DR. BABASAHEB AMBEDKAR TECHNOLOGICAL UNIVERSITY, LONERE – 402 103, Raigad (MS)

REPORT on

ONE WEEK FACULTY DEVELOPMENT PROGRAMME ON "DATA SCIENCE" SPONSORED BY TEQIP-III

Organized at DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING,
CSMSS, CHH. SHAHU COLLEGE OF ENGINEERING, AURANGABAD

During 12th to 16 June 2018

A one week Faculty Development Program (FDP) On "**Data Science**" was successfully conducted in CSMSS, Chh. Shahu College of Engineering, Aurangabad from 12th to 16th June 2018. FDP was sponsored by TEQIP III of Dr. Babasaheb Ambedkar Technological University, Lonere.

The participants from various engineering colleges and affiliated to our University were attended the FDP. Most of the sessions were conducted by data analytics domain specific Industry "Grid Analytics" and few sessions also conducted by DBATU faculty. The main beauty of program was, 30% theory and 70% lab sessions. All the participants individually practiced assignments on desktop pc. Also, all the participants solved the complex problem of various operations on DATA manually as well as using analytics software.

In the entire FDP all participants have gain the knowledge and hands on experience to solve the data exploration, data preparation, machine learning, neural networks, analytical modeling and model deployment problems using STATISTICA software.

Objectives of the course:

- ✓ To provide participants with a conceptual overview of data science and its many facets
- ✓ To develop familiarity with advance analytics platform software enabler tool STATISTICA
- ✓ To introduce participants to large selection of models within STATISTICA and their application to various types of real-world problems
- ✓ To provide experience using STATISTICA's project workflow environment

- ✓ To demonstrate how STATISTICA can be used to solve common data science problems such as data preparation, variable selection, deployment of trained & building models, and more.

Key Takeaways:

- ✓ Learn how to develop Data Science Technical Architecture to Design Solutions.
- ✓ How do you Manage Data Science IT Projects
- ✓ Learn Testing Strategy for dealing with Data Science Projects.
- ✓ Learn Data Science Related techniques (Data Mining, Machine Learning, and Neural Networks etc.)
- ✓ Explore future Data Science technologies in for collective intelligence.
- ✓ Learn Data integration & management solutions
- ✓ Understand Visualization and Interface of Data Science.
- ✓ Learn quick and efficient ways to apply Modeling techniques.
- ✓ Automatic and interactive report generation.

Following Training Material is Provided to the every Participant:

- ✓ Statistica installation DVD's for participants.
- ✓ E-Books "Statistica Methods & Applications by Thomas Hill" & "Handbook of Statistical Analysis and Data Mining Applications by Gary Miner".
- ✓ How-to Video Tutorials for Learning and to perform hands on practice.

Schedule

Day 1 (12th June 2018): Data Science/Advanced analytics trends, technologies & application overview.
Objective: To provide participants with a conceptual overview of analytics & data science and its many facets. Also develop familiarity with *STATISTICA*'s interactive analytical environment.

Session	Title / Topics	
10:00 AM – 1:00 PM	**Data Science/Advanced analytics trends, technologies & application Overview.**	
	Statistica™ Data science platform Overview.	**Statistica™** GUI interface (Classic Menu, Ribbon Bar & Workspace)
	Data Connectivity & Integration	Data Acquisition (Structured, Semi Structured & External Database Connectivity) Extract Transform & Load (ETL)
1:00 PM – 2:00 PM	**Lunch Break**	
01:00 PM – 4:30 PM	**Data Preparation**	Data Filtering/ Recoding Techniques, Data Treatment (Missing Values, Outlier Treatment Data Transformation, Data Sampling Techniques
	Graphs, Charts & Visualization	Creating Graphs 2D & 3D Graphs (Histogram, Scatter Plot, Normal Probability Plot, Variability Plot, Surface Plot, Contour Plot etc.) Customizing Graphs Brushing Techniques
	Reporting & Output Management.	Statistica Workbook, Reports & Documents management.
04:30 PM – 05:00 PM	**Review of the Day**	**Q&A Session**

Day 2 (13th June 2018): Descriptive Statistics, Hypothesis Testing, Correlation, Multivariate Analysis, Regression Technique.

Session	Title / Topics	
10:00 PM – 01:00 PM	**Descriptive Statistics, Hypothesis Testing, Correlation & Cross Tabulation**	Descriptive Statistics, Correlation, t-test, ANOVA,Frequency Tables Cross-Tabulations
1:00 PM – 02:00 PM	**Lunch Break**	
02:00 PM – 04:30 PM	**Regression Analysis.**	Simple/Multiple Regression. Verifying Assumptions Residual Analysis. **Example: Case studies** (Poverty Prediction).
	Multivariate Exploratory Techniques	Cluster Analysis. **Example: Case studies** (Market segmentation- Automobile). Factor Analysis. **Example: Case studies** (Work Satisfaction). Discriminant Analysis. Principle Component Analysis
04:30 PM – 05:00 PM	**Review of the Day**	**Q&A Session**

Day 3 (14th June 2018): Model Management life cycle management.

