

D.K.T.E. Society's
TEXTILE & ENGINEERING
INSTITUTE
(An Autonomous Institute)
Rajwada, Ichalkaranji – 416115.



Syllabus of
Applied Data Science (Minor)
(With effect from June 2020)

D.K.T.E. Society's

TEXTILE & ENGINEERING INSTITUTE

(An Autonomous Institute)

Rajwada , Ichalkaranji – 416115.

Department of Computer Science and Engineering

Applied Data Science (Minor)

(With effect from Academic Year 2020-21)

Syllabus Structure

Sr. No.	Course Code	Course Name	Sem	Teaching Scheme Hours/Week				Examination Scheme and Marks						Credits
				Theory	Tutorial	Practical	Total	Theory			Practical		Total	
								SE-I	SE-II	SEE	CIE	SEE		
01	CSL801	Basic Statistics (Online Course)	IV	2	1	-	3	25	25	50	-	-	100	03
02	CSL802	Programming Fundamentals	V	3	-	2	5	25	25	50	50	-	150	04
03	CSP805	Programming for Everybody (Online Course)	V	1	-	2	3	-	-	-	50	50	-	02
04	CSL804	Data Structures and Algorithms	VI	3	-	-	3	25	25	50	-	-	100	03
05	CSP806	Applied Machine Learning in Python (Online Course)	VI	2	-	2	4	-	-	-	50	50	-	03
06	CSP807	Introduction to Data Science in Python (Online Course)	VII	2	-	2	4	-	-	-	50	50	-	03
07	CSD808	Capstone Project	VII	-	-	2	2	-	-	-	50	50	100	02
Total				6	1	2	24	75	75	150	250	200	450	20

DKTES Textile and Engineering Institute, Ichalkaranji

(An Autonomous Institute)

Teaching and evaluation Scheme for year 2020-21

Second Year B. Tech. (Semester – IV) for Minor Degree in Applied Data Science

Sr. No.	Course Code	Course Title	Course Category	Teaching scheme				Course Credits	Evaluation scheme					
				L	T	P	Contact Hrs/wk		Theory		Practical		TOTAL	
									CIE		SEE	CIE		SEE
									SE-I	SE-II				
1	CSL801	Basic Statistics (Online Course)	BSC	2	1	-	3	3	25	25	50			100
		Total		2	1	-	3	3	25	25	50			100

L- Lecture
T-Tutorial
P-Practical

SE-I: Semester Examination-I
SE-II: Semester Examination-II

CIE – Continuous In Semester Evaluation
SEE- Semester End Examination

Course Category	HSMC (Hum. & Social Sc., Mgt)	BSC (Basic Sc.)	ESC (Engg. Sc.)	PCC (Prof. Core Courses)	PEC (Prof. Elect. Courses)	OEC (Open Elct. Courses)	MC (Mandatory Courses)	PST (Project / Seminar / Ind. Training)
Credits	-	03	--	--	--	--	--	--
Cumulative Sum	-	--	--	--	--	--	--	--

Progressive Total Credits = 03

DKTES Textile and Engineering Institute , Ichalkaranji
Second Year B. Tech.(Semester – IV)
CSL801: Basic Statistics

Teaching Scheme: Lectures: 02 Hrs./Week Tutorials: 01Hrs./Week Practicals: 00 Hrs./Week	Credits 03	Evaluation Scheme: SE-I: 25 Marks SE-II: 25 Marks SEE: 50 Marks
--	-------------------	--

Course Outcomes:

On completion of the course, student will be able to–

- describe the basics of statistics
- illustrate the methods of descriptive statistics.
- explain basics of probability
- illustrate methods of inferential statistics

