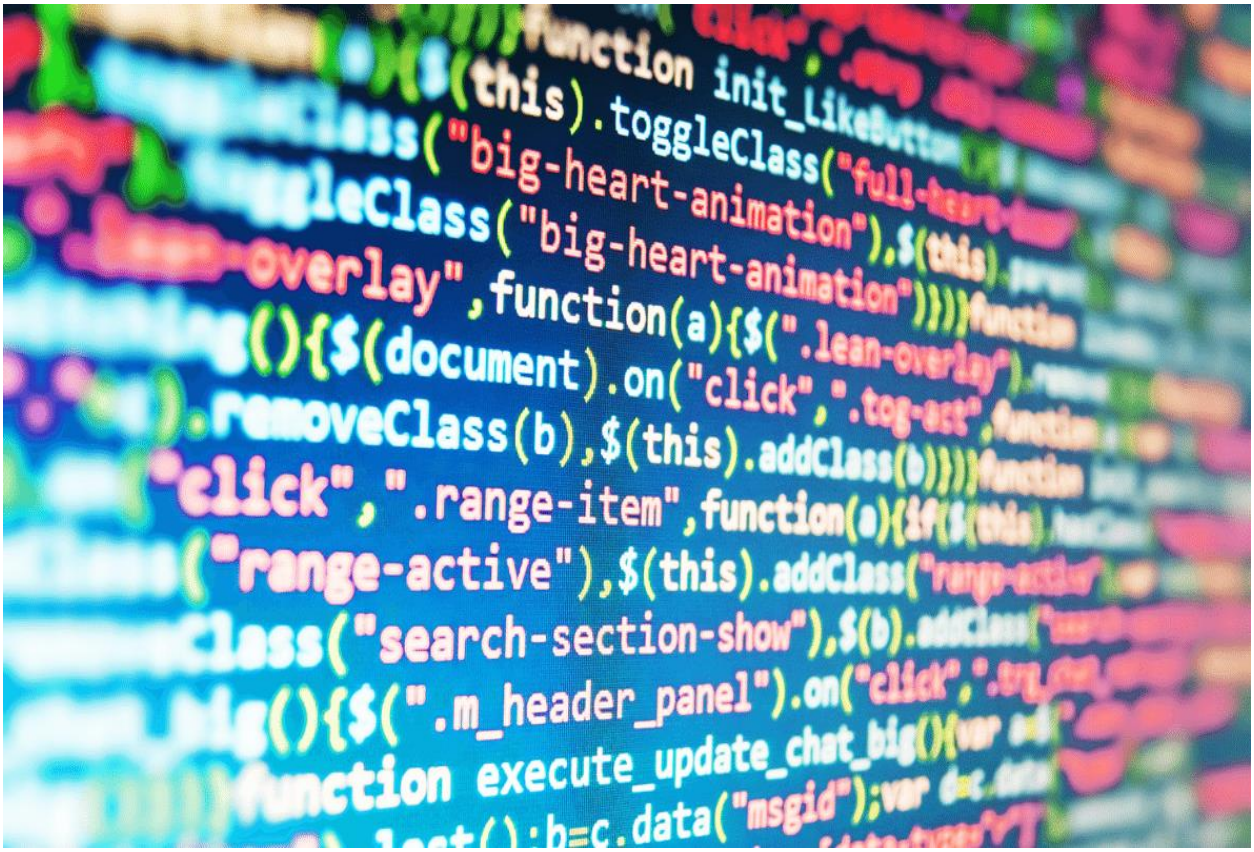


Detailed Syllabus for Computer Science & Engineering with Specialization in Data Science

UEM
श्रद्धावान लभते ज्ञानम्
UNIVERSITY OF ENGINEERING & MANAGEMENT
Good Education, Good Jobs

DEPT. OF COMPUTER SCIENCE & ENGINEERING

UNIVERSITY OF ENGINEERING & MANAGEMENT, JAIPUR

PREAMBLE

Education plays an enormously significant role in the building of a nation. There are quite a large number of educational institutions, engaged in imparting education in our country. Majority of them have entered recently into semester system to match with international educational pattern. However, our present education system is churning out youth who have to compete locally, regionally, nationally as well as globally. The present alarming situation necessitates transformation and/or redesigning of system, not only by introducing innovations but developing "learner-centric approach.

Majority of Indian higher education institutions have been following marks or percentage-based evaluation system, which obstructs the flexibility for the students to study the subjects/courses of their choice and their mobility to different institutions. There is need to allow the flexibility in education system, so that students depending upon their interests can choose inter-disciplinary, intra-disciplinary and skill-based courses. This can only be possible when choice based credit system (CBCS), an internationally acknowledged system, is adopted. The choice based credit system not only offers opportunities and avenues to learn core subjects but also explore additional avenues of learning beyond the core subjects for holistic development of an individual. The CBCS will undoubtedly facilitate benchmarking of our courses with best international academic practices.

Advantages of the choice based credit system:

•Shift in focus from the teacher-centric to student-centric education.

•Student may undertake as many credits as they can cope with (without repeating all courses in a given semester if they fail in one/more courses).

•CBCS allows students to choose inter-disciplinary, intra-disciplinary courses, skill oriented papers (even from other disciplines according to their learning needs, interests and aptitude) and more flexibility for students.

•CBCS makes education broad-based and at par with global standards. One can take credits by combining unique combinations.

•CBCS offers flexibility for students to study at different times and at different institutions to complete one course (ease mobility of students). Credits earned at one institution can be transferred to another institution.

CHOICE BASED CREDIT SYSTEM

The Indian Higher Education Institutions have been moving from the conventional annual system to semester system. Currently many of the institutions have already introduced the Choice Based Credit System. The semester system accelerates the teaching-learning process and enables vertical and horizontal mobility in learning. The credit based semester system provides flexibility in designing curriculum and assigning credits based on the course content and hours of teaching. The Choice Based Credit System provides a 'cafeteria' type approach in which the students can take courses of their choice, learn at their own pace, undergo additional courses to acquire more than the required credits and adopt an interdisciplinary approach to learning.

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

It has been a necessity to align higher education with the emerging needs of the economy so as to ensure that the graduates of higher education system have adequate knowledge and skills for employment and entrepreneurship since last few years. The higher education system has to incorporate the requirements of various industries in its curriculum, in an innovative and flexible manner while developing a well-groomed graduate. CSE department aims to encourage research and innovation in the field of Machine Learning, Cyber security, Artificial Intelligence and other allied areas such as Computational Theory, Cloud Computing, Blockchain Technology, Data Science, Big Data Analytics and many more. The objective of the Computer Science & Engineering Programme with Specialization in Block Data Science is to prepare students to undertake careers involving innovative technologies, develop a problem solving capability, or to opt for advanced studies for research oriented careers.

In order to give due importance to practical applications as well as theoretical aspects of CSE, the curriculum for the Computer Science & Engineering Programme with Specialization in Data Science covers most of the foundational aspects as well as develops engineering skills for problem solving.

JOB OPORTUNITIES

Booming IT sector in India has plenty of jobs for fresh computer science graduates. Candidates with a high percentage of mark and good communication skills as well as sound computer knowledge do not face problem in getting a job. Computer engineers can get jobs in non-IT companies like universities, research, private and public industries, government departments, business organizations, commercial organizations and the manufacturing sector, etc. Besides the Computer Engineers have plenty of options to work in IT companies in departments such as design, development, assembly, manufacture, and maintenance, etc. Software Developers: Software developers are professionals who are concerned with facets of the software development process which involves activities such as design and coding, computer programming, project management, etc. The People who are having the skills and interest to develop then Data science is a very good career option because of the reasons like: Soaring demand analytics professionals and there are a lot of job opportunities and meeting the skill gap. A couple of years ago think about the Computer Science.

PROGRAMME EDUCATIONAL OBJECTIVES (PEO)

PEO 01: High Quality Engineering Design and Development Work: Graduates of the program will engage in the effective practice of computer science and engineering to identify and solve important problems in a diverse range of application areas.

PEO 02: Real Life Problem Solving: To educate students with proficiency in core areas of Computer science & Engineering and related engineering so as to comprehend engineering trade-offs, analyses, design, and synthesize data and technical concepts to create novel products and solutions for the real life problems.

PEO 03: Leadership: Graduates of the program will engage in successful careers in industry, academia and attain positions of importance where they have impact on their business, profession and community.

PEO 04: Lifelong Learning: Graduates of the program will adapt to contemporary technologies, tools and methodologies to remain at the frontier of computer science and engineering practice with the ability to respond to the need of a challenging environment.

PROGRAM OUTCOME (PO)

PO	Summary	Description
PO1	**Engineering knowledge**	Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
PO2	**Problem analysis**	Identify, formulate, research literature, and analyse complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
PO3	**Design /development of solutions**	Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
PO4	**Conduct investigations of complex problems**	Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
PO5	**Modern tool usage**	Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modelling to complex engineering activities with an understanding of the limitations.
PO6	**The engineer and society**	Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities

		relevant to the professional engineering practice.
PO7	**Environment and sustainability**	Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.
PO8	**Ethics**	Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
PO9	**Individual and team work**	Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
PO10	**Communication**	Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
PO11	**Project management and finance**	Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.
PO12	**Life-Long Learning**	Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change

TYPES OF COURSES

1. Courses are the subjects that comprise the Computer Science & Engineering Programme with Specialization in Data Science.
2. A course may be designed to comprise lectures, tutorials, laboratory work, fieldwork, outreach activities, project work, vocational training, viva, seminars, term papers, assignments, presentations, self-study etc. or a combination of some of these components.
3. The learning outcomes of each course will be defined before the start of a semester.
4. Following are the course types:

   i. **Core Course (CC):** This is a course, which is to be compulsorily studied by a student as a core requirement to complete the requirement of B.Tech in Computer Science & Engineering with Specialization in Data Science.

   ii. **Elective Course:** An elective course is a course, which can be chosen from a pool of courses. It is intended to support the discipline of study by providing an expanded scope, enabling exposure to another discipline/domain and nurturing a student's proficiency and skill. An elective may be of following types:

   a) **Discipline Elective Courses (DE):** It is an elective course that adds proficiency to the students in the discipline.

   b) **Open Elective Courses (OE):** It is an open elective course taken from other engineering disciplines and enhances the generic proficiency and interdisciplinary perspective of students.

   c) **Specialization Elective Courses (SE):** This is a course, which is to be compulsorily studied by a student as a core requirement to complete the requirement of B.Tech in Computer Science & Engineering with Specialization in Data Science.

   iii. **Obligatory Courses:**

   a) **Mandatory Courses (MC)**: It can be taken from among a pool of foundation courses, which aim at value-based education. They may provide hands-on training to improve competencies and skills or provide education on human, societal, environmental and national values.

   b) **Internship/Training/Project/Dissertation (PTI):** Course designed to acquire special/advanced knowledge, such as supplement study/support study to a project work, and a candidate studies such a course on his own with an advisory support by a teacher/faculty member is called dissertation/project

   c) **Humanities, Social Sciences & Management (HSM):** It is an elective course taken from non-engineering disciplines (humanities, social sciences and management) that broadens the perspective of an engineering student.

   d) **Basic Science Courses (BSC):** It is based upon content that leads to fundamental knowledge enhancement in sciences, and basic engineering principles.

   e) **Engineering Science Courses (ESC):** It is based upon content that leads to fundamental knowledge enhancement in basic Engineering Principles.

   f) **NPTEL (NPT):** "Essential Studies for Professionals Skill & Skill Development for Professionals" courses designed to encourage and enrich the students for the technical and professional exams.

g) **General Studies Courses (GSC):** "Essential Studies for Professionals Skill & Skill Development for Professionals" courses designed to encourage and enrich the students for the technical and professional exams.

h) Mandatory Additional Requirements (MAR): A student has to do the following things to achieve the MAR points: The student should engage herself / himself in activities outside the curriculum. Join different types of Clubs of NSCBIP, write something for the wall magazine, remain active in outer society, participate in Tech Fests activities, etc.

5. Each credit course contributes certain credits to the programme. A course can be offered either as a full course (4 credits) or as a half course (2 credits). A full course is conducted with 3 hours of lectures and either 1 hour of tutorial or 2 hours of practical work per week. A half course is conducted with 2 hours of lectures. There are also some exceptional electives with 3 credits and 1 credit.

Definition of Credit:

1 Hr. Lecture (L) per week	1 Credit
1 Hr. Tutorial (T) per week	1 Credit
1 Hr. Practical (P) per week Or 2 Hr. Practical (Lab)/week	0.5 Credits Or 1 Credit

6. A project work/dissertation is considered as a special course involving application of the knowledge gained during the course of study in exploring, analyzing and solving complex problems in real life applications. A candidate completes such a course with an advisory support by a faculty member.
7. **Mandatory Courses** may be offered. They do not carry credits but aim at expanding knowledge or bridging deficiency in knowledge or skill.
8. A course may have pre-requisite course(s) that are given in the Semester-wise Course Allocation scheme.
9. A student can opt for a course only if he/she has successfully passed its pre- requisite(s).
10. A student has to register for all courses before the start of a semester.
11. **Program codes:** The codes for various undergraduate programmes are as follows:
    i. Civil Engineering: CE
    ii. Computer Science & Engineering: CS
    (Specialization in Data Science (CSDS))
    iii. Electronics and Communication Engineering: EC
    iv. Electrical Engineering: EE
    v. Mechanical Engineering: ME
12. **Departmental Course Codes:** The codes for departmental core courses and discipline-specific electives are specific to each discipline. The first two characters are derived from departmental codes listed above. The third character is 'C' for core courses and 'D' for discipline-specific courses and 'INT' for Dissertation/Project/Training/Internship. This is followed by a digit sequence number:
    i. CSCyyy: Core Course
    ii. CSDyyy: Discipline-Specific Elective Courses

iii. DSyyy: Specialization Elective Courses

iv. XXXyyy: Open Elective Courses (Depends on the respective Dept.)

v. INTyyy: Project/Training/Internship/ Dissertation

13. **Common Elective Course Codes:** All disciplines will follow a common code as shown below. The 3-digit sequence number 'yyy' is taken from the respective tables of different types of courses.

i. HSMyyy: Humanities, Social Sciences & Management Course

ii. BSCyyy: Basic Science Course

iii. MCyyy: Mandatory Course

iv. GSCyyy: General Studies Courses

v. MARyyy: Mandatory Additional Requirements

Here, yyy will be follow by a sequence of digit.

14. **General Electives:** A student may take a course under the category of General Elective (GE) offered by any other Department of the Institute under the categories of Core Course (CC) and Discipline Specific Electives (DE). However, such options shall be offered to a student as per prescribed guidelines of the Institute.

15. The opting of a course by the student will depend upon the requisites for that course and with the consent of the course advisor.

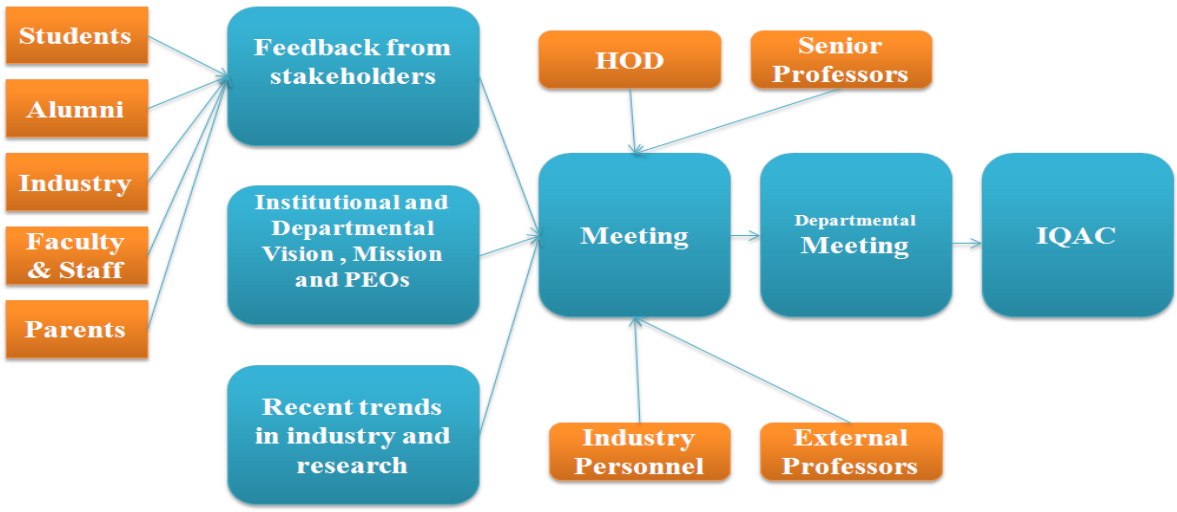
PROCESS FOR DESIGNING THE PROGRAM CURRICULUM

The process for designing the program curriculum involves consideration of the following aspects:

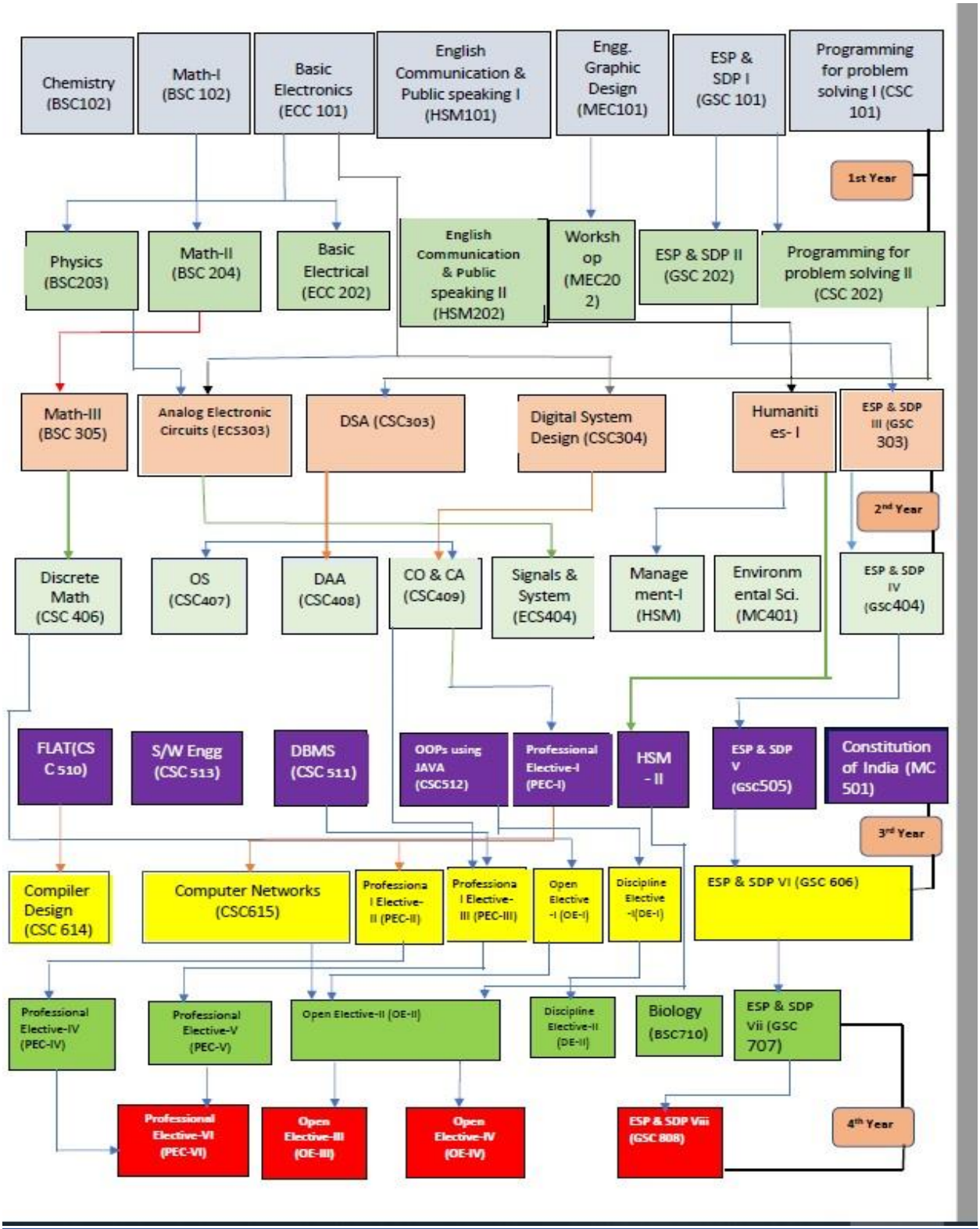
i) Feedback from stakeholders

ii) Institutional and Departmental Vision, Mission and PEOs

iii) Recent trends in industry and research



PREREQUISITE TREE



SCHEME – SEMESTER WISE COURSE ALLOCATION

First Semester Syllabus

Sl No.	Type	Subject Code	Topic	L	T	P	S	Credit Points
1.	BSC	BSC101	Chemistry	3	1	3	0	5.5
2.	BSC	BSC102	Mathematics-I (Calculus & Linear Algebra)	3	1	0	0	4
3.	ESC	ECC101	Basic Electronics Engineering	2	0	2	0	3
4.	ESC	MEC101	Engineering Graphics & Design	1	0	2	0	2
5.	ESC	CSC101	Programming for Problem Solving-I (C)	2	0	0	2	2
6.	HSM	HSM101	English Communication & Public Speaking Skills-I	0	0	2	0	1
7.	GSC	GSC101	ESP & SDP-I	2	0	0	2	2
			Total	13	2	9	4	19.5/28

#Students will undergo a mandatory Induction Program

Second Semester Syllabus

Sl No.	Type	Subject Code	Topic	L	T	P	S	Credit Points
1.	BSC	BSC203	Physics (Semi-Conductor Physics)	3	1	3	0	5.5
2.	BSC	BSC204	Mathematics-II(Probability & Statistics)	3	1	0	0	4
3.	ESC	EEC202	Basic Electrical Engineering	2	0	2	0	3
4.	ESC	MEC202	Workshop & Manufacturing Practices	1	0	2	0	2
5.	ESC	CSC202	Programming for Problem Solving-II (Python)	2	0	0	2	2
6.	HSM	HSM202	English Communication & Public Speaking Skills-II	1	0	3	0	2
7.	GSC	GSC202	ESP & SDP-II	2	0	0	2	2
8.	NPT	NPT201	(NPTEL/MOOCs)	-	-	-	-	2
			Total	15	0	9	4	22.5/28

#(NPT201)NPTEL/MOOCs are based on the respective year's offered courses.

Third Semester Syllabus

Sl No.	Type	Subject Code	Topic	L	T	P	S	Credit Points
1.	BSC	BSC305	Mathematics-III (Differential Calculus)	2	0	0	0	2
2.	PCC	CSC303	Data Structure & Algorithms	3	0	3	0	4.5
3.	ESC	ECS303	Analog Electronic Circuits	2	0	3	0	3.5
4.	PCC	CSC304	Digital Electronics	3	0	3	0	4.5
5.	PCC	CSC305	IT Workshop (Sci Lab/MATLAB)	1	0	4	0	3
6.	HSM	HSM---	Humanities-I	3	0	0	0	3
7.	GSC	GSC303	ESP & SDP-III	2	0	0	2	2
8.	MAR	MAR381	Mandatory Additional Requirements (MAR)	0	0	0	1	0.5
9.	NPT	NPT302	(NPTEL/MOOCs)	-	-	-	-	2
			Total	17	0	13	3	25/33

#(NPT302): NPTEL/MOOCs are based on the respective year's offered courses.

Suggestive Choice Based Subjects

Sl No	Type	Subject Code	Topic	L	T	P	Credit Points
1.	HSM	**HSM303**	Organizational Behavior	3	0	0	3
2.	**HSM**	**HSM304**	**Values and Ethics in Profession**	3	0	0	3
3	HSM	HSM305	Industrial Psychology	3	0	0	3

Fourth Semester Syllabus

Sl No.	Type	Subject Code	Topic	L	T	P	S	Credit Points
1.	PCC	CSC406	Discrete Mathematics	3	0	0	0	3
2.	PCC	CSC407	Operating System	3	0	3	0	4.5
3.	PCC	CSC408	Design & Analysis of Algorithms	3	0	3	0	4.5
4.	PCC	CSC409	Computer Organization & Architecture	3	0	3	0	4.5
5.	ESC	ECS404	Signals & System	3	0	0	0	3
6.	HSM	HSM---	Management-I	3	0	0	0	3
7.	GSC	GSC404	ESP & SDP-IV	2	0	0	2	2
8.	MC	MC401/402	Environmental Sciences/Disaster Management	0	0	0	2	0
9.	MAR	MAR484	Mandatory Additional Requirements (MAR)	0	0	0	1	0.5
10.	NPT	NPT403	(NPTEL/MOOCs)	-	-	-	-	2
			Total	20	0	9	5	27/34

#(NPT403): NPTEL/MOOCs are based on the respective year's offered courses.

