

Program Structure for Third Year Information Technology
Semester V & VI
UNIVERSITY OF MUMBAI
(With Effect from 2021-2022)

Semester VI

Course Code	Course Name	Teaching Scheme (Contact Hours)			Credits Assigned				
		Theory		Pract. Tut.	Theory	Pract.	Total		
ITC601	Data Mining & Business Intelligence	3		--	3	--	3		
ITC602	Web X.0	3		--	3		3		
ITC603	Wireless Technology	3		--	3	--	3		
ITC604	AI and DS – 1	3		--	3	--	3		
ITDO601 X	Department Optional Course – 2	3		--	3	--	3		
ITL601	BI Lab	--		2	--	1	1		
ITL602	Web Lab	--		2	--	1	1		
ITL603	Sensor Lab	--		2	--	1	1		
ITL604	MAD & PWA Lab	--		2	--	1	1		
ITL605	DS using Python Skill based Lab	--		2	--	1	1		
ITM601	Mini Project – 2 B Based on ML	--		4 ^s	--	2	2		
Total		15		14	15	07	22		
Course Code	Course Name	Examination Scheme							
		Theory					Term Work	Prac /oral	Total
		Internal Assessment			End Sem Exam	Exam. Duration (in Hrs)			
		Test1	Test2	Avg					
ITC601	Data Mining & Business Intelligence	20	20	20	80	3	--	--	100
ITC602	Web X.0	20	20	20	80	3	--	--	100
ITC603	Wireless Technology	20	20	20	80	3	--	--	100
ITC604	AI and DS – 1	20	20	20	80	3	--	--	100
ITDO601 X	Department Optional Course – 2	20	20	20	80	3	--	--	100
ITL601	BI Lab	--	--	--	--	--	25	25	50
ITL602	Web Lab	--	--	--	--	--	25	25	50
ITL603	Sensor Lab	--	--	--	--	--	25	25	50
ITL604	MAD & PWA Lab	--	--	--	--	--	25	25	50
ITL605	DS using Python Lab (SBL)	--	--	--	--	--	25	25	50

1. To study the requirement for progressive web application for Ecommerce using the concept of service worker, Webapp Manifest and framework tools
2. To Design a wireframe for simple PWA for E-commerce website
3. Case study for successful real life implementation of PWA.

Term Work:

Term Work shall consist of at least 10 to 12 practical's based on the above list. Also Term Work Journal must include at least 2 assignments as mentioned in above syllabus.

Term Work Marks: 25 Marks (Total marks) = 15 Marks (Experiment) + 5 Marks (Assignments) + 5 Marks (Attendance)

Practical & Oral Exam: An Practical & Oral exam will be held based on the above syllabus.

Course Code	Course Name	Teaching Scheme (Contact Hours)		Credits Assigned		
		Theory	Practical	Theory	Practical	Total
ITL605	DS using Python Lab	--	02	--	01	01

Course Code	Course Name	Examination Scheme							
		Theory					Term Work	Pract / Oral	Total
		Internal Assessment			End Sem Exam	Exam Duration (in Hrs)			
		Test1	Test 2	Avg.					
ITL605	DS using Python Lab	--	--	--	--	--	25	25	50

Lab Objectives:

Sr. No.	Lab Objectives
The Lab experiments aims:	
1	To know the fundamental concepts of data science and analytics
2	To learn data collection, preprocessing and visualization techniques for data science
3	To Understand and practice analytical methods for solving real life problems based on Statistical analysis
4	To learn various machine learning techniques to solve complex real-world problems
5	To learn streaming and batch data processing using Apache Spark
6	To map the elements of data science to perceive information

Lab Outcomes:

Sr. No.	Lab Outcomes	Cognitive levels of attainment as per Bloom's Taxonomy
---------	--------------	--

On successful completion, of course, learner/student will be able to:		
1	Understand the concept of Data science process and associated terminologies to solve real-world problems	L1
2	Analyze the data using different statistical techniques and visualize the outcome using different types of plots.	L1, L2, L3, L4
3	Analyze and apply the supervised machine learning techniques like Classification, Regression or Support Vector Machine on data for building the models of data and solve the problems.	L1,L2, L3, L4
4	Apply the different unsupervised machine learning algorithms like Clustering, Decision Trees, Random Forests or Association to solve the problems.	L1, L2,L3
5	Design and Build an application that performs exploratory data analysis using Apache Spark	L1,L2,L3,L4,L5,L6
6	Design and develop a data science application that can have data acquisition, processing, visualization and statistical analysis methods with supported machine learning technique to solve the real-world problem	L1,L2,L3,L4,L5,L6

Prerequisite: Basics of Python programming and Database management system.

DETAILED SYLLABUS:

Sr. No.	Module	Detailed Content	Hours	LO Mapping
I	Introduction to Data Science and Data Processing using Pandas	i. Introduction, Benefits and uses of data science ii. Data Science tasks iii. Introduction to Pandas iv. Data preparation: Data cleansing, Data transformation, Combine/Merge /Join data, Data loading & preprocessing with pandas v. Data aggregation vi. Querying data in Pandas vii. Statistics with Pandas Data Frames viii. Working with categorical and text data ix. Data Indexing and Selection x. Handling Missing Data	04	LO1
II	Data Visualization and Statistics	i. Visualization with Matplotlib and Seaborn ii. Plotting Line Plots, Bar Plots, Histograms Density Plots, Paths, 3Dplot, Stream plot, Logarithmic plots, Pie chart, Scatter Plots and Image visualization using Matplotlib iii. Plotting scatter plot, box plot, Violin plot, swarm plot, Heatmap, Bar Plot using seaborn iv. Introduction to scikit-learn and SciPy v. Statistics using python: Linear algebra, Eigen value, Eigen Vector, Determinant, Singular Value Decomposition, Integration, Correlation, Central Tendency, Variability, Hypothesis testing, Anova, z-test, t-test and chi-square test.	04	LO2
III	Machine Learning	i. What is Machine Learning? ii. Applications of Machine Learning; iii. Introduction to Supervised Learning iv. Overview of Regression v. Support Vector Machine vi. Classification algorithms	05	LO3

IV	Unsupervised Learning	i. Introduction to Unsupervised Learning ii. Overview of Clustering iii. Decision Trees iv. Random Forests v. Association	05	LO4
V	Data analytics using Apache Spark	i. Introduction to Apache Spark ii. Architecture of Apache Spark iii. Modes and components iv. Basics of PySpark	04	LO5
VI	Case Studies	i. Understanding the different data science phases used in selected case study ii. Implementation of Machine learning algorithm for selected case study	04	LO1, LO6

Text Books:

1. Jake VanderPlas, “Python Data Science Handbook”, O’Reilly publication
2. Frank Kane, “Hands-On Data Science and Python Machine Learning”, packt publication
3. M.T. Savaliya, R.K. Maurya, G.M.Magar, “Programming with Python”, 2nd Edition, Sybgen Learning.

References:

1. Armando Fandango, “Python Data Analysis”, Second Edition, Packt publication.
2. Alberto Boschetti, Luca Massaron, “Python Data Science Essentials Second Edition”, Packt Publishing
3. Davy Cielen, Arno D. B. Meysman, Mohamed Ali, “Introducing Data Science”, Manning Publications.

Online References:

Sr. No.	Website/Reference link
1.	https://www.w3schools.com/python/pandas/default.asp
2.	https://matplotlib.org/stable/gallery/index.html
3.	https://seaborn.pydata.org/examples/index.html
4.	https://docs.scipy.org/doc/scipy/reference/linalg.html#module-scipy.linalg
5.	https://scikit-learn.org/stable/auto_examples/index.html
6	https://www.tutorialspoint.com/scipy/scipy_integrate.htm
7	https://machinelearningmastery.com/statistical-hypothesis-tests-in-python-cheat-sheet/
8	https://data-flair.training/blogs/data-science-project-ideas/

Suggested List of Experiments

For the following Experiments, use any available data set or download it from Kaggle/UCI or other repositories and use Python to solve each problem.

1. Data preparation using NumPy and Pandas
 - a. Derive an index field and add it to the data set.
 - b. Find out the missing values.
 - c. Obtain a listing of all records that are outliers according to the any field. Print out a listing of the 10 largest values for that field.
 - d. Do the following for the any field.
 - i. Standardize the variable.
 - ii. Identify how many outliers there are and identify the most extreme outlier.

2. Data Visualization / Exploratory Data Analysis for the selected data set using Matplotlib and Seaborn
 - a. Create a bar graph, contingency table using any 2 variables.
 - b. Create normalized histogram.
 - c. Describe what this graphs and tables indicates?
3. Data Modeling
 - a. Partition the data set, for example 75% of the records are included in the training data set and 25% are included in the test data set. Use a bar graph to confirm your proportions.
 - b. Identify the total number of records in the training data set.
 - c. Validate your partition by performing a two-sample Z-test.
4. Implementation of Statistical Hypothesis Test using Scipy and Sci-kit learn [Any one]

1. Normality Tests
 1. Shapiro-Wilk Test
 2. D'Agostino's K^2 Test
 3. Anderson-Darling Test
2. Correlation Tests
 1. Pearson's Correlation Coefficient
 2. Spearman's Rank Correlation
 3. Kendall's Rank Correlation
 4. Chi-Squared Test
3. Stationary Tests
 1. Augmented Dickey-Fuller
 2. Kwiatkowski-Phillips-Schmidt-Shin
4. Parametric Statistical Hypothesis Tests
 1. Student's t-test
 2. Paired Student's t-test
 3. Analysis of Variance Test (ANOVA)
 4. Repeated Measures ANOVA Test
5. Nonparametric Statistical Hypothesis Tests
 1. Mann-Whitney U Test
 2. Wilcoxon Signed-Rank Test
 3. Kruskal-Wallis H Test
 4. Friedman Test

5. Regression Analysis
 - a. Perform Logistic Regression to find out relation between variables.
 - b. Apply regression Model techniques to predict the data on above dataset
 6. Classification modelling
 - a. Choose classifier for classification problem.
 - b. Evaluate the performance of classifier.
 7. Clustering
 - a. Clustering algorithms for unsupervised classification.
 - b. Plot the cluster data.
 8. Using any machine learning techniques using available data set to develop a recommendation system.
 9. Exploratory data analysis using Apache Spark and Pandas
 10. Batch and Streamed Data Analysis using Spark
 11. Implementation of Mini project based on following case study using Data science and Machine learning [Any one]
-

List of Case Studies		
Fake News Detection	Road Lane Line Detection	Sentiment Analysis
Detecting Parkinson's Disease	Brain Tumor Detection with Data Science	Leaf Disease Detection
Speech Emotion Recognition	Gender Detection and Age prediction	Diabetic Retinopathy
Uber Data Analysis	Driver Drowsiness detection	Chatbot Project
Credit Card Fraud Detection	Movie/ Web Show Recommendation System	Customer Segmentation
Cancer Classification	Traffic Signs Recognition	Exploratory Data Analysis for Housing price prediction
Coronavirus visualizations	Visualizing climate change	Predictive policing
Uber's pickup analysis	Earth Surface Temperature Visualization	Web traffic forecasting using time series
Pokemon Data Exploration	Impact of Climate Change on Global Food Supply	Used Car Price Estimator
Skin Cancer Image Detection	World University Rankings	and so on

Assignments:

- 1) Recent trends in Data science
- 2) Comparative analysis between Batch and Streamed data processing tools like Map-reduce, Apache spark, Apache Flink, Apache Samza, Apache Kafka and Apache Storm.

Term Work:

- Term work shall consist of at least 10 experiments and a case study.
- Journal must include 2 assignments.
- The final certification and acceptance of term work indicates that performance in laboratory work is satisfactory and minimum passing marks may be given in term work.
- The distribution of marks for term work shall be as follows:
- Laboratory work (Experiments) (15) Marks.
- Mini project (Implementation) (05) Marks.
- Attendance..... (05) Marks

TOTAL:.....(25) Marks.

Oral examination will be based on Laboratory work, mini project and above syllabus.