Stock Market Trading using Machine Learning

Course Code: EE769

Indian Institute of Technology Bombay

Submitted By:

Harsh Maheshwari (173050015)

Kedar Anavardekar (17305R006)

CONTENTS

INTRODUCTION

- Trading is the process of buying and selling of financial instruments.
- Stock market
  - market for the trading
  - one of the most important sources for companies to raise money
  - allows businesses to go public, or raise additional capital for expansion

MOTIVATION

- Predicting stock performance is a very large and profitable area of study
- Many companies have developed stock predictors based on machine learning algorithms
- This technique may prove useful in aiding the decisions of investors
- Can give an edge to beginning investors who don’t have a lifetime of experience

PROBLEM STATEMENT

The project is primarily concerned with predicting future returns of the stock market by applying various machine learning algorithms such as MLPRegressor and Adaboost and performing their comparative analysis.

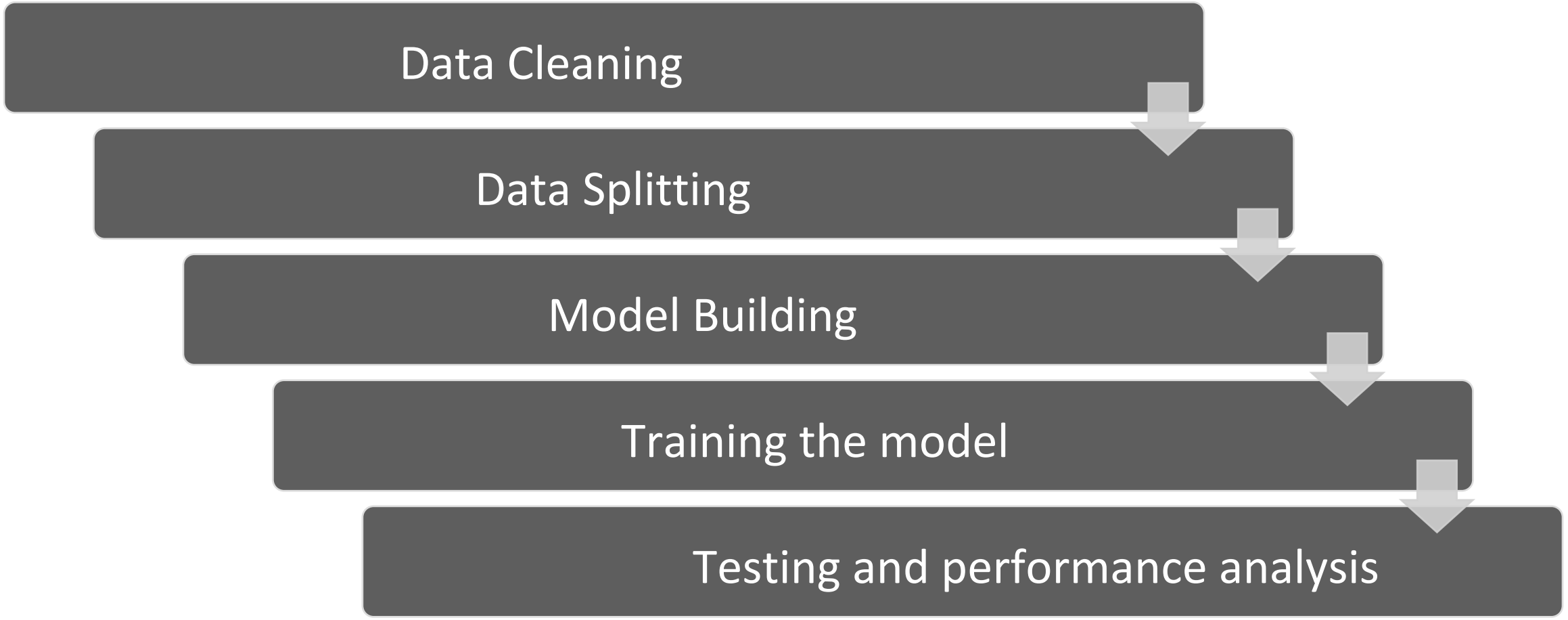
DATA SET (Winton Stock Market Challenge)

- **train.csv - the training set, including the columns of:**
  - Feature_1 - Feature_25
  - Ret_MinusTwo, Ret_MinusOne
  - Ret_2 - Ret_120
  - Ret_121 - Ret_180: target variables
  - Ret_PlusOne, Ret_PlusTwo: target variables
  - Weight_Intraday, Weight_Daily
- **test.csv - the test set, including the columns of:**
  - Feature_1 - Feature_25
  - Ret_MinusTwo, Ret_MinusOne
  - Ret_2 - Ret_120

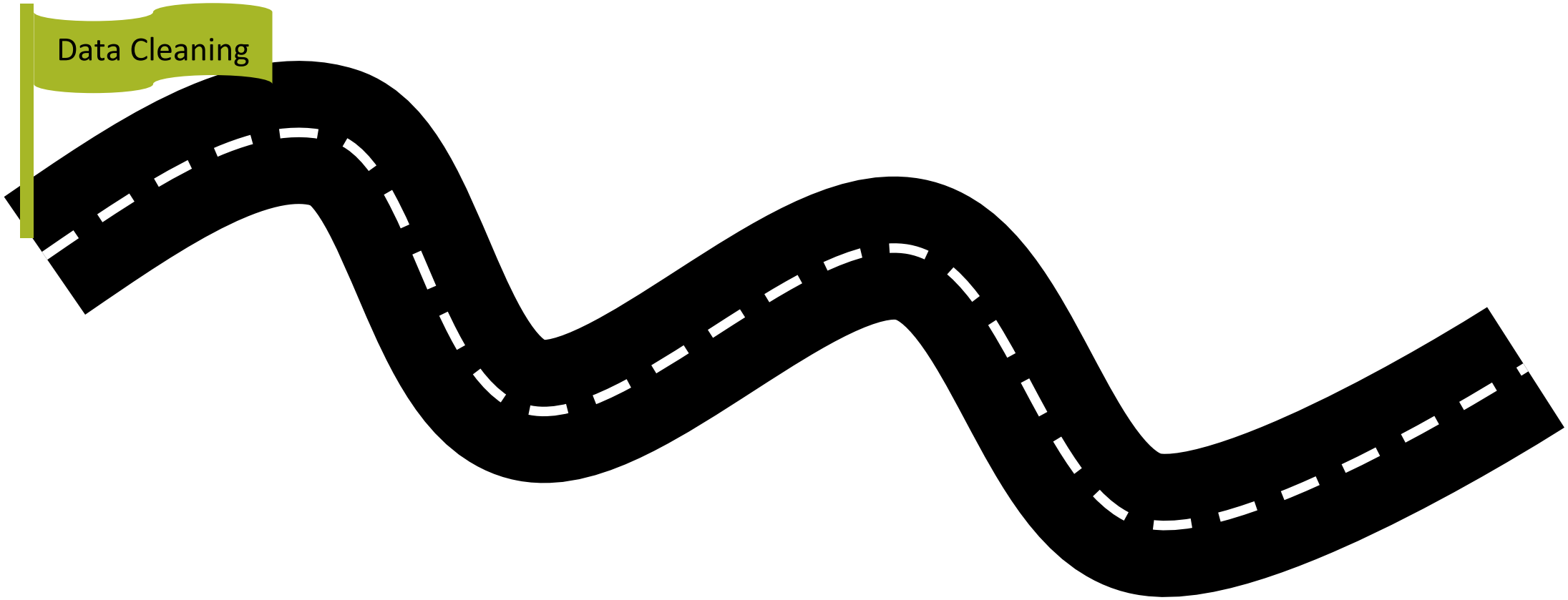
DATA SET (Winton Stock Market Challenge)

- **Feature_1 to Feature_25**: different features relevant to prediction.
- **Ret_MinusTwo**: this is the return from the close of trading on day D-2 to the close of trading on day D-1 (i.e. 1 day).
- **Ret_MinusOne:** this is the return from the close of trading on day D-1 to the point at which the intraday returns start on day D (approximately 1/2 day).
- **Ret_2 to Ret_120:** these are returns over approximately one minute on day D. Ret_2 is the return between t=1 and t=2.
- **Ret_121 to Ret_180:** intraday returns over approximately one minute on day D. These are the target variables needed to be predicted as {id}_{1-60}.
- **Ret_PlusOne:** this is the return from the time Ret_180 is measured on day D to the close of trading on day D+1. (Approximately 1day). This is a target variable needed to be predicted as {id}_61.
- **Ret_PlusTwo:** this is the return from the close of trading on day D+1 to the close of trading on day D+2 (i.e. 1 day). This is a target variable needed to be predicted as {id}_62.
- **Weight_Intraday:** weight used to evaluate intraday return predictions Ret 121 to 180.
- **Weight_Daily:** weight used to evaluate daily return predictions (Ret_PlusOne and Ret_PlusTwo)

PROCEDURE


Data Cleaning
Data Splitting
Model Building
Training the model
Testing and performance analysis

