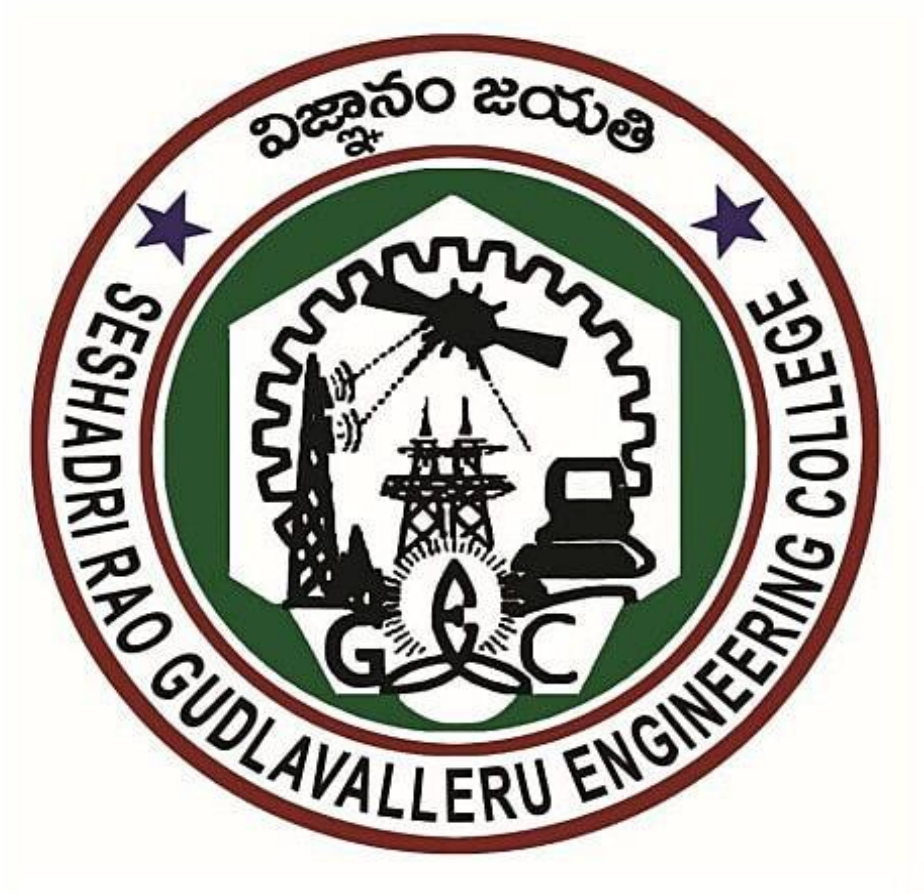
SHESHADRI RAO

GUDLAVALLERU ENGINEERING COLLEGE



Cereal Analysis Based On Ratings By Using Machine Learning Techniques With IBM Watson

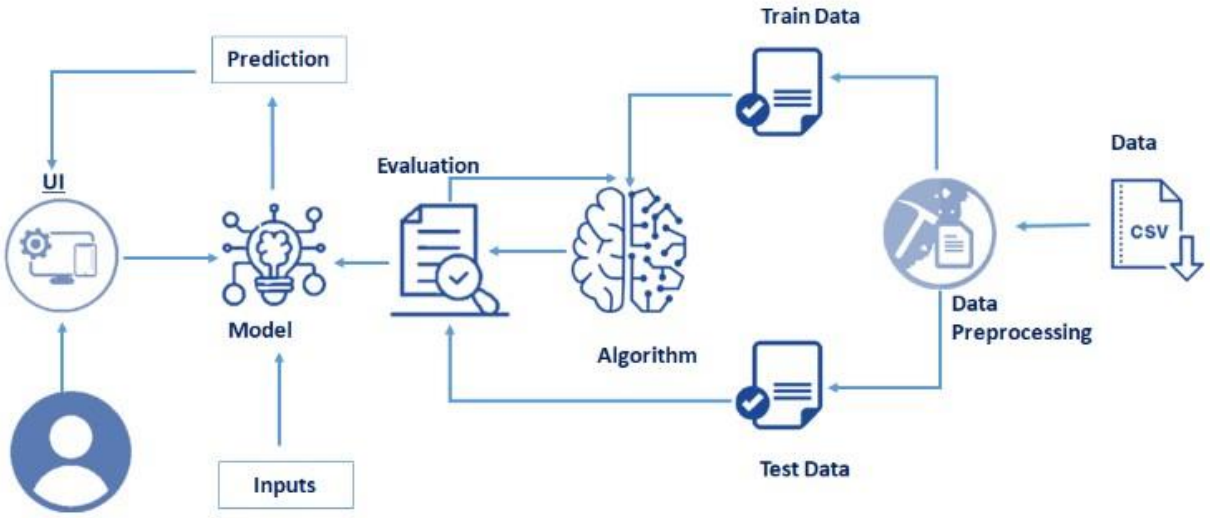
Submitted By :

