

Punyashlok Ahilyadevi Holkar Solapur University, Solapur



Name of the Faculty: Science & Technology

(As per New Education Policy 2020)

Syllabus: - Data Science

Name of the Course: B.Sc. I (Sem. I &II)

(To be implemented from Academic Year 2024)

Semester-wise Structure for
B. Sc. Data Science (Honors/Research) Programme
as per NEP-2020
(w.e.f. – June 2024)

B.Sc. Part-I (Semester-I) Data Science						
Course Type	Course Code	Course Title	Credits	Teaching hours/week		
				T	P	Total
DSC	DSC-1	Introduction to DBMS	2	2	--	2
	DSC-1 (P)	Data Science Practical-I	2	--	4	4
SEC	SEC-1	Introduction to R Programming	2	2	--	2
IKS	IKS	Evolution of Statistical Science in India	2	2	--	2
B.Sc. Part-I (Semester-II) Data Science						
DSC	DSC-2	Introduction to Data Science	2	2	--	2
	DSC-2 (P)	Data Science Practical-II	2	--	4	4
OE	OE-1	Fundamentals of Data Science	2	2	--	2
VSC	VSC-1	Introduction to Python Programming	2	2	--	2

B. Sc. Part-I (Data Science) Semester-I

DSC-1	Theory	INTRODUCTION TO DATA BASE MANAGEMENT SYSTEM	Credits: 02 Hours: 30
--------------	---------------	--	----------------------------------

Course Objectives:

- This course focuses on describing a sound introduction to the discipline of database management systems,
- This course gives a good formal foundation on the relational model of data and usage of Relational Algebra.

Course Outcomes:

After successful completion of this course, students are expected to:

- Explain the features of database management systems and Relational database.
- Create and populate a RDBMS for a real-life application, with constraints and keys, using SQL.
- Retrieve any type of information from a data base by formulating complex queries in SQL.
- Analyze the existing design of a database schema and apply concepts of normalization to design an optimal database.
- Build indexing mechanisms for efficient retrieval of information from a database.

Course Content

Unit-1: Basics of DBMS (15 hrs.)

Characteristics of database approach, data models, database system architecture and data independence. Entity types, relationships, constraints. Relational model concepts, relational constraints, relational algebra, SQL queries Mapping ER/EER model to relational database, functional dependencies, Lossless decomposition, Normal forms (upto BCNF).

Unit-2: File organization in Databases (15 hrs.)

ACID properties, concurrency control Operations on files, File of Unordered and ordered records, overview of File organizations, indexing structures for files (Primary index, secondary index, clustering index), Multilevel indexing using B and B+ trees.

Reference Books:

1. R. Elmasri, S.B. Navathe, *Fundamentals of Database Systems* 6 Edition, Pearson Education, 2010.
2. R. Ramakrishnan, J. Gehrke, *Database Management Systems* 3 Edition, McGraw-Hill, 2002.
3. A. Silberschatz, H.F. Korth, S. Sudarshan, *Database System Concepts* 6 Hill, 2010. Edition, McGraw-Hill
4. R. Elmasri, S.B. Navathe *Database Systems Models, Languages, Design and application Programming*, 6th Edition, Pearson Education, 2013.

DSC-1 (P)	Practical	DATA SCIENCE PRACTICAL-I	Credits: 02 Hours: 60
----------------------	------------------	---------------------------------	----------------------------------

Course Outcomes:

At the end of this course students are expected to be able to

1. Understand the framework of database management system.
2. Perform different operations on databases.

At least 10 practicals to be conducted based on DSC-1

SEC-1	Theory	INTRODUCTION TO R PROGRAMMING	Credits: 02 Hours: 30
--------------	---------------	--------------------------------------	----------------------------------

Course Objectives: The main objective of this course is to introduce the R software to the students. The course is aimed to inculcate analytical skills in students in order for them to fulfill current industry standards. At the end of the course, students are expected to be able,

1. To identify object types in R.
2. To input and edit data in R.
3. To format data in R.
4. To analyze data statistically in R.

