CAPSTONE

PROJECT

REPORT

Kerala University of Digital Sciences, Innovation & Technology

(Digital University Kerala)

AJMAL M S (213208)

ACKNOWLEDGMENT

I'm conveying my hearty appreciation, respect to my professor and mentor, Dr. T.K Manoj Kumar, Associate professor, Digital University of Kerala, Trivandrum, for his motivated learning and irreplaceable support. His thought and advisement always made me learn without boundaries, which helped me complete this project.

I would like to thank Prof. Saji Gopinath for providing me with an avant-garde atmosphere, DUK. The effect of exposure, mentoring, facilities that you providing gave me a potential to enjoy the education, technology and the people more. This filled my gap of social importance of technology. It helped me to mould my project into a valuable idea.

I must thank Anand Sir, who helped me a lot in allocating the time needed for this project by understanding my situations. That helped me a lot in cross checking my work. I would also like to thank my friends, family for their valuable presence, encouragement, and support during this project's development phase.

CONTENTS

ABSTRACT

Around 11 billion tons of products are supplied around the world according to HIS Markit. It increases year by year. In fact, it increases day to day. Just like real estate is booming according to the demand rate and population, supply chain industry is too chasing to be with the world.

Business organizations face day to day competitions around their organization. A need for business is a need of a customer, so what and how customer like to buy next will be the target of every organization. This understanding can help us preventing over supply, under, supply and logistics cost. For finding a customer need we need some data management like big data.

nowadays, businesses adopt ever-increasing precision marketing efforts to remain competitive and to maintain or grow their margin of profit.

As such, forecasting models have been widely applied in precision marketing to understand and fulfill customer needs and expectations.

Supply chain management (SCM) focuses on customer demands, flow of supply, entity-relations, goods management.

Sometimes there will be demand uncertainties, organizational risks, supplies transportation and leas times, so that a data driven approach is needed.

Supply chain management refers to designing and controlling the system of organizations, people, information and resources that go into producing and moving a product or service from supplier to customer. Or differently put, managing the process from design to re-purposing

It involves a moving life cycle of Supply chain strategy, Supply chain planning, Supply chain demands, Behaviour of customers, Recent trends, demand forecasting which brings a lot of saving and productive economic growth of an organization. for. This, includes also marketing, in particular the demand planning and forecasting.

Nowadays, Forth Industrial Revolution technologies are opening up new possibilities that provide advantages for businesses, people and the environment. We can see a lot of benefits with the platform economy – for example, digital platforms that facilitate collaboration between supply chain operators for truck capacity and combined volumes, and the reuse, remake and recycle in the circular economy. Here my implementation is about customer demands in a certain demographic location. For this, a standard data set from an online or retail store is crucial. The demands are actually analysed from core market areas. These demands are taken into account for forecasting the sales. Forecasting the sales is important as without it, the economy, climate, man power, productivity, resources are all will be affected.
`

I used various forecasting models profound in the world of python itself. One is Multilayer Perceptrons or MLP. MLPs for short, can be used to model univariate time series forecasting problems. Univariate time series are a dataset comprised of a single series of observations

with a temporal ordering. Here a model is required to learn from the series of past observations to predict the next or future value in a sequence. The other model I designed is called PuLp modelling. PuLp modelling is a framework in python for Linear and Integer

programming problems. This model is maintained by Computational Infrastructure for Operations Research Foundation (COIN-OR). For PuLp, the data sets used are, Manufacturing Facility Fixed Costs, Shipping Variable costs, Manufacturing Capacity by site, Customers demand per market. For MLP, the data set used is a sample data set having columns date, store, item and sales. During the analysis of codes, ensemble methods boosted the forecasting. In classification models, Linear SVM performed well. For prediction, the Random Forest algorithm performed well. My aim in this project is to make supply chain efficient to avoid business delays, financial lose, to improve marketing, and to observe the trend and customer behaviour association.

INTRODUCTION

BACKGROUND

Nowadays, business became global and the network associated through out the world. Supply chain helps all of the business output possible. For example, in India we cannot make silicon integrated chips without the raw material from Samsung, Japan. If other administration cannot bear a supply chain system, we cannot make electronic gadgets in India. Until another resource exploration, we need to wait. This effects almost all industries from electronics to health care elastomers [1]. The biggest challenge, a business organization now facing is a good forecasting about how much to supply, when to supply, what to supply, and whom to supply. These are all related to the Scarcity Principle. The scarcity principle is an economic theory that explains the price relationship between dynamic supply and demand**.** According to the scarcity principle, the price of a good, which has low supply and high demand, rises to meet the expected demand [2]. Everywhere, demand meets supply. For instance, in Bangalore, consider a cluster of people are buying energy drink every evening. Like wise in Mumbai a group of population are interested in brand Puma. Finally in Indore a cluster of population are more interested in Brown bread rather than White. From the above demand, an energy drink company should be aware of how much the sale will increase. Will it depend on season? Will it depend on certain tourists? Will it depend on cost? Then we want to make a data and find on average how much is the demand. If the average is 1000 sold outs, for next time we cannot exactly believe that digit. We should use an exact historical data of a store to conclude the sales. So, after the model accuracy test, the data is followed about how to manufacture and deliver.

For Mumbai, the Puma company should invest more or introduce according to the customer expectations. Then for Indore, more wheat raw material should be imported/exported than the Maida. If we have a demand for a particular product on all those above said cities, still we can arrange them hierarchically according to demand using machine learning models and good data sets.

The shipping cost can also be controlled after following this method. The forecasting methods in supply chain management also helps the production and manufacturing unit about how much to produce on upcoming sales batch.

PROBLEM STATEMENT

The supply chain management (SCM) networks always wait for a forecasted information. If a company want to redefine its SCM for the next 5 years taking into account the recently increased shipping expenditure. This company too want to forecast it's future demands also.

Stating that I have Supply Chain Network as follows;

5 markets in Brazil, USA, INDIA, GERMANY. And they all have a high and low-capacity manufacturing sites. Shipping cost would be Rupees (₹)/Container. Customers demand (Units/Year).

For Manufacturing Facility Charges the Capital Expenditure for the equipment, Utilities, Factory Management, Administration, Space rental should be considered.

The Cost depends on country and the type of plant.

Fixed Cost(₹/Month)	Low-Capacity Sites	High-Capacity sites
USA	6500	9500
INDIA	4980	7270
GERMANY	6230	9100
BRAZIL	3230	4730
JAPAN	2110	3080

Manufacturing Costs

7

For the production costs, we want to consider production lines operators and Raw materials.

Cost (₹/Unit)	USA	Germany	Japan	Brazil	India
USA	12	12	12	12	12
GERMANY	13	13	13	13	13
JAPAN	10	10	10	10	10
INDIA	8	8	8	8	8
BRAZIL	5	5	5	5	5

From the table, the production cost of a unit produced in India is 5 ₹/Unit.

There is Shipping cost. Here the measurement is ₹/container. For simplicity we can assume one container can contain 1000 units.

Shipping cost (₹/container)	USA	GERMANY	JAPAN	BRAZIL	INDIA
USA	0	1750	1100	2300	1254
GERMANY	1905	0	1231	2892	1439
JAPAN	2200	3250	0	6230	2050
BRAZIL	2350	3150	4000	0	4250
INDIA	1950	2200	3500	4200	0

Shipping Costs (₹/container)

If I got a supply from Germany to India, it will cost 1439 (₹/container).

Manufacturing capacity by site is another factor I have. That is, how much raw materials can I put into one manufacturing period.

Capacity(kUnits/Month)	Low	High
USA	500	1500
GERMANY	500	1500
JAPAN	500	1500
BRAZIL	500	1500
INDIA	500	1500

For example, a high-capacity factory in USA can produce 1,500,000 units per month.

The next factor is Customers demand per market.

Units/Month	Demand
USA	2 800 000
GERMANY	90 000
JAPAN	1 700 000
BRAZIL	145 000
INDIA	160 000

Manufacturing site capacity

If company's Brazil plant is planning to produce for the local market and Germany;

Facilities: 1 high-capacity plant

(BRAZIL, BRAZIL) = 145,000 (Units/Month)

(BRAZIL, USA) = 1,355,000 (Unit/Month)

If India plant produce for all countries except Brazil;

Facilities: 1 high and 1 low-capacity plant

(India, India) = 160,000 (Units/Month)

(India, Germany) = 90,000 (Units/Month)

(India, Japan) = 200,000 (Units/Month)

(India, USA) = 1,550,000 (Units/Month)

Japan produces locally since a shortage;

Facilities: 1 high-capacity plant

(Japan, Japan) = 1,500,000 (Units/Month)

The total costs will be = 62,038,000 (₹/Month)

We can implement an outsourcing to low-cost regions. Like in the sense, moving the production to a less expensive area and doubling the production there.

