Savitribai Phule Pune University T.Y.B.Sc. (Computer Science) – Sem - V Course Type:DSEC – II Course Code: CS - 354 Paper Title : Foundations of Data Science					
Teaching Scheme	No. of Credits	Examination So	cheme		
03 lectures / week	2	IE : 15 mar	rks		
		UE: 35 mar	rks		
Prerequisites					
Problem solving using of the solving using of the solving using of the solving using of the solving using the solving usin the solving using the solving using the solving using the solvin	computers				
• Basic mathematics and	statistics				
Knowledge of Database	28				
Course Objectives					
 Provide students with k scientific discovery 	nowledge and skills for data-inte	ensive problem solvin	ng and		
• Be prepared with a vari	ed range of expertise in different	aspects of data scien	nce such as		
data collection, visualiz	ation, processing and modeling	of large data sets.			
• Acquire good understar	ding of both the theory and appl	ication of applied stat	tistics and		
computer science based	l existing data science models to	analyze huge data set	ets		
originating from divers	ified application areas.	,			
• Be better trained profes	sionals to cater the growing dem	and for data scientists	s in		
industry.					
Course Outcomes					
On completion of the course, s	tudent will be able to-				
Perform Exploratory Da	ata Analysis				
• Obtain, clean/process, a	and transform data.				
• Detect and diagnose co	mmon data issues, such as missi	ng values, special valu	ues.		
outliers, inconsistencies	• Detect and diagnose common data issues, such as missing values, special values, outliers inconsistencies and localization				
Demonstrate proficienc	 Demonstrate proficiency with statistical analysis of data 				
 Present results using data 	ta visualization techniques.				
 Prepare data for use with 	h a variety of statistical methods	and models and reco	ognize how		
the quality of the data a	nd the means of data collection i	nav affect conclusion			
Course Contents	ind the means of data concerton i	nay arreet conclusion	15.		
Chapter 1 Introduction to	o Data Science	6	6 lectures		
Introduction to data science. T	he 3 V's: Volume, Velocity, Var	ietv			
Why learn Data Science?					
Applications of Data Science					
The Data Science Lifecycle					
Data Scientist's Toolbox	Data Scientist's Toolbox				
Types of Data					
Structured, sem	i-structured, Unstructured Data,	Problems with unstru	ictured		
data					
Data sources					
Open Data, Social Media Data, Multimodal Data, standard datasets					
Data Formats					
Integers, Floats, Text Data, Text Files, Dense Numerical Arrays, Compressed or					
Archived Data,	CSV Files, JSON Files, XML Fi	les, HTML Files , Ta	ar Files,		
GZip Files, Zip	Files, Image Files: Rasterized, V	vectorized, and/or Con	mpressed		

Chapter 2	Statistical Data Analysis	10 lectures			
2.1.Role of	of statistics in data science				
2.2.Descr	iptive statistics				
	Measuring the Frequency				
	Measuring the Central Tendency: Mean, Median, and Mode				
	Measuring the Dispersion: Range, Standard deviation, Variance, Int	erquartile			
	Range				
2.3.Infere	ntial statistics				
	Hypothesis testing, Multiple hypothesis testing, Parameter Estimation	on methods,			
2.4.Meas	uring Data Similarity and Dissimilarity				
	Data Matrix versus Dissimilarity Matrix, Proximity Measures for N	ominal			
	Attributes, Proximity Measures for Binary Attributes, Dissimilarity	of Numeric			
	Data: Euclidean, Manhattan, and Minkowski distances, Proximity M	leasures for			
	Ordinal Attributes				
2.5.Conce	ept of Outlier, types of outliers, outlier detection methods				
Chantor 3	Data Proprocessing	10 lootures			
Chapter 5	Data Preprocessing				
Data Objects	and Attribute Types: what Is an Attribute?, Nominal, Binary, Ordina	al			
Attributes, N	imeric Attributes, Discrete versus Continuous Attributes				
Data Qua	lity: Why Preprocess the Data?				
3.3.Data	nunging/wrangling operations				
Cleaning Dat	a - Missing Values, Noisy Data (Duplicate Entries, Multiple				
	Entries for a Single Entity, Missing Entries, NULLs, Huge Outliers	, Out-of-			
	Date Data, Artificial Entries, Irregular Spacings, Formatting Issues	- Irregular			
	between Different Tables/Columns, Extra Whitespace, Irregular Ca	pitalization,			
	Inconsistent Delimiters, Irregular NULL Format, Invalid Characters	3,			
	Incompatible Datetimes)	1.0			
Data Transfo	rmation – Rescaling, Normalizing, Binarizing, Standardizing, Label ar	nd One			
Hot Encoding					
Data reductio	n .				
Data discretiz	ation				
Chanter 4	Data Visualization	10 loctures			
Introduction t	o Exploratory Data Analysis	10 lectures			
Data visualiz	o Exploratory Data Analysis				
Data visualiz	ation libraries				
Basic data vis	ultimitiates				
Dasic uata vis	Histograms Bar charts/graphs Scatter plots Line charts Area plot	· Dia charte			
	Histograms, Bar charts/graphs, Scatter plots, Line charts, Area plots, Pie charts,				
Donut charts Specialized data visualization tools					
Specialized data visualization tools Revelete Rubble plote Heat man Dendrogram Vonn diagram Treeman 2D					
Boxpiols, Buddle piols, Heat map, Dendrogram, Venn diagram, Treemap, 5D					
Advanced data visualization tools. Wordslouds					
	Visualization of geospatial data				
Data Visualization types					
Reference Books.					
1) Data Saianaa Eundamentale and Dreatical Approaches. Correst Mardi Dourous					
1) Data Science Fundamentals and Practical Approaches, Gypsy Nandi, Kupam Sharma, BDB Publications, 2020					
	2) The Data Science Handbook Field Cady John Wiley & Sons Inc 2017				
	The Data Science Handbook, Fleid Cady, John Wiley & Sons, Inc, 20.	1/			
3) Data Mining Concepts and Techniques, Third Edition, Jiawei Han, Micheline					

Kamber, Jian Pei, Morgan Kaufmann, 2012.4) A Hands-On Introduction to Data Science, Chirag Shah, University of Washington Cambridge University Press

Savitribai Phule Pune University T.Y.B.Sc. (Computer Science) – Sem VI Course Type:DSEC – V Course Code: CS - 364 Course Title : Data Analytics				
Teaching Scheme	No. of Credits	Examina	ation Scheme	
03 lectures / week	2	IE :	15 marks	
		UE:	35 marks	
 Prerequisites Basic of mathematics and statistics Basic programming Knowledge of python Knowledge of databases 				
 Course Objectives Deploy the Data Analytics Lifecycle to address data analytics projects. Develop in depth understanding of the key technologies in data analytics. Apply appropriate analytic techniques and tools to analyze data, create models, and identify insights that can lead to actionable results. 				
Course Outcomes				
 On completion of the course, student will be able to- Use appropriate models of analysis, assess the quality of input, and derive insight from results. Analyze data, choose relevant models and algorithms for respective applications Understand different data mining techniques like classification, prediction, clustering and association rule mining 				
problems	data analysis teeninques to th	le solution of real	i world busiliess	
	Course Contents			
Chapter 1 Introduction	to Data Analytics		6 lectures	
Concept of data analytics Data analysis vs Data analytics Types of analytics Diagnostic Analytics, Predictive Analytics, Prescriptive Analytics, Exploratory Analysis, Mechanistic Analysis Mathematical models - Concept Model evaluation: metrics for evaluating classifiers - Class imbalance - AUC, ROC (Receiver-Operator Characteristic) curves, Evaluating value prediction models				
Chapter 2 Machine Lea	arning Overview		6 Lectures	
Introduction to Machine Learning, deep learning, Artificial intelligence Applications for machine learning in data science The modeling process Engineering features and selecting a model, Training the model, Validating the model, Predicting new observations Types of machine learning				
Supervised learning, Unsupervised learning, Semi-supervised learning, ensemble techniques Regression models Linear Regression				

	Polynomial Regression				
	Logistic Regression				
2.6. Cond	2.6. Concept of classification, clustering and reinforcement learning.				
Chapter 3	Mining Frequent Patterns, Associations, and	12 lectures			
	Correlations				
What kind of	patterns can be mined				
	Class/Concept Description: Characterization and Discrimi	nation, Mining			
	Frequent Patterns, Associations, and Correlations, Classifi	cation and Regression			
	for Predictive Analysis, Cluster Analysis, Outlier Analysis	8			
Mining frequ	ent patterns - Market Basket Analysis.				
Frequent Iter	nsets, Closed Itemsets, and Association Rules				
Apriori Algo	rithm				
Generating A	sociation Rules from Frequent Itemsets				
Improving of	ficiency of apriori algorithm				
Frequent nat	ern growth (FP-growth) algorithm				
r requein pau					
Chapter 4	Social Media and Text Analytics	12 lectures			
Overview of	social media analytics				
Social Media Analytics Process, Seven layers of social media analytics,					
accessing social media data					
Key social media analytics methods					
Social network analysis					
	Link prediction, Community detection, Influence maximization, Expert finding,				
Prediction of trust and distrust among individuals					
Introduction to Natural Language Processing					
Text Analytics : Tokenization, Bag of words, Word weighting : TF-IDF, n-Grams, stop					
words, Stemi	ning and lemmatization, synonyms and parts of speech tagg	ing			
Sentiment Analysis					
Trend analytics					
Challenges to social modia analytics					
Chancinges to	social media analytics				
Reference B	ooks:				
1) Data Science Fundamentals and Practical Approaches, Gypsy Nandi, Rupam Sharma,					
BPB Publications, 2020.					
2) The Data Science Handbook, Field Cady, John Wiley & Sons, Inc, 2017					
3) Data Mining Concepts and Techniques, Jiawei Han, Micheline Kamber, Jian Pei,					
Morgan Kaufmann, Third Edition, 2012.					
4) A Hands-On Introduction to Data Science, Chirag Shah, University of Washington					
Camb	ridge University Press				

- 5) The Data Science Design Manual, Steven S. Skiena, Springer, 2017
- 6) Introducing data science: big data, machine learning, and more, using Python tools, Cielen D., Meysman A. D., & Ali M., Manning Publications Co., 2016