SSN COLLEGE OF ENGINEERING
Department of CSE
COURSE PLAN

SUBJECT NAME	**:**	**DATA ANALYTICS**
SUBJECT CODE	**:**	**IT6006**
DEGREE / YEAR	**:**	**B.E. CSE / IV YEAR/ A Section**
BATCH	**:**	**2014-2018**
SEMESTER	**:**	**VII (2017-18: Odd)**
NAME OF THE STAFF	**:**	**S.RAJALAKSHMI**
DESIGNATION	**:**	**ASSISTANT PROFESSOR**

Teaching Methodology and aids (Content Delivery Methods(CDM)) : **Powerpoint presentations\Projector\Use of ICT\Chalk and Blackboard (for all topics)**
Content Delivery Methods (CDM): T-Tutorial, S-Seminar, D-Demo

Sl.No	Unit No	Topic	CDM	No of Hrs (plan)	No of Hrs (actual)	Remarks
1	**UNIT 1 (8 Hrs)**	**INTRODUCTION TO BIG DATA** Introduction to Big Data Platform – Challenges of conventional systems		1		
2		Web data – Evolution of Analytic scalability, analytic processes and tools, Analysis vs reporting		1		
3		Modern data analytic tools		1		
4		Statistical concepts: Sampling distributions	T	2		
5		Resampling, statistical inference		2		
6		Prediction error		1		
		Planned Hours		**8**		
1	**UNIT 2 (12 Hrs)**	**DATA ANALYSIS** Regression modeling	T	2		
2		Multivariate analysis,		1		
3		Bayesian modeling, inference and Bayesian networks,	D	2		
4		Support vector and kernel methods	D	2		
5		Analysis of time series: linear systems analysis, nonlinear		2		
6		Rule induction		1		
7		Neural networks: learning and generalization, competitive learning, principal component analysis and neural networks	D	3		
8		Fuzzy logic: extracting fuzzy models from data, fuzzy decision trees		1		
9		Stochastic search methods.		1		
		Planned Hours		**15**		

Sl.No	Unit No	Topic		No of Hrs (plan)	No of Hrs (actual)	Remarks
1	**UNIT 3 (8 Hrs)**	**MINING DATA STREAMS** Introduction to Streams Concepts – Stream data model and architecture - Stream Computing		1		
2		Sampling data in a stream – Filtering streams		1		
3		Counting distinct elements in a stream		1		
4		Estimating moments – Counting oneness in a window		2		
5		Decaying window - Realtime Analytics Platform(RTAP) applications		1		
6		case studies - real time sentiment analysis, stock market predictions	S	2		
		Planned Hours		**8**		
1	**UNIT 4 (9 Hrs)**	**FREQUENT ITEMSETS AND CLUSTERING** Mining Frequent itemsets - Market based model – Apriori Algorithm	T	2		
2		Handling large data sets in Main memory – Limited Pass algorithm		1		
3		Counting frequent itemsets in a stream		1		
4		Clustering Techniques – Hierarchical		1		
5		K- Means	D	1		
6		Clustering high dimensional data – CLIQUE and PROCLUS		1		
7		Frequent pattern based clustering methods – Clustering in non-euclidean space		1		
8		Clustering for streams and Parallelism.		1		
9		**Planned Hours**		**9**		

Sl. No	Unit No	Topic		No of Hrs (plan)	No of Hrs (actual)	Remark
1	**UNIT 5 (8 Hrs)**	**FRAMEWORKS AND VISUALIZATION** Hadoop - Hadoop Distributed file systems		2		
2		MapReduce , Hive, MapR		2		
3		Sharding – NoSQL Databases - S3		2		
4		Visualizations - Visual data analysis techniques, interaction techniques; Systems and applications:		2		
		Planned Hours		**8**		

Total Number of Syllabus Hours : 45
Total Number of Planned Hours : 48

PREPARED BY	REVIEWED BY	APPROVED BY
S.Rajalakshmi	PAC team	Dr.Chitra Babu HOD-CSE