*IOSR Journal of Computer Engineering (IOSR-JCE)*
*e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 22, Issue 5, Ser. II (Sep. – Oct. 2020), PP 53-68*
*www.iosrjournals.org*

# A Review on Data Analytics in Supply Chain Management using Forecasting and Product Portfolio techniques

Sanskriti Upadhyaya[1], Ayusha Kakkad[2]
[1](Computer Engineering, NMIMS University, India)
[2](Computer Engineering, NMIMS University, India)

***Abstract:***
*India is becoming a global manufacturing hub. Increasing demand in international markets is opening a new world of opportunities for the Indian Industry. Data analysis and supply chain is a tool for improving overall performance in today's global competitive environment. Supply chain management affects practically every part of the economy and is based on increased collaboration between companies, customers and governmental organizations which are addressed with big data solutions. Therefore, Big Data concepts and technologies play a key role. Big Data-related research has attracted a growing number of Business Intelligence and Analytics (BI&A) researchers by combining relevant frameworks from supply chain management. There are various issues related to SCM in this paper which are resolved by different analytical techniques like forecasting and product portfolio through effective algorithms.*
***Keywords**:Supply chain; Data analytics; Bigdata; Forecasting; Product portfolio.*

---

Date of Submission: 10-10-2020 Date of Acceptance: 26-10-2020

---

## I. Introduction

A supply chain consists of all parties involved, directly or indirectly, in fulfilling a customer request. The supply chain includes not only the manufacturer and suppliers, but also transporters, warehouses, retailers, and even customers themselves. Within each organization, such as manufacturer, the supply chain includes all the functions involved in receiving and filling a customer request. These functions include, but are not only limited to new product development, marketing, operations, distributions, finance, and customer services.
A typical supply chain may involve a variety of stages, including the following:[1,5]

1. Customers
2. Retailers
3. Wholesalers/Distributors
4. Manufacturers
5. Component/Raw Material Suppliers

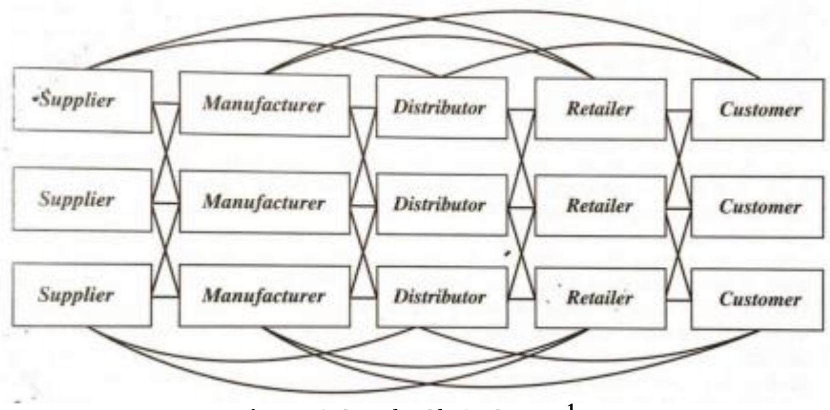


**Figure 1:**Supply Chain Stages.[1]

Each stage in supply chain is connected through the flow of product, information, and funds. These flows often occur in both directions and may be managed by one of the stages or an intermediary. For example, Dell has two supply chain structure that it uses to serve its customers. For Its server business, dell builds to order; that is, a customer order initiates manufacturing at dell. For the sale of servers, dell does not have a separate retailer, distributor, or wholesaler in the supply chain. Dell also sells consumer products such as PCs and tablets through retailer such as Walmart, which carry dell product in inventory. This supply chain thus contains an extra stage (the retailer), compared with the direct sales model used by sell for server. In the case of other retail stores, the supply chain may also contain a wholesaler or distributor between the store and the manufacturer.[1]

**What is Data Analytics?**

Data analytics (DA) is the process of examining data sets to draw conclusions about the information they contain, increasingly with the aid of specialized systems and software. Data analytics technologies and techniques are widely used in commercial industries to enable organizations to make more-informed business decisions. [2,7]

It is the responsibility of managers to plan, coordinate, organize, and lead their organizations to better performance. Ultimately, managers' responsibilities require that they make strategic, tactical, or operational decisions. [2,8]

There are a number of approaches to making decisions: tradition ("We've always done it this way"), intuition ("gut feeling"), and rules of thumb ("As the restaurant owner, I schedule twice the number of waiters and cooks on holidays"). [2]

Business analytics is the scientific process of transforming data into insight for making better decisions. Business analytics is used for data-driven or fact-based decision making, which is often seen as more objective than other alternatives for decision making.[2]

Business decision making processes today are also overwhelmed by massive amount of information where the realistic situation has gone beyond the natural cognitive ability of humans to cope. However, by embracing and making best use of big data analytical techniques, we can create value that have never been known before and that can significantly help achieve improved operational efficiency, gain competitive advantages over business rivals, generate or increase new revenue stream, deliver cost reductions, better storage and transport facility and drive agile decision making from predictive insights and thus helping in supply chain. [2,8]

The enablement of data driven decision support to overcome big data challenges by taking advantage of analytic technologies and techniques that can attain significant achievements to do more with less, in a smarter manner that provides intellectual decision support. Other well documented benefits include quicker time to market, improved process lead time, improved adaptability, and flexibility within limited resources. The data analytics approach to be discussed ranges from business understanding, data understanding, data preparation,modelling and visualization through dashboards that include three different types of analytics, being descriptive, predictive, and prescriptive analytics.[2,7]

**Use of Data Analytics in Supply Chain Management:**

One of the earliest applications of analytics was in supply chain management.[2]

Companies can benefit from better inventory and processing control and more efficient supply chains. Analytic tools used in this area span the entire spectrum of analytics.[2]

For example, the women's apparel manufacturer Bernard Claus, Inc., has successfully used descriptive analytics to present the status of its supply chain to managers visually. ConAgra Foods uses predictive and prescriptive analytics to better plan capacity utilization by incorporating the inherent uncertainty in commodities pricing.[2]

The supply chain process includes the total lead time from identifying opportunities to making or procuring the product to getting the product on the shelves to align with the forecasted demand; this can potentially take several months, so the accuracy of forecasts is critical throughout each step of the supply chain.[2]

Adding to this challenge is the risk of obsolescence. Companies sell many dated items, such as planners and calendars, which have a natural, built-in obsolescence. In addition, many of the products feature designs that are fashion-conscious or contain pop culture images, and these products can also become obsolete very quickly as tastes and popularity change. An overly optimistic forecast for these products can be very costly, but an overly pessimistic forecast can result in lost sales potential and give the competitors an opportunity to take market share.[2]

**Analytic techniques in Supply chain:**

Analytics techniques can be categorized into three types:

1. Descriptive:

Descriptive analytics derives information from significant amounts of data and answers the question of what is happening. Real-time information about the location and quantities of goods in the supply chain provides managers with tools to adjust delivery schedules, place replenishment orders, place emergency orders, change transportation modes, and so forth.[8]

2. Predictive:

Predictive analytics in supply chains derives demand forecasts from past data and answers the question of what will be happening. [8]

3. Prescriptive:

Prescriptive analytics derives decision recommendations based on descriptive and predictive analytics models and mathematical optimization models.[8]

**Table no 1:** Analytic techniques used in Supply Chain Management.[8]

| Analytics Techniques | Source | Make | Deliver | Return |
|---|---|---|---|---|
| Descriptive | • Supply chain mapping | • Supply chain visualization | | |
| Predictive | • Time series methods (e.g., moving average, exponential smoothing, autoregressive models)<br>• Linear, non-linear, and logistic regression<br>• Data-mining techniques (e.g., cluster analysis, market basket analysis) | | | |
| Prescriptive | • Analytic hierarchy process<br>• Game theory (e.g., auction design, contract design) | • Mixed-integer linear programming (MILP)<br>• Non-linear programming | • Network flow algorithms<br>• MILP<br>• Stochastic dynamic programming | |

The Supply Chain Operations Reference (SCOR) model developed by the Supply Chain Council provides a good framework for classifying the analytics applications in supply chain management. The SCOR model outlines four domains of supply chain activities: source, make, deliver, and return. A fifth domain of the SCOR model plan is behind all four activity domains.[3]

**Table no 2:** SCOR model and example of decision at three levels.

| SCOR Domain | Source | Make | Deliver | Return |
|---|---|---|---|---|
| Activities | Order and receive materials and products | Schedule and manufacture, repair, remanufacture, or recycle materials and products | Receive, schedule, pick, pack, and ship orders | Request, approve, and determine disposal of products and assets |
| **Strategic** (time frame: years) | • Strategic sourcing<br>• Supply chain mapping | • Location of plants<br>• Product line mix at plants | • Location of distribution centers<br>• Fleet planning | • Location of return centers |
| **Tactical** (time frame: months) | • Tactical sourcing<br>• Supply chain contracts | • Product line rationalization<br>• Sales and operations planning | • Transportation and distribution planning<br>• Inventory policies at locations | • Reverse distribution plan |
| **Operational** (time frame: days) | • Materials requirement planning and inventory replenishment orders | • Workforce scheduling<br>• Manufacturing, order tracking, and scheduling | • Vehicle routing (for deliveries) | • Vehicle routing (for returns collection) |
| **Plan** | Demand forecasting (long term, mid term, and short term) | | | |

**Supply Chain Data Analysis consist of 2 types:**

1. Forecasting Analysis which is the understanding of qualitative and quantitative predictions.[2,8,9]
2. Product Mix/Portfolios is basically for optimal solutions for profitable productions.[2,8,9]

## II. Big Data

Big data describes a way of collecting, managing, and analyzing large amounts of data. Therefore, big data is mostly referenced with the three Vs as described by. They are defined as volume, velocity, and variety.[10,11]

The volume characterizes the vast amounts of data stored within the IT infrastructure. In general, storing a large amount of data already represents a big issue for the data storage itself. One of the main challenges for IT infrastructures dealing with big data is to ensure the availability of storage space and an efficient accessibility. Velocity describes the large amounts of data that arrive in real-time irregularly. The fast arriving data must be handled if a further usage is planned. The last one – variety – relates to the various structure of information handled within the big data environment. The stored data can either be based on a structure (datatype) or consist of unstructured information. The value of big data is often described as the 4th V. This is related to the fact that the data needs to be processed and analyzed for further usage. Handling large amounts of data requires both, concepts for managing these datasets and concepts for processing the amounts of data.[10,11]

The framework differentiates between internal (focusing on the company) and external (focusing on the overall cross-organizational supply chain) impact of ACBs. For this purpose, we use different codes to indicate whether ACBs relate to the value dimensions internally "I", externally "E", both ways "I, E", or have no impact in the respective value dimension "- ".[10,11]

**Table no 3:** BD Applications, Challenges and Benefits-Key findings.[11]

| A – Applications C – Challenges B - Benefits | Value Dimension Impact: SCM-Specific Operations: Procurement | Production | Distribution | Sales | Generic (SCM-related) | Summary of Findings |
|---|---|---|---|---|---|---|
| A1. Material flows | I | I | I E | - | - | Exemplary applications: Pre-production check, internet of things, research and development, process and quality monitoring, stock handling, logistics, crowd solutions |
| A2.Info flows | E | I E | E | I E | I | Exemplary applications: Product lifecycle management, design-to-value, virtual collaboration sites, demand forecast, supply chain event management, supplier negotiations, big data analytics, risk management, issue identification, automated decision support, customer management, geo-targeting, 4PL integrator |
| A3.Financial flows | - | - | - | I E | I E | Exemplary applications: Demand shaping, design of new business model, pricing and assortment, financial aspects of human resources, optimization, strategic network planning, customer segmentation |
| C1. BD requirements | - | - | - | - | I | Key requirements: Technology, software, human capital, culture, financial investments |
| C2. BD complexity | - | - | - | - | I E | Complexity caused by processing, access, sharing, transparency, quality vs. quantity of data |
| C3. BD use | - | - | - | - | I | Key challenges of BD use: BD strategy, real-time processing, managerial perception of, and support for innovation, internal regulations, application areas, insights into action, lack of business cases |
| B1. Products and services | I | I | I E | I E | - | Key benefits: Cost reduction, resource efficiency, purchase, quality improvement, sensor-driven, distribution, faster go to market, virtual prototyping, simulations |
| B2. BD-enabled Analysis | I | I | I E | I E | I E | Benefits of BD-enabled Analysis: Bottleneck analysis, new strategic directions, new business model, marketing and sales, opportunity recognition, internal decisions, automated real-time decisions, transparency, risk management, cost identification, (near) real-time analysis, responsiveness, research and development |
| B3. SC relationships | I E | - | - | I E | - | Benefits related to improved and new SC relationships: Improved customer service, increased bargaining power, selection of suppliers, new forms of collaboration |

**Table no 4:**Example of cause of Big Data.[7]

| Type of data | Volume | Velocity | Variety |
|---|---|---|---|
| Sales | More detail around the sale, including price, quantity, items sold, time of day, date, and customer data | From monthly and weekly to daily and hourly | Direct sales, sales of distributors, Internet sales, international sales, and competitor sales |
| Consumer | More detail regarding decision and purchasing behavior, including items browsed and bought, frequency, dollar value, and timing | From click through to card usage | Face profiling data for shopper identification and emotion detection; eye-tracking data; customer sentiment about products purchased based on "Likes," "Tweets," and product reviews |
| Inventory | Perpetual inventory at more locations, at a more disaggregate level (e.g., style/color/size) | From monthly updates to hourly updates | Inventory in warehouses, stores, Internet stores, and a wide variety of vendors online |
| Location and time | Sensor data to detect location in store, including misplaced inventory, in distribution center (picking, racks, staging, etc.), in transportation unit | Frequent updates for new location and movement | Not only where it is, but what is close to it, who moved it, its path to get there, and its predicted path forward; location positions that are time stamped from mobile devices |

## III. Issues and Advantages

**A. The issues identified are in following aspects of Supply chain management:**

1. Supply chain strategy
2. Risk management
3. Performance measurement
4. Framework and standards implementation
5. Informatization of supply chain
6. Integration in supply chain management.

All of these are important for Business Process Renovation and ensuring an effective supply chain.

**Issues at Strategic Level:**

To compete in the current economic scenario, proper implementation of supply chain system is vital for the organizations. Profits are getting lower due to high cost of goods and services forming a major part of the sales revenue. These costs must be reduced by effective supply chain management. For this, supply chains need to be designed to suit to the specific needs of the situation, as one solution may not fit all.[5,6]

There are five critical components of supply chain implementation – strategies for Operations, outsourcing, channel, customer service and effective asset network. The four main approaches towards production are “make to stock”, “make to order”, “configure to order” and “engineer to order”. The strategies for supply chain are affected by choice of these approaches. Modern supply chains also face problems because of communication across departments, external partners and misaligned business goals and IT implementations.[5,6]

For this reason, strategy formulation is of paramount importance. Business process renovation is required to be done once the strategies are formulated. For this, Supply chain processes must be examined across the departments as well as across external parties involved, to identify possible areas for changes to improve quality, reduce cost or time. The changes can be transformational such as Business Process Reengineering (BPR) or transitional such as Continuous Process Improvement (CPI). IT strategies must also be integrated in these changes.[5,6]

Business models are developed to understand how various components of the business are related to each other and how they interact with each other. It helps to understand the existing status and set the future goals.[5,6]

The MNC’s have various branches in various countries and the business strategy for each country varies according to the economic and political conditions. Hence business strategies for each branch varies creating challenges for the integrated supply chain management.[5,6]

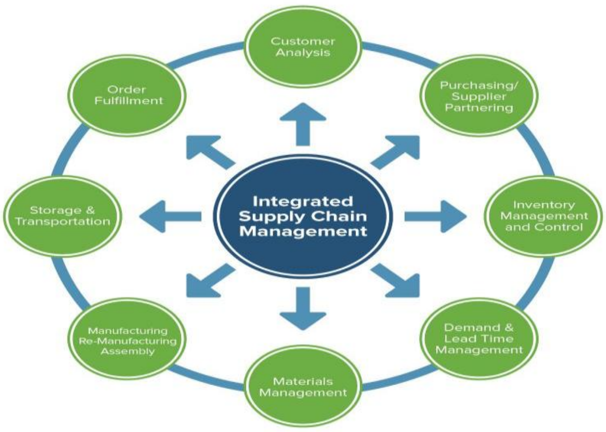


**Figure 2:**Integrated SupplyChain Management**.**

**Issues in supply chain risk management:**

In recent times, risks factors have increased in Supply chain. This has happened due to producers increasingly outsourcing components, increased distances from suppliers due to globalization, reduction of number of suppliers, reduced material reserves to reduce costs, customers demanding on-time delivery and shorter product lifecycle etc. To effectively address these risks, risk management has become an important aspect in supply chain. SCRM consists of identification, analysis and response planning and continuous monitoring of the risk factors.[5,6]

Authors discuss a recently developed model that predicts risks connected to the suppliers. Different suppliers and their sub-suppliers operate in different environments and hence face different levels of risk from various factors. E.g. the strategies of ordering large batch to save ordering and transportation costs works in a stable environment, but same strategy may become risky in a turbulent environment where technological advances render the stored material useless. Hence different strategies need to be adopted for different suppliers. In this model, the sources of risk have been clearly demarcated into two different types:[5,6]

1.Endogenous risks are due to factors within the supply chain. These risks may change the arrangements between firm and its suppliers. Changing customer preference and competitors are also risk factors. Changing technology may also impact the supply chain.[5,6]

2.Exogenous risks are due to factors outside the supply chain. These risks can happen due to isolated events such as strike, natural calamities, outbreak of diseases etc. or they can happen due to environmental economic factors such as inflation, consumer price index changes.[5,6]

**Issues in supply chain frameworks and standards implementation:**

Although different supply chain frameworks and standards have been developed, organizations need to decide the extent to which they need to implement the standards. It also binds the organizations to a certain methodology that becomes difficult to change – which is desirable in the changing business environment. Supply-Chain Operations Reference (SCOR) model by the Supply-Chain Council offers a process-oriented approach to organizations to increase the effectiveness of their supply chains. This model provides a common language for communicating among supply-chain partners in the decision areas of PLAN, SOURCE, MAKE, DELIVER and RETURN. Lock Amy & McCormack have shown that though SCOR planning practices are important for supply chain performance, supply chain partners mostly do not implement them. Also, some of the best practices documented in literature help in improving supply chain performance only in specific decision areas.[5,6]

**Issues in supply chain Performance measurement:**

To manage the supply chain well, we need to measure the performance in metrices which – are related directly to the manufacturing strategy, primarily use non-financial measures, vary between locations, change over time, as needs change, are simple and easy to use, provide fast feedback to operators and managers, are intended to improve rather than just monitor. Replenishment lead time, on-time performance, supply flexibility, delivery frequency, quality, viability, information coordination capability etc. are classic metrics whereas in recent time's quality of delivered goods, on time delivery and flexibility of supply are considered more important. This shows shift towards customer orientation. Performance measurement is closely connected with process renovation. Authors have suggested different metrics for performance measurements such as supplier delivery performance and purchase order time for "SOURCE", utilization of resources or percentage of defects for "MAKE" phase. They have also suggested use of Simulations for performance measurements and risk assessment.[5,6,8]

**Issues in supply chain informatization:**

Information technology is an important enabler of effective supply chain management. There are enormous possibilities due to abundance of data and the savings inherent in sophisticated analysis of these data. Supply chain systems consists of various component systems that are involved in supply chain planning.

These systems must be standardized to work together. This is a major challenge. These are typically systems that combine short-term and long-term decision support system elements. SCM software such as group decision support systems, EDI and e-business are implemented to increase information flows through collaboration, but increasing collaboration requires much more than just a software. Supply chain participants must use the tool effectively to ensure its effective use.[5,6]

**B. ADVANTAGES OF USING SUPPLY CHAIN MANAGEMENT:**

1. **Higher Efficiency Rate:**
   Running a small business could be a tough task. But when your business comes up with product innovation strategies, integrated logistics, and effective supply chain management, one can forecast demand and act accordingly. To keep up with the ever-changing economy supply chain management is a must.[4]

2. **Storage Buffers:**
   Industries must manage their storage goods in such a way that there is no shortage or wastage of goods and minimizes holding costs, fulfilling the customer demands as one may also lose out on business and revenue

if their customers get impatient waiting to receive the product. As a result, they may shop somewhere else. An efficient supply chain management system eliminates the issue.[4]

3. **Operation Flexibility:**
With the help of thoughtful supply chain management, integrated logistics, and product innovation strategies, prediction is done while having a look at the past records and produce and store for their product accordingly. Integrated supply chain management also includes increased flexibility. "Tight supply chain integration gives management operational flexibility to respond rapidly to external environment, such as the actions of competitors and changes in customer demand," "By data collection through their supply chains, allows the company to be generally aware of what their competitors are planning months in advance."[4]

## IV. Literature Review

### A. METHOD I: FORECASTING

Demand forecasting is a part of the machine learning intelligence known as" prediction" and plays a vital role and act as a critical input to supply chain planning. Different time frames for demand forecasting require different analytics techniques. It not only brings profit but also maintain the right quantity of products at the right time. Long-term demand forecasting is used at the strategic level and may use macro-economic data, demographic trends, technological trends, and competitive intelligence. For example, demand factors for commercial aircraft at Boeing include energy prices, discretionary spending, population growth, and inflation, whereas demand factors for military aircraft include geo-political changes, congressional spending, budgetary constraints, and government regulations (Safavi, 2005). Causal forecasting methods——called such because they analyze the underlying factors that drive demand for a product——are used at this level. Analytics causal forecasting methods include linear, non-linear, and logistic regression.[8,9]

Demand forecasting for independent demand items is usually performed using time-series methods, for which the only predictor of demand is time. Time series methods include moving average, exponential smoothing, and autoregressive models. For example, winter's exponential smoothing method incorporates both trend and seasonality and can be used for both short-term and mid-term forecasting. In an autoregressive model, demand forecast in one period is a weighted sum of realized demands in the previous periods.[9]

Mid-term forecasting can also benefit from causal forecasting methods, especially in nonmanufacturing industries or the manufacturing of non-discrete items. They used stepwise regression to identify the most relevant indexes and found parsimonious models for predicting TL demand for specific industries and regions. Their model only predicts industry-wide demand for TL services (nationally or by region); the connection to demand forecasts at the firm level was made using historical market shares.[9]

The efficiency or accuracy of the demand forecasting is taking as the major account for taking any major decisions such as capacity building, resource allocation, expansion and Forward or backward integration etc. As the accuracy goes on increasing the deviation from the actual sales also decreases. Now the major concern is the under or the over forecasting in the demand forecasting. The accuracy curve can be assumed to be a bell-curve in which for a particular accuracy, there can be two forecasting values i.e. on either side of the hundred percent accuracy. Thus, the major factor is the kind of forecast the machine is generating.[9]

**Forecast accuracy check based on time series prediction process:**
There are lot of methods or metrics to check the accuracy on the time series prediction process. In this case we are following two methods to measure the accuracy of the models:[9]

$$FACC = 1 - \frac{\sum_{i=1}^{k}(forecast_i - actuals_i)}{\sum_{i=1}^{k} forecast_i}$$

$$Accuracy = 1 - \frac{\sum_{i=1}^{k}(forecast_i - actuals_i)}{\sum_{i=1}^{k} actuals_i}$$

Here k denotes the range of the months in which the forecast accuracy is checked. If k = 3, then the forecast accuracy checked is termed as MT or the Medium-Term. This is very important in the product management. When k = 1, the forecast accuracy is termed as N-1 Accuracy. For the MT, the accuracy is checked after two of the forecast submissions. For example, the actuals that we have to input into the model is till August-17, then for checking the MT accuracy, the months that will be considered are cumulative sum of November-17, December-17 and January-18. Thus,[9]

$$FACC_{MT_{Jan-18}} = 1 - \frac{|(f_{Nov-17} + f_{Dec-17} + f_{Jan-18}) - (A_{Nov-17} + A_{Dec-17} + A_{Jan-18})|}{f_{Nov-17} + f_{Dec-17} + f_{Jan-18}}$$

In the other hand, when the N-1 Accuracy is taken into the consideration, for the same above case the accuracy will be checked in November-17. Thus,[9]

$$FACC_{N-1} = 1 - \frac{|f_{Nov-17} - A_{Nov-17}|}{f_{Nov-17}}$$

Apart from the above forecasting measures, some of the industries also use the state-of art accuracy measurement metrics which is defined as below:[9]

$$Accuracy = 1 - \frac{|f_{Sep-17} - A_{Sep-17}|}{f_{Sep-17}}$$

Here, we will show the results of our work in both the accuracy measures and how our improved methodology will prove to be better in both the cases. Instead of checking the model performance in only one month, we will also concentrate in the stability of the models selected for the products or stock keeping-unit (SKU). The most important is the model selected should be good in bias and variance at the same time.[9]

**Two types of forecasting:**

**Under Forecasting:** [9]

$$\Rightarrow FACC = 1 - \frac{a - f}{f}$$

$$\Rightarrow FACC = \frac{f - a + f}{f}$$

$$\Rightarrow FACC = \frac{2f - a}{f}$$

**Over Forecasting:**[9]

$$\Rightarrow FACC = 1 - \frac{(f - a)}{f}$$

$$\Rightarrow FACC = \frac{f - f + a}{f}$$

$$\Rightarrow FACC = \frac{a}{f}$$

**There are two types of model in forecasting used in paper to analyze over and under forecasting:**[9]

1. Time series based
2. Regression based

**If both are over-forecasting:** This sum takes the actual forecast more deviating from the actuals.[9]

**If both are under-forecasting:** This takes the sum more below the actuals and towards the negative forecast accuracy.[9]