Course Outcomes:

At the end of the course, students are able to

1. Use R in to solve real-world problems.
2. Create graphical representation of the data in using R.
3. Handle data using existing R functions.
4. Perform statistical analysis of the data using R.

Course Content**Unit 1: Getting Acquainted with R (15 hrs.)**

Introduction: R interpreter, Introduction to major R data structures like vectors, matrices, arrays, list and data frames, Control Structures, vectorized if and multiple selection, functions.

Unit 2. Statistical Analysis using R (15 hrs.)

Installing, loading and using packages: Read/write data from/in files, extracting data from web-sites, Clean data, Transform data by sorting, adding/removing new/existing columns, centering, scaling and normalizing the data values, converting types of values, using string in-

built functions, Statistical analysis of data for summarizing and understanding data, Visualizing data using scatter plot, line plot, bar chart, histogram and box plot.

Reference Books:

1. Gardener, M.(2017). Beginning R: The statistical programming language, WILEY
2. Lawrence, M., & Verzani, J. (2016). Programming Graphical User Interfaces in R. CRC press. (ebook)

IKS	Theory	EVOLUTION OF STATISTICAL SCIENCE IN INDIA	Credits: 02 Hours: 30
------------	---------------	--	----------------------------------

Course Objectives:

1. To trace the evolution of Statistics as a subject in India.
2. To learn about renowned Indian Statisticians and their works.
3. To understand the working of various Statistical organizations in India.

Course Outcomes:

At the end of this course, students are able to

1. Acknowledge, appreciate and value the rich heritage offered by India in areas of Science and Technology.
2. Gauge the immensity of the contributions made by Indian scientists to world knowledge.
3. Identify the erstwhile lesser known applications of Statistics since ancient times in India.
4. Recognize the significance of contributions of Indian Statisticians.
5. Identify the role of Statistical organizations towards the progress and development of India.

Course Content

Unit 1: Overview of IKS (10 hrs)

- 1.1 Introduction to IKS
- 1.2 Need of IKS
- 1.3 Scope of IKS
- 1.4 Living examples of IKS
- 1.5 Traditional IKS

Unit 2: IKS and Modern Knowledge (10 hrs.)

- 2.1 Indian Mathematics.
- 2.2 Indian Statistics
- 2.3 The outreach of Indian knowledge System
- 2.4 Indian Education
- 2.5 Foundational Literature of Indian Civilization

Unit 3: Discipline Related IKS (Statistical Heritage of India) (10 hrs.)

- 3.1 Contributions of Indian Statisticians: P. C. Mahalanobis, P. V. Sukhatme, Raghu Raj Bahadur, Debabrata Basu, Gopinath Kallianpur, Keshav Raghavan Nair, Calyampudi Radhakrishna Rao, Vasant Shankar Huzurbazar.
- 3.2 Official Statistics in India - Past and Present
- 3.3 Historical perspective of Official Statistics in India.
- 3.4 Statistical organizations in India and their functions: CSO, Indian Statistical Institute (ISI), NSSO, IIPS, Bureau of Economics and Statistics.

Reference Books:

