Soft Sensor Model in Debutanizer Column

CH512 - Course Project

Submitted to: Dr. Jayaram Valluru

Presented By -

- Rahul (2021CHB1052)
- Susanth (2021CHB1053)
- Akhil (2021CHB1364)

Soft Sensor System

The Debutanizer column is a crucial component in the distillation process. It is responsible for separating propane (C3) and butane (C4) from the naphtha stream. The main objectives of the column is to increase the concentration of stabilized gasoline (C5) in the overheads, which is used as the feed for the LP gas splitter, and to decrease the concentration of butane (C4) in the bottoms, which is fed into the naphtha splitter.

Two sensors, N1 and N2, provide real-time estimates of key variables: N1 measures the percentage of C5 in C4 (F_C5), while N2 assesses the percentage of C4 in C5

Debutanizer System
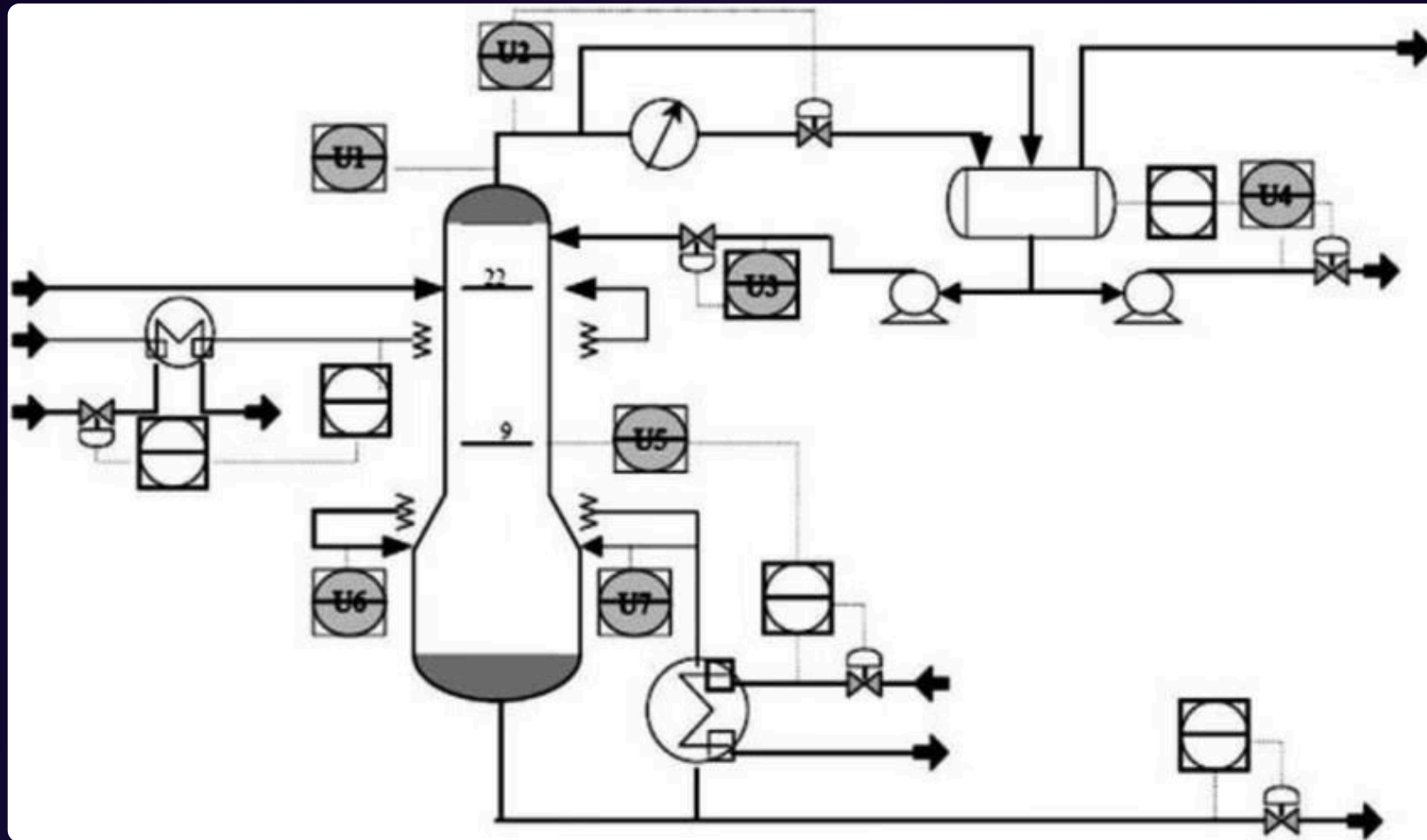
System Parameters

Variable	Description
u1	Top Temperature
u2	Top Pressure
u3	Reflux Flow
u4	Flow to Next Process
u5	6th Tray Temperature
u6	Bottom Temperature (repeated)
u7	Bottom Temperature (repeated)

To predict the soft sensor

- y - Percentage of C5 in C4 (F_C5)

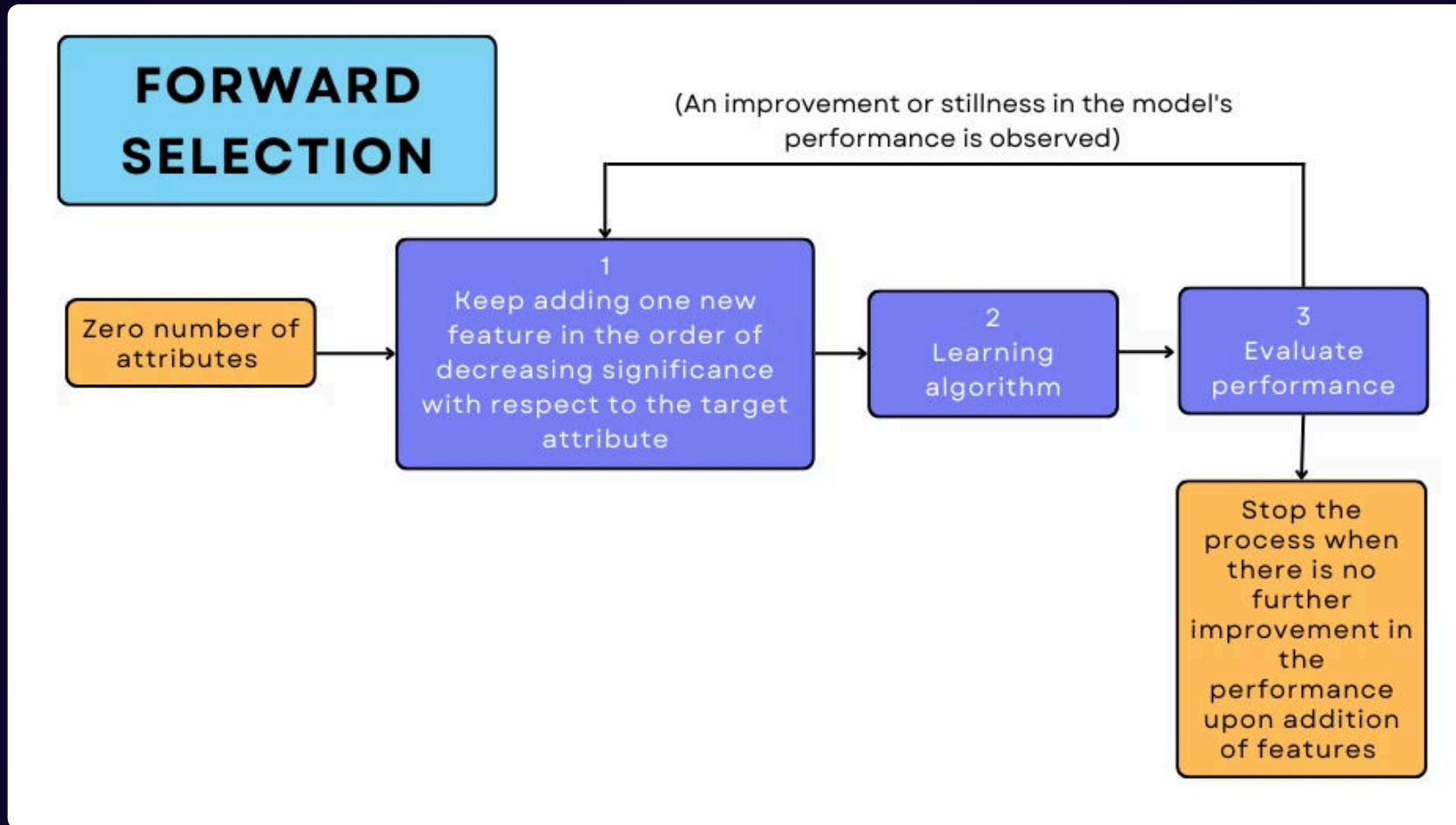# Linear Function Approximations

We have utilized the use of following preprocessing methods

- Forward Selection Method

- Lasso regression technique.

- PCA based feature selection.

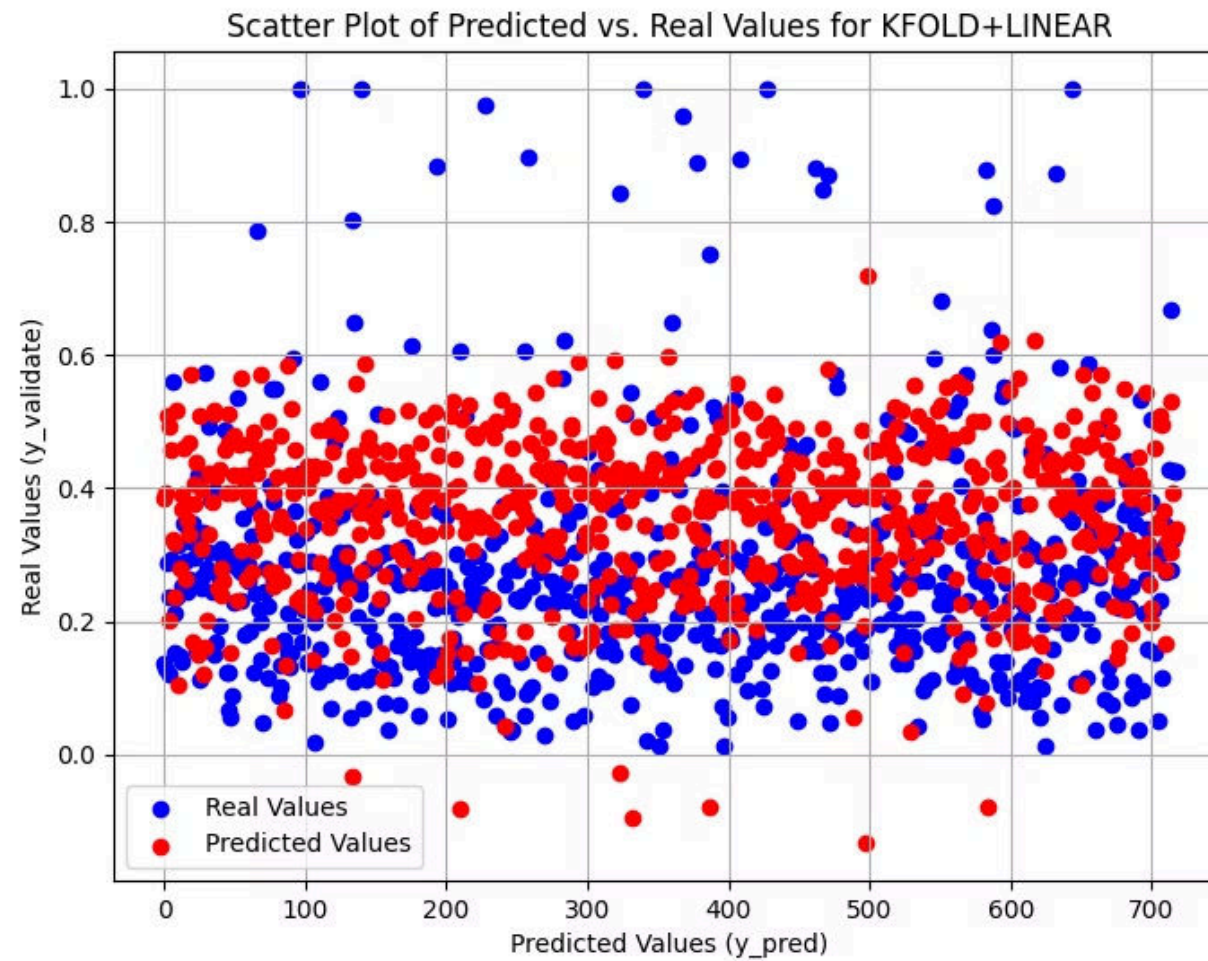We have also built following Models to train and evaluate the model and to predict the soft sensor value

- Linear Regression Model

- Polynomial Regression Model Degree 2

- Polynomial Regression Model Degree 3
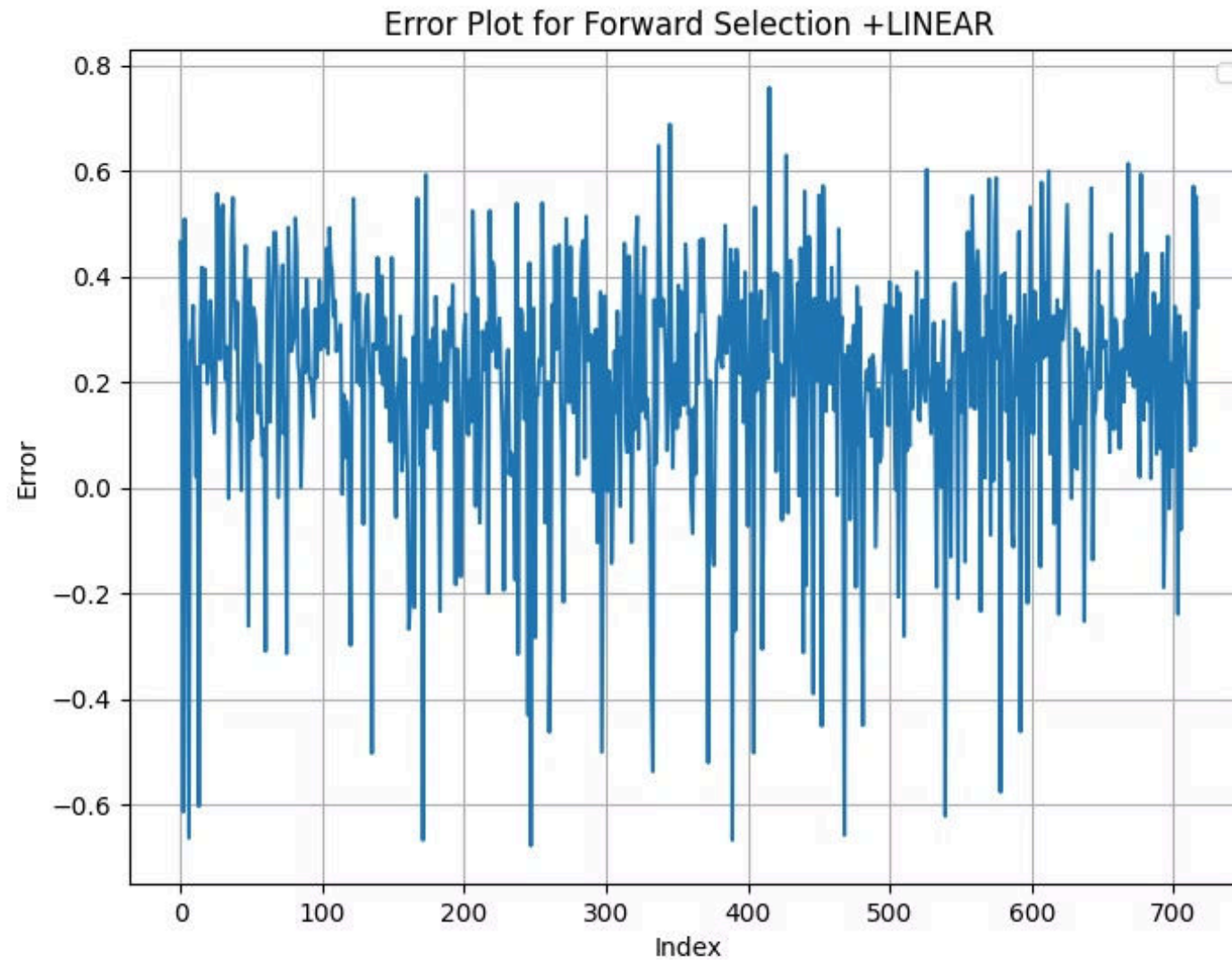
Forward Selection Method

FORWARD SELECTION

(An improvement or stillness in the model's performance is observed)

Zero number of attributes

1
Keep adding one new feature in the order of decreasing significance with respect to the target attribute

2
Learning algorithm

3
Evaluate performance

Stop the process when there is no further improvement in the performance upon addition of features

Final features u1, u2, u5

Forward Selection Method

Scatter Plot of Predicted vs. Real Values for KFOLD+LINEAR

Forward Selection Method Error

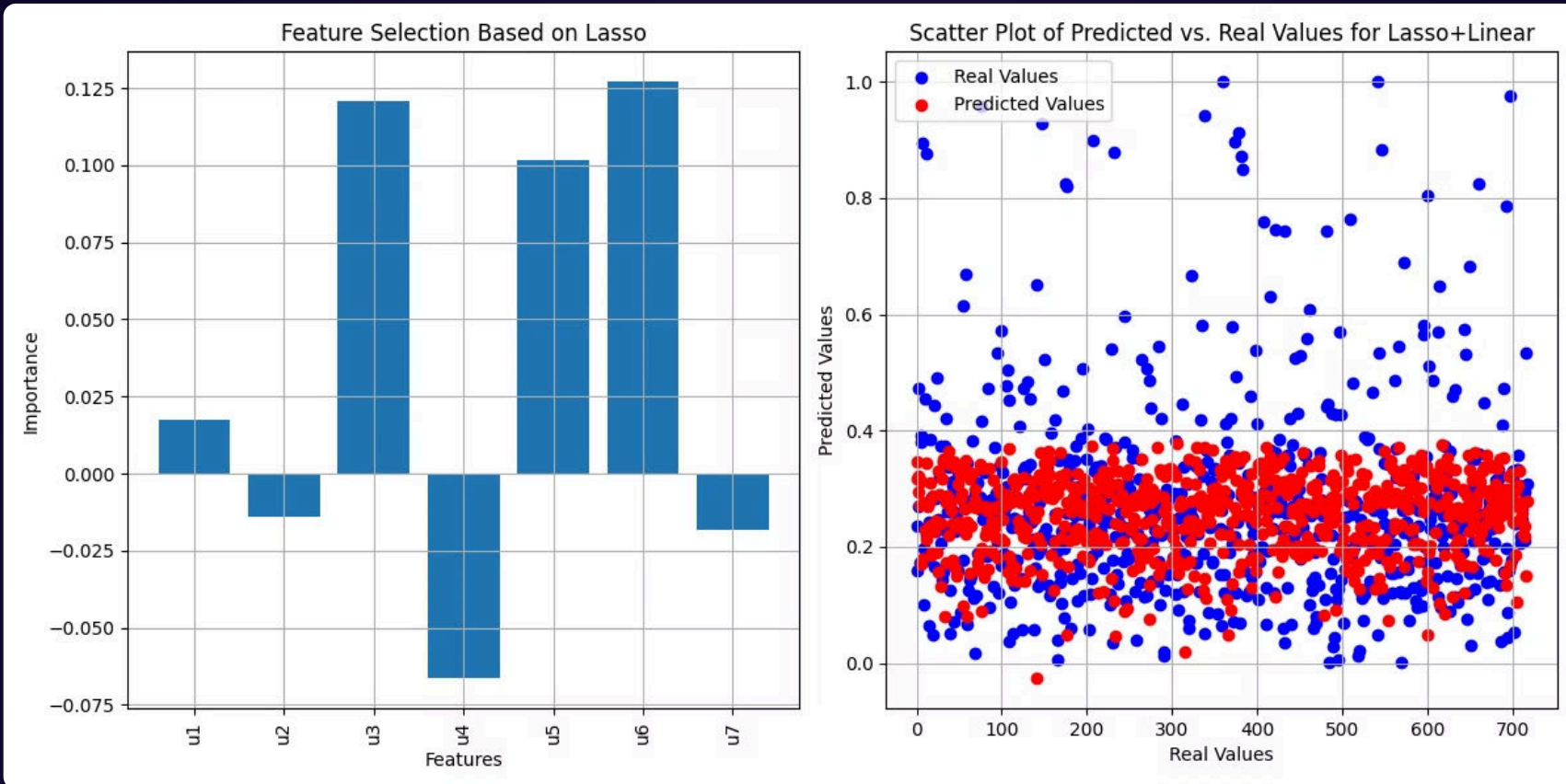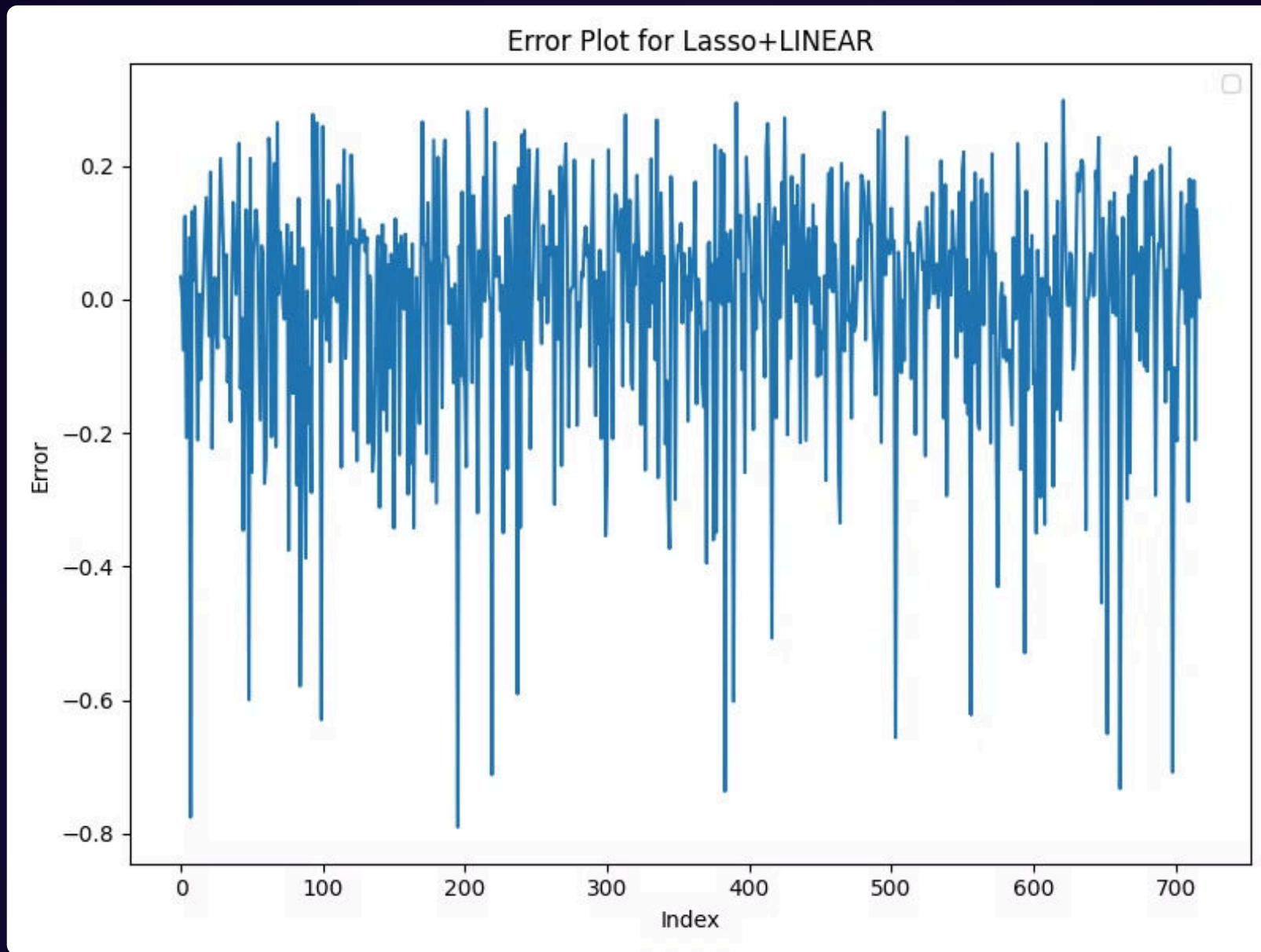Error Plot for Forward Selection +LINEAR

Lasso Regression Technique

$$\hat{y}_i = w_0 + \sum_{j=1}^{m} X_{ij} w_j$$

$$J(w) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \alpha \sum_{j=1}^{m} |w_j|$$
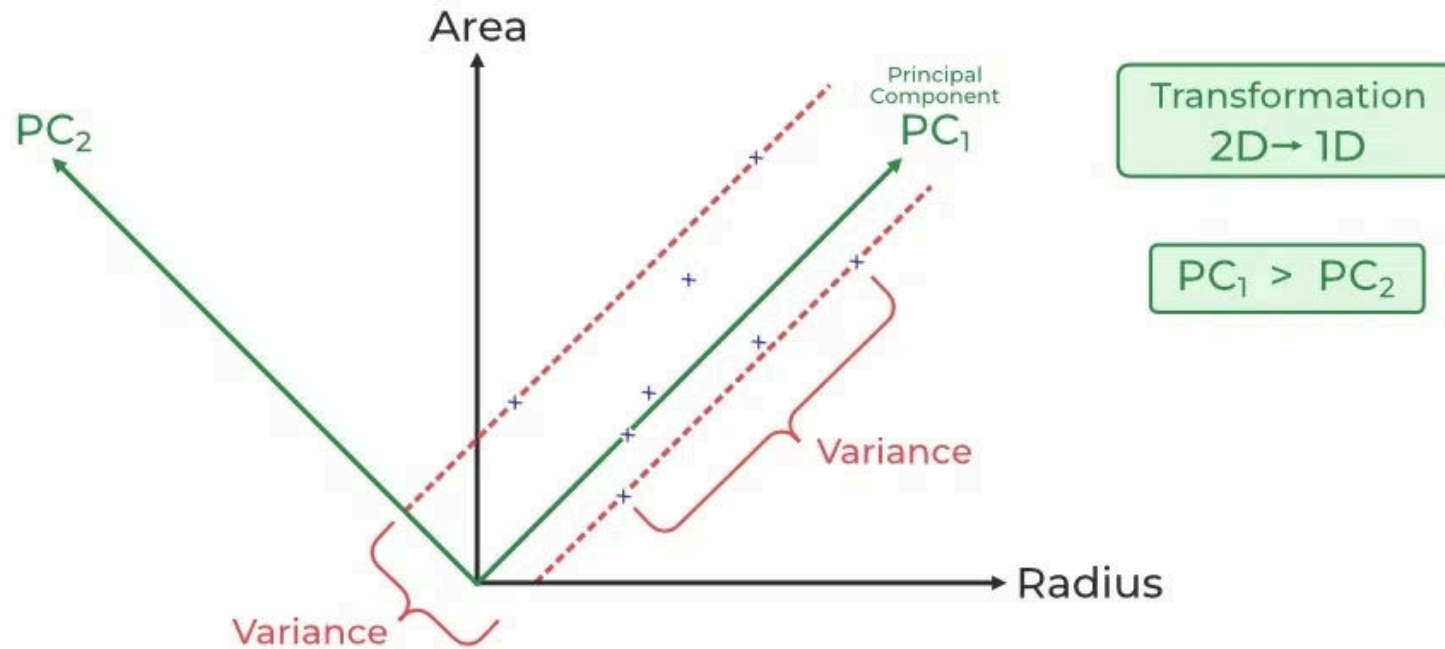
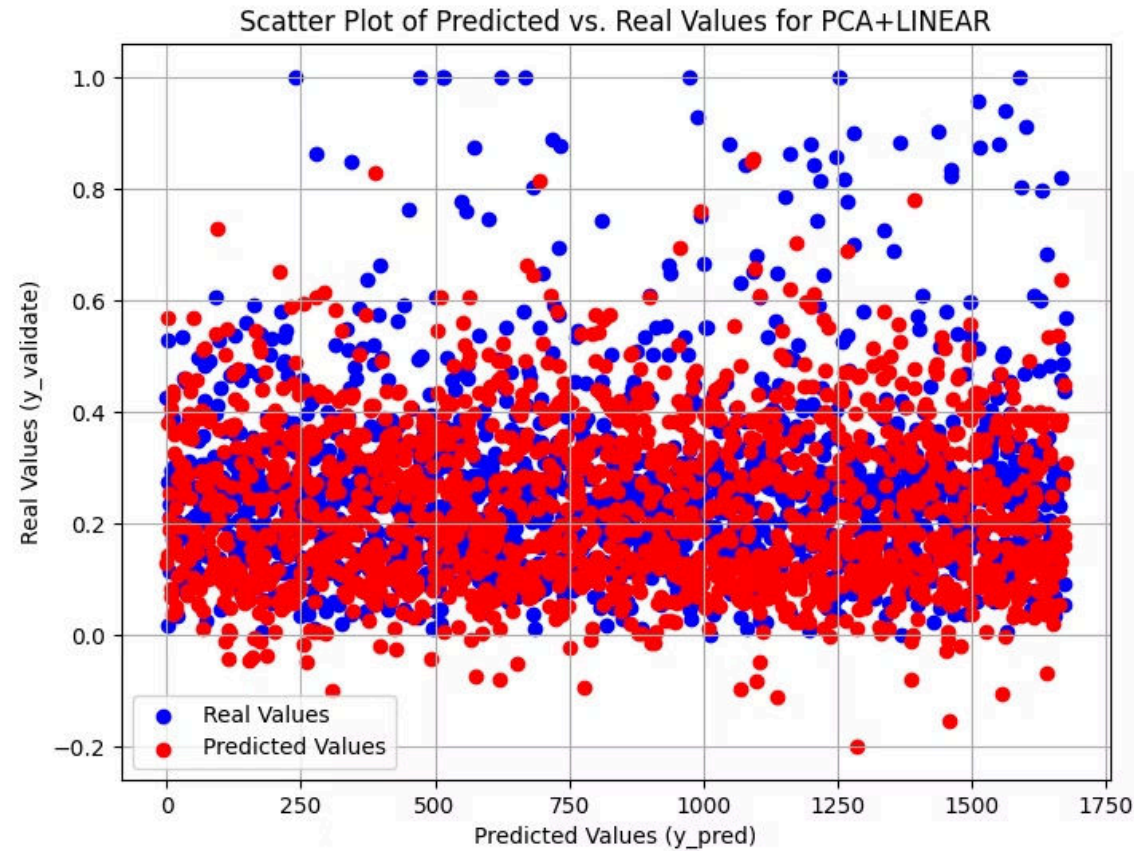$$\|w\|^2 = \sum_{j=1}^{m} |w_j|^2$$
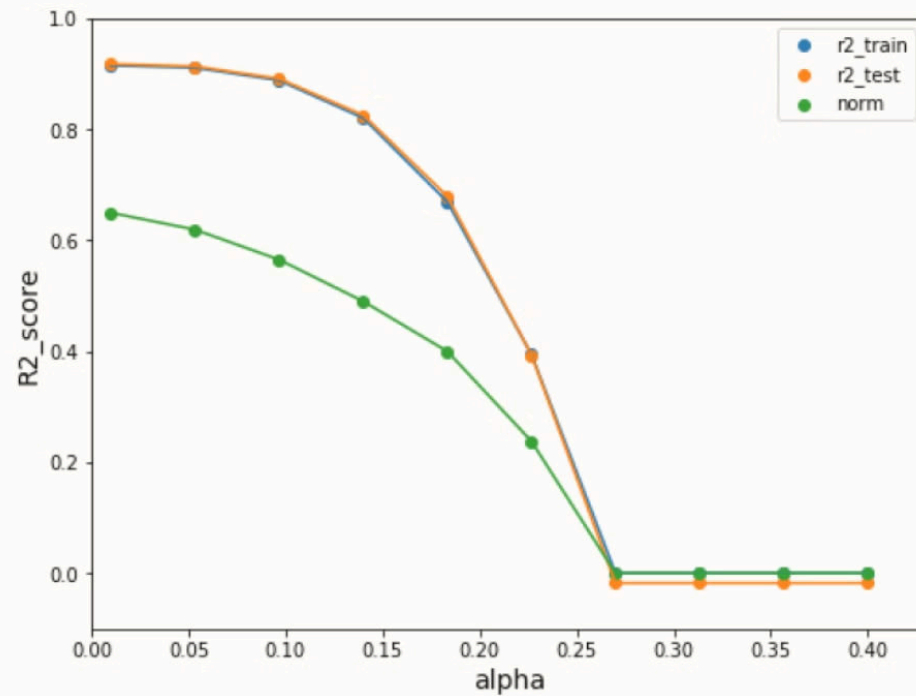
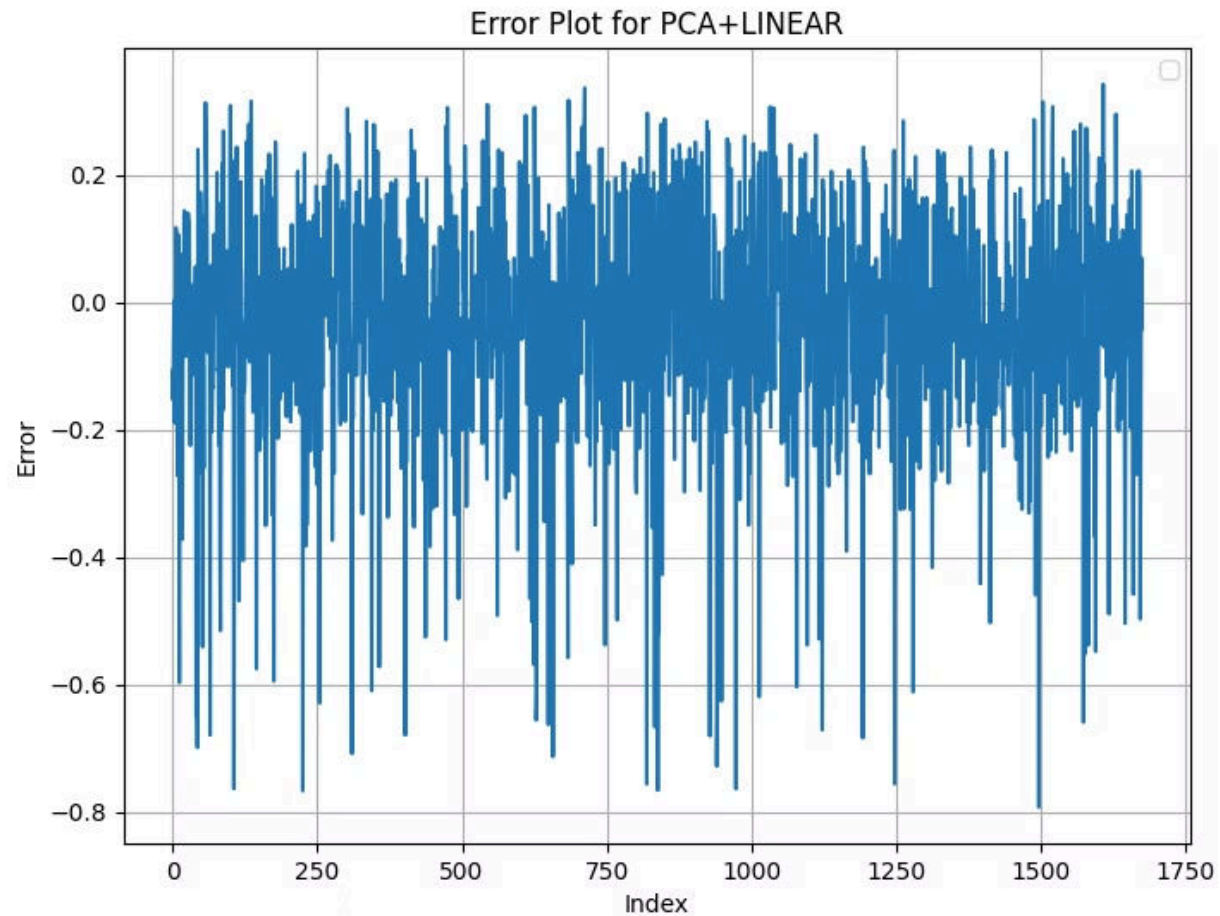# Lasso Regression Technique

