CSEN2161	INTRODUCTION TO DATA SCIENCE	L	T	P	S	J	C
		2	1	0	0	0	3
Pre-requisite	Programming, DBMS						
Co-requisite	**None**						
Preferable exposure	**None**						

Course Description:

This course is designed to provide an introduction into the field of Data Science. Students will develop skills in appropriate technology and basic statistical methods by completing hands-on projects focused on real-world data and addresses the social consequences of data analysis and application.

Course Educational Objectives:

The Course Enables Students to

- Provide a strong foundation for data science and application areas related to it.
- Understand the underlying core concepts and emerging technologies in data science.
- Learn the process of working with data on large scale.
- Explore the concepts of Data Processing.
- Learn basic concepts of Machine Learning.
- Prepare students for advanced courses in Data Science.

UNIT 1 Data Evolution 9 hours, P – 2 hours

Data Evolution: Data to Data Science – Understanding data: Introduction – Type of Data, Data Evolution – Data Sources. Preparing and gathering data and knowledge - Philosophies of data science - data all around us: the virtual wilderness - Data wrangling: from capture to domestication - Data science in a big data world - Benefits and uses of data science and big data - facets of data.

UNIT 2 Digital Data-An Imprint 9 hours, P - 2 hours

Type of data analytics (Descriptive, diagnostic, perspective, predictive, Prescriptive.) Exploratory Data Analysis (EDA), EDA-Quantitative Technique, EDA - Graphical Technique. Data Types for Plotting, Data Types and Plotting, Simple Line Plots, Simple Scatter Plots, Visualizing Errors, Density and Contour Plots, Histograms,
Binnings, and Density, Customizing Plot Legends, Customizing Color bars, Multiple Subplots, Text and Annotation, Customizing Ticks.

UNIT 3 Descriptive statistics 9 hours, P - 2 hours

Measures of Central Tendency – Measures of Variation – Quartiles and Percentiles – Moments – Skewness and Kurtosis. Exploratory Data Analytics Descriptive Statistics – Mean, Standard

Deviation, Skewness and Kurtosis – Box Plots – Pivot Table – Heat Map – Correlation Statistics – ANOVA, Random variable, Variance, covariance, and correlation-Linear transformations of random variables, Regression.

UNIT 4 Data Science tools 9 hours, P - 2 hours

Overview and Demonstration of Open source tools such as R, Octave, Scilab. Python libraries: SciPy and sci- kitLearn, PyBrain, Pylearn2; Weka.

UNIT 5 Ethics and Data Science LTP 902 9 hours, P - 2 hours

Data Ownership, The Five Cs, Implementing the Five Cs, Ethics and Security Training, Developing Guiding Principles, Building Ethics into a Data-Driven Culture, Regulation, Building Our Future, Case Study.

TextBooks:

1. Introducing Data Science, Davy Cielen, Arno D. B. Meysman and Mohamed Ali, Manning Publications,2016.

2. Think Like a Data Scientist, Brian Godsey, Manning Publications, 2017.

3. Ethics and Data Science, Mike Loukides, Hilary Mason and D J Patil, O'Reilly, 1st edition, 2018.

Reference Book(s):

1. Data Science from Scratch: First Principles with Python, Joel Grus, O'Reilly, 1st edition, 2015.
2. Doing Data Science, Straight Talk from the Frontline, Cathy O'Neil, Rachel Schutt,2 O' Reilly, 1st edition, 2013.
3. Mining of Massive Datasets, Jure Leskovec, Anand Rajaraman, Jeffrey David Ullman, Cambridge University Press, 2nd edition, 2014.
4. An Introduction to Statistical Learning: with Applications in R, Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Springer, 1st edition, 2013.
5. Data science Handbook – Field cady- wiley,2017
6. Statistical inference for data science-Brian Caffo,2016
7. Introducing Data science by Davy cielen, Arno D.B.Meysmen, Mohamed Ali.2020
8. 1.https://www.coursera.org/learn/what-is-datascience?specialization=introduction-data-science
9. 2.https://www.coursera.org/learn/open-source-tools-for-data-ience?specialization=introduction- data-science
10. 3.https://www.coursera.org/learn/data-science-methodology?specialization=introduction-data- science

Course Outcomes:

After successful completion of the course the student will be able to:

1. Understand the fundamental concepts of data science.
2. Evaluate the data analysis techniques for applications handling large data and Demonstrate the data science process.
3. Understand concept of machine learning used in the data science process.
4. Visualize and present the inference using various tools.
5. Learn to think through the ethics surrounding privacy, data sharing.

CO-PO Mapping:

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0
CO2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1
CO3	1	1	1	1	0	0	0	0	0	0	1	1	0	0	0
CO4	1	1	1	1	0	0	0	1	1	1	1	0	1	0	0
CO5	1	1	0	0	0	0	0	0	1	1	1	1	0	1	1

Note: 1 - Low Correlation 2 - Medium Correlation 3 - High Correlation

APPROVED IN:
BOS : 06-09-2021 **ACADEMIC COUNCIL: 01-04-2022**

SDG No. & Statement:

SDG Justification: