***Project report on***

**DemandEst - AI Powered Food Demand Forecaster**

|  |  |
| --- | --- |
| **Team Id**  **Industry Mentor**  **Faculty Mentor** | **PNT2022TMID41126**  **Prof Swetha**  **Dhanasekar.p** |
| **Nandhini.S** | **612719106016** |
| **Priya.M** | **612719106019** |
| **Mahalakshmi.M** | **612719106011** |
| **Anitha.P** | **612719106002** |

**CONTENTS**

# 1. INTRODUCTION

Project Overview

Purpose

# 2. LITERATURE SURVEY

Existing problem

References

Problem Statement Definition

# 3. IDEATION & PROPOSED SOLUTION

Empathy Map Canvas

Ideation & Brainstorming

Proposed Solution

Problem Solution fit

# 4. REQUIREMENT ANALYSIS

Functional requirement

Non-Functional requirements

# 5. PROJECT DESIGN

Data Flow Diagrams

Solution & Technical Architecture

User Stories

# 6. PROJECT PLANNING & SCHEDULING

Sprint Planning & Estimation

Sprint Delivery Schedule

Reports from JIRA

# 7. CODING & SOLUTIONING

Data Dictionary

Libraries Used

Data Pre-Processing

Feature Engineering

Data Transformation

Evaluation Metric

Initial Approach

Advanced Models

# 8. TESTING

Test Cases

User Acceptance Testing

**9. RESULTS**

Performance Metrics

# 10. ADVANTAGES & DISADVANTAGES 11. APPLICATIONS 12. CONCLUSION 13. FUTURE SCOPE 14. APPENDIX Source Code Output

Screenshots

GitHub & Project Demo Link

**1. INTRODUCTION**

# OVERVIEW

A food delivery service has to dealwith a lot of perishable raw materials which makes it all, the most important factor for such a company is to accurately forecast daily and weekly demand. Too much inventory in the warehouse means more risk of wastage, and not enough could lead to out-of-stocks - and push customers to seek solutions from your competitors. The replenishment of the majority of raw materials isdone on weekly basis and since the raw material is perishable, the procurement planning is of utmost importance, the task is to predict the demand for the next 10 weeks.

# PURPOSE

The main aim of this project is to create an appropriate machine learning model to forecast then number of orders to gather raw materials for next ten weeks. To achieve this, we should know the information about of fulfillment center like area, city etc., and meal information like category of food, sub category of food, price of the foodor discount in particular week. By using this data, we can use any classification algorithm to forecast the quantity for 10 weeks. For this a web application is built whichis integrated with the model.

**2. LITERATURE SURVEY**

# EXISTING PROBLEM

The replenishment of the majority of raw materials is done on weekly basis and since the raw material is perishable, the procurement planning is of utmost importance. Also the recruiting of staff members at the fulfillment center is an prospect wherein the prediction of orders would be beneficial. Although this is a process that can be done manually.

# REFERENCE

* Adi, G. N. (2018, March 9). Thousands of GO-CAR Drivers on Strike in Surakarta.The

Jakarta Post. https:/[/www.thejakartapost.com/news/2018/03/08/thousands-of-go-car-](http://www.thejakartapost.com/news/2018/03/08/thousands-of-go-car-) drivers-on-strike-in-surak arta.html

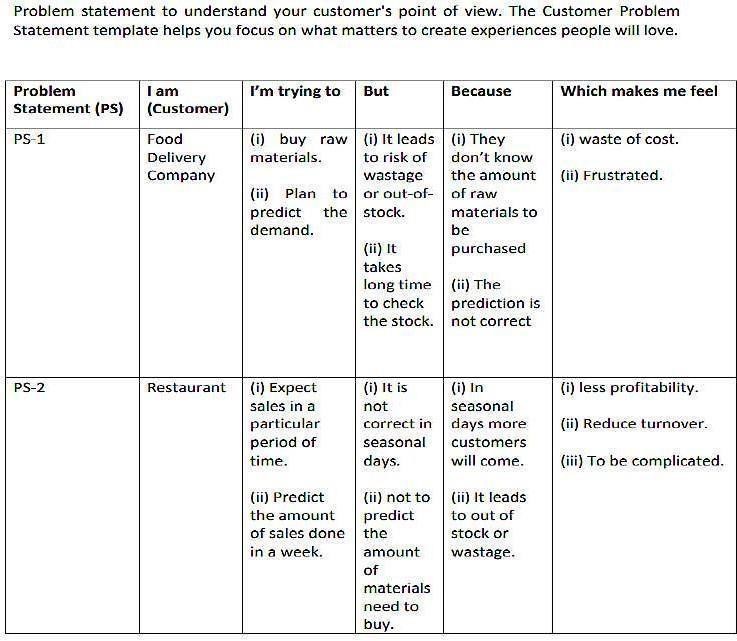
* Alkhatib, A., & Bernstein, M. (2019, May). Street-level algorithms: A theory at thegaps between policy and decisions. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1-13).

* Brown, Tim. (2008). Design Thinking. Harvard Business Review. 86. 84-92, 141.

* Colley, A., & Häkkilä, J. (2018, November). Service Design Methods for Human Computer Interaction. In Proceedings of the 17th International Conference onMobile and Ubiquitous Multimedia (pp. 563-566).

* Clarke, S. (2006). Transformation Lessons from Coca-Cola Enterprises Inc.: Managing the Introduction of a Structured Forecast Process. Foresight: The International Journal of Applied Forecasting, (4), 21-25.

# PROBLEM STATEMENT DEFINITION



## 3. IDEATION & PROPOSED SOLUTION

### Empathy Map Canvas

An empathy map is a collaborative visualization used to articulate what we know about a particular type of user. It externalizes knowledge about users in order to 1) create a shared understanding of user needs, and 2) aid in decision making.

Traditional empathy maps are split into 4 quadrants (Says, Thinks, Does, and Feels), with the user or persona in the middle. Empathy maps provide a glance into whoa user is as a whole and are not chronological or sequential.

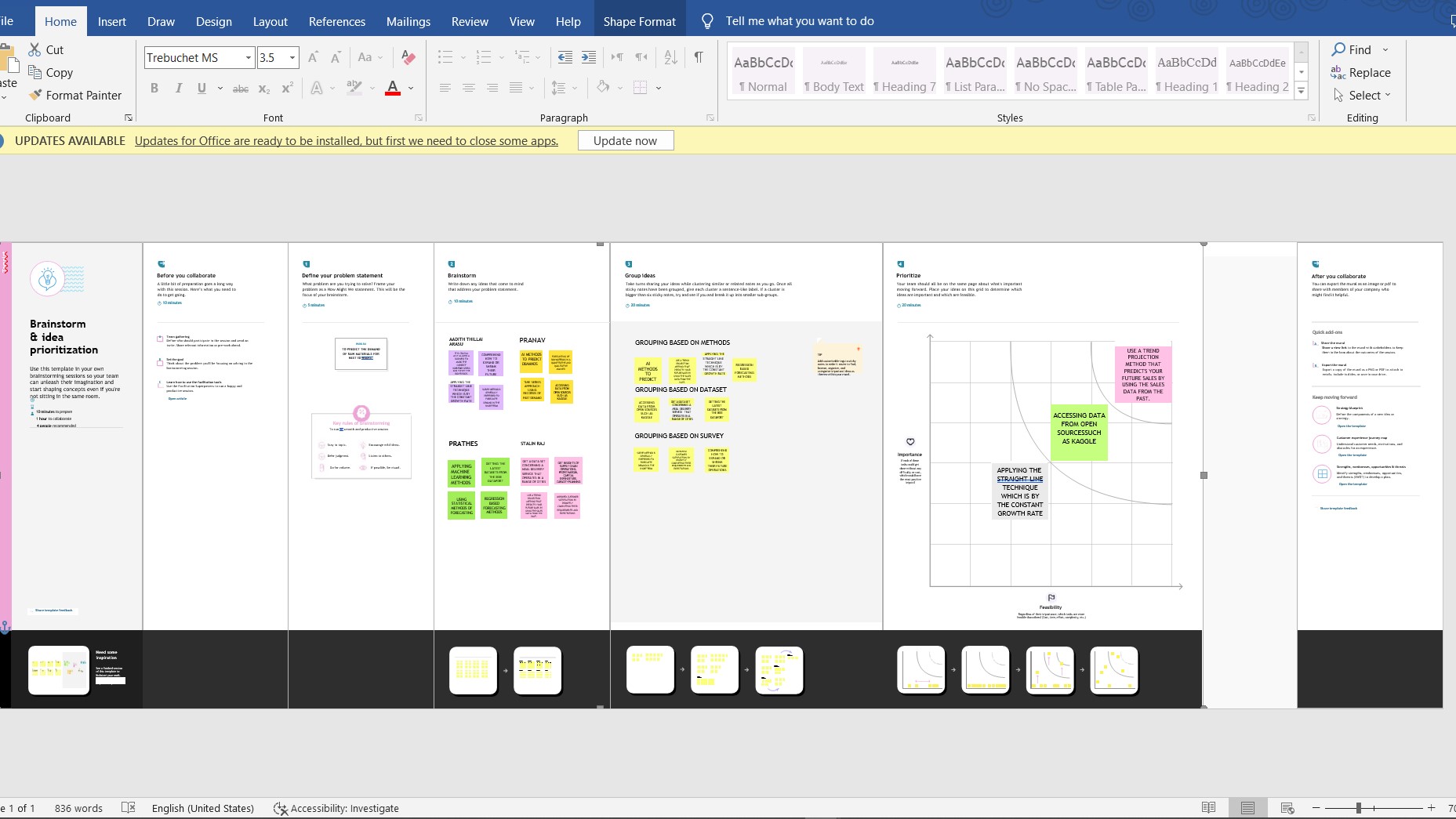


### Ideation & Brainstorming

Brainstorming is a method design teams use to generate ideas to solve clearly defined design problems. In controlled conditions and a free-thinking environment, teams approach a problem by such means as “How Might We” questions. They produce a vast array of ideas

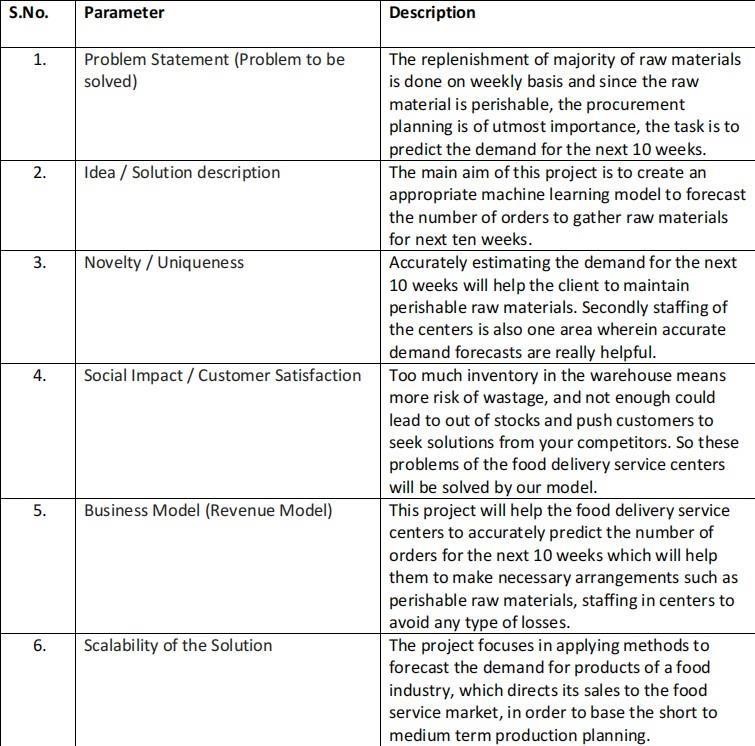
|  |  |
| --- | --- |
| and | draw links between them to find potential solutions |

Everyone in a design team should have a *clear* definition of the target problem. They typically gather for a brainstorming session in a room with a large board/wall forpictures/Post- Its. A good mix of participants will expand the experience pool and therefore broaden the idea space.



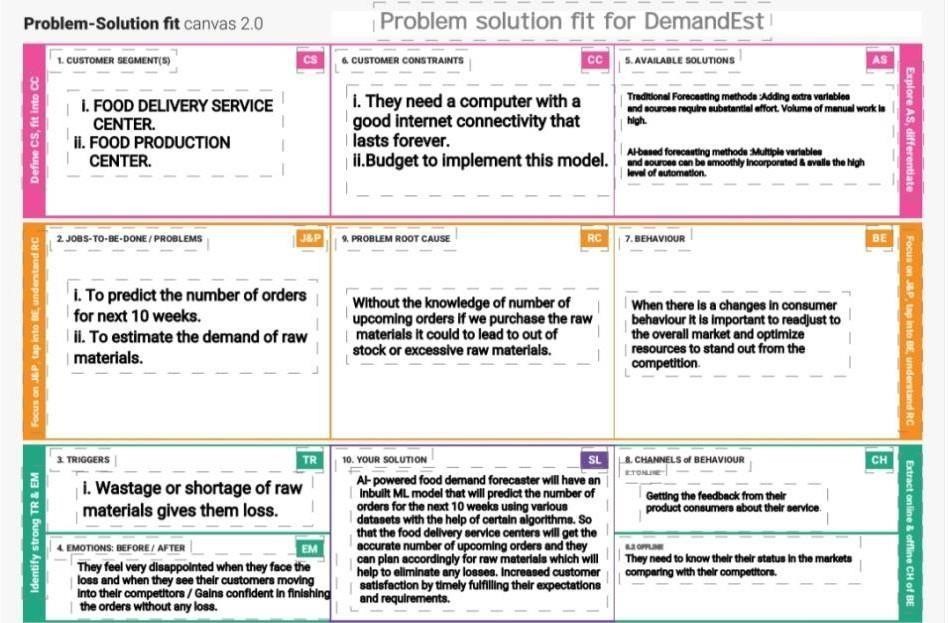
### 3.2 Proposed Solution

[Proposed Solution m](https://www.lawinsider.com/dictionary/proposed-solution)eans the technical solution to be provided by the Implementation agency in response to the requirements and the objectives of the Project.[Proposed Solution](https://www.lawinsider.com/dictionary/proposed-solution) means the Proposed System with modifications that meet the Agency’s requirements as set forth in this RFP.[Proposed Solution m](https://www.lawinsider.com/dictionary/proposed-solution)eans the combination of software, hardware, other products or equipment, and any and all services (including any installation, implementation, training, maintenance and supportservices) necessary to implement the solution described by Vendor in its Proposal.



### 3.4 Problem Solution Fit

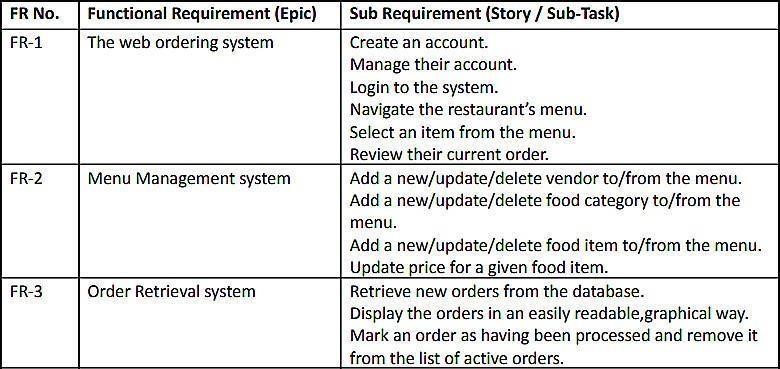
The Problem-Solution Fit simply means that you have found a problem with your customer and that the solution you have realized for it actually solves the customer's problem.



## 4. REQUIREMENT ANALYSIS

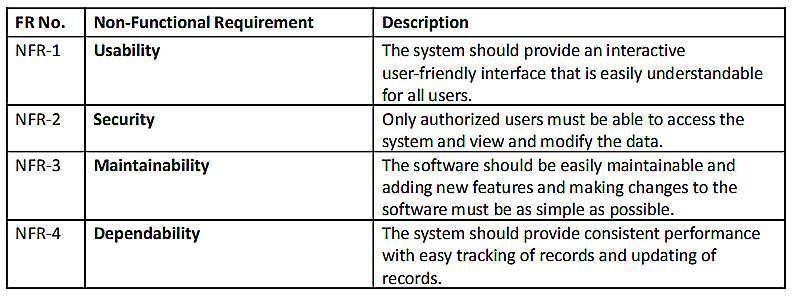
### Functional requirement

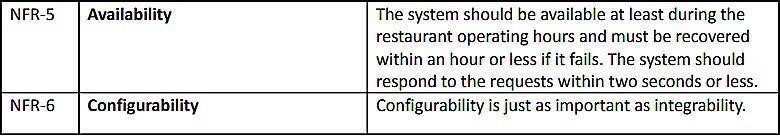
Functional requirements may involve calculations, technical details, data manipulation and processing, and other specific functionality that define what a systemis supposed to accomplish. Behavioral requirements describe all the cases where the system uses the functional requirements, these are captured in use cases.



### Non-Functional requirements

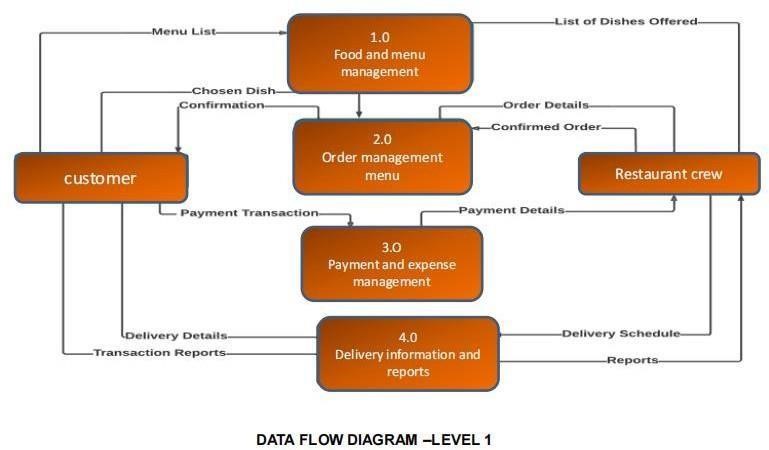
In [systems engineering a](https://en.wikipedia.org/wiki/Systems_engineering)nd [requirements engineering, a](https://en.wikipedia.org/wiki/Requirements_engineering) non-functional requirement (NFR) is [a requirement t](https://en.wikipedia.org/wiki/Requirement)hat specifies criteria that can be used to judge the operation of a system, rather than specific behaviours.

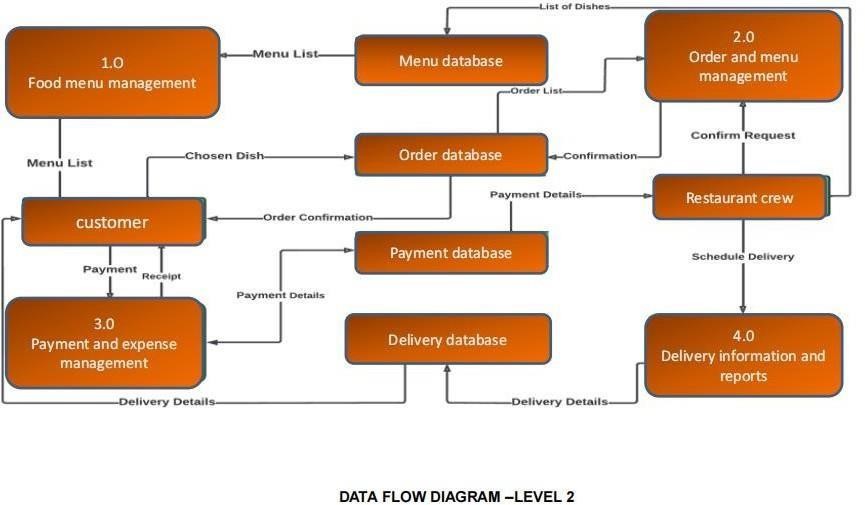




## 5. PROJECT DESIGN Data Flow Diagrams

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.

