FOOD PRESERVATION THROUGH DATA SCIENCE

Many industries can manufacture food, But only some can preserve food efficiently and profitably.

FEBRUARY 10, 2023

BATCH 1
Sagi Rama Krishnam Raju Engineering College

Contents

Pain-Points 2
Pain-point sources: 2
Problem statement 3
Identified problem: 3
Feature Engineering 4
Feature creation: 4
Feature selection sources 5
Feature Selection: 6
Feature extraction 7
Feature Transformation: 8
Dataset Preparation 8
Sources: 8
Sample dataset: 8
Creation of one million data set 8
First trail: 9
Second trail: 9
Third trail: 10
Calculation of average weights from three trails: 10
Program to create one million data set: 10
One million Data Set output: 11
Regression Model 11
Importing the modules 11
Data Handling 11
Data Analysis 12
Model Selection 12
Observations 12
Outputs: 12
APPLICATION AND FEATURE THOUGHTS 13

Industry: Food

Department: Manufacturing pain-points

Pain-Points

1. Supply chain crisis
2. Increased costs of manufacturing
3. Material shortage
4. Backlog of orders
5. Regulations
6. Ineffective workflow system
7. Environmental conditions
8. Microbial production
9. Equipment maintenance and upkeep
10. Work force shortage challenges
11. Food safety
12. Minimising machine down time
13. Technology challenges
14. Material Preservation

Pain-point sources:

1,2

https://taia.io/blog/food-beverage-industry-translation/#:~:text=Solving%20Top%206%20Food%20Manufacturing%20Pain%20Points%20Through,increased%20pressure%20on%20the%20food%20%26%20beverage%20industry

3,4,5

https://industryselect.com/blog/top-manufacturing-pain-points-and-how-you-can-help

6,12

https://eemits.co.uk/articles/food-manufacturing-pain-points-for-productivity-safety-efficiency

7

https://academic.oup.com/nutritionreviews/article/59/3/S21/1866047

8

https://books.google.co.in/books?hl=en&lr=&id=Po1wAgAAQBAJ&oi=fnd&pg=PP1&dq=pain+points+in+manufacturing+of+food+indu&ots=ob6yqsEmMk&sig=x9piYYIQC-Va98mEMjTBU8_OQbM&redir_esc=y#v=onepage&q&f=false

9,10

https://www.aiscorp.com/blog/5-challenges-to-overcome-to-increase-manufacturing-productivity/

11

13

https://www.sciencedirect.com/science/article/abs/pii/S0958694601000991

14

https://www.britannica.com/topic/food-preservation

Problem statement

Many industries can manufacture food , But only some can preserve food efficiently and profitably.

Outcome: Preservability of Material

Identified problem:

Many companies and industries are struggling with the best way to store its raw material and all the processed materials this pain point is the reason for many drawbacks in the industry such as food material wastage, cost manufacturing industries.

Globally, around 14 percent of food produced is lost between harvest and retail, while an estimated 17 percent of total global food production is wasted (11 percent in households, 5 percent in the food service and 2 percent in retail).

(source: https://www.un.org/en/observances/end-food-waste-day#:~:text=Did%20you%20know%3F-,Globally%2C%20around%2014%20percent%20of%20food%20produced,lost%20between%20harvest%20and%20retail)

(Source: https://th.bing.com/th/id/OIP.ea5juyEir7aomOOeDypIRQHaHt?pid=ImgDet&rs=1)

Feature Engineering

Feature creation:

1. Temperature
2. Moisture
3. Oxygen
4. Light
5. Container Material
6. Infestation (several common insects infest into home stored dry food)
7. Shelf date
8. Microbial growth
9. Dryness
10. Airflow
11. Preservative chemicals
12. Weather conditions
13. Quality of air
14. Type of raw material
15. Pest control
16. Chemical exposure
17. Transportation time
18. Transportation distance
19. Over crowding of materials
20. Stacking of materials
21. Mould prevention
22. Flood history
23. Earthquake history
24. Location history
25. Storage technology
26. Rodent control
27. Managing
28. Monitoring
29. Packaging
30. Political factors
31. Policies
32. Government regulations
33. Biological factors
34. Cultural factors
35. Social factors
36. Crop choices
37. Low-capacity soil
38. Taxes
39. Investments
40. Harvest time
41. Distribution
42. Economic factors
43. Staff
44. Control
45. Safety of employees
46. Experience of labour
47. No strategic planning
48. Inefficient teamwork
49. Technical automation
50. Demand

Feature selection sources

1-7

https://extension.usu.edu/news_sections/home_family_and_food/foodstorage

8

http://ecoursesonline.iasri.res.in/mod/page/view.php?id=111412

9,10

https://www.terrafoodtech.com/en/food-preservation-methods-comparison/

11

https://www.canr.msu.edu/news/preservatives-keeping-our-foods-safe-fresh

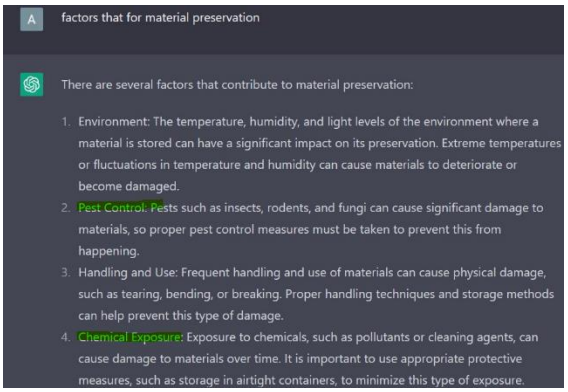
12,13

https://basicmedicalkey.com/preservation-of-industrial-products-and-processes/#:~:text=A%20successful%20industrial%20manufacturing%20hygiene%20program%20consists%20of,5%20Microbiological%20testing%20of%20finished%20products%20More%20items

14

https://www.sciencedirect.com/topics/food-science/food-preservation

15,16



17,18

https://www.iso9001help.co.uk/755.html

19,20

https://sciencing.com/types-used-storing-acids-bases-8456845.html

21

https://www.army.mil/article/219276/ventilation_eliminating_moisture_keys_to_stopping_mold_in_its_tracks#:~:text=The%20following%20tips%20will%20assist%20with%20proper%20ventilation,not%20be%20placed%20against%20walls.%20...%20More%20items

22,23,24

https://www.msha.gov/news-media/announcements/2016/04/26/materials-storage-and-warehouse-safety-best-practices

25

https://www.bbc.co.uk/bitesize/guides/zsdhbk7/revision/3

26

https://timesofindia.indiatimes.com/city/chennai/chennai-godowns-overrun-with-rats-pests-cag/articleshow/53999108.cms

27,28

https://www.deskera.com/blog/raw-materials-inventory-management-guide/#:~:text=Tips%20to%20Manage%20Raw%20Materials%20Inventory%201%20Utilize,Special%20Tips%20for%20Small%20and%20Medium%20Businesses%20

29

https://www.paramount.ie/2021/10/how-does-packaging-protect-and-preserve-food/#:~:text=Packaging%20is%20essential%20in%20maintaining,%2C%20taste%2C%20smell%20and%20look

30-41

https://www.customwritingservice.org/factors-affecting-food-production/

42-44

https://www.hotelmanagementtips.com/environmental-factors-affecting-food-industry/

45

https://rb.gy/sxs32r

46-48

https://www.tayanasolutions.com/factors-that-affect-the-quality-in-process-manufacturing/#:~:text=Manufacturing%20i

49,50

https://pestleanalysis.com/pestle-analysis-of-the-food-industry/

Feature Selection:

1. Temperature
2. Moisture
3. Humidity
4. Light
5. Container materials
6. Shelf date
7. Microbial growth
8. Dryness
9. Pest control
10. Preservative chemicals
11. Air flow
12. Chemical exposure
13. Transportation time
14. Transportation distance
15. Overcrowding of materials
16. Stacking of materials
17. Flood history
18. Earthquake history
19. Rodent control
20. Managing
21. Monitoring
22. Packaging

