



Calory Predictor Linear Regression

By:

Shahad Alkaltham
Amjad Althinyyan

TABLE OF CONTENTS

01

Problem Statement

What is calory predictors ?

02

Data Preprocessing

From data collection to ready for modeling data

03

Best Model Selection

What is the optimal model for calory prediction ?

04

Feature Engineering

How score improved ?

05

Reporting the Result

What is the test set score and MAE ?

06

Calory Predictor Website



01

Problem Statement

Predicting food calories based on some nutrition's such as: Carbohydrates, Iron, and Sodium. To help Healthy people monitoring their daily calories



Data Preprocessing

02



Data Collecting



Data Cleaning



Data Log Transformation

02

Data Preprocessing

Data Collecting



SPROUTS
FARMERS MARKET

+

**TRADER
JOE'S®**

Web Scrabing on

Using



BeautifulSoup

02

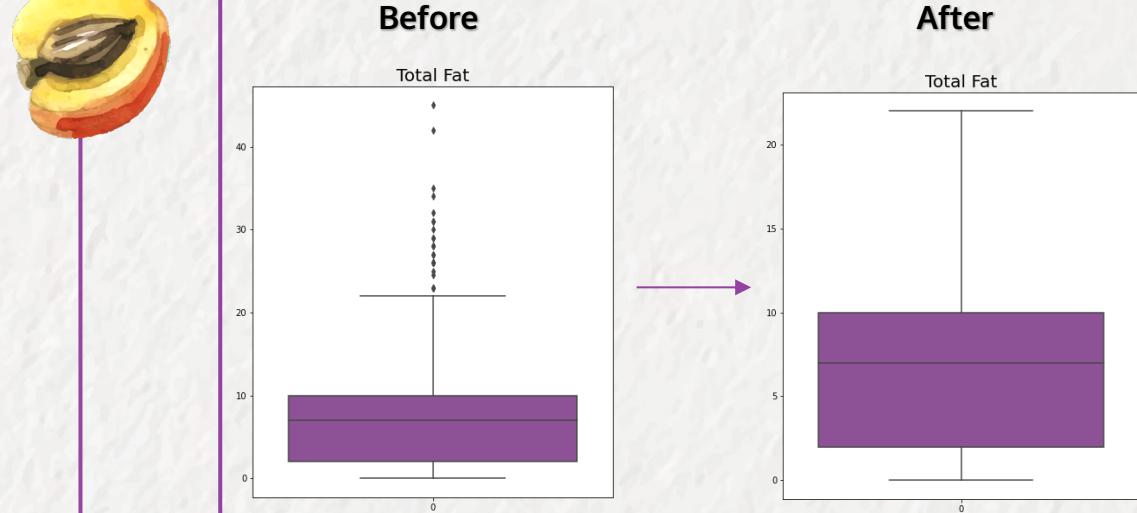
Data Preprocessing

02



Data Cleaning

- 1 Drop unnecessary columns
- 2 Fill nulls with zero
- 3 Replace the outliers with mean



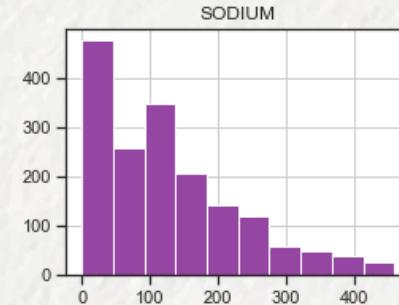
Data Preprocessing

02

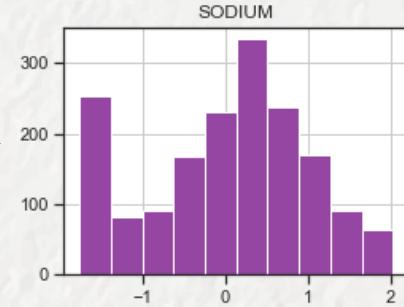


Data Log Transformation

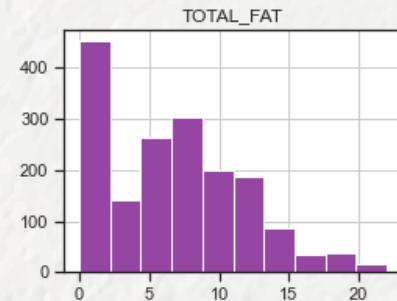
Before



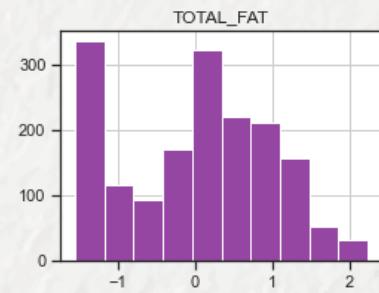
After



Before



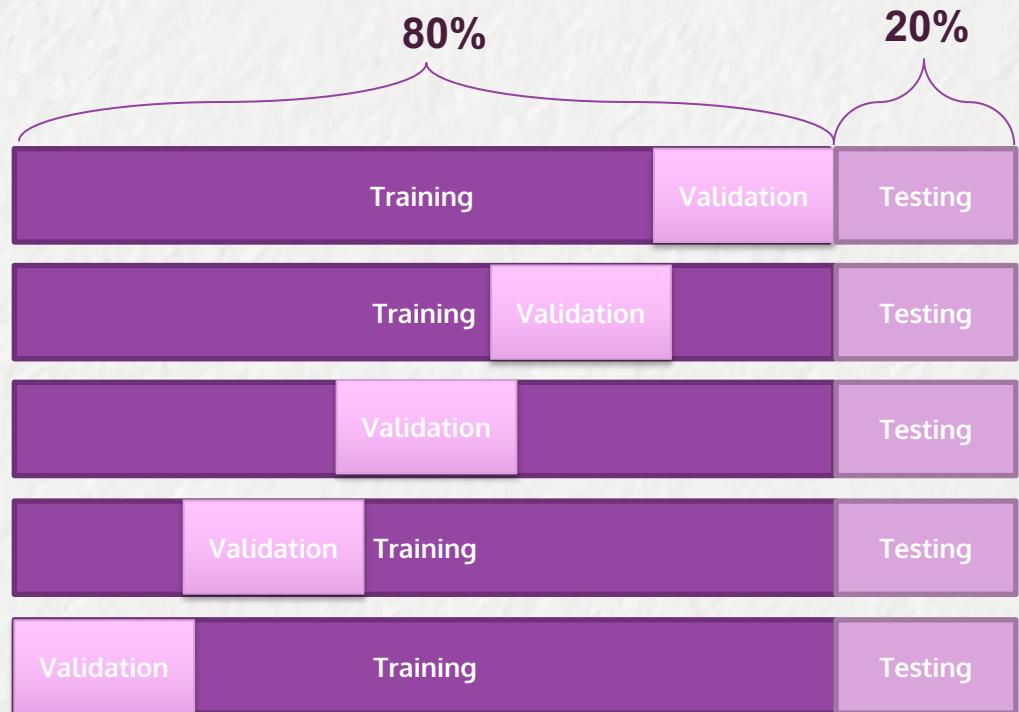
After



Best Model Selection

03

Cross-Validation Approach



Best Model Selection

03

Models Scores



Linear Regression

82.81%



Lasso

82.57%



Ridge

82.81%



Elastic Net

78.26%



Polynomial of Degree 2

82.98%

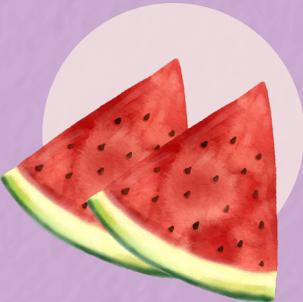


Gradient Boosting Regressor

