Role Data Scientist in Biopharmaceutical Industries

Amit Kumar

Bharat Immunologicals and Biologicals Corporation, Limited (BIBCOL)

Abstract

Biopharmaceutical Industries and data science are rapidly growing industries worldwide and data scientists are using data science tools with the help of biopharmaceutical professionals to develop novel molecules for human and animal healthcare. In current article, author emphasised on the role of data scientists with job responsibilities in major thrust areas of biopharmaceutical Industries.

Keywords: *Data Science, Data Scientist, Job Responsibilities, Thrust areas, Biopharmaceutical Industry.*

Introduction

In the current scenario of the pandemic COVID-19 situation, biopharmaceuticals is an emerging and demanding field of life science, and data science is equally another popular discipline. Abundant literature is available on data science and biopharmaceutical industries with their achievements for public health benefits. Despite that, the author emphasizes using data science skills in developing problem-solving tools in the biopharmaceutical industry. Because there is litter literature available on combined studies on data science and the biopharmaceutical industry. This article provides a new opportunity for academic and industrial professionals about the newly combined discipline of data science and biopharmaceutical. The professionals may develop new applications by applying data science as a tool in biopharmaceutical industries.

Corresponding Author:
Dr. Amit Kumar
Deputy General Manager, (Quality Assurance and Training), Bharat Immunologicals and Biologicals Corporation Limited (BIBCOL), Oral Vaccine Plant, Chola, Bulandshahr, Pin Code – 203203, Uttar Pradesh, India
e-mail: akbibcol@gmail.com

Presently, data science uses as a tool for the collection, preparation, analysis, visualization, management, and preservation of an abundant amount of information in the form of data as described in figure 1.
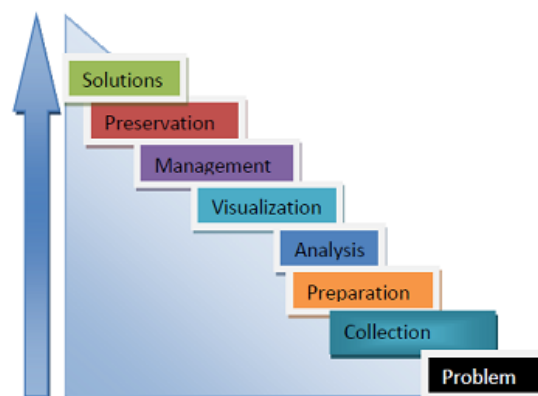

Figure 1: Concept and working of data science use as tool from problem to solution

Data science is the only way to extract knowledge and insights from structure/unstructured data by employing various scientific methods, processes, algorithms, and system in an automated manner. Data science is a progressive, sequential, and upward tool to solve task-specific problem. Data scientists are acquainted with statistical and computational competencies, paired with

domain knowledge, to analyze and interpret raw data and assist in decision-making. In the current scenario, data science is an interdisciplinary field. Due to this interdisciplinary collaborative work, the new discipline has a broad range of applications in the academic (Medical, Biomedical, Physics, Chemistry, & Biology) and industries (Biotechnology, Clinical, and Biopharmaceuticals). Graduate, Master's, and doctoral degree qualified professionals in Biology, Medical science, Biomedical science, Biotechnology, Microbiology, Biochemistry, Bio-static, Biophysics, Vaccine biology, Pharmaceutical science, and Healthcare [1]. They are actively engaged in the industrial production of life-saving drugs, preventive medicines, and therapeutic drugs & vaccines for human purposes. Author stamps that biopharmaceutical industries play a dynamic role in this field of health care for humans and animals.

Biopharmaceutical Industries are fastest growing industries and is also considering as basic needed for the healthy life of human and animal. On the basis of work requirement, It can describe in various field such as Pharmaceutics, Pharmacology, Novel Drug Delivery System, Targeted Drug Delivery System, Pharmaceutical Nanotechnology, Nanoparticles, *In vitro* evaluation, *In vivo* evaluation, Ex-vivo evaluation of plant extracts and formulations, Prodrug design, Pharmacokinetic studies, Pharmacodynamic studies, Pharmacogenomics, Pharmacoepidemiology, Physical pharmacy, Toxicology, Chemotherapy, Comparative pharmacology, Animal pharmacology, Posology etc [2]. We can divide these fields in three groups to make easy and well understandable and it is describing in the below table 1.

Table 1: Division of working fields in groups of Biopharmaceutical industry

S.No.	Group	Fields
1.	Drug Designing	Prodrug design and Drug designing.
2.	Drug Development	Pharmaceutics, Pharmacology, Novel Drug Delivery System, Targeted Drug Delivery System, Pharmaceutical Nanotechnology, Nanoparticles, Physical pharmacy, Chemotherapy, and Posology.
3.	Drug Evaluation	In vitro evaluation, In vivo evaluation, Ex-vivo evaluation of plant extracts and formulations, Pharmacokinetic studies, Pharmacodynamic studies, Pharmacogenomics, Pharmacoepidemiology, Toxicology, Comparative pharmacology, and Animal pharmacology.

After successfully completion of various offline and online courses of data science, biopharmaceutical professionals can do the R&D for innovation and development of novel molecules as drugs and vaccines against the emerging and re-emerging diseases and outbreaks. Nowadays, online mode is the most suitable way to acquire knowledge and skill through the certificate & degree courses and also take 100% free of cost certificate courses like Analytics Vidhya, Great Learning Academy, ICMR-NIOH, Cognitive Classes, Credly, NEPTEL etc[3,4,5,6,7,8]. You can choose your free certificate courses and start learning as per your requirement. The author has already experienced learning about data science tools through free and online mode courses from the above-mentioned portal of the organization like Machine Learning, Artificial Intelligence, Deep Learning & Natural Language Processing. I am giving below list of some data science courses for reference purpose.

List of data science courses

1. Successfully completed free online course on **"Introduction to Natural Language Processing"** with certification ID: rciquyt5pp from Analytics Vidhya in the month of April 2022.

2. Successfully completed free online course on **"Data Science Foundations"** in the month of May 2022 and certificate provided by Great Learning Academy, verify CBQQONAS at mygratelearning.com.

3. Successfully completed free online course on **"Statistics with SAS"** in the month of May 2022 and certificate verify at http://coursera.Org/verify/YNPP85POAAA3.

4. Successfully completed free online course on **"Machine Learning Certification Course for Beginners"** with certification ID: 97mo4kuerz from Analytics Vidhya in the month of May 2022.

5. Successfully completed free online course on **"Getting Started with Neural Networks"** with certification ID: vkytuebswx from Analytics Vidhya in the month of May 2022.

6. Successfully completed free online course on **"Getting Started with Decision Trees"** with certification ID: bdre29zaqa from Analytics Vidhya in the month of May 2022.

7. Successfully completed free online course on **"Application of AI"** in the month of May 2022 and certificate provided by Great Learning Academy, verify FPVDWDGT at mygratelearning.com.

8. Successfully completed free online course on **"AI in Healthcare"** in the month of June 2022 and certificate provided by Great Learning Academy, verify FPVDWDGT at mygratelearning.com.

9. Successfully completed free online course on **"Data Visualization using Python"** in the month of June 2022 and certificate provided by Great Learning Academy, verify MUUEQCOR at mygratelearning.com.

10. Successfully completed free online course on **"Supervised Machine Learning with Tree Based Models"** in the month of May 2022 and certificate provided by Great Learning Academy, verify JZHCMKMP at mygratelearning.com.

11. Successfully completed free online course on **"Introduction to Neural Networks and Deep Learning"** in the month of May 2022 and certificate provided by Great Learning Academy, verify YYMFPZDN at mygratelearning.com.

12. Successfully completed free online course on **"Cloud Foundations"** in the month of June 2022 and certificate provided by Great Learning Academy, verify WDOHWEJM at mygratelearning.com.

13. Successfully completed free online course on **"Introduction to Deep Learning"** in the month of June 2022 and certificate provided by Great Learning Academy, verify NHHIJXVX at mygratelearning.com.

14. Successfully completed free online course on **"Statistics for Data Science"** in the month of June 2022 and certificate provided by Great Learning Academy, verify MNXDGCPT at mygratelearning.com.

15. Successfully completed free online course on **"Introduction to Python"** with certification ID: oywxoqnnxx from Analytics Vidhya in the month of February 2022.

16. ICMR-NIOH, Ahmedabad for successful participation in Expert consultation Webinar on **"Spreadsheet Data Management: Simple, Convenient yet Powerful"** with Certificate on 09th March 2022.

Key roles of data scientist in various industries

Most industries are adopting data analysis for their growth. Data Scientists are in a growing demand not just in technology, but also in all other major sectors, including FMCG, logistics, and more. Out of the world, Google, Amazon, Apple, Microsoft, and Facebook have employed one-half of the total data scientists the world has. Key roles of the Data Scientists in various industries are describing below:

Key role as data scientists: A Data Scientist explores various data patterns to measure the impact on an organization. A key role of a data scientist is the ability to explain the importance of data in a simpler method to be understood by others. They are supposed to have a statistical knowledge of different programming languages required for solving complex problems.

Key role as data analyst: Data analyst has major role for analyzing data and/or figures to figure out a marketing trend in the current situation. He helps in providing a clear picture of the company's standing in the market. Once the desired goal is set by a company, a data analyst provides datasets to achieve the required goal. The role of a data analyst may change as per the requirement of a company. For example - Marketing division of company is required the services of data analyst at initial stage for consumer behaviour, reaction and understanding about the product to set different marketing strategies.

Key role as data engineer: Data Engineer works with the core of the organization and can be considered as the backbone of a company. They are the builder, designer, and manager of a large database. They are in charge of building data pipelines, enabling correct data flow, ensuring the data to reach the relevant departments. A data engineer has to work in collaboration with other data experts to communicate results with his colleagues. In a nutshell, a data engineer has to share his insights with the company through data visualization, helping the organization grow.

Key role as business intelligence analyst: A business intelligence analyst helps in analyzing the collected data to maximize the company's efficiency, hence generating more profits. Their role is more technical in nature then analytical, requiring more knowledge of popular machines. They have to serve as a bridge between business and IT, helping them improves. They are required in different industries only for knowledge specific industry.

Key role as marketing analyst: The role of a marketing analyst is to assist companies in their marketing division. They analyze and suggest which product to produce in large quantities and which product to discontinue. Monitoring customer satisfaction reports help in improving existing products and services. They decide which products to sell with the targeted customers and at which price.

Applications of data science in industries:

Data science is a well-established field and plays a crucial role in different industries because each industry has unprocessed, un-utilized, and unstructured/semi-structured data. According to a study, 2.5 quintillion bytes of data has generated in a single day. There is a need to process the massive amount of the data for information in systematic algorithms. The data should be used and then analyzed to develop new findings and applications to nourish the industrial outcome. In the present situation, data science is the only way of utilizing data and developing possible applications through applying data analysis for structure, statistical and mathematical techniques on collected data to detect underlying patterns and make predictions. Data science professionals can improve efficiency and control risks in the industries and are also crucial in the cost cuttings. Apart from this, the professional can help in various ways like Data Architect, Applications Architect, Infrastructure Architect, Enterprise Architect, Data Analyst, and Scientist.

Data science professionals should be acquainted with a broad range of qualities, which may be helpful in different industries[1]. Therefore, the professionals have more chances to get a job in various industries for their services and job responsibilities described in table 2.

