SEMESTER S5

DATA ANALYTICS

Course Code	PECST523	CIE Marks	40
Teaching Hours/Week (L: T:P: R)	3:0:0:0	ESE Marks	60
Credits	3	Exam Hours	2 Hrs. 30 Min.
Prerequisites (if any)	None	Course Type	Theory

Course Objectives:

1. To help the learner to understand the basic concepts of data analytics.
2. To cover the mathematics for data analytics, predictive and descriptive analytics of data, classification, and clustering & text analytics.
3. To enable the learners to perform data analysis on a real world scenario using appropriate tools.

SYLLABUS

Module No.	Syllabus Description	Contact Hours
1	**Introduction to Data Analytics:-** Analytics Process Model, Analytical Model Requirements, Data Analytics Life Cycle overview; Association of two variables - Discrete variables, Ordinal and Continuous variable; Probability calculus - probability distributions; Hypothesis Testing - Basic definitions. **Proximity Measures** - Data Objects, Attribute types, Dissimilarity and Similarity measures.	9
2	**Association of Two Variables:-** Summarizing the Distribution of Two Discrete Variables, Contingency Tables for Discrete Data, Joint, Marginal, and Conditional Frequency Distributions, Graphical Representation of Two Nominal or Ordinal Variables, Measures of Association for Two Discrete Variables,	9

	Association Between Ordinal and Continuous Variables, Visualization of Variables from Different Scales.	
3	**Statistical Description of data** - Central tendency, Dispersion, Range, Quartiles, Variance, Standard Deviation, and Interquartile Range. **Data Preprocessing** - Cleaning, Integration, Reduction, Transformation, Discretization. **Mining Frequent Patterns** - Associations, Correlations, and Apriori Algorithms. **Classification** - General approach to classification, ID3, Attribute selection measures, Naive Bayesian Classification. **Clustering** - K-Means, Agglomerative versus Divisive Hierarchical Clustering, BIRCH, DBSCAN.	**9**
4	**Text Processing** :- Boolean retrieval, Example IR problem, inverted index, processing Boolean queries, tokenization, stemming, phrase queries, vector space model, finite automata and language model, query likelihood model, naïve bayes text classification.	**9**

Course Assessment Method

(CIE: 40 marks, ESE: 60 marks)

Continuous Internal Evaluation Marks (CIE):

Attendance	Assignment/ Microproject	Internal Examination-1 (Written)	Internal Examination- 2 (Written)	Total
5	**15**	**10**	**10**	**40**

End Semester Examination Marks (ESE)

In Part A, all questions need to be answered and in Part B, each student can choose any one full question out of two questions

Part A	Part B	Total
• 2 Questions from each module. • Total of 8 Questions, each carrying 3 marks **(8x3 =24 marks)**	• Each question carries 9 marks. • Two questions will be given from each module, out of which 1 question should be answered. • Each question can have a maximum of 3 subdivisions. **(4x9 = 36 marks)**	**60**

Course Outcomes (COs)

At the end of the course students should be able to:

Course Outcome		Bloom's Knowledge Level (KL)
CO1	Explain the key concepts of data analytics	**K2**
CO2	Apply appropriate techniques to convert raw data into suitable format for practical data analytics tasks	**K3**
CO3	Extend the concept of association rule mining in real world scenario	**K3**
CO4	Select appropriate clustering and classification algorithms for various applications and extend data analytics methods to the new domains of data.	**K4**
CO5	Understand the basics of text analytics and text classification	**K3**

Note*: K1- Remember, K2- Understand, K3- Apply, K4- Analyse, K5- Evaluate, K6- Create*

CO-PO Mapping Table (Mapping of Course Outcomes to Program Outcomes)

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	3	3										3
CO2	3	3	3									3
CO3	3	3	3									3
CO4	3	3	3									3
CO5	3	3	3									3

Note: *1: Slight (Low), 2: Moderate (Medium), 3: Substantial (High), -: No Correlation*

Text Books				
Sl. No	**Title of the Book**	**Name of the Author/s**	**Name of the Publisher**	**Edition and Year**
1	Introduction to Statistics and Data Analysis	Christian Heumann and Michael Schomaker	Springer	1/e, 2016
2	Jiawei Han and Micheline Kamber	Data Mining Concepts and Techniques	Elsevier	3/e, 2012

Reference Books				
Sl. No	**Title of the Book**	**Name of the Author/s**	**Name of the Publisher**	**Edition and Year**
1	Introduction to Information Retrieval	Christopher D. Manning, Raghavan, P., Schutze, H.	Cambridge University Press	1/e, 2008
2	Mining Text Data	Charu C. Aggarwal, Cheng Xiang Zhai	Springer	1/e, 2012
3	Analytics in a Big Data World: The Essential Guide to Data Science and its Business Intelligence and Analytic Trends	Bart Baesens	John Wiley	1/e, 2013
4	Introduction to Data Mining	Pang-Ning Tan, Michael Steinbach and Vipin Kumar	Pearson Education	1/e, 2007

Video Links (NPTEL, SWAYAM…)	
No.	**Link ID**
1	https://archive.nptel.ac.in/noc/courses/noc19/SEM1/noc19-cs15/
2	https://onlinecourses.swayam2.ac.in/cec19_cs01/preview