23. Government regulations
24. Staff
25. Experience of labour
26. Inefficient teamwork
27. Technical automation
28. Storage technology
29. Economic factors
30. Investment
31. Location history

Feature extraction

1. Natural factors
    a. Temperature
    b. Moisture
    c. Humidity
    d. Light
    e. Air flow
    f. dryness
2. Infestation control
    a. Rodent control
    b. Pest control
    c. Microbial growth
3. Shipping
    a. Transportation distance
    b. Transportation time
    c. Packaging
4. Location history
    a. Flood history
    b. Earthquake history
    c. Location history
5. Workforce
    a. Experience of labour
    b. Inefficient teamwork
    c. Managing
    d. Monitoring
    e. Staff
6. Storage
    a. Stacking of materials
    b. Overcrowding of materials
    c. Container materials
    d. Shelf date
    e. Chemical exposure
    f. Preservative chemicals
    g. Storage technology
    h. Technical automation
7. Finance
    a. Government regulations
    b. Economic factors
    c. Investments

Feature Transformation:

1. Natural_Factors (1to10)
2. Infestation_control (0to80)
3. Shipping (1to5)
4. Location_history (1,2) (2 for better location)
5. Work_Force (1to10)
6. Storage (1to10)
7. Finance (1,2) (2 for good finance)

Dataset Preparation

Sources:

1. Dharmastala (https://www.youtube.com/watch?v=iqeTx7700k8)
2. Akshaya patra (https://www.youtube.com/watch?v=drGeDmBXNuU)
3. Kit Kat factory(https://youtu.be/2Yd2HhKIsiE)
4. Sri racha hot sauce (https://dplus.app.link/pTdxT2CVixb)
5. Gokul chat (https://www.youtube.com/watch?v=aprPDS7yb1g)
6. Rajahmundry Rose Milk(https://www.youtube.com/watch?v=grXg0MreB6s)
7. Baskin Robbins main kitchen(https://en.wikipedia.org/wiki/Baskin-Robbins)
8. Call recording Naresh Hostel (https://drive.google.com/drive/folders/1SQT6sOQ5uCBQ9Q1b8UtsUesO3JMBuqrw?usp=share_link)
9. Kerala potato chips, Nashik (https://youtu.be/YKGHhxg-whk)
10. Sikinder Kitchen,Karachi(https://youtu.be/j5T4R3yO960)
11. Sri sai pootarekulu shop (https://drive.google.com/file/d/1taD5ZC_epVYBroGGlnS93MXl4-LUcvKW/view?usp=share_link)

Sample dataset:

	A	B	C	D	E	F	G	H	I	J
1	NATURAL_FACTORS	INFESTATION_CONTROL	SHIPPING	LOCATION_HISTORY	WORK_FORCE	STORAGE	FINANCE	PRESERVABILITY		
2	1 to 10	0 to 80	1 to 5	1/2	1 to 10	1 to 10	1/2	119		
3	4	53	4	2	7	8	2	67.22689076		
4	5	63	4	1	8	7	2	75.6302521		
5	7	60	5	2	4	5	1	70.58823529		
6	4	45	3	1	5	4	1	52.94117647		
7	9	76	3	2	8	6	2	89.07563025		
8	8	44	3	2	4	4	1	55.46218487		
9	6	67	3	2	3	4	1	72.26890756		
10	3	42	4	1	4	4	1	49.57983193		
11	5	34	4	1	4	3	1	43.69747899		
12	4	66	3	2	5	9	2	76.47058824		
13	6	65	4	2	8	8	2	79.83193277	sr1	
14	8	75	4	2	9	10	2	92.43697479	sr2	
15	7	70	4	2	9	9	2	86.55462185	sr3	
16	6	72	3	1	7	8	2	83.19327731	sr4	
17	8	55	3	1	8	6	2	69.74789916	sr5	
18	7	55	3	2	5	4	1	64.70588235	sr6	
19	6	65	4	2	9	8	2	80.67226891	sr7	
20	8	70	4	2	9	9	2	87.39495798	sr8	
21	6	45	2	1	4	5	1	53.78151261	sr9	
22	7	40	3	1	5	6	1	52.94117647	sr10	
23	8	60	4	2	7	9	2	77.31092437	sr11	
24										

**sr1 indicates source 1 from the above sources list*

Creation of one million data set

We will consider 3 sets of 7 tupels of data from the sample data set we created from different sources,we solve those three sets of equations respectively and find the weights of each feature ,we select these 7 tuples randomly.

First trail:

System of equations:

$$\begin{cases} 4x_1 + 53x_2 + 4x_3 + 2x_4 + 7 + 8 + 2 = 67.22689076 \\ 7x_1 + 60x_2 + 5x_3 + 2x_4 + 4x_5 + 5x_6 + 1x_7 = 70.58823529 \\ 4x_1 + 45x_2 + 3x_3 + 1x_4 + 5x_5 + 4x_6 + 1x_7 = 52.94117647 \\ 5x_1 + 34x_2 + 4x_3 + 1x_4 + 4x_5 + 3x_6 + 1x_7 = 43.69747899 \\ 7x_1 + 55x_2 + 3x_3 + 2x_4 + 5x_5 + 4x_6 + 1x_7 = 64.70588235 \\ 7x_1 + 40x_2 + 3x_3 + 1x_4 + 5x_5 + 6x_6 + 1x_7 = 52.94117647 \\ 8x_1 + 60x_2 + 4x_3 + 2x_4 + 7x_5 + 9x_6 + 2x_7 = 77.31092437 \end{cases}$$

Solutions of the above equations are:

Answer:

$$x_1 = 0.8403361337$$
$$x_2 = 0.8403361343$$
$$x_3 = 0.8403361345$$
$$x_4 = 0.8403361360$$
$$x_5 = 0.8403361358$$
$$x_6 = 0.8403361352$$
$$x_7 = 0.8403361327$$

Second trail:

System of equations:

$$\begin{cases} 5x_1 + 63x_2 + 4x_3 + 1x_4 + 8 + 7 + 2 = 75.6302521 \\ 7x_1 + 60x_2 + 5x_3 + 2x_4 + 4x_5 + 5x_6 + 1x_7 = 70.58823529 \\ 5x_1 + 34x_2 + 4x_3 + 1x_4 + 4x_5 + 6x_6 + 1x_7 = 46.21848739 \\ 8x_1 + 44x_2 + 3x_3 + 2x_4 + 4x_5 + 4x_6 + 1x_7 = 55.46218487 \\ 6x_1 + 65x_2 + 4x_3 + 2x_4 + 9x_5 + 8x_6 + 2x_7 = 80.67226891 \\ 6x_1 + 45x_2 + 2x_3 + 1x_4 + 4x_5 + 5x_6 + 1x_7 = 53.78151261 \\ 6x_1 + 72x_2 + 3x_3 + 1x_4 + 7x_5 + 8x_6 + 2x_7 = 83.19327731 \end{cases}$$

The solution of the above equations are:

Answer:

$$x_1 = 0.8403361353$$
$$x_2 = 0.8403361348$$
$$x_3 = 0.8403361316$$
$$x_4 = 0.8403361279$$
$$x_5 = 0.8403361421$$
$$x_6 = 0.840336135\overline{1}$$
$$x_7 = 0.840336096\overline{7}$$

Third trail:

System of equations:

$$\begin{cases} 5x_1 + 63x_2 + 4x_3 + 1x_4 + 8 + 7 + 2 = 75.6302521 \\ 6x_1 + 67x_2 + 3x_3 + 2x_4 + 3x_5 + 4x_6 + 1x_7 = 72.26890756 \\ 6x_1 + 65x_2 + 4x_3 + 2x_4 + 9x_5 + 8x_6 + 2x_7 = 80.67226891 \\ 8x_1 + 70x_2 + 4x_3 + 2x_4 + 9x_5 + 9x_6 + 2x_7 = 87.39495798 \\ 8x_1 + 75x_2 + 4x_3 + 2x_4 + 9x_5 + 10x_6 + 2x_7 = 92.43697479 \\ 3x_1 + 42x_2 + 4x_3 + 1x_4 + 4x_5 + 4x_6 + 1x_7 = 49.57983193 \\ 4x_1 + 66x_2 + 3x_3 + 2x_4 + 5x_5 + 9x_6 + 2x_7 = 76.47058824 \end{cases}$$

The solutions of the above equations are:

Answer:

$$x_1 = 0.84033613$$
$$x_2 = 0.8403361349$$
$$x_3 = 0.8403361313$$
$$x_4 = 0.8403361372$$
$$x_5 = 0.8403361373$$
$$x_6 = 0.8403361357$$
$$x_7 = 0.8403361216$$

Calculation of average weights from three trails:

Trail -1 solutions	0.8403361337	0.8403361343	0.8403361345	0.840336136	0.8403361358	0.8403361352	0.8403361327
Trail -2 solutions	0.84033613	0.8403361349	0.8403361313	0.8403361372	0.8403361373	0.8403361357	0.8403361216
Trail -3 solutions	0.8403361353	0.8403361348	0.8403361316	0.8403361279	0.8403361421	0.8403361351	0.8403360967
AVERAGE OF 3 SOLUTIONS	0.840336133	0.8403361347	0.8403361325	0.8403361337	0.8403361384	0.8403361353	0.840336117

Program to create one million data set:

```
import random
import pandas as pd
w1,w2,w3,w4,w5,w6,w7=[0.840336133,0.8403361347,0.8403361325,0.8403361337,0.840
3361384,0.8403361353,0.840336117]
vals=[]
for i in range(1000000):
    x1=random.randint(1,10)
    x2=random.randint(0,80)
    x3=random.randint(1,5)
    x4=random.randint(1,2)
    x5=random.randint(1,10)
    x6=random.randint(1,10)
    x7=random.randint(1,2)
    eq=x1*w1+x2*w2+x3*w3+x4*w4+x5*w5+x6*w6+x7*w7
    if eq>65:
        cls=1
    else:
        cls=0
    vals.append([x1,x2,x3,x4,x5,x6,x7,eq,cls])
df=pd.DataFrame(vals,columns=['Natural_Factors','Infestation_Control','Shippin
g','Location_history','Work_Force','Storage','Finance','Preservability_Factor'
,'Efficiency'])
df.to_csv('preserve.csv',index=False)
```

Through this program we created a random data set of 1,000,000 tuples by using the weights we calculated through the data we collected from different sources. The data we created is now written into a ".csv file"(here preserve.csv) which we later use to train our machine learning algorithm to get the accurate regression outcome. As our dataset is randomly generated from the accurate data, we can expect that the regression outcome to be nearly accurate.

One million Data Set output:

	A	B	C	D	E	F	G	H	I
1	Natural_Factors	Infestation_Control	Shipping	Location_history	Work_Force	Storage	Finance	Preservability_Factor	Efficiency
2	3	29	3	1	8	1	1	38.6554622	0
3	9	75	2	2	1	8	2	83.19327729	1
4	1	9	4	1	1	9	2	22.6890756	0
5	3	44	4	1	10	3	1	55.4621849	0
6	7	51	1	2	3	3	2	57.98319326	0
7	9	19	4	2	5	9	1	41.17647058	0
8	6	23	3	1	1	3	1	31.93277309	0
9	1	19	4	2	8	5	2	34.45378151	0
10	3	74	4	1	8	7	2	83.19327732	1
11	7	58	3	1	9	7	2	73.1092437	1
12	7	17	5	1	4	8	1	36.13445377	0
13	1	74	4	1	7	3	2	77.31092437	1
14	1	59	2	1	6	8	2	66.38655463	1
15	1	6	2	2	6	5	2	20.16806721	0
16	3	71	1	2	9	5	2	78.15126052	1
17	7	24	2	1	1	2	1	31.93277309	0
18	8	47	2	1	4	9	2	61.3445378	0
19	4	43	2	2	3	1	2	47.89915964	0
20	8	36	3	1	5	9	1	52.94117647	0
21	4	18	2	2	9	10	2	39.49579832	0
22	5	59	4	2	4	8	1	69.74789916	1
23	4	79	5	2	5	6	1	85.71428572	1
24	2	58	4	2	5	6	2	66.38655461	1
25	6	75	4	1	8	1	1	80.67226892	1
26	4	17	4	2	4	6	2	32.77310922	0
27	6	65	5	2	7	3	1	74.78991598	1
28	7	58	3	2	7	5	2	70.58823529	1
29	9	22	3	1	5	10	1	42.85714285	0

