



E&ICT Academy, IIT Kanpur
A Joint Initiative of MeitY & IIT Kanpur



Advanced Certification in AI & Machine Learning

The **Advanced Certification in AI & Machine Learning** course by the **E&ICT Academy of IIT Kanpur** in association with **Edvancer** is one of the most comprehensive and state of the art programs in this domain. This 500 learning hours program for working professionals & freshers is designed, curated and delivered by faculty from IIT's and Industry. This program is a blend of academic rigour and significant industry exposure equipping you with highly advanced conceptual knowledge and futuristic skillsets **unmatched by any other such programs**. This program once completed satisfactorily strives to prepare you for seamless accelerated transition to a high-end career in data science & AI. On successful completion of the program you will **receive a certificate from E&ICT Academy, IIT Kanpur**, a joint initiative of IIT Kanpur and Ministry of Electronics & Information Technology, Government of India.

Full Curriculum

Module 1: Machine Learning in Python

What is this module about?: Through this Machine Learning module, you will learn how to process, clean, visualize and automate decision making through data science by using Python, one of the most popular machine learning tools. You will learn cutting edge machine learning techniques in Python.

Tools to be learnt: Python (Libraries like Pandas, Numpy, Scipy, Scikit-learn, Seaborn, Beautifulsoup)

Class Duration: 78 hours (13 weekends)

Topic	What does it mean?
Introduction to Machine Learning in Python <ul style="list-style-type: none"> • What is machine learning & why is it so important? • Applications of machine learning across industries • Machine Learning methodology • Machine Learning Toolbox • Tool of choice- Python: what & why? • How to approach a machine learning problem using a case study • Course Components 	<p>In this section we shall provide you an overview into the world of machine learning (ML). You will learn about the various applications of machine learning, how companies from all sort of domains are solving their day to day to long term business problems. We'll learn about required skill sets of a machine learning expert which make them capable of filling up this vital role. Once the stage is set and we understand where we are heading we discuss why Python is the tool of choice in ML.</p>
Python Training	
Introduction to Python <ul style="list-style-type: none"> • Installation of Python framework and packages: Anaconda and pip • Writing/Running python programs using Spyder, Command Prompt • Working with Jupyter Notebooks • Creating Python variables: Numeric, string and logical operations • Basic Data containers: Lists, Dictionaries, Tuples & sets • Importing & exporting data • Practice assignment 	<p>Python is one of the most popular & powerful languages for data science used by most top companies like Facebook, Amazon, Google, Yahoo etc. It is free and open source. This module is all about learning how to start working with Python. We shall teach you how to use the Python language to work with data.</p>

