Semester:	4				
Course Code:	ME2050				
Course Name:	Introduction to Statistical Learning				
Credits Value:	3 (Notional hours: 150)				
Pre-requisites:	EM2020				
Core/ Optional:	Core				
Hourly Breakdown	Lectures (hours)	Tutorials (hours)	Practical classes (hours)	Assignments (hours)	Independent Learning & Assessment (hours)
	30			30	90

Course Aim:

The aim of this course is to introduce the essential elements of data science. Specifically, by providing an introduction to the ideas and techniques of data collection and management, summarizing and visualizing of data, statistical inference, and machine learning.

Intended Learning Outcomes:

On successful completion of the course, the students should be able to;

- **manage**, summarize and visualize data using the Python programming language and Jupyter notebooks
- **use** Bayes theorem to assign probabilities to a given event based on observations
- **use** Bayesian inference methods for decision making and assess the associated risks
- **apply** machine learning methods for prediction and classification

Course Content:

- **Introduction and Data visualization using Python**: A holistic overview of the process of data analysis, Data Wrangling, data types, basic statistics, visualization tools
- **Randomness**: Assigning probabilities, random variables, distributions, conditional probability and Bayes theorem, empirical distributions, law of large numbers, the significance of the normal distribution and the central limit theorem, frequentist vs Bayesian approach
- **Bayesian Inference:** modeling, estimation, model assessment, A/B testing, the loss function and risk, choosing the prior, Bayesian multi arm bandit (MAB) problem.
- **Predication and Classification**: properties of multivariate Gaussians and their joint and conditional distributions, Gaussian Process (GP) regression, linear regression from GP, linear regression from a Bayesian approach, linear regression as least squares fit, logistic regression, nearest neighbor clustering, Gaussian mixture models (GMM) and GMM clustering

Teaching/ Learning Methods:

Classroom lectures, in-class exercises and assignments

Assessment Strategy:

Continuous Assessment 50%		**Final Assessment** 50%		
Details:		Theory (%)	Practical (%)	Other (%)
Assignments/Quizzes	30%	50%		
Mid semester examination	20%			

Recommended Reading:

- Abha Belorkar, Sharath Chandra Guntuku, Shubhangi Hora (2020), *Anshu Kumar Interactive Data Visualization with Python: Present your data as an effective and compelling story* (2nd Edition), Publisher Packt, ISBN 9781800200944.
- Trevor Hastie, Robert Tibshirani, Jerome Friedman (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd Edition), Springer Series in Statistics, DOI-10.1007/978-0-387-84858-7
- Cameron Davidson-Pilon (2015), *Bayesian Methods for Hackers: Probabilistic Programming and Bayesian Inference (Addison-Wesley Data & Analytics)* (1st Edition), Addison-Wesley Professional, ISBN-13: 978-0133902839.