USE OF MACHINE LEARNING FOR AUTOMATED TRADING

Gaurav Jain

Vihaan Thora

AGENDA

- Introduction
- Dataset
- Model
- Algorithm
- Back Testing
- Further Scope

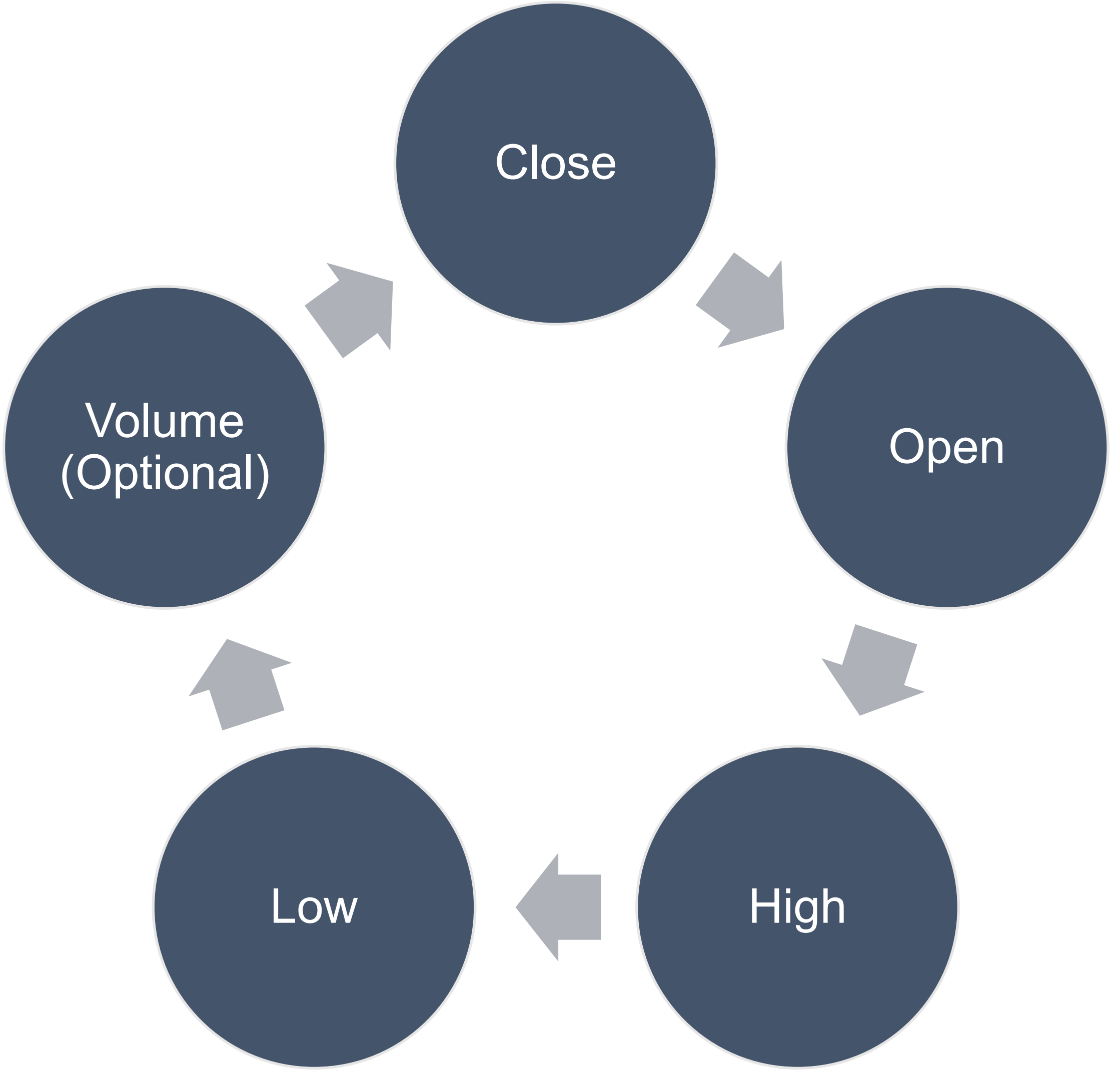
INTRODUCTION

Our project focuses on algorithmic trading through machine learning. We have used various algorithms such as Random Forest, Extra Trees Classifier, and Gradient Boosters to predict price trends. The technical analyzers have been used as the features of the dataset to train these models. Our aim is to achieve accurate predictions for profitable trading decisions.