V.NAGA SIVA GANESH

ROLL NO. - 20485A0334

CONTENT

TECHNICAL ARCHITECHTURE:



1.Introduction:

For kids who grew up in the 1980s and 1990s, sugary cereals flooded the television landscape and many of these commercials shared a common theme of advocating cereal as the "perfect and complete breakfast". Overtime, as more people became health-conscious, fortified cereal turned out to be the new solution to address any health concerns. We knew cereal was the real deal when supermarkets started to dedicate an aisle for it. Walking along these aisles, we observed labels often screaming "25% more iron", "10% more calcium!", "High in Vitamin A, C and E!" and the list goes on.. However, the question to ask is: on what basis does our consumers choose their cereals? Is it based on taste, nutritional values or price?
In this analysis, we will disclose the main attributes that influence cereal ratings. We will also be predicting ratings using a simple linear regression model.

2.Literature survey:

Existing Problem:

A customer wants to buy some food items with high dietary benefits so that he wants to know which food item has high dietary benefits. It is so difficult to choose an item .Usually a customer expects to consume dietary cereals with high proteins, fiber and low sugars, fats. Predicting a brand with high dietary cereals became a big issue. The project objective is to find the high dietary food that is predicted on the basis of rating of the food.

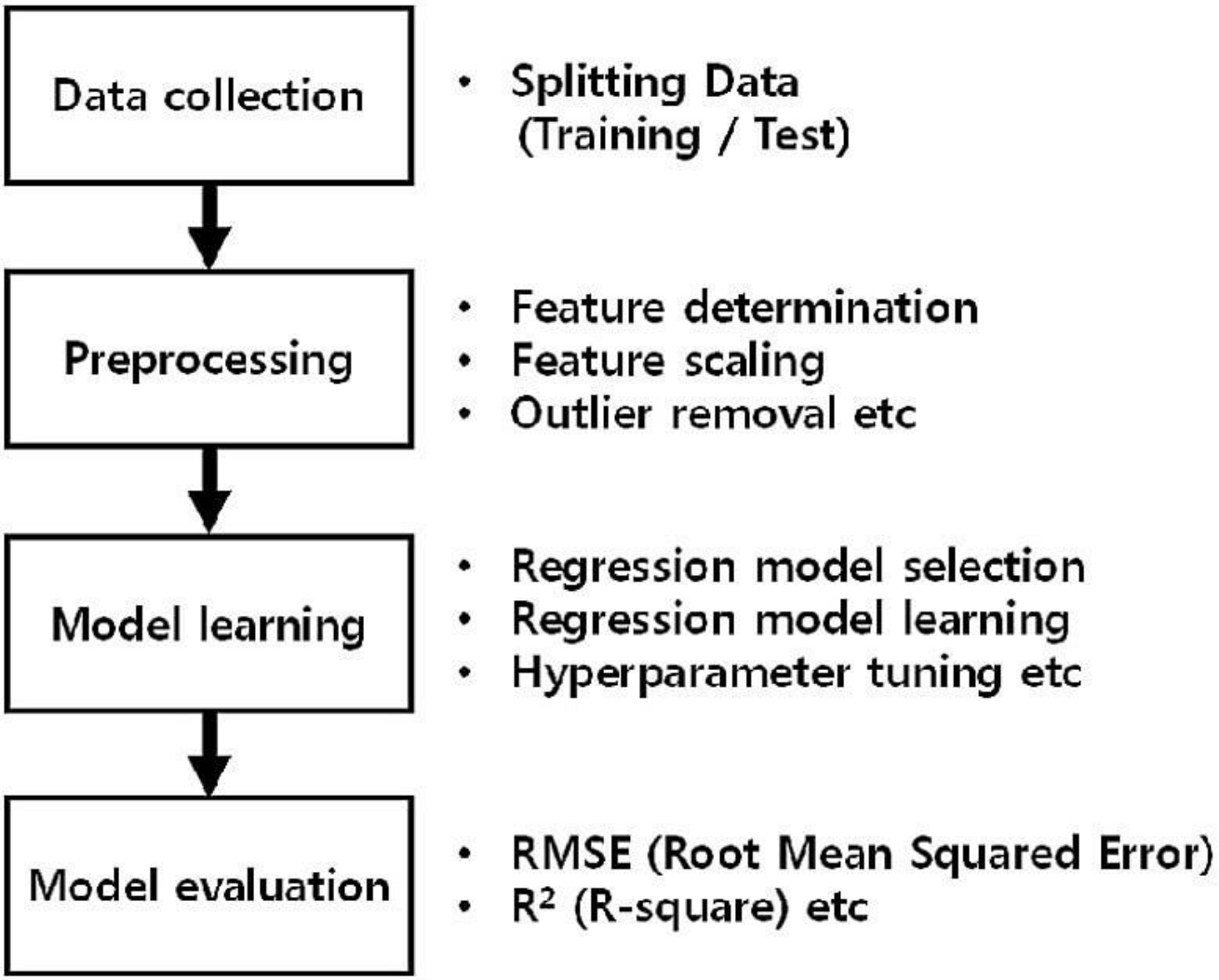
1. To find which quantities are showing more impact on the rating of food.
2. To show the food which is impacting less on the rating of food?

