Implementing Automation on Warehouse Optimization and Administrating Machine Learning Strategies for its Maintenance

MSc Research Project
Data Analytics

Shrikant Samarth
Student ID: x18129137

School of Computing
National College of Ireland

Supervisor: Anu Sahani

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Shrikant Samarth
Student ID:	x18129137
Programme:	Data Analytics
Year:	2019-2020
Module:	MSc Research Project
Supervisor:	Anu Sahani
Submission Due Date:	11/08/2019
Project Title:	Implementing Automation on Warehouse Optimization and Administrating Machine Learning Strategies for its Maintenance
Word Count:	XXX
Page Count:	25

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	11th August 2019

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Implementing Automation on Warehouse Optimization and Administrating Machine Learning Strategies for its Maintenance

Shrikant Samarth
x18129137

Abstract

A Business is defined as an integration or enterprise occupying commercial or professional exercises. Consumer and the producer share a give and take relationship. Producers business can be small, medium and large scale depending upon the targeted consumers. For the small and medium scale business, consumer requirements are easily met, and customer satisfaction is maintained. The large-scale business however targets a larger audience and it turns out to be difficult to manage the requirements of a variety of consumers. To meet the requirements, the producer needs a good infrastructure in place for customers. The warehouse plays an important role as it stores all the elements to be delivered to consumers, be it a B2B or B2C entity. To maintain the wear and tear of warehouse, the warehouse manager must invest a large amount of resources to keep business running. This plan of action speaks about the automation of warehouse activities, so that the optimization of warehouse is achieved; and with the help of data science principles the maintenance of warehouse is carried on, which brings about smoothness in warehouse governance. To meet this objective, data science models will be used to identify the patterns from the dataset and the resultant will then be used as an input to the pricing mechanisms, which will keep the wearing-out products from the warehouse at bay.

Keywords: E-commerce, Warehouse Optimization, Enhancement, Machine Learning

Contents

1 Introduction

E-Commerce industry is the most nourishing industry with buying and selling of products over internet and brings in various subsidiary industries like mobile commerce, supply chain management, warehouse maintenance etc. In the words of Bloomenthal (n.d.), Electronic commerce or e-commerce is a business model that lets firms and individuals conduct business over electronic networks, most notably: the Internet. This e-commerce operates in 4 different divisions namely Business-to-business (ex: Newegg for business), business-to-consumer (sellers on Walmart, Amazon etc..), consumer-to-business (ex: Netscape) and consumer-to-consumer (ex: Ebay). The e-commerce conducted at every division has slight difference with related to the dealing with the people. The focus of this paper is business-to-business area where Kłodawski et al. (2017) mentions the warehouse manager face issues related with movement of products which results in storage space issues. Based on the market trend and by sales forecasting, warehouse manager brings in products to the warehouse. But usually these products get more time for its movement through Ecommerce platforms than forecasted, because of various reasons like change in the trend or seasonal change or any new competitor product introduced to the market. Ecommerce facilitates online business to the customers and makes available the products and services needed to the customer so as to provide best possible assistance and satisfaction.

1.1 Motivation

Due to increased amount of competition in the market and customer satisfaction practices, the analysts are ready to gadget recommendation and machine learning strategies to improve user experience. It is easy to understand the working model of warehouse management for a small-scale industry as the tasks can be taken care with manual intervention. But this can lead to disadvantage for a large-scale industry as amount of defects would increase exponentially giving rise to waste miss-management and incurring monetary losses Laudon and Laudon (2015). To improve the efficiency of operational flow in the warehouse, a research was conducted by Reyes et al. (2019) which directs us to tackle operations like volume of products, uncertainty in demands and rapid customer service demand. Ai-min and Jia (2011) research study provides us an idea of how optimization is achieved with the help of genetic algorithm and implements simulation with the help of decimal coding, sum of weight method to solve fitness function and uniform mutation operation. To perform any type of automation with the help of AI, we need to understand different administered examples over warehouse management, Mccrea (2019) gives us targeted idea about return-to-inventory management and possess internal technical expertise to build system functionality that make implementation easier and better. One of the strategies is discussed in the research work done by Kartal et al. (2016) guides us about employment of specific machine learning algorithms which are discussed later in literature review to predict classes of initially determined stock items. In this paper, researchers proposed a product placement optimization in medium-to-large logistic company for both new and already operating warehouse. A machine learning has a primary aim to allow computers learn automatically without human intervention or assistance and adjust actions accordingly. The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide.

With the emergence of warehouse optimization for large scale industries, it becomes very important to automatize the income and outgo of the products to maintain the stability of the warehouse and market as well. In such situations, customer feedback becomes very important factor which adds on to the data that can sentimentally work in the favor of warehouse manager. A study conducted by Zunic et al. (2018) concept can be elaborated with the implementation of machine and deep learning concepts where the system will be trained to an extent that it will identify the wear-off time period of a specific product in the warehouse and let the manager know the quantity to be ordered in the next release. This research can help the warehouse manager to understand the patterns of different products in the warehouse and then with the application of different types of pricing techniques the quantities of products to be ordered and prices of products can be manipulated. A research collectively done by Minga et al. (2003) consists of application of different dynamic pricing strategies which can help individual to further optimize the stock management in a warehouse. This research would help the seller understand the stock management with the help of different machine learning techniques and would help him understand a pattern of specific products blow out time period and with the help of which the seller would get an idea of clearing the stock before it goes out of trend and product is no longer a heavy subject in the warehouse.

1.2 Research Question

- How to predict the blowout time period of seasonal products in advance via implementation of Machine learning methodology?
- To what extent different pricing strategies can be applied on the slow-moving products to maintain the warehouse space?

1.2.1 Research Objective

The aim of this paper is to predict blowout products which are going to be out of trend by applying certain machine learning techniques; which would help warehouse manager to get advance information to clear the excessive stock from the warehouse as and when required, through certain online business to business strategies.

1.3 Plan of paper

This research proposal is divided in four sections i.e Introduction, Literature Review, Research Methodology and Summary. All sections collectively address an explanation towards the research objective; which is to optimize the warehouse space by identifying the blow-out products. Section 2 focuses on the related work done in previous literature studies which is divided into five sub sections. This section will mainly consist of work done by different research scientists who have already paved their way towards warehouse maintenance and other related areas. Section 3 is divided into seven sub sections which describe the defined procedure to attain the objective. This section also discusses about the proposed methodology that will be used to answer the research questions and also present a series of evaluation methods that will be used to assess the models and techniques. The last section 4 summarizes work done in this research proposal and gives a brief overview of all sections.

2 Literature Review

The main purpose behind this section is to give a detailed outline of previous studies which addressed in predicting the issues in the warehouse management using various techniques; also, based on previous studies how this research undertaken to improvise those issues can be seen in this section. Before going to the main section first we need to investigate the previous history of warehouse management to get acquainted with the topic.

2.1 History of Warehouse Management

Over the last two decades Logistics Service Providers have turned out to be significant players in numerous chains and businesses. New difficulties have emerged because of the advancement technology. So, to make a proper decision it is important to have an insight how decision should be made and what are the important parameters for a warehouse management. Krauth et al. (2005) paper introduces a framework which clusters a performance measurement of different streams. Based on the researchers empirical validation, they suggest a list of key performance indicators that could help industry to reexamine their operations and let them think apart from cost minimization. Additionally, in the process of improvement in the decision making, one of the key performance indicators is the warehousing management problems which includes inventory management under space restriction, storage allocation and scheduling the warehouse operations. Van den Berg and Zijm (1999) discussed some examples of models which shows insights of relation between inventory control decisions and product allocation and assignment problems. With the introduction in the electronic shopping and ordering has profoundly changed the logistics of the supply chain that lead to a dramatic change in the warehouse inventory management. The study shows, a sophisticated class assignment approach can lead to a better service since it provides a proper storage space allocation. For instance, in the e-commerce sector, where sellers have given just in time deliveries there is a constant pressure of maintaining a trend product as well as to improve response times of the warehouse. As a result, the enthusiasm for new, complex arranging strategies is quickly developing in this sector.

