# Improving Supply Chain Efficiency in the Market via Predictive Analysis Techniques

Deekshita Prakash Savanur, Gouri Benni, Krishna Sameera Surapaneni, Sonali Arcot

Department of Applied Data Science, San Jose State University

DATA 270 : Data Analyst Process

Dr. Eduardo Chan

February 17, 2023

## 1. Proposed Research Problem

### *1.1 Problem Statement*

In the current unbounded global universe supply chain management is key to ensure proper shipment of goods and delivery to the end user efficiently with minimized operational risks. However, many companies fail to achieve these end-to-end operations, as they depend upon multiple suppliers and vendors, each with their own data process and governance leading to poor collaboration and customer dissatisfaction.

Therefore, the problem is how to achieve flawless supply chain visibility from raw materials to the end consumer with readily available goods in the warehouse when in demand. This could culminate in long lead times, late shipment, inaccurate forecasting, poor data visibility and high logistics costs. This research aims to provide insights on how to improve supply chain efficiency through the use of predictive analytics, dashboard visibility and ultimately building a forecasting model which can predict demand and efficiently regulate the supply of goods in the warehouse, leveraging technologies like IoT and machine learning.

### *1.2. Proposed ML Models*

Statistical forecasting models like exponential smoothing, double exponential smoothing, and ARIMA models, along with regression analysis models such as linear regression and adaptive boosting, can help to improve supply chain efficiency in the market via predictive analytics techniques. These models can be applied to historical data on sales, inventory, and production to identify trends, seasonality, and other patterns. Exponential smoothing and double exponential smoothing are time-series models that are commonly used to forecast demand, while ARIMA models can capture both time-series and external factors that influence demand.

Regression analysis models like linear regression can be used to identify the relationships between different variables, such as demand and price, while adaptive boosting can help to improve the accuracy of these models by combining multiple weaker models. By analyzing these models, companies can make data-driven decisions about inventory levels, production schedules, and supply chain optimization, reducing the risk of stockouts and overstocking. Furthermore, these predictive analytics techniques can help companies to identify potential supply chain disruptions and enable them to take preventive measures, such as alternate sourcing or safety stock planning, to minimize the impact of such disruptions. Overall, the application of statistical forecasting and regression analysis models in predictive analytics can significantly improve supply chain efficiency and enable companies to make better-informed decisions in the market.

## 2. Background Information

The supply chain industry is an essential part of modern life, facilitating the efficient and timely transportation of goods from manufacturers to consumers. This industry encompasses a range of activities, including procurement, transportation, warehousing, and distribution, that work in tandem to bring products to market. Our daily lives rely heavily on the supply chain to deliver the food we eat, the clothes we wear, the electronics we use, and many other products and services. Furthermore, the supply chain is a vital contributor to the economy, providing jobs, revenue, and supporting international trade. By optimizing the supply chain, companies can improve efficiency, reduce costs, and increase customer satisfaction, ultimately leading to a better quality of life for everyone.

To stay competitive, manufacturers and retailers across all industries must accurately predict product demand. Machine learning can be especially useful in forecasting sales for short-shelf-life and highly-perishable items, including those found in the fresh food, technology, and fashion sectors. Compared to traditional statistical methods, machine learning can provide a greater level of accuracy in sales forecasting, leading to improved inventory management throughout the supply chain. This, in turn, can reduce the incidence of stockouts, improve product availability for consumers, and increase profitability. Therefore, machine learning offers a considerable advantage in predicting demand for various products.

Previously, to ensure prompt stock replenishment, improved capacity management, and optimum sales and revenue, demand forecasting was crucial to effective supply chain management. Additionally, it enhances management and decision-making while advancing future growth and expansion plans. An in-depth examination of the variables that may have an impact on a company's supply infrastructure is necessary for accurate demand forecasting. To maintain the business preparedness, continuity, and a great end-user experience, anticipating demand necessitates rigorous research of a number of variables, from sales history patterns to specific occasions in the commercial calendar (such as Christmas).

Tarallo et al. (2019) in their study discovered that makers and merchants of fast-moving consumer items can gain advantages from machine learning approaches, including deep learning. The key advantage is improved demand forecasting accuracy, which can result in cost savings, increased sales, and more customer satisfaction. Comparisons with conventional statistical methods revealed improved sales forecasting, adaptability to handle more data variables, and processing power for big data volumes.

In supply chains with abundant demand data, Aviv (2003) suggests a unified time-series paradigm for forecasting and stock management. It permits the development of time-series models which can capture enhanced information spaces and offers forecasting and replenishment techniques that can handle a broad range of product demand characteristics. It also provides a simple framework for examining the advantages of various information-sharing strategies. The development of cost estimating processes and the analysis of information-rich supply chain systems with imperfect knowledge of the system characteristics are covered in this work. We can see from the cost expressions that the performance of a system, and in particular the performance of coordinated supply chains, is dependent on both the accuracy of the individual members' forecasts and the correlation between them.

## 3. Data Source/Datasets

### *3.1. Sample 1*

http://data.un.org/Data.aspx?d=ComTrade&f=_l1Code%3a19

UNdata_Export_20230213_030535077

| Country or Area | Year | Commodity | Flow | Trade (USD) | Weight (kg) | Quantity Name | Quantity |
|---|---|---|---|---|---|---|---|
| **Afghanistan** | 2018 | Generators and alternators | Import | 11395053.13 | 831888.682346215 | Number of items | 47830.29 |
| **Afghanistan** | 2018 | Electric instant, storage and immersion water heaters | Import | 14121694.08 | 1975848.36300013 | Number of items | 291781.99 |
| **Afghanistan** | 2018 | Electric space heating nes and soil heating apparatus | Export | 155902.25 | 0 | Number of items | 1650 |
| **Afghanistan** | 2018 | Electric space heating nes and soil heating apparatus | Re-Export | 502.25 | 0 | Number of items | 600 |
| **Afghanistan** | 2018 | Electric smoothing irons | Import | 4040529.56 | 240154.816843458 | Number of items | 802058 |
| **Afghanistan** | 2018 | Electric smoothing irons | Export | 1272.23 | 0 | Number of items | 250 |
| **Afghanistan** | 2018 | Electric smoothing irons | Re-Export | 1272.23 | 0 | Number of items | 250 |
| **Afghanistan** | 2018 | Electro-thermic appliances, domestic, nes | Import | 1975720.14 | 154764.675817025 | Number of items | 74633.86 |
| **Afghanistan** | 2018 | Telephone sets | Import | 6344.26 | 124.097909558753 | Number of items | 352 |
| **Afghanistan** | 2018 | Unrecorded sound recording media except photo/magneti | Import | 234148.06 | | No Quantity | 0 |
| **Afghanistan** | 2018 | Transmit-receive apparatus for radio, TV, etc. | Import | 13846962.1 | 21675.9561241523 | Number of items | 436345.4 |
| **Afghanistan** | 2018 | Transmit-receive apparatus for radio, TV, etc. | Export | 23663.67 | 0 | Number of items | 181 |

***Fig 3.1.*** *Dataset snapshot*

The aforementioned dataset, gives users access to information about global trade for different nations. The Commodity Trade Statistics Database (COMTRADE), a database of information about international trade, served as the basis for the dataset, which was created by the United Nations Statistics Division. This dataset has 100,000 rows and 8 columns which includes both categorical and numerical variables such as the country or territory, year, data flow, commodity, trade, weight, quantity name, and quantity.

### *3.2. Sample 2*

https://archive.ics.uci.edu/ml/datasets/Daily+Demand+Forecasting+Orders

Daily_Demand_Forecasting_Orders

| Day of the week (Monday to Friday) | Non-urgent order | Urgent order | Order type A | Order type B | Order type C | Fiscal sector orders | Orders from the traffic controller sector | Banking orders (1) | Banking orders (2) |
|---|---|---|---|---|---|---|---|---|---|
| 4 | 316.307 | 223.270 | 61.543 | 175.586 | 302.448 | 0 | 65556 | 44914 | 188411 |
| 5 | 128.633 | 96.042 | 38.058 | 56.037 | 130.580 | 0 | 40419 | 21399 | 89461 |
| 6 | 43.651 | 84.375 | 21.826 | 25.125 | 82.461 | 1.386 | 11992 | 3452 | 21305 |
| 2 | 171.297 | 127.667 | 41.542 | 113.294 | 162.284 | 18.156 | 49971 | 33703 | 69054 |
| 3 | 90.532 | 113.526 | 37.679 | 56.618 | 116.220 | 6.459 | 48534 | 19646 | 16411 |
| 4 | 110.925 | 96.360 | 30.792 | 50.704 | 125.868 | 79 | 52042 | 8773 | 47522 |
| 5 | 144.124 | 118.919 | 43.304 | 66.371 | 153.368 | 0 | 46573 | 33597 | 48269 |
| 6 | 119.379 | 113.870 | 38.584 | 85.961 | 124.413 | 15.709 | 35033 | 26278 | 56665 |
| 2 | 218.856 | 124.381 | 33.973 | 148.274 | 162.044 | 1.054 | 66612 | 19461 | 103376 |
| 3 | 146.518 | 101.045 | 36.399 | 43.306 | 168.723 | 865 | 58224 | 7742 | 82395 |
| 4 | 178.433 | 102.793 | 45.706 | 111.036 | 124.678 | 194 | 47046 | 17299 | 108719 |
| 5 | 145.865 | 91.180 | 43.851 | 66.277 | 133.440 | 6.523 | 66910 | 17768 | 36693 |
| 6 | 170.566 | 114.412 | 43.339 | 136.434 | 128.405 | 23.200 | 32529 | 34002 | 78153 |
| 2 | 220.343 | 141.406 | 46.241 | 120.865 | 196.296 | 1.653 | 34878 | 32905 | 117137 |
| 3 | 193.768 | 141.854 | 56.519 | 136.709 | 143.644 | 1.250 | 57858 | 23956 | 101048 |

***Fig 3.2***. *Dataset snapshot*

The Daily Demand Forecasting Orders dataset is a structured dataset containing data related to daily demand for products in a Brazilian company. The data is stored in CSV format and contains 60 rows and 13 columns. The dataset includes both numerical and categorical data types, including columns such as order date, product ID, order quantity, and delivery date. This dataset is intended for use in forecasting demand for products in an e-commerce business and includes data collected over a period of five months. It can be used to explore various aspects of demand forecasting, such as time-series analysis and regression modeling.

### *3.3. Sample 3*

https://data.mendeley.com/datasets/8gx2fvg2k6/5#:~:text=Dataset%20of%20Supply%20Chain%2C%20which%20allows%20the%20use,Structured%20Data%20with%20Unstructured%20Data%20for%20knowledge%20generation

