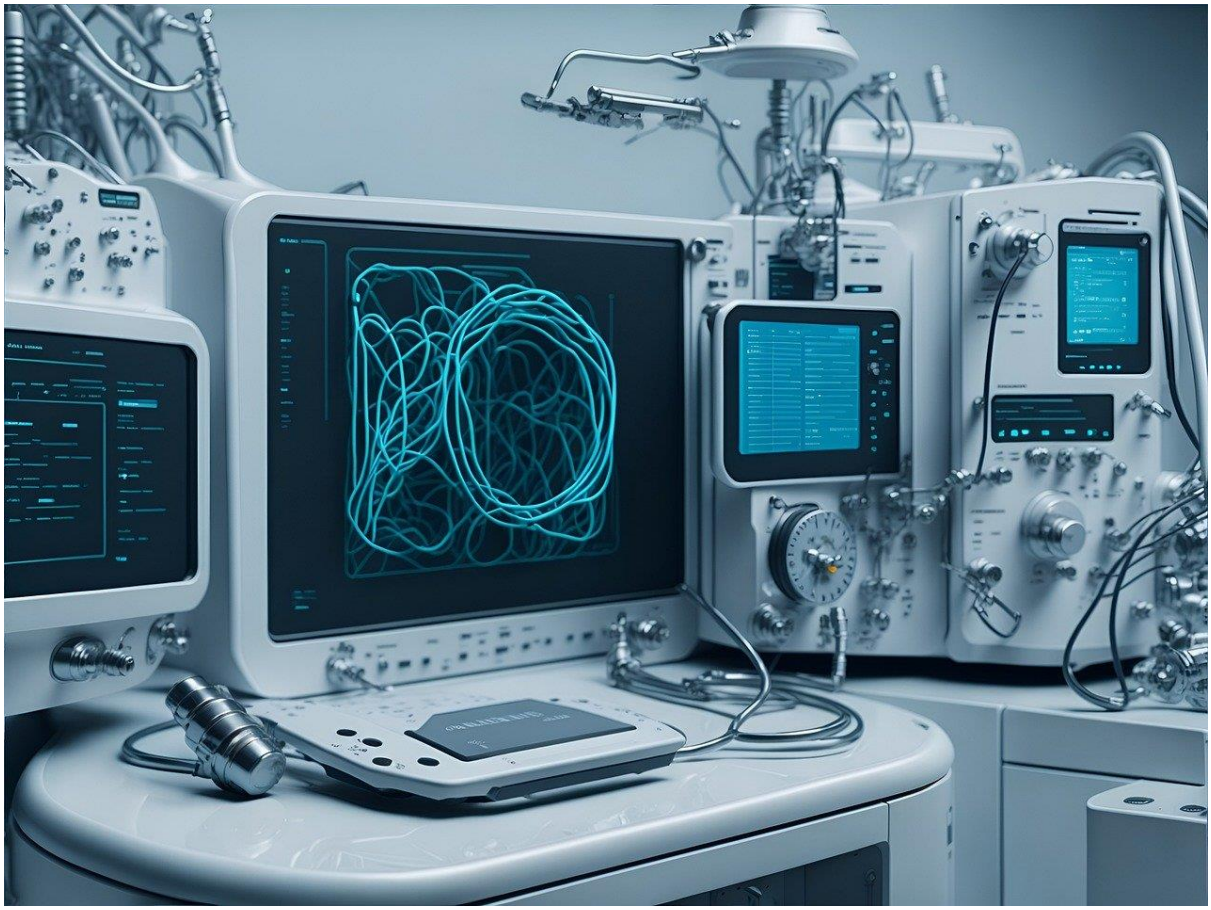
Program in Healthcare Analytics

Online, live, hands-on classes on analysing Healthcare Data using Machine Learning & AI
(135-hours)