Flow Control, Iterative Operations & Functions in Python <ul style="list-style-type: none"> • Writing for loops in Python • List & Dictionary Comprehension • While loops and conditional blocks • List/Dictionary comprehensions with loops • Writing your own functions in Python • Writing your own classes and functions as class objects • Practice assignment 	<p>This is where we move beyond simple data containers and learn about amazing possibilities and functionalities hidden in various associated operators. We get introduced to wonderful world of loops, list and dictionary comprehensions. In addition to already existing functions and classes we learn to write our own custom functions and classes. This module sets the stage for handling data and ML algorithm implementation in python.</p>
Basic Statistics & Maths <ul style="list-style-type: none"> • Introducing statistical inference • Estimators and confidence intervals • Central tendency measures • Variability measures • Statistical distributions • Matrices & Linear Equations 	<p>You will get a refresher here on the college level statistics & maths required for machine learning & AI.</p>
Data Summary, Numerical and Visual EDA in Python <ul style="list-style-type: none"> • Need & ways for data summary • Introduction to pandas • Summarising numeric data in Pandas • Summarising categorical data • Group wise summary of mixed data • Need for visual summary • Introduction to Seaborn • Visual summary of different data combinations • Practice Exercise 	<p>Data summary is extremely important to understand what the data is saying and gain insights in just one glance. Visualization of data is a strong point of the Python software using the latest Seaborn package.</p>
Data Handling in Python using NumPy & Pandas <ul style="list-style-type: none"> • Introduction to NumPy arrays, functions & properties • Dataframe functions and properties • Creating, Modifying, Renaming dataframe columns as Pandas series • Sub-setting data on conditions, row and column indices, data types • Creating flag features for categorical columns • Imputing missing values, modifying column values with dictionary replacements • Reading and writing external data • Creating summary features • Handling datetime data • Manipulating data columns 	<p>Python is a very versatile language and in this module we expand on its capabilities related to data handling. Focusing on packages Numpy and Pandas we learn how to manipulate data & undertake feature engineering which will be eventually useful in converting raw data suitable for machine learning algorithms.</p>
Hypothesis Testing <ul style="list-style-type: none"> • What is a hypothesis? • Statistical distribution & confidence intervals • Central Limit Theorem • Single & Two sample T-Tests • ANOVA • Chi-Square test • Non-parametric tests 	<p>With 95% confidence we can say that there is an 55% chance, people visiting this site twice will enroll for the course 😊. Here, you learn how to create a hypothesis like this, test and validate it through data within a statistical framework and present it with clear and formal numbers to support decision making.</p>
Machine Learning Models in Python	
Basics of Machine Learning 1 <ul style="list-style-type: none"> • Converting Business Problems to Data Problems • Broad Categories of Business Problems • Supervised and Unsupervised Machine Learning Algorithm • Drivers of ML algorithms • Cost Functions 	<p>In this module we understand how we can transform our business problems to data problems so that we can use machine learning algorithms to solve them. We will further get into discovering what all categories of business problems and subsequently which machine learning algorithms are there.</p>

Basics of Machine Learning 2 <ul style="list-style-type: none"> • Brief introduction to Gradient Descent • Importance of Model Validation • Methods of Model Validation • Introduction to Cross Validation and Average Error • Parameter Estimation for a simple linear model with gradient descent • Model evaluation metrics for regression problems • Model evaluation metrics for classification problems 	<p>We'll learn what is the ultimate goal of any machine learning algorithm and go through a brief description of the mother of many modern optimisation methods- Gradient Descent. We'll wrap up this module with discussion on importance and methods of validation of our results.</p>
Generalised Linear Models in Python <ul style="list-style-type: none"> • Linear Regression • Limitation of simple linear models and need of regularisation • Ridge and Lasso Regression (L1 & L2 Penalties) • Getting Started with sklearn API • Introduction to Classification with Logistic Regression • Methods of threshold determination and performance measures for classification score models • Case Studies 	<p>We start with implementing machine learning algorithms in this module. We also get exposed to some important concepts related to regression and classification which we will be using in the later modules as well. Also this is where we get introduced to scikit-learn, the legendary python library famous for its machine learning prowess.</p>
Decision Tree Models using Python <ul style="list-style-type: none"> • Introduction to decision trees • Tuning tree size with cross validation • Introduction to bagging algorithm • Random Forests • Grid search and randomized grid search • ExtraTrees (Extremely Randomised Trees) • Partial Dependence Plots • Case Studies • Assignments 	<p>In this module we will learn a very popular class of machine learning models, rule based tree structures also known as Decision Trees. We'll examine their biased nature and learn how to use bagging methodologies to arrive at a new technique known as Random Forest to analyse data. We'll further extend the idea of randomness to decrease bias in ExtraTrees algorithm. In addition, we learn about powerful tools used with all kind of machine learning algorithms, gridSearchCV and RandomizedSearchCV.</p>
Boosting Algorithms using Python <ul style="list-style-type: none"> • Concept of weak learners • Introduction to boosting algorithms • Gradient descent in functional space • Gradient Boosting Machines (GBM) for regression and classification • Extreme Gradient Boosting (XGBoost) • Case study • Home exercise 	<p>Want to win a data science contest on Kaggle or data hackathons or be known as a top data scientist? Then learning boosting algorithms is a must as they provide a very powerful way of analysing data and solving hard to crack problems.</p>
Support Vector Machines (SVM) and KNN in Python <ul style="list-style-type: none"> • Introduction to idea of observation based learning • Distances and Similarities • K Nearest Neighbours (KNN) for classification • Introduction to SVM for classification • Regression and classification with SVM • Case study • Home exercises 	<p>We step in a powerful world of "observation based algorithms" which can capture patterns in the data which otherwise go undetected. We start this discussion with KNN which is fairly simple. After that we move to SVM which is very powerful at capturing non-linear patterns in the data.</p>
Unsupervised learning in Python <ul style="list-style-type: none"> • Need for dimensionality reduction • Introduction to Principal Component Analysis (PCA) • Difference between PCAs and Latent Factors • Patterns in the data in absence of a target • Segmentation with Hierarchical Clustering and K-means • Measure of goodness of clusters • Limitations of K-means • Introduction to density based clustering (DBSCAN) 	<p>Many machine learning algorithms become difficult to work with when dealing with many variables in the data. It comes to rescue PCA which solves problems arising from data which has highly correlated variables. The same idea can be extended to find out hidden factors in our data with Factor Analysis which is used extensively in surveys and marketing analytics.</p> <p>We also learn about two very important</p>