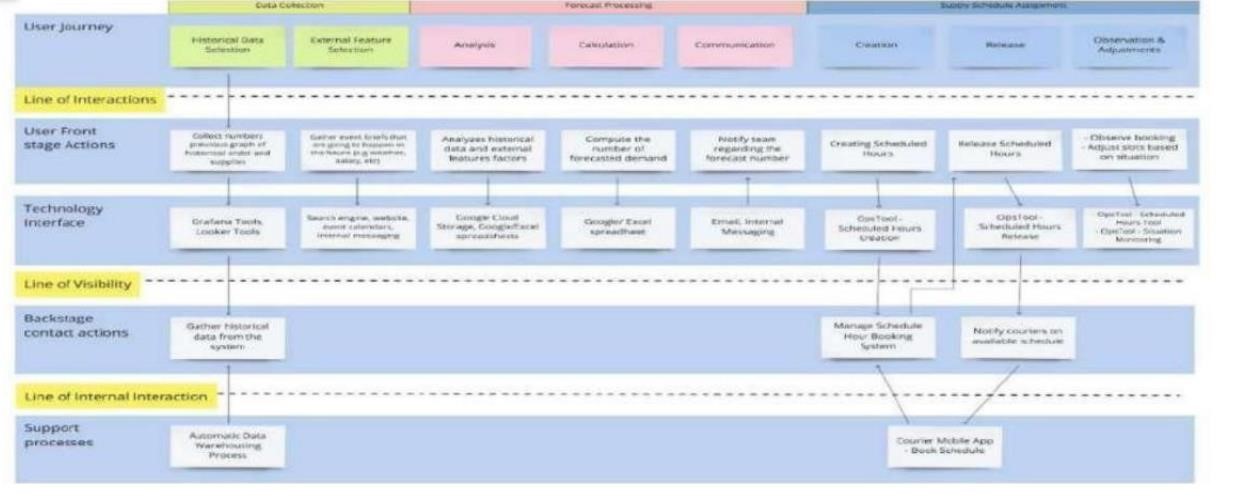




### Solution & Technical Architecture

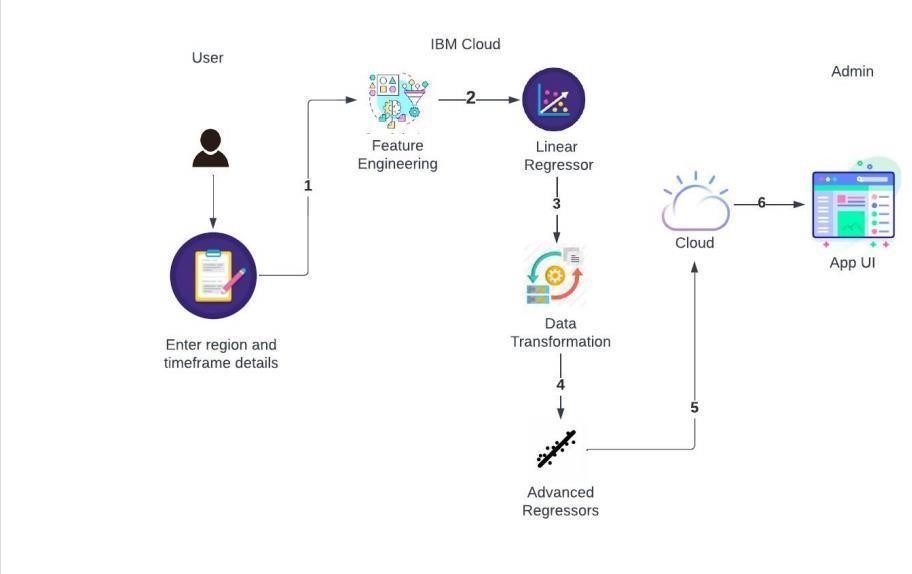
Solution Architecture:

A solution architecture (SA) is an architectural description of a specific solution. SAs combine guidance from different enterprise architecture viewpoints (business, information and technical), as well as from the enterprise solution architecture (ESA).



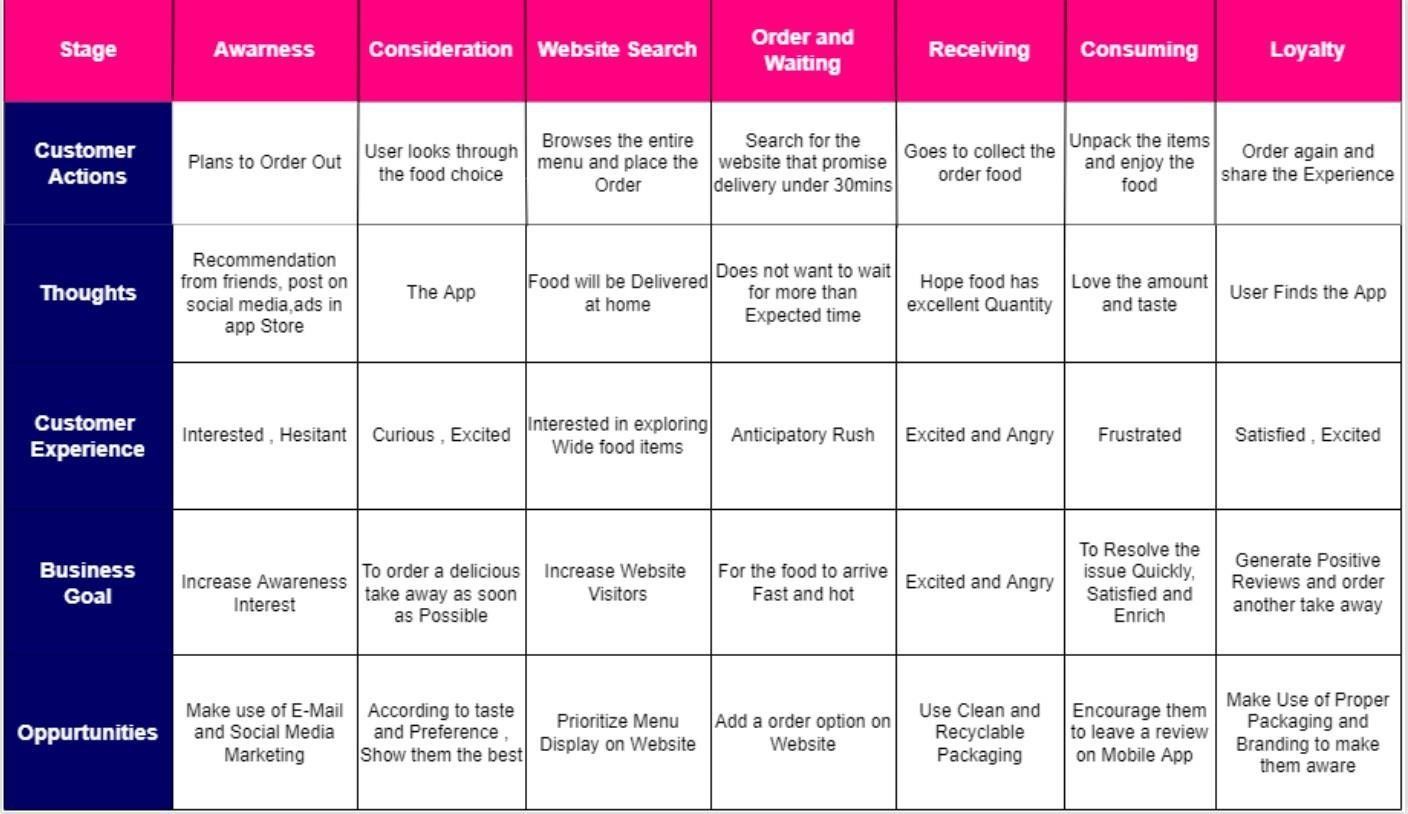
Technical Architecture:

Technical Architecture (TA) is a form of IT architecture that is used to design computer systems. It involves the development of a technical blueprint with regard tothe arrangement, interaction, and interdependence of all elements so that system- relevant requirements are met.



### User Stories

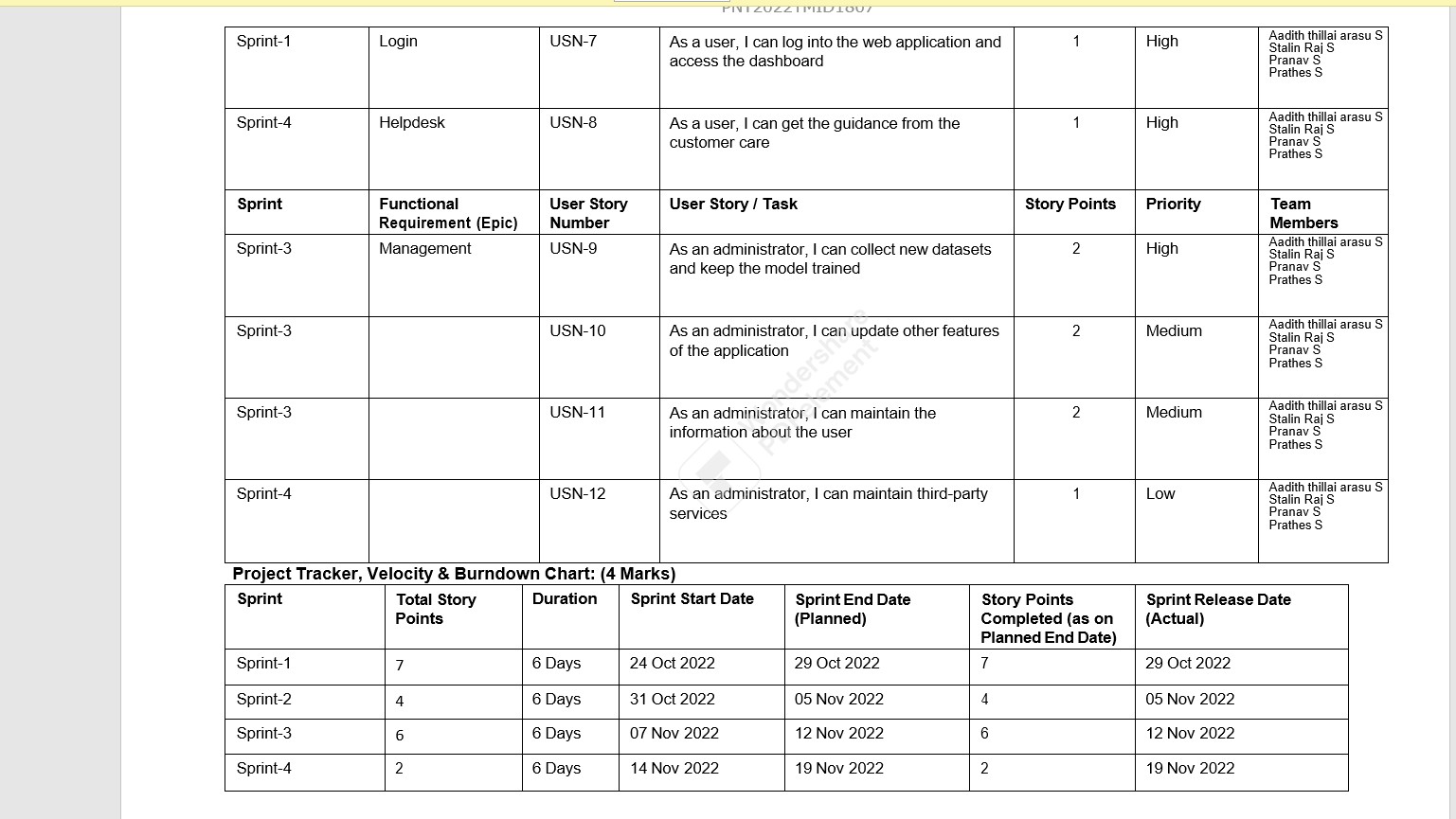
A user story is an informal, general explanation of a software feature written fromthe perspective of the end user or customer. The purpose of a user story is to articulate how a piece of work will deliver a particular value back to the customer.



## 6. PROJECT PLANNING & SCHEDULING

### Sprint Planning & Estimation

In Scrum Projects, Estimation is done by the entire team during Sprint Planning Meeting. The objective of the Estimation would be to consider the User Stories for theSprint by Priority and by the Ability of the team to deliver during the Time Box of the Sprint.

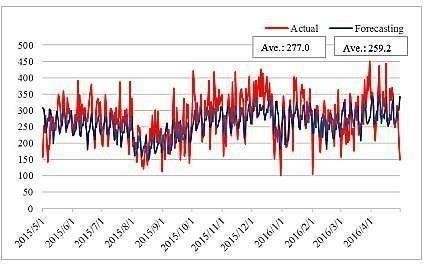


### Sprint Delivery Schedule

A sprint schedule is a document that outlines sprint planning from end to end. It'sone of the first steps in the agile sprint planning process—and something that requiresadequate research, planning, and communication.

### Reports From JIRA

Jira helps teams plan, assign, track, report, and manage work and brings teams together for everything from agile software development and customer support to start-ups and enterprises. Software teams build better with Jira Software.



## 7. CODING & SOLUTIONING

### Data Dictionary

Our base data consists of four csv files containing information about test data,train data and other required information.

* train.csv: Contains information like id, week, center id, meal id, checkout price, base price, emailer for promotion, homepage featured, number of orders. This fileis used for training.



* test.csv: Contains information like id, week, center id, meal id, checkout price,

base price, emailer for promotion, homepage featured. This file is used fortesting. ● fulfilment\_center\_info.csv: Contains information of each fulfilment center.