89.03%

Feature Engineering

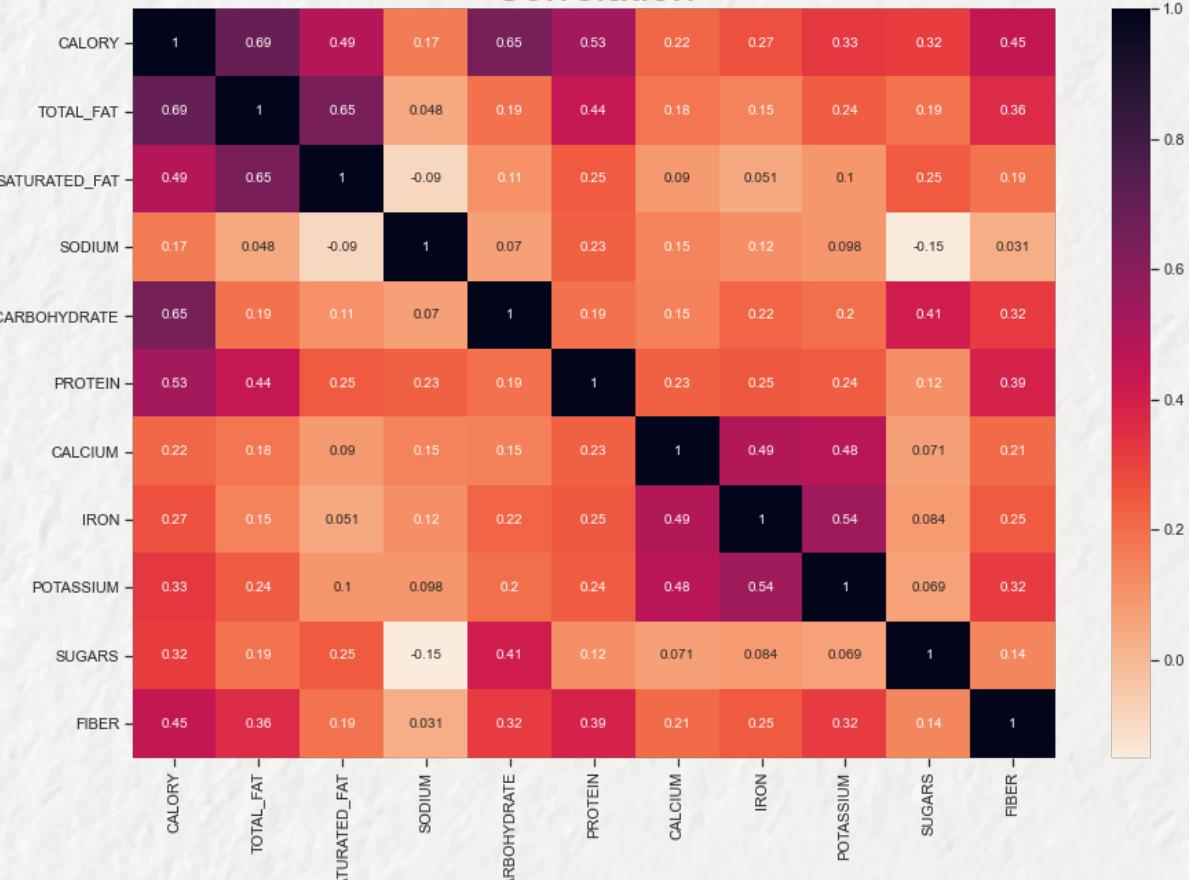
04



1 2 3

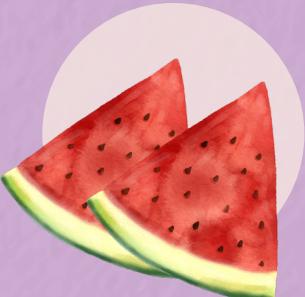
Features and Target Correlation

Correlation

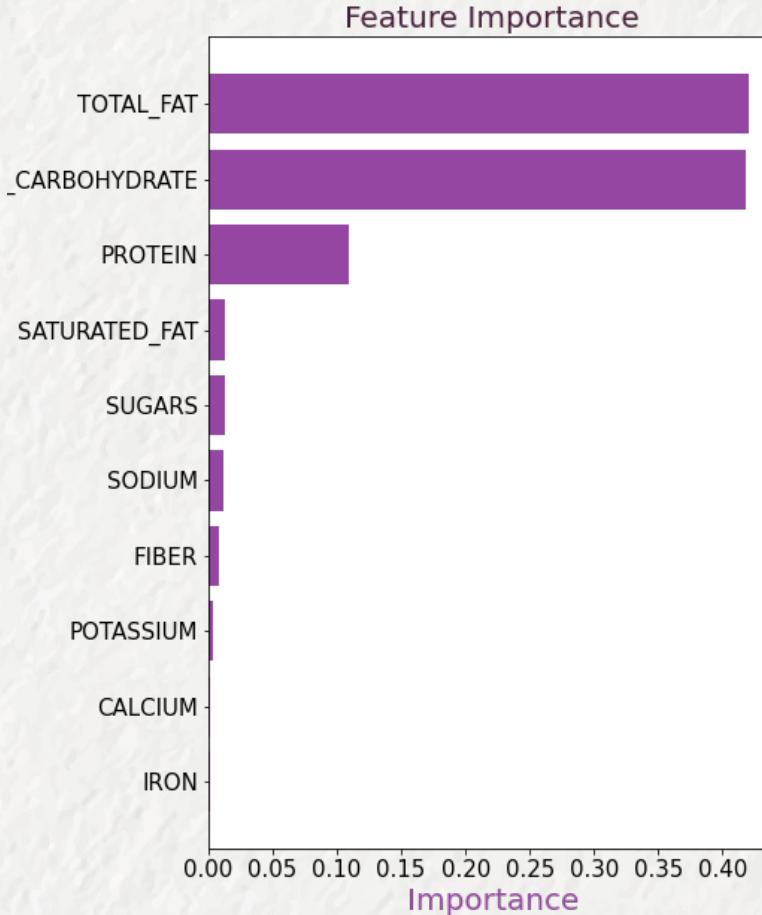


Feature Engineering

04



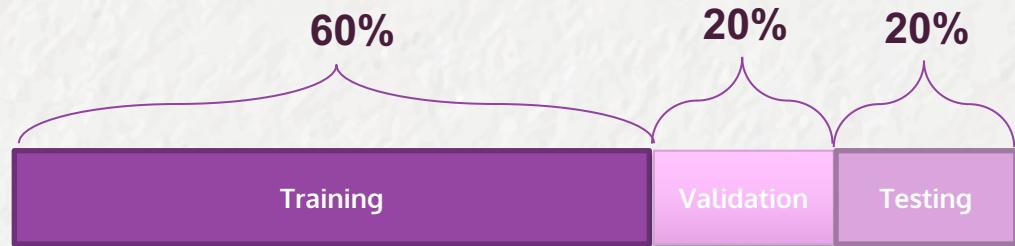
Features Importance



Feature Engineering

04

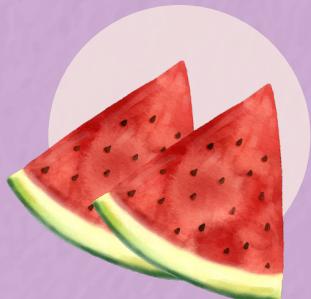
Simple Validation Approach



Scores Before Feature Engineering

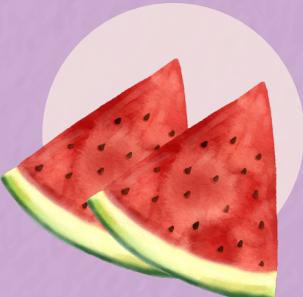
Training Score **94.21%**

Validation Score **89.20%**



Feature Engineering

04



Heatmap

