



ICDS2021



Center for Data Science

Center for Data Science

University of Colombo, Sri Lanka

Department of Statistics,
University of Colombo

PROCEEDINGS OF THE INTERNATIONAL CONFERENCE IN DATA SCIENCE 2021

Data Science: Reshaping the Future



ICDS 2021

29th - 30th June 2021

Colombo, Sri Lanka

Data Science: Reshaping the Future



Proceedings of the International Conference in Data Science 2021 (ICDS2021)

29th - 30th June 2021

Colombo, Sri Lanka

**ICDS2021 is organized by the Center for Data Science
jointly with the
Department of Statistics, University of Colombo**

Publisher: Center for Data Science, University of Colombo

ISBN 978-624-5873-02-9

Proceedings of the International Conference in Data Science 2021
29th - 30th June 2021, Colombo, Sri Lanka

Editors:

Dr. Dilhari Attygalle,
Department of Statistics, University of Colombo

Dr. Rushan Abeygunawardana,
Department of Statistics, University of Colombo

Dr. H. A. S. G. Dharmarathne,
Department of Statistics, University of Colombo

Dr. K. A. D. Deshani,
Department of Statistics, University of Colombo

Date published: 29th June 2021

Published by:

Center for Data Science,
Department of Statistics, Faculty of Science, University of Colombo, Colombo 03,
Sri Lanka

©Center for Data Science, University of Colombo, Sri Lanka
ISBN 978-624-5873-02-9

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means without the permission of the editors.

Responsibility for the contents of these articles rests squarely upon their respective author(s) and not upon the editors or the reviewers. Data presented and conclusions reached by the author(s) are for information only and are not for use without independent substantiating investigation on the part of the user.

Welcome Message by the Conference Chair and Director Center for Data Science



On behalf of the organizing committee, it is my great pleasure to welcome you to the first ever International Conference in Data Science (ICDS 2021) in Sri Lanka. This two-day virtual conference is organized by the Center for Data Science jointly with the Department of Statistics, Faculty of Science, University of Colombo to celebrate the 5th anniversary of the Center for Data Science, 20th anniversary of the Department of Statistics and as part of the Centenary celebration of the Faculty of Science, University of Colombo. Our theme for the conference “Data Science: Reshaping the Future” has been carefully chosen to mark these milestones.

We have an exciting program lined up for this virtual conference, with a keynote speech, two guest speeches and six technical sessions spanning Data Visualization to Deep Learning and AI, facilitating collaborations between academia, industries and students. On the second day of the conference, we have organized two workshops followed by the final round of the mini-hackathon where more than eighty undergraduate teams gave their best in solving a real life problem utilizing cutting edge data science methodologies. I hope you will have an educational, productive and enjoyable time at this very special conference.

My sincere gratitude goes to the invited speakers, presenters and co-authors, workshop attendees and resource personnel, and organizers of the mini-hackathon. I would like to thank our sponsors; strategic and knowledge partner – Creative Software and technical partner Altria for providing generous financial support.

To organize a conference of this magnitude in the midst of a pandemic is not an easy task. The organizing committee planned and executed all activities virtually. We are now rolling out the proceedings virtually as well. I would like to thank the Vice-Chancellor of the University of Colombo, Senior Professor Chandrika N. Wijeyaratne for her wisdom and guidance. Also I would like to thank Dean of the Faculty of Science, Senior Professor and Chair of Physics Upul Sonnadara for his

continuous and invaluable support. A very special thank you goes to the Board of Management of the Center for Data Science for the encouragement and the blessings given. Last but not least, I must thank the Head of the Department of Statistics, Professor Chandima Tillakaratne and all the members of the staff for their untiring effort to make this conference a success.

Dr. Sameera Viswakula

Conference Chair/ ICDS 2021

Director/ Center for Data Science

Message by the Vice-Chancellor, University of Colombo



I convey my sincere congratulations on the occasion of the International Conference in Data Sciences (ICDS 2021) organized by the Centre for Data Sciences and the Department of Statistics in celebration of the centenary year of the Faculty of Science, University of Colombo (UoC).

