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

-Lecture by Ram Narasimhan

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 Industrial Data Science. Industrial data science is a critical tool for optimizing industrial processes, improving efficiency, and reducing costs by the application of mathematical & physics-based analysis in various industries .Industrial analytics is the practice of using advanced analytical methods and tools to extract insights and knowledge from industrial data. This data may come from a variety of sources such as sensors, equipment, machinery, and production systems, and can be used to optimize industrial processes, reduce costs, and improve product quality. By analyzing data from equipment, sensors, and production systems, industrial data science can help predict when equipment failures might occur, enabling organizations to perform maintenance before a breakdown occurs, reducing downtime and maintenance costs. As discussed by the speaker earlier we use to have very little amount of data being generated such as 1kb per flight having only very few parameters in consideration, but now we have seen vast amount of increase in sensors being used in every electronic part of the flight resulting in collection of terabytes of data from thousands of parameters like ground speed, cruising angle , Exhaust gas temperature etc…. from each flight .The role of data science in few industries are very huge , even 1% at times results in huge savings across industries like Aviation, Power, Healthcare ,Rail ,Oil & gas.

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

In Industrial Data Science sector , data science related skills and technologies plays an very important role .Few of the Data science skills which are very essential in this sector are Theoretical & Applied Statistics ,Machine Learning & Artificial Intelligence , Knowledge of

Contextual Analysis , Prognostic systems &methods, Image Processing & Analysis ,Physics and expert based modeling ,Process Engineering and operations science , Verification, Validation and Metrology . Knowledge of machine learning algorithms, such as decision trees, random forests, and neural networks, is important for building predictive models and making data-driven decisions, Familiarity with big data technologies, such as Hadoop, Spark, and NoSQL databases, is important for handling and processing large volumes of data. Understanding of statistical concepts and methods, such as regression analysis, hypothesis testing, and probability distributions, is necessary for analyzing data and testing hypotheses. Also, Knowledge of data management concepts, such as data warehousing, data lakes, and data governance, is important for managing and organizing industrial data. Also, the domain knowledge plays a vital role in industrial data science .Major time series tasks include Summarization ,segmentation, prediction, Anomaly Detection, indexing .Predix is a cloud-based platform developed by General Electric (GE) for the Industrial Internet of Things (IIoT). Predix is designed to help organizations build and deploy IIoT applications that can connect to and gather data from various industrial systems, such as equipment, machines, various sensors, and production lines. In the industrial data science sector, Predix can be used to create and deploy IIoT applications that can assist businesses in streamlining industrial processes, decreasing downtime, and raising the caliber of their products . Predix, for instance, may be used to create programs that track and examine data from production systems in order to spot inefficiencies and recommend process enhancements. Predix may be used to create apps for predictive maintenance that foretell equipment problems and save downtime using machine learning techniques. Main features of Predix are It is scalable from machines to cloud, Orchestrates data, control and analytics for effective interoperation, Enables quick response to customer requirements , Accelerates the industrial app economy through a service catalog, ingests and processes data faster, Analyzes data using industrial and material science, securely connecting all operating assets etc….

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

In general, data and computational methods are critical to the success of industrial data science ,from data collection and exploration to Analytic Development, Analytic Hardening, User Acceptance testing and project closure. The ability to work with large data sets, use machine learning and predictive modeling, having great domain knowledge, Maintenance Planning and Fault Diagnosis is key to success in industrial data science. Data and data processing techniques are widely used to collect and analyze data, identify patterns and anomalies, and optimize system performance and then Fault diagnosis , Maintenance planning , maintenance development and performance monitoring takes place . These methods are used to collect, clean, process, and analyze data to identify patterns and trends, make predictions, and optimize industrial processes. Through these methods, companies can improve the reliability and usability of its industrial systems, reduce maintenance costs and optimize its operations. Regarding calculation methods, time series analysis would help us in many cases to identify anomalies or make future predictions. Domain Knowledge is highly necessary in the industrial

data science sector. GE wind turbines generate large amounts of data such as wind speed, temperature and power. Using data analytics and machine learning algorithms, GE can analyze this data in real time and predict when a turbine is likely to go up. This allows GE to plan maintenance activities before a failure occurs, reducing downtime and maintenance costs. GE is able to improve the reliability and availability of its equipment, replacements in the power industry, reduce maintenance costs, and optimize its operations. Similarly as discussed in the video ,it can also be helpful in increasing the profits of aviation industry by reducing fuel consumption from terabytes of data being generated by 1000 of parameters per each flight.

