Introduction to PMMA Reactor Unit Modeling

In this presentation, we will explore the development of a predictive model for the Poly-Methyl Methacrylate (PMMA) Reactor Unit. The goal is to accurately predict the monomer concentration, a critical target variable, based on several key input variables.

2022chb1055 - Purab Gupta

2022chb1038- Aditya Vikram Sharma

Background on PMMA Reactor Unit

The PMMA (Poly-Methyl Methacrylate) Reactor Unit is a critical component in the production of PMMA, a versatile thermoplastic material widely used in various industries. This reactor system is responsible for the polymerization of methyl methacrylate monomers to form PMMA polymer chains. Therefore, accurate prediction of the monomer concentration is important for optimizing the reactor performance and product quality.

We will now adhere to this problem by various regression and deep learning techniques

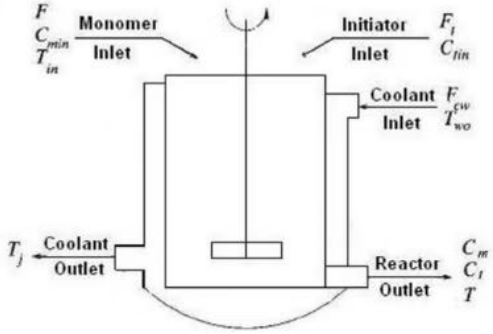


Fig. 2. MMA polymerization reactor flow sheet

System Parameters

INPUT VARIABLES

1. **Coolant water flowrate (Fcw)**: It controls the rate at which heat is removed from the system, ensuring optimal temperature conditions for the polymerization process. Adjusting the coolant water flowrate can have a direct impact on the overall reaction kinetics and the quality of the PMMA polymer produced.
2. **Monomer inlet Flowrate (F) :** It determines the rate at which the methyl methacrylate monomers are fed into the system for polymerization.
3. **Initiator Concentration(CI) :** It determines the rate at which the polymerization reaction is initiated and affects the overall reaction kinetics.

Variables

4.Temperature (T) : it determines the reaction rate and the extent of polymerization.

5.**Jacket Temperature (T):** It refers to the temperature of the reactor's outer jacket, which helps to maintain a specific temperature range throughout the reaction.

6.Coolant inlet temp.water (Two) : The coolant inlet temperature (TWC) in the PMMA Reactor Unit refers to the temperature of the water used as a cooling agent in the system. It is a crucial parameter to preventing any potential overheating.

7.**Inlet temperature of Feed (Tin) :** This refers to the temperature of the monomer and initiator solution that is fed into the reactor to start the polymerization process.

OUTPUT VARIABLES

Monomer Concentration : The primary target variable in this PMMA reactor modeling task is the monomer concentration.

Data Preprocessing

The raw data was cleaned, validated, and preprocessed to handle missing values, outliers, and ensure consistency across the dataset.

Removing outliers

and filtering data that may skew the analysis . This ensures the accuracy and reliability of the analysis by eliminating any anomalies that may interfere with the modeling process.

Test - Train Split

: The dataset is then split into a test set (30 % of dataset) and a train set (70% of dataset) .

Algorithms Applied

1. Feature Selection
   - Linear Regression
   - Polynomial Regression
2. Applying Principal Component Analysis +PCR
   - Linear Regression
   - Polynomial Regression
   - Lasso Regression
3. ANN (single pass)
4. ANN (multiple hidden layer)
5. KNN

Feature Selection

These techniques help to identify the most significant features, reducing computational complexity and improving model performance.