Proposed Solution:

We use machine learning algorithms to predict the food with a high beneficiary diet. The model can predict the rating of the food more accurately by giving the inputs which are the cereals and ingredients present in the food. Thus a customer can get high dietary food by the rating of the food given to it from the cereals and ingredients present. The rating is predicted using the neural networks model. This will helps the customers to predict the best cereal based on ratings. For this we mainly used this algorithms to predict the rating of cereals based on manufacturing and type.Finally we can find out the good or bad based on prediction score.If customers uses more cereals their usage is high as well as prediction is also very high.so, we evaluate the prediction based on ratings given by customer after buying it from the shop.

3.Theoretical Analysis:

Block Diagram:



4.Hardware / Software Requirements:

Recommended System Requirements

● Processors:Intel® Core™ i5 processor 4300M at 2.60 GHz or 2.59 GHz (1 socket, 2 cores, 2 threads per core), 8 GB of DRAMIntel® Xeon® processor E5-2698 v3 at 2.30 GHz (2 sockets, 16 cores each, 1 thread per core), 64 GB of DRAMIntel® Xeon Phi™ processor 7210 at 1.30 GHz (1 socket, 64 cores, 4 threads per core), 32 GB of DRAM, 16 GB of MCDRAM (flat mode enabled)

● Disk space: 2 to 3 GB

● Operating systems: Windows® 10, macOS*, and Linux* Minimum System Requirements

● Processors: Intel Atom® processor or Intel® Core™ i3 processor

● Disk space: 1 GB

● Operating systems: Windows* 7 or later, macOS, and Linux

● Python* versions: 3.9

softwares we used:

