



DCT's
Dhempe College of Arts and Science
Miramar, Goa

DEPARTMENT OF MATHEMATICS

Report on the 6 Days State Level Training Programme on Data Science Using R under DBT Star College Scheme

Date: 3rd February to 8th February

Resource Persons:

1. Dr. Amiya Bhowmick, Assistant Professor, ICT Mumbai (Lead Trainer),
2. Dr. Abhishek Mukherjee, Associate Professor, ISI Kolkata
3. Dr. Sabyasachi Battacharya, Professor, Indian Statistical Institute, Kolkata
4. Dr. Buddhananda Banerjee, Assistant Professor, IIT Kharagpur
5. Dr. Achut Kumar Banarjee, Assistant Professor, Ajim Premji University, Bhopal
6. Ms. Riddhi Bharani (Teaching Assistant),
7. Ms. Dipali Mestry (Teaching Assistant)

Participants: 32

Location: AV Hall

Program Type: State Level Training Programme

Sponsored by: DBT Star College Scheme

Introduction:

A 6-day State Level Training Programme on Data Science using R was organized as part of the DBT Star College Scheme, which aims to enhance scientific knowledge and research skills. The workshop, held from 3rd February to 8th February, was designed to provide participants with hands-on training in various aspects of data science using R programming, from basic introduction to advanced methods. Dr. Amiya Bhowmick, an expert in the field, was the lead trainer, supported by his teaching assistants Ms. Riddhi and Ms. Dipali. The program included interactive sessions, hands-on exercises, and specialized lectures on various data science techniques.

Day 1: Introduction to R and Graphical Methods

The training began with an introduction to the basics of R programming. Dr. Bhowmick provided an overview of R's capabilities in data science, highlighting its usage in statistical analysis and graphical representation of data. He covered the fundamentals such as R syntax, data types, and data structures. The session also focused on graphical methods in R, enabling

participants to visualize data effectively. To ensure active engagement and understanding, quizzes were conducted after each session to test participants' knowledge.

Day 2: Introduction to Probability Distribution

On the second day, Dr. Bhowmick delved into probability distributions, a crucial concept in data science. The session covered different types of probability distributions, including normal, binomial, and Poisson distributions. Participants learned how to apply these distributions in R and interpret the results. The teaching assistants, Ms. Riddhi and Ms. Dipali, assisted in guiding students through practical exercises to reinforce the theoretical concepts.

Day 3: Regression Analysis

The third day focused on regression analysis, an essential statistical method in data science. Dr. Bhowmick explained both simple and multiple regression models, emphasizing how to analyze relationships between variables using R. Participants worked through practical examples, using R to implement regression models and interpret the outputs. The session aimed to provide participants with the skills needed to apply regression analysis to real-world data.

Day 4: Special Lectures

On Day 4, the training program featured a special lecture on **Environmental Science** by Dr. Abhishek Mukherjee highlighting the importance of data science in environmental studies. This lecture provided insight into how data science is applied to environmental research and conservation. Later, Dr. Sabyasachi Battacharya, gave an in-depth lecture on **Logistic Regression: Theory and Implementation in R**. Professor Battacharya's lecture provided valuable insights into logistic regression, a key tool in data science for classification tasks.

Additionally, Ms. Dipali Mestry conducted a session on **Quarto**, an open-source tool used to create dynamic documents with R. This session helped participants understand how to integrate R code, output, and narrative into professional reports and presentations.

Day 5: Multivariate Methods in R

The fifth day focused on advanced statistical methods, particularly **Multivariate Methods** in R. Dr. Buddhanda Banerjee, led the session, where he introduced techniques such as Principal Component Analysis (PCA) and Factor Analysis. These methods are essential for analyzing datasets with multiple variables. Dr. Banerjee provided practical examples and demonstrations, helping participants grasp the concepts of multivariate analysis.

In the second half of the day, Dr. Amiya Bhowmick continued with the session, elaborating on the application of these methods in real-life data science problems. The day ended with another quiz to assess the participants' understanding of multivariate techniques.

