CENTRAL UNIVERSITY OF PUNJAB BATHINDA

SYLLABUS

PG Diploma in Data Science for Bioinformatics

Session – 2025 Onwards

Department of Computer Science & Technology
Central University of Punjab, VPO-Ghudda, Bathinda, Punjab, India- 151401

Proposed Course Curriculum
PG Diploma in Data Science for Bioinformatics

Course Description: The Course provides a broad and practical overview of selected techniques and concepts in rapidly developing areas such as Computational Biology, systems biology, network biology, synthetic biology, data analytics, predictive modelling, machine learning, and machine intelligence. Topics relevant for the student and faculty of computer fraternity, working biologists, computational scientists, and applied investigator (Biotechnology and engineering).

Objective of Programme

The programme provides a broad and practical overview of available techniques and concepts in the area of computational bioscience and machine learning. The aims of the diploma course are:

1. To support the development of Data Sciences field in Life Science and Computer Engineering students. This will be for students either coming from a bioscience or computer science background.
2. To support the analysis of complex biological systems at difference scales, ranging from molecules, cells, organs, to organisms.
3. To acquire a working knowledge on which techniques from machine learning, network theory, modelling, and statistics are useful and how to apply them in the analysis of biological systems.
4. To advance from data to information about significant correlated features hidden in the data and advancing to insights into biological mechanisms.
5. To provide the student with a comprehensive conceptual and in part practical overview of this inter-disciplinary area. Yet programme will include practical computer exercises, in order to ground the lectures in hands-on knowledge.
6. To emphasize the computational technique in life sciences. Computational techniques are necessary, not yet sufficient, to understand and analyse data produced from living systems.

Relevance of the Course:

The course is relevant for the researchers in life-science in academia or industry, and data-scientists developing new analytic techniques in research or industry. Upon the completion of the course the student can readily advance to further studies in selected sub-areas. Alternatively, the student develops the background and ability to collaborate and identify the relevant specialists in the appropriate sub-area.

Pre-Requisite:

Knowledge and skills corresponding to basic Mathematics and Statistics.

Eligibility: Minimum qualification as UG Courses in Life Sciences/Pharmacy, B.E./B.Tech(Computer Science & Technology, Information Technology, Bio-Technology)/BCA/B.Sc(Information Technology, Computer Science).

Duration of Program: 1 Year (2 Semester)
No. of seats: 30

Course Structure of the Programme

Semester-I

Semester -I				
Course Code	Course Name	Type	Credit	Credit hours
BAI-701	Statistics in Data Sciences	Core/Theory	2	2
BAI-702	Omics in Health Informatics	Core/Theory	3	3
BAI-703	Artificial Neural Network	Theory	3	3
BAI-704	Python Programming for ANN -Lab	Practical	2	4
BAI-741	Seminar	Practical	2	4
BAI-798	Project	Practical	8	16
	Total		**20**	**32**

Semester-II

Semester -II				
Course Code	Course Name	Type	Credit	Credit hours
BAI-751	Machine Learning for Data Science	Theory	4	4
BAI-752 BAI-753 BAI-75X	**Elective-I(Any One)** 1.Biological Intelligence 2.Neuronal Networking 3.Any other Equivalent from Health science School as Elective	Theory	4	4
BAI-760	Machine Learning for Data Science Lab	Practical	2	4
BAI-742	Seminar	Practical	2	4
BAI-799	Project	Practical	8	12
	Total		**20**	**28**

SEMESTER-I

Course Code: BAI-701
Course Title: Statistics in Data Science
Total Hours: 30

L	T	P	Cr
2	0	0	2

Course Objective:
The objective of this course is to provide students with a solid foundation in statistical methods essential for data science. It focuses on key concepts such as descriptive statistics, probability, inferential techniques, hypothesis testing, and regression analysis. The course aims to develop analytical thinking and the ability to apply statistical tools to real-world data, supporting effective decision-making and laying the groundwork for advanced data science and machine learning applications.

