



INDRAPRASTHA INSTITUTE of
INFORMATION TECHNOLOGY
DELHI



PG DIPLOMA

Data Science in Health and
Climate Change for Social
Impact

dshcs.iiitd.ac.in



In association with

data.org

 J-PAL
ABDUL LATIF JAMEEL POVERTY ACTION LAB
SOUTH ASIA AT IFMR

 ART PARK
AI & Robotics Technology Park, I-Hub @ IISc



TABLE OF CONTENTS

- Welcome Message
- About IIIT-Delhi
- About data.org
- Program Overview
- Program Structure
- Program Coordinators
- Teaching Faculty
- Admission Process
- Program Fees
- Our Partners
- Contact Information

WELCOME MESSAGE

Welcome to the Data Science in Health and Climate Change for Social Impact Program. Your journey towards becoming a data-driven change-maker starts here!

We are thrilled to introduce you to an exceptional learning experience designed to equip you with the skills and knowledge needed to make data-driven decisions in the fields of health, climate, and social sciences. This Postgraduate Diploma program, offered in collaboration with data.org, ARTPARK, and J-PAL South Asia at IFMR, promises to provide you with a comprehensive education that combines academic excellence and practical application.

At the core of our program are global data science experts who will guide you through the intricacies of utilising data to address critical challenges in health and climate change. With their deep expertise and industry experience, they will provide you with valuable insights, techniques, and best practices to navigate the ever-evolving data landscape. Our curriculum is specifically designed to cater to students and professionals who aspire to gain hands-on experience in data science. You will have the opportunity to work on real-time projects, analyzing and interpreting datasets to develop practical solutions. By immersing yourself in applied learning, you will develop the skills necessary to thrive in the data-driven world.

The faculty members from IIIT-Delhi, distinguished in their respective fields, will be your mentors throughout this program. Their guidance and support will ensure that you receive a well-rounded education, covering both the theoretical foundations and practical applications of data science in health, climate, and social sciences.

Additionally, we understand the importance of real-world exposure and the value of industry connections. As part of this program, you will have the option to pursue internships with leading organizations in the domain. These internships will provide you with invaluable opportunities to apply your skills in a professional setting and expand your network.

We invite you to embark on this exciting journey of learning, growth, and making a positive impact. Together, let's harness the power of data science to tackle the pressing challenges in health and climate change and create a better, more sustainable future for all.





ABOUT IIIT-Delhi

IIIT-Delhi is a comprehensive research-led teaching Institute where faculty, students, staff, and alumni work together in a spirit of discovery and creativity to build a better world. The Institute, set up by an Act of Govt. of NCT of Delhi (GNCTD), awards B.Tech., M.Tech., and Doctoral degrees. It emphasizes quality education, curiosity-led research, and innovation leading to entrepreneurship. The teaching activities at the Institute derive strength from the cutting-edge research conducted by faculty, strong industry linkages, and international collaborations.

IIIT-Delhi has six academic departments - Computer Science and Engineering, Electronics and Communications Engineering, Computational Biology, Human-Centered Design, Mathematics, and Social Sciences & Humanities. Its research centres include the Infosys Centre for Artificial Intelligence, Centre for Design and New Media, Centre of Technology in Policing, Centre of Excellence on Sustainable Mobility, Centre of Excellence in Healthcare, Centre of Excellence on Light Fidelity, Datakart Centre of Excellence, Centre of Excellence (CoE) in Human-Centered Computing, and Center for Quantum Technologies.

The IIITD-Incubation Centre on campus is instrumental in driving the growth of entrepreneurial activities. IIIT-Delhi has also set up a Technology Innovation Hub called iHub Anubhuti under the National Mission on Interdisciplinary Cyber-Physical Systems. There is also an Electropreneur Park focused on electronics startups.

IIIT-Delhi is committed to playing a leadership role by creating new knowledge, defining the future of higher education, preparing leaders for tomorrow, and making a tangible impact for the betterment of society.

More details can be found at iiitd.ac.in.