Your chosen theme “Data Science: Reshaping the Future” is time appropriate and aptly addressed by the excellent mix of lecture topics, thematic workshops and mini hackathons for students. I salute the academic and administrative leads of the Faculty of Science for recognizing the emerging importance of Data Sciences, conceptualizing UoC’s strategic development of the field and your strong commitment to enhance the required human capital in Sri Lanka. Your hard work has paid dividends through the establishment of the state-of-the-art Centre, an enabling infrastructure and the vibrant academic programs for undergraduates and postgraduates. Your commitment extends further into the development of partnerships with the corporate sector and the industry.

The upscaling of pedagogy, research and scholarship in Data Sciences witnessed in the recent past has encouraged intramural and extramural multidisciplinary and multisector engagement, which is very progressive. The need for a fully-fledged higher educational facility in big data analysis and the related subjects was championed by visionaries and policy makers of our University. The current academic staff are aptly qualified with doctoral degrees from highly recognized universities to be excellent trainers, supervisors and public intellectuals with international recognition. I take this opportunity to thank the head and members of the Department of Statistics for your lively engagement in the holistic professional development of every graduate of the Faculty of Science.

I am confident that this international conference, held during unprecedented times of the COVID19 pandemic, will add greater value to the learning environment of the Faculty of Science.

I extend my warmest wishes to all the speakers, organizers and participants of this conference. I am certain that this virtual conference will yield fruitful and productive deliberations entwined with the establishment of strong academic networks, while preserving the best of health and safety.

May you have a fulfilling future in Data Sciences at the University of Colombo!

Senior Professor Chandrika N Wijeyaratne
Vice Chancellor, University of Colombo



On behalf of the staff and students of the Faculty of Science, University of Colombo, I extend my warmest welcome to all the participants of the 1st International Conference on Data Science 2021, 'Reshaping the Future'. This conference is organized by the Center for Data Science in collaboration with the Department of Statistics to mark the 20th anniversary of the establishment of a separate department for Statistics under the Faculty of Science, University of Colombo.

The hosting of the ICDS 2021 conference at the University of Colombo is a very special occasion for us as we are celebrating 100 years of excellence in science education this year. ICDS 2021 offers an excellent platform for the exchange of scientific and technical knowledge and information related to the emerging field of Data Science, across many application areas. With the wide expanse of conference tracks ranging from machine learning to ethics of data use and with experts coming together from the industry and academia, we expect the knowledge sharing to be at a high level at this conference.

A total of 6 separate sessions are being conducted during the 2 days of the conference, including an inaugural session with keynote addresses, presentation of technical papers and post conference workshops focused on machine learning and deep learning followed by a mini hackathon.

I would like to take this opportunity to thank all the presenters and their co-authors for contributing to the dissemination of their research findings and the participants for registering at the conference. On behalf of the Faculty of Science, I extend my sincere gratitude to the organizing committee for working under difficult conditions due to the covid-19 pandemic, to make this event a success.

Senior Professor Upul Sonnadara
Dean of the Faculty of Science, University of Colombo



The International Conference in Data Science 2021 (ICDS 2021) is a joint endeavor of the Center for Data Science (CDS) and the Department of Statistics of the University of Colombo. This is the first conference in Data Science organized in Sri Lanka. It is organized to commemorate the 100 years of excellence in Science at the University of Colombo, 20th anniversary of the Department of Statistics and 5th anniversary of the CDS. Being able to organize an international conference in the same year which it celebrates the 5th anniversary, is a great achievement of the CDS.

This conference provides a platform for data scientists to communicate the novel concepts in data science as well as the applications of data science techniques to solve real world problems. It also enhances the collaboration between academia and industry.

ICDS 2021 includes keynote and guest speeches, several contributed sessions, two workshops and a mini hackathon. CDS collaborates with the Acuity Knowledge Partners to conduct one of the workshops while collaborating with Creative Software and the Stat Circle, the student society associated with the Department of Statistics, to conduct the mini hackathon.

I take this opportunity to extend my sincere thanks to the chairperson and the organizing committee of the conference and wish all the best to the participants of the conference. I am confident that the ICDS 2021 will open doors to new developments and enhance the research in the discipline of Data Science. Furthermore, it will provide a platform to strengthen the ties between academia and the industry.

Professor Chandima Tilakaratne
Head, Department of Statistics, University of Colombo

Keynote Speaker



Jay Emerson is Director of Graduate Studies in the Department of Statistics and Data Science at Yale University. His academic work has included Bayesian change point analysis, statistics in sports, computational statistics and graphics, and environmental science. He is the lead statistician of the Yale Environmental Performance Index, and is working to apply similar rigor to the analysis of ESG metrics for applications in finance.

Guest Speaker



Bibhas Chakraborty is an Associate Professor and Ex-Director of the Centre for Quantitative Medicine at the Duke-National University of Singapore Medical School (Duke-NUS), as well as an Associate Professor of Statistics and Applied Probability at the National University of Singapore. He also holds an Adjunct Associate Professor position at the Department of Biostatistics and Bioinformatics at Duke University. Previously (2009-13), he was an Assistant Professor of Biostatistics at Columbia University. He completed his Ph.D. in Statistics from the University of Michigan, under the supervision of Prof. Susan A. Murphy in 2009. He is the recipient of the Calderone Research Prize for Junior Faculty from Columbia University's Mailman School of Public Health in 2011, and the Young Statistical Scientist Award from the International Indian Statistical Association (IISA) in 2017. His core areas of research include statistical reinforcement learning, precision medicine, dynamic treatment regimens, mobile/digital health, adaptive clinical trial designs, and a variety of applications in clinical and behavioral sciences. He wrote the first textbook on dynamic treatment regimens.

Guest Speaker



Umashanger Thayasivam, PhD, earned his doctoral degree for his research work on Mixture distribution at the University of Georgia. He is currently a Professor in Statistics & Data Science at the Rowan University New Jersey. At the university he lectures and trains both undergraduate and graduate students. He received the Teaching Wall of Fame award for his excellence in teaching in 2018. His interdisciplinary statistical research has spanned diverse areas including mixture distribution, robust estimation, high dimensional data analysis, statistical data mining, biomarker discovery, and educational data mining. He has an excessive experience with collaborating scientist engineers. He has several journal publications and numerous conference presentations. He has been PI/co-PI for several internal and external grants and collaborations, including the recent data mining project with Bristol Myers Squibb pharmaceutical company as well as the NIH grant on blood-based biomarkers for early-stage Alzheimer's disease. He also has extensive experience in mentoring student research in statistical/data mining research within the last ten years mentored more than 40 undergraduate students, many of which have presented their works at regional and national conferences. Several of his publications are with undergraduate students.

International Conference in Data Science 2021 (ICDS2021)

29th - 30th June 2021, Colombo, Sri Lanka

Conference Committee

Advisory Committee

- **Senior Professor Chandrika N. Wijeyaratne,**
Vice-Chancellor, University of Colombo, Sri Lanka
- **Senior Professor Upul Sonnadara,**
Dean, Faculty of Science, University of Colombo, Sri Lanka
- **Senior Professor M. R. Sooriyarachchi,**
Department of Statistics, University of Colombo, Sri Lanka
- **Dr. Nimal Wickremasinghe,**
Former Professor in Statistics, University of Colombo, Sri Lanka
- **Dr. Ruwan Weerasinghe,**
Senior Lecturer, University of Colombo, School of Computing, University of Colombo, Sri Lanka

Members of the Board of Management - Center for Data Science:

- Prof. Upul Sonnadara - Dean/Faculty of Science (Chairman)
- Prof. C.D. Tilakaratne (Head/Department of Statistics)
- Dr. Sameera Viswakula (Director/Center for Data Science)
- Prof. S. Karunaratne (Representative from the Council)
- Mr. J.M.U.B. Jayasekara (Representative from the Council)
- Prof. Pavithra Kailasapathy (Representative from the Senate)
- Prof. Chamari Hettiarachchi (Representative from the Senate)
- Dr. Pemantha Lakraj (Representatives from the Department of Statistics)
- Dr. Rasika Jayatillake (Representatives from the Department of Statistics)
- Dr. Nadeeka Basnayake (Representatives from the Department of Statistics)
- Dr. Rohitha Wijewardena -Director, Altria Consulting (Pvt) Ltd
(Representative of Industry experts)

- Dr. Rashmi Bomiriya, Managing Director - R S Metrics Asia Holdings & Chief Data Scientist, Remote Sensing Metrics Pvt Ltd (Representative of Industry experts)

Members of the Department of Statistics:

- Professor C. D. Tilakaratne – Head of the Department (Professor)
- Senior Professor M. R. Sooriyarachchi (Senior Professor)
- Dr. Dilhari Attygalle (Senior Lecturer)
- Dr. Rushan Abeygunawardana (Senior Lecturer)
- Dr. J. H. D. S. P. Tissera (Senior Lecturer)
- Dr. C. H. Magalla (Senior Lecturer)
- Mr. E. R. A. D. Bandara (Senior Lecturer)
- Dr. G. P. Lakraj (Senior Lecturer)
- Dr. R. V. Jayatillake (Senior Lecturer)
- Dr. S. D. Viswakula - Conference Chair and Director CDS (Senior Lecturer)
- Dr. A. A. Sunethra (Senior Lecturer)
- Dr. K. A. D. Deshani (Senior Lecturer)
- Dr. N. D. Basnayake (Senior Lecturer)
- Dr. H. A. S. G. Dharmarathne (Senior Lecturer)
- Ms. Lakmini K. N. Wijesekara (Lecturer)
- Mr. A. R. Munasinghe (Lecturer)
- Dr. I. T. Jayamanne (Lecturer)
- Ms. H. S. Karunarathna (Lecturer)
- Ms. D S Wickramarachchi (Lecturer)
- Ms. Avanthi Saumyamala (Lecturer)
- Ms. K. M. Manushi Hansani Siriwardana (Temporary Assistant Lecturer)
- Ms. S. Ishanka Randini Fernando (Temporary Assistant Lecturer)
- Mr. M. D. B. P. P. Badullahewage (Temporary Assistant Lecturer)
- Ms. K. D. O. Virajini (Temporary Assistant Lecturer)
- Ms. A. P. N. G. Adhikari (Temporary Assistant Lecturer)
- Ms. M. R. R. G. Maha Ranawaka (Temporary Assistant Lecturer)

- Ms. E. S. P. H. Rajadasa (Temporary Instructor)
- Ms. T. M. M. Jayaweera (Temporary Instructor)
- Ms. Lankadinee W. Rathuwadu (Temporary Instructor)
- Mr. A. A. K. T. Amarasinghe (Temporary Instructor)
- Ms. .G. C. J. Piyatilake (Temporary Instructor)
- Ms. Jayamini C. Liyanage (Temporary Instructor)
- Ms. S. D. Daluwatumulle (Temporary Instructor)
- Ms. N.G.N. Abirami (Temporary Instructor)
- Mr. H. K. T. Nanayakkara (Management Assistant)
- Mr. N.D Suduwella (Assistant Network Manager)
- Ms. R.M.N.E.K. Rathnayake (Management Assistant)
- Ms. R. A. K. Kithmini (Lab Attendant)
- Mr. W. D. M. C. Withanage (Support Staff)

Panel of Reviewers

All abstracts included in the Proceedings of the International Conference in Data Science 2021 have been independently reviewed through a double-blind process. The Advisory Committee and the Staff of the Department of Statistics would like to thank the following reviewers for their valuable services.

- **Dr. Rushan Abeygunawardana**
Senior Lecturer, Department of Statistics, University of Colombo
- **Dr. Radheeka Abeyweera**
Manager, Research and Development in Marketing Science, Facebook Inc.
- **Prof. Dhammika Amaratunga**
Independent Researcher
- **Dr. Anuradha Ariyaratne**
Senior Lecturer, Department of Computer Science, University of Sri Jayewardenepura
- **Dr. Dilhari Attygalle**
Senior Lecturer, Department of Statistics, University of Colombo
- **Mr. E.R.A.D. Bandara**
Senior Lecturer, Department of Statistics, University of Colombo
- **Dr. N.D. Basnayake**
Senior Lecturer, Department of Statistics, University of Colombo
- **Dr. Rashmi Bomiriya**
Managing Director, Asia and Chief Data Scientist, RS Metrics Holdings
- **Dr. Vasana Chandrasekara**
Senior Lecturer, Department of Statistics & Computer Science, University of Kelaniya
- **Prof. Sanjay Chaudhuri**
Associate Professor, Department of Statistics and Applied Probability, National University of Singapore.
- **Dr. Mahasen Dehideniya**
Senior Lecturer, Department of Statistics & Computer Science, University of Peradeniya
- **Dr. K.A.D. Deshani**
Senior Lecturer, Department of Statistics, University of Colombo
- **Dr. Neluka Devpura**
Senior Lecturer, Department of Statistics, University of Sri Jayewardenepura
- **Dr. H. A. S. G. Dharmarathne**
Senior Lecturer, Department of Statistics, University of Colombo

- **Dr. Malathi Imiyage Dona**
Research Officer, Cardiac Cellular Systems, Baker Heart and Diabetes Institute, Australia
- **Dr. Nusrath Hameed**
Senior Lecturer, Department of Computer Science, University of Ruhuna
- **Dr. Gayan Hettiarachchi**
Chief Scientist, OpenDNA Inc. Japan. Executive External Advisor, Sierra Consulting Inc. USA. Visiting academic staff, Osaka University and Rutgers University.
- **Dr. Isuru Hewapathirana**
Senior Lecturer, Software Engineering Teaching Unit, Faculty of Science, University of Kelaniya
- **Dr. I.T. Jayamanne**
Lecturer, Department of Statistics, University of Colombo
- **Ms. Mandu JayasundeDera**
Research Scientist, Facebook Inc.
- **Dr. R.V. Jayatillake**
Senior Lecturer, Department of Statistics, University of Colombo
- **Dr. J.A. Jeewani**
Senior Lecturer, Department of Computer Science, University of Ruhuna
- **Dr. Kasun Karunanayaka**
Senior Lecturer, University of Colombo School of Computing
- **Dr. Prasanna Karunanayake**
Assistant Professor ,Department of Radiology, Department of Public Health Sciences, Department of Neural and Behavioral Sciences, Penn State Neuroscience Institute, Penn State University, USA
- **Ms. G.H.S. Karunarathna**
Lecturer, Department of Statistics, University of Colombo
- **Dr. Chamath Keppitiyagama**
Senior Lecturer, University of Colombo School of Computing
- **Dr. Chinthaka Kuruwita**
Associate Professor of Statistics, Hamilton College, New York, USA
- **Dr. G.P Lakraj**
Senior Lecturer, Department of Statistics, University of Colombo
- **Dr. C. H. Magalla**
Senior Lecturer, Department of Statistics, University of Colombo

- **Mr. Rajith Munasinghe**
Lecturer, Department of Statistics, University of Colombo
- **Dr. Hasanthi Pathberiya**
Senior Lecturer, Department of Statistics, University of Sri Jayewardenepura
- **Dr. Dilina Perera**
Senior Lecturer, Department of Physics, University of Colombo
- **Mr. Shanaka Perera**
PhD candidate, Department of Computer science, University of Warwick.
- **Dr. Charith Peris**
Research Scientist, Alexa AI at Amazon
- **Dr. S. P. Pitigala**
Senior Lecturer, Department of Statistics & Computer Science, University of Kelaniya
- **Dr. Chathura Rajapakse**
Senior Lecturer, Department of Industrial Management, University of Kelaniya
- **Dr. Nishath Rajiv**
Postdoctoral Researcher, Prairie View A&M University, USA
- **Ms. Avanthi Saumyamala**
Lecturer, Department of Statistics, University of Colombo
- **Mr. Nuwan Senaratna**
Independent Consultant
- **Ms. Devini Senaratna**
Data Scientist, Booking.com
- **Dr. Sadun Malpriya Silva**
Biostatistician, NHMRC Clinical Trials Centre, University of Sydney, Australia
- **Dr. A.A. Sunethra**
Senior Lecturer, Department of Statistics, University of Colombo
- **Dr. Priyanga Talagala**
Senior Lecturer, Department of Computational Mathematics, University of Moratuwa
- **Dr. Thiyanga Talagala**
Senior Lecturer, Department of Statistics, University of Sri Jayewardenepura
- **Prof. Umashanger Thayasivam**
Professor of Statistics and Data Science, Department of Mathematics, Rowan University, USA