Suggestive Choice Based Subjects

Sl No.	Type	Subject Code	Topic	L	T	P	Credit Points
1.	HSM	HSM406	Human Resource Development and Organizational Behavior	3	0	0	3
2.	HSM	**HSM407**	**Economics & Financial Accounting**	3	0	0	3
3.	HSM	HSM408	Economics for Engineers	3	0	0	3

Fifth Semester Syllabus

Sl No.	Type	Subject Code	Topic	L	T	P	S	Credit Points
1.	PCC	CSC510	Formal Language & Automata Theory	3	0	0	0	3
2.	PCC	CSC511	Data Base Management System	3	0	3	0	4.5
3.	PCC	CSC512	Object Oriented Programming Using Java	2	0	3	0	3.5
4.	PCC	CSC513	Software Engineering	2	0	2	0	3
5.	PEC/SEC	---	Professional/Specialization Elective-I	3	0	0	0	3
6.	HSM	HSM---	Humanities-II	3	0	0	0	3
7.	MC	MC501	Constitution of India/Essence of Indian Knowledge Tradition	0	0	0	0	0
8.	GSC	GSC505	ESP & SDP-V	2	0	0	2	2
9.	PTI	INT501	Internship/Project-I	0	0	0	1	1
10.	NPT	NPT504	(NPTEL/MOOCs)	-	-	-	-	2
			Total	19	0	6	3	25/28

#(NPT504): NPTEL/MOOCs are based on the respective year's offered courses.

Suggestive Choice Based Subjects

Sl No.	Type	Subject Code	Topic	L	T	P	Credit Points
1.	PEC*	CSP501	Embedded Systems	2	0	2	3
2.	PEC**	CSP502	AI & Machine Learning	2	0	2	3
3.	HSM	HSM509	Industrial Psychology	2	0	0	2
4.	HSM	**HSM510**	**Principle of Management**	2	0	0	2
5.	HSM	HSM511	Total Quality Management	2	0	0	2

Note: Students can opt any Professional Track from 5th Sem Onwards/Data Science Specialization

***Track:** IOT, Cybersecurity & Blockchain Track

****Track:** AI & Machine Learning Track

Sixth Semester Syllabus

Sl No.	Type	Subject Code	Topic	L	T	P	S	Credit Points
1	PCC	CSC614	Compiler Design	3	0	3	0	4.5
2	PCC	CSC615	Computer Networks	3	0	3	0	4.5
3	PEC/SEC	---	Professional/Specialization Elective-I -II	2	0	2	0	3
4	PEC/SEC	---	Professional/Specialization Elective-I -III	3	0	0	0	3
5	OE	---	Open Elective-I	2	0	0	0	3
6	DE	CSD---	Discipline Elective-I	2	0	2	0	3
7.	GSC	GSC606	ESP & SDP-VI	2	0	0	2	2
8.	PTI	INT604	Internship/Industrial Training/Project-II	0	0	0	1	1
9.	NPT	NPT605	(NPTEL/MOOCs)	-	-	-	-	2
			Total	17	0	10	3	26/30

#(NPT605): NPTEL/MOOCs are based on the respective year's offered courses.

Suggestive Choice Based Subjects

Sl No.	Type	Subject Code	Topic	L	T	P	Credit Points
1.	PEC*	CSP5603	Blockchain Technology	3	0	0	3
2.	PEC**	CSP604	Soft Computing	3	0	0	3
3.	PEC*	CSP605	AI & Machine Learning	3	0	2	4
4.	PEC**	CSP606	Big Data Analytics	3	0	2	4
5.	OE	BSC607	Numerical Methods & Operation Research	2	0	2	3
6.	OE	BSC608	Operations Research	2	0	2	3
7.	OE	BSC609	Statistics for Data Analysis	2	0	2	3
8.	OE*	CSD601	Blockchain Technology	3	0	0	3
9.	OE**	CSD602	Big Data Analytics	3	0	2	4
10.	DE	CSD603	Web Technology	2	0	2	3
11.	DE	CSD604	Computer Graphics	2	0	2	3
13.	DE	**CSD605**	**Software Project Management**	2	0	2	3
14.	DE	CSD606	E-Commerce	2	0	2	3

Seventh Semester Syllabus

Sl No.	Type	Subject Code	Topic	L	T	P	Credit Points
1	PEC/SEC	---	Professional/Specialization Elective-I -IV	3	0	0	3
2	PEC/SEC	---	Professional/Specialization Elective-I -V	3	0	0	3
3.	OE	---	Open Elective-II	3	0	0	3
4.	BSC	BSC710	Biology	2	1	0	3
5.	GSC	GSC707	ESP & SDP-VII	2	0	2	2
6.	PTI	INT705	Internship/Industrial Training/Project-III	0	0	8	4
			Total	13	1	8	18/21

#Students will undergo project/training/internship in the industry / research organization / reputed Institute during the vacation

Suggestive Choice Based Subjects

Sl No.	Type	Subject Code	Topic	L	T	P	Credit Points
1.	PE*	CSP707	Blockchain Technology	3	0	0	3
2.	PE**	CSP708	Natural Language Processing	3	0	0	3
3.	PEC*	CSP709	Digital Forensics	3	0	0	3
4.	PEC**	CSP710	Computer Vision	3	0	0	3
5.	CSD*	CSD707	Cyber Law, IPR & Ethics	3	0	0	3
6.	CSD**	CSD708	Natural Language Processing	3	0	0	3
7.	OE	CSD709	Wireless Sensor Network & Network Security	3	0	0	3
8.	OE	CSD710	Neural Network and Application	3	0	0	3
9.	OE	CSD711	Real Time Operating System	3	0	0	3
10.	OE	CSD712	Distributed System	3	0	0	3
11.	BSC	BSC711	Statistics For Data Analytics	3	0	0	3
12.	BSC	BSC712	Statistical Methods For Decision Making	3	0	0	3
13.	BSC	BSC713	Exploratory Data Analysis	3	0	0	3
14.	BSC	BSC714	Graph Theory	3	0	0	3

Eighth Semester Syllabus

Sl No.	Type	Subject Code	Topic	L	T	P	Credit Points
1.	PEC/SEC	---	Professional/Specialization Elective-I -VI	3	0	0	3
2.	OE	---	Open Elective-III	3	0	0	3
3.	OE	---	Open Elective-IV	3	0	0	3
4.	GSC	GSC808	ESP & SDP - VIII	2	0	2	2
5.	PTI	INT806	Internship Industrial Training/Project-IV	0	0	12	6
6.	CC	CSC881	Grand Viva	0	0	0	2
			Total	11	0	14	19/25

#Students will undergo project/training/internship in the industry / research organization / reputed Institute during the vacation.

Suggestive Choice Based Subjects

Sl No.	Type	Subject Code	Topic	L	T	P	Credit Points
1.	PEC*	CSP811	Digital Forensics	3	0	0	3
2.	PEC**	CSP812	Deep Learning	3	0	0	3
3.	OE	CSD813	Data Mining & Data Ware Housing	3	0	0	3
4.	OE	CSD814	Information Theory & Coding	3	0	0	3
5.	OE	CSD815	Advanced Algorithms	3	0	0	3
6.	OE	CSD816	Digital Image Processing	3	0	0	3

Professional Elective Courses:

Subject Code	IOT, Cybersecurity & Blockchain Track	Subject Code	AI & Machine Learning Track
CSP501	Embedded Systems	CSP502	AI & Machine Learning
CSP603	Blockchain Technology	CSP504	Soft Computing
CSP605	AI & Machine Learning	CSP606	Big Data Analytics
CSP707	Cyber Security	CSP608	Natural Language Processing
CSP709	Cyber Law, IPR & Ethics	CSP710	Computer Vision
CSP811	Digital Forensics	CSP712	Deep Learning

Specialization Course:

Subject Code	Data Science
DS501	Introduction to Data Science
DS602	Introduction to AI and ML
DS603	Computational Data analytics
DS704	Web Data Mining
DS705	Analyzing, Visualizing and Applying data science with python
DS806	Exploratory Data Analysis
DS807	Data Scientist's Tool Box

Note: Refer to Computer Science & Engineering Syllabus for "syllabus of other subject".

TITLE OF COURSE: Introduction to Data Science
COURSE CODE: DS501
L-T-P: 3-0-0
CREDITS: 3

Pre-requisite: Knowledge is also assumed of basic concepts in data base management system, and mathematics.

Introduction:
Data science is an inter-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from many structural and unstructured data.
• To Provide the knowledge and expertise to become a proficient data scientist;
• Demonstrate an understanding of statistics and machine learning concepts that are vital for data science;
• Produce Python code to statistically analyse a dataset;
• Critically evaluate data visualisations based on their design and use for communicating stories from data;

Course Outcomes (CO):
After completion of course, students would be able:
CO1: To explain how data is collected, managed and stored for data science;
CO2: To understand the key concepts in data science, including their real-world applications and the toolkit used by data scientists;
CO3: To implement data collection and management scripts using MongoDB.

Mapping of Course Outcomes (CO) and Program Outcomes (PO):

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓			✓								✓
CO2	✓	✓		✓							✓	✓
CO3	✓	✓	✓		✓				✓		✓	✓

Course Contents:
Module 1:
Introduction to Data Science, Different Sectors using Data science, Purpose and Components of Python in Data Science.
Module 2:
Data Analytics Process, Knowledge Check, Exploratory Data Analysis (EDA), EDA Quantitative technique, EDA- Graphical Technique, Data Analytics Conclusion and Predictions.
Module 3:
Feature Generation and Feature Selection (Extracting Meaning from Data)- Motivating application: user (customer) retention- Feature Generation (brainstorming, role of domain expertise, and place for

imagination)- Feature Selection algorithms.

Module 4:

Data Visualization- Basic principles, ideas and tools for data visualization, Examples of inspiring (industry) projects- Exercise: create your own visualization of a complex dataset.

Module 5:

Applications of Data Science, Data Science and Ethical Issues- Discussions on privacy, security, ethics- A look back at Data Science- Next-generation data scientists.

Lab Work:

1. Python Environment setup and Essentials.
2. Mathematical computing with Python (NumPy).
3. Scientific Computing with Python (SciPy).
4. Data Manipulation with Pandas.
5. Prediction using Scikit-Learn
6. Data Visualization in python using matplotlib

Textbooks:

1. Data Sciences & Analytics, V.K. Jain, Khanna Publishing House.
2. Business Analytics: The Science of Data - Driven Decision Making, U Dinesh Kumar, John Wiley & Sons.
3. Introducing Data Science: Big Data, Machine Learning, and More, Using Python Tools, Davy Cielen, John Wiley & Sons.
4. Joel Grus, Data Science from Scratch, Shroff Publisher/O'Reilly Publisher Media
5. Annalyn Ng, Kenneth Soo, Numsense! Data Science for the Layman, Shroff Publisher
6. Cathy O'Neil and Rachel Schutt. Doing Data Science, Straight Talk from The Frontline. O'Reilly Publisher.
7. Jure Leskovek, Anand Rajaraman and Jeffrey Ullman. Mining of Massive Datasets. v2.1, Cambridge University Press.

References Books:

1. Jake VanderPlas, Python Data Science Handbook, Shroff Publisher/O'Reilly Publisher Media.
2. Philipp Janert, Data Analysis with Open Source Tools, Shroff Publisher/O'Reilly Publisher Media.

TITLE OF COURSE: Introduction to AI and ML
COURSE CODE: DS602
L-T-P: 3-0-2
CREDITS: 4

Pre-requisite: Knowledge is also assumed of basic concepts in data base management system, and mathematics.

Introduction:

- To understand basics of machine learning in data science.
- To understand various basic machine learning algorithm that can be used with various type of data.

Course Outcomes (CO):

After completion of course, students would be able:

CO1: To explain how data is collected, managed and stored for data science;

CO2: To use various type of Machine learning model

CO3: To implement various ML algorithms on data models

Mapping of Course Outcomes (CO) and Program Outcomes (PO):

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓			✓								✓
CO2	✓	✓		✓							✓	✓
CO3	✓	✓	✓		✓				✓		✓	✓

Course Contents:

Module 1:

Linear Regression: Basic facts of linear regression, implementation of linear regression, case studies of linear regression using data set.

Module 2:

Logistic Regression: Basic facts and implementation of logistic regression, solve a case study to predict output using existing data set

Module 3:

Clustering and Principle Component Analysis: K means and hierarchical clustering, how to make market strategies using clustering, recommendation and PCA.

Module 4:

Support Vector Machine: basics of SVM and use it to detect the spam emails and recognize alphabets.

Model Selection and advanced regression: use of Lasso and Ridge

Lab Work:

1. Use python to predict employee attrition in a firm and help them plan their manpower. (take data set from kaggle).
2. Create customer clusters using different market strategies on a data set.
3. Make a movie recommendation system.
4. Develop a prediction mechanism to predict which employee can go on leave in a company in near future.
5. Recognizing alphabets using SVM.

Textbooks:

1. Machine Learning using Python , U Dinesh Kumar and Manaranjan Pradhan, John Wiley & Sons.
2. A Classical Approach to Artificial Intelligence, M.C. Trivedi, Khanna Publishing House.
3. Machine Learning, V.K. Jain, Khanna Publishing House.
4. Advanced Data Analytics Using Python: With Machine Learning, Deep Learning by By Sayan Mukhopadhyay, Apress.

5. Practical Data Mining" by Monte F. Hancock, Auerbach Publication.
6. "Machine Learning for Absolute Beginners: A Plain English Introduction (Second Edition)" by Oliver Theobald.
7. Practical Data Science with R, Nina Zumel, John Wiley & Sons.

References Books:

1. Python for Data Science for Dummies, John Paul Mueller, Luca Massaron, John Wiley & Sons.
2. Big Data and Analytics, Seema Acharya and Subhashini Chellappan, Wiley Publication.
3. Introduction to Machine Learning, Jeeva Jose, Khanna Publishing House.

TITLE OF COURSE: Computational Data analytics
COURSE CODE: DS603
L-T-P: 3-0-0
CREDITS: 3

Pre-requisite: Knowledge is also assumed of basic concepts in data base management system, and mathematics.

Introduction:

• To learn how to think about your study system and research question of interest in a systematic way in order to design an efficient sampling and experimental research program.

• To understand how to analyze collected data to derive the most information possible about your research questions.

Course Outcomes (CO):

After completion of course, students would be able to:
CO1: Explain how data is collected, managed and stored for data science;
CO2: When to use which type of Machine learning model.
CO3: Implement various ML algorithms on data models.

Mapping of Course Outcomes (CO) and Program Outcomes (PO):

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓			✓								✓
CO2	✓	✓		✓							✓	✓
CO3	✓	✓	✓		✓				✓		✓	✓

Course Contents:

Module 1:

Introduction to R Computing language. Best practices in executing Reproducible Research in data science, Sampling and Simulation. Descriptive statistics, and the creation of good observational sampling designs.

Module 2:
Data visualization, Data import and visualization, Introduction to various plots

Module 3:
Frequentist Hypothesis Testing, Z-Tests, Power Analysis.

Module 4:
Linear regression, diagnostics, visualization, Likelihoodist Inference, Fitting a line with Likelihood, Model Selection with one predictor

Module 5:
Bayesian Inference, Fitting a line with Bayesian techniques, Multiple Regression and Interaction Effects, Information Theoretic Approaches

Lab Work:
1. To give a basic insight of R and its various libraries.
2. Libraries in R. R as a Data Importing Tool, Dplyr. Forcats.
3. Simulation and Frequentist Hypothesis testing, Simulation and Power.
4. Bayesian computation in R, Fitting a line with Bayesian techniques.

Textbooks:
1. Beginner's Guide for Data Analysis using R Programming, Khanna Publishing House
2. Practical Data Science with R, Nina Zumel, John Wiley & Sons.
3. Big Data & Hadoop, V.K. Jain, Khanna Publishing House.
4. N. C. Das, Experimental Designs in Data Science with Least Resources, Shroff Publisher Publisher.
43
5. Hadley Wickham, Garret Grolemund, R for Data Science, Shroff Publisher/O'Reilly Publisher Publisher.

References Books:
1. Benjamin M. Bolker. Ecological Models and Data in R. Princeton University Press, 2008. ISBN 978-0-691-12522-0.
2. John Fox and Sanford Weisberg. An R Companion to Applied Regression. Sage Publications, Thousand Oaks, CA, USA, second edition, 2011. ISBN 978-1-4129-7514-8.

TITLE OF COURSE: Web Data Mining
COURSE CODE: DS704
L-T-P: 3-0-2
CREDITS: 4

Pre-requisite: Knowledge is also assumed of basic concepts in data base management system, and mathematics.

Introduction:
• To learn how to extract data from the Web.
• To understand how to analyze collected data to derive the most information

Course Outcomes (CO):

After completion of course, students would be able to:

CO1: To explain how data is can be collected from the Web.

CO2: To extract data and information from the webpages.

CO3: To make decision based on the data collected.

Mapping of Course Outcomes (CO) and Program Outcomes (PO):

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓			✓								✓
CO2	✓	✓		✓							✓	✓
CO3	✓	✓	✓		✓				✓		✓	✓

Course Contents:

Module 1:

Introduction to internet and WWW, Data Mining Foundations, Association Rules and Sequential Patterns, Basic Concepts of Association Rules, Apriori Algorithm, Frequent Itemset Generation, Association Rule Generation, Data Formats for Association Rule Mining, Mining with multiple minimum supports, Extended Model, Mining Algorithm, Rule Generation.