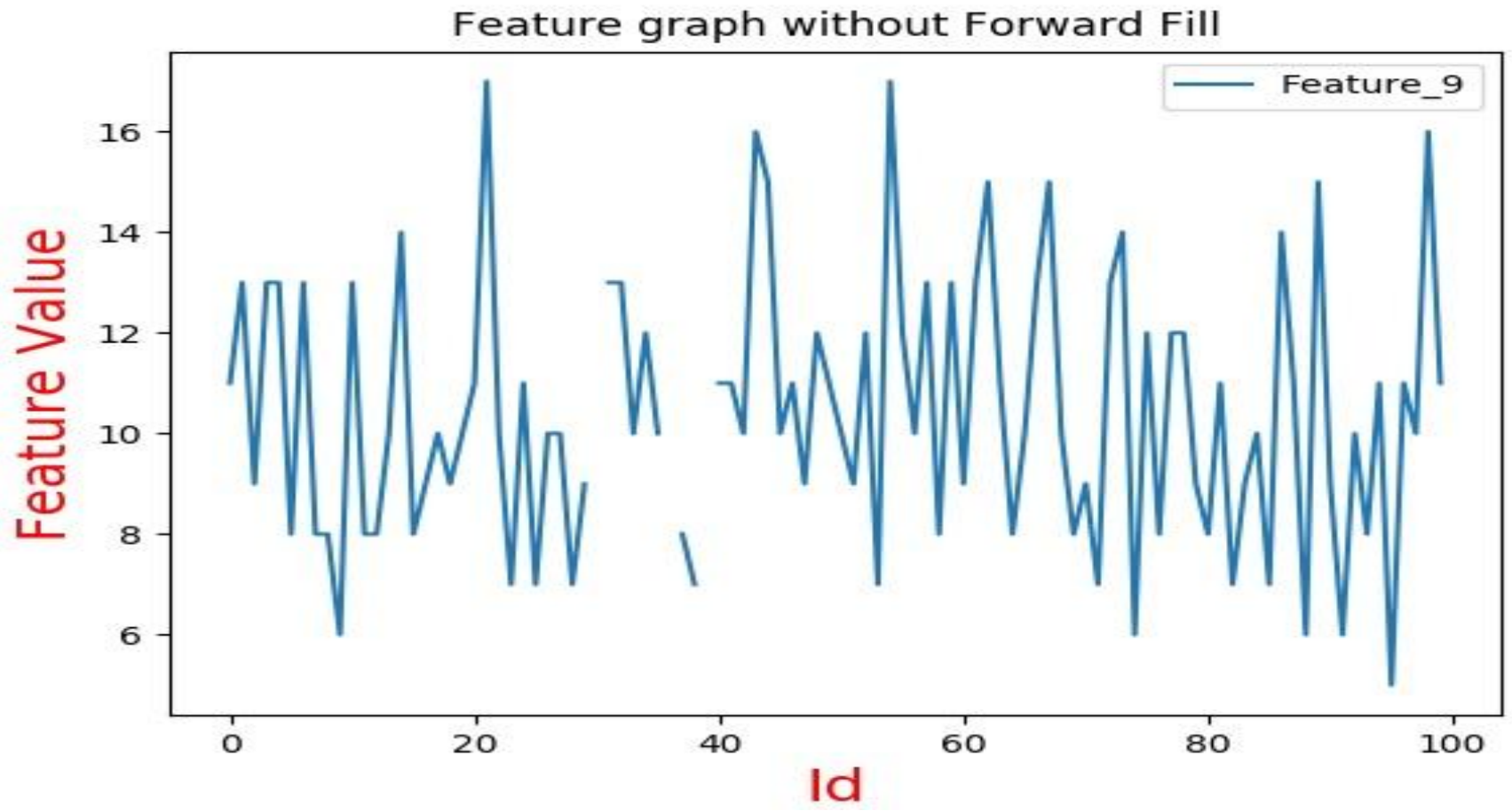
OBJECTIVE



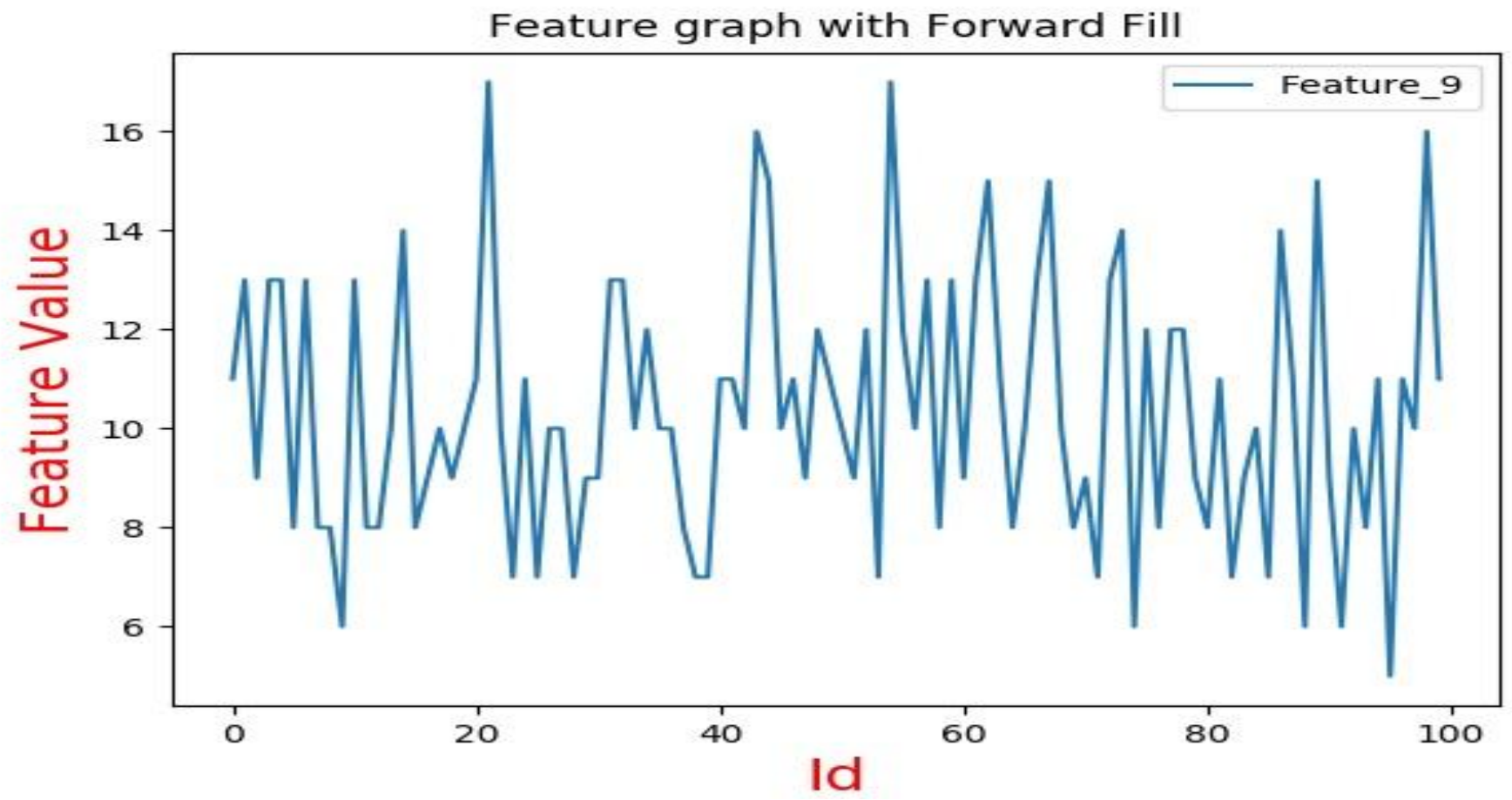
DATA CLEANING

- Handling Missing values by filling NaN values with:
  - Mean
  - Forward fill
  - Backward fill
- Min max normalization

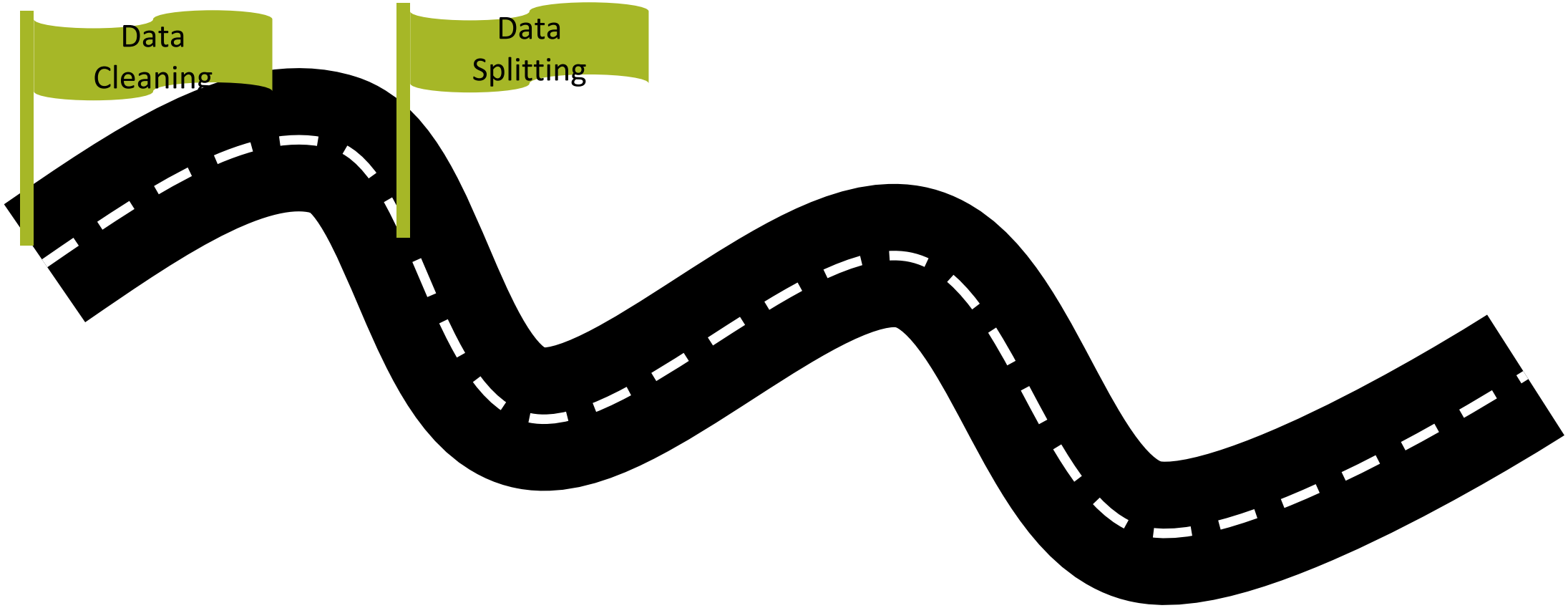
DATA CLEANING



DATA CLEANING



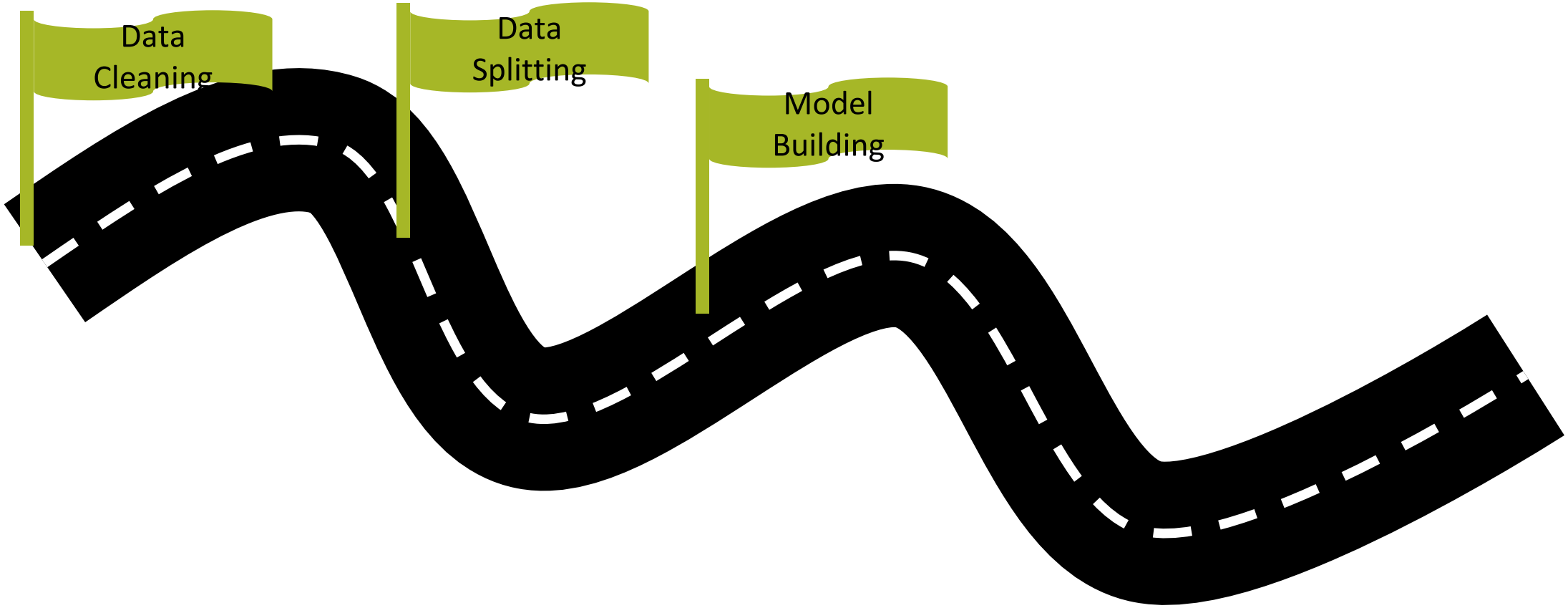
OBJECTIVE



DATA SPLITTING

- Splitting training data into input matrix and output matrix.
- Input matrix:
  - Feature_1 to Feature_25, Ret_2 to Ret_120
- Output matrix:
  - Ret_121 to Ret_180, Ret_PlusOne, Ret_PlusTwo

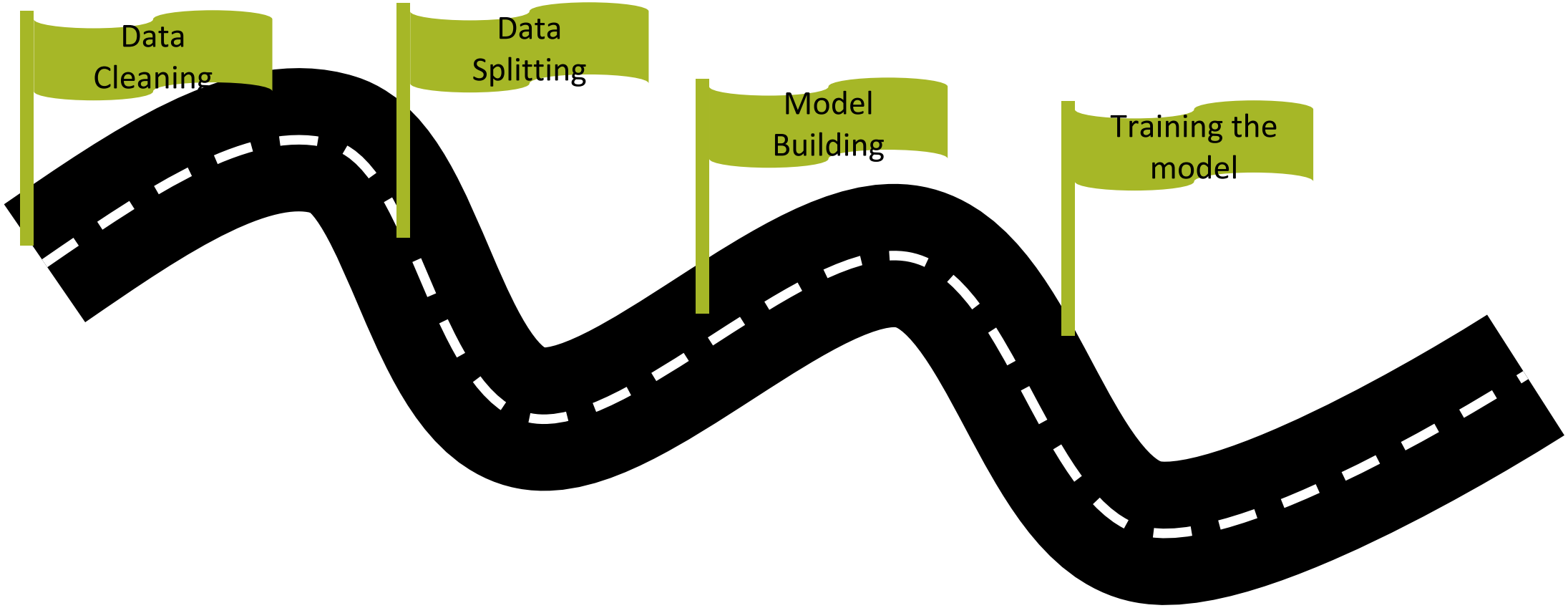
OBJECTIVE



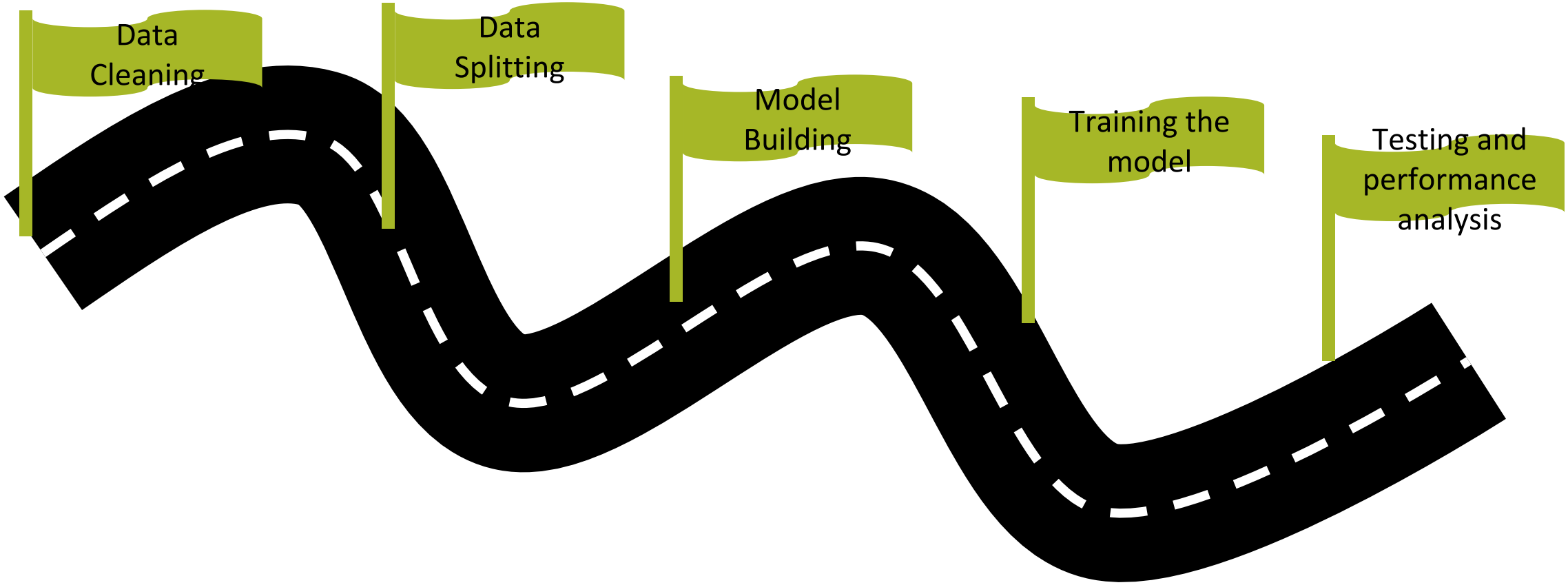
MODEL BUILDING

- **Using MLPRegressor**
  - solver='adam', alpha=0.0001, hidden_layer_sizes=(100,100),random_state=1, max_iter=200
  - Solver adam is used because more than one value has to be predicted for each sample data
- Using xgboost
  - Model_parameters: max_depth=10, eta=0.1, gamma=0, alpha=400, lambda=500
  - alpha:L1 regularization term on weights
  - lambda:L2 regularization term on weights
  - gamma: minimum loss required to make further partition on leaf node of a tree
  - eta: learning rate

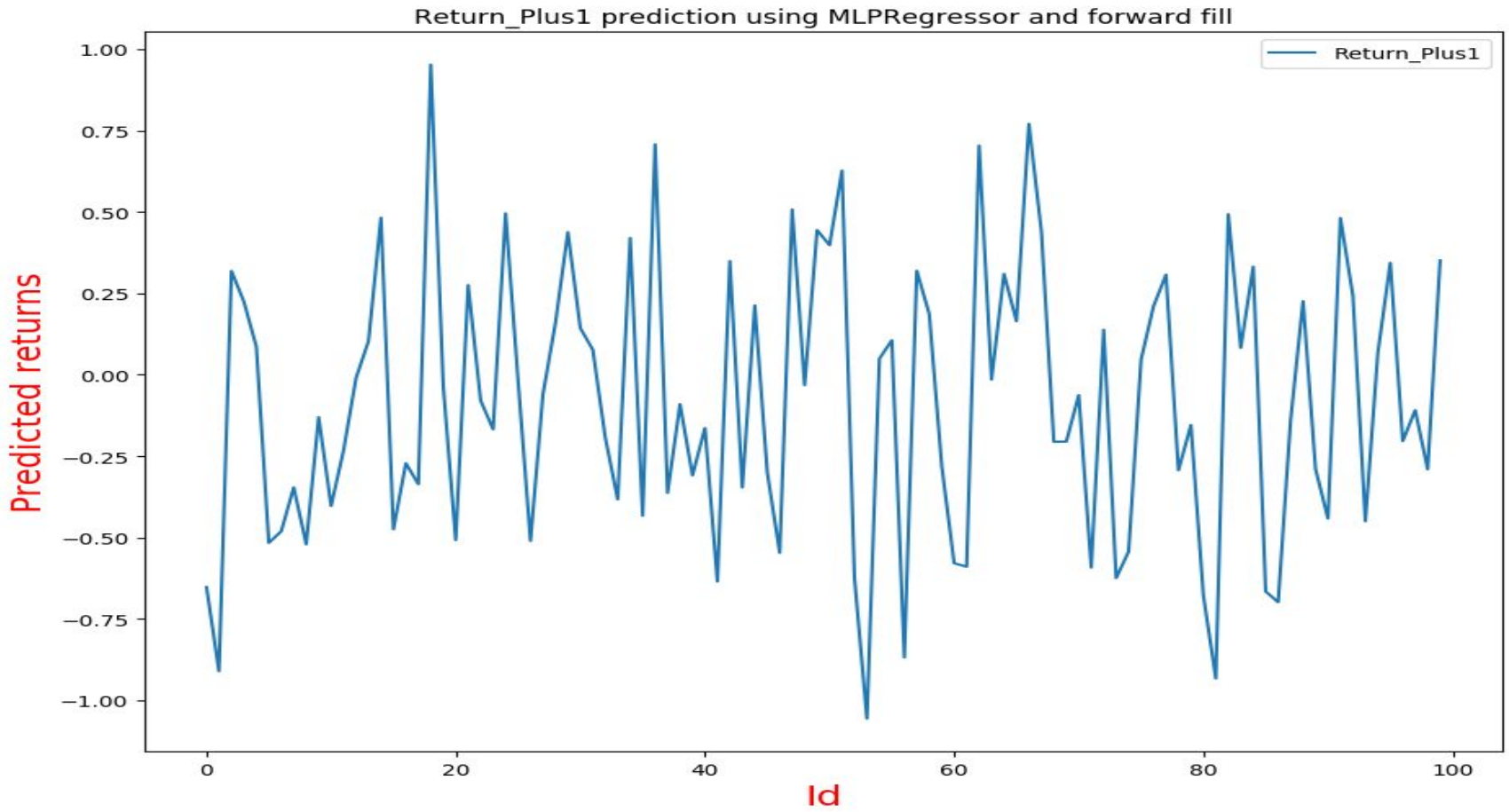
OBJECTIVE


Data Cleaning
Data Splitting
Model Building
Training the model

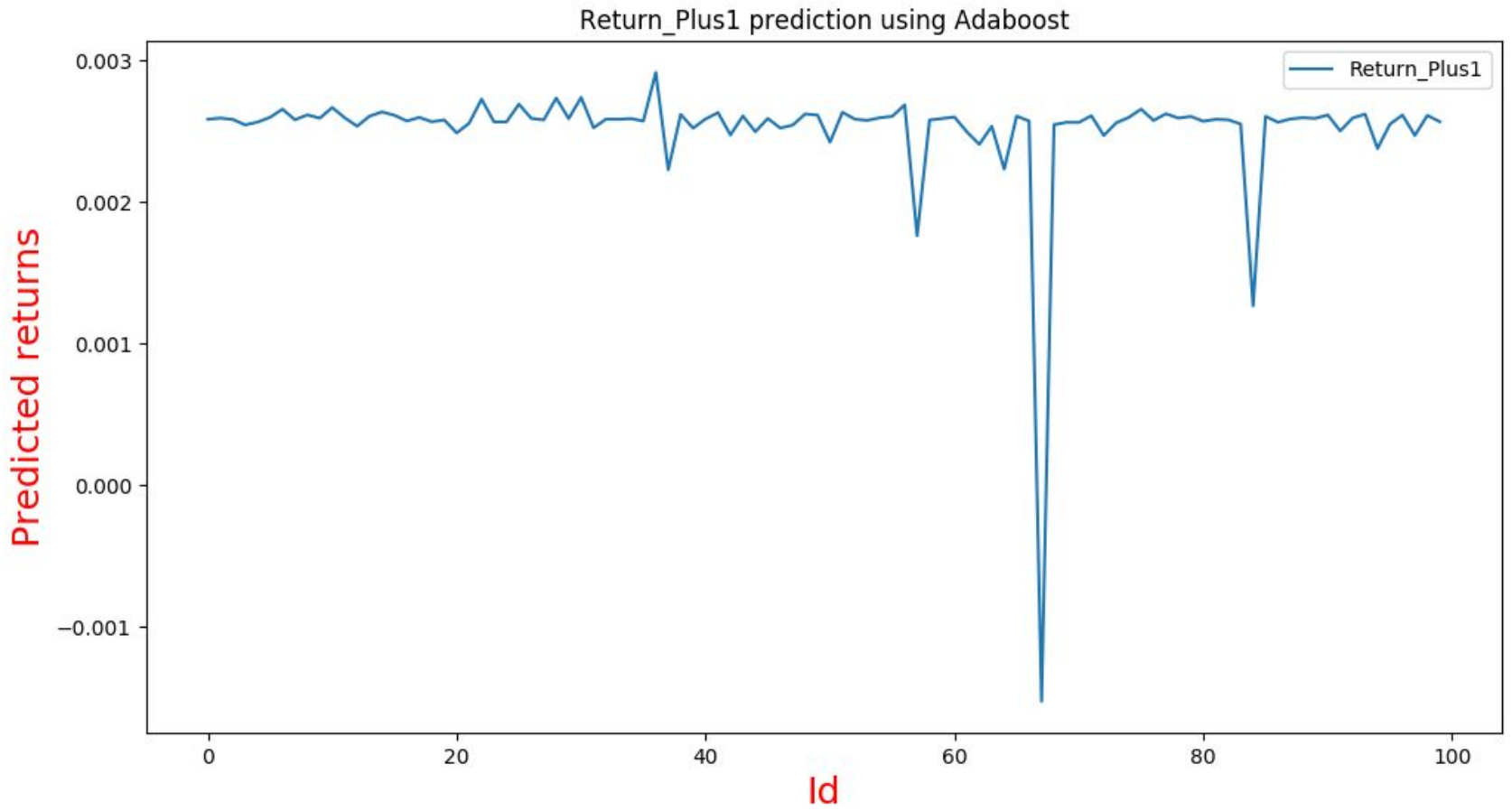
OBJECTIVE


Data Cleaning
Data Splitting
Model Building
Training the model
Testing and performance analysis