In our case we have used the following feature selection methods

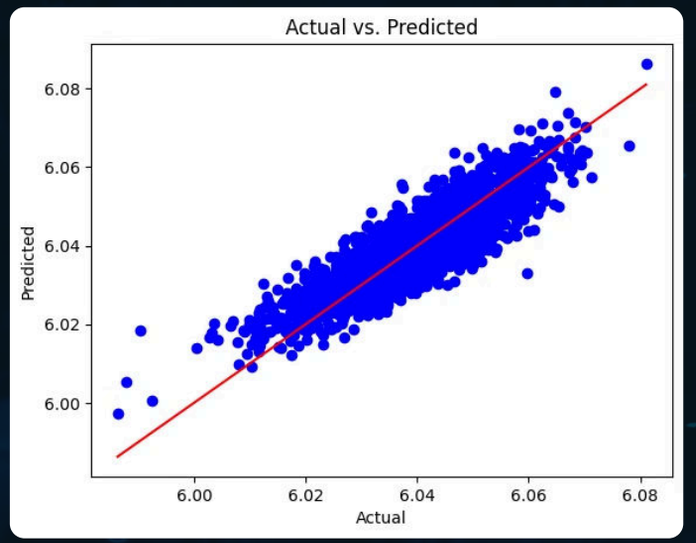
1. PCA - Regression Models
2. Backward Selection Model - ANN (using RFE)

important variables identified

1. Initiator Concentration
2. Monomer Inlet Flowrate

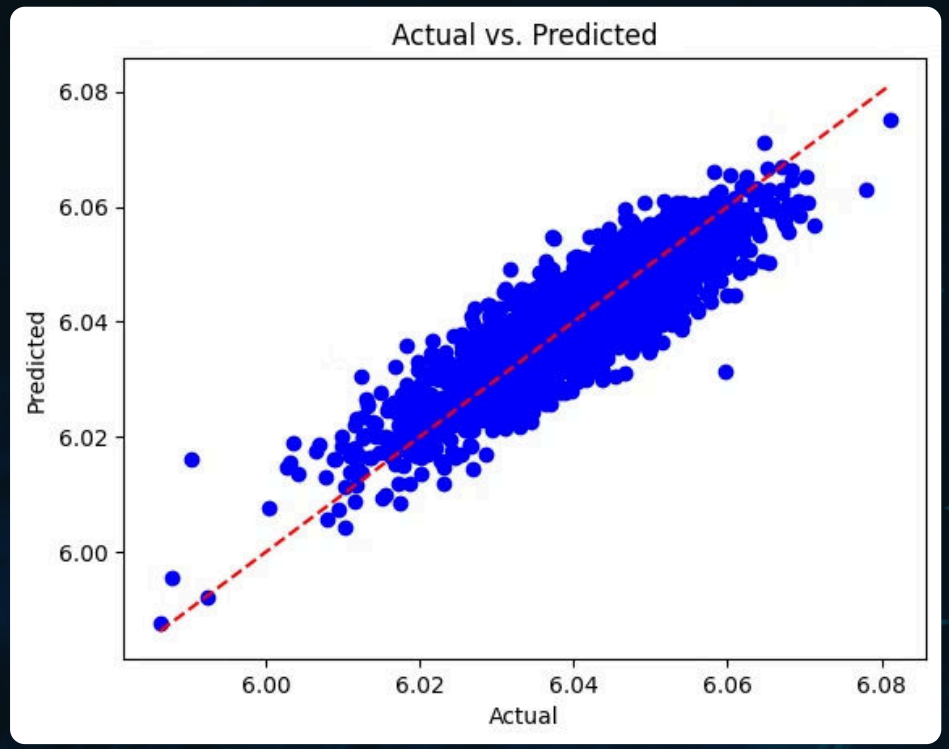
Linear Regression Model

MSE	3.726761654178907e-05
R^2	0.7590920922526123
AIC	-61178.314785823335
coefficients	[223.691907016.23043635]



Polynomial Regression Model

MSE	3.726761654178907e-05
R^2	0.8685903868061529
AIC	-42823.02035007634



We begin by rewriting $\mathbf{C_Y}$ in terms of the unknown variable.

$$\begin{aligned}
\mathbf{C_Y} &= \frac{1}{n}\mathbf{Y}\mathbf{Y}^T \\
&= \frac{1}{n}(\mathbf{P}\mathbf{X})(\mathbf{P}\mathbf{X})^T \\
&= \frac{1}{n}\mathbf{P}\mathbf{X}\mathbf{X}^T\mathbf{P}^T \\
&= \mathbf{P}(\frac{1}{n}\mathbf{X}\mathbf{X}^T)\mathbf{P}^T \\
\mathbf{C_Y} &= \mathbf{P}\mathbf{C_X}\mathbf{P}^T
\end{aligned}$$

Note that we have identified the covariance matrix of $\mathbf{X}$ in the last line.

Our plan is to recognize that any symmetric matrix $\mathbf{A}$ is diagonalized by an orthogonal matrix of its eigenvectors (by Theorems 3 and 4 from Appendix A). For a symmetric matrix $\mathbf{A}$ Theorem 4 provides $\mathbf{A} = \mathbf{E}\mathbf{D}\mathbf{E}^T$, where $\mathbf{D}$ is a diagonal matrix and $\mathbf{E}$ is a matrix of eigenvectors of $\mathbf{A}$ arranged as columns.[3]

Now comes the trick. *We select the matrix* $\mathbf{P}$ *to be a matrix where each row* $\mathbf{p_i}$ *is an eigenvector of* $\frac{1}{n}\mathbf{X}\mathbf{X}^T$. By this selection, $\mathbf{P} \equiv \mathbf{E}^\mathbf{T}$. With this relation and Theorem 1 of Appendix A ($\mathbf{P}^{-1} = \mathbf{P}^T$) we can finish evaluating $\mathbf{C_Y}$.

$$\begin{aligned}
\mathbf{C_Y} &= \mathbf{P}\mathbf{C_X}\mathbf{P}^T \\
&= \mathbf{P}(\mathbf{E}^T\mathbf{D}\mathbf{E})\mathbf{P}^T \\
&= \mathbf{P}(\mathbf{P}^T\mathbf{D}\mathbf{P})\mathbf{P}^T \\
&= (\mathbf{P}\mathbf{P}^T)\mathbf{D}(\mathbf{P}\mathbf{P}^T) \\
&= (\mathbf{P}\mathbf{P}^{-1})\mathbf{D}(\mathbf{P}\mathbf{P}^{-1}) \\
\mathbf{C_Y} &= \mathbf{D}
\end{aligned}$$

It is evident that the choice of $\mathbf{P}$ diagonalizes $\mathbf{C_Y}$. This was the goal for PCA. We can summarize the results of PCA in the matrices $\mathbf{P}$ and $\mathbf{C_Y}$.

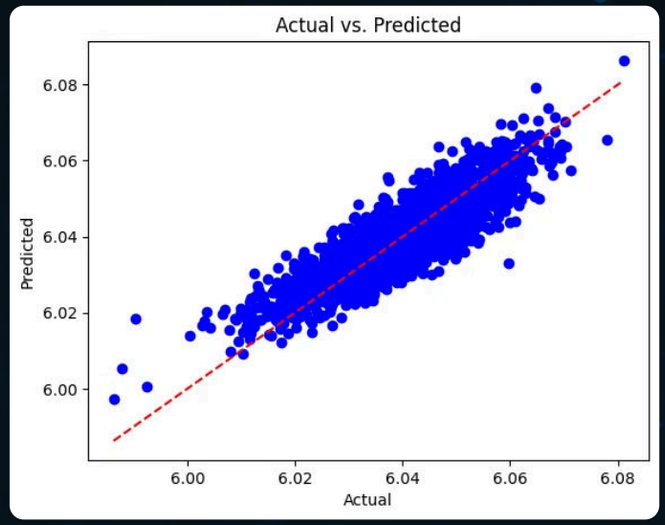
- The principal components of $\mathbf{X}$ are the eigenvectors of $\mathbf{C_X} = \frac{1}{n}\mathbf{X}\mathbf{X}^T$.
- The i^{th} diagonal value of $\mathbf{C_Y}$ is the variance of $\mathbf{X}$ along $\mathbf{p_i}$.

Applying Principal Component Analysis

applied to the preprocessed dataset in order to reduce the dimensionality of the data and identify the most relevant features that contribute to the variation in the dataset.

