

CHRIST

(DEEMED TO BE UNIVERSITY)
BANGALORE · INDIA

Department of Computer Science and Engineering
Computer Society of India
Presents
Workshop
On
Data Science



Date & Time : 10th Feb 2020 @ 4.30 - 6.00 PM



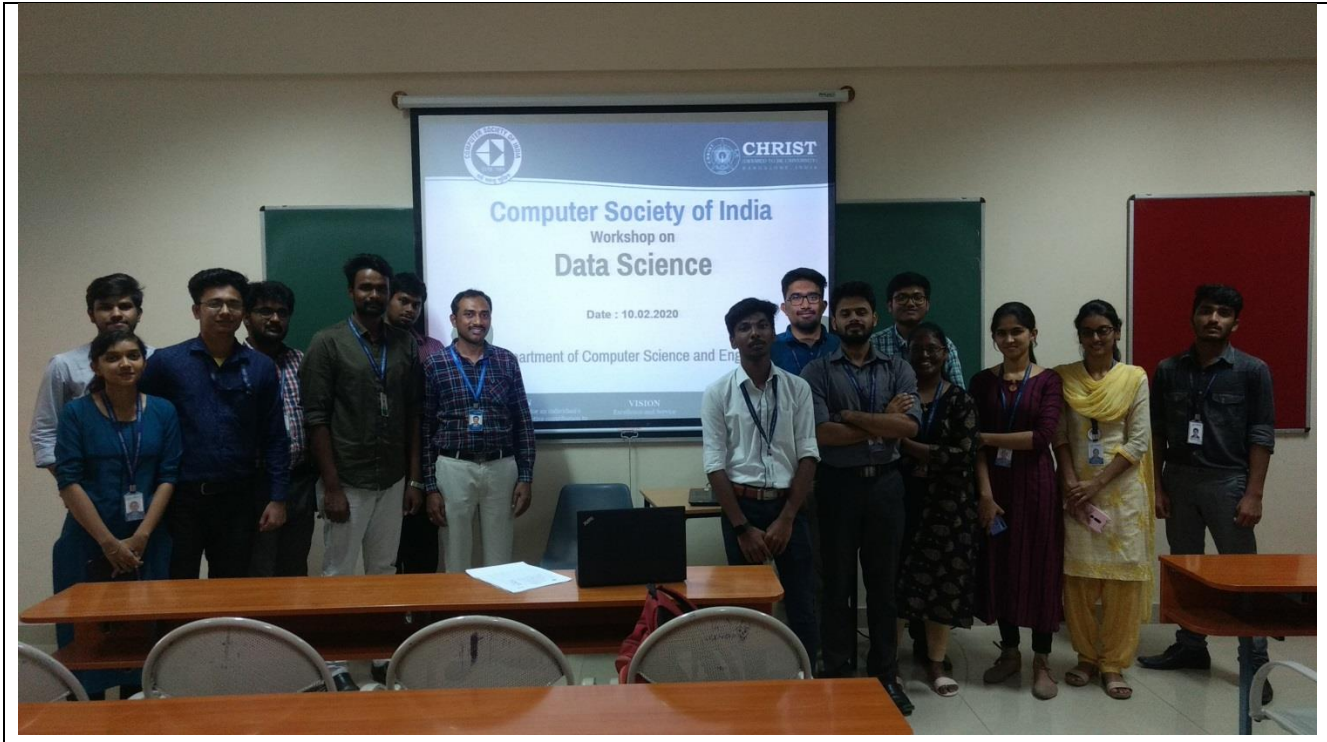
Scan to Register

Venue

Room No 218, First Floor, Second Block,
Kengeri Campus.

Event Coordinator : Dr. Jayapandian N (jayapandian.n@christuniversity.in)

Department Vision: " To Fortify Ethical Computational Excellence"





CHRIST
(DEEMED TO BE UNIVERSITY)
BANGALORE · INDIA

Department of Computer Science and Engineering
Computer Society of India
Event Report
Workshop
On
DATA SCIENCE

CONDUCTED BY : COMPUTER SOCIETY OF INDIA
DATE & TIME : 10TH FEB, 2020. 4:30-6:00 PM
VENUE : #218, 2ND BLOCK, CHRIST - KENGERI
CAMPUS
RESOURCE PERSONS: Blessin Raju, Darel V Johny, Prabavathi
S
EVENT COORDINATOR: Dr JAYAPANDIAN N.
Total No. of Participants: 16

Data science is an inter-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data. Data science is related to data mining and big data.

Data science is a "concept to unify statistics, data analysis, machine learning and their related methods" in order to "understand and analyze actual phenomena" with data. It employs techniques and theories drawn from many fields within the context of mathematics, statistics, computer science, and information science. Turing award winner Jim Gray imagined data science as a

"fourth paradigm" of science (empirical, theoretical, computational and now data-driven) and asserted that "everything about science is changing because of the impact of information technology" and the data deluge. In 2015, the American Statistical Association identified database management, statistics and machine learning, and distributed and parallel systems as the three emerging foundational professional communities.

