

NASSCOM Think Tank Discussion on Talent and Skill Development for Data Science in India: A Report

Saumyadipta Pyne and Meghana Aruru

Abstract— NASSCOM held its ‘Data Science – Talent & Skill Development’ think-tank meeting on Thursday, June 22, 2017, at the Hyderabad International Convention Centre in Hyderabad, Telangana State, India. The meeting brought together industry and academic representatives to participate in a discussion on employment opportunities and training needs in the areas of Big Data Analytics (BDA) and Artificial Intelligence (AI). Following is a report on this significant event.

Index Terms—Data Science, Big Data Analytics (BDA), and Artificial Intelligence (AI).

1 INTRODUCTION

NASSCOM think tank focused on a crucial aspect of advancing Data Science in India. Specific objectives of this meeting were:

1. To identify specific job roles in BDA/AI;
2. To develop an understanding of how companies are skilling and reskilling their workforce;
3. To evaluate the BDA/Data Science needs in the market.

Following introduction of all participants, the meeting began with the NASSCOM representatives presenting an overview of the analytics market in India and projected growth. Currently, there are 600+ analytics firms operating in India hiring 120,000+ professionals. Approximately, 60% of these firms are directly engaged in BDA.

In terms of the market, key emerging forces such as ageing populations, disruptive technologies, geo-politics, etc., provide a push to build capacity and train professionals in the tools and techniques of BDA and AI. An estimated 150 million direct jobs were identified in 40 sectors along with 200 million indirect jobs indicating the need to train and develop analysts to meet the needs of a growing market.

The following sections describe issues and opportunities across different areas as identified in the meeting.

2 ROLES AND TITLES IN BDA AND AI

The primary agenda of this discussion section was to identify the new roles and skills needed beyond the traditional roles, the numbers of existing and potential jobs, and challenges associated with timely recruitment and training.

The NASSCOM team presented their results from an exhaustive search of BDA/AI job listings on the professional networking website LinkedIn. Role descriptions and pathways to promotion were identified. Specifically, 6 new job roles in Big Data space and 12 under AI space were mentioned. Meeting participants were asked to look into each position title and the corresponding required skill-sets as well as role ambiguity or role overlap, if any, with other positions.

A discussion on what constitutes a ‘data scientist’ ensued followed by role descriptions across various sectors. In terms of BDA/AI roles, each sector uses its own terminology and definition of what constitutes a particular role. Such definitions may not be transferrable across sectors and may present challenges in recruitment and skilling/reskilling. For example, it was proposed that an employee moving from the finance sector to the supply chain/logistics sector may need to be reskilled in specific domains. Such reskilling is not often easy nor is it readily available, thus making it harder to map the role in the latter case. There was agreement around the table to develop a mechanism for mapping roles across industries and making job descriptions more consistent. Participants expressed a need to identify training and assessment modalities for capacity building.

3 CAPACITY BUILDING AND TRAINING

This discussion section was centered around three major questions:

1. How would industry propose to handle unskilled workers and train them?
2. How would industry propose to reskill workers into

specific domains?

3. What are the common basic skills necessary for a data scientist?

One of the first and foremost areas of concern mentioned was a lack of depth in Machine Learning in general. Some industry representatives mentioned that they would prefer cross-domain ability; i.e., a blend of domain and technology and a wider talent mix over focused expertise in single domains. Other representatives noted that they found gaps in skills, e.g., an expert in Natural Language Processing may not be as skilled in statistical modeling. To this, some expressed that reskilling could be difficult as compared to hiring freshers and skilling them. Yet, other representatives suggested filling the gap of data scientists/analysts by skilling non-analysts/non-technical employees.

In terms of training and capacity building, partnerships with academia were suggested. Outcomes based performance through programs such as the U.S. Certification program INFORMS was discussed. It was proposed that job descriptions and roles be consistent across industries to make capacity building easier. Industry representatives identified courses on MOOCs, Coursera and edX that may be helpful but would need to be mapped with existing courses and curricula within the country to identify gaps and opportunities.

The need to have a common baseline program to prepare for data science positions covering basic skills like mathematics, statistics and programming languages as well as structured thinking and client orientation was discussed. Some experts highlighted the need for special skills (such as streams, graphs, etc.) required for innovating analytical products and Big Data solutions, and the tests for benchmarking the same.

Participants agreed that the Indian market would need significant capacity building and training at the academic level to develop BDA/AI professionals who would be market ready by 2020. An urgency to develop such training was expressed along with quality metrics and standards for certification/licensure, etc. Participants agreed that job descriptions and roles found on LinkedIn were highly varied and needed mapping across industries/sectors. The meeting ended with a resolve from experts to continue the role mapping process in both academia and industry towards reducing the employment gaps in BDA/AI for the Indian market.

Prof. Saumyadipta Pyne and Dr. Meghana Aruru are faculty members at the Indian Institute of Public Health, Hyderabad, focusing on the areas of health analytics and health communications. They attended the NASSCOM think-tank meeting on 'Data Science – Talent & Skill Development' in Hyderabad on June 22, 2017.