1. Buchanan, F. (1807) Survey of Eastern India, Report submitted to the Court of Directors, London.
2. Chaudhuri, S. B. (1964) History of the Gazetteers of India, Publication Division, New Delhi.
3. Dutta A. K., The Concept of Arithmetic Mean in Ancient India, in *25 Years Gone By*, ISIREA(2017):158-192.
4. Ghosh, J. K., Mitra, S. K., and Parthasarathy, K. R. (1992) Glimpses of India's Statistical Heritage, Wiley Eastern, New Delhi.
5. Ghosh, J. K, Maiti, P., Rao, T. J., and Sinha, B. K. (1999) Evolution of Statistics in India, International Statistical Review, 67, 13-34.
6. Glaz Sarah, Poetry Inspired by Mathematics: A Brief Journey through History, J. of Mathematics and the Arts, 5, 171-183, 2011
7. Glaz Sarah, Mathematical Ideas in Ancient Indian Poetry, Proceedings of Bridges Enschede, July 2013
8. Indian Contributions to Science - Compiled by Vijnana Bharati.
9. Jarret, H. S. (1894) Translation of Ain-i-Akbari, Asiatic Society of Bengal, Vol. II, p. vii.
10. Mukhopadhyay, N. (1997) A conversation with Sujit Kumar Mitra, Statistical Science, 12, 61-75.
11. <https://pragyata.com/the-untold-foundations-of-modern-economics-did-adam-smith-plagiarise-kautilya/>
12. https://unstats.un.org/unsd/wsd/docs/India_wsd_history.pdf
13. Raju, C. K. 'Probability in Ancient India'. In Handbook of Philosophy of Statistics, edited by Paul Thagard Dov M. Gabbay and John Woods, 7:1175–96. Handbook of Philosophy of Science. Elsevier, 2011.
14. Rangarajan L. N., Kautilya - The Arthashastra, Penguin Books India (P) Ltd.
15. Rao, Talluri. (2010). Official Statistics in India: The Past and the Present. Journal of Official Statistics. 26. 215-231.
16. Sihag Balbir S., Kautilya's Arthashastra: The Origin of Statistical Economics During -4CE, JRSA, Vol. 2, No. 1, June-2013, pp 1-14

B. Sc. Part-I (Data Science) Semester-II

DSC-2	Theory	INTRODUCTION TO DATA SCIENCE	Credits: 02 Hours: 30
--------------	---------------	-------------------------------------	--

Course Objectives:

- Introduction to Data Science and Analytics gives students a solid knowledge of the ideas, tools, and applications of data science across several domains.
- The course covers core concepts such data collection, preparation, analysis, visualization, and interpretation.
- Through hands-on activities and real-world case studies, students will obtain practical expertise with data analytics tools and programming languages typically used in data science.
- This course is designed to help students who want to advance their education or pursue professions in data science, machine learning, or artificial intelligence.

Course Outcomes:

Upon completion of Introduction to Data Science and Analytics, students will:

- Understand the fundamental principles and techniques of data science and analytics.
- Acquire proficiency in data acquisition, preprocessing, analysis, and visualization.
- Apply statistical methods and hypothesis testing techniques to analyze data and derive insights.
- Gain exposure to machine learning algorithms and their applications in solving real-world problems.
- Explore career opportunities in data science, machine learning, and artificial intelligence across various industries.

Course Content

Unit-1: Overview of data science and its applications.

(15 hrs.)

- Role of data scientists and data analysts
- Introduction to data lifecycle and data-driven decision-making

Data Acquisition and Preprocessing

- Techniques for data collection and storage
- Data cleaning, transformation, and normalization
- Introduction to data wrangling tools (e.g., Pandas, SQL)

Unit-2: Exploratory Data Analysis (EDA)

(15 hrs.)

Descriptive statistics and data visualization techniques

- Exploratory analysis using statistical plots and charts
- Hands-on exercises with data visualization.

Reference Books:

1. *"Data Science for Business"* book by Foster Provost and Tom Fawcett
2. *"Introduction to Statistical Learning"* book by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani

3. *"Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow"* book by Aurélien Géron
4. *"Data Science from Scratch"* book by Joel Grus

DSC-2 (P)	Practical	DATA SCIENCE PRACTICAL-II	Credits: 02 Hours: 60
----------------------	------------------	----------------------------------	----------------------------------

Course Outcomes:

At the end of this course students are expected to be able to

1. Apply basic concepts of data science to real life problems.
2. Store and extract information from databases.

At least 10 practicals to be conducted based on DSC-2

OE-1	Theory	FUNDAMENTALS OF DATA SCIENCE	Credits: 02 Hours: 30
-------------	---------------	-------------------------------------	----------------------------------

Course Objectives:

- Introduction to Data Science and Analytics gives students a solid knowledge of the ideas, tools, and applications of data science across several domains.
- The course covers core concepts such data collection, preparation, analysis, visualization, and interpretation.
- Through hands-on activities and real-world case studies, students will obtain practical expertise with data analytics tools and programming languages typically used in data science.
- This course is designed to help students who want to advance their education or pursue professions in data science, machine learning, or artificial intelligence.

