1.	Title of the course	Algorithms for Big Data
2.	Course number	CS524L
3.	Structure of credits	3-0-0-3
4.	Offered to	PG
5.	New course/modification to	Modification To CS5229/14
6.	To be offered by	Department of Computer Science and Engineering
7.	To take effect from	July 2022
8.	Prerequisite	CoT
9.	**Course Objective(s):** To introduce scalable, probabilistic algorithms with sub-linear space complexities for processing massive, heterogeneous, streaming data.	
10.	**Course Content:** Review of basic probability; Tail bounds: Markov, Chebyshev, Chernoff, Bernstein and Hoeffding; Streaming models; Streaming algorithms: Morris algorithm, Flajolet-Martin algorithm, reservoir sampling; Hashing: universal hash functions, min-hashing, locality sensitive hashing, Bloom filter, Jaccard similarity estimation, load balancing, power of two choices, consistent hashing; Sketching: Alon-Matias-Szegedy (AMS) algorithm, Indyk's algorithm, count-min sketch, majority and heavy-hitters problem; Dimensionality reduction: random projection, Johnson-Lindenstrauss lemma; Graph and geometric streaming algorithms; Randomized numerical linear algebra algorithms.	
11.	**Textbook(s):** 1. Akerkar R, *Models of Computation for Big Data*, 1st Edition, Springer (2018). 2. Blum A, , Hopcroft J and Kannan R,, *Foundations of Data Science*, 1st Edition, Cambridge University Press (2020).	
12.	**Reference(s):** 1. Mitzenmacher M and Upfal E, *Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data analysis*, 2nd Edition, Cambridge University Press (2017). 2. Muthukrishnan S, *Data Streams: Algorithms and Applications*, 1st Edition, Now Publishers Inc (2005).	