# Intelligent Coin Trading (IST) Platform


Mohammad Saim

# CONTENT

Dataset

Regression Models

Conclusion

Pre Processing

Evaluation

# Problem Statement

1

## Data Preprocessing

Using suitable cleaning and preprocessing techniques, we need to obtain a refined dataset.

2

## Feature Engineering

Select relevant features from all the parameters given.

3

## Choosing the Model

Obtaining optimal dataset based on the nature and features selected from the dataset

# Dataset

ID

Name

Symbol

Num_Market_
Pair

Date Added

Max_Supply

Circulating
Supply

Total
Supply

Platform_ID

Platform
Name

Platform
Symbol

Platform Token
Address

CMC
Rank

Last
Updated

Price

Volume_24H

Percent Change( 1hr,
24h,7d, 30d, 60d, 90d)

Market
Cap

Extracted
Time

Count

# Data Pre-Processing

1. Handle Missing Data: Identify missing values: Check for missing values in each column of the dataset
2. Remove Irrelevant Columns: Remove irrelevant features based on the relevance analysis mentioned earlier, remove columns
3. Cleaning Up the Dataset: This includes removing duplicates, handling outliers and deciding whether to remove outliers or apply appropriate transformations to handle them, and checking for inconsistent data types.
4. Data Normalization: This includes normalizing numerical features like Market_cap, Volume_24h, etc to ensure they are on a similar scale.
5. Categorical Variable Encoding: For categorical features like Name or Symbol, encode them using techniques like one-hot encoding or label encoding to convert them into numerical values.

# Feature Engineering

1

**Date Based Features**

Extract useful features such as day of the week, month, or year.

2

**Historical Price Features**

Create features representing percentage change in price over different time intervals)

3

**Rolling Statistics**

Calculate rolling averages of price and volume over a specific time window

4

**Lagged Features**

Create features representing price from previous time steps (e.g., previous day's price)

