



LAKSHYA INSTITUTE OF TECHNOLOGY

BCA, BSc ITM and BSc CS

Affiliated to Utkal University

SHANTI SWAROOP

Ph - 8984363414



Machine Learning / Data Science with Bigdata and Python

Python Basics

- Introduction to Python
- Variables and Data Types
- Control Structures (if, else, loops)
- Functions and Modules
- File Handling
- Exception Handling

Pandas

- Introduction to Pandas
- Series
- Data Frame
- Grouping
- Ranking
- Slicing
- Apply, Assign, Lambda, Map
- Concatenating and Merging or joining Data Frames

Multivariate Regression

- Understanding Regression Analysis
- Multivariate Regression Techniques
- Implementation in Python





LAKSHYA INSTITUTE OF TECHNOLOGY

Affiliated to Utkal University

SHANTI SWAROOP

Ph - 8984363414

Machine Learning / Data Science with Bigdata and Python

Evaluation Metrics and Confusion Matrix

- Precision, Recall, F1 Score
- ROC-AUC Curve
- Confusion Matrix and its interpretation

Lasso , Ridge

- Regularization Techniques in Machine Learning
- Lasso Regression
- Ridge Regression

Logistic Regression

- Binary and Multinomial Logistic Regression
- Odds Ratio and Log Odds
- Logistic Regression in Python

Feature Selection

- Importance of Feature Selection
- Techniques for Feature Selection

Normalization, Corelation, Variance Bias-Variance Trade-off

- Normalization and Standardization
- Corelation Analysis
- Variance, Bias, and Bias-Variance Trade-off

Decision Tree

- Decision Tree for Classification
- Decision Tree for Regression
- Implementation in Python





LAKSHYA INSTITUTE OF TECHNOLOGY

Affiliated to Utkal University

SHANTI SWAROOP

Ph - 8984363414

Machine Learning / Data Science with Bigdata and Python

Support Vector Machine

- Linear SVM
- Soft margin SVM
- Kernels
- SVM Regression

Bagging / Random Forest

- Introduction to Bagging
- Random Forest Algorithm
- Implementation in Python

Perceptron

- Introduction to Perceptron
- Perceptron Algorithm

Neural Network

- Introduction to Neural Networks
- Feedforward and Backpropagation
- Implementation in Python

Clustering

- K-means
- K-modes
- Density-based Clustering
- Hierarchical Clustering

Principal Component Analysis

- Dimensionality Reduction
- PCA in Python





LAKSHYA INSTITUTE OF TECHNOLOGY

Affiliated to Utkal University

SHANTI SWAROOP

Ph - 8984363414

Machine Learning / Data Science with Bigdata and Python

Outliers Detection

- Identification and Handling of Outliers

SMOTE Imbalance Data Set

- Techniques to Handle Imbalanced Datasets

FP Growth

- Frequent Pattern Growth Algorithm
- Association Rule Mining

Collaborative Filtering

- Matrix Factorization
- K Nearest Neighbors

Boosting

- Ada Boost
- Gradient Boosting Machines

Deep Learning

- Feed Forward Network
- Backpropagation
- Versions of Gradient Descent
- Recurrent Neural Network (RNN)
- Long Short-Term Memory (LSTM)
- Convolutional Neural Networks (CNN)
- Attention Mechanism
- Transformer Model
- Generative Adversarial Network (GAN)
- Implementation in Google Collab





LAKSHYA INSTITUTE OF TECHNOLOGY

Affiliated to Utkal University

SHANTI SWAROOP

Ph - 8984363414

BIG DATA

Introduction to Pyspark

- Overview of Pyspark
- Setting up Pyspark Environment

Distributer Spark Programming

- Understanding Distributed Computing
- Spark Programming Basics

RDD (Resilient Distributed Datasets)

- Introduction to RDD
- RDD Operations

Partitioning

- Data Partitioning in Spark

Transformation / Actions

- Transformation and Action Operations on RDD

Paired RDD

- Operations on Paired RDD

Join in RDD

- Different types of Joins in RDD

Shuffling

- Shuffling in Spark

Distributed Variable

- Accumulator and Broadcast Variable in Spark





LAKSHYA INSTITUTE OF TECHNOLOGY

Affiliated to Utkal University

SHANTI SWAROOP

Ph - 8984363414

BIG DATA

SQL Basics

- Basics of SQL in Spark
- Group by, Joins, Subquery, Correlated Subquery
Inline View, Window Function

Data Frame

- Introduction to Spark DataFrames
- Operations on DataFrames

Operations on Data Frame

- Creation
- Joins
- Explode
- Flat Map
- Aggregations
- Coalesce , Repartition
- User Defined Functions (UDFs)

Data Skewness and Salting Technique

- Understanding Data Skewness
- Salting Technique in Spark

Data Sets

- Introduction to Spark Datasets

Spark ML

- Machine Learning in Spark
- Spark MLlib Overview





LAKSHYA INSTITUTE OF TECHNOLOGY

Affiliated to Utkal University

SHANTI SWAROOP

Ph - 8984363414

BIG DATA

Hadoop Architecture

- Basics of Hadoop
- Hadoop Ecosystem

Hadoop Map Reduce

- Introduction to MapReduce
- MapReduce Programming Model

Reduce Side Join

- Understanding Reduce Side Join in Hadoop

HDFS Commands

- Basic Hadoop Commands



Address: M4/46, Acharya Vihar, Bhubaneswar-751013, Odisha
Ph – 0674-2544690 , www.litindia.ac.in