STOCK PREDICTION

Project Mentor :
Prof. Arnab Chakraborty

TEAM MEMBERS :

- AYUSH KUMAR
- ADITYA RAJ
- ADITYA ANAND
- SHUBHAM KUMAR RAY
- DEBOSHREE PRAMANIK

CONTENT

PROJECT OBJECTIVE & SCOPE

Project Objectives

- Predict future stock prices using historical market data.
- Analyze and compare the performance of various machine learning models.
- Build a user-friendly web application for interactive forecasting.
- Demonstrate the effectiveness of ML in financial time-series analysis.

Goals

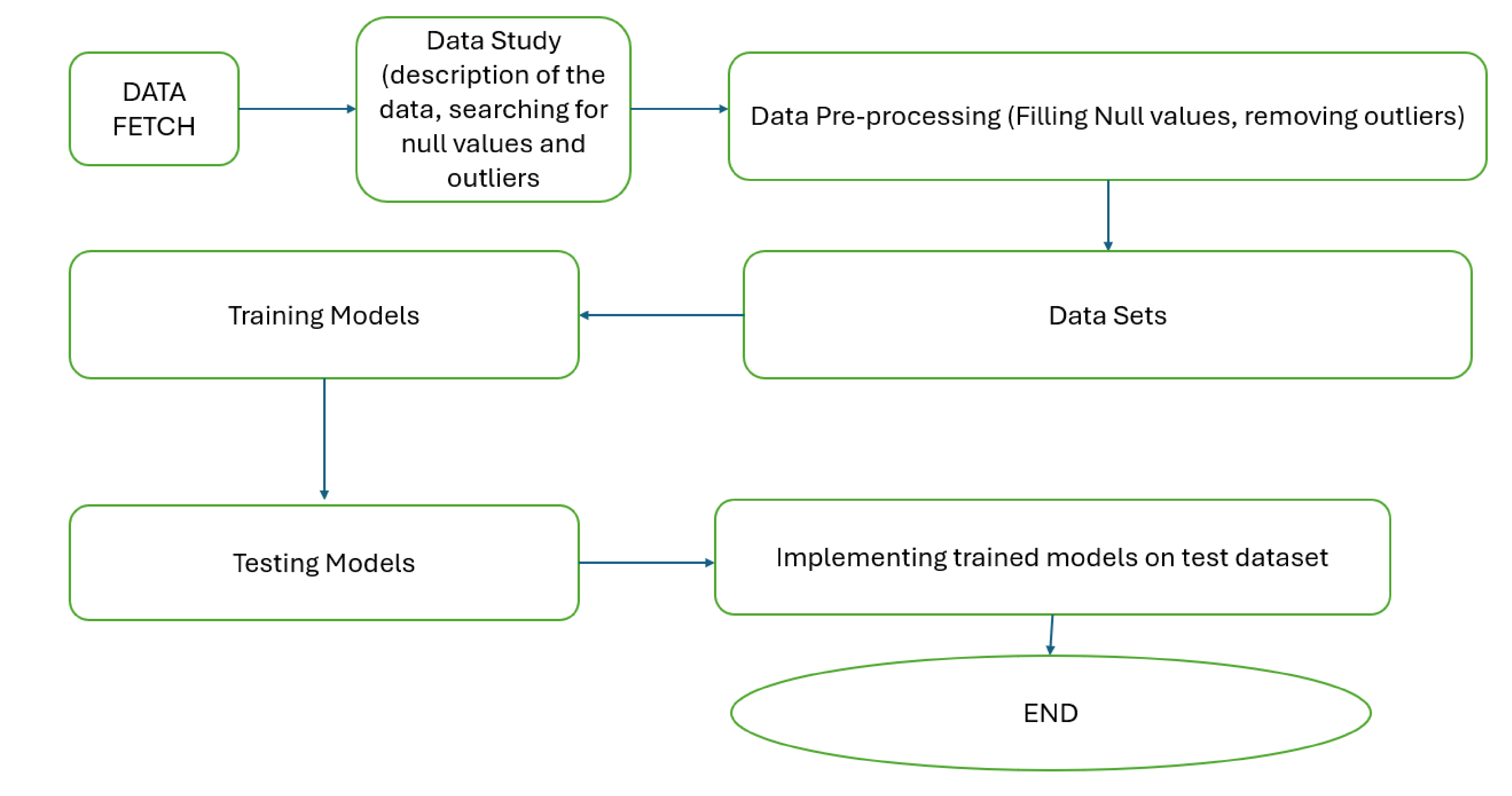
- Develop accurate ML models to forecast stock prices using historical data and technical indicators.
- Compare model performance (e.g., Linear Regression, Random Forest, LSTM) using metrics like RMSE and R^2 to identify the most reliable approach.
- Deploy an interactive web app

DATA DESCRIPTION

The description of the data with type and description of each Attribute is given/shown in the below.

Field	Description
Open	Price at market open on given day
High	Highest price during the trading day
Low	Lowest price during the trading day
Close	Price at market close (or adjusted close when adjusted)
Volume	Number of shares traded during the day
Adj Close	Close price adjusted for splits and dividends (if used; often renamed to Close)

METHODOLOGY



DATA PREPROCESSING

Term	Definition
Moving Average (MA_n)	Average of closing prices over the last *n* days (e.g. MA_5 = 5-day moving average).
Volatility_20	Standard deviation of daily returns over the previous 20 days.
RSI (Relative Strength Index)	Momentum indicator that measures upward vs downward movements over a period (usually 14 days); identifies overbought / oversold conditions.
MACD (Moving Average Convergence Divergence)	Difference between two exponential moving averages (EMAs), typically 12-day and 26-day, plus signal line and histogram.
Bollinger Bands	Volatility bands placed above and below a moving average, usually at ±2 standard deviations.
ADX (Average Directional Index)	Measures the strength of a trend (regardless of direction).
OBV (On Balance Volume)	A volume-based indicator that relates volume flow to price change.
CCI (Commodity Channel Index)	Measures the difference between a security's typical price and its moving average, scaled by mean deviation.
Stochastic Oscillator (K, D)	Compares a security's closing price to its price range over a recent period, giving %K and smoothed %D lines.
Lagged Features	Features from prior days (e.g. lag1, lag2, etc.), used for capturing temporal autocorrelation.
Classification Model	Machine learning model that predicts categories (Up vs Down), not continuous values.
True Positive (TP)	A case where the model predicted "Up" and the stock actually went Up.
False Positive (FP)	Model predicted "Up" but actual was "Down".
True Negative (TN)	Model predicted "Down" and actual was "Down".
False Negative (FN)	Model predicted "Down" but actual was "Up".

MODELS USED

THE MACHINE LEARNING MODELS USED FOR THIS PROJECT ARE:

- LOGISTIC REGRESSION ACCURACY
- RANDOM FOREST ACCURACY
- GRADIENT BOOSTING ACCURACY
- K-NEAREST NEIGHBORS ACCURACY
- SVC ACCURACY

LOGISTIC REGRESSION ACCURACY

- Type: Binary classification (e.g., price up or down)
- Simple and interpretable
- Fast to train and test
- Limitations:
- Assumes linear relationships

Logistic Regression Accuracy: 1.0000

Classification Report:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	128
1	1.00	1.00	1.00	149
accuracy			1.00	277
macro avg	1.00	1.00	1.00	277
weighted avg	1.00	1.00	1.00	277

Confusion Matrix:

0	1
128	0
0	149

RANDOM FOREST ACCURACY

- Type: Ensemble classifier or regressor

Strengths:
- Handles non-linear data well
- Robust to noise and overfitting
- Provides feature importance

Limitations:
- Can be slower with large datasets
- Less interpretable than simpler models

Random Forest Accuracy: 1.0000

Classification Report:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	128
1	1.00	1.00	1.00	149
accuracy			1.00	277
macro avg	1.00	1.00	1.00	277
weighted avg	1.00	1.00	1.00	277

Confusion Matrix:

GRADIENT BOOSTING ACCURACY

Type: Ensemble boosting method

Strengths:

- High predictive power
- Good at capturing complex patterns
- Performs well on imbalanced datasets

Limitations:

- Sensitive to hyperparameters
- Longer training time

Gradient Boosting Accuracy: 1.0000

Classification Report:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	128
1	1.00	1.00	1.00	149
accuracy			1.00	277
macro avg	1.00	1.00	1.00	277
weighted avg	1.00	1.00	1.00	277

Confusion Matrix:

K-NEAREST NEIGHBORS ACCURACY

- Type: Instance-based learning

Strengths:

- Simple and intuitive
- No training phase

Limitations:

- Computationally expensive for large datasets
- Sensitive to irrelevant features and scaling

K-Nearest Neighbors Accuracy: 0.7870

Classification Report:

	precision	recall	f1-score	support
0	0.71	0.92	0.80	128
1	0.91	0.67	0.77	149
accuracy			0.79	277
macro avg	0.81	0.80	0.79	277
weighted avg	0.82	0.79	0.79	277

Confusion Matrix:

RESULT

- Multiple machine learning models were trained to predict stock price movement direction (up/dow using historical data and technical indicators.
- The models achieved the following accuracies:
- Gradient Boosting delivered the highest accuracy, effectively capturing complex patterns in the data.
- Random Forest also performed well, offering robust predictions with minimal tuning.
- Logistic Regression, while simple and fast, showed lower accuracy due to its linear assumptions.
- KNN provided decent results but was sensitive to feature scaling and less efficient with larger datasets.
- Visual comparisons of actual vs predicted trends confirmed that ensemble models (Random Forest and Gradient Boosting) were more reliable for short-term forecasting.

CONCLUSION

- Machine learning models can effectively predict stock price movements when trained on well-preprocessed historical data and engineered features.
- Among the models tested, Gradient Boosting achieved the highest accuracy, followed closely by Random Forest, demonstrating their strength in capturing complex market patterns.
- Simpler models like Logistic Regression and KNN provided baseline performance but lacked the precision needed for volatile financial data.
- The deployment of a Streamlit-based web app makes the prediction tool accessible and interactive, allowing users to explore forecasts in real time.
- This project highlights the potential of ML in financial forecasting and sets the foundation for future enhancements like sentiment analysis and real-time data integration.

THANK YOU