Course Learning Outcomes:
After completion of course, students would be able to:
CLO1: Able to perform inferential statistics for the data analysis of biological data.
CLO2: Learn about the computer skills for biological data management; Able to learn statistics software for graphical presentation.
CLO3: Students will able to implement the statistical techniques over biological data.

Units/Hours	Contents	Mapping with Course Learning Outcome
I **15 Hours**	**Overview of Statistics:** Types of Studies, Levels of Measurements, Presentation of Data: Frequency tables and diagrams, Descriptive statistics: Measures of central tendency and dispersal, Kurtosis and Skewness, Error Bars, Moments, Normality Tests and Outliers	**CLO1** **CLO2** **CLO3**
II **15 Hours**	**Inferential Statistics-I:** t-Distribution and tests of significance based on t-distribution, F-distribution and tests of significance based on F distribution, $\chi 2$ Distribution and tests of significance based on $\chi 2$ distribution, Comparing Proportions, Gaussian, Binomial, Lognormal and Poisson Distributions, Pearson's Correlation, Simple Linear Regression, Non-Linear Regression, Nonparametric tests, One-way and two-way analysis of variance (ANOVA) **Inferential Statistics-II:** Standard errors of regression coefficients, Comparing two regression lines, Pearson Product-Moment Correlation Coefficient, Spearman Rank Correlation Coefficient, Power and sampling size in correlation and regression.	**CLO2** **CLO3**

Transactional Modes:
- Lecture cum Demonstration
- Peer Learning/Teaching
- E-tutorial
- Self-Learning

Text Books:

1. Motulsky, Harvey. *Intuitive biostatistics: a nonmathematical guide to statistical thinking*. Oxford University Press, USA, 2014.
2. Van Belle, Gerald, et al. *Biostatistics: a methodology for the health sciences*. Vol. 519. John Wiley & Sons, 2004.

Suggested Readings:

1. Le, Chap T., and Lynn E. Eberly. *Introductory biostatistics*. John Wiley & Sons, 2016.
2. Norman, Geoffrey R., and David L. Streiner. *Biostatistics: the bare essentials*. PMPH USA (BC Decker publishers), 2008.
3. Rohlf, F. James, and Robert R. Sokal. *Biometry: the principles and practice of statistics in biological research*. New York: Freeman publishers, 1

L	T	P	Cr
3	0	0	3

Course Code: BAI-702
Course Title: Artificial Neural Networks
Total Hours:45

Course Objective:
The objective of this course is to introduce students to the fundamental concepts and architectures of Artificial Neural Networks (ANN). It aims to develop a strong understanding of how neural networks learn from data, including topics such as perceptrons, multilayer networks, backpropagation, and activation functions. The course prepares students to apply ANN techniques to solve real-world problems in classification, regression, and pattern recognition, forming a basis for advanced study in deep learning and intelligent systems.

Course Learning Outcomes:
After completion of course, students would be able to:
CLO1: Define and describe the concepts of artificial neural networks.
CLO2: Use the different types of neural networks.
CLO3: Breakdown a practical problem of neural network learning and exploitation.
CLO4: Propose a solution to the given problem with different ANN techniques.

Units/Hours	Contents	Mapping with Course Learning Outcome
I **12 Hours**	Artificial Intelligence, Introduction to Artificial Neural Networks, Model of ANN, Building Blocks, Network Topology, Activation Functions and Types, Neural Network Learning Rules, Adaptation in ANN, Types of Learning in ANN.	**CLO1**
II **13 Hours**	Supervised Learning: McCullah Pitts Neuron, Perceptron, Adaline and Madaline Networks, Backpropagation Neural Network, Generalized Delta Learning Rule, Learning Vector Quantization, Adaptive Resonance Theory. Feedforward neural networks –-Linear responsibility X-OR problem and solution. -Analysis of pattern mapping networks summary of basic gradient search methods. Feed back neural networks Pattern storage networks, stochastic networks and simulated annealing	**CLO2**
III **12 Hours**	Kohonen Self Organizing Feature Maps, Associative Memory Network: Auto Associative Memory, Hetero associative Memory, Hopfield Networks: Discrete and continuous Hopfield networks, Boltzmann machine, Brain-State-In-Box Network.	**CLO3**
IV **13 Hours**	Deep Neural Networking: Introduction to learning process of deep neural networks, Challenges in training the deep learning models, Tuning Deep Learning Models and Trends in Deep Learning	**CLO4**

Transactional Modes:

- Lecture cum Demonstration
- Peer Learning/Teaching
- E-tutorial
- Self-Learning

Text Book:

1. Yegnanarayana, B. Artificial Neural Networks, Prentice-Hall of India. *New Delhi* (2006)

Suggested Readings:

1. Norvig, P. and Russel S. Artificial Intelligence. *A Modern Approach*. Upper Saddle River, NJ, USA:Prentice Hall, 2002
2. Negnevitsky, Michael. *Artificial Intelligence*. Pearson Education India, 2005
3. Freeman, James A., and David M. Skapura. *Neural networks: algorithms, applications, and programming techniques*. Addison Wesley Longman Publishing Co., Inc., 1991

Course Code: BAI-703
Course Title: Omics in Health Informatics
Total Hours: 45

L	T	P	Cr
3	0	0	3

Course Objective:
The objective of this course is to provide students with a foundational understanding of omics technologies and their applications in health informatics. It focuses on the integration and analysis of genomics, proteomics, and metabolomics data to advance personalized medicine, disease prediction, and healthcare decision-making.

Course Learning Outcomes:
After completion of course, students would be able to:
CLO1: Learn genomic sequences and knowledge about online databases of genomic sequences
CLO2: Observe the gene expression and functional interpretation of array data
CLO3: Explain the variations in genes and human diseases, phenology, comparative genomics
CLO4: Apply the alignment tools, NAR databases tools and software

Is going to be develop by Human Genetics Department

Transactional Modes:
- Lecture cum Demonstration
- Peer Learning/Teaching
- E-tutorial
- Self-Learning

Suggested Readings:

Course Code: BAI-704
Course Title: Python Programming for ANN -Lab
Total Hours: 60

L	T	P	Cr
0	0	4	2

Course Objective:
The objective of this lab course is to provide hands-on experience in using Python for implementing Artificial Neural Networks. It focuses on developing practical skills in coding neural network models, using libraries like NumPy, TensorFlow, and Keras, and applying them to solve real-world problems in classification, prediction, and pattern recognition.

Course Learning Outcomes:
After completion of course, students would be able to:
CLO1: Implementation of basic operations in python.
CLO2: Implementation of lists, operations, tuples and user defined functions along with file handling in python
CLO3: Deploy the various Supervised and Un-supervised models of ANN

List of Practical (but not limited to:)

1. Variables and Variables type, Data types, Data Types Conversion, Operators
2. Python Decision making (if, el if, else, nested if),
3. Python loops (while, for, nested loops), Break and continue statements.
4. Sequence introduction, Number operations, String Operations
5. Implementation of List and operations.
6. Implementation of Tuples and its operation
7. User defined functions, Functions with parameters, Keywords and optional parameters
8. Modules, Standard Modules (Sys, Math, Time)
9. Python File handling: Read Write and Update operations
10. Implementation of Supervised ANN
11. Implementation of Un-Supervised ANN

Suggested Readings

1. Paul Gries, Jennifer Campbell, Jason Montojo, Practical Programming- An Introduction to Computer Science Using Python 3.6, Shroff Publications and Distributors, 2018
2. Guttag, John V. *Introduction to computation and programming using Python*. MIT Press, 2013.

