**Forecasting Application for Inventory Management**

**Siddhi Kate**

2982006

Submitted in partial fulfillment for the degree of

Master of Big Data Management and Analytics

Griffith College Dublin

September, 2019

Under the supervision of Prof. Aqeel Kazmi

**Disclaimer**

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the Degree of Master of Science Big Data Management and Analytics at Griffith College Dublin, is entirely my own work and has not been submitted for assessment for an academic purpose at this or any other academic institution other than in partial fulfilment of the requirements of that stated above.

**Signed: Date:**

Acknowledgements

The completion of this project would not be achieved without the support and guidance of many people. First and foremost, I would like to thank Prof. Aqeel Kazmi for being my mentor and assisting me to achieve my goals in this project.

I would also like to thank my family and friends without whose encouragement and support , it would have been impossible to reach the finish line. Finally, I would like to thank the entire Big Data Management and Analytics faculty at Griffith College Dublin , for helping me to move towards my career in the right direction. Thank you.

**Table of Contents**

[Acknowledgements iii](#_Toc297629416)

[List of Equations v](#_Toc297629417)

[List of Figures v](#_Toc297629418)i

[List of Tables v](#_Toc297629419)ii

[Abstract vi](#_Toc297629420)ii

[Chapter 1. Introduction 1](#_Toc297629421)

[1.1 The Field of Inventory Management 1](#_Toc297629422)

[1.2 Aims of the Project 2](#_Toc297629423)

[1.3 Overview of Approach 3](#_Toc297629424)

[1.4 Document Structure 4](#_Toc297629425)

[Chapter 2. Background 5](#_Toc297629426)

[2.1 Literature Review 5](#_Toc297629427)

[2.2 Related Work 8](#_Toc297629428)

[Chapter 3. Methodology 11](#_Toc297629429)

[Chapter 4. System Design and Specifications 27](#_Toc297629430)

[Chapter 5. Implementation 29](#_Toc297629431)

[Chapter 6. Testing and Evaluation 41](#_Toc297629432)

[Chapter 7. Conclusions and Future Work 43](#_Toc297629433)

[References 45](#_Toc297629434)

# List of Equations

Equation 5.1 SMA Equation…………………………………………………………32

Equation 5.2 ARIMA Equation………………………………………………………33

Equation 5.3 Holt Winters Equation…………………………………………………34

Equation 5.4 LSTM Model Equation………………………………………………...35

# List of Figures

Figure 3.1 Gantt Chart of Project Schedule………………………………………….17

Figure 3.2 Comparison of Buy Price and Sell Price………………………………....18

Figure 3.3 Comparison of Stock on order and Stock on Hand………………………19

Figure 3.4 Comparison of Supplier Lead times……………………………………...20

Figure 3.5 Quantity of products ordered over time……………………………….....21

Figure 3.6 Plotting the difference between order and delivery dates………………..22

Figure 3.7 Prediction of Buy Price…………………………………………………..23

Figure 3.8 Prediction of Sell Prices…………………………………………………24

Figure 3.9 Comparison of Quantity of products ordered and supplier lead time…….25

Figure 5.1 Forecasts from SMA model……………………………………………...32

Figure 5.2 Forecasts from ARIMA model…………………………………………...33

Figure 5.3 Forecasts from Holt Winters Model……………………………………...34

Figure 5.4 Results of LSTM model…………………………………………………..35

Figure 5.5 Forecasts from KNN model………………………………………………36

# List of Tables

Table 5.1 Accuracy Table………………………………………………………..38

Table 6.1 Predictions of Lead time………………………………………………41

Table 6.2 Predictions of Price……………………………………………………41

Table 6.3 Predictions of Demand…………………………………………………42

# Abstract

Time is an important factor in driving businesses. Many companies sell different types of products, that are supplied to them by different suppliers. It is a necessity that all products , are delivered in time and in required amount. Along with this, one has to keep a regular check on the buy prices of these products, profit margins as well as monitor the inventory to avoid shortage as well as wastage. Inventory management can be a tough task if not carried out with proper insights. Delivery time, demand of products and buy price of products play an important role in inventory management.

In this particular project, a profile of company that has four different suppliers is being studied. Product level analysis is done with respect to attributes like, number of days required to deliver particular product, cost of the total shipment as well as demand of the product. The main aim of this project is to study historical data available, observe patterns and then predict what the company should expect for their next shipment. It focuses on time series forecasting , which is achieved by implementing various different algorithms available at hand. An effort is made to provide accurate predictions for the company, based on which they can make informed decisions.

# Chapter 1. Introduction

## 1.1 The Field of Inventory Management

All big retail companies have their inventories that are major part of their business. However, maintaining and managing these inventories can be a major task. Sometimes, there are situations where a company faces shortage due to unavailability of products, which can lead to company’s loss and also sometimes the company may order excessive goods which can also lead to the same.

Achieving a balance between the above two is necessary, so as to ensure that the company has greater profit margins and minimal loss.. Initially, companies used to maintain excel sheets to keep a track of their products and manage their inventory. A fixed amount of goods would be stocked and the unused would end up in waste. A powerful tool such as data analytics is now being used for managing the inventory.

Using data analysis techniques such as prediction, time series forecasting, machine learning a company can keep their inventories up to date and also increase the profit margin. Using such techniques, one can find hidden insights which can be used for actionable intelligence. For example :

1. In food industry, studying the market and finding out a trending food item, companies can stock up that and initially offer it at slightly lower price than other companies and increase profits.eg. the current trend on including **‘Avacados’** in the diet.
2. In technology, finding out the maximum selling electronic device based on sales report, stocking up that and eliminating the one which does not sell at all.
3. Weather forecast can also play an important part in inventory management when food has to be transported over large distances. Choosing the right day and avoiding food wastage due to temperature fluctuations can save companies from unseen losses.

## 1.2 Aims of the project

**Motivation :**

The motivation of this project rises from personal experience. During the course of my study, I also worked at a coffee shop as a part-time team member. There would be days, where there was extreme shortage and sometimes extreme wastage.

Data analytics is a powerful tool which if employed correctly can help businesses save a lot of money. Hence I decided to take up a project in the field of inventory management and deploy all the knowledge that I have gained so far to improve the system.

**Aims:**

The aims of the project are as follows :

* To use predictive analysis, for finding out the demand of the product.
* Monitoring inventory levels of various products and examining trends.
* Based on trend observations, gain insights of the data and find out variables that are related to each other and can be used for predictions.
* Develop an application, that has an easy to use user interface and requires minimum maintenance.

**Objectives :**

The objectives of the project are as follows :

* Study different algorithms that can be used for the purpose of the project and finalize one based on accuracy.
* Decide a web application framework that can be easy to use and convenient.
* Study the company profile , and make decision about what strategies can be used to improve the overall company performance.
* Based on gained insights, decide on actionable intelligence.

## Overview of Approach

The steps taken to design and implement this project are as follows :

1. **Choosing the dataset**

The first and foremost step was to find a dataset for the required purpose of the project. The dataset used is a sample dataset, and has many interesting features that will be explained in the later part of the chapter.

1. **Exploratory data Analysis**

This step was achieved using R and Tableau interface, which involved cleaning the data, examining various variables etc and analytics.

1. **Deciding the strategy used for prediction**

By studying various data analytics techniques available at hand, using time series forecasting was finalized.

1. **Experimenting with algorithms**

Developing various algorithms available at hand for time series forecasting and finalizing one based on accuracy.

1. **Developing the Application**

The Shiny Application Framework that is compatible with R studio , is used to develop the application.

1. **Testing and Evaluation**

Testing the performance of the application by giving different inputs and recording the accuracy results.

