Appendix 1: Proposed Course Syllabi

Proposed Common Core Syllabus for Data Analysis core course:

Total Credits	(L,T, P, C): 3, 0, 2, 8
Type	Theory
Description	1. *Motivation for data analysis and interpretation*. Examples from science, engineering, management and social sciences. Visualization of data. Different plots and graphs. Elementary statistical concepts related to data analysis, including Measures of Location and Measures of Variability. 2. *Introduction to Probability Theory*. Sample Spaces and events Probability axioms. Properties of Probability, Counting Techniques. Random Variables. Expectations and Variances. Conditional probabilities and conditional expectation. Independence. Important discrete and continuous distributions including the ones derived from the Normal (t, chi-squared and F). Random vectors. Joint Probability distributions. Bivariate and Multivariate Normal Distributions with the corresponding mean vectors, variance-covariance matrices and correlation matrices. Multinomial distribution. Distribution of the sample mean. 3. *Estimation and Testing*. Point and interval estimates of the unknown parameters. Unbiased estimators. Methods of point Estimation including maximum likelihood estimation. Confidence levels and confidence intervals. Hypothesis testing. Type 1 and Type 2 errors. Testing for parameters of a normal distribution and for percentages based on a single sample and based on two samples. Introduction to the chi-squared test. The concept of p-value. Mean-square estimation and Kalman filtering. 4. *Regression and ANOVA*. Simple linear regression. The least squares error criterion. Relation to maximum likelihood. The R^2 measure for fit of a regression model. Correlation coefficient. Regression models for prediction. Introduction to one way and two way ANOVA. 5. Overview of classification from a usage viewpoint a. Training, validation, and testing process in model building. 6. Dimensionality reduction and clustering 7. Lab exercises (using tools such as R/Matlab/Scilab) related to the above topics with emphasis on visualization, sampling, and basic ML techniques such as classification, and regression.

References	1. S. M. Ross (2014) *Introduction to Probability and Statistics for Engineers and Scientists*, 5th Edition, Academic Press, 2. J. L. Devore (2015) *Probability and Statistics for Engineers*, 9th Edition, Brooks Cole. 3. D. Freedman, Robert Pisani, and Roger Purves (2007) *Statistics, 4th International Student Edition*, W.W. Norton & Company. 4. R.A. Irizarry (2019) *Introduction to Data Science, Data Analysis and Prediction Algorithms with R*, Chapman & Hall. 5. A.B. Downey (2014) *Think Stats* (*Exploratory Data Analysis in Python*), Green Tea Press.

Proposed syllabus for Minor Course: DS 2XX: Programming for Data Science

Total Credits	(L,T, P, C): 0, 2, 2, 6
Type	Lab
Description	Programming Basics (Python programming, R, Data Structures), Visualization/Plotting, Data Science Libraries (Pandas, PyPlot, matplotlib) Databases, GPUs/CUDA programming, Parallel/distributed computing for data science (Map/Reduce, Spark/Hadoop), working on the cloud (Amazon Web services, Google Cloud Platform, Azure, etc). The course will be programming heavy, with in-class and take-home programming exercises A project can be optionally included.
References	1. Principles and Techniques of Data Science, By Sam Lau, Joey Gonzalez, and Deb Nolan, 2019, available online at https://www.textbook.ds100.org/intro 2. Online tutorials on Python and R 3. Learning Python, Mark Lutz, O Reilly, 2005 4. Python for data analysis, Wes Mckinney, O Reilly, 2013 5. CUDA by Example: An Introduction to General-Purpose GPU Programming, Jason Sanders, Nvidia, 2010 6. NORMAN MATLOFF. Parallel Computing for Data Science: With Examples in R, C++, and CUDA. Boca Raton: CRC Press.

Proposed syllabus for Minor Course: DS 3XX: Introduction to Machine Learning

Total Credits	(L,T, P, C): 3, 0, 0, 6

Type	
Description	Introduction to machine learning: What is learning, learning objectives, data needed. Supervised Learning. Bayesian inference and learning: Inference, naïve Bayes. Measures of success and loss functions. Generalization and model complexity, bias-variance tradeoff. Training, validation, and testing. Introduction to convex optimization. Convergence and training time. Objective functions for classification, regression, and ranking. Linear regression, Perceptron and logistic regression. MLP and backpropagation. Deep learning, CNN and RNN. SVM, support vector regression, increase in dimensionality through simple kernels. Decision trees. Role of randomization and model combination, bagging and boosting. Unsupervised Learning. Clustering criteria, K-means, DB-scan. Kernel Density estimation. EM-algorithm for mixture of Gaussians. Dimensionality reduction using PCA and Kernel-PCA. Other Topics: Overview of Reinforcement Learning. Bias and Ethics in ML. Optional topics: Introduction to one or more of the following topics: Active/Transfer Learning, Bootstrapping, Semi-supervised learning, Generative and probabilistic graphical models, Online/incremental learning.
References	Pattern Recognition and Machine Learning, by Christopher Bishop, Springer 2011 The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, by Trevor Hastie and Robert Tibshirani (Springer Series in Statistics) 2016 Supplementary material available online, e.g. Dive into Deep Learning by Aston Zhang, Zack C. Lipton, Mu Li and Alexander Smola, 2020 (https://d2l.ai)

Appendix 2: Elective Details

There are a large number of existing courses in IITB that are relevant to the AI/ML and Data Science areas. To better understand their relationship, we list these courses below organized into 3 baskets: Mathematics basket, AI/ML basket, and Applications basket. We also provide a further subdivision into basket-specific sub-areas.

Currently these courses predominantly run as department electives (or in a few cases, as department core courses). Course structure is assumed to be 3-0-0-6 unless otherwise specified. Prerequisites are also listed where required, but instructor/ASC is the final word on prerequisite.

In addition to the existing courses listed below, further course(s) will be created specifically targeted at the AI/ML and Data Science application areas.

Mathematics Basket

Probability and Statistics	**Linear algebra, Vector space etc.**	**Optimization**	**Related subjects**
IE502 Probabilistic Models	SC 639 Mathematical structures for Control (linear algebra, vector spaces, convexity)	IE501 Optimization Models	IE692 Applied Time Series Analysis Prereq: IE501, IE601
SC 629 Introduction to Probability and Random Processes	CS 740 Mathematics for Visual Computing	CS 709 Convex Optimization	SI 402 Statistical inference 3-1-0-8
IE605 Engineering Statistics 3-1-0-8	SI 423 Linear algebra and applications 3-1-0-8	IE601 Optimization Techniques	SI 422 Regression analysis 3-0-2-8 Prereq: SI 417 Introduction to Probability Theory
SI 417 Introduction to Probability Theory 3-1-0-8		SC 607 Optimization	SI 509 Time series analysis 3-1-0-8 Prereq: SI 402 Statistical Inference

SI 505 Multivariate Analysis 3-1-0-8		EE 736 Stoch Opt	IE614 Linear Systems 3-1-0-8
IE611 Introduction to Stochastic Models 3-1-0-8		SI 416 Optimization 2-1-0-6	SC 640 Applied Predictive Analytics
SI 514 Statistical modelling 2-1-0-6 Prereq: SI 402 Statistical Inference		EE 659 A first course on optimization	IE616 Decision Analysis and Game Theory 3-0-0-6
		CL 603 Optimization	SC 631 Games and Information
			EE 601 Statistical Signal Processing
			EE 737 Introduction to Stochastic Control
			SC 612 Introduction to Linear Filtering and Beyond
			SI 509 Categorical data analysis 3-1-0-8
			CL 672 Applied Multivariate Statistics in Chemical Engineering Prereq: IC 102 or CL 602

AI/ML basket

Basic AI/ML	**Advanced AI/ML**	**Related to AI/ML**
CS 747 Foundations of Intelligent and Learning	IE712 Topics in Machine Learning (requires IE611)	IE702 Neural Nets, Fuzzy Systems and Applications

Agents		Prereq: IE611
	CS 729 Topics in machine learning	EE 763 Science of Information Statistics and Learning
	IE643 Deep Learning Theory and Practice	SC 403 System Modelling and Simulation
	GNR 638 Machine Learning for Remote Sensing II	IE708 Markov Decision Processes Prereq: IE611
	EE 782 Adv ML	IE704 Selected topics in A.I. for Operations Research Prereq: IE611
	CS 726 Advanced machine learning	IE613 Online machine learning Prereq: IE611
	CS 748 Advances in Intelligent and Learning Agents	SC 643 Stochastic and Networked Control.

Application basket

Visual Comp	**NLP/Speech/IR**	**Database and Mining**	**Robotics & Mechatronics**	**Bio and Chemical Engg**
EE 610 Image processing	CS 460 Natural Language Processing	CS 632 Advanced Database Management Systems	ME 402 Mechatronics II	SI 528 Biostatistics 2-1-0-6 Prereq: SI 402 Statistical Inference
EE 678 Wavelets	CS 635 Information Retrieval and Mining for Hypertext and the Web	CS 631 Implementation Techniques for Relational Database Systems	SC 634 Introduction to Mobile robotics	CL 688 Artificial Intelligence in Process Engineering
EE 779	CS 728	SI 515 Statistical	SC 635 Advance	CL 625 Process

Advanced Signal Processing	Organization of Web information	techniques in data mining 2-1-0-6 Prereq: SI 402 Statistical Inference	Topics in Mobile Robotics	Modeling and Identification
CS 736 Medical Image Computing	CS 753 Automatic Speech Recognition	SI 536 Analysis of multi-type and big data 2-1-0-6 Prereqs: SI 505 Multi-variate Analysis SI 515 Statistical Techniques in Data Mining		CL 686 Advanced Process Control
CS 749 Digital Geometry Processing	CS 730 Natural language processing			CL 712 Process Data Analysis & Design of Experiments
CS 754 Advanced Image Processing	CS 626 Speech, Natural Language Processing and the Web			
CS 763 Computer Vision	EE 679 Speech Processing			
CS 775 Advanced Computer Graphics				