MACHINE LEARNING PROJECT

Architecture Report

On

BACKORDER PREDICTION

Made by: Gaurav Singh

Domain: E-commerce

ABSTRACT

Backorder is an order which can't be fulfilled at the given time due to lack of supply or the product is currently out of stock or not in inventory but can guarantee delivery of the goods or service requested by a certain date in the future because the production of goods or replenishment of inventory is underway. Unlike in the situation of Out-of-stock where the delivery date of the goods can't be promised , in the Backorder scenario the customers are allowed to shop for the products and order. Simply put Backorder can be thought of as an order with a delayed delivery date.

It is a common supply chain problem, impacting an inventory system service level and effectiveness. Identifying parts with the highest chances of shortage prior its occurrence can present a high opportunity to improve an overall company's performance.
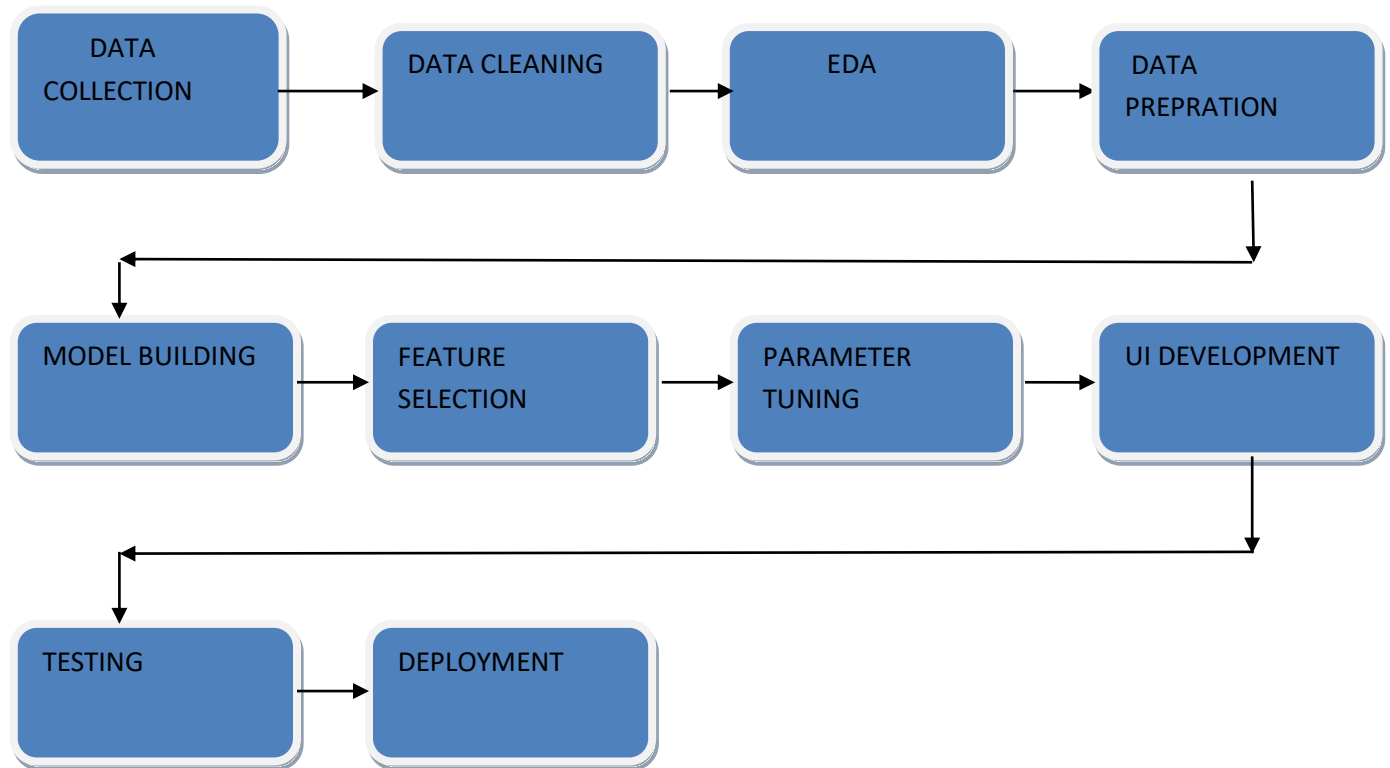
Table Of Contents

1.Introduction:

The purpose of preparing a low-level design report for the project is to impart necessary information to concerned and prospective stakeholders and management personnel, in respect to the design adopted to build the project, the interface created to interact with the model and the system & input requirements that need to be satisfied in order for the model to perform as needed.

Also, the document will act as a source of guidance to potential programmers who wish to replicate the project for their personal or commercial use as they could follow the steps mentioned to rebuild the machine learning model and user interface from scratch to the final state.

2.Architecture:

```
DATA          →   DATA CLEANING   →   EDA   →   DATA
COLLECTION                                        PREPRATION
                                                      ↓
MODEL BUILDING →  FEATURE    →  PARAMETER  →  UI DEVELOPMENT
                  SELECTION      TUNING
                                                      ↓
TESTING    →   DEPLOYMENT
```

3. ARCHITECTURE DESCRIPTION

1. Data:

- **sku** : unique id for a product
- **national_inv**: present national level of inventory of the product
- **lead_time** : the amount of time between when a purchase order is placed to replenish products and when the order is received in the warehouse.
- **in_transit_qty** : qty of goods in transit
- **forecast_3_month** : Forecasted sales of the product for the next 3 months.
- **forecast_6_month** : Forecasted sales of the product for the next 6 months.
- **forecast_9_month** : Forecasted sales of the product for the next 9 months.
- **sales_1_month** : Actual Sales of the product in the last 1 month.
- **sales_3_month** : Actual Sales of the product in the last 3 months.
- **sales_6_month** : Actual Sales of the product in the last 6 months.
- **sales_9_month** : Actual Sales of the product in the last 9 months.
- **min_bank**: Minimum amount of stock recommended to have.
- **potential_issue**: Any problem identified with the product or part.
- **pieces_past_due** : product kept for long time, past their expiry date.
- **perf_6_month_avg** : Average performance of product over last 6 months.

- **perf_12_month_avg**: Average performance of product over last 12 months.

- **local_bo_qty** : (undeliverable orders / total number of orders)*100.

- **deck_risk** : risk associated with keeping the items in stock

- **ppap_risk** : used to determine whether a production will produce parts with consistency and repeatability

- **stop_auto_buy** : Has the auto buy for the product, which was back ordered, cancelled.

TARGET FEATURE : went_on_backorder - Whether an items was backordered or not

2. Data Cleaning:

In this process, we have cleaned up all the data because data is present in very bad format which was can not recognized by machine .So data Cleaning is done very first by data validation methods. In which we are create a json file in which name of file , numbers of columns etc information present. File name is given in "BackOrder_08012020_120000" Format

Dataset Size – 1929935 X 23

3. EDA:

In this we perfom univariate , bivariate , multivariate analysis & try to understand the relationship of features. The intent of the EDA is to determine whether a predictive model is a viable analytical tool for a particular business problem, and if so, which type of modeling is most appropriate.

The deliverable is a low-risk, low-cost comprehensive report of findings of the univariate data and recommendations about how the company should use additional modeling.

At the very least, the EDA may reveal aspects of your company's performance that others may not have seen.

4. Model Building:

To predict, for given set of information related to a product, whether a product could become a backorder for the e-commerce platform or not, a Machine Learning model is used. Since there are 2 outcomes, we are concerned with i.e if a product is a backorder or not, the Supervised Classification Machine Learning algorithms are used.

There were several Classification models that were tested such as Logistic Regression, Decision Tree and Ensemble techniques such as Bagging (Random Forest) & Boosting (Gradient Boosting).

From the several baseline models built, Random Forest was chosen and taken ahead for feature selection & Hyperparameter tuning. The final model built was a baseline Random Forest model with 10 features.

5. UI- Development:

The user interface i.e the web application, designed to enable interaction of the user with the model has been built using HTML, CSS, JavaScript & Flask The user interface contains the following

• Landing Page: This is the first page that is visible to the user on starting the application. Consists of navigation pane and buttons to allow users to scroll to desired sections.

• Backorders Description Section: This is the 1st section on the web page, which provides important information to users about backorders and its significance/ risks posed to e-commerce platforms.

• Prediction Section: This is the 2nd section, with a form containing several fields that need to be filled by the users according to the questions asked. On successfully filling all information, the user can hit the submit button which will enable to form to process the information provided and pass it to the model

. • Results Page: This is a 2nd and final page of the application where the result generated by the model is displayed to the users.

How the following technologies were used to build the interface are given below: HTML – The front-end interface for the application i.e webpages, have been designed using Hyper Text Markup Language. HTML forms have been used to create a form that consists of inputs and options for the user to pass information about the product, using which the model would produce the required output CSS – The webpages for the application have been designed using Cascading Style Sheets. In order to facilitate scrolling across the web application to different sections and provide a navigation pane for the users JavaScript has been used.

6. Deployment:

The model has been deployed on the web server through Heroku. This is accessible to everyone by using the URL for the web application.