**If one over-forecasting and another under-forecasting:** This tries to minimize the error or deviation.[9]

Here in this scenario, the two different levels of forecasting (one under and another over) drive the accuracy badly. In the second case, if the accuracy measure is:[9]

$$FACC = 1 - \frac{abs(\sum_{i=1}^{k}(forecast_i - actuals_i))}{\sum_{i=1}^{k} actual_i}$$

Here the only non-constant term will be the deviation part i.e. as the forecast is the variable term. Also, the above accuracy measure if k=1 is the accuracy measure for most of the industries doing the demand planning.[9]

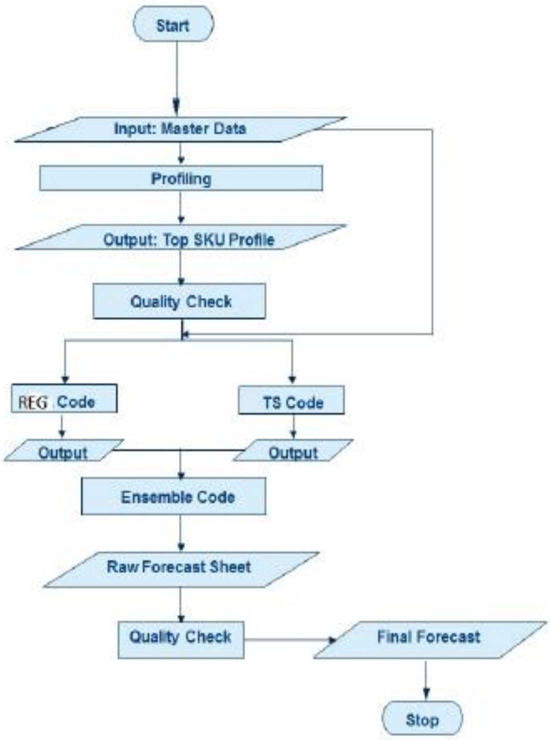


**Figure 3:**Forecasting flowchart model**.**

In the above figure, the model is basically a kind of neural network that decides the weightage for each model based upon the deviation/errors generated in the previous months. Here the weightage is decided on the errors generated in the last three months of the check of the forecast against the actuals. Hence, unlike the usual ensemble where the weightage is decided in the validation part of the model, the weightage here decided on the errors the model generates in the previous history. Thus, the weightage comes into the picture after the greedy selection of the model. Below are the steps followed in this method:[9]

**Step 1: Data Processing:**

The outliers are treated using the Median-Absolute Deviation Technique (MAD). We check whether the sales of a month are outliers or not and not take seasonal points into consideration. The first 6 points and the last 6 points are considered, and MAD is applied. The points below and above the median are marked. Seasonality is checked.[9]

The points below median are calculated using: -
Median -1.5*(median-absolute-deviation)

The extreme points are calculated using: -
Median + 2* (median – absolute - deviation).

**Step 2: Dataset Creation:**
If the forecast is to be generated for the month of September'17, the files to be considered are: [9]

1. Dataset with actuals till September'17: dt
2. Dataset with actuals till August'17: dt-1
3. Dataset with actuals till July'17: dt-2
4. Dataset with actuals till June'17: dt-3

**Step 3: Model Run:**
Model Run- All the datasets di, for i = t, t −1, t − 2, t − 3, is input to the two defined above models.[9]

**Step 4: Forecast generation:**
The output generated for the two models are:[9]

1. f1i, i = t + 1,..., t + 17 from Time Series Model [9]
2. f2i, i = t + 1,..., t + 17 from Regression Based Model[9]

**Step 5: Weight Generation:**
Weights for each of the model is found out from the actuals in the model. Let a1 be the actuals for the data. [9]

$$w_{ki} = \frac{\frac{1}{error_{ki}}}{\sum_{k=1}^{2}(error_{ki})^{-1}} \qquad w_{k_{t+1}} = average(\sum_{i=1}^{3} w_{ki})$$

**Step 6: Final Forecast Generation:**
The weights generated in the above step is multiplied with the respective forecast generated by the model to give the final forecast output. In the next final step for the final forecast generation, some market intelligence is incorporated to give the Final Forecast. [9]

**Table no 5:** Data mining.[7]

| Comparative discipline | Dimension of interest | Predictive analytics research (examples) | |
|---|---|---|---|
| | | Relevant | Less relevant |
| Forecasting | Predicting the future | Using forecasting techniques for evaluating what would have happened under different circumstances | Deriving generalized estimators of seasonal factors |
| Data mining | Search for patterns and relationships between a large number of variables with lots of data | Data mining preceded by logical and theoretical descriptions of possible relationships and patterns | Gibbs posterior for variable selection in data mining |

Data mining has also been used for demand forecasting in conjunction with traditional forecasting techniques (Rey, Kordon, & Wells, 2012). Usually, the data-mining step precedes the use of causal forecasting techniques by finding appropriate demand drivers (i.e., independent variables) for a product that can be used in regression analysis. For example, Dow Chemical uses a combination of data mining and regression techniques to forecast demand at the strategic and tactical levels (e.g., identifying demand trends), which is useful for its pricing strategy and for configuring and designing its supply chain to respond to these trends (Rey & Wells, 2013). Data-mining methods usually involve clustering techniques. So, if a retailer finds out, for example, that demand for cereal is strongly related to milk sales, then the retailer may build a causal forecasting model that predicts cereal sales with milk sales as one of the predicting variables. Market basket analysis is a specific data-mining technique that provides an analysis of purchasing patterns at the individual transaction level, so a retailer can analyze the frequency with which two product categories (e.g., DVDs and baby products) are purchased together. Lift for a combination of items is equal to the actual number of times the combination occurs in a given number of transactions divided by the predicted number of times the combination occurs if items in the combination were independent Lift values above 1 indicate that items tend to be purchased together. This kind of analysis can be useful when building causal regression models for demand forecasting. It can also aid in promotion activities because the retailer can predict how much sales of Product 1 would increase if there is a promotion for Product 2 if the two products are often purchased together.[8]

**B. METHOD II: PORTFOLIO**

Expanding the offered product portfolio is a typically chosen measure to address varying customer requirements

From different market segments. This adjustment to market requirements to expand product portfolio increases internal complexity. The more variants a company provides to the market, the more complex and difficult the planning and control of its supply chain and value creation process may get. The main objective of Product Variety Management (PVM) is to reduce variety-induced complexities and their associated costs. [12,13]
Data mining methods have been applied in many applications in PVM, especially in the design of product Architectures, product families, and product platforms, as well as in group technology for part family and Manufacturing cellular formation. Clustering methods could make a significant contribution to reduce decision making complexity in production network design.[12]
Clustering methods could significantly contribute to the ability of handling and reducing complexities in Strategic and tactical planning decisions associated with the design of a company's production network. The idea is to partition the product portfolio into clusters of variants with similar requirements regarding production Capacities and capabilities in order to obtain a simplified but more practical decision-making basis.[12]

