International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072

Demand Forecasting using Machine Learning Methods

Akash Singh, Nikhil Singh

Dept. of Computer Science and Engineering, RV College of Engineering, Bangalore, India
Data Science- Lead, o9 Solutions Inc

Abstract – Demand Forecasting is a technical term to depict the Demand Planning for any product which is associated or subjected to sale in a live market. It plays a vital role to organise and plan the supply chain management of a company. Forecasting means to predict the future performance of the product depending on the historical data which is technically referred to the term Actuals throughout this project. The product is uniquely identified using a term SKU which stands for stock keeping unit, it is the aggregation of customer group, material and region and acts as any other barcode while the data is being identified and treated uniquely. The trivial data which has being worked upon is received from a petrochemical giant and has been treated optimally to obtain the forecast which corelates with the historical sale of the products.

Key Words: SKU, SCM, SSIS, IBPL, SMA, SES, DES, TES, STLF, TBATS, ARMA, ARIMA, TBATS

1. INTRODUCTION

Supply Chain Management is commonly referred to the use of processes and techniques used to maintain the continuous flow of goods and services over different areas for different time intervals. Making sure that no demand goes unmet or delayed. Every industry heavily relies on their Supply chain management methodology to keep their daily processes going. They must deal with huge datasets to keep track of everything. Hence, the client needs tools which they can use efficiently to their advantage to maintain their data base and view the data in the required format to make necessary decision for their company.

Based on the company's structure, data is obtained regarding the various sectors and there planning processes and a plan is prepared as to how the desired results will be met. This is generally followed by cleaning of data before uploading them making sure there are no erroneous data which might cause problems later. This ranges from creating groups and relations between the data, writing logics to perform calculations also using an in built tool for performing the demand meeting calculations to make sure most of the demands are met on time and also tracing which inventory is linked to which demand which adds another data to the pre-existing data which is then used to predict trends, a AI functionality of the platform.

2. LITREATURE SURVEY

Punam Khobragade et al. [1] discusses about the mechanisms used for inventory management and the losses incurred due to bad inventory management. She talks about the use automation in generating invoices, printing invoice and viewing invoices where we can get all the information about the supplier and the clients at one location and also the information about where the part or the good is located at. She goes on further to talk about the optimum level of inventory to be ordered at a point all which can be achieved with the help of a desktop application.

Mohamed Awwad et al. [2] talks about what Big Data is and how its increase will affect the industries. He talks about the role where this big data plays a huge including the sources of it, one of it being the demand planning industry. He goes on to talk about the amount of data generated and how these companies reply on data to for decision making processes. He talks about these will increase the sales of the company and reduces loses. He suggests the use of data for better forecast prediction and business management.

Aziz Muysinaliyev et al. [3] performs a literature review of many papers that were written on demand planning management and he discusses about how each of them provide insights as to how demand planning proves to be a lead differentiator in business and how each company spends a lot of resources in making them strong and error free. How they adopt practices like value-based demand planning practices instead of quantity based. The paper also states that most of the papers on this topic are similar in nature and provide almost a similar result.

Shilpa Parkhi et al. [4] discusses about the concept of demand planning management has evolved over time and how researchers describe them. They go on to talk about the dimensions in demand planning management and organizational theories such as the Institutional theory, organisation theory and network theory. They provide key dissimilarities between demand planning management, operations management and also what differentiates demand planning from value chain.

Lv Yanchang et al. [5] discusses about the benefits of cross docking and its effect on reducing the warehousing cost. She discusses about the different computation methods and algorithms that can be used to for achieving cross docking efficiently in docking houses. They studied the effect of the cross-dock shape, the size of the goods to be segregated and the size of the warehouse, also the use of conveyor belt was suggested over others such as cross shape or rectangle.

John J. Bartholdi, III et al. [6] in their paper "Best Shape for a cross dock discuss the various factors that effect the shape of cross dock such as the size of the good, the man power, machinery and also the relative area at loading and unloading facilities available for receiving goods. In their paper they concluded that as size of the goods increases, the most labor effective shapes for a cross dock are I, T and X respectively. I - shape is the most effective shape for docks with lesser than 150 doors in the warehouse, Also the size at which T is preferred over an I- shape and X over an T depends on the receiving number of doors and the concentration of flow. T is preferred over smaller ranges of door when fraction of doors devoted to receiving is higher.

Huang Yijun and Wang Jingjing [7] discusses about the variance in safety stock level with different parameters such as the inventory cost, company service level, shortage cost, uncertainty of demand. They concluded that if the lead time is reduced by a factor of 'x' then even the safety stock requirements can reduce by a factor of 'x'. In this paper different levels of customer service levels were given for various market demand and lead time to simulate the changes of total cost. When the service rates are lower than there is a relatively lower rate of increase in safety stock whereas if the service levels are very large then the rate of increase in safety stock is very large.

Yiping CAI et al. [8] tells that there is a total of four different ways to reduce the safety stock levels, which are reducing lead time, reducing lead time variability, reducing demand uncertainty, and reducing availability uncertainty. The paper suggests in order to do so we can plan to improve on logistical processes being undertaken, Adoption of postponement strategy, Employment of quick response QR and Collaborative Planning, Forecasting and replenishment CPFR, and lastly by making use of information technology (IT) by sing big data science and modern technologies such as AI and ML.

Nazar Sohail et al. [9] discusses a case study of a Steel manufacturing industry (small scaled industry), the relationship between the inventory management and the company's performance was determined based on its return on investment category. The research showed that the company had few issues in its inventory management system, also there was a lot of problem related to unorganized workers. They concluded that there were problems in various departments like marketing, planning and procurement and he suggested that company invest more in better inventory management techniques in a better manner.

Jaein Choi et al. [10] developed a mathematical model for various aspects such as operational cost, customer service level and investment cost. They studied problems in designing a distribution network in a demand planning that involves determining optimal DC locations. They developed a mixed integer nonlinear programming model. It was considering the cost and trying to minimize the total cost based on the cost inputs for demand shortage and inventory storage cost.

3. PROPOSED SYSTEM

The proposed system is unique compared to the existing forecasting methodology adopted in the industry. The data is initially pre-processed, where all the unwanted anomaly is cleaned and now it is ready to be worked upon. Products are uniquely segmented based on its mathematical attributes derived by analysing the historical sales data. Forecasting is then performed using various time series analysis algorithm uniquely allocated to each segment over which the products were divided. The forecasting is just not performed casually but every product is uniquely subjected to various algorithms belonging to its segment and the best fit is calculated by comparing with the closest matching forecasted data to the testing set and further the algorithm giving the best result is selected. The accuracy analysis is being

3.1 Data Pre-Processing

Data in today world is very important as most of the supply chain management processes work on data. Once a client's requirements are finalized, he provides the required data in a format to o9 solutions team. This is done to maintain homogeneity of data and its structure and if the data is not in a format then it becomes difficult to manipulate it and analyse it. Even when the data is received in a format it still requires cleaning up, this can be due to several reasons such as null values, dummy values, multiple entries of the same data, inconsistent data and erroneous data. To perform all this, we use an SSIS tool using SQL which is a data base query language used to manipulate data as preferred. This process is called as data integration. This data can be either got directly from the client or it can also be obtained from the already existing systems of SAP maintained by the client,, hence the platform needs to have functionality to read from all the different sources. This obtained data is then uploaded onto the o9 platform.

