	National Institute of Technology Meghalaya An Institute of National Importance	CURRICULUM

Programme	**Bachelor of Technology in Computer Science and Engineering**	Academic Year of Regulation	**2018-19**
Department	**Computer Science and Engineering**	Semester	**VII**

Course Code	Course Name	Credit Structure				Marks Distribution			
		L	T	P	C	INT	MID	END	Total
CS 471	**Data Analytics using Python**	**2**	**0**	**0**	**2**	**50**	**50**	**100**	**200**

Course Objectives		Course Outcomes		
	This course introduces understand the importance of data analytics		CO1	**Able to analyse the different data representation and data pre-processing techniques**
	This course explains the different types of data analytics techniques		CO2	**Able to assess and compare different data analytics techniques**
	This course familiarizes the data analytics techniques using python programming for publically available datasets		CO3	**Able to determine data analytics techniques using python libraries for real life applications**

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**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**2**	**1**	**-**
2	CO2	**1**	**1**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**2**	**1**	**-**
3	CO3	**1**	**1**	**1**	**1**	**1**	**1**	**-**	**-**	**-**	**-**	**-**	**-**	**2**	**1**	**-**

SYLLABUS

No.	Content	Hours	COs
I	**Introduction: Data analytics and its importance, introduction of python programming and installing Python, understanding operators, variables, data types, conditional statements , looping constructs , functions, lists and dictionaries in Python, Importing and exporting data in python**	**06**	**CO1**
II	**Data pre-processing : Handling missing values, data transformation, normalization, discretization** **Data Analysis Techniques: Supervised and unsupervised learning, Unsupervised techniques - K-means, Hierarchical clustering, Density based clustering, evaluation of clustering, Supervised techniques - Linear Regression, Logistic Regression, K-nearest neighbor, naive Bayes, support vector machine, artificial neural networks (ANNs)**	**10**	**CO2**
III	**Learn and installing Jupyter Notebook, Understanding the concept of Standard Libraries in python : Numpy, Pandas, sci-kit learn, MatplotLib,** **Case studies: Predicting loan defaulters, Customer segmentation, Time series forecasting etc.**	**08**	**CO3**
Total Hours		**24**	

Essential Readings

1. A.C. Müller and S. Guido. "*Introduction to machine learning with Python: a guide for data scientists*". O'Reilly Media, Inc. 1st edition, 2016
2. D. Beazley and B.K. Jones. " *Python Cookbook: Recipes for Mastering Python".* O'Reilly Media, Inc. 2nd edition, 2013
3. J. Han, J. Pei, and M. Kamber. "*Data mining: concepts and techniques".* Elsevier, 3rd edition, 2011

Supplementary Readings

1. W. McKinney. "*Python for data analysis: Data wrangling with Pandas. NumPy, and IPython".* O'Reilly Media, Inc. 2nd edition, 2017
2. P.N.Tan, M. Steinbach, A. Karpatne, and Vipin Kumar. "*Introduction to data mining"*. Pearson Education India, 2nd edition, 2016.
3. S. Raschka and V. Mirjalili. "*Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow"*. Packt Publishing Ltd. 2nd edition, 2019.