MCA- E312 Data Science				
	L	T	P	C
	4	0	0	4

Course objective:

1. To introduce students to the basic concepts of Data Science.
2. To become familiar with data preprocessing methods.
3. To learn and demonstrate the application of statistical data analysis techniques in business decision making.
4. To gain knowledge on strategies for data transformation.
5. To learn to implement Data Science related applications in Python.

Course outcomes:

1. Learn methods in integrating and interpreting the data sets and improving effectiveness, efficiency and quality for data analysis.
2. Gain practical, hands-on experience with principles of Data Science to the analysis of business problems

Introduction to Data Science: What is data science? Basic terminology, Why data science? The data science Venn diagram. Domain knowledge, Data Science case studies

Data Representation: Data Objects and Attribute Types: Nominal, Binary, Ordinal, Numeric, Discrete and Continuous, Types of data: Record, Temporal, Spatial Temporal, Graph, Unstructured and Semi structured data, Basic Statistical Descriptions of Data.

Introduction to Data Analysis: Probability and Random Variables, Correlation, Regression.

Exploratory Data Analysis: What is Exploratory Data Analysis? EDA Techniques, EDA Classification, Univariate Non-graphical EDA, Univariate Graphical EDA, Bivariate Non-graphical EDA, Bivariate Graphical EDA, Philosophy of EDA, The Data Science Process

Data Preprocessing: Meaning, Data Quality, Major Tasks in Data Preprocessing, *Data Cleaning* - Missing Values, Noisy Data, Data Cleaning as a Process, *Data Integration* - Entity Identification Problem, Redundancy and Correlation Analysis, Tuple Duplication, Data Value Conflict Detection and Resolution, *Data Reduction Strategies* - Overview, Wavelet Transforms, Principal Components Analysis, Attribute Subset Selection, Regression and Log-Linear Models: Parametric Data Reduction, Histograms, Clustering, Sampling, Data Cube Aggregation, *Data Transformation Strategies* : Overview, Normalization, *Data Discretization* and Binning, Histogram Analysis, Cluster, Decision Tree, and Correlation Analyses, Concept Hierarchy Generation for Nominal Data

Introduction to Probability: Common terminology of Probability, Events, Union, Intersection, Types of Events, Rules of Probability, Types of Probability: Marginal, Joint, Conditional, Bayesian Inference: Bayes Theorem, Random Variable, General Product Rule

Introduction to Statistics: Standardization (Z-score), Data Distribution (Poisson Distribution and Normal Distribution), Sampling Theorem, Hypothesis Testing, ANOVA test, Chi-square test, p-values

Machine Learning: What is Machine Learning? AI vs ML vs DL, Significance of Machine Learning, Applications of Machine Learning in data science, Supervised, Unsupervised and Reinforcement Learning, Classification and Regression, Cross Validation.

HEAD
Department of Computer Science
Gurukul Kangri Vishwavidyalaya
Haridwar (UK) - 249404

Data Visualization: Traditional Visualization, Multivariate Data Visualization, Principles of Perception, Color, Design, and Evaluation, Text Data Visualization, Network Data Visualization, Temporal Data Visualization and visualization Case Studies.
Recommended Books: 1. Cathy O'Neil and Rachel Schutt. Doing Data Science, Straight Talk from The Frontline, O'Reilly 2. Sinan Ozdemir, Principles of Data Science, Packt Publishing 3. Han, J., Kamber, M. and Pei, J., Data Mining Concepts and Techniques, Morgan Kaufmann 4. Hastie, T., Tibshirani, R. and Friedman, J., The Elements of Statistical Learning, Springer 5. Simon, P., The Visual Organization: Data Visualization, Big Data, and the Quest for Better Decisions, John Wiley & Sons

HEAD
Department of Computer Science
Gurukul Kangri Vishwavidyalaya
Haridwar (UK) - 249404