

DBT-NCCS holds popular webinar on machine learning for life sciences

New Delhi, Oct 23: Machine learning (ML) and Computational Modeling (CM) are increasingly gaining importance in the field of life sciences research. They offer invaluable tools to predict and understand how various molecules in the cells, the basic units of life, are structured, and how they function.



They have been pivotal in gaining essential insights into what kind of structures and functions molecules like proteins, the building blocks of life, could have. They also help predict and understand how these proteins might operate in the cell, how they might interact with other molecules like drugs, and also to design proteins with desired properties. Given that proteins are essential molecules that participate in almost all aspects of cellular processes, these modern technologies are vital components of a scientist's toolbox.

To give science enthusiasts an opportunity to learn how ML and CM are becoming an integral part of research in biology, DBT- National Centre for Cell Science (DBT-NCCS), Pune held a webinar on the topic, 'Combining Machine Learning & Modeling Approaches to Map Protein Structure-Function Relationships: Applications to Prediction & Design'.

This session was presented by Prof. Richard Bonneau, Director of the Centre for Genomics & Systems Biology, New York University, and the Group Leader for Systems Biology at the Centre for Computational Biology, Flatiron Institute, New York, USA. He discussed recent advances in the intersection of machine learning approaches to protein sequence, language models and their use in settings that incorporate protein three-dimensional structure.

He dwelt on his work on DeepFRI, a method for using language models and graph convolutional neural network models to predict, model and ultimately design protein structure and function. He also discussed new methods for using these models to find similar proteins and implications for scale in an ever increasing protein sequence universe. Lastly, he focussed on how these ML methods can be integrated with generative (MCMC/modeling) approaches like Rosetta to build design and prediction pipelines that scale to meta genomics and beyond.

The webinar was open to all and was free of charge. It was organized under the auspices of the “Manav: Human Atlas Initiative”, a collaborative project between the National Centre for Cell Science (DBT-NCCS), the Indian Institute of Science Education and Research (IISER-Pune) and Persistent Systems. Funded by the Department of Biotechnology (DBT), Government of India, and co-funded by Persistent Systems, this initiative is an example of a public-private partnership.

The “Manav: Human Atlas Initiative” aims to annotate the extensive information related to the human body that is available in the scientific literature and databases, and initiate steps towards creating a human atlas eventually. The project aims to help students across India to learn how to comprehend and analyze scientific literature, and extract relevant information from it using a digital annotation tool. Online workshops on how to read scientific literature are also routinely conducted. Colleges interested in having a session organized for their students can write to manav.iiserpune@gmail.com. Students, faculty members and researchers interested in participating in the project can learn more from the project’s website (<https://manav.gov.in/>), Twitter (Manav Human Atlas; @ManavAtlas) & Facebook (MANAV Human Atlas). The Rajya Sabha TV has also featured this initiative on Gyaan Vigyaan & Science Monitor.

This webinar is one of several in the data science webinar series which was initiated to help increase awareness about various aspects of data science. This includes applications of data science across disciplines, from public health and pandemics to astronomy, as well as the entrepreneurial opportunities available in this field.

These webinars also give the participants an opportunity to interact with experts from diverse fields around the world, during the Q&A session. This series is aimed mainly at college students, but is open to everyone.

The data science webinar series was initiated to meaningfully engage students and other science enthusiasts in an interactive educational experience during the lockdown. It was kick-started at the end of April with a webinar on “R0: How scientists quantify the intensity of an outbreak”. This was followed by webinars on diverse topics, including “Data Science Approaches for Genomics and Pandemics”, “Biomedical NLP at Scale: Automated inference of Biological Networks”, “The Future Research Scientist in the Age of AI and Big Data”, “When and where to divide to conquer: Personalized medicine using systems and machine-based biology approaches”, “Game Theory: Applications in biology & in a pandemic”, “Networks matter! Estimating the spread of disease in heterogenous, connected environments”, “Machine Learning & Deep Learning Applications in Astronomy & Biology”, “Personalized Precision Cancer Therapy”, “Entrepreneurial Promises of Data

Science”, and “Introduction to AI & Machine Learning for Biologists”. This series has been hugely successful, having attracted thousands of registrations, mainly from students. Given its popularity, the previous webinars have now been made available on the project’s YouTube channel, to anyone interested in listening to these talks.

Contact details: Jyoti Rao (jyoti@nccs.res.in)

Link: <https://www.nccs.res.in/>