Assignment 1

EAS 504

Anish Shah

Base Questions:

- What are the principal uses of data sciences in this domain?
    1. Some of the principal uses of data sciences in this domain are Attribute Mining, Text Mining, Natural Language Processing(NLP), Speech to Text conversion, Sentiment analysis, Recommender systems.

- How are data and computing related methods used in the organizational workflow?
    1. Data and computing methods used in the organizational flow really depends on the type of data (speech, tabular, text) we have.
    2. Neural networks are used on tabular data.
    3. CNN's on image processing related tasks.
    4. Long Short-Term Memory on speech and text data is used.

- What data science related skills and technologies are commonly used in this sector?
    1. Data Processing was carried using Spark. R language was used for building the respective models.
    2. This sector requires lot of in-depth knowledge of various models because models vary from industry to industry.
    3. Techniques such as Random Forests, Gaussian Mixture Models, Decision Trees and other regression models are also used apart from techniques mentioned in previous question.

- What are the primary opportunities for growth?
    1. As the speaker mentioned health care industry is going to have an exponential of growth with projected market value being over 5.5 Trillion by 2030. So, there is huge opportunity for growth here as the usage of BPS will also grow with the industry.

BPO (Business Process Outsourcing) vs BPS (Business Process Service):

- Every company has bunch of core businesses and non-core businesses called processes. For Example: IT, HR or Customer Support.
- Outsourcing a non-core business component of a company with cost saving as a main objective is called BPO. It provides flexibility in managing finances as well as operations.
- BPO is typically categorized into back office outsourcing, which includes internal business functions such as HR or finance and accounting, and front office outsourcing, which includes customer-related services such as contact-center services.
- BPS is typically deals with core businesses. Instead of becoming third party service provider they kind of become your business partners.
- BPS focuses on how to add significant value, or direct business impact, beyond the bottom line.
- It also focuses on optimizing how knowledge workers and advanced algorithms can be combined to automate key operational activities.

Case Studies:

1. Prognostics for Low-cost Devices:
    - The main idea behind the use case is to collect the streams of data from the machines and use it to draw some insights using Machine Learning.
    - Survival analysis which is used by Insurance companies is an application of this use case.
    - One approach used in this use case was to screen all machine faults for association with service calls. Identify relevant predictive indicators for device subsystems and service call problem descriptions.
    - Another approach used was to coerce the problem into a classification problem for predicting the likelihood of a service call within a certain time window with the use of random forest classification algorithm.
    - Random Forest model was a good final model as it gave higher accuracy (80%).

2. Transportation:
    - This use case deals with characterizing connectivity in Public Transportation.
    - It uses attributes such as Walking time between transit stops and origin/destination, and during transfers, due to discrete services in space, Waiting time at origin, transfer point and destination, due to discrete services in time, Travel time in vehicle travel time and Buffer time due to variability of travel time.
    - Network Connectivity captured many important factors such as physical network, service schedule and reliability. There was significant difference when services were reliable and not reliable.
    - Spatial model was used to predict total number of transit riders per transportation analysis zone.
    - Connectivity of a city area, area covered by the core network, compared to areas of the city, connectivity of a city population: population with access to core network, compared to total population of the city was also analyzed.