Day 6: Data Science Applications in Ecology and Science

On the final day, the session began with a lecture on **Data Science Applications in Ecology and Science** by Dr. Achut Kumar Banarjee. This lecture focused on how data science and

statistical methods can be applied to ecological studies and environmental sciences, showcasing the broad applicability of data science.

The program concluded with a session for participants to present their learnings. The Heads of Departments (HODs) from various departments also attended the last day's session, showing their support for the initiative and offering feedback on the program.

Conclusion:

The 6-day training programme successfully equipped 30 participants with a strong foundation in data science using R. Through a combination of theoretical lectures, practical exercises, and special sessions by renowned experts, participants gained essential skills in statistical analysis, regression, probability distributions, and more. The program not only helped students develop technical skills but also fostered a deeper understanding of how data science can be applied in various scientific fields, including environmental science and ecology.

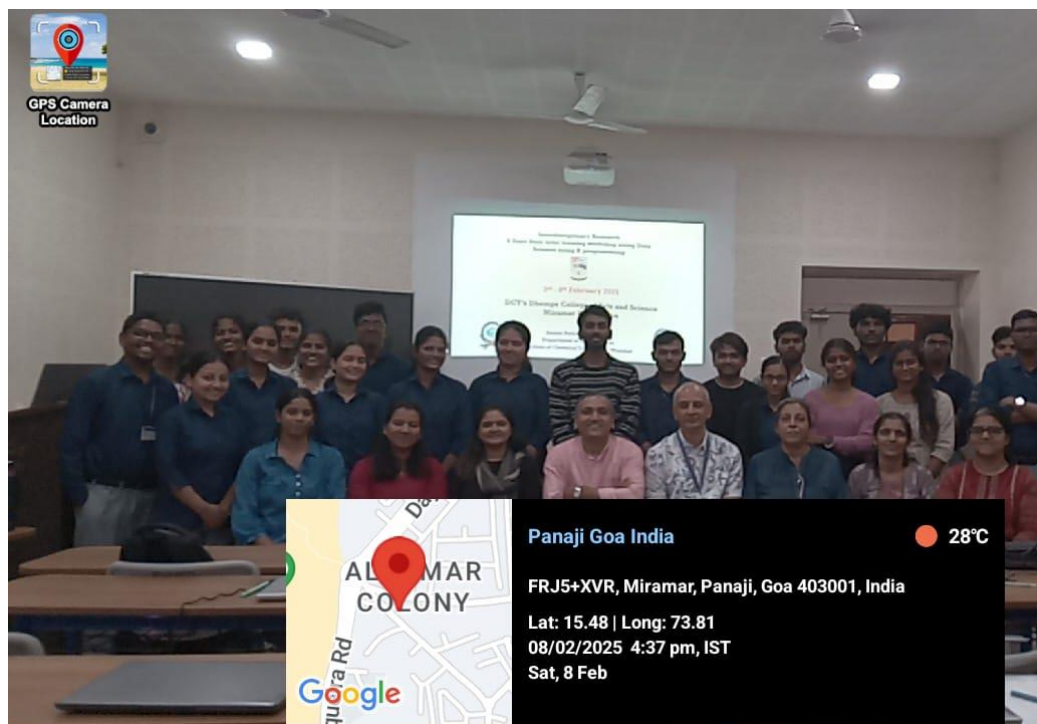
The participation of HODs and guest lecturers enhanced the overall learning experience, ensuring that the program was not just a technical training but also a collaborative and intellectually enriching experience. The quizzes conducted throughout the sessions helped reinforce the concepts and allowed instructors to gauge participants' progress.

In conclusion, this training programme was a successful and enriching experience for all involved, providing students with the knowledge and skills to pursue further work in data science and its applications in various domains.

Acknowledgments: We would like to express our sincere gratitude to Dr. Amiya Bhowmick, Ms. Riddhi, Ms. Dipali, and all the guest speakers for their valuable contributions to this program. Special thanks to the DBT Star College Scheme for supporting this initiative and making it possible.



During Session



Group Picture