



SOMAIYA
VIDYAVIHAR

K J Somaiya Institute of Technology
(Formerly known as K J Somaiya Institute of Engineering and Information Technology)
An Autonomous Institute Permanently Affiliated to University of Mumbai.



Report of Webinar on “Dataset Creation”



SOMAIYA
VIDYAVIHAR

K J Somaiya Institute of Technology
(Formerly known as K J Somaiya Institute of Engineering and Information Technology)
An Autonomous Institute Permanently Affiliated to University of Mumbai.



S4DS-KJSIT STUDENT CHAPTER IN ASSOCIATION WITH
DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA
SCIENCE IS ORGANIZING A WEBINAR ON

DATASET CREATION



About Speaker:

Dr. Amol Goje is the President of Society for Data Science, Director at VIIT Baramati and Founder of Cognier Academy which is an Educational institution for Data Analytics. He is also the recipient of the Ashoka Fellow award in 2002, Government of Maharashtra Award for IT entrepreneurship, and the Manthan Award for Technology content for schools and community radio.

5 May 2023 | Friday | 9.30 am | Zoom

Name of the event: Dataset Creation

Date & Time: 5th May 2023, 9:30 AM

Description: The student chapter of KJSIT-Society for Data Science KJSIT-S4DS in association with Department of Artificial Intelligence and Data Science organized a Webinar on Dataset Creation. The speaker personage for this webinar was Dr Amol Goje, the President of Society for Data Science(S4DS). The webinar was conducted on Zoom platform on 5th May 2023 and provided an overview on the creation of datasets.

Objectives:

- To educate students about various methods to create a dataset.
- To educate students about accessing various online data sources.

Beneficiaries: Students of KJSIT

No. of Participants: 60



SOMAIYA
VIDYAVIHAR

K J Somaiya Institute of Technology
(Formerly known as K J Somaiya Institute of Engineering and Information Technology)
An Autonomous Institute Permanently Affiliated to University of Mumbai.



Venue/Mode: Online-Zoom Meeting

Duration: 2 hours

Speaker Personage: Dr. Amol C. Goje is the President of Society of Data Science. He was the Director at Vidya Pratishthan's Institute of Information Technology (VIIT), Baramati, Pune for the last 20 Years. Currently he consulting companies in the field of technology inclusion and infrastructure. He has a total of over twenty five years of experience in the field of Information and Computer Technology (ICT). He has developed many systems for the University. Dr. Amol's main area of interest is to work for underprivileged people in the rural part of India. In his nineteen years as a Director, he has designed and implemented numerous path-breaking, innovative and cost effective solutions. Some solutions to name a few are – number of computer labs, which are economically sustainable to the rural schools and colleges. Dr Amol's main innovation is Computer Mobile Van. He has done lot of research work in Information Technology and its application for rural community. In appreciation to his exemplary work, Dr Amol has received the Ashoka Fellow award in the year 2002. He is engaged as a technical advisor on many government and non-government organizations.

He was also a member of the Planning Commission, Government of India in Information Technology division. In his tenure as the Director at VIIT, he has organized 8 international conferences in Baramati under the aegis of Baramati Initiatives. Dr. Goje has received the Marathwada Bhushan Award for spreading the IT education to the masses. He is the member of the Working group on Agricultural Extension Constituted by Planning Commission government of India. He is key player in setting up the Community Training and Learning Centers (CTLC) in Maharashtra. In this project, VIIT is providing Computer Training to the women from the self help groups (SHG). Dr Goje has received the Manthan (AIF) award successively for two years in Year 2005 and 2006. Government of Maharashtra "Maharashtra IT award (IT HRD) in 2008. He is also the President of Community Radio Association (CRA) of India. Recently he has been bestowed with the honour of 'Best Director' of Pune University -2016-17.



SOMAIYA
VIDYAVIHAR

K J Somaiya Institute of Technology

(Formerly known as K J Somaiya Institute of Engineering and Information Technology)
An Autonomous Institute Permanently Affiliated to University of Mumbai.



Key Points: The webinar on “Dataset Creation” started with the host, Mr. Nishant Moghe, welcoming all the dignitaries and the students followed by the Somaiya prayer. Then, the Principal, Dr Suresh Ukrande, spoke a few words about the college and welcomed all the official S4DS members. The event proceeded further with Dr Milind Nemade, Convener of KJSIT-S4DS and Head of Department of Artificial Intelligence and Data Science, introducing the speaker for the event, Dr Amol Goje and gave a brief overview of the topic “Dataset Creation”. Then Dr Amol started the webinar by giving a gist of the flow of the presentation. The webinar comprised of the introduction to datasets, definition of dataset, type of datasets, purpose and scope of dataset following by the data preprocessing methods, data labelling, data integration and finally concluding with data quality control and documentation. Further there was a doubt solving session where Dr Amol addressed few doubts of the students regarding the session. The event concluded with the vote of thanks given by Prof. Sejal Shah followed by the National Anthem.