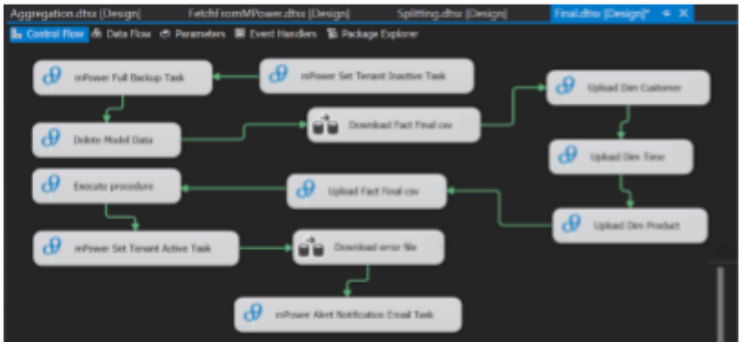

Figure -1: Master Package for Data Pre-Processing

3.2 Segmentation

• Segment allocation– All the data for a various intersection cannot be worked upon. To get an appropriate forecast, level of forecast needs to be decided. Initially segments is allocated to the data which talks a lot about its characteristics and attributes.

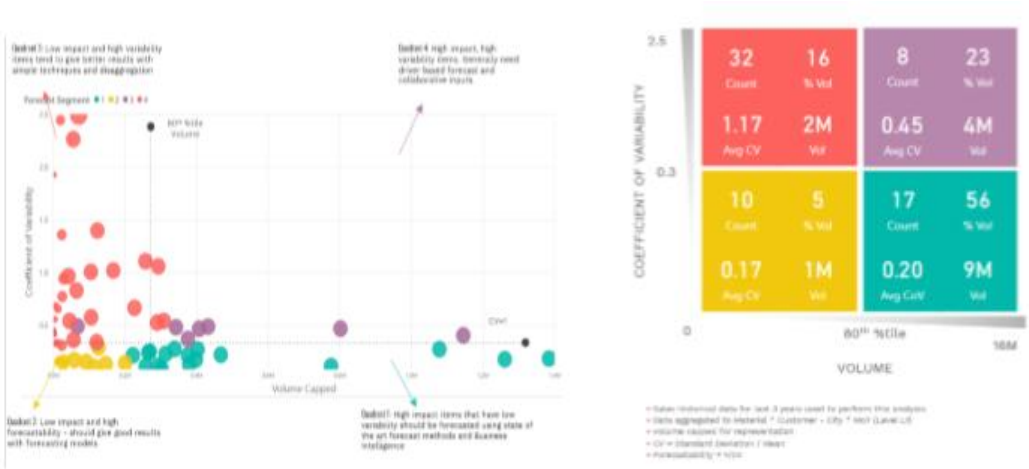

Figure -2: Intermidiate Visualisation

• Attribute Generation- Mathematical attributes about the data is being generated along with the PowerBI input file and the forecast input file in every iteration.

• Deciding the appropriate forecast level– The quad plot and the scatter plot gives an vivid idea of the most appropriate intersection of the attributes where the forecast can be conducted in order to attain a high level of accuracy. It is an iterative process and accurate level is obtained over several iterations.

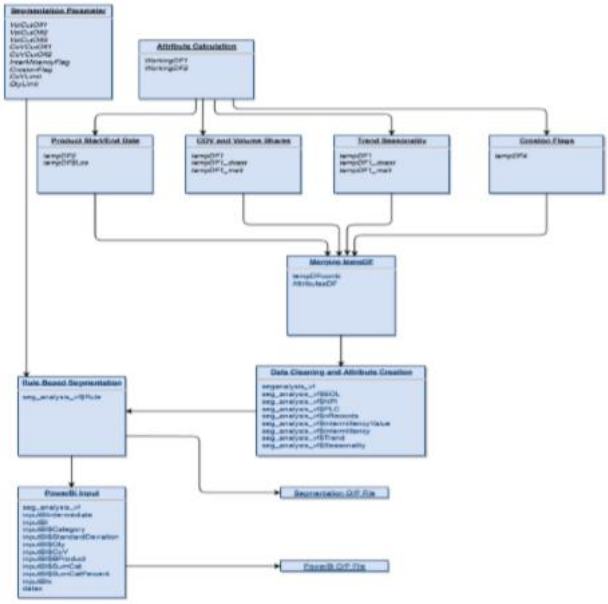

Figure -3: Segmentation Flow

3.3 Forecasting

• Rule allocation–To get an appropriate forecast, level of forecast needs to be decided. Initially rule is allocated to the data which further decides which time series algorithm will be used to calculate the forecast.

