Machine Learning (ML) For Compound Property Prediction Training Course

Professional Development Training Course Outline

Category	Biotechnology and Pharmaceutical Development	**Duration**	10 Days
Base Fee	$4,000 USD	**Delivery Mode**	Online / On-site Hybrid

Course Introduction

In the era of computational chemistry, materials informatics, and drug discovery, the ability to accurately predict compound properties using machine learning is a game changer. Machine Learning (ML) For Compound Property Prediction Training Course empowers participants with the latest data driven modeling, feature engineering, and predictive modeling techniques tailored for molecular and materials data, bridging the gap between raw data and actionable predictions. Participants will walk away with hands on capabilities to build, validate, and deploy models for compound property prediction using state-of-the-art algorithms, graph neural networks, and ensemble learning.

Throughout the training, emphasis will be placed on combining domain knowledge and machine learning best practices including cross validation, model interpretability, uncertainty quantification, and model deployment so that the predictive models are not only accurate but also robust and explainable. By the end, attendees will be ready to tackle real world tasks in materials design, pharmaceutical development, chemical engineering, or advanced materials R&D using scalable machine learning pipelines.

Programme Curriculum

Machine Learning for Compound Property Prediction Training Course

Introduction

This course delves into the transformative application of **Machine Learning (ML)** in the field of **computational chemistry** and **drug discovery**. By leveraging vast chemical datasets and powerful ML algorithms, we can **accelerate** the identification of new compounds with desired properties, such as therapeutic activity, solubility, or toxicity. This **paradigm shift** moves beyond traditional, resource-intensive experimental methods, enabling

predictive modeling and **virtual screening** to rapidly filter and prioritize molecular candidates. The integration of advanced techniques like **deep learning** and **graph neural networks** with **cheminformatics** is revolutionizing how we design, optimize, and discover new materials and medicines.

Machine Learning for Compound Property Prediction Training Course provides a comprehensive, hands-on experience, bridging the gap between theoretical ML concepts and their practical implementation in molecular science. Participants will gain proficiency in **data curation**, **feature engineering**, and **model deployment** for **molecular property prediction**. We will explore state-of-the-art architectures and frameworks, focusing on their application to real-world challenges in areas like **ADME/Tox prediction** and **rational drug design**. The course culminates in a project-based approach, empowering learners to build robust predictive pipelines and contribute to the next generation of data-driven scientific innovation.

Course Duration

10 Days

Course Objectives

1. Understand the fundamentals of **machine learning for cheminformatics**.
2. Master data preprocessing and **curation of chemical datasets**.
3. Apply **molecular representation** techniques, including SMILES, fingerprints, and molecular graphs.
4. Build and evaluate **predictive models** for various chemical properties.
5. Implement **deep learning models**, such as Graph Neural Networks (GNNs), for molecular property prediction.
6. Perform **feature engineering** to enhance model performance.
7. Conduct **virtual screening** and **de novo drug design** using ML models.
8. Assess and mitigate **model bias** and data-related challenges.
9. Interpret and explain model predictions using **Explainable AI (XAI)** techniques.
10. Utilize cloud computing platforms for scalable **model training and deployment**.
11. Explore applications in **ADME/Tox prediction** and **pharmacokinetics**.
12. Integrate ML models into a complete **drug discovery pipeline**.
13. Stay current with **trending research** in ML-driven molecular science.

Target Audience

1. Computational Chemists and Bioinformaticians seeking to expand their skill sets.
2. Data Scientists and Machine Learning Engineers interested in applying their skills to molecular science.
3. Medicinal Chemists and Pharmacologists looking to integrate AI into their research workflows.
4. Graduate students and postdocs in chemistry, biology, or data science.
5. Researchers in pharmaceutical and biotech companies.
6. Professionals involved in materials science and agrochemicals.
7. Software developers and engineers building tools for molecular science.
8. Project managers and R&D leaders overseeing data-driven discovery initiatives.

Course Modules

Module 1: Introduction to ML in Cheminformatics

- Understanding chemical representations (SMILES, SDF).
- From problem definition to model deployment.
- Overview of regression, classification, and clustering.

- Introduction to Python, RDKit, and NumPy.
- **Case Study:** Predicting drug-likeness for a new compound library.

Module 2: Data Curation & Molecular Descriptors

- Accessing public databases like ChEMBL and PubChem.
- Handling missing values, duplicates, and standardization.
- Generating molecular fingerprints and physiochemical properties.
- Strategies for creating robust training and test sets.
- **Case Study:** Curating a dataset for solubility prediction and generating descriptors.

Module 3: Foundational Machine Learning Models

- Predictive modeling for simple properties.
- Decision Trees, Random Forests, and Gradient Boosting.
- Application in binary classification.
- Metrics for regression (RMSE) and classification (AUC-ROC).
- **Case Study:** Predicting the blood-brain barrier permeability of a small molecule using Random Forest.

