



Innovation & Leadership  
www.i2it.edu.in

Hope Foundation's

**International Institute of Information Technology (I²IT)**

**Hinjawadi, Pune- 411057**

## Report on Guest/Expert Lecture

Organized by

**Department of Electronics & Telecommunication**

1. **Program type:** Guest/Expert Lecture on Data Science
2. **Date & Time:** 05/10/2020 & 9.30am to 11.30 am
3. **Venue:** online-Google meet
4. **No. of students registered/appeared:** 201
5. **Target students:** SE E&TC, TE E&TC, BE E&TC
6. **Fee Details:** FREE
7. **Instructor details:** Ms. Hima Patel, Ms. Nitin Gupta, Mr. Shashank Mujumdar  
IBM Research Lab, India  
himapatel@in.ibm.com, ngupta47@in.ibm.com, shamujum@in.ibm.com  
9611133668.
8. **Objectives :**
  - 1) Explore supervised and unsupervised learning paradigms of machine learning used for-regression and classification.
  - 2) To design and analyze various machine learning algorithms using neural networks
  - 3) To explore Deep learning techniques and various feature extraction strategies.
9. **Outcomes:**
  - 1) To compare and contrast pros and cons of various machine learning techniques and to get an insight of when to apply a particular machine learning approach.
  - 2) To mathematically analyze various data cleaning approaches
  - 3) To implement unstructured data and applications.

**Photo gallery:**

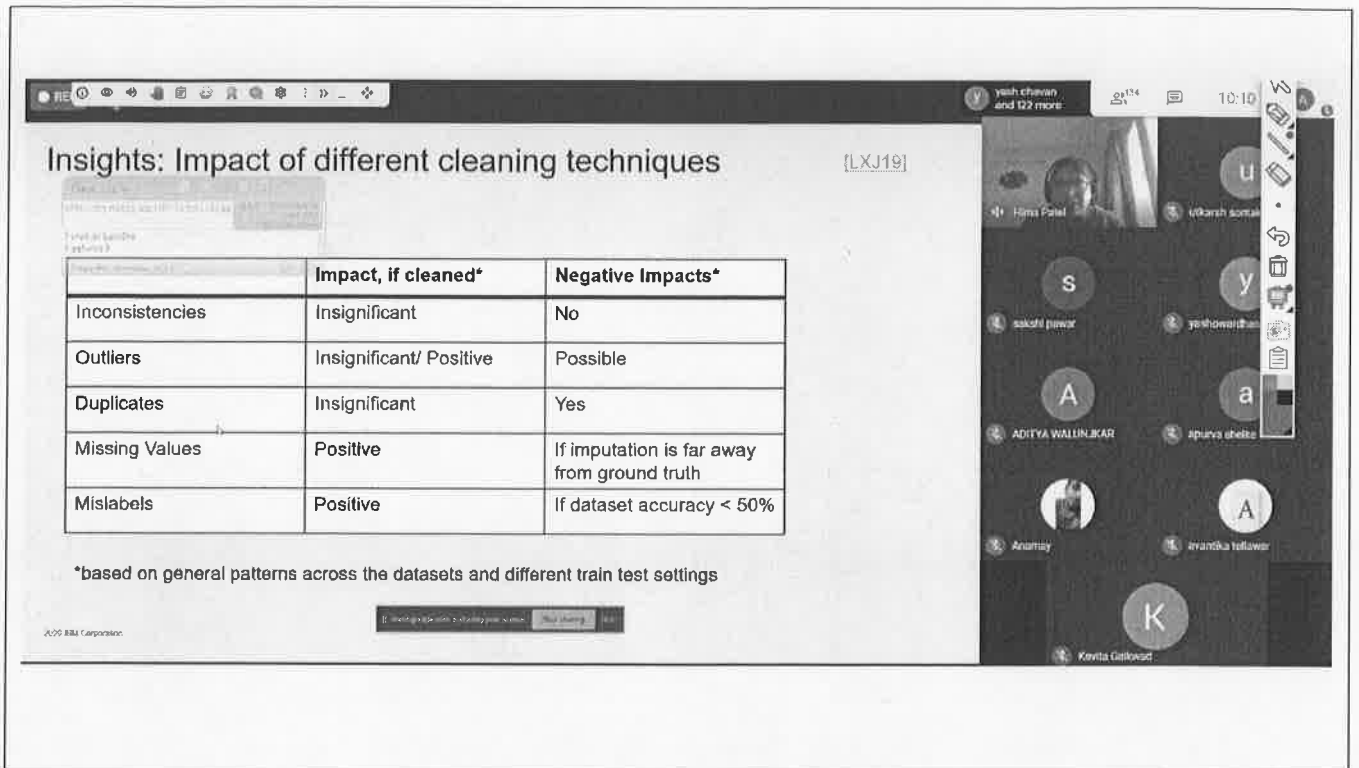


Image 1: Ms. Hima Patel explaining data cleaning techniques

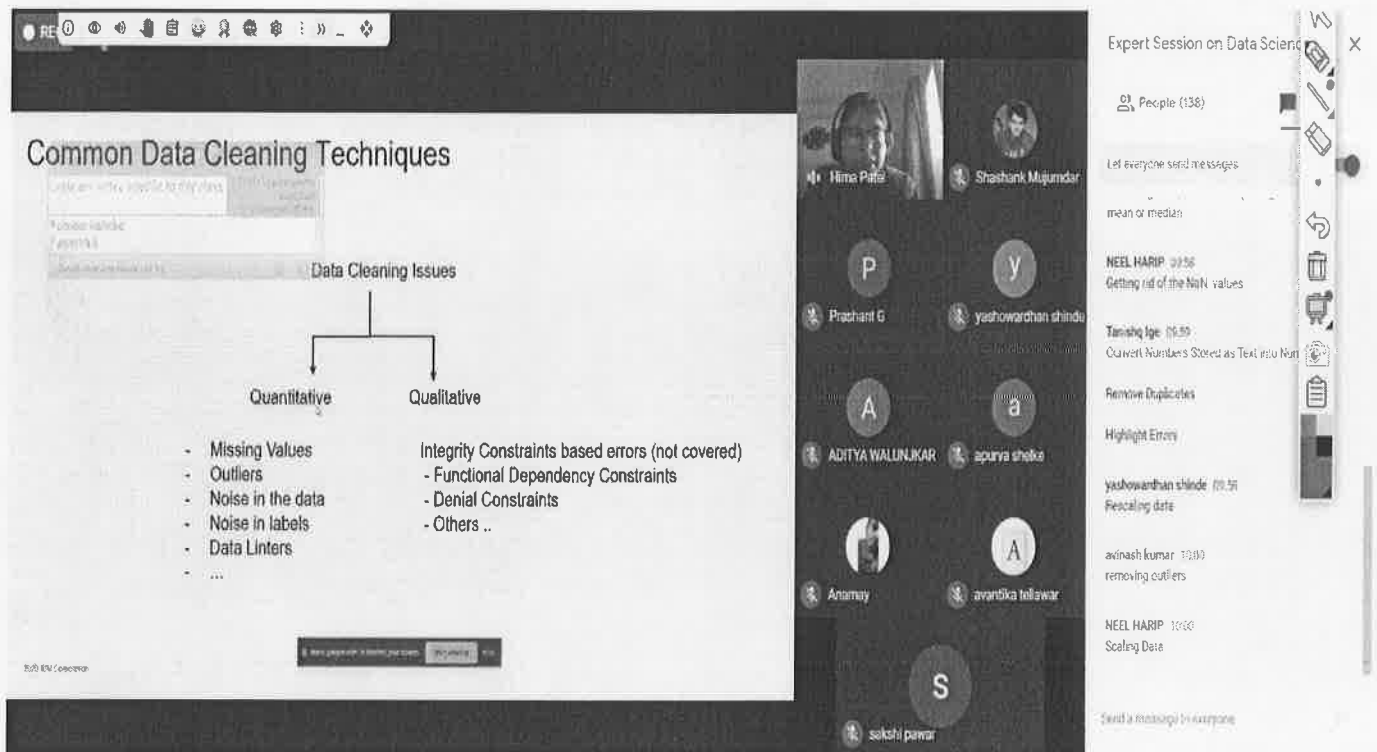


Image 2: Ms. Hima Patel Explaining common cleaning techniques

**Filter Based Approaches**

Training Instances → Filter → "Correctly Labeled" Training Instances → Learning Algorithm → Classifier

Source: Brodley et al. 1998, Identifying mislabeled training Data

- Identifying mislabelled training data [EF99]**
  - Published in Journal of artificial intelligence research, 1999
  - Train filters on parts of the training data in order to identify mislabeled examples in the remaining training data
  - Voting schemes used:
    - Majority Voting
    - Consensus Voting
- Finding label noise examples in large scale datasets [EGH17]**
  - Published in International Conference on Systems, Man, and Cybernetics, 2017
  - Two-level approach
    - Level 1 – Train SVM on the unfiltered data. All support vectors can be potential candidates for noisy points.
    - Level 2 – Train classifier on the original data without the support vectors. Samples for which the original and the predicted label does not match, are marked as label noise.
- On the labeling correctness in computer vision datasets [ARK18]**
  - Published in Interactive Adaptive Learning 2018
  - Train CNN ensembles to detect incorrect labels
  - Voting schemes used:
    - Majority Voting

Image 3: Mr. Nitin gupta explain the approaches for filtering in Data Science

*all*

Prof. Ashvini Kulkarni  
Event Coordinator

*Rhatekala*

Dr. Risil Chhatrala  
Head of Department