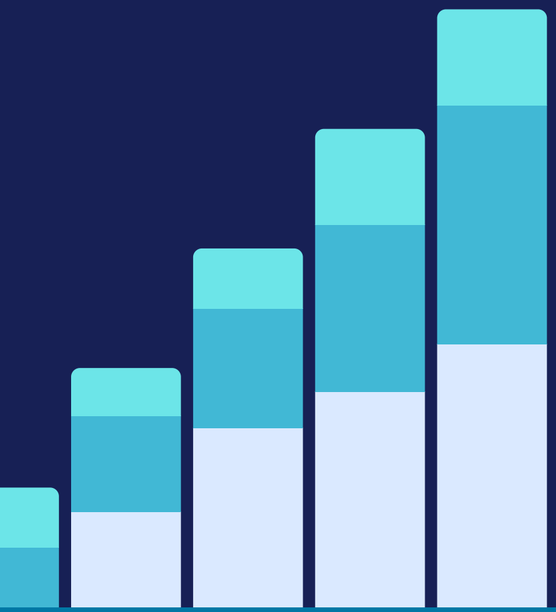
LASSO REGRSSION USING PYSPARK MLLIB

JYOTHI MUDALU-2309039
TOM GEORGE-2309044

In many data modeling tasks, the challenge lies in managing high-dimensional datasets, where irrelevant or redundant features can lead to overfitting and hinder the model's ability to generalize effectively to unseen data



WHAT IS LASSO REGRESSION?

Lasso Regression (Least Absolute Shrinkage and Selection Operator) is a type of linear regression that performs both variable selection and regularization. It helps in preventing overfitting and improving model interpretation by penalizing the absolute size of the regression coefficients.

HOW LASSO REGRESSION WORKS?

$$\text{Lasso Objective} = \text{Residual Sum of Squares} + \lambda \sum |\beta_j|$$

Where λ is a tuning parameter and βj are the coefficients.

- Lasso Loss combines Residual Sum of Squares (RSS) with an L1 penalty.
- It promotes feature selection by shrinking less important coefficients to zero.
- Increasing λ reduces variance but too high a value can cause underfitting.

LASSO VS RIDGE

LASSO	RIDGE
Uses L1 regularization, which adds the absolute values of the coefficients to the penalty term.	Uses L2 regularization, which adds the squared values of the coefficients to the penalty term.
Can shrink some coefficients to zero, thus performing feature selection.	Shrinks coefficients but does not set any of them to zero, so it does not perform feature selection.

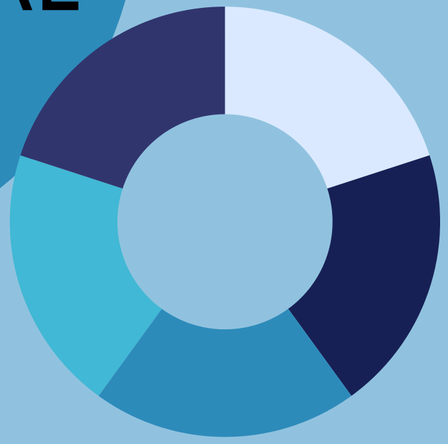
Application of Lasso Regression in Predicting the height of wave

ADVANTAGES OF LASSO REGRESSION

PREVENTS OVERFITTING

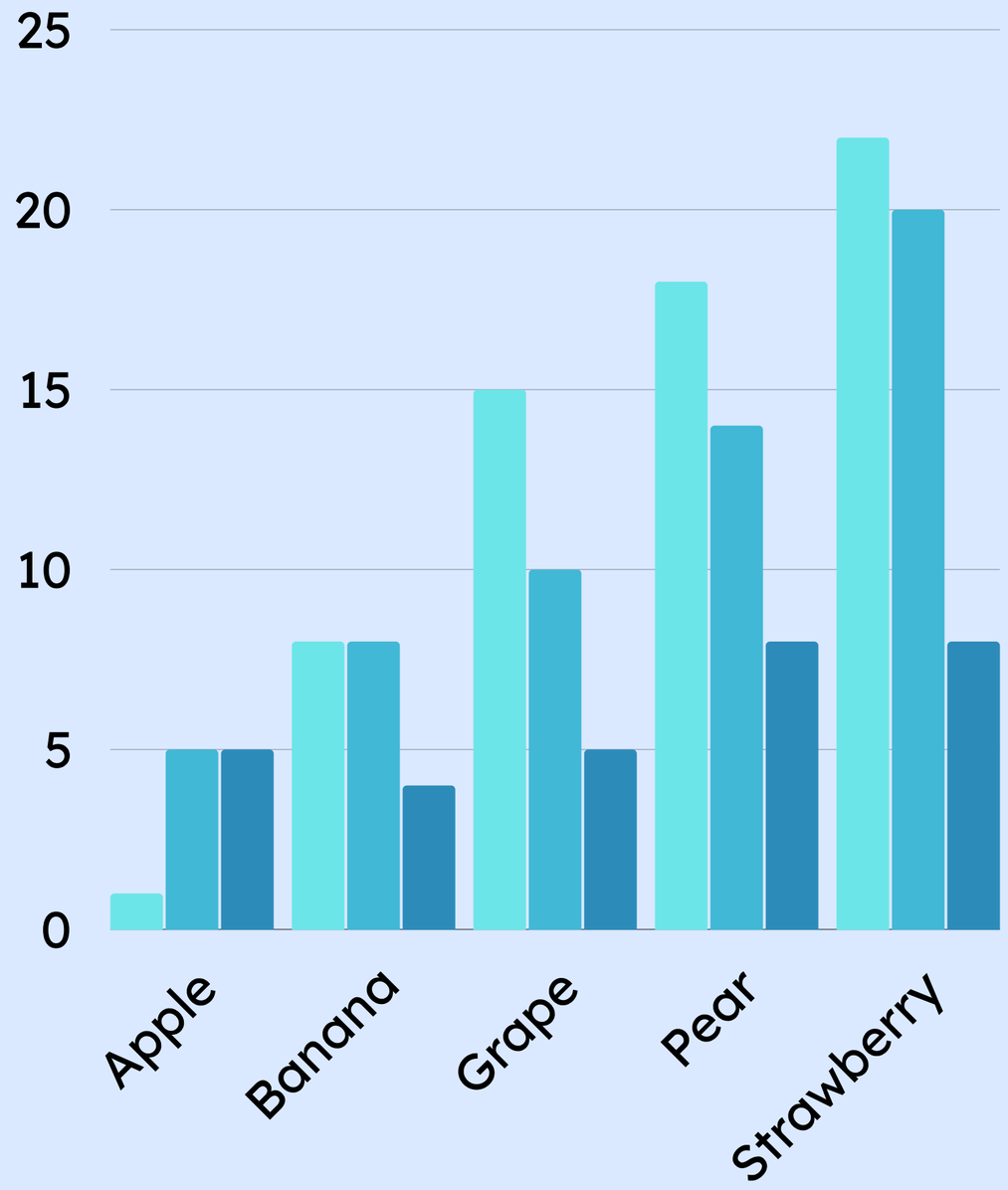
FEATURE SELECTION

HANDLES HIGH-DIMENSIONAL DATA



LIMITATIONS

- Lasso may perform poorly when many predictors are strongly correlated.
- In such cases, Elastic Net might be a better choice as it combines L1 and L2 regularization.




THANK YOU.