Training Course on Feature Engineering and Selection (Advanced)

Professional Development Training Course Outline

Category	Data Science	**Duration**	10 Days
Base Fee	$3,000 USD	**Delivery Mode**	Online / On-site Hybrid

Course Introduction

In the rapidly evolving landscape of **Machine Learning (ML)** and **Artificial Intelligence (AI)**, the quality and relevance of input features often dictate the success of predictive models more than the choice of algorithm itself. This advanced training course on **Feature Engineering** and **Feature Selection** delves deep into the art and science of transforming raw data into powerful, informative features and strategically reducing dimensionality. Participants will master cutting-edge techniques to enhance model accuracy, improve interpretability, and optimize computational efficiency, equipping them with indispensable skills for building robust and deployable AI/ML solutions in real-world scenarios.

Training Course on Feature Engineering & Selection (Advanced) goes beyond foundational concepts, exploring advanced methodologies for **data preprocessing**, **feature creation**, and **dimensionality reduction**. Through practical, hands-on exercises and real-world case studies, attendees will learn to identify key data patterns, handle complex data types including **time-series**, **text**, and **image data**, and employ sophisticated **feature extraction** and **selection algorithms**. By the end of this course, professionals will possess the expertise to significantly elevate the performance and generalizability of their machine learning models, driving impactful data-driven decisions within their organizations.

Programme Curriculum

Training Course on Feature Engineering & Selection (Advanced): Techniques for creating powerful features and dimensionality reduction

Introduction

In the rapidly evolving landscape of **Machine Learning (ML)** and **Artificial Intelligence (AI)**, the quality and relevance of input features often dictate the success of predictive models more than the

choice of algorithm itself. This advanced training course on **Feature Engineering** and **Feature Selection** delves deep into the art and science of transforming raw data into powerful, informative features and strategically reducing dimensionality. Participants will master cutting-edge techniques to enhance model accuracy, improve interpretability, and optimize computational efficiency, equipping them with indispensable skills for building robust and deployable AI/ML solutions in real-world scenarios.

Training Course on Feature Engineering & Selection (Advanced) goes beyond foundational concepts, exploring advanced methodologies for **data preprocessing**, **feature creation**, and **dimensionality reduction**. Through practical, hands-on exercises and real-world case studies, attendees will learn to identify key data patterns, handle complex data types including **time-series**, **text**, and **image data**, and employ sophisticated **feature extraction** and **selection algorithms**. By the end of this course, professionals will possess the expertise to significantly elevate the performance and generalizability of their machine learning models, driving impactful data-driven decisions within their organizations.

Course Duration

10 days

Course Objectives

1. Implement sophisticated techniques for handling missing values, outliers, and noisy data in complex datasets.
2. Apply Exploratory Data Analysis (EDA) with advanced visualization tools to reveal critical relationships and anomalies for effective feature design.
3. Develop expert-level skills in creating transformative features from raw data, including polynomial, interaction, and domain-specific features.
4. Utilize automated feature engineering (AutoFE) tools and techniques to efficiently generate and evaluate a high volume of potential features.
5. Apply cutting-edge methods for encoding categorical variables with high cardinality, such as target encoding and embedding techniques.
6. Master time-series feature engineering, including lag features, rolling statistics, and seasonality extraction for forecasting models.
7. Learn advanced feature extraction techniques for Natural Language Processing (NLP) (e.g., word embeddings, TF-IDF) and Computer Vision (e.g., deep features).
8. Effectively reduce data complexity using advanced methods like PCA, t-SNE, UMAP, and factor analysis while preserving essential information.
9. Implement a range of filter, wrapper, and embedded methods for selecting the most relevant features and combating the curse of dimensionality.
10. Design and implement end-to-end feature pipelines using frameworks like scikit-learn pipelines and specialized feature stores for efficient deployment.
11. Critically assess the contribution of engineered features using various model evaluation metrics, feature importance plots, and A/B testing.
12. Understand strategies for monitoring and managing feature drift in production environments and ensuring the long-term viability of feature sets.
13. Explore techniques that enhance model interpretability through judicious feature creation and selection, aligning with Responsible AI principles.