For any warehouse, optimal operation is achieved when each customer is satisfied by fulfilling order in due time; which could be possible only when all logistic warehouse operations are done in the shortest possible time utilizing all the resources with minimal cost under dramatically changing conditions. Karásek (2013) discussed the issues regarding the warehouse layout which lies in the viable utilization of space. Further the researcher utilize the shop rescheduling technique with the vehicle routing problems to ensure the optimization of warehouse. Automated Guided Vehicles (AGV) techniques converted from open space Vehicle Routing Problem (VRP) techniques to adjust in environment of warehouse. The transportation in the warehouse was studied and how it can be improvised is discussed in this paper.

It can be seen through a case study on warehouse which currently utilizes the earliest due date (EDD) strategy for sequencing its order fulfillment jobs which is easy to implement because of the static nature of the job priority. The study done by Kim (2018) shows that critical ratio (CR) priority rule performs well where delayed delivery spoils products virtue by rescheduling pickup times. The simulation results dependent on the warehouse specific cost parameters demonstrate potential advantages of the CR rule. Compared with the other rules for which the need is either fixed on entry refreshed as per stages, CR has the special property that it adjusts priorities by examining the dynamic lining status of jobs. Considering real work process examples could refine these appraisals by product lining information from the warehouse process by forecasting the queuing times using statistical and Machine learning techniques. This study holds importance, as it directs us to investigate further in the space and product optimization in a warehouse.

The above-mentioned literature has shown the difficulties in the advancement, warehouse space and suggests a need for optimization in this sector. Now that the warehouse optimization and its importance is explained, its time to review some studies regarding what more challenges and issues in handling large scale warehouses needs to be further explored in the next section.

2.2 Issues and Problems in Handling Large Scale Warehouses

The decision-making process is one of the entangled procedures associated with both for the large and the small-scale warehouse operation for productively satisfying customer orders. This is particularly valid if the orders require cross-border delivery activities. For such complex problems, case-based reasoning is an intelligent method is useful which uses previous cases to find a solution to new problems. However, recent case retrieval methods are based on a fixed set of attributes for various types of orders in which particular order features for a case groups are ignored. Therefore, Lam et al. (2012) in his research work a hybrid approach called case genetic algorithm based on decision support model (C-GADS) is proposed in characterizing new customer orders into groups with similar values which helps in selecting the most comparable cases among the group. For the research, three categories of data was collected i.e. customer request which includes order and product information, warehouse operation includes pick-up sequence, operation time and available resources and cross border regulatory policies. The GA applied on each clusters in the case retrieval process shows enhancement in delivery operation planning process and also increases effectiveness of the warehouse operation and customer order practices based on grouping of similar cases.

In any warehouse, replenishment of stock is a complicated process as it needs to be dependent upon production and the status of the supplier. Moreover, any warehouse faces problems regarding the maintenance of the stock as the products faces uncertain demands over a certain period. It is therefore important and a need to find an efficient storage of a warehouse system. According to Ang et al. (2012) heuristics methods applied on the turnover-based and the duration-of-stay based policy do not consider the changeability of both inflow and outflow of products. For this, 48 weeks of arriving and departing pallet groups of each product data was used. To handle the changeability in the product flow, researchers considered a factor-based demand model where demand of a product in each period is uncertain. So, they have developed a storage assignment problem in a warehouse as a robust optimization problem limits the expected worst-case scenario to the capacity constraints of the storage classes. From a case study it has been found that using weekly planning linear storage retrieval policy can be obtained for certain amount of time and saving with this method can be as high as 36%. Moreover, the distribution center is informed regarding the arrival of pallets with 98% of accuracy ahead of time with the analysis. Interestingly, in spite of uncertain demand distributions, the confined linear policy achieves near the expected value. Hong et al. (2012) introduces Route selecting and order batching (RSB) model bound with the heuristic procedure to resolve large scale order batching problems. The data set generator and comparison of FCFS, Seed, CWII and LB (linear relaxation of RPP) algorithm was developed to evaluate performance. The result evaluation shows Seed and CWII algorithm performs better for large number of orders. Also, the performed evaluation shows the heuristic route selection based order batching algorithm delivers close to the optimal solution arrangements in a narrow aisle order picking system where there are ten aisles and more orders.

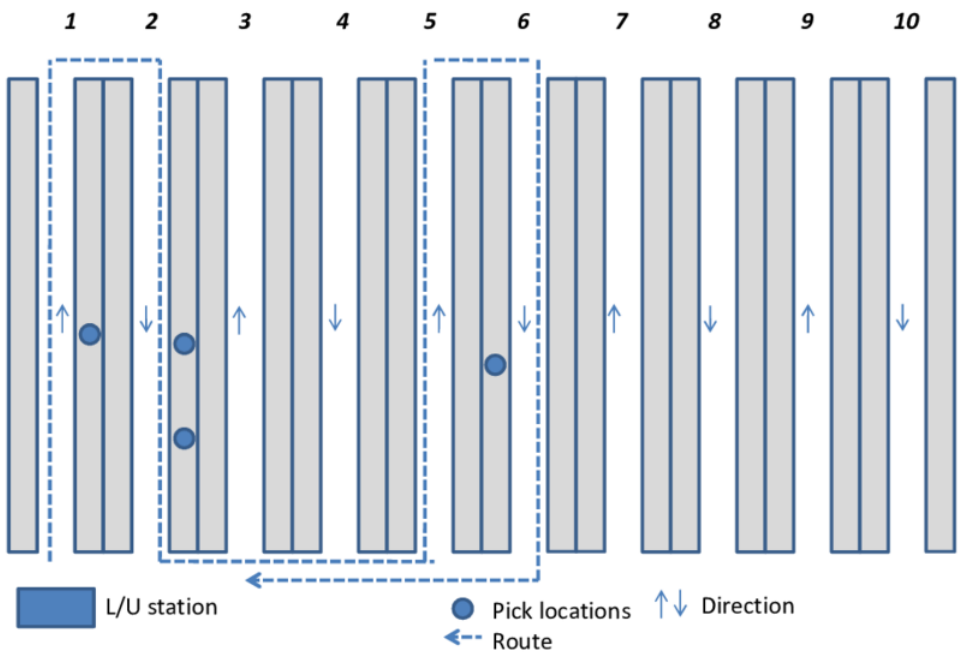

Figure 1: Data Source: Large-scale order batching in parallel-aisle picking systems.Hong et al. (2012)

The developed procedure contributes to effective distribution centers i.e warehouse operation where both space usage and operational throughput are basic. Figure 1 illustrates a proposed ten aisle shelve order picking with parallel-aisle system to improve the picking time per item.

The above literature studies used genetic algorithm based on decision support model and heuristic algorithm on the data for large and a small distribution centers to achieve the effective warehouse optimization. These studies emphasized further to look into the automation for warehouse complex data sets in order to achieve the effective space utilization in large warehouses.

2.3 Automation Today in Warehouse Optimization

There are many studies which worked on the automation strategies in the different areas of distribution center to effectively organize the warehouse management. One such study conducted by Zunic et al. (2018) on predictive analysis on efficient product placement for already operational warehouse which are not productively able to process the warehouse stock. For this, the product position for about a year worth of data was taken based on the daily item picking frequency and item grade. Position Grade Calculation Pseudo Code algorithm was developed which results in the 17.34% average picking route length reduction for more than 40 orders. The optimization module was implemented in medium to large warehouse in Bonsia and Herzegovina which was compared with the initially empty warehouse. Zunic et al. (2017) in his earlier study uses a statistic based fitting algorithm (picking zone prediction filter) for the optimal and strategic product placement in the picking zone for medium to a large logistics warehouse with the help of real-world case study. His study addressed the issue that in the event if the product is not available in the picking zone, the worker needs to depend on the forklift will put the total palettes down from the top racks which decrease the efficiency of the warehouse. As we know from the above literature review that the picking process takes almost 50% of the total cost. Hence, any optimization in this area will significantly increase the total efficiency of the warehouse. Hence, in this paper for optimization, one month of data was considered as the bench-marking period, as a result, 91.73% of accuracy achieved with shifted median whereas, 93.08% for rounded shifted median.

