National Institute of Technology Meghalaya

An Institute of National Importance

CURRICULUM

Programme	**Bachelor of Technology in Computer Science and Engineering**	Year of Regulation	**2019-20**
Department	**Computer Science and Engineering**	Semester	**VI**

Course Code	Course Name	Credit Structure				Marks Distribution			
		L	T	P	C	INT	MID	END	Total
CS 324	**Data Analysis and Visualization**	**3**	**0**	**0**	**3**	**50**	**50**	**100**	**200**

Course Objectives	Course Outcomes	
To understand the need of data analysis and visualization techniques	CO1	**Able to analyse the different data representation and data pre-processing techniques**
To learn the different types of data analysis and visualization tools and techniques	CO2	**Able to assess and compare different data analysis and visualization techniques**
To apply the concept of data analysis and visualization to real life problems	CO3	**Able to implement data analysis and visualization based solutions for real life problems**

No.	COs	Mapping with Program Outcomes (POs)												Mapping with PSOs		
		PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
1	CO1	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**2**	**1**	**0**
2	CO2	**1**	**1**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**2**	**1**	**0**
3	CO3	**1**	**1**	**2**	**1**	**1**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**2**	**1**	**1**

SYLLABUS

No.	Content	Hours	COs
I	**Introduction** **Concepts and Need of data analysis and visualization in the era of data abundance** **Data Representation - Nominal, Binary, Ordinal, Numeric, Discrete and Continuous, Types of data - Record, Temporal, Spatial Temporal, Graph, Unstructured and Semi structured data**	**04**	**CO1**
II	**Data Statistical Properties and Data Pre-Processing** **Basic Statistical Descriptions of Data (mean, median, standard deviation, maximum, minimum, tests of significance), Probability and Random Variables, introduction to estimation theory , Correlation, Regression** **Data pre-processing- Attribute transformation, Sampling, Dimensionality reduction, Feature subset selection, Distance and Similarity calculation**	**08**	**CO1**
III	**Data Analysis Techniques** **Supervised and unsupervised learning, gradient descent, over fitting, regularization** **Unsupervised techniques - K-means, Gaussian mixture models and expectation-maximization, evaluation of clustering** **Supervised techniques - K-nearest neighbor, naive Bayes, logistic regression and Regularization, support vector machine, artificial neural networks (ANNs)**	**12**	**CO2**
IV	**Visualization and Applications** **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** **Data visualization in Python and R**	**12**	**CO2 & CO3**
	Total Hours	**36**	

Essential Readings

1. Han, Jiawei, Jian Pei, and Micheline Kamber. "*Data mining: concepts and techniques"*. Elsevier, 3rd edition, 2011
2. Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. "*The elements of statistical learning: data mining, inference, and prediction"*. Springer Science & Business Media, 2nd edition, 2009.
3. Embarak, Ossama. "*Data Analysis and Visualization Using Python: Analyze Data to Create Visualizations for BI Systems"*. Apress, 1st edition, 2018.

Supplementary Readings

1. Bishop, Christopher M. "*Pattern recognition and machine learning"*. springer, 1st edition, 2006.
2. Tan, Pang-Ning, Michael Steinbach, and Vipin Kumar. "*Introduction to data mining"*. Pearson Education India, 2nd edition, 2016.
3. Knaflic, Cole Nussbaumer. "*Storytelling with data: A data visualization guide for business professionals"*. John Wiley & Sons,1st edition, 2015.