

horizontal line

Project Summary

| Batch details | DSE–GURUGRAM-AUGUST(2022) |
| --- | --- |
| Team members | RITESH KUMAR  VATAN KUMAR  AMAN SINGH SANJAY KUMAR GUPTA  YASHARTH SRIVASTAVA |
| Domain of Project | SUPPLY CHAIN ANALYTICS |
| Proposed project title | FOOD DEMAND FORECASTING |
| Group Number | 5 |
| Team Leader | RITESH KUMAR |
| Mentor Name | MR CHANDRAN VENKATESAN |

Date: 22-12-2022

 Signature of the the Mentor :- Signature of the Team Leader :





**Table of Contents**

| Sl NO | Topic | Page No |
| --- | --- | --- |
| 1 | Overview | 3 |
| 2 | Business problem goals | 3-4 |
| 3 | Topic survey in depth | 5-6 |
| 4 | Critical assessment of topic survey | 6-7 |
| 5 | Methodology to be followed | 8 - 12 |
| 6 | References |  |





Project Details

# OVERVIEW

Our client is a meal delivery company that operates in multiple cities. They have various fulfillment centers in these cities for dispatching meal orders to their customers.

The client wants us to help these centers with demand forecasting for upcoming weeks so that these centers will plan the stock of raw materials accordingly.

The replenishment of the majority of raw materials is done on a weekly basis and since the raw material is perishable, procurement planning is of utmost importance. Secondly, staffing of the centers is also one area wherein accurate demand forecasts are really helpful.

# Business problem statement (GOALS)

1. What would you achieve by this project ?

* 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.

Hence, through this project we would optimize the inventory in the warehouse in order to reduce the wastage as well as to avoid the situations leading to out-of-stocks scenario.





1. How would this help the business or clients ?

* providing insight into the data in order to optimize inventory which would lead cost minimization and thus increase a better resource planning of man and materials.
* Increased customer satisfaction by timely fulfilling their expectations and requirements.

1. What is the further scope of the project ?

Demand forecasting helps reduce risks and make efficient financial decisions that impact profit margins, cash flow, allocation of resources, opportunities for expansion, inventory accounting, operating costs, staffing, and overall spend. All strategic and operational plans are formulated around forecasting demand.

1. Limitation of the project

* Our model is not a generalized model but a business oriented, realistic and specific model meaning needs to be calibrated and updated with time and business.





# TOPIC SURVEY IN BRIEF

1. Problem understanding

* Your client is a meal delivery company which operates in multiple cities.
* The replenishment of the majority of raw materials is done on a weekly basis and since the raw material is perishable, the procurement planning is of utmost importance.
* Secondly, staffing of the centers is also one area wherein accurate demand forecasts are really helpful.

1. Current solution to the problem

* Manual intervention - Guessing the number of orders on the basis of previous day orders count. And, thus utilizing the inventory on the basis of that guess.

1. Proposed solution to the problem

* Algorithm and ML driven intervention in order to predict the number of order for future sale and business growth.







4. Reference to the problem

* <https://blog.3ds.com/brands/delmia/food-waste-is-a-big-problem-that-can-be-solved/>
* <https://smartinternz.com/guided-project/food-demand-forecasting-for-food-delivery-company-using-ibm-cloud-659>
* <https://www.aquare.la/en/ai-for-demand-forecasting-in-the-food-industry/>

# CRITICAL ASSESSMENT OF TOPIC SURVEY

KEY AREAS:-

1. Checkout price ranges vary similar to base price with a little discount.

2. Maximum orders placed from restaurants are those having operation area

within 3.5 to 4.5 SqKm.

3. Beverages and Rice Bowl are the food category which has the highest number

of orders and Soup and Fish are the food category with the least number of orders.

4. In the Food Category we could see that Indian-Rice Bowl has the

highest number of orders while Indian-Biryani has the least number of orders.

 5. We can see that No. of Orders remain constant across the week. This Shows a





steady Demand for the Service Provider.

6. Most of the Orders have Checkout\_Price in range from 100 to 300. It shows

that there are less bulk orders.

7. Most of the Orders have Base Price less than 500. It shows customers are

ordering for 1 or 2 people only at once.

8. Beverages are the Most Ordered Item.Vs while Biryani, Soup, Fish are the

LeastOrdered Dishes.

9. Italian is the Most Ordered Cuisine, followed by Thai and Indian.

10. Type\_A Centers are catering to most of the Orders, while Type\_B center deals

with the Lowest no. of orders.

GAPS:-

1. The correlation between the variable with each order is very high so we

believe that the multicollinearity is also low in them.

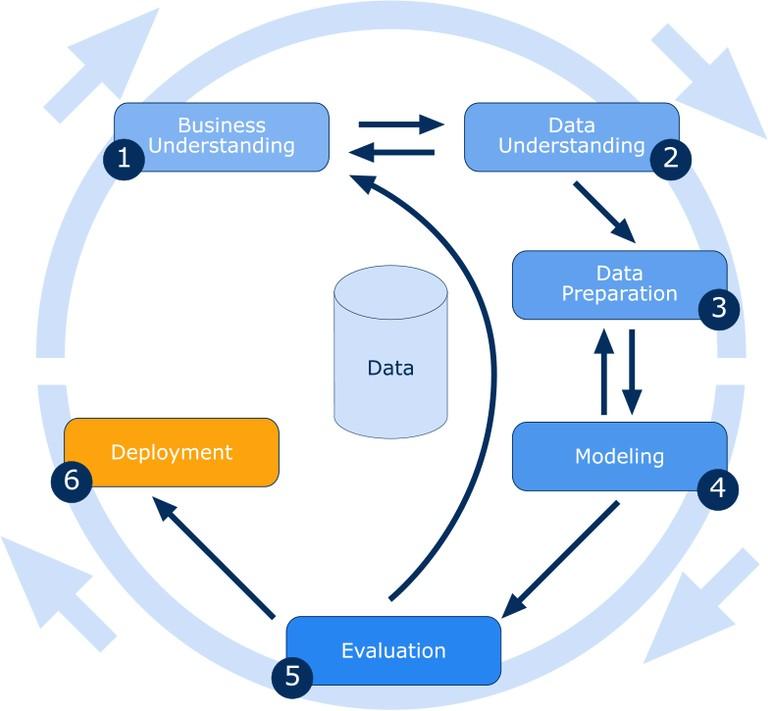
1. But base\_price & checkout\_price shows high correlation.
2. Missing Value There are no missing value In the Dataset
3. Type\_B center deals with the Lowest no. of orders.







# METHODOLOGY TO BE FOLLOWED



BUSINESS UNDERSTANDING

Our client is a meal delivery company that operates in multiple cities. They have various fulfillment centers in these cities for dispatching meal orders to their customers.

The client wants us to help these centers with demand forecasting for upcoming weeks so that these centers will plan the stock of raw materials accordingly.

The replenishment of the majority of raw materials is done on a weekly basis and since the raw material is perishable, procurement planning is of utmost importance. Secondly, staffing of the centers is also one area wherein accurate demand forecasts are really helpful.





DATA UNDERSTANDING

The data is retrieved using 3 different tables.

Weekly Demand data (train.csv): Contains the historical demand data for all centers,

test.csv contains all the following features except the target variable.

| Variable | Definition |
| --- | --- |
| id | Unique ID |
| week | Week No |
| center\_id | Unique ID for fulfillment center |
| meal\_id | Unique ID for Meal |
| checkout\_price | Final price including discount, taxes & delivery charges |
| base\_price | Base price of the meal |
| emailer\_for\_promotion | Mailer sent for promotion of meal |
| homepage\_featured | Meal featured at homepage |
| num\_orders | (Target) Orders Count |

fulfilment\_center\_info.csv: Contains information for each fulfillment center





| Variable | Definition |
| --- | --- |
| center\_id | Unique ID for fulfillment center |
| city\_code | Unique code for city |
| region\_code | Unique code for region |
| center\_type | Anonymized center type |
| op\_area | Area of operation (in km^2) |

meal\_info.csv: Contains information for each meal being served

| Variable | Definition |
| --- | --- |
| meal\_id | Unique ID for the meal |
| category | Type of meal (beverages/snacks/soups….) |
| cuisine | Meal cuisine (Indian/Italian/…) |





# DATA PREPARATION

Exploratory Data Analysis(EDA) :

⮚Aggregation of Data

⮚Descriptive Statistics

⮚Univariate Analysis- Numerical and Categorical Variable

⮚Bivariate Analysis- Numerical and Categorical

⮚Data Scaling and Transformation

⮚Missing Value

⮚ Outlier Analysis

# MODELING

Base Model

* Linear regression

Intermediate Model

* Decision Tree Regressor
* Random Forest Regressor
* XGBoost Regressor

Final model

* Random Forest Regressor (Hypertuned using GridSearchCV)





EVAlUATION

RMSE

**Root Mean Square Error** is the measure of how well a regression line fits the data points. RMSE can also be construed as Standard Deviation in the residuals.

Adjusted R2

Adjusted R2 is **a corrected goodness-of-fit (model accuracy) measure for linear models**. It identifies the percentage of variance in the target field that is explained by the input or inputs. R2 tends to optimistically estimate the fit of the linear regression.

R-Squared

R-Squared (R² or the coefficient of determination) is **a statistical measure in a regression model that determines the proportion of variance in the dependent variable that can be explained by the independent variable**. In other words, r-squared shows how well the data fit the regression model (the goodness of fit)

# REFERENCES

[1] Patrick Meulstee and Mykola Pechenizkiy,” Food Sales Prediction: “If Only It Knew What We Know”” 2008 IEEE International Conference on Data Mining Workshop.

[2] Yoichi Motomura, Baysian network, Technical Report of IEICE, Vol.103, No.285, pp.25-30, 2003.

[3] Yoichi Motomura, Baysian Network Softwares, Journal of the Japanese Society for Artificial Intelligence, Vol.17 No.5, pp.1-6, 2002.

[4] D. Adebanjo and R. Mann. Identifying problems in forecast-ing consumer demand within the fast paced commodity sector. Benchmarking: An International Journal, 7(3):223– 230, 2000.

[5] Bohdan M. Pavlyshenko,” Machine-Learning Models for Sales Time Series Forecasting”, 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), Lviv, Ukraine, 21–25 August 2018.

[6] İrem İşlek and Şule Gündüz Öğüdücü,” A Retail Demand Forecasting Model Based on Data Mining Techniques”.

[7] <https://statisticshowto.com/lasso/regression>

[8] <https://en.wikipedia.org/wiki/Random_forest>

[9]<https://towardsdatascience.com/supportvectormachines-svm-c9ef22815589>

[10] <https://towardsdatascience.com/httpsmedium-comvishalmorde-xgboost-algorithmlong-she-may-reinedd9f99be63d>