In one of the earlier study on improving the productivity of order picking and multi-level rack distribution conducted by Chan and Chan (2011) uses a real case warehouse data. Auto replenishment orders were categorized into different picking density groups and for each time was recorded. Based on this aisle number, station number and rack number are arranged, when order arrives the system would assign a picking location for respective items. The classification of products is checked using the COI index (i.e cube per order index) which gives the 60% of the total turnover and EIQ method (i.e Entry Item Quantity) analysis shows that order quantity of each item (IQ analysis) and frequency of order of each time (IK analysis) gives better results to reflect inventory turnover of warehouse. A multiple rack warehouse method used to improve the performance based on storage assignments, routing and picking density.

The advancements and the benefits of machine learning technology will boost the automation in the warehouse industry which will further contribute towards optimization of warehouse. There are some researches conducted in this area which is drafted in the next section.

2.4 Warehouse Management using Machine Learning

Many studies by now have utilized both Supervised and Unsupervised Machine learning techniques in various areas of the warehouse to increase its effectiveness and optimize its workflow.

In any logistic warehouse, inventory management plays a major role in overall warehouse performance. Until now, there are various traditional methods based on mathematical functions available that give a good amount of success in predicting a reordering time. But these traditional methods are time-consuming for any warehouse officer. Hence, Inprasit and Tanachutiwat (2018) worked on safety stock management and reordering point determination using Machine learning technique. For training purpose, the data was collected from various companies which were publicly available online. By applying various training algorithms, testing and comparison of Artificial Neural Network model (ANN), evaluation of safety stock and reordering point for products based on 5 factors. It has been found out that, MSE (mean squared error) performed best with the algorithm in order of $[10^{-5}]$ and 0.999 adjusted R^2 value. Also, most accurate neural network model was identified. To which researchers conclude to further research work in the inventory management field. On the other hand, for the analysis of multi-attribute inventory classification problems Kartal et al. (2016) integrates a "genetic" hybrid methodology of multi-criteria decision making (MCDM) models with machine learning methods. First ABC analysis is employed using simple additive weighting, analytical hierarchy and VICOR MCDM techniques for determination exact classes of inventory items. For performance evaluation, Machine learning algorithms such as naive Bayes, Bayesian network, artificial neural network (ANN) and support vector machine (SVM) are implemented and predicted classes of stock items. It has been found out that SVM outperformed all models and demonstrated 91.93% of highest accuracy followed by the SVM-normalized poly kernel with 89.89% of average accuracy in classifying inventory items. Using the balanced data set, the novel hybrid approach increases the ANN's performance from 81.6% to 94.70%. However, this study shows unlike ANN, SVMs did not exhibit any huge change over the balanced and unbalanced data sets.

In one of the contemporary studies, de Santis et al. (2017) addressed an issue of material backorder in a supply chain warehousing. In any company, identifying products with the highest chance of shortage/backorder before its occurrence, conferring the business a suitable time to react and could increase its overall performance. This study proposed a predictive model for imbalance class problem, where the frequency of backorder items is low compared to the items that do not. This study uses a real-world imbalanced dataset which was taken from Kaggle competition. The dataset contains 8 weeks of historical data before the week which they were predicting. First logistic Regression (LOGIST) was used for binomial classification and to provide a baseline score; followed by the Classification Tree (CART) model to present outlier robustness and feature selection. Then after sampling techniques, machine learning techniques such as Random Forest (FOREST), Gradient Tree Boosting and Blagging (BLAG) were applied. As a result, RF gives 0.9441 AUC (area under the curve) score by adopting a bagging ensemble of the tree whereas the Gradient Tree Boosting gives the highest of 0.9482 AUC score. Whereas, in a practical approach, BLAG give 20% of the products which could go into backorder and correctly identified 60% of positive class items. This research work motivates my study to work in the direction of improving the identification of backorder before its occurrence.

In another study, a genetic algorithm was used by Ai-min and Jia (2011) on storage management of pharmaceutical products and characteristic of automated warehouse for slotting optimization. There are different issues occurs in a pharmaceutical warehouse than a normal warehouse. To avoid confusion between the products which are placed together, a bar code technology is used. The weight of the products as well as the distance of the products are considered to avoid the time consumption. This paper used MATLAB and genetic algorithm to resolve multi-objective optimization problem. From the analysis it has been found out that, a specific distribution of goods before and after the optimization could be calculated by genetic algorithm to solve the efficiency of such warehouse. The researcher further cites to use this scenario in the real-life data for further improvement.

In recent years, some of the research work done on RFID enabled warehouse. Zhou et al. (2017) in his research focuses on warehouse environment where both shelves and products are tracked by the RFID-based system. A novel concept of flexible warehousing dependent on real-time decision support system was introduced in this paper. A Knowledge-based self-adjusting mechanism was also proposed to satisfy inventory demands at the lowest cost for multiple periods. The results show the significant performance improvement with this method of selected parameters of location and capacity of the warehouse. For further improvisation by relaxing some of the traditional guidelines and practices proposed in this paper. The author suggests further improvement and suggests

KNN, VM, and heuristic algorithm be used to handle more complex task specialized classification. Alfian et al. (2019) extends this research work and addressed an issue on false-positive data capturing in RFID using Nave Bayes (NB), multilayer perceptron (MLP), Support Vector Machine (SVM), Decision Tree, and Random Forest algorithms. These algorithms were then integrated with IQR or LOF as it improved the accuracy. As a result, conventional Random forest model outperformed other models and achieved 97% of accuracy. Whereas, previous studies done by Keller et al. (2015) and Ma et al. (2018) comparing ML algorithms, anticipated RFID utilizing Received Signal Strength (RSS) demonstrated that, high classication accuracy up to 95.5% could be attained with Support Vector Machine (SVM). However, these studies focus on static tags, whereas, Alfian et al. (2019) research study considered static labels just as labels moving near the door tag as false-positive readings. These research work can help my research-project in developing the product classification and identifying the blow-out product using machine learning techniques.

Another important parameter for any warehouse is the sales forecasting. To ensure products momentum of warehouse, it is important to maintain the effective sales forecasting and manage the warehouse inventory. Chen and Lu (2017) proposed clustering based forecasting model by combining it with Machine learning techniques for retailing sales. For this, three computer retailer data on notebook computer (NB), Personal computer and liquid crystal display (LCD) was collected. In this paper, using clustering technique is clusters with similar features into groups and then machine learning methods are used to train the forecasting models of these groups. Clustering technique such as self organizing map (SOM), K-mean and growing hierarchical self-organizing map (GHSOM) and support vector regression (SVM) and extreme learning machine (ELM) machine learning technique was used in this analysis. Using the MAPE indicator GHSOM-ELM method shows most promising performance for forecasting of three computer products over other methods with the smallest prediction error. In the same period, Bohanec et al. (2017) addressed an issue of weak performance in prediction of B2B sales forecasting in decision making process. An organizational model was developed using machine learning models combined with) general explanation (IME and EXPLAIN) methods. The classification accuracy performance was evaluated for Random Forest (RF), Naive Bayes (NB), Decision Tree (DT), neural nets (NN) and Support Vector Machine (SVM) ML algorithms. Random Forest performed well in terms of AUC with 0.853 average score and 0.782 classification accuracy performance. The monthly B2B sales forecasts monthly and compared them with ML forecast models, the actual outcomes for the next month was evaluated with this approach. The above two sales forecasting studies could help my project to build a strategy for clearing the excessive stock from the warehouse.

However, there are very few studies who worked to optimize warehouse space by identifying those products which are going out of trend and how to get rid of such products effectively which makes this project novel. Now after identifying the challenges from the above literature work which are straight from basic Inventory management, Backorder challenges to the automated RFID enabled warehouse challenges and sales forecasting methods; it is time to review research work on the effective dynamic pricing strategies to sell the products along with the blowout products effectively.

2.5 Dynamic Pricing for Warehouse Maintenance

Pricing strategy is another important aspect of supply chain industry, where competitive pricing plays a vital role in making a good product momentum. While selling the blowout products from B2B Ecommerce platform like Newegg for Business a demand sensitive pricing model is useful for fast changing in such electronic commerce business. One such study conducted by Minga et al. (2003) proposed a application of dynamic pricing strategies which can help individual to further optimize the stock management in a warehouse. Price setting algorithm was introduced for demand sensitive models which could help decision makers to maximize the profit for the bulk quantity ordered by the buyers. The researchers mentions some websites which allows buyers to register for the purchase and notifies them as soon as the price drops. When enough people have agreed for the purchase the price drops further. Hence, this proposed price setting algorithm helps seller to make good amount of profit even with the increase in the demand. The price elasticity of demand method mentioned in the research work examines the relationship between price and the unit sale. The result of analysis shows the price and marginal cost of a product drops and follows the curve with the demand aggregation purchased higher volumes. The relation and the methodology mentioned in this literature could help my project to formulate further effective strategy for selling goods on the Ecommerce platform in order to manage and optimize space in the warehouse.

Another effort to improve the accuracy and business performance in Ecommerce, an extension of price levels and price dispersion within and across multiple retailer was studied. Ancarani and Shankar (2004) proposed a hypothesis on how price and price dispersion compare among traditional and multi-channel retailers. The data was collected from the retailers and e-retailers of Italy which contains 13,720 price quotes of books and CD category. Price Sensitive multidimensional recommender system (PSRS) and collaborative filtering was used and the analysis was done with a statistical methods. The researchers used user based and item based engines by two different criteria to set different price category i.e. low price and high price. In $P_T, balance$ user sparsity of user with item price matrices were balanced whereas, in P_T, avg the item categories were divided into two equal intervals by splitting the price range of products. As a result in the analysis the observations were found out as the low-priced recommendation increase the accuracy whereas, the high priced recommendation decreases the accuracy. In this analysis the researchers demonstrates the way price setting affects the business performance. This analysis is useful for my project and gives the direction in order to sell bulk products how the price setting can be done effectively.

Now after the detailed analysis of literature review, now lets look into the methodology (section 3) which describes about the approach to carry out the research questions (section 1.2).

A brief summary of the above literature review is provided in the table given below.

Study	Data Source	Model / Technique	Result/observation
Krauth et al. (2005)	New dataset created - multiple sources	New Frame Work to Cluster different Stream	Business Under- -standing
Van den Berg and Zijm (1999)	ERP Dataset from multiple sources	Polynomial time dynamic-programming	Storage space improved
Karásek (2013)		AVG technique combined with VRP technique	Warehouse layout understanding and optimization
Kim (2018)	Warehouse data	Critical ratio (CR) rule and SLACK rule	Critical ratio (CR) priority rule performs well
Lam et al. (2012)	Company Data	Hybrid Approach (C-GADS)	Delivery planning process improved
Ang et al. (2012)	Client's 17 Weeks of data was used	ADAPTIVE algorithm & Heuristics	Improve- Storage saving by 36% of improvi- -sation) & 98% pallets one week in advance
Hong et al. (2012)	Dataset generator was used	FCFS, Seed, CWII & LB (linear relaxation of RPP) algorithm	Seed and CWII algorithm performs better
Zunic et al. (2018)	1 year product positioning data	Position Grade Calculation Pseudo Code	Reduction upto 17.34% avg picking route length
Zunic et al. (2017)	Warehouse's 1-Month of dataset	Statistics-Based Fitting algorithm	Shifted median=91.73% Round shifted median = 93.08%
Chan and Chan (2011)	Real case warehouse data	ABC classification, CQI-Index,EIQ analysis	IQ and IK analysis performs better
Inprasit and Tanachutiwat (2018)	Various companies data was collected	Artificial Neural Network model	MSE performed best and Accurate ANN model was identified
Kartal et al. (2016)	Company dataset	Genetic Hybrid Methodology, naive Bayes, Bayesian network, ANN & SVM)	SVM outperformed (91.93%) & w/ hybrid approach ANN accuracy 94.70%
de Santis et al. (2017)	Kaggle competition dataset	LR,Statistical method,RF,Gradient Tree Boosting & BLAG	RF= 0.9441 AUC score & Gradient Tree Boosting = 0.9482 AUC score. BLAG = 20% Backorder items predict
Ai-min and Jia (2011)	Company dataset used	MATLAB and employs genetic algorithms	Specific distribution of goods solved by genetic algorithm

Study	Data Source	Model / Technique	Result/observation
Zhou et al. (2017)	Real time warehouse data	Decision support system	Wh. location & capacity Improvement
Alfian et al. (2019)	RFID enabled warehouse data collected	NB, MLP, SVM, DT and RF algorithms	RF outperformed w/ 97% accuracy
Keller et al. (2015)	Warehouse RFID dataset collected	LR,DT, ANN,Time series analysis	Single-attribute classier 95.83% accuracy
Ma et al. (2018)	Real time RFID data collected	LR, SVM, DT	SVM = 95.3% accuracy
Chen and Lu (2017)	Three computer products sales data collected	SOM, K-mean, GHSOM, SVM, ELM	GHSOM-ELM shows most promising forecasting performance
Bohanec et al. (2017)	Data collected by users was used	RF, NB, DT, ANN & SVM	RF performed(AUC= 0.853&classification accuracy = 0.782)
Minga et al. (2003)		Statistical techniques	Business Understanding, effective pricing strategy
Ancarani and Shankar (2004)	Retailers & e-retailers of Italy	Statistical techniques	Price settings understanding

3 Research Methodology and Specification

From the above literature, it has come to notice that the researchers have done a relentless effort to understand the concept of warehouse management and have put forth different sets of algorithm that have proved beneficial from time to time. With a deep and an in-depth research on the data, an analyst can understand patterns in the data consumption and this will in turn help the user on how to deal with the products that are available in the market. It becomes very important to understand and discuss about the quality of data before the data is used for machine learning purpose. The 4 Big-Vs of the Big Data Programming suggests that, volume, variety, velocity and veracity are very important factors for understanding the variations in the data. It turns out to be confusing if these variations intrude the quality of the data. So it becomes very important to comprehend the data quality and then process the data for future use. In the words of Hissibini (2019) data mining is defined as "The process of finding anomalies, patterns and correlations within large data sets to predict outcomes. Using a broad range of techniques, you can use this information to increase revenues, cut costs, improve customer relationships, and reduce risks and more." It is very important to pre-process the data to improve the quality of it and CRISP-DM comes to rescue.

