Preparationg Data Scientists for the Country: What Do We Have Here?

by Norizan binti Mohamad

Starting 2015, Malaysia is introducing new courses related to the development of data scientist. The CEO of MDeC Corporation, Datuk Yasmin Mahmood said that several universities have been identified to offer the related courses, which have not been offered by any of the local universities. More data scientists are needed due to the importance of data management for various purposes. The sectors that are expected to demand highly for the data scientists are banking and finance sector (Afiq Anif, 2014). The use of data scientists have also proven very useful in politics and elections (Sharda, Delen, & Turban, 2014, p. 598).

So what is data scientist? Data scientist is a job frequently associated with Big Data or data science. In a very short time, it has become one of the most sought-after roles in the marketplace (Sharda, et al., 2014, p. 595). Data scientists use a combination of their business and technical skills to investigate Big Data looking for ways to improve current business analytics practices (from descriptive to predictive and prescriptive) and hence to improve decisions for new business opportunities. Data scientists use creative tools and techniques to turn data into actionable information for others to use for better decision making.

What about the educational background of a data scientist? Currently, there is no consensus on what educational background a data scientist has to have. Graduates in Computer Science, MIS, Industrial Engineering may be necessary but not sufficient to call someone a data scientist. However, one of the most sought-out characteristics of a data scientist is expertise in both technical and business application domains (Sharda, et al., 2014). Some companies have had good luck in recruiting some of the brightest data scientists with educational and work backgrounds in the physical or social sciences such as in esoteric fields like ecology and systems biology (Davenport & Patil, 2012).

Although there is no consensus on where data scientists come from, there is common understanding of what skills and qualities they are expected to possess.

A typical job post for data scientists (Sharda, et al., 2014, p. 596) is shown below.

ADVERTISEMENT: Our company is seeking a Data Scientist to join our Big Data Analytics team. Individuals in this role are expected to be comfortable working as a software engineer and a quantitative researcher. The ideal candidate will have a keen interest in the study of an online social network and a passion for identifying and answering questions that help us build the best products.

The main responsibilities of a data scientist are listed below:

- Work closely with a product engineering team to identify and answer important product questions
- Answer product questions by using appropriate statistical techniques on available data
- Communicate findings to product managers and engineers
- Drive the collection of new data and the refinement of existing data sources
- Analyze and interpret the results of product experiments
- Develop best practices for instrumentation and experimentation and communicate those to product engineering teams
Requirements

- M.S. or PhD in a relevant technical field or 4+ years of experience in a relevant role
- Extensive experience solving analytical problems using quantitative approaches
- Comfort with manipulating and analyzing complex, high-volume, high-dimensionality data from varying sources
- A strong passion for empirical research and for answering hard questions with data
- A flexible analytic approach that allows for results at varying levels of precision
- Ability to communicate complex quantitative analysis in a clear, precise and actionable manner
- Fluency with at least one scripting language such Python or PHP
- Familiarity with relational databases and SQL
- Expert knowledge of an analysis tool such as Research, Matlab or SAS
- Experience working with large data sets, experience working with distributed computing tools a plus (MapReduce, Hadoop, Hive, etc)

The demand for data scientist is not new. In 2013, our Prime Minister projected that data scientist gets the highest paid salary (Afiq Anif, 2014). Since people with this range of skills are rare, that explains why data scientists are in short supply. Because of the high demand, the starting salaries for data scientists are well above six figures (Sharda, et al., 2014).

Considering its great impact to the country, Minister of Education II, Datuk Seri Idris Jusoh has launched the ‘Big Data Competency Center’ at five higher learning institutes i.e. Universiti Utara Malaysia (UUM), Universiti Sains Malaysia (USM), Universiti Malaysia Terengganu (UMT), Universiti Malaysia Sarawak (UNIMAS) and Politek Ibrahim Sultan (PIS).

At UiTM, in order to prepare its graduates to be fully-employed and sought-out in this direction of the industry, almost a decade ago, the Faculty of Computer and Mathematical Sciences has started its own degree program known as Degree in Information Technology (Business Computing) and the program code is CS244. One of the subjects in the study plan is Business Intelligence. In this subject, students learn the theoretical concept of data warehouse, data mining and data analytics which are the important characteristics in data scientist. Interestingly this subject opens the path for the students to become a data scientist. However, given the short time frame, the single subject alone will not be sufficient to allow the students to obtain good grasp in some practical areas such as in data mining techniques, which is a vital skill a data scientist should possess. Therefore, one of the subjects, seen as the most suitable avenue for the students to gain hands-on experience in data mining tasks is the final year project or the FYP.

The use of data mining tasks such as prediction, association and clustering will be the heart of the data analysis. Popular algorithms such as Genetic Algorithms, Graph-based Matching, K-means, ANN to name a few will be applied as and when needed. As a result, students will gain some understanding in these areas and at the same time will be equipped with reasonable technical skills.

Recent anecdotes about data scientists reveal that (Sharda, et al., 2014):

- Data scientists turn Big Data into big value, delivering products that delight users and insight that informs business decisions.
- A data scientist is not only proficient to work with data, but also appreciates data itself as an invaluable asset.
- Today’s data scientists are the quants of the financial markets of the 1980s.
- By the year 2020, there will be 4.5 million new data scientist jobs, of which only one-third will be filled because of the lack of available personnel.

Looking at the current trend in the industry, higher institutions have to act upon this urgency and find ways to meet the industry’s demand. It is believed that FYP is the right avenue to prepare the students the necessary background of data mining task which is an important concept a data scientist must master.

References: