

## Course Name: Machine Learning Using Python

Course Code: EMBSBIO64

Duration: 1 hour daily- 1 month

To apply online : <http://www.eminentbio.com/home/Onlineform>

Trainer : Dr. Anuraj Nayariseri

<http://www.eminentbio.com/home/members>

[https://www.researchgate.net/profile/Anuraj\\_Nayariseri](https://www.researchgate.net/profile/Anuraj_Nayariseri)

Eminent Biosciences  
91, Sector-A,  
Mahalakshmi Nagar,  
Indore - 452010,  
Madhya Pradesh, India.  
[www.eminentbio.com](http://www.eminentbio.com)  
[info@eminentbio.com](mailto:info@eminentbio.com)  
Mob:+91 9752295342  
+91 8720000784

### Course Content

#### Introduction to Numpy

- Create 1D and 2D array
- Fetch elements
- Array operations
- Shape, size, axis concept

#### Introduction to Pandas

- Series
- DataFrames
- Create series from ndarray/dict
- Fetch elements using Position/Label
- Vectorized operations and label alignment with Series
- Dataframes from Series/Dict of Dicts/Dict of lists
- Read from csv/tsv file
- Column selection, addition, deletion
- Indexing / selection
- Loc, iloc
- Data alignment and arithmetic
- Merge tables
- Play with IRIS dataset
- Groupby

#### Introduction to Matplotlib

- Scatter plot
- Histogram
- Sub plots
- Line plot
- Bar plot
- Box plot

#### Introduction to Seaborn

- Visualizing statistical relationships
- relplot / Scatter plot

- Line Plot
- Faceting
- Categorical scatterplots
- catplot
- Boxplot
- Beeswarm plot
- Boxen plot
- Violinplots
- Count plot
- Bar plots
- Wide/Long format data
- Pair plot
- Pair plot
- Hexbin plots
- Kernel density estimation based plots

#### Explore inbuilt Datasets

- Explore datasets package
- Explore IRIS dataset
- Understand bunch (Features, Targets)
- Visualize data using seaborn

#### Prepare Training and Test Datasets

- Explore model\_selection.train\_test\_split module
- Prepare training and test dataset

#### Simple model building process

- Define model/estimator
- Fit training datasets into model
- Test using test dataset
- Evaluate Model performance
- score() function of estimator
- confusion\_matrix
- Accuracy

- precision\_score
- recall\_score
- f1\_score
- classification\_report
- confusion matrix plot
- Explore classifiers
- Nearest Neighbors
- Linear SVM
- RBF SVM
- DecisionTreeClassifier
- RandomForestClassifier
- LogisticRegression
- Neural Net
- AdaBoost
- Naive Bayes
- Gaussian Process
- Explore cross validation strategy
- Explore model\_selection.cross\_val\_score module
- k fold cross validation

#### Data Preprocessing

- Standardization (Mean removal and variance scaling)
- Understand why standardization is important?
- Explore StandardScaler class
- Scaling features to fall within a range
- MinMaxScaler
- MaxAbsScaler
- Scaling data with outliers
- Non-linear transformation
- Mapping to a Uniform distribution
- Mapping to a Gaussian distribution
- Compare the effect of different scalers on data with outliers

# Course Name: Machine Learning Using Python

## Feature Selection

Removing features with low variance  
Explore VarianceThreshold module  
Univariate feature selection  
SelectKBest  
SelectPercentile  
SelectFdr  
SelectFwe  
SelectFpr  
Univariate scoring methods  
chi2  
f\_classif  
mutual\_info\_classif

## Tuning the hyper-parameters of an estimator

Exhaustive Grid Search

## Clustering

k-Means Clustering  
Hierarchical Clustering

## Regression

Introduction to Regression  
Simple Linear Regression  
Model Evaluation in Regression Models  
Evaluation Metrics in Regression Models  
Multiple Linear Regression

## Dimensionality Reduction

Principal component analysis (PCA)

## Case study

Breast cancer classification based on omics data  
from TCGA

EMINENT<sup>®</sup>

BIO SCIENCES

Enlightening the Modern Bioinformatics