

Department of Statistics

Revised Syllabus of II Year Advanced Diploma Program (PG)

Title of Program: Advanced Diploma on Python Software

Syllabus Structure (PG)

Year	Semester	Course No.	Course Code	Contact Hours	Credits (1Credit=15 H)	Total Marks	
2	III	CT III	ADST 303	30	2	75	
		CL III	ADSL303	60	2	150	
	IV	CT IV	ADST 404	30	2	75	
		CL IV	ADSL404	60	2	150	
	Annual	CP II	ADSP202	60	2	150	
	Industrial and or Incubation and or Research and or Field Training				60	2	-
	Total				270	12	600

AD: Advanced Diploma, S : Statistics, C: Course, T: Theory, L: Lab (Practical), P: Project

Total No. of Papers: 05 (Theory: 02, Practical: 02, Project: 01)

Theory and Practical: Semester, **Project: Annual**

Semester III

**CT-III: DST 303: Title: Python for Machine Learning
(Contact Hrs: 30 Credits: 2)**

Learning Objectives:

Students will be able to

1. Implement Classification and Regression algorithm using Python
2. Perform unsupervised learning through python

Unit I: Multivariate Analysis

(15)

Discrimination and Classification, Cluster Analysis, Hierarchical methods, Complete, Average linkage method and non-hierarchical clustering methods K means clustering. Principal component analysis and related results, introduction to factor analysis

Unit II: Supervised and Unsupervised learning

(15)

Regression - Multiple linear regression, Tree based regression model, Support vector regression
Classification - Logistic regression, Naïve-Bayes, KNN, Decision tree, Random forest Xgboost, SVM , ANN

Learning Outcomes:

After completion of the unit, the Student is able to

1. Implement Machine Learning algorithm in Python Software
2. Applied Multivariate Analysis techniques on Data

Reference Books:

1. Applied Univariate, Bivariate and Multivariate Statistics Using Python, Daniel J. Denis, Willey, 2021
2. An Introduction to Statistics with Python, Hnslwanter, Thomas, Springer, 2016
3. In Introduction to Machine Learning with Python, Andreas C. Muller & Sarah Guido, OREILLY
4. Python for Data Analysis, Wes McKinney, OREILLY, 2017

**CL-III: DSL303: Title: Practical - III
(Contact Hrs: 60 Credits: 02)**

Learning Objectives:

Students will be able to

1. Understand use python to perform statistical analysis
2. Solve real life problem by implanting ML algorithm in Python

List of Practical's (15)

1. Hierarchical Clustering
2. K – mean Clustering
3. Principal component Analysis
4. Factor Analysis
5. Multiple Linear Regression
6. Decision tree Regression
7. Random Forest Regression
8. Xgboost Regression
9. Support Vector Regression
10. Logistic Regression
11. Decision Tree Classification
12. Random Forest Classification
13. Xgboost Classification
14. Support Vector Classification
15. Artificial Neural Network

Learning Outcomes:

After completion of the unit, Student is able to

1. Implement Classification algorithms in Python
2. Implement Regression algorithms in Python

Reference Books:

1. Applied Univariate, Bivariate and Multivariate Statistics Using Python, Daniel J. Denis, Willey, 2021
2. An Introduction to Statistics with Python, Hnslwanter, Thomas, Springer, 2016
3. In Introduction to Machine Learning with Python, Andreas C. Muller & Sarah Guido, OREILLY
4. Python for Data Analysis, Wes McKinney, OREILLY, 2017

Semester IV

**CT-IV: DST 404: Title: Python for NLP and Image Processing
(Contact Hrs: 30 Credits: 2)**

Learning Objectives:

Students will be able to

1. Understand Natural Language Processing using Python
2. Classify Image Data into different classes

Unit I: Natural Language Processing(NLP)

(15)

Natural Language Toolkit (NLTK)in Python. Data Preparation: punctuation removal, stop-words removal, numeric value removal, frequent words removal, rare words removal, spelling correction, tokenization, stemming, lemmatization. Feature Engineering: count vectors as features, Term Frequency-Inverse Document Frequency (TF-IDF), TF-IDF vectors as features, word level TF-IDF, N-Gram level TF-IDF, Character level TF-IDF, Inverse Document Frequency, word embedding as features, Text/NLP based features, Topic Models as features, word2vec.

Unit II: Opencv Librray In Python

(15)

Getting Started with images, Basic Operations on Images, Arithmetic Operations on Images, Image Preprocessing: changing color spaces, geometric transformations, thresholding, smoothing, morphological transformations, gradients, Canny Edge detection, image pyramids, image segmentation with Watershed algorithm, Feature Detection and Description. Image Detection and recognition examples.

Learning Outcomes:

After completion of the unit, the Student is able to

1. Build a Text Classification Model in Python software
2. Implement Image Classification Model through python software

Reference Books:

1. Natural language processing with Python: analyzing text with the natural language toolkit. " Bird, S., Klein, E., & Loper, E. (2009).
2. Deep Learning for Computer Vision: Expert techniques to train advanced neural networks using TensorFlow and Keras. Shanmugamani R. (2018). Packt Publishing Ltd
3. Python for data science for dummies. Mueller, J. P., & Massaron, L. (2015). John Wiley & Sons.
4. Natural Language Processing: Python and NLTK. Hardeniya, N., Perkins, J., Chopra, D., Joshi, N., & Mathur, I. (2016). Packt Publishing Ltd.

**CL-IV:DSL404: Title (Practical): Practical – IV
(Contact Hrs: 60 Credits: 02)**

Learning Objectives:

Students will be able to

1. Solve real-life problem by implementing text classification algorithms
2. Perform sentiment analysis on text data

List of Practical's (15)

1. Regular Expression
2. Optical Character Recognition
3. Data Preparation – Count Vector
4. Data Preparation – TF-IDF
5. Word Cloud Creation
6. Sentiment Analysis
7. Text Classification
8. Part of Speech tagging
9. Name Entity Recognition
10. Basic operation Image
11. Image Processing
12. Canny Edge Detection
13. Image Classification
14. Image Recognition
15. Object Detection

Learning Outcomes:

After completion of the unit, Student is able to

1. Build word cloud on text data
2. Implement object detection model using Python

Reference Books:

1. Natural language processing with Python:analyzing text with the natural language toolkit. " Bird, S., Klein, E., & Loper, E. (2009).
2. Deep Learning for Computer Vision: Expert techniques to train advanced neural networks using TensorFlow and Keras. Shanmugamani R. (2018). Packt Publishing Ltd
3. Python for data science for dummies. Mueller, J. P., & Massaron, L. (2015). John Wiley & amp; Sons.
4. Natural Language Processing: Python and NLTK. Hardeniya, N., Perkins, J., Chopra, D., Joshi, N., & Mathur, I. (2016). Packt Publishing Ltd.

**CP-II: DSP202: Project
(Contact Hrs. 60, Credits: 2)**

**Industrial and or Incubation and or Research and or Field Training
(Contact Hrs. 60, Credits: 2)**

BOS Sub-Committee

1. Dr. S. M. Nimbale Chairman
2. Dr. D. S. Jadhav Member

Expert committee

1. Dr. S. P. Gite
Head, Dept. of Statistics,
Mumbai university, Mumbai
2. Mr. S. M. Nevase,
Senior Data Scientist,
Birlasoft, Pune