropout
  - **Output**: Single value (predicted low price)
- **Training Configuration**:
  - Same hyperparameters as HIGH model
  - Trained independently on daily low prices
  - Uses same callbacks and optimization strategy



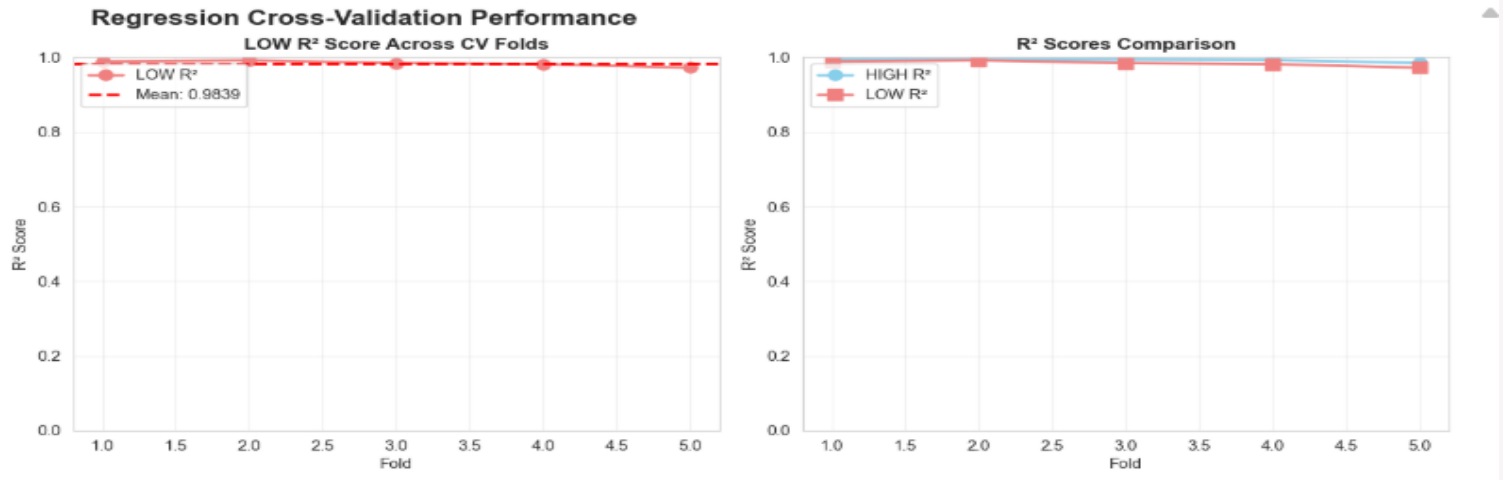
Results : Classification

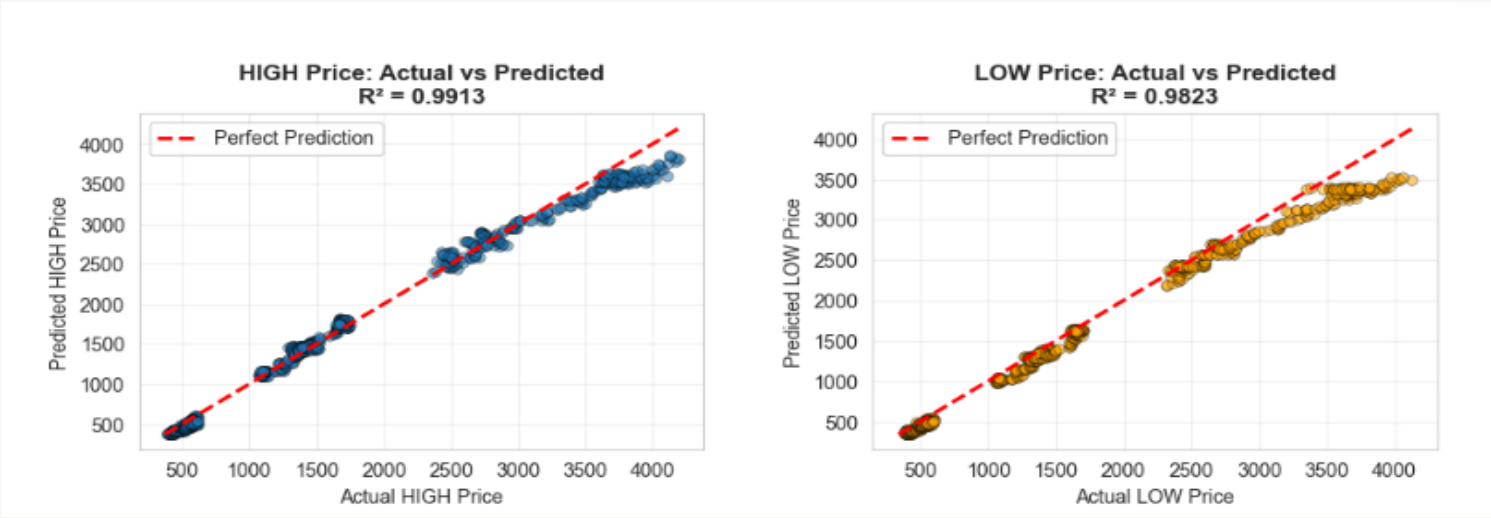
5-fold cross validation





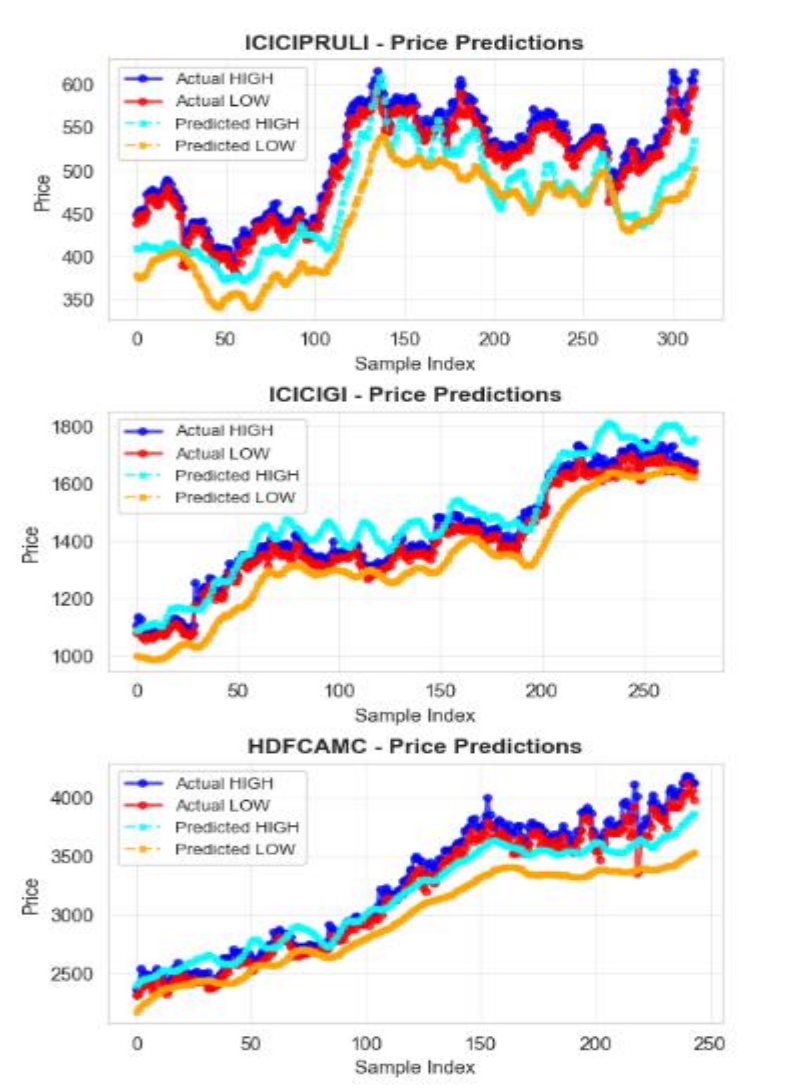
Result : Regression

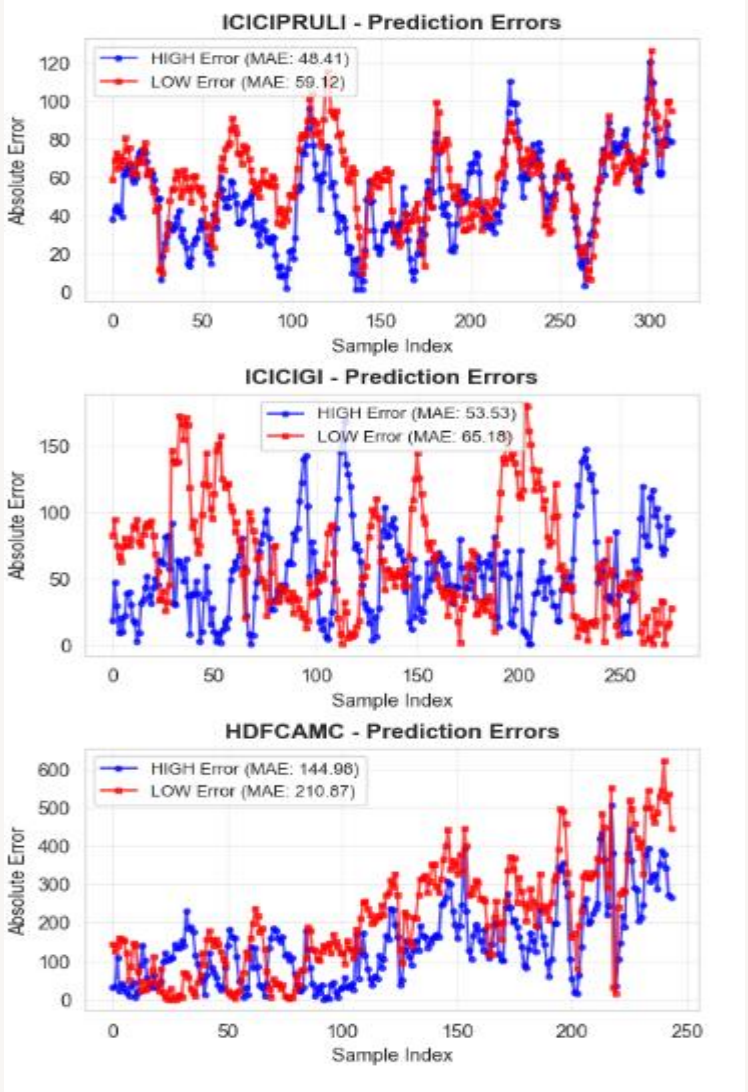




5-fold cross validation

Prediction


ICICIPRULI - Price Predictions
Actual HIGH
Actual LOW
Predicted HIGH
Predicted LOW
Price
Sample Index
ICICIGI - Price Predictions
Actual HIGH
Actual LOW
Predicted HIGH
Predicted LOW
Price
Sample Index
HDFCAMC - Price Predictions
Actual HIGH
Actual LOW
Predicted HIGH
Predicted LOW
Price
Sample Index


ICICIPRULI - Prediction Errors
HIGH Error (MAE: 48.41)
LOW Error (MAE: 59.12)
Absolute Error
Sample Index
ICICIGI - Prediction Errors
HIGH Error (MAE: 53.53)
LOW Error (MAE: 65.18)
Absolute Error
Sample Index
HDFCAMC - Prediction Errors
HIGH Error (MAE: 144.98)
LOW Error (MAE: 210.87)
Absolute Error
Sample Index

THANK YOU

Team -20