| Type | Delivery_Status | Customer_City | Customer_Country | Customer_Segment | Customer_State | Customer_Street | Department_Name | Market | Order_City | Order_Country | Order_Regio |
|---|---|---|---|---|---|---|---|---|---|---|---|
| DEBIT | Advance shipping | Caguas | Puerto Rico | Consumer | PR | 5365 Noble Nectar Island | Fitness | Pacific Asia | Bekasi | Indonesia | Southeast Asi |
| TRANSFER | Late delivery | Caguas | Puerto Rico | Consumer | PR | 2679 Rustic Loop | Fitness | Pacific Asia | Bikaner | India | South Asia |
| CASH | Shipping on time | San Jose | EE. UU. | Consumer | CA | 8510 Round Bear Gate | Fitness | Pacific Asia | Bikaner | India | South Asia |
| DEBIT | Advance shipping | Los Angeles | EE. UU. | Home Office | CA | 3200 Amber Bend | Fitness | Pacific Asia | Townsville | Australia | Oceania |
| PAYMENT | Advance shipping | Caguas | Puerto Rico | Corporate | PR | 8671 Iron Anchor Corners | Fitness | Pacific Asia | Townsville | Australia | Oceania |
| TRANSFER | Shipping canceled | Tonawanda | EE. UU. | Consumer | NY | 2122 Hazy Corner | Fitness | Pacific Asia | Toowoomba | Australia | Oceania |
| DEBIT | Late delivery | Caguas | Puerto Rico | Home Office | PR | 1879 Green Pine Bank | Fitness | Pacific Asia | Guangzhou | China | Eastern Asia |
| TRANSFER | Late delivery | Miami | EE. UU. | Corporate | FL | 7595 Cotton Log Row | Fitness | Pacific Asia | Guangzhou | China | Eastern Asia |
| CASH | Late delivery | Caguas | Puerto Rico | Corporate | PR | 2051 Dusty Route | Fitness | Pacific Asia | Guangzhou | China | Eastern Asia |
| CASH | Late delivery | San Ramon | EE. UU. | Corporate | CA | 9139 Blue Blossom Court | Fitness | Pacific Asia | Guangzhou | China | Eastern Asia |
| TRANSFER | Shipping canceled | Caguas | Puerto Rico | Corporate | PR | 4058 Quiet Heights | Fitness | Pacific Asia | Tokio | Japón | Eastern Asia |
| TRANSFER | Late delivery | Freeport | EE. UU. | Consumer | NY | 3243 Shady Corner | Fitness | Pacific Asia | Manado | Indonesia | Southeast Asi |
| TRANSFER | Late delivery | Salinas | EE. UU. | Corporate | CA | 131 Sunny Treasure Green | Fitness | Pacific Asia | Manado | Indonesia | Southeast Asi |
| DEBIT | Late delivery | Caguas | Puerto Rico | Corporate | PR | 2531 Wishing Square | Fitness | Pacific Asia | Sangli | India | South Asia |
| TRANSFER | Late delivery | Peabody | EE. UU. | Corporate | MA | 6417 Silver Towers | Fitness | Pacific Asia | Sangli | India | South Asia |
| DEBIT | Late delivery | Caguas | Puerto Rico | Corporate | PR | 257 Harvest Close | Fitness | Pacific Asia | Sangli | India | South Asia |
| PAYMENT | Late delivery | Canovanas | Puerto Rico | Corporate | PR | 7342 Hazy Beacon Park | Fitness | Pacific Asia | Seúl | Corea del Sur | Eastern Asia |

***Fig 3.3.*** *Dataset snapshot*

The provided dataset contains comprehensive information on order details, spanning from the pricing of products to the specifics of shipping. This data is structured in CSV format, comprising 50+ columns and 180,520 rows. It encompasses both categorical and quantitative data, with columns including order status, product name, shipping method, and product price. However, there are a few instances of missing data that will need to be addressed. Once this data has been cleaned, it will be a valuable resource for accurate forecasting and analyzing past trends.

### *3.4. Sample 4*

https://www.census.gov/data/tables/2015/econ/e-stats/2015-e-stats.html

| NAICS Code | Description | Value of Sales: 2015 Total Sales | Value of Sales: 2015 E-commerce | Value of Sales: 2014 Revised Total Sales | Value of Sales: 2014 Revised E-Commerce | Y/Y Percent Change: Total Sales | Y/Y Percent Change: E-commerce Sales | E-commerce as Percent of Total Sales: 2015 | E-commerce as Percent of Total Sales: 2014 | Percent Distribution of E-commerce Sales: 2015 |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Total Retail Trade** | **4,727,427** | **340,415** | **4,639,440** | **298,682** | **1.9** | **14.0** | **7.2** | **6.4** | **100.0** |
| 441 | Motor vehicles and parts dealers | 1,095,412 | 29,716 | 1,020,851 | 27,362 | 7.3 | 8.6 | 2.7 | 2.7 | 8.7 |
| 442 | Furniture and home furnishings stores | 106,779 | 691 | 99,718 | 599 | 7.1 | 15.4 | 0.6 | 0.6 | 0.2 |
| 443 | Electronics and appliance stores | 102,108 | 1,331 | 103,518 | 1,304 | -1.4 | 2.1 | 1.3 | 1.3 | 0.4 |
| 444 | Building materials and garden equipment and supplies stores | 331,644 | 1,708 | 318,352 | S | 4.2 | S | 0.5 | S | 0.5 |
| 445 | Food and beverage stores | 685,568 | 1,166 | 669,165 | 1,099 | 2.5 | 6.1 | 0.2 | 0.2 | 0.3 |
| 446 | Health and personal care stores | 315,257 | D | 299,263 | D | 5.3 | D | D | D | D |
| 447 | Gasoline stations | 443,817 | D | 538,790 | D | -17.6 | D | D | D | D |
| 448 | Clothing and clothing accessories stores | 255,831 | 4,107 | 250,409 | 3,801 | 2.2 | 8.1 | 1.6 | 1.5 | 1.2 |
| 451 | Sporting goods, hobby, book, and music stores | 87,355 | 2,265 | 85,466 | 2,329 | 2.2 | -2.7 | 2.6 | 2.7 | 0.7 |

***Fig 3.4.*** *Dataset snapshot*

The dataset available through the link provided contains economic statistics data that has been compiled by the U.S. Census Bureau. The data is presented in tables and charts and covers various economic indicators for different sectors of the economy, including retail trade, wholesale trade, services, manufacturing, and more. The dataset is organized into numerous tables and charts that present the economic statistics in various formats such as percentages, ratios, and absolute numbers. The dataset can be downloaded in various file formats, including Excel and CSV, to enable users to analyze and manipulate the data using statistical software or programming languages.

**References**

Aviv, Y., 2003. A time-series framework for supply-chain inventory management. Operations *Research*, *51*(2), pp.210-227. https://www.jstor.org/stable/4132401

Cadavid, J.P.U., Lamouri, S. and Grabot, B., 2018, July. Trends in machine learning applied to demand & sales forecasting: A review. In *International conference on information systems, logistics and supply chain*. https://hal.science/hal-01881362/document

Dhanush, G.A., Raj, K.S. and Kumar, P., 2021. Blockchain aided predictive time series analysis in supply chain systems. In *Innovations in Electrical and Electronic Engineering: Proceedings of ICEEE 2021* (pp. 913-925). Springer Singapore. https://www.springerprofessional.de/en/blockchain-aided-predictive-time-series-analysis-in-supply-chain/19192386

Khan, M.A., Saqib, S., Alyas, T., Rehman, A.U., Saeed, Y., Zeb, A., Zareei, M. and Mohamed, E.M., 2020. Effective demand forecasting model using business intelligence empowered with machine learning. *IEEE Access*, *8*, pp.116013-116023. https://www.researchgate.net/publication/342326031_Effective_Demand_Forecasting_Model_Using_Business_Intelligence_Empowered_With_Machine_Learning

Parzen, E., 1961. An approach to time series analysis. *The Annals of Mathematical Statistics*, *32*(4), pp.951-989. https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-32/issue-4/An-Approach-to-Time-Series-Analysis/10.1214/aoms/1177704840.full

Schlegel, G.L., 2014. Utilizing big data and predictive analytics to manage supply chain risk. *The Journal of Business Forecasting*, *33*(4), p.11. https://www.proquest.com/docview/1657419126/fulltextPDF/C7AAA560ECD8424CPQ/1?accountid=10361

Sheremetov, L.B., González-Sánchez, A., López-Yáñez, I. and Ponomarev, A.V., 2013. Time series forecasting: applications to the upstream oil and gas supply chain. *IFAC Proceedings Volumes*, *46*(9), pp.957-962. https://www.sciencedirect.com/science/article/pii/S1474667016344123

Tarallo, E., Akabane, G.K., Shimabukuro, C.I., Mello, J. and Amancio, D., 2019. Machine learning in predicting demand for fast-moving consumer goods: An exploratory research. *IFAC-PapersOnLine*, *52*(13), pp.737-742. https://www.sciencedirect.com/science/article/pii/S240589631931153X

Zwißler, F. and Hermann, M., 2012. Supply chain risk management in the electronics industry. *Risk Management for the Future–Theory and Cases*, pp.467-496. http://www.doiserbia.nb.rs/Article.aspx?ID=0013-32641714121A#.Y-896i-U35g