

Course Durations: 40 Hours

Course Mode: Online/Offline

About Company:

EduNextgen extended arm of Product Innovation Academy is a growing entity in education and career transformation, specializing in today's most in-demand skills. A platform with blended learning programs supported by in-trend technology platforms for learning. Engaging organizations for learning development objectives.

Training courses are designed and updated by renowned industry experts. Our blended learning approach combines online classes, instructor-led live virtual classrooms and virtual teaching assistance.

About The Course:

The Data Science and Machine Learning is designed the depth of Data Science and Machine Learning Curriculum using Python. This course will help you to move your first step on Data Science world. You will get in depth knowledge on Visualization and Metrics, Sampling and Estimating, Probability, Clustering, Classification, KNN/Tree Method, Regression etc. The Course is bundled with Industry Grade Projects, Assignments and Use Cases.

Why This Course:

- Covers Visualization and Metrics, Sampling and Estimating, Probability, Clustering, Classification, KNN/Tree Method, Regression etc
- Hands-on Experience
- Industry Grade Use Cases
- Online Practical Platform
- Live Support (24x7)

Participants will get the Access to:

- LMS Access
- Cloud Lab
- 50+ Assignments
- 100+ Quizzes
- 5+ Industry Grade Projects
- Live Support via Mail, Call and Screen Sharing
- Course Completion Certificate

Batch Schedule (Online):

Weekend: 3 Hours per day

Weekday: 2 Hours per day

Batch Schedule (Offline):

Weekend: 4 Hours per day

Weekday: 2 Hours per day

Course Curriculum

Module 1: Introduction to Data Science and Machine Learning (2 Hours)

This module will help you to understand what Data Science is and how it solves modern day challenges. You will also get to know what makes a Data Scientist and the job family. You will also be introduced to evolution of Data Science role into Machine Learning:

- What is Data Science?
- What Problems are Solved with Data Science?
- Who is Data Scientist?
- Data Science Family of Jobs
- What is Machine Learning?
- What kind of Problems are Solved with Machine Learning?
- What are broad family of Machine Learning Techniques?
- Unsupervised / Supervised – Classification and Regression
- What is difference between ML and AI? What role do they play in Data Science?

Hands on/Programs/Practical:

- List Down Applications of Data Science in your area of work.
 - Choose an Application and Document Requirement for one challenge
 - List Down Data needed to solve the challenge
-

Module 2: Python Primer – if needed (3 Hours)

This module will help you get hands on with Python as Programming Language for Data Science and Machine Learning:

- Installing Python and Jupyter
- Numpy, Pandas, matplotlib
- Basic Code Execution in .ipynb and .py files
- Variables, Data Structures (List, Tuple, Dictionaries)
- Functions and working with files

Hands on/Programs/Practical:

- All the Topics are hands on
 - Functions and working with files
-

Module 3: Data Comprehension through Visualization and Metrics (2 Hours)

This module will help you to read data in context, visualize, calculate basic metrics and draw insights about data:

- What is different data types? (Nominal, Ordinal, Interval, Ratio)
 - Standard Visuals for different Data Types
 - Standard Metrics for different Data Types
-

- Histograms, Frequency Distribution, Probability Distributions
- Interpreting Various Distributions Forms

Hands on/Programs/Practical:

- Calculate Standard Metrics using Python
 - Interpreting Distributions
-

Module 4: Sampling and Estimating (3 Hours)

This module will help you to appreciate the need for Right Sampling and it's Utility in ML:

- Case for Sampling
- What are different Sampling Techniques?
- What is Estimation and how do we Estimate?
- Validating Estimates with Hypotheses Testing
- Types of Error and Confusion Matrix
- Sampling for Machine Learning (Train/Test/Validate)

Hands on/Programs/Practical:

- Representative Sampling Plan and Drawing Inferences for sample
 - Using Python to Validate Estimates
-

Module 5: Probability Refresher (3 Hours)

Probability is key to arrive at and Validate Estimates. ML is all about Estimating. This is refresher for basic probability required for ML:

- Modeling Uncertainty and Case for Probability
- Probability Distribution
- Joint Probability
- Conditional Probability
- Prior Probability
- Posterior Probability

Hands on/Programs/Practical:

- Drawing Probabilistic Inferences
 - Calculating Probability with Python
-

Module 6: Recommendation Systems - Market Basket Analysis (3 Hours)

Market Basket Analysis is carried out to understand what goes in the basket of Retail Customers. The direct benefit of it is planning inventory and recommending retailer on what clubbing-offers will make sense for customers. Market Basket Analysis is the case of applying Probability and ONLY Probability to build a valuable Recommendations System:

- Market Basket Analysis Case Study
 - Application of Probability to identify items in basket
-

- Variants of Market Basket Analysis (Conjoint Analysis / Discrete Choice Modeling)
- Other Recommendation Systems (collaborative filtering / content-based filtering)

Hands on/Programs/Practical:

- Market Basket Analysis
 - Recommendation from Market Basket Analysis
-

Module 7: Clustering - Market Segmentation (5 Hours)

Clustering is a Technique to Reduce Data to Manageable Chunks. You will be exposed various Clustering Techniques and their Applications. Market Segmentation is one of them:

- Case for Clustering
- Clustering Techniques - Hierarchical
- Clustering Techniques – K-Means
- Feature Engineering
- Choosing Appropriate Technique

Hands on/Programs/Practical:

- Segmentation Case Study
-

Module 8: Principal Component Analysis (3 Hours)

PCA is a popular technique to reduce data when there are large number of features at our disposal to enable machine learning:

- Working with multiple features that are Correlated
- Correlation and Related Concepts
- Feature Filtering
- Math behind PCA – Linear algebra (Eigen Values / Eigen Vectors)

Hands on/Programs/Practical:

- Feature Engineering Case Study
-

Module 9: Classification – Exploring KNN/Tree Method (6 Hours)

Classification problems are well addressed ML techniques. In this, we will learn Probabilistic/Tree/Regression/NN based Classification Technique:

- K-NN Classification Technique
 - Tree Based Classification Technique
 - Math behind Tree Based Classification Techniques (Entropy)
 - Naïve Bayes – Probabilistic Classification Technique
 - Logistic Regression Overview – Regression based Classification Technique
 - Appreciating need for different Techniques
 - Measuring Accuracy with Confusion Matrix
 - Trade off – Bias and Variance
-

Hands on/Programs/Practical:

- Tree Based Classification – Case Study
 - KNN Based Classification – Case Study
-

Module 10: Regression (6 Hours)

Regression is Technique to Predict Continuous Values. Regression finds application in many areas and the most challenging in ML. In this you will learn different Techniques, Challenges and Solutions along with Applications:

- Relationship between Variables/Features
- Linear/Non-Linear Regression
- Cost Function to Measure Effectiveness of Models
- Logistic Regression – Regression for Classification
- Problem of Overfitting and Regularization
- Trade off – Bias and Variance

Hands on/Programs/Practical:

- Linear Regression – Case Study
 - Non-Linear Classification – Case Study
-

Module 11: ML Model Evaluation and Boosting Accuracy (2 Hours)

ML Models works with a level of Accuracy. This module will give methods to Improve Accuracy:

- Ensemble Techniques – Boosting / Bagging
- Random Forest
- Cross Validation – K-Fold
- Trade off – Bias and Variance

Hands on/Programs/Practical:

- Bagging and Boosting Exercises
-

Module 12: Neural Networks and other Advanced Concepts (2 Hour)

This module will give overview of Neural Network and other Advanced Concepts. We will also discuss deploying ML Model in Production:

- Neural Network
- Support Vector Machines
- Natural Language Processing
- Time Series Techniques
- Deploying ML Model in Production

Hands on/Programs/Practical: N/A