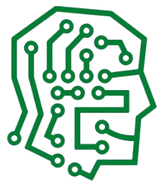
1

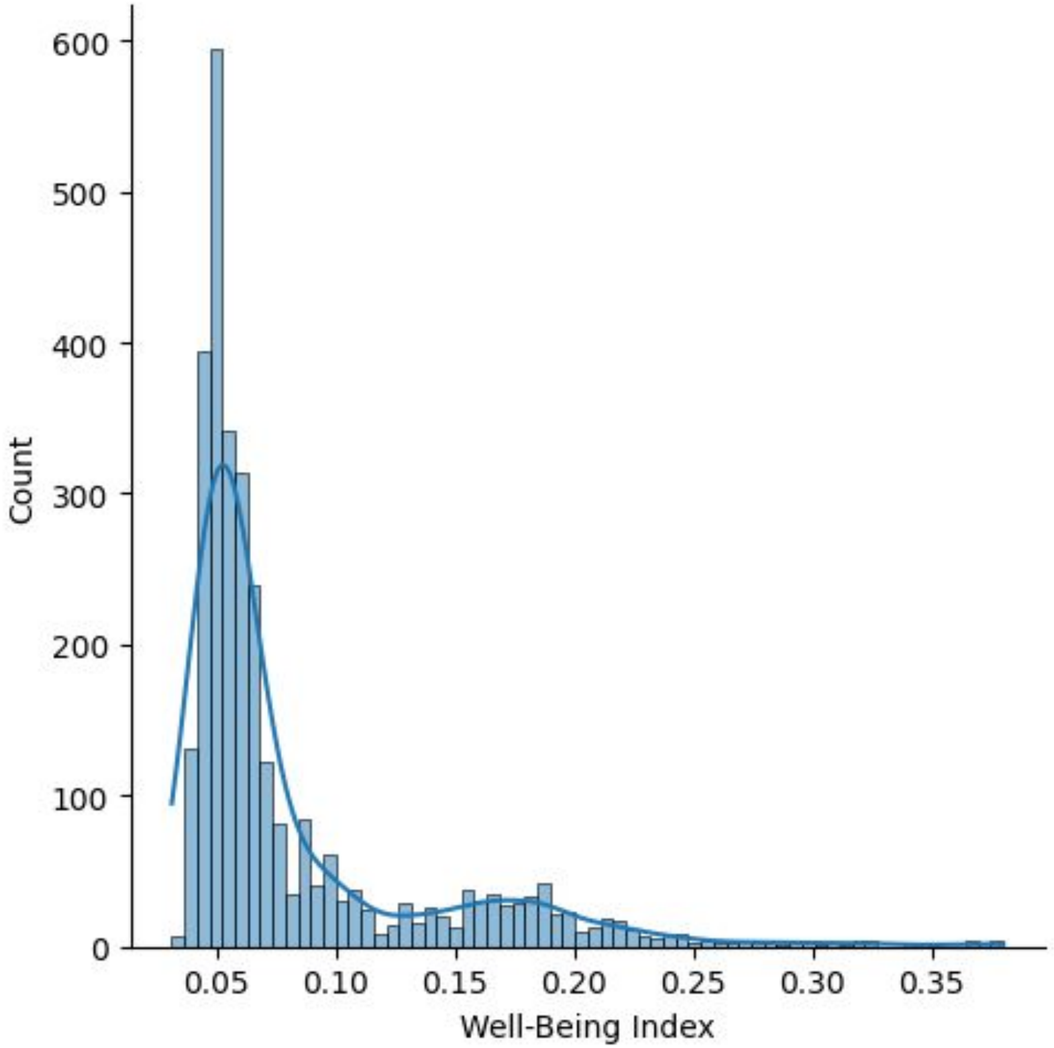
## Introduction

# What Makes the Galaxy Better

- **Galaxies, which features best explains the existence of them**
- **A Regression Analysis Model that predicts the future well being of the galaxies**



# Exploratory Data Analysis




- **3098 rows**

- **continuous label**
- **80 features**

**Distribution of Well-Being Index**

- Skewd to the right therefore the need to scale the values

# Data Engineering

## ENGINEERING PROCESSES

Too much missing values in some columns, dropped the columns

**Eliminating Noise**

**Data Imputation**

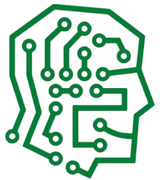
Used Median for replacing Missing values in aach column

