EXPLORATORY DATA ANALYSIS FOR BUSINESS			
Course Code	**22MBABA304**	**CIE Marks**	**50**
Teaching Hours/Week (L:P:SDA)	**4:0:0**	**SEE Marks**	**50**
Total Hours of Pedagogy	**50**	**Total Marks**	**100**
Credits	**04**	**Exam Hours**	**03**

Course Learning objectives:

- To make students to understand data analysis and visualization process
- To make students to use tools to extract trends from existing data to drive business decisions
- To make students to analyze and to visualize different characteristics of a data set.
- To make students to develop summarized reports

Module-1 8 Hours

Introduction to Data Mining: Applications- Nature of The Problem- Classification Problems in Real Life- Email Spam, Handwritten Digit Recognition, Image segmentation, Speech Recognition, DNA Expression Microarray, DNA Sequence Classification. Exploratory Data Analysis (EDA)- What is Data- Numerical Summarization - Measures of Similarity and Dissimilarity, Proximity- Distance- Euclidean Distance, Minkowski Distance, Mahalanobis Distance Visualization- Tools for Displaying Single Variables - Tools for Displaying Relationships Between Two Variables - Tools for Displaying More Than Two Variables

R Scripts- R Library: ggplot2-R Markdown

Module-2 8 Hours

Statistical Learning and Model Selection: Prediction Accuracy - Prediction Error, Training and Test Error as A Function of Model Complexity, Over fitting a Model, Bias-Variance Trade-off, Cross Validation- Holdout Sample: Training and Test Data, Three-way Split: Training, Validation and Test Data, Cross-Validation, Random Sub sampling, K-fold Cross-Validation, Leave-One-Out Cross-Validation with examples for each.

Module-3 8 Hours

Linear Regression and Variable Selection: Meaning- Review Expectation, Variance, Frequentist Basics, Parameter Estimation, Linear Methods, Point Estimate, Example Results, Theoretical Justification, R Scripts.

Variable Selection- Variable Selection for the Linear Model, R Scripts.

Module-4 9 Hours

Regression Shrinkage Methods and Tree based method: Meaning, Types- Ridge Regression, Compare Squared Loss for Ridge Regression, More on Coefficient Shrinkage, The Lasso.

Tree Based Methods- Construct the Tree, The Impurity Function, Estimate the Posterior Probabilities of Classes in Each Node, Advantages of the Tree-Structured Approach, Variable Combinations, Missing Values, Right Sized Tree via Pruning, Bagging and Random Forests, R Scripts, Bagging, From Bagging to Random Forests, Boosting

Module-5 10 Hours

Principal Components Analysis and Classification: Singular Value Decomposition (SVD), Principal Components, Principal Components Analysis(PCA), Geometric Interpretation, Acquire Data, Classification - Classification Error Rate, Bayes Classification Rule, Linear Methods for Classification, Logistic Regression - Assumptions, Comparison with Linear Regression on Indicators- Fitting based on Optimization Criterion, Binary Classification, Multiclass Case ($K \geq 3$), Discriminant Analysis - Class Density Estimation, Linear Discriminant Analysis, Optimal Classification

Module-6 7 Hours

Support Vector Machines: Overview, When Data is Linearly Separable, Support Vector Classifier, When Data is NOT Linearly Separable, Kernel Functions, Multiclass SVM.

Assessment Details (both CIE and SEE)

The weightage of Continuous Internal Evaluation (CIE) is 50% and for Semester End Exam (SEE) is 50%. The minimum passing marks for the CIE is 50% of the maximum marks. Minimum passing marks in SEE is 40% of the maximum marks of SEE. A student shall be deemed to have satisfied the academic requirements (passed) and earned the credits allotted to each course if the student secures not less than 50% in the sum total of the CIE (Continuous Internal Evaluation) and SEE (Semester End Examination) taken together.

Continuous Internal Evaluation:

There shall be a maximum of 50 CIE Marks. A candidate shall obtain not less than 50% of the maximum marks prescribed for the CIE.

CIE Marks shall be based on:

a) Tests (for 25Marks) and

b) Assignments, presentations, Quiz, Simulation, Experimentation, Mini project, oral examination, field work and class participation etc., (for 25 Marks) conducted in the respective course. Course instructors are given autonomy in choosing a few of the above based on the subject relevance and should maintain necessary supporting documents for same.

Semester End Examination:

The SEE question paper will be set for 100 marks and the marks scored will be proportionately reduced to 50.

- The question paper will have 8 full questions carrying equal marks.
- Each full question is for 20 marks with 3 sub questions.
- Each full question will have sub question covering all the topics.
- The students will have to answer five full questions; selecting four full questions from question number one to seven in the pattern of 3, 7 & 10 Marks and question number eight is compulsory.

Suggested Learning Resources:Books

1. John W. Tukey “Exploratory Data Analysis”, 1st Edition, ISBN13: 978-0201076165, ISBN-10: 0201076160
2. Foster Provost and Tom Fawcett. “Data Science for Business: What you need to know about data mining and data-analytic thinking”. O'Reilly Media, latest edition, ISBN-13: 978-1449361327
3. Hadley Wickham, Garrett Grolemund."R for Data Science: Import, Tidy, Transform, Visualize, and Model Data", Publisher: "O'Reilly Media, Inc.", 2016, ISBN 1491910364, 9781491910368
4. Cathy O'Neil, Rachel Schutt. "Doing Data Science: Straight Talk from the Frontline", Publisher: "O'Reilly Media, Inc.", 2013, ISBN 144936389X, 9781449363895

Web links and Video Lectures (e-Resources):

- https://r.search.yahoo.com/_ylt=AwrKEtWRgvdiaAQVgLW7HAx.;_ylu=Y29sbwNzZzMEcG9zAzQEdnRpZAMEc2VjA3Ny/RV=2/RE=1660416785/RO=10/RU=https%3a%2f%2farchive.org%2fdetails%2fexploratorydataa0000tuke_7616/RK=2/RS=5BVZN5konKiTESg8jAvGKFu9qtw-
- https://r.search.yahoo.com/_ylt=AwrKEtWRgvdiaAQVf7W7HAx.;_ylu=Y29sbwNzZzMEcG9zAzMEdnRpZAMEc2VjA3Ny/RV=2/RE=1660416785/RO=10/RU=https%3a%2f%2fjhu-advdatasci.github.io%2f2019%2flectures%2fEDA.pdf/RK=2/RS=pupbdtbn2rtanCvRHfwBi9lWxMk-

Note: The aforesaid links and study materials are suggestive in nature, they may be used with due regards to copy rights, patenting and other IPR rules.

Skill Development Activities Suggested:

- To conduct explorative research to collect data and analyse using statistical tools like excel and SPSS
- Interpret the data objectively and prepare report

Course outcome :

At the end of the course the student will be able to :

Sl. No.	Description	Blooms Level
CO1	Understand Data Mining and its importance .	L2
CO2	Apply knowledge of research design for business problems	L3
CO3	Analyze the cause and effect relationship between the variables from the analysis	L4
CO4	Evaluate Regression and decision tree based methodsto solve business problems	L5

Mapping of COs and POs

	PO1	PO2	PO3	PO4	PO5	PSO 1	PSO 2	PSO 3	PSO 4
CO1	1				2	3			
CO2		2	2				2		
CO3				3		3		2	
CO4		2		2			1		2

HOD
PG Dept. of Business Administration
Alva's Institute of Engg. & Technology
Mijar - 574225

01.02.2023

Mapping of COs and Pos

	PO1	PO2	PO3	PO4	PO5	PSO 1	PSO 2	PSO 3	PSO 4
CO1	1				2	3			
CO2		2	2				2		
CO3				3		3		2	
CO4		2		2			1		2

HOD
PG Dept. of Business Administration
Alva's Institute of Engg. & Technology
Mijar - 574225

- www.Programiz.
- www.CodeCademy
- www.FreeCodeCamp

Note: The aforesaid links and study materials are suggestive in nature, they may be used with due regards to copy rights, patenting and other IPR rules.

Skill Development Activities Suggested to:

- Practice on Python software to become Expertise in data visualization process.
- Access to web-frameworks and get motivated to work on analytical tools
- Analyse any big retail chain data using python

Course outcome:

At the end of the course the student will be able to :

Sl. No.	Description	Blooms Level
CO1	Understand the concepts of python programming	L1
CO2	Structure a simple Python programs for solving problems.	L2
CO3	Apply the knowledge to decompose a Python program into functions.	L3
CO4	Analyse and Represent compound data using Python lists, tuples, dictionaries.	L4
CO5	Read and write data form/to files in Python Program.	

Mapping of COS and Pos

	PO1	PO2	PO3	PO4	PO5	PSO1	PSO2	PSO3	PSO4
CO1	1				2	3			
CO2		2	2				2		
CO3				3		3		2	
CO4		2		2			1		2

HOD
PG Dept. of Business Administration
Alva's Institute of Engg. & Technology
Mijar - 574...

RUBRICS FOR INTERNSHIP 22MBAIN 307

Sl.No.	Evaluation Type	Particulars	Marks
1	CIE	Assessment by the Guide- Interaction with the student by Seminars, etc,.	25
2	CIE	Report Evaluation by the Guide	25
3	SEE	Viva-Voce Examination to be conducted by the Guide and an External examiner from the Industry/Institute	50
Total			100

MARK SHEET FOR VIVA VOCE EXAMINATION (SEE)

Visvesvaraya Technological University
Name of the Institution
Name of the Department
Course Code: 22MBA IN 307 and Course Title: Internship

Sl.No.	Aspects	Marks
1	Introduction and Understanding the Industry	5
2	Understanding the Corporate Functions/Company profile	10
3	Mckensy's 7S framework and Porter's Five Force Model	10
4	SWOT/SWOC analysis justification	10
5	Financial statement analysis	5
6	Learning experience	10
Total		50

Marks Sheet for Internship Viva Voce examination

Sl.No.	USN	1	2	3	4	5	6	Total
1								
2								
3								
4								
5								

Signature of Internal Examiner
Name and Designation with affiliation

Signature of External Examiner
Name and Designation with affiliation

HOD
PG Dept. of Business Administration
Alva's Institute of Engg. & Technology
Mijar - 574225

01.02.2023

Mapping of COS and Pos

	PO1	PO2	PO3	PO4	PO5	PSO1	PSO2	PSO3	PSO4
CO1	1				2	3			
CO2		2	2				2		
CO3				3		3		2	
CO4		2		2			1		2

HOD
PG Dept. of Business Administration
Alva's Institute of Engg. & Technology
Mijar - 574225