



Gandhinagar Institute of Technology

A Report on
“Demystifying Data Science using Machine Learning”
Sponsored by GUJCOST and DST
(27-28 October, 2020)

Objective:

Machine Learning is not only for effective data collection, but to make use of the ever increasing amounts being gathered by manipulating and analyzing it without heavy human input. This webinar aims to give hands on training to the faculty members, research scholars and PG students on various aspects of Machine Learning and Data Science.

About Webinar:

The program will provide a valued platform for the industry specialists and academicians from institutes to exchange and explore the latest views on technological developments in the field of Information Technology. 21st century poses a major technological advancement in computer science area. It is being very difficult for Researcher to define the scope of research and to choose the correct technology for their work. This webinar is primarily targeted to benefit motivated Researchers, Faculty members, Masters Students and PhD candidates who work in the field of Cloud Computing, Machine Learning and Deep Learning. These technologies have been buzz word now a days but the main challenge is to understand the actual working and implementation in each of them. This webinar has provided theoretical knowledge regarding subject but also provide hands on sessions in each area. Total 38 participants from different colleges like L D College of Engineering, Indus University, Ganpat University, Shankersinh Vaghela Babu Institute of Technology and many more. The attendees have actively participated, appreciated and gave excellent feedback about experts and program. Experts of this webinar also gave their positive reviews for organizing such program.

A leader is someone who creates an event and the team members essentially bring life to it, No event can shape without a hardworking leader. It was the dedication and enthusiasm of the coordinators Prof. Prakash B. Patel and Prof. Anirudhdha M. Nayak who have successfully coordinated the whole event under the patronage of Dr. H N Shah, Director. Above and beyond, let's not forget the backbone of this event conveners Prof. Rahul A. Vaghela (HOD, IT), Prof. Archana A. Singh (HOD, CE), Prof. Madhuri R. Chopade (GIT, IQAC) and department faculties. Prof. Alpa K. Oza has well moderated the overall program. GIT is highly obliged with Sponsors of this event GUJCOST and DST. The event got enlighten with knowledge of experts Dr. Hitesh R. Chhinkaniwala (Associate Professor, HOD-ICT, AIIE, Ahmedabad), Prof. Priyank B. Thakkar (Associate Professor, Nirma University), Dr. Chintan V. Shah (Data Scientist, Corpository, Ahmedabad) and Dr. Mayuri A. Mehta (Professor and PG In-Charge, CE Department, Sarvajank College of Engineering and Technology, Surat). The entire event was not possible without support of honorable trustees and management of Gandhinagar Institute of Technology.

Day-1 (27/10/2020)

Time (09:00 am to 03:45 pm)

Webinar started with Inaugural Speech of Prof. Alpa K. Oza, Assistant Professor of Gandhinagar Institute of Technology. Prof. Alpa K. Oza has welcomed all dignitaries. Prof. Rahul Vaghela, HOD IT brief about Data Science and data as one of the important features of every organization because it helps business leaders to make decisions based on facts, statistical numbers and trends. This growing scope of data, data science came into picture which is a multidisciplinary field. It uses scientific approaches, procedure, algorithms, and framework to extract the knowledge and insight from a huge amount of data. The extracted data can be either structured or unstructured. Data science is a concept to bring together ideas, data examination, Machine Learning, and their related strategies to comprehend and dissect genuine phenomena with data. Data science is an extension of various data analysis fields such as data mining, statistics, predictive analysis and many more. Data

Science is a huge field that uses a lot of methods and concepts which belongs to other fields like information science, statistics, mathematics, and computer science. Some of the techniques utilized in Data Science encompasses machine learning, visualization, pattern recognition, probability model, data engineering etc.

Speaker 1: Prof Anirudhdha M. Nayak (Assistant Professor, Information Technology Department, Gandhinagar Institute of Technology)

Time: 9:30 am to 10:30 am

Prof. Anirudhdha M. Nayak is working as Assistant Professor, Information Technology Department, Gandhinagar Institute of Technology. He has more than 7 years of experience in academia. He has expertise in the area of Blockchain, Image Processing, Data Mining, Machine learning and Data Science. Prof. Anirudhdha M. Nayak discussed about Data Science, its overview and its applications. Starting with introduction, he brief about Big Data and its types. Then introduce Data Science and Data Scientists. He brief about what are the roles of data scientists and how data science can be used with AI, Machine Learning and Deep Learning. Also brief the various data science tools and its applications. At last he concluded with the issues of data science and what will be scope of data science in future.

The screenshot shows a Google Meet interface. The main window displays a presentation slide titled "Two Day State Level Webinar On 'Demystifying Data Science using Machine Learning'". The slide lists sponsors: Gujarat Council on Science and Technology, Internal Quality Assurance Cell (IQAC, GIT), and Department of Science & Technology, Government of Gujarat. It also mentions the organizing department: Information Technology Department & Computer Engineering Department, Gandhinagar Institute of Technology. The event is scheduled for Day 1, 27th & 28th October 2020, from 9:30 AM onwards. A registration link is provided: <https://forms.gle/Sfupa2NSjwvz7Lp6>. The bottom of the screen shows a row of participant avatars, including Hardikkumar B., Dr. Nisha Khur..., Bhumi Shah, guest 681, Sejal Bhavsar, and Alpa Oza. The right sidebar shows meeting details, including a list of participants: Krishna Hingrajliya (You), Alpa Oza, Alpa Oza (Presentation), Anirudhdha Nayak, Bhavana Hotchandani, Bhumi Shah, Birendrasinh Zala, and DARSHIL PANCHAL.

The screenshot shows a Google Meet interface. The main window displays a presentation slide titled "Data Science Data Science Is Multidisciplinary". The slide features a diagram illustrating the interdisciplinary nature of data science, with overlapping circles representing various fields: Statistics, Pattern Recognition, Neurocomputing, Machine Learning, AI, Data Mining, Databases & Data Processing, KDD, Problem Solving, Business Analysis, Business Strategy, and Domain Knowledge. The diagram is credited to Brendan Tierney, 2012. The bottom of the screen shows a row of participant avatars, including Dr. Nisha Khur..., DHIRUVISH PA..., Khushbu Maur..., Sejal Bhavsar, Anirudhdha Na..., and Anirudhdha Na... The right sidebar shows meeting details, including a list of participants: Krishna Hingrajliya (You), Alpa Oza, Alpa Oza, Anirudhdha Nayak, Anirudhdha Nayak, Anirudhdha Nayak (Presentation), Bhavana Hotchandani, Bhumi Shah, and Birendrasinh Zala.

Speaker 2: Dr. Mayuri A. Mehta, Professor and PG In-Charge, Computer Engineering Department, Sarvajanik College of Engineering and Technology, Surat

Time: 10:45 am to 12:45 pm

Dr. Mayuri is a passionate learner, teacher and researcher. She is working as a Professor and PG In-Charge in the Department of Computer Engineering, Sarvajanik College of Engineering and Technology, Surat, Gujarat. She has 20 years of teaching experience including 11 years of research experience. Her areas of teaching and research include Data Science, Machine Learning & Deep Learning, Health Informatics, Computer Algorithms, and Python Programming. Her AI-powered Healthcare project was approved by Indian Council of Medical Research (ICMR) for funds. She has received funds several times from Gujarat Council on Science and Technology (GUJCOST) for organizing various workshops and trainings. She is invited to conduct tutorial/workshop, to deliver a talk or to organize a special session on 'AI in Healthcare' in several International Conferences (Recent conferences: IEEE CONECCT 2019 (Bangalore), ODSC 2019 (Bangalore), IGNIS 2019 (Vidyanagar), IEEE TENCON 2019 (Kochi), BDA 2019 (Ahmedabad), WiDs 2020 (Noida), IEEE INDISCON 2020 (Visakhapatnam, BDA 2020 (Delhi)) as well as in several FDPs/STTPs.

The screenshot shows a Google Meet window with the title "Dr Mayuri Mehta is presenting". The main content is a slide titled "Outline" with five numbered points: 01 My Background, 02 Data Science: Definition, Applications & Process, 03 Machine Learning for Data Science, 04 Machine Learning based Use Cases, and 05 Q & A. The meeting interface shows 27 participants and the time 10:49 AM.

The screenshot shows a Google Meet window with the title "Dr Mayuri Mehta is presenting". The main content is a slide titled "Applications of ML" showing various industries and their applications: Manufacturing (Search Engines, Smart Phones, Video Game), Insurance (Email Programs, Robots, Adaptive Telescopes), Transportation (Banks, Autonomous Car, Disease Diagnosis), and Automobile (Healthcare, Customer Service, E-commerce). The meeting interface shows 36 participants and the time 11:51 AM.

Dr. Mayuri A. Mehta discussed about Intro to Machine Learning for Data Science and analytics. She brief Data Science, its process definition and various applications. Also brief importance and need of Machine learning and various ML applications. Machine learning makes easier to conduct analyses such as predictive analysis, which can then serve as helpful visualizations to present. Data visualization is not only important for data scientists and data analysts, it is necessary to understand data visualization in any career. Whether you work in finance, marketing, tech, design, or anything else, you need to visualize data. That fact showcases the importance of data visualization. She also given introduction of ANN usually simply called neural networks, are computing systems vaguely inspired by the biological neural networks that constitute animal brains. An ANN is based on a collection of connected units or nodes called artificial neurons, which loosely model the neurons in a biological brain.

Speaker 3: Dr. Chintan V. Shah (Data Scientist, Corpository, Ahmedabad)

Time: 1:45 pm to 3:45 pm

Dr. Chintan Shah is Data Scientist at Corpository, Ahmedabad. He has expertise in information retrieval using BM25 score technologies and NLP. His focused area is to provide solutions to real world problems using cutting edge AI/ML/Data Science/NLP/BI technologies. He has completed his PhD from MS University, Baroda. He is also having more than 9 years of experience in academia. His research areas of interest are Machine Learning, Text Summarization with Semantic Analysis.

The screenshot shows a Google Meet interface. The main window displays a slide titled "Plan of Attack" with the following content:

What we will learn in this section:

- The Neuron
- The Activation Function
- How do Neural Networks work? (example)
- How do Neural Networks learn?
- Gradient Descent
- Stochastic Gradient Descent
- Backpropagation

The right sidebar shows a grid of participants, including Chintan Shah, Birendrasinh Zala, and others. The bottom status bar indicates the time is 1:51 PM on 27-10-2020.

The screenshot shows a Google Meet interface. The main window displays a slide titled "How do Neural Networks learn?" with a diagram of a perceptron and the following content:

Input value 1 X_1 , Input value 2 X_2 , Input value m X_m

Perceptron

Output value \hat{y} , Actual value y

Cost function: $C = \frac{1}{2}(\hat{y} - y)^2$

The right sidebar shows a grid of participants, including Chintan Shah, Alpa Oza, Bhavana Hotc..., Jignesh Patel, and others. The bottom status bar indicates the time is 2:19 PM on 27-10-2020.

Dr. Chintan Shah has discussed about ANN (Artificial Neural Network) and CNN (Convolution Neural Network). Artificial Neural Network (ANN) uses the processing of the brain as a basis to develop algorithms that can be used to model complex patterns and prediction problems. In our brain, there are billions of cells called neurons, which processes information in the form of electric signals. External information/stimuli is received by the dendrites of the neuron, processed in the neuron cell body, converted to an output and passed through the Axon to the next neuron. The next neuron can choose to either accept it or reject it depending on the strength of the signal. The network architecture has an input layer, hidden layer (there can be more than 1) and the output layer. It is also called MLP (Multi-Layer Perceptron) because of the multiple layers. The hidden layer can be seen as a “distillation layer” that distills some of the important patterns from the inputs and passes it onto the next layer to see. It makes the network faster and efficient by identifying only the important information from the inputs leaving out the redundant information. Convolution layers are used to extract the features from input training samples. Each convolution layer has a set of filters that helps in feature extraction. In general, as the depth of CNN model increases, complexity of features learnt by convolution layers increases. For example, first convolution layer captures simple features while the last convolution layer captures complex features of training samples.

Day-2 (28/10/2020) Time (09:30 am to 3:45 pm)

Speaker 1: Prof. Priyank B. Thakkar (Associate Professor, Computer Science and Engineering Department, Institute of Technology, Nirma University)

Time: 9:30 am to 11:30 am

Dr. Priyank Thakkar has done his Ph.D. from Nirma University and at present, he is working as an associate professor in the Computer Science and Engineering Department, Institute of Technology, Nirma University. He has more than 20 years of teaching experience. He has published more than 20 research papers in international journals and conferences. His research interests include Data and Web Mining, Machine Learning, Soft Computing, Deep Learning, and their applications. He has served as a program committee member of many international conferences. He is also serving as a reviewer for many international journals of Elsevier, Springer, PLOS, World Scientific, Hindawi, Taylor and Francis, and IEEE Transactions.

The screenshot shows a Google Meet interface. At the top, the browser tabs include 'Gandhinagar Institute of Technol...', 'Search results - alpa.oza@git.org', 'Cumulative Attendance 3rd,5th &...', and 'Meet - qbv-qgjiw-oxv'. The Meet header shows 'Priyank Thakkar is presenting' with a 'REC' button. The main content area displays a Jupyter Notebook titled '#K-means Clustering of IRIS Dataset' with the following Python code:

```
import numpy
from sklearn import datasets, metrics
from sklearn.cluster import KMeans

data, target=datasets.load_iris(return_X_y=True)
X_train=data[range(0,150,2),:]
X_test=data[range(1,150,2),:]

model=KMeans(n_clusters=3) #max_iter = 300
model.fit(X_train)
print(model.labels_)

prediction = model.predict(X_test)
print(prediction)
```

The right sidebar shows a grid of participants: Priyank Thakkar (presenting), Bhavana Hotc..., Alpa Oza, Madhuri Chop..., URVISH SOJIL, Vaidehi Patel, PATEL JEEL, Nirali Kapadia, Dhaval Vaja, Birendrasinh..., Rohit Singh, and Hardikkumar... The bottom of the screen shows the Windows taskbar with the search bar and various application icons.

Unsupervised Learning NN

Simple Competitive Learning Neural Network

Output = $\begin{cases} 1 & \text{if node is a winner} \\ 0 & \text{otherwise} \end{cases}$

Output layer with inhibitory connections

Input layer

$w_{new} \text{ vector} = w_{old} \text{ vector} + \eta(l/p \text{ vector} - w_{old} \text{ weight vector})$

2 layers

Dr. Priyank Thakkar has discussed about basics of Supervised Learning and Unsupervised learning. Unsupervised learning is a machine learning technique, where you do not need to supervise the model. Supervised learning allows you to collect data or produce a data output from the previous experience. Unsupervised machine learning helps you to find all kinds of unknown patterns in data. Unsupervised learning is a type of machine learning that looks for previously undetected patterns in a data set with no pre-existing labels and with a minimum of human supervision. Some use cases for unsupervised learning — more specifically, clustering — include: Customer segmentation, or understanding different customer groups around which to build marketing or other business strategies. Dr. Priyank sir has taken data set and given a hands on session on unsupervised learning.

Speaker 2: Prof. Rahul A. Vaghela (Assistant Professor & HOD, Information Technology Department, Gandhinagar Institute of Technology, Gandhinagar)

Time: 11:45 am to 12:45 pm

Prof. Rahul A. Vaghela is an Assistant Professor & HOD Information Technology Department, Gandhinagar Institute of Technology, Gandhinagar. He is an active GTU OSTC NODAL & Inst. Coordinator, He has coordinated and participated in many STTPSs, seminars & author of national & international papers. He was awarded with Pedagogical Innovation Award by GTU & letter of appreciation by IIT-Bombay in 2014. His area of interest are Data Mining, Software Engineering, and Mobile Computing.

Deep Learning = Training Multistage Machines

Traditional Pattern Recognition: Fixed/Handcrafted Feature

Feature Extractor → Trainable Classifier

Mainstream Pattern Recognition

Feature Extractor → Mid-Level Features → Trainable Classifier

Deep Learning: Multiple stages/layers trained end to end

Low-Level Features → Mid-Level Features → High-Level Features → Trainable Classifier

Prof. Rahul A. Vaghela has discussed related to Deep Learning and concepts of CNN. He has discussed a convolutional neural network (CNN, or ConvNet) which is a class of deep neural networks, most commonly applied to analyzing visual imagery. They are also known as shift invariant or space invariant artificial neural networks (SIANN), based on their shared-weights architecture and translation invariance characteristics. They have applications in image and video recognition, recommender systems, image classification, medical image analysis, natural language processing, and financial time series. CNNs are regularized versions of multilayer perceptrons. Multilayer perceptrons usually mean fully connected networks, that is, each neuron in one layer is connected to all neurons in the next layer. The "fully-connectedness" of these networks makes them prone to overfitting data. Typical ways of regularization include adding some form of magnitude measurement of weights to the loss function. CNNs take a different approach towards regularization: they take advantage of the hierarchical pattern in data and assemble more complex patterns using smaller and simpler patterns. He concluded the session with advantages and future scope of CNN.

Speaker 3: Dr. Hitesh R. Chhinkaniwala (Associate Professor & Head, Information & Communication Technology, Adani Institute of Infrastructure Engineering, Ahmedabad)

Time: 1:45 pm to 3:45 pm

Dr. Hitesh Chhinkaniwala is Associate Professor, Computer Engineering at AIIE since August 2015. Dr. Chhinkaniwala has 18+ years' academic and 2+ years' industry experience. Dr. Chhinkaniwala is a Board of Studies member for M.E./B.E./Diploma courses of Computer Engineering, Information Technology, Computer Science & Engineering and Information & Communication Technology by Gujarat Technological University. Dr. Chhinkaniwala has been invited to deliver expert talk at many engineering colleges. He is also a peer reviewer of several national as well as international journals/conferences. He is also a Life Member of ISTE.

The screenshot shows a Google Meet interface. At the top, the browser tabs include 'Gandhinagar', 'Day 2 Session', 'Cumulative At...', 'Meet - qibv-qgijw-oxv', 'Day 1 Feedback', 'Feedback form', 'valedictory sp...', and '10 Tips to Cre...'. The address bar shows the meeting URL: meet.google.com/qibv-qgijw-oxv. The main window displays 'Dr. Hitesh Chhinkaniwala is presenting'. Below this, a code editor window is open, showing Python code for a simple calculator and a number guessing game. The code includes comments and function calls like `print()`, `input()`, and `range()`. On the right side of the screen, a grid of participant avatars is visible, including Dr. Hitesh Chh..., Rajan Patel, Sneha Jani, Nirali Kapadia, Krishna Hing..., Sejal Bhavsar, Pooja Shah, Birendrasinh..., HOD IT, Dhaval Vaja, Vaibhavi Ghar..., and Rahul Vaghela. At the bottom, the Windows taskbar is visible with the search bar and various application icons. The system clock shows 14:21 on 28-10-2020.

Dr. Hitesh Chhinkaniwala is presenting

Exercise (Set-1)

```
thislist = ["India", "Canada", "USA"]
print(thislist)

print(thislist[1])

thislist[1] = "UAE"
print(thislist)

thislist = ["India", "Canada", "USA"]
for x in thislist:
    print(x)

if "Canada" in thislist:
    print("Yes, 'Canada' is in the country list")
else:
    print("No, 'Canada' is not in the country list")

print(len(thislist))

thislist.append("Japan")
print(thislist)

thislist.insert(1, "Australia")
print(thislist)

thislist.remove("Canada")
print(thislist)

thislist.pop()
print(thislist)
```

Participants: Dr. Hitesh Chhinkaniwala, Rajan Patel, VICKY PANCHAL, Sneha Jani

14:31 28-10-2020

Dr. Hitesh Chhinkaniwala has started session with introduction to data preprocessing in machine learning. Pre-processing refers to the transformations applied to our data before feeding it to the algorithm. Data Preprocessing is a technique that is used to convert the raw data into a clean data set. In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis. He has also delivered the hands on session on top python libraries like matplotlib, Numpy and pandas. He has also taken some exercises sets and implemented with the datasets with the help of libraries.