<ul style="list-style-type: none"> • Anomaly detection with DBSCAN and one-SVM • Case Studies • Practice Assignments 	segmentation algorithms; K-means and DBSCAN and understand their differences and strengths.
Neural Networks <ul style="list-style-type: none"> • Introduction to Neural Networks • Single layer neural network • Multiple layer Neural network • Activation functions • Back propagation Algorithm • Multi-layer perceptrons in sklearn • Practice assignment 	Artificial Neural Networks are the building blocks of artificial intelligence. Learn the techniques which replicate how the human brain works and create machines which can solve problems like humans.
Text Mining in Python <ul style="list-style-type: none"> • Quick Recap of string data functions • Gathering text data using web scraping with urllib • Processing raw web data with BeautifulSoup • Collecting twitter data with Twitter API • Introduction to Naive Bayes • Feature Engineering for text Data • Feature creation with TFIDF for text data • Case Studies 	Unstructured text data accounts for more and more interaction records as most of our daily life moves online. In this module we start our discussion by looking at ways to collect all that data. In addition to scraping simple web data; we'll also learn to use data APIs with example of Twitter API, right from the point of creating a developer account on twitter. Further we discuss one of the very powerful algorithm when it comes to text data; Naive Bayes. Then we see how we can mine the text data.
Machine Learning Beyond Traditional Model Building	
Ensemble Methods for mixed algorithms <ul style="list-style-type: none"> • Combining multiple models with simple and weighted averaging • Simple Majority vote and weighted majority vote • Introduction to Blending • Introduction to Stacking • Stacking gotchas and 2 layer model in Python • Case Study 	Individual machine learning models extract pattern from the data in different ways, which at times results in them extracting different patterns from the data. Rather than sticking to just one algorithm and not making use of other's results is what we move past in this module. We learn to make use of multiple ML models taken together to make our predictive modelling solutions even more powerful.
Python pipelines and model in production <ul style="list-style-type: none"> • Discussion on elements of model deployment • Introduction to sklearn pipelines for consistent data processing • Introduction to joblib for model saving and loading • Basic flask API app for understanding model deployment 	For making quick prototypes of your solutions which can be scaled later as interactive visualisation in the form of standalone or hosted web pages, we introduce you to Bokeh, an evolving library in python which has all the tools that you'll need to do the same.
Version Control using Git <ul style="list-style-type: none"> • Need and Importance of Version Control • Setting up git and Github accounts on local machine • Creating and uploading GitHub Repos • Git commands • Push and pull requests with GitHub repos • Merging and forking projects 	We finish this module with a discussion on two very important aspects of a data scientist's work. First is version control which enables you to work on large projects with multiple team members scattered across the globe. We learn about git and most widely used public platform version control that is GitHub.
Advanced Ideas on Feature Engineering, Model Interpretation, Parameter Tuning, Genetic Algorithm <ul style="list-style-type: none"> • Representing cyclic date time features • Revisiting dummies creation for categorical data • Categorical embeddings • Permutation feature importance with rfpimp and feature selection • Experiments on capabilities of algorithms for extracting non-linear relationships [polynomial vs log vs inverse vs ratio etc.] • Global model interpretation with partial dependence plots with package skater 	Learn the tips and tricks used by highly experienced machine learning engineers to create more relevant data features, interpret models in a better manner, simplify complex models and find optimal values. This module is designed to give you skills equivalent to a very experienced data scientist.

- Surrogate Trees as methods for simplifying a complex model
- Local interpretation [Observation level] with LIME
- Finding optimal values of features for min/max target given an ML model [using genetic algorithm]

Module 2: Deep Learning (Artificial Intelligence) Using Tensorflow and Keras

What is this module about?: Through this module, you will learn the various techniques used in the world of artificial intelligence like deep learning techniques, reinforcement learning, NLP and computer vision using latest Python libraries Tensorflow and Keras. This module will put you on the cutting edge of technology and make you future proof.

Tools to be learnt: Tensorflow and Keras

Class Duration: 66 hours (11 weekends)

Topic	What does it mean?
Introduction to AI and Deep Learning <ul style="list-style-type: none"> • What is AI? • How will AI change the world? • What is Deep Learning? • Uses of Deep Learning? • Examples of Deep Learning & AI 	Get introduced to the world of Artificial Intelligence which is poised to change the entire world/ Understand what is deep learning and how it is used in AI
Introduction to parameter optimisation <ul style="list-style-type: none"> • Neural Network representation • Gradient descent with backpropagation for a generic feedforward network • Optimisers in deep learning [SGD, SGD with momentum, RMSProp, Adam] • Understanding difference in optimisers with implementation in python from scratch 	Learn the basics of deep learning and how to change the attributes of your neural network such as weights and learning rate to reduce the losses. Optimization is the most essential ingredient in the recipe of deep learning algorithms.
Getting Started with Tensorflow <ul style="list-style-type: none"> • What is a tensor? • Setting up Tensorflow and the GPU instance • Understanding computation graph and basics of Tensorflow • Implementing simple perceptron in Tensorflow • Simple linear Regression with Gradient Tape • Working with Google Collab • Visualising training with tensor board • Classification with a deep feedforward network using base Tensorflow and Gradient tape 	TensorFlow™ is an open source software library in Python for high performance numerical computation. Originally developed by researchers and engineers from the Google Brain team within Google's AI organization, it comes with strong support for machine learning and deep learning.
Deep Learning Algorithms	
Deep Feed Forward & Convolutional Neural Networks <ul style="list-style-type: none"> • Limitations of simple feed forward networks • How vision works for humans • Understanding convolutions, strides, padding, filters etc. • Convolutions for extracting features from images • How receptive fields change in sequence of convolution networks • Pooling for parameter reductions [max , average , sum] • Implementing CNN with Tensorflow • Regularizing with dropout 	A feedforward neural network is an artificial neural network wherein connections between the nodes do not form a cycle and information only flows in one direction. A convolutional neural network (CNN, or ConvNet) is a class of deep, feed-forward artificial neural networks, most commonly applied to analysing visual imagery. Learn these techniques for classifying images.

<ul style="list-style-type: none"> • Learning rate decay and its effects • Batch normalisation and its effects 	
Introduction to Keras <ul style="list-style-type: none"> • Basics of Keras • Getting started with model building in keras • Sequential and Functional API • Regularisation with dropout and batch-normalisation • L1 and L2 penalties on weights • Weight initialisation [e.g. xavier] • Model saving and loading • Data generators and data loaders in keras • Transfer learning in keras • Early stopping • Model checkpoint • Custom call backs 	<p>Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow. It was developed with a focus on enabling fast experimentation and allows for easy and fast prototyping.</p>
Sequence to Sequence models with Recurrent Neural Networks, Long-Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) <ul style="list-style-type: none"> • Introduction to seq2seq problems • Intro to RNN architecture • Backpropagation through time and vanishing gradient • Modelling sequences • Limitations of RNNs and data reshaping for RNNs • Introduction to LSTM • Univariate times series with LSTM • Multidimensional time series with LSTM • Action sequences with LSTM • Introduction to GRU and implementation on text classification 	<p>A recurrent neural network (RNN) is a class of artificial neural network where connections between nodes form a directed graph along a sequence. Long Short Term Memory networks – usually just called “LSTMs” – are a special kind of RNN, capable of learning long-term dependencies. Gated recurrent units (GRUs) are a gating mechanism in recurrent neural networks. These techniques are very popular for Natural Language Processing.</p>
Unsupervised Deep Learning <ul style="list-style-type: none"> • Autoencoders • Embeddings • Autoencoders and embeddings implementation • Generative Adversarial Networks • GAN Implementation 	<p>An autoencoder is a type of artificial neural network used to learn efficient data codings in an unsupervised manner. The aim of an autoencoder is to learn a representation (encoding) for a set of data, typically for the purpose of dimensionality reduction. Generative adversarial networks (GANs) are, implemented by a system of two neural networks contesting with each other.</p>
Applications of AI	
Computer Vision	
Object Detection and Localisation <ul style="list-style-type: none"> • Object detection and localisation as combination of classification and regression problem • RCNN, fastRCNN and Faster RCNN • Semantic Segmentation, Mask R-CNN • YOLO, SSD • Object detection with pre trained models • Annotating custom data with vott • Training YOLOv3 for custom data • Using trained YOLOv3 with a camera 	<p>Learn various CNN and RNN based special algorithms that make it possible to create AI systems to detect and localize objects in videos and images.</p>
Face Detection and Recognition <ul style="list-style-type: none"> • Face detection with MTCNN • Face Extraction • Face embeddings with VGGFace • Face Recognition 	<p>Create AI systems that will help detect faces in videos and images.</p>
Natural Language Processing	
Deep Learning for Natural Language Processing <ul style="list-style-type: none"> • Word embeddings • How to train and use word2vec 	<p>Learn how to process and understand natural language using multiple algorithms. This is how Alexa and Siri are created.</p>

<ul style="list-style-type: none"> • Fasttext • Glove • Bi-Directional RNNs • Encoder Decoder Architecture • Attention models • Transformers • Context Aware word embeddings with BERT • GPT-2 • QA models with memory network • Transformer model on SQUAD • Custom data QnA with cdqa • Sentence embeddings • Text Summarisation 	
Chatbots with Rasa <ul style="list-style-type: none"> • Intention and entities • Rasa installation and project setup • NLU data • Model pipelines • Dialogue management • Dialogue Policies 	Build industry level chatbots using the Rasa open source framework
Other Applications	
Working with Audio Data <ul style="list-style-type: none"> • Audio data processing in python • Using spectrogram as features for audio data • Using CNN for audio data • using 1D CNNs with raw audio data 	Process audio data using CNNs
Image Captioning And Style Transfer <ul style="list-style-type: none"> • Image encoding • Word embeddings • Hybrid model • Mapping output to caption • Revisiting receptive fields • Using different layers to learn style • Understanding loss function for style transfer • Implementation 	This module will help you to create an AI system which can caption images and transfer style of one image to another. Imagine making a Da Vinci painting look like Picasso's!

Module 3: Data Analysis in SQL (Videos only)

What is this module about?: This Data Analyst using SQL video tutorial teaches you how to use the popular SQL language to analyse data stored in databases. SQL is a requirement in almost all analytics roles & this module will make you eligible to work as a data analyst. In this SQL tutorial you will learn how to communicate with databases, extract data from them, manipulate the data, analyse it and create reports.

Tools to be learnt: MS SQL

Class Duration: 6 hours of pre-recorded videos

Topic	What does it mean?
Introduction To SQL <ul style="list-style-type: none"> • What is SQL? • Why SQL? • What are relational databases? • SQL command group • MS SQL Server installation 	Structured Query Language (SQL) is a standard language for storing, manipulating and retrieving data in databases. It is a heavily used language and a must know for every data scientist. Here we will introduce you to SQL using MS SQL.

SQL Data Types & Operators <ul style="list-style-type: none"> • SQL Data Types • Filtering Data • Arithmetic Operators • Comparison operators • Logical Operators • Exercises 	Learn about various types of data and how to filter and conduct basic operations on data in databases using SQL.
Useful Operations in SQL <ul style="list-style-type: none"> • Distinct Operation • Top N Operation • Sorting results • Combine results using Union • Null comparison • Alias • Exercises 	Learn more advanced operations on data.
Aggregating Data in SQL <ul style="list-style-type: none"> • Aggregate functions • Group By clause • Having clause • Over clause • Exercises 	Aggregate data using various conditions and clauses in SQL to gain the answers you are looking for.
Writing Sub-Queries in SQL <ul style="list-style-type: none"> • What are sub-queries? • Sub-query rules • Writing sub-queries • Exercises 	A subquery is a SQL query within a query. Subqueries are nested queries that provide data to the enclosing query. In this module you will learn how to write various sub-queries.
Common function in SQL <ul style="list-style-type: none"> • Ranking functions • Date & time functions • Logical functions • String functions • Conversion functions • Mathematical functions • Exercises 	Learn some of the common functions available in SQL to transform the data into more meaningful data.
Analytic Functions in SQL <ul style="list-style-type: none"> • What are analytic functions? • Various analytic functions • SQL syntax for analytic functions • Exercises 	Here you will learn various analytics function in SQL to undertake data analysis in SQL.
Writing DML Statements <ul style="list-style-type: none"> • What are DML Statements? • Insert statement • Update statement • Delete statement • Exercises 	DML is abbreviation of Data Manipulation Language in SQL. It is used to retrieve, store, modify, delete, insert and update data in databases.
Writing DDL Statements <ul style="list-style-type: none"> • What are DDL Statements? • Create statement • Alter statement • Drop statement • Exercises 	DDL refers to "Data Definition Language", a subset of SQL statements that change the structure of the database schema in some way, typically by creating, deleting, or modifying schema objects such as databases, tables, and views.
Using Constraints in SQL <ul style="list-style-type: none"> • What are constraints? • Not Null Constraint • Unique constraint 	Constraints provide a standard mechanism to maintain the accuracy and integrity of the data inside table. There are several different types of constraints in SQL which you will learn here.

<ul style="list-style-type: none"> • Primary key constraint • Foreign key constraint • Check constraint • Default Constraint • Exercises 	
SQL Joins <ul style="list-style-type: none"> • What are joins? • Cartesian Join • Inner Join • Left & Right Join • Full Join • Self Join 	<p>A SQL Join statement is used to combine data or rows from two or more tables. Learn the various joins in SQL in this module.</p>
Views in SQL <ul style="list-style-type: none"> • What are views? • Create View • Drop view • Update view 	<p>A view is a virtual table that consists of columns from one or more tables. Though it is similar to a table, it is not stored in the database. It is a query stored as an object. Hence, a view is an object that derives its data from one or more tables. Learn how to create these views in this module.</p>

Contact Us

For further details & career counselling please reach out at:

📞 +91 8080928948

✉ ict_mlai@iitk.ac.in

💻 : <https://ict.iitk.ac.in/ai-ml-certification>