Project Title:

Predict First, Calculate Never: Machine Learning in Chemistry

Project Supervisor:

Dr. Saikat Mukherjee,

Asst. Prof. (Adjunct), Molecular Modeling Group, Faculty of Chemistry,

Room 213A, Building A

Project Overview:

The field of chemistry is undergoing a paradigm shift with the integration of Machine Learning (ML). From drug discovery to materials design, ML is being used to make predictions that traditionally required time-consuming quantum mechanical calculations. This project offers an exciting, hands-on introduction to how data-driven models can be trained to "learn chemistry" and make predictions of molecular properties instantly in no time.

Why is this Important?

The emergence of Machine Learning (ML) in quantum chemistry has finally created a way to advance the state of the art. ML predicts quantum chemical quantities very fast, bypassing the computationally expensive electronic structure calculations. Algorithms based on kernel-ridge regression models and neural networks are becoming popular in the supervised learning of highly nonlinear chemical data. Once the ML model has been trained to the reference quantum chemical data to reproduce the molecular properties, the costly quantum chemistry calculation is skipped, resulting in an extremely fast prediction of properties for new molecular configurations.

What Will Students Do?

This project is designed for beginners. The students will:

- **Learn the fundamentals of machine learning** – including supervised learning and model evaluation.
- **Get introduced to chemical database**
- **Explore chemical representations** for ML, such as SMILES strings or molecular fingerprints.
- **Implement and train simple models**, like kernel-ridge regression or feedforward neural networks.
- **Predict molecular properties** like atomization energies, dipole moments, or HOMO-LUMO gaps.
- **Visualize and interpret** model results to understand chemical trends.

Expected Outcomes:

By the end of the project, the students will:

- Gain practical experience with ML workflows.
- Build working ML models that predict properties of unseen molecules.
- Understand how ML complements traditional quantum chemistry.
- Prepare themselves for more advanced study or research in cheminformatics and computational chemistry.

Prerequisites:

- Basic knowledge of chemistry (molecular structure, bonding).
- No prior ML and coding experience is necessary—just curiosity and a willingness to learn.

Tools and Platforms:

- Python (Jupyter notebooks)
- Scikit-learn, TensorFlow or PyTorch (for ML)
- RDKit (for handling molecular structures)