EAS 504 ASSIGNMENT-6

By: - Katyayni Shankar Kaushik (50289158)

General Information: -

The following Lecture was held by Mr. Abhishek Singh Tomar, he currently works as Engineering Manager in Linkedin. He talks about various applications of Data Science in and outside of Linkedin, especially from Search, Ranking perspective.

Base Questions: -

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

Ans) Following are principal uses of data science in this domain:

- Page Ranking/ Document Ranking
- Search Engine Optimization
- Information Retrieval (especially in Query processing)
- Web Search/ Web Crawling
- Enterprise Search for e.g.
    - HR services
    - Self-Driving Customer Services
    - Conversational Bots
    - Student Services (Registering Course, Getting Transcripts)

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

Ans) Data and computing related methods are used in following ways in organizational workflow: -

- Recommender Systems are used to identify the Intent of the query made by the user and give best/ nearest possible result
- Page Ranking algorithms are used for search engine optimization. N-phase ranking is used to make search simpler, for e.g. 1st Phase Ranking and Complex Ranking.
- ML algorithms are used for spam classifier, adult classifier, text quality estimation.
- Spell Correction, Segmentation, Stemming, Annotation, Term expansion, vector space models are all data science related computing methods.
- Knowledge Graphs are used for structured data to make query processing much faster.
- ML algorithms are also used to differentiate between Original vs Mirror Site vs Spam detection.

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

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

- Recommender Systems used in different type of searches.
- Python, R programming and other ML platforms for Machine Learning algorithms.

- Natural Language Processing technique for text data (Spell Correction, Segmentation, Stemming, Annotation, Term Expansion etc).
- Deep Convolutional Neural Networks, Decision Trees based Ensemble methods used for ranking of algorithms.
- Information Retrieval, Data/ Text Mining

4.) What are the primary opportunities for growth?

Ans) Speaker explained us about various areas where there are opportunities for growth, they are follows: -

- Linkedin is currently working on identifying skills of each user profile and providing expert ranking based on their relevant skillset. It also is working on making recommendations for missing skill set.
- Other next generation features Linkedin is working on and where there is primary opportunity of growth are as follows: -
    - Recommend relevant content
    - Location based search result
    - Skill based expert search system
    - Integrate with Chatbot/ Slack etc
    - Develop friendly UI's and filters
    - Natural Language Processing implementation for understanding Unstructured data.