EAS 504
ASSIGNMENT-1

By: - Katyayni Shankar Kaushik

General Information: -

The following Lecture was held by Mr. Sriganesh Madhvanath, he currently holds the position of Director, Buyer Experience Applied Research, at eBay.

Base Questions: -

1.) What are principal uses of data sciences in this domain?

Data Science, in this domain, is used in various areas like: -

- Recommender Systems (for Buyer/ Seller Experience)
- Advertisement & Seller guidance
  - How to price products
  - What should be the advertisement rates
  - How to make items more visible
- Query understanding & Graph Linking for Search Engine Optimization

For all such work techniques used from data science domain are **Natural Language Processing, Text/ Attribute Mining, Recommender Systems, statistical modeling techniques** etc.

2.) How are data and computing related methods used in the organizational workflow?

Within organizational workflow, data can be present in multiple formats like time series, speech, image, text, tabulated format (with features and response variable). There are multiple varied techniques that are used for such types of data, for e.g.: -

- For Time series data, computing related methods used are ARIMA, SARIMAX etc.
- For Speech recognition (voice data) and handwriting recognition, techniques used are LSTM to deal with exploding/ vanishing gradient problems.
- For Text data, methods implemented are NLP techniques.
- For image data, CNNs (with Autoencoders) having different variants like ResNet, GoogLeNet, AlexNet are used. Metric Learning is also implemented in some cases.
- For Tabulated data, regular statistical modeling techniques are used that comes under supervised learning domain.

3.) What data science related skills and technologies are commonly used in this sector?

The data science related skills and technologies commonly used in this sector are as follows: -

- Data Mining, Statistical Modeling and Machine Learning (Decision Trees, Random Forest, Logistic Regression, SVM etc)
- Spatio-temporal modeling, Operations Research and Optimization, Text Analytics, Multimedia Analysis etc

- Technologies used are: -
  - Big Data: - Hadoop, Spark, Apache, Storm
  - Languages like R programming, Python (libraries i.e. Numpy, Pandas, Scikit Learn, Scipy, OpenCV etc.)
  - Visualization tools like Tableau, PowerBI, Qlickview etc.
- Domain expertise and general inquisitiveness

4.) What are the primary opportunities for growth?

The speaker talked about different case studies such as Healthcare, Transportation, Customer Care, retail etc.

These domains provide various opportunities for e.g.

- Healthcare would value to 5.5 trillion USD or 20% of US economy. With the correct use of data this cost be optimized to maximum value.
- Public Transportation can be segregated to Supply and Demand. With the absolute monitoring of data generated from vehicular movement, there can be improvement of 62% in the overall connectivity (use case discussed in case studies).
- Customer experience can be improved using techniques like sentiment analysis and analysis through visualization of dashboards.
- Recommender Systems can used to improve Buyer and Seller experience.

<u>Other 3 Questions with respect to this Lecture</u>: -

1.) BPO vs BPS vs BPaaS: -

- Businesses have 2 main types of work i.e. Core Business work and Non-Core Business work. Non-Core Business work are further sub-divided into 2 categories i.e. front end and backend. Examples of Back-end work are HR, Finance, accounting, whereas example of front-end work is Customer Care.
- Companies usually prefer to get the Non-Core Businesses through third party vendors, in order to optimize the cost. This is called as BPO or Business Process Outsourcing.
- BPS, also known as Business Process Services, are offered by companies to its client as specific set of expertise in order to automate the work and add value. BPS companies act as business partners of the client rather than third party vendor. Services offered by BPS can vary from takeover to managed captive, build-operate-transfer model etc. BPS also offers custom services according to the need of the client.
- BPaaS, also known as Business Process as a Service, are services offered as a standard set of operations. It employs Cloud computing service model where services are offered through cloud platforms, in order to enhance automation.

2.) Increasing Role of Data Science in BPS Domain?

The BPS industry has become more competitive, in terms of prices, as the industry now offers outcome based and gain sharing revenue model. The end goal is to drive savings, increase business value, deliver transformational benefits through process re-engineering through standardization, bringing in efficiency drivers through platform-based solutions. Data Science plays a key role in doing so, since the end focus is not only the cost arbitrage but intellectual arbitrage.

Case Studies: -

1.) Transportation

Characterizing connectivity in Public Transportation: -

- This use case deals with Transit Network Connectivity (e.g. buses) in an area and looks at the problem from 2 perspective i.e. Supply and Demand. Supply is related to provider and considers spatial/ temporal measures, Demand is related to Consumer. Motive is to improve the functioning and optimize the cost so as enhance the provider and consumer experience.
- Many significant features are taken into consideration such as physical network, service schedule, reliability etc. Relevant information like walking time, waiting time, travel time, buffer time, in-vehicle time, stop-level connectivity, stops, routes, service schedule etc is considered to optimize the movement of vehicles.
- Upon applying data science techniques. Visualization, it is observed that if all the services ran on schedule there will be an overall improvement of 62% in connectivity.

Automated License Plate Reidentification: -

- Another use of Data Science Techniques is to read the license plate of vehicles, running on highways and charge them toll accordingly. This removes the need of physical toll booth structures that slows the movement of traffic.
- Various techniques such SURF, SIFT features, CNN, Metric Learning etc is used to achieve this automation.

2.) Retail

Healthy Food Recommendation: -

- The idea is to reduce the problem of obesity and recommend foods to people that have nutritional value in it and will make them healthier.
- Grocery Data from Wegman's is collected and analysis is performed to see the total nutrition someone gets from their respective purchase and compared it to daily average.
- Problem statement is converted to a Knapsack problem (here it is caloric limit) to see the ideal way for people to reduce the caloric gap between current intake and ideal intake. Cosine similarity is used to make recommendation to people, on how should their daily intake be!