

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

Assignment 3

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

Lecture by Ram Narasimhan - Industrial Data Science

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

Ans: The market sector covered in this lecture is Industrial Data Science. As discussed in the video Industrial Data Science is outcome-oriented application of mathematical & physics-based analysis & models to real-world problems in industrial operations. Industrial data science is the application of data science methods and techniques to solve real-world problems in industries. It involves the collection, cleaning, analysis, and interpretation of data from industrial systems to gain insights, optimize processes, and improve efficiency. Industrial analytics involves the collection, cleaning, integration, and analysis of large volumes of data generated by industrial systems. It uses techniques such as machine learning, statistical analysis, data mining, and predictive modeling to identify patterns, anomalies, and trends in the data. Reports show that more than 50B Machines will be connected on the internet by 2020. Also the amount of data getting collected by each sensor will be increased enormously, more than 35GB Data will be collected from each Smart meter, more than 1 TB of data will be collected from each flight, 500 GB data to be collected per blade by gas turbines etc. Also, industries that are all about IT & data had seen huge productivity and profits like Telecom, Financial services.

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

Ans: Data science skills which are very important in this sector are Knowledge of Contextual Analysis, Theoretical & Applied Statistics, Domain knowledge, Machine Learning & Artificial Intelligence, Prognostic systems & methods, Verification, Validation and Metrology, Image Processing & Analysis, Physics and expert based modeling, Process Engineering and operations science. Knowledge of sensor technologies, such as IoT sensors and RFID tags, is necessary for collecting data from industrial systems and understanding of control systems, such as PLCs and SCADA systems, is necessary for integrating data from various sources and optimizing industrial processes. Also, have knowledge regarding Major time series tasks like Summarization, segmentation, prediction, Anomaly Detection, indexing. Knowledge of data management concepts, such as data warehousing, data lakes, and data governance, is important for managing and organizing industrial data and also having knowledge with technologies like Hadoop, spark for handling huge chunks of data is very much essential. Building predictive models and making data-driven choices need an understanding of machine learning methods like decision trees, random forests, and neural networks and also must have a understanding of statistical concepts and techniques including hypothesis testing and regression analysis. In the industrial data science

sector, Predix is a cloud-based platform developed by General Electric (GE) for the Industrial Internet of Things (IIoT) which can be used to build and deploy IIoT applications that can help organizations optimize industrial processes, reduce downtime, and improve product quality. Few Prime characteristics of Predix includes Enables quick response to customer requirements, Analyzes data using industrial and material science, Accelerates the industrial app economy through a service catalog, ingests and processes data faster and much more. Predix can be used to create customized applications for predictive maintenance that forecast equipment problems and save downtime using machine learning techniques.

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

Ans: In the industrial data science sector, data and computing-related methods are used extensively in typical workflows to gain insight into the large volumes of data produced by industrial systems basically in few steps data gathering and exploration, Analytic Development, Analytic Hardening, User Acceptance testing and project closure .The first step in this workflow is to detect and diagnose failures in industrial systems. This involves collecting data from various sensors and devices and analyzing it to detect abnormal activity and then data pre-processing takes place, then machine learning is used to identify patterns and anomalies. Once the fault is detected , root cause analysis comes to the picture to identify underlying cause and suggest corrective actions. Maintenance activities are then planned and scheduled to perform the necessary repairs. This workflow makes extensive use of data and computational methods to identify and diagnose faults, plan, and execute maintenance activities, and monitor system performance. Using these methods, industrial data science organizations can improve the reliability and usability of their systems, reduce maintenance costs, and optimize their operations. Predix allows GE to collect data from its various industrial systems, such as wind turbines and jet engines, and use it to predict and prevent equipment failures before they happen. GE uses data analytics and machine learning to optimize the performance of its jet engines. By collecting data from sensors built into engines, GE can identify patterns and anomalies in the data and use that information to improve engine efficiency and reduce fuel consumption. Similarly, Midwest power generation - Storm preparedness during severe weather situations on where to deploy the crew during storms to minimize disruption and recovery time.