Table 2: Jobs of data science professionals in Industries

S.No.	Jobs of data science professionals in Industries
1	Banking services
2	Financial services
3	Energy and Utilities
4	Pharmaceuticals and Healthcare
5	E-commerce
6	Media and Entertainment
7	Automobile Telecommunication
8	Retail
9	Travels
10	Hospitality

Out of these industries, the only healthcare industry has been selected and discussed for this article. Based on available literature on various website portals, it has been proved that data science has been used to extract information from daily data generated in healthcare institutions and industries. The healthcare industry is being used the data science as a tool for exploiting data for decision-making and meaningful information as data science applications in healthcare like Medical Image Analytics, Quantified Health, Data Science for Post-care Monitoring, Diagnosis, and Disease Prevention, Data Science for Drug Discovery, Natural Language Processing (NLP) for Electronic Health Records, and Bio-metrics in Healthcare [1,2].

Major thrust areas of the Biopharmaceutical industry

Based on the available literature, the author has summarized six major thrust areas of the Biopharmaceutical industry in the current situation as given below table 3 [2]..

Table 3: Tagged thrust areas of Biopharmaceutical industry

S.No.	Thrust area of Biopharmaceutical Industry
1	Designing and Development of new & novel candidates of vaccines
2	Designing and Development of new & novel molecules of drugs
3	Improvement in processing steps for vaccine production
4	Monitoring of vaccine safety data
5	Clinical trials of drugs and vaccines
6	Supply chain management of drugs and vaccines

Job responsibilities of data scientist in biopharmaceutical industry

Data scientists are acquainted with statistical and computational competencies, paired with domain knowledge, to analyze and interpret raw data and assist in decision-making. In the current scenario, data science is an interdisciplinary field. Due to this interdisciplinary collaborative work, the new discipline has a broad range of applications in the academic (Medical, Biomedical, Physics, Chemistry, & Biology) and industries (Biotechnology, Clinical, and Biopharmaceuticals). The continued rapid growth in data science, generation of big data sets, and advancement in algorithms have increased the demand for data science tools such as machine learning, artificial intelligence, Deep Learning etc [9,10,11].

In addition of the above-mentioned thrust areas, author is describing about key job responsibilities of the data scientists include, but not limited to:

- Develop user-friendly web applications and dashboards for purpose of data visualization, data analysis, documentation, validation and user testing of the applications.

- Application maintenance and develop data Science solutions through their lifecycle, from ideation to deployment.

- Uses data and machine learning methods to solve the problems

- Data cleaning and data wrangling

- Set-up databases and data pipelines

- Clear communication of work

- Support the development and transfer of oral drug products by evaluation of material, formulation and process parameters by means of applied statistics.

- Utilize Design of Experiment (DoE) concepts and Quality Based Designing (QbD) principles to achieve robust formulations and processes.

- Generation, management, and preservation of the parameter-base data and the process-based data for further use.

- Analyze the data inter-and intra-parameters.

- Retrieve and analyze the laboratory quality management system.

- Optimization and standardization of the production.

- Production process steps through preserving data in the database.

- Validation of the process steps based on available data set of the specific instrument or equipment used in the industries

- Management of the data generated from clinical trials.

- Preserved data use and analysis parameter-wise / process-wise / product-wise / other requirement basis.

- Inventory Management and Demand forecasting.

- Predictive analytics or Real-time data of performance and quality.

- Preventive maintenance and fault prediction.

- Price optimization.

- Optimization of the supply chain.

- Designing and development of the product.

- Automation and Robotization in the factory.

- Efficiency and Computer Vision Applications.

- Work required to investigate the other aspects

- Conducting the cost-effective clinical trials of drugs and vaccines: Data scientists can help to reduce the costs of clinical trials by enabling drug companies to implement as given below:

Data-Based Patient Selection: Pharmas use multiple data sources including social media and public health databases and more targeted criteria (e.g., genetic information) to identify which populations would work best in trials.

Real-Time Monitoring: Companies now monitor real-time data from trials to identify safety or operational risks and nip problems in the bud.

Drug Safety Assurance: Data scientists can even tap into side-effect data to predict whether a compound will provoke an adverse reaction before it even reaches trial. Working the University of California-San Francisco, researchers at Novartis have built computer models to do just that.

