

Techniques of Data Handling and Visualization (24M22MA112)

Course Description

Course Code	24M22MA112	Semester Even (specify Odd/Even)	Semester II Session 2023-24 Month from Jan-May 2024
Course Name	Techniques of Data Handling and Visualization		
Credits	3	Contact Hours	3-0-0
Faculty (Names)	Coordinator(s)		
	Teacher(s) (Alphabetically)		
COURSE OUTCOMES: After the successful completion of this course, the student will be able to			COGNITIVE LEVELS
CO 1	explain important terms related to the data handling.		Understanding (C2)
CO 2	organize data using visualization, cleaning and management techniques.		Applying (C3)
CO 3	select data models by error analysis and data size.		Applying (C3)
CO 4	compare different techniques of data analysis and presentation.		Analyzing (C4)
Module No.	Title of the Module	Topics in the Module	No. of Lectures for the module
1.	Introduction	Literature on applications of Data Analytics in science and engineering, examples of supervised, unsupervised and reinforcement learning, Variable Types and Terminology, Data Analytics in cloud computing.	8
2.	Data Collection and Storage techniques	Direct involvement of software engineers, indirect involvement of software engineers, study of work artifacts, cloud storage, software defined storage, file storage, block storage, object storage.	6
3	Data Visualization and Data Cleaning Techniques	Histogram, Box plot, scatter plot, pixel-based techniques, quality metrics for data presentation, data filtering, missing data, duplicates and outliers, scaling, normalization.	7
4	Functionalities of data storage management	Data Availability, Easy and Quick Access, automatic tiering, data backup solutions, encryption for both in-transit	7

		and stored data, multifactor authentication to restrict unauthorized access, Control and compliance.	
5	Model Assessment and Selection	Statistical Decision Theory for regression and classification tasks, training and testing errors, regularization and stability, Bias variance trade-off, Optimism of the Training Error Rate, Estimates of In-Sample Prediction Error, Bayesian information criterion.	8
6	Dimension and Size issues in Data	Curse of dimensionality, Principal component analysis, factor analysis, bagging, boosting.	6
Total number of Lectures			42
Evaluation Criteria			
Components		Maximum Marks	
T1		20	
T2		20	
End Semester Examination		35	
TA		25 (Quiz, Assignments, Tutorials, Project)	
Total		100	
Recommended Reading material:			
Project based learning: Students in a small group will collect sample data set. The data presentation techniques will be applied to explain all the data in use and the data management techniques will be explored for efficient data storage. In this way, students will be able to learn presentation of data and its management effectively.			
Text Books			
1.	T. Hastie, R. Tibshirani and J. Friedman , The Elements of Statistical Learning, 2 nd Ed., Springer, 2008.		
2.	S. R. Das , Data Science Theories, Models, Algorithms, and Analytics, Apache License, 2016.		
3.	W. Stallings and L. Brown , Computer Security: Principles and Practice, 2nd edition, Pearson/Prentice Hall, 2012. ISBN-10: 0136004245, ISBN-13: 9780136004240.		
4.	W. Olsen , Data Collection: Key Debates and Methods in Social Research, SAGE Publications Ltd, 2012. ISBN:9781847872562 Online ISBN:9781473914230.		
5.	S. Gnanasundaram and A. Shrivastava , Information Storage and Management Storing, Managing, and Protecting Digital Information in Classic, Virtualized, and Cloud Environments 2nd Edition, 2012. ISBN: 978-1-118-23696-3.		
6	C. O. Wilke , Fundamentals of Data Visualization A Primer on Making Informative and Compelling Figures, O'Reilly Media, 2019.ISBN-13: 978-1-492-03108-6.		

Regression Models for Data Inference and Prediction (24M22MA212)

Course Description

Course Code	24M22MA212	Semester Odd	Semester III Session 2024-25, Month from July- Dec 2024
Course Name	Regression Models for Data Inference and Prediction		
Credits	3	Contact Hours	3-0-0
Faculty (Names)	Coordinator(s)		
	Teacher(s) (Alphabetically)		
COURSE OUTCOMES			COGNITIVE LEVELS
After pursuing the above-mentioned course, the students will be able to:			
CO1	interpret the basic concepts of regression models, multicollinearity and model building.	Understanding (C2)	
CO2	apply parameter estimation techniques on given data set.	Applying (C3)	
CO3	analyze data and make predictions and inferences using appropriate regression models	Analyzing (C4)	
CO4	evaluate important variables to be included in order to make a regression model expressive.	Evaluating (C5)	
Module No.	Title of the Module	Topics in the Module	No. of Lectures
1.	Introduction	Regression and model building, Data collection and uses of regression	2
2.	Simple Linear Regression	Simple linear regression model, Least-Squares Estimation of the model parameters, Inference about the slope and the intercept and the slope parameters, Prediction of new observations, Estimation by maximum likelihood method.	6
3.	Multiple Linear Regression	Multiple linear regression models, Least-Squares Estimation of the model parameters, Inference in multiple linear regression	8
4.	Model Adequacy Checking	Residual analysis, Detection and treatment of outliers, Lack of fit of the regression model	6
5.	Multicollinearity	Source of multicollinearity, Consequences of multicollinearity, Multicollinearity diagnostics, Remedies for multicollinearity	5

6.	Logistic regression Model	Logistic Regression Models its linear Predictions, Prediction of new observations, Maximum likelihood estimation of parameters, Interpretation of parameters	6
7.	Variable Selection and Model Building	Introduction: the model building problem, Model misspecification, Criteria for evaluating subset regression, Computational techniques for variable selection: all possible regressions, Stepwise regression methods	9
Total number of Lectures			42
Evaluation Criteria			
Components		Maximum Marks	
T1		20	
T2		20	
End Semester Examination		35	
TA		25 (Quiz, Assignments, PBL etc.)	
Total		100	
Project based learning: Each student in a group of 4-5 will collect data and apply appropriate regression models using software for prediction purpose. The students will be able to use various regression models to achieve the defined objectives in different fields.			
Recommended Reading material: Author(s), Title, Edition, Publisher, Year of Publication etc. (Text books, Reference Books, Journals, Reports, Websites etc. in the IEEE format)			
1.	Montgomery, D.C., Peck, E.A. and Vining, G.G. (2012). Introduction to Linear Regression Analysis (3rd Edition). John Wiley & Sons, Inc., New York.		
2.	Bingham, N. H. and Fry, J. M. (2010). Regression: Linear Models in Statistics. Springer, USA		
3.	Myrers, R.H. (1990). Classical and Modern Regression with Applications (2nd Edition). PWS-Kent Publishers, Boston.		
4.	Draper, N.R. and Smith, H. (1998). Applied Regression Analysis (3rd Edition). John Wiley & Sons, Inc., New York.		
5.	Golberg, M. A. and Cho, H. A. (2010) : Introduction to Regression Analysis, WIT press, USA		