Course Code: BAI-741
Course Title: Seminar
Total Hours: 60

L	T	P	Cr
0	0	4	2

Course Objective:
The objective of this course is to enhance student's research, analytical, and communication skills through the exploration and presentation of recent developments in Data Science and Bio-Informatics. Students will learn to critically review current research, understand the working of advanced tools, and effectively present their insights in both oral and written formats.

Course Learning Outcomes:
After completion of course, students would be able to:
CLO1: Learn to present the latest techniques in Data Sciences and Bio-Informatics
CLO2: Working of latest tools for Data Sciences and Bio-Informatics
CLO3: Review of research in the field of Data Sciences and Bio-Informatics

Content

Student should give the seminar of half an Hour along with Seminar Report.

Course Code: BAI-798
Course Title: Project
Total Hours: 120

L	T	P	Cr
0	0	16	8

Course Objective:
To enable students to identify real-world problems in Life Sciences, Bio-Informatics, or Computer Science and develop practical solutions through systematic design, implementation, and analysis, culminating in a project report and viva presentation.

Course Learning Outcomes:
After completion of course, students would be able to:
CLO1: Prepare the problem in the field of Life Science/Bio-informatics/Computer Science
CLO2: Designing of the solution of problem
CLO3: Implementation and analysis of solution designed for the problem

Content

Student should prepare the solution of the designed problem and give the Viva at the end of semester.

SEMESTER – II

Course Code: BAI-751
Course Title: Machine Learning for Data Sciences
Total Hours: 60

L	T	P	Cr
4	0	0	4

Course Objective:
The objective of this lab course is to equip students with practical skills in Python programming for building and training Artificial Neural Networks. It emphasizes hands-on experience with libraries such as NumPy, TensorFlow, and Keras to design, implement, and evaluate neural network models for real-world data science and AI applications.

Course Learning Outcomes:
After completion of course, students would be able to:
CLO1: Apply basic concepts of machine learning such as types of learning, bias, evaluation cross validation
CLO2: Implement the linear regression, Bayesian learning, Naïve baise classifier over the data
CLO3: Use classification techniques such as k nearest neighbour, support vector machines, linear SVM formulation, Non-linear SVM, neural networks and recurrent networks

Units/Hours	Contents	Mapping with Course Learning Outcome
I **15 Hours**	Introduction to Machine Learning: Basic definitions, types of learning, hypothesis space and inductive bias, evaluation, cross-validation. Description of Data: Usage of Python Libraries for Machine learning like Pandas, Numpy, Matploytlib and Scipy	**CLO1**
II **15 Hours**	Linear Regression: Introduction, Definition, (Linear functions and other functions), Various Types of regression models, multiple linear regression, Assumption for errors, The least square regression line Bayesian Learning: Bayes theorem, Bayes Optimal Classifier, Naïve Bayes classifier, Gibbs algorithm, Bayesian belief networks.	**CLO2**
III **15 Hours**	Classification Techniques: K-Nearest Neighbors, Support Vector Machines, Linear SVM formulation, Nonlinear SVM, Feature Space Kernel Function, KNN modeling, SVM modeling, Decision Trees Neural networks: Perceptron, Multilayer Neural Network, Backpropagation Algorithm, Recurrent Networks.	**CLO3**
IV **15 Hours**	Unsupervised Learning: Hierarchical Clustering, K-means Clustering, Expectation Maximization (EM) Algorithm. Combining Multiple Classifiers: Voting, Bagging, Boosting, AdaBoost	**CLO4**

Transactional Modes:

- Lecture cum Demonstration
- Peer Learning/Teaching
- E-tutorial
- Self-Learning

Text Book

1. Cory Lesmeister, Mastering Machine Learning with R, Packt Publishing, 3rd Edition, January 2019
2. Abhijit Ghatak, Machine Learning with R, Springer Nature Singapore Pte Ltd. 2017

Suggested Readings:

1. Kevin Murphy, Machine Learning: Probabilistic Perspective, MIT Press. August 2012
2. Sebastian Raschka and Vahid Mirjalili, Python Machine Learning, Packt Publishing, Sept 2017
3. Shai Shalev-Shwartz, Shai Ben-David, Understanding Machine Learning: From Theory to Algorithms, Cambridge University Press, 2014
4. Stephen Marsland, Machine Learning: An Algorithmic Perspective, CRC Press, 2nd Edition, 2015
5. Machine Learning Online Course: http://nptel.ac.in/courses/106105152/

Course Code: BAI-752
Course Title: Neuron Networking
Total Hours: 45

L	T	P	Cr
3	0	0	3

Course Objective:
The objective of this course is to provide a comprehensive understanding of the structure, function, and communication mechanisms of biological and artificial neurons. It explores neural signaling, synaptic transmission, and network architectures to help students understand how neurons interact within networks, forming the basis for neural computation and brain-inspired artificial intelligence systems.

Course Learning Outcomes:
After completion of course, students would be able to:
CLO1: To illustrate the identification of the Big Data problem.
CLO2: To differentiate structured data from unstructured data.

Content is prepared by Department of Pharmacology

Transactional Modes:
- Lecture cum Demonstration
- Peer Learning/Teaching
- E-tutorial
- Self-Learning

Course Code: BAI-760
Course Title: Machine Learning for Data Science Lab
Total Hours: 60

L	T	P	Cr
0	0	4	2

Course Objective:
The objective of this lab course is to provide hands-on experience in applying machine learning algorithms to real-world datasets using programming tools such as Python and libraries like scikit-learn, pandas, and matplotlib. It focuses on practical implementation, model evaluation, and performance tuning to develop data-driven solutions in the context of data science applications.

Course Learning Outcomes:
After completion of course, students would be able to:
CLO1: Learn to clean of structured data using various operations
CLO2: Implementation of various operations for handling missing data and modeling of data
CLO3: Working with semi structured and unstructured data for prediction model building and analysis

List of Practical (But not limited to:)
1. Descriptive analysis and operations over the structured data.
2. Operations for Data Cleaning over structured data
3. Operations for Handling Missing Data for structural data
4. Modelling of Big Data for inferential information
5. Handling Semi-Structured data
6. Conversion of Semi-structured data to Structured data with No-Sql Techniques
7. Operations over Semi-structured data with MongoDB.
8. Handling Un-structured Data
9. Conversion of Un-structured data into Structured format
10. Operation over unstructured data with MongoDB

Suggested Reading:

1. Lab Manual

Course Code: BAI-742
Course Title: Seminar
Total Hours: 60

L	T	P	Cr
0	0	4	2

Course Objective:
The objective of this course is to enhance student's research, analytical, and communication skills through the exploration and presentation of recent developments in Data Science and Bio-Informatics. Students will learn to critically review current research, understand the working of advanced tools, and effectively present their insights in both oral and written formats.

Course Learning Outcomes:
After completion of course, students would be able to:
CLO1: Learn to present the latest techniques in Data Sciences and Bio-Informatics
CLO2: Working of latest tools for Data Sciences and Bio-Informatics
CLO3: Review of research in the field of Data Sciences and Bio-Informatics

Content

Student should give the seminar of half an Hour along with Seminar Report.

Course Code: BAI-799
Course Title: Project
Total Hours: 120

L	T	P	Cr
0	0	12	8

Course Objective:
To enable students to identify real-world problems in Life Sciences, Bio-Informatics, or Computer Science and develop practical solutions through systematic design, implementation, and analysis, culminating in a project report and viva presentation.

Course Learning Outcomes:
After completion of course, students would be able to:
CLO1: Prepare the problem in the field of Life Science/Bio-informatics/Computer Science
CLO2: Designing of the solution of problem
CLO3: Implementation and analysis of solution designed for the problem

Content

Student should prepare the solution of the designed problem and give the Viva at the end of semester.