We can also implement a surge in shipping cost due to container shortage if one is already happening.

LITERATURE REVIEW

The literature works I have observed throughout the development of this project will be discussed now under this review.

General resources used to forecast

Data in the context of supply chains (SC) can be categorized into customer, shipping, delivery, order, sale, store, and product data.SC data originate from sources such as sales, inventory, manufacturing, warehouse and transportation.

Data like competition, price volatilities, technological development, and varying customer commitments could lead to underestimation or overestimation of demand.

Therefore, to increase the precision of demand forecast, supply chain data shall be carefully analysed to enhance knowledge about "market trends", "customer behaviour", "Suppliers and technologies**".**

Complexity of SCM

Analysis of supply chain data has become a complex task due to;

1) Increasing multiplicity of SC entities.

2) Growing diversity of SC configurations depending on the homogeneity or heterogeneity of products.

3) Interdependencies among these entities.

4) Uncertainties in dynamical behavior of these components

5) lack of information as relate to SC entities.

6) Networked manufacturing/production entities due to their increasing coordination and cooperation to achieve a high-level customization and adaptation to varying customers' needs.

7) The increasing adoption of supply chain digitization practices (and use of Blockchain technologies) to track the activities across supply chains

1) Forecasting Using Big Data/Machine learning Techniques

Time-series forecasting

Time series forecasting is the process of analyzing time series data using statistics and modeling to make predictions and inform strategic decision-making.Time series forecasting occurs when you make scientific predictions based on historical time stamped data

Least square method in forecasting can be used with all prices to estimate future demands, but can also be used with volumes and other indicators.

Understand the least-square method let assume that all points (values) which are used as historical data to predict the future belong to the unknown function *f(x)*

The main goal is to find function *f(x)* which is in many cases almost impossible, or to approximate the *f(x)* function with another function *q(x)*.

K-nearest-neighbour (KNN)

KNN is a method of classification that has been widely used for pattern recognition.

KNN algorithm identifies the similarity of a given object to the surrounding objects (called tuples) by generating a similarity index. These tuples are described by n attributes. Thus, each tuple corresponds to a point in an n-dimensional space. The KNN algorithm searches for k tuples that are closest to a given tuple. These similarity-based classifications will lead to formation of clusters containing similar objects.[3] [48].

1.2) Forecasting Using Big Data/Machine learning Techniques

Artificial neural networks

In artificial neural networks, a set of neurons (input/output units) are connected to one another in different layers in order to establish mapping of the inputs to outputs by finding the underlying correlations between them.

The configuration of such networks could become a complex problem, due to a high number of layers and neurons, as well as variability of their types (linear or nonlinear), which needs to follow a data-driven learning process to be established

In doing so, each unit (neuron) will correspond to a weight, that is tuned through a training step. At the end, a weighted network with minimum number of neurons, that could map the inputs to outputs with a minimum fitting error (deviation), is identified

Regression analysis

These methods are used to predict the value of a response (dependent) variable with respect to one or more predictor (independent) variables

There are various forms of regression analysis, such as linear, multiple, weighted, symbolic (random), polynomial, nonparametric, and robust.

The latter approach is useful when errors fail to satisfy normalcy conditions or when we deal with big data that could contain significant number of outliers [4] [48]

1.3) Forecasting Using Big Data/Machine learning Techniques

Support vector regression (SVR)

Continuous variable classification problems can be solved by support vector regression (SVR), which is a regression implementation of SVM.

The main idea behind SVR regression is the computation of a linear regression function within a high-dimensional feature space. SVR has been applied in financial/cost

prediction problems, handwritten digit recognition, and speaker identification, object recognition, etc.

2)Comparison of Techniques

Most of the studies examined, developed and used a certain data-mining algorithm for their case studies. However, there are very few comparative studies available in the literature to provide a benchmark for understanding of the advantages and disadvantages of these methodologies. Additionally, as depicted by Table (right), there is no clear trend between the choice of the BDA algorithm/method and the application domain or category. The neural networks and regression analysis are observed as the two mostly employed techniques, among others.