Used Robust Scaling on the values since it is not sensitive to outliers

**Scaling The Values**

80% - 20%

split the data into 80% for training and 20% for testing



# Data Modeling

## MODELS USED

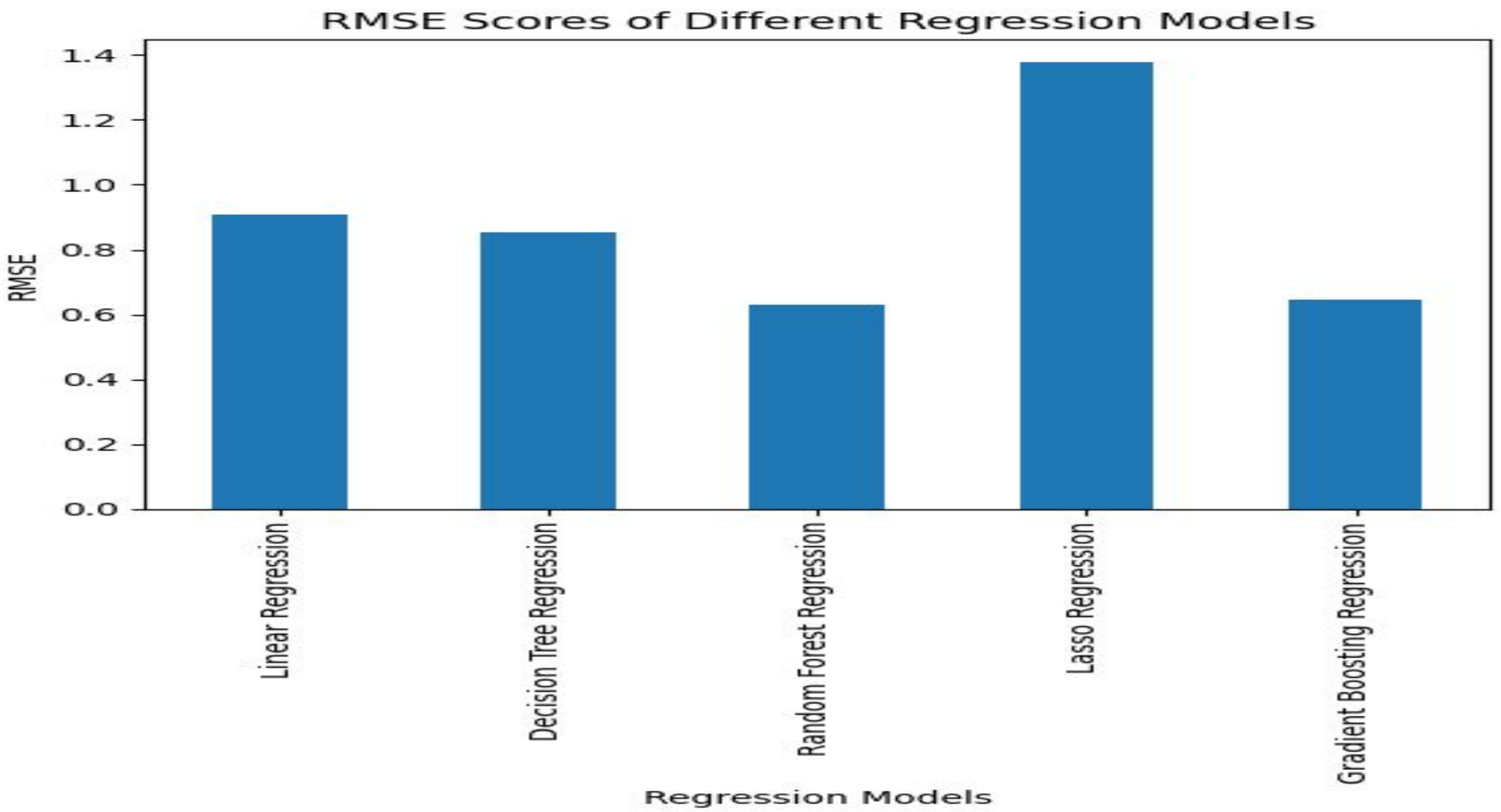
1. Random Forest Regressor
2. Decision Tree Regressor
3. Linear Regression
4. Gradient Boost Regressor

## Accuracy_Metrics

1. RMSE

The smaller the RMSE The better the model performance

# Best Model


RMSE Scores of Different Regression Models
RMSE
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Linear Regression
Decision Tree Regression
Random Forest Regression
Lasso Regression
Gradient Boosting Regression
Regression Models