## Document Structure

The layout of this document is divided into various chapters, they are as follows :

1. **Chapter 1 : Introduction**

This includes a brief introduction to my project , my field of research and the steps taken to achieve it.

1. **Chapter 2 : Background study and Literature Review**

This includes all the sources and the references studied to accomplish the project.

1. **Chapter 3 : Methodology**

This provides, the plan and timeline of my project, and the design specification of my project.

1. **Chapter 4 : System Design and Specifications**

This includes description of technologies and frameworks used.

1. **Chapter 5 : Implementation Details**

This included describing the algorithms used, the results etc and also deployment of the application.

1. **Chapter 6 : Testing and Evaluation**

This includes results from testing and evaluation of my project and an explanation of how it is achieving the required accuracy.

1. **Chapter 7 : Conclusion and Future Work**

This includes, what the project has achieved and how can it be extended to achieve bigger goals in the future.

# Chapter 2. Background

## 2.1 Literature Review

**Understanding important terms related to Inventory Management :**

**Inventory forecasting :**

Inventory forecasting is the process of predicting when and how much you’ll need to order. The future prediction depends on various factors like previous sales history, trends in demand for different products and the average lead time of suppliers to deliver products [1]. The combination of everything specified above, followed by their detailed study , helps in making an accurate decision about inventory forecasting.

**Demand forecasting :**

Demand forecasting is the process of making estimations of the value of the product in the market in future. [2] An accurate demand forecasting technique can give a company an insight about their potential in the future market. It can help managers to plan inventory, design pricing strategies and re-ordering inventory.

**Supplier Lead time :**

Supplier lead time can be defined as , the average time taken buy a supplier to deliver products one an order has been placed buy a company [3]. Every supplier has a different lead time. Factors affecting lead time include distance, transportation costs, quantity of products ordered etc. Even though it is a simple term, it can be used to gain actionable insights, if studied properly. Accurate forecasting of this, can help managers to place orders in such a way that they can rarely run out of stock as well take measures so that they do not order extra.

**Stock on hand :**

This refers to the quantity of a particular product available in the inventory of the company.

**Stock on order :**

This refers to the quantity of the a particular product , after the order has been received and the product has been stocked up.

**Need for forecasting in Inventory Management :**

Accurately predicting availability of products can lead to adequate stocking of inventory , and ensures that there will be no problem in sales. However, there can be a problem even if the supplier delivers products before time as well after time [4]. Relatively small variations in demand can also lead to a huge effect on the profit and loss margins of the company over a larger period of time [5]. The necessity to minimize unsold stock products can always lead to complications in the process of re-ordering and re-stocking [6]. Inventory management also hugely depends on the lifecycle of products. Planning for product with short life cycle of weeks cannot be same as the one used for mass market products [7]. The company can face huge losses over time, if the inventory is not managed properly. Hence, there is a need to use historical data available for forecasting.

Businesses have to be always be one step ahead when it comes to satisfying customer demand. Forecasting is important for businesses on different levels. It not only can play an important role in planning, maintaining finance as well as some corporate activities but can also be an important factor for marketing strategies [8]. Forecasting can always be put to good use in this fields. With an efficient technique for forecasting companies can efficiently deal and cope with flow of products, warehousing, maintaining and depletion of stock etc. It can help companies achieve high levels of customer service standard [9].A company that uses inventory forecasting can minimise the risk of bottlenecks in production, maintain the investment value required to minimize loss, use a proactive strategy , and maintain a balance between shortage and wastage [10]. Hence forecasting plays an important role in Inventory Management.

**Techniques that are being used for Inventory Forecasting :**

A range of different forecasting techniques are available for inventory management. These techniques can be chosen based on data patterns, seasonality of data , available attributes etc. Some of the techniques that are currently being used are as follows :

1. Rule Based Forecasting

This technique can use casual knowledge in structured format. It has set of 28 series features, one hast to identify what series in the dataset in order to apply RBF. There are a set of 99 rules for combining the forecasts [11].

1. Artificial Neural Networks

These are non linear models. These models being complex can also give a better accuracy and performance as compared to other linear models such as regression [12]

1. Statistical Models

There are different types of model such as moving average, holt Winters, Arima model etc. The widely used model is the Arima Model [13]

1. Support Vector Machines

It has the ability to solve non-linear regression estimation problems and hence can outperform ANN. It gives good performance for time series forecasting[14]

The above mentioned techniques are the ones that are widely used currently for forecasting purposes. However each of this technique has its own set of advantages and disadvantages . Everything largely depends on the data that is available at hand to forecast. Most of the simple techniques are available for forecasting smaller datasets whereas complex model can give great performance against large datasets.

**2.2 Background Study**

**Case Studies :**

1. Using the ABC analysis for Inventory Management [15]

An ABC classification or also known as ABC analysis , is a technique that is used by inventory managers. This process is performed at the starting point of inventory control. The method is implemented by setting priorities for management time and the financial aspects.

The base of this analysis technique is Pareto Analysis. The rule of Pareto Analysis is stated as follows:

The 20 % items of the inventory contribute to 80 % of sales. That means the minimum number of product generate maximum sales of the business.

The products are divided into 3 categories as follows :

1. Class A : They consist of 20 % of the inventory and contribute 80 % towards the sales or the revenue.
2. Class B : They consist of 30 % of the inventory and contribute 15 % towards the sales or the revenue.
3. Class c : They consist of 50 % of the inventory and contribute 5 % towards the sales or the revenue.

Class A items are the most critical items. They require high level analysis, inventory control , regular price checks as well constant count of the products.

Class B items are somewhat critical as compared to class A . They require nominal control techniques and a few checks on stock.

Class C items require minimum attention and control as they have least effect on the data stocking and financial status of the company.

This is a popular technique and is used by many businesses. However this is a primitive technique and can be replaced by powerful data analytics.

1. A Data Mining Approach Used by a Pharmaceutical Company [16]

This case study is based on pharmaceutical company in Iran. The industry generally uses a wholesale approach. The sellers buy the stock of their own from manufacturers and sell to big customers such as hospitals, large medical stores etc.

To predict the sales forecast, time series forecasting was carried out. Different algorithms were implemented. The techniques under focus were :

1. ARIMA
2. Hybrid neural networks using the past record of each medicine
3. Hybrid neural networks using the past record of each medicine and its co members.

The above three models were implemented and the accuracy was tested. The results of the system were that, the second model performed better than the first and the third model

**Approach of this project :**

After researching and studying the field of Inventory Management, It has come to attention that forecasting plays an important role in it . The field is vast and there are many approaches that can be taken towards it. One of the main approach is the time series forecasting.

Time is a driving factor for businesses and it is essential to always plan ahead of time. Major variables on which the profit of a particular company depends is the supplier lead time, demand of product in the market, stock levels as well as buy price and sell price of products. There is a need that all these values are monitored correctly.

Various techniques are available at hand and each of them has its own specific advantage. During this project multiple techniques will be developed and a final technique will be chosen on how it handles the data , what ate the predictions results, accuracy as well as time in which the results are returned.

The approach of this project is different in such a away that it will suit small businesses. Such businesses cannot afford commercial inventory management software that are widely available. Also , there data is not that huge which would require such huge systems to deploy.

The cost of development of this project is minimal as it does not require any hardware specifications . Hence , small case businesses will benefit from this project.