The CRISP-DM (CRoss Industry Standard Process for Data Mining) project[1] addressed parts of these problems by defining a process model which provides a framework for carrying out data. mining projects which is independent of both the industry sector

[1]The CRISP-DM process model is being developed by a consortium of leading data mining users and suppliers: DaimlerChrysler AG, SPSS, NCR, and OHRA. The project was partly sponsored by the European Commission under the ESPRIT program (Project number 24959)

and the technology used. The CRISP-DM process model aims to make large data mining projects, less costly, more reliable, repeatable and manageable.
Wirth and Hipp (2000)

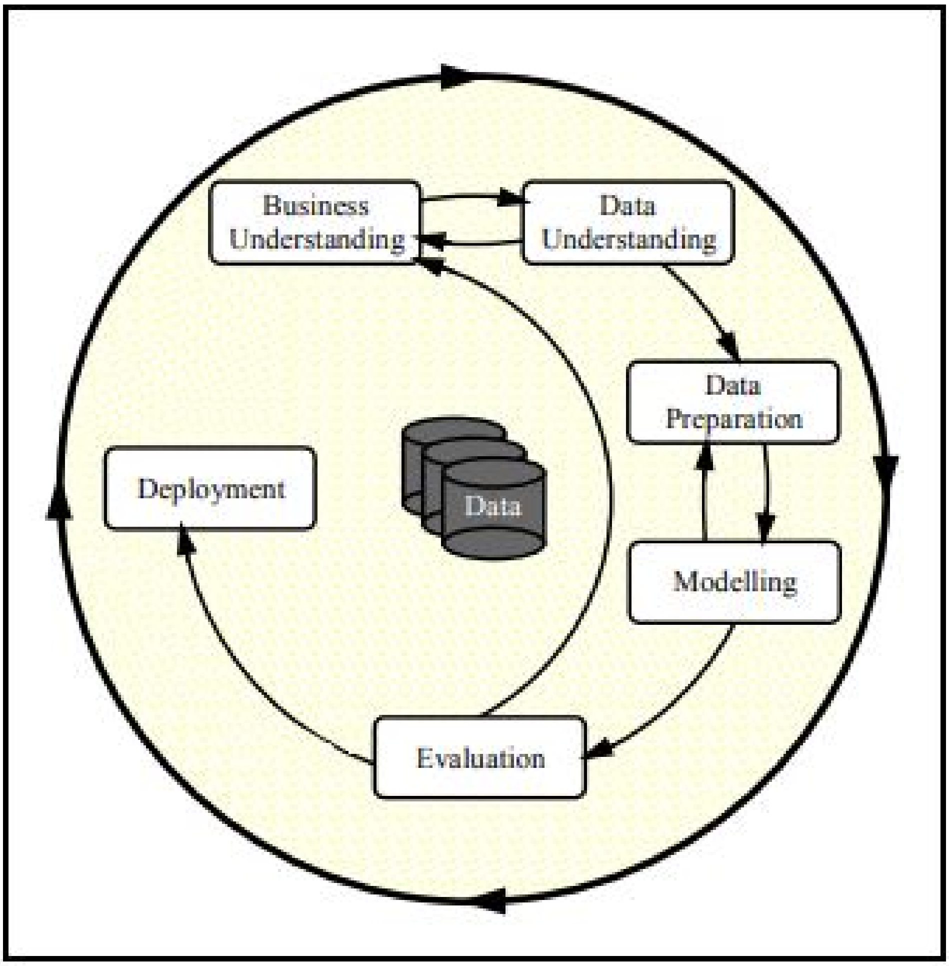

Figure 2: CRISP-DM Process Model for Data Mining, Data Source: Wirth and Hipp (2000)

3.1 Business Understanding

The first and foremost thing is to understand what customers needs to succeed from a business perspective. As the customers always have competitive objectives, it is the part of analyst to understand the business requirement well that will help to influence the outcome of the project. The outputs of the business understanding would be "setting objectives" from a business perspective, "producing the project plan" to achieve the data mining and the business goals and the "business success criteria" which describes the criteria for a successful outcome to the project from business point of view. When it comes to understanding the business before the data to be extracted, heuristics of ecommerce business would provide a helping hand and will also lead us to understanding of pre-defined goals that industry expects and provide an output that would be more than helpful in running the ecommerce warehouse business.

3.2 Data Understanding

Data collection is a very important step as it gives an ice breaker to the planning that has been done. This topic has been dedicated to the improvement of thewarehouse business, data must be a part of a functioning organization which is raw and unfurnished. This will help in understanding the real-world challenges possessed in the data and a motivation to tackle them. To achieve this objective, the data is gathered, with the consent of the start-up organization EasternStar which is involved in warehouse management and affiliate marketing, via electronic mail transfer. With personal experience and expertise in the field of affiliate marketing, it is possible to understand the business throughout and comprehend on what grounds the organization can improve and more efficiently perform business operations with different organizations. The data will generally consist of all the products that the organization sells in 3 months with products divided into 4 types; liquidation, recurring, blow-out and mapped. The data consists of 1000 products for each day for 90 days, which brings about 90000 rows of data. As the data is already labeled, this data will directly use for training algorithms to bring about the expectation of blow out time-period of products and applying different pricing strategies to understand and optimize the warehouse well and improve product income.

3.3 Data Preparation and Transformation

Data pre-processing is an important step in Machine Learning as the nature of information and quality of data that can be derived, directly influences the capacity of the model to adapt; in this manner, it is important that we pre-process our data before proceeding it to the model Nalić and Švraka (2018). The data would consist of all types of values that would decrease its quality. The outlined ways are generally used to remove any discrepancies in the data. Kumar (2019)

- Handling Null and Missing values
- Standardization
- Handling Categorical Values
- One-Hot Encoding
- Multi-collinearity

3.3.1 Handling Null and Missing values

The real-world datasets will always have null values or missing values that occurred due to data entry or due to default settings of the database. These values are not very helpful when it comes to training a model with an algorithm. The best way is to remove such entities or replace them with a default value which does not make any significant variance to the dataset Kumar (2019). In the terms of E-commerce, the null values or the missing values will be handled either by replacing them with zero.

3.3.2 Standardization

The process of standardization is one another concept of data pre-processing where the values are converted into the mean of 0 and the standard deviation to 1 Kumar (2019).

The formula of standardization is given by,

$$Z = (X_i - \mu)/\sigma \tag{1}$$

Where,
Z = Z-score aka standardization value
X = the observation (a specific value)
the observation (a specific value that you are calculating the z-score for).
Mew/μ = the mean
Sigma/μ = the standard deviation

3.3.3 Handling Categorical values

Categorical variables are discrete and not continuous. The categorical variables can also be ordinal i.e. the variables can be ordered in a series. The other form of categorical variable is nominal i.e. the variables cannot be arranged in an order. These non-numeric categorical variables needs to be converted into numerical categories so that we can apply analytic on the dataset Kumar (2019).

3.3.4 One-Hot Encoding

This is a unique way of dealing with the nominal categorical variables. Here, the variables will be divided into m unique columns where m being the number of unique nominal categorical values in the dataset. These can thus be given logical values whenever the column is set with that specific value. For the ecommerce dataset, we have to deal with many categorical values like the payment_type or the product_type which will have theoretical values and have to be handled with one-hot encoding Kumar (2019).

This process is very good and neat process to deal with categorical values but issue arises when we have the problem of multicollinearity. This can be handled in the next topic.

3.3.5 Multicollinearity

Multi-collinearity is when the features are strongly dependent on each other. This can impact the model by changing the decision boundary which can cause a huge change in the result of the model. Also, the weight factor would prove helpless to calculate feature importance. The ecommerce dataset is validated with multi-collinearity and would be handled if found after the one-hot encoding is implemented.

The data transformation and exploration can be normal or complicated depending upon the changes that are needed on the initial and the final data. For the ecommerce dataset, as we have many different columns, it is not very necessary to use all for them while we train the algorithm. The most important concept that is to be recollected while training an algorithm is to not use large number of dimensions which can cause limitations in a model which would cause it to overfit. One of the procedures that will be followed in the training would be dimension reduction which will help us accurately identify number dimensions needed to train the algorithm. It is the most important step as we need to define the training columns which will be product_price and item_status. The columns like MAP and SRP would only cause in increasing the number dimension which is not beneficial for the model to build. This technique would help us rectify the value of data by decreasing the variance.

3.4 Modelling

As we have shortly understood the concept of application of different pre-processing strategies, the data that is available is a finished product and now can be used for training different machine learning algorithms. Machine Learning can be defined as a response that a machine can give which is an improved version of the earlier response based on training with a patterned data Macro V (2017). The machine can reciprocate a response that humans can think of and the accuracy can be monitored with varied quality and quantity of data. This human like feature can help machines judge a specific situation and perform operations which are out of reach for humans, thus leading to a comfortable life. The wider categories of machine learning have been already declared in literature review and the dataset available is a labelled data. This labelled data can be fed to machine learning algorithms. There is a variety of machine learning algorithms that have been discussed in the literature review which can prove beneficial in identifying the blown-outs products. Below are the algorithms which will be implemented with the present dataset,

Support Vector Machine: Support Vector Machine provides decision planes that give decision boundaries. A decision plane bifurcates a set of objects having separate class members. The dataset is classified by SVM via constructing hyper-planes in a multiple dimensional space that separates different labels of the class. SVM supports classification as well as regression and has capability to handle multiple categorical and continuous data variables. With researches conducted with SVM over ecommerce, this algorithm had responded well with good results in terms of accuracy and efficiency Kartal et al. (2016) Thus, SVM would prove to be better suited when implemented on warehouse management data.