5

**Seasonality and Holiday Effects**

Include binary or categorical features to represent fixed periods of seasonality

# Machine Learning Models

Random Forest Regressor

Extra Trees Regressor

Facebook's Prophet Model

Lasso Regressor

# Random Forest Regressor

## Suitability

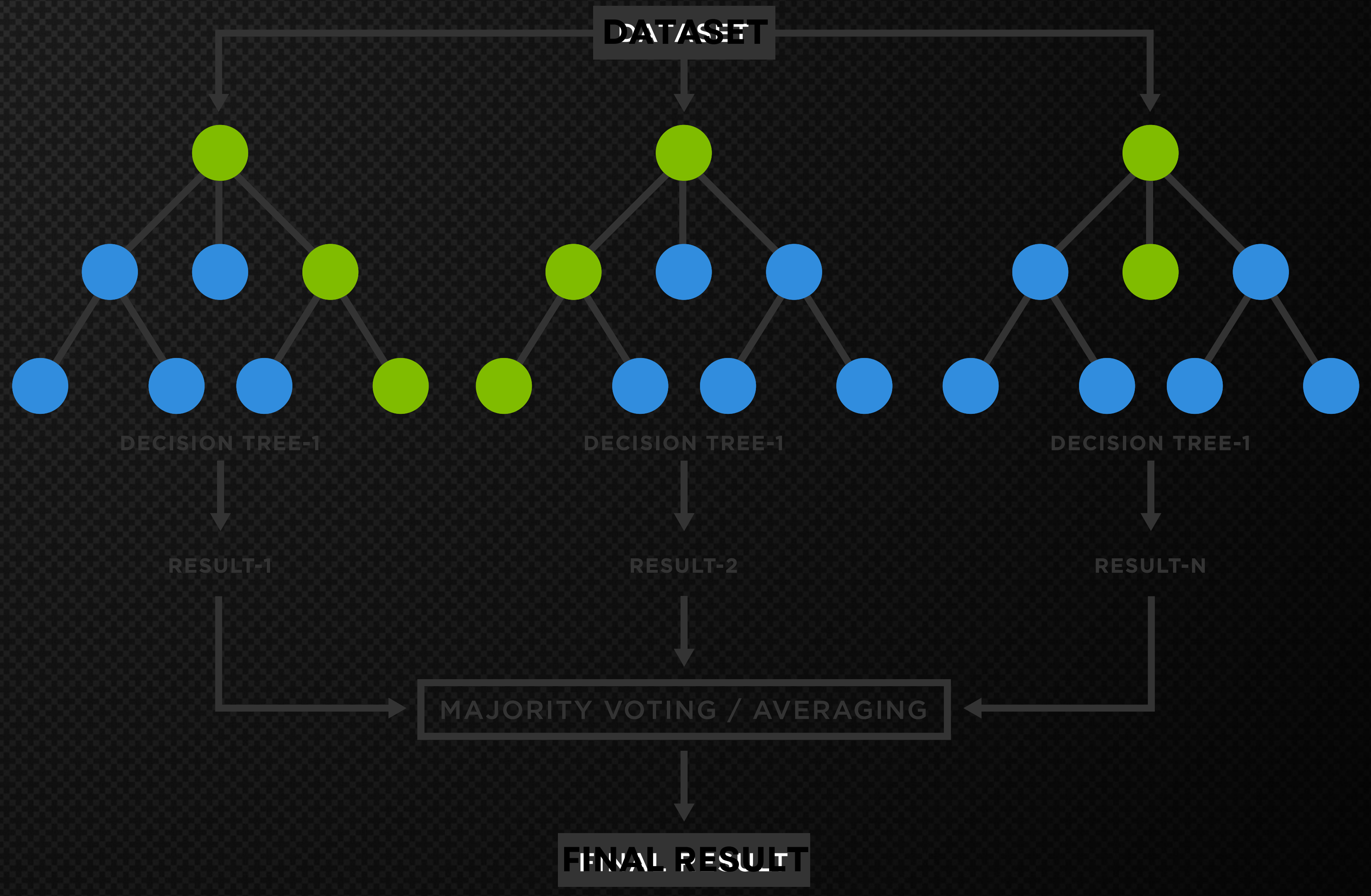
It handles non-linearity, robustness to outliers, employs feature importance and can handle high dimension dataet.

## Algorithm

It starts with sampling to create a tree subset, build decision trees, conduct ensemble voting and predict final result along with handling overfitting.

## Applications

Predominantly used in stock market prediction, real estate price prediction, energy demand forecasting, etc.



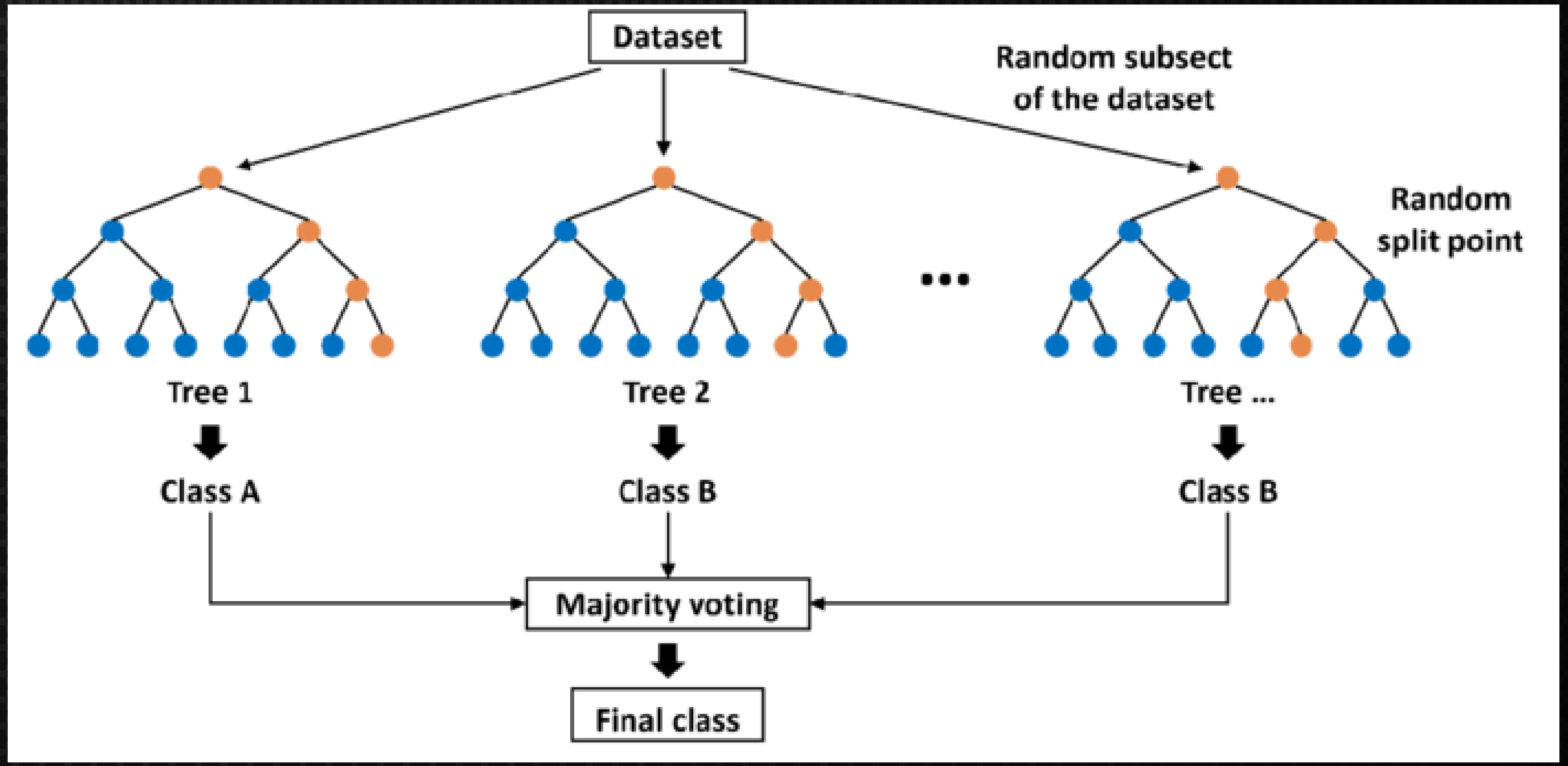

## LIMITATION

- Prone to overfitting
- Computationally Expensive

# Extra Trees Regressor

## Suitability

It handles non-linearity, robustness to outliers, employs feature importance and is efficient in training process




## Algorithm

Introduces additional randomness compared to Random Forest, making it faster to train while still maintaining good predictive performance.

FINAL RESULT

## Applications

Predominantly used in demand forecasting, financial time series analysis, credit risk assessment, etc.

## LIMITATION

- High Variance
- No built-in feature importance

# Facebook's Prophet Model

## Suitability

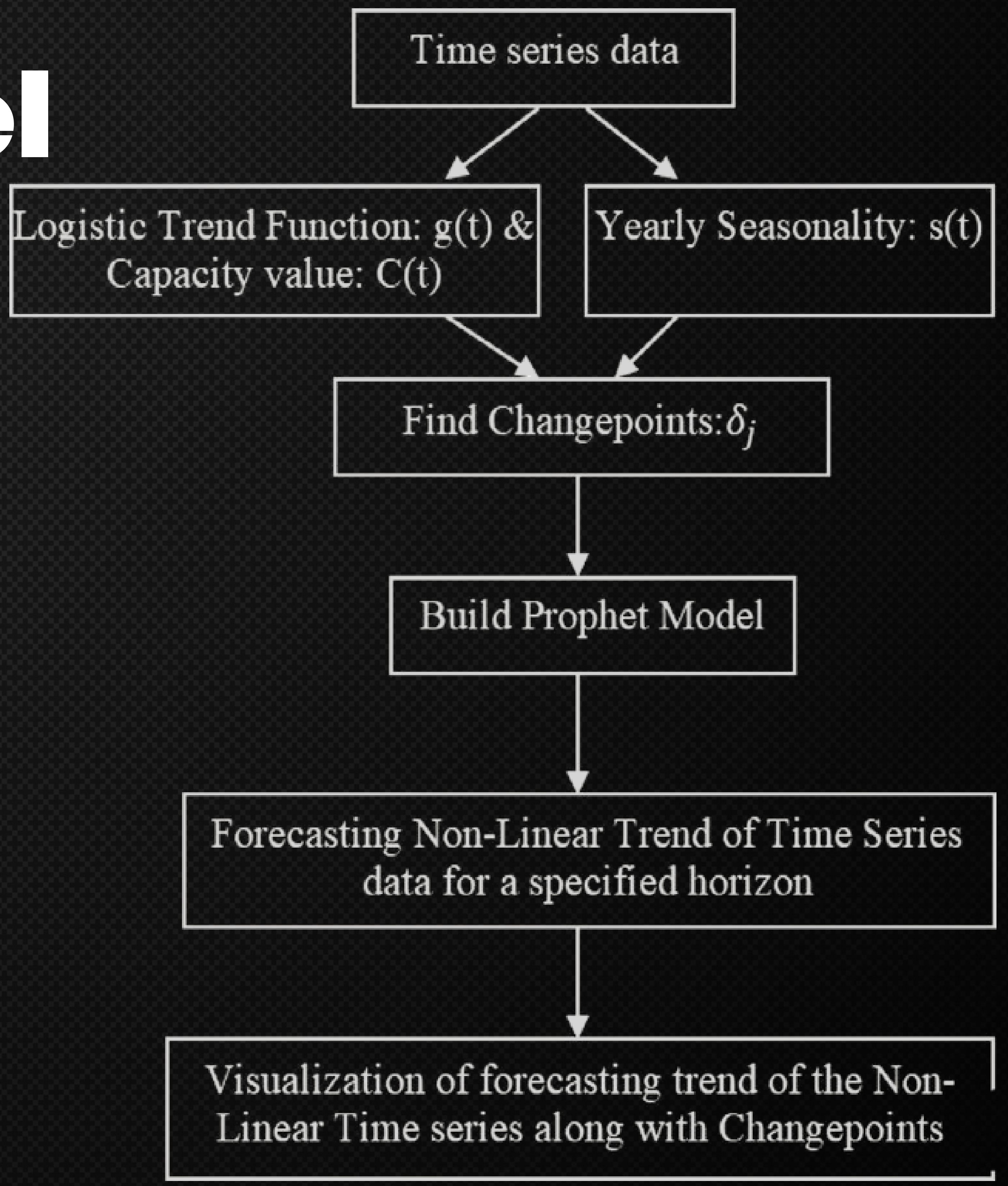
It's suitable for time series forecasting, accumulates trends and seasonality mode, gives uncertainty estimates Designed to be user-friendly and accessible

## Algorithm

It starts with trend estimation (linear or logistic), and adds seasonality mode, holiday effect, additive decomposition and Bayesian Forecasting.

## Applications

Predominantly used in demand forecasting, electricity forecasting, web traffic prediction, etc.




## LIMITATION

- Limited Exogenous Factors
- Linearity Assumption

# Lasso Regressor

## Suitability

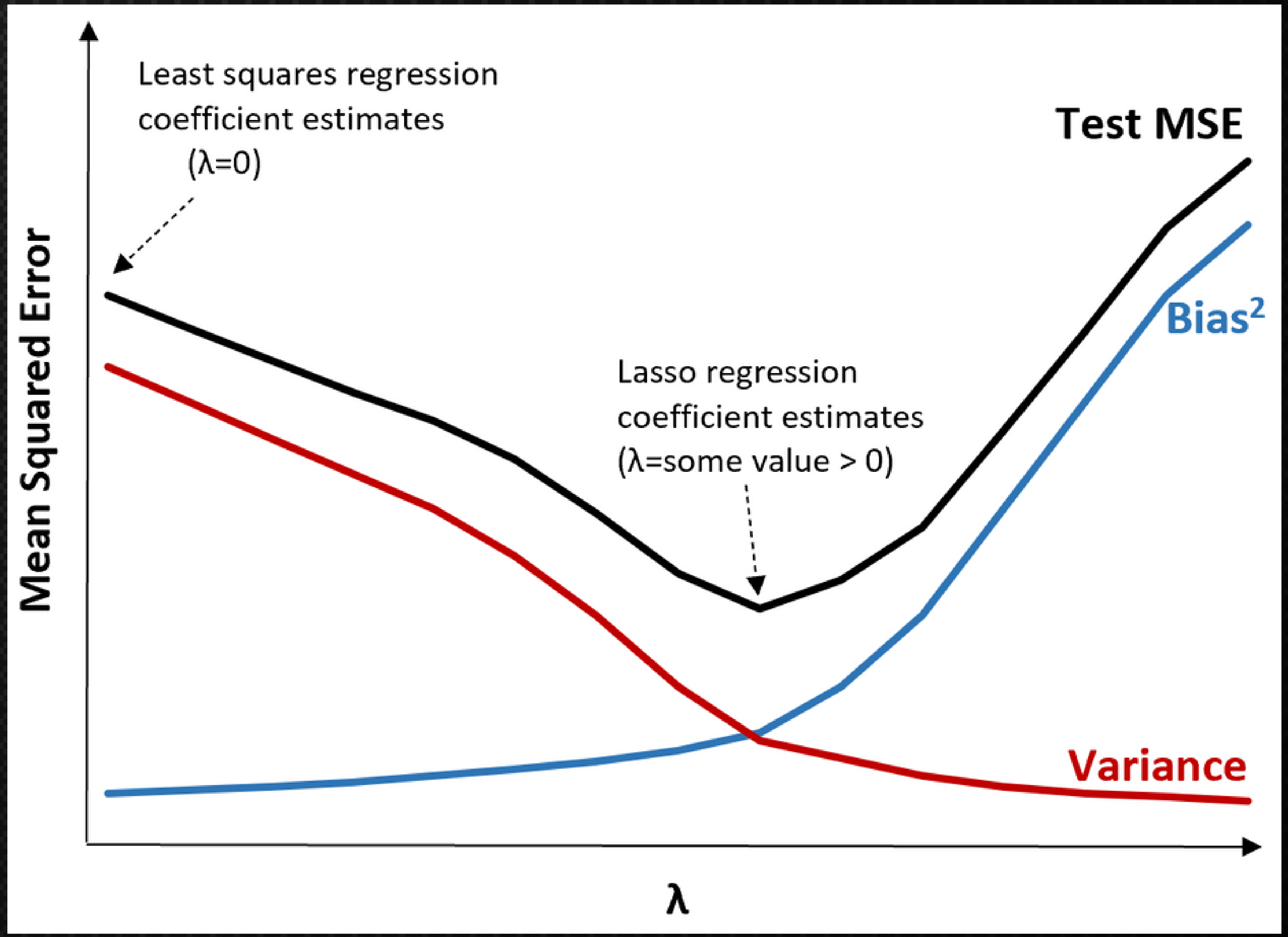
It's known for its feature selection and regularization, handling multicollinearity when working with high-dimension data and producing an interpretable model.

## Algorithm

Initializes weights of the model with small random values, defines a loss function and minimizes it, and performs feature selection for optimal result.

## Applications

Used in sectors of Finance, Healthcare and Economics along with forecasting energy consumption.




## LIMITATION

- Deterministic Feature Selection
- Feature Selection Bias

# Evaluation Metrics

## R-squared (R^2) Score

R-squared measures the proportion of variance in the target variable that is explained by the model. It ranges from 0 to 1, where 1 indicates a perfect fit and 0 indicates that the model does not explain any of the variance.

## Mean Squared Error (MSE)

MSE calculates the average squared difference between the predicted values and the actual values. It gives a measure of the average magnitude of the error. Lower MSE values indicate better performance.

## Root Mean Squared Error (RMSE)

RMSE is the square root of MSE and provides a more interpretable metric in the same unit as the target variable. It gives a measure of the average magnitude of the error with the same scale as the target variable.

## Mean Absolute Error (MAE)

MAE calculates the average absolute difference between the predicted values and the actual values. Lower MAE values indicate better performance.

# CONCLUSION

1

**DATASET**

Display of the provided dataset along with preprocessing techniques

2

**Feature Engg.**

Employed five most relevant feature engineering methods to optimize the datasete.

3

**Regression**

Overview and discussion of the chosen machine learning models for regression analysis.

# THANK YOU!

**Questions are Welcome.**