COURSE TABLE OF CONTENTS

AI/ML Applications in Life Sciences

OEM: Koenig Original • Duration: 6 Days (45 hrs)

COURSE MODULES & TOPICS

Module 1: Foundations of AI, Biological Data and Cloud Computing

- AI vs. ML vs. Deep Learning in life sciences
- Historical milestones: BLAST to AlphaFold 3
- Biological data types: tabular, sequence, structural, chemical, imaging
- FAIR data principles and public repositories
- Introduction to cloud notebooks for biology
- Python basics for biological data manipulation
- Biopython and pandas for parsing biological files
- No-code bioinformatics and ML platforms (Galaxy)

Module 2: Biological Data Preprocessing and Exploratory Analysis

- Data cleaning: missing values, duplicates, batch effects, outliers
- Feature scaling: normalization, standardization, TPM/FPKM
- Encoding biological data: one-hot, k-mer frequency
- Exploratory data analysis: plots and trend discovery
- Dimensionality reduction: PCA, t-SNE, UMAP
- Data splitting and validation strategies
- Preventing data leakage in biological datasets

Module 3: Classical Machine Learning in Life Sciences

- Supervised learning: regression (Linear, Ridge, Lasso)
- Supervised learning: classification (Logistic, k-NN, SVM)
- Tree-based models: Decision Trees, Random Forests, XGBoost
- Unsupervised learning: clustering (K-Means, Hierarchical, DBSCAN)
- Model evaluation: accuracy, precision, recall, ROC-AUC, PR-AUC
- Handling class imbalance: SMOTE, class weighting
- Explainable AI: SHAP values and feature attribution

Module 4: Deep Learning Foundations for Biology

- Artificial neural networks: perceptrons, MLPs, activation functions
- Training and optimization: gradient descent, Adam, SGD
- Regularization: dropout, batch normalization, early stopping
- Convolutional neural networks: 1D/2D CNNs for biological data
- Sequence modeling: RNNs, LSTMs, attention mechanisms
- Transformers and biological foundation models

Module 5: AI in Genomics, Transcriptomics and Single-Cell Analysis

- AI applications in genomics and variant calling
- Transcriptomics: gene expression and RNA-seq analysis
- Single-cell omics: clustering and trajectory inference
- Integration of multi-omics data

© Koenig Solutions

- Case studies in genomics and transcriptomics
- Hands-on: AI pipelines for omics data

Module 6: Structural Biology and Protein Language Models

- Protein structure prediction with AI
- Protein language models and embeddings
- Structural bioinformatics: PDB, mmCIF data
- AI for protein function and interaction prediction
- Hands-on: exploring AlphaFold and ESM models

Module 7: AI in Drug Discovery and Cheminformatics

- AI-driven drug discovery workflows
- Cheminformatics: SMILES, molecular descriptors
- Predicting drug-target interactions
- Virtual screening and compound prioritization
- Hands-on: ML models for drug potency prediction

Module 8: Bioimage Analysis, Clinical AI and Multi-Omics

- AI for bioimage analysis: microscopy and radiology
- Clinical AI: diagnostics and patient stratification
- Multi-omics data integration
- Case studies in clinical and imaging AI

Module 9: Generative Bio-AI, Ethics and Capstone Project

- Generative models in biology
- Ethical considerations in AI for life sciences
- Capstone project: end-to-end AI/ML pipeline