Nave Bayes: The Naive Bayesian classifier is based on Bayes theorem which is easy to build and has no complicated parameter estimates which makes the algorithm useful for large datasets. With the help of Bayes theorem, it is possible to provide a way of calculating the posterior probability, $P(c\|x)$, from P(c), P(x), and likelihood probability $P(x\|c)$. Naive Bayes classifier naively assumes that the value of a predictor (x) on a class (c) is independent of the values of other predictors Sayad (2019).

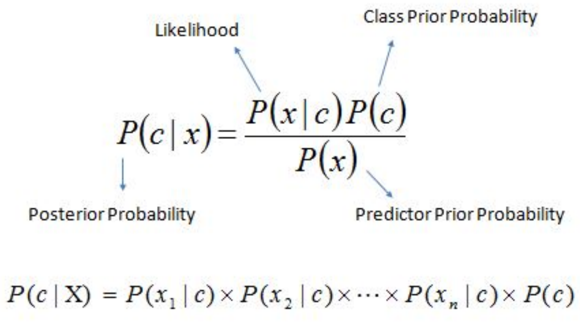

Figure 3: Naive Bayesian, Sayad (2019)

$P(c\|x)$ is the probability of class c when attribute x is given

P(c) is the probability of class
$P(x\|c)$ is the probability of attribute x given class c.
P(x) is the prior probability of attribute x.

This research paper would have an implementation of Nave Bayes because it performs exceptionally well with large datasets and has independent assumptions in predictors.

Decision Tree: The decision tree has influenced a wide range of machine learning covering both classification and regression. A decision tree is a flowchart in which each internal node depicts a test on an attribute, each branch becomes the outcome of the test, and each leaf becomes a class label Brid (2018). Tree-based learning algorithms are mostly used supervised learning algorithms. They also provide predictive models with a high level of accuracy, stability and provides easy to understand interpretation. They are adaptable and can solve problems at hand. This research will be benefited with the implementation of Decision Tree, as it can prove unexpectedly well which gives rise to good accuracy and efficient classification algorithm for understanding the patterns in warehouse optimization dataset.

Random Forest: Random forest consists of large number of unique decision trees acting as an ensemble. Each tree gives out the class prediction and the class with most prediction becomes final model. The concept behind the random forest is very powerful and simple which is based on crowd wisdom. When looked as whole, many non-correlated models operate as a committee to overpower a single constituent model out of all the models. The dataset of warehouse can prove to be very volatile when it comes to training of data, as a slight change in training set might result into different output. This can easily be averted with the help of a concept called bootstrap aggregation where random forest allows each tree to be randomly sampled from the dataset with replacement, resulting into separate and unique trees. Thus, it brings us to inference that this algorithm can lead us to more accuracy while dealing with a warehouse dataset Yiu (2018).

Artificial Neural Network: ANNs are computational algorithms that simulate the behavior of neurons which are biological entities in a living being. A neural network consists of input and output layer and in most cases a hidden layer. It consists of units that are helpful in converting the input to a usable feedback product which is then provided to the output layer. It uses the technique of back-propagation which allows fine-tuning of weights of neural networks based on loss received in previous iteration. The results of the research showed that ANN can be used to manage a company warehouse's order cycle. Optimal neural networks show suitable results for subsequent prediction of the number of items to be ordered and for achieving reduced inventory purchase while keeping costs down. This application of ANN to supply chain management where many organized neural networks were in place to optimize the companys inventory, provided a good amount of motivation for applying this out-of-box algorithm to the warehouse dataset Sustrova (2016).

3.5 Application of Efficient Pricing Strategies for Warehouse Optimization

This is an additional part of research which is very important in terms of achieving the objective of automating storage clearance in a warehouse. When a comparison output is achieved, the same output is applied with different pricing strategies which gives rise to optimized market sale and would help in stock clearance before the wear-off of the specific product. Literature review has discussed many strategies of implementing product pricing which produces a differential formula. This formula then can be programmatically applied to the output received from predictive models, which further will help organizations to handle products more profoundly and accurately.

3.6 Evaluation

The main purpose of evaluation is to verify whether the trained algorithms have provided legitimate output to the user It is very important to apply evaluation strategies on the data to be pre-processed, most popular being the simple split and k-fold cross validation strategies. The simple split or hold-out strategy focuses on the division of data into training and testing sets. The training set is used to train the model and the testing set is used to test the trained model to verify how well the model performs for an unseen data. The common split that is performed during the hold-out or simple split strategy is 80% training data and the 20% of testing data. Hold-out strategy works well with fully independent data which in turn decreases computational cost. For small datasets, the performance of the algorithm outputs high variance which could be disadvantageous for applying hold-out strategy Robert Kelley (2017).

3.6.1 K-Fold Cross Validation

The K-Fold cross validation divides the dataset into K number of sets or subsections and then trains the model with remaining K-1 subsets. This iteration is similarly carried out for all remaining k-1 subsets leaving behind the one to test the model. For the warehouse data, this methodology would prove beneficial as the training data is prone to less variation as it uses entire dataset for training the model. This method has a high computational cost as the model needs to be trained K times at various validation steps. Diagrammatically it can be depicted as below,

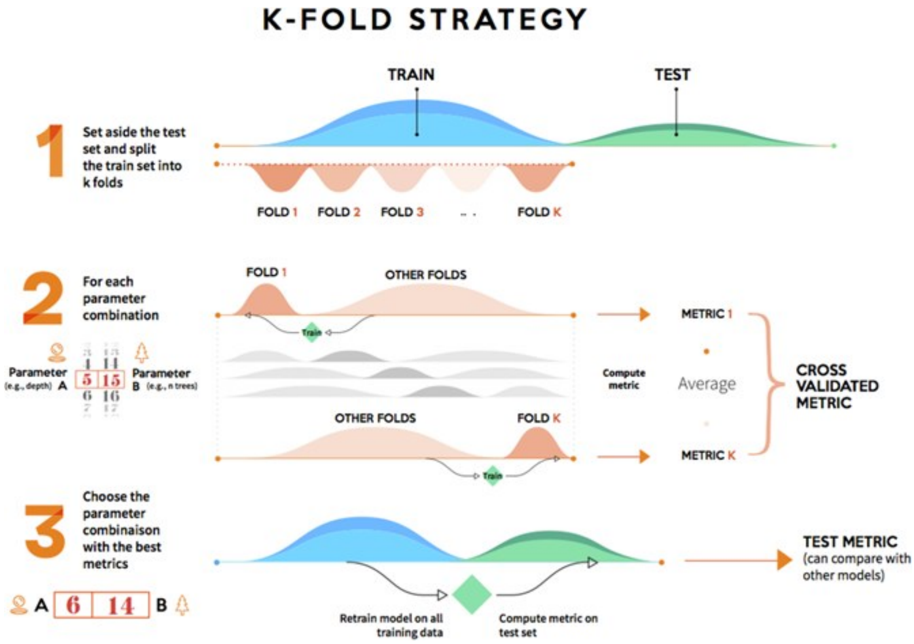

Figure 4: K-Fold Cross Validation, Robert Kelley (2017)

3.6.2 R-Square Criterion

Statisticians are of the view that regression model fits the data perfectly, if the difference between the observed and predicted values is small and unbiased i.e. fit values should not be very high or very low in the observation space. R-Squared criterion evaluates the outlying data points from the fit regression line. For a dataset, higher the R-Square value, smaller will be the differences between the observed and fit values. R-squared criterion can be formulated as below,

$$(R^2) = (Varianceexplainedbythemodel)/TotalVariance \tag{2}$$

The value of R^2 lies between 0% and 100%, where 0 represents that the model does not explain variance in response, whereas 100 represents variation in response variable around its mean. This research paper will specifically mention the usage of this evaluation procedure to understand the accuracy with which the model was predicted and provides a comparison measure for the other competitive evaluation procedures Frost (2018).

