



**JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY ANANTAPUR**

**COLLEGE OF ENGINEERING:: ANANTHAPURAMU (Autonomous)**

**TEOIP III**

1. **Name of the Staff:** Dr. C. Shoba Bindu
2. **Designation:** Professor.
3. **Department:** Computer Science and Engineering.
4. **Contact No:** +91 8143289089.
5. **Name of the Conference/Workshop/Laboratory:** Data Analytics and Machine Learning using R and Python – Phase-II
6. **Venue:** JNTUA College of Engineering, Ananthapuramu.

**Date:** 3<sup>rd</sup>-August-2020 to 08- August-2020.

**Organized by:** Department of Computer Science and Engineering, JNTUA CEA in association with IEEE Ananthapuramu Subsection & IEEE CIS/GRSS Joint Chapter, Hyderabad.

7. **AICTE Sponsored One Week Short Term Training Programme (STTP) on:** Data Analytics and Machine Learning using R and Python – Phase-II
- 8.

**REPORT**

**Brief Description of Event / Visit:**



Python is a multi-purpose programming language which is simple to use which can be used for web development, machine learning, browser automation etc. It is introduced in many colleges to make the students industry ready. In this context, this workshop was planned to train the students from various Engineering Colleges by inviting experts in this field to impart knowledge from them.

Machine Learning is continually transforming the day to day lives of people and has become so pervasive today. It is an application of Artificial Intelligence. The continuous research and developments in the field of machine learning has given us self-driving cars, practical speech recognition, effective web search, weather prediction and many others. Python is the new Fortran in the scientific world and this language is great to use when working with machine learning algorithms and has easy syntax relatively.

This workshop begins with a primary introduction to basic machine learning methods. Supervised and unsupervised learning methods will be discussed in detail with its applications. New trends like Deep Learning, evolutionary optimization and soft computing also will be explored. One day complete hands-on sessions are planned on introduction to Python programming, machine learning library scikit-learn and demonstration of deep learning concepts. The workshop will conclude with a discussion on a number of open problems, which the participants will be encouraged to take up as their future research endeavor.

**REGISTRAR**  
**J.N.T.U. Anantapur**  
**ANANTAPURAMU-515002**

**Program Schedule:**

	<b>DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING</b> <b>Short Term Training Programme</b> <b>On</b> <b>Data Analytics and Machine Learning using R and Python - Phase-II</b> <b>(03 ~ 08, August 2020)</b>	
---	---	---

DATE/TIME	10:00 AM – 11:30 AM		11:45 AM – 01:15 PM		02:15 PM – 03:45 PM		04:00 PM – 05:30 PM
03-08-2020	Inauguration & Welcome Address	Tea Break	Digital Image Processing: Basics to Advances (Prof. S. Srinivas Kumar)	Lunch Break	Introduction to Data Science (Dr. C. Shoba Bindu)	Tea Break	Applications of Linear Algebra in Machine Learning (Dr. E. Keshava Reddy)
04-08-2020	Data Analytics using R (Dr. T. Sree Ganesh)		Data Wrangling (Dr. Salman Abdul Moiz)		Exploratory Data Analysis (Dr. Salman Abdul Moiz)		Data Analytics using Python (Dr. T. Sree Ganesh)
05-08-2020	Role of Big Data in Healthcare (Prof. V. Chandrasekaran)		Pattern Recognition Principles & Approaches (Prof. V. Chandrasekaran)		Bayes, Decision Trees (Dr. T. Hitendra Sarma)		KNN, K-means, Kernel K-Means and its improvements (Dr. T. Hitendra Sarma)
06-08-2020	Object Detection using Faster RCNN (Dr. V. Anantha Natarajan)		Hands on Session (Dr. V. Anantha Natarajan)		Sequential Models (Dr. Chandrajit Pal)		Identifying Suitable Model to fit the given use case and Modeling the Business case : Property_Value_prediction (Dr. K. Dhana Sree Devi)
07-08-2020	Statistical Learning (Debasis Samanta)		Regression and correlation analysis (Debasis Samanta)		Reinforcement Learning (Dr. Chandrajit Pal)		Use case: Text mining (Dr. K. Dhana Sree Devi)
08-08-2020	Probabilistic classification (Debasis Samanta)		Geospatial Deep Learning using Python (Dr. Raghavendra Kune)		Hands-on Training of CNN models (Dr. Chandrajit Pal)		Assessment Test & Valedictory

**DAY 1-SESSION 1&2&3:**

Digital Image Processing: Basics to Advances (Prof. S. Srinivas Kumar)

Digital Image Processing is a field that focuses on the manipulation of digital images using various algorithms and techniques. It plays a crucial role in several applications, including medical imaging, computer vision, and image analysis. Some key points to consider for your report on this topic might include:

**Introduction to Digital Image Processing:** Explain the basic concepts of digital image processing, including image representation, pixel values, and image enhancement.

**Image Filtering:** Discuss techniques like convolution and filtering, which are used for image enhancement and noise reduction.

**Image Transformation:** Cover topics such as Fourier Transform and image registration, which are essential for frequency domain analysis and image alignment.

**Feature Extraction:** Explain how features are extracted from images for pattern recognition and object detection tasks.

**Applications:** Highlight real-world applications of digital image processing, such as medical image analysis, facial recognition, and satellite image processing.

  
**REGISTRAR**  
**J.N.T.U. Anantapur**  
**ANANTAPURAMU-515002**

## Introduction to Data Science (Dr. C. Shoba Bindu)

Data Science is an interdisciplinary field that combines various techniques from statistics, computer science, and domain knowledge to extract insights and knowledge from data. Here are some key points to include in your report:

**Data Collection and Cleaning:** Discuss the importance of data collection and data preprocessing, which involve handling missing values and outliers.

**Exploratory Data Analysis:** Explain how data scientists use techniques like data visualization and statistical analysis to gain insights from the data.

**Machine Learning:** Introduce machine learning algorithms and their applications in data science, including supervised and unsupervised learning.

**Big Data:** Discuss the challenges and tools associated with processing and analyzing large datasets, including technologies like Hadoop and Spark.

**Data Science Applications:** Provide examples of data science applications in various industries, such as finance, healthcare, and e-commerce.

## Applications of Linear Algebra in Machine Learning (Dr. E. Keshava Reddy)

Linear algebra is a fundamental mathematical tool in machine learning and data analysis. Here are some key points for your report on the applications of linear algebra in machine learning:

**Vectors and Matrices:** Explain the concepts of vectors and matrices, which are fundamental to representing data and transformations in machine learning.

**Matrix Operations:** Discuss common matrix operations such as matrix multiplication, transpose, and inversion, and their relevance in machine learning algorithms.

**Eigenvalues and Eigenvectors:** Explain how eigenvalues and eigenvectors are used in dimensionality reduction techniques like Principal Component Analysis (PCA).

**Singular Value Decomposition (SVD):** Discuss the importance of SVD in matrix factorization and recommendation systems.

**Neural Networks:** Show how linear algebra is used in the architecture and computations of neural networks, including feedforward and backpropagation.

**Applications:** Provide examples of how linear algebra is applied in real-world machine learning tasks, such as image recognition and natural language processing.

### **DAY 2-SESSION1,2, 3&4:**

## Data Analytics using R (Dr. T. Sree Ganesh)

R is a popular programming language and environment for statistical computing and graphics. It is widely used for data analysis, data visualization, and statistical modeling. Here are some key points to include in your report:

**Introduction to R:** Provide an overview of the R programming language, its history, and its importance in data analytics.

**Data Import and Manipulation:** Explain how R is used to import and manipulate data from various sources, including CSV files, databases, and web scraping.

  
**REGISTRAR**  
**J.N.T.U. Anantapur**  
**ANANTAPURAMU-515002**



**Data Visualization:** Discuss the powerful data visualization capabilities of R, including packages like ggplot2 for creating informative charts and plots.

**Statistical Analysis:** Describe how R is used for statistical analysis, hypothesis testing, and regression modeling.

**Packages and Libraries:** Highlight essential R packages and libraries commonly used for data analytics, such as dplyr, tidyr, and caret.

**Case Studies:** Include real-world case studies or examples that demonstrate the application of R in data analytics projects.

**Data Wrangling (Dr. Salman Abdul Moiz)**

Data wrangling refers to the process of cleaning, transforming, and preparing raw data into a format suitable for analysis. Here are key points to cover in your report:

**Data Cleaning:** Explain the importance of data cleaning, which involves handling missing values, outliers, and inconsistencies in datasets.

**Data Transformation:** Describe how data is transformed to meet the requirements of a specific analysis, including techniques like normalization and standardization.

**Data Integration:** Discuss methods for integrating data from multiple sources and merging datasets.

**Data Reshaping:** Explain how to reshape data using techniques like pivoting, melting, and stacking.

**Automated Data Wrangling Tools:** Mention tools and libraries used for automating data wrangling processes, such as pandas in Python or dplyr in R.

**Exploratory Data Analysis (Dr. Salman Abdul Moiz)**

Exploratory Data Analysis (EDA) is the process of visually and statistically summarizing and exploring datasets to understand their underlying structure and patterns. Here are key points for your report:

**EDA Techniques:** Describe common EDA techniques, including data visualization, summary statistics, and distribution analysis.

**Data Visualization:** Discuss the use of various graphs, charts, and plots to explore data, such as histograms, scatter plots, and box plots.

**Data Summarization:** Explain how to calculate and interpret summary statistics like mean, median, and standard deviation.

**Outlier Detection:** Discuss methods for identifying and handling outliers in EDA.

**Correlation Analysis:** Explain how to assess relationships between variables using correlation coefficients.

**Case Studies:** Include examples of EDA in real-world datasets to illustrate the importance and impact of EDA on decision-making.

**Data Analytics using Python (Dr. T. Sree Ganesh)**

Similar to the section on R, provide an overview of data analytics using Python. You can cover topics such as data manipulation with libraries like pandas, data visualization with libraries like Matplotlib and Seaborn, and statistical



analysis with libraries like SciPy.

### **DAY 3-SESSION 1,2,3&4:**

Role of Big Data in Healthcare (Prof. V Chandrasekaran)

The role of big data in healthcare is becoming increasingly important for improving patient care, disease prevention, and healthcare management. Here are some key points to include in your report:

**Introduction to Big Data in Healthcare:** Explain what big data is and why it's crucial in the healthcare sector.

**Data Sources in Healthcare:** Discuss the various sources of healthcare data, including electronic health records (EHRs), medical imaging, wearable devices, and genomics.

**Applications of Big Data:** Highlight how big data analytics is used for disease prediction, treatment optimization, patient monitoring, and epidemiological studies.

**Challenges and Ethical Considerations:** Discuss the challenges associated with handling sensitive patient data and the ethical considerations of using big data in healthcare.

**Real-World Examples:** Include case studies and examples of how big data has made a significant impact in healthcare, such as early disease detection and personalized medicine.

Pattern Recognition Principles & Approaches (Prof. V Chandrasekaran)

Pattern recognition is a fundamental concept in machine learning and data analysis. Here's what you can cover in your report:

**Introduction to Pattern Recognition:** Explain the concept of pattern recognition and its significance in various fields, including healthcare, finance, and image processing.

**Pattern Recognition Approaches:** Discuss different approaches to pattern recognition, such as supervised learning, unsupervised learning, and semi-supervised learning.

**Feature Extraction:** Explain the importance of feature extraction in pattern recognition and provide examples of feature selection techniques.

**Classification Algorithms:** Introduce common classification algorithms used in pattern recognition, such as support vector machines, neural networks, and decision trees.

**Applications in Healthcare:** Highlight how pattern recognition is applied in healthcare, including disease diagnosis, medical image analysis, and patient monitoring.

Bayes, Decision Trees (Dr. T. Hitendra Sarma)

Bayesian methods and decision trees are essential tools in machine learning and data analysis. Here are some points to consider for your report:

  
**REGISTRAR**  
**J.N.T.U. Anantapur**  
**ANANTAPURAMU-515002**

**Bayesian Methods:** Explain the principles of Bayesian probability and how it is used in machine learning for classification and probabilistic modeling.

**Decision Trees:** Introduce decision trees as a popular supervised learning algorithm for classification and regression tasks.

**Bayesian Decision Trees:** Discuss how Bayesian decision trees combine the principles of both Bayesian methods and decision trees for improved model performance.

**Applications:** Provide examples of how Bayesian methods and decision trees are applied in real-world scenarios, including healthcare applications like disease prediction and treatment recommendations.

**KNN, K-means, Kernel K-Means and its Improvements (Dr. T. Hitendra Sarma)**

These are fundamental machine learning algorithms used for clustering and classification. Here are some key points for your report:

**K-Nearest Neighbors (KNN):** Explain how the KNN algorithm works for both classification and regression tasks, including the concept of finding the K nearest data points.

**K-Means Clustering:** Introduce K-means as an unsupervised clustering algorithm and explain how it divides data into clusters based on similarity.

**Kernel K-Means:** Discuss the use of kernel functions to extend K-means for nonlinear data.

**Improvements and Variations:** Mention improvements and variations of these algorithms, such as weighted KNN or hierarchical K-means.

**Applications:** Provide examples of how these algorithms are applied in healthcare, such as patient clustering for personalized treatment plans or disease subtype identification.

#### **DAY 4-SESSION 1,2,3&4:**

**Object Detection using Faster R-CNN (Dr. V. Anantha Natarajan)**

Faster R-CNN (Region-based Convolutional Neural Network) is a deep learning model used for object detection in images. Key points for your report might include:

**Introduction to Object Detection:** Explain the concept of object detection and its applications, such as autonomous vehicles, surveillance, and image captioning.

**Faster R-CNN:** Provide an overview of the Faster R-CNN architecture, which combines convolutional neural networks (CNNs) with region proposal networks (RPNs).

**Training and Inference:** Discuss the training process for Faster R-CNN, including dataset preparation and fine-tuning. Explain how inference is performed to detect objects in images.

**Applications:** Highlight real-world applications of Faster R-CNN, such as pedestrian detection, object tracking, and anomaly detection.

Challenges and Limitations: Discuss the challenges and limitations of object detection using Faster R-CNN, including computational requirements and fine-tuning difficulties.

Hands-on Session (Dr. V. Anantha Natarajan)

This appears to be a practical session related to object detection using Faster R-CNN. You can expect to learn and practice the concepts covered in the Object Detection using Faster R-CNN topic.

Sequential Models (Dr. Chandrajit Pal)

Sequential models refer to a class of machine learning models that are particularly useful for sequential data, such as time series or text. Key points for your report might include:

Introduction to Sequential Models: Explain the importance of sequential data and the need for specialized models.

Recurrent Neural Networks (RNNs): Introduce RNNs as a class of models that can handle sequential data through recurrent connections. Discuss applications like natural language processing and speech recognition.

Long Short-Term Memory (LSTM) Networks: Explain LSTM networks as an improvement over RNNs, capable of capturing long-range dependencies in sequential data.

Gated Recurrent Unit (GRU): Mention GRU as another variant of RNNs with reduced computational complexity compared to LSTMs.

Applications: Discuss real-world applications of sequential models, including time series forecasting, language modeling, and sentiment analysis.

Identifying Suitable Model to Fit the Given Use Case and Modeling the Business Case: Property Value Prediction (Dr. K. Dhana Sree Devi)

This topic seems to focus on the process of selecting an appropriate machine learning model for a specific use case, in this case, property value prediction. Key points for your report might include:

Problem Statement: Describe the problem of property value prediction and its relevance in real estate.

Model Selection: Explain the process of selecting a suitable machine learning model for the given problem, considering factors like data characteristics and business goals.

Data Preprocessing: Discuss the steps involved in preparing and cleaning the data for modeling, including feature engineering and handling missing values.

Model Building: Describe the process of building and training the chosen model for property value prediction. This could involve regression techniques or other predictive modeling approaches.

Model Evaluation: Explain how the model's performance is assessed and validated, including metrics like Mean Absolute Error (MAE) or Root Mean Square Error (RMSE).

Business Impact: Discuss how the model's predictions can be used to make informed business decisions, such as property valuation or investment decisions.

#### **DAY 5 -SESSION 1,2,3&4:**

Statistical Learning (Debasis Samanta)

  
**REGISTRAR**  
**J.N.T.U. Anantapur**  
**ANANTAPURAMU-515002**



Statistical learning refers to the use of statistical techniques to make predictions or infer relationships from data. Key points for your report might include:

**Introduction to Statistical Learning:** Explain the concept of statistical learning, including supervised and unsupervised learning, and their applications.