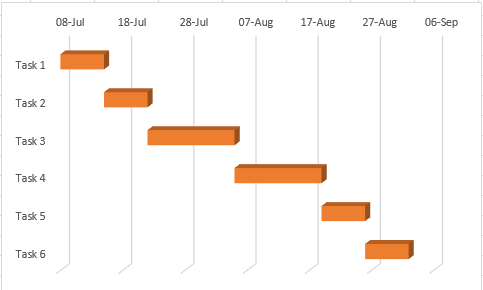
This application will be developed so as to give insights to decision makers within a click, help them to view the trends of their data that has been in the past few years, make current decisions on predictions available and pave a path for the overall development of the business.

# Chapter 3. Methodology

**Project Schedule :**

The following gantt chart explains the schedule of the project. The project was divided into small goals, each goal was set to be completed in a week, so that the final goal of completion of project could be achieved over time.

Figure 3.1 : Gantt Chart of Project Schedule



Task 1 : Research of project idea

Task 2 : Selection of dataset and exploratory data analysis

Task 3 : Implementing Algorithms

Task 4 : Implementing Application

Task 5 : Testing, Evaluation and Documentation

Task 6 : Testing, Evaluation and Documentation

**Dataset Overview :**

The dataset chosen for this project, is provided by The Lokad Technical Documentation [2]. Lokad is a quantitative supply chain firm, that provides high level solutions for supply chain management.

The dataset is a collection of historical data of a retailer. The data is provided as tabular files which are formatted as flat text files.

The sample dataset contains the following files :

* Lokad\_Items.tsv : Contains the list of products sold by the retailer.
* Lokad\_Orders.tsv : Contains a list of historical client orders
* Lokad\_PurchaseOrders.tsv : Contains a list of orders made from the suppliers.
* Lokad\_Suppliers.tsv : Contains a list of suppliers and their description.
* Lokad\_BOM.tsv : Contains the bills of material for bundles.

This dataset has various different variables. Each of this variable can be used as an interesting piece of information. Examples of such variables are product categories, stock levels, lead times, buy and sell prices etc. The data is similar to a data that can be obtained from a company that uses not-too-old version of ERP.

**Dataset Description :**

In this section, each an every field of the dataset is explained with a one-line description.

1. Lokad\_items.tsv

This part of the dataset contains a list of products that are sold by the retailer whose profile is under study.

* Id : A unique code of every product also can be called as an SKU.
* Name : The name of the product as appears on the lable.
* Category : This is a categorical attribute that has 3 values , namely :

1. Hardware
2. Software
3. Accessories

* Subcategory : Also a categorical attribute, but has many different values such as speakers, headsets etc.
* Brand : The name of the brand which sells this product.
* Colour Code : Each product has a specific colour that has a set value.
* Supplier : This retailer has four different suppliers, namely :

1. Computer & Co
2. Digital Age
3. Office Supplies
4. Techno First

* Buy price : The price at which a particular product is purchased from the supplier.
* Sell Price : The price at which the product is sold after purchasing it.
* Sell Currency : The currency in which the product is sold.
* Supplier Lead time : This is a default value expressed in number of days.
* Stock on Hand : The number of products that are currently available in the inventory
* Stock on Order : The number of products that are ordered from the supplier but not yet received.

1. Lokad\_Orders.tsv

This file contains transactional level data. It includes information about historical orders of products from customers.

* Id : It is the same value as described in the dataset above. It can be used as a foreign key for connecting the datasets.
* Date : It specifies a date on which the product was ordered by the customer.
* Quantity : Number of units ordered by the customer.
* Currency : The currency code in 3 letters in which the payment was made.
* Net Amount : A amount paid for the entire shipment , without taxes.

1. Lokad\_PurchaseOrders.tsv

This file also contains transactional level data of every shipment from the supplier.

* Id : A foreign key field available in above datasets.
* Date : The date on which the order of the product was placed.
* Delivery date : The date on which the products were delivered.
* Quantity : Number of units ordered.
* Currency : The currency used for payment.
* Supplier : Supplier from whom the product was ordered.
* Net Amount : The total amount paid to the supplier for the shipment.

**Exploratory Data Analysis :**

Technology used : **Tableau**

Reasons for using Tableau :

1. It is fast, helps to create data visualization quickly.
2. Easy to implement
3. It can be connected to any file formats of the data
4. The size of the data does not matter, it handles small as well huge data with great effectiveness.
5. The drag and drop feature makes it easy to work with.

**Data Analysis Report :**

After studying the chosen dataset, various variables were explored for their relationship with each other. The company has 4 major suppliers, from whom it orders its products. There is particular buy price and sell price for every product.

**Company Profile :**

The company that is being studied for this particular project , sells various types of electronics and the side products associated with that. The products sold are broadly classified into three major categories , namely :

* Hardware
* Software
* Accessories

This company gets product delivery from four different suppliers. Each supplier delivers a set of products. Following are the companies and the products that they deliver.

**Company A : Computer & Co :**

The list of products delivered buy this company is as follows :

1. Speaker
2. Cover
3. Charger
4. Adapter

**Company B : Digital Age :**

The list of products delivered by this company is as follows :

1. Table
2. Stand
3. Projector
4. Gaming
5. Keyboard
6. Educational
7. Desktop
8. Connector
9. Case
10. Accounting

**Company C : Office Supplies**

The list of products delivered by this company include :

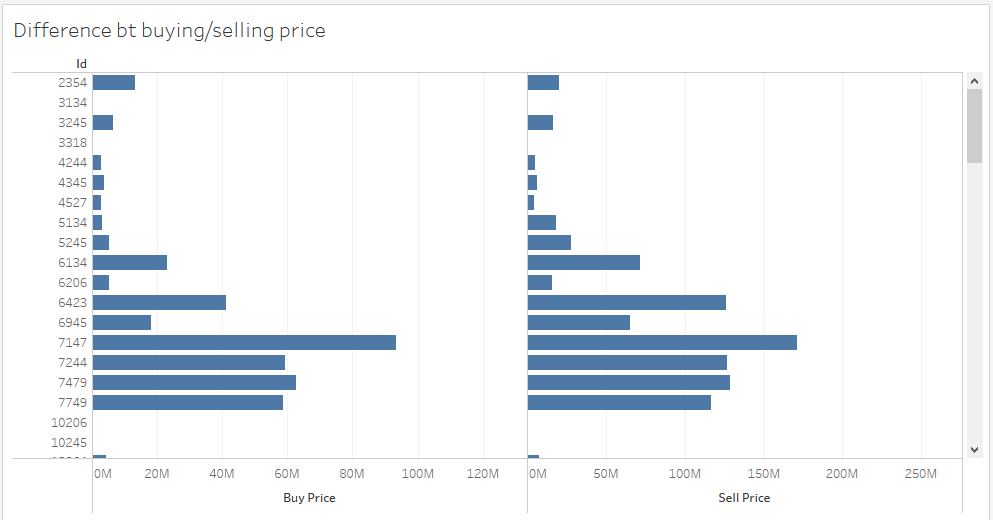
1. Headset
2. Editing

**Company D : Techno First**

The list of products delivered buy this company :

1. Speaker
2. Office Software

**Graph 1 :**



**Figure 3.2 : Comparison of Buying and Selling Price**

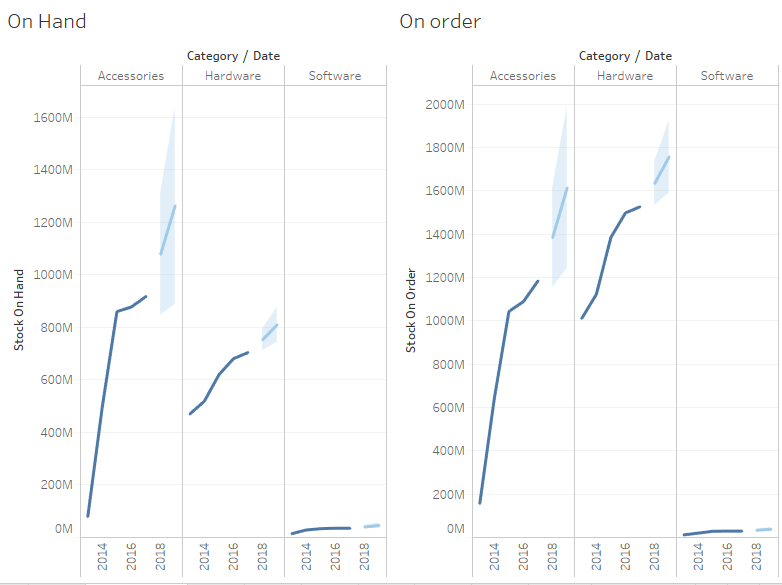
**Insight :**

This graph compares the buying and selling price of different products by taken into consideration individual product ids.