* meal\_info.csv: Contains information of each meal being served.



**Libraries Used** pandas, numpy, scikit learn, matplotlib, seaborn, xgboost, lightgbm, catboost

### Data Pre-Processing

* There are no Missing/Null Values in any of the three datasets.
* Before proceeding with the prediction process, all the three data sheets need tobe merged into a single dataset. Before performing the merging operation, primary feature for combining the datasets needs to be validated.
* The number of Center IDs in train dataset is matching with the number of CenterIDs in the Centers Dataset i.e 77 unique records. Hence, there won't be any missing values while merging the datasets together.
* The number of Meal IDs in train dataset is matching with the number of Meal IDsin the Meals Dataset i.e 51 unique records. Hence, there won't be any missing values while merging the datasets together.
* As checked earlier, there were no Null/Missing values even after merging the datasets.

### Feature Engineering

Feature engineering is the process of using domain knowledge of the data tocreate features that improves the performance of the machine learning models.

With the given data, We have derived the below features to improve our model performance.

* Discount Amount : This defines the difference between the “base\_Price” and

“checkout\_price”.

* Discount Percent : This defines the % discount offer to customer.
* Discount Y/N : This defines whether Discount is provided or not - 1 if there is Discount and 0 if there is no Discount.
* Compare Week Price : This defines the increase / decrease in price of a Meal fora particular center compared to the previous week.
* Compare Week Price Y/N : Price increased or decreased - 1 if the Price increasedand 0 if the price decreased compared to the previous week.
* Quarter : Based on the given number of weeks, derived a new feature named asQuarter which defines the Quarter of the year.
* Year : Based on the given number of weeks, derived a new feature named as Yearwhich

defines the Year.

### Data Transformation

* Logarithm transformation (or log transform) is one of the most commonly used mathematical transformations in feature engineering. It helps to handle skeweddata and after transformation, the distribution becomes more approximate to normal.
* In our data, the target variable ‘num\_orders’ is not normally distributed. Using thiswithout applying any transformation techniques will downgrade the performanceof our model.
* Therefore, we have applied Logarithm transformation on our Target feature

‘num\_orders’ post which the data seems to be more approximate to normal distribution.

* After Log transformation, We have observed 0% of Outlier data being presentwithin the Target Variable – num\_orders using 3 IQR Method.

### Evaluation Metric

The evaluation metric for this competition is 100\*RMSLE where RMSLE is Root ofMean Squared Logarithmic Error across all entries in the test set.

### Initial Approach

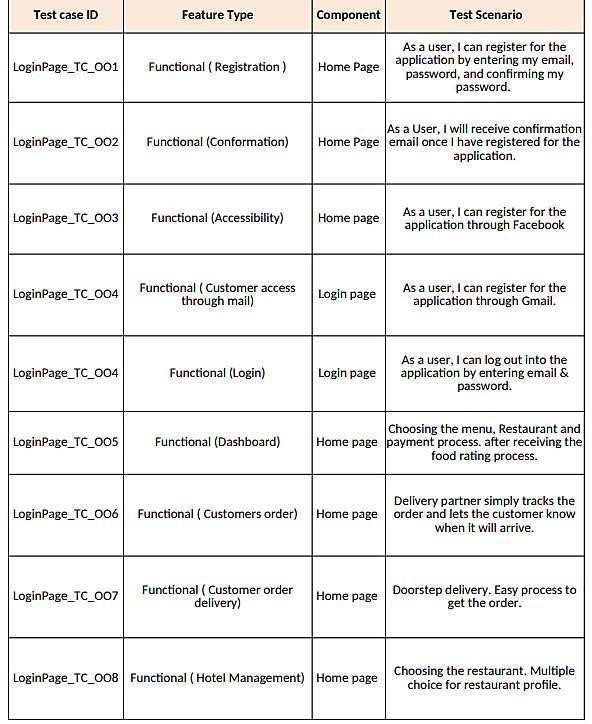
* Simple Linear Regression modelwithout any feature engineering and data transformation which gave a RMSE : 194.402
* Without feature engineering and datatransformation, the model did not performwell and could'nt give a good score.
* Post applying feature engineering and data transformation (log and log1p transformation), Linear Regression model gave a RMSLE score of 0.634.

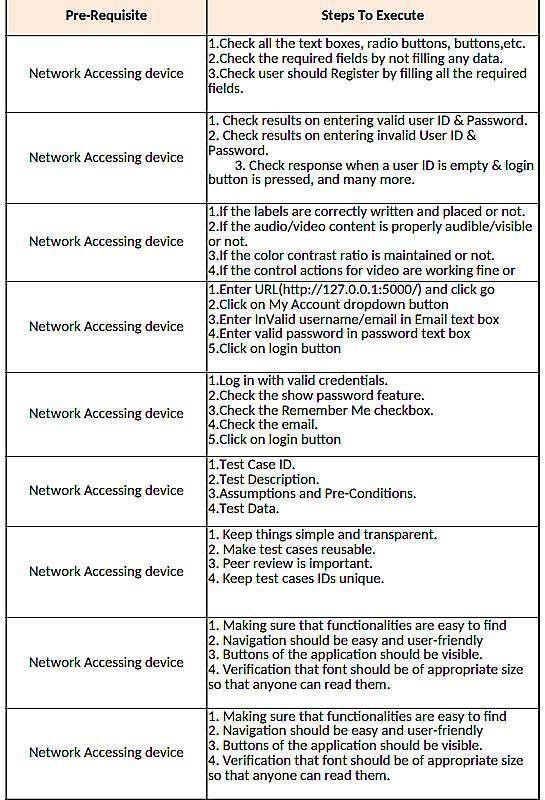
### Advanced Models

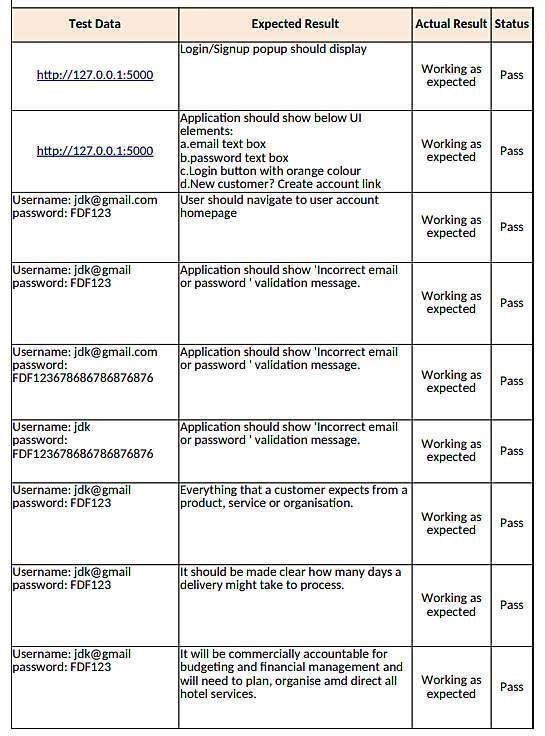
* With improvised feature engineering, built advanced models using Ensemble techniques and other Regressor algorithms.
* Decision Tree Regressors performed well on the model which gave muchreduced RMSLE.
* With proper hyper-parameter tuning, Decision Tree Regressor performed well onthe model and gave the lease RMSLE of 0.5237

## 8. TESTING Test Cases

A test case includes information such as test steps, expected results and datawhile a test scenario only includes the functionality to be tested.



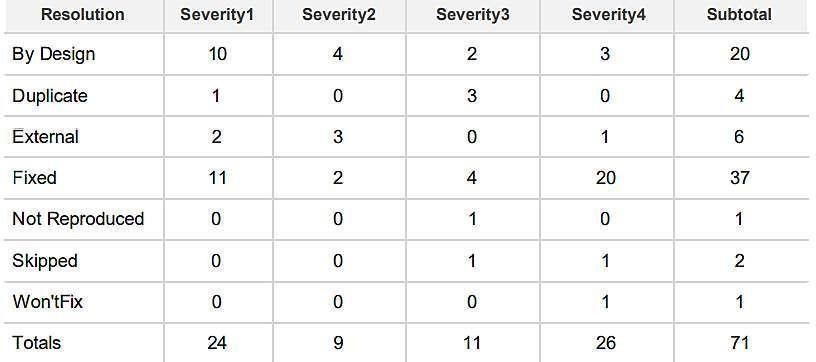




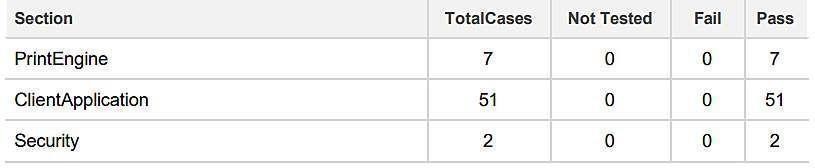
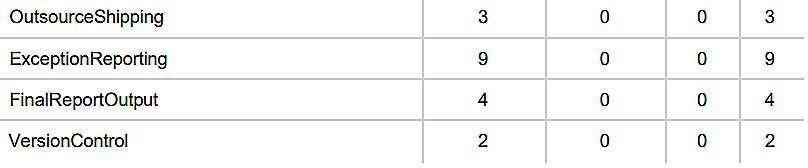
### User Acceptance Testing

User Acceptance Testing (UAT), which is performed on most UIT projects, sometimes called beta testing or end-user testing, is a phase of software developmentin which the software is tested in the "real world" by the intended audience or businessrepresentative.

Defect Analysis:



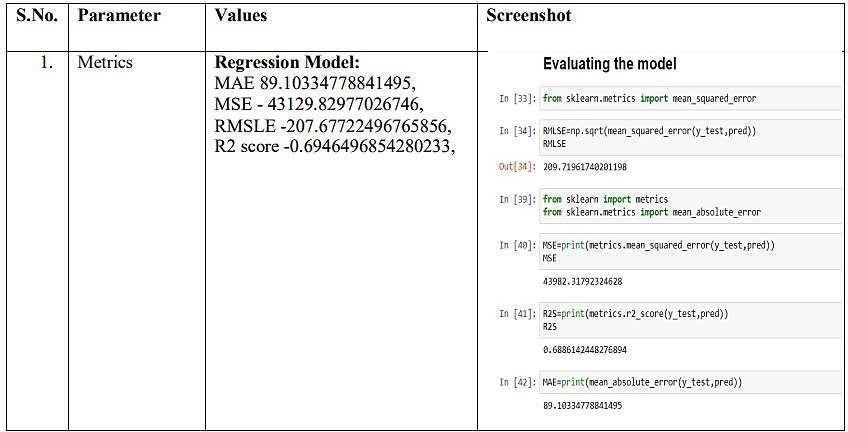
Test Case Analysis:

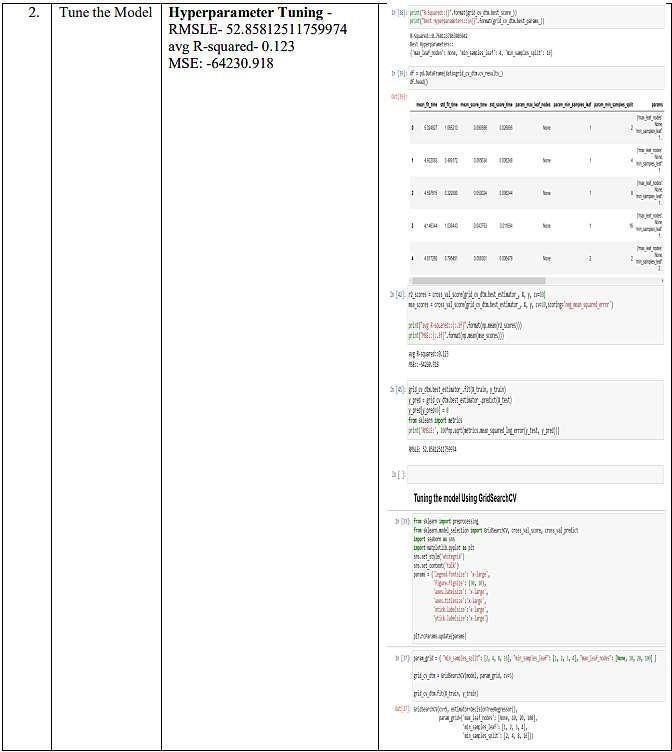


## 9. RESULTS

### Performance Metrics

Performance testing is the practice of evaluating how a system performs in terms of responsiveness and stability under a particular workload. Performance testsare typically executed to examine speed, robustness, reliability, and application size.





### 10. ADVANTAGES & DISADVANTAGES

**Advantages:**

1. Food wastage will be minimized.
2. Simple and easy to use framework.

**Disadvantages:**

1. The output obtained may not be precised, due to the use of limited datasets.

### 11. APPLICATIONS

This project focuses on one food deliveryclient, which delivers food in many differentcities through distribution networks and fulfillment centers.

## 12. CONCLUSION

The main moto behind this project is to reduce food wastage.The availability ofthe food items makes the society better. Our purposed model would definitely come handy to a company for predicting then number of food orders and help them to servetheir customers better.

### 13. FUTURE SCOPE

1. Working on the frontend to make the framework more dynamic.
2. In the future, we also plan to improve forecasting accuracyand research onthe efficiency of store management.

