INSIGHTS INTERPRETATION SUMMARY

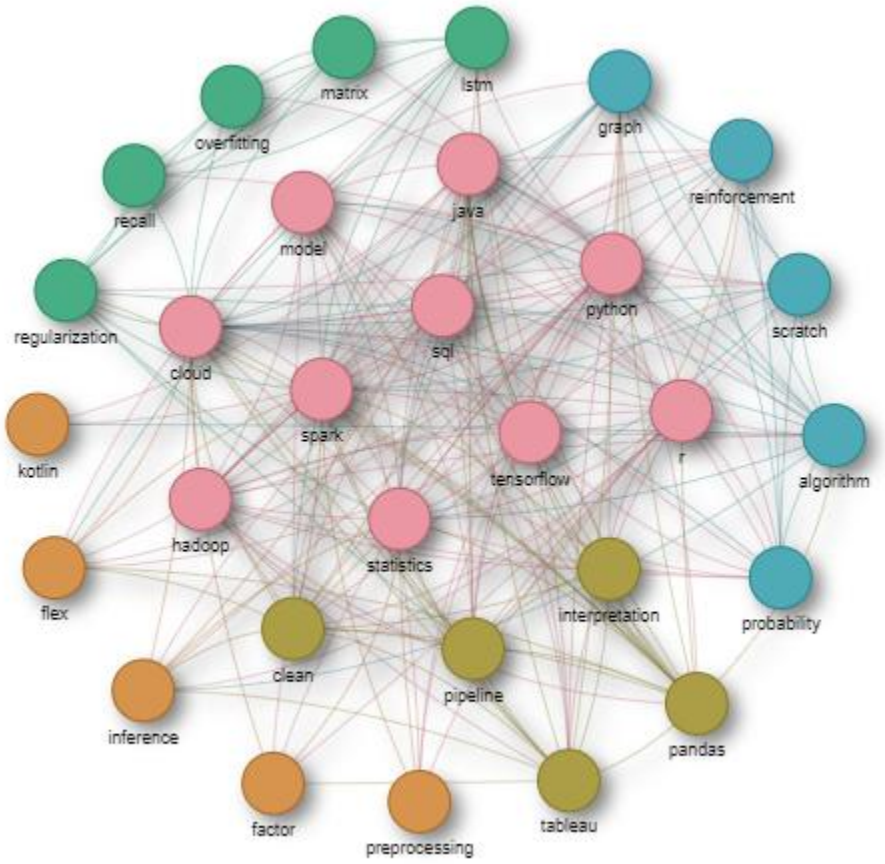

Fig 1.1: Hard Skills Dependency Graph of Records with Job Title of 'Data Scientist' [1, 2]

[1] Nodes in Pink illustrate the 10 most commonly occurring hard skills found in records within the job title on Indeed.ca. Other node colors represents the clusters or, the typical groups a job title belongs to as per an implementation of the Spectral Clustering algorithm

[2] The cluster node labels are the top 5 terms that best characterize a cluster, ranked based on Laplace Smoothed Positive Pointwise Mutual Information (PPMI)

Current industry needs (represented as per job postings on Indeed.ca available between Nov. 2020 – Feb. 2021) can largely be classified into 3 types of data scientists. Job role narratives inferred from Fig 1.1 is as provided below:

1. **The data pipeline focussed roles** performing **preprocessing** using **Kotlin**
2. **The "full-stack" data scientist roles** with wide-ranging skills involving data **cleaning** (using **pandas**), creating **pipelines** and **interpreting** findings using visualization tools such as **tableau**
3. **Two more research based roles** distinctly separated into two categories:
    a. **Graph** and **Reinforcement** learning focussed roles with ability to develop **algorithms** from **scratch**
    b. **Deep Learning** focussed roles with skills in developing deep neural networks (i.e. commonly **LSTM**) along with **other** machine learning related know-how