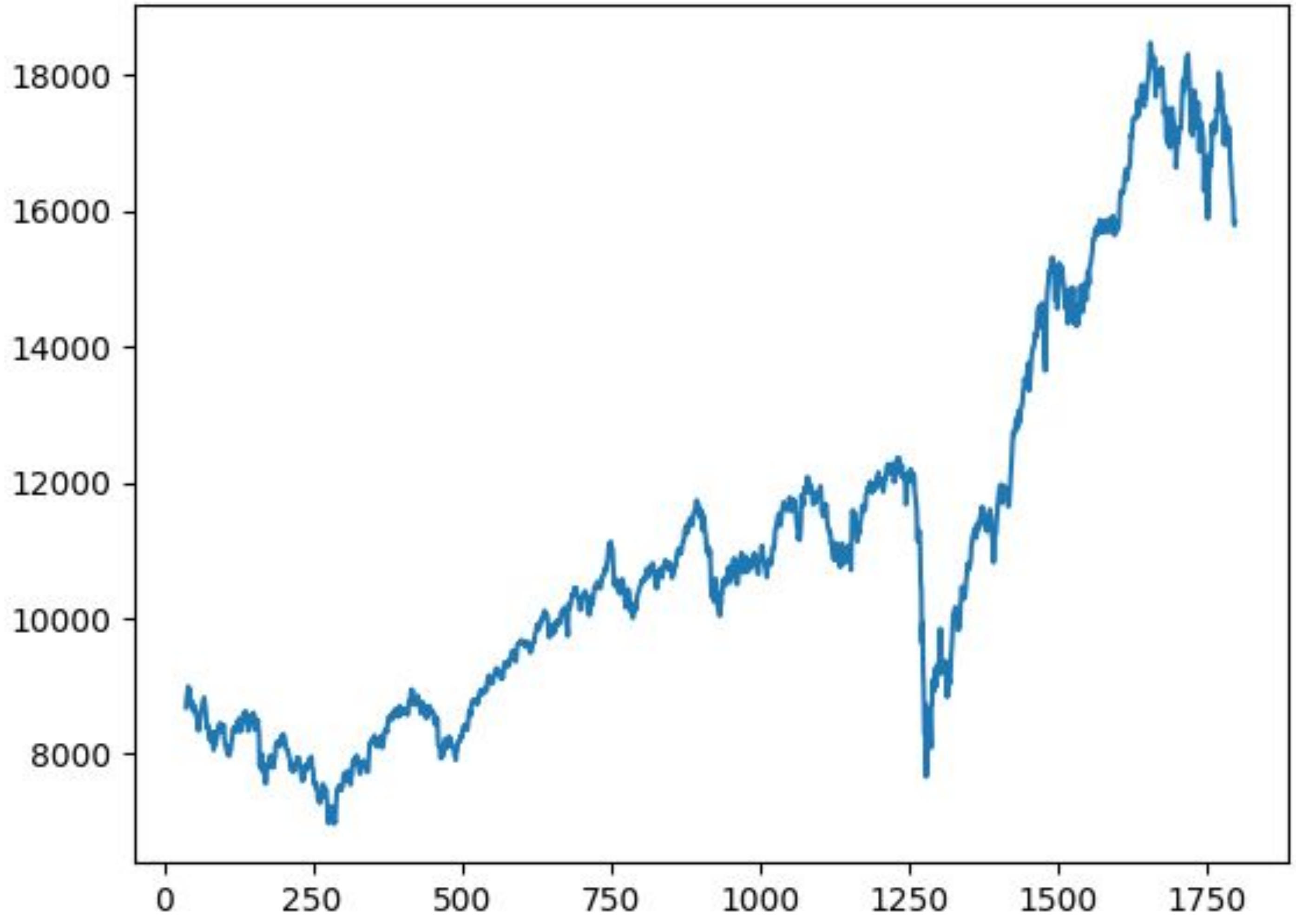
DATASET

NIFTY 50 and more...

INPUT DATA FEATURES



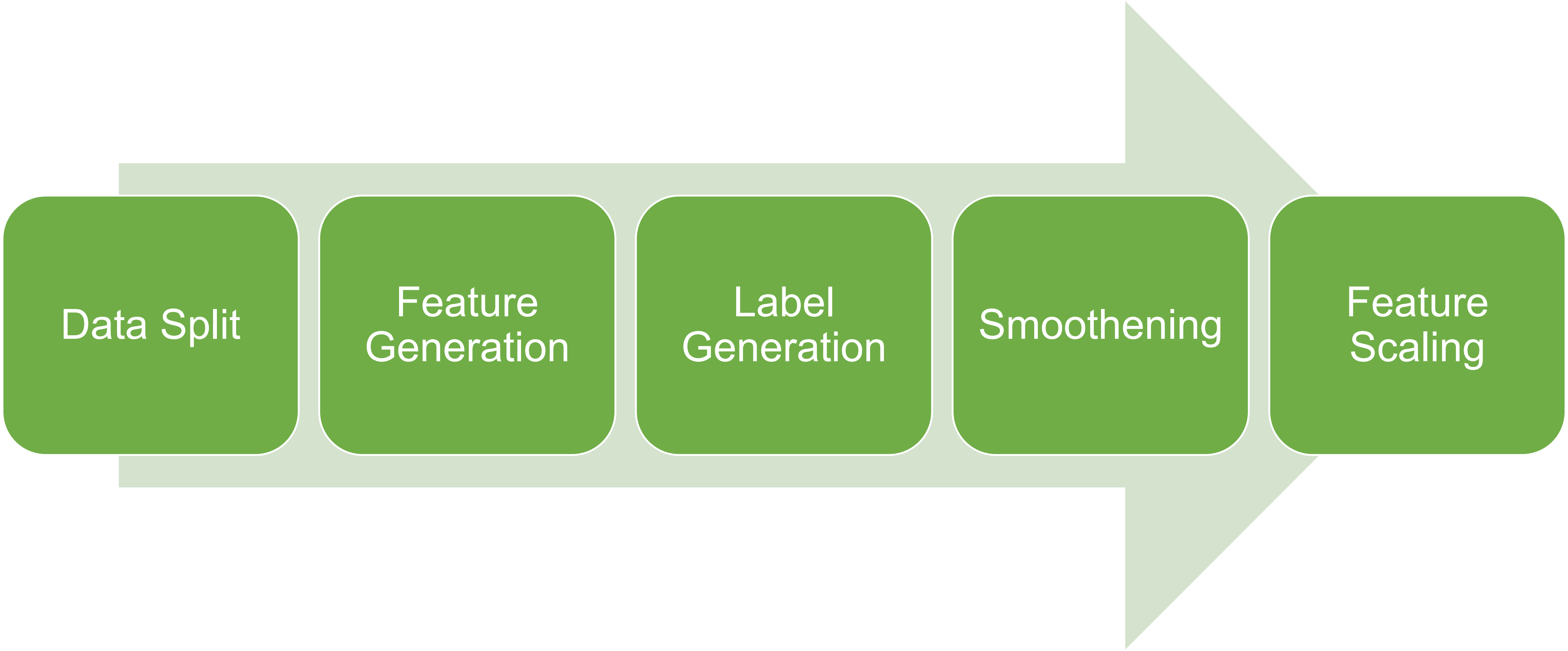
NIFTY 50 DATA (CLOSE VS DAYS)



SYNTHESIZED DATA FEATURES

- Simple Moving Averages (SMA)
- Exponential Moving Averages (EMA)
- Bollinger Bands
- Moving Average Convergence Divergence (MACD)
- Rate of Change (ROC)
- Relative Strength Index (RSI)
- Stochastic Oscillators
  - STOCHK
  - STOCHD
- Average True Range (ATR)
- Kaufman's Adaptive Moving Average (KAMA)
- Vortex Indicators
  - VIm
  - VIp