Even though the graphs appear similar, if seen carefully on the scale of the graph, the selling price ranges from 50 to 200 million, whereas the buying price ranges from 20 to 80 million.

From this graph , it can be said that the company is usually making profits with very rare loss cases.

**Graph 2 :**



**Figure 3.3 Comparison of Stock on Order and Stock on Hand**

**Description :**

The company has 3 major categories of product, namely accessories, hardware and software.

Stock in hand defines what is the current amount of products in the inventory and stock on order, determines what is the resultant quantity of the products after stocking.

The graph shows how these numbers vary over the period from 2014-2018.

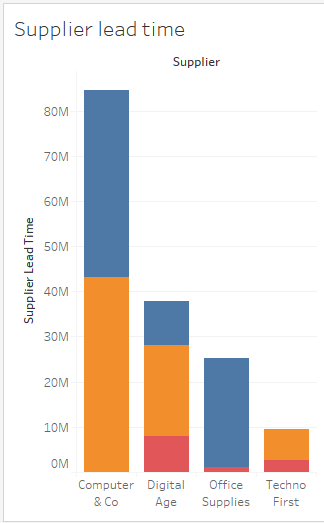
**Insight :**

There is a steep linear rise in the accessories stock, this means that the company has sold a lot more accessories over this period.

There is a steady increase in the amount of hardware ordered whereas software ordered is very low throughout the year.

When a forecast is performed, the stock is shown to increase over the years as there was no drop in the stocks for past 4 years.

**Graph 3 :**



**Figure 3.4 Companies and their Supplier Lead time**

**Description :**

In this graph you can see the 4 different suppliers that this company has.

Supplier lead time is the difference between the date that the product is ordered and the date the product is actually delivered by the company.

The colour code indicates different categories of the product.

The company takes maximum number of deliveries from Computer & Co and least from TechnoFirst.

**Insight :**

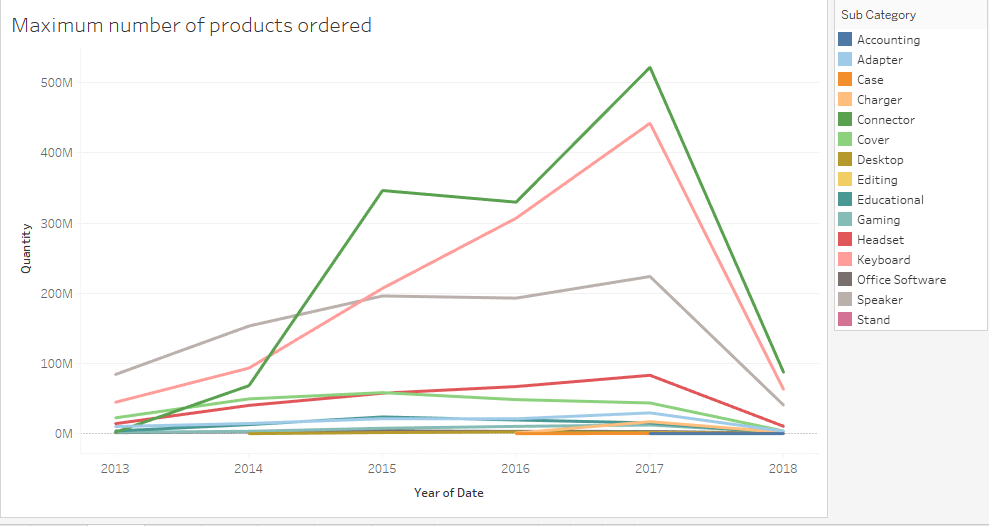
More the supplier lead time more is delay.

Based on this graph, we can conclude which company delivers products in the least amount of time.

But the fact that amount of products ordered may also affect the supplier lead time .

That is if more the number of products more is the supplier lead time.

**Graph 4 :**



**Figure 3.5 Quantity of products ordered over time**

**Description :**

This graph is plotted between the type of the product ordered, and its quantity over the period between 2013 to 2018.

Generally the more number of products that are ordered are accessories rather than hardware and software.

**Insight :**

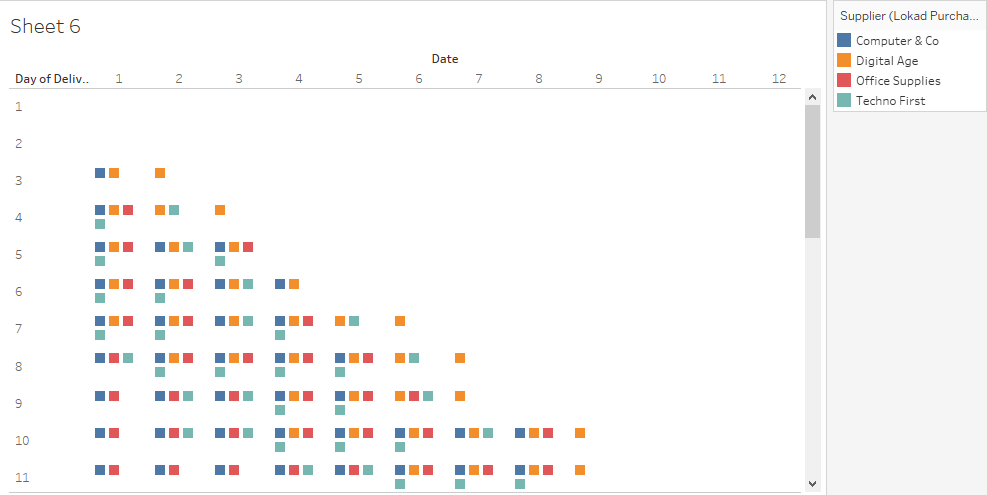
A steep rise in the quantity of products is seen in 2017.

The amount of products steeply increases till 2017, and then there is a sudden fall after that.

The maximum amount of products that are ordered are connectors.

The least number of products that are ordered are desktops.

**Graph 5 :**



**Figure 3.6 Plotting the difference between Order and Delivery Dates**

**Description :**

In this graph, we track the date of ordering the product against the date the product is actually delivered.

The colour code is to find out what different companies take different days to deliver their products.

**Insight :**

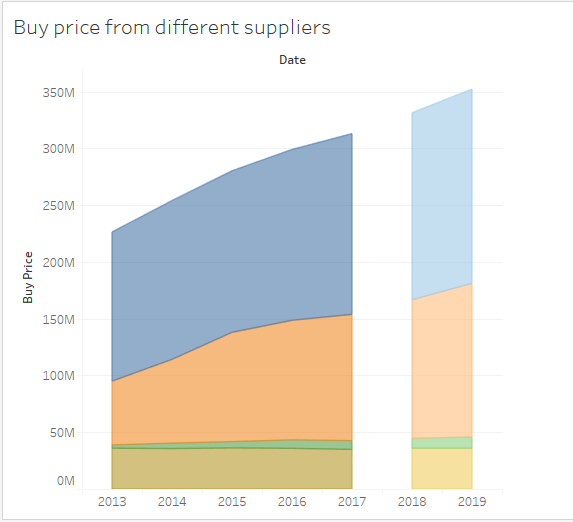
There are very few companies that deliver the products within 2-3 days of ordering.

Out of the 4 companies, we can see from the graph that Digital Age gives the best delivery where the time factor is considered.

Computer & Co and Office Supplies have a larger gap between there ordering date and delivery date

Even though minimal products are ordered from Techno First , they provide a average good supplier lead time.

**Graph 6 :**



**Figure 3.7 : Prediction of Buy Price**

**Description :**

In this graph, we track the buy price of different products from different suppliers.

We also forecast what will it the buy price be in the coming year.

**Insight :**

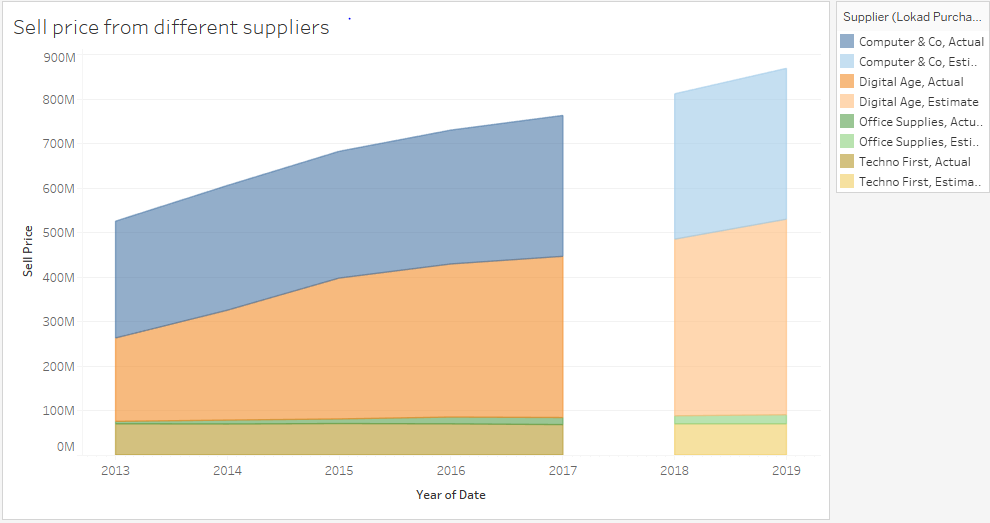
In the above graph we can see that, an exponential increase takes place in the buy price of Computer & Co and Digital Age.

In the future forecasts it is also shown to keep increasing.

The buy prices from Office Supplies has minutely increased.

Unnaturally , the buy prices of products remain same throughout the entire period and have predicted to remain the same for the next year.

**Graph 7 :**



**Figure 3.8 : Prediction od Sell Price**

**Description :**

In this graph, we track the sell price of different products from different suppliers.

We also forecast what will it the buy price be in the coming year.

**Insight :**

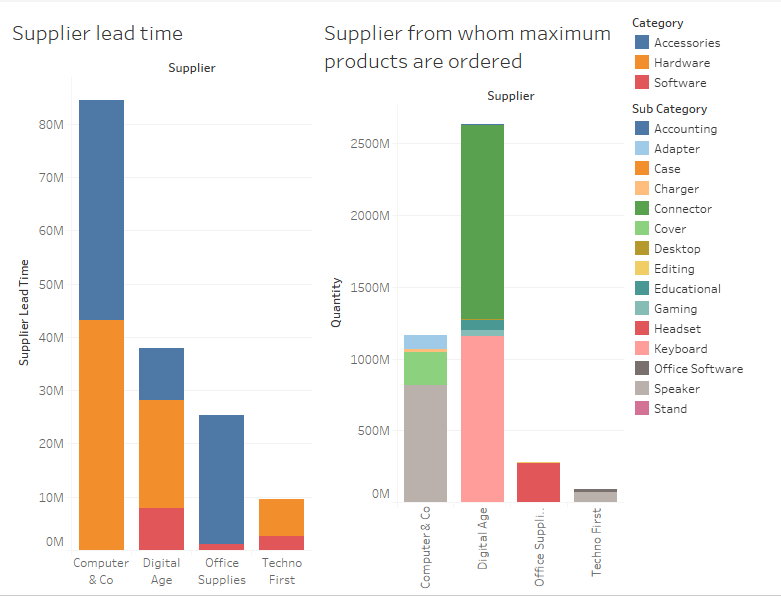
In the above graph we can see that, an exponential increase takes place in the sell price of Computer & Co and Digital Age.

In the future forecasts it is also shown to keep increasing.

The sell prices from Office Supplies has minutely increased.

Unnaturally , the sell prices of products remain same throughout the entire period and have predicted to remain the same for the next year.

**Graph 8 :**



**Figure 3.9 Comparison of Quantity of products and Supplier Lead time**

**Description :**

In this particular graph, we compare 2 different graphs.

One graph is the supplier lead time of every company.

And other graph is the quantity and sub category of products that are ordered from different companies.

**Insight :**

From the above graph , we can see that maximum number of products are ordered from Digital Age.

Least number of products are ordered from Techno First.

Even though maximum number of products are ordered from Digital Age, their supplier lead time is comparatively less to Computer and Co.

Even though not much products are ordered, the have a very large supply time.

**Results of Data Analysis :**

After doing the necessary analytics, the following important decisions regarding the project were made :

1. Out of all the subparts of the dataset provided, the dataset with itemset description i.e. Lokad\_items.tsv and the dataset with description of orders from suppliers i.e. Lokad\_PurchaseOrders.tsv are the most important .
2. These two datasets can be joined by using the ‘ID’ field as the foreign key.
3. Out of all the attributes , the following is the list of attributes on which the entire project will be built on :

* Order date
* Delivery date
* Supplier Name
* Product Category
* Product Subcategory
* Quantity
* Buy price
* Sell price

All the other unnecessary attributes will be removed from the dataset.

**Choice of Programming Language**

R and Python are two major programming languages that can be used for data analytics. In this project I have chosen R over Python and the reasons are as follows :

* R has many different inbuilt libraries and corresponding functions that can be used directly rather than writing complex codes.
* R is easier to code as compared to python( personal view point ).
* R has advanced graphical and visual capabilities , for example plots.
* R is better than statistical analysis for python.

**Choice of application development framework**

In this project I have used the Shiny Development Framework and the reasons are as follows :

* It is completely compatible with R, no other tools need to be installed to run the code.
* It can be used to develop fully interactive dashboards.
* Shiny produces a web application that can be deployed centrally and can be shared as a URL like web pages.
* It is similar to HTML, and if one knows HTML basics it can be easy to takeover it.

# Chapter 4. System Design and Specifications

R studio and the libraries used to developed the system :

1. The version of R studio used is 3.6.1 . Release Date : 2019-07-05
2. This is the latest version of R studio, and is used so that all the new libraries that are available with this version can be loaded and utilized as per requirement
3. The libraries used are as follows :

* lubridate()

It is library that deals with date-time objects and allows easy manipulation of the same. The syntax is consistent and memorable and hence it could be understood and implemented easily

* dplyr()

This library provides easy manipulation of the data. The reason of using this was that the function are in form of verbs with their literal meanings which makes coding easier . For example , summarise(), filter(), arrange() etc.

* forecast()

Most important library. Forecasting models such as SMA, ARIMA etc were built using the functions available under this library.