Module 4: Deep Learning for Molecular Structures

- Architecture, activation functions, and backpropagation.
- Using images of molecular structures.
- Handling sequential data like SMILES strings.
- A deep dive into modeling molecules as graphs.
- **Case Study:** Implementing a GNN to predict molecular toxicity from a graph representation.

Module 5: Advanced Predictive Modeling

- Predicting multiple properties with a single model.
- Leveraging pre-trained models for new tasks.
- Combining multiple models for improved accuracy.
- Quantifying uncertainty in predictions.
- **Case Study:** Building a multi-task learning model for ADME properties (Absorption, Distribution, Metabolism, Excretion).

Module 6: De Novo Drug Design & Generative Models

- Creating novel molecules.
- Designing molecules with specific properties.
- Optimizing molecule generation based on desired rewards.
- Iteratively improving a candidate compound.
- **Case Study:** Generating a new library of molecules with a generative model, optimized for a specific target activity.

Module 7: Explainable AI in Molecular Science

- Understanding the need for model interpretability.
- Local and global interpretability methods.
- Identifying key molecular features driving predictions.
- Visualizing predictions and model outputs on a molecular level.

- **Case Study:** Interpreting the predictions of a GNN to understand which parts of a molecule contribute to its bioactivity.

Module 8: Practical Deployment and MLOps

- Preparing models for production use.
- Building a predictive service with Flask or FastAPI.
- **Cloud Deployment:** Deploying a model on platforms like AWS, GCP, or Azure.
- Ensuring the model's performance over time.
- **Case Study:** Deploying a toxicity prediction model as a web service for real-time inference.

Module 9: Specialized Applications in ADME/Tox

- The critical role of these properties in drug development.
- Using ML for pharmacokinetics.
- Identifying potential adverse effects (e.g., hERG inhibition).
- Building an ADME/Tox dataset from literature.
- **Case Study:** Developing a model to predict the hepatotoxicity of novel compounds.

Module 10: Quantitative Structure-Activity Relationships (QSAR)

- A brief overview of the classical approach.
- Combining ML with QSAR principles.
- **:** The importance of statistical significance in QSAR.
- Following industry standards for QSAR model publication.
- **Case Study:** Building and validating a QSAR model for a specific enzyme inhibitor.

Module 11: High-Throughput Screening with ML

- Managing and preprocessing large screening datasets.
- Assessing the efficiency of virtual screening.
- Iterative model improvement with new data.
- Using models to guide chemical optimization.
- **Case Study:** Using a predictive model to re-rank compounds from a virtual screen and identify new hits.

Module 12: Machine Learning for Protein & Ligand Interactions

- Understanding 3D structural data.
- Integrating traditional docking with ML.
- Predicting how strongly a ligand binds to a protein.
- GNNs on protein structures.
- **Case Study:** Developing a model to predict the binding affinity of small molecules to a specific protein target.

Module 13: The Future of AI in Drug Discovery

- An introduction to quantum computing for molecular problems.
- The impact of large language models (LLMs) and foundation models.
- Automated synthesis and experimentation.
- Navigating the challenges of AI in regulated industries.
- **Case Study:** Exploring a recent breakthrough paper in AI for molecular science and discussing its implications.

Module 14: Capstone Project I: Project Scoping and Data

- Choosing a project from a curated list of challenges.
- Locating and preparing relevant datasets.
- Choosing appropriate molecular representations.
- Creating a simple, initial model to set a benchmark.
- **Case Study:** Scoping a project to predict the melting point of organic compounds.

Module 15: Capstone Project II: Model Building & Presentation

- Implementing a deep learning or ensemble model.
- Optimizing model performance.
- Rigorous testing and validation of the final model.
- Communicating results, insights, and future work.
- **Case Study:** Presenting a final project on a novel anti-cancer drug candidate identified via an ML pipeline.

Training Methodology

- Interactive Lectures
- Hands-on Workshops.
- Case Study Analysis.
- Mock Negotiations.
- Guest Speaker Sessions

Register as a group from 3 participants for a Discount

Send us an email: info@fineskilltrainingcenter.org or call *+254769199797*

Certification

Upon successful completion of this training, participants will be issued with a globally- recognized certificate.

Tailor-Made Course

We also offer tailor-made courses based on your needs.

Key Notes

a. The participant must be conversant with English.

b. Upon completion of training the participant will be issued with an Authorized Training Certificate

c. Course duration is flexible and the contents can be modified to fit any number of days.

d. The course fee includes facilitation training materials, 2 coffee breaks, buffet lunch and A Certificate upon successful completion of Training.

e. One-year post-training support Consultation and Coaching provided after the course.

f. Payment should be done at least a week before commence of the training, to Fineskill Training Center account, as indicated in the invoice so as to enable us prepare better for you.

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