CHE F226

TITLE- AI/ML ANALYSIS ON GLUCOSE SENNSOR DATA

Overview- Performing AI/ML models on glucose sensor data to get the patten and correlation between the input and output parameters.

Input parameter-

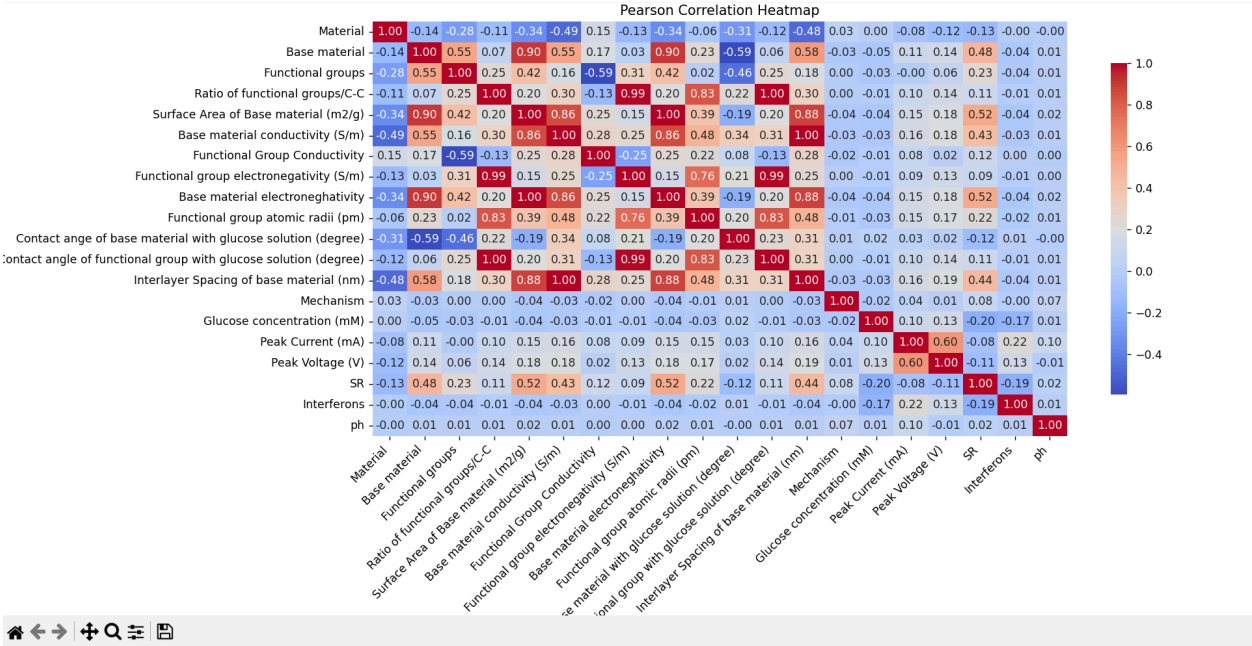
- Material
- Base material
- Functional groups
- Ratio of functional groups/C-C
- Surface Area of Base material (m^2/g)
- Base material conductivity (S/m)
- Functional Group Conductivity
- Functional group electronegativity (S/m)
- Base material electroneghativity
- Functional group atomic radii (pm)
- Contact angle of base material with glucose solution (degree)
- Contact angle of functional group with glucose solution (degree)
- Interlayer Spacing of base material (nm)
- Mechanism
- Glucose concentration (mM)
- Peak Voltage (V)
- SR
- Interferons

- pH

Output Parameter- Peak Current (mA)

ML/AI models used: Pearson correlation, Bayleigh Regression, Polynomial Regression, Random Forest Regression, SVR model, ANN model

PEARSON CORRELATION

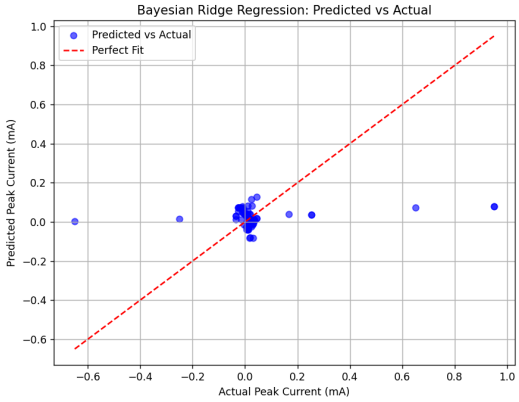


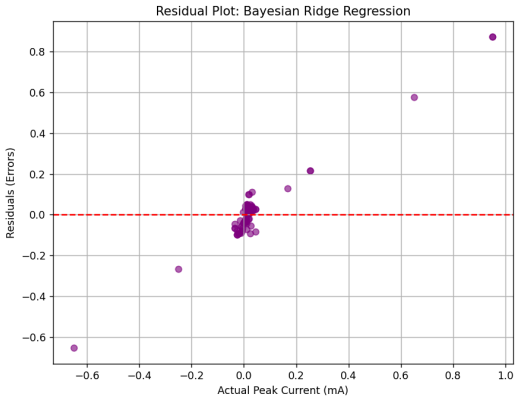
OUTCOME:

- Peak current has the highest +ve correlation with Peak Voltage(V) and highest -ve correlation with Scan Rate.
- Peak current has low correlation with the features.

BAYLEIGH REGRESSION

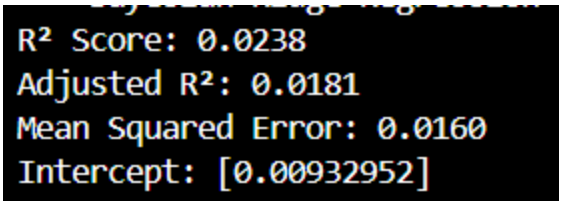
OUTCOME: R square value (-ve except for peak voltage) signified that the model is performing worse than simply predicting the mean of the target variable.



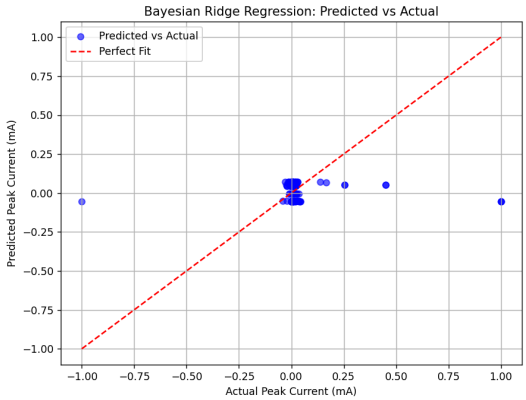


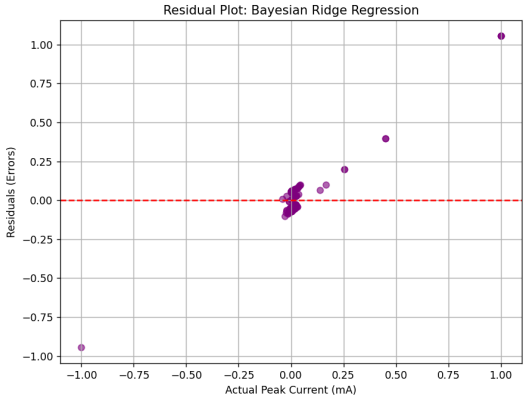
45 degree line: Perfect Prediction

180 degree line: Zero error



(high +ve correlation)Peak Voltage vs Peak current: It shows that though the model is better than simple mean, it is inefficient.