Linear Regression Model

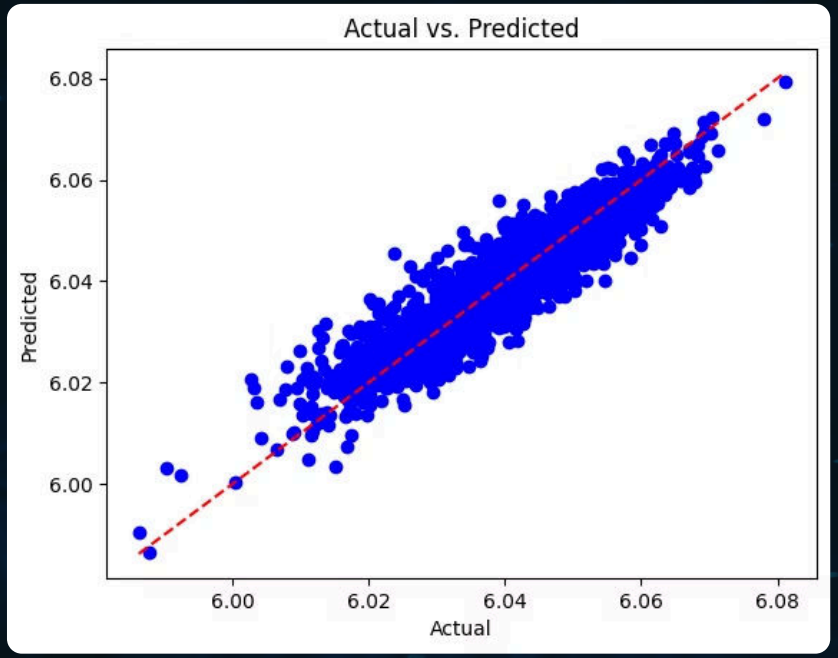
MSE	2.1107854969007497e-05
R^2	0.8635531421249624
AIC	-45210.63431605374



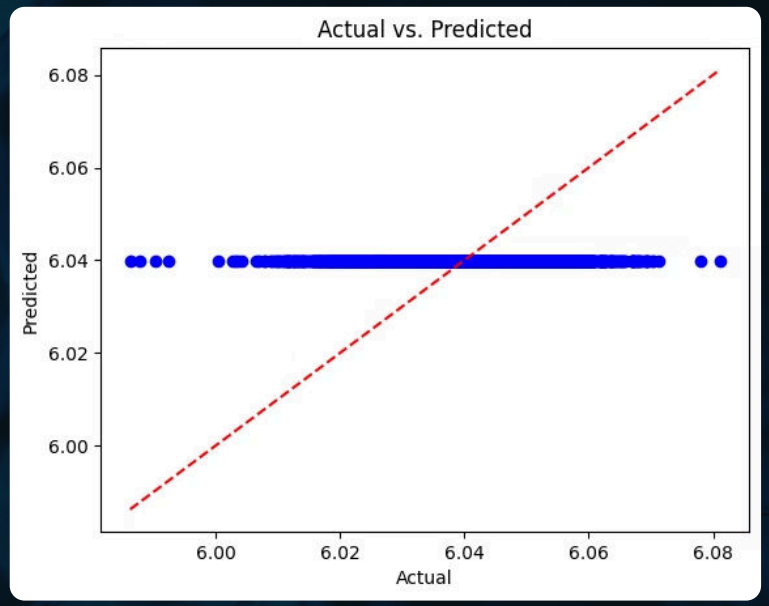
β 0	β1	β2	β3	β4	β5	β6
1.02802240e+02	3.93313771e-05	-8.84973937e-06	-5.25112382e-07	3.35969893e+00	1.85848710e-03	6.35570582e-02

