**Course Title:** Machine Learning in Drug Design

**Course Code:** BI-811

**Course Objectives:**

This course aims to introduce the course participants to different machine learning methods, with particular emphasis on applications of machine learning methodologies in modern drug discovery pipeline. The course will comprise a mixture of hands-on training sections, lectures, and informal discussions. A key focus will be on the underlying rationale/theory, rather than teaching which buttons to press in a particular piece of software

**Course Outcomes:**

By the end of the course, students should be able to:

- Understand complexity of Machine Learning algorithms and their limitations
- Understand how to apply a variety of learning algorithms to biological and chemical data.
- Understand how to perform evaluation of learning algorithms and model selection

**Course Contents**

a. Lessons learned from R&D Failures
    i. Any recent study about R&D failures and success for example
        1. Learning lessons from Pfizer's $800 million failure
        2. Lessons learned from the fate of AstraZeneca's drug pipeline: a five dimensional framework
        3. Oncology Case Studies and Lesson learned
        4. Lessons Learned and Potentials for Improvement in CNS Drug Development
        5. Success Stories of Computer- Aided Drug Design
b. Publicly available chemical and biological space (databases)
c. Characterization of the Chemical and Biological Space (databases)
        1. Protein Target Descriptors
        2. Ligands descriptors
        3. Cross-term descriptors
        4. Applicability Domain
d. Feature Extraction (Data Preprocessing)
        1. Feature Collection
        2. Feature Computation
        3. Normalization
        4. Discretization
e. Feature Selection
        1. Feature subset selection

2. Dimensionality reduction

f. Data Classification Algorithms)
    1. Algorithm 1 – k-Nearest Neighbor Classifier
    2. Algorithm 2 – Decision Tree
    3. Algorithm 3 – Support Vector Machine
    4. Algorithm 4 – Artificial Neural Network
    5. Algorithm 6- Self organizing maps
    6. Proteochemometrics Modeling
    7. Ensemble (Classification) Methods: Bagging, Random Forests, Boosting, Stacking
g. Post-processing
    1. Visualization /
    2. Model/ Pattern interpretation
h. Model Evaluation
    1. N-fold cross validation
    2. Leave One out Cross Validation
    3. Y-scrambling
    4. Classification Algorithms

**Recommended / Reference Books:**

Following reference books can be consulted. Additional supporting material will be provided, as and when required.


a. Pharmaceutical Data mining by Konstin V Balakin, Wiley interscience, 2010
b. Neural Networks in QSAR and Drug design by James Devillers, 2010
c. Virtual screening in Drug discovery Susan Napier · Matilda Bingham, Springer Science, 2009