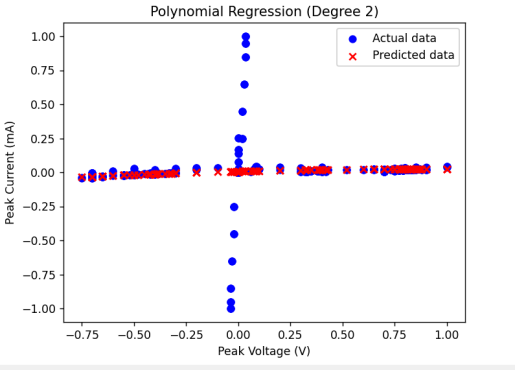
```
R² Score: -0.1584
Adjusted R²: -0.1651
Mean Squared Error: 0.0297
Intercept: [-0.01786931]
```

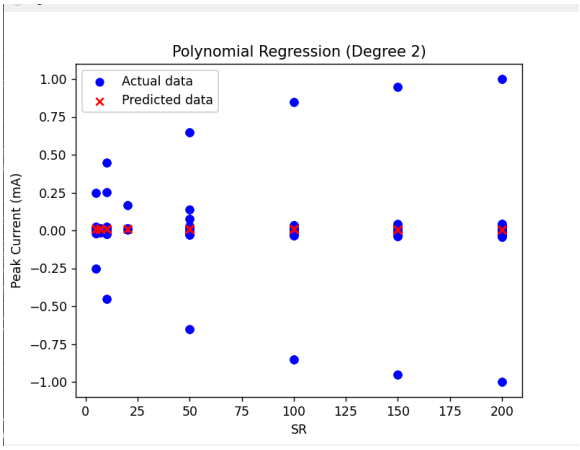
(high -ve correlation)Scan Rate vs Peak Current: It shows that though the model is worse than simple mean.

*OTHER OUTCOMES: All other features showed -ve R square suggesting an unfit model.

POLYNOMIAL REGRESSION

OUTCOME: R square (+ve but <10%) signified that the model is performing better than simply predicting the mean of the target variable but is very inefficient.





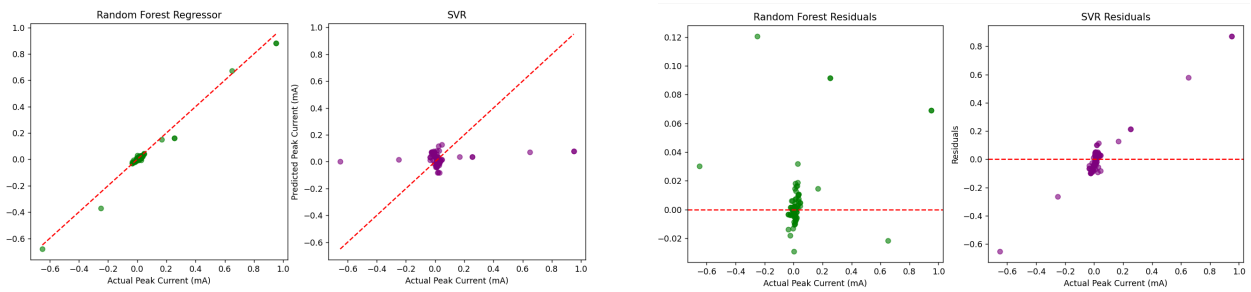

```
R² Score: 0.0066
Adjusted R²: 0.0043
```

(high +ve correlation)Peak Voltage vs Peak Current: It shows that though the model is better than simple mean, it is inefficient.

```
R² Score: 0.0003
Adjusted R²: -0.0020
```

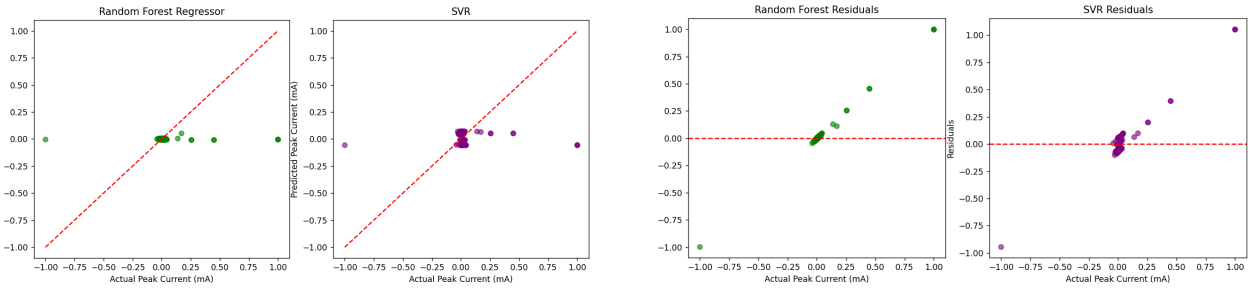
(high -ve correlation)Scan Rate vs Peak Current: It shows that though the model is better than simple mean, it is inefficient.

SVR & RANDOM FOREST REGRESSION




```
SVR R²: 0.0238
Random Forest R²: 0.0238
```

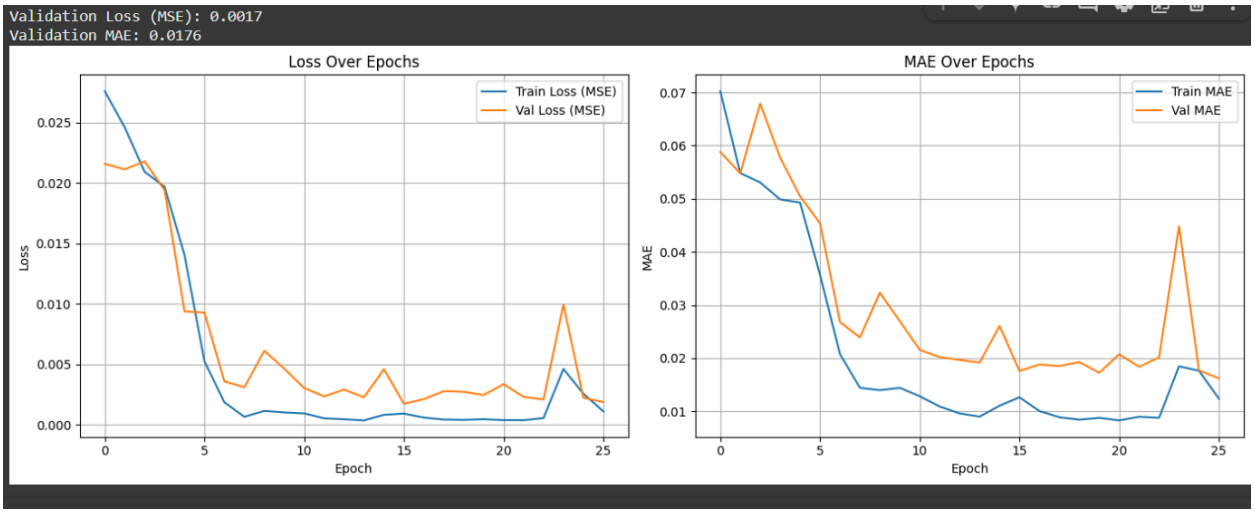
(high +ve correlation)Peak Voltage vs Peak current: It shows that though the models is better than simple mean, it is inefficient.




```
SVR R²: -0.1584
Random Forest R²: -0.1584
```

(high -ve correlation)Scan Rate vs Peak Current: It shows that though the models is worse than simple mean.

Artificial Neural Network (ANN model)



OUTCOME (for Peak Current as Output): The ANN model's performance plots signifies that both training and validation losses (MSE and MAE) decrease over epochs, indicating effective **learning**. However, the consistently lower training loss compared to validation loss suggests potential **overfitting** which is **not** a **major** issue, where the model may be capturing some noise in the training data. Notably, the **final low error values (converged ends of graph)** demonstrate a decent fit.