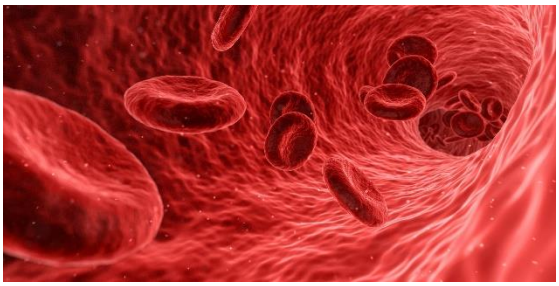
FORE School of Management
New Delhi

FORE School of Management, New Delhi

Contents

Program

If anything, COVID-19 has given a massive fillip to analytics in healthcare sector. It is not that healthcare sector was not using analytical tools earlier, just that use of analytics has become more pronounced after the emergence of the global pandemic. Massive amount of data has been generated the World over on COVID-19 and such large data can only be analysed though Machine Learning techniques. ML technology helps to understand numerous aspects of this malady.



Health sector has generally been generating large amount of data. Today, many hospitals are computerised leading to massive volumes of patient data, sensor data coming out of electronic instruments, claims and cost data, inventory & drugs data, pathology and X-ray data. Machine Learning and Deep Learning techniques offer a way to explore and make sense out of this data and make very useful predictions or rather assist in fast and accurate diagnosis. Some questions that we try to answer pertain to Clinical problems, pharmaceutical and research challenges, patient behaviour or insurance and costs and even related to many other aspects. Here is a list of some of the possible questions.

The program is divided into three modules. Details about the Modules and the Type of Questions answered are given below.

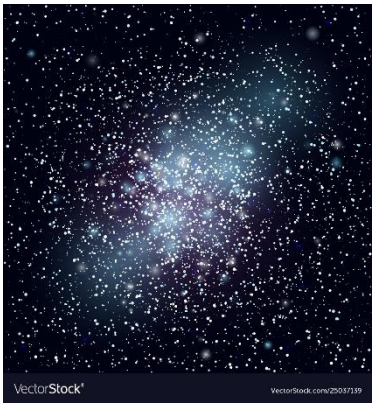
No Programming

We would like to highlight, at the very outset, that we cover ML&AI techniques using '*No-code*' approach. We use the best, highly reputed and industry standard Visual frameworks that generally use drag-and-drop approach to build ML workflows—to process data, build models, test them and then deploy for production use. All these tools are open-source, have very liberal licencing policies (*copyleft*, so to say) and can be utilised even with very large data. We fully realise that many of our students are deeply busy in their core professional work and have little time to learn the intricacies of a programming language (such as python or R). And at the same time, they would like to apply power of analytics to assist them in their work. About the Visual tools used, please see below(here).

We would also like to mention that very few Institutions in India offer program in *Healthcare Analytics* (though many do in *Healthcare Management*). And among those very few none offers a program using '*No-code*' approach.

This program is also unique in its breadth of coverage. We cover Machine Learning, Deep Learning & Natural Processing, Generative AI and LLMs, almost in their entirety. Of the very few programs in Healthcare Analytics, none has this broad coverage.

About the Modules



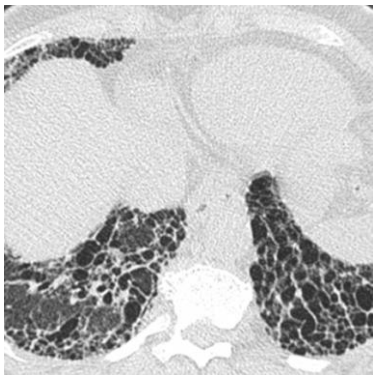
We have four modules. The first covers Statistical Analysis and the other two cover Machine Learning and AI Techniques (ML & AI) as applied to health sector. These modules are **totally hands-on and practise based.** These are online, live, and totally interactive **lab oriented Modules** with the primary objective of disseminating techniques of Healthcare Analytics using Statistics, Data Visualization, Machine Learning & AI. These technologies enable a practitioner to apply them on data in numerous ways.

Primary Objectives

i. Develop insights into healthcare data through visual analytics
ii. Discover if data has any structure
iii. Learn techniques to group/segment data
iv. Develop models for predictive analytics
v. Optimize model performance, and
vi. Single-out attributes that contribute most towards higher performance-->Explainable AI

Types of questions answered

Here are several typical or atypical questions that we strive to answer in our classes. We will perform Statistical analysis, Segmentation analysis, Classification analysis and Regression analysis.



Clinical Problems

- Classify fetal health in order to prevent child and maternal mortality
- Predict lung function decline—Pulmonary Fibrosis Progression
- Predict Possibility of Heart Attack
- Classify Pulmonary Embolism cases in chest CT scans
- Predict the onset of diabetes based on diagnostic measures
- Predict Age from X-rays
- Predict if an infant is likely to develop autistic tendencies
- Predict severity of epileptic seizure
- Detect Malaria through Infected Cell Images
- Detect Autism from a facial image
- Identify acute intracranial haemorrhage and its subtypes
- MRI Imaging Comparisons of Demented and Nondemented Adults
- Create an accurate model to predict the stage of Alzheimer.
- Distinguishing Different Stages of Parkinson's Disease

Pharmaceutical and R&D Problems

- Predict a biological response of molecules from their chemical properties

Patient behaviour related

- Can you predict if a patient will keep his appointment?
- Prevalence and attitudes towards mental health among tech workers

Insurance and Costs related

- Can you accurately predict medical insurance costs?
- Healthcare Provider Fraud Detection Analysis
- Explore Health Insurance Marketplace
- Predict length of stay in hospital
- Predict medical insurance costs
- Predict hospital readmission for diabetes patients

Retail purchases and sale

- Forecast sales of drugs using store, promotion, and competitor data

Course Modules

The plan for the four modules is as below. Teaching sequence will also be in this order. Details about each Module are given below under respective Module heads.

Module	Theme	Hours
Module---**I**	Statistical Analysis and Data Visualization	30
Module---**II**	Machine Learning	45
Module---**III**	Deep Learning, and NLP	35
Module---**IV**	Generative AI and LLMs	25
Total Hours		135

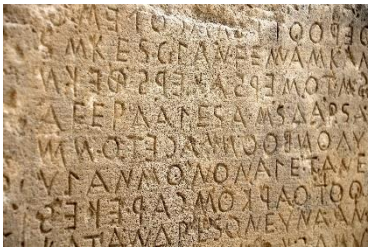
About Visual Tools used

We use the following Visual frameworks:

Software	Applied in
KNIME	Machine Learning & Deep Learning
H_2O	Deep Learning
Deep Learning Studio	Deep Learning
Ollama/anythingLLM	Generative AI & LLM

All these have highly intuitive user interface to perform analytics.

Detailed Contents



Module—I is the foundation module. Concepts taught here are universally used in ML and AI. Modules--II and III differ in their approach to Machine Learning. Module--II pertains to what may be called *Traditional Machine Learning*, Module-III pertains to *Deep-Learning and Natural Language Processing*, Module-IV pertains to *Generative AI and Large Language Models*. Classical ML techniques require much less data than Deep-Learning or GenAI techniques. Both approaches have their pros and cons. Traditional ML techniques generally deal with tabular data sets while deep-learning/GenAI techniques also process unstructured data as text, images or video datasets.

Module—I

Statistical Analysis and Data Visualization

Go to top

Data Mining is intimately intertwined with Statistics. Knowledge of basic statistics is essential for a successful analyst. Many ‘Small data’ techniques such as correlation, testing of hypothesis, data-transformation and others need to be learnt to fully understand data. Concepts of inferential statistics are used in comparing machine learning models. Descriptive statistics is invariably used in data pre-processing. In this Course we refresh as also learn statistical fundamentals and essential inferential statistics.

1. Measures of Central Tendency and Dispersion
2. Probability Theory (Different Approaches, Rules of Probability, Bayes' Theorem)
3. Random Variables and Probability Distributions Discrete Probability Distributions
4. Continuous Probability Distributions – Normal Distribution
5. Correlation and Regression Analysis: Simple & Multiple Regression
6. Concept Of Hypotheses Testing, Type I & Type II Errors, Power Of The Test, Hypothesis Testing of Mean and Proportion, Two Sample Tests, Tests for Difference in Means and Proportions.
7. Chi-Square Goodness-of-Fit Test, Test of Independence

Module—II

Machine Learning (Classical)

Go to top

We practice those modelling techniques that consistently garner high performance, are relatively fast and are well known in ML community. Thus, these will be of immense use in many predictive applications. These techniques do not perform that well with image or video data.

