THE TITLE OF THE TIME OF THE T	Course Title: DATA SCIENCE							
Dr.	Course Code:	No. of Credits: 3: 0: 0	No. of lecture hours/week:					
AND A PEETHA WELFARE TRUST	18CSE024	(L-T-P)	3					
Aided By Govt. of Karnataka	Exam Duration:	CIE + Assignment + SEE =	Total No. of Contact					
	3 hours	45 + 5 + 50 = 100	Hours: 42					

Course	Description						
Objectives:	1. Determine the appropriate natural language processing, machine learning and deep learning models to solve the business-related challenges.						
	2. Indicate proficiency with statistical analysis of data to derive insight from results and interpret the data findings visually.						
	3. Demonstrate skills in data management by obtaining, cleaning and transforming the data.						
	4. Discuss how social networks appraise the ways in which the social clustering shape individuals and groups in contemporary society.						

Unit No	Syllabus Content	No of Hours				
1.	Visualizing Data, matplotlib, Bar Charts, Line Charts, Scatterplots, Linear					
	Algebra, Vectors, Matrices, Statistics, Describing a Single Set of Data,					
	Correlation, Simpson's Paradox, Some Other Correlational Caveats, Correlation					
	and Causation, Probability , Dependence and Independence, Conditional					
	Probability, Bayes's Theorem, Random Variables, Continuous Distributions, The					
	Normal Distribution, The Central Limit Theorem.					
2.	Hypothesis and Inference , Statistical Hypothesis Testing, Example: Flipping a					
	Coin, p-Values, Confidence Intervals, p-Hacking, Example: Running an A/B					
	Test, Bayesian Inference, Gradient Descent , The Idea Behind Gradient Descent					
	Estimating the Gradient, Using the Gradient, Choosing the Right Step Size, Using					
	Gradient Descent to Fit Models, Minibatch and Stochastic Gradient Descent, Getting Data, stdin and stdout, Reading Files, Scraping the Web, Using APIs, Example: Using the Twitter APIs, Working with Data, Exploring Your Data, Using NamedTuples, Dataclasses, Cleaning and Munging, Manipulating Data, Rescaling, An Aside: tqdm, Dimensionality Reduction.					
3.	Machine Learning, Modeling, What Is Machine Learning?, Overfitting and					
	Underfitting, Correctness, The Bias-Variance Tradeoff, Feature Extraction and					
	Selection, k-Nearest Neighbors, The Model, Example: The Iris Dataset, The					
	Curse of Dimensionality, Naive Bayes, A Really Dumb Spam Filter, A More					
	Sophisticated Spam Filter, Implementation, Testing Our Model, Using Our					
	Model, Simple Linear Regression, The Model, Using Gradient Descent,					
	Maximum Likelihood Estimation, Multiple Regression, The Model, Further					
	Assumptions of the Least Squares Model, Fitting the Model, Interpreting the					
	Model, Goodness of Fit, Digression: The Bootstrap, Standard Errors of					
	Regression Coefficients, Regularization, Logistic Regression, The Problem, The					

Strong	rong -3 Medium -2 Weak -1												
CO		3	2	2	2	3	-	-	-	-	-	-	-
CO	3	3	3	3	3	3	-	-	-	-	-	-	-
CO	2	2	2	2	2	3	-	-	-	-	-	-	-
CO	1	3	3	2	2	3	-	-	-	-	-	-	-
CO-P Mappi		PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO	Illustrate how network analysis and recommender systems can contribute to increasing knowledge about diverse aspects of societal clustering.							L3					
СО	for real world problems.							L3					
СО	Examine visualize curate and prepare data and recognize how the quality						L3						
CO	1	Interpret the concepts and methods of mathematical disciplines relevant to data analytics and statistical modeling.							L3				
Cour Outco	Description							RBT Levels					
	Popular, User-Based Collaborative Filtering, Item-Based Collaborative Filtering, Matrix Factorization.												
	Analysis, Betweenness Centrality, Eigenvector Centrality, Directed Graphs and PageRank, Recommender Systems, Manual Curation, Recommending What's												
	Grammars, An Aside: Gibbs Sampling, Topic Modeling, Word Vectors, Recurrent Neural Networks, Example: Using a Character-Level RNN, Network												
•	Natural Language Processing, Word Clouds, n-Gram Language Models,								00				
5.	Choosing k, Example: Clustering Colors, Bottom-Up Hierarchical Clustering. SELF-STUDY								08				
	Revisited, Softmaxes and Cross-Entropy, Dropout, Example: MNIST, Saving and Loading Models, Clustering, The Idea, The Model, Example: Meetups,												
	Linear Layer, Neural Networks as a Sequence of Layers, Loss and Optimization, Example: XOR Revisited, Other Activation Functions, Example: FizzBuzz												
	Example: Fizz Buzz, Deep Learning, The Tensor, The Layer Abstraction, The												
	Creating a Decision Tree, Putting It All Together, Random Forests, Neural Networks , Perceptrons, Feed-Forward Neural Networks, Backpropagation,												
4.	Decision Trees, What Is a Decision Tree?, Entropy, The Entropy of a Partition,								09				
	Logistic Function, Applying the Model, Goodness of Fit, Support Vector Machines.												

TEXT BOOKS:

1. Joel Grus, "Data Science from Scratch", 2nd Edition, O'Reilly Publications/Shroff Publishers and Distributors Pvt. Ltd., 2019. ISBN-13: 978-9352138326.

REFERENCE BOOKS:

- 1. Emily Robinson and Jacqueline Nolis, "Build a Career in Data Science", 1st Edition, Manning Publications, 2020. ISBN: 978-1617296246.
- 2. Aurélien Géron, "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems", 2nd Edition, O'Reilly Publications/Shroff Publishers and Distributors Pvt. Ltd., 2019. ISBN-13: 978-1492032649.
- 3. François Chollet, "**Deep Learning with Python**", 1st Edition, Manning Publications, 2017. ISBN-13: 978-1617294433
- 4. Jeremy Howard and Sylvain Gugger, "Deep Learning for Coders with fastai and PyTorch", 1st Edition, O'Reilly Publications/Shroff Publishers and Distributors Pvt. Ltd., 2020. ISBN-13: 978-1492045526.
- 5. Sebastian Raschka and Vahid Mirjalili, "Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2", 3rd Edition, Packt Publishing Limited, 2019. ISBN-13: 978-1789955750

SELF-STUDY REFERENCES/WEBLINKS:

1. Natural Language Processing

https://www.youtube.com/watch?v=xvqsFTUsOmc

2. Network Analysis

https://www.youtube.com/watch?v=K5xiFDClgjo

3. Recommender Systems

https://www.youtube.com/watch?v=39vJRxIPSxw

COURSE COORDINATOR:

Dr.Gowrishankar S.