Forecasting Stock Price Movements: A Comparative Analysis of Machine Learning Techniques

AISHIK PAL, HIMMAT PATEL, ANOOP BEHERA,
HARDIK SINGH, JAGDISH SUTHAR
{b22ee085, b22ee033, b22cs011,
b22ee092, b22ai064}@iitj.ac.in

भारतीय प्रौद्योगिकी संस्थान जोधपुर
Indian Institute of Technology Jodhpur
॥ त्वं ज्ञानमयो विज्ञानमयोऽसि ॥

Problem of Market Unpredictability

Stock market trends are highly complex due to a multitude of factors, including:

- Market Volatility: economic indicators, geopolitical events, and investor sentiment.
- Nonlinear Relationships
- Interconnected Global Markets
- Information Asymmetry
- Human Behaviour

- Traditional statistical methods and econometric models have limitations in capturing nonlinear relationships and adapting to changing market conditions.

Related works

- **Stock Closing Price Prediction using Machine Learning Techniques**

  Methods- Forest Regression, Artificial Neural Network, Stock market prediction

- **Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques**

  Methods- Artificial Neural Network (ANN), Support Vector Machine (SVM), random forest and naive-Bayes.

- **Stock Market Prediction Using LSTM Recurrent Neural Network**

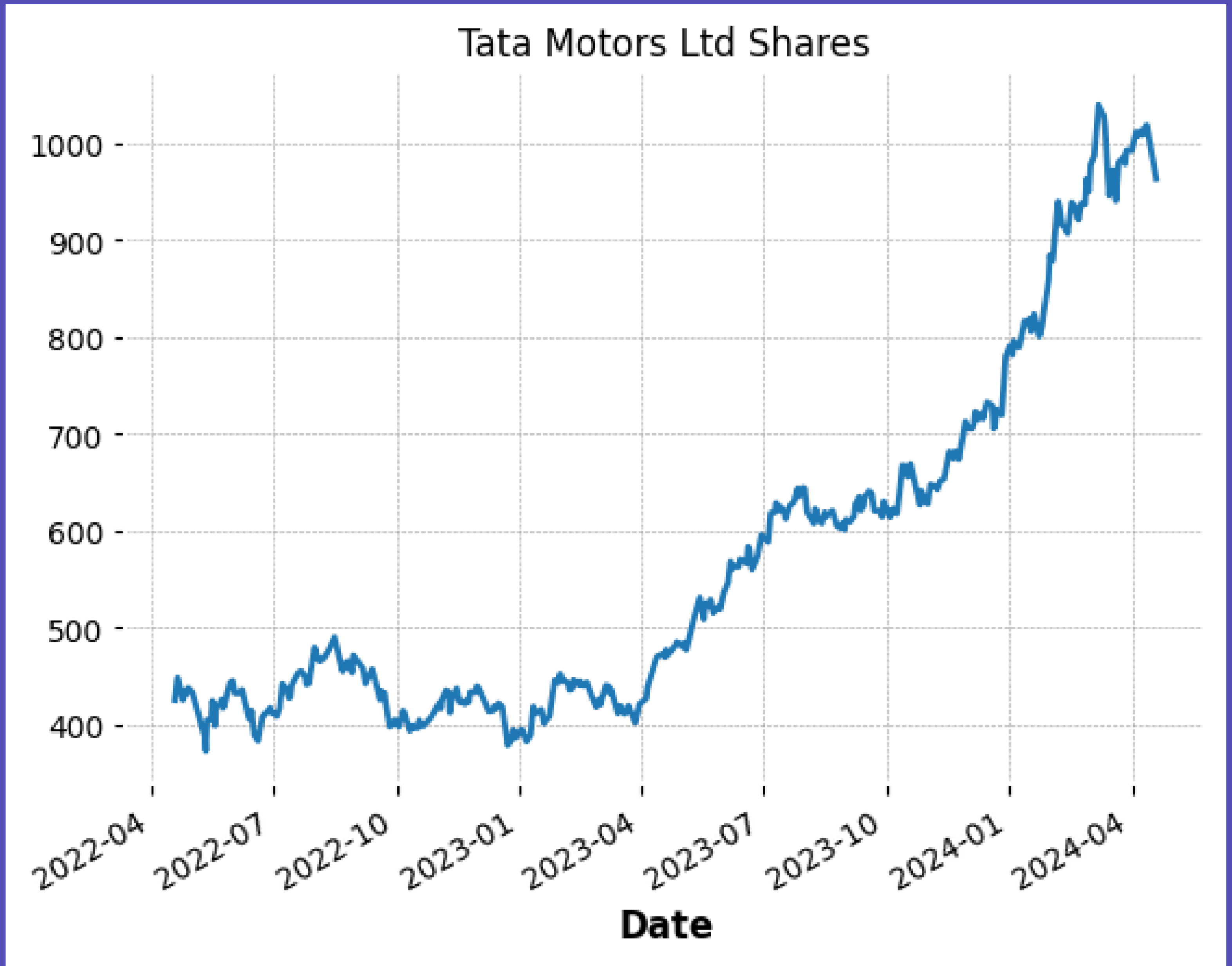
  Methods- Recurrent Neural Networks (RNN) and especially Long-Short Term Memory model (LSTM)