**Types of Models:** Describe different types of statistical learning models, such as linear regression, logistic regression, and decision trees.

**Model Evaluation:** Discuss methods for evaluating the performance of statistical learning models, including metrics like mean squared error (MSE) and accuracy.

**Cross-Validation:** Explain the importance of cross-validation in model selection and performance assessment.

**Applications:** Provide examples of real-world applications of statistical learning, such as predicting stock prices or customer churn.

#### Regression and Correlation Analysis (Debasis Samanta)

Regression and correlation analysis are statistical techniques used to understand relationships between variables. Key points might include:

**Regression Analysis:** Explain the concept of regression analysis, including linear regression and multiple regression, for predicting outcomes based on input variables.

**Correlation Analysis:** Discuss how correlation analysis measures the strength and direction of relationships between two variables.

**Model Interpretation:** Explain how to interpret regression coefficients and correlation coefficients.

**Use Cases:** Provide examples of when and how regression and correlation analysis are used in real-life scenarios, such as market research or medical studies.

#### Reinforcement Learning (Dr. Chandrajit Pal)

Reinforcement learning is a type of machine learning where an agent learns to make decisions by interacting with an environment. Key points for your report might include:

**Introduction to Reinforcement Learning:** Explain the basic concepts of reinforcement learning, including agents, environments, and rewards.

**Markov Decision Processes (MDPs):** Discuss the mathematical framework of MDPs that underlies reinforcement learning.

**Q-Learning and Policy Gradient Methods:** Introduce common reinforcement learning algorithms like Q-learning and policy gradients.

**Applications:** Provide examples of applications of reinforcement learning, such as game playing (e.g., AlphaGo), robotics, and autonomous driving.

#### Use Case: Text Mining (Dr. K. Dhana Sree Devi)

Text mining involves extracting useful information and insights from text data. Key points for your report might include:

Introduction to Text Mining: Explain what text mining is and why it's important, especially in dealing with large volumes of unstructured text data.

Text Preprocessing: Discuss the steps involved in text preprocessing, including tokenization, stemming, and stop word removal.

Text Classification: Explain how text classification is used to categorize text documents into predefined categories.

Sentiment Analysis: Describe sentiment analysis as a specific application of text mining for understanding opinions and emotions expressed in text.

Use Cases: Provide examples of use cases for text mining, such as sentiment analysis for customer reviews, topic modeling for document clustering, and information retrieval.

### **DAY 5 -SESSION 1,2,3:**

Probabilistic Classification (Debasis Samanta)

Probabilistic classification is a machine learning technique that assigns class labels to data points while providing a probability score indicating the confidence in the assigned label. Key points might include:

Introduction to Probabilistic Classification: Explain the concept of probabilistic classification and its applications in machine learning.

Probabilistic Models: Discuss different probabilistic models used for classification, such as Naive Bayes, logistic regression, and support vector machines.

Probabilistic Scores: Explain how probability scores are generated for each class, and how they can be used for decision-making and model evaluation.

Applications: Provide examples of real-world applications of probabilistic classification, such as spam email detection or medical diagnosis.

Geospatial Deep Learning using Python (Dr. Raghavendra Kune)

Geospatial deep learning involves using deep neural networks for tasks related to geographic and spatial data. Key points might include:

Introduction to Geospatial Deep Learning: Explain the significance of geospatial data and how deep learning techniques can be applied to analyze such data.

Spatial Data Processing: Discuss techniques for processing and handling spatial data, including geospatial libraries in Python.

Convolutional Neural Networks (CNNs) for Geospatial Data: Explore how CNNs are adapted for geospatial data analysis, including applications like remote sensing and GIS.

Applications: Provide examples of applications of geospatial deep learning, such as land cover classification, object detection in satellite imagery, and disaster monitoring.

Hands-on Training of CNN Models (Dr. Chandrajit Pal)

This topic appears to be a practical session on training Convolutional Neural Network (CNN) models, likely building on the

  
**REGISTRAR**  
**J.N.T.U. Anantapur**  
**ANANTAPURAMU-515002**

previous topic. Expect to learn and practice implementing CNN models during this session.

#### Assessment Test & Valedictory

The assessment test is likely a way to evaluate your understanding of the topics covered in the course or workshop. The valedictory session is typically a closing ceremony where participants receive certificates or acknowledgments for completing the course.

For the assessment test, it's important to review the materials and concepts covered in the course to prepare adequately. You may expect questions, quizzes, or assignments to assess your knowledge.

During the valedictory session, there may be speeches, awards, and acknowledgments for participants and instructors. It's an opportunity to celebrate the completion of the course or workshop.

#### **Photographs:**



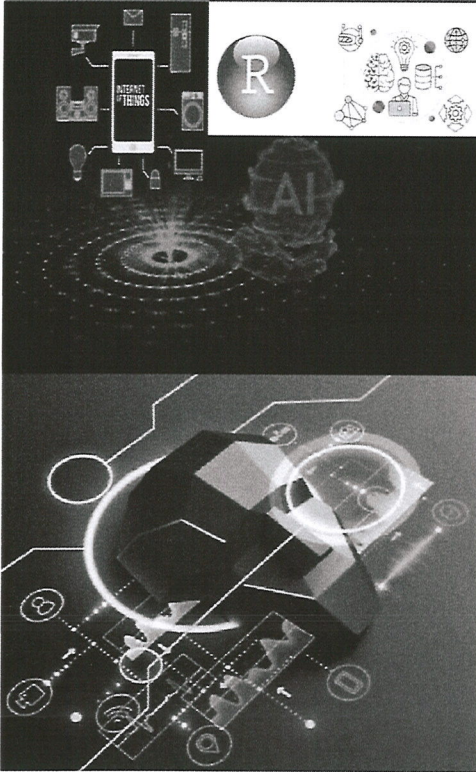
  
**REGISTRAR**  
**J.N.T.U. Anantapur**  
**ANANTAPURAMU-515002**



**AICTE Sponsored One Week Short Term  
Training Programme (STTP)**

on

**Data Analytics and Machine Learning using  
R and Python – Phase-II**  
August 3<sup>rd</sup> to 8<sup>th</sup>, 2020



**CHIEF PATRON**

Prof. S. Srinivas Kumar  
Hon'ble Vice-Chancellor, JNTUA, Ananthapuramu

**PATRONS**

Prof. M. Vijay Kumar  
Registrar, JNTUA, Ananthapuramu  
Prof. K. Govinda Rajulu  
Principal, JNTUACEA

**CO-PATRON**

Prof. P.Sujatha  
Vice-Principal, JNTUACEA

**CONVENER**

Prof. A. Suresh Babu  
HoD Computer Science and Engineering, JNTUACEA

**ORGANIZING COMMITTEE**

**Coordinator**

Prof. C. Shoba Bindu  
Department of Computer Science and Engineering  
JNTUACEA, Ananthapuramu

**MEMBERS**

Prof. A. Ananda Rao, Director R&D, JNTUA  
Prof. B. Eswara Reddy, Director SDC, JNTUA  
Prof. P. Chenna Reddy, Professor of CSE  
Prof. S. Vasundra, NSS Coordinator, JNTUA  
Dr. K. Madhavi, Associate Professor of CSE  
Dr. R. Rajasekhar, Associate Professor of CSE  
Dr. K. F. Bharati, Assistant Professor of CSE  
Dr. B. Lalitha, Assistant Professor of CSE  
Dr. A. P. Siva Kumar, Assistant Professor of CSE

**RESOURCE PERSONS**

Experienced faculty from renowned universities/premier institutes, industry experts and host Institution.



**AICTE Sponsored One Week Short Term  
Training Programme (STTP)**

on

**Data Analytics and Machine Learning using  
R and Python – Phase-II**  
August 3<sup>rd</sup> to 8<sup>th</sup>, 2020

*Organized by*

**Department of Computer Science and  
Engineering**



**JNTUA College of Engineering (Autonomous)  
Ananthapuramu**

Web Site: <http://jntuacee.ac.in/fdp2020>

Online Registration:

<https://forms.gle/AZwwiWQEaKEqvX3p9>



  
**REGISTRAR**  
**J.N.T.U. Anantapur**  
**ANANTAPURAMU-515002**