- **Dr. Pushpika Thilakarathne**
Statistician, Janssen, Pharmaceutical Companies of Johnson and Johnson
- **Prof. C.D.T. Tilakarathne**
Professor, Department of Statistics, University of Colombo
- **Dr. J.H.D.S.P.Tissera**
Senior Lecturer, Department of Statistics, University of Colombo
- **Dr. Hakim. A. Usoof**
Senior Lecturer, Department of Statistics & Computer Science, University of Peradeniya
- **Dr. S.D. Viswakula**
Senior Lecturer, Department of Statistics, University of Colombo
- **Dr. Ruwan Weerasinghe**
Senior Lecturer, University of Colombo School of Computing
- **Mr. Viraj Welgama**
Senior Lecturer, University of Colombo School of Computing
- **Dr. Dilani Wickramaarachchi**
Senior Lecturer, Department of Industrial Management, University of Kelaniya
- **Ms. D. S. Wickramarchchi**
Lecturer, Department of Statistics, University of Colombo
- **Dr. Shanika Wickramasuriya**
Lecturer, Department of Statistics, University of Auckland, Australia
- **Ms. Lakmini K. N. Wijesekara**
Lecturer, Department of Statistics, University of Colombo
- **Ms. Rupika Wijesinghe**
Senior Lecturer, University of Colombo School of Computing
- **Mr. Hiran Wijesinghe**
Assistant Director IT, Sri Lanka Tea Board

Table of Contents

| Title and Author/s | Page No: |
|--|----------|
| Data Science: From 2015 to 2021 and Beyond (Keynote speech) John W. Emerson ("Jay") | 1 |
| Data Science in Mobile Health (Guest speech) Bibhas Chakraborty | 2 |
| Future (Challenges) of Data Science- Quantum Computing (Guest speech) Umashanger Thayasivam | 3 |
| <u>Technical Sessions</u> | |
| Predicting Career Satisfaction of IT Employees: A Case Study on Software Developers Nimasha Arambepola, Lankeshwara Munasinghe | 5 |
| Improving Time Series Anomaly Detection Using an Ensemble Feature Model H. Asela | 6 |
| A Case Study in Financial Fraud Detection using Big Data Analytics W. P. A. Boteju, I. U. Hewapathirana | 7 |
| Prediction of Contact Maps and Family of Proteins using Deep Learning N. H. Charuka , W. A. M. Madhavi | 8 |
| A Study on Vehicle Emission Test in Sri Lanka K. S. P. Fernando, R. A. B. Abeygunawardana | 9 |
| Impact of the dollar rate in prediction of the Colombo Stock Exchange Performance using Machine Learning Techniques Sachith Fernando, N. A. I. Supasan, Ranjan Dissanayake | 10 |
| A Machine Learning Algorithm to Validate Present Data with Historical Data S. R. N. Gunaratne, K. A. D. Deshani, S. Srikathirkamanathan | 11 |

| Title and Author/s | Page No: |
|---|-----------------|
| Identifying the Factors Associated with Infant Mortality in Galle MOH Areas M. Gunawardhana, R. V. Jayatillake, K.T. Samararathna | 12 |
| Seismicity And Stress Patterns Along The Central Tunisia Makrem Harzali, Abdelmalak Mohamed Mansour | 13 |
| Spatial Changes and Relationship Analysis of NDVI and LST Using Satellite Image Datasets in Pottuvil Divisional Secretariat Division, Ampara M. L. M. Hicmathulla, A. W. F. Nafla | 14 |
| Gamified Feedback and Job Satisfaction of Generation Z Software Professional in Post-COVID-19 Environment H. D. Jayathilake, N. P. Barupala, J. A. Jayathilake, H. M. Jayathilake | 15 |
| A Deep Learning Approach to Hate Speech Detection in Sinhala B. A. S. S. B. Jayawardhana, M. R. D. S. G. Punchihewa | 16 |
| Quantitative Evaluation of Equity Research for Accurate Investment Decisions in Future D. R. S. Kalubowila, R. Fernando, S. D. Viswakula | 17 |
| Improving Sinhala Language Modelling through Deep Learning D. S. Kulatunga, A. R. Weerasinghe, E. R. A. D. Bandara | 18 |
| Forecasting the Directional Movement and Value of S & P SL 20 Index Using News and Macro-Economic Indicators M. P. Maha Arachchi, G. P. Lakraj, I. Wijesekara, P. Wijesiriwardana | 19 |
| Multi Target Regression for Click Through Rate Prediction Rajitha Manellanga | 20 |
| Investigating the Chest Imaging Features of Covid-19 Patients using Deep Neural Networks A. B. Meepaganithage, M. G. N. A. S. Fernando | 21 |
| Topic modeling in Youtube Data A. C. Nanayakkara, G. A. D. M. Tennakoon | 22 |

