Diploma in Machine Learning and Data Science

Introduction
Young IT professionals, recent graduates, and senior undergrad students should adopt this technology because it provides a backdoor to compete with the developed world. This course is useful for those who want to pursue their career is artificial intelligence, machine learning, and data science fields.

Duration
The duration for this diploma in six Months.

Days in a week
Only Sunday from 10 AM to 5 PM

Number of Modules
This course includes eight modules

Description of Each Module
The detailed description of each module is given below:

Module 1
Introducing tools for building AI systems

This introductory module is designed specifically to provide student with current tool kits, libraries, programming languages and practices which are required for data scientists and machine learning experts. Initially, an overview of the Python programming language is provided to familiarize students with current tools and techniques used in building automated intelligent systems. Secondly, practical aspects of data science are introduced in R programming language and Weka toolkit. Finally, introduction to Neural Designer is provided in which students are familiarized with easy to use and rapid implementation of custom Artificial Neural Networks ANNs.

Module Contents:
Introduction to Python programming
- Why to use python
- Installing python
- Variables and data types

- Operators
- Type conversion
- Working with lists
- List manipulation
- Working with functions
- Working with packages
- Using Numpy package
- Numpy basic statistics
- Data visualization using matplotlib

R programming language

- Installing R and R Studio
- Data types, Control Structures, Functions, and List
- Using the built-in datasets
- Importing data
- Creating bar and pie charts for categorical variables
- Creating histograms and box plots for quantitative variables

Introduction to Weka toolkit

- Installing Weka
- Importing data
- Visualizing data
- Exploring data

Introduction to Neural Designer for Neural Networks

- Installing ND
- Loading datasets
- Using statistical measures to analyze data

Introduction to KNIME

- Installing KNIME
- Understanding the KNIME environment
- Understanding the data workflow using KNIME

Module 2
Data science using R programming

R is a popular programming language for data analytics, statistical analysis and data visualization. The purpose of this workshop is to provide basic training in understanding the core techniques and concepts in Data science and R Programming. In this module, you will learn how to use R to model statistical relationships using graphs, calculations, tests, and other analysis tools. In addition, one will also learn how to enter and modify data; create charts, scatter plots, and histograms; examine outliers; calculate correlations; and compute regressions. Furthermore, this module will go through some advanced data analysis concepts techniques such as data extraction, data pre-processing, data visualization, data modeling and finally data presentation.

Module Contents:

- Introduction to Data Science
- Using data to make decisions
- Role of Data Scientist
- Data Scientist's Tools
- Real-world data
- Data collection
- Data pre-processing
- Data visualization
- Importing data
- Creating bar and pie charts for categorical variables
- Creating histograms and box plots for quantitative variables
- Calculating frequencies and performing descriptive analysis
- Transforming variables
- Coding missing data
- Analyzing by subgroups
- Creating charts for associations
- Calculating correlations
- Creating charts and statistics for three or more variables
- Creating crosstabs for categorical variables
- qqplots for data visualization
- download data from file or web
- data analysis using regression
- data analysis using clustering
- data analysis using classification
- shiny for data presentation

Module 3
Machine learning for business intelligence

In this module, we introduce the field of machine learning and describe the well-known processes, algorithms, and tools for one to be a successful machine learning practitioner. This module will help to build skills in data acquisition and modeling, classification, and regression using Java Weka API and Weka Toolkit. In addition, one will also get to explore very important tasks such as model validation, optimization, scalability, and real-time streaming.

Module Contents:
Introduction to AIand machine learning

- What is AI
- AI Applications
- AI online demo

- What is machine learning and deep learning
- How machines learn?
- The machine learning workflow: from data o deployment
- AI impact on jobs
- Types of AI

Supervised Machine Learning - Classification

- Step by step on classification problem
- Data cleaning
- Handling missing data
- Classification models
- Modeling and prediction using Weka Toolkit
- Naïve Bayes
- Support Vector Machines
- Decision Trees
- Random Forest
- Hands on practice using Weka Toolkit

Feature engineering using Weka Toolkit

- Importance of feature engineering
- Feature engineering process
- Feature selection
- Working with text features
- Working with image features
- Working with time-series features

Supervised Machine Learning - Regression

- Step by step on regression problem
- Hands on practice using Weka Toolkit/Python

Unsupervised Machine Learning- Clustering

- Clustering models
- Hands on practice using Weka Toolkit/Python

Working with Artificial Neural Network (ANN)

- What is ANN
- Hands on practice using Weka Toolkit/Python

Model evaluation and optimization using Weka Toolkit

- Evaluating prediction accuracy for unseen data
- Evaluation of classification models
- Evaluation of regression models
- Model optimization through parameter tuning

Module 4
Big Data Analytics

Big Data Analytics has become a competitive and sustainable advantage for many organizations. To harness the benefits of big data and machine learning, however, business leaders face the pressing challenge of not only acquiring the right technologies and talent to analyze and interpret the data, but also to weave a data-centric mindset into the organization's structure and cultural fabric. This module will empower with the skills and confidence to tackle data-driven opportunities and accelerate data-analysis transformation in the organization.