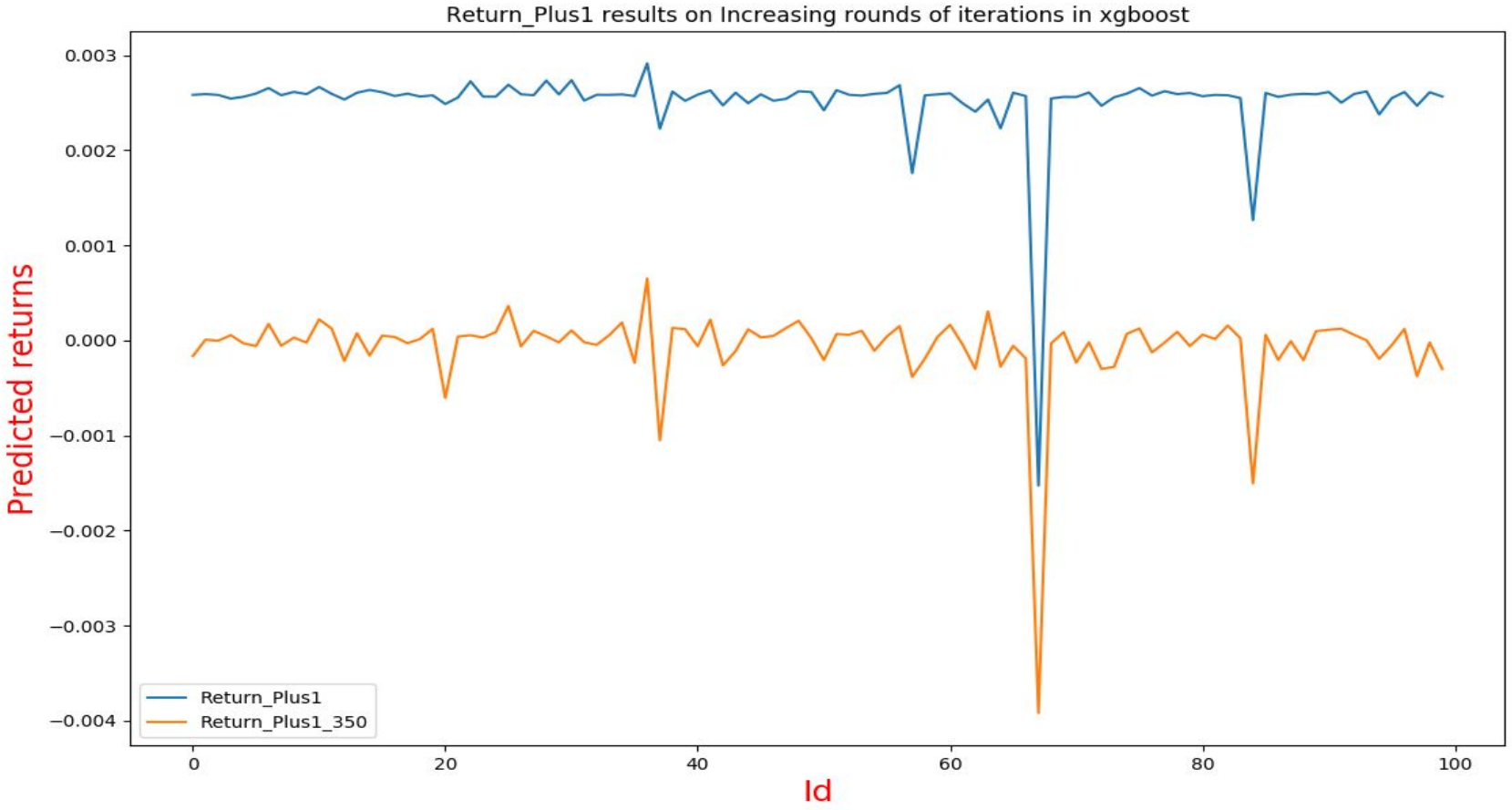
MLPRegressor PREDICTIONS



Adaboost PREDICTIONS



Adaboost PREDICTIONS

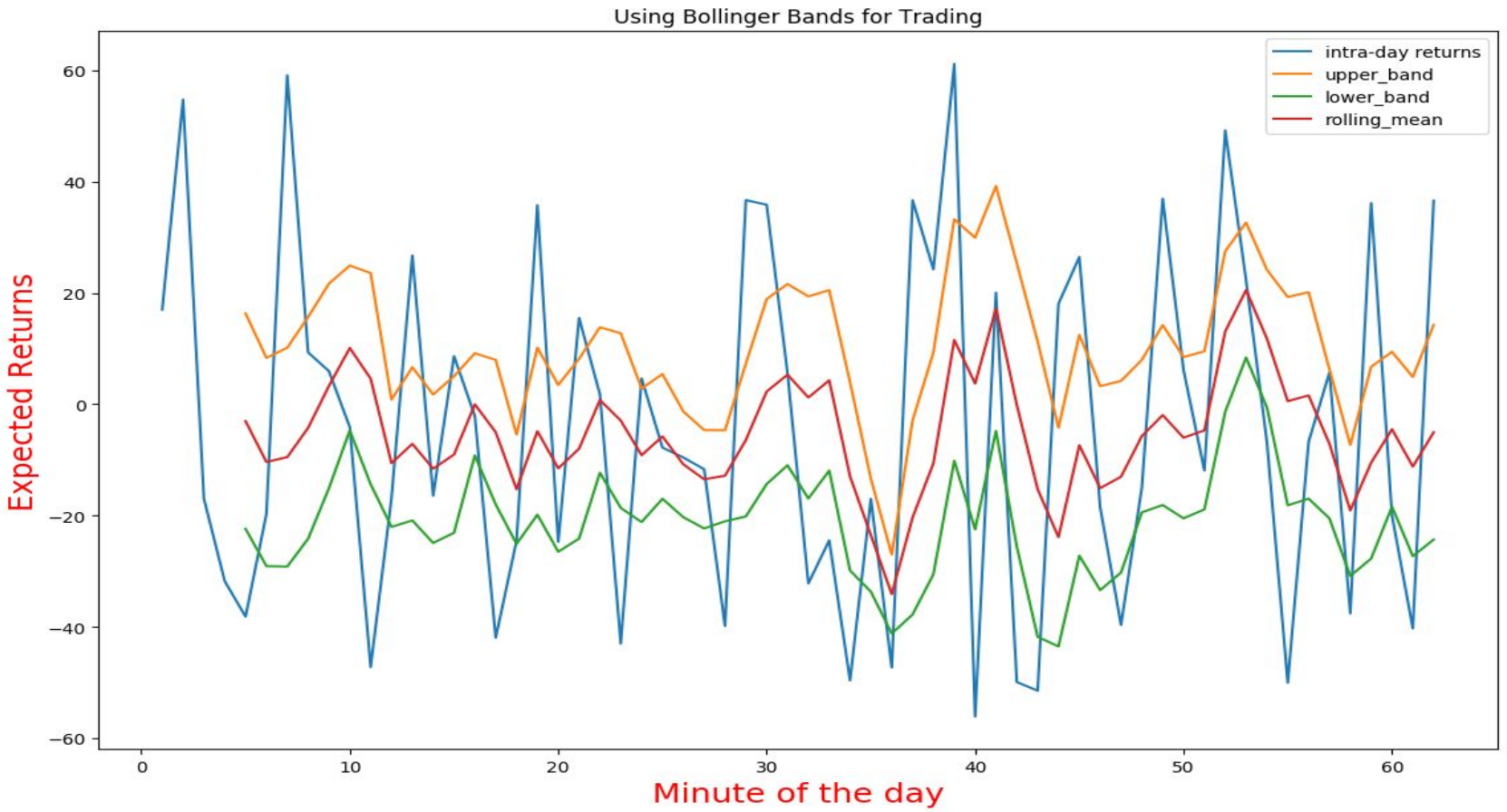


PERFORMANCE ANALYSIS

- The Weighted Mean Absolute Error according to kaggle were

Algoritm	Kaggle Score (weighted mean absolute error)
MLPRegressor (L2 Norm = 0.0001)	456780.57527
MLPRegressor (L2 Norm = 0.001)	444785.56787
Adaboost (iterations = 50)	4828.94362
Adaboost (iterations = 350)	2044.88105

USING BOLLINGER BANDS FOR TRADING



USING BOLLINGER BANDS FOR TRADING

- Represents the prediction of the returns calculated by the Bollinger Band algorithm
- used to calculate the fluctuations in the market
- used to predict the points in time when one should buy or sell a stock
- used the concept of rolling mean and standard deviation
- Rolling mean is the mean calculated by sliding a window of fixed size over whole data.
- We have used the window of size 5 minutes.

$$UpperBand = RollingMean + 2 * StandardDeviation$$
$$LowerBand = RollingMean - 2 * StandardDeviation$$

REFERENCES

- [1] Yusuf Perwej , Asif Perwej , “Prediction of the Bombay Stock Exchange (BSE) Market Returns Using Artificial Neural Network and Genetic Algorithm”, Computer Science & Information System, Jazan University, Jazan, Kingdom of Saudi Arabia (KSA); Department of Management, Singhnia University, Rajasthan, India, February 13th, 2012.

- [2] Zahid Iqbal, R. Ilyas, W. Shahzad, Z. Mahmood and J. Anjum, Member, IEEE, “Efficient Machine Learning Techniques for Stock Market Prediction”, Zahid Iqbal et al Int. Journal of Engineering Research and Applications ISSN : 2248-9622, Vol. 3, Issue 6, Nov-Dec 2013, pp.855- 867.

- [3] Waqas Ahmad NUST, College of Electrical and Mechanical engineering Rawalpindi Pakistan, “Analyzing Different Machine Learning Techniques for Stock Market Prediction”, (IJCSIS) International Journal of Computer Science and Information Security, Vol. 12, No. 12, December 2014.

THANK YOU