Course Outcomes:

Upon completion of Introduction to Data Science and Analytics, students will:

- Understand the fundamental principles and techniques of data science and analytics.
- Acquire proficiency in data acquisition, preprocessing, analysis, and visualization.
- Apply statistical methods and hypothesis testing techniques to analyze data and derive insights.
- Gain exposure to machine learning algorithms and their applications in solving real-world problems.
- Explore career opportunities in data science, machine learning, and artificial intelligence across various industries.

Course Content

Unit-1: Overview of data science and its applications.**(15 hrs.)**

- Role of data scientists and data analysts
- Introduction to data lifecycle and data-driven decision-making

Data Acquisition and Preprocessing

- Techniques for data collection and storage
- Data cleaning, transformation, and normalization
- Introduction to data wrangling tools (e.g., Pandas, SQL)

Unit-2: Exploratory Data Analysis (EDA)**(15 hrs.)****Descriptive statistics and data visualization techniques**

- Exploratory analysis using statistical plots and charts
- Hands-on exercises with data visualization.

Reference Books:

1. Agarwal B. L. (2003). Programmed Statistics, Second Edition, New Age International Publisher, New Delhi.
2. Bhat B. R., Srivenkatramana T. and Rao Madhava, K. S. (1996). Statistics: A Beginner's Text: Vol. I, New Age International (P) Ltd.
3. Goon A. M., Gupta M.K. and Dasgupta B. (2002). Fundamentals of Statistics, Vol. I and II, 8th Edition, The World Press Pvt. Ltd. Kolkata.
4. Gupta S. P. (2002): Statistical Methods, Sultan Chand and Sons, New Delhi.
5. Gupta V.K. and Kapoor S.C.: Fundamentals of Mathematical Statistics-Sultan & Chand.

VSC-1	Theory	INTRODUCTION TO PYTHON PROGRAMMING	Credits: 02
--------------	---------------	---	--------------------

Course Objectives:

1. To get acquainted with Python programming.
2. To understand objects and structures in python.

Course Outcomes: Upon successful completion of this course, a student will be able to:

- i) Develop programmes in Python.
- ii) Perform data organization, data manipulation, statistical and mathematical computations, and data analysis using Python.

Course Content

Unit 1: Introduction, installation, keywords, identifiers: variables, constants, literals; comments, Operators, statements and expressions, data types with methods: numbers, string, lists, tuple, dictionary, set; indexing and slicing of each data type, data type conversion, built-in functions, control statements and loops, list comprehensions, user defined functions, anonymous/lambda function, local and global variables, modules: math, stat, random; creating own modules.

(15 L)

Unit 2: Concept of library and its working, Data storage, manipulation, visualization and analysis using the libraries: Numpy, Pandas, Scipy, statsmodels, Matplotlib, Seaborn, Regular Expressions (RegEx), Ski-kit learn. **(15 L)**

Reference Books:

1. Gowrishankar S. & Veena, A. (2018). Introduction to Python programming. CRC Press.
2. Guttag J. V. (2021). Introduction to Computation and Programming Using Python: With Application to Computational Modeling and Understanding Data. Mit Press.
3. Haslwanter T. (2016). An Introduction to Statistics with Python. With Applications in the Life Sciences. Switzerland: Springer International Publishing.
4. Nelli F. (2018). Python data analytics with Pandas, NumPy, and Matplotlib.
5. Unpingco J. (2016). Python for probability, statistics, and machine learning (Vol. 1). Springer International Publishing.
6. Vander Plas, J. (2016). Python data science handbook: Essential tools for working with data. " O'Reilly Media, Inc."