Semester	**EIGHT**
Course	**Major**
Paper Code	**C4ST230821T**
Paper Title	**Econometrics & Machine Learning**
No. of Credits	**6**
Theory / Practical /Composite	**Theory**
Classes per week	**Module I: 2 periods/week** **Module II: 2 periods /week**
Modules	**2**

Course Outcomes for Module I

1. **Remember** the fundamental concepts of simple and multiple linear regression, model assumptions, regression diagnostics, model selection criteria, and sources of econometric problems.
2. **Understand** the assumptions underlying classical linear regression models, the causes and consequences of heteroscedasticity, autocorrelation, multicollinearity, and measurement errors, along with appropriate diagnostic methods.
3. **Apply** regression diagnostics, model selection techniques, statistical tests, and remedial methods such as generalized least squares (GLS), ridge regression, variable transformation, and instrumental variable estimation to econometric data
4. **Analyze** regression models by identifying influential observations, violations of model assumptions, and the effects of heteroscedasticity, autocorrelation, multicollinearity, and measurement errors on parameter estimation.
5. **Evaluate** alternative econometric models and remedial techniques using diagnostic measures, model selection criteria, and statistical evidence to determine the most appropriate model for a given dataset
6. **Create** appropriate econometric modelling strategies by integrating diagnostic procedures, corrective methods, and model selection techniques to solve real-world economic and social science problems

Course Outcomes for Module II

1. **Remember** the basic concepts, terminology, and categories of machine learning including supervised, unsupervised, and semi-supervised learning.
2. **Understand** performance evaluation metrics, bias-variance trade-off, train-test split, cross-validation, and dimensionality reduction techniques.
3. **Apply** machine learning algorithms such as PCA, clustering methods, Ridge Regression, LASSO, k-NN, Logistic Regression, and Discriminant Analysis to data analysis problems.
4. **Analyze** datasets using classification, regression, clustering, and dimensionality reduction techniques and interpret model outputs.
5. **Evaluate** the performance of machine learning models using accuracy measures, ROC-AUC, confusion matrix, and model validation procedures.
6. **Create** machine learning solutions by selecting, tuning, and implementing appropriate predictive and descriptive models for real-world datasets.

Syllabus

Unit/Module	Content	No. of lectures	CO mapping	Cognitive levels
Module I	**Econometrics**			
Unit 1	**Introduction** Review of simple and multiple linear regression models and related tests of significance. Use of nominal and ordinal predictors. *Regression diagnostics:* Residual plots, outliers, hat matrix and leverage points, influential points, DFBETA, DFFIT and Cook's distance. *Model Selection:* Mallows Cp, AIC. Best subset selection.	3 3 2	CO1 CO2 CO3 CO4	K1 K2 K3 K4
	Violation of classical assumptions and remedies **Heteroscedasticity** Nature of heteroscedasticity – illustrative examples, OLS method under heteroscedasticity and its consequences, detecting			

Unit 2	heteroscedasticity- residual plot, Glejser test, Goldfeld-Quandt test, remedial measure through variable transformation and generalized least squares (GLS) and estimated GLS. **Autocorrelation** Nature of autocorrelation – illustrative examples, OLS method under autocorrelation – AR(1) model, detecting autocorrelation – residual plot, Durbin-Watson test, estimated GLS method for correcting autocorrelation. **Multicollinearity** OLS method under perfect multicollinearity and its consequences. Detecting multicollinearity – thumb rules based on R^2 , pair-wise and partial correlations, VIF and condition number. Remedial measures -- more data, dropping and transformation of variables, Ridge Regression. **Measurement error in covariates** Berkson and classical measurement error models, Bias and inconsistency caused by measurement error in simple linear regression. Remedial measures – using restrictions in maximum likelihood methods, Instrumental Variable technique.	5 4 5 4	CO2 CO3 CO4 CO5 CO6	K2 K3 K4 K5 K6
Module II	**Machine Learning**			
Unit 1	*Machine Learning Overview:* Supervised, unsupervised and semi-supervised techniques. Regression and Classification problems. Curse of Dimensionality. Bias-variance trade-off. Train-test split. Accuracy, Precision, Recall, F1 score. Confusion matrix. ROC curve & AUC. Cross	7	CO1 CO2 CO5	K1 K2 K5

	Validation (Hyperparameter tuning).			
Unit 2	*Unsupervised Learning techniques:* Clustering - Partitioning and hierarchical clustering. Dimensionality Reduction - Principal component Analysis. Fator Analysis.	9	CO2 CO3 CO4	K2 K3 K4
Unit 3	*Supervised Learning techniques:* Ridge Regression, LASSO. Classification - k-Nearest Neighbors, linear Discriminant Analysis, Logistic Regression. Decision Trees (CART), Random Forests.	10	CO3 CO4 CO5 CO6	K3 K4 K5 K6

Reading/Reference list

1. Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics*, 5th ed., McGraw-Hill
2. Maddala, G. S., & Lahiri, K. (2009). *Introduction to econometrics* (4th ed.). John Wiley & Sons.
3. Johnston, J., & Dinardo, J. (1997). *Econometric methods* (4th ed.). McGraw-Hill.
4. Wooldridge, J. M. (2022). *Introductory econometrics: A modern approach* (7th ed.). Cengage Learning
5. Koutsoyiannis, A. (1977). *Theory of econometrics: An introductory exposition of econometric methods* (2nd ed.). Macmillan Press
6. Carroll, R. J., Ruppert, D., Stefanski, L. A., & Crainiceanu, C. M. (2006). *Measurement error in nonlinear models: A modern perspective* (2nd ed.). Chapman & Hall/CRC.
7. James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). Statistical learning. In *An introduction to statistical learning: With applications in Python* (pp. 15-67). Cham: Springer International Publishing
8. Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2, pp. 1-758). New York: springer.
9. Breiman, L., Friedman, J., Olshen, R. A., & Stone, C. J. (2017). *Classification and*

regression trees. Chapman and Hall/CRC.

10. Johnson, R. A., & Wichern, D. W. (2002). Applied multivariate statistical analysis.

Evaluation

	CIA: 30 End-Sem: 70 Total: 100	
Paper Structure for Semester Exam	**Module 1 (35 marks)**	**Module II (35 marks)**
	Short questions (5 marks each): 4 out of 6 Long questions (15 marks each): 1 out of 2	Short questions (5 marks each): 4 out of 6 Long questions (15 marks each):1 out of 2

CO	CO Description for Module I	Cognitive levels
CO1	**Remember** the fundamental concepts of simple and multiple linear regression, model assumptions, regression diagnostics, model selection criteria, and sources of econometric problems	K1
CO2	**Understand** the assumptions underlying classical linear regression models, the causes and consequences of heteroscedasticity, autocorrelation, multicollinearity, and measurement errors, along with appropriate diagnostic methods.	K2
CO3	**Apply** regression diagnostics, model selection techniques, statistical tests, and remedial methods such as generalized least squares (GLS), ridge regression, variable transformation, and instrumental variable estimation to econometric data	K3
CO4	**Analyze** regression models by identifying influential observations, violations of model assumptions, and the effects of heteroscedasticity, autocorrelation, multicollinearity, and measurement errors on parameter estimation	K4
CO5	**Evaluate** alternative econometric models and remedial techniques using diagnostic measures, model selection criteria, and statistical evidence to determine the most appropriate model for a given dataset	K5
CO6	**Create** appropriate econometric modelling strategies by integrating diagnostic procedures, corrective methods, and model selection techniques to solve real-world economic and social science problems.	K6

CO	CO Description for Module II	Cognitive levels
CO1	**Remember** the basic concepts, terminology, and categories of machine learning including supervised, unsupervised, and semi-supervised learning.	K1
CO2	**Understand** performance evaluation metrics, bias-variance trade-off, train-test split, cross-validation, and dimensionality reduction techniques.	K2
CO3	**Apply** machine learning algorithms such as PCA, clustering methods, Ridge Regression, LASSO, k-NN, Logistic Regression, and Discriminant Analysis to data analysis problems.	K3
CO4	**Analyze** datasets using classification, regression, clustering, and dimensionality reduction techniques and interpret model outputs.	K4
CO5	**Evaluate** the performance of machine learning models using accuracy measures, ROC-AUC, confusion matrix, and model validation procedures.	K5
CO6	**Create** machine learning solutions by selecting, tuning, and implementing appropriate predictive and descriptive models for real-world datasets.	K6