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

Ans: Industrial data scientists must be able to work with complex data sets, develop scalable algorithms, and interpret results in a way that makes sense to decision makers. They must also be aware of privacy and security issues and have a deep understanding of a specific industry and field. Effective industrial data science requires a great understanding of a specific domain. Finding people with the right combination of data science skillset and domain knowledge is hard. Unorganized and chaotic industrial data might make it challenging to get actionable insights. Data that is difficult to gather, data that is incorrect, or data that is missing. Also, the data which is collected from various sensors is often spread across multiple systems which makes it difficult to integrate and analyze .Complex models and algorithms are frequently used in industrial data

analysis, and it can be hard to comprehend the results in a way that is useful to domain knowledge experts and decision-makers. Large and complicated industrial data sets sometimes demand a lot of processing power to examine. The challenge is in creating scalable algorithms that can rapidly process enormous amounts of data. For example, it can be difficult to gain meaningful insights from vast amounts of data collected from many different sources due to data quality issues such as missing data, data errors, and hard-to-collect data. Therefore, identifying the root cause of problems or predicting device failures can be difficult.

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

Ans: Industrial data science is still a relatively new field, which means there is a lot of potential for innovation and discovery. Data scientists in this field have the opportunity to explore new techniques and approaches that can push the boundaries of what is possible. Industrial data science can be applied to many different fields and use cases, from predicting equipment failures in the aerospace industry to optimizing energy use in manufacturing facilities. This means that data scientists in this field have the opportunity to work on diverse and interesting projects. The industrial sector generates vast amounts of data that can be used to improve efficiency, reduce costs and improve safety. By applying data science techniques, we can gain insights and make predictions that can significantly impact operations and results. Industrial operations can be transformed across multiple industries like Aviation, Healthcare ,Power distribution , Power Generation , Manufacturing , Wind, Mining ,Water , Oil & gas , Rail .Sometimes, 1% improvement in Efficiency and cost savings , new customers , risk avoidance cuts almost 276B dollars across industries. Resource optimization and maintenance optimization also can be done for increasing profit gains.

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

Data science has been increasingly utilized in different areas such as computational advertising and BPO, but the importance of industrial data science should not be overlooked. Although there may be some overlap in the techniques and methods used in industrial data science and other fields such as computer advertising or BPO, industrial data science problems have unique characteristics that set them apart. Industrial data often have much larger and more complex data sets than those of other industries. The data collected from various sensors is often spread across multiple systems, making it difficult to integrate and analyze. Industrial data analysis requires the use of complex models and algorithms, and it can be hard to comprehend the results in a way that is useful to domain knowledge experts and decision-makers. Industrial data science deals with complex global problems that are influenced by multiple variables, such as weather, temperature, humidity, earthquakes, Hurricanes etc., and these problems require efficient reactions to ensure optimal outcomes. The data used in industrial data science is massive and complex, making it challenging to manage and analyze, and it includes data from sensors, machines, and other sources. Industrial data science deals with day-to-day operations that are

critical to running a smooth operation, and failures in this field can have a catastrophic impact on society which can lead to the loss of human lives. Therefore, the standards that must be met in industrial data science are significantly higher, and it is imperative to ensure that the data is reliable, accurate, and that the results are comprehensible to domain experts and decision-makers. Industrial data science problems often require collaboration across multiple domains and teams. Data scientists must work closely with engineers, industry experts, and other stakeholders to understand the problem, collect and annotate data, and validate results. In summary, industrial data science is an essential field with a lot of potential for future data scientists. It deals with global problems that affect millions of people and requires collaboration across multiple domains and teams. The data sets used in industrial data science are larger and more complex than those of other industries, making it challenging to manage and analyze the data. Industrial data science problems require the use of complex models and algorithms, and they must be comprehensible to domain knowledge experts and decision-makers. In contrast, computer advertising or BPO data science problems may not require such collaboration.

(ii): Also, answer the following multiple-choice questions: You can list the question number and the letter corresponding to the correct choice as Answer in your report, (2x5 = 10 pts of the 80 C+R points in the rubric)

Ans:

Q1) A

Q2) C

Q3) D

Q4) E

Q5) E