

Course Code	PIT21G301J	Course Name	Social Media and Text Analytics	Course Category	G	Generic Elective Courses	L	T	P	C
							3	0	2	4

Pre-requisite Courses	Nil	Co-requisite Courses	Nil	Progressive Courses	Nil
Course Offering Department	Computer Science	Data Book / Codes/Standards	Nil		

Course Learning Rationale (CLR):		The purpose of learning this course is to,	Learning			Program Learning Outcomes (PLO)														
CLR-1 :	Familiarize the software lifecycle models and software development process		1	2	3	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
CLR-2 :	Understand the various techniques for requirements, planning and managing a technology project		Level of Thinking (Bloom)	Expected Proficiency (%)	Expected Attainment (%)	Disciplinary Knowledge	Critical Thinking	Problem Solving	Analytical Reasoning	Research Skills	Team Work	Scientific Reasoning	Reflective Thinking	Self-Directed Learning	Multicultural Competence	Ethical Reasoning	Community Engagement	ICT Skills	Leadership Skills	Life Long Learning
CLR-3 :	Examine basic methodologies for software design, development, testing, closure and implementation																			
CLR-4 :	Understand manage users expectations and the software development team																			
CLR-5 :	Acquire the latest industry knowledge, tools and comply to the latest global standards for project management																			
Course Learning Outcomes (CLO):		At the end of this course, learners will be able to:																		
CLO-1 :	Identify the process of life cycle model and process project		3	80	70	L	H	-	H	L	-	-	-	L	L	-	H	-	-	-
CLO-2 :	Analyze and specify software requirements through a productive working Relationship with project stakeholders		3	85	75	M	H	L	M	L	-	-	-	M	L	-	H	-	-	-
CLO-3 :	Design the system based on Functional Oriented and Object Oriented Approach for Software Design.		3	75	70	M	H	M	H	L	-	-	-	M	L	-	H	-	-	-
CLO-4 :	Develop the correct and robust code for the software products		3	85	80	M	H	M	H	L	-	-	-	M	L	-	H	-	-	-
CLO-5 :	Perform by applying the test plan and various testing techniques		3	85	75	H	H	M	H	L	-	-	-	M	L	-	H	-	-	-

Duration (hour)		15	15	15	15	15
S-1	SLO-1	Getting Started with R and Social Media Analytics, Understanding Social Media	Visualizing data , Managing packages	Overview of Text Mining , What's Special About Text Mining?	Using Text for Prediction	Finding Structure in a Document Collection
	SLO-2	Advantages and Significance of Social Media	Data analytics - Analytics workflow	Structured or Unstructured Data	Recognizing that Documents Fit a Pattern	
S-2	SLO-1	Disadvantages and Pitfalls of Social Media	Machine learning techniques	Is Text Different from Numbers?	How Many Documents Are Enough?	Clustering Documents by Similarity
	SLO-2	Social media analytics	Supervised learning, Unsupervised learning	What Types of Problems Can Be Solved?	Document Classification , Learning to Predict from Text	Similarity of Composite Documents
S-3	SLO-1	A typical social media analytics workflow	Text analytics , Understanding Twitter, APIs	Document Classification, Information Retrieval	Similarity and Nearest-Neighbor Methods	k-Means Clustering , Hierarchical clustering
	SLO-2	Data access, Data processing and normalization, Data analysis and Insights, Opportunities and Challenges	Registering an application	Clustering and Organizing Documents	Document Similarity , Decision Rules	
S4 - 5	SLO-1	Laboratory 1 : Practice elementary mathematical operations and control statements	Laboratory 4 : Creating Various types of plots /charts from various data source	Laboratory 7 : Implementation of Linear regression with multiple regression	Laboratory 10: Implementation of classifier problem	Laboratory 13 : Implementation of decision tree
	SLO-2					
S-6	SLO-1	Getting started with R, Environment setup	Connecting to Twitter using R	Information Extraction	Decision Trees, Scoring by Probabilities	The EM Algorithm
	SLO-2	Data types	Extracting sample Tweets	Prediction and Evaluation	Linear Scoring Methods	