Lasso Regression Technique Error

Error Plot for Lasso+LINEAR

PCA based Feature Selection

PCA based Feature Selection

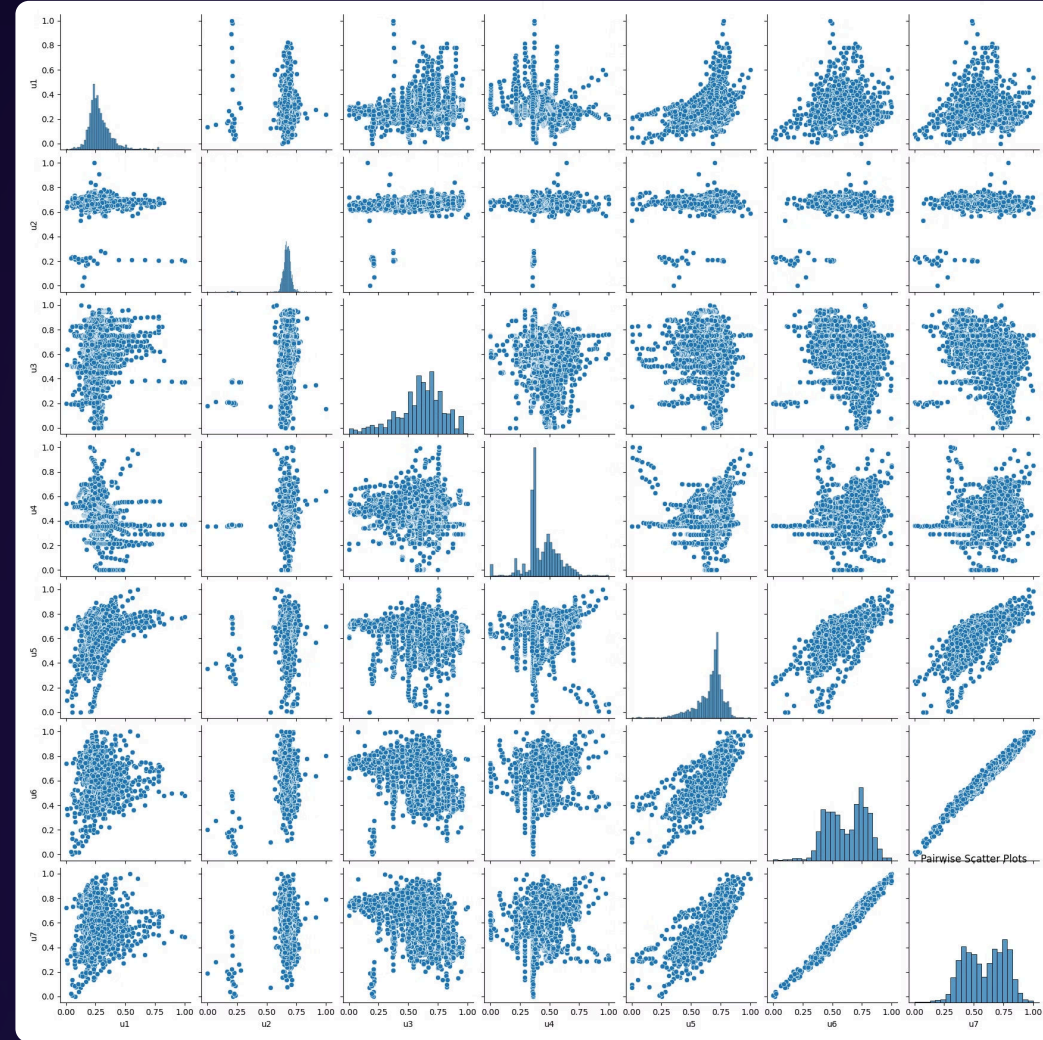Scatter Plot of Predicted vs. Real Values for PCA+LINEAR

