Indian Institute of Technology Kanpur

Proposal for a New Course

1. **Course No:** CHE6XX
2. **Course Title:** Machine Learning in Chemical Engineering
3. **Per Week Lectures:** 3 (L), Tutorial: 0 (T), Laboratory: 0 (P), Additional Hours [0-2]: 0 (A),
   Credits (3-0-0-0): **Duration of Course:** Full semester
4. **Proposing Department:** Department of Chemical Engineering

   Other Departments/IDPs which may be interested in the proposed course: MSE, ME, AE

   Other faculty members interested in teaching the proposed course: Indranil Saha Dalal, Ishan Bajaj, and Akash Chaudhry

5. **Proposing Instructor(s):** Salman Ahmad Khan (salman@iitk.ac.in)
6. **Course Description:**

   A) **Objectives:** This course will introduce students to machine learning (ML) applications across various areas of chemical engineering. Students will also be introduced to various ML Python libraries and most lectures will involve hands-on case studies. The course is particularly targeted towards postgraduate students and has been designed to help both, experimentalists and computational researchers.

   B) Contents (preferably in the form of 5 to 10 broad titles):

   Lecture-wise break-up (considering the duration of each lecture is 50 minutes)

S. No.	Broad Title	Topics	No. of Lectures
1.	Basics of machine learning (ML) and its application to chemical engineering	Broad overview of ML and its application to various areas of chemical engineering. Introduction to Python and Python libraries to be used in the course.	3
2.	Data pre-processing for chemical engineering data	Introduction to different types of data in chemical engineering (molecular datasets and spatiotemporal data). Handling missing data and feature scaling. Introduction to the scikit-learn Python library.	3
3.	Dimensionality reduction techniques	Feature selection and linear dimensionality reduction (PCA, LDA). Non-linear dimensionality reduction (diffusion maps).	3
4.	Classification methods applied to chemical engineering data	Logistic regression, k-nearest neighbours, decision trees, and random forest. Introduction to the RDKIT python library for cheminformatics. Implementing classification methods for molecular data with RDKIT and scikit-learn.	3
3.	Regression methods applied to chemical engineering data	Linear regression, polynomial regression, and lasso regression. Implementing regression methods with the scikit-learn Python library. Regression methods applied to predicting chemical properties from chemical/material structure (e.g. solubility, heat capacities etc.).	4

4.	Neural networks and deep learning in chemical engineering	Multi-layer perceptron, activation functions, and backpropagation. Training neural networks with the PyTorch Python library. Modeling reactor dynamics with neural networks.	3
5.	Physics informed neural networks (PINNs) for solving PDEs for chemical engineering problems	Introduction to PINNs and training PINNs with the PINA Python library. Solving reaction diffusion PDEs with PINNs.	3
6.	Uncertainty quantification of ML predicted properties	Ensemble methods, Gini importance, and Bayesian optimization. Introduction to the PyStan Python library for Bayesian inference and its application to molecular properties predicted from structure.	5
7.	Natural Language Processing (NLP) and Large language models (LLMs)	Introduction to NLP and LLMs. Application of LLMs to mining and organizing chemical data from unstructured data sources.	3
8.	Transfer learning and domain adaptation	Introduction to transfer learning and domain adaptation. Introduction to the ADAPT Python library applied to learn molecular properties with limited data.	4
9.	Symbolic regression for discovering equations from data	Genetic programming for symbolic regression. Introduction to the gplearn Python package to discover symbolic expressions from thermodynamic data.	3
10.	Material and chemical discovery with generative AI	Generative adversarial networks and variational autoencoders. Application of generative AI for material discovery.	3
Total			40

C) **Recommended pre-requisites, if any (examples: a- PSO201A, or b- PSO201A or equivalent):** None

D) **Short summary for including in the Courses of Study Booklet:** Machine Learning in Chemical Engineering, Chemical Informatics, Python for Machine Learning, Classification and Regression methods, Uncertainty Quantification.

7. Recommended text/reference books:

A) Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, Jonathan Taylor, *An Introduction to Statistical Learning: with Applications in Python*, 1st Edition, Springer, 2023.

B) Andrew Ng, *Machine Learning Yearning*, 2018.

C) Dan Jurafsky, James H. Martin, *Speech and Language Processing*, 2nd Edition, Prentice Hall, 2009.

D) Jeremy Watt, Reza Bohrani, Aggelos K. Kastaggelos, *Machine Learning Refined*, 2nd Edition, Cambridge University Press, 2020.

8. Any other remarks: None

Dated: 20/02/2025 Proposer: Salman Ahmad Khan

Dated: DPGC Convener:

The course is approved / not approved

Chairman, SUGC

Dated: