



**Department of Information Technology**

**Report**

**Academic Year: 2020 - 2021**

**Name of Event:**Add On Course

**Topic:**“ Python and Data Science with Machine Learning”

**Date & Time of Conduction:** 24/02/2021 to 27/02/2021 9:00 AM to 5:00 PM

**Venue:**International Institute of Information Technology, Pune

**Coordinator:**Prof. Keshav Tambre

**Number of Participants:**85

**Activity Description:**

Due to the Covid-19 pandemic, the Add on Course was conducted through online mode The link is shared with all the students.. Mr. Sagar Nangare was the resource person for the Course.

First Expert Mr Sagar Nangare instructed to install the latest Python version if not installed & to install the Anaconda software with Jupyter Notebook. Apart from this told about version control system. He helped Students to install all the things required and gave a demonstration of how to install the softwares and use the Github platform.

About Github:

Distribution Version Control System

How internally, GIT Manages Version Control on

Changesets. Creating Repository

Basic Commands like, git status, git add, git re-move, git branch,

git checkout, git log, git cat-file, git pull, git push, git commit

Managing Configuration – System Level, User Lev-el, Repository level

The next we were introduced was the ‘Numpy’ library. We were taught all the basics of the

Numpy module and its uses to create multidimensional array.

e.g.:

Numpy Vector and Matrix

Functions – arange (), zeros (), ones (), linspace (), eye (),

reshape (), random (), max (), min (),

argmax (), argmin (), shape and dtype

attribute Indexing and Selection

Numpy Operations – Array with Array, Array with Scalars,

Universal Array Functions

‘Pandas’ is an another python module used for data analysis and data arrangement and etc. Some of the functions that were taught at the time are as follows:

Pandas Series

Pandas Data-Frame

Missing Data

(Imputation) Group by

Operations

Merging, Joining and Concatenating Data-

Frame. Pandas Operations

Data Input and Output from wide variety of formats like csv, excel, db and html etc.

After Data framing and analysis comes the visual representation part. It is very important to have a good visual representation of the data analysis performed, to get the better insights on how the data explains itself for better results. For that we were introduced to the ‘Matplotlib’, ‘Seaborn’ and ‘Pandas-inbuilt’.

Matplotlib functions:

plot() using Functional

approach multi-plot using

subplot() plt.figure() using OO

API Methods

add\_axes(), set\_xlabel(), set\_ylabel(), set\_title()

Methods Customization – figure size, improving dpi, Plot

appearance, Markers, Control over axis appearance and

special Plot Types Seaborn functions:

Distribution Plots using `distplot()`, `jointplot()`, `pair-plot()`, `rugplot()`, `kdeplot()`

Categorical Plots using `barplot()`, `countplot()`, `box-plot()`, `violinplot()`, `stripplot()`, `swarmplot()`, `fac- torplot()`

Matrix Plots using `heatmap()`,

`clustermap()` Grid Plots using

`PairGrid()`, `FacetGrid()` Regression Plots

using `lmlplot()`

Styles and Colors customization

Pandas Built-in:

Histogram, Area Plot, Bar Plot, Scatter Plot, Box-plot, Hex-plot, Kde-plot, Density Plot e.  
Choro-pleth Maps

Interactive World Map and US Map using Plotly and Cufflinks Module

After Data Science he introduced to the topic of Machine Learning & taught about some basic things and then FAQs session.

Machine Learning session involved following things:

What is Machine Learning?

Machine Learning Process Flow-

Diagram Different Categories of

Machine Learning –

1. Supervised
2. Unsupervised
3. Reinforcement

Learning Scikit-Learn

Overview Scikit-Learn

cheat-sheet

Some important Machine Learning algorithms like Regression were taught. Here are some of the ‘regression Algorithms’:

Linear Regression

Robust Regression (RANSAC

Algorithm) Logistic Regression,

Decision Trees.

At last taught the NLP (Natural Language Processing) : Computational Linguistic, History of NLP, Why NLP, Use of NLP, NLP Components, NLP Ambiguity, Lexical Ambiguity, Syntactic Ambiguity, Referential Ambiguity, Natural Language Understanding, Natural Language Generation, Text Planning, Sentence Planning, Text Realization

Some Topics like Feature Engineering and RNN, CNN were briefed.

After that the question answer session and useful resources were shared for better understanding and practise of the learners.

Basic and frequently used algorithmic techniques including sorting, searching, greedy algorithms and dynamic programming. Familiarity with Python installation, syntax and design. At the end of this workshop, Mr. Sagar demonstrates the designing of small running projects. He conducted a doubt session & performance of students.