MASTERY IN FEATURE ENGINEERING

Introduction

- Start your Feature Engineering journey today

What You'll Learn

This course is designed to provide a comprehensive understanding of feature engineering for data scientists, covering both fundamental and advanced concepts with practical implementation.

Highlights

- **Course Duration - 3 days (24 Hours)**
- **Number of Modules - 16**

Course Content

Module 1: Introduction to Feature Engineering

- Welcome and course overview
- Importance of feature engineering in machine learning
- Exploring the GitHub repository for resources
- Data preprocessing and cleaning
- Handling missing values and outliers
- Practical exercises using the GitHub resources
- Define what feature engineering is and its critical role in machine learning.
- Explain the importance of selecting and transforming features.
- Highlight how good feature engineering can lead to better model performance.

Module 2: Feature Creation and Extraction

- Feature creation through mathematical operations
- Date and time feature engineering
- Utilising domain knowledge for feature extraction
- Text feature engineering (NLP)
- Image feature engineering (Computer Vision)
- Hands-on exercises and coding using GitHub resources

Module 3: Handling Imbalanced Datasets

- Understanding imbalanced datasets
- Resampling techniques (oversampling and undersampling)

- Synthetic data generation (SMOTE)
- Cost-sensitive learning
- Evaluation metrics for imbalanced datasets
- Practical exercises and case studies from the GitHub repository

Module 4: Exploratory Data Analysis (EDA) for Feature Engineering

- Discuss the process of EDA and its connection to feature engineering.
- Demonstrate techniques to visualise and understand the data.
- Identify patterns and relationships that inform feature selection and creation.

Module 5: Handling Missing Data

- Strategies for dealing with missing data in features.
- Techniques like imputation, deletion, and advanced methods like K-nearest neighbours imputation.
- The impact of different approaches on model performance.

Module 6: Categorical Feature Encoding

- Explain how to handle categorical variables in machine learning models.
- Discuss one-hot encoding, label encoding, and target encoding.
- Best practices for encoding depending on the nature of the data.

Module 7: Feature Scaling and Normalisation

- Detail the importance of scaling and normalizing features.
- Discuss techniques like Min-Max scaling, Z-score normalization, and robust scaling.
- Demonstrate how improper scaling can affect model performance.

Module 8: Feature Transformation

- Feature scaling and normalization
- Encoding categorical variables
- One-hot encoding vs. label encoding
- Dimensionality reduction techniques (PCA, t-SNE)
- Feature selection methods
- Practical exercises using the provided GitHub resources

Module 9: Feature Selection Techniques

- Explore feature selection methods such as univariate selection, recursive feature elimination, and feature importance from tree-based models.
- Discuss the concept of dimensionality reduction and when to apply it.

Module 10: Feature Extraction from Text and Images

- Explain techniques for extracting meaningful features from unstructured data like text and images.
- Discuss text vectorization methods (e.g., TF-IDF, word embeddings) and image feature extraction methods (e.g., CNN-based features).
- Real-world examples of text and image feature engineering.

Module 11: Feature Engineering for Time Series Data

- Introduction to time series data
- Time-based features (lag, rolling statistics)
- Handling seasonality and trends
- Feature extraction from textual time series data
- Hands-on coding with time series datasets from the GitHub repository
- Q&A and discussion
- Feature engineering strategies for time series data.
- Discuss lag features, rolling statistics, and seasonal decomposition.
- Demonstrate the use of time-related features in time series forecasting.

Module 12: Feature Engineering for Specialized Domains

- Explore feature engineering considerations for specialized domains like natural language processing (NLP), computer vision, and recommendation systems.
- Techniques such as word embeddings, image embeddings, and collaborative filtering features.

Module 13: Feature Engineering for Model Interpretability

- Importance of interpretable features
- SHAP values and feature importance
- Building interpretable feature sets
- Model-agnostic interpretability techniques
- Hands-on practice using GitHub resources
- Case studies and real-world applications

Module 14: Advanced Techniques in Feature Engineering

- Feature engineering for geospatial data
- Feature hashing and binning
- Embedding and deep learning-based feature extraction
- Handling time-based variables and cyclical features
- Practical implementation using provided resources
- Project discussion

Module 15: Feature Engineering in Machine Learning Frameworks

- Feature engineering in scikit-learn
- Feature tools and automated feature engineering
- Utilising TensorFlow and PyTorch for deep learning-based feature engineering
- Guest lecture or expert panel discussion on industry best practices
- Real-world feature engineering challenges and solutions

Module 16: Hands-On Feature Engineering Project

- Walk through a practical feature engineering project.
- Apply various feature engineering techniques to a real-world dataset.
- Evaluate the impact of feature engineering on model performance.

_____ _____ _____ _____ _____ _____