Project Summary

Batch details	DSE – MARCH 2022 (CHENNAI)
Team members	Vigneshwar K TamilSelvan M Balaji Prasanth S Rahul R Adarsh R
Domain of Project	E-Commerce (Supply Chain)
Proposed project title	Backorder Prediction
Group Number	Team 11
Team Leader	Vigneshwar K
Mentor Name	Pratik Sonar

Date: 05/06/2022

Pratik Sonar   Vigneshwar K

Signature of the Mentor   **Signature of the Team Leader**

Table of Contents

S.No	Topic	Page No
1	Overview	3
2	Business Problem Goals	3
3	Topic Survey In Depth	4
4	Critical Assessment Of Topic Survey	4
5	Methodology To Be Followed	5
6	References	7

Project Details

Overview

For industries, Inventory management has to be monitored and optimised based on supply and demand and it is one of important factors in customer satisfaction simultaneously company has to perform stocking of products efficiently to cut costs associated with it. Backorder translates that an item maynot be held in the company's available inventory but could still be in production, or the company may need to still manufacture, it is an indication that demand for a company's product outweighs its supply, if not optimised company loses it customer base as it could not provide the product required by customer resulting in dissatisfaction and initiation of refunds from order cancellation by customer. Identifying parts with the highest chances of shortage prior its occurrence can present a high opportunity to improve an overall company's performance. In this project we worked on the dataset of the company Sigma Retail Ltd (E-Commerce company) for Backorder prediction.

Business Problem Statement (GOALS)

i. **Business Problem Understanding:**

The company wants to predict which of its products has a chance of getting backordered using Machine Learning Models. Based on the past data from supply chain, inventories and sales given to us, we need to classify a particular product will go into backorder.

ii. **Business Objective:**

The Business wants to improve the effectiveness of inventory management by avoiding backorders and overstocking of all products leading huge overhead cost. In order to minimize cost and improve customer satisfaction, we will deploy a classification machine learning model empowering the company in effective stocking and backorder handling.

iii. **Approach:**

Studying past data provided by Supply Chain, Sales and Inventory, understanding the dependency of these features onto prediction of backorder of a product by deploying classification machine learning algorithms striving for a higher accuracy in prediction.

iv. **Conclusions**:

By implementing prediction model (Logistic Regression, Random Forest, Decision tree, XGBoost) company would be able achieve better efficiency in product stock and customer satisfaction.

TOPIC SURVEY IN BRIEF

i. Problem Understanding

Backorders are unavoidable, but by anticipating which things will be backordered, planning can be streamlined at several levels, preventing unexpected strain on production, logistics, and transportation. ERP systems generate a lot of data (mainly structured) and contain a lot of historical data; if this data can be properly utilized, a predictive model to forecast backorders and plan accordingly can be constructed. Based on past data from inventories, supply chain, and sales, classify the products as going into backorder (Yes or No).

ii. Current solution to the problem

Despite having a good sales forecasting system sometimes these situations are inevitable because of the factors which can't be controlled or unpredictable events. If many items are going on Backorders consistently it is a sign that companies' operations are not properly planned and also there is a very high chance of missing out business on the products.

iii. Proposed solution to the problem

The solution here is a Classification based Machine Learning model. It can be implemented by different classification algorithms (like Logistic Regression, Random Forest, Decision tree, XGBoost and so on. Here First we are performing Data pre-processing step, in which Data Profiling, feature engineering, feature selection, feature scaling, VIF steps are performed and then we are going to build model.

iv. Reference to the problem: https://www.kaggle.com/datasets/chandanareddy12/back-order-prediction

# CRITICAL ASSESSMENT OF TOPIC SURVEY

1. Find the key area, gaps identified in the topic survey where the project can add value to the customers and business

   Part backorders is a common supply chain problem, wherein a customer places an order for a product that is temporarily out of stock. The percentage of items backordered and the number of backorder days are important measures of the quality of a company's customer service and the effectiveness of its inventory management.

2. What key gaps are you trying to solve?

> A company can manage its inventory more efficiently using a prediction on the backorder risk for the products. Goal here is to use the past data and metadata around the backorders, and provide a prediction on the potential products for backorders.