The above attached screenshot is some of the random data that is generated through the program we have designed using the weights that we calculated. There are one million of the data similar to above in the preserve.csv file. We use this csv file as out preliminary prerequisite to train the model for the regression outcome.

Regression Model

Importing the modules

```
# IMPORTING MODULES
import pandas as pd
from sklearn.linear_model import LogisticRegression
from matplotlib import pyplot as plt
```

Data Handling

```
# Data Handling
data = pd.read_csv('preserve.csv')
print(data.head())
print(data.describe())
X = data[['Natural_Factors', 'Infestation_Control', 'Shipping', 'Location_his-
tory', 'Work_Force', 'Storage', 'Finance']]
Y = data["Preservability_Factor"]
Y_ = data["Efficiency"]
```

While working on the regression we first start by handling the data

In the list *X* we consider all our features

In the *Y* set we consider the regression outcome

In the *Y_* set we consider the classification of the data.

Data Analysis

```
# DATA ANALYSIS
plt.scatter(X['NATURAL_FACTORS'], Y, color='b')
plt.xlabel('NATURAL_FACTORS')
plt.ylabel('PRESERVE_FACTOR')
plt.show()
```

Model Selection

Based on the analysis of data we did in the above step we have selected the logistic regression model.

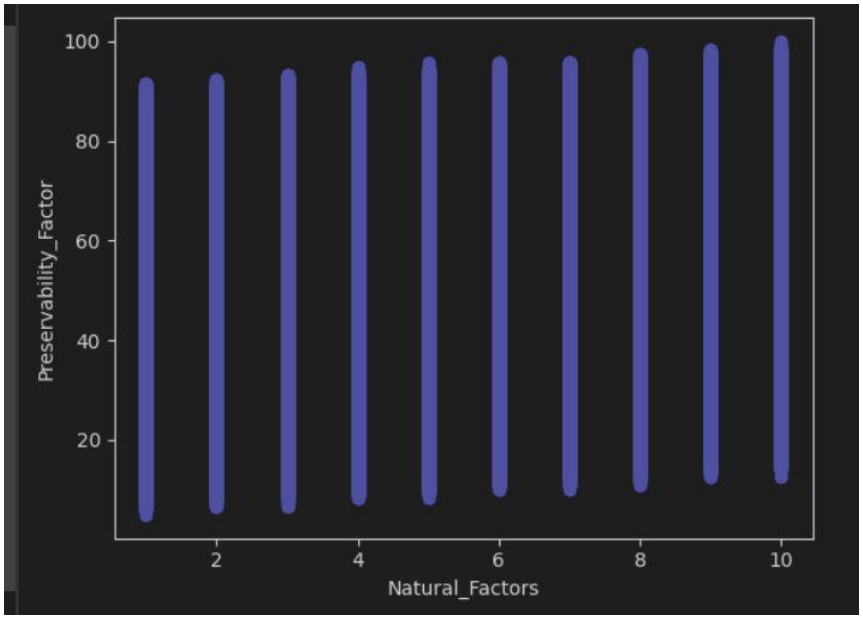
The points we have take in consideration while selecting the logistic regression model are:

- There is no necessity of dimensional reduction to the data in our use case.
- Our program requires giving a responses.
- We require speed over accuracy in the food industry as some foods may get damaged easily in just a small time span.
- The outcome we are going to get is explainable that if we can use that method to preserve food or not.

