Course Code	Name of the Course			
216H03C501	**Data Analysis**			
Teaching Scheme (Hrs./Week)	**TH**	**P**	**TUT**	**Total**
	03	**--**	**--**	**03**
Credits Assigned	**03**	**--**	**--**	**03**
Evaluation Scheme	**Marks**			
	LAB/TUT CA	**CA (TH)**	**ESE**	**Total**
		IA | **ISE**		
	--	**20** | **30**	**50**	**100**

Course pre-requisites:
- Concepts of DBMS, Probability and statistics, knowledge of programming language (C/C++/Java/ Python).

Course Objectives:

Introduction to the fundamental concepts of Data Analytics , analyse real world case studies by applying mining algorithms and visualization for decision-making in Geospatial, social media ,healthcare and text mining business applications

Course Outcomes (CO):

At the end of successful completion of the course the student will be able to

CO1	Understand basic concepts of data analytics to solve real world case studies
CO2	Apply the data analysis on geospatial system
CO3	Perform the graph data analysis
CO4	Perform Time series Analysis
CO5	Apply the text data analytics in the field of Health care

Module No.	Unit No.	Details	Hrs.	CO
1		**Introduction to Data Analysis**	05	CO1
	1.1	Introduction to Data Analytics, Different types of data analytics: Descriptive analytics, Diagnostics Analytics, Predictive Analysis, Prescriptive Analysis		
		# Self-Learning: LinkedIn Analysis, Netflix Analysis, Cricket and FIFA Analysis.		
2		**Spatial Data Analysis**	10	CO2 CO1
	2.1	Contents and characteristics of Spatial data, Spatial data formats, spatial data bases		
	2.2	Introduction, Definition of GIS, Evolution of GIS , components of GIS Spatial Association rule mining, spatial hierarchal clustering, Set based decision and knowledge recovery		
	2.3	Case study: GIS application for Spatial data mining using open source software tools		
		# Self-learning: QGIS, Hadoop, GeoSpark R PostgreSQL, PostGIS, Python		
3		**Graph Analysis**	10	CO3
	3.1	Introduction to the Social Network, Clustering of Social-Network Graphs, Direct Discovery of Communities		
	3.2	Partitioning of Graphs, Finding Overlapping Communities, Simrank, Counting Triangles, Neighborhood Properties of Graphs		
	3.3	# Self-learning: GraphX tools of Apache.		
4		**Time series Analysis for prediction and forecasting**	10	CO4
	4.1	Introduction, Finding and Wrangling Time Series Data, Exploratory Data Analysis for Time Series, Simulating Time Series Data, Storing Temporal Data,		
	4.2	Statistical Models for Time Series, State Space Models for Time Series, forecasting methods, Testing for randomness, Regression based trend model :AR,MA,ARIMA, random walk model, moving average forecast, exponential smoothing forecast, seasonal models,		
5		**Data Analysis in Health Care Application**	10	CO5
	5.1	Introduction, Components of HER, Benefits of EHR-		
	5.2	Natural Language Processing and Analysis Of Clinical Text : Introduction , report analyser, text analyser, Core NLP Components Morphological Analysis , Lexical Analysis , Syntactic Analysis ,Semantic Analysis , Data Encoding .		
	5.3	Mining Information from Clinical Text: Rule-Based Approaches. Pattern-Based Algorithms ,Machine Learning Algorithms		

Module No.	Unit No.	Details	Hrs.	CO
		Self-learning: Introduction to Social media Analysis for healthcare,		
		Total	**45**	

 # Students should prepare all Self Learning topics on their own. Self-learning topics will enable students to gain extended knowledge of the topic. Assessment of these topics may be included in IA and Laboratory Experiments.

Reference Books:

Sr. No.	Name/s of Author/s	Title of Book	Name of Publisher with country	Edition and Year of Publication
1.	Michael J. de Smith, Michael F. Goodchild and Paul A. Longley	Geospatial Analysis: A Comprehensive Guide to Principles, Techniques, and Software Tools,	Wiley, Second Edition	2019
2.	Anil Maheshwari	*Data Analysis*	Mc Graw Hill	2017
3.	James, G., Witten, D., Hastie, T., Tibshirani, R.	*An introductionto statistical learning with applications in R*	Springer	2013
4.	Chandan K. Reddy and Charu C Aggarwal	*Healthcare data Analysis*	Taylor & Francis	2015
5	U. Dinesh Kumar	*Business Analysis*	Wiley	2017
6	Li, Deren., Wang, Shuliang., Li,Deyi	Spatial Data Mining: Theory and Application.	Spatial Data Mining: Theory and Application.	2016
7.	Albright and Winston	Bussiness Analysis	Cengage Publication	5th edition, 2015
8.	Aileen Nielsen	Practical Time Series Analysis	O'Reilly Media, Inc.	1st edition October 2019

*In addition to printed books, faculty can suggest (authentic) urls or e-books, e-contents etc.

Course Code	Name of the Course			
216H03L501	**Data Analysis Laboratory**			

Teaching Scheme (Hrs./Week)	TH	P	TUT	Total
	-	02	-	02
Credits Assigned	-	01	--	01

Evaluation Scheme	Marks				
	LAB/TUT CA	CA (TH)		ESE	Total
		IA	ISE		
	50	–	–	–	50

Course Objectives
- Comprehend methodologies of data analysis

Course Outcomes

At the end of the course students will be able to

CO 1	Apply and Implement spatial data analysis
CO 2	Apply and Implement graph data analysis
CO 3	Apply and Implement time seriesl data analysis
CO 4	Apply and Implement data analysis for inferring the results

Term-Work:

Term work will consist of minimum 8 experiments/ tutorials covering entire syllabus of the course 'Data Analysis Laboratory. Students will be graded based on continuous assessment of their term work.