Probability Statistics for Data Science	Code & No:	CS 472
	Credits:	3 (2-2-0)
	Pre-requisite:	STAT 102
	Co-requisite:	
	Level:	9 /10

Course Description: This course is aimed to provide the probability and statistics for data scientist with the application of programming languages. Topics covered include Exploratory Data Analysis, Data and Sampling Distributions, Probability Theory, Random Variables, Stochastic Processes, Statistical Experiments and Significance Testing, Regression and Prediction, Classification, and Clustering. Statistical Machine Learning, Unsupervised Learning.

Course Aims:

1) To provide the conceptual knowledge data science.
2) To provide the important and useful from the data science perspective.
3) To develop the skills of applying the techniques & tools of statistical practice and empirical research.
4) To provide the knowledge and applications of software package (R- Language).

Course Learning Outcomes (CLOs):

After completing this course, the students will be able

1. To explore and analyze the data.
2. To model projects (whether in data science or in research) with the statistical tool among predictors, and between predictors and a target variable.
3. To apply the sampling techniques from the Big Data projects.
4. To design an experiment for test of the hypothesis.
5. To take the automated decision faced with a problem.
6. To apply the basic statistical techniques on data, using statistical software package (R).

No.	Topics	Weeks	Teaching hours
1	Probability review	1	2

2	Random variables (Review)	1	2
3	Non-parametric inference, Parametric inference	1	2
4	Exploratory Data Analysis	3	6
5	Data and Sampling Distributions	2	4
6	Statistical Experiments and Significance Testing	3	6
7	Regression and Prediction	2	4
8	Classification, Naive Bayes, Discriminant Analysis	2	4
	Total	**14**	**28**

Textbook:

- Peter Bruce and Andrew Bruce, Practical Statistics for Data Scientists, O'Reilly Media, 2017, 978-1-491-95296-2

Essential References:

- Colin O. Wu Xin Tian, Nonparametric Models for Longitudinal Data with Implementation in R, CRC Press, 978-1-4665-1600-7
- Hongshik Ahn, Probability and Statistics for Science and Engineering with Examples in R [2nd ed.] Cognella, 2018 978-1-5165-3111-0
- Jay L. Devore, Probability and Statistics for Engineering and the Sciences. 9th Edition, Cengage Learning. ISBN: 1305251806.