EAS 504: Applications of Data Science – Industrial Overview – Spring 2023

-Lecture by Andrew Loeb and George Baggs

Name: Tananki Ranga Sai Saran Rohit

UB ID: 50441793

UB Email: rangasai@buffalo.edu

Ques 1 : Describe the market sector or sub-space covered in this lecture :

The market sector or sub-space covered in this lecture is Manufacturing Industry. The manufacturing industry is under growing pressure to decrease its environmental footprint and shift to a circular economy model that prioritizes material reuse and recycling. This includes creating more sustainable manufacturing methods, minimizing waste, and designing more eco-friendly and resource-efficient goods. To manufacture their goods, the aviation, space, and defense sectors require a complex and highly specialized supply chain. To guarantee that suppliers satisfy the specified standards and deliver components on time, the manufacturing sector has created complex supply chain management systems. Manufacturing procedures include designing, prototyping, engineering, testing, and manufacturing things. To manufacture completed items, these operations often involve a combination of physical labor and mechanized technology. Materials that are lightweight, sturdy, and resistant to severe heat and pressure are required for the aviation, space, and defense sectors. To address these criteria, the manufacturing sector has produced novel materials such as carbon fiber composites and sophisticated metals.

Ques 2 : What data science related skills and technologies are commonly used in this sector?

Data analytics is a critical skill for data scientists in the manufacturing business since it requires the ability to collect, convert, and analyze data from diverse sources in order to uncover patterns and insights that may assist optimize manufacturing processes, quality control, and supply chain management. Cloud computing is a critical technology for data scientists in the manufacturing business because it allows them to swiftly and efficiently store and analyze massive volumes of data. They must be knowledgeable with cloud-based data storage and analytics systems such as Amazon Web Services (AWS), Azure, and Google Cloud. Deep learning is also being utilized to

improve robots and automation in the industrial business. It may be used to create algorithms that enable robots to learn from their surroundings and accomplish difficult jobs more effectively. Deep learning is used extensively in the industrial industry for predictive maintenance. It entails evaluating data from sensors and other sources to forecast when equipment problems are likely to occur, allowing for proactive maintenance and downtime reduction.

Ques 3 : How are data and computing related methods used in typical workflows in this sector? Illustrate with an example.

The sensors capture information on a variety of characteristics, including temperature, pressure, and vibration. The manufacturer wishes to increase product quality by optimizing the manufacturing process. Data is collected from sensors mounted on the machine and pre-processed and cleaned by the manufacturer to eliminate outliers, impute missing values, and scale the data. Machine learning algorithms are used to evaluate data and discover patterns that may be utilized to improve the manufacturing process. Decision Making is used to fine-tune the manufacturing process, such as changing machine settings to maintain appropriate temperature and pressure levels during production. This can lead to improved product quality and less waste. CNNs and RNNs are used in robotics and automation to improve robotics and automation by allowing robots to learn from their surroundings and adapt to new conditions. For example, a robot may utilize a CNN to evaluate photos of things it needs to operate, allowing it to more properly recognize and classify them. RNNs may be used to anticipate product demand based on previous sales data. Manufacturers may use this to optimize inventory levels and enhance production planning.

Ques 4 : What are the data science related challenges one might encounter in this domain?

The growing demand for data science-related skills and technology is critical to the success of the Manufacturing Industry domain sector in meeting the expanding data size on a daily basis. They have vast warehouses and huge inventories, as well as large equipment and machinery, which results in massive amounts of data. Inconsistencies in data collecting caused by human mistake might have an influence on the accuracy and quality of the data. A worker, for example, may forget to log a certain parameter or record an inaccurate value, resulting in incorrect data being utilized for analysis. One of the most difficult tasks in the manufacturing business is assuring data quality. Data acquired from many sources, such as sensors and machines, might be noisy, incomplete, or contain mistakes, lowering the accuracy of any analysis or predictions based on the data. In the manufacturing business, implementing data science efforts sometimes necessitates adjustments to current processes and workflows. Change management may be difficult since it requires stakeholder buy-in and may disrupt current procedures.

Ques 5 : What do you find interesting about the nature of data science opportunities in this domain?

Data consumption in online transactions, mobile phone apps, including sales transactions, customer interactions, online banking, bill payments, statement printing, and more, has increased exponentially over the previous several years. In 2010, for example, data was created every two days; however, by 2021, data was generated just every 40 minutes. This data may be used by data scientists to find insights and patterns that can help drive business decisions and enhance performance. The manufacturing business creates a massive quantity of data from a variety of sources, including sensors, equipment, and processes. This opens up intriguing potential for data scientists to evaluate the data and extract insights that can assist optimize production processes, save costs, and increase product quality. The manufacturing sector is rapidly developing, with new technology and ideas being introduced on a regular basis. This opens up new possibilities for data scientists to use cutting-edge tools and approaches to optimize industrial processes and drive innovation.

(i) In the case studies, the speaker illustrates applications of ML for analyzing process and experimental outputs. Describe one application in your own words, the problem to be solved, and the solution presented

The visual classification process of metal additive manufacturing parts using a convolutional neural network (CNN) is a type of deep learning algorithm that can automatically identify defects or anomalies in the output of the manufacturing process. The first step is to collect a dataset of images of metal additive manufacturing parts, labeled with defect categories. A convolutional neural network is trained to learn patterns and features that are indicative of each defect category. The CNN model is deployed to classify new images of metal additive manufacturing parts and outputs a probability distribution over the defect categories. . Maximum pooling is employed in the convolutional layers, whereas random drops are used in the fully connected dense output layers. The RELU function was used to train the CNN. The CNN model is integrated into the inspection system, where it automatically analyzes the output parts and identifies any defects. If a defect is detected, the part is flagged for further inspection or removal from the production line.

(ii) Both Supervised and Unsupervised Machine Learning approaches are used in the case studies presented by the speaker. Explain in your own words the differences between Supervised and Unsupervised Machine Learning approaches and provide examples.

An unsupervised machine learning approach is used to detect and isolate elevated lumpy structures. This means that the algorithm is provided with unlabeled data, and it must identify patterns and structures in the data without any prior knowledge of what it is looking for. To accomplish this, the algorithm is used to identify two classes of data based on photos and color

classification. The algorithm looks at the raw images and color-classifies pixels as either green or red. This helps to identify the areas of interest, which are the elevated lumpy structures. By identifying these areas, the algorithm can then separate them from the rest of the data, which can help to isolate and analyze them further. To model the response surface, a Design of Experiments (DoE) matrix is used. A DoE matrix is a systematic approach to designing experiments that allows researchers to identify the most important factors that affect the outcome of the experiment. By using a DoE matrix, researchers can systematically vary the input parameters of the experiment and observe the output response. This can help to identify the optimal conditions for achieving the desired outcome. Overall, this approach combines unsupervised machine learning with a DoE matrix to detect and isolate elevated lumpy structures. By using an unsupervised approach, the algorithm can identify patterns and structures in the data without any prior knowledge of what it is looking for. By using a DoE matrix, researchers can systematically vary the input parameters and observe the response, which can help to identify the op Deep-learning CNN is used to identify seams in supervised machine learning systems.The raw data was stripped of its picture boundaries, and the remaining field was divided into four quadrants. Before it can be utilized to construct the CNN dataset, the dataset must be trained and verified. The picture and brightness of the metal coupons must be constant for optimum CNN behavior. Images were greyscaled using weighted average on RGB channels for CNN input normalization to increase CNN coverage.timal conditions for achieving the desired outcome.