Objective: To introduce participants to Data Mining large selection of models building, evaluation & Deployment.

Session	Title / Topics	
10:00 AM – 01:00 PM	Decision Trees	Classification & Regression Tree (CART). Chi Automatic Interaction Detection (CHAID), Random Forest. Model Evaluation (Confusion Matrix, ROC, Lift Chart & Gain chart. **Example: Case studies** (Heart Disease Prediction)
	Cluster Analysis	Hierarchical Clustering K- Means Clustering EM (Expectation Maximization) Clustering
01:00 PM – 2:00 PM		**Lunch Break**
01:00 PM – 04:30 PM	**Machine Learning Algorithms**	Support Vector Machine K-Nearest Neighbours Naïve Bayes Classifier
	Association Rules	Market Basket Analysis
04:30 PM – 05:00 PM	**Review of the Day**	**Q&A Session**

Day 4 (15th June 2018): Introduction to AI, Overview of Text Mining & application.

Objective: To introduce participants to STATISTICA Automated Neural Network, Advance Clustering, Text & Web Mining and their applications.

Session	Title / Topics	
10:00 AM – 1:00 PM	**Introduction to AI. Automated Neural Networks (ANN).**	Introduction & Overview or Neural Network Approaches & application. Implementation Example Solutions 1. Classification 2. Regression 3. Clustering 4. Time Series
01:00 PM – 2:00 PM		**Lunch Break**
01:00 PM – 04:30 PM	**Introduction to Text Analytics & Web mining.**	Introduction & Overview Web Mining Social Media Analytics Sentiment Analysis
04:30 PM – 05:00 PM	**Review of the Day**	**Q&A Session**

Day 5 (16th June 2018): Introduction to Citizen data scientist & collective Intelligence.

Objective: To provide experience using STATISTICA's project workflow environment

Session	Title / Topics	
10:00 AM - 1:00 PM	**Data Mining Workspace**	Introduction Case Study: 1. Multiple Model Creation 2. Model Comparison & Deployment 3. Working with R & Python
1:00 PM - 01:00 PM	**Lunch Break**	
02:00 PM - 3:00 PM	**Data Mining Recipe**	**Quick Model Management Life Cycle**
02:00 PM - 5:00 PM	**Review of the Week**	**Q&A Session**

DAY WISE BRIEF REPORT

Date: 12th June 2018 (First Day)

Inaugural Session

The FDP was inaugurated by our Hon'ble Vice Chancellor Prof. Dr. V. G. Gaikar and delivered keynote address to the participants through Video Conferencing.

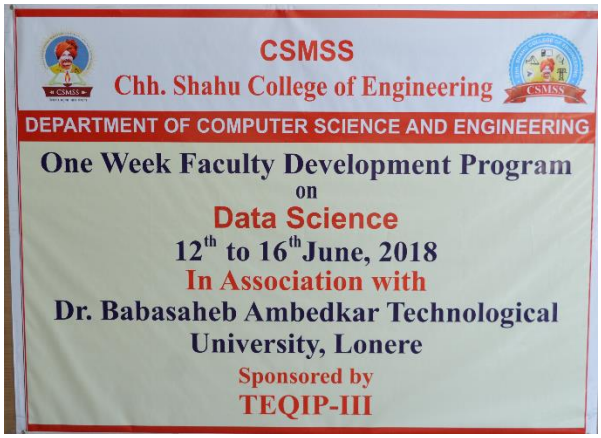
In inaugural function Hon'ble Vice Chancellor, said faculty should equip himself with concepts and technology by attending such technology oriented fdp. Also, told the importance of faculty training and mentioned that every University department should recognized with some specialized laboratory.

Dr. L. D. Netak (Coordinator FDP-DBATU, Lonere), Mr. Bishwajit Nayak (Resource Person, Grid Analytics India Pvt. Ltd.), Mr. Ayush Rastogi (Resource Person, Grid Analytics India Pvt. Ltd.), Dr. U.B. Shinde (Principal, CSMSS, CSCOE, Aurangabad), Dr. R. S. Pawar (Principal, SYCET, Aurangabad), and Dr. S. P. Abhang (Coordinator FDP-CSMSS, Aurangabad) were present for the inaugural function.

Address by Hon'ble Vice Chancellor Prof. Dr. V. G. Gaikar (Through Videoconferencing)

Address by Dr. L. D. Netak FDP, Coordinator, DR. BATU

Felicitation of Mr. Bishwajit Nayak



FDP Banner

Felicitation

Session Photograph: Mr. Bishwajit Nayak

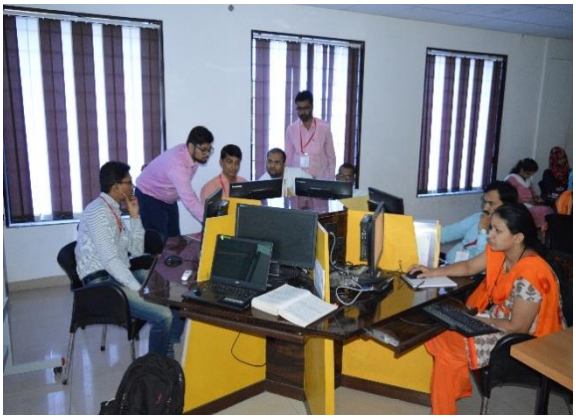
Photographs: Day One

Mr. Bishwajit Nayak has given focus on advanced analytics trends, technologies and application overview.