Author has gone through various websites to find out about the biopharmaceuticals industries, where data science was used to develop applications and utilized as a problem-solving tool in the industry[9,10,11].. Out of this, I am summarizing here only the top five industrial examples in below table 4.

Table 4: Description of the top five biopharmaceuticals industries in Artificial Intelligence (AI) as data science tool

S.No.	Company name and application in biopharmaceuticals industries
1	**M/s Sanofi** - Automation in medical literature review with minimum time (01 Second per article).
2	**M/s Bayer** - Digital transformation with digital technologies like AI in Research and Development to make simple, effective, and speed up the discovery and development of new drugs for patients to treat noncommunicable diseases like cardiovascular and oncological diseases
3	**M/s Janssen Pharmaceutica** - An AI-powered used to develop a new drug design system
4	**M/s Novartis** - Establishment of a new **AI Empowerment lab** for leveraging data and Artificial Intelligence to transform the developed medicines discovered, developed, and commercialised.
5	**M/s Pfizer** - For the advancement in Precision Oncology using AI and real-world data.

Conclusion

Based on need, problem-solving & knowledge gaps, Biopharmaceutical and Data Science industries are growing and most demanding industries nowadays. Therefore present article will help to provide knowledge for tagging the thrush areas of biopharmaceuticals and encourage the budding scientists and actively engaged scientists. Those are working in any field of the biopharmaceutical industry with knowledge of data science tools. In conclusion this article, the author strongly suggests and concludes the following points:

1. Biopharmaceutical professional to acquire knowledge of data science tools through various online and offline certificates, diplomas, nano-degree, graduate, and post-graduate courses.

2. Based on the acquired knowledge, the professionals will select and start working on the thrust area.

3. Author has already tagged six major thrust areas of the biopharmaceutical industry designing & develops novel vaccine candidates, designing & develops novel drug molecules, improving the involved procedure of vaccine production, and monitoring vaccine safety.

4. Author has also encouraged the professionals to contribute to public health issues through this newly developed field.

5. As discussed the above, data scientists need with the key responsibilities in Biopharmaceutical industries to do crucial jobs such as designing and conducting the clinical trials of drugs and vaccines.

At the end of this article, the author strongly recommends that the new combined disciplines will play a crucial role in filling the gaps and demands of biopharmaceutical industries through newly developed data science tools for novel applications as a new beginning.

Conflict of Interest: Author declare that there is no conflict of interest.

Source of Funding: Nil

Ethical Clearance: Nil

References

1. Amit Kumar. Data Science as a Tool in Biopharmaceutical Industry. In: Data Science Blogathon Edition -19; 2022. 05th May 2022: p. 1-6. Available online at: https://datahack.analyticsvidhya.com.

2. Amit Kumar. Applications of Data Science Tools in Biopharmaceutical Industry. In: Data Science Blogathon Edition – 21; 2022. 28th June 2022: p. 1-6. Available online at: https://datahack.analyticsvidhya.com.

3. Website of Analytics Vidhya – https://datahack.analyticsvidhya.com.

4. Website of Great Learning Academy – https://www.mygratelearning.com.

5. Website of ICMR-NIOH – http://www.nioh.org.

6. Website of Cognitive Classes –https://cognitiveclass.ai.

7. Website of Credly – https://info.credly.com.

8. Website of NEPTEL – https://nptel.ac.in.

9. Wikimedia Foundation. History of artificial intelligence. Available online at: https://en.wikipedia.org/wiki/History_of_artificial_intelligence. 2020. (Accessed April 19, 2020).

10. Z. Qin. Using artificial intelligence to read chest radiographs for tuberculosis detection: a multi-site evaluation of the diagnostic accuracy of three deep learning systems. Nat. Sci. Rep.; 2019. vol. 9: p. 15000. (doi: 10.1038/s41598-019-51503-3).

11. USAID. Artificial Intelligence in Global Health: Defining a Collective Path Forward. Available online at: https://www.usaid.gov/cii/ai-in-global-health. 2019. (Accessed April 12, 2021).