Subject Description Form

Subject Code	DSAI3101
Subject Title	Bayesian Methods for Data Science
Credit Value	3
Level	3
Pre-requisite/ Co-requisite/ Exclusion	Pre-requisite: DSAI1103 Principles of Data Science
Objectives	The objective of this subject is to: (a) introduce the principle of Bayesian inference, how to select the prior and calculate the posterior. (b) introduce some effective current approaches to Bayesian modeling and computation in statistics. (c) introduce some traditional Bayesian models, Bayesian classification and the fundamental concepts of Bayesian networks. (d) introduce RStan/PyStan and integration with data science workflows (e.g., rmarkdown, Jupyter notebook). (e) demonstrate the flexibility and practicality of traditional Bayesian inference, and expand to Bayesian classification and Bayesian networks in real-world applications through industry-aligned examples (e.g., testing, predictive modeling) and case studies in domains like healthcare, finance, and tech.
Intended Learning Outcomes	Upon completion of the subject, students will be able to: (a) understand the foundations of Bayesian statistical; and how to select the prior and calculate the posterior; (b) presents effective current approaches to Bayesian modeling and computation in statistics; (c) understand the foundations of traditional Bayesian statistical models, Bayesian Classification, Bayesian Network; (d) Master software tools for scalable Bayesian analysis (e.g., RStan, PyStan) and apply Bayesian models to solve data science challenges such as uncertainty quantification, decision-making under noise, and high-dimensional inference.
Subject Synopsis/ Indicative Syllabus	**Preliminaries (2L+1T)** Random variables; point estimation: moment, maximum likelihood, least squares; interval estimation; hypothesis testing; regression models: linear, logistic

	Basics of Bayesian inference (6L+3T) Basic theory: prior, posterior, Bayes' rule; Prior select: conjugate priors, improper priors, objective priors; Model selection and diagnostics: cross validation, Bayes factors, stochastic search variable selection, model selection criteria and Goodness-of-fit checks; hierarchical modeling **Bayesian modeling and computation in statistics (6L+3T)** Posterior computational approaches: maximum a posteriori estimation (MAP), Bayesian central limit theorem (CLT), Markov chain Monte Carlo (MCMC), Metropolis-Hastings (MH) **Bayesian Classification (4L+2T)** Bayesian classifier: Bayesian classification rule, Bayesian classification error, average risk; decision surfaces: minimum distance classifiers; naive Bayes classifier **Bayesian Networks (4L+2T)** Graphical structure: nodes, arcs(links), dags; conditional probabilities; Markov property; Bayesian networks: probability propagation, evidence nodes, diagnostic reasoning, predictive reasoning; sparse Bayesian networks: chain rule of probability theory, Pearl's network construction algorithm, node ordering **Case studies (2L+1T)** Programming languages or packages: RStan/PyStan; other relevant languages or packages: JAGS, PyMC3, BUGS, RNimBle; Case studies
Teaching/Learning Methodology	The subject will mainly be delivered through lectures and tutorials. The lectures will be conducted to introduce the Bayesian theoretical background, and practical problems will be discussed in the tutorial sessions to illustrate how the theory developed can be applied in practice. Encourage students to use R/Python language with RStan/PyStan and various Bayesian statistical models to conduct exploratory analysis by using real data.

Assessment Methods in Alignment with Intended Learning Outcomes

Specific assessment methods/tasks	% weighting	Intended subject learning outcomes to be assessed			
		a	b	c	d
1. Assignments	30%	✓	✓	✓	✓
2. Project	10%	✓	✓	✓	
3. Examination	60%	✓	✓	✓	✓
Total	100 %				

Explanation of the appropriateness of the assessment methods in assessing the intended learning outcomes:

This subject focuses on the mathematical foundation of data science. Many of these topics are based on theory in statistics and machine learning. **Exam-based assessment** is an appropriate assessment method, including a 60% examination.

Since this subject also emphasizes on understanding the implementation of various Bayesian methods in data science, a mini-project that takes a weight of 10% is appropriate for assessing the intended learning outcomes (c) and (d), in which students will be encouraged to analyze datasets using various Bayesian models study in this class and communicate their findings. A 30% worth of assignments including a midterm test are a component of continuous assessment in order to keep students in progress.

Continuous Assessment comprises assignments, mini-project and test. A written examination is held at the end of the semester.

Student Study Effort Expected		
	Class contact:	
	▪ Lectures/Tutorials	39 Hrs.
	Other student study effort:	
	▪ Assignments/Projects	36 Hrs.
	▪ Self-study	30 Hrs.
	Total student study effort	105 Hrs.

Reading List and References

Textbooks:

Reich, Brian J., and Sujit K. Ghosh.	Bayesian statistical methods	Chapman and Hall, 1st edition

	Theodoridis, Sergios.	Machine learning: a Bayesian and optimization perspective	Academic press, 1st edition
	Korb, Kevin B., and Ann E. Nicholson.	Bayesian artificial intelligence	Chapman and Hall , 2nd edition
	Deskbooks:		
	Carlin, Bradley P., and Thomas A. Louis.	Bayesian methods for data analysis	Chapman and Hall, 3rd edition
	Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B.	Bayesian data analysis	Chapman and Hall, 3rd edition
	Kruschke, John.	Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan	Available online