KOENIG SOLUTIONS · COURSE OUTLINE

Feature Engineering: Encoding, SMOTE & Time Series

Start your Feature Engineering journey today

3 Days	16	74
24 HOURS	MODULES	TOPICS

COURSE OVERVIEW

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

WHO THIS COURSE IS FOR

- **Data Scientists**
- **Machine Learning Engineers**
- **Data Analysts**
- **NLP Engineers**
- **Computer Vision Engineers**

LEARNING OBJECTIVES

- Explain the role of feature engineering in machine learning
- Apply data cleaning and missing value handling techniques
- Create features using mathematical, date and domain-based methods
- Implement resampling and SMOTE for imbalanced datasets
- Compare categorical encoding methods such as one-hot and target encoding
- Apply feature scaling and normalisation techniques
- Evaluate feature selection and dimensionality reduction methods
- Develop time series features such as lag and rolling statistics

COURSE SYLLABUS

16 Modules, 74 Topics

01

Introduction to Feature Engineering

TOPICS

- 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

02

Feature Creation and Extraction

TOPICS

- 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

03

Handling Imbalanced Datasets

TOPICS

- 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

04

Exploratory Data Analysis (EDA) for Feature Engineering

TOPICS

- 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

05

Handling Missing Data

TOPICS

- 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

06

Categorical Feature Encoding

TOPICS

- 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

07

Feature Scaling and Normalisation

TOPICS

- 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

08

Feature Transformation

TOPICS

- 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

09

Feature Selection Techniques

TOPICS

- 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

10

Feature Extraction from Text and Images

TOPICS

- 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

11

Feature Engineering for Time Series Data

TOPICS

- 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
- 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

12

Feature Engineering for Specialized Domains

TOPICS

- 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

13

Feature Engineering for Model Interpretability

TOPICS

- 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

14

Advanced Techniques in Feature Engineering

TOPICS

- 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

15

Feature Engineering in Machine Learning Frameworks

TOPICS

- 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

16

Hands-On Feature Engineering Project

TOPICS

- 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

WHAT IS INCLUDED

Everything You Need to Succeed

- Koenig course material for every module
- Delivery by certified instructors and subject matter experts
- Hands-on exercises throughout the course
- Course completion certificate from Koenig

ABOUT KOENIG

Who We Are

Established in 1993, Koenig Solutions is a globally recognised training provider with 35+ technology partnerships, 5,000+ training programmes and a network of certified instructors and subject matter experts, supporting individuals and enterprise teams worldwide.

FLEXIBLE DELIVERY

Learn the Way That Suits You

Classroom Training | Live Online Classes (ILO) | Fly-Me-A-Trainer (FMAT) | Flexi (Self-Paced) | 1-on-1 Training

Ready to Get Started?

Talk to a learning advisor about schedules, group training and customised delivery.

Visit koenig-solutions.com | info@koenig-solutions.com