**Algorithm to assign data object to cluster:**[12]

```
Algorithm
Input:
    •   k: # clusters
        A: set of data objects
Output:
    •   data object assignment to clusters
Method:
 1:  randomly select k objects as initial
     cluster representatives
 2:  assign all non-representative objects to
     nearest cluster representative
 3:  calculate cost and swap cluster
     representative with object that minimizes
     cost function within the cluster
 4:  repeat
 5:     assign all non-representative objects
        to nearest cluster representative
 6:     recalculate cost and swap cluster
        representative with object that
        minimizes cost function within the
        cluster
 7:  until cost does not change anymore
 8:  return data object assignment to clusters
```

The cost computed in lines 3 and 6 is calculated as follows: Let CR be the set of current clusters representatives $\{r_1, \dots, r_k\}$ and $NR = V \setminus CR$ the set of non-representative objects $nr_1, \dots, nr_{p-k}$. for all clusters $C_1, \dots, C_k$ calculate the sum of distances between all nonrepresentative objects within the cluster and the cluster representative. The sum over all clusters defines the current cost of clustering, as illustrated below

$$Cost = \sum_{i=1}^{k} \sum_{j=1}^{p-k} DM(r_i, nr_j) * \theta(r_i, nr_j)$$

With,

$$\theta(r_i, nr_j) = \begin{cases} 1, if\ nr_j\ is\ currently\ assigned\ to\ r_i \\ 0, otherwise \end{cases}$$

The algorithm works differently compared to existing k-medoids algorithms such as the "Partitioning around Medoids"-Algorithm (PAM) described by Kaufmann and Rousseeuw. The main difference is when updating the representative object, no random non-representative object is chosen as the new representative, and rather the object that minimizes clustering cost within the cluster is selected instead. Therefore, the developed clustering algorithm can be entitled "k-Minimizing-Medoids" (kMM).[12]

We propose to integrate the characteristics of TRIZ into the product design process and establish “product development portfolio” to manipulate the design of key issues and quality tracking.[13]

TRIZ is a Russian acronym, translated in English as “Theory of Inventive Problem Solving”. The TRIZ theory was mainly developed by Russian scientist G. Altshuller in 1946.[13]

There are various methods and tools in TRIZ, including 40 Inventive Principles, Contradiction Matrix, Separation Principles, Patterns of Evolutions, Substance-Field, Ideality, ARIZ, etc.[13]

TRIZ theory is founded on four pillars of philosophy which are “Functionality”, “Resource”, “Contradiction”, and "Ideality". To integrate TRIZ into the product development process, the framework must be considered from these four pillars.[13]

**Functionality:** Product design has to meet the design requests to fulfil the functionality of the product. Therefore, the way to express design requests should be planned, so that product development team members may be easy to communicate each other and analyze the product contents. We propose levels of expression as design objectives, design requirements and design parameters.[13]

**Resources:** TRIZ emphasizes the concept of resource utilization to solve problems. We therefore utilize TRIZ solving tools, such as D. Mann’s 37 trend lines and scientific effects, to be the applicable resources for product analysis; In addition, group technology can be used to help find the relevant factors and clustering relationship in the design.[13]

**Contradiction:** In the contradiction analysis, 48 engineering parameters and 31 management parameters are considered as the gene of product design such that 40 innovative principles can be applied whenever needed to solve the problem.[13]

**Ideality:** The product development is processed with the “Ideal Final Results (IFR)” in mind, and then gradually deal with the conflicts in between. The overall design quality is improved according to the ideality which is increasing as well.[13]

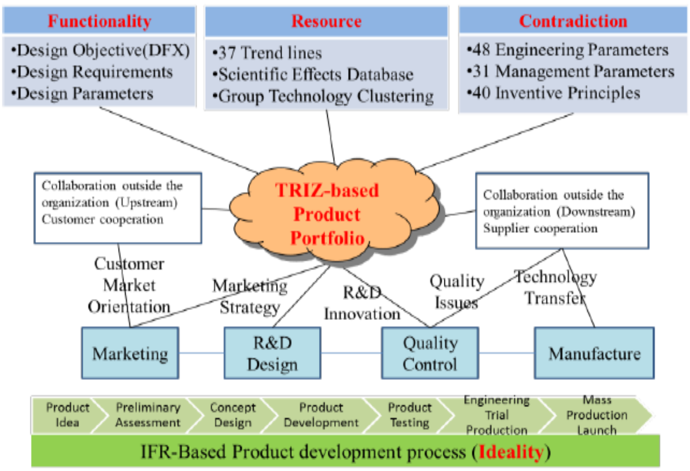


**Figure 4**: TRIZ-based framework for product development.[13]

**Parameterization of product design portfolio:**

Product design is considered and connected to the engineering/management parameters which are developed based on TRIZ theory. The scope involved in the product development process, such as marketing, design, quality, etc. are linked to the TRIZ parameters to deliberate. The 39 engineering parameters originally defined in classical TRIZ may be insufficient to describe current technology, so we adopt 48 engineering parameters. These parameters are applied to describe the engineering and managerial issues faced during the product development process, and then contradiction matrix can be applied whenever required to analyze or solve the problem. As for how to describe the content of product development, prior classification and definition play a very important role. We must establish the relationships between these parameters and the factors during product development process, so the construction of “product design portfolio” is discussed below:[13]

**Table no 6:** 48 Engineering Parameters in TRIZ.[13]

| | | | |
|---|---|---|---|
| 1 | Weight of moving object | 25 | Loss of Substance |
| 2 | Weight of non-moving object | 26 | Loss of Time |
| 3 | Length of moving object | 27 | Loss of Energy |
| 4 | Length of stationary object | 28 | Loss of Information |
| 5 | Area of moving object | 29 | Noise |
| 6 | Area of stationary object | 30 | Harmful Emissions |
| 7 | Volume of moving object | 31 | Other Harmful Effects Generated by System |
| 8 | Volume of stationary object | 32 | Adaptability/Versatility |
| 9 | Shape | 33 | Compatibility/Connectivity |
| 10 | Amount of Substance | 34 | Trainability/Operability/Controllability |
| 11 | Amount of Information | 35 | Reliability/Robustness |
| 12 | Duration of Action of Moving Object | 36 | Repairability |
| 13 | Duration of Action of Stationary Object | 37 | Security |
| 14 | Speed | 38 | Safety/Vulnerability |
| 15 | Force/Torque | 39 | Aesthetics/Appearance |
| 16 | Energy used by moving object | 40 | Other Harmful Effects Acting on System |
| 17 | Energy used by stationary object | 41 | Manufacturability |
| 18 | Power | 42 | Manufacturing Precision/Consistency |
| 19 | Stress/Pressure | 43 | Automation |
| 20 | Strength | 44 | Productivity |
| 21 | Stability | 45 | System Complexity |
| 22 | Temperature | 46 | Control Complexity |
| 23 | Illumination intensity | 47 | Ability to Detect/Measure |
| 24 | Function Efficiency | 48 | Measurement Precision |

