Semester	**FOUR**
Paper Number	**13**
Paper Code	**MDTS 4411**
Paper Title	**Bayesian Data Analysis and Data Governance & Compliance**
No. of Credits	**6**
Course description	CORE Composite Paper Module 1: 2 classes/week Module 2: 2classes/week No. of classes assigned Theory: 4 classes per week Practical: 3 classes per week
Course Objective	At the end of the course, a student is expected to understand the 1. At the end of the course, a student is expected to understand the 2. Bayesian estimation and credible regions. 3. Hierarchical and Empirical Bayesian analysis. 4. Complex posterior computation using Acceptance-rejection, importance sampling and simulation methods such as MCMC. 5. Bayesian Linear and logistic regression. 6. How privacy, security and transparency standards in using data are met. 7. How misuse of data may impact the society at large. 8. Have an insight into the consequences of violation of property rights.
Syllabus	**Module I** **Bayesian Data Analysis** **Unit1:** Subjective definition of probability, Conditional Probability, Marginal Probability, Bayes theorem, Applications of Bayes theorem in Spam filter, Bayesian search, etc. [3] **Unit2:** Prior and posterior distributions, Posterior estimates, credible intervals, highest posterior density regions; Hierarchical models, Hierarchical Bayes and Empirical Bayes. [7] **Unit3:** Acceptance-rejection sampling, importance sampling; Markov chain basics, Markov chain Monte Carlo (MCMC), Gibbs sampling, Metropolis-Hastings MCMC, MCMC diagnostics. [10] **Unit4:** Bayesian Linear regression; Bayesian logistic regression. [6]

	Module-II **Data Governance & Compliance** **Data Sources:** Different sources of data and their interrelations. Database, Data Lake, Data Mesh (6L) **Data Quality :** Different aspects of data quality measures and their statistical relevance (6L) **Data and Information**: Theory of Information. Data Governance, Ownership and consent. Use/Overuse/Misuse of Data. Equilibrium. The FAT Flow framework of Data Science Ethics. (5L) **Data Privacy:** Privacy and Confidentiality. Aspects and Challenges. PII information. Methods of maintaining data privacy. Privacy Models. Trade-off between protecting privacy and loss of information. (6L) **Data Privacy & Security:** Importance of data security and its solutions. Governance and compliance. Data Security Audits. (3L) In all the above topics, case studies can be discussed to elaborate on the ideas.	
List of Practical	Based on Bayesian Data Analysis.	
Reading/Reference Lists	1. An Introduction to Bayesian analysis: theory and Methods; Jayanta Kumar Ghosh, Mohan Delampady, Tapas Samanta. 2. Bayes and Empirical Bayes Methods for Data Analysis; Bradley P. Carlin, homas A Louis 3. Bayesian Data Analysis; Andrew Gelman, John B. Carlin, Hel S. Stern, David B. Dunson, Aki Vehtari and Donald B. Rubin. 4. Theory of Statistics; Mark J. Schervish 5. Monte Carlo Statistical Methods; Christian Robert and George Casella. 6. Markov Chain Monte Carlo in Practice; W. R. Gilks, S. Richardson, D. J. Speilgelhalter. 7. Handbook of Markov Chain Monte Carlo; Steve Brooks, Andrew Gelman, Galin L. Jones, Xiao-Li Meng. 8. Guide to Data Privacy Models, Technologies, Solutions; Vicenç Torra. 9. Data Science Ethics- Concepts, Techniques and Cautionary Tales: David Martens, Oxford University Press, 2022. 10. Data Ethics and Challenges: Shukla S. etal, Springer Series in Applied Sciences and Technologies, 2022. 11. Ethics of Data and Analytics- Concepts and Cases: Kristen Martin, CRC Press, 2022. 12. Data Governance and Data Management: Contextualizing Data Governance Drivers, Technologies, and Tools; Rupa Mahanti.	
Evaluation	Theory CIA: 10 End Sem Exam: 50 (30+20) Total : 60	Practical(Based on Bayesian Inference) Continuous Assessment: 30 End Sem Viva: 10 Total: 40

Paper Structure for End Semester Theory	Short questions: 5 marks each	Long questions: 10 marks each
Module I	2 out of 4	2 out of 3
Module II	2 out of 4	1 out of 2