Course Title: Data Science Fundamentals			
Course code:	**No. of credits:** 2	**L-T-P: 20-10-0**	**Learning hours: 30**
			L: Lectures**; T:** Tutorials**; P:** Practicals
Pre-requisite course code and title (if any): None			
Department: Natural and Applied Sciences			
Course coordinator:		**Course instructor:**	
Contact details:			
Course type: Core		**Course offered in:** Semester 1	

Course Description

The course introduces the multidisciplinary nature of data science and develops a conceptual understanding on what constitutes data science. The course gives premise of not only the composites of *data analysis with data processing methods and domain expertise*, cultivates data science but also deals with data transformation leading to understandable and actionable knowledge for informed decision making. The course develops an understanding to overcome data complexity and the limitations of classical statistics through data science. It develops a wider understanding on how data science is different from statistics, machine learning, computer science, data analytics despite all of them being interrelated. The course introduces the students on how Data Science develops the ability to overcome data complexity and the limitations of classical statistics and machine learning techniques where a data scientist deals simultaneously with heterogeneous data sources coping with non-independencies, non-normalities, and hypothesis on variable's distributions. The students get a complete understanding of Data Sciences broader scope and arena and its associated elements.

Course objectives

- To develop broader lens of data science and its fundamentals.
- To understand how data science is different from several data based conventional disciplines.
- To understand the basic components of Data Science what essentials cultivate a data scientist with introductory functions of data science and data scientists

Course content

Module	Topic	L	T	P
1	**Data Science Concepts**			
	The objective of this module is to gain familiarity with the phrases, terms and jargons used in Data Science. The module introduces data science concepts, its development and future scope. The module further strengthens the on what data specialization makes a data scientist different from data specialists in classical data based disciplines. What is data science, what is included in data science in its broader context, what are current challenges and its futuristic scope of data science. The thematics in this module are: What is data science, what is included in data science in its broader context, what are current challenges and its futuristic scope of data science. Data Science: Definition, benefits and uses, issues and challenges, big data and future of data science; Data scientist's role and responsibility.	5		
2	**Data Science Process Life Cycle**			
	The purpose of this module is to showcase the centrality of data science process life cycle, discovering the answers for basic questions including requirements, priorities of data science projects, collecting structured or	5		

	unstructured data, processing and fine tuning the raw data, factors influencing data life cycle and ultimately leading to model development and deployment. Various expertise and components required to build a data scientist. It also discusses attributes and qualities of data and data scientists. This module addresses the concepts of: Data Science Process: identify research goals, data retrieving and preparation, statistical inference, exploratory data analysis and model building and deployment, Datafication, Data scientist, Data science profile, Meta-definition, Data warehousing.			
3	**Data modelling and Data mining**			
	This module connects various themes of 'data modelling'. The previous two modules are building blocks for this module. The goal is to illustrate the types of data used and stored within the system, the relationships among these data types, the ways the data can be grouped and organized and its formats and attributes. data modelling employs standardized syllogistic figures and techniques. The module also introduces the concept of sorting the data to find a patterns within the datasets leading to a logical conclusion. The following topics are introduced in this module: Modelling of data and pattern recognition from data mining; case study involving data modelling and data mining.	5	5	
4	**Data Munging**			
	After making the students familiar with the theoretic basis of the components of data science in the previous modules, this module touches upon the transforming the erroneous data or unusable data into useful and usable data. The module discusses the stages of data munging from data discovery, data structuring, data cleansing, data enrichment to data validation along with benefits and challenges with data munging. The module also introduces the concepts of data collection via internet, the crowdsourcing. Following topics are addressed in this module: Properties, Languages; Data collections and cleaning, crowdsourcing and its application.	5	5	
	Total	20	10	0

Evaluation criteria

- Minor Test 1: Written test [at the end of teaching of modules 1 and 2] -- 25%
- Minor Test 2: Written test [at the end of teaching of module 3] -- 25%
- Major Test: Written test [at the end of the semester, full syllabus] -- 50%

Learning outcomes

By the end of the course, students will:

- command a critical understanding of the key concepts of data science, attributes of data scientist and understand the components of data science and data science life cycle. [Module 1 and 2; Minor Test 1]
- develop knowhow of data modelling and its requirement. Understanding the foundation of pattern recognition through data mining [Module3; Minor Test 2]
- understand the conceptual, theoretical, methodical, requirements of data collection, cleaning and analysis of data collected through multiple platforms [Module 1, 2, 3, 4; Major Test]

Pedagogical approach

- The course critically evaluates the concepts of data sciences and develops discussion in classroom through lectures, case studies and tutorials.
- The course doesn't dwell into the hardware or software based analysis rather analyzes case studies to develop understanding of data science components

Reading Resources (* = compulsory readings)

*Cielen, D., Meysman, D.B.A., Ali, M. (2016). *Introducing data science: big data, machine learning, and more, using Python tools*. Simon and Schuster.

Grus, J. (2019). *Data science from scratch: first principles with python*. O'Reilly Media.

O'Neil, C., Schutt, R. (2013). *Doing data science: Straight talk from the frontline*. O'Reilly Media, Inc.

Peng, R.D. (2016). *R programming for data science* (pp. 86-181). Victoria, BC, Canada: Leanpub.

*Pierson, L. (2021). *Data science for dummies*. John Wiley & Son.

Skiena, S.S. (2017). *The data science design manual*. Springer.

VanderPlas, J. (2016). *Python data science handbook: Essential tools for working with data*. O'Reilly Media, Inc.

Student Responsibilities

The students are required to come prepared with readings that would be given in the class. The students are required to participate in the discussion.

Course Reviewers

1. **Dr Om Prakash Rishi**, Associate Professor, Department of Computer Science & Informatics, University of Kota, Rajasthan
2. **Dr Varun Gupta**, Professor (CSE) & Head (AS), Chandigarh College of Engineering & Technology, Chandigarh