**Table no 7:** Management parameters in TRIZ.[13]

| | | | |
|---|---|---|---|
| 1 | R&D Spec/ Quality/ Capability | 17 | Support Cost |
| 2 | R&D Cost | 18 | Support Time |
| 3 | R&D Time | 19 | Support Risk |
| 4 | R&D Risk | 20 | Support Interfaces |
| 5 | R&D Interfaces | 21 | Revenue/ Demand/ Feedback from Customer |
| 6 | Production Spec/ Quality/ Means | 22 | Amount of Information |
| 7 | Production Cost | 23 | Communication Flow |
| 8 | Production Time (8) | 24 | System Affected Harmful Effects |
| 9 | Production Risk | 25 | System Generated Harmful Effects |
| 10 | Production Interfaces | 26 | Convenience |
| 11 | Supply Spec/ Quality/ Means | 27 | Adaptability/ Versatility |
| 12 | Supply Cost | 28 | System Complexity |
| 13 | Supply Time | 29 | Control Complexity |
| 14 | Supply Risk | 30 | Tension/ Stress |
| 15 | Supply Interfaces | 31 | Stability |
| 16 | Product Reliability | | |

The classification hierarchy contains four levels: Design Intention, Design Requirement, Design Parameter and Engineering Parameter. These levels are explained as follows:[13]

First level- Design Intentions (DIs): We apply the concept of Design for excellence (DFX) as the first level to catch the collaborative team member's purposes on his/her design demands. The common DFXs used in this level are set as Design for Assembly (DFA), Design for Manufacturability (DFM), Design for Environment (DFE) and Design for Customer (DFC). However, the DFXs can be expanded if necessary. [13]

Second level- Design Requirements (DRs): These are the explicit requests that certain team member demands along with his/her purposes on DFXs. The requirements should be concrete and meaningful.[13]

Third level- Design Parameters (DPs): Considering the DRs, these are the product specifications which are usually related to the parts or components. In such way the design conflicts may be identified accordingly.[13]

Fourth level- Engineering Parameters (EPs): These are the same as those EPs defined in TRIZ theory as shown in Table I. The further contradiction analysis may be applied to resolve the possible design problems. [13]

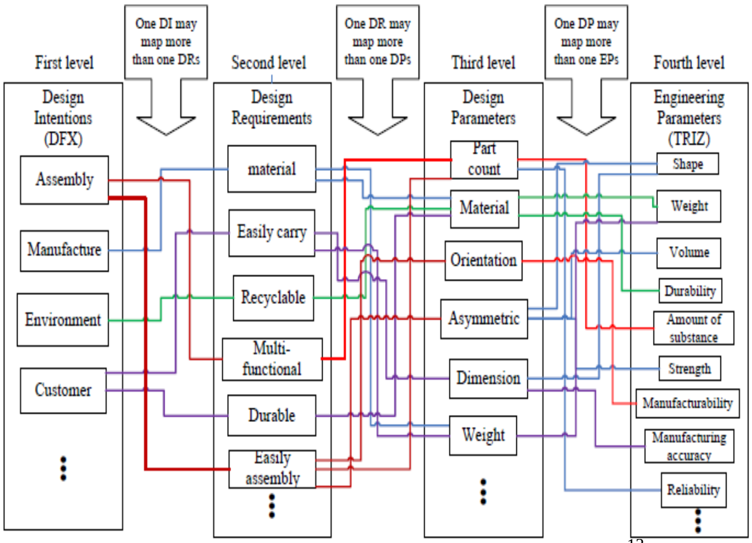


**Figure 5:** The Four Level Structure Classification.[13]

## V. Result and Discussion

### A. Method 1: Forecasting

**Table no 8:** Accuracy comparison using FACC measure.[9]

| Month-MT | #SKU's | TS FACC | Reg FACC | Ensemble FACC |
|---|---|---|---|---|
| Jun-MT 17 | | 69% | 65% | 71% |
| Jul-MT 17 | 2909 | 62% | 62% | 67% |
| Aug-MT 17 | | 62% | 63% | 65% |

In the above table, the total number of SKU's that we have taken into the consideration is 2909. The accuracy measure in the above table is

$$FACC = 1 - \frac{\sum_{i=1}^{k}(forecast_i - actuals_i)}{\sum_{i=1}^{k} forecast_i}$$

Here, k is 3. For the Jun-MT 17, the months considered for the weight's generation is the previous months. Thus, for the results generation, if the actuals will be till January'17 and the forecast will be February'17 onwards and the accuracy will be calculated for the months of April'17, May'17 and June'17 absolute sum of deviations from the consecutive actuals sum. From the above results we can see that that the proposed method is proving to be better than the individual model with a significant difference. Also, the stability is better as it is consistent in the time frame as well.[9]

**Table no 9:** Accuracy comparison using accuracy measure.[9]

| Month-MT | #SKU's | TS Accuracy | Reg Accuracy | Ensemble Accuracy |
|---|---|---|---|---|
| Jun-MT 17 | | 61% | 57% | 64% |
| Jul-MT 17 | 2909 | 54% | 51% | 58% |
| Aug-MT 17 | | 63% | 60% | 64% |

Here in the above table the accuracy measure that has been taken into the account is

$$Accuracy = 1 - \frac{\sum_{i=1}^{k}(forecast_i - actuals_i)}{\sum_{i=1}^{k} actual_i}.$$

Thus, we can see that even in the above table, our proposed methodology for Ensemble outperforms in comparison to the individual models which are Time Series Model and Regression Model. Also, the above results prove that the proposed solution is stable in the span of the three months.[9]

**Table no 10:** Accuracy comparison for optimizing the weight measure.[9]

| N | N-1 | MT |
|---|---|---|
| 62.47% | 66.20% | 66.95% |

For considering which time zone or time period to be taken care for the optimization of the generation of the weights, the above results show that optimizing using the N-1 i.e. two months after the forecast generation and for the MT, are comparable. In this case we have taken the results using the N-1 optimizing the weights.