Polynomial Regression Model

MSE	2.1107854969007497e-05
R^2	0.8685903868061529
AIC	-45210.63431605374



Lasso Regression Technique

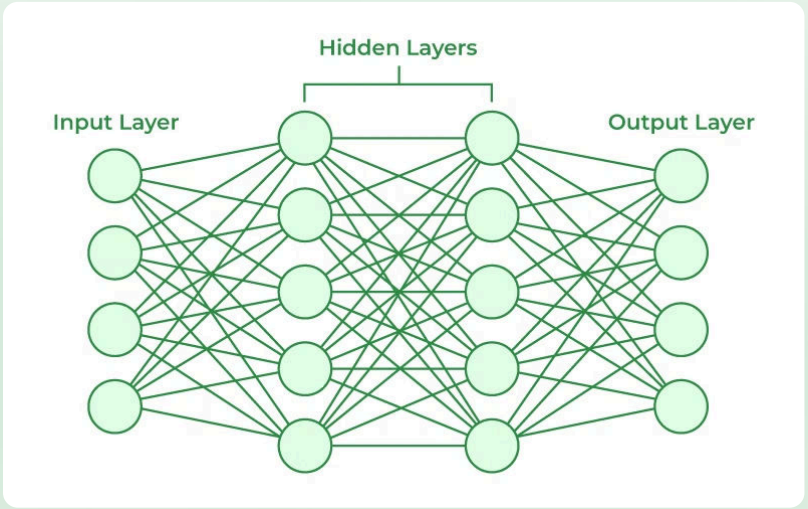


MSE	0.000154757733035724 54
R^2	0.000395655342207312 8
AIC	-36833.32864128485

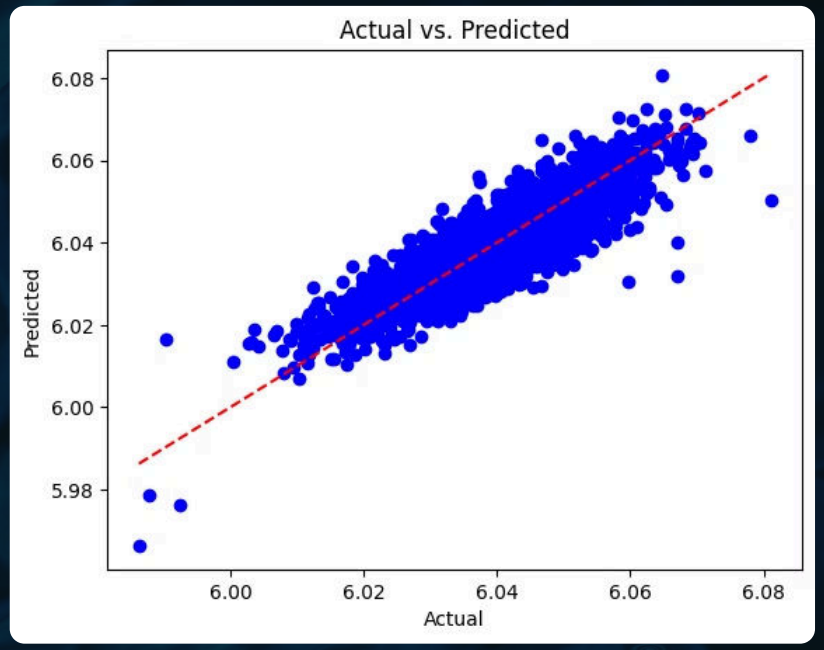
Artificial Neural Network (ANN)

we have used **backward selection model** in ANN

By iteratively removing features based on their contribution to the model, the backward selection model in ANN helps to find the optimal subset of features for prediction.