• Algorithm allocation- From a pool of variety of algorithms the best fit is allocated to a certain SKU depending on the rules allocated to it. This is an optimal case as the data is not generalized and is treated according to its attributes.

Segment #	Segment Rule	Stat Segment Algorithm List
1	Volume Segment = "A" or "B" & CoV <= 0.3	ARIMA, TES, TES Damped, DES, DES Damped, TBATS, NNET, STLF
2	Volume Segment = "A" or "B" & CoV > 0.3	ARIMA, TES, TES Damped, DES, DES Damped, TBATS, NNET, STLF
3	Volume Segment = "C" or "D" & CoV <= 0.6	TES, TES Damped, DES, DES Damped, SES, Seaive, Moving Average
4	Volume Segment "C" or "D" & CoV > 0.6	TES, TES Damped, DES, DES Damped, SES, Moving Average
5	Intermittency == "Yes"	Croston, SES, Moving Average, Naive
6	PLC == "NPI"	SES, Naive, Moving Average, DES Damped, TES Damped

Table -4: Algorithm Allocation

3.4 Accuracy Analysis

Accuracy analysis is performed to understand the improvement of the of the forecast that it being performed. It is an iterative process and in every step of iteration the accuracy is improved using qualitative analysis methods. The attributes of the low accuracy products are studied and a method from the pool of algorithms is suggested to be force fitted on the product. The sole goal of qualitative analysis is to increase the accuracy. Weighted MAPE and biased is calculated and is being improved over every iteration and therefore is compared against the benchmark forecast and a final optimized forecast is derived.

3.5 Disaggregation

There are multiple approaches for disaggregation such as top-down approach, bottom up approach and middle out approach. In this project the bottom up approach is followed where the higher-level data is disaggregated using the lower-level data. The forecast proportions, forecast disaggregation and the forecast for both lower and higher-level is being obtained. The disaggregation is done with minimal noise and inhibits most of the demand patterns. The forecast proportion is much more

forward looking and shows variation in a long run which is necessary to obtain an accurate forecast whereas the historical proportion is used as a benchmark and is very easy to obtain but shows less variations for future data.

4. RESULTS

The experiments were conducted on the client's data by using various methods such as force-fitting the time series algorithms to check if the level, trend, and seasonality were caught by the manual models or not. Comparison was done by subjecting the data to various of these combinations and the best ones among those were visualized and checked.

Further there were two types of data which could be worked upon to improve the accuracy that was historical proportion and forecast proportions. The historical proportion seemed very stable much showed less variety in the future predictions and was gracing by the mean of the actual data. Whereas the forecast proportion was not very stable but decided to be worked upon as it showed variety in the prediction and was accurately catching the level, tend and seasonality.