Improvement Attempts



Total Attempts 11

Approved Attempts 3

First Approved Attempt

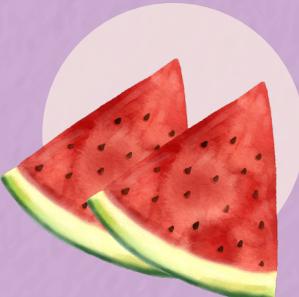


Training Score From 94.58% to 95.04%

Validation Score From 89.20% to 90.09%

Feature Engineering

04



Heatmap

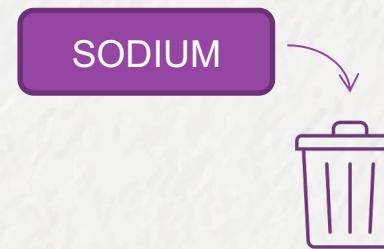
Improvement Attempts



Total Attempts 11

Approved Attempts 3

Second Approved Attempt

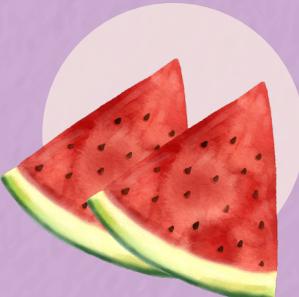


Training Score From 95.04% to 94.63%

Validation Score From 90.09% to 90.95%

Feature Engineering

04



Heatmap

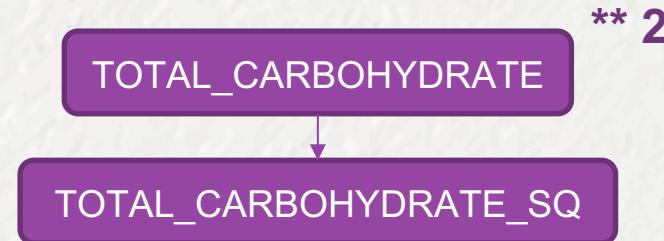
Improvement Attempts