Module 2:

Mining Class Association Rules, Basic Concepts of Sequential Patterns, Mining Sequential Patterns on GSP, Mining Sequential Patterns on Prefix Span, Generating Rules from Sequential Patterns

Module 3:

Concepts of Information Retrieval, IR Methods, Boolean Model, Vector Space Model and Statistical Language Model, Relevance Feedback, Evaluation Measures, Text and Web Page Pre-processing, Stopword Removal, Stemming, Web Page Preprocessing, Duplicate Detection, Inverted Index and Its Compression, Inverted Index, Search using Inverted Index, Index Construction, Index Compression, Latent Semantic Indexing, Singular Value Decomposition, Query and Retrieval, Web Search, Meta Search, Web Spamming.

Module 4:

Link Analysis, Social Network Analysis, Co-Citation and Bibliographic Coupling, Page Rank Algorithm, HITS Algorithm, CommModuley Discovery, Problem Definition, Bipartite Core CommModuleies, Maximum Flow CommModuleies, Email CommModuleies, Web Crawling, A Basic Crawler Algorithm – Breadth First Crawlers, Preferential Crawlers, Implementation Issues – Fetching, Parsing, Stopword Removal, Link Extraction, Spider Traps, Page Repository, Universal Crawlers, Focused Crawlers, Topical Crawlers, Crawler Ethics and Conflicts.

Module 5:

Opinion Mining, Sentiment Classification, Classification based on Sentiment Phrases, Classification Using Text Classification Methods, Feature based Opinion Mining and Summarization, Problem Definition, Object feature extraction, Comparative Sentence and Relation Mining, Opinion Search and Opinion Spam. Web Usage Mining, Data Collection and Preprocessing, Sources and Types of Data, Key Elements of Web Usage Data Preprocessing, Data Modeling for Web Usage Mining, Discovery and Analysis of Web

Usage Patterns, Session and Visitor Analysis, Cluster Analysis and Visitor Segmentation, Association and Correlation Analysis, Analysis of Sequential and Navigation Patterns.

Lab Work:

1. To give a basic insight of R and its various libraries.
2. Libraries in R. R as a Data Importing Tool, Dplyr. Forcats.
3. Simulation and Frequentist Hypothesis testing, Simulation and Power.
4. Bayesian computation in R, Fitting a line with Bayesian techniques.

Textbooks:

1. Mining the Web: Discovering Knowledge from Hypertext Data, Soumen Chakrabarti, Morgan Kaufmann Publishers.
2. Bing Liu, Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, Springer Publications, 2011.
3. Jiawei Han, Micheline Kamber, Data Mining: Concepts and Techniques, Second Edition, Elsevier Publications 2010.
4. Anthony Scime, Web Mining: Applications and Techniques, 2005.
5. Kowalski, Gerald, Mark T Maybury: Information Retrieval Systems: Theory and Implementation, Kluwer Academic Press, 1997.
6. Mathew Russell, Mining the Social Web 2nd Edition, Shroff Publisher/O'Reilly Publisher Publication.

References Books:

1. Data Mining and Data Warehousing Principles and Practical Techniques, Parteek Bhatia, Cambridge University Press.
2. Data Mining & Business Intelligence, Balram Krishan, Khanna Publishing House

TITLE OF COURSE: Analysing, Visualizing and Applying data science with python
COURSE CODE: DS705
L-T-P: 3-0-2
CREDITS: 4

Pre-requisite: Knowledge is also assumed of basic concepts in data base management system, and mathematics.

Introduction:

- To learn how to use python for data science.
- To understand and use all the tools and libraries of python for data science.

Course Outcomes (CO):

After completion of course, students would be able to:

CO1: To explain how data is can be collected from the Web.
CO2: To extract data and information from the webpages.
CO3: To make decision based on the data collected.

Mapping of Course Outcomes (CO) and Program Outcomes (PO):

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓			✓								✓
CO2	✓	✓		✓							✓	✓
CO3	✓	✓	✓		✓				✓		✓	✓

Course Contents:

Module 1:

Data Analysis libraries: will learn to use Pandas DataFrames, Numpy multi-dimentional arrays, and SciPy libraries to work with a various dataset.

Module 2:

Pandas, an open-source library, and we will use it to load, manipulate, analyze, and visualize various datasets.

Module 3:

Scikit-learn, and we will use some of its machine learning algorithms to build smart models and make predictions, various parameters that can be used to compare various parameters.

Module 4:

Descriptive Statistics, Basic of Grouping, ANOVA, Correlation, Polynomial Regression and Pipelines, R-squared and MSE for In-Sample Evaluation, Prediction and Decision Making.

Module 5:

Grid Search, Model Refinement, Binning, Indicator variables.

Lab Work:

1. Demonstrate knowledge of Data Science and Machine Learning.
2. Apply Data Science process to a real-life scenario.
3. Explore New York City - 311 Complaints and Housing datasets.
4. Analyze and Visualize data using Python.
5. Perform feature engineering exercise using Python.
6. Build and validate predictive machine learning model using Python.
7. Create and share Actionable Insights to real life data problems.

Textbooks:

1. Taming Python by Programming, Jeeva Jose, Khanna Publishing House.
2. Data Visualization with Python and JavaScript, Kyran Dale, Shroff Publisher/O'Reilly Publisher Publication.
3. Data Science Using Python and R by Chantal D. Larose and Daniel T. Larose, Wiley Publication.

References Books:

1. Data Science & Analytics (with Python, R, SPSS Programming), V.K. Jain, Khanna Publishing House.
2. Python for Data Science and Visualization -Beginners to Pro, Udemy.

TITLE OF COURSE: EXPLORATORY DATA ANALYSIS
COURSE CODE: DS806
L-T-P: 3-0-0
CREDITS: 3

Pre-requisite: Calculus, Probability and Statistics for Computer Science

Introduction:

Learn how to use graphical and numerical techniques to begin uncovering the structure of your data. When your dataset is represented as a table or a database, it's difficult to observe much about it beyond its size and the types of variables it contains. In this course, you'll learn how to use graphical and numerical techniques to begin uncovering the structure of your data. Which variables suggest interesting relationships? Which observations are unusual? By the end of the course, you'll be able to answer these questions and more, while generating graphics that are both insightful and beautiful. To learn the essential exploratory techniques for analyzing and visualizing data, and to gain hands-on experience of using software tools for data analytics.

Course Outcomes (CO):
After completion of the course the students will able to
CO1: Describe exploratory data analysis and visualization concepts
CO2: Describe data analysis and visualization models and algorithms
CO3: Describe applicability of different data analysis and visualization models techniques to solve real world problems
CO4: Acquire and pre-process data
CO5: Apply exploratory data analysis to some real data sets and provide interpretations via relevant visualization

Mapping of Course Outcomes (CO) and Program Outcomes (PO):

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓	✓									✓
CO2	✓			✓								✓
CO3	✓		✓									✓
CO4	✓				✓							✓
CO5	✓			✓							✓	✓

Course Contents:

Module 1: Introduction to Exploratory Data Analysis and Visualization, Overview of the exploratory aspect of data analysis, Data acquisition from on-line data sources and preprocessing techniques.

Module 2: Pattern Discovery, Dimensionality Reduction – Linear and Non-Linear Model, Clustering and Classification, Smoothing Scatterplots and Regression.

Module 3: Graphical Visualization, Visualizing Clusters, Visualization Data Distributions, Multivariate Visualization, Graph Data Visualization.

Module 4: Case Studies in Exploratory Data Analysis for Different Application Domains

Text Books

1. W.L. Martinez and A.R. Martinez. Exploratory Data Analysis with MATLAB, Chapman & Hall/CRC, 2011
2. B. Everitt. An Introduction to Applied Multivariate Analysis with R (Use R!), Springer, New York, 2011

References Books

1. W. McKinney. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, O'Reilly, 2012
2. M.A. Russell. Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub and More, O'Reilly, 2013

TITLE OF COURSE: DATA SCIENTIST'S TOOL BOX
COURSE CODE: DS807
L-T-P: 3-0-0
CREDITS: 3

Pre-requisite: Data Science and Data Visualization Basic Knowledge.

Introduction:

In this course you will get an introduction to the main tools and ideas in the data scientist's toolbox. The course gives an overview of the data, questions, and tools that data analysts and data scientists work with. There are two components to this course. The first is a conceptual introduction to the ideas behind turning data into actionable knowledge. The second is a practical introduction to the tools that will be used in the program like version control, markdown, git, GitHub, R, and RStudio.

