

1.	Title of the course	Data Science and Engineering
2.	Course number	CS5109
3.	Status of the course	Core
4.	Structure of credits	3-0-0-3
5.	Offered to	PG
6.	New course/modification to	New course
7.	To be offered by	Department of Computer Science and Engineering
8.	To take effect from	July 2020
9.	Prerequisite	Nil
10.	Whether approved by the Department	Yes
11.	<b>Course Objective(s):</b> To impart knowledge on large scale systems for data life cycle. To impart knowledge on best practices and processes for building data science applications.	
12.	<b>Course Content:</b> 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.	
13.	<b>Textbook(s):</b> 1. Chakrabarti S, Cox E, Frank E et al., <i>Data Mining Know It All</i> , 1st Edition, Elsevier (2009). 2. García S, Luengo J and Herrera F, <i>Data Preprocessing in Data Mining</i> , 1st Edition, Springer (2014).	
14.	<b>Reference(s):</b> 1. Bahga A and Madiseti V, <i>Big Data Analytics: A Hands-On Approach</i> , 1st Edition, VPT (2018). 2. Hair J, Black W, Babin B and Anderson R, <i>Multivariate Data Analysis</i> , 1st Edition, Pearson (2014). 3. Patterson J, <i>Deep Learning: A Practitioner's Approach</i> , 1st Edition, O' Reilly (2007). 4. Sharda R, Delen D and Turban E, <i>Business Intelligence: A Managerial Perspective on Analytics</i> , 1st Edition, Pearson (2017).	