

MASTERY IN DATA PRE-PROCESSING

Introduction

- Start your Data Preprocessing journey today

What You'll Learn

This course is designed to provide practical examples and hands-on exercises to reinforce the concepts. Additionally, encourage students to work on a mini-project or real datasets to gain practical experience in data preprocessing.

Highlights

- **Course Duration - 3 days (24 Hours)**
 - **Number of Modules - 10**
-

Module 1: Introduction to Data Preprocessing

- What is Data Preprocessing?
- Why Data Preprocessing is Essential in Machine Learning
- The Data Science Workflow
- Common Challenges in Real-World Data

Module 2: Data Collection and Acquisition

- Data Sources and Formats
- Web Scraping and APIs
- Handling Structured and Unstructured Data
- Data Privacy and Ethics

Module 3: Data Cleaning

- Identifying Missing Data
- Handling Missing Values: Imputation Techniques
- Outlier Detection and Handling
- Data Duplicates and Noise Removal

Module 4: Data Transformation

- Scaling and Normalisation
- Encoding Categorical Data: One-Hot Encoding vs. Label Encoding
- Feature Engineering: Creating New Features
- Date and Time Handling

Module 5: Data Visualization and Exploration

- Data Visualization Libraries (Matplotlib, Seaborn)
- Univariate and Bivariate Analysis
- Correlation Analysis
- Data Distribution and Transformation

Module 6: Feature Selection and Dimensionality Reduction

- Importance of Feature Selection
- Common Feature Selection Techniques
- Principal Component Analysis (PCA)
- t-Distributed Stochastic Neighbour Embedding (t-SNE)

Module 7: Data Splitting and Sampling Techniques

- Training, Validation, and Test Sets
- Cross-Validation
- Stratified Sampling
- Oversampling and Under-sampling for Imbalanced Data

Module 8: Handling Time Series Data

- Understanding Time Series Data
- Resampling and Interpolation
- Time Series Feature Engineering
- Seasonal Decomposition

Module 9: Text Data Preprocessing

- Tokenization and Stopword Removal
- Text Vectorization (TF-IDF, Word Embeddings)
- Handling Text Data Challenges (Spelling Errors, Noise)
- Sentiment Analysis

Module 10: Handling Image and Audio Data

- Image Preprocessing (Resizing, Normalisation)
- Image Augmentation
- Audio Data Preprocessing
- Spectrogram and MFCC Features
- Challenges in Image and Audio Data