* zoo() and xts()

Both these libraries make working with time series easy. They were used to convert the data into time series format correctly.

* tsfknn()

This particular library was used for time series forecasting using the KNN model. It has all the necessary functions that can be programmed to get accurate predictions.

* keras()

This particular library was used to develop neural network model. It made developing the deep learning model easier.

* tensorflow()

This is the library that is used developing machine learning algorithms. The Estimator API of this library helps to build high level models for classification and regression.

* shiny()

It is web development framework provided by R studio. Easy to understand code, similar to HTML and many built in functions that helped in developing a interactive web app.

* shinydashboard()

This feature of shiny helped in designing interactive dashboards. The results appeared on a single page and on a single click.

# Chapter 5. Implementation

**Dataset Cleaning :**

The dataset understudy did not include any missing values, hence it did not require any pre-processing.

**Preparing the data :**

As mentioned earlier in the Methodology chapter, **Lokad\_items.tsv** and **Lokad\_PurchaseOrders.tsv** are the datasets that will be considered for implementing this project, a **‘merge’** operation is performed so as to combine the two.

* Both the datasets are combined using the **‘ID’** field which is common to the dataset.
* Lokad\_items.tsv has 123 rows of 13 variables.
* Lokad\_PurchaseOrders.tsv has 15681 rows of 7 variables.
* The final dataset has 15681 rows of 19 variables.

**Calculating the lead time :**

Lead time is the number of delivery days required to deliver the particular product. However this field is not available in the dataset. The fields that are available to us are **Order Date** and **Delivery Date**. The lead time is calculated by introducing a new field called **‘diff’**. This field is obtained by subtracting the Order Date from the Delivery Date.

Use of the **as.Date()** function in R

* It is necessary to tell R to treat a Date attribute as a date attribute otherwise it gives an error. Hence the **as.Date()** function is used while subtracting the two dates.
* After the above step, it is necessary to convert the obtained result to a numeric attribute so as to achieve the required format. i.e. Number of days.
* Hence the **as.numeric()** function is used to convert it into numeric format.

**Calculating the profit :**

As similar to above, the profit field is not available. But, what is given is **Buy Price** and **Sell Price** of the products. By subtracting the buy price from the sell price, the profit field is calculated. It has a positive value if there is profit and a negative value if there is loss.

**Creating data subsets as per requirement :**

The project focuses on time series forecasting of different attributes against time. Hence necessary subsets are created so that they can be easily used for further implementation.

1. **Lead time dataset** : This dataset is creating a subset of the final dataset by including the following attributes :

* Date
* Supplier
* Sub Category
* Diff

1. **Profit dataset** : This dataset is creating a subset of the final dataset by including the following attributes :

* Date
* Supplier
* Sub Category
* Buy Price
* Sell Price
* Profit

1. **Demand dataset** : This dataset is creating a subset of the final dataset by including the following attributes :

* Date
* Supplier
* Sub Category
* Quantity

**Converting the obtain datasets to a time series object :**

For doing time series analysis it is necessary that the data available is in proper time series format. This is because all the libraries and functions in R, that are available for time series formatting run only if the data is in correct format. The following steps were taken to convert the dataset into proper times series format :

* Selecting the Date and the attribute to forecast.
* Ordering the data by Date.
* Converting the data to an ‘xts’ object.
* Calculating the start and end positions of the date.
* Converting the xts object to ‘ts’ object by using the ts function.

**Required libraries :**

1. **zoo**

This library in R is used to handle irregular time series.

1. **xts :** Extensible time series

This is an extension to the zoo library. It helps in making the time series fast and reliable.

1. **ts :** Time series

It is used to create time series objects.

**Experimenting with different algorithms :**

The entire idea of this project is based on time series forecasting. It is necessary that there is strong underlying algorithm that can be a strong base and give accurate results. The following algorithms were implemented and tested for their accuracy :

1. **Simple Moving Average (SMA)**

A simple moving average is a simple arithmetic forecasting technique available in R. The past data is taken into consideration, previous averages are calculated, added and then divided by the number of time periods. Moving averages are widely used for stock forecasts [3].

The equation of SMA is as follows

SMA = A1 + A2 + ……+An / n ………………………………………..(5.1)

Where :

An = the value of an asset at period n

n = number of time periods

Graph and results obtained from SMA :

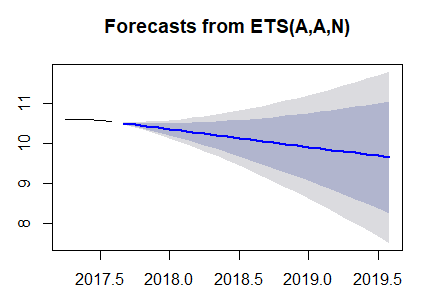


Figure 5.1 : Forecasts from SMA model

1. **ARIMA Forecasting**

Auto Regressive Integrated Moving model is used for forecasting a stationary time series. It is an advanced model which can be used to forecast different trends in data such seasonal, non-seasonal and irregular [4].

**auto.arima() function in R**

The function helps to generate a set of best values that can be used as an input while developing the Arima model by taking the univariate time series as an input.

Inputs :

In Arima(p, d, q) model :

p = number of autoregressive terms

d = number pf non seasonal differences needed for stationarity

q = number of lagged forecast errors in prediction

For this model, the values given as input were, (1,1,0) respectively. This set of values is also known as differenced first order auto regressive model . Equations are as follows :

Ŷt - Yt-1 =  μ  +  ϕ1(Yt-1- Yt-2)

Ŷt  - Yt-1=  μ

Which can be rearranged to

Ŷt  =  μ  + Yt-1  +  ϕ1 (Yt-1- Yt-2)…………………………….(5.2)

Forecasting graph obtained :

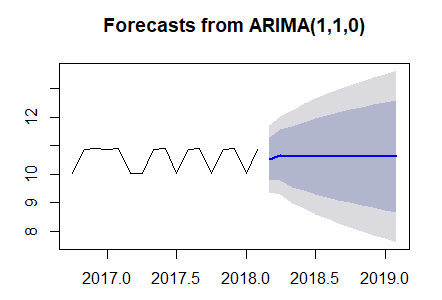


Figure 5.2 : Forecasts from ARIMA model

1. **Holt Winters Forecasting Model**

The Holt Winters model is one of the best models to forecast seasonal trends in the data. The algorithm performs best when data has seasonality. The model is suitable for daily, monthly as well as yearly data. [5] . The equation of Holt Winters Model is as follows :

Forecast equation : y^t+h|tℓtbt=ℓt+hbt………………………………………………………………..(5.3)

Where l^t = estimate of the time series at time t

b^t = estimate of the trend at time t

Graph and results obtained from Holt Winters method :

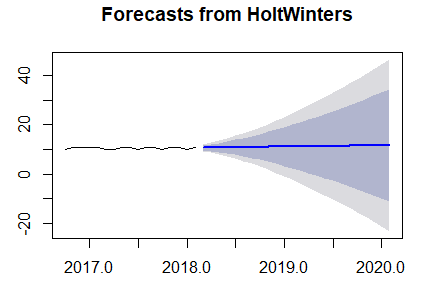
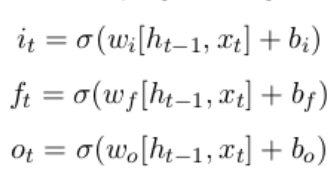
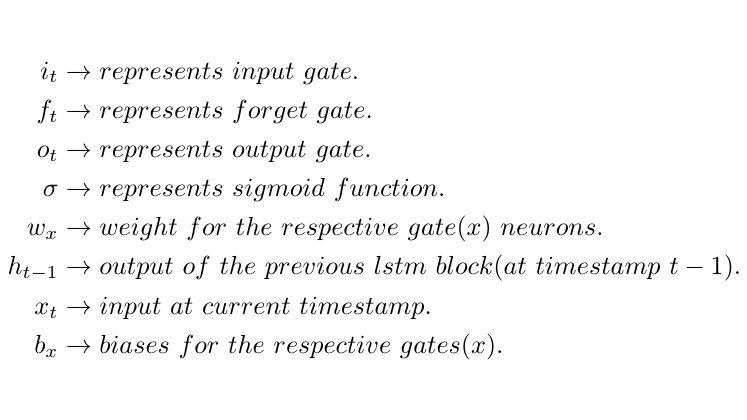


