

Mathematics is the foundation of Data Science

**ABHISHEK CHAKRABORTY
& AMITAVA MUKHERJEE**

Within the purview of Data Management, a new discipline called "Data Science" has emerged. It should not be confused with Analytics—the buzzword of the 21st century. Unfortunately, however, Data Analysis, Data Analytics and Data Science are used interchangeably in the Indian context. It is an emerging, highly impacted and a rapidly expanding area of studies. Its usage can be found in various industries across different sectors, including but not limited to banking, telecom, manufacturing, logistics and healthcare.

As far as pure Data Science is concerned, the National Institute of Securities Markets in collaboration with the Department of Economics, University of Mumbai, has launched a part-time Post-Graduate Certificate Programme in Data Science. Nevertheless, there are various courses on Analytics offered by several reputed institutes which may serve certain purposes. One can attend some of these programmes even in distance learning mode. To name a few, IIM Lucknow offers a one-year distance learning course. IIM Bangalore also offers the same, which includes few campus

visits. ISB Hyderabad also offers a course on Data Analytics, which is again a combination of both distance learning and campus visits.

To develop adequate skills in the field of Data Science, one must have a clear understanding of certain statistical and mathematical tools, along with knowledge of software packages like R, Python etc. Students are expected to have a fair idea of the business or applied problems, where they are interested to implement the knowledge of Analytics and Data Science. Students with a good analytical ability in combination with business knowledge can make the best use

of far as imparting the software package knowledge is concerned.

Coming to the point of fees for such a course, it varies from Rs 4-6 lakh, as against Rs 16 lakh for a full-time two-year course being offered by IIM Calcutta in collaboration with ISI Calcutta and IIT Kharagpur. As far as the scope of employability is concerned, there is a dearth of quality human resources in this field. Students with reasonable knowledge of Data Science ended up being recruited by reputed firms on hefty pay packages. However, the situation is not that great for those who ended up doing a distance learning course. But they can still leverage this knowledge in moving up the ladder in the same organisation.

A course in Data Science may be lead to a great potential of earning initially, but an employee may struggle to perform well and consequently lose the job if the person does not learn the subject from the base and try to pick up jargon. There is a potential threat that many of these courses, espe-

cially those that are offered in distance learning mode, will lead to more and more misuses of statistics. A classic example is the misuse of the concept of p-value in validating hypothesis and models. On March 7 this year, the American Statistical Association for the first time in its 177 years

**If you want to excel
in Data Science,
before venturing
into distance
learning, you must
pursue graduation
or PG in maths,
operation research,
statistics, economics**

of existence had to issue a warning to scientific community and the business researchers on such misuse of p-value and to set up a guideline for the same. Similar misuse of concepts may complicate business problems even more, rather than providing sustainable solutions. As a result, the so-called Analytics bubble may burst sooner or later in India. Perhaps the only way to prevent this is an increasing dedication towards mathematical sciences.

If you want to excel in Data Science, before venturing into distance learning, it is desirable that you start preparing by pursuing

cludes few campus visits. ISB Hyderabad also offers a course on Data Analytics, which is again a combination of both distance learning and campus visits.

To develop adequate skills in the field of Data Science, one must have a clear understanding of certain statistical and mathematical tools, along with knowledge of software packages like R, Python etc. Students are expected to have a fair idea of the business or applied problems, where they are interested to implement the knowledge of Analytics and Data Science. Students with a good analytical ability in combination with business knowledge can make the best use of the same. The primary advantage of doing a distance learning course is driven by the fact that students are already into some job and thus are expected to have a good exposure to the business environment. It is a cheaper option than a full-time course on Data Analytics. The disadvantage is learning software packages. Teaching a software package through online mode might not be viable or could be extremely difficult for students to grab as well as for faculty members to teach. A classroom environment, thus, is a better option as

tion for the first time in its 177 years

of existence had to issue a warning to scientific community and the business researchers on such misuse of p-value and to set up a guideline for the same. Similar misuse of concepts may complicate business problems even more, rather than providing sustainable solutions. As a result, the so-called Analytics bubble may burst sooner or later in India. Perhaps the only way to prevent this is an increasing dedication towards mathematical sciences.

If you want to excel in Data Science, before venturing into distance learning, it is desirable that you start preparing by pursuing graduation/PG in mathematics, operations research, statistics or economics. Further, for engineering graduates and students from other discipline, it is desirable they have gone through several advanced courses in mathematical sciences during their graduation, apart from standard core courses.

Abhishek Chakraborty is assistant professor, Amitava Mukherjee is associate professor, Production, Operations & Decision Science area, XLRI-Xavier School of Management, Jamshedpur