DATA PIPELINE



MODEL

Random Forest & Extra Trees Classifier

RANDOM FOREST CLASSIFIER

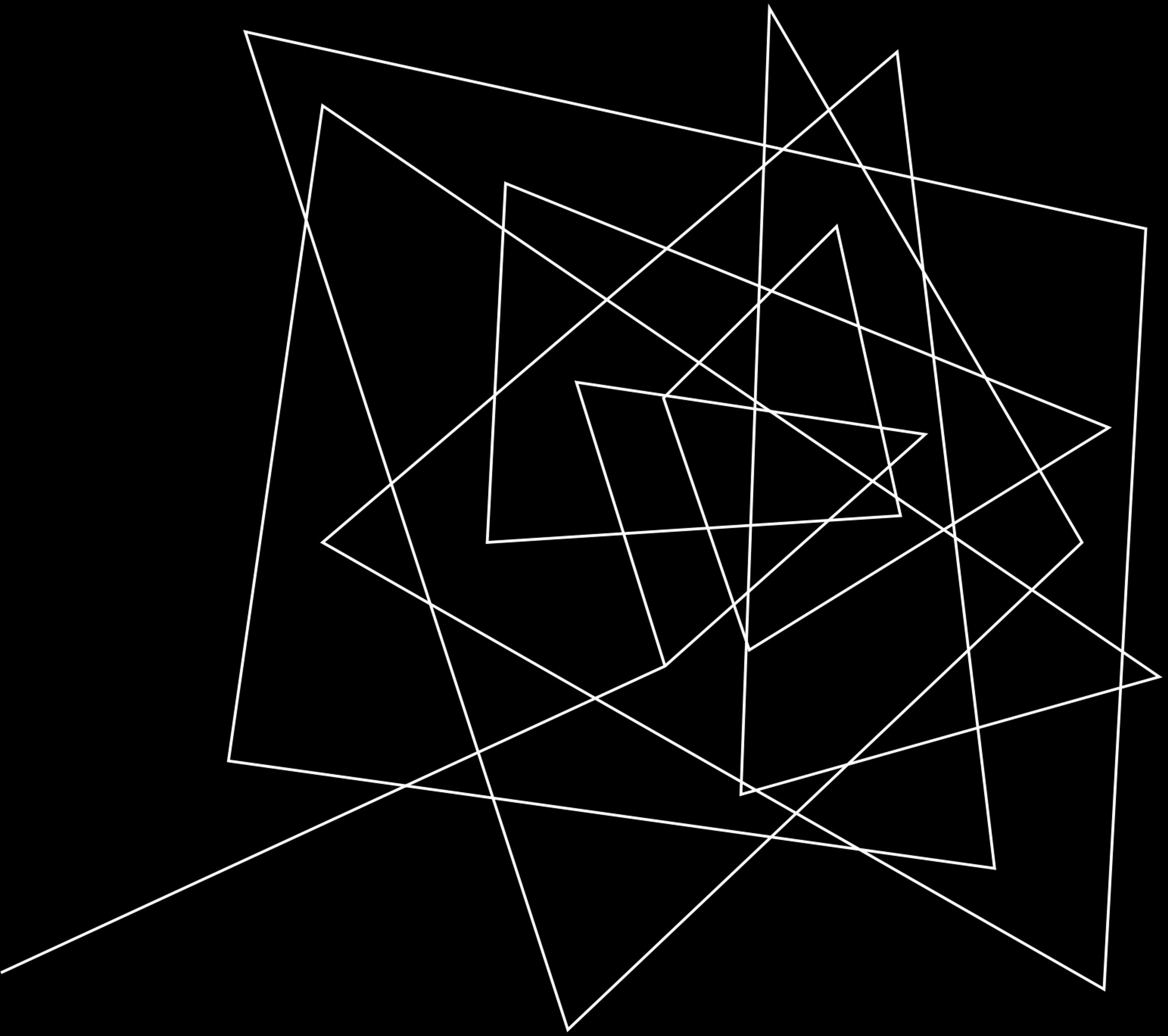
Overview	Creating the individual decision trees	Making predictions	Evaluating the model
Supervised learning algorithm used for classification tasks. An ensemble method that creates multiple decision trees and combines their predictions to obtain a final prediction.	Creates multiple decision trees by selecting a random subset of features and a random subset of training data for each tree. Each tree created using a different random subset of features and training data. The trees are grown using recursive binary splitting, to find the feature and threshold that best separates the data.	Individual decision trees used to make predictions. A new data point is passed through each decision tree in the forest, and each tree makes a prediction. The predictions from all trees combined to obtain a final prediction. This can be done using a simple majority vote.	Evaluation can be done using metrics such as accuracy, precision, recall, and F1 score. Cross-validation ensures model is not overfitting to the training data. Feature importance can be calculated to understand which features are most important for making accurate predictions.

EXTRA TREES CLASSIFIER

Decision Trees	Ensemble of Trees	Extra Trees
Randomly select subset of features. Pick best split using subset of data.	Train multiple decision trees. Combine to make predictions.	Randomly select features and split points. Train multiple trees with randomized splits. Combine to make predictions

PERFORMANCE METRICS

Returns	The **Annual Return** measures the total percentage change in an investment's value over a year, including capital gains and losses. It is calculated by dividing the current value of the investment by its original value, subtracting 1, and multiplying by 100 to get the percentage return.
Sharpe Ratio	The **Sharpe Ratio** is a measure of risk-adjusted return. It is calculated by subtracting the risk-free rate of return from the investment's average return and dividing the result by the standard deviation of the investment's return. The higher the Sharpe ratio, the better the risk-adjusted performance of the investment.
Maximum Drawdown	**Maximum Drawdown** is the maximum percentage decline in an investment's value from a previous high. It is a measure of the investment's risk and is used to calculate the Calmar ratio.
Calmar Ratio	The **Calmar Ratio** measures risk-adjusted return that compares an investment's average annual return to its maximum drawdown, the largest percentage decline in its value from a previous high. The higher the Calmar ratio, the better the risk-adjusted performance of the investment.



ALGORITHM

The Strategy

TRAINING

- We have used a sliding window approach to training and testing that enables a fair and accurate fitting result.
- The default window size taken is of 40 days reflects two financial months of the dataset.
- The default lookahead of the labels is taken as 10 days (2 weeks), which means that our model will be trained to predict whether the stock will go up or down after two weeks.

Training accuracy results -

Random Forest	Extra Trees	Gradient Boosting	Bagging	Stacking Ensemble
66.9%	69.9%	63%	61%	67.8%

TRADING STRATEGY - INPUTS

TRAINED MODEL

TEST DATASET

PRINCIPAL AMOUNT

TRADING WINDOW

MAXIMUM VOLUME

INITIAL CONFIDENC
E

CONFIDENC
E RANK

TRADING STRATEGY

- We have currently used the Extra Trees Classifier's trained model as an input to our algorithm.
- The dataset that is passed is unlabeled but contains all the synthesized features.
- The principal amount denotes the amount of money available to purchase the shares. The trading window size denotes the days after which the predicted label is obeyed.
- Maximum volume is set such that any major fault in the price would not lead to a large drawdown value.
- Confidence is an indicator that determines the volume of the stock that needs to be traded in a position.
- Confidence is updated according to the following rule:

```
confidence =  (1-alpha)*confidence + alpha*result
```

  (where result signifies if the current prediction was correct or not)
- The max shares that can be traded on a day is determined by the rule:

```
max_shares = volume * confidence^n
```

- Hence the number of shares that can be bought or sold is determined by the balance available and the above value.

BACK TESTING

The Results

BACK TESTING

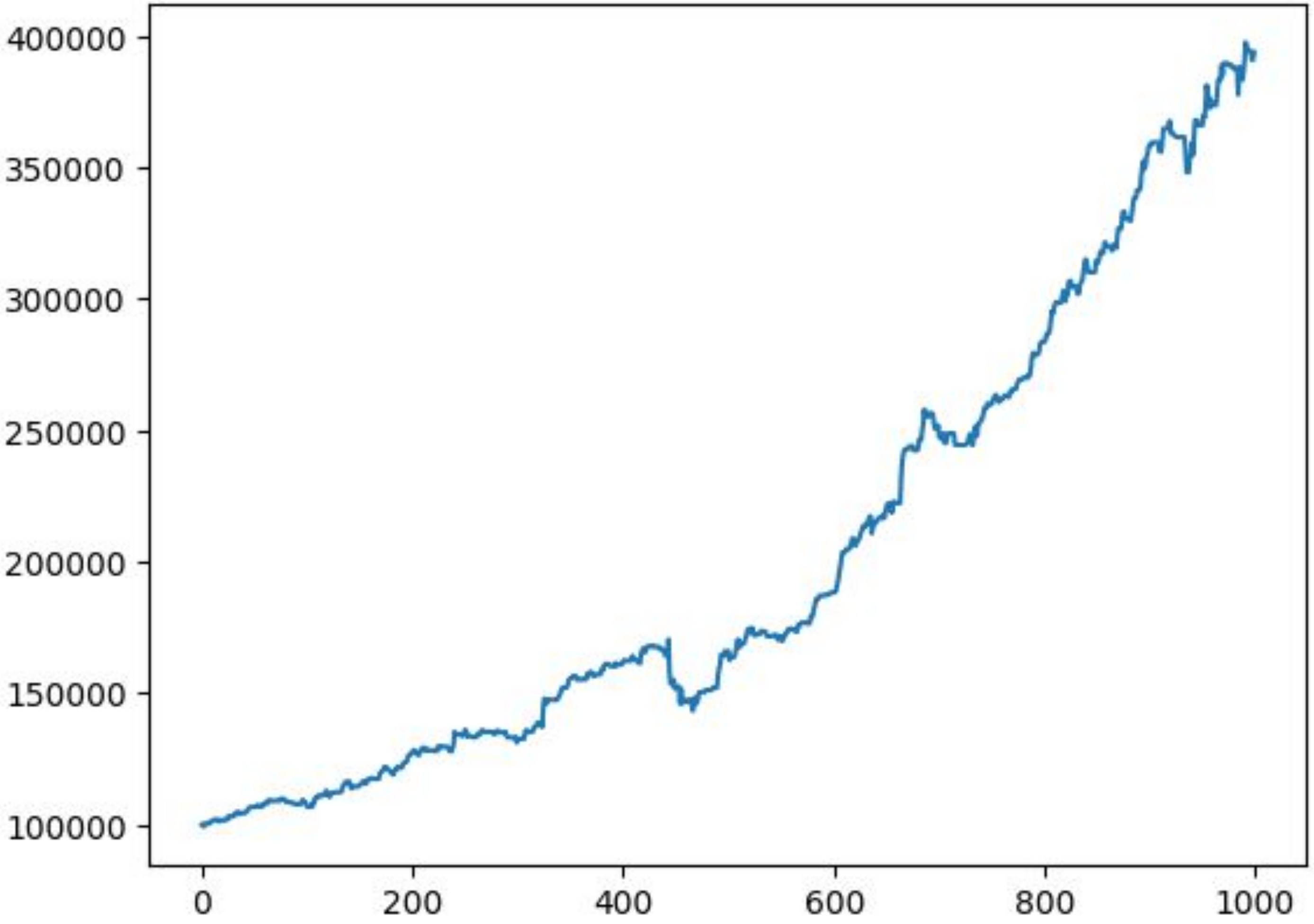
- We back test our strategy on **four years** of unseen data, and calculate the annual return produced. The training is performed on 18 years of data (2000-18).
- Example test case: Amount = INR **I00000**, Max Volume = **70**, Window = **5** days, Confidence Rank = **0.7,** n = **2.4**, Model = Extra Trees Classifier
- We take a risk-free rate of **15%** while calculating Sharpe Ratio as it denotes the average increment in NIFTY 50 price annually.

BACK TESTING RESULTS

Back tested over 2018-22 NIFTY 50 unseen data

Annual Returns	Cumulative Returns	Sharpe Ratio	Max Drawdown
41.26%	293.85%	1.72	-15.87%

PNL CHART GENERATED



FURTHER SCOPE

ML and Finance Perspective

ML PERSPECTIVE

Feature correlation	Backwards elimination	Feature normalization
Identify the degree of correlation between features. Eliminate redundant features to avoid multicollinearity.	Iteratively eliminate features to reduce model complexity. Based on the least significant feature with p-value threshold.	Scale features to a common range for comparison. Avoid bias towards features with higher magnitudes.

FINANCIAL PERSPECTIVE

Neutralization	Shorting	Volume Function
Take a neutral position with respect to a sector This will diversify the portfolio and cut losses in an unexpected market event.	Short stocks in order to improve returns and Sharpe.	Calculate share volume using polynomial regression. Polynomial takes confidence as argument for resource allocation. Grid search used to optimize polynomial coefficients

REFERENCES

- https://towardsdatascience.com/predicting-future-stock-market-trends-with-python-machine-learning-2bf3f1633b3c
- https://towardsdatascience.com/deep-reinforcement-learning-for-automated-stock-trading-f1dad0126a02
- https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html
- https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.ExtraTreesClassifier.html
- https://blog.quantinsti.com/mean-reversion-time-series/
- https://irep.ntu.ac.uk/id/eprint/32787/1/PubSub10294_702a_McGinnity.pdf

THANK YOU