Module Contents:
- Theoretical and practical foundations of Big Data Analytics
- Machine Learning for Big Data
- Business Intelligence and Big Data Visualization
- Big Database Management System with No SQL Data Stores
- Infrastructure Development for Real-time Big Data Analytics

Module 5
Computer vision using OpenCV and Python

OpenCV is an open-source toolkit for Advanced Computer Vision. It is one of the most popular tools for facial recognition, used in a wide variety of security, marketing, and photography applications, and it powers a lot of cutting-edge tech, including augmented reality and Robotics. This module introduces OpenCV for Python. It will show how to leverage the image-processing power of OpenCV to identify and recognize features.

Module Contents:
- Installing and configuring OpenCV
- Data types and structures
- Image types
- Image pre-processing
- Image denoising
- Manipulating pixels
- Scaling and rotating images
- Using video inputs
- Creating custom interfaces
- Thresholding
- Object detection
- Face and feature detection
- Template matching
- CNN

- Transfer Learning
- Object Detection
- Image Classification

Module 6
Natural language processing using NLTK and Python

NLTK is the most popular Python natural language processing package that is widely used in industry and academia. It provides an easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for text classification, tokenization, stemming, lemmatization, par-of-speech (POS) tagging, parsing, and semantic reasoning. This module will teach you how to use NLTK effectively.

Module Contents:
- Basic text analysis with NLTK
- Text pre-processing
- Stopword removal
- Stemming and lemmatization
- Parts of speech tagging
- Chunking
- Named entity recognition
- Wordnet with NLTK
- Text classification
- Converting word to features
- Classifying text documents using NLTK
- Integration with scikit-learn classifiers
- Gensim
- Word2vec
- Investigating data biasness using NLTK
- Twitter sentiment analysis using NLTK

Module 7
Building AI systems through Deep Learning and NVIDIA GPUs: Keras and Tensorflow

Deep Learning (DL) is the fastest-growing field in Machine Learning and Artificial Intelligence (AI) that enables machines to be far more efficient, advanced and intelligent at predicting things. It uses many layered Deep Neural Networks (DNNs) to make sense of data such as images, sound and text, and powers some of the most interesting applications in the world like autonomous vehicles (driverless cars), speech recognition, image recognition, preventive healthcare, and more. Today's advanced DNNs use algorithms, big data, and the computational power of Graphic Processing Units (GPUs) so machines can learn at the speed, accuracy and

scale that are driving true AI computing. During the Module, you'll learn the latest techniques on how to design, train and deploy neural network-powered machine learning in your applications. You'll also explore widely used open-source frameworks and NVIDIA's latest GPU-accelerated deep learning platforms.

Module Contents:

- What is Deep Learning and what are Neural Networks?
- Artificial Neural Networks (ANN) Intuition
- Building an ANN
- Evaluating Performance of an ANN
- Hands-On Exercise
- Introduction to Keras and TenserFlow
- Convolutional Neural Networks (CNN) Intuition
- Building a CNN
- Evaluating Performance of a CNN
- Hands-On Exercise
- Recurrent Neural Networks (RNN) Intuition
- Building a RNN
- Evaluating Performance of a RNN
- Hands-On Exercise
- Image Classification with DIGITS
- Object Detection with DIGITS
- Neutral Network Deployment with DIGITS and TensorRT

Module 8
Automatic speech recognition

In this module, the process of automatic speech recognition is introduced from both hardware and software standpoints. Historical perspective of speech recognition and the problem of vocabulary representation is introduced. Afterwards, modern systems with higher level capabilities in python are presented. The process of digitization from analogue signal followed by applying Hidden Markov Model (HMM) is presented. Voice activity detectors (VADs) are presented by isolating audio signals containing speech. Finally, hands-on experience is provided to ensure practical implementation of the acquired knowledge.

Module Contents:

- How Speech Recognition Works – An Overview
- Picking a Python Speech Recognition Package
- Installing SpeechRecognition
- The Recognizer Class
- Working With Audio Files
- Supported File Types
- Using record() to Capture Data From a File

- Capturing Segments With offset and duration
- The Effect of Noise on Speech Recognition
- Working With Microphones
- Installing PyAudio
- The Microphone Class
- Using listen() to Capture Microphone Input
- Handling Unrecognizable Speech
- Putting It all Together: A "Guess the Word" Game

Target Audience for this Diploma course

Leaders, senior officials, academicians, and computer science students interested in building analytics capabilities to drive change within their Departments/Organization.

Program Fees for this Diploma course

Each month fee is 10000 PKR.

Trainers for this Diploma course

Dr. Ghulam Mujtaba Shaikh
Dr. Asif Ali Rajput
Dr. Abdul Rehman Soomrani
Dr. Javed Ahmed Shahani
Dr. Sajid Khan
Dr. Sher Mohammad Daudpota
Mr. Asim Kaaghzi