Organizational Benefits

- Organizations will build more precise and reliable machine learning models, leading to better predictions and improved decision-making across various business functions.
- Optimized feature sets decrease the dimensionality of data, resulting in quicker model training times and lower computational resource consumption.
- Well-engineered features enable a deeper understanding of model predictions, fostering trust and facilitating compliance with regulatory requirements (e.g., **XAI**).
- Teams will be equipped with advanced skills to tackle complex data challenges, leading to the development of sophisticated and highly performant AI solutions.
- By identifying and leveraging the most impactful features, organizations can allocate resources more effectively towards data collection and engineering efforts.
- A strong grasp of feature engineering empowers teams to derive maximum value from their data assets, fostering innovation and the creation of new data products and services.
- Strategic feature selection mitigates the risk of overfitting, ensuring models perform robustly on unseen data, which is crucial for real-world deployment.

Target Audience

1. Data Scientists.
2. Machine Learning Engineers.
3. AI Researchers.
4. Data Analysts with ML Interest.
5. Software Engineers (with ML exposure).
6. Statisticians.
7. Deep Learning Practitioners
8. Researchers and Academics.

Course Outline

Module 1: Foundations of Advanced Feature Engineering & Selection

- Revisiting Feature Engineering Principles: The critical role of features in model accuracy and interpretability.
- Data Quality Assessment: Advanced techniques for identifying and diagnosing issues in raw data.
- The Curse of Dimensionality in modern datasets and its implications for model building.
- Feature Engineering Workflow: A structured approach from raw data to production-ready features.
- Introduction to advanced feature generation and feature reduction strategies.
- **Case Study:** Analyzing a public healthcare dataset to identify initial data quality challenges and potential feature engineering opportunities for disease prediction.

Module 2: Advanced Data Cleaning and Imputation

- Robust methods for handling diverse types of missing values (MCAR, MAR, MNAR).
- Outlier Detection and Treatment: Statistical and ML-based approaches (e.g., Isolation Forest, IQR Rule).
- Data Normalization and Standardization: Beyond basic scaling, understanding when and why different methods are crucial.
- Handling Inconsistent Data Formats and ensuring data integrity.
- Data Augmentation techniques for enhancing dataset richness.
- **Case Study:** Cleaning a noisy e-commerce transaction dataset, including imputing missing purchase values and identifying fraudulent outliers to improve customer segmentation.

Module 3: Feature Creation from Numerical Data

- Polynomial Features: Capturing non-linear relationships and interaction terms.
- Binning and Discretization: Optimal strategies for transforming continuous variables.
- Feature Interactions: Systematically discovering and creating meaningful combinations of features.
- Statistical Features: Deriving descriptive statistics (e.g., mean, std, quantiles) for groups or windows.
- Ratio and Difference Features: Creating expressive features from existing numerical attributes.
- **Case Study:** Enhancing a housing price prediction model by engineering polynomial features of living area and interaction terms between neighborhood and property age.

Module 4: Advanced Categorical Feature Engineering

- High-Cardinality Encoding: Strategies for handling categories with many unique values (e.g., Target Encoding, Leave-One-Out Encoding).
- Ordinal Encoding vs. One-Hot Encoding: Deeper dive into appropriate use cases and limitations.
- Feature Hashing: A scalable approach for high-dimensional categorical data.
- Categorical Feature Interactions: Combining categorical features for richer representations.
- Dealing with rare categories and their impact on model stability.
- **Case Study:** Improving a customer churn prediction model by effectively encoding customer demographic and subscription history features, including handling product categories with many unique items.

Module 5: Time-Series Feature Engineering

- Lag Features: Creating features from past observations to capture temporal dependencies.
- Rolling Window Statistics: Aggregating data over time windows (e.g., moving averages, standard deviations).
- Date and Time-based Features: Extracting components like day of week, hour of day, and cyclical features.
- Trend and Seasonality Extraction: Decomposing time series for predictive modeling.
- Handling irregular time series and missing temporal data.
- **Case Study:** Developing features for a stock price prediction model by incorporating daily lag features, weekly rolling averages, and identifying seasonal patterns.

Module 6: Text Feature Engineering (NLP)

- Advanced Text Preprocessing: Tokenization, lemmatization, stemming, and custom stop word removal.
- TF-IDF (Term Frequency-Inverse Document Frequency): Understanding its nuances and applications.
- Word Embeddings (Word2Vec, GloVe, FastText): Leveraging pre-trained and custom embeddings.
- Sentence and Document Embeddings: Aggregating word embeddings for higher-level representations.
- Introduction to Transformer-based embeddings (e.g., BERT, GPT-style features) for text data.
- **Case Study:** Building features from customer review text data to improve a product sentiment analysis classifier, using both TF-IDF and word embeddings.