Figure 5.3 Forecasts from Holt Winters Model

1. **LSTM model**

A Long Short-Term memory model is a deep learning model. As compared to other neural networks, LSTM is different in a way that it has feedback connections. Hence, it is also known as a recurrent neural network. [6]. The equation of LSTM model is as follows :

………………………5.4



Inputs given to the model :

No of layers = 1

Batch size = 1

Learning rate = 0.02

Number of Epochs = 5

Decay rate = 1e – 6

Output of model :

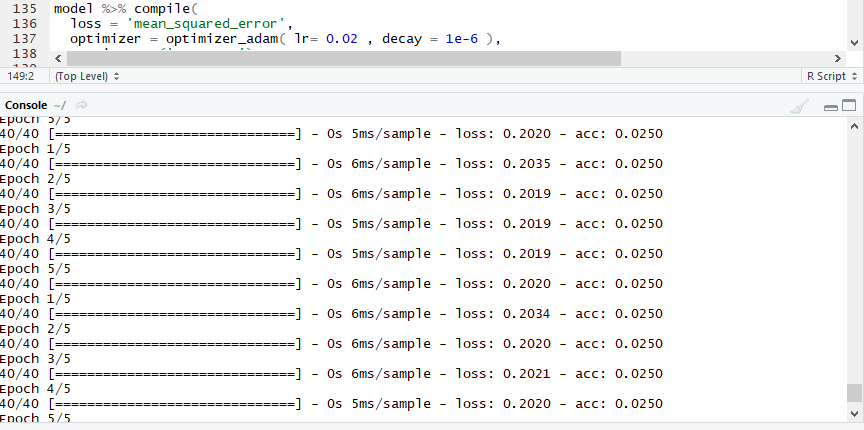


Figure 5.4 : Results of LSTM model

loss : Root Mean Squared error

acc : Accuracy

1. **KNN forecasting**

The K nearest neighbour machine learning algorithm is one of the simplest machine learning algorithm. It is a supervised learning algorithm which assumes that objects within close proximity of each other are similar [7]. The algorithm is versatile and can be used for classification , regression and search problems as well.

In this project, KNN algorithm is used for regression purposes. Distance measure used is the Euclidean distance. Inputs for the KNN model are as follows :

k (number of neighbours ) = 4

h(number of values to predict ) = 3

lag factors = 1: 12

The values of lag factors are such because , the dataset is a monthly dataset, and hence h = 12.

By giving these inputs, the model will predict values for the next 3 months.

Results obtained from KNN forecasting :

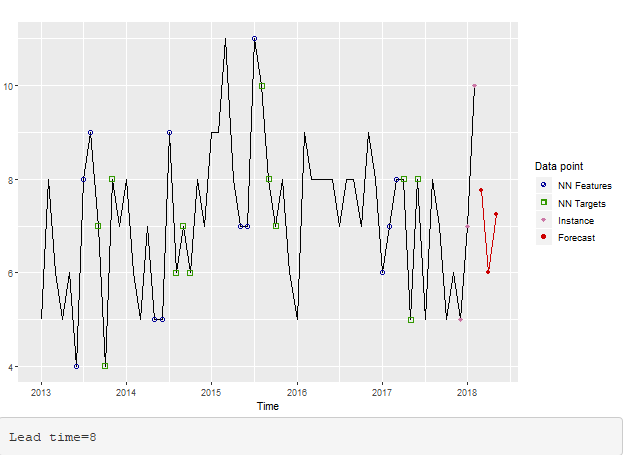


Figure 5.5 : Forecasts from KNN model

**Measures used for comparing accuracy :**

1. Root Mean Squared Error(RMSE)
2. Mean Absolute Error (MAE)
3. Mean Absolute Percentage error (MAPE)

**Accuracy Table :**

|  |  |  |  |
| --- | --- | --- | --- |
| Algorithms | RMSE | MAE | MAPE |
| SMA | 0.3421 | 0.30076 | 2.56 |
| ARIMA | 0.58183 | 0.5003 | 4.7737 |
| Holt Winters | 0.76108 | 0.6945 | 6.59 |
| LSTM | 0.3085 |  |  |
| KNN | 0.2273 | 0.2225 | 2.1007 |

Table 5.1 : Accuracy Table

**Results :**

As seen above, KNN performs the best among all algorithms. It gives the best accuracy as compared to other models . Hence the final model chosen for time series forecasting in K nearest neighbour machine learning model.

**Application Development :**

The framework used for building the application is Shiny. As stated above, the finalized algorithm for forecasting is KNN, we embed this model in the application.

Shiny applications are built in R studio itself. The two main libraries required for development are :

1. shiny
2. shinydashboard

**Shiny Client Development :**

1. This section of the application is also known as user interface . The entire code for the client side is written with the **‘ui()’** tag.
2. The results of the application will be described in the format of a dashboard. Hence the format of the page is set to dashboard , by using **dashboardPage()** function .
3. As we have four suppliers, four tabs are created by using **tabPanel()** with following features :
4. Each tab is named after the supplier name
5. Each tab has a **‘width = 3’.** This is because the entire page has width 12, and all four boxes fit in single line.
6. Within each tab , a select input box is provided. This is similar to a **dropdown menu in HTML**. On clicking it , the a list of products sold by that particular supplier drops down, and the user can select which product he wants to.
7. A **submit** button is provided for each tab.
8. The results are displayed in the **dashboard** format. There are four dashboards which display the following:

* Dashboard 1 : Leadtime forecast
* Dashboard 2 : Buy Price forecast
* Dashboard 3 : Profit forecast
* Dashboard 4 : Results from predictions

1. The colour combination is simple white and blue, which looks decent and has minimum strain on the eyes.

**Shiny Server Development :**

This side of the application is written in the server() function . The values are accessed by their id’s that are set in the client side. The important line of code in the server development is function(input, output). The use of this is explained as follows :

* While setting id’s for values on the client side, all of them are set in the format input$value = “value\_name”, for example input$submit = “Submit”, this is so that shiny understands that the value is coming as an input.
* While writing output functions, they are written as :

output$function\_name 🡨 render\_function(………line of code….)

This is because to understand the difference between input and output.

1. The use of render functions :

The render functions are used while writing output functions.

The type of render functions depends on the type of output that has to be generated. The functions used in this application are as follows :

* renderPlot() : to plot output of forecasting graphs for dashboards 1,2,3.
* renderText() : to print results of prediction in text format in dashboard 4.

1. The use of reactive values. :

This application is dynamic , i.e. the output values will change as the input changes. Hence the function reactiveValues() is used :

* By using this function , we can create multiple values of the same variable .
* For example , in this project . The main value is set as

Predictions 🡨reactiveValues()

and the sub-values are set as

1. predictions$lead\_time
2. predictions$profit
3. predictions$demand
4. The use of plot\_click() function :

The plot\_click() function is used to make the graph interactive . When the user hovers over the graph and clicks on a point, the x and y value is printed below. Example : Lead time = 7 , Month , Year = January, 2015.