S-7	SLO-1	Data structures- Vectors	Trend analysis , Sentiment analysis	From Textual Information to Numerical Vectors	Evaluation of Performance - Estimating Current and Future Performance	What Do a Cluster's Labels Mean?
	SLO-2	Arrays, Matrices		Collecting Documents	Getting the Most from a Learning Method, Errors and Pitfalls in Big Data Evaluation	Applications, Evaluation of Performance
S-8	SLO-1	Lists	Key concepts of sentiment analysis –Subjectivity, Sentiment polarity	Document Standardization, Tokenization	Information Retrieval and Text Mining	Case Study: Market Intelligence from the Web
	SLO-2	DataFrames	Opinion summarization , Features	Lemmatization- Inflectional Stemming, Stemming to a Root	Is Information Retrieval a Form of Text Mining?	
S9-S10	SLO-1	Laboratory 2 : Operations on Matrices and Vectors	Laboratory 5 : Create subplots and color plots	Laboratory 8 : Implementation of Data preprocessing methods , Correlation matrix	Laboratory 11 : Implementation of K-Mean Clustering	Laboratory 14 : Implementation of Random Forest
	SLO-2					
S-11	SLO-1	Functions - Built-in functions	Sentiment analysis in R , Follower graph analysis	Vector Generation for Prediction, Multiword Features	Key Word Search , Nearest-Neighbor Methods	Mining Social Media , E-mail Filtering
	SLO-2	User-defined functions		Labels for the Right Answers, Feature Selection by Attribute Ranking	Measuring Similarity -Shared Word Count	
S-12	SLO-1	Controlling code flow - Looping constructs	Flickr Data Analysis , Accessing Flickr's data	Sentence Boundary Determination	Word Count and Bonus, Cosine Similarity	Emerging Directions
	SLO-2	Conditional constructs	Understanding Flickr data	Part-of-Speech Tagging	Web-Based Document Search - Link Analysis	
S-13	SLO-1	Advanced operations	Understanding interestingness – similarities	Word Sense Disambiguation	Document Matching	Summarization
	SLO-2	apply, lapply, sapply, tapply, mapply	Are your photos interesting? - Preparing the data - Building the classifier	Phrase Recognition, Named Entity Recognition, Parsing, Feature Generation	Inverted Lists, Evaluation of Performance	Active Learning , Learning with Unlabeled Data
	SLO-1					

S14-15	SLO-2	Laboratory 3 : Vectorized operation on simple matrix operations	Laboratory 6 : Implement Linear regression problem	Laboratory 9 : Implementation of spam and non-spam classification problem.	Laboratory 12 : Implementation of K- Mean Clustering	Laboratory 15 : Implementation of CART
---------------	--------------	--	---	---	--	---

Learning Resources	<ol style="list-style-type: none"> Raghav Bali, Dipanjan Sarkar, Tushar Sharma, (2017), "Learning Social Media Analytics with R", Packt Publishing. Sholom M. Weiss, Nitin Indurkha, Tong Zhang, (2015), "Fundamentals of Predictive Text Mining", Second Edition, Springer London.
---------------------------	---

Learning Assessment											
Bloom's Level of Thinking		Continuous Learning Assessment (50% weightage)								Final Examination (50% weightage)	
		CLA – 1 (10%)		CLA – 2 (15%)		CLA – 3 (15%)		CLA – 4 (10%)#			
		Theory	Practice	Theory	Practice	Theory	Practice	Theory	Practice	Theory	Practice
Level 1	Remember Understand	20%	20%	15%	15%	15%	15%	15%	15%	15%	15%
Level 2	Apply Analyze	20%	20%	20%	20%	20%	20%	20%	20%	20%	20%
Level 3	Evaluate Create	10%	10%	15%	15%	15%	15%	15%	15%	15%	15%
Total		100 %		100 %		100 %		100 %		100%	

CLA – 4 can be from any combination of these: Assignments, Seminars, Short Talks, Mini-Projects, Case-Studies, Self-Study, MOOCs, Certifications, Conf. Paper etc.,

Course Designers		
Experts from Industry	Experts from Higher Technical Institutions	Internal Experts
Mr. S. Karthik, IT Analyst, Tata Consultancy Services	Dr. Neelanarayanan,, Professor, School of Computer Science and Engineering, VIT Chennai	Mrs. Sweety Bakiarani
		Mr. M. Ramesh