Dataset

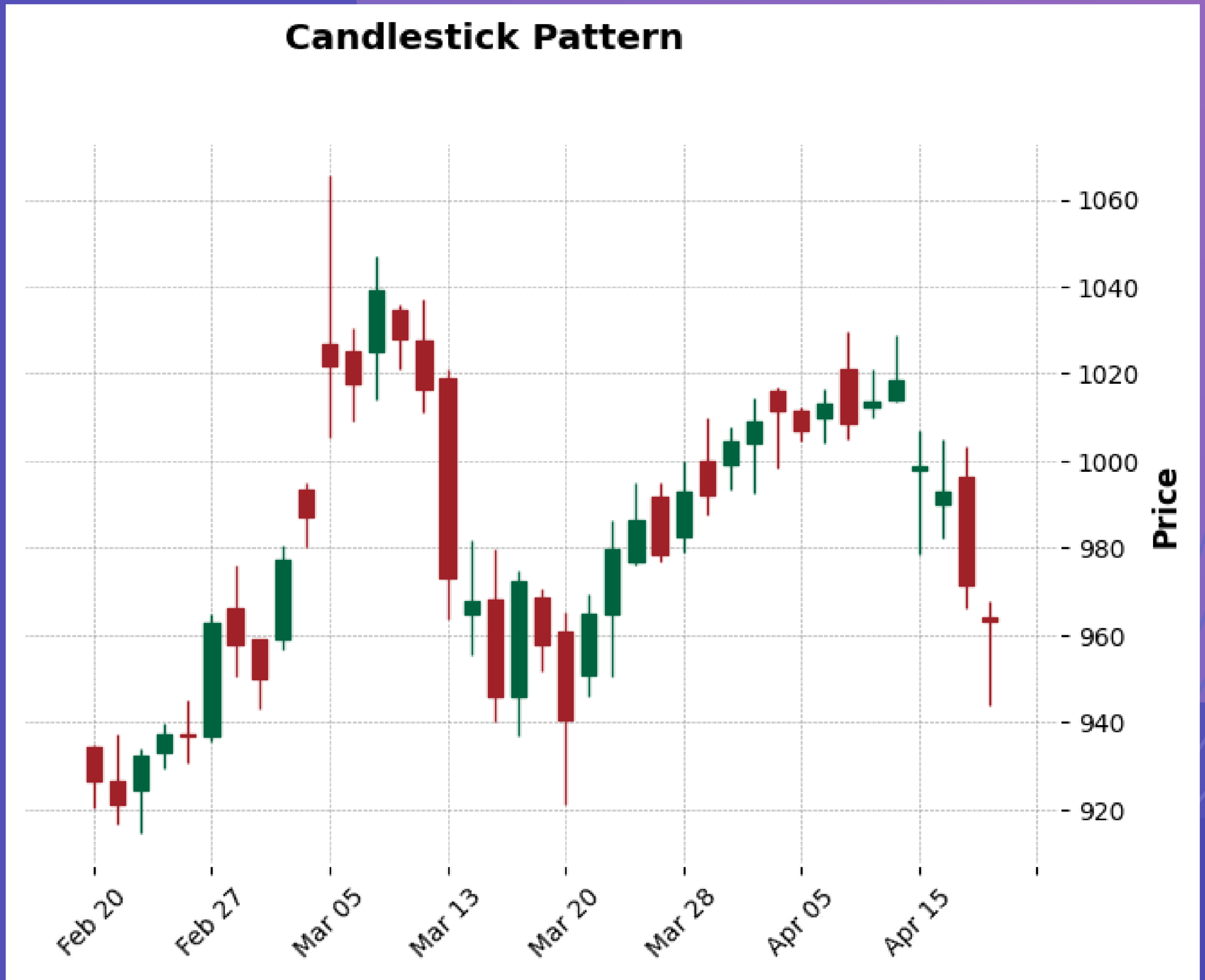
- **yfinance** : This data includes information like historical prices such as open, high, low, close, trading volumes, market capitalization, and other financial metrics.
- It gives real time data making it ideal for making realtime.

	Open	High	Low	Close	Adj Close	Volume
Date						
2022-04-19	436.750000	440.250000	417.000000	424.950012	423.629700	12669041
2022-04-20	431.000000	441.700012	428.250000	440.549988	439.181183	19639290
2022-04-21	444.500000	449.350006	441.049988	448.049988	446.657898	14124872
2022-04-22	442.049988	447.250000	438.000000	438.850006	437.486481	14362036
2022-04-25	432.750000	436.000000	423.149994	425.149994	423.829041	14729009

Stock of a company



2 YEARS DATA SET



2 MONTHS DATA SET

+TECHNICAL INDICATORS:

We have utilised traditional statistical methods, namely, Moving Average Convergence Divergence (MACD), Moving Average (MA), Relative Strength Index (RSI), Stochastic Oscillator (%K and %D), Volume Moving Average (Volume MA), Volume Rate of Change (Volume ROC), Average True Range (ATR), volatility indicators, Williams %R (%R), and candlestick patterns, with machine learning techniques.

The candlestick patten was recgonised by using open source candlestick library. and the indicators were calculated as follows

$$MA = \frac{Sum of Closing Prices over\,(n) period}{n}$$

$$\text{EMA} = (\text{P} * \alpha) + (\text{Previous EMA} * (1 - \alpha))$$

$$\text{MACD Line} = \text{12-day EMA} - \text{26-day EMA}$$

Bollinger %b = (Closing Price - Lower Band) / (Upper Band - Lower Band).

$$\%R = \left(\frac{\text{Highest High} - \text{Current Close}}{\text{Highest High} - \text{Lowest Low}}\right) \times -100$$

$$\text{RSI} = 100 - (100 \div (1 + \text{RS})), \text{ where RS} = \frac{\text{Average of upward price changes over 'n' periods}}{\text{Average of downward price changes over 'n' periods}}$$

$$\%K = \left(\frac{\text{Current Close} - \text{Lowest Low}}{\text{Highest High} - \text{Lowest Low}}\right) \times 100$$

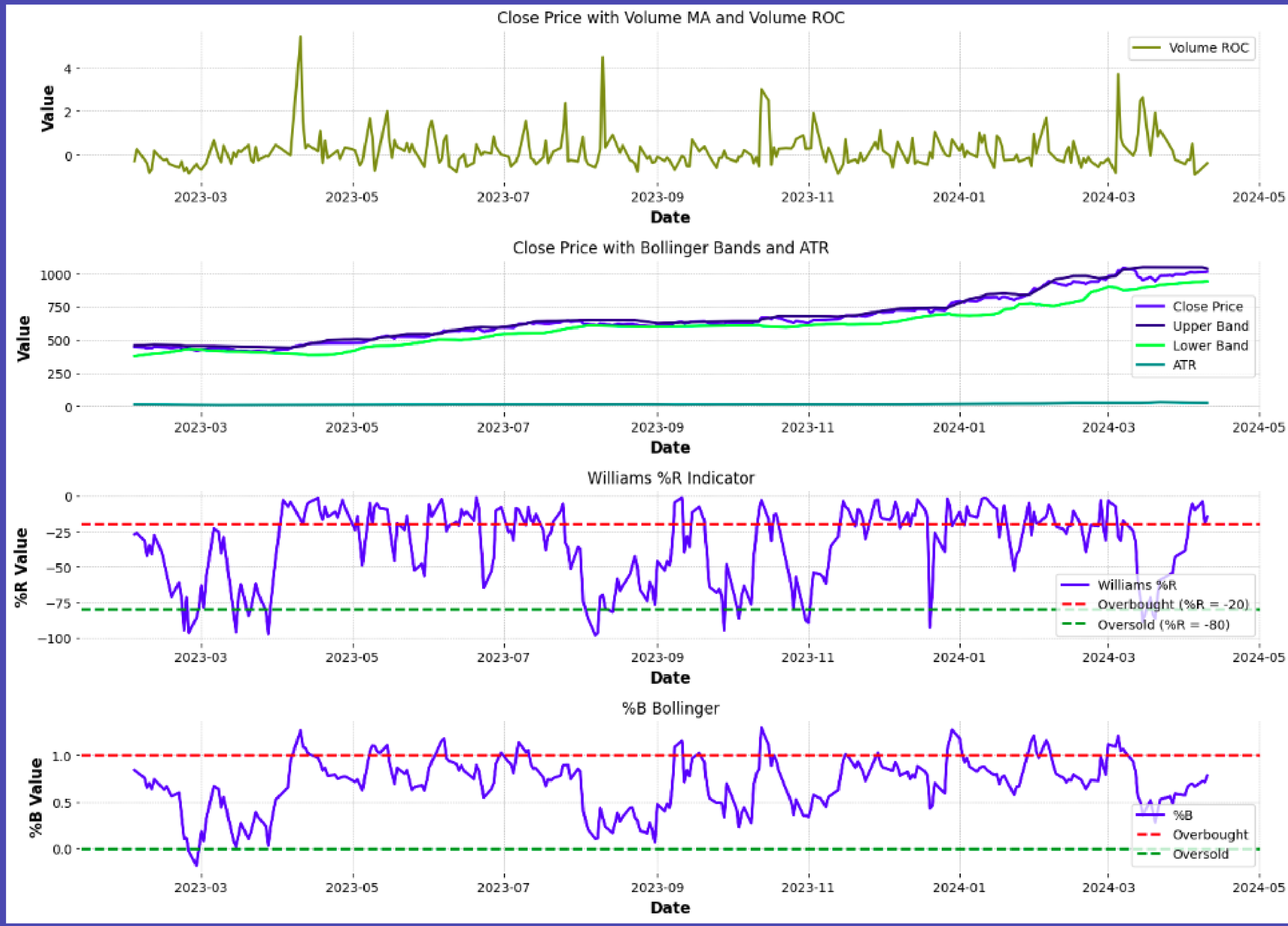
$$\%D = \text{3-day SMA of \%K}$$

$$\text{Volume MA} = \frac{\text{Sum of Trading Volume over 'n' periods}}{n}$$

$$\text{ATR} = \text{Average of (True Range over 'n' periods)}$$

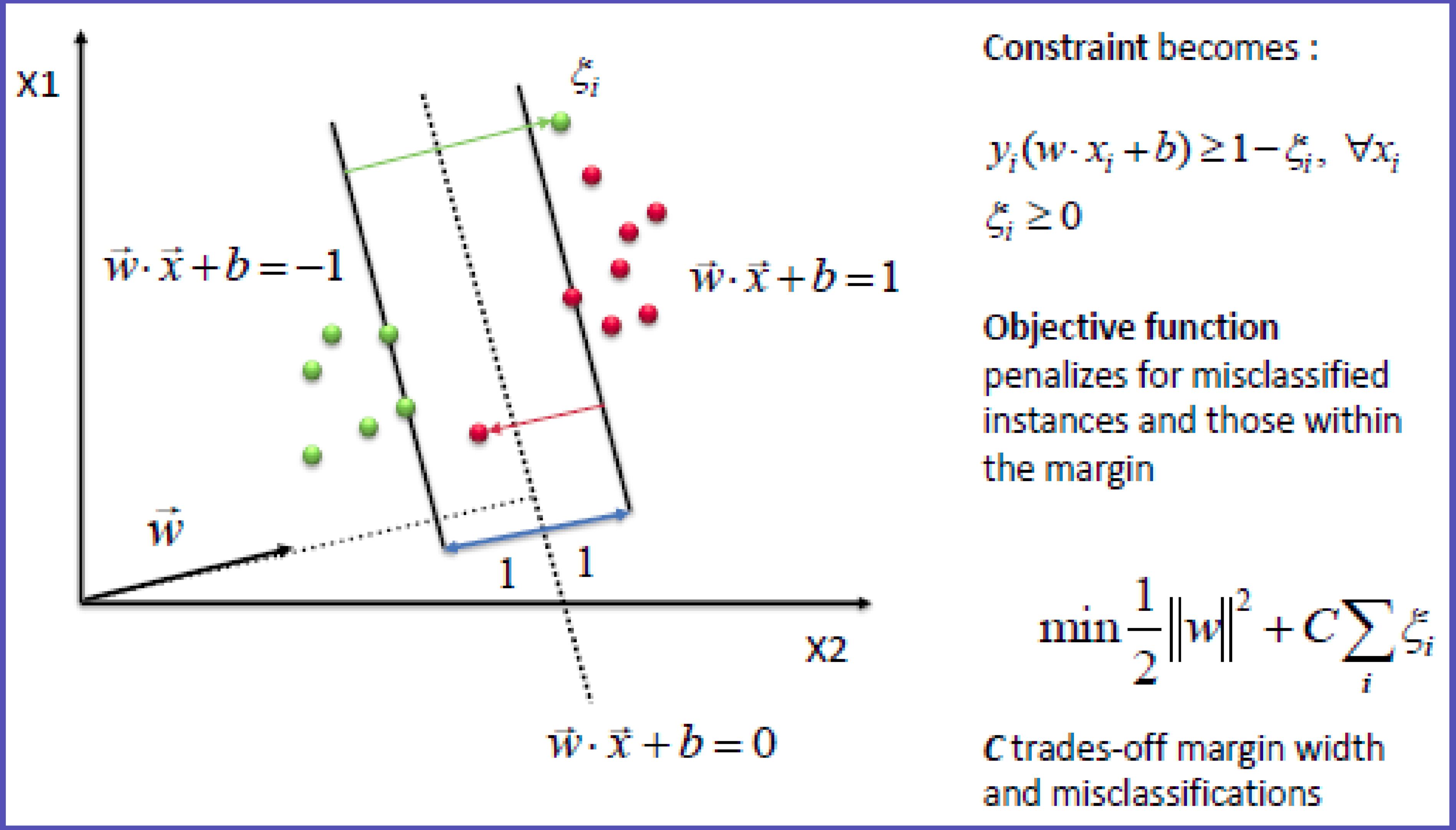
$$\text{VolumeROC} = \left(\frac{\text{Current Volume} - \text{Volume 'n' periods ago}}{\text{Volume 'n' periods ago}}\right) \times 100$$

