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

Assignment 1

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

Lecture by Sriganesh Madhvanath, eBay:

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

Ans: Business Process Services (BPS) and the use of Data Analytics in this industry is the market sector or sub-space covered in this lecture. BPS process, enables firms to continuously improve the efficacy and efficiency of their processes by integrating technology into essential business activities. It mainly focuses on how to have a direct business impact or bring significant value and maximize the utilization of knowledge workers and sophisticated algorithms combinedly to automate essential operational tasks. A notable illustration of this type is the tolling system in New York State. Someone with knowledge of toll booths is given control of the tolling system by the government because this combines expert people and cutting-edge algorithms to improve their business operations, efficiency while also delivering value to the company. BPS first focused on operations like human resources, finance, customer service etc., but it is now widely spread across all the sectors and values more than US $150 billion. Bps offers a variety of business model customizations that are best suited to the client's long-term strategic objectives. A client may now assess what their business goals are, including cost optimization, scalability, enabling maximum efficiency, adherence to controls and compliance, and risk minimization, and determine which operating model is most appropriate for them.

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

Ans: Data science is used in Business Process Services (BPS) to solve business issues, support, and assist in the development of sensible business decisions. Descriptive

Analytics which include statistical analysis, Visualizations, Root cause Analysis, Predictive, Prescriptive Analytics, Statistical modeling, Data mining, Machine Learning, big data technologies which include Hadoop, Apache, Spark, Storm, Dynamic control and systems, operations research and optimization, multimedia analysis – text, video, audio, Spatio-temporal modelling, ethnography, behavioral economics, text analytics etc., are the data science related technologies and skills most commonly used in this BPS sector which help in dealing with business challenges, answering business queries and supporting intelligent business decisions.

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

Ans: The best way to demonstrate the uses of information data and computing-related approaches is with an example, like predicting what would be the diabetes prevalence 20 years from now in New York state. We would firstly start by collecting historical data, deciding the contributing factors, such as demographics, type of food intake, education, age, sex, medical history etc., we wish to employ to forecast the prevalence of diabetes. After gathering data based on these features, the data is cleaned and pre-processed to remove any discrepancies and null values, and then an appropriate model is created for instance a regression model or a stochastic data model or algorithmic model which tries to model the underlying processes which leads to the outcome and provide reliable predictions with higher accuracy and interpretability.

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

Ans: Imbalanced data is one of the major problems encountered for instance, if we want to categorize whether a new animal is a goat or a sheep, we need atleast 1,000 rows of training data each for goats and sheep to help the model in distinguishing the two. But, If the training data had 1850 goats and 150 sheep, we would falsely believe that the majority of new animals should be goats which is a problem. Varied data, abundance of data sources, large volumes of the data, processing speed for the data, missing values in the data, Overfitting, less features, data uniformity, and data authenticity are some of the challenges related to data science that one could encounter in BPS.

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

Ans: Data science enables businesses to effectively comprehend enormous amounts of data from several sources and helps to gain insightful information for more informed decisions. I feel Data science is in huge demand and has lot of opportunities in BPS because of the massive amounts of data being produced by electronic devices, different day-day applications, sensors, virtual assistants, Robots, smart phones, Artificial Intelligence devices, health sector machines etc., which needs to be utilized to make effective predictions for business decisions. I also believe that the wide usage of data Science in the fields of retail, health care, banking, finance, construction, transportation, communication, media, entertainment, education, manufacturing, defense, security etc., to solve problems and make data driven decisions has also created lot of opportunities. Nearly all businesses are attempting to utilize data science's potential today; this number will only grow as data science technology gets more dependable and affordable. Data science is currently a force to be reckoned with.

Q6 (i): Please comment on the BPO vs BPS vs BPaaS paradigms and the increasing role of Data Science in the BPS domain. (15 pts of the 80 C+R points in the rubric)

Business Process Outsourcing (BPO): To save time, money and focus on core business aspects, businesses might outsource non-core front-office and back-office functions to other companies through business process outsourcing (BPO). The business outsourcing company would provide the standards, and the company providing the services would manage the business operations in accordance with those standard guidelines. Commercial operations could involve contact center services, human resources, financial services, Accounting and BPO mainly provides flexibility in handling operations and finances.

Business Process Services (BPS): Businesses that are skilled in one or more business activities offer business process services (BPS) to organizations that need them. BPS process, enables firms to continuously improve the efficacy and efficiency of their processes by integrating technology into essential business activities. These commercial operations could involve contact center services, Document management, Life sciences, insurance, human resources, F&A, financial services etc. Customers/ Clients can select from a range of models including customizations from prominent BPS providers including ADP, Conduent, Accenture, Teleperformance, Xerox, Convergys, Paychex, Arvato Bertelsmann, DXC

Technologies, Capita etc., that are best suited for their long-term strategic objectives.

