Mahatma Gandhi University

Kottayam

Programme	**BSc (Hons) Statistics**					
Course Name	**Applied Statistics and Data Mining**					
Type of Course	**SEC**					
Course Code	**MG5SECSTA300**					
Course Level	**300-399**					
Course Summary	The course explores the basic concepts of Index Numbers, Vital Statistics and Data Mining techniques					
Semester	5	Credits			3	Total Hours
Course Details	Learning Approach	Lecture	Tutorial	Practical	Others	
		3				45
Pre-requisites						

CO No.	Expected Course Outcome	Learning Domains	PO No
1	Construction and Evaluation of Index Numbers.	E	2
2	Understanding and Application of Vital Statistics.	An	2
3	Design, implement, and evaluate Data Mining solutions across diverse domains.	An	2
4	Apply Data Mining techniques in healthcare, finance, marketing, and social sciences.	A	2

***Remember (K), Understand (U), Apply (A), Analyze (An), Evaluate I, Create (C), Skill (S),** Interest (I) and Appreciation (Ap)*

COURSE CONTENT

Content for Classroom Transaction (Sub-units)

Module	Course Description	Hours	CO No.
	Modified	**Modified**	**Modified**
Module 1	**Index Numbers**	**15**	
1.1	Classification of Index Numbers-Price, Quantity and Value indices, Methods of constructing Index Numbers, Unweighted Index Numbers, Weighted Index Numbers	**3**	**1**
1.2	Laspeyre's, Paasche's, Fisher's Marshal-Edgeworth, Dorbish - Bowley's, Kellys Index numbers	**3**	**1**
1.3	Different tests of a good Index numbers: - Unit test, Time Reversal, Test-Factor Reversal Test- Circular test	**3**	**1**
1.4	Fixed base and chain base Index. Numbers, Splicing	**3**	**1**
1.5	Cost of Living Index Numbers- Consumer Price Index Numbers, wholesale price index and index of industrial production	**3**	**1**
Module 2	**Vital Statistics**	**15**	
2.1	Introduction to Vital Statistics, Sources of Vital Statistics: Sample Registration System (SRS), Civil Registration System (CRS).	**3**	**2**
2.2	Fertility rate-Crude Birth Rate (CBR), Age- Specific Fertility Rate (ASFR), Total Fertility Rate (TFR), General Fertility Rate (GFR)	**3**	**2**
2.3	Mortality rate-Crude Death Rate (CDR), Age- Specific Death Rate (ASDR), Standardized Death Rate (SDR), Infant Mortality Rate (IMR)	**3**	**2**
2.4	Population growth- Net Reproduction Rate(NRR), Gross Reproduction Rate (GRR).	**3**	**2**
2.5	Construction of simple life tables-Complete life tables, Abridged life tables, Characteristics of life tables	**3**	**2**
3	**Data Mining**	**15**	
3.1	Introduction to Data Mining and its Applications. Data Pre-processing Techniques: Data Cleaning, Data Integration, Data Transformation, and Data Reduction.	**3**	**3**
3.2	Classification and Prediction Techniques: Decision Trees, Naive Bayes, k-Nearest Neighbors (k-NN),	**3**	**3**

	Linear Regression, Multiple Regression, Time Series Analysis		
3.3	Clustering Techniques: K-Means, Hierarchical Clustering	**3**	**3**
3.4	Association Rule Mining: **(**ARM)-Market basket analysis, Apriori algorithm, FP-growth algorithm. Evolution and Deviation Analysis- Time-Series Analysis, Periodicity Pattern Matching, Similarity-Based Analysis, Outlier Detection	**3**	**3**
3.5	Evaluation of Data Mining Models. Application of Data mining in healthcare, finance, marketing, and social sciences.	**3**	**4**
Module 4	**Teacher Specific Content**		

Teaching and Learning Approach	**Classroom Procedure (Mode of transaction)** Direct Instruction: Brainstorming lecture, E-learning, Interactive Instruction, Seminar, Group Assignments, Authentic learning, Presentation by students by group.
Assessment Types	**MODE OF ASSESSMENT** **A. Continuous Comprehensive Assessment (CCA)** ***Formative assessment*** Quiz, Assignments, Seminar, Viva **Theory:10 marks** ***Summative assessment*** Written tests **Theory: 15 marks**
	B. End Semester Evaluation (ESE) **Theory**: **50 marks (Duration : 1.5 Hrs)** **i) Short answer type questions: Answer any 7 questions out of 10 (7*2=14).** **ii) Short essay type questions: Answer any 4 questions out of 6 (4*6=24).** **iii) Essay type questions: Answer any 1 question out of 2 (1*12=12)**

References

1. Gupta, S.C. and. Kapoor, V.K. (2018). Fundamentals of Applied Statistics, Sultan Chand & Co. New Delhi.
2. Srivastava, O.S. (1983). A Text Book of Demography, Vikas Publishing House, New Delhi.
3. Parimal Mukhopadhyay. (2005). Applied Statistics. Books & Allied (p) Ltd.
4. Han, J., Kamber, M., & Pei, J. (2012). Data mining concepts and techniques third edition. University of Illinois at Urbana-Champaign Micheline Kamber Jian Pei Simon Fraser University.
5. Tan, P. N., Steinbach, M., & Kumar, V. (2016). Introduction to data mining. Pearson Education India

Suggested Readings

1. Goon, A.M. Gupta, M.K. and Das Gupta, B. (2016): Fundamentals of Statistics, Vol. II, World press, Calcutta.
2. Newsholme, A. (2021). The Elements of Vital Statistics, Routledge, Taylor & Francis Group.
3. Keyfitz, N, and Beekman, J.A. (2010), Demography through Problems, 1st Edition, Springer- Verlag.
4. Jhingan, M.L., Bhatt, B.K. and Desai, J.N. (2016). Demography, 3rdEdition, Vrinda Publications (P) Ltd, Delhi.
5. Benjamin B (1960). Elements of Vital Statistics, Quadrangle Books.
6. Whipