

Feature Engineering for Omics & Clinical Data — Hands-on

Design high-signal, trustworthy features for genomic, proteomic, metabolomic, imaging-derived, and clinical tabular data. This module focuses on rigorous preprocessing, leakage-safe splits, domain-aware transformations (e.g., gene-set activity scores), selection and stability, and packaging features into reproducible pipelines ready for modeling and deployment.

Feature Engineering for Omics & Clinical Data

[Help Desk](#) · [WhatsApp](#)

Session Index

[Session 1 — Preprocessing, Encoders & Normalization](#) [Session 2 — Leakage-Safe Splits & Feature Selection](#) [Session 3 — Domain Features: Gene-Set & Clinical Scores](#) [Session 4 — Mini Capstone: Reproducible FE Pipeline](#)

Session 1

Fee: Rs 12,800 [Apply Now](#)

Preprocessing, Encoders & Normalization

Data quality & missingness handling

[MCAR/MAR/MNAR](#) [imputation \(KNN/MICE\)](#) [outliers & winsorization](#)

Encoders & scalers

[one-hot/target/WOE](#) [standard/robust/quantile](#) [Yeo-Johnson/Box-Cox](#)

Omics-specific normalization

TPM/CPM/RPKM log/CLR batch-effect correction

Session 2

Fee: Rs 17,200 Apply Now

Leakage-Safe Splits & Feature Selection

Leakage traps & proper CV design

pipeline-first splits nested CV group/time-aware CV

Selection methods

filter: correlation/MI/F-test wrapper: RFE/SFS

embedded: L1/trees/GBMs

Stability & dimensionality reduction

bootstrap stability PCA/UMAP variance & sparsity controls

Session 3

Fee: Rs 22,400 Apply Now

Domain Features: Gene-Set & Clinical Scores

Pathway-level features for expression data

GSEA/GSVA/ssGSEA KEGG/Reactome/GO sets meta-genes & modules

Clinical feature crafting

composite risk scores time-window aggregations interactions & polynomials

Feature drift & monitoring signals

PSI/KS population shift alerts threshold design

Session 4

Fee: Rs 28,800 Apply Now

Mini Capstone: Reproducible FE Pipeline

Build an end-to-end feature pipeline for a bio dataset

Theory + Practical

Package & persist features

sklearn pipelines **feature store basics** **data dictionary**

Deliverables: notebook, pipeline artifact & report

.ipynb/.py **serialized pipeline** **PDF/HTML**