PCA based Feature Selection

PCA based Feature Selection

Pairwise Scatter Plots

Comparing MSE LOSS for all Models

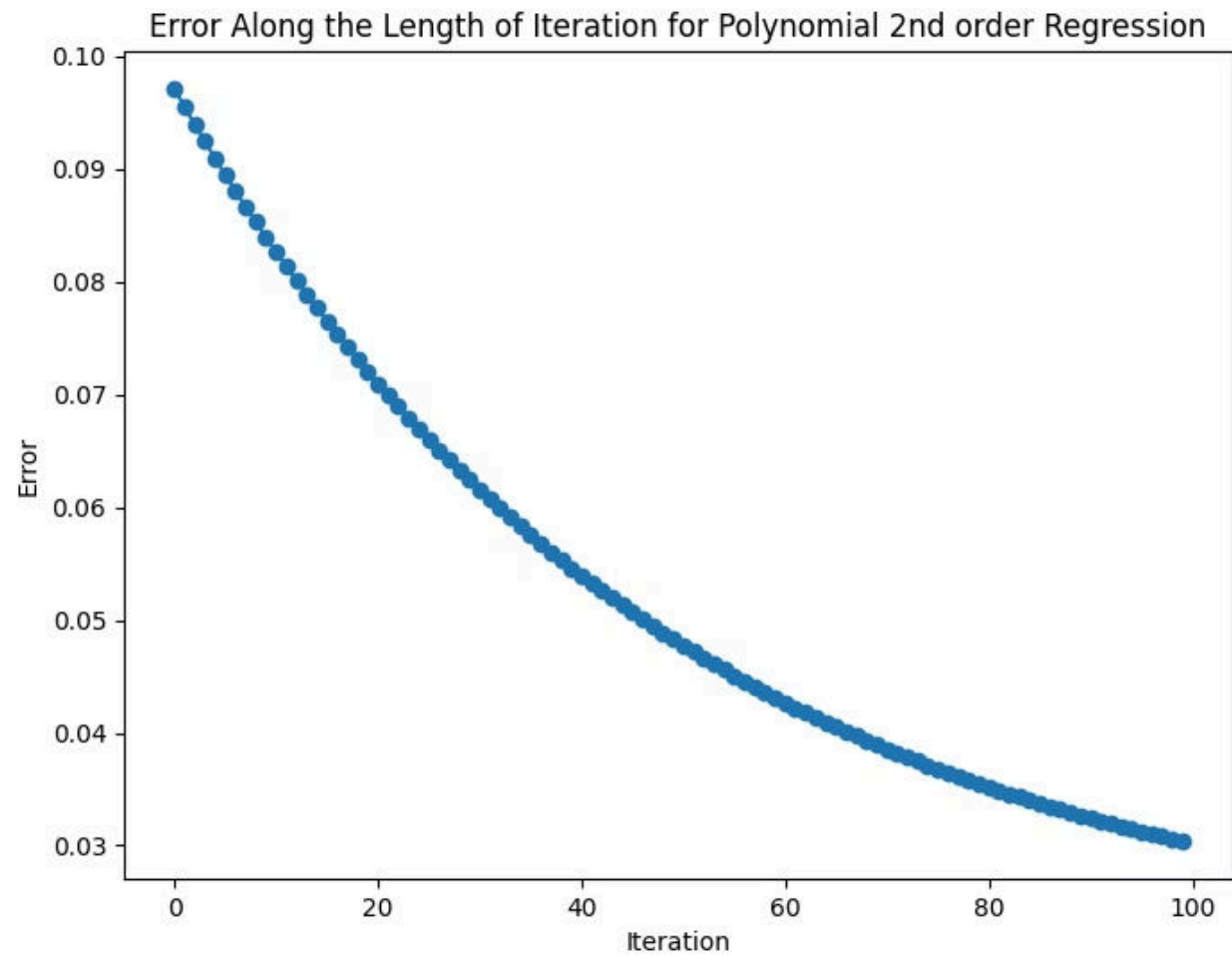| Model	Mean Squared Error	AIC
Lasso Regulation with Linear Regression	0.03241180293154075	16.858
Lasso Regulation with 2nd Order Regression	0.02481301269715393	19.392
Lasso Regulation with 3rd Order Regression	0.02244005206459875	13.423
PCA with Linear Regression	**0.044719823733125205**	**10.214**
PCA with 2nd Order Regression	0.023987154250022288	11.460
PCA with 3rd Order Regression	0.024285145721552204	11.435
Forward Feature Selection with Linear Regression	0.057633911176724555	19.809
Forward Feature Selection with 2nd Order Regression	0.030892300847618723	21.070
Forward Feature Selection with 3rd Order Regression	0.02971214984037660 5	20.918

Linear Regression Model

Residuals Plot

Error Along the Length of Iteration for Linear Regression

Polynomial Regression Model

Residuals Plot

Error Along the Length of Iteration for Polynomial 2nd order Regression
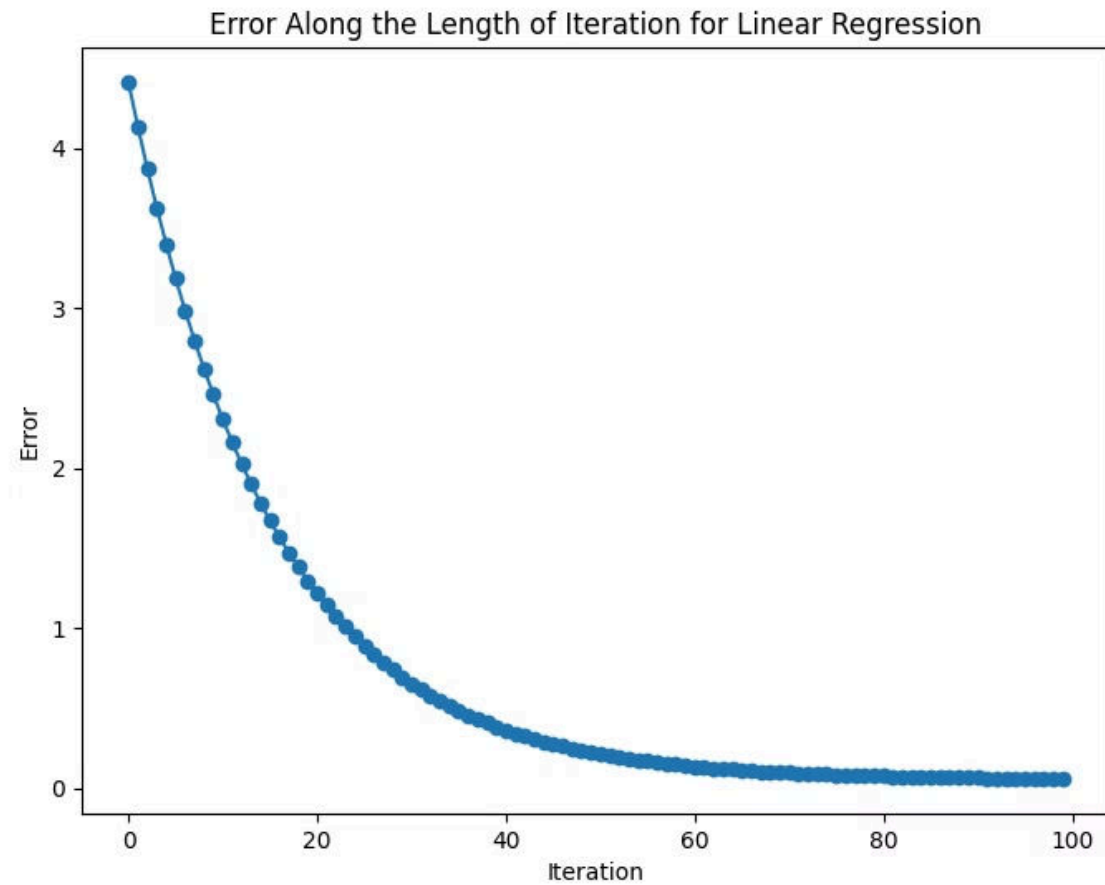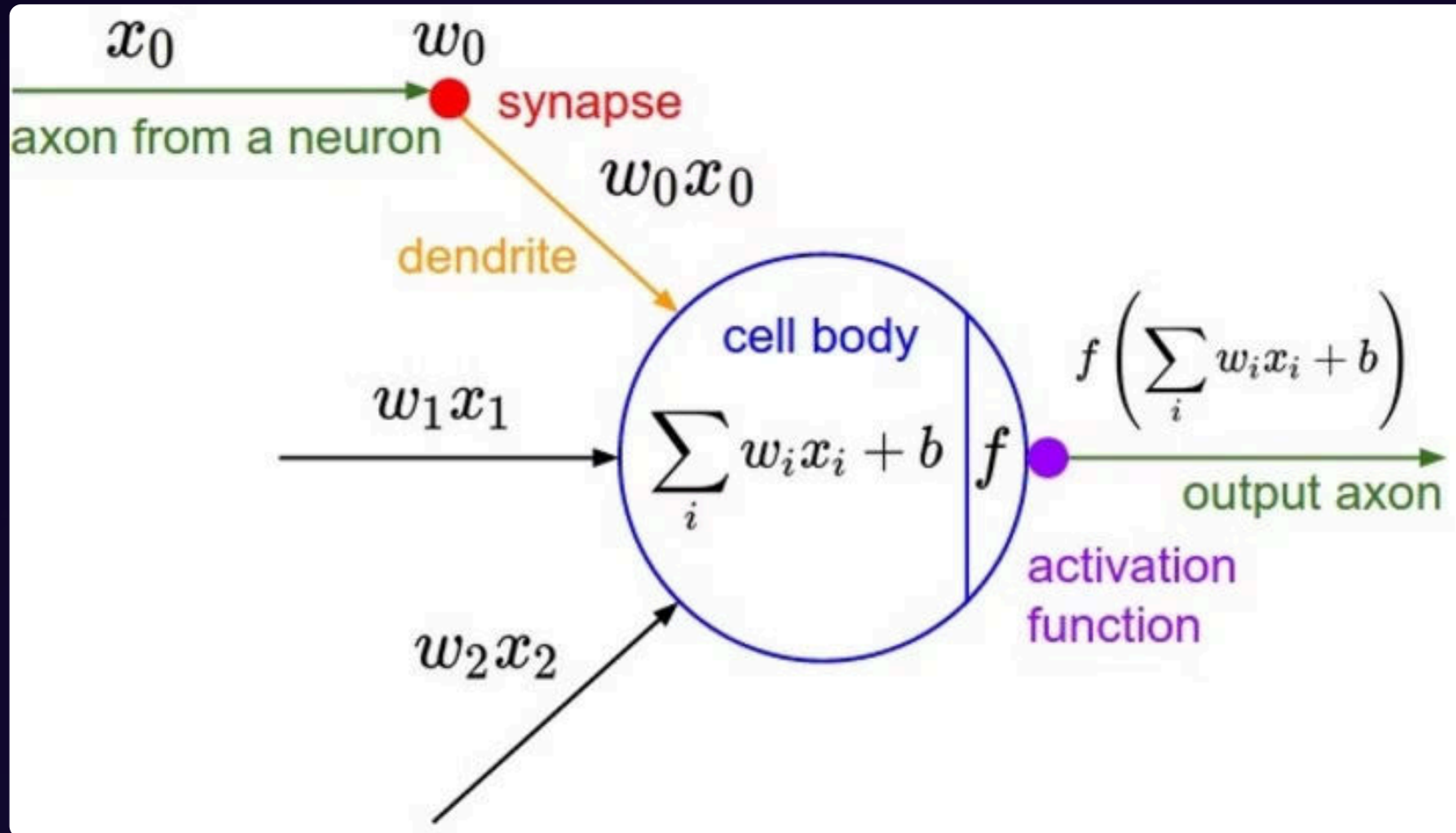
Non-Linear Function Approximations

We have utilized the use of following preprocessing methods
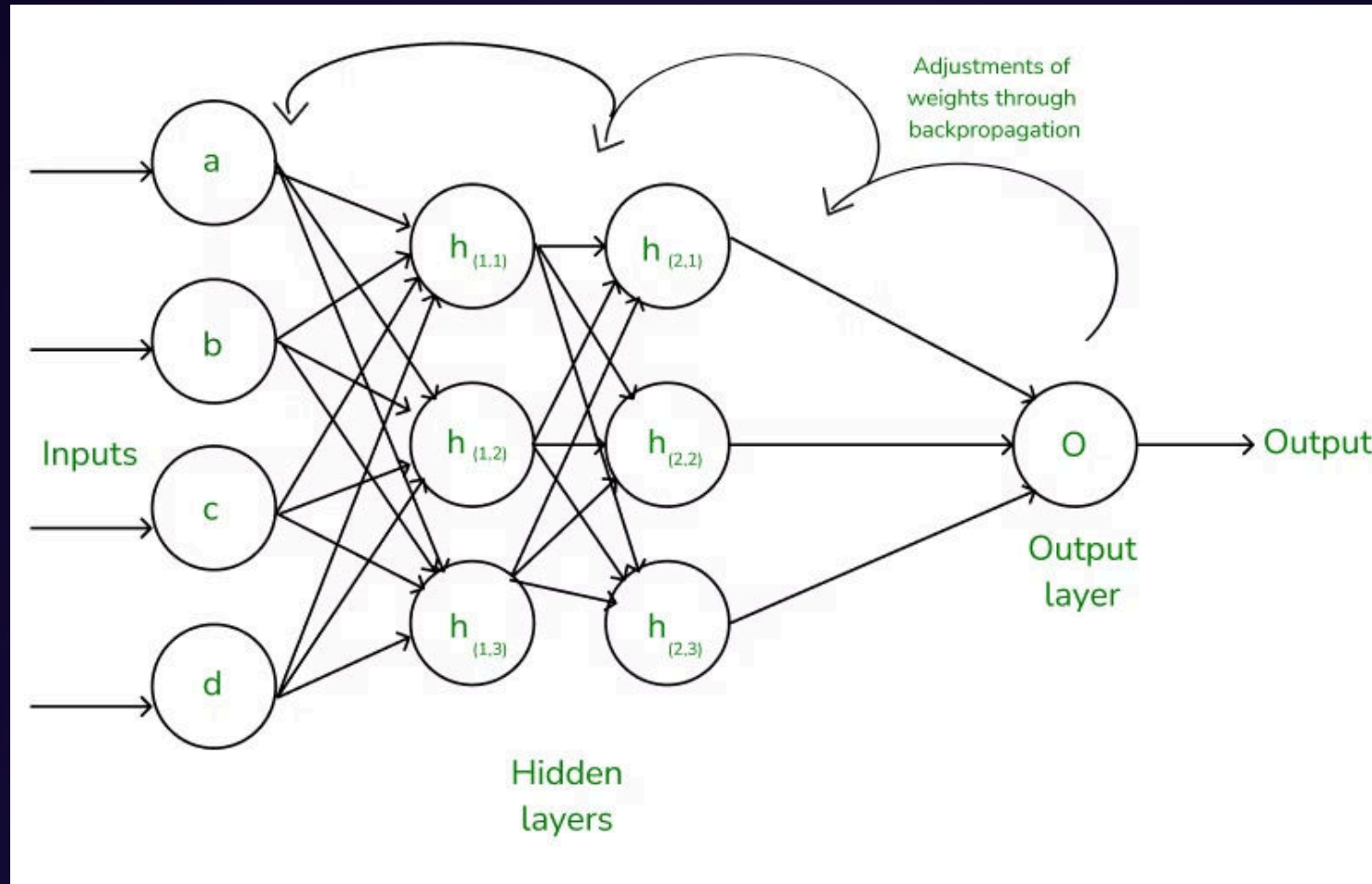
- PCA based feature selection.

We have also built following Models to train and evaluate the model and to predict the soft sensor value

- Neural Network with Single Hidden Layer
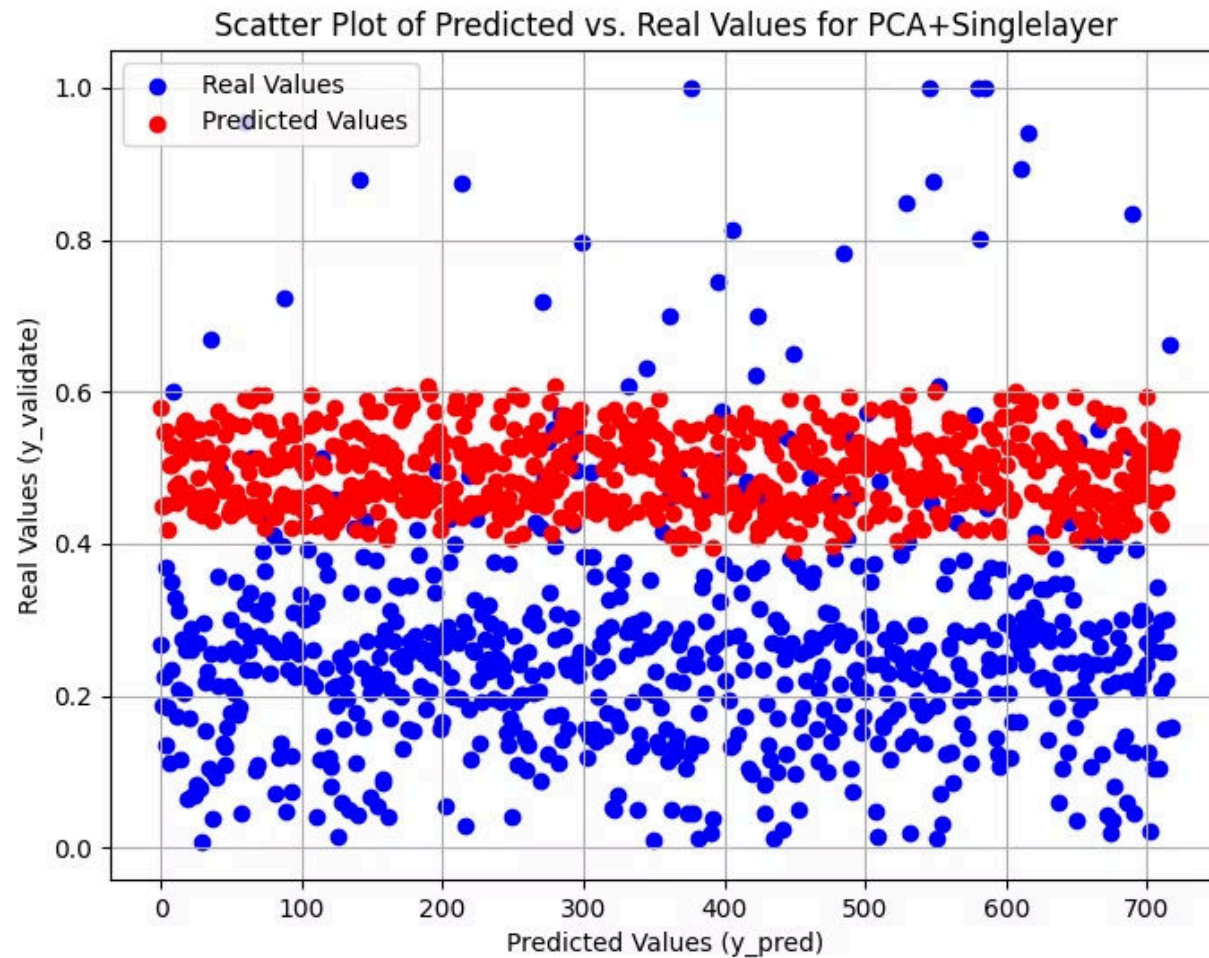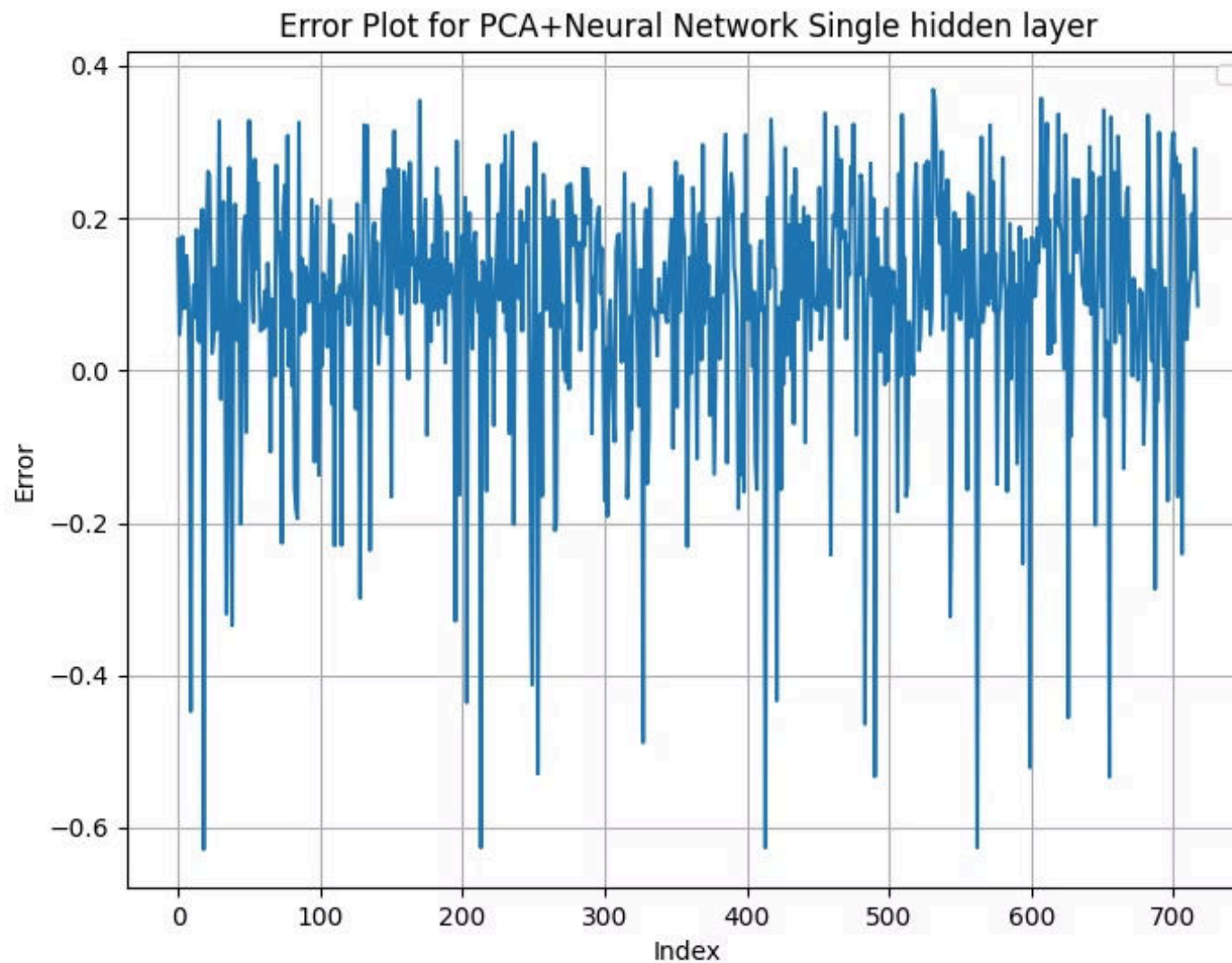- Neural Network with Multi Hidden Layer

