1.	Title of the course	Data Science and Engineering
2.	Course number	CS516L
3.	Structure of credits	3-0-0-3
4.	Offered to	PG
5.	New course/modification to	Modification To CS5109/12
6.	To be offered by	Department of Computer Science and Engineering
7.	To take effect from	July 2022
8.	Prerequisite	Nil
9.	**Course Objective(s):** To impart knowledge on large scale systems for data life cycle. To impart knowledge on best practices and processes for building data science applications.	
10.	**Course Content:** Review of concepts in probability; Data life cycle using CRoss Industry Standard Process for Data Mining (CRISP-DM); Data pre-processing: data wrangling, cleaning, handling structured and unstructured data including missing values, noise in data, normalization, multimedia, feature engineering and dimensionality handling; Best practices in model design and development; Explainability: balancing, visualization, story-telling and hyperparameter tuning, automatic model learning, ensembling, concept drift and model maintenance; Advanced topics: methodologies of active, meta and transfer learning paradigms and stream data mining; Introduction to big data including sharding, deduplication and map-reduce.	
11.	**Textbook(s):** 1. Chakrabarti S, Cox E, Frank E et al., *Data Mining Know It All*, 1st Edition, Elsevier (2009). 2. GarcÃa S, Luengo J and Herrera F, *Data Preprocessing in Data Mining*, 1st Edition, Springer (2014).	
12.	**Reference(s):** 1. Bahga A and Madisetti V, *Big Data Analytics: A Hands-On Approach*, 1st Edition, VPT (2018). 2. Hair J, Black W, Babin B and Anderson R, *Multivariate Data Analysis*, 1st Edition, Pearson (2014). 3. Patterson J, *Deep Learning: A Practitioner's Approach*, 1st Edition, O' Reilly (2007). 4. Sharda R, Delen D and Turban E, *Business Intelligence: A Managerial Perspective on Analytics*, 1st Edition, Pearson (2017).	