Total Attempts 11

Approved Attempts 3

Third Approved Attempt

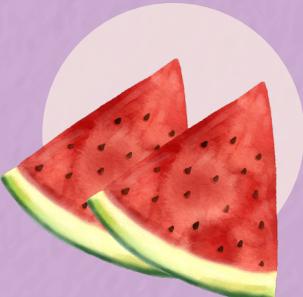


Training Score From 94.63% to 94.97%

Validation Score From 90.95% to 91.38%

Feature Engineering

04



The Resultant Dataset



Columns Names

Basic



Added

SATURATED_FAT+TOTAL_FAT TOTAL_CARBOHYDRATE_SQ

Target

CALORY

Number of rows

1720

Number of columns

12

Reporting The Results

05

Merge Training and Validation Sets



By Simple Validation Approach

Training Score 94.21%

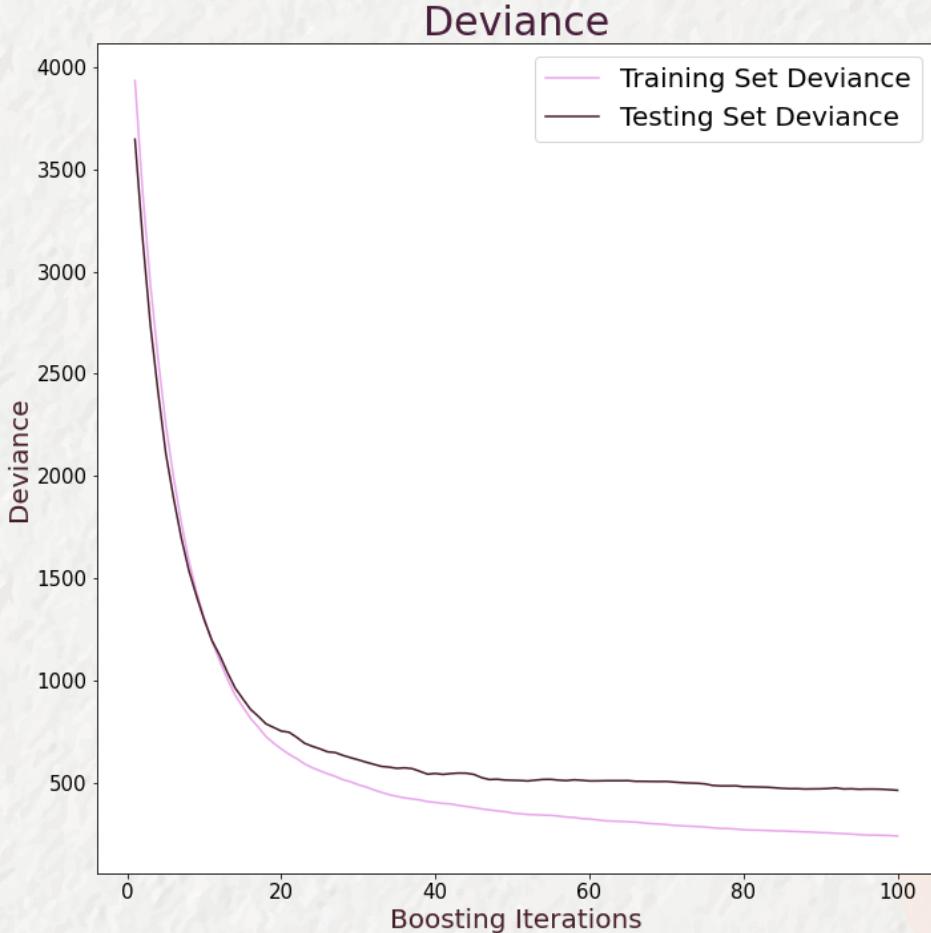
Testing Score 89.16%



05

Reporting The Results