3.6.3 Mean absolute error

The MAE measures the mean magnitude of the faults in a set of forecasts, without considering their direction. It measures the accuracy and efficiency of continuous variables. MAE is an average over the sample of the absolute values of the differences between forecast and the corresponding observation. The MAE is a score which tells us that all the individual differences are weighted on equal platform in the average eumetrain (2018). There is a comparison done between the employment of root mean square and mean absolute error to verify the evaluation processes, but after experimental proofs it has been found that, mean absolute error outperforms RME. Thus, it would prove beneficial to imply mean absolute error evaluation strategy to scrutinize the output received from the ecommerce dataset Chai and Draxler (2014).

3.6.4 Receiver Operating Characteristics (ROC) Curve

Receiver operating characteristics (ROC) analysis is a method for evaluating, comparing and selecting classiers based on their performance Majnik and Bosnić (2013). It demonstrates the following things

- Trade-off between sensitivity and specificity (increase in sensitivity will cause decrease in specificity).
- Trade-off between sensitivity and specificity (increase in sensitivity will cause decrease in specificity).
- Nearer the curve comes to the 450 diagonal of the ROC space, the less accurate the classifier.

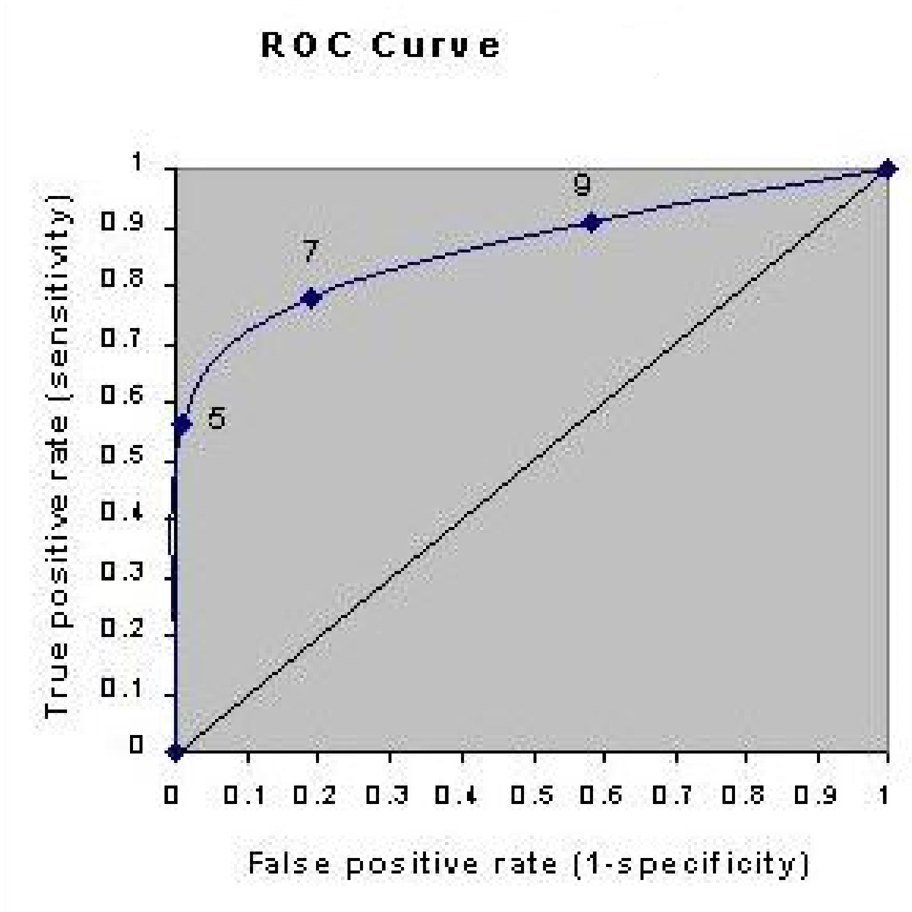

Figure 5: ROC Curve, Tape (2017)

3.6.5 Project Plan

The below Gantt chart gives an overview of project planning which will be carried out in the next semester.

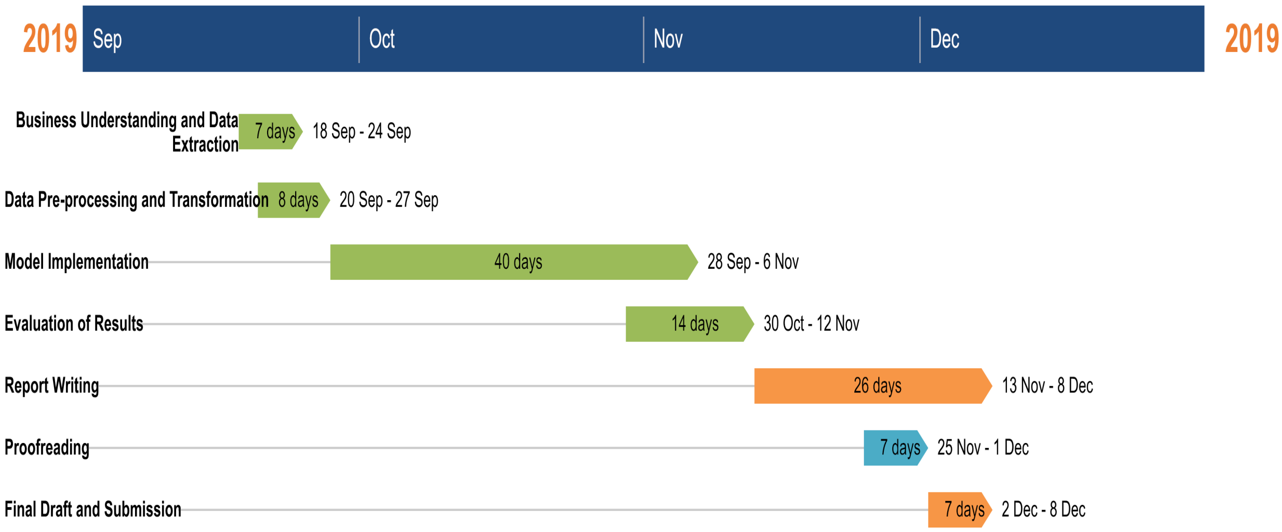

Figure 6: Project Plan

3.7 Deployment

The deployment is a final phase of CRISP-DM methodology where the best practices are been followed with the help of tasks like planning deployment, monitoring and maintenance, reporting the resultant and reviewing the final outputs. This phase cares least about how well the research is conducted or how perfectly the models have been fit till they do not contribute towards the growth of the business. The gist to be understood from it is the closure is an important phase of any task which can either be a research or a practical. The dataset has been exploited carefully to understand the sequences and the final product adds value to the organizational growth.

4 Summary

Numerous studies are conducted with regards to the warehouse optimization, management and ecommerce domain which just started off with a question of improving the business of warehouse optimization. It has come to a standstill that manual processes conducted based on heuristic judgment can provide better results provided the industry under the deal is a small scale. But when it comes to dealing with large scale organizations, automation comes to rescue. This research paper has discussed different ways of predicting the duration in which the products fall into blow out sequence and also gives insight on how those products can be well managed by applying different pricing strategies. Thus, a warehouse is always ready with products that are a market ready rather than products that turn dead and become heavy. A systematic plan has been chalked out with best practices of business to achieve objective within timelines without hindering the quality.

References

Ai-min, D. and Jia, C. (2011). Research on slotting optimization in automated warehouse of pharmaceutical logistics center, *2011 International Conference on Management Science & Engineering 18th Annual Conference Proceedings*, IEEE, pp. 135–139.

Alfian, G., Syafrudin, M., Yoon, B. and Rhee, J. (2019). False positive rfid detection using classification models, *Applied Sciences* **9**(6): 1154.

Ancarani, F. and Shankar, V. (2004). Price levels and price dispersion within and across multiple retailer types: Further evidence and extension, *Journal of the academy of marketing Science* **32**(2): 176.

Ang, M., Lim, Y. F. and Sim, M. (2012). Robust storage assignment in unit-load warehouses, *Management Science* **58**(11): 2114–2130.

Bloomenthal, A. (n.d.). What you should know about electronic commerce.
URL: *https://www.investopedia.com/terms/e/ecommerce.asp*

Bohanec, M., Robnik-Šikonja, M. and Borštnar, M. K. (2017). Organizational learning supported by machine learning models coupled with general explanation methods: A case of b2b sales forecasting, *Organizacija* **50**(3): 217–233.

Brid, R. S. (2018). Decision tree.
URL: *https://medium.com/greyatom/decision-trees-a-simple-way-to-visualize-a-decision-dc506a403aeb*

Chai, T. and Draxler, R. R. (2014). Root mean square error (rmse) or mean absolute error (mae)?–arguments against avoiding rmse in the literature, *Geoscientific model development* **7**(3): 1247–1250.

Chan, F. T. and Chan, H. K. (2011). Improving the productivity of order picking of a manual-pick and multi-level rack distribution warehouse through the implementation of class-based storage, *Expert systems with applicatiostockns* **38**(3): 2686–2700.

Chen, I.-F. and Lu, C.-J. (2017). Sales forecasting by combining clustering and machine-learning techniques for computer retailing, *Neural Computing and Applications* **28**(9): 2633–2647.

de Santis, R. B., de Aguiar, E. P. and Goliatt, L. (2017). Predicting material backorders in inventory management using machine learning, *2017 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, IEEE, pp. 1–6.

eumetrain (2018). Mean absolute error (mae) and root mean squared error (rmse).
URL: *http://www.eumetrain.org/data/4/451/english/msg/ver*$_c$*ont*$_v$*ar/uos3/uos3*$_k$*o1.htm*

Frost, J. (2018). How to interpret r-squared in regression analysis.
URL: *https://statisticsbyjim.com/regression/interpret-r-squared-regression/*

Hissibini, C. (2019). Data mining: What it is and why it matters.
URL: *https://histechup.com/data-mining-what-it-is-and-why-it-matters/*

Hong, S., Johnson, A. L. and Peters, B. A. (2012). Large-scale order batching in parallel-aisle picking systems, *IIE Transactions* **44**(2): 88–106.

Inprasit, T. and Tanachutiwat, S. (2018). Reordering point determination using machine learning technique for inventory management, *2018 International Conference on Engineering, Applied Sciences, and Technology (ICEAST)*, IEEE, pp. 1–4.

Karásek, J. (2013). An overview of warehouse optimization, *International journal of advances in telecommunications, electrotechnics, signals and systems* **2**(3): 111–117.

Kartal, H., Oztekin, A., Gunasekaran, A. and Cebi, F. (2016). An integrated decision analytic framework of machine learning with multi-criteria decision making for multi-attribute inventory classification, *Computers & Industrial Engineering* **101**: 599–613.

Keller, T., Thiesse, F. and Fleisch, E. (2015). Classification models for rfid-based real-time detection of process events in the supply chain: an empirical study, *ACM Transactions on Management Information Systems (TMIS)* **5**(4): 25.

Kim, T. Y. (2018). Improving warehouse responsiveness by job priority management: A european distribution centre field study, *Computers & Industrial Engineering* .

Kłodawski, M., Jacyna, M., Lewczuk, K. and Wasiak, M. (2017). The issues of selection warehouse process strategies, *Procedia Engineering* **187**: 451–457.

Krauth, E., Moonen, H., Popova, V. and Schut, M. (2005). Performance indicators in logistics service provision and warehouse management–a literature review and framework, *Euroma international conference*, pp. 19–22.

Kumar, D. (2019). Introduction to data preprocessing in machine learning.
URL: *https://towardsdatascience.com/introduction-to-data-preprocessing-in-machine-learning-a9fa83a5dc9d*

Lam, C. H., Choy, K. L., Ho, G. T. and Chung, S. H. (2012). A hybrid case-ga-based decision support model for warehouse operation in fulfilling cross-border orders, *Expert Systems with Applications* **39**(8): 7015–7028.

Laudon, K. C. and Laudon, J. P. (2015). E-commerce: Digital markets, digital goods, *in* K. C. Laudon and J. P. Laudon (eds), *Management Information Systems: Managing the Digital Firm Plus MyMISLab with Pearson eText–Access Card Package*, Prentice Hall Press, chapter 10, pp. 415–425.

Ma, H., Wang, Y. and Wang, K. (2018). Automatic detection of false positive rfid readings using machine learning algorithms, *Expert Systems with Applications* **91**: 442–451.

Macro V, D. M. (2017). What is machine learning? a definition.
URL: *https://www.expertsystem.com/machine-learning-definition/*

Majnik, M. and Bosnić, Z. (2013). Roc analysis of classifiers in machine learning: A survey, *Intelligent data analysis* **17**(3): 531–558.

Mccrea, B. (2019). Six warehouse management trends to watch in 2019, *Logistics management (Highlands Ranch, Colo.: 2002)* .

Minga, L. M., Feng, Y.-Q. and Li, Y.-J. (2003). Dynamic pricing: ecommerce-oriented price setting algorithm, *Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No. 03EX693)*, Vol. 2, IEEE, pp. 893–898.

Nalić, J. and Švraka, A. (2018). Importance of data pre-processing in credit scoring models based on data mining approaches, *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, IEEE, pp. 1046–1051.

Reyes, J., Solano-Charris, E. and Montoya-Torres, J. (2019). The storage location assignment problem: A literature review, *International Journal of Industrial Engineering Computations* **10**(2): 199–224.

Robert Kelley, D. (2017). Making predictive models robust: Holdout vs cross-validation.
URL: *https://www.kdnuggets.com/2017/08/dataiku-predictive-model-holdout-cross-validation.html*

Sayad, D. S. (2019). Naive bayesian.
URL: *https://www.saedsayad.com/naive$_b$ayesian.htm*

Sustrova, T. (2016). An artificial neural network model for a wholesale company's order-cycle management.
URL: *https://journals.sagepub.com/doi/full/10.5772/63727*

Tape, T. G. (2017). Plotting and intrepretating an roc curve.
URL: *http://gim.unmc.edu/dxtests/roc2.htm*

Van den Berg, J. P. and Zijm, W. H. (1999). Models for warehouse management: Classification and examples, *International journal of production economics* **59**(1-3): 519–528.

Wirth, R. and Hipp, J. (2000). Crisp-dm: Towards a standard process model for data mining, *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, Citeseer, pp. 29–39.

Yiu, T. (2018). Understanding random forest.
URL: *https://towardsdatascience.com/understanding-random-forest-58381e0602d2*

Zhou, W., Piramuthu, S., Chu, F. and Chu, C. (2017). Rfid-enabled flexible warehousing, *Decision Support Systems* **98**: 99 – 112.
URL: *http://www.sciencedirect.com/science/article/pii/S0167923617300805*

Zunic, E., Hasic, H., Hodzic, K., Delalic, S. and Besirevic, A. (2018). Predictive analysis based approach for optimal warehouse product positioning, *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, IEEE, pp. 0950–0954.

Zunic, E., Hodzic, K., Hasic, H., Skrobo, R., Besirevic, A. and Donko, D. (2017). Application of advanced analysis and predictive algorithm for warehouse picking zone capacity and content prediction, *2017 XXVI International Conference on Information, Communication and Automation Technologies (ICAT)*, IEEE, pp. 1–6.