Development of a data-driven approach for understanding discrepancies between design and implementation of inventory policies.

Sudharshan Bindiganavale Ramesh
Master of Science in Data Science
sb1501@rit.edu

Dr. Guiping Hu
Department Head
Golisano Institute for Sustainability
ghgis@rit.edu

Dr. Nenad Nenadic
Research Faculty
Golisano Institute for Sustainability
nxnasp@rit.edu

The goal of this project is to develop a framework which assists in minimizing the discrepancies that occur during implementation and the practice of existing inventory policies for an industrial partner.

Background: A business is essentially looking to come up with products to serve its customers. To build these products, several goods and raw-materials around the world have to be aggregated at a specific location where the materials can be processed and assembled into finished goods. Inventory management, a vital component of the Supply Chain is the systemic approach of sourcing, storing, and efficiently distributing raw materials and finished goods. Inventory management [1] helps in controlling and tracking of the materials involved from the suppliers, to the intermediate warehouses and all the way to the point of sale. Inventory management has a direct impact on the success of a business as too little inventory of a product when it is going through its high demand cycle can lead to unhappy customers which has an adverse effect on the business, and hence it is hypothesized that, that could potentially be one of the reasons for discrepancies between the design and implementation of inventory polices. Therefore, it is natural for businesses to invest a significant portion of their resources in research and development of optimizing inventory policy designs to mitigate uncertainty and maximize throughput with minimal overhead costs.

Broader Impact: Businesses before the pandemic (2020) had a supply chain which essentially was a broad network encompassing several continents and relied on third party suppliers to help them source some of the raw materials required to build their product. The pandemic was a major disruptor to this harmonious network, due to the travel and transport restrictions put in place by multiple countries and even the UN. Businesses lost customers, revenue and some even had to let its employees go. Supply Chain and Inventory Management now had a complex problem to tackle. In these past few years top businesses like Apple, Amazon, Walmart [2] etc., have invested enormous amounts into integrating their supply chain to effectively manage all aspects of their network including their inventory policies in order to be robust to external factors and establish contingency plans in place, in the case of unaccounted disruptions in the future. This indicates that, companies are investigating their existing supply chain networks and are carefully evaluating to essentially determine ways to optimise the network in order to help the business generate a net positive return-on-investment. Regarded as one of the top chemical companies in the world, our industrial partner is one among many businesses looking to implement effective strategies in balancing inventory targets to minimize the overhead costs.

Scientific Merit: The proposed work will focus on development of an intelligent analytic framework that will enable the inference and the interpretation needed to understand the current inventory management system in place for the industrial partner. Once a digital thread has been established and we have significant visibility of the ground truth - data analysis will be performed to extract meaningful information about the discrepancies which occur in the practice of the existing policies. Since our work is directly dependent on data provided by our partner and with the help of data manipulation libraries like Pandas and NumPy in Python along with Matplotlib for data representation we will be generating probabilistic models for hypothesis testing, with integrated model comparison to describe inventory policies that are actually practices by the inventory managers, with Bayesian model comparison to enable hypotheses testing. Therefore, this project will contain all the essential steps involved in an applied Data Science project.

Tasks and Goals:

1. Conduct literature survey on inventory optimization policies, methods of replenishment, and data-driven methods in probabilistic modeling. This activity will also include:

- Identifying and exploring existing datasets related to supply chain.
- Identifying and exploring existing libraries in Python scientific ecosystem (e.g. PyTorch-based Pyro [3], or TensorFlow-based Edward [4])

2. Using the available datasets from the existing literature, the aim is to generate synthetic sawtooth replenishment data which closely follows the expected dataset from our industrial partner. The reason for generating synthetic data is that most of the available datasets contain limited number of samples. We wish to generate these samples using popular methods such as generative adversarial networks (GANs) [5] or variational autoencoders (VAEs). [6]

3. Apply various state-of-art probabilistic models like Logistic Regression, Bayesian models, Hidden Markov Models [7] etc., on the generated synthetic data for hypothesis testing.

4. These trained models will then be applied to real-world data and the performance will be evaluated by formulating hypothetical discrepancies between design and implementation of inventory policies.

5. It is possible that the models trained on synthetic data will not perform optimally on the real-data. To further improve performance, we will fine tune the model using real-world data.

References:

[1] IBM. What is inventory management and how does it work? | IBM.

[2] Steve Banker. Walmart's massive investment in a supply chain transformation, 2021.

[3] Eli Bingham, Jonathan P Chen, Martin Jankowiak, Fritz Obermeyer, Neeraj Pradhan, Theofanis Karaletsos, Rohit Singh, Paul Szerlip, Paul Horsfall, and Noah D Goodman. Pyro: Deep universal probabilistic programming. *The Journal of Machine Learning Research*, 20(1):973–978, 2019.

[4] Dustin Tran, Alp Kucukelbir, Adji B Dieng, Maja Rudolph, Dawen Liang, and David M Blei. Edward: A library for probabilistic modeling, inference, and criticism. *arXiv preprint arXiv:1610.09787*, 2016.

[5] James McCaffrey. Generating synthetic data using a generative adversarial network (GAN) with PyTorch, 2021.

[6] James McCaffrey. Generating synthetic data using a variational autoencoder with PyTorch, 2021.

[7] Gathika Ratnayaka. Probability and Machine Learning? — Part 1- Probabilistic vs Non-Probabilistic Machine Learning Models, 2020.