Course Contents

Unit I	Exploring Data	06 Hours
Cases, variables and levels of measurement, Data matrix and frequency table, Graphs and shapes of distributions, Mode, median and mean, Range, interquartile range and box plot, Variance and standard deviation, Z-scores, Example		
Unit II	Correlation and Regression	06 Hours
Crosstabs and scatterplots, ,Pearson's r, Regression - Finding the line, Regression - Describing the line, Regression - How good is the line?, Correlation is not causation, Example contingency table, Example Pearson's r and regression		
Unit III	Probability	06 Hours
Randomness, Probability, Sample space, event, probability of event and tree diagram, Quantifying probabilities with tree diagram, Basic set-theoretic concepts, Practice with sets ,Union, Joint and marginal probabilities, Conditional probability, Independence between random events ,More conditional probability, decision trees and Bayes' Law		
Unit IV	Probability Distributions	06 Hours
Random variables and probability distributions, Cumulative probability distribution , The mean of a random variable, Variance of a random variable, Functional form of the normal distribution, The normal distribution: probability calculations, The standard normal distribution, The binomial distribution		
Unit V	Sampling Distributions	06 Hours
Sample and population, Sampling, The sampling distribution, The central limit theorem, Three distributions, Sampling distribution proportion , Example		

Unit VI	Confidence Intervals and Significance Tests	09 Hours
Confidence Intervals-Statistical inference, CI for mean with known population sd, CI for mean with unknown population sd, CI for proportion, Confidence levels, Choosing the sample size, Example. Significance Tests - Hypotheses, Test about proportion, Test about mean, Step-by-step plan, Significance test and confidence interval, Type I and Type II errors, Example.		
Text Books:		
--		
References Books:		
--		
Useful Links:		
1. https://www.coursera.org/learn/basic-statistics		

DKTES Textile and Engineering Institute, Ichalkaranji

(An Autonomous Institute)

Teaching and evaluation Scheme for year 2020-21

Third Year B. Tech. (Semester – V) for Minor Degree in Applied Data Science

Sr. No.	Course Code	Course Title	Course Category	Teaching scheme				Course Credits	Evaluation scheme					
									Theory			Practical		TOTAL
									CIE		SEE	CIE	SEE	
				L	T	P	Contact Hrs/wk		SE-I	SE-II				
1	CSL802	Programming Fundamentals	PCC	3	-	2	5	4	25	25	50	50	-	150
2	CSP805	Programming for Everybody (Online Course)	PCC	1	-	2	3	2	-	-	-	50	50	100
Total				4	-	4	8	6	25	25	50	100	50	250

L- Lecture

T-Tutorial

P-Practical

SE-I: Semester Examination-I

SE-II: Semester Examination-II

CIE – Continuous In Semester Evaluation

SEE- Semester End Examination

Course Category	HSMC (Hum. & Social Sc., Mgt)	BSC (Basic Sc.)	ESC (Engg. Sc.)	PCC (Prof. Core Courses)	PEC (Prof. Elect. Courses)	OEC (Open Elct. Courses)	MC (Mandatory Courses)	PST (Project / Seminar / Ind. Training)
Credits	-	--	--	06	--	--	--	--
Cumulative Sum	-	03	--	--	--	--	--	--

Progressive Total Credits 06 + 03 = 09

DKTES Textile and Engineering Institute, Ichalkaranji**Third Year B. Tech. (Semester – V)****CSL802 : Programming Fundamentals**

Teaching Scheme: Lectures: 03 Hrs/Week Tutorials: 00Hrs/Week Practicals: 00 Hrs/Week	Credits 03	Evaluation Scheme: SE-I: 25 Marks SE-II: 25 Marks SEE: 50 Marks
---	---------------	--

Course Objectives:

- To learn basic fundamental of Programming.
- To introduce concept in C Programming language
- To develop program applying concept in C Language

Course Outcomes:

On completion of the course, student will be able to–

- Explain terminology in Programming Language
- Design algorithm to solve the problem
- Build a program for developed algorithm in C Language.

