AMITY UNIVERSITY
UTTAR PRADESH

FORMAT FOR COURSE CURRICULUM

Course Title: Machine learning and data mining
Course Code: BIOF301 **Units: 03**

L	T	P/ S	SW/F W	TOTAL CREDIT UNITS
3	0	0	0	3

Course Objectives:

Theory: To understand machine learning and data mining utility in bioinformatics.

Pre-requisites: Basic knowledge about bioinformatics & data structure.

Student Learning Outcomes:

- The student will be able to do data mining from large source of data.
- The student will acquire data warehouse basics, its life cycle and implementation of same.
- The student will acquire skills to classify and cluster data, also can do outlier analysis.

Course Contents/Syllabus- Theory:

	Weightage (%)
Module I	**20**
Descriptors/Topics: Introduction Need for data warehouse, definition, goals of data warehouse, Data Mart, Data warehouse, architecture, extract and load process, clean and transform data. Star, snowflake and galaxy schemas for multidimensional databases, fact and dimension data, Designing fact tables, Partitioning. Statistical measures in large Databases. Measuring Central Tendency, Measuring Dispersion of Data, Biological data warehouses and their life cycle.	
Module II	**20**
Descriptors/Topics: Data Warehouse Data warehouse and OLAP technology, multidimensional data models and different OLAP operations, OLAP Server: ROLAP, MOLAP and HOLAP. Data warehouse implementation, efficient computation of data cubes, processing of	

OLAP queries, indexing OLAP data, Machine learning, methods, Supervised and unsupervised learning applications of machine learning. Pattern recognition and pattern discovery.	
Module III	**30**
Descriptors/Topics: Data Mining Data mining primitives, Need of data mining, Types of Data Mining, sources of data mining Data Mining query language, Architectures of data mining. Data preprocessing, data integration and transformation, data reduction, Discretization and concept Hierarchy Generation, Data mining of biological data and report generations.	
Module IV	**30**
Descriptors/Topics : Classification Algorithms Classification and Prediction: Classification by decision tree induction, Back propagation, Bayesian classification, and classification based in association rules, Prediction, classifier accuracy, Cluster analysis, partitioning and hierarchical methods, Density based methods, and Grid based methods, Outlier Analysis, web mining, Temporal and spatial data mining. Classification and prediction of biological data through data mining methods and tools.	

Pedagogy for Course Delivery:

Lectures: 42
Tutorial:
Presentation/ Seminar: 2
Class Test: 1
Total: 45

Assessment/ Examination Scheme:

Theory L/T (%)	Lab/Practical/Studio (%)	Total
100%	NA	100%

Theory Assessment (L&T):

End Term Examination					
Components (Drop down)	**Class Test 1**	**Home Assignment**	**Presentation/ Seminar**	**Attendance**	

Weightage (%)	15	5	5	5	70

Lab/ Practical/ Studio Assessment:

	Continuous Assessment/Internal Assessment				**End Term Examination**			
Components (Drop down	**Class Test (Practical based)**	**Lab record**	**viva**	**Attendance**	**Lab record**	**Performance**	**Viva**	**Total**
Weightage (%)	0	0	0	0	0	0	0	0

Text:

- **Data Mining, Southeast Asia Edition: Concepts and Techniques by Jiawei Han, Micheline Kamber 2006 by Elsevier publication. ISBN: 978-1558609013.**
- **I.H. Whiffen: Data Mining, Practical Machine learning tools & techniques with Java (Morgan Kanffmen) ISBN: 978-1558605527.**

References

- **Pieter Adriaans Dolf Zantinge: Data Mining, Addition Wesley.**
- **David Hand, Heikki Mannila, and Padhraic Smyth: Principles of Data mining, PHI, Publication.**

Any other Study Material:

- **Online soft ware's and tools for data mining,**