## 14. APPENDIX

### SOURCE CODE: home.html

<!DOCTYPE html>

<html lang="en">

<head>

<meta charset="UTF-8" />

<meta http-equiv="X-UA-Compatible" content="IE=edge" />

<meta name="viewport" content="width=device-width, initial-scale=1.0" />

<link rel="stylesheet" href="home.css" />

<title>Document</title>

</head>

<body>

<div class="main">

<img class="bg" src="img1.jpg" />

</div>

<div class="nav"></div>

<div class="n">

<a>Home</a>

<a>Results</a>

</div>

</body>

</html>

### upload.html

<!DOCTYPE html>

<html lang="en">

<head>

<meta charset="UTF-8">

<meta http-equiv="X-UA-Compatible" content="IE=edge">

<meta name="viewport" content="width=device-width, initial-scale=1.0">

<link rel="stylesheet" href="result.css">

<title>Document</title>

</head>

<body>

<div class="main">

<img class="bg" src="img1.jpg" />

</div>

<div class="nav"></div>

<div class="n">

<a>Home</a>

<a>Results</a>

</div>

<div class="form">

<form action = "http://localhost:5000/pred" method = "post">

<span>Enter homepage\_featured</span><input type="text" name="homepage\_featured"/>

<br/>

<span>Enter email\_for\_promotions</span><input type="text" name="email\_for\_promotions"/>

<br/>

<span>Enter op\_area</span><input type="text" name="op\_area"/>

<br/>

<span>Enter food\_type</span><input type="text" name="food\_type"/>

<br/>

<span>Enter city\_code</span><input type="text" name="city\_code"/>

<br/>

<span>Enter region\_code</span><input type="text" name="region\_code"/>

<br/>

<span>Enter category</span><input type="text" name="category"/>

<br/>

<br/>

<input type="submit" />

</form>

</div>

</body>

</html>

**Result.css** bg{ width:100vw; height: 25vw;; position:absolute;

} .nav{ width:100vw; height:50px; position:absolute; background-color: grey; z-index:99; opacity:0.6;

}

.n{ right:0; position:absolute; z-index:99999;

} a{ margin:30px; cursor: pointer; font-size: xx-large; font-weight: bolder; color:red;

}

.form{ position:absolute; margin-top: 25vw;; margin-left:30%;

}

### app.py

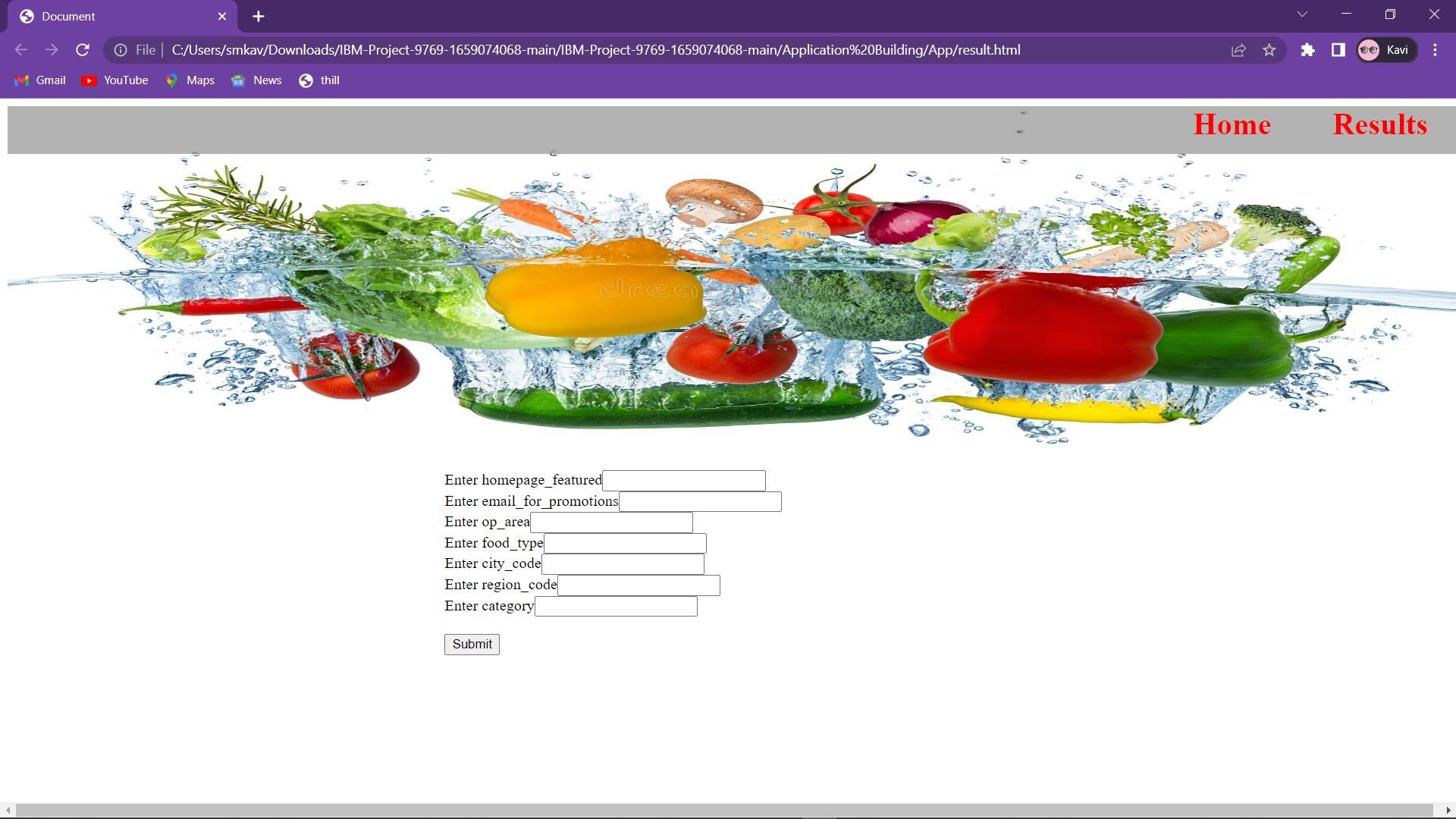
from flask import \* import pickle import os import pandas as pd import sklearn import numpy as np app = Flask(\_\_name\_\_) #creating the Flask class object @app.route('/pred',methods = ['GET','POST']) #decorator drfines the def home():

homepage\_featured=request.form['homepage\_featured'] food\_type=request.form['food\_type'] model = pickle.loads(open('fdemand.pkl','rb').read()) features = [1 for x in request.form.values()] feature\_value = [np.array(features)] prediction = model.predict(feature\_value) output = prediction[0]

return "The Number of Orders is : "+str(output)

if \_\_name\_\_ =='\_\_main\_\_': app.run(debug = True)

**OUTPUT SCREENSHOTS**:



**GITHUB LINK:**

[https://github.com/IBM-EPBL/IBM-Project-](https://github.com/IBM-EPBL/IBM-Project-9769-1659074068.git)48763-1660812765