



Advanced Certification in Data Science and Machine Learning (ML)

Duration: 30-35 days/60-70 Hours

Course Syllabus

INTRODUCTION

This industry oriented course is developed by both the Software development division & Training division of **ipsr solutions limited**. IPSR is a **public limited IT company** with 22+ years of expertise in [Software product development](#), [Training services](#), [Placement services](#) & [Digital Marketing services](#). During the past 2 decades, IPSR has trained candidates from **60+ countries** and helped **more than One lakh candidates** to build their IT career. Our IT services division is a pioneer in development of **Academic solution products**, incorporating cutting edge technologies like Artificial Intelligence, Data Analytics and Machine learning. Live industry experts from this IT division contribute a major role in delivering this course. Our placement division is having **1600+ placement tie-up companies** and we are conducting [recruitment on all days](#).

The Course curriculum is designed and developed by a team of expertise panel lead by following academicians

- ❑ **Dr. Mendus Jacob, M.Sc., M.Phil., Ph.D., MloD**
 - ❑ M.D & C.E.O - IPSR & Valin Technologies, U.K.
 - ❑ Director - MCA, Marian College, Kuttikkanam (Autonomous)
 - ❑ Former Director of School of Applicable Mathematics, M.G. University.
 - ❑ Academician and Entrepreneur with 30+ years experience
- ❑ **Dr. Sunil Job K.A, M.Sc, M.Ed, M.Phil, Ph.D., RHCE**
 - ❑ Chief of Academic Solutions - IPSR
 - ❑ Former college Principal and a Specialist in Data Analytics & Machine Learning
 - ❑ Blogger and a Resource person for National conferences
 - ❑ Academician with 25+ years experience



What you'll learn

- Python Language Basics, Sequences, Functions, OOPS, RegEx, Database
- Programming using Python and MySQL
- Python Packages for Basic Data Analytics - numpy, pandas, matplotlib, seaborn
- Exploratory and Explanatory data analysis using Python packages.
- data processing and post processing using python
- Machine Learning using Python packages.

Description

This course covers the Python Language basics and upto database programming using MySQL. It even covers important data analytics packages as numpy, pandas, matplotlib and seaborn and also covers data pre-processing for machine learning and prediction using real world datasets.

Course Outcome (CO)



While successfully completing this course, the learner will be able to:

- Data analysis using Python packages.
- Data processing and post processing in structured and unstructured data using python
- Demonstrate ability to engage in various roles and responsibilities of a Machine learning engineer
- Practice analytical skills in data preparation, analysis and visualization using Python
- Practice analytical skills in data preprocessing, machine learning with Python
- Prepare Webapps with machine learning prediction tasks.

What does this course give you?

1. Python skills needed for basic Data Analytics and the skill required to do Data Analytics using Python packages in real world data
2. Skill required to implement Machine Learning in the real world.

SECTION 1:

Course content

1. Python Basics



- Language Structure
- Control & Loop Constructs
- Sequences
- Functions
- OOPS
- Database programming with MySQL

2. Python Data Analytics packages

1. Numpy

- Ndimensional array
- Datatypes
- Random numbers
- Matrix operations

2. Pandas

- Series,Dataframe
- Importing csv, Exporting csv
- Groupby
- Describe,Info
- Iloc,loc
- Filtering
- Slicing

3. Matplotlib

- Line plot
- Scatter plot



- Histogram
- Box plot

4. Seaborn

- Heatmap

5. Mini Project

- a. Mini Project Using Data Visualization Tools.
- b. Mini Project Using Python Data Analytics Package

SECTION 2:

1. What is Data Analytics

- Data Analytics vs Data Analysis
- What is a dataset
- Need for Dataset structure

2. Data Analytics Workflow Theory

- Data Requirement Specification
- Data Collection
- Optional Data Integration
- Data Selection

3. Exploratory Data Analysis with Descriptive statistics

- Central tendency



- Standard deviation
- Interquartile range
- Histograms
- Distributions
- Skew
- Kurtosis
- Correlation

4. Pattern Processing from Data

- Data Pre-processing
- Data Processing
- Data Post-processing

5. Code running environment

- Introduction to google colab
- connecting google colab to local data and drive data

6. Data Analytics Workflow Practical

1. Data Collection using Pycaret and Pandas
2. Exploratory Data analysis using Pandas profiling==2.8.0 library and seaborn heatmap(correlation)



- Drill down in profile Report
- Obtaining irregularities in dataset

Note:For more Dataset with columns more than 30 correlation checks in pandas profiling should be turned off or minimal mode should be turned on.

3.Data pre-processing Using Python ml libraries(Pycaret setup function and pandas drop duplicates function)

- Missing value treatment
- Normalization
- transformation
- multicollinearity issue
- outliers issue

- Low variance issue
- Dimensionality Reduction for large column datasets
- setting for GPU usage
- Using EDA Report for deciding must do data cleaning steps
- Application of pre-processing steps totally

SECTION 3:



1. What is Machine Learning

- Machine Learning relation with Data Analytics
- What is Prediction
- Need for Prediction Task

2. Supervised Learning

- What is Training and Testing?
- Accuracy
- Data splitting as Training and Testing
- Independent Columns and Dependent Target/Label column
- Cross -Validation splitting
- Target Data Imbalance
- Data Sampling

4. Introduction To Github account

- Creating Github Account
- Repository creation, Push Pull, commit using Github onlinetools
- Creating streamlit sharing account
- Requesting for streamlit sharing invite

5. Machine Learning Prediction Algorithms

1. Regression



- Linear Regression
- Parameters
- Mean Absolute Error, MSE, RMSE
- OLS Regression
- Decision Tree Regression
- Random Forest Regression

2. Classification

- Decision Boundary
- Logistic Regression
- Hyper Parameters
- Accuracy, AUC, Recall, Precision, F1-score
- Decision Tree (CART)
- Random Forest Classification
- KNN
- Linear SVM
- Radial SVM
- Naïve Bayes

6. Data Processing -code working



- Data Collection, Task setting as Regression/Prediction
- Separating Data for unseen data prediction
- Data Pre-processing
- Model comparison study
- Selecting Best model
- Tuning Hyperparameters -optimizing for Accuracy measures
- Finalizing model
- Pickling model
- Loading model and testing in new notebook

7. Data Post Processing -code working

- What is a Webapp?
- Introduction to streamlit
- Streamlit template file for webapp
- Testing webapp in colab with pyngrok hosting

8. Permanent deployment of tested webapp

- creating Requirements.txt file
- Uploading project to Github
- Streamlit sharing invite mail for web hosting
- Deployment of webapp



9. Mini Project

Contact Us

ipsr solutions limited

Merchant's Association Building

M.L. Road, Kottayam - 686001

Kerala, India, Pin-686001

Phone: +91-481 2561410, 2561420, 2301085

Mobile: +91 9447294635, +91 9447169776

Email: training@ipsrsolutions.com

Website: <https://www.ipsr.org>

We have branches at Kochi, Calicut, Trivandrum and Bengaluru.