Course Contents

Unit I	Introduction to digital computers and Number systems	04 Hours
Computer Software and Hardware, CPU, memory and its type, peripherals, I/O devices, Number Systems: Binary, Octal, Hexadecimal		
Unit II	Introduction to programming	05 Hours
Programming Language, Why Programming, Different programming languages, Systems Development Life Cycle, Program Design, Writing algorithm, representing it with flowcharts, Coding, testing, and debugging.		
Unit III	Imperative languages	06 Hours
Introduction to imperative language - syntax and constructs of a specific language (preferably C); variables, data types, Operators, Expression, assignment, input/output.		
Unit IV	Control Flow	06 Hours
Statements and Blocks ,Conditionals and branching: If-Else, Else-If, Switch, Loops for Repetition: While, Do-While and For Loop, Break and Continue, Goto and Labels		
Unit V	Data handling	10 Hours
One Dimensional and Two Dimensional arrays, Character and String, Pointers, Structures and Union, dynamic allocation.		
Unit VI	Functions and Recursion	05 Hours
Function - parameter passing, procedure call, call by value, call by reference; Recursion		

Textbooks:

- 1 Kenneth Leroy Busbee and Dave Braunschweig “Programming Fundamentals: A Modular Structured Approach”, 2nd Edition
- 2 C Programming Language 2nd Edition, Brian W. Kernighan , Dennis Ritchie
- 3 C How to Program 7e, by Deitel

References Books:

- 1 Let Us C, 14 Edition, Yashavant Kanetkar, BPB Publication
- 2 Problem Solving & Programming Concepts, Maureen Sprankle, Jim Hubbard, PHI Publication

Useful Links:

1. <https://www.coursera.org/specializations/c-programming>
2. https://swayam.gov.in/nd1_noc19_cs42/preview
3. <https://www.tutorialspoint.com/cprogramming/index.htm>
4. https://www.tutorialspoint.com/learn_c_by_examples/simple_programs_in_c.htm

DKTES Textile and Engineering Institute, Ichalkaranji
Third Year B. Tech. (Semester – V)
CSP805: Programming for Everybody

Lab Scheme:	Credits	Evaluation Scheme:
Lectures: 01 Hrs./Week	02	CIE: 50 Marks
Practicals: 02 Hrs./Week		SEE: 50 Marks

Course Outcomes:

On completion of the course, student will be able to–

- describe the basics of python.
- explain the procedural and object-oriented approaches used in python.
- develop various application programs using python.

UNIT-I: Why we Program?, Installing and Using Python 06 Hours

This module introduces the field of data science, review common Python functionality and features which data scientists use, and be introduced to the Jupyter Notebook for the lectures.

UNIT-II: Variables and Expressions 06 Hours

Fundamentals of one of the most important toolkits Python has for data cleaning and processing -- pandas. You'll learn how to read in data into DataFrame structures, how to query these structures, and the details about such structures are indexed.

UNIT-III: Conditional Code, Functions 06 Hours

This module covers python pandas library by learning how to merge DataFrames, generate summary tables, group data into logical pieces, and manipulate dates. We'll also refresh your understanding of scales of data, and discuss issues with creating metrics for analysis.

UNIT-IV: Loops and Iteration 06 Hours

Loops and iteration complete our four basic programming patterns. Loops are the way we tell Python to do something over and over. Loops are the way we build programs that stay with a problem until the problem is solved.

List of Experiments

(It should consist of 10-12 experiments based on the following topics.)

1	Installation & configuration of python on various IDE's.
2	Write a python program to take 2 numbers as command line arguments and perform their addition.
3	Write a python program to print 'n terms of Fibonacci series using iteration.
4	Write a python program using function to find all prime numbers within a given range.
5	Write a python program to demonstrate use of slicing in string.
6	Write a Python program to demonstrate working of classes and objects
7	Write a python program to demonstrate constructors.
8	Write a python program to demonstrate inheritance.
9	Write a python program to demonstrate tuples and lists using python.
10	Write a python program to demonstrate dictionaries & related functions using python.

DKTES Textile and Engineering Institute, Ichalkaranji

(An Autonomous Institute)

Teaching and evaluation Scheme for year 2020-21

Third Year B. Tech. (Semester – VI) for Minor Degree in Applied Data Science

Sr. No.	Course Code	Course Title	Course Category	Teaching scheme				Course Credits	Evaluation scheme					
									Theory			Practical		TOTAL
									CIE		SEE	CIE	SEE	
				L	T	P	Contact Hrs./wk.		SE-I	SE-II				
1	CSL804	Data Structures and Algorithms	PCC	3	-	-	3	3	25	25	50	-	-	100
2	CSP806	Applied Machine Learning in Python (Online Course)	PCC	2	-	2	4	3	-	-	-	50	50	100
Total				5	-	2	7	6	25	25	50	50	50	200

L- Lecture

T-Tutorial

P-Practical

SE-I : Semester Examination-I

SE-II : Semester Examination-II

CIE – Continuous In Semester Evaluation

SEE- Semester End Examination

Course Category	HSMC (Hum. & Social Sc., Mgt)	BSC (Basic Sc.)	ESC (Engg. Sc.)	PCC (Prof. Core Courses)	PEC (Prof. Elect. Courses)	OEC (Open Elct. Courses)	MC (Mandatory Courses)	PST (Project / Seminar / Ind. Training)
Credits	-	--	--	06	--	--	--	--
Cumulative Sum	-	03	--	06	--	--	--	--

Progressive Total Credits 06 + 09 = 15

DKTES Textile and Engineering Institute, Ichalkaranji

Third Year B. Tech. (Semester – VI)

CSL804 : Data Structures and Algorithms

Teaching Scheme: Lectures: 03 Hrs./Week Tutorials: 00 Hrs./Week Practicals: 00 Hrs./Week	Credits 03	Evaluation Scheme: SE-I: 25 Marks SE-II: 25 Marks SEE: 50 Marks
---	-------------------	--

Course Objectives:

- To get familiar with different data structures like stack, queue, list, tree, graph.
- To introduce different operations on data structures like insert, delete, search, sort.
- To learn to analyze performance of algorithms.

Course Outcomes:

On completion of the course, student will be able to–

- Explain the basics of different data structures like stack, queue, list, tree, graph, and their implementation.
- Perform different operations on data structures like insert, delete, search, sort
- Compute the complexity of algorithms.

Course Contents

Unit I	Introduction	06 Hours
Abstract Data Types (ADT), Data Types, Data Structures, Classification of Data Structures, Algorithms, Different approaches to design algorithms, Analysis of Algorithms, Asymptotic Notations.		
Unit II	Stack and Queue	07 Hours
Stack – ADT, Implementation, Operations and Applications. Queue – ADT, Implementation, Operations and Applications. Circular Queue, Priority Queue		
Unit III	Linked List	07 Hours
List - ADT, Implementation Applications, Implementation of Linear Linked List, Circular Linked List, Doubly Linked List, Operations on linked list- Insert, Delete, Search, Performance Analysis of Algorithms.		
Unit IV	Searching and Hashing	06 Hours
Linear and Binary Search, Performance Analysis of Searching Algorithms. Hashing – concept, hashing methods, hash collision, hash collision resolution techniques.		
Unit V	Sorting	06 Hours
Sorting Techniques: Bubble Sort, Insertion Sort, Selection Sort, Quick Sort, Merge Sort, Radix Sort, Performance Analysis of Sorting Algorithms.		
Unit VI	Tree and Graph	07 Hours
Basic tree concepts, binary tree – properties, implementation, traversal, binary search tree - concept, implementation, traversal, Search, Insertion, deletion. Graph- Definition and storage, traversal – depth first and breadth first algorithm		

Textbooks:

1. Data Structures using C – ISRD Group (TMH publication)
2. Data Structures with C (Schaum's Outline Series) – Seymour Lipschutz (MGH)