Neural Networks

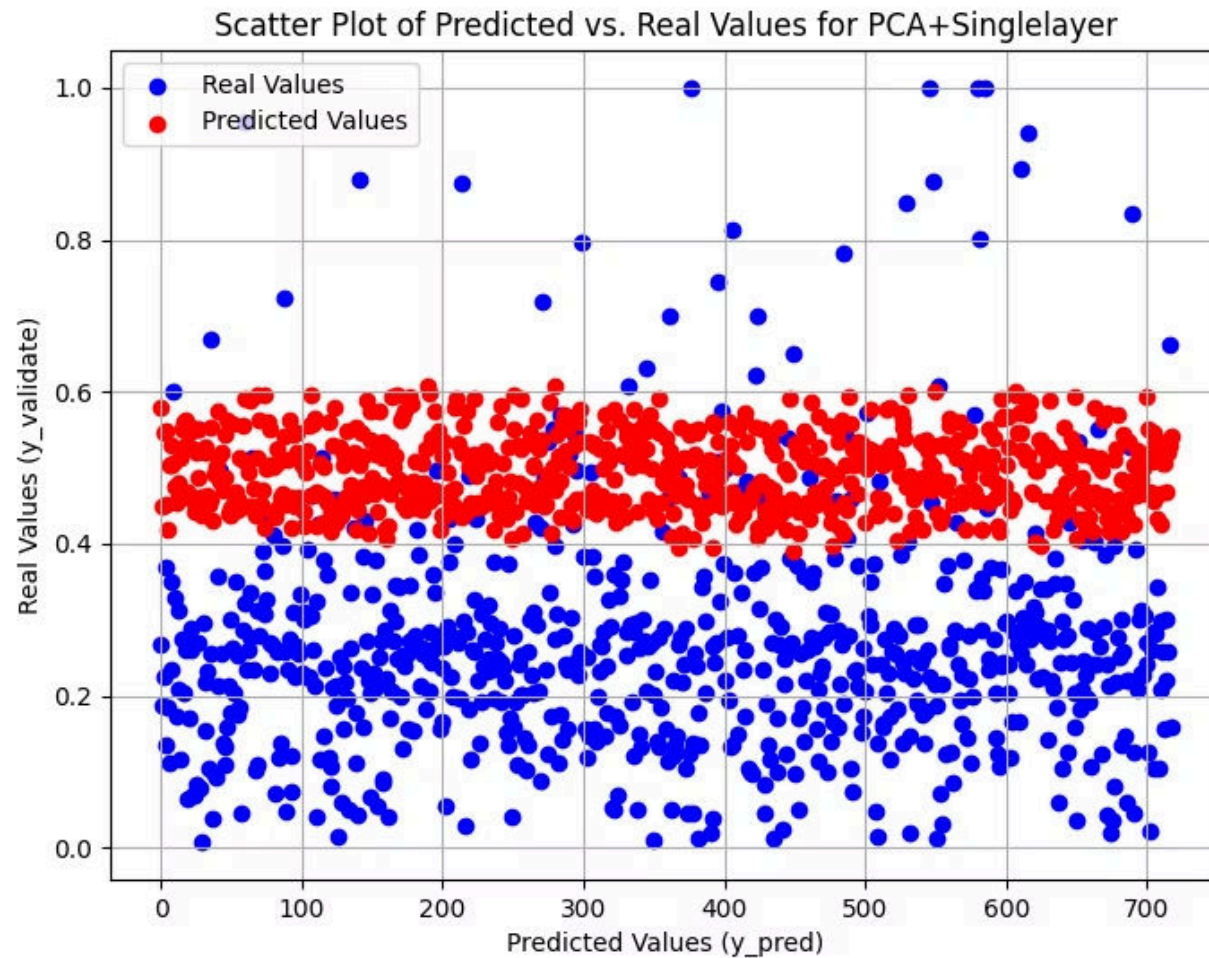# Back Propagation in Neural Networks

PCA with Single Hidden Layer Neural Network

Scatter Plot of Predicted vs. Real Values for PCA+Singlelayer

PCA with Single Hidden Layer Neural Network Error

Error Plot for PCA+Neural Network Single hidden layer

PCA with Multi Hidden Layer Neural Network

Scatter Plot of Predicted vs. Real Values for PCA+Singlelayer
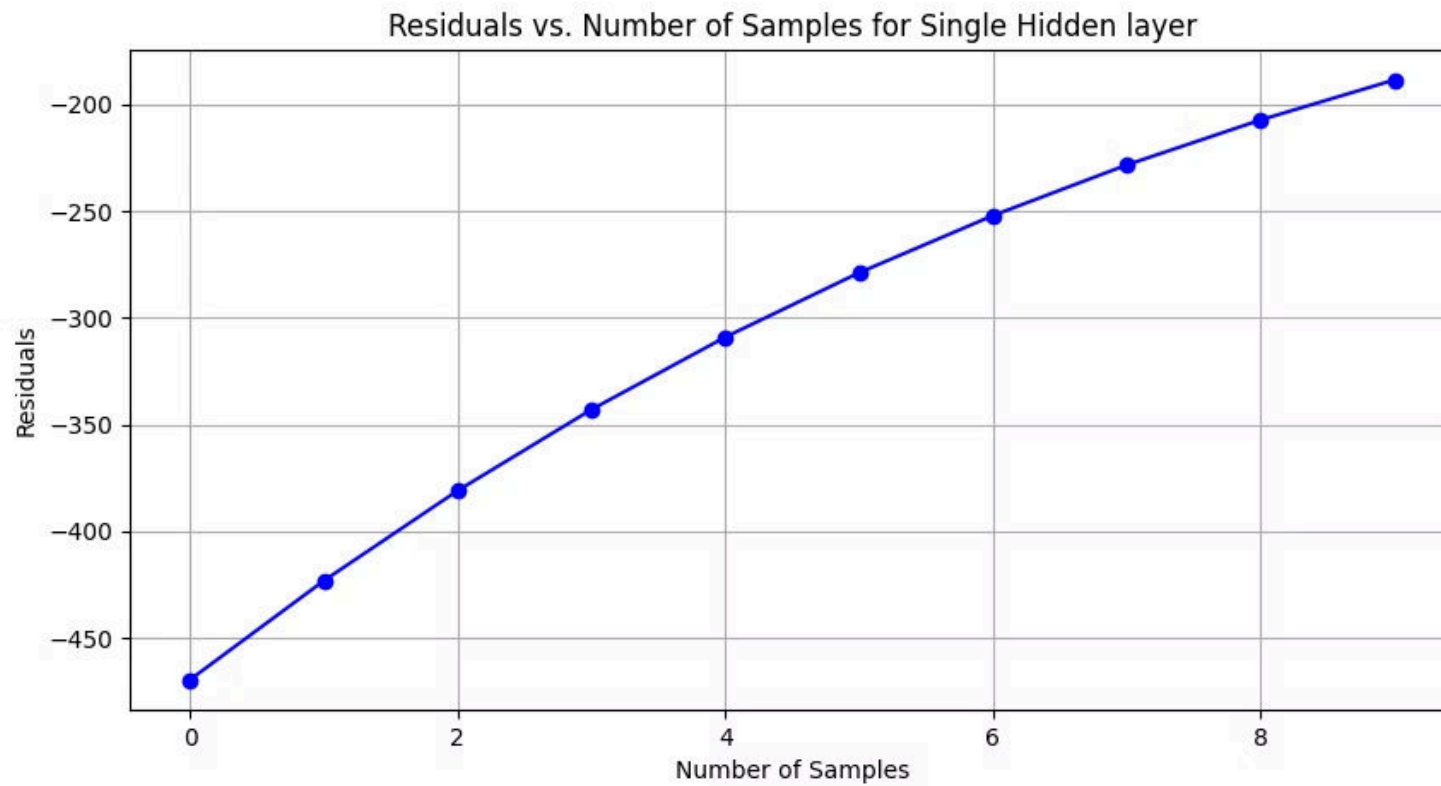
PCA with Multi Hidden Layer Neural Network Error