Graphs of Indicators used as feature

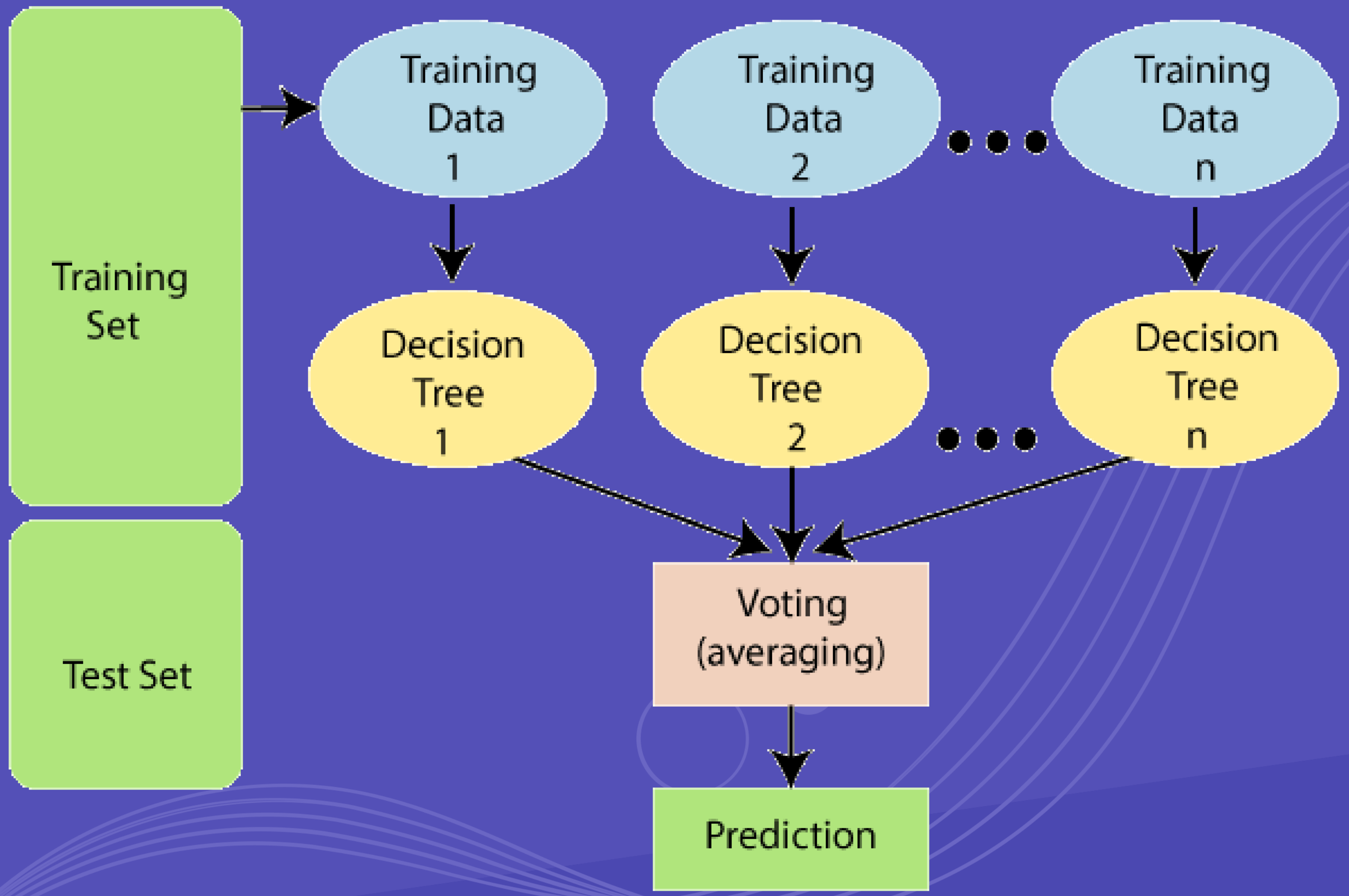


MACHINE LEARNING MODELS:

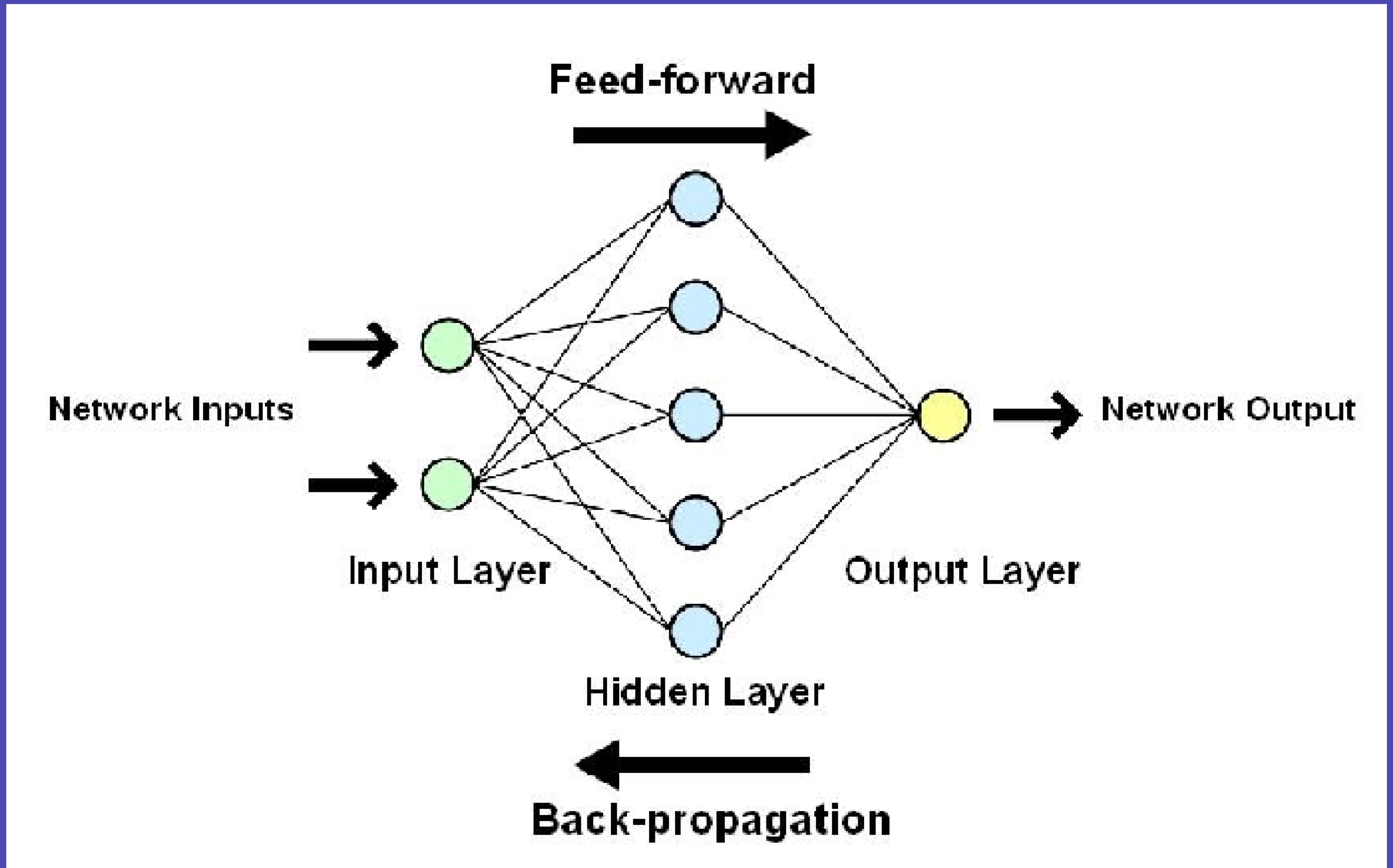
1.SVM

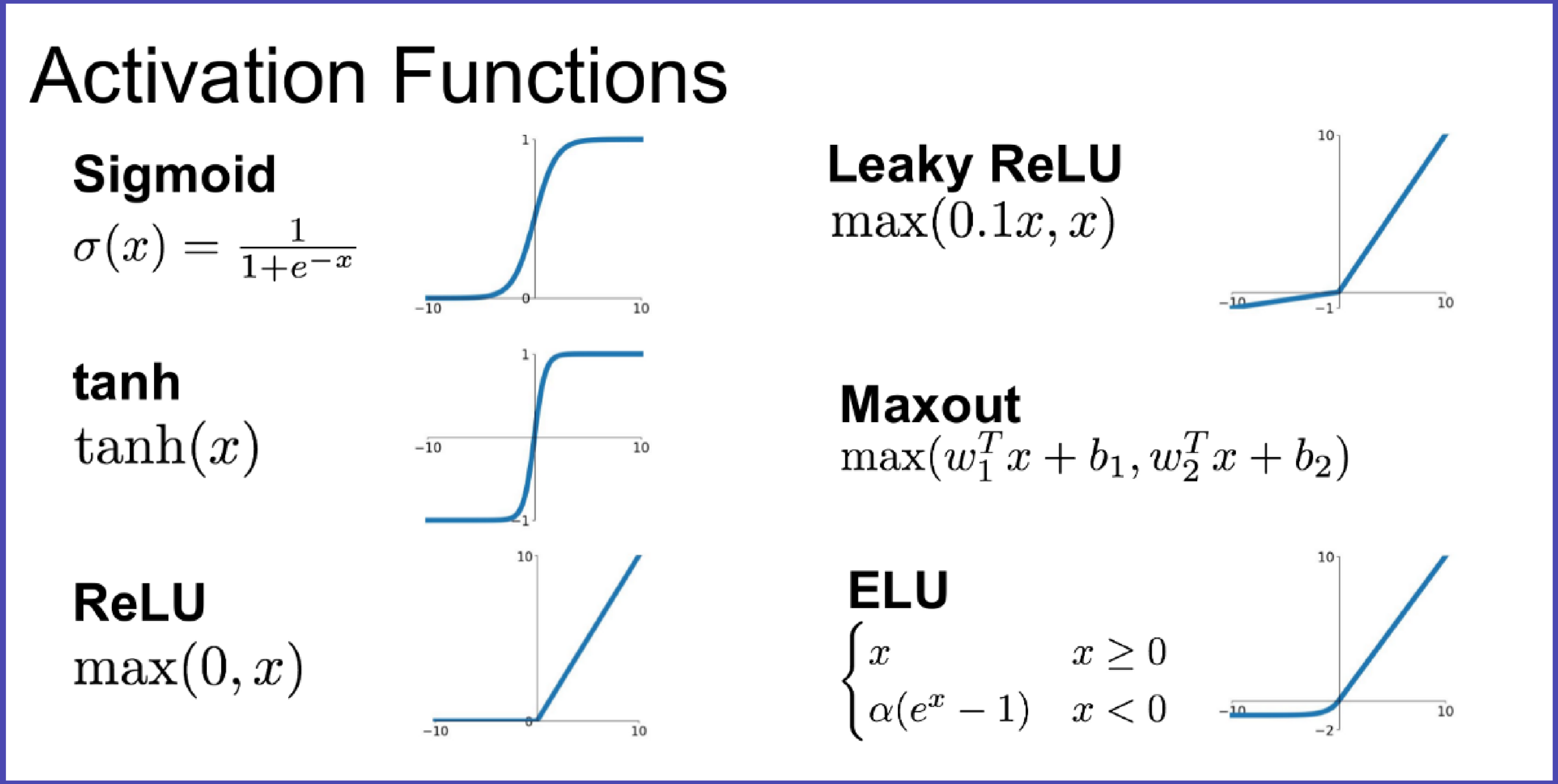


2. Random Forest



3. ANN


Feed-forward
Network Inputs
Network Output
Input Layer
Output Layer
Hidden Layer
Back-propagation

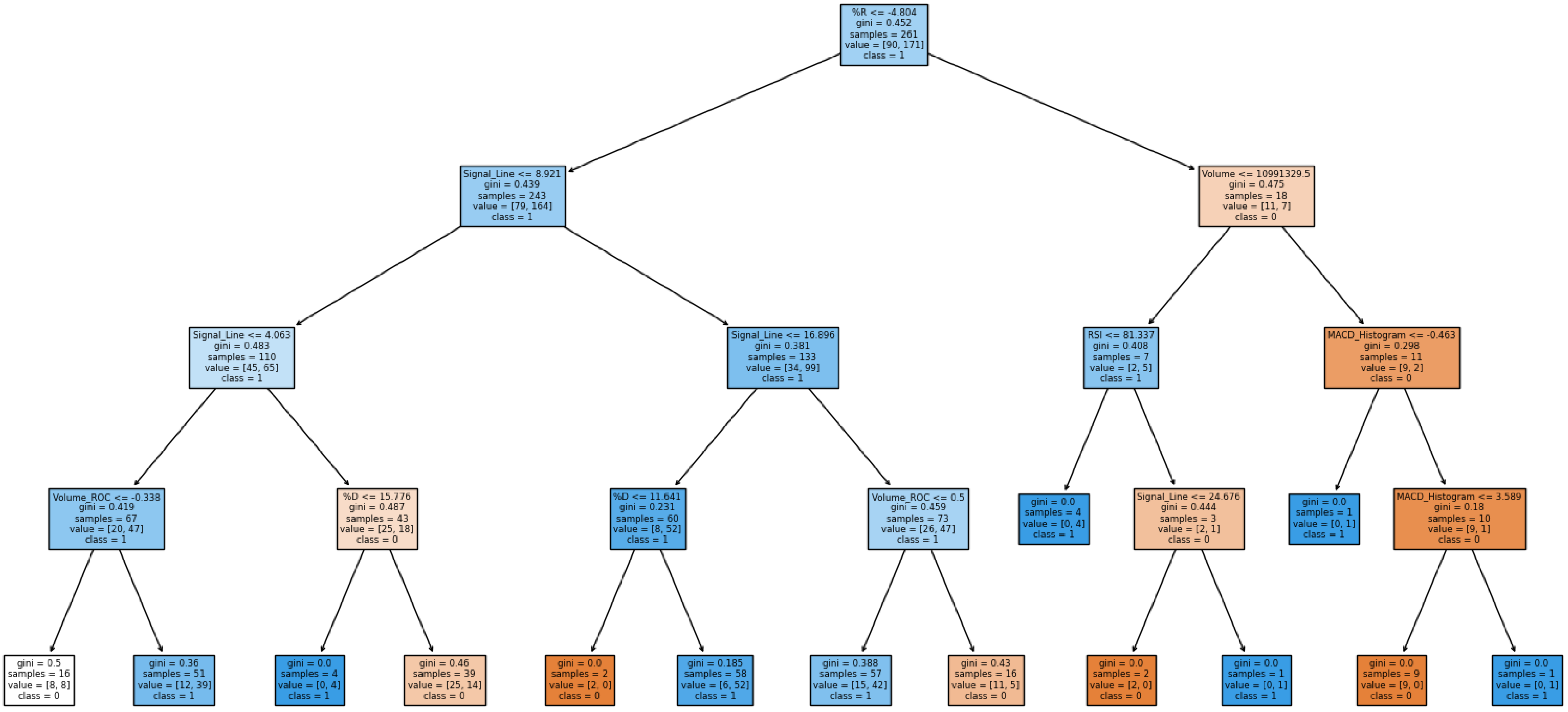

Activation Functions
Sigmoid
$\sigma(x) = \frac{1}{1+e^{-x}}$
tanh
$\tanh(x)$
ReLU
$\max(0, x)$
Leaky ReLU
$\max(0.1x, x)$
Maxout
$\max(w_1^T x + b_1, w_2^T x + b_2)$
ELU
$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$

COMPARISON OF RESULTS:

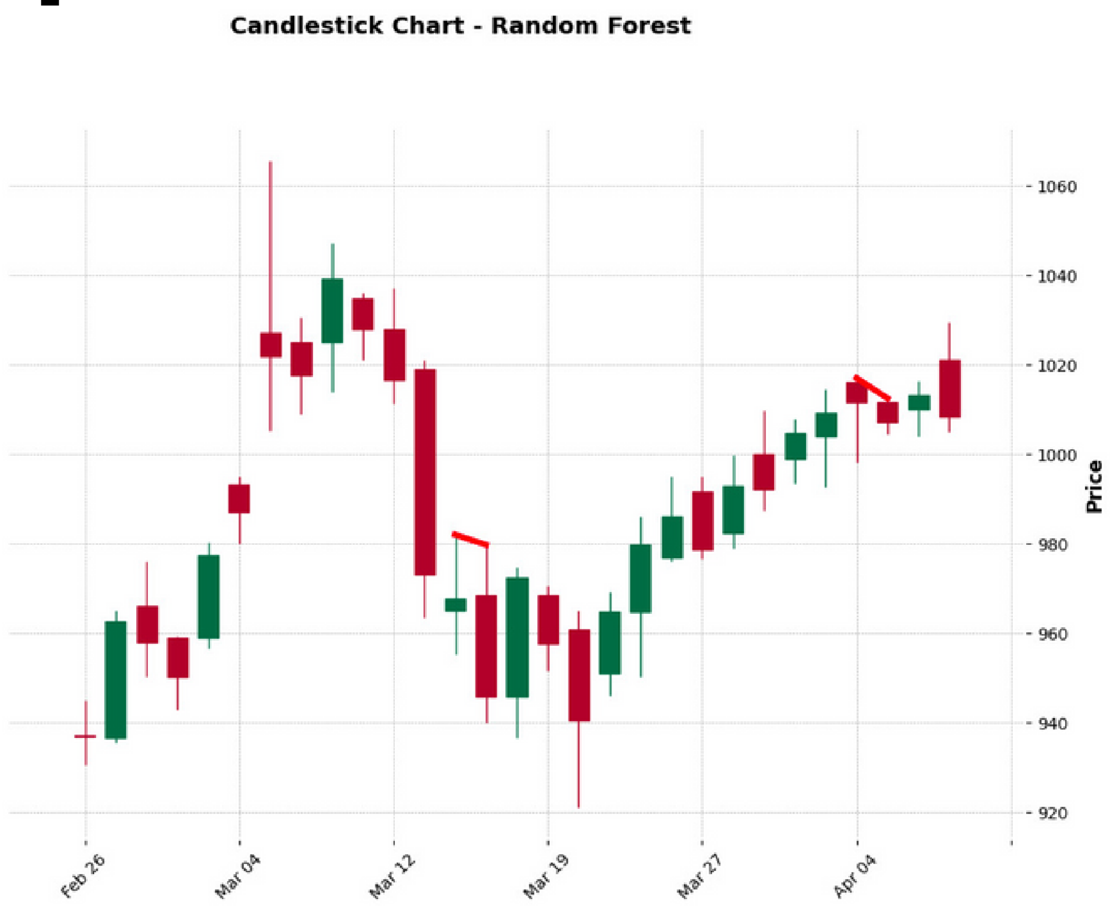
- The Random Forest's superior accuracy may stem from its capability to handle obscure relationships between features and target variables.
- Unlike classifiers such as KNN and linear SVM, Random Forests are less sensitive to data scaling, contributing to their robust performance in noisy environments.