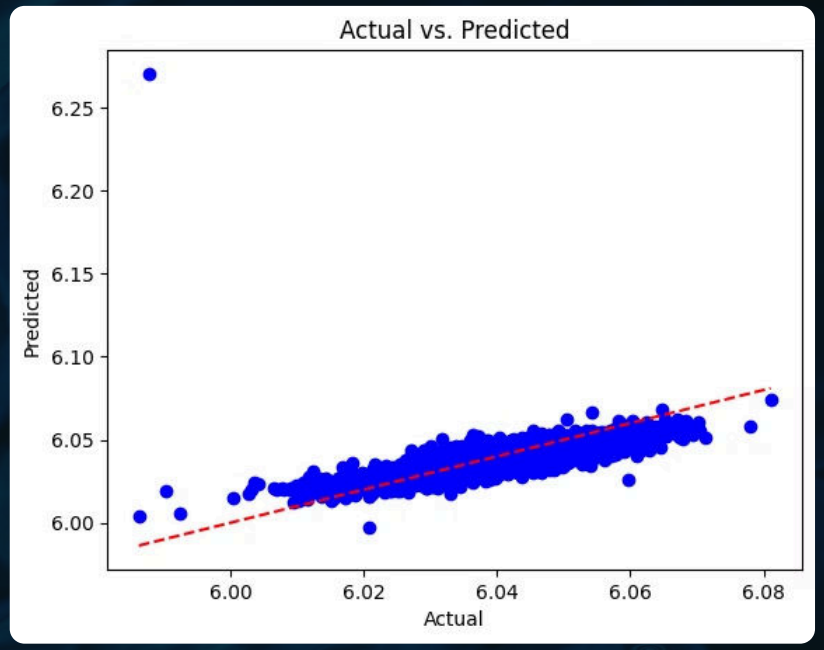


ANN (Single Layer)



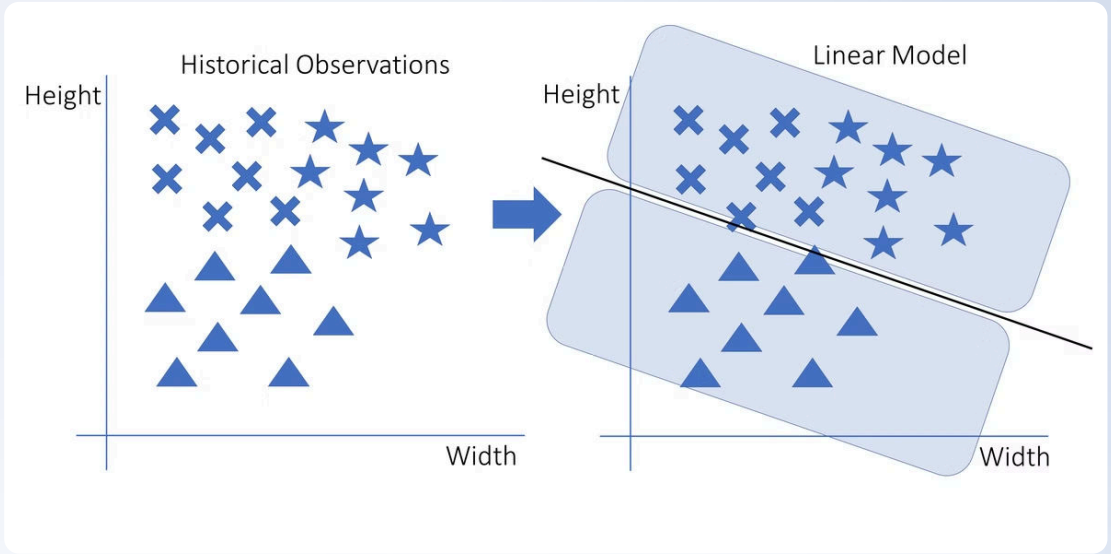
MSE	4.0238493966171 52e-05
R^2	0.7398871354836 924
AIC	-42500.88312847 685

ANN (Multiple Layer)



MSE	4.0238493966171 52e-05
R^2	0.5288274281180 687
AIC	-42500.88312847 685

KNN


Historical Observations
Height
Width
Linear Model
Height
Width

CONCLUSION

Algorithm	MSE	AIC
Linear	3.726761654178907e-05	-61178.31478 5823335
Polynomial	3.726761654178907e-05	-42823.02035 007634
Multi-linear	2.1107854969007497e-05	-45210.63431 605374
Polynomial	2.1107854969007497e-05	-45210.63431 605374
Lasso	0.00015475773303572454	-36833.32864 128485
ANN	4.023849396617152e-05	-42500.88312 847685
ANN (multiple layer)	4.023849396617152e-05	-42500.88312 847685
KNN		

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 PCA with Polynomial Regression with respect to R^2 metric.