ABOUT **data.org**

[data.org](#) is a platform for partnerships to build the field of data for social impact (DSI). Launched in 2020 by the Mastercard Center for Inclusive Growth and The Rockefeller Foundation, data.org bring together philanthropy, private sector technology, academia, and social impact organizations to build the field of data for social impact. The organization do this work in three ways:

- Prove the Cases
- Strengthen Capacity
- Transform the Commons

Through one of their key initiatives, the Capacity Accelerator Network (CAN), they have committed to help train one million, purpose-driven data practitioners by 2032. As a part of CAN, following the launch of our Workforce Wanted report, data.org is currently building Data Capacity Accelerators in low- and middle-income countries that will train the next generation of data practitioners with the interdisciplinary skills needed to be successfully working in social impact. The India Data Capacity Accelerator, comprising partners like JPAL South Asia, Ashoka University, BITS Pilani and IIIT-D, is envisioned to catalyze the application of data to address systemic challenges at the intersection of climate and health.

PROGRAM OVERVIEW

This Postgraduate Diploma program, offered in collaboration with data.org, ARTPARK, and J-PAL South Asia at IFMR, combines the expertise of global data science leaders with hands-on practical experience.

In an increasingly data-driven world, this program stands at the forefront of innovation, empowering participants to make data-driven decisions in the domains of health, climate, and social sciences. Led by renowned experts, our curriculum offers a comprehensive blend of theoretical foundations and applied learning, ensuring that students gain a deep understanding of data science principles and the ability to apply them to real-world challenges.

- Global data science experts lead training in the area of health, climate and social sciences
- Specifically designed for students/professionals aiming for hands-on experience in data science
- Classes delivered by distinguished IIIT-Delhi faculty members
- Applied learning with hands-on projects using real-time data
- Fellowship and Internship opportunities with leading organisations in the domain (optional)

Program Duration

**37
WEEKS**



Classes in
**HYBRID
MODE**

PROGRAM



Module 1: Introduction and working with data

- **Introduction to Statistics and Data Science**
 - Overview of the Program: Unique
 - Opportunities and Challenges
 - Overview of Data Science Approaches
 - Types of Data
 - Open Sources of Data, Introduction to the Key Datasets for the Curriculum
 - Generating Your Own Data: Study Design



Module 2: Working with Data

- **Introduction to Excel, R and Python**
 - Introduction to Excel, R & Python
 - Reading and Writing Data using R & Python
 - Basic Exploration of Data
 - Data Quality (Seeing and Analyzing)
 - Remediation Approaches (Filtering, Transformations, Imputation)
 - Real World Example Summarizing the module



Module 3: Inferential Statistics and Comparative Data Sciences

- **Basics of Biostatistics**
 - Getting Data into Shape (Long and Wide Data)
 - Random variable and probability
 - Measures of Centrality and Deviation
 - Probability Distributions: Discrete and Continuous
 - Moments of distributions (mean, variance, skewness, kurtosis)
 - Probability Density Function and Cumulative Density Function
 - CLT and Normal Distribution
 - Sampling distributions (e.g., sampling distribution of the mean)
 - Central Limit Theorem
 - Histograms, Density Plots, Boxplots, Violin Plots
- **Beginning Data Analytics**
 - Relationship between distributions (e.g., normal approximation to binomial)
 - Quantiles and Rank Statistics
 - Testing for Normality and Data Transformations
 - Statistical Inference: Parametric Tests
 - Statistical Inference: Non-parametric Tests
 - Real World Example Summarizing the module





Module 4: Modelling and Basic Machine Learning for Health and Climate Applications

- **Linear Algebra and Regression Models**
 - Introductory Linear Algebra
 - Supervised and Unsupervised Learning
 - Regression: Linear Model General Linear Model
 - Classification: Logistic Regression and Multinomial Model
 - Variable Selection: Stepwise approaches
- **Machine Learning**
 - Introduction to Machine Learning & Bayes Theorem
 - High Dimensional Data: Principal Component Analysis and t-SNE
 - Machine Learning Approaches: Decision Tree, Neural Networks. SVM, Random Forest, Bayesian Networks
 - Overfitting and Generalization
 - Regularization: Lasso, Ridge and Elastic Net Regression
 - Unsupervised Learning: Clustering
 - Real World Example Summarizing the module



Module 5: Visualization of Health and Climate Data

- **Advanced Visualization Techniques in Data Sciences**
 - Introduction to ggplot (R) and seaborn (python)
 - Advanced Visualization Strategies in Data Sciences e.g., heatmaps, sankey plots, radar charts, sunbursts, word clouds, waterplots, icicle plots
 - Interactive Data Visualization (using plotly, networkD3)
 - 3D data visualization
 - Creating Dashboard
- **Introduction to Geographical Data, Spatial Statistics**
 - Introduction to Spatial Data and Spatial Models: Geostatistical data; Lattice sata; and Point data.
 - Characterising Spatial Autocorrelation (Metrics) and Relevant Issues for Classical Regression Analysis;
 - Exploratory Spatial Data Analysis and Stationarity of Spatial Random Processes
 - Measuring Spatial Dependence and Spatial Heterogeneity
 - Environmental Pollution and Economic Growth application with hands-on exercises on ArcGIS and R.



Module 6: Advanced Data Science Applications in Health and Climate

- **Artificial Intelligence and Deep Neural Networks: Models in Health and Climate Data**
 - Introduction and Multi-layer perceptron
 - Introduction to TensorFlow and Feedforward Neural Networks
 - Activation Functions and Training Neural Networks
 - Learning from images: Convolutional Neural Networks (CNNs)
 - Learning from sequences: Recurrent Neural Networks (RNNs)
 - Applications of AI and Deep Neural Networks
 - Interpretability
- **Nature Language Processing**
 - Parts of Speech, Parsing, Ontologies, WordSense
 - Introduction to Modelling with Text: Real World Examples
 - Modelling Text: Hidden Markov Model, Recurrent Neural Networks, LSTMs
 - Word Embeddings & word2Vec
 - Transformer
 - BERT and its Applications
 - Social Media, Emotion Modeling, Chatbots & ChatGPT



Module 7: Data Science in Health Care

- **Data Ethics**
 - Introduction to Data Ethics
- **Data Driven Health Sciences with Social Impact Case Studies**
 - Each participating faculty is expected to demonstrate an actual health - related case study with demonstrated social impact. Preference will be on standard publicly available data sets that can be reused.



Module 8: Data Science in Climate Emergency

- **Data and Climate Science**
 - Different types of publicly available data products for studying climate change and emergency. Data driven models for climate predictions and projections
- **Data Driven Climate Sciences with Social Impact Case Studies**
 - The module will cover data-driven case studies investigating the impact of climate change on rural communities, urban development, health and nutrition as well as studies linking climate change impact and public policy designs



Module 9: Data Science: Intersectionality and Representativeness of Data

- **Intersectionality and representativeness of data science**
 - Intersectionality and social datasets
 - Data gathering and representativeness of data
 - Intersectionality and representativeness in datasets and its impact on data driven strategies
 - Case studies about representativeness and intersectionality in datasets
- **Case Studies**



Module 10: Futuristic Data Science Perspectives in Health and Climate

- **Future of Data Science**
 - Futuristic developments in Data Science applications in Health, Climate and Social Sciences



PROGRAM COORDINATORS



Tarini Shankar Ghosh

Assistant Professor (Computational Biology)

Tarini Shankar Ghosh is a computational biologist focusing on human microbiome research. He received his B.Tech from Indian Institute of Technology (IIT) Kharagpur, his M.Tech from IIT Kanpur and his Ph.D from the University of Hyderabad (while working at the TCS Research of Tata Consultancy Services Ltd). He completed his post-doctoral training from APC Microbiome Ireland. His research interests include microbiome informatics, integrated multi-OMICs, machine learning and statistical learning.

Jaspreet Kaur Dhanjal

Assistant Professor (Computational Biology)

Dr. Jaspreet Kaur Dhanjal is an Assistant professor in the Department of Computational Biology at IIIT-Delhi. She received her PhD degree from Indian Institute of Technology Delhi in 2019. She has then worked as a Postdoctoral Fellow at the National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba Japan. Her research interests have been at the interface of computation, cellular and molecular biology. In particular, she is interested in the field of personalized therapeutics. She is also interested in exploring the potential of traditional medicines using modern science. She has recently been awarded the Har Govind Khorana- Innovative Young Biotechnologist Award from the Department of Biotechnology, Govt. of India.

Venkata Ratnadeep Suri

Assistant Professor (Social Science and Humanities)

Ratnadeep Suri (Ratan), Assistant Professor, SSH, IIIT-Delhi, has completed his Post Doctoral research at the Nanyang Technological University, Singapore in 2016. Earlier he earned his PhD from The Indiana Media School, Indiana University in 2013, with a Major in Communication and a Minor in Information Science. He also has an MA from Western Michigan University, Michigan, and an MA from University of Hyderabad, India.

Tavpritesh Sethi

Associate Professor (Computational Biology)

Dr. Tavpritesh Sethi is a physician-scientist and Associate Professor of Computational Biology at Indraprastha Institute of Information Technology Delhi, India. He is the Founding Head of Center of Excellence in Healthcare at IIIT-Delhi, has been a fellow of the DBT/Wellcome India Alliance at AIIMS, New Delhi, India and a visiting faculty member at Stanford University. He received his M.B.B.S from Government Medical College, Amritsar and PhD from CSIR-Institute of Genomics and Integrative Biology. His research focuses on AI/ML for Medicine and Public Health.





TEACHING FACULTY

Computational Biology

Dr. N. Arul Murugan

Dr. Debarka Sengupta

Prof. Ganesh Bagler

Dr. Gaurav Ahuja

Dr. Vibhor Kumar

Dr. Tarini Shankar Ghosh

Dr. Jaspreet Kaur

Dr. Tavpritesh Sethi

Dr. Arjun Ray

Prof. GPS Raghava

Social Sciences & Humanities

Dr. Venkata Ratnadeep Suri

Dr. Manohar Kumar

Dr. Gaurav Arora

Computer Science

Dr. Ojaswa Sharma

Prof. Vikram Goyal

Dr. Pravesh Biyani

ADMISSION PROCESS



* Interview process will only be initiated based on the number of applications received from deserving candidates.

PROGRAM FEE

₹
25,000

-  1 Highly subsidised fee for the first two years
-  2 Access to IIITD library resources
-  3 Inclusion into the IIITD alumni network
-  4 Exposure to hands-on projects
-  5 Internship opportunities with leading organisations

OUR PARTNERS

As part of the program's mission to strengthen the data capacity of the social sector in India, we are collaborating with prominent social impact organizations to place a limited number of students in 1-year fellowships (via J-PAL), or 3-month internships (via J-PAL or ARTPARK). For the candidates, these partnerships offer a unique opportunity to apply their data skills in real-world scenarios and contribute to projects with a profound social impact.

Internship Partner



ARTPARK @ IISc (AI & Robotics Technology Park) is a unique non-profit organization promoted by the Indian Institute of Science (IISc) to foster innovations in AI & Robotics by bringing together the best of the startup, industry, research, and government ecosystem. It is seed funded by the Department of Science & Technology (DST), Govt. of India, under the National Mission on Interdisciplinary Cyber-Physical Systems (NM-ICPS) and the Govt. of Karnataka.

ARTPARK @ IISc is driving advances in robotics, autonomous systems and AI through translational R&D in areas of Intelligent Healthcare, Automation for Logistics and Skilling for the AI age. Our work spans open tools, standards, IP, technologies, databanks and path-breaking companies.

Fellowship Partner



The Abdul Latif Jameel Poverty Action Lab (J-PAL) is a global research center working to reduce poverty by ensuring that policy is informed by scientific evidence. Anchored by a network of over 750 affiliated researchers at universities around the world, J-PAL conducts randomized impact evaluations to answer critical questions in the fight against poverty.

J-PAL was launched at the Massachusetts Institute of Technology in 2003 and has 7 regional offices around the world. J-PAL South Asia, hosted by the Institute for Financial Management and Research (IFMR), Chennai, and has built partnerships for the generation of new research and advancement of evidence-informed policymaking with 20 Indian state and central government ministries, as well as training partnerships with the governments of Nepal, Bhutan, Bangladesh, and Pakistan. For more information, visit povertyactionlab.org/south-asia



POWERUP, ACHIEVE, & ELEVATE



CONTACT
INFORMATION

dshcs.iiitd.ac.in
dshcs@iiitd.ac.in
IIIT-Delhi, Okhla Phase -III
New Delhi-110020