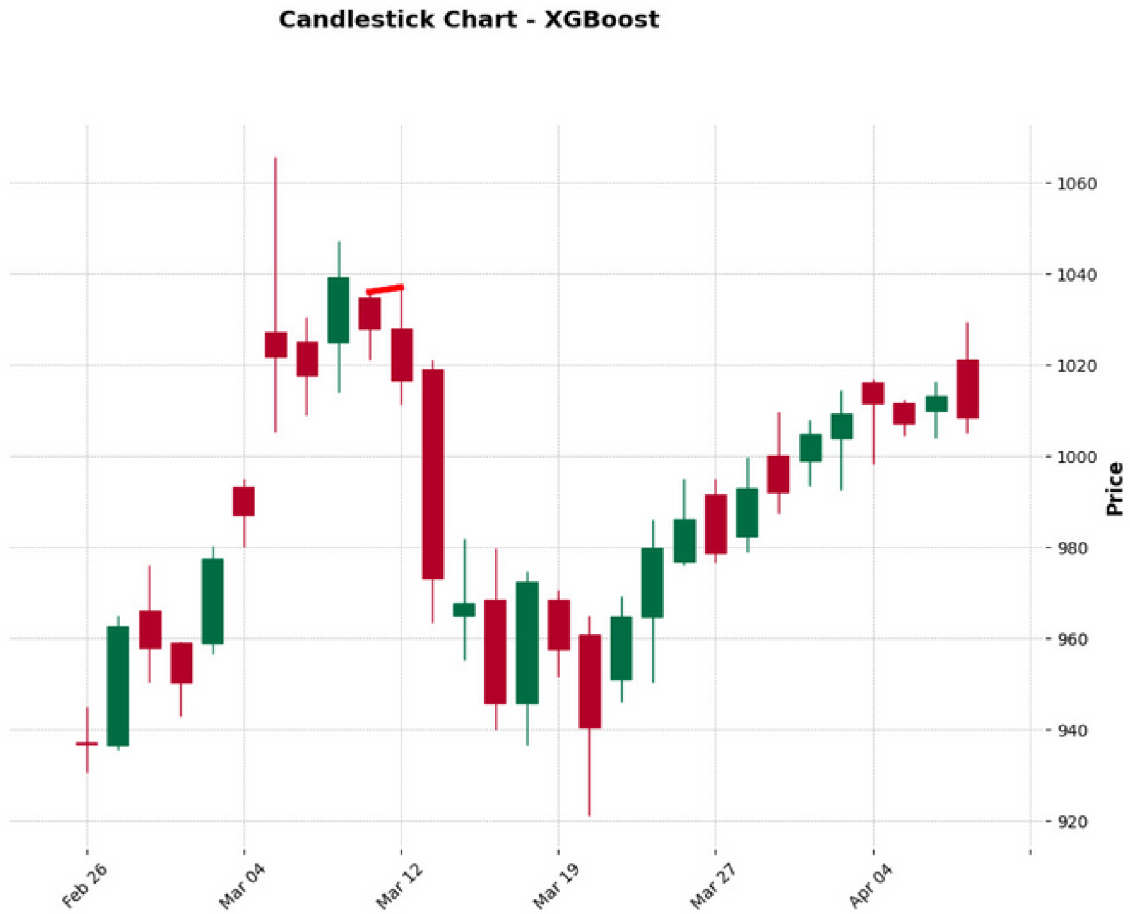
Classifier	Accuracy	Precision	Recall	F1-score
KNN	0.6206896551724138	0.7913618948101706	0.6206896551724138	0.678321304464047
SVM (poly Kernel)	0.6206896551724138	1.0	0.6206896551724138	0.7659574468085107
SVM (RBF Kernel)	0.6206896551724138	1.0	0.6206896551724138	0.7659574468085107
Decision Tree	0.6896551724137931	0.6959247648902821	0.6896551724137931	0.685141065830721
ANN	0.4827586206896552	0.47614071752002785	0.4827586206896552	0.47890340544013715
Bayesian Learning	0.6551724137931034	0.8655520724486242	0.6551724137931034	0.7223466188983431
Perceptron	0.5172413793103449	0.5444096133751305	0.5172413793103449	0.5286751361161525
Regression	0.6206896551724138	1.0	0.6206896551724138	0.7659574468085107
Random Forest	0.7931034482758621	0.7931034482758621	0.7931034482758621	0.7931034482758621
XGBoost	0.8275862068965517	0.8594566353187042	0.8275862068965517	0.8345665224068127
Gradient Boosting	0.7586206896551724	0.7941483803552768	0.7586206896551724	0.7683931313695378

Decision Tree With Depth- 5



Some of the misclassified predictions





Conclusion

- Our study demonstrates the feasibility of using machine learning techniques and technical indicators to predict stock market trends with reasonable accuracy.

- Random Forest emerged as the best ML model for predicting as it can handle complex non linear pattern for predicting.

- Future research endeavors may focus on refining existing models, exploring ensemble learning techniques, and integrating alternative data sources such as sentiment analysis from news and social media feeds to improve prediction accuracy and reliability.

Link To Github

THANK YOU FOR WATCHING!