COURSE TITLE	**MACHINE LEARNING FOR BIOINFORMATICS**
COURSE CODE	**01CB0607**
COURSE CREDITS	**4**

Objective:

1 To develop a foundational understanding of Machine Learning in Bioinformatics.
2 To build students' skills in applying and evaluating machine learning models effectively.

Course Outcomes: After completion of this course, student will be able to:

1 Describe and classify the major types of machine learning, supervised, unsupervised, and reinforcement learning.
2 Analyse and compare the suitability of different ML paradigms for biological data problems.
3 Apply regression, classification, and clustering models to bioinformatics datasets.
4 Evaluate and optimize model performance using metrics such as precision, recall, F1-score, and confusion matrices.

Pre-requisite of course: Computer Programming, Python and Mathematics

Teaching and Examination Scheme

Theory Hours	Tutorial Hours	Practical Hours	ESE	IA	CSE	Viva	Term Work
3	0	2	50	30	20	25	25

Contents : Unit	Topics	Contact Hours
1	**Foundations of Machine Learning** Introduction to Machine Learning: scope, applications, and types, Supervised vs Unsupervised Learning, Reinforcement Learning: basic concepts and use cases	4
2	**Regression Models & Evaluation Metrics** Linear Regression: cost function, gradient descent, learning rate, Multiple Linear Regression (MLR): vectorization, feature scaling, feature engineering, Model Evaluation Metrics: accuracy, precision, recall, F1-score, sensitivity, specificity, confusion matrix. Bias-Variance Trade-off	10
3	**Classification, Clustering & Dimensionality Reduction** Logistic Regression: cost function, gradient descent, Overfitting and Regularization: L1/L2 for linear and logistic models, Unsupervised Learning: k-Means Clustering: choosing optimal k, Principal Component Analysis (PCA): dimensionality reduction	9

Contents : Unit	Topics	Contact Hours
4	**Decision Trees & Ensemble Methods** Decision Tree Learning: impurity measures (Gini, entropy), information gain, Handling categorical features: one-hot encoding, Ensemble Techniques: Random Forest, XGBoost: boosting fundamentals and implementation	9
5	**Artificial Neural Networks (ANN)** Neural Network Architecture: layers, weights, biases, Activation Functions: sigmoid, ReLU, tanh, Derivatives and backpropagation, Gradient Descent for Neural Networks, Parameter vs Hyperparameter tuning	7
6	**Application of Machine Learning in Bioinformatics** k-mean clustering, and classification and simple neural network for modelling of Gene expression dataset etc	3
Total Hours		**42**

Suggested List of Experiments:

Contents : Unit	Topics	Contact Hours
1	**Experiment 1** Apply basic Pandas functions to load, manipulate, and summarize tabular datasets.	2
2	**Experiment 2** Construct visual representations of data using Matplotlib and Seaborn to identify patterns and trends.	2
3	**Experiment 3** Demonstrate the use of NumPy for performing numerical computations and array operations.	2
4	**Experiment 4** Develop a linear regression model using gradient descent and evaluate its performance on sample data.	2
5	**Experiment 5** Implement multiple linear regression with regularization techniques and analyze the impact of regularization on model accuracy.	2
6	**Experiment 6** Construct a logistic regression model using gradient descent and interpret its classification results	2
7	**Experiment 7** Apply k-means clustering to a dataset and evaluate the optimal number of clusters using the elbow method.	2
8	**Experiment 8** Build a decision tree classifier and explain its decision-making process using sample data.	2
9	**Experiment 9** Perform random forest model and compare its performance with a single decision tree.	2

Suggested List of Experiments:

Contents : Unit	Topics	Contact Hours
10	**Experiment 10** Apply XGBoost to a dataset and assess its effectiveness in improving predictive accuracy.	2
11	**Experiment 11** Perform one-hot encoding and feature engineering to prepare data for machine learning models.	2
12	**Experiment 12** Conduct hyperparameter tuning to optimize model performance using sample datasets.	2
13	**Experiment 13** Design and train an artificial neural network using sample dataset-1 and evaluate its accuracy.	2
14	**Experiment 14** Modify and retrain the neural network using sample dataset-2 to compare performance across datasets.	2
Total Hours		**28**

Textbook :

1 Introduction to Machine Learning , Ethem Alpaydin , The MIT Press, 2014

References:

1 Engineering optimization: Theory and Practice, Engineering optimization: Theory and Practice, S. S. Rao, New Age publication, 2019

2 Optimization for Engineering Design: Algorithms and Examples, Optimization for Engineering Design: Algorithms and Examples, Kalyanmoy Deb, PHI Learning Private Limited, 2012

3 Engineering optimization: Methods and Applications, Engineering optimization: Methods and Applications, A. Ravindran, K. Ragsdell and G. Reklaiti, , John Wiley and Sons, 2006

4 Machine Learning an algorithmic perspective, Machine Learning an algorithmic perspective, Stephen Marsland, CRC Press, 2015

Suggested Theory Distribution:

The suggested theory distribution as per Bloom's taxonomy is as follows. This distribution serves as guidelines for teachers and students to achieve effective teaching-learning process

Distribution of Theory for course delivery					
Remember / Knowledge	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Higher order Thinking / Creative**
0.00	0.00	35.00	35.00	30.00	0.00

Instructional Method:

1. The course delivery method will depend upon the requirement of content and need of students. The teacher in addition to conventional teaching method by black board, may also use any of tools such as demonstration, role play, Quiz, brainstorming, MOOCs etc.
2. The internal evaluation will be done on the basis of continuous evaluation of students in the laboratory and class-room.
3. Practical examination will be conducted at the end of semester for evaluation of performance of students in laboratory.
4. Students will use supplementary resources such as online videos, NPTEL videos, e-courses, Virtual Laboratory.

Supplementary Resources:

1. https://machinelearningmastery.com/
2. https://www.geeksforgeeks.org/machine-learning/
3. https://www.javatpoint.com/machine-learning
4. https://huggingface.co/

DR. SEJAL SAHILKRISHNA SHAH

Digitally signed by (Name of HOD)

DR. RAJENDRASINH BAHADURSINH JADEJA

Digitally signed by (Name of Dean/ Principal)