Residuals Neural Network Single Hidden Layer
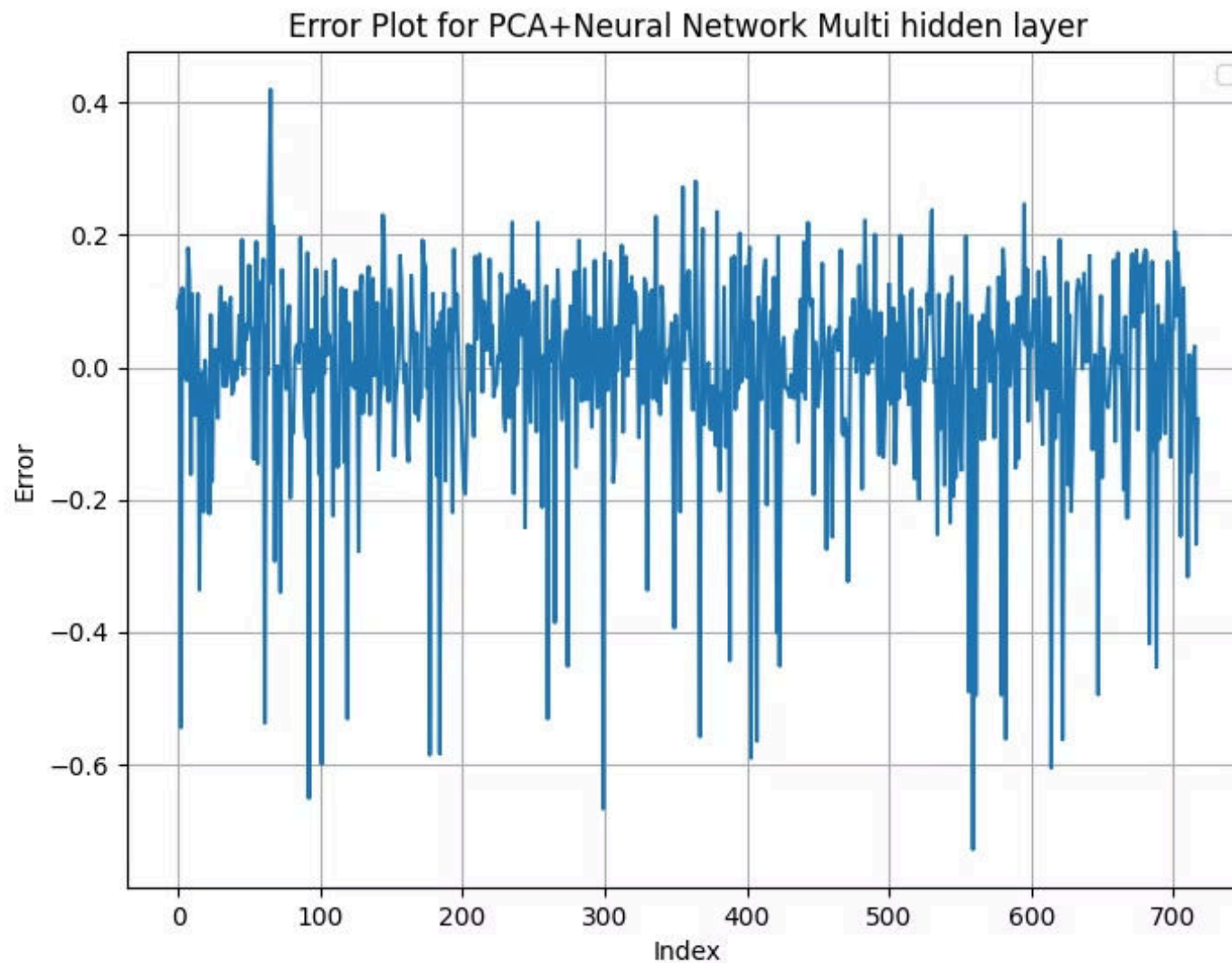
Residuals vs. Number of Samples for Single Hidden layer

Residuals Neural Network Multi Hidden Layer

Residuals vs. Number of Samples for Multi Hidden layer
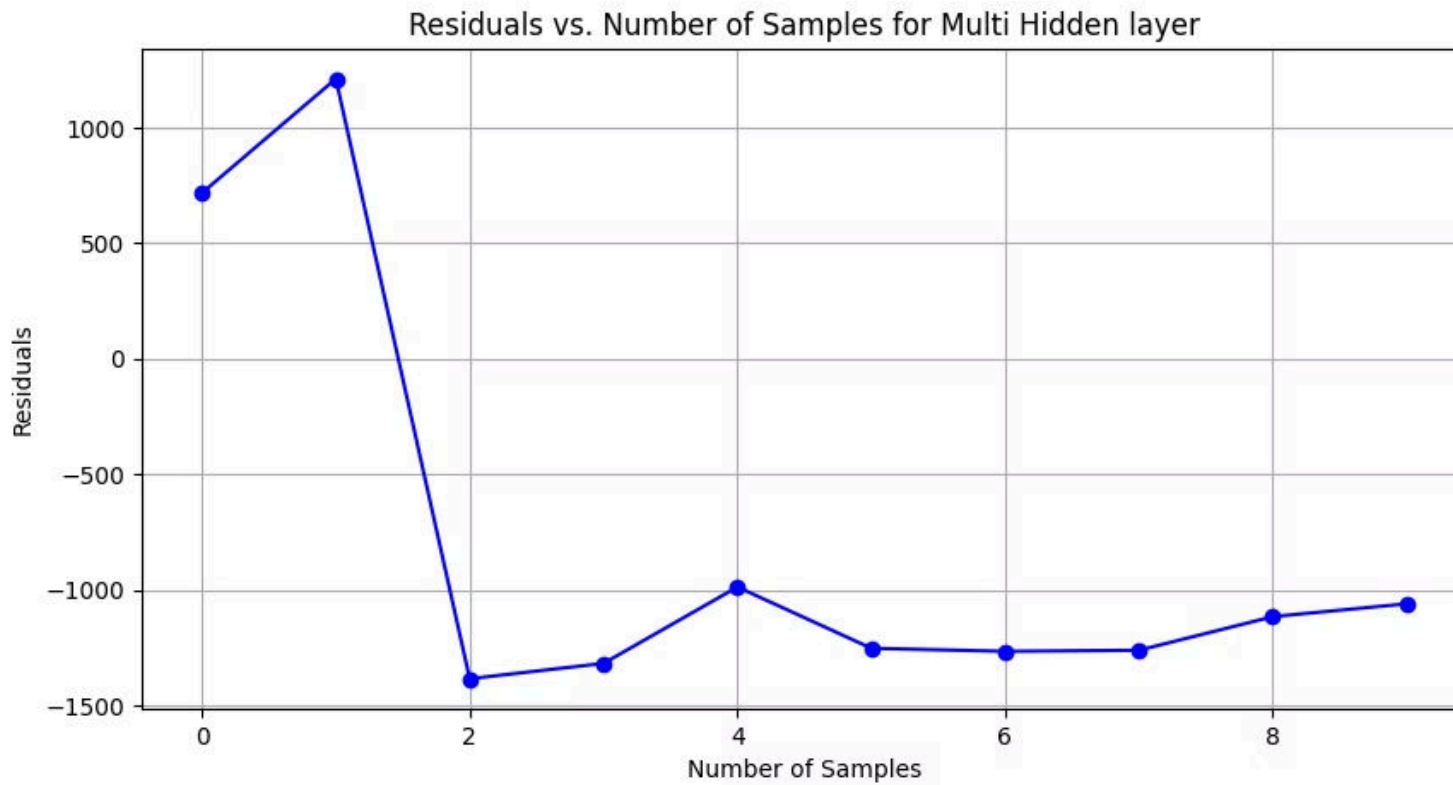
Comparing the Non Linear Model

Model	Mean Squared Error	AIC
NN with Single Hidden Layer	0.028388499570481444	11.123
NN with Mutli Hidden Layer	**0.03391075838059221**	**10.768**

Just In Time Model

- Average MSE loss = 0.09105176840288019
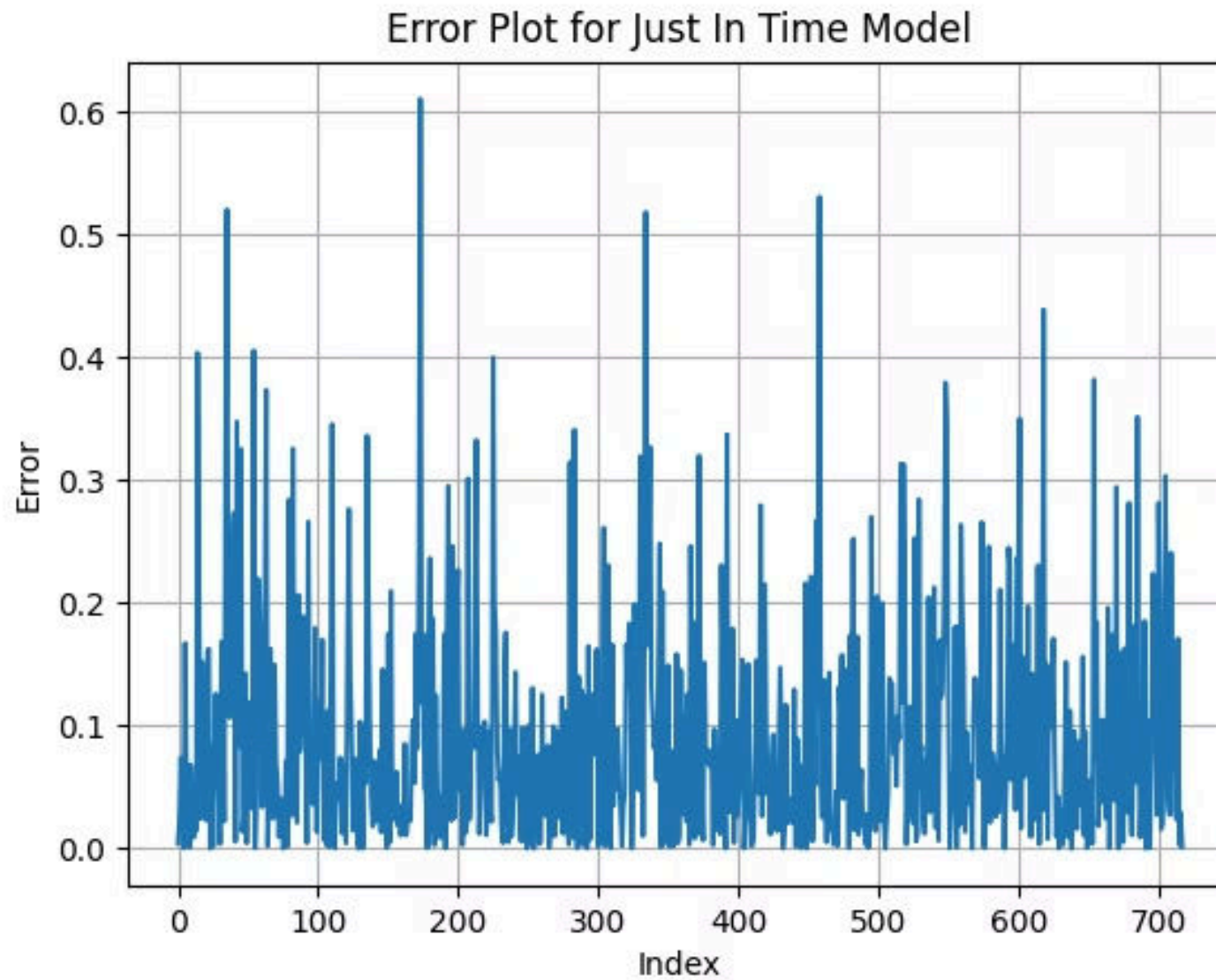
Conclusion

We have successfully implemented all the preprocessing methods along with both linear and non linear models, and have validated them with metrics such as R2, MSE, AIC, BIC and compared them accordingly.

The best model came out to be lasso regression with polynomial model with 3 degree with respect to MSE metric

The best model came out to be PCA with Linear Regression with respect to AIC metric.