

# Applications of Data Science in Respective Engineering Domains

Motashim Rasool<sup>1</sup>, Vipin Kumar Chaudhary<sup>2</sup>

<sup>1</sup>Assistant Professor, SR Institute of Management & Technology, Lucknow, India

<sup>2</sup>Assistant Professor, BBD University, Lucknow, India

## Article Info

Volume 9, Issue 5

Page Number : 71-75

## Publication Issue

September-October-2022

## Article History

Accepted : 01 Sep 2022

Published : 08 Sep 2022

## ABSTRACT

Data Science is answerable to the every field of technology domain even it is Electronics and Communication Engineering, Electrical Engineering, Mechanical Engineering or Civil Engineering. For example in Civil Engineering following applications like Advance Predictions in construction works, Project risk Analysis, availability of water resources, Traffic forecasting and many more similarly we have many applications from Engineering Domain. Data Science approaches has adopted many technical techniques within the Engineering domain ranging from data analysis to neural network to deep learning and it has very vast variety of particular engineering domain solutions. Data scientists today draw largely from extensions of the “analyst” of years past trained in traditional disciplines. As data science becomes an integral part of many industries and enriches research and development, there will be an increased demand for more holistic and more nuanced data science roles.

Keywords : Data Science, Data Analysis, Data Wrangling, Data Visualization, IDEs, Python Libraries.

## I. INTRODUCTION

As per Today’s Scenario Data Science is widely used in various fields of Engineering as well as Business and Financial Analysis. Data science requires some tools and programming languages like Python, R, etc. There are many IDEs used for to work on these languages. Python is most popular and very strong language for the data analysis, the reason is Python has very huge set of Libraries.

Data Science Algorithms plays a most important role to map solutions of given problems like Risk Analysis, Sales Analysis, Financial Analysis and many more. To get expertise in data science one should know

Supervised Learning and Unsupervised Learning. These algorithms categorized in other languages based on type of problem.

The most common libraries for Data analysis are Numpy, Pandas, Matplotlib, and Seaborn. Basically Data Science profile is just like a story telling job but its require a very vast knowledge of Data Science Tools and Data Wrangling along with strong knowledge of Mathematics( Statistics and Probability). For Data visualization data correction and elimination of corrupt data is also important. A critical task in the education of future data scientists is to instill data acumen. This requires exposure to key concepts in data science, real-world data and problems that can

reinforce the limitations of tools, and ethical considerations that permeate many applications. Key concepts involved in developing data acumen include the following:

Mathematical foundations,  
 Computational foundations,  
 Statistical foundations,  
 Data management and curation,  
 Data description and visualization,  
 Data modeling and assessment,  
 Workflow and reproducibility,  
 Communication and teamwork,  
 Domain-specific considerations, and  
 Ethical problem solving.

## II. Data Science Algorithms

In Data Science we used specialy Supervised Learning and Unsupervised Learning algorithms along with many other algorithms for different types of problems.

- **Supervised Learning:**

- 1) In Supervised Learning used a model which is trained with labeled data for example using a historical data to check the CIBIL score for loan application of a candidate to grant permission or not.

Types of Supervised Learning

- 2) **K-Nearest Neighbors:**

K-nearest neighbors (KNN) is one of the simplest and powerful Supervised Learning algorithm where the classification is done based on k nearest data points. The idea behind KNN is that similar points are grouped together (clustered) by measuring the nearest data point's properties, we can classify a test data point. For example, we solve a standard classification problem where we want to predict a data point belongs to class X or class Y. Let  $k=3$ ; now we will test 3 nearest data point of the test data point, if two of them belongs to class X, we will declare the test data point as class X otherwise

class Y. The right value of K is found through cross-validation. It has a linear time complexity hence cannot be used for low latency applications.

- 3) **Linear Regression:**

Linear regression is a supervised data science algorithm. The objective is to find a Linear Space (hyper plane) where the maximum number of points lies in the hyper plane. For example, predicting rain is a standard regression problem where linear regression can be used. Linear regression assumes that the relation between the independent and dependent variables is linear, and there is very little or no multi co-linearity.

- 4) **Logistic Regression**

The geometric intuition is that we can separate different Class labels using a linear decision boundary. The output variable of logistic regression is categorical. In logistic regression, our main motto was to find a separating linear surface.

- 5) **Support Vector Machine**

We can consider the Support vector machine as an extension of this idea to find a hyperplane that maximizes the margin.

- 6) **Decision Tree**

Decision Tree is a nested If-Else based classifier that uses a tree-like graph structure to make the decision. Decision Trees are trendy and one of the most used supervised machine learning algorithms in the whole area of data science. It provides better stability and accuracy in most cases than other supervised algorithms and robust to outliers. The decision tree's output variable is usually categorical, but it also can be used to solve regression problems.

- **Unsupervised Algorithms**

Unsupervised algorithms are used for the tasks where the data is unlabelled. The most popular use case of unsupervised algorithms is clustering.

Clustering is the task of grouping together similar data points without manual intervention. Let's discuss some of the popular unsupervised algorithms here

• **K Means**

K Means is a randomized unsupervised algorithm used for clustering. K Means follows the below steps

1. Initialize K points randomly (c1,c2..ck)
2. For each point (Xi) in the data set Select nearest Ci {i=1,2,3..k} Add Xi to Ci
3. Recomputed the centric using proper metrics (i.e. intracluster distance)
4. Repeat step (2)(3) until converges

2) **K-Means++**

The initialization step in K means is purely random, and based on the initialization, the clustering changes drastically. K means++ solves this problem by initializing k in a probabilistic way instead of pure randomization. K means++ is more stable than classic K means.