| Title and Author/s | Page No: |
|---|-----------------|
| Predictive Analytics for Machinery Maintenance in the Production of Quality Garments U.M.M.P.K.Nawarathne, Vijani Piyawardana, Jesuthsan Alosius | 23 |
| Study on Part-Time Employment of Science Undergraduates, University of Colombo K. G. J. Nishadini, R. A. B. Abeygunawardana | 24 |
| Machine Learning Approaches to Forecast Inflation in Sri Lanka W. P. I. S. Pathirana, C. D. Thilakaratne | 25 |
| A Comparative Study on Artificial Neural Network Techniques for Short-term Rainfall Forecasting in Colombo, Sri Lanka G. C. J. Piyatilake, K. A. D. Deshani | 26 |
| Game Outcome Predictor and Game Plan Generator for Twenty20 Cricket L. C. P. Pussella, R. M. Silva, W. C. P. Egodawatta | 27 |
| Machine Learning Approach to Detect Tea Plant Diseases in Sri Lanka S. N. K. Rajakaruna, Rangana Jayashanka, H. A. S. G. Dharmarathne | 28 |
| Question Answering for Sinhala Language Using Deep NN Architectures M. W. K. S. Rasantha, Dr. A. R. Weerasinghe, Dr. J. H. D. S. P. Tissera | 29 |
| Algorithm to Detect Doctored Images R. M. D. S. Rathnayaka, S. D. Viswakula, T. A. M. C. Thantriwatte | 30 |
| Developing the Modelling Framework of an Intelligent News Recommendation Engine for Financial Analysts H. M. P. P. K. H. Samarasekara, W. M. N. D. Basnayake, D. Hureekaduwa, B. Weralupitiya | 31 |
| Improving Bank-Customer Efficiency through Automatic Speech Recognition L. D. P. S. Senaratne, H. N. D. Thilini, C. H. Magalla | 32 |
| Vehicle Detection In Satellite Images K. Sutharsan, S. Gamika, Dr. G. P. Lakraj | 33 |

| Title and Author/s | Page No: |
|---|-----------------|
| Linkage Comparison on Agglomerative Clustering for Online Retail Dataset R. M. B. P. M. Uduweriya, N. A. D. N. Napagoda | 34 |
| Building an OCR Tool with a Better Text Recognition for Financial Domain B. L. A. S. Upekshika, S. D. Viswakula, M. F. S. Ahamed, F. L. Muthukumarasamy | 35 |
| Spotting Railway Signs to Build Smart Decision Support Tools in Railway Management Systems P. T. Weerasinghe, M. Shaheer, R. Rimalshan, P. Gunathilake, E. Dayarathne | 36 |
| A Case Study of Creditworthiness Prediction at the Loan Approval B. N. Weralupitiya and R. V. Jayatillake | 37 |
| Quantifying the Heterogeneous Close Contact Patterns in Sri Lanka W. A. P. S. Wickramarachchi, L. Munasinghe | 38 |
| Workflow Composition with Predictive Modelling C. R. Wijesinghe, A. R. Weerasinghe | 39 |
| Developing a Model to Decrease Call Volume Cost of a Telecommunication Company by Analysing Time-to-Failure since the Installment of the SetupBox K. A. C. N. Wijewardana, H. A. S. G. Dharmarathne, P. A. D. L. Samarakoon | 40 |

Data Science: From 2015 to 2021 and Beyond

John W. Emerson (“Jay”)

Department of Statistics and Data Science, Yale University

This talk offers commentary and perspective on the evolution of Data Science through personal examples from Yale and around the world. Some observations I made in 2015 – at a Keynote Address right here in Colombo! – were accurate and, I hope, helpful. But I didn’t place sufficient emphasis on the importance of collaboration. And I didn’t predict the explosion of interest and activity in Data Science that we’ve witnessed in the last 6 years. I will correct those mistakes today in this non-technical talk that I hope will be accessible and enjoyable for everyone in attendance, whether in-person or remote. And unlike my 2015 address, I will not present code, run simulations, or consider topics in high-performance computing. My primary example will be the collaborative Environmental Performance Index project, ranking countries on environmental health and ecosystem vitality. It is a biennial analyses with roots going back to the 2000 Environmental Sustainability Index. I will leave more technical topics in Data Science to other speakers, your technical sessions, and conference workshops.

Data Science in Mobile Health

Bibhas Chakraborty

Centre for Quantitative Medicine and Program in Health Services and Systems Research,
Duke-NUS Medical School

Mobile health (mHealth) is a modern field in health sciences. mHealth interventions, broadly defined as health and behavioural interventions delivered via mobile and/or wearable devices, offer a powerful vehicle to improve health outcomes over a broad spectrum of the population in today's technology-enabled world. In recent years, healthcare providers and governments are excited about the opportunities mHealth offers in terms of improving efficiency and reducing cost of healthcare systems. With the increasing volume, quality and accessibility of mHealth data, statistics/data science has a key role to play in the development of data-driven, personalized, just-in-time adaptive interventions (JITAs), and thereby in advancing human health. JITAs often consist of sequences of text-messages or other prompts that aim to adapt to an individual's time-varying context (e.g., location, weather, past behaviour, past treatments, evolving health status, etc.), and thus optimizing them is a multi-stage decision problem.

In this talk, we will present an innovative experimental design called the micro-randomized trial (MRT), the current gold-standard data source for developing optimal JITAs. This design involves sequential, within-individual randomizations, and aims to estimate proximal causal effects of "push"-type interventions that constitute JITAs, e.g., motivational text-messages to promote physical activity or other healthy behaviours. JITAs are usually learned via offline longitudinal data analysis after the MRTs are completed. While this is a sound randomization-based approach, trial participants do not get the benefit of receiving the optimized JITAs. So, in order to make the MRT design more adaptive and thereby to deliver better interventions to the trial participants, we will consider contextual multi-armed bandit (MAB) type online reinforcement learning algorithms from the field of artificial intelligence, wherein a system must repeatedly choose among a set of actions (interventions) in an effort to maximize a reward (outcome). We will focus on a popular MAB algorithm called Thompson Sampling (TS) – deeply grounded in Bayesian statistics, particularly useful in case of small amounts of data, and often operationalized via a regression model for the reward. We will illustrate these ideas by discussing an mHealth case study in detail, and presenting real data analysis.

Future (Challenges) of Data Science- Quantum Computing

Umashanger Thayasivam

Department of Mathematics, Rowan University

Every day we create more than 2.5 quintillion data bytes of data, and this number is expected to grow to 3.5 quintillion data by 2025 especially with the rise of IoT (Internet of Things) and 5G capabilities. Data Science and artificial intelligence (AI) are some of the ways to help manage and analyze data for competitive advantage, however continued revolution and the desire for profound intuitions may make data more complex for organizations to collect and investigate. Quantum computing has the potential to be the most disruptive technology of the 21st century. This is a different form of computation that builds from quantum mechanics, and it promises to solve problems we cannot solve with classic computers.

Recently, Google announced that it had achieved quantum supremacy with its “Sycamore” quantum computer that can solve complex algorithms unsolvable by any other computer today. This milestone raises fundamental questions about how quantum computing can be used and how it will affect initiatives in the digital era. As classical binary computing reaches its performance limits, quantum computing is becoming one of the fastest-growing digital trends and is predicted to be the solution for the future’s big data challenges. Though quantum computing is still just on the horizon, the U.S. plans to invest more than \$1.2 billion toward quantum information over the next 10 years in a race to build the world’s best quantum technology. Quantum computers will perform incredibly complex calculations in a couple of seconds when it would take a traditional computer a few thousand years to do the same. Google’s Sycamore was able to perform a calculation in 200 seconds that would have taken today’s fastest supercomputer 10,000 years to settle.

The future of data science will be impacted by quantum computing and some radical transformations in data science are necessary to develop quantum computing algorithms. Scientists hope quantum computing can help pave the way for the future of data science.

ICDS2021 - Technical Sessions

Center for Data Science (CDS)

Data Science is an emerging field that has capacity to grow and provide many opportunities for research and collaborative projects. Therefore, in 2016, **Center for Data Science** was established under the Department of Statistics, Faculty of Science, University of Colombo. The Center strives to facilitate research and development in Data Science in Sri Lanka through collaborations with local and international expertise both from academia and industry. It also conducts training programs, workshops and public talks to disseminate knowledge in Data Science and increase awareness of Data Science among the community. The Center promotes partnerships with local industry through consultancy projects providing them with technical expertise while enhancing skills of the students and academics in the application of Data Science techniques in the real world.