The term "data science" has appeared in various contexts over the past thirty years but did not become an established term until recently. In an early usage, it was used as a substitute for computer science by Peter Naur in 1960. Naur later introduced the term "datalogy". In 1974, Naur published *Concise Survey of Computer Methods*, which freely used the term data science in its survey of the contemporary data processing methods that are used in a wide range of applications.

The modern definition of "data science" was first sketched during the second Japanese-French statistics symposium organized at the University of Montpellier II (France) in 1992. The attendees acknowledged the emergence of a new discipline with a specific focus on data from various origins, dimensions, types and structures. They shaped the contour of this new science based on established concepts and principles of statistics and data analysis with the extensive use of the increasing power of computer tools.

In 1996, members of the International Federation of Classification Societies (IFCS) met in Kobe for their biennial conference. Here, for the first time, the term data science is included in the title of the conference ("Data Science, classification, and related methods"), after the term was introduced in a roundtable discussion by Chikio Hayashi.

There are several data science applications that completely changed the way those things were handled or managed prior to the data-science era. Data science has left an indelible mark on several industries, including finance and banking, healthcare, transportation, manufacturing, and advertising amongst other. However, its impact on healthcare and continuous contribution in improving patient outcomes and healthcare delivery has been too immense to not get a mention. And there are more Data Science Applications in the pipeline that are going to transform the way the healthcare industry functions across the world. Some of the areas in healthcare that data science has had a huge influence on include:

Medical image analysis

Drug recovery

Genomics and genetics

Predictive diagnosis modelling

Health bots

Going by what we have been to do using data science, the future seems very bright. Data science could have a huge role to play in how care is provided to patients, how their records are securely maintained and how different stakeholders work together to improve the entire healthcare landscape. Data science is the technology of the future and is here to stay. The growth it has in recent years is testimony to its relevance and effectiveness in the current times but in the future as well.

Web Scraping (also termed Screen Scraping, Web Data Extraction, Web Harvesting etc.) is a technique employed to extract large amounts of data from websites whereby the data is extracted and saved to a local file in your computer or to a database in table (spreadsheet) format.

Data displayed by most websites can only be viewed using a web browser. They do not offer the functionality to save a copy of this data for personal use. The only option then is to manually copy and paste the data - a very tedious job which can take many hours or sometimes days to complete. Web Scraping is the technique of automating this process, so that instead of manually copying the data from websites, the Web Scraping software will perform the same task within a fraction of the time.

The image shows a screenshot of the RStudio software interface. The main window is divided into four panes:

- Source Editor (R script):** Contains R code for biomass calculation per tree across various plots. The code includes comments and function calls like `plot(kalimantanSdbh, kalimantanSdbh, col=1)` and `bio.plot.brown<-as.data.frame(tapply(kalimantanSdbh, list(kalimantanSdbh$plot_id, kalimantanSdbh$tree_id), FUN=function(x) {sum(x)}, MARGIN=2))`.
- Console (R console):** Shows the execution of the R script, with output indicating the calculation of biomass per plot and the writing of a CSV file named 'kalimantanSdbh.csv'.
- Environment (R environment):** Lists the objects created in the R session, including `h1.trees` (716 obs. of 23 variables), `kal.plot` (94 obs. of 18 variables), `kalimantan` (1993 obs. of 44 variables), and `list.plots` (59 obs. of 19 variables).
- Plots (Graphical output):** Displays a box plot titled 'Biomass estimation per plot with different models'. The y-axis is labeled 'Biomass (Mg/ha)' and ranges from 100 to 500. The plot shows the distribution of biomass for different plots, with individual data points overlaid on the box plots.

K. S. Lakshman
 Head of the Department
 Computer Science and Engineering
 CHRIST (Deemed to be University)
 Bengaluru - 560 074



Workshop on
Data Science



CHRIST
BANGALORE - INDIA

Computer Society of India & Department of CSE
CHRIST (Deemed to be University), Bangalore-560074, India
10th Feb 2020

S.NO	Name of the Participants	Signature
1	ANUSHA-P.V (1860382)	
2	Sushma.S (1860401)	
3.	Sahansh Parmecha (1860424)	
4.	Sonam Pal (1861028)	
5.	Shivam Khattri (1660381)	
6.	Kartikaya Namdev (1760493)	
7.	ALLAM VINEETH REDDY [1760474]	
8.	ANNAAJUALA ABHIRAM SAI 1760661	
9	Harish prabu-S.P	
10.	G.Thejo Vamsi	
11.	BASIL BENEDICT VICTOR (1960318)	
12.	SHAIK SADIQ ALI (1960282).	
13,	S.M.d.k.N.U. Affan Ahmed	
14	Sushma Patil (1961044)	
15	Soujradip.	
16	Uttara Nambiar (1761026)	