3) **K -Medoids**

K medoids is also a clustering algorithm based on K means. The main difference between the two is the centroids of K means does not necessarily exist in the data set, which is not the case for K medoids. K medoids offer better interpretability of clusters. K means minimizes the total squared error, while K medoids minimize the dissimilarity between points.

**Applications of Data Science in Engineering and Industrial Domains:**

• **Mechanical Engineering:**

Data Science helps to forecast the problems in mechanical engineering like Equipments performance, Accuracy of output of machines results, factors that are directly or indirectly affects the workflow. The future of machine learning will be deeply rooted in physics- The

more we want to use machine learning in solving real-world problems, the more we will need professionals who can work with non-linear and dynamic systems in an environment with constantly changing factors.

Mechanical engineers have long been working in different areas. They have much to offer due to their experience with sensors, systems, and fluids. For instance, they can collect and convert energy from sources such as the sun, wind, and waves and account for turbulence in airplane designs.

• **Electrical Engineering:**

The field of Data Science is indebted to electrical engineering, Data science has adopted many techniques within the signal processing field ranging from signal analysis to neural network to deep learning and many more. Many consider machine learning as an outgrowth of statistical signal processing techniques

• **Civil and Construction Engineering:**

The world is overloaded with data. It results in a steady improvement in technologies. The construction industry has always been a victim of poor planning, management, budgeting, miscalculations, cost overruns, low return on construction assets, mistakes in proportions, and insufficient means for the building. Data science is called upon to make these problems miserable and facilitate construction on each of its levels. The construction companies use the benefit of data science to improve construction sites and manage the building process.

• **Finance:**

Use cases of Data Science in Financial Solutions:

- 1) Automating risk management
- 2) Managing customer data
- 3) Predictive analytics
- 4) Real-time analytics
- 5) Fraud detection
- 6) Consumer analytics

- 7) Algorithmic Trading
- 8) Deep personalization and customization

- **Pharmaceutical Industries:**

Use cases of Data Science in Pharma Industries:

- 1) Personalized Medication Plans
- 2) Marketing and Sales
- 3) Enhanced Drug Discovery and Development
- 4) Improved Drug Trials
- 5) Genomics
- 6) Genome Editing
- 7) Machine Learning
- 8) Patient Follow-ups
- 9) Safety and Risk Management
- 10) Operational Optimization

### III. CONCLUSION

As per future of Technologies and the problems associated with the performance and accuracy of results in various different fields, Data Science is the solution and Data Analytics also plays an important role in analyzing the problem and target the audience for marketing.

Data science is emerging as a field that is revolutionizing science and industries alike. Work across nearly all domains is becoming more data driven, affecting both the jobs that are available and the skills that are required. As more data and ways of analyzing them become available, more aspects of the economy, society, and daily life will become dependent on data. It is imperative that educators, administrators, and students begin today to consider how to best prepare for and keep pace with this data-driven era of tomorrow. Undergraduate teaching, in particular, offers a critical link in offering more data science exposure to students and expanding the supply of data science talent.

Data science education is well into its formative stages of development; it is evolving into a self-supporting

discipline and producing professionals with distinct and complementary skills relative to professionals in the computer, information, and statistical sciences. However, regardless of its potential eventual disciplinary status, the evidence points to robust growth of data science education that will indelibly shape the undergraduate students of the future. In fact, fueled by growing student interest and industry demand, data science education will likely become a staple of the undergraduate experience. There will be an increase in the number of students majoring, minoring, earning certificates, or just taking courses in data science as the value of data skills becomes even more widely recognized. The adoption of a general education requirement in data science for all undergraduates will endow future generations of students with the basic understanding of data science that they need to become responsible citizens. Continuing education programs such as data science boot camps, career accelerators, summer schools, and incubators will provide another stream of talent. This constitutes the emerging watershed of data science education that feeds multiple streams of generalists and specialists in society; citizens are empowered by their basic skills to examine, interpret, and draw value from data.

### IV. REFERENCES

- [1]. Python Data Science Handbook, J. VanderPlas, O'Reilly (2016)
- [2]. Abu-Mostafa YS, Magdon-Ismail M, Lin H-T. Learning from data: a short course. [United States]: : AMLBook.com 2012
- [3]. Practical Statistics for Data Scientists O'Reilly
- [4]. <https://www.kdnuggets.com/2019/03/top-8-data-science-use-cases-manufacturing.html>
- [5]. <https://towardsdatascience.com/>
- [6]. Data Science For Dummies Paperback – by Lillian Pierson
- [7]. Data Science from Scratch 2e: First Principles with Python Paperback – by Joel Grus

- [8]. Storytelling with Data: A Data Visualization Guide for Business Professionals Paperback – by Cole Nussbaumer Knaflic

**Cite this article as :**

Motashim Rasool, Vipin Kumar Chaudhary, "Applications of Data Science in Respective Engineering Domains", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 9 Issue 5, pp. 71-75, September-October 2022. Available at doi : <https://doi.org/10.32628/IJSRST22958>  
Journal URL : <https://ijsrst.com/IJSRST22958>