1. Introduction to Machine Learning Technology
2. Data visualization and discovering structure in data. (Techniques include t-sne, parallel coordinates, mosaic plots) and Feature importance
3. Unsupervised learning techniques
    1. K-means clustering
    2. Hierarchical clustering
    3. Expectation-Maximization algorithm
    4. T-SNE & UMAP manifold learning technique
    5. Dimensionality reduction
    6. Principal Component Analysis (PCA)
4. Supervised learning techniques for Classification and Regression
    1. Decision trees
    2. Ensemble modelling using Random Forest
    3. Gradient Boosting Techniques
        a. Gradient Boosting Learner
        b. XGBoost
        c. LightGBM
    4. Performance measures: Accuracy, Precision and Recall, F-measure; Area Under the Curve, Cohen's Kappa, Sensitivity, Specificity
5. Hyper-parameter optimisation techniques—Bayes Optimization;
6. Interpreting Machine Learning Models

Module—III

Deep Learning and Natural Language Processing

Go to top

In Module-III we practise Deep-learning techniques and Natural Language Processing. Deep Learning techniques are especially useful for image datasets, for example, chest-X-rays and CAT-Scans. These techniques are also used with sensor data (for example, ECG). If tabular data is sufficiently large, deep learning techniques can be applied for making predictions. Many a time, structured data has one or more columns of text data (for example describing

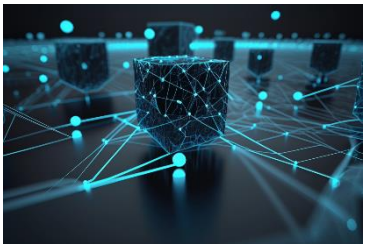
patients' state of health etc). Natural Language processing techniques can be applied on such columns. Here is what we cover and learn:

1. Introduction to Neural Networks (NN)
2. Experiments with MLP networks
3. Regularising NN
    a. Dropouts
    b. Batch-normalization
    c. l1 and l2 regularization
    d. Starting weight initialization
4. Deep Learning with Convolution Neural Networks
    a. Data Augmentation
5. Using very Deep Convolution Networks
    a. Transfer learning with VGG16
    b. Transfer learning with ResNet50
    c. Transfer learning with InceptionV3
6. Recurrent Neural Networks
    a. LSTM, GRUs and Bi-directional LSTM

Module—IV

Generative AI and LLMs

Go to top



Generative AI and Large Language Models (LLMs) are integral components of the AI landscape, each with distinct roles and capabilities. Generative AI encompasses a range of tools that leverage information from LLMs and other AI models to create new content through machine learning. On the other hand, LLMs are a specific type of AI model that utilizes machine learning with billions of parameters to understand and generate text.

Generative AI and Large Language Models (LLMs) are increasingly being utilized in healthcare for various applications. A survey conducted in 2024 revealed insights into the adoption and impact of these technologies in the healthcare industry. The survey highlighted the importance of healthcare-specific, task-oriented language models, particularly in improving patient interactions, streamlining clinical processes, and advancing research. Respondents expressed a preference for LLMs in applications like transcribing doctor-patient conversations, medical chatbots, and answering patient questions, indicating their potential transformative impact on patient-facing tasks.

Furthermore, the integration of generative AI and LLMs in healthcare is seen as a significant advancement that can revolutionize the industry. These technologies can assist in tasks such as clinical note documentation and transcription, enhancing patient engagement through chatbots and virtual assistants, synthesizing biomedical literature, and optimizing clinical trials. Despite challenges related to data privacy and regulatory compliance, the benefits of generative AI in healthcare are evident, with applications that can streamline operations, improve patient care, and drive innovation in the field.

1. General Architecture of Transformers

2. Zero-shot classification and few-shot learning
3. Ollama and anythingLLM installation
4. Embedding, vector databases and search
5. Prompt Engineering
6. Developing knowledge products in healthcare using web-UIs for LLM

Modules Pedagogy:

We strongly believe that a course in Healthcare Analytics can only be practice-based rather than pure theory based. We also believe that a practice-based course requires constant interaction with the teacher during lecture hours in real time. Our teaching pedagogy is like this: First, the algorithm (or theory part) is conceptually explained without getting into mathematics and then a project is undertaken to implement the techniques. Healthcare datasets for implementation are made available in advance. During the lecture, we go through one of the visual frameworks such as KNIME workflow and explain the steps. At his end, the student goes through the same steps on his laptop. Consequently, results are available at our end as also with the Students immediately. In short, *both the teacher and students are working on their respective laptops simultaneously*; students solve their problems and ask any questions to clarify. The whole experience is just as if everyone is sitting in a laboratory and working together. Students are required to have a laptop with minimum of 16GB of RAM.

Who Should Attend?

Healthcare industry generates lots of data and this data is analyzed by professionals specializing in numerous fields. The program would be very useful for Medical Practitioners, Bio-technologists, bio-informatics or in general, students of Life –Sciences-- Biology, Biotechnology, Biochemistry, Bioinformatics, Cell biology (cytology), Ecology, Molecular biology, Microbiology, Marine Sciences —will find the program very useful.

Officers Managing Hospitals, Data Scientists or programmers, Doctors or those in academics or Healthcare workers will find program extremely beneficial.

Program Timings and Duration

Total Duration for all Modules is 135 hours spread over 4.5-months. Program will be delivered on Saturdays and Sundays. There are 5 hours of teaching per week. Students are expected to perform exercises. This methodology of "*learning concept->performing class projects-->Do self-exercise*s" leads to better and stress-free absorption.

Program Requirements (for students)

Participating students should be having a laptop or desktop with minimum 16gb of RAM. More RAM is advisable. Preferably the processor should be not lower than i5.

Program Faculty

Prof. Ashok Kumar Harnal

Ashok Kumar Harnal has worked extensively at multiple facets of Big Data Systems--Machine Learning, Generative AI and LLM, Big-Data storage systems (Hadoop and NoSQL databases), Graph Databases, Streaming Analytics using Apache Spark, Apache Kafka, Confluent and Reinforcement Learning. He has been teaching Big Data technology since around last twelve years. Since last nine years Prof Harnal has been collaborating closely with University of California, Riverside, in a program on taking sessions on Big Data for Executives from around the World. We have trained officers from several organizations including RITES, NABARD, TechMahindra, Punjab National Bank, Central Bank of India and Union Bank of India Presently we are training officers in one another Bank. What is a matter of pride for us is that many of our students are at very high positions in Industry. My GitHub site is here. We have successfully conducted three programs on Healthcare Analytics; two programs were of three months duration and one of nine months duration. During his stay in Min of Defence, he has executed three country-wide projects on Information Systems: (a) *Raksha-Bhoomi* to computerize land records (as old as 150 years); (b) Knowledge Management of land-title related files/maps in all Defence Estates offices; and (c) Setting up of a Disaster Management organization: Archival Unit and Resource Center (AU&RC), at Delhi and Pune for safe storage of land-title related records in paper and digital forms. He has published two books (both by Tata McGraw-Hill); One on *How to program games on Computers* and the other on *Linux Administration and Applications*.

Prof. Amarnath Mitra

Dr. Amarnath Mitra is working as an Associate Professor in the area of Information Technology and Big Data Analytics at FORE School of Management, New Delhi. Prior to joining FORE, Dr. Mitra worked as Senior Quant Analyst at BioUrja Power LLC (Texas, USA). Dr. Mitra has over five years of industry experience as an analyst and researcher with substantial exposure of working with big & high frequency data and analytics. In academics, Dr. Mitra worked as full-time faculty for over six years in management institutes such as BML Munjal University Gurugram, IMI New Delhi and IBS Hyderabad. As guest/visiting faculty he has taught in several reputed institutions like SIBM Pune, NMIMS Hyderabad, IIIT Bhubaneswar among others. Dr. Mitra has taught subjects like Data Science, Predictive Analytics, Business Analytics, Quantitative Methods, Business Research Methods, Operations Research, Econometrics, among others.