Literature (2005–2019)	Predictive BDA techniques							Demand forecasting SC application
	ANN	Clustering	KNN	Regression	SVM	SVR	TSF	
Chang et al. [91]	✓							Printed circuit boards
Levis and Papageorgiou [116]						✓		Electrical appliances
Chi et al. [117]					✓			Calendars
Sun et al. [118]	✓							Fashion retails
Efendigil et al. [119]	✓							Durable consumer goods
Lee and Ou-Yang [120]	✓							Supplier bidding
Chen et al. [121]	✓							Wafer tests
Thomassey [105]	✓							Clothings
Wu [95]					✓			Cars
Wu [122]					✓			Cars
Guanghui [96]	✓					✓		Paper and Cards
Babai et al. [123]							✓	Superstore retails
Kouremtzes [124]	✓							Automotive spare parts
Lau et al. [125]	✓							Gasoline
Da Veiga et al. [73]							✓	Dairy products
Jun et al. [37]							✓	Hybrid cars
Ma et al. [71]							✓	Smartphones
Arunraj and Ahrens [126]	✓						✓	Food retails
Blackburn et al. [7]							✓	Chemicals
Gaur et al. [81]			✓					Walmart SC
Islek and Oguducu [100]		✓						Dried nuts and fruits
Murray et al. [74]		✓						Bulk materials
Ramos et al. [75]							✓	Retails sector (footwear)
Saha et al. [90]	✓							Server manufacturing
Di Pillo et al. [127]					✓		✓	Retails sector
Liu et al. [86]	✓							Furniture
Nikolopoulos et al. [80]			✓					Automotive spare parts
da Veiga et al. [128]	✓						✓	Foodstuff retails
Vhatkar and Dias [84]	✓						✓	Oral care products
Amirkolaii et al. [88]	✓							Aircraft spare parts
Bohanec et al. [17]	✓				✓			Toys
Chen and Lu [98]						✓		Computer retails
Huber et al. [19]		✓					✓	Bakery products
Chawla et al. [129]	✓							Walmart
Huang et al. [89]	✓			✓				E-logistics
Loureiro et al. [102]	✓			✓		✓		Fashion retails
Pereira et al. [130]				✓				Scrap tires
Yang and Sutrisno [93]	✓			✓				Perishable goods
Villegas et al. [94]					✓		✓	Personal care products
Yuan et al. [87]	✓							E-logistics
Fanoodi et al. [131]	✓						✓	Blood cells (platelet)
Kilimci et al. [101]	✓					✓	✓	Discount stores
Merkuryeva et al. [92]				✓			✓	Pharmaceuticals
Puspita et al. [104]		✓				✓		Forklifts
Punam et al. [103]			✓	✓		✓		Online grocery
Sharma and Singhal [132]	✓			✓				Automotive/industrial lubricants

COMPUTATIONAL MODELS USED

Multilayer perceptron (MLP) Neural Network

Multilayer perceptron or MLP is a type of Artificial Neural Network. MLP is commonly used for regression problems. Its is one of the advanced models of Deep learning too. MLP's are not ideal for processing patterns with sequential and multidimensional data.

The basic building blocks of a deep learning algorithm is a perceptron which is a single neuron in neural networks. An artificial neuron inspired by biological neuron is called as a perceptron.

MLP is a feed forwarding ANN that generates a set of outputs from a set of inputs. MLP connects multiple layers in a direct graph, which means that the signal path through the nodes only go unidirectional. MLP consists of input, output and hidden layers. Hidden layer keeps

the perceptron. It performs computations and transfer information from the input layer to the output layer. Hidden layer has no direct connection with the outside world.

Random Forest

RF is a supervised learning algorithm which is used for classifying and as well as regression. It is but mainly used for classifying problem. Like a forest with many trees and less trees have difference. Even from a dense forest, RF will apply decision tree and make decision by means of voting. It is also an ensemble model. This model is better than a single decision tree because it will be able to cut out over-fitting by the averaging result.

Ensemble

In ensemble model, we use a diverse number of models to predict an outcome. The model is used in most of the cases and situations to avoid generalization error in the prediction. As long as all diverse models are not interlinked or dependent, the prediction error decreases after applying ensemble model.

Long Short-Term Memory (LSTM)

Long short-term memory model is a kind of special model which is capable of learning long term dependencies in data. This achievement is because the module of the model has a combination of four layers interacting with one another. Basically, this can be added up to the time series analysis of data.

KNN

K-Nearest Neighbor or KNN is complex less algorithm based on supervised learning technique. KNN decide similarities of data points between new and available data. It puts data to a new category where new and available data are more similar. It stores new data point based on the similarity like mean of desired points. Used mostly for classification, but act as a regression model too. It is a non-parametric data.

SVC

Support vector classifier or SVC is a linear model which aims to fit the data we provide, by placing a best fit returns given by a hyper plane geometry on data points.

AdaBoostClassifier

This is another ensemble model like RF. In Adaptive Boosting or Ada boosting, the model combines multiple ‘weak classifiers’ into a single ‘strong classifier’.

Logistic regression

Logistic regression is a probabilistic supervised classifying algorithm, which uses input data to predict the outcome. It has only two or binary outcomes like yes/no. it is mainly used for fraud detection and classification.

GausianNB

This model is developed using Naïve Bayes and is a probabilistic model with strong independence. This NB algorithm manly used when the features have continuous values. It finds the distribution across the points.

DesicionTreeClassifier

These are a type of supervised Machine learning model. It is a regression and classification model and a tree structured classifier where the internal nodes are features of dataset and branches represent rules for decision and each leaf nodes represent the outcome. That is we will be getting more outcomes.

GradientBoostingClassifier

It is a decision tree classifier used for preventing over fitting. This histogram-based classification tree, is a meta estimator. That is, it fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictivity.

DATASET AND DESCRIPTION

The dataset I mainly used is the csv file consisting of production cost, sale data, shipping, and demand.

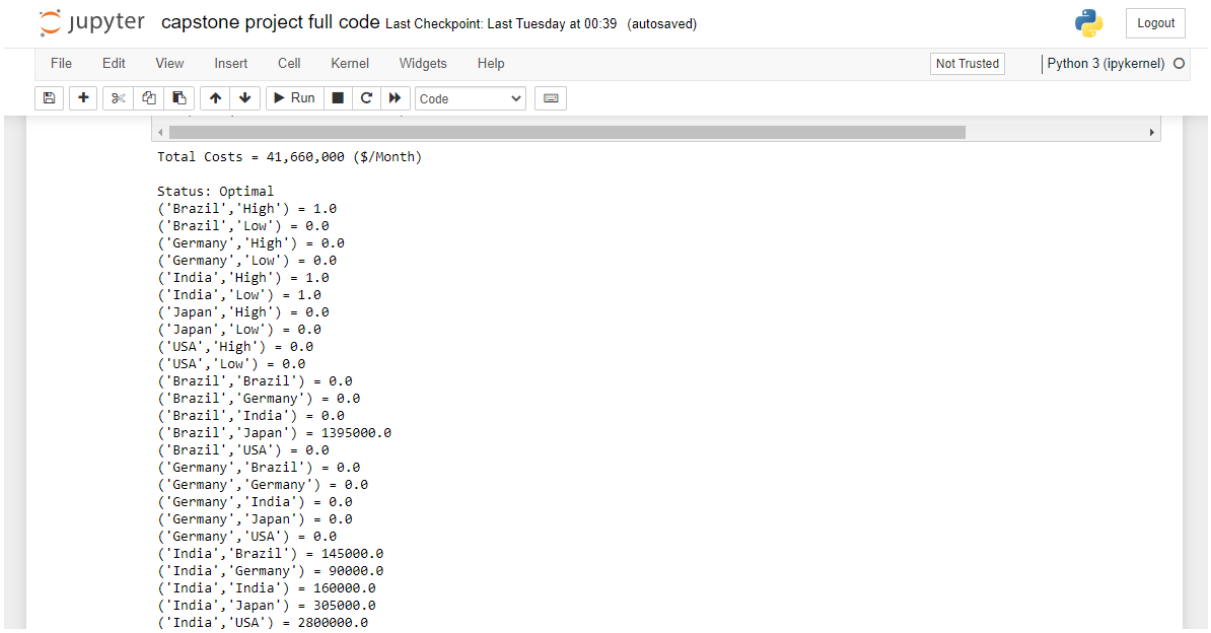
It’s a regular no complex dataset which can be achieved and used by any historical information by a store of production unit.

METHODOLOGY

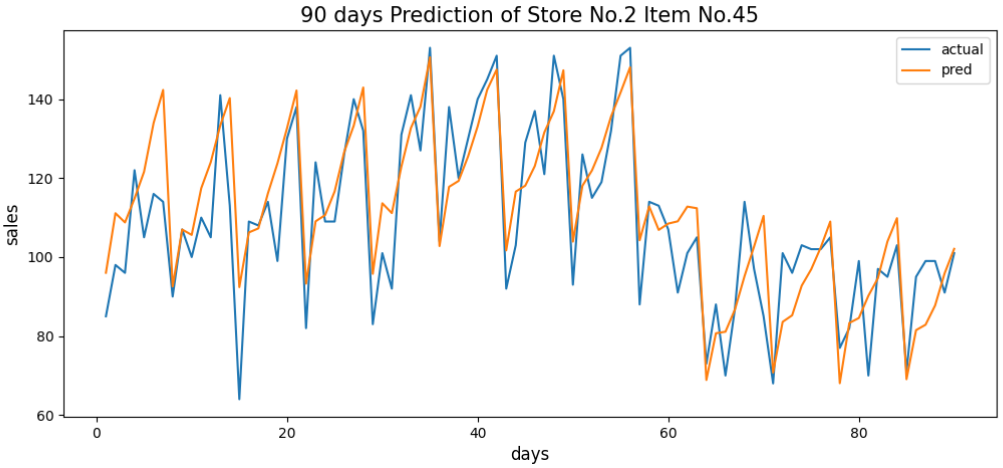
The project is started first by importing the necessary libraries we needed. The models we discussed above are needed to implement. Since it is a regression related project, Data visualization is very important.

The visualization part is done using pandas library. After implementing the PuLp model, we have discussed earlier, since its data part is very strong (Manufacturing cost data, Production cost data, Shipping charges, and Demand of product.).

The predicted cost for all high and low-capacity sites are attached below.



The sales demand forecasting predicted can be viewed as a plot as seen below;



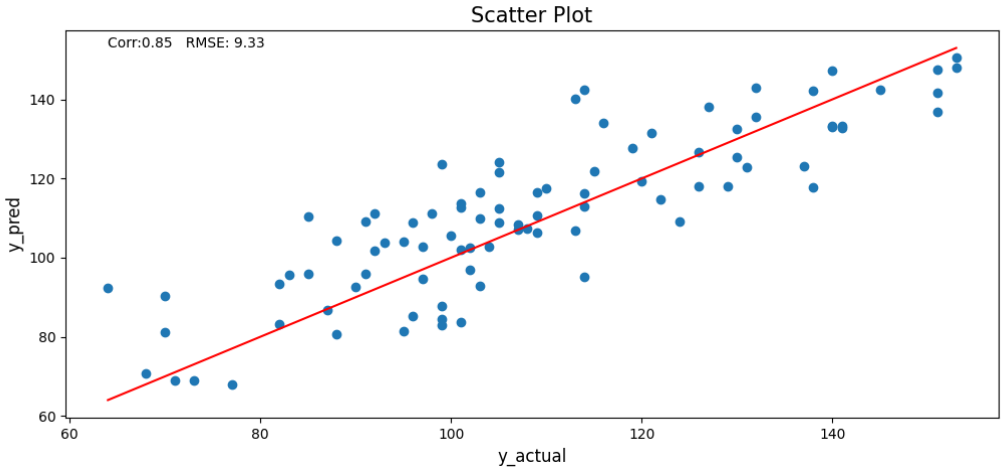
This is a prediction model using MLP, which is a well performing neural network. The blue line in the graph is actually the historic data of the store and the orange-coloured line is forecasted data. It follows the regression points fairly and using this, the supply chain can be altered.

OBSERVATIONS

Since MLP is a deep neural network, the code took a lot of time and few CPU crashes to during implementation. Since it risks more memory and GPU, the results are fair and the model's accuracy rate convinced more than error rate.

The implementation of the project using PuLp libraries are not that much crashing and struggling. It relied on data sets to a great extent. The PuLp libraries struggled in regression part, since it cannot give a detailed prediction outcome.

The ensemble model helped a lot for this project. From comparing the models to using it concurrently, the prediction outcomes increased in a fair manner.



The above scatter plot is generated using MLP model and it have a correlation of 0.85 and a Root Mean Square Error value of 9.33.

CONCLUSION

Since proctoring is needed in every department, the need for forecasting is very crucial in SCM systems. Because in SCM systems, people cannot have access after the delivery or during the delivery. Once a delivery is scheduled, its gone. All the ground works are needed to be done on the office or on the yard.

By using data science and machine learning we can utilize the forecasted data from manufacturing to delivery. All of the e-commerce websites use supply chain management system and demand forecasting using the customer behaviour and every click we are doing.