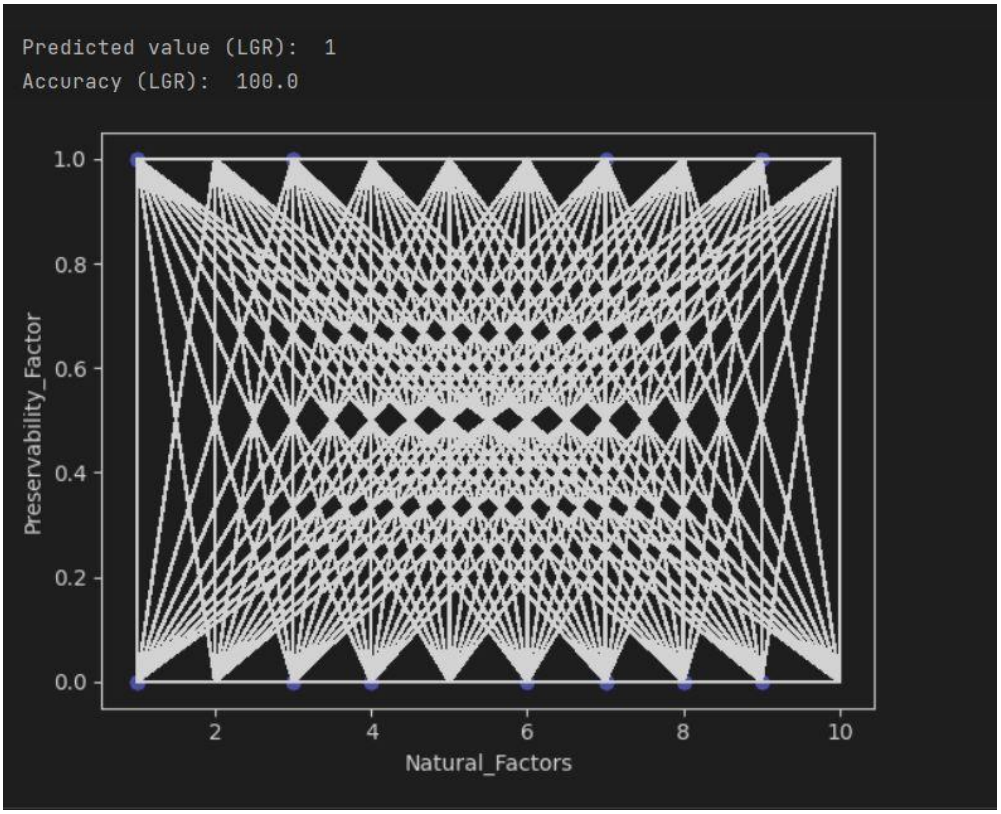
```
#implementation and plotting
mdl = LogisticRegression()
mdl.fit(X, Y_)
predi = mdl.predict(inputs)
print("Predicted value (LGR): ", predi[0])
print("Accuracy (LGR): ", mdl.score(X[:100], Y_[:100]) * 100)

plt.scatter(X['Natural_Factors'][:20], Y_[:20], color='b')
plt.rcParams['agg.path.chunksize'] = 10000
plt.plot(X['Natural_Factors'], mdl.predict(X), color='black', linewidth=3)
plt.xlabel('Natural_Factors')
plt.ylabel('Preservability_Factor')
plt.show()
```

Observations

Outputs:



Here in this graph we are showing that as the preservability_factor is increasing when we are increasing Natural_factors



Here we have used matploit lib to plot the logistic regression between the preservability_factor and Natural_factors.

According to our input the predicted outcome is 1 and the value is completely accurate ie,100%

APPLICATION AND FEATURE THOUGHTS

- Preservability can be found easily.
- We can classify the factors into preservable and non preservable conditions
- We are looking forward to make this dataset much more relevant to out real world and train the algorithm much more effectively
- It might be a sensation If any industry can implement this data science in their business field