Business Process as a Service (BPaaS): Business Process as a Service (BPaaS) is a type of BPO delivered based on a cloud services model. BPaaS offers many business benefits such as Product/service deliverability, provides cutting edge technologies to help reduce costs, Accommodates fluctuating business needs etc. BPaaS utilizes the accessibility and effectiveness of a cloud-based system to provide businesses the personnel, operational procedures, and technological infrastructure required to function as a pay-per-use service. By offering an on-demand solution depending on the services required as opposed to purchasing a package deal tied into a single application, this method of operating significantly lowers total cost of ownership, improves flexibility and agility. Some companies that offer BPaaS are Wipro, Capgemini, Fujitsu, Genpact, IBM, Cognizant , TCS, Oracle, SAP SE, HCL, LLC, DXC Technology, Tech Mahindra, NGA Human Resources, eBuilder Sweden, WNS, Verecloud, Adaptive Insights and Red Hat etc.

Data Science's Role in BPS: By using various statistical modeling, data mining, and machine learning, big data approaches to business data, data science helps businesses overcome problems. Data science is increasingly being used in BPS and some areas where data Science is used widely are in the fields of retail, health care, banking, finance, construction, transportation, facial recognition, fraud prevention, device failure detection, vehicle number plate recognition, communication, media, entertainment, education, manufacturing, defense, security etc., to solve problems and make data driven decisions and which in turn increase an organization's business value.

Q7 (ii): Pick two of the case studies from the lecture to discuss how different data science techniques are used to solve these problems. (15 pts of the 80 C+R points in the rubric)

Case study 1: Customer Churn Analysis - Churn analysis measures a company's customer attrition rate with the goal of lowering it. A company with a high churn rate must compete while struggling to attract enough new clients to patch the ship's holes. Churn, also known as customer attrition rate, may be reduced by analyzing our product and how users interact with it. For this case study, Churn in Health care Insurance is used for illustration purpose. Cost of churn is very high for the Insurance service provider; it is almost 5 times costly to acquire a new customer as to retain one. So, the question here is to predict which of the customers are

going to churn and provide them with feasible solutions to prevent the churn. So firstly, the problem is divided in to 3 parts- Identifying risk level for different customers, characteristics of churners, prevention. Users were having multi way relationships and these are modelled by hyper graphs and each hyper edge may have 2 or more nodes and shows a shared characteristic between users like age group, language, service, job status, referral sources, relationships between customers etc. Historical data of churned people is taken, and frequent pattern mining technique is used in which subsets of features of people who churned which have support(occurred atleast many times) are found and features found in people who are not churned are deleted leaving us with unique features of churners. Also, for churn prediction refinement feedback loop is used. For churn prevention, Net Promoter score (0-10 scale of how likely you would recommend the service) in which 0-6 is considered not likely to recommend, 7-8 is neutral and 9-10 is considered highly recommended i.e., promoters is calculated by subtracting the percentage of people in 0-6 range and percentage of people in 9-10 range. Influencer identification is done with all the data for churn prevention because those are the people who retain and promote the service. Net promoter scores are calculated on the people with the influence they have and using decision trees, random forest, support vector machines and net scores approach all the predictive models recall scores are compared and 93% of churners have been detected and also with the feedback loop churn prediction accuracy is improved significantly. Model built is also good at early prediction i.e., if a person churns at 6 months, how much in advance can we predict the churn. Accuracy achieved is high even for early predictions.

Case study 2: Public Health Service - The high cost of public healthcare in the US has caused a transition in the health-care paradigm from greater expenditures for patients, to one of value, which is characterized by better outcomes and cheaper prices. The goal is to predict future instances of poor health in the broader population. The information provided was derived from Medicaid encounter insurance claims data for 7 years and consists of data such as race, ethnicity, gender, age, insurer information, drug information, encounter history, care accessibility etc. ER visits and inpatient stays make up a considerable chunk of the overall expense. Injury, drug and alcohol use, and chronic health conditions accounted for most ED visits. Predicting the likelihood that a patient would visit the ED is addressed as it would indicate whether a patient will visit the emergency room in the upcoming six months, this is a binary classification problem. The type of data used was that for ED visits over long and short time periods. Logistic

Regression and Random Forest models have been used for greater accuracy. Different thresholds were employed for various purposes for example, we require data with low false positives for ED pop monitoring and high false positives for case management. The Hospital Readmission Model, which essentially predicts whether patient will be readmitted to the hospital 30 days after release using characteristics like inpatient stay, encounter history, prior illnesses, etc., was the subject of the second problem that was brought up. The patient's health, post-discharge patient self-management, and post-discharge outpatient etc., are other variables which can be considered. Logistic regression is used in this case, to divide the patients who were at risk and not at risk.