STOCK PRICE PREDICTION USING

MACHINE LEARNING

GROUP 11 :

INTRODUCTION

Stock markets are dynamic, complex, and influenced by a wide range of factors — from global events and company performance to investor sentiment. Predicting stock prices accurately is a huge challenge, especially because the data is highly non-linear and often unpredictable.

In recent years, Machine Learning has emerged as a powerful tool to tackle this problem. Unlike traditional methods, ML can analyze large volumes of historical stock data and detect patterns that aren't obvious to the human eye. By learning from past trends, it helps us make more informed predictions about future stock prices.

MOTIVATION

Why Predict Stock Prices?

For Investors:

- Make smarter and more confident buy/sell decisions
- Reduce risk of unexpected losses by forecasting market movements
- Better timing of market entries and exits to maximize returns
- Gain a competitive edge in a highly volatile market

For Researchers:

- Evaluate whether AI and Machine Learning can outperform traditional statistical methods
- Understand which algorithms are more accurate, efficient, and stable in real market conditions
- Analyze the practical usefulness of technical indicators like moving averages and volatility measures
- Contribute to the growing body of research in financial forecasting using modern data-driven techniques

The Real Problem

While stock price prediction has been widely explored, most existing studies focus on only one machine learning method and often lack proper comparisons across models. Many also overlook the impact of creating new technical features like moving averages or volatility indicators. This study aims to fill that gap by comparing two powerful ML algorithms — Artificial Neural Network (ANN) and Random Forest (RF) — on the same real-world dataset, using newly generated features and objective performance metrics like RMSE and MAPE to determine which model predicts stock closing prices more accurately.

OBJECTIVES

1. To build Machine Learning models that can predict the next day's stock closing price.

2. To compare the performance of Artificial Neural Network (ANN) and Random Forest (RF) models for stock price prediction.

3. To generate new input variables (like moving averages, high-low difference, etc.) from stock market data to improve prediction accuracy.

4. To evaluate the models using metrics like RMSE (Root Mean Squared Error), MAPE (Mean Absolute Percentage Error), and MBE (Mean Bias Error).

5. To analyze and identify which model gives more accurate and reliable results.

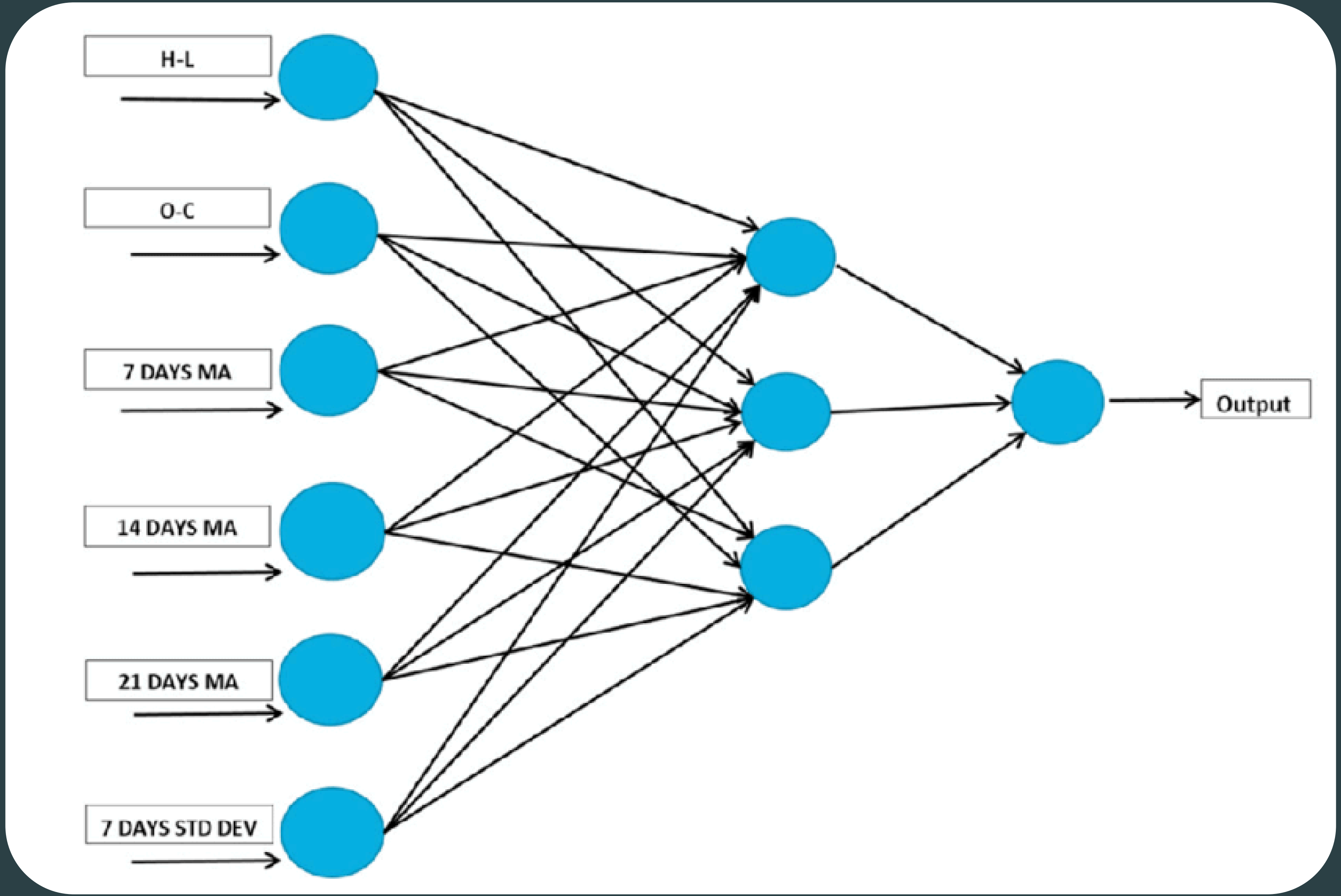
RESEARCH METHODOLOGY – OVERVIEW

What This Study Did:

- Collected 10 years of daily stock data (2009–2019) for five companies across different sectors: Nike, Pfizer, Johnson & Johnson, JP Morgan, and Goldman Sachs.
- Engineered new technical features from basic stock data like open, high, low, close, and volume.
- Built and trained two machine learning models:
  - Artificial Neural Network (ANN)
  - Random Forest (RF)
- Compared both models by evaluating how accurately they predict the next day's closing stock price using performance metrics like RMSE, MAPE, and MBE.

MODEL 1 - ARTIFICIAL NEURAL NETWORK (ANN)

- ANN is a type of machine learning model inspired by the human brain.
- It uses layers of "neurons" to process input data and learn patterns over time.
- Architecture used in this study:
  - Input Layer with 7 features
  - Hidden Layer that processes the inputs using weighted connections
  - Output Layer that gives the predicted closing price
- It’s particularly good at capturing non-linear relationships in data, which is important because stock markets are not always predictable or linear.

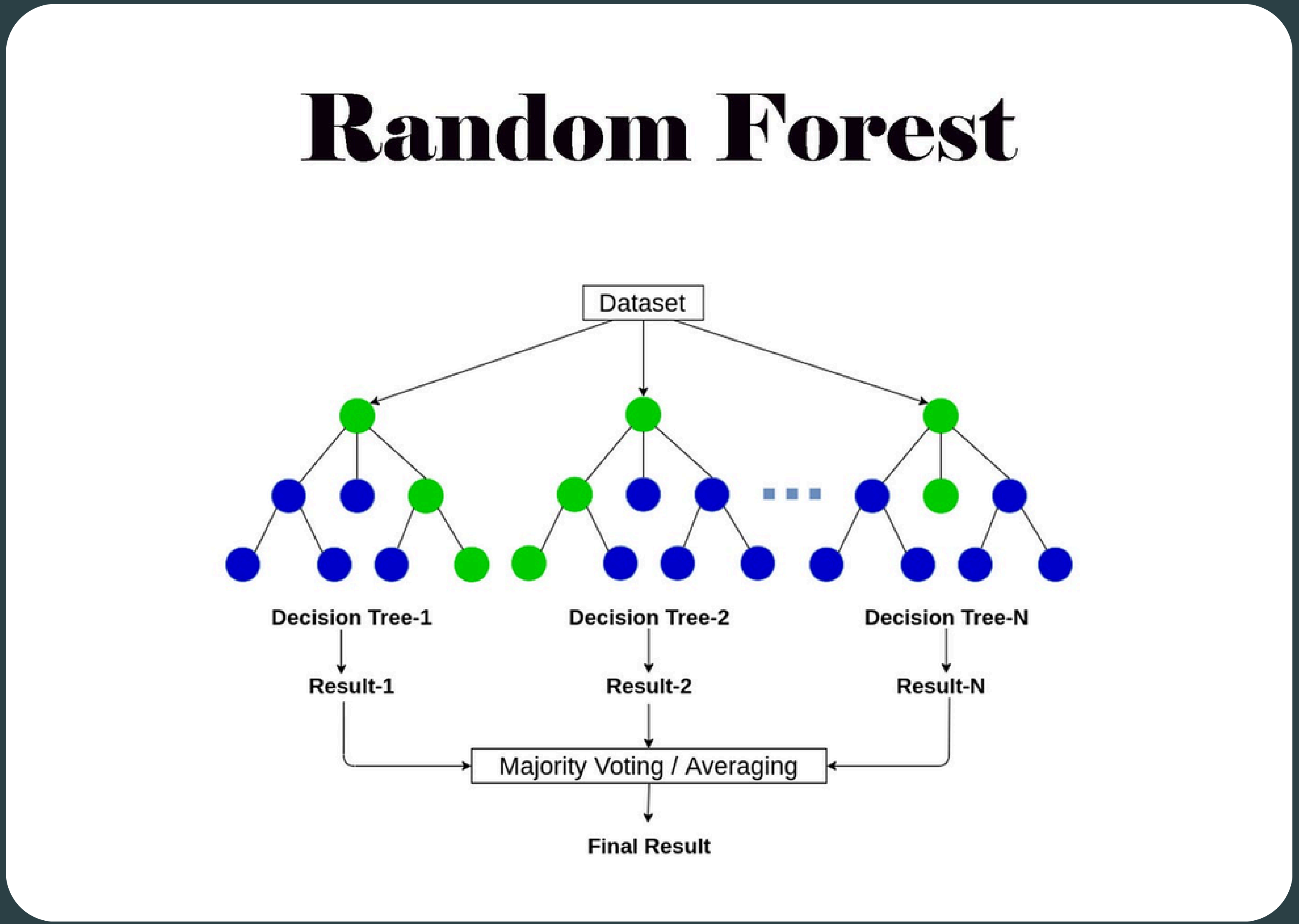


Why ANN?

Because it learns complex, hidden patterns in large datasets and adapts over time.

MODEL 2 –
RANDOM FOREST (RF)

- Random Forest is an ensemble learning method made up of many decision trees.
- Each tree is trained on a random subset of the data and makes its own prediction.
- The final prediction is the average of all the trees' outputs.
- It's highly effective for regression tasks like stock prediction because it can handle large, noisy datasets and reduces overfitting compared to single decision trees.



Why RF?
It's reliable, fast, and gives stable predictions even with messy financial data.

FEATURE ENGINEERING – INPUTS USED

From the raw data (Open, High, Low, Close, Volume), the study created 7 new features to improve model performance:

FEATURE	EXPLANATION
High - Low (H-L)	Measures how much the stock fluctuated that day
Close - Open (O-C)	Shows whether the price rose or fell during the day
7-Day Moving Average	Short-term trend of closing prices
14-Day Moving Average	Medium-term trend indicator
21-Day Moving Average	Long-term price trend
7-Day Standard Deviation	Measures recent price volatility
Volume	Total number of shares traded on a given day

These inputs give the model better context about market behavior and price movement trends.

EVALUATION METRICS USED

RMSE (Root Mean Squared Error)

- Measures average error in predicted price (in actual currency units like ₹ or $)
- Sensitive to large errors (penalizes big mistakes)
- Formula:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Predicted_i - Actual_i)^2}{N}}$$

- N: Total number of predictions (data points)
- Oi: Actual (original) closing price on day i
- Fi: Forecasted (predicted) closing price on day i
- (Oi-Fi): Square of the error for that day

- RMSE: The square root of the average of squared errors → gives the error in original price units
- Lower RMSE = Better prediction accuracy

MAPE (Mean Absolute Percentage Error)

- Measures average percentage error between predicted and actual values
- Makes it easier to compare across different stocks (standardized)
- Formula:

$$MAPE = \frac{1}{n} \times \sum \left| \frac{actual\ value\ - forecast\ value}{actual\ value} \right|$$

- n: Total number of predictions
- Oi: Actual closing price on day i
- Fi: Predicted closing price on day i
- (Oi-Fi)\Oi: Percentage error for that day

- MAPE: Average of the absolute percentage errors across all days (expressed as a %)
- Lower MAPE (<5%) is considered excellent in finance

MBE (Mean Bias Error)

- Measures whether the model tends to overpredict or underpredict
- Helps understand if predictions are biased in a specific direction
- Formula:

$$\text{MBE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)$$

- n: Number of predictions
- yi: Actual closing price
- y^i: Predicted closing price
- (yi-y^i): Positive if model underpredicts, negative if it overpredicts

- MBE: Shows average direction of error (not absolute)
- MBE > 0 → underprediction; MBE < 0 → overprediction

THANK YOU