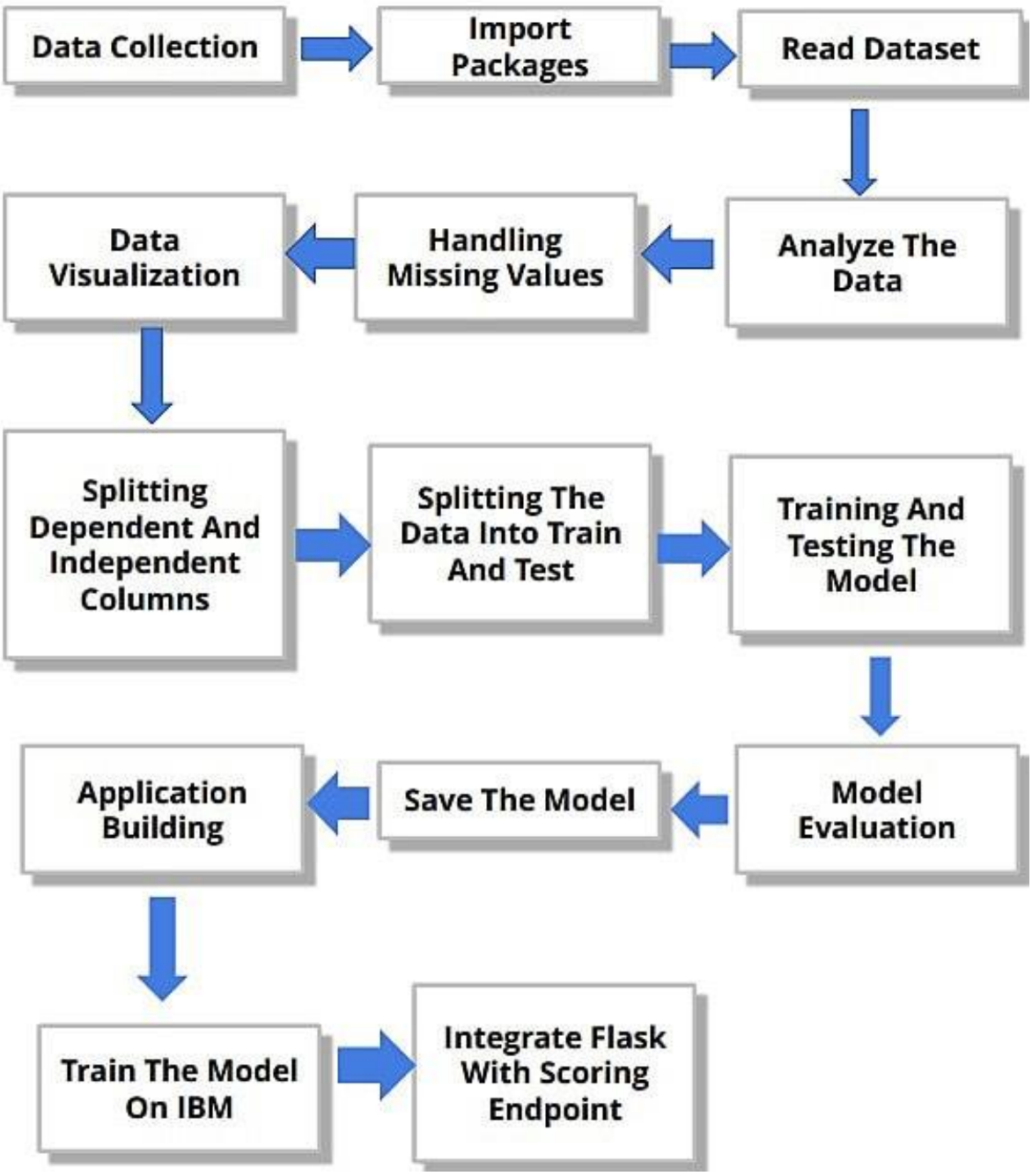
Anaconda navigator:

Anaconda is an open-source distribution for python and R. It is used for data science, machine learning, deep learning, etc. With the availability of more than 300 libraries for data science, it becomes fairly optimal for any programmer to work on anaconda for data science.