METHODOLOGY TO BE FOLLOWED
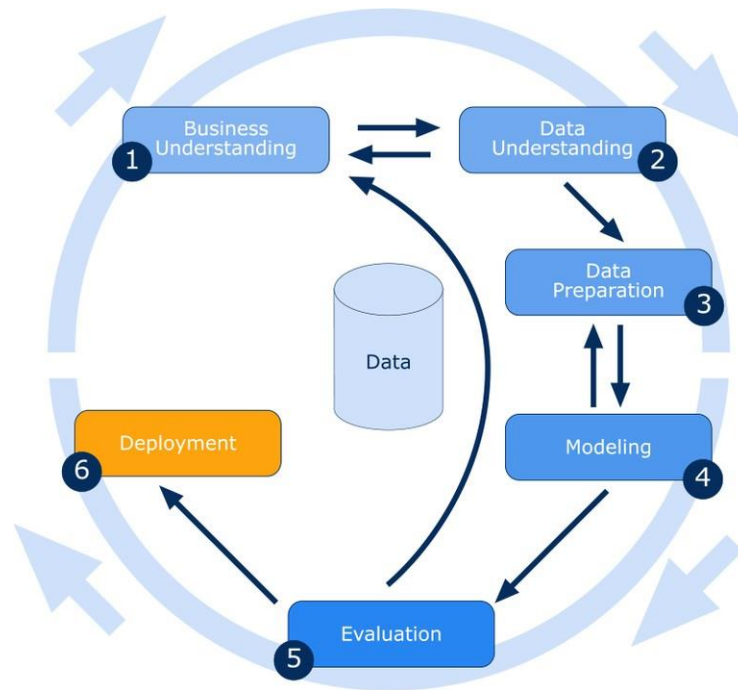
Business Understanding:

> Sigma Retails Ltd is a leading online store providing products ranging from clothing, home improvements to grocery. Recently, client has been finding it difficult to manage the backorders, resulting into increasing customer issues and a decline in customer satisfaction.

> Client is looking for ways to improve backorders handling. With the help of this analysis, a reasonable prediction on the products that can go on backorder is expected. Such a prediction could immensely help client to plan for a more effective stocking and backorder handling

Data Understanding:

> Dataset consists of the historical data around the backorders. It has 23 features and 1687861 observations.
> Dataset can be found at: https://www.kaggle.com/datasets/chandanareddy12/back-order-prediction

INDEPENDENT VARIABLES:

national_inv	Current inventory level for the part
lead_time	Transit time for product (if available)
in_transit_qty	Amount of product in transit from source
forecast_3_month	Forecast sales for the next 3 months
forecast_6_month	Forecast sales for the next 6 months
forecast_9_month	Forecast sales for the next 9 months
sales_1_month	Sales quantity for the prior 1-month time period
sales_3_month	Sales quantity for the prior 3-month time period
sales_6_month	Sales quantity for the prior 6-month time period
sales_9_month	Sales quantity for the prior 9-month time period
min_bank	Minimum recommend amount to stock
potential_issue	Source issue for part identified
pieces_past_due	Parts overdue from source
perf_6_month_avg	Source performance for prior 6 month period
perf_12_month_avg	Source performance for prior 12-month period
local_bo_qty	Amount of stock orders overdue
deck_risk	Part risk flag
oe_constraint	Part risk flag
ppap_risk	Part risk flag
stop_auto_buy	Part risk flag
rev_stop	Part risk flag

DEPENDENT VARIABLE (TARGET VARIABLE):

went_on_backorder - Product actually went on backorder or not. This is the target variable.

Data Preparation:

Goal: Prepare the backorder dataset for EDA and Modelling

Tasks performed:
• Handling inconsistent column names and datatype

• Missing Data handling

• Removal of duplicate rows

• Handling columns with repetitive values

• Handling the outliers

• Write the clean data into a new file for further steps

Modelling:

The modelling process is actually very iterative and general goal of data analysis is to acquire knowledge from data so as to improve business decision towards sustainability and profitability in this case.

Statistical model provides a convenient framework for achieving this project objective because they make it possible to know importance of each survey variables and how they collectively influence Backorder through predictions.

Evaluation:

Based on the performance of all the predictive model, choosing the most suitable predictive model we give accurate and more precise predictions.
we will recommend this model to client.

Timeline Chart:

Date	Project Start Date	Synopsis Date	1st Work in Progress (EDA Analysis)	Interim Report	Presen tation	2nd Work Progress (Model Building)	Final Report	Final Presentation
24-05-2022	■							
05-06-2022		■						
13-06-2022			■					
15-06-2022				■				
17-06-2022					■			
18-07-2022						■		
20-07-2022							■	
22-07-2022								■

Reference documents of CRISP-DM
- https://paginas.fe.up.pt/~ec/files_0405/slides/02%20CRISP.pdf
- https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining

REFERENCES

The references can be blogs, articles or even social media news relevant to explain the importance of the projects.
- Classification concepts - https://www.analyticsvidhya.com/blog/2021/09/a-complete-guide-to-understand-classification-in-machine-learning/
- Key strategies - https://www.minderest.com/blog/2018/03/23/do-out-stock-strategies-work-increase-sales

Notes For Project Team

Original owner of data	Sigma Retails Ltd
Data set information	Product Backorder Prediction
Any past relevant articles using the dataset	https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00345-2
Reference	Kaggle
Link to web page	https://www.kaggle.com/datasets/chandanareddy12/back-order-prediction

****