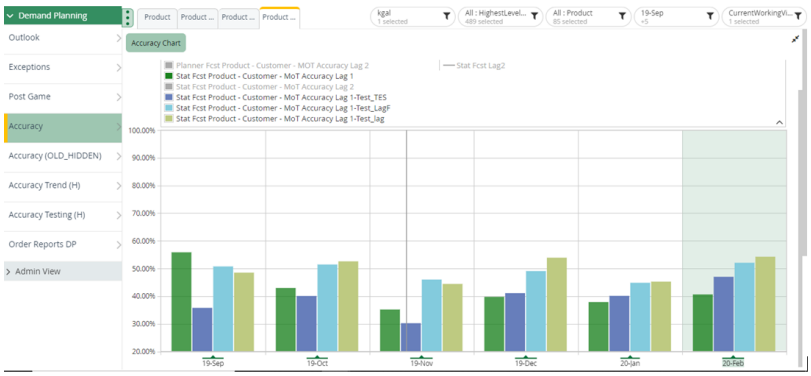

Figure -5: Graph on M-Power

5. FUTURE SCOPE

This time series problem can be converted into a regression problem and multiple machine learning algorithms can be subjected to it for example a deep learning or a boosting algorithm to improve and boost the accuracy. Machine learning grids can be used to automate the hyper parameter tuning of the forecast proportions which is very repetitive. The model for each time series algorithm can be chosen by looking at the behavior of the historical data. Similarly, the weightage can be given to the level, trend, and seasonality according to the nature of the forecast shown in the graph.

6. CONCLUSIONS

Overall, this is a very intelligent forecast approach to predict the demand on such a large scale for huge number of products. Applying the machine learning algorithms initially would give a very inaccurate forecast as the benchmarks are not set. This project tends to prepare the data by calculating the forecast by obtaining the best fit using various time series analysis method. The forecast calculation is only performed after the SKU's are grouped and segmented using the mathematical attributes of the historical data such a volume, co-variance etc. After the initial forecast is obtained it is being fine-tuned by performing Accuracy Analysis and Hyper Parameterized Tuning. The obtained result is finally visualized the m-power platform and is client ready.

REFERENCES

[1] Punam Khobragade, Roshni Selokar, Rina Maraskolhe Prof.Manjusha Talmale, "Research paper on Inventory management system ", Department of Computer Science and Engineering, G.N.I.T., Kalmeshwar Road , Nagpur, Maharashtra, India International Research Journal of Engineering and Technology (IRJET), Vol 5, Issue 4, 2018

[2] Mohamed Awwad, Pranav Kulkarni, Rachit Bapna and Aniket Marathe, "Big Data Analytics in Supply Chain: A Literature Review", Proceedings of the International Conference on Industrial Engineering and Operations Management Washington DC, USA, September 27-29, 2018

[3] Aziz Muysinaliyev, Sherzod Aktamov, "Supply chain management concepts: literature review", IOSR Journal of Business and Management (IOSR-JBM), Volume 5, issue 6, Jan 2014

[4] Shilpa Parkhi, Sourabh Joshi, Shubham Gupta, Mridul Sharma, " A Study of Evolution and Future of Supply Chain Management ", AIMS International, Volume 9, Number 2, May 2015

[5] Lv Yanchang, Dong Min, "Discussion of Cross Docking Design in modern Logistical System", World Congress on Computer Science and Information Engineering, 2009.

[6] John J. Bartholdi - III, Kevin R Gue, "The Best Shape for a Crossdock", , Jornal of Transporation Science, Volume 38, No 2, May 2004.

[7] Huang Yijun, Wang Jingjing, "Study on Safety Stock Control Based on System Dynamics", 11th International Conference on Service Systems and Service Management (ICSSSM), 2016.

[8] Yiping CAI et al. "Safety Stock Management Based on Lead Time Optimization", International Conference on Electric Technology and Civil Engineering (ICETCE), 2011.

[9] Nazar Sohail, Tariq Hussain Sheik, "A Study of Inventory Management System Case Study", Jour of Adv Research in Dynamical & Control Systems, Vol. 10, 10-Special Issue, 2018.

[10] Jaein Choi, Hokyung Lee, Soon-Ki Heo, and Jongku Lee, "A Mathematical Programming for Supply Chain Network Design", Supply Chain Planning Team, Process Technology R&D, LG Chem. Ltd Research Park, Deajon, SICE-ICASE International Joint Conference, 2006.

BIOGRAPHIES

Akash Singh
B.E. CSE
RVCE, Bangalore
Data Scientist, o9 Solutions Inc

Nikhil Singh
B.E. ECE, M.S. Economics
BITS, Pilani
Data Science-Lead, o9 Solutions Inc