Centre for Data Science
Faculty of Engineering & Technology (ITER), SOA Deemed to be University

LESSON PLAN

Introduction to Data Science using Python (CSE 3054)

Session: April' 2022 - July' 2022

1. **Course Number and Name:**

   CSE 3054, `Introduction to Data Science using Python`

2. **Credits and Course Format:**

   Grading Pattern = 2

   Credits = 3

   Course format: 6 hours/week (2 labs/week, 2hrs/lab, 1 Problem solving class/week, 2hrs/class)

3. **Target Students:**

   Programme: B.Tech. (6^{th} Semester)

   Branch: CSE

4. **Instructor's Names:**

(1) Dr. Manoranjan Parhi, Associate Professor, ITER
`manoranjanparhi@soa.ac.in`

(2) Ashish Ranjan, Assistant Professor, ITER
`ashishranjan@soa.ac.in`

(3) Chiranjibi Mahapatra, Assistant Professor, ITER
`chiranjibimahapatra@soa.ac.in`

(4) Jagseer Singh, Assistant Professor, ITER
`jagseersingh@soa.ac.in`

(5) Dr. Madhurima Datta, Assistant Professor, ITER
`madhurimadatta@soa.ac.in`

(6) Nikhil Jaiswal, Assistant Professor, ITER
`nikhiljaiswal@soa.ac.in`

(7) Prabhat Dansena, Assistant Professor, ITER
`prabhatdansena@soa.ac.in`

(8) Kumuda Sharma, Assistant Professor, ITER
`kumudasharma@soa.ac.in`

(9) Jahanvi Rajput, Assistant Professor, ITER
`jahanvirajput@soa.ac.in`

Lesson Plan

(10) Priya, Assistant Professor, ITER
priya@soa.ac.in

(11) Shivendu Prasad, Assistant Professor, ITER
shivenduprasad@soa.ac.in

5. **Text Book(s):**

   (1) Data Science from Scratch Joel Grus, Shroff/O'reilly, Second Edition

6. **Specific Course Information:**

   (a) **Course Description:** Preliminaries, Introduction, A Crash Course in Python, Visualizing Data, Linear Algebra, Statistics, Probability, Hypothesis and Inference, Gradient Descent, Getting Data, Working with Data, Machine Learning, k-Nearest Neighbors, Naive Bayes, Simple Linear Regression, Multiple Regression, Logistic Regression, Decision Trees, Neural Networks, Deep Learning, Clustering, Natural Language Processing, Network Analysis, Databases and SQL, MapReduce, Data Ethics

   (b) **Prerequisites:**

   ☞ Programming in Python (CSE 3142)

7. **Course Outcomes (COs) :**
   By the end of course through lectures, readings, home works, laboratory, assignments and exams students will be able to:

   CO 1. Understand how to use data to predict most accurate results.

   CO 2. Develop efficient data science models for several problems.

   CO 3. Analyze, debug and test the models on various sets of training data to make the model as accurate as possible.

   CO 4. Illustrate the process of extracting only the useful data out of the huge datasets that are available and make use of only the relevant data to create a model.

   CO 5. Solve the real life problems by preprocessing the datasets given and applying the model on them to predict the best output.

   CO 6. Retrieve and analyse the information from a computerized database using the concepts of SQL.

8. **Brief List of Topics to Be Covered: (L: Lecture, P: Laboratory)**

Contact hour	***Topics to be covered***	***Remarks (if any)***
Week#1:		

L-01	Introduction to the course and Evaluation procedure, Introduction: The Ascendance of Data, What Is Data Science?, A Crash Course in Python	
P-01	A Crash Course in Python (Cont.)	
P-02	Visualizing Data, Matplotlib, Bar Charts, Line Charts, Scatter Plots Minor Assignment-1	
Week#2:		
L-02	Linear Algebra: Vectors, Matrices. Statistics: Describing a Single Set of Data	
P-03	Statistics: Correlation. Probability: Dependence and Independence, Conditional Probability	
P-04	Bayes's Theorem, Random Variables, Continuous Distributions, The Normal Distribution, The Central Limit Theorem Minor Assignment-2	
Week#3:		
L-03	Hypothesis and Inference: Statistical Hypothesis Testing, p-Values, Confidence Intervals	
P-05	Gradient Descent: The Idea Behind Gradient Descent, Estimating the Gradient, Using the Gradient Descent to Fit Models	
P-06	Minibatch and Stochastic Gradient Descent, Minor Assignment-3	
Week#4:		
L-04	Getting Data: Reading files Minor Assignment-4 Working with Data: Exploring Your Data	
P-07	Using NamedTuples, Dataclasses, Cleaning and Munging, Manipulating Data, Rescaling, Dimensionality Reduction	
P-08	Minor Assignment-5 Machine Learning: Modeling, What Is Machine Learning?	
Week#5:		

Lesson Plan

L-05	Machine Learning (Contd.) : Overfitting and Underfitting, Correctness The Bias-Variance Tradeoff	
P-09	Machine Learning (Contd.) : Feature Extraction and Selection Minor Assignment-6	
P-10	k-Nearest Neighbors: The Model, Example: The Iris Dataset, The Curse of Dimensionality	
Week#6:		
L-06	Naive Bayes : Spam Filtering, Implementation	
P-11	Naive Bayes(Contd.) : Testing and using model Minor Assignment-7	
P-12	Simple Linear Regression: The Model, Maximum Likelihood Estimation	
Week#7:		
L-07	Multiple regression: The model, Fitting and interpreting the model	
P-13	Multiple regression(contd.) : Regularization, Digression	
P-14	Logistic regression : The problem, Logistic function, Applying the model	
Week#8:		
L-08	Logistic regression: SVM Minor Assignment - 8	
P-14	Decision Tree : Entropy, Random forests	
P-15	Neural Networks : Perceptron, Feed forward neural network	
Week#9:		
L-09	Backpropagation Minor Assignment-9	
P-17	Deep learning : Tensor, Layer abstraction, Linear layer, neural network as a sequence of layers	
P-18	Loss and optimization, activation function, Cross entropy	
Week#10:		

L-10	Dropout, Saving and loading models Minor Assignment-10	
P-19	Clustering : The idea, The model, choosing k	
P-20	Example: Clustering colors, Bottom up hierarchical clustering	
Week#11:		
L-11	Natural language processing : Word clouds, Grammars, Topic modeling	
P-21	Recurrent neural network Minor Assignment-11	
P-22	Network Analysis : Betweenness centrality, Directed graphs and Pagerank	
Week#12:		
L-12	Recommender System: Manual Curation, User based and item based Collaborative Filtering, Matrix Factorization	
P-23	Databases and SQL: Create table, insert, update, delete, select, group by, order by, join, Query Optimization	
P-24	Map Reduce Minor Assignment-12 Data Ethics	

Lesson Plan

9. **Evaluation scheme (under Grading Pattern-2) out of** 100%**:**

Attendance:	05%
Major Lab/Session Assignments/Quizzes :	13%
Minor Assignments :	7%
Mid-Term :	15%
In Lab Examination:	20%
Theory Examination:	40%

10. **Program Outcomes & Program Specific Outcomes**
There are twelve program outcomes (1-12) and Two program specific outcomes for the Computer science & Engineering B. Tech program:
Program Outcomes (POs)

1. **Engineering knowledge:** Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.
2. **Problem analysis:** Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.
3. **Design/development of solutions:** Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.
4. **Conduct investigations of complex problems:** Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
5. **Modern tool usage:** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.
6. **The engineer and society:** Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
7. **Environment and sustainability:** Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.
8. **Ethics:** Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
9. **Individual and team work:** Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
10. **Communication:** Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
11. **Project management and finance:** Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.
12. **Life-long learning:** Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

Program Specifis Outcomes (PSOs)

PSO 1. The ability to understand, analyze and develop computer programs in the areas related to business intelligence, web design and networking for efficient design of computer-based systems of varying complexities.

PSO 2. The ability to apply standard practices and strategies in software development using open-ended programming environments to deliver a quality product for business success.

11. **Correlation between the Course Outcomes(COs), the Program Outcomes(POs) , and the Program Specific Outcomes(PSOs)**

Course Articulation Matrix:

Introduction to Data Science using Python (CSE 3054)													Session: 2021-2022	
	POs												PSOs	
	1	2	3	4	5	6	7	8	9	10	11	12	1	2
CO 1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CO 2	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CO 3	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CO 4	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CO 5	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CO 6	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Average	-	-	-	-	-	-	-	-	-	-	-	-	-	-