Sandip Foundation's
Sandip Institute of Technology & Research Center, Nashik.

<u>E-BULLETIN MAY-2016</u>

Staff Coordinator:-
Prof. P.R.Pachorkar

Student Coordinator:-
Mr. Pratik Takle

Day 1: Big Data Analytics session inaugurated with a
great zeal along with sessions on
"Data Science Basics"

Inauguration Event for 10 days Department of Science and Technology, Govt. of India

sponsored training programme started with lightening the Lamp in front of Saraswati Devi (Godess of Knowledge).

Dignitories present on the Dias included:

◦ Prof. (Dr.) S. G. Bhirud, Professor, Computer Engineering Department, VJTI, Mumbai.
(Chief Guest)

◦ Prof. (Dr.) V. M. Thakare, Professor and Head in Computer Science, Faculty of Engineering & Technology, Post Graduate Department of Computer Science, SGB Amravati University, Amravati(Guest of Honour)

◦ Prof. (Dr.) S. T. Gandhe, Principal, SITRC, Nashik

◦ Prof. (Dr.) A. G. Jadhav, Principal, SIPS, Nashik

◦ Prof. S. D. Pawar, Principal, SIP, Nashik

◦ Prof. A. D. Potgantwar, Head, Department of Computer Engineering and Information Technology, SITRC, Nashik

◦ Prof. (Mrs.) N.D. Ghuse, Convener, Big Data Analytics Training Programme2016, SITRC, Nashik

Felicitation of Chief Guest and Guest of Honour done by Prof. (Dr.) S. T. Gandhe and Prof. (Dr.) A. G. Jadhav respectively along with intorduction to their portfolio by Prof. (Mrs.) S. N. Patil as follows:

◦ Dr. S. G. Bhirud, Professor, Computer Engineering Department, VJTI, Mumbai.

Pursued Ph.D. in EC and Computer Science from Swami Ramanad Teerth Marathwada University, Nanded in 2001 and continuted to serve the education sector with all of possbile skills. Sir has numerous publications in Electronics and Computer, Image processing, Neural network and Digital Signal processing.

AICTE's AdviserI for eGOVERNANCE CELL and LEGAL CELL. Also AICTE's AdvisorI and Chief Vigilance Officer for VIGILANCE CELL.

His continous contribution in diversified research fields motivates all of us for thinking

about the current buzzwords SMAC(Social, Mobility, Analytics and Cloud).
◦

Dr. V. M. Thakare, Professor and Head in Computer Science, Faculty of Engineering & Technology, Post Graduate Department of Computer Science, SGB Amravati University, Amravati

Awarded with 2 UGC fellowships in 10th Plan of Govt of India.

Statistical Views on his Career Excellence:

21+ years of Experience

Member of Expert Committee, Advisory Board, Board of Studies and Selection Committees in eminent universities across the nation.

10+ Research Scholars awarded Ph.D.

Guided 300+ projects at M.E./M.S./M.Phil./M.C.A. level

85 + Research Paper published in Internatinal Journals and International/National Conferences

55+ Keynote addresses and invited talks delivered in India and overseas.

Day 2: The Training Intensifies with Heavy Expert Lectures

Handouts from Prof N M Shahane's Expert Talk:

Contents to Cover:

a) Linear regression modeling and diagnostics

b) Multiple linear regression modeling

c) Logistic regression and binary classification

The use of probability to measure uncertainty and variability dates back hundreds of years.

Probability has found application in areas as diverse as medicine, gambling, weather forecasting, and the law.

Handouts From Dr. Parikshit N. Mahalle's Expert Talk:

DATA MODEL PERFORMANCE-"ROC PERSPECTIVE"

IN 2015   TODAY, A PETABYTE OF DATA 1,024 TERABYTES, TO BE EXACT PROBABLY MEETS MANY PEOPLE'S DEFINITION OF "BIG DATA."

IN 2025   FASTFORWARD 10 YEARS, HOWEVER, AND A PETABYTE NO LONGER WILL QUALIFY AS BIG AT LEAST NOT IN THE ENTERPRISE.

DATA EXPLOSION

Each day the following happens:

There are 1440 minutes per day...that means there are approximately

• 294 BILLION emails sent every day!

• 6 BILLION Google Searches each day!

• 3.5 BILLION Facebook messages posted daily!

• 40 Million Tweets shared each day!

• Average Retailer Generates 2.5 Petabytes Per Day

Source: https://www.gwava.com/blog/internetdatacreateddaily2014

When the size of the data, itself, becomes a problem

When the "old way" of processing data just doesn't work effectively

It's "big" when we have to rethink:

• How we store that much data?

• How we move that much data?

• How we extract, load & transform that much data?

• How we explore and analyze that much data?

• How we process and get meaningful insights from that much data?

• Data Ingestion, Storage, Processing, Reporting and Action

DAY 4   Live Demonstration of Finding structures in data through Clustering

Handouts from Prof Dr S.R. Dhore Expert Talk:

Contents to Cover:

a) Clustering methods

• Kmeans, Partitioninng around medoids(PAM),
• Visualization of clustering results,
• clustering evaluation using Silhouette coeficients and other indices,
• BIRCH clustering for large datasets, Anomoly detection, Finding frequent Itemsets using APriori Algorithm and variants, e.g.SON; Basics of networks centrality measures, Network link analysis, PageRank algorithm.

b) Outlier analysis

• Anomoly detection,
• Finding frequent Itemsets using APriori Algorithm and variants,

c) Association analysis

• e.g.SON

d) Network analysis and optimization

•
•
•

Basics of networks centrality measures,
Network link analysis
 PageRank algorithm.

What is Clustering?

Clustering, in the context of databases, refers to the ability of several servers or instances to connect to a single database. An instance is the collection of memory and processes that interacts with a database, which is the set of physical files that actually store data. Clustering offers two major advantages, especially in highvolume database environments:

  Fault tolerance: Because there is more than one server or instance for users to connect to, clustering offers an alternative, in the event of individual server failure.

  Load balancing: The clustering feature is usually set up to allow users to be automatically allocated to the server with the least load.

DAY 5 Industry Showers its Expertise in the Hadoop Ecosystem

Handouts from Dr. M. R. Sanghavi Expert Talk:

Contents to Cover:
a) MapReduce basics
b) Hadoop MapReduce
c) HDFS basics
d) Haddop Ecosystem

DAY 6 Showcasing of Applications of large data in High performance

Handouts from Prof T. B. Kute Expert Talk:

Contents to Cover:
a) Integrating R and Hadoop
b) RHIPE and RHadoop
c) Applications on large data
d) Highperformance and parallel R

DAY 7 Trainees experienced the Analysis of data
in motion
Handouts from Mr Prasad Chandane Expert Lecture\

DAY 8   Cloud in the context of Big Data eNightCloud
Handouts from   Prof.Amol Kalugade   Expert Talk:

Contents to Cover:
a) Relational and Nonrelational databases
b) R interface to databases
c) Managing data security and variety
d) Cloud in the context of Big Data