Year: M. Tech. I (Semester – II)

Subject Name: Privacy Preserving Data Publishing and Data Mining **Subject Code:** MTCO14206
Type of course: Professional Elective – IV
Prerequisite (if any): Exposure to fundamentals of Data Mining

List of Courses where this course will be prerequisite: --

Rationale: Data mining and data publishing have been used for inferring vital information from the data stored in a database. In recent years, however, there is an increase in the need to preserve the privacy of the involved parties. In this course we study the numerous techniques to ensure privacy preservation in data mining and data publishing for various applications and data mining algorithms.

Teaching and Examination Scheme:

Teaching Scheme				Theory Marks			Practical Marks		Total
L	T	P	C	TEE	CA1	CA2	TEP	CA3	
3	0	2	4	60	25	15	30	20	150

CA1: Continuous Assessment (assignments/projects/open book tests/closed book tests. CA2: Sincerity in attending classes/class tests/ timely submissions of assignments/self-learning attitude/solving advanced problems TEE: Term End Examination TEP: Term End Practical Exam (Performance and viva on practical skills learned in course) CA3: Regular submission of Lab work/Quality of work submitted/Active participation in lab sessions/viva on practical skills learned in course

Content:

Sr. No.	Content	Total Hrs
1	Security, Privacy and Data Mining; An Introduction to Privacy-Preserving Data Publishing; An Introduction to Privacy-Preserving Data Mining; Need for Privacy Preservation; Applications of Privacy Preservation; Laws for Privacy Preservation	4
2	Privacy Preserving Data Publishing - Overview and Applications; Attack Models and Privacy Models; Anonymization Operations; Information Metrics; Anonymization Algorithms; Multiple Release Publishing; Collaborative Data Publishing	9

3	Privacy Preserving Data Mining - Introduction; Applications of Privacy-Preserving Data Mining - Medical Databases, Bioterrorism Applications, Homeland Security Applications, Genomic Privacy; The Randomization Method; Group Based Anonymization; Quantification of Privacy Preserving Data Mining Algorithms	8
4	Distributed Privacy-Preserving Data Mining - Introduction; Basic Cryptographic Techniques; Privacy definitions; Homomorphic Encryption; Secure Sub-protocols; Data partitioning models; Adversary Models	8
5	Privacy-Preserving Methods Across Horizontally Partitioned Data - Introduction; Decision Tree Mininig; Association Rule Mining; k-means clustering; Extension to Malicious Parties; Limitations; Privacy Issues Related to Data Mining Results	6
6	Privacy-Preserving Methods Across Vertically Partitioned Data - Introduction; Decision Tree Mininig; Association Rule Mining; k-means clustering; Challenges; Comparison to Horizontally Partitioned Data Model	6
7	Advanced Topics - differential privacy, social network data publishing - introduction, attacks, privacy models	4

Reference Books:

Sr.No.	Title of book /article	Author(s)	Publisher and details like ISBN	Publication Year	Publication Edition
1	Privacy-Preserving Data Mining Models and Algorithms-	Editors: Aggarwal, Charu C., Yu, Philip S.	Springer, 978-0-387-70992-5	2008	1
2	Privacy-preserving Data Mining	Jaideep Vaidya, Christopher W. Clifton, Yu Michael Zhu	Springer, 9780387294896	2006	1
3	Privacy-Preserving Data Publishing	Bee-Chung Chen, Daniel Kifer, Ashwin Machanavajjhala, Kristen LeFevre	Now Publishers,9781601982766	2009	1

4	Introduction to Privacy-Preserving Data Publishing: Concepts and Techniques	Benjamin C.M. Fung, Ke Wang, Ada Wai-Chee Fu, Philip S. Yu	CRC Press, Taylor & Francis Group	2010	1

Course Outcomes:

Sr.No.	CO statement	Marks % weightage
1	Examine the fundamental cryptographic concepts for privacy preserving data mining	15%
2	Analyze various attack models and privacy for privacy-preserving data publishing.	20%
3	Integrate privacy-preservation in data publishing and data mining with the help of various techniques for different applications.	25%
4	Construct privacy-preserving models for horizontally and vertically partitioned data.	25%
5	Integrate the privacy-preserving concepts for social network data.	15%

List of Open learning website:

1. Privacy-Preserving Data Publishing: A Survey of Recent Developments by Fung, B. C., Wang, K., Chen, R., & Yu, P. S (https://dl.acm.org/doi/10.1145/1749603.1749605)

List of Open Source Software:

1. Weka - https://www.cs.waikato.ac.nz
2. Simjava - http://www.icsa.inf.ed.ac.uk/research/groups/hase/simjava/
3. Python libraries for Cryptography and Privacy

List of Experiments:

Sr. No.	Practical
1	Implement the Datafly algorithm for privacy-preserving data anonymization.
2	Implement the Mondrian algorithm for privacy-preserving data anonymization.

3	Implement the multiplicative Elgamal homomorphic encryption scheme.
4	Implement the additive Benaloh and Paillier homomorphic encryption scheme.
5	Implement Distributed Privacy preserving Frequent Itemset mining using a homomorphic scheme on the transaction data from http://fimi.uantwerpen.be/data/ by creating a horizontally partitioned setup.
6	Implement Distributed Privacy preserving Frequent Itemset mining by creating a vertically partitioned setup.
7	Implement Distributed Privacy preserving Clustering using a homomorphic scheme on the dataset of your choice by creating a horizontally partitioned setup.
8	Implement Distributed Privacy preserving Classification using a homomorphic scheme on the dataset of your choice by creating a horizontally partitioned setup.