The most win probable model is clustering the data. Clustered data can be useful in forecasting, like market basket analysis.

Most of the data driven model considered in literature use historical data. It ignores current trends. Organizational factors are also dependent like organizations reputation, availability of SCM resources to forecast demand. Incorporating existing driving factors outside the historical data, such as economic instability, inflation, and purchasing power, could help adjust the predictions with respect to unseen future scenarios of demand.

Combining predictive algorithms with optimization or simulation can equip the models with prescriptive capabilities in response to future scenarios and expectations.

The main backend advantage of this system is we can analyze how a customer is changing according to trends and how we can move the business according to the customer.

We can save a large amount from implementing this project, like we can utilize the low and high-capacity manufacturing sites, we can avoid shipping charges by analyzing local importance of supply.

Another advantage we get from this model is advance production plans. Producing according to demand can help everyone in many ways.

BIBILIOGRAPHY

- Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities Mahya Seyedan and Fereshteh Mafakheri - Seyedan and Mafakheri J Big Data (2020) - https://doi.org/10.1186/s40537-020-00329-2
- Blackburn R, Lurz K, Priese B, Göb R, Darkow IL. A predictive analytics approach for demand forecasting in the process industry. Int Trans Oper Res. 2015;22(3):407–28. https://doi.org/10.1111/itor.12122.
- Varela IR, Tjahjono B. Big data analytics in supply chain management: trends and related research. In: 6th international conference on operations and supply chain management, vol. 1, no. 1, p. 2013–4; 2014 – https://doi.org/10.13140/RG.2.1.4935.2563
- [1]https://www.bing.com/ck/a?!&&p=201c761d05d823a7f0007f3552a591edfc5eda54e1074ee54dd9d8afa58c8249JmltdHM9MTY1NTQ1NzM0OCZpZ3VpZD02YjYzOTcwNC0yNTJlLTQwMzQtYTFlZC03ZjQyODg2YTc3MGUmaW5zaWQ9NTM5Mg&ptn=3&fclid=136468b4-ee1e-11ec-bcfa-2558d748fbe8&u=a1aHR0cHM6Ly93d3cudGVjaG5vYWQuY29tL3J1YmJlci1pbmR1c3RyeS1ydWJiZXItc3VwcGx5LWRpc3J1cHRpb25zLyM6fjp0ZXh0PVRoZSUyMHNpbGljb25lJTIwc2hvcnRhZ2UlMjBpcyUyMGF0dHJpYnV0ZWQlMjB0byUyMGElMjBudW1iZXIsZm9yJTIwcHJvZHVjdHMlMjBtYW51ZmFjdHVyZWQlMjBmcm9tJTIwc2lsaWNvbmUlMkMlMjBlc3BlY2lhbGx5JTIwaGVhbHRoY2FyZSUyMGVsYXN0b21lcnMu&ntb=1
- [2]https://www.bing.com/ck/a?!&&p=1af7aafc6d1db2a3d12cd1d9f380e0bbf24187205c9613f9e44d524a44a9bbfeJmltdHM9MTY1NTQ1NzY4MiZpZ3VpZD00OGUzYzk5ZS1mNWU1LTQxYjMtOGUxYi1iMzhmMDJmYTM1NTkmaW5zaWQ9NTQxNg&ptn=3&fclid=da310165-ee1e-11ec-b4e5-182c2b933475&u=a1aHR0cHM6Ly93d3cuaW52ZXN0b3BlZGlhLmNvbS90ZXJtcy9zL3NjYXJjaXR5LXByaW5jaXBsZS5hc3AjOn46dGV4dD1UaGUlMjBzY2FyY2l0eSUyMHByaW5jaXBsZSUyMGlzJTIwYW4lMjBlY29ub21pYyUyMHRoZW9yeSUyMHRoYXQsaGlnaCUyMGRlbWFuZCUyQyUyMHJpc2VzJTIwdG8lMjBtZWV0JTIwdGhlJTIwZXhwZWN0ZWQlMjBkZW1hbmQu&ntb=1
- https://www.researchgate.net/publication/288228334_A_study_of_forecasting_practices_in_supply_chain_management
- https://www.researchgate.net/publication/286677185_Forecasting_Supply_Chain_Demand_Using_Machine_Learning_Algorithms
- https://ieeexplore.ieee.org/document/9580006