Course Outcome:

CO1: Set up R, R-Studio, Github and other useful tools

CO2: Understand the data, problems, and tools that data analysts use

CO3: Explain essential study design concepts

CO4: Create a Github repository

Mapping of Course Outcomes (CO) and Program Outcomes (PO):

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓				✓							✓
CO2	✓	✓		✓								✓
CO3	✓		✓									✓
CO4	✓				✓				✓		✓	✓

Module 1: Data Science Fundamentals: Why Automated Videos?, What is Data Science? What is Data? Getting Help, The Data Science Process.

Module 2: R and RStudio, Installing R, Installing R Studio, RStudio Tour, R Packages, Projects in R.

Module 3: Version Control and GitHub, Version Control, Github and Git, Linking Github and R Studio, Projects under Version Control.

Module 4: R Markdown, Scientific Thinking, and Big Data, R Markdown, Types of Data Science Questions, Experimental Design, Big Data.

Text Books:

1. Data Science Mindset, Methodologies, and Misconceptions By Zacharias Voulgaris
2. Domain-Specific Languages in R, Advanced Statistical Programming By Thomas Mailund

Reference Books:

1. R programming for data science by roger d. Peng
2. The analytics lifecycle toolkit a practical guide for an effective analytics capability, by greg nelson

Some Important subjects in Data Science

TITLE OF COURSE: DATA MINING & DATA WARE HOUSING
COURSE CODE:
L-T-P: 3-0-0
CREDITS: 3

Pre-requisite: Knowledge is also assumed of basic concepts in data base management system, and mathematics.

Introduction:
The recent years have generated explosive expansion of digital data stored in computer databases as well as increased pressure on companies to keep competitive advantage. This has put Data Mining (DM) as a key method for extracting meaningful information from the flood of digital data collected by businesses, government, and scientific agencies.

Course Outcomes (CO):
This course will serve to broaden the student's understanding of the issues and latest developments in the area of data mining. To reach this goal, the following objectives need to be met:
CO1: To understand the basic principles, concepts and applications of data warehousing and data mining
CO2: To introduce the task of data mining as an important phase of knowledge recovery process.
CO3: Ability to do Conceptual, Logical, and Physical design of Data Warehouses OLAP applications and OLAP deployment.
CO4: Have a good knowledge of the fundamental concepts that provide the foundation of data mining.
CO5: Design a data warehouse or data mart to present information needed by management in a form that issuable for management client.

Mapping of Course Outcomes (CO) and Program Outcomes (PO):

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓			✓								✓
CO2	✓	✓		✓							✓	✓
CO3	✓	✓	✓		✓				✓		✓	✓
CO4	✓											✓
CO5	✓				✓					✓		✓

Course Contents:
Module 1:
Overview of Data warehousing, Strategic information and the need for Data warehousing, Defining

a Data warehouse, Evolution of Data warehousing, Data warehousing and Business Intelligence.

Module 2:

Concepts of Data warehouse architecture – Definition and architecture in the areas of Data acquisition, Data storage, and Information delivery Distinguishing characteristics – Different objectives and scope, Data content, Complex analysis for faster response, Flexible and Dynamic, Metadata-driven etc Architectural Framework – supporting flow of data, and the Management and Control module Technical architecture – Data acquisition, Data storage, and Information delivery.

Module 3:

Distinction between architecture and infrastructure, understanding of how data warehouse infrastructure supports its architecture Components of physical infrastructure, Hardware and Operating systems for data warehouse, Database Software, Collection of Tools, Data warehouse Appliances – evolution and benefits.

Module 4:

Understanding the importance of Metadata, Metadata types by functional areas – Data acquisition, Data storage, and Information delivery, Business Metadata – overview of content and examples, Technical Metadata – overview of content and examples, Metadata Requirements, Sources of Metadata, Metadata management – challenges, Metadata Repository, Metadata, integration and standards.

Module 5:

Data Design – Design decisions, Basics of Dimensional modeling, E-R modeling versus Dimensional modeling, The STAR schema – illustration, Dimension Table, Fact Table, Factless Fact Table, Data granularity, STAR schema keys – Primary, Surrogate, and Foreign, Advantages of the STAR schema, STAR schema examples. Overview of ETL, Requirements of ETL and steps Data extraction – identification of sources and techniques Data transformation – Basic tasks, Transformation types, Data integration and consolidation, Transformation for dimension attributes, Data loading – Techniques and processes, Data refresh versus update, Procedures for Dimension tables, Fact tables : History and incremental loads ETL Tool options.

Module 6:

Overall concept of Online Analytical Processing (OLAP), OLAP definitions and rules, OLAP characteristics Major features and functions of OLAP – General features, Dimensional analysis, Hypercubes, Drill Down and Roll Up, Slice and Dice, Rotation, Uses and Benefits Familiarity with OLAP models – Overview of variations, MOLAP, ROLAP, HOLAP, DOLAP, Database OLAP, Web OLAP. Web-enabled Data Warehouse – adapting data warehouse for the web Web- based information delivery – Browser technology for data warehouse and Security issues OLAP and Web – Enterprise OLAP, Web-OLAP approaches, and OLAP Engine design.

Module 7:

Overview of Data mining – Definition, Knowledge Discovery Process (Relationships, Patterns, Phases of the process), OLAP versus Data mining, Some aspects of Data mining – Association rules, Outlier analysis, Predictive analytics etc), Concepts of Data mining in a Data warehouse environment, Major Data Mining techniques – Cluster Detection using R Language, Decision Trees, Memory-based Reasoning, Link Analysis, Neural, Networks, Genetic Algorithms etc, Data Mining Applications in industry – Benefits of Data mining using R Language, Discussion on applications in Customer Relationship, Management (CRM), Retail, Telecommunication, Biotechnology, Banking and Finance etc.

Module 8:

Introduction to Big Data Topics, Rise of Big Data, Compare Hadoop vs traditional systems, Limitations and Solutions of existing Data Analytics Architecture, Attributes of Big Data, Types of data, other technologies vs Big Data, Idea of using R Language.

Module-9:
Hadoop Architecture and HDFS Topics - What is Hadoop? Hadoop History, Distributing Processing System, Core Components of Hadoop, HDFS Architecture, Hadoop Master – Slave Architecture, Daemon types - Learn Name node, Data node, Secondary Name node. Hadoop Clusters and the Hadoop Ecosystem: Topics - Hadoop Cluster, Pseudo Distributed mode, Type of clusters, Hadoop Ecosystem, Pig, Hive, Oozie, Flume, SQOOP. Hadoop MapReduce Framework: Topics - Overview of MapReduce Framework, MapReduce Architecture, Learn about Job tracker and Task tracker, Use cases of MapReduce, Anatomy of MapReduce Program.

Textbooks:
1. Data Mining Technology, Third Edition by Arun K Pujari, Universities Press, India
2. Data Warehousing Fundamentals for IT Professionals, Second Edition by Paulraj Ponniah, Wiley India
3. Alex Berson, Stephen J. Smith, "Data Warehousing Data Mining & OLAP", Tata McGrow- Hill References

References Books:
1. Data Warehousing, Data Mining, & OLAP – Second Edition by Alex Berson and Stephen J. Smith, Tata McGraw Hill
2. Data warehouse Toolkit by Ralph Kimball, Wiley India
3. Gajendra Sharma,"Data Mining Data Warehousing and OLAP", S.K.KATARIA & SONS.
4. Sam Anahory, Dennis Murray, "Data Warehousing in the Real World", PEARSON

TITLE OF COURSE: DATA SCIENCE WITH PYTHON
COURSE CODE:
L-T-P: 3-0-2
CREDITS: 4

Pre-requisite: This course is intended for learners who have a basic knowledge of programming in any language (Java, C, C++, Pascal, Fortran, Javascript, PHP, python, etc.).

Introduction:

This course will introduce the learner to the basics of the python programming environment, including fundamental python programming techniques such as lambdas, reading and manipulating csv files, and the numpy library. The course will introduce data manipulation and cleaning techniques using the popular python pandas data science library and introduce the abstraction of the Series and DataFrame as the central data structures for data analysis, along with tutorials on how to use functions such as groupby, merge, and pivot tables effectively. By the end of this course, students will be able to take tabular data, clean it, manipulate it, and run basic inferential statistical analyses.

Course Outcomes (CO):

After Completion of this course student able to understand:

CO1: Basic process of data science

CO2: Python and Jupyter notebooks

CO3: An applied understanding of how to manipulate and analyze unsaturated datasets

CO4: Basic statistical analysis and machine learning methods

CO5: How to effectively visualize results

Mapping of Course Outcomes (CO) and Program Outcomes (PO):

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓				✓							✓
CO2	✓			✓	✓							✓
CO3	✓	✓	✓						✓			✓
CO4	✓	✓	✓		✓							✓
CO5	✓			✓							✓	✓

Module 1: Data Science, Jupyter Notebook System, Python Functions, Python Types and Sequences, Python More on Strings

Module 2: Python Demonstration: Reading and Writing CSV files, Python Dates and Times, Advanced Python Objects, map (), Advanced Python Lambda and List Comprehensions, Advanced Python Demonstration: The Numerical Python Library (NumPy).

Module 3: The Series Data Structure, querying a Series, The Data Frame Data Structure, Data Frame Indexing and Loading, querying a Data Frame, Indexing Data frames, Missing Values.

Module 4: Merging Data frames, Pandas Idioms, Group by, Scales, Pivot Tables, Date Functionality.

Module 5: introduced to a variety of statistical techniques such a distributions, sampling and t-tests, Distributions, More Distributions, Hypothesis Testing in Python.

Text Books

1. Learning Python, 5th Edition by Mark Lutz, O'Reilly Media, 2013. ISBN 978-1-4493-5573-9
2. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython by Wes McKinny, O'Reilly Media, 2012. ISBN 978-1-4493-1979-3

Reference Books:

1. Clean Code: A Handbook of Agile Software Craftsmanship by Robert C. Martin, Prentice Hall, 2008. ISBN 000-0-1323-5088-2
2. The Linux Command Line: A Complete Introduction by William E. Shotts, Jr., No Starch Press, 2012. ISBN 978-1-5932-7389-7

TITLE OF COURSE: STATICS FOR DATA ANALYSIS
COURSE CODE:
L-T-P: 3-0-0
CREDITS: 3

Pre-requisite: This course requires that you are familiar with high-school level linear algebra, and calculus. Knowledge of probability theory, statistics, and programming is desirable.

Introduction:

This course will expose you to the data analytics practices executed in the business world. We will

explore such key areas as the analytical process, how data is created, stored, accessed, and how the organization works with data and creates the environment in which analytics can flourish.

Course Outcomes (CO):

After Completion of this course student able to understand:

CO1: Strong foundation in all the areas that support analytics

CO2: Basis for going deeper into advanced investigative and computational methods

CO3: Use a simple but powerful language called SQL to extract analytical data sets

CO4: Machine learning utilization in Data Analysis.

Mapping of Course Outcomes (CO) and Program Outcomes (PO):

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓	✓		✓							✓
CO2	✓			✓								✓
CO3	✓				✓						✓	✓
CO4	✓	✓	✓		✓						✓	✓

Module 1: Introduction to Data & Analysis in Real World, Thinking about Analytical Problems, Conceptual Business Models, The Information-Action Value Chain, Real World Events and Characteristics, Data Capture by Source Systems.

Module 2: Introduction - Analytical Technologies, Data Storage and Databases, Big Data & the Cloud, Virtualization, Federation, and In-Memory Computing, the Relational Database, Data Tools Landscape, the Tools of the Data Analyst.

Module 3: 1. Introduction to SQL, Aggregating and Sorting Data in SQL, Extracting Data from Multiple Tables, Stacking Data with UNION Command, Extending SQL Queries Using Operators, Using SQL Subqueries.

Module 4: Introduction to Real World Analytical Orgs, Analytical Organizations – Roles, Analytical Organizations – Structures, Data Governance, Data Privacy, Data Quality.

Module 5: Descriptive Statistics, Inferential Statistics through hypothesis tests Permutation & Randomization Test, Regression & ANOVA, and Machine Learning: Introduction and Concepts, Supervised and Unsupervised Learning Technique.

Textbooks:

1. Montgomery, Douglas C., and George C. Runger., Applied statistics and probability for engineers. John Wiley & Sons, 2010

TITLE OF COURSE: STATISTICAL METHODS FOR DECISION MAKING

COURSE CODE:

L-T-P: 3-0-0

CREDITS: 3

Prerequisites: The students are expected to have knowledge of basic mathematics at the plus two level

Introduction:

Introduction to probabilistic and statistical techniques for decision making, including inferential statistics, hypothesis tests, analysis of variance, regression analysis, and statistical quality control. Using computer software and data in statistical analysis. Emphasis on formal modeling and the use of data for managerial decision making and problem solving.

Course Outcome:

CO1: Understand and appreciate the most widely used tools of business statistics which form the basis for rational and sound business decisions

CO2: Focus on problem recognition and test hypothesis/model in the context of managerial decision-making.

CO3: Develop skills in analysis and interpretation of data

CO4: Handle challenging problems using appropriate analysis tool

Mapping of Course Outcomes (CO) and Program Outcomes (PO):

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓	✓	✓		✓							✓
CO2	✓			✓								✓
CO3	✓	✓	✓									✓
CO4	✓	✓	✓		✓						✓	✓

Course Content:

Module 1: Introduction to Descriptive Statistics, Descriptive and Inferential Statistics, Types of measurements, Descriptive Statistics (Using Graphs), Descriptive Statistics (Using Numbers), Measures of location, variability, and relative standing

Module 2: probability and sampling applications, and rules, conditional probability, discrete distributions (binomial, poisson, hypergeometric, geometric), continuous distribution, normal and standard normal distribution, sampling distributions, sampling distribution parameters, central limit theorem, applying sampling distribution theory

Module 3: Confidence Intervals & Hypothesis Testing: Estimation and Hypothesis Testing, point estimators, interval estimation, t-distribution, Hypothesis testing, p-values, Estimation of Population Proportion.

Module 4: Statistical Process/Quality Control: Common Causes And Special Causes Of Variation, X-Bar Chart, R Chart, P Chart, Comparing Two Population Means (Confidence Intervals)

Module 5: Regression Analysis: Simple And Multiple Linear Regression, Relationship Between Two(Simple), Three Or More(Multiple) Variables, Model Estimation, Model Inference

Textbooks:

1. Anderson, Sweeney, and Williams, Statistics for Business and Economics, Seventh Edition, West Publishing Co., available at Hammes Bookstore.
2. Statistical Techniques for Data Analysis, By John K. Taylor, Cheryl Cihon

Reference Books:

1. An Introduction to Statistical Learning: with Applications in R By Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani
2. Statistical Models for Data Analysis, edited by Paolo Giudici, Salvatore Ingrassia, Maurizio Vichi

TITLE OF COURSE: DATA VISUALIZATION
COURSE CODE:
L-T-P: 3-0-0
CREDITS: 3

Prerequisites

Students should have taken a course in algorithms and data structures. While the computer graphics is not required, it is useful background. Familiarity with Web technologies and JavaScript is also useful.

Introduction:

Visualization is increasingly important in this era where the use of data is growing in many different fields. Data visualization techniques allow people to use their perception to better understand this data. The goal of this course is to introduce students to data visualization including both the principles and techniques. Students will learn the value of visualization, specific techniques in information visualization and scientific visualization, and how understand how to best leverage visualization methods.

Course Outcome:

CO1: Students will be able to prepare data for visualization.
CO2: Students will be able to design visualizations.
CO3: Students will be able to use web technology to create visualizations.
CO4: Understand the type of data impacts the type of visualization.

Mapping of Course Outcomes (CO) and Program Outcomes (PO):

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	✓											✓
CO2	✓		✓									✓
CO3	✓	✓	✓								✓	✓
CO4	✓			✓								✓

Course Content:

Module 1: The Computer and the Human: Overview of Visualization, 2-D Graphics, SVG-example, 2-D Drawing, 3-D Graphics, Photorealism, Non-Photorealism, The Human, Memory, Reasoning, The Human Retina, Perceiving Two Dimensions, Perceiving Perspective.

Module 2: Visualization of Numerical Data Introduction, Data, Mapping, Charts, Glyphs, Parallel Coordinates, Stacked Graphs ,Tufte's Design Rules, Using Color.

Module 3: Visualization of Non-Numerical Data Introduction, Graphs and Networks, Embedding Planar Graphs, Graph Visualization, Tree Maps, Principal Component Analysis, Multidimensional Scaling, Packing.

Module 4: Introduction to Visualization Systems, The Information Visualization Mantra, Database Visualization Part, Visualization System Design

Textbooks:

1. Data Visualization: A Practical Introduction By Kieran Healy
2. Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures By Claus O. Wilke

Reference Books:

1. Data Visualization: A Handbook for Data Driven Design, By Andy Kirk
2. Effective Data Visualization: The Right Chart for the Right Data, Book by Stephanie Evergreen