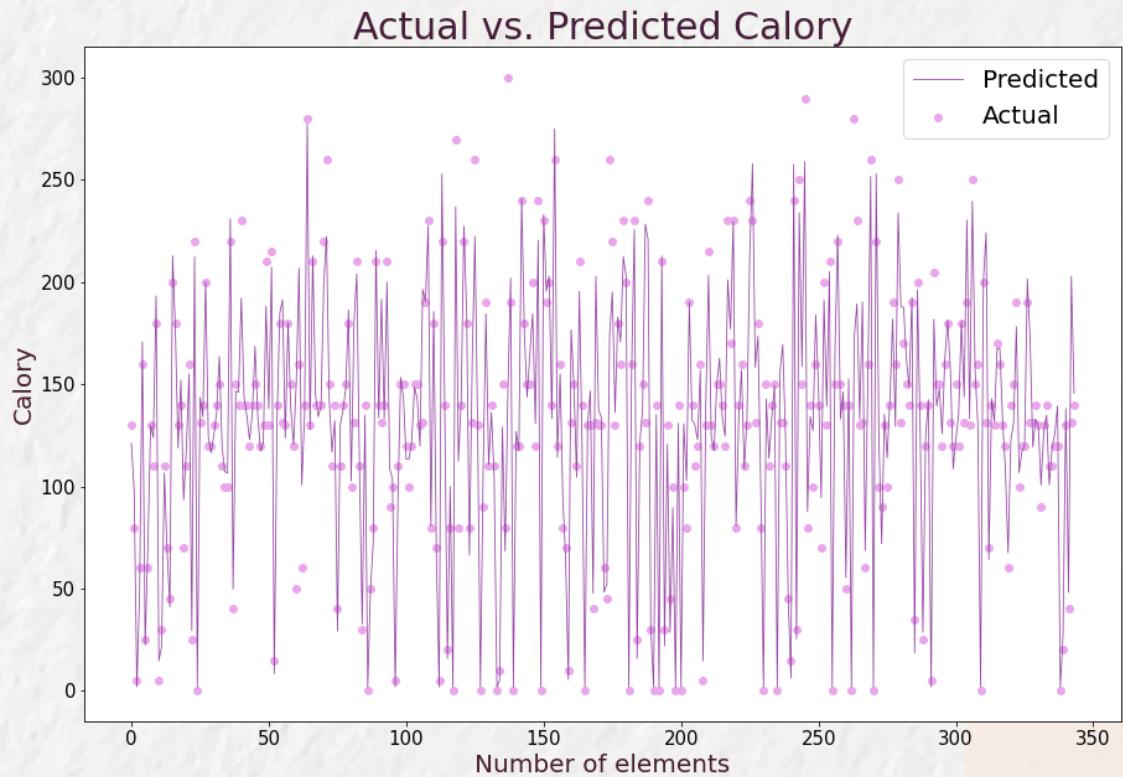
Training vs Validation Sets Deviance



Reporting The Results

05

Actual vs Predicted Calory



Calory Predictor Website

06

Actual vs Predicted Calory

Enter the following nutritions, 0 if not mentioned

Total Fat	Saturated Fat
The amount of total fat	The amount of Saturated Fat
Total Carbohydrate	Protein
The amount of Total Carbohydrate	The amount of Protein
Calcium	Iron
The amount of Calcium	The amount of Iron
Sugars	Fiber
The amount of Sugars	The amount of Fiber

The Results

Detected Calory	True Calory



Thank You