Date: 13th June 2018 (Second Day)

On second day, started discussion on methods of Descriptive Statistics, Correlation, t-test, ANOVA, Frequency Tables Cross-Tabulations.

After the post lunch session focused on Simple/Multiple Regression, Verifying Assumptions, and Residual Analysis along with case studies on Poverty Prediction, Work Satisfaction, Market segmentation- Automobile etc.

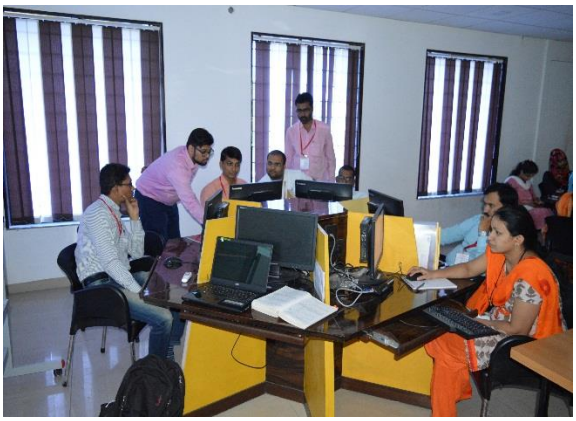


Hands-on practice by Participants

Photographs: Day Two

Date: 14th June 2018 (Third Day)

On third day, began with Model Management life cycle management. The main objective of this session was to introduce the participants to Data Mining large selection of models building, evaluation & Deployment.



Hands-on practice by Participants

Photographs: Day Third

In further sessions, discussed the different Machine learning algorithms including Support Vector Machine, K-Nearest Neighbours, and Naïve Bayes Classifier.

Date: 15th June 2018 (Fourth Day)

On forth day, started session with introduction to automated Neural Network. Introduction to the concepts of ANN with practically classification, regression, clustering, and time series.

Also demonstrated problems using STATISTICA tool on Web mining typically focusing on Social Media Analytics and Sentiment Analysis.

Hands-on Session Conducted by Mr. Nayak

Group Photo at Ellora Caves

Photographs: Day Forth

In the afternoon session, field visit was organized to the world heritage place called Ellora caves and Ghrushneshwar temple at Ellora.

Date: 16th June 2018 (Fifth Day)

On fifth day, Prof. Dr. A. W. Kiwelekar, DBATU Lonere has conducted one session on "An architectural perspective of learning analytics".

Felicitation of Prof Dr. A. W. Kiwelekar

Session by Dr. A. W. Kiwelekar

Photographs: Day Fifth

The main theme of Prof. Kiwelekar's talk was, fundamentals of analytics and different tasks performed by data analyst/scientist. In his talk, he briefly discussed concepts of data analytics and co-related with the real life scenarios such as analysis of student progress, etc.

Further, Prof. Kiwelekar has discussed the architectural perspective of learning analytics and concluded the session by sharing the research activities initiated in the department of computer science & engineering, at Dr. Babasaheb Ambedkar Technological University, Lonere.

After post lunch session, valedictory function was conducted in the presence of Dr. S. B. Deosarkar (Institute Project Director, TEQIP-III, Dr. BATU, Lonere), Dr. L. D. Netak (Coordinator FDP, Dr. BATU, Lonere), Dr. Ulhas Shiurkar (Principal, DIEMS, Aurangabad), Dr. U. B. Shinde (Principal, CSMSS CSCOE, Aurangabad), and Dr. S. P. Abhang (HoD and Coordinator FDP, CSMSS , Aurangabad)

Few participants expressed their views on overall preparation and execution of the fdp and also submitted their feedback about FDP.

Prof. Dr. S. B. Deosarkar discussed about the TEQIP III project plan and activities, opportunities to the faculty members, Institutions for conducting workshops, training session by Industry experts and research work.

The Certificate distributed by the hands of dignitaries to all participants.
Dr. S. P. Abhang presented vote of thanks to Hon.'ble Vice-Chancellor, Prof. Dr. V. G. Gaikar, Dr. BATU, Lonere, Management of CSMSS, all participants, dignitaries present for the function, DBATU TEQIP-III office bearers and those are directly or indirectly involved to make fdp successful.

The valedictory function ends with the group photograph of participants with all dignitaries.

Felicitation of

Prof. Dr. S. B. Deosarkar

Felicitation of

Dr. Ulhas Shiurkar

Certificate Distribution

FDP Group Photograph

Photographs: Day Fifth

Coordinators

1. DR. L. D. NETAK, DR. B.A.T.U. LONERE
2. DR. S. P. ABHANG, CSMSS, CSCOE AURANGABAD