Outcome(s): The event ended on a high note with all the students aware about the creation of dataset. The insightful session gave the students a deep understanding of the various aspects of creating datasets, from gathering, preprocessing data to organizing and analyzing it. The speaker's expertise and experience in the field was evident, and their ability to simplify complex concepts was truly commendable. The webinar has provided our students with practical knowledge and skills that are highly relevant in today's data driven world. The concepts taught during the webinar will undoubtedly help them in their academic and professional pursuits.

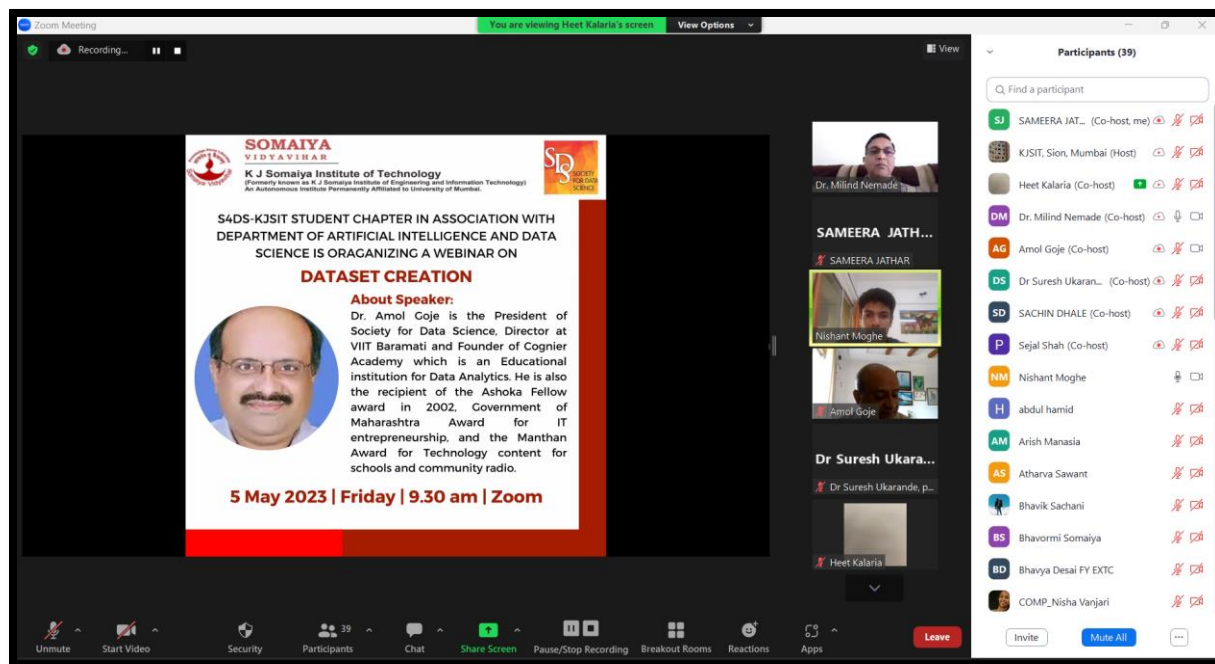


SOMAIYA
VIDYAVIHAR

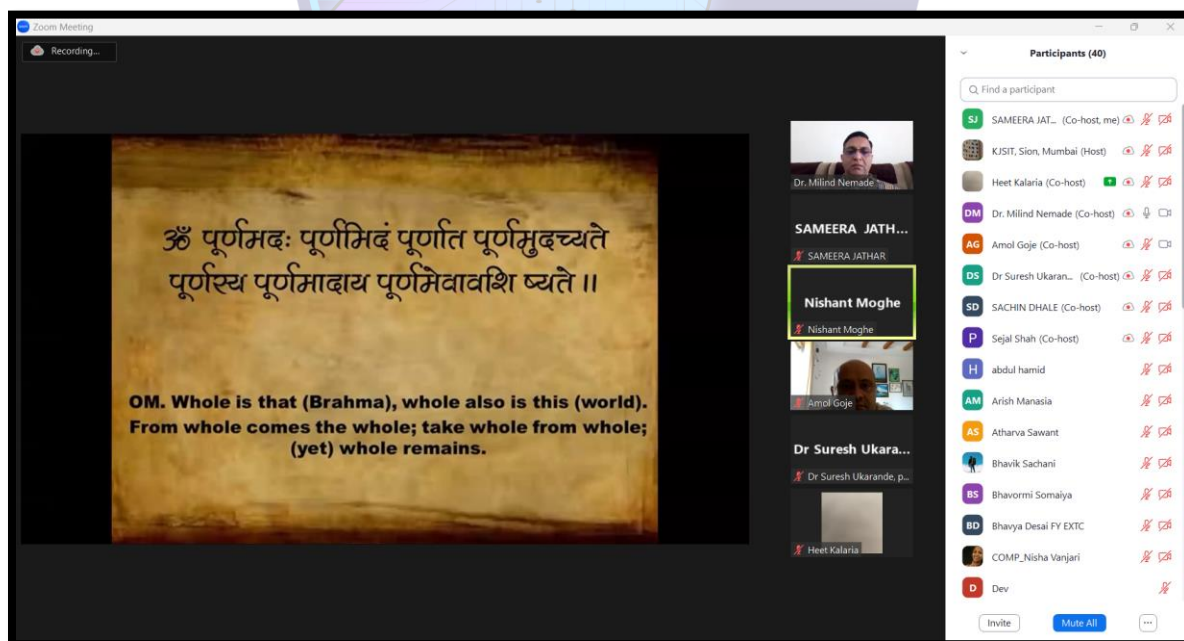
K J Somaiya Institute of Technology
(Formerly known as K J Somaiya Institute of Engineering and Information Technology)
An Autonomous Institute Permanently Affiliated to University of Mumbai.



Glimpse of the Event:



Host welcoming everyone

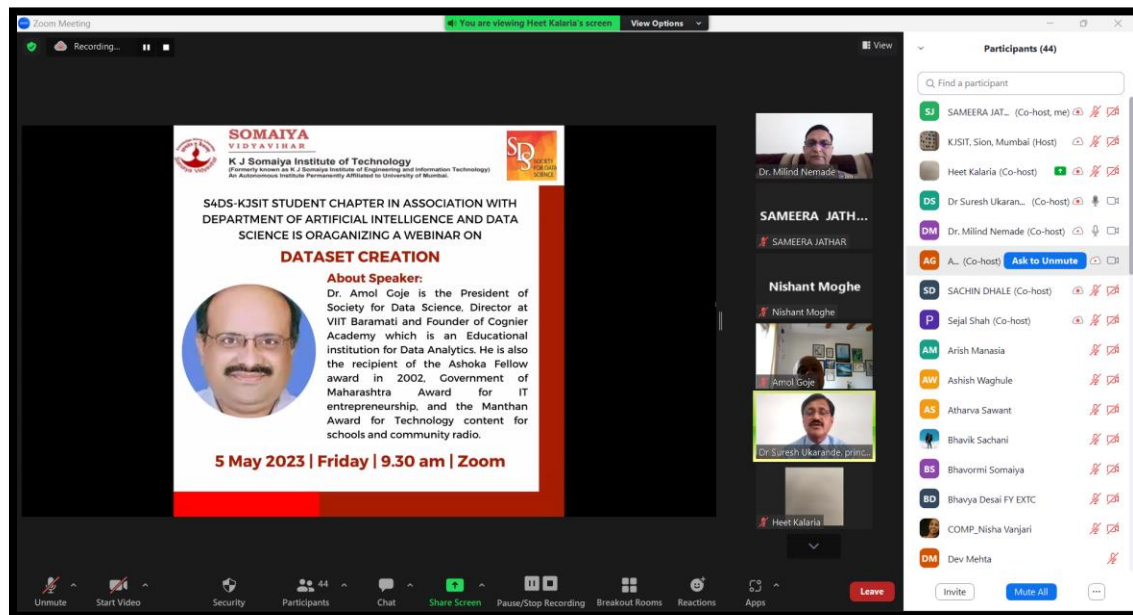


Somaiya Prayer

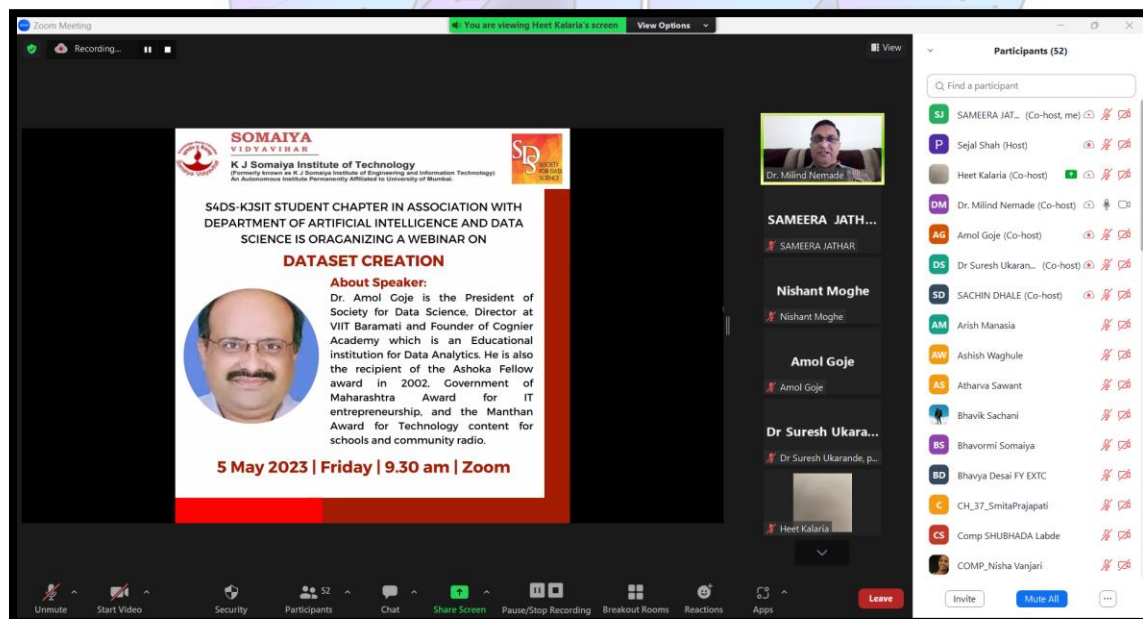


SOMAIYA
VIDYAVIHAR

K J Somaiya Institute of Technology
(Formerly known as K J Somaiya Institute of Engineering and Information Technology)
An Autonomous Institute Permanently Affiliated to University of Mumbai.



Principal addressing the students



Introduction of Chief Guest by Dr Milind Nemade



SOMAIYA
VIDYAVIHAR

K J Somaiya Institute of Technology
(Formerly known as K J Somaiya Institute of Engineering and Information Technology)
An Autonomous Institute Permanently Affiliated to University of Mumbai.



Zoom Meeting You are viewing Amol Goje's screen View Options

Recording... View

Introduction

Creating a dataset involves collecting, preprocessing, transforming, integrating, and labeling data to create a structured format that can be used for data analysis, research, or machine learning.

Here's a brief overview of the steps involved in creating a dataset:

1. Identify the purpose and scope of the dataset: Determine what the dataset will be used for and what data will be included.