References Books:

1. Data Structures: A Pseudocode Approach with C - Richard F. Gilberg, Behrouz A. Forouzan (Cengage Learning)
2. Data Structures and Algorithm Analysis in C, 2 Edition, by Weiss, Pearson Education India

Useful Links:

1. https://www.tutorialspoint.com/data_structures_algorithms/data_structures_basics.htm
2. <https://www.coursera.org/specializations/data-structures-algorithms>
3. <https://www.geeksforgeeks.org/data-structures/>
4. <https://towardsdatascience.com/8-common-data-structures-every-programmer-must-know-171acf6a1a42>

DKTES Textile and Engineering Institute, Ichalkaranji
Third Year B. Tech. (Semester – IV)
CSP806: Applied Machine Learning in Python

Lab Scheme: Lectures: 02 Hrs./Week Practicals: 02 Hrs./Week	Credits 03	Evaluation Scheme: CIE: 50 Marks SEE: 50 Marks
--	--------------------------	---

Course Outcomes:

On completion of the course, student will be able to–

- describe the basics of machine learning.
- use various python libraries for machine learning.
- apply machine learning algorithms on sample data.

UNIT-I: Fundamentals of Machine Learning - Intro to SciKit Learn 06 Hours

This module introduces basic machine learning concepts, tasks, and workflow using an example classification problem based on the K-nearest neighbors method, and implemented using the scikit-learn library.

UNIT-II: Supervised Machine Learning - Part 1 06 Hours

This module delves into a wider variety of supervised learning methods for both classification and regression, learning about the connection between model complexity and generalization performance, the importance of proper feature scaling, and how to control model complexity by applying techniques like regularization to avoid overfitting. In addition to k-nearest neighbors, this week covers linear regression (least-squares, ridge, lasso, and polynomial regression), logistic regression, support vector machines, the use of cross-validation for model evaluation, and decision trees.

UNIT-III: Evaluation 06 Hours

This module covers evaluation and model selection methods that you can use to help understand and optimize the performance of your machine learning models.

UNIT-IV: Supervised Machine Learning - Part 2 06 Hours

This module covers more advanced supervised learning methods that include ensembles of trees (random forests, gradient boosted trees), and neural networks (with an optional summary on deep learning). You will also learn about the critical problem of data leakage in machine learning and how to detect and avoid it.

List of Experiments

(It should consist of 10-12 experiments based on the following topics.)

1	Demonstration of various open source tools and technologies used for machine learning.
2	Implementation of tuples, lists & dictionaries using python.
3	Implementation of array processing using NumPy.
4	Implementation of exploratory data analysis using Pandas dataframes.
5	Demonstration of data visualization using matplotlib.
6	Write a program for implementation of simple linear regression.
7	Write a program to implement Linear Regression with Multiple Variables.
8	Write a program for implementation of logistic regression.
9	Implementation of classification using neural networks.
10	Implementation of clustering using K-means.

11	Implementation of Naïve Bays Classifier .
12	Implementation of Decision Tree.

DKTES Textile and Engineering Institute, Ichalkaranji
(An Autonomous Institute)

Teaching and evaluation Scheme for year 2021-22
Third Year B. Tech. (Semester – VII) for Minor Degree in Applied Data Science

Sr. No.	Course Code	Course Title	Course Category	Teaching scheme				Course Credits	Evaluation scheme					
									Theory			Practical		TOTAL
									CIE		SEE	CIE	SEE	
				L	T	P	Contact Hrs./wk.		SE-I	SE-II				
1	CSP807	Introduction to Data Science in Python (Online Course)	PCC	2	-	2	4	3	-	-	-	50	50	100
2	CSD808	Capstone Project	PCC	-	-	2	2	2	-	-	-	50	50	100
		Total		2	-	4	6	5	-	-	-	100	100	200

L- Lecture
T-Tutorial
P-Practical

SE-I: Semester Examination-I
SE-II: Semester Examination-II

CIE – Continuous In Semester Evaluation
SEE- Semester End Examination

Course Category	HSMC (Hum. & Social Sc., Mgt)	BSC (Basic Sc.)	ESC (Engg. Sc.)	PCC (Prof. Core Courses)	PEC (Prof. Elect. Courses)	OEC (Open Elct. Courses)	MC (Mandatory Courses)	PST (Project / Seminar / Ind. Training)
Credits	-	--	--	05	--	--	--	--
Cumulative Sum	-	03	--	12	--	--	--	--

Progressive Total Credits 15 + 05= 20

DKTES Textile and Engineering Institute, Ichalkaranji
Third Year B. Tech. (Semester – IV)
CSP807: Introduction to Data Science in Python

Lab Scheme: Lectures: 02 Hrs./Week Practicals: 02 Hrs./Week	Credits 03	Evaluation Scheme: CIE: 50 Marks SEE: 50 Marks
---	-------------------	--

Course Outcomes:

On completion of the course, student will be able to–

- visualize, curate, and prepare data for use with a variety of statistical methods and models
- analyze the and apply predictive models on the data to provide insights
- evaluation of predictive models.

UNIT-I: Fundamentals of Data Manipulation with Python 06 Hours

This module introduces the field of data science, review common Python functionality and features which data scientists use, and be introduced to the Coursera Jupyter Notebook for the lectures.

UNIT-II: Basic Data Processing with Pandas 06 Hours

Fundamentals of one of the most important toolkits Python has for data cleaning and processing -- pandas. You'll learn how to read in data into DataFrame structures, how to query these structures, and the details about such structures are indexed.

UNIT-III: More Data Processing with Pandas 06 Hours

This module covers python pandas library by learning how to merge DataFrames, generate summary tables, group data into logical pieces, and manipulate dates. We'll also refresh your understanding of scales of data, and discuss issues with creating metrics for analysis.

UNIT-IV: Answering Questions with Messy Data 06 Hours

variety of statistical techniques such a distributions, sampling and t-tests. two discussions of science and the rise of the fourth paradigm -- data driven discovery.

List of Experiments

(It should consist of 10-12 experiments based on the following topics.)

1	Write a program to import and export, clean and prepare the data.
2	Write a program to visualize data using data visualization library seaborn.
3	Write a program to perform Exploratory Data Analysis (EDA) on data set.
4	Write a program to build a predictive model using regression techniques.
5	Write a program to build predictive model using classification techniques.
6	Write a program to build predictive model using Support Vector Machine (SVM) algorithm.
7	Write a program to evaluate predictive models.
8	Write a program to compare different classification algorithms.
9	Write a program to improve accuracy based on feature selection.
10	Write a program to cluster documents using clustering algorithms. Compare the performance of clustering algorithms.

DKTES Textile and Engineering Institute, Ichalkaranji
Third Year B. Tech. (Semester – V)
CSP808: Capstone Project

Lab Scheme: Practical: 02 Hrs/Week	Credits 02	Evaluation Scheme: CIE: 50 Marks SEE: 50 Marks
--	----------------------	---

Course Outcomes:

On completion of the course, student will be able to–

- A team of student will analyze the problem statement
- A team of student will build the SRS and design document
- A team of student will develop the code according to the design
- A team of student will test the developed software
- A team of student will write the report.

Student will form the group for the capstone project. The group will submit the completed project work to the department at the end of semester VII as mentioned below.

1. The workable project.
2. The project report in all respect with the following: -
 - i. Problem specifications
 - ii. System definition – requirement analysis.
 - iii. System design – dataflow diagrams, database design
 - iv. System implementation – algorithm, code documentation
 - v. Test results and test report.
 - vi. In case of object oriented approach – appropriate process be followed.

CIE will be jointly assessed by a panel of teachers appointed by head of the institution. SEE examination will be conducted by internal and external examiners as appointed by the CoE.