# Justification for Choosing Lasso Regression

In this project, multiple regression models were evaluated to predict Brix and Pol values from sugarcane juice sensor data (channels R, S, T, U, V, W). The goal was to identify the most suitable regression model with the lowest possible percent error while maintaining interpretability and simplicity for a college-level thesis.

## Models Evaluated

The following regression models were tested:  
- Linear Regression  
- Ridge Regression  
- Lasso Regression  
- Polynomial Regression (degree 2)  
- Random Forest Regressor

## Evaluation Metrics

Each model was evaluated using the following metrics:  
- MAE (Mean Absolute Error)  
- MAPE (Mean Absolute Percentage Error)  
- R² Score (Goodness of Fit)

## Results Summary

### Brix Prediction

|  |  |  |  |
| --- | --- | --- | --- |
| Model | MAE | MAPE (%) | R² Score |
| Linear Regression | 1.62 | 11.21 | -2.06 |
| Ridge Regression | 1.51 | 10.45 | -1.69 |
| Lasso Regression | 1.08 | 7.64 | -0.39 |
| Random Forest | 2.54 | 17.34 | -4.47 |
| Polynomial (Deg 2) | 9.69 | 64.61 | -157.62 |

### Pol Prediction

|  |  |  |  |
| --- | --- | --- | --- |
| Model | MAE | MAPE (%) | R² Score |
| Linear Regression | 2.42 | 20.76 | -1.51 |
| Ridge Regression | 2.21 | 19.16 | -1.10 |
| Lasso Regression | 1.41 | 13.50 | 0.31 |
| Random Forest | 4.66 | 43.39 | -4.57 |
| Polynomial (Deg 2) | 60.13 | 450.90 | -3113.73 |

## Why Lasso Regression Was Chosen

Lasso Regression provided the best trade-off between accuracy and simplicity. It produced the lowest MAPE for Brix (7.64%) and an acceptable MAPE for Pol (13.50%) while keeping the model interpretable through a simple linear formula with 6 coefficients and 1 intercept.

Unlike more complex models like Random Forest and Polynomial Regression, Lasso Regression avoids overfitting and allows us to extract a human-readable equation, which is more suitable for undergraduate thesis projects.