Name: Sahithi Dodda
Person Number: 50441731
UB Email: sahithid@buffalo.edu

Assignment 10

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

Lecture by Andrew Loeb and George Baggs on Data Science in the Manufacturing Sector

Q1): Describe the market sector or sub-space covered in this lecture.

Ans: This lecture will examine the Manufacturing Industry market sector or sub-space. The manufacturing industry continues to play an important role in the global economy, producing a diverse variety of goods and supporting innovation and economic progress. However, in today's world, the sector is also confronting tremendous problems that are forcing considerable adjustments and transformation. Reshoring and localization of production have been increasingly popular in recent years as firms attempt to decrease supply chain risk, boost agility, and adapt to changing consumer expectations. Bringing production closer to end markets, minimizing reliance on international suppliers, and building local supply chains are all part of this strategy. The emergence of digital technology is altering the manufacturing business, allowing companies to streamline their operations, cut costs, and enhance quality. This involves the use of automation, robots, artificial intelligence (AI), machine learning, and the Internet of Things (IoT) to improve manufacturing processes, product design, and supply chain management. To cut costs and enhance efficiency, automation has become more vital in the aviation, space, and defense industries. For constructing and testing complicated items such as aircraft engines and satellites, the manufacturing sector has created specialized robots and automated systems.

Q2): What data science related skills and technologies are commonly used in this sector?

Ans: Data science is fast expanding in importance in the Manufacturing Industry sector, and the Manufacturing Industry sector is now realizing the potential of data analytics to promote organizational development and profitability while also minimizing risks. ML and data science have created a multitude of opportunities in the Manufacturing Industry sector by developing tools to monitor, control, and manage huge volumes of data. Statistical analysis is an essential ability for data scientists, especially in the manufacturing business. It entails applying statistical tools to evaluate data and develop conclusions about product quality, manufacturing process performance, and other relevant variables. Data visualization is an important ability for data scientists in the manufacturing business because it helps them to display complicated data in a form that decision-makers can understand. They must be knowledgeable in data visualization technologies such as Tableau, Qlik, Power BI. Deep learning may be used to enhance manufacturing processes by evaluating data from sensors and other sources to discover improvement possibilities. This technology may be used to increase overall efficiency by optimizing production processes, reducing waste, and reducing waste. Natural language processing (NLP) is another use of deep learning in the manufacturing business. It may be used to analyze text data from customer feedback, maintenance reports, and other sources to find patterns and insights that might aid in the improvement of product design and manufacturing processes.

Q3): How are data and computing related methods used in typical workflows in this sector? Illustrate with an example.

Ans: The workflow entails gathering data from a variety of sources, including sensors, equipment, and other systems. This information comprises machine performance, product quality, and other vital indicators. Data preparation entails eliminating outliers, filling in blanks, and scaling the data. Data analysis encompasses a wide range of data analysis techniques, including statistical analysis, machine learning, and data visualization. Using the insights gathered from data analysis to inform decision-making processes is what decision-making entails. A company, for example, may employ data analysis findings to optimize production schedules, modify machine settings, or prioritize maintenance chores. CNNs are often employed in quality control processes to detect product faults. A company, for example, can use CNN to evaluate product photos and discover faults such as fractures, scratches, or discoloration. RNNs may be used to forecast when a piece of equipment will fail. An RNN can recognize patterns that suggest when equipment breakdown is probable by examining sensor data over time. This enables manufacturers to do preventative maintenance, decreasing downtime and increasing overall efficiency. CNNs and RNNs may both be used to detect abnormalities in manufacturing processes. An RNN, for example, may evaluate sensor data from a manufacturing line and recognize whether a machine is running outside of its typical operating range. This can assist producers in detecting problems early and taking appropriate action.

Q4): What are the data science related challenges one might encounter in this domain?

Ans: Manufacturing firms frequently have several data-generating systems and platforms, making data integration and analysis complex. Integrating data from several sources takes time and can be difficult, especially if the data is in different forms and structures. In the manufacturing industry, sensitive information such as intellectual property, trade secrets, and confidential customer information is handled. It is vital to protect sensitive data and ensure its privacy and security, especially in the context of regulatory compliance. Temperature variations can also have an influence on sensor data accuracy, especially in temperature-sensitive operations. Temperature variations can cause data discrepancies, which can lead to erroneous analysis and forecasts. Environmental elements like humidity, light, and vibrations can all have an influence on sensor data accuracy. High humidity, for example, might cause sensor corrosion, resulting in false results. Overall, in this manufacturing sector data science faces challenges such as data integration, data privacy and security, data volume, data quality, domain expertise, organizational culture, and technological advancements. Addressing these challenges requires a combination of technical expertise, domain knowledge, and organizational readiness for data-driven decision-making.

<u>Q5)</u>: What do you find interesting about the nature of data science opportunities in this domain?

<u>Ans</u>: Data usage is increasing at an exponential rate over the last few years. For example, in 2010, data was generated every two days, however in 2021, data was generated just every 40 minutes. The quantity of data generated by apps, sensors, electronic devices, virtual assistants, smart phones, self-driving automobiles, AI gadgets, and so on... However, in today's world, the sector is also confronting tremendous problems that are forcing considerable adjustments and transformation. Manufacturing processes are frequently complicated, with several variables and influences influencing the result. This creates interesting prospects for data scientists to model these processes and uncover trends that can lead to process improvements and cost savings using advanced analytical approaches like machine learning and artificial intelligence. The emergence of digital technology is altering the manufacturing business, allowing companies to streamline their operations, cut costs, and enhance quality.

(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 (15 pts of the 80 C+R points in the rubric)

<u>Ans</u>: The use of a CNN model in visual classification can considerably increase the efficiency and accuracy of fault identification in the metal additive manufacturing process. Manufacturers may cut costs, enhance product quality, and boost production throughput by automating the inspection process. The initial stage is to gather a collection of photos of metal additive manufacturing items, both excellent and bad. These photos are identified with fault types like porosity, cracking, or warping. The labeled information is then used to train a convolutional neural network to learn patterns and features that are indicative of each fault category. Optimizer for Adam gradient descent with adjustable learning rate. A test accuracy of 90% was achieved in the 200th epoch. Individually recognized tiny picture partitions were color colored and reassembled to provide a visual signal for the ensemble coupon image using CNN classification. CNN is tuned to optimum accuracy in detecting errors in training data. Once trained, the CNN model is used to categorize fresh photos of metal additive manufacturing components. The CNN model is fed a picture of a metal part and generates a probability distribution across the defect categories, indicating the chance of each defect being present.

(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. (15 pts of the 80 C+R points in the rubric)

<u>Ans</u>: This approach combines unsupervised machine learning with a Design of Experiments (DoE) matrix to detect and isolate elevated lumpy structures. The algorithm looks at the raw images and color-classifies pixels as either green or red to identify the areas of interest, which are the elevated lumpy structures. 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 and observe the output response, which can help to identify the optimal conditions for achieving the desired outcome. For image classification, supervised learning can be employed, in which the algorithm is taught on labeled photos to detect and categorize objects in fresh images. Unsupervised learning may be used to reduce dimensionality, in which the algorithm determines the most significant characteristics in a dataset and reduces the number of dimensions to facilitate analysis. The availability of labeled data is the primary distinction between supervised and unsupervised learning. In the presence of labeled data, supervised learning may be used to predict a target variable based on input attributes. Unsupervised learning may be used to find patterns and structures in data if no labeled data is provided. Deep learning data augmentation techniques including random width and height shifts, random shears and zooms, and random rotations of dataset photos were used to create a sparse training and validation dataset. The dataset was divided 50/50 into rows with and without seams, and the apparent size of the available data samples was raised from 40 to 200 (140 for training and 6- for validation).