# Chapter 6. Testing and Evaluation

The last step of the development of the application is to check whether it is giving accurate results and how it treats different inputs. As this system has many types of products and three different variables to forecast, the attributes of the product were tested for their actual value and predicted value. Also ,the system was checked for errors.

The following tables explain the testing and evaluation results :

1. Checking the prediction for lead time

|  |  |  |
| --- | --- | --- |
| Product Name | Actual Value | Predicted Value |
| Speaker | 8 | 7.75 |
| Adapter | 5 | 5.25 |
| Charger | 6 | 5.85 |

Table 6.1 Prediction for lead time

The algorithm predicts values in point format, the final value is obtained by rounding of the value.

1. Checking for prediction of buy price

|  |  |  |
| --- | --- | --- |
| Product Name | Actual Value | Predicted Value |
| Speaker | 112.56 | 112.89 |
| Adapter | 9.56 | 9.05 |
| Charger | 10.93 | 10.68 |

Table 6.2 Prediction for Buy Price

The algorithm predicts values that are almost close to the actual values.

1. Checking the predictions for demand of products

|  |  |  |
| --- | --- | --- |
| Product Name | Actual Value | Predicted Value |
| Speaker | 34 | 33 |
| Adapter | 25 | 27 |
| Charger | 52 | 54 |

Table 6.3 Prediction for demand

As seen from the above table predictions are made with minimal errors.

**Results of Testing and Evaluation :**

The system is pretty much accurate and is making predictions that are very much similar to actual values. However there are a few errors in the amount of 15 to 20 percentage that can be used to improvised in the future.

# Chapter 7. Conclusions and Future Work

# Conclusion :

Inventory Management is utmost necessary factor that can take businesses to new peaks of glory . Combining the time factor available in historical data and the powerful algorithms of data and statistical analytics, insights can be achieved as well a future strategy can be developed.

Forecasting technique helps businesses to be proactive. They can be prepared well in advance for the problems that might rise in the future and can tackle them with ease. KNN , even though a simple model, is powerful and has the ability to achieve great results. However, the accuracy can be improved. The application developed will help businesses to visualize a data within a click and get predictions within no time.

Data Analytics is a powerful tool , that can enhance the future of businesses and lead to their overall development.

**Future Work :**

There are a set of improvements that can be done to this system in the future. Exploring other variables of the dataset that can be used for analysis and how these affect the nature of sales.

Implementing more complex and different types of algorithm to improve the accuracy can also be carried out in the future.

# 

# Bibliography

|  |  |
| --- | --- |
| [1] | tradegecko.com, “TradeGecko-eBook-sales-and-inventory-forecasting-for-small-businesses,” [Online]. Available: https://www.tradegecko.com/hubfs/eBooks/TradeGecko-eBook-sales-and-inventory-forecasting-for-small-businesses.pdf. [Accessed 15 August 2019]. |
| [2] | N. C. D. A. R. G. Nishanth Damakonda, “An Intelligent Approach to Demand Forecasting,” *International Conference on Inventive Computation Technology (ICICT),* vol. 2, 2017. |
| [3] | X. Z. Kut C. So, “Impact of Supplier's Lead Time and Forecast Demand Updating On Retailer's Order Quantity Variability in a two - level Supply Chain.,” *International Journal of Production Economics,* vol. 86, no. 2, pp. 169-179, 2003. |
| [4] | W. Y. D. Y. P. K. J. C. J. S. Ashis Gopal Banerjee, “A Hybrid Statistical Method For Accurate Prediction Of Supplier Delivery Times Of Aircraft Engine Parts,” *International Design Engineering Technical Conferences & Computers and Information in Engineering Conference,* pp. 2-5, 2015. |
| [5] | A. B. J. a. D. S. M. Syntetos, “Forecasting for Inventory Planning : A 50- Year Review,” *Journal of Operations Research Society,* vol. 60, no. S1, pp. 149-160, 2009. |
| [6] | J. G. Christian Mascle, “Inventory Management Maximization Based on Sales Forecast : Case Study,” *Journal Of Production Planning and Control,* vol. 25, no. 12, pp. 1039-1057, 2014. |
| [7] | R. B. P. V. Sabrina Berbain, “Forecasting, Production and Inventory Management of Short Life Cycle Products : A review of the literature and case studies,” *Supply Chain Forum,* vol. 12, no. 4, 2011. |
| [8] | S. C. W. Spyros Makridakis, “Forecasting : Issues and Challenges in Marketing Management.,” *Journal of Marketing,* pp. 24-28, 1977. |
| [9] | G. K. ,. S. R. Kot S., “Theory of Inventory Management Based on Demand Forecasting,” *Polish Journal Of Management Studies,* vol. 3, 2011. |
| [10] | G. N. Naliaka V.W., “Role of Inventory Management on Competitive Advantage among Manufacturing Firms in Kenya : A case study of Unga Group Limited,” *International Journal of Academic Research in Business and Social Sciences ,* vol. 5, no. 5, pp. 2222-6990, 2015. |
| [11] | K. C. G. J. Scott Armstrong, “Demand Forecasting II : Evidence Based Methods and Checklists,” *Working Paper 89-clean,* 2017. |
| [12] | A. K. L. S. Sanjeev Kumar Agarwal, “Short term price forecasting in deregulated electricity markets,” *International Journal of Energy Sector Management,* vol. 3, no. 4, pp. 333-358, 2009. |
| [13] | H. P. Margaretha Ohyver, “Arima Model for Forecasting the Price of Medium Quality Rice,” *International Conference of Computer Science and Computational Intelligence,* vol. 135, pp. 707-711, 2018. |
| [14] | R. K. A. Ratnadip Adhikari, “An Introductory study on time series forecasting and modelling,” *Lambert Publications,* 2009. |
| [15] | A. B. Darya Plinere, “Case Study on Inventory Management Improvemnt,” *Information Technology and management Science,* pp. 91-96, 2015. |
| [16] | M. M. S. H. F. Neda Khalil Zedah, “Intelligent Sales Prediction for Pharmaceutical Distribution Companies,” *Mathematical Problems in Engineering ,* vol. 2014, p. 15, 2014. |
| [17] | “The sample dataset,” Lokad Technical Documentation, [Online]. Available: https://docs.lokad.com/getting-started/the-sample-dataset/. [Accessed 15 July 2019]. |
| [18] | H. A. X. X. X. L. X. S. Lijun Wang, “Generating Moving Average Trading Rules on the oil futures market with Genetic Algorithms,” *Mathematical Problems in Engineering,* vol. 2014, p. 10, 2014. |
| [19] | L. S. ,. S. G. Prapanna Mondel, “Study of Effectiveness of Time Series Modelling(Arima) in forecasting Stock Prices,” *International Journal of Computer Science, Engineeering and Applications (IJCSEA),* vol. 4, no. 2, 2014. |
| [20] | M. B. Eimutus Valakevicius, “Application of seasonal Holt Winters Model to study Exchange Rate Volatility,” *Engineering Economics,* vol. 4, no. 26, pp. 384-390, 2015. |
| [21] | N. T. A. S. N. Sima Samini Namini, “A Comparison of ARIMA and LSTM in Forecasting,” *IEEE International Conference on Machine Learning and and Applications (ICMLA),* no. 17, 2018. |
| [22] | V. S. Suresh Kumar Sharma, “Time series Prediction Using KNN Algorithms Via Euclidean Distance,” *Asian Journal of Computer Science And Information Technology,* vol. 2, no. 7, pp. 219-221, 2012. |