Pycharm:

PyCharm is a dedicated Python Integrated Development Environment (IDE) providing a wide range of essential tools for Python developers, tightly integrated to create a convenient environment for productive Python, web, and data science development

5.Flow chart:


Data Collection
Import Packages
Read Dataset
Analyze The Data
Handling Missing Values
Data Visualization
Splitting Dependent And Independent Columns
Splitting The Data Into Train And Test
Training And Testing The Model
Model Evaluation
Save The Model
Application Building
Train The Model On IBM
Integrate Flask With Scoring Endpoint

6.Result:

- For checking the prediction we need to give the input values.
- Then click on predict to show the prediction value

Cereal Prediction Analysis

A machine learning Web App, Built with Flask

Prediction : **80.61603554949039**

7.Advantages and disadvantages of project:

Advantages:

There is an endless number of advantages of ML. We can take a look at the one. which are really helpful. The advantages of Machine Learning tell us how using ML would benefit us. So, let's have a look at the advantages of Machine Learning

- ➤ Automation of Everything
- ➤ Wide Range of Applications
- ➤ Scope ofImprovement
- ➤ Efficient Handling of Data
- ➤ Best for Education

Disadvantages:

Similar to the advantages of Machine Learning, we should also know the disadvantages of Machine Learning. If you don't know the cons, you won't know the risks of ML. So, let's have a look at these disadvantages:

- ➤ Possibility of High Error
- ➤ Algorithm Selection
- ➤ Data Acquisition
- ➤ Time and Space
- ➤ Internet Issues

8.Conclusion:

There is a positive correlation between calories and sugars in cereal. Most cereals do not have relatively high potassium values. Kellogg's offers the most cereals out of any manufacturer that are above the median calorie count (110). The more calories that a cereal has, the less likely it is to receive a high rating. Manufacturers that want to bring in high ratings should create cereals that are high in fiber, protein, and potassium and avoid creating cereals with high calorie counts or lots of sugar or fat. Cereals with high ratings are more likely to be placed on the first or third shelf, because that is generally where the consumers' eyes gravitate. Using a linear regression model can allow for accurate predictions of future cereal with less than ten percent error on average. o For instance, a cereal that has thirteen grams of sugars, one-hundred and ten calories, and two grams of protein is projected to receive a rating of 80.61

9.Biblography:

Installation of Anaconda Navigator:
https://www.youtube.com/embed/5mDYijMfSzs
Installation Of Py charm Professionals:
https://www.youtube.com/embed/z73PyNDgVyQ
Installation Of Python packages:
https://www.youtube.com/embed/akj3_wTploU
Data collection:
https://www.kaggle.com/crawford/80-cereals
Data Preprocessing :
https://thesmartbridge.com/documents/spsaimldocs/Datapreprocessing.pdf
Handling null values:
https://towardsdatascience.com/7-ways-to-handle-missing-values-in-machine-learning-1a6326adf79e
Data Visualization:
https://www.youtube.com/embed/TLdXM0A7SR8

Splitting Dependent And Independent Columns:
https://www.youtube.com/embed/A_V6daPQZIU
Splitting The Data Into Train And Test:
https://www.youtube.com/embed/xgDs0scjuuQ
Training and testing the model:
https://www.youtube.com/embed/yIYKR4sgzI8
Model Evaluation:
https://towardsdatascience.com/the-5-classification-evaluation-metrics?you-must-know-aa97784ff226
Flask Frame Work Reference:
https://www.youtube.com/embed/lj4I_CvBnt0
Flask Refarance To Run:
https://www.youtube.com/embed/UbCWoMf80PY
Train The Model On IBM:
Account Creation:
https://www.youtube.com/embed/4y_zD-0Q3F8
Train Model On IBM Watson:
https://www.youtube.com/embed/TysuP3KgSzc
Integrate Flask With Scoring Endpoint:
https://www.youtube.com/embed/ST1ZYLmYw2U