2. Collect data: Collect data from various sources, such as surveys, experiments, or web scraping.
3. Preprocess data: Clean and preprocess the data to ensure it's high-quality and relevant to the dataset's purpose and scope. This may involve data cleaning, normalization, and feature selection.
4. Transform data: Transform the data into the desired format for the dataset. This may involve encoding categorical data to numerical data or converting data to a specific file format.
5. Integrate data: Integrate data from different sources into a single dataset. This may involve merging or joining different data sources.
6. Label and annotate data: Label and annotate the data to ensure it's usable for the dataset's purpose and scope. This may involve using human annotation or automated labeling techniques.
7. Quality control: Ensure the dataset is error-free, complete, and consistent to ensure it's suitable for accurate data analysis and machine learning.

Amol Goje

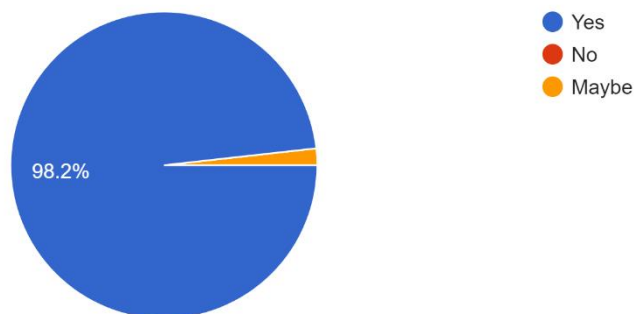
Unmute Start Video Security Participants (59) Chat Share Screen Pause/Stop Recording Breakout Rooms Reactions Apps Leave

The speaker addressing the students

Feedback:

Was the session useful?

57 responses





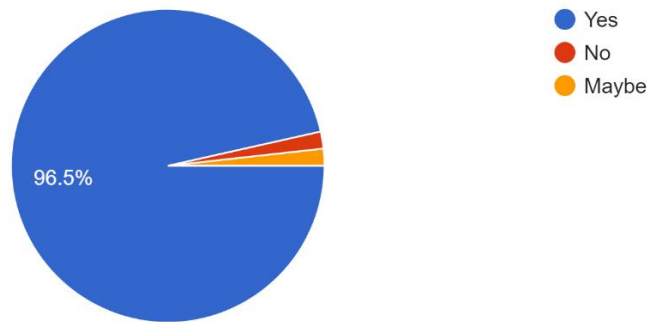
SOMAIYA
VIDYAVIHAR

K J Somaiya Institute of Technology
(Formerly known as K J Somaiya Institute of Engineering and Information Technology)
An Autonomous Institute Permanently Affiliated to University of Mumbai.



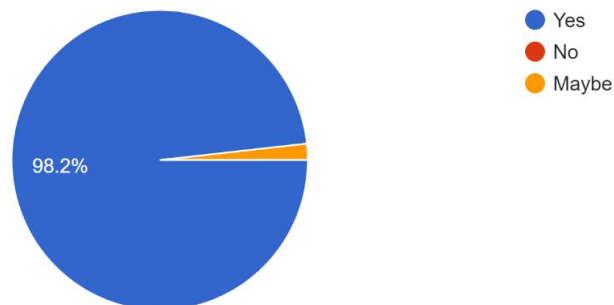
Were the objectives of the session achieved?

57 responses



Did you find the session useful for your professional/academic activities/career?

57 responses





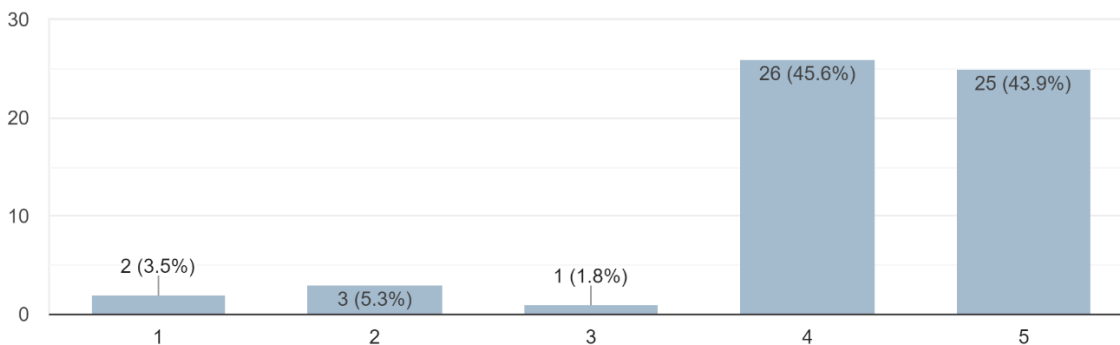
SOMAIYA
VIDYAVIHAR

K J Somaiya Institute of Technology
(Formerly known as K J Somaiya Institute of Engineering and Information Technology)
An Autonomous Institute Permanently Affiliated to University of Mumbai.



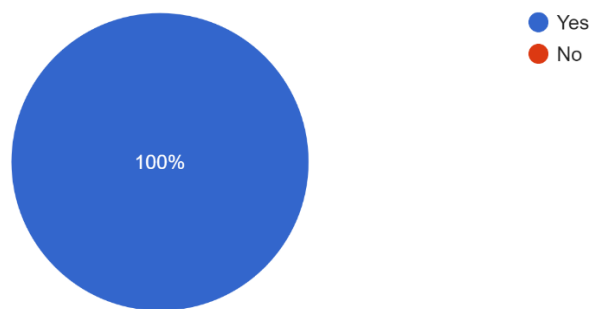
On a scale of 5, how would you rate this session?

57 responses



Would you like S4DS-KJSIT to conduct more such sessions?

57 responses



Event Coordinator(s): Prof Sejal Shah, Faculty Coordinator, KJSIT-S4DS

Report Prepared By: Ms. Sameera Jathar, Chief Executive Officer, KJSIT-S4DS

Report Verified By: Mr. Heet Kalaria, President, KJSIT-S4DS

Report Approved By: Dr Milind Nemade, Counsellor, KJSIT-S4DS



SOMAIYA
VIDYAVIHAR

K J Somaiya Institute of Technology

(Formerly known as K J Somaiya Institute of Engineering and Information Technology)
An Autonomous Institute Permanently Affiliated to University of Mumbai.