Module 7: Feature Engineering for Image and Unstructured Data

- Feature Extraction from Images: Introduction to traditional methods (e.g., SIFT, HOG) and transfer learning with Convolutional Neural Networks (CNNs).
- Deep Features: Utilizing pre-trained CNNs as feature extractors for various computer vision tasks.
- Feature Engineering from Audio Data: Basic signal processing techniques and mel-frequency cepstral coefficients (MFCCs).
- Feature Extraction from Graph Data: Graph embeddings and node features for network analysis.
- Challenges and best practices for unstructured data feature engineering.
- **Case Study:** Applying transfer learning from a pre-trained CNN (e.g., ResNet) to extract features from medical images for diagnostic purposes.

Module 8: Filter Methods for Feature Selection

- Univariate Feature Selection: Correlation, Chi-squared, ANOVA F-value, Mutual Information.
- Variance Thresholding: Removing features with low variance.
- Correlation-based Feature Selection (CFS): Identifying and removing highly correlated features.
- Information Gain and Gini Impurity: Feature importance measures from tree-based models.
- Comparison and appropriate use cases for different filter methods.
- **Case Study:** Using filter methods to reduce the dimensionality of a genetic dataset for a disease susceptibility prediction model, focusing on highly informative genes.

Module 9: Wrapper Methods for Feature Selection

- Forward Selection: Iteratively adding features based on model performance.
- Backward Elimination: Iteratively removing features to improve model performance.
- Recursive Feature Elimination (RFE): Feature ranking based on model coefficients.
- Stochastic Search Methods: Genetic algorithms and simulated annealing for feature selection.
- Computational cost and advantages/disadvantages of wrapper methods.
- **Case Study:** Optimizing a fraud detection model using RFE to select a subset of features that maximizes the F1-score while minimizing computational overhead.

Module 10: Embedded Methods for Feature Selection

- L1 Regularization (Lasso): Feature selection through sparse model coefficients.
- Tree-based Feature Importance: Gaining insights from decision trees, Random Forests, and Gradient Boosting Machines (GBMs).
- Feature Importance from Neural Networks: Interpreting weights and activation patterns.
- SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations): Advanced interpretability for feature contribution.
- Combining embedded methods with other selection strategies.
- **Case Study:** Utilizing XGBoost's feature importance to identify the most critical factors influencing customer purchase behavior in a retail dataset.

Module 11: Dimensionality Reduction Techniques

- Principal Component Analysis (PCA): Theory, implementation, and interpretation of principal components.
- Linear Discriminant Analysis (LDA): Supervised dimensionality reduction for classification.

- Non-linear Dimensionality Reduction: t-SNE and UMAP for visualization and feature extraction.
- Factor Analysis: Uncovering latent variables and their relationships.
- When to use dimensionality reduction versus feature selection.
- **Case Study:** Reducing the high-dimensional feature space of a bioinformatics dataset using PCA to enable better visualization and downstream clustering.

Module 12: Automated Feature Engineering (AutoFE) and Feature Stores

- Introduction to AutoML frameworks and their AutoFE capabilities (e.g., Featuretools).
- Deep Feature Synthesis: Automatically generating complex features from relational data.
- Feature Stores: Concepts, benefits, and practical considerations for managing and sharing features in production.
- Operationalizing Feature Engineering: Best practices for robust and reproducible feature pipelines.
- Monitoring feature drift and maintaining feature quality in real-time systems.
- **Case Study:** Designing a feature store for a recommendation system to ensure consistent feature availability and reuse across different models.

Module 13: Feature Engineering for Interpretable AI & Responsible AI

- The relationship between feature engineering and

Upcoming Scheduled Sessions

Start Date	End Date	Location	Price
23 Nov 2026	04 Dec 2026	Online	$2,000
23 Nov 2026	04 Dec 2026	Physical	$3,000
23 Nov 2026	04 Dec 2026	Physical	$6,000
23 Nov 2026	04 Dec 2026	Physical	$6,000
23 Nov 2026	04 Dec 2026	Physical	$12,000
23 Nov 2026	04 Dec 2026	Physical	$9,000
23 Nov 2026	04 Dec 2026	Physical	$12,000
23 Nov 2026	04 Dec 2026	Physical	$12,000

Ready to enroll or request a customized program? Book online or contact us:

Register Online Now

Email: info@fineskilltrainingcenter.com | Website: www.fineskilltrainingcenter.com