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

In industrial Data science , huge amounts of data is getting continuously generated from many IOT devices, sensors etc… Some of the data science challenges that can be encountered are data variability - different types of data , quantity - data size, speed - processing speed where data can often be complex and difficult to process and data can contain missing values, which are some of the obstacles that can be encountered in industrial Data science.Some times the data collected from sensors have errors, missing values due to irregular troubleshooting and network delays. This can make it difficult to collect and use the information needed for effective steps for the faults detected in industrial data science. Industrial data analysis often involves complex models and algorithms, and it can be challenging to interpret the results in a way that is meaningful to domain experts and decision-makers. Industrial data is often large and complex, and requires significant computing power to analyze. The challenge is to develop scalable algorithms that can process large amounts of data quickly. Effective industrial data science requires a deep understanding of a specific industry and domain. Finding data scientists with the right combination of technical skills and domain knowledge can be difficult .

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

Increasing demand in the use of data is exponentially growing from past few years in every industry . For instance, the amount of data in 2010, which was being created every two days, and in 2021 it was being created just for every 40 minutes. The amount of data ,which is being created by applications, sensors ,electronic devices, machines ,turbines, flights etc… had become massive and the need for use of this data to enhance machines, increase profit gains, take timely business decisions lead to huge demand in data science and analytics .Data science is a rapidly evolving in every industry like aviation , power generation, Oil and gas ,Wind,Rail with new technologies and methods emerging regularly. This creates exciting opportunities for data scientists to stay on the cutting edge of the field and explore new domains and innovative approaches increase profits of the organization. As discussed in the video internet of things, intelligent machines , Big data & analytics helps shaping the industrial data science . It is also different from company to company ,industry to industry as discussed in the lecture based on

many events like cost/value per event, frequency of event, Data quality, Data Quantity and process integration.

Ques6:

(i)   Please discuss some of the characteristics of industrial data science problems alluded to in the talk and how they differ qualitatively from data science problems in other domains ?

In industrial Data science , huge amounts of data is getting continuously generated from many IOT devices, sensors etc… Some of the data science challenges that can be encountered are data variability - different types of data , quantity - data size, speed - processing speed where data can often be complex and difficult to process and data can contain missing values, which are some of the obstacles that can be encountered in industrial Data science. In industrial data science we usually deal with huge chunks of data which is often stored across different systems which complicates data storage, management and analysis. In contrast, data for computer advertising or BPO can be relatively smaller and more structured. While there could be a few similarities between the techniques and methods used in industrial data science and those used in other fields, such as computer advertising or business process outsourcing (BPO), there are a number of distinctive features that distinguish problems in industrial data science. The data obtained from the sensors will be more sensitive and errors might occur easily due to troubleshooting issues and network delays. In many industrial environments, such as factories or the aerospace industry, the consequences of a data science error can be severe ,any small errors from electronic sensors which are not dealt correctly by data scientists , might result in flight crashes at times. Data scientists must be aware of the potential consequences of their work and take appropriate measures to ensure that their models are safe, reliable and accurate. In contrast, data science in other sectors like computer advertising or BPO doesn't have human loss or major damages. Similarly in power generation and power distribution , Data science in this industry also should keep the adverse effects in mind while dealing with sensitive data. Data science problems in computerized advertising or BPO may focus more on optimizing customer engagement or improving decision making whereas in industrial data science problems often focus on optimizing processes, reducing costs, and improving efficiency.