**B. Method 2: Product Portfolio**

To validate the developed clustering method, it is executed on a real-world heart disease data set taken from the UCI machine learning repository. This data features a mixed data set with eight categorical and five numeric features and contains 297 data objects belonging to two clusters. To apply the developed clustering method, categorical attributes have been transformed into binary ones resulting in 19 auxiliary variables. Table I depicts the results for silhouette coefficient, Rand index, and BCubed metrics for the kMM application on the heart disease data set.[12]

**Table no 11:** Clustering quality on generated data set for different values of α.[9]

| α | Sil. Coeff. | Rand index | Precision | Recall |
|---|---|---|---|---|
| 0.1 | 0.8254 | 0.8957 | 0.5215 | 0.6309 |
| 0.2 | 0.8169 | 0.9003 | 0.5443 | 0.6419 |
| 0.3 | 0.8267 | 0.9027 | 0.5611 | 0.6424 |
| 0.4 | 0.9542 | 0.9074 | 0.5861 | 0.6587 |
| 0.5 | 0.9558 | 0.9074 | 0.5861 | 0.6587 |
| 0.6 | 0.9561 | 0.9095 | 0.5925 | 0.6673 |
| 0.7 | 0.9656 | 0.9320 | 0.6942 | 0.7204 |
| 0.8 | 0.9828 | 0.9514 | 0.7795 | 0.7898 |
| 0.9 | 0.9816 | 0.9527 | 0.7803 | 0.7996 |
| 1.0 | 0.9803 | 0.9094 | 0.5983 | 0.6339 |

The Silhouette coefficient ranges from 82.54% to 98.28%. With increasing α from 0.1 to 0.8 it increases continuously, reaching its peak α= 0.8. Coefficient values for α= 0.9 and α= 1 are slightly lower. Rand index ranges between 89.57% and 95.27%, peaking at α= 0.9. For values of α between 0.7 and 0.9 the Rand index is significantly higher than for $0.1 \leq \alpha \leq 0.7$ and $\alpha \geq 0.9$.[12]

Precision ranges between 52.15% and 78.03%, peaking at α= 0.9. However, with a span of 25.88% precision features large variations. Recall, reaching its maximum value also at α= 0.9 ranging between 63.09% and 79.96% with a span of 16.87%. The clustering results for α= 0.9 are plotted in Figure 6 using multidimensional scaling:[12]

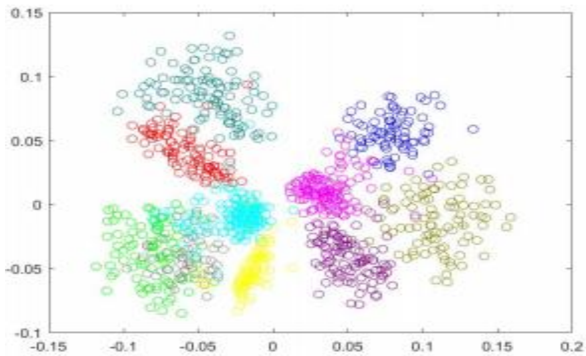


**Figure 6:**Clustering result for α= 0.9[12]

**Table no 13:**Final Result.

| | Method 1 | | | Method 2 | | |
|---|---|---|---|---|---|---|
| Parameters | Forecasting | | | Product portfolio | | |
| | H | M | L | H | M | L |
| Accuracy | ✓ | | | | ✓ | |
| Time | | ✓ | | | ✓ | |
| Cost | ✓ | | | | | ✓ |
| Speed | | | ✓ | | ✓ | |
| Amount of information | ✓ | | | ✓ | | |

## VI. Conclusion

As observed from literature survey of research paper, it can be seen that people are finding more and more areas where analytics can be applied for solving business as well as technology issues in supply chain management. This application is in multiple domains such as manufacturing pharmaceutical etc. but like every immerging technology people are defining the application and a lot of work is being done and as time progresses analytic is likely to find a lot of usage too solve chain problem as it has potential to solve this issue effectively.

## References

[1]. Chopra, S et. al., 2016, Supply Chain Managements – Strategy, Planning and Operation-Sixth Edition, India, Pearson Publishing.
[2]. Camm, et.al., 2015, Essentials of Business Analytics –First Edition, Stamford,CT, Cengage Learning.
[3]. Supply Chain Business Solutions,ASCM 2020. Accessed on: January 2020. [Online]. Available: http://www.supply-chain.org/.
[4]. Inspirage -The Integrated Supply Chain Specialists, Inspirage, 2020.Accessed on: January 2020. [Online].Available: http://www.inspirage.com.
[5]. Kiran Bala, Supply Chain Management: Some Issues and Challenges-A Review. International Journal of Current Engineering and Technology, Vol.4, No.2, April 2014, 946-953.
[6]. Ales Groznik, Peter Trkman, "Current issues and challenges of supply chain management",International Conference on automation and information,2017,243-248.
[7]. Fawcett, Waller, "Journal of Business Logistics",2013,34(2): 77-84.
[8]. Gilvan C. Souza, "Supply Chain Analytics", HBS Publishing, 57, 2014, 595-605.
[9]. Nimai Chand Das Adhikari, Rajat Garg, ShaivyaDatt, Lalit Das, Shrinivas Deshpande and Dr. Ashutosh Mishra. "Ensemble Methodology for Demand Forecasting." International Conference on Intelligent Sustainable Systems, 2017, 846-851.
[10]. Jens Leveling, Matthias Edelbrock and Boris Otto. "Big Data Analytics for Supply Chain Management", Audi-Endowed Chair of Supply Net Order Management: Technical University of Dormund, 14, 2014, 918-922.
[11]. Moritz Kynast, Olivera Marjanovic "Big data in Supply Chain Management- Applications, Challenges and Benefits" Twenty-second Americas conference on information system, San Diego,2016,1-10.
[12]. J. Hochdorffer, C. Laule and G. Lanza. "Product variety management using Data-mining Methods-Reducing Planning Complexity by Applying Clustering Analysis on Product Portfolios", Wbk institute of Production Science, Karlsruhe Institute of Technology (KIT), Karlsruhe Germany, 17,2017,593-597
[13]. Tien-Lun Liu and Jhih-Wei Chen" A Novel Scheme for Constructing Product Development Portfolio Based on TRIZ" IEEE International Conference on Applied System Innovation, 2017, 1289-1292.

Sanskriti Upadhyaya, Ayusha Kakkad, et. al. "A Review on Data Analytics in Supply Chain Management using Forecasting and Product Portfolio techniques." *IOSR Journal of Computer Engineering (IOSR-JCE)*, 22(5), 2020, pp. 53-68.