

Proposal for New Course		
Course Number	:	MB510
Course Name	:	Probability and Statistics for Data Science and AI
Credits	:	2-0-0-2 (L-T-P-C) ¹
Prerequisites	:	None
Intended for	:	MBA
Distribution	:	Compulsory
Semester	:	Odd

Preamble

Increasingly more acceptability of data science and artificial intelligence (DSAI) in industry urges for proper explanation of the working of complex DSAI tools and techniques. The explainability of AI models requires sound statistics and probability knowledge. Also, today's businesses are becoming more complex and complicated. The AI techniques as a solution to this complex and complicated business requires proper understanding of associated uncertainty and risk. Statistics and probability theories are used to deal with these uncertainties and risks. Growing availability of data and pressure on using these data for decision making is an industrial norm. So, it is paramount to use well developed theories of statistics and probability to make sense of data. This course is focusing on understanding the basic concepts of probability and statistics to gain a mastery over the discipline of data science and artificial intelligence.

Objective

On completion of this course, the student should be able to:

- understand intuitively the complex concepts of statistics and probability.
- interpret and apply these concepts in real-life business problem solving and decision making.
- appreciate the intimate link between business and mathematics in general.

¹ L= Lectures per week, T=Tutorials per week – P = Practical/Lab session per week – C = Credits for course

Course Modules with Quantitative lecture hours		
Module 1	Introduction	(4)
Uncertainty and probability, Probability and its types, Conditional, joint and marginal probability, Problem classification with emphasis on random problems, Probability and statistics concepts for DSAI.		
Module 2	Random Variables and Probability Distributions	(4)
Describing Randomness, Random Variables and Probability Distributions- Continuous and Discrete Distributions, Normal distribution, lognormal distribution and Power-Law Distributions, Bernoulli distributions.		
Module 3	Statistics	(4)
Collections of Random Values-Expected Value, variance and standard deviation, Independent and Identically Distributed Variables, law of large numbers, central limit theorem.		
Module 4	Sampling	(4)
Sampling and Replacement-Selection with Replacement, Selection Without Replacement, Bootstrapping.		
Module 5	Bivariate Statistics	(4)
Covariance and Correlation-Pearson and Spearman.		
Module 6	Baye's Rule	(4)
Frequentist vs. Bayesian Probability, Bays Rule and Confusion Matrices, Repeating Bayes' Rule, Multiple Hypothesis.		
Module 7	Information Theory	(4)
Entropy, Measuring uncertainty, information, and surprise, Maximal entropy distribution, Cross Entropy, KL Divergence		

Lab Exercises (If applicable):
Nil

Textbooks:	
1.	Nil
2.	
Reference Book:	
1.	Thomas Nield (2022), Essential Math for Data Science: Take Control of Your Data with Fundamental Linear Algebra, Probability, and Statistics, Shroff Publishers & Distributors Pvt. Ltd., Mumbai
2.	Trevor Hastie, Robert Tibshirani, Jerome Friedman (2017), The Elements of Statistical Learning: Data Mining, Inference, And Prediction, Springer (2 nd ed.)
3	Anirban Das Gupta (2011), Probability for Statistics and Machine Learning: Fundamentals and Advanced Topics, Springer
4	Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong (2020), Mathematics for Machine Learning, Cambridge University Press (https://mml-book.com)
5	Scott E. Page (2018), The Model Thinker: What You Need to Know to Make Data Work for You, Basic Books.
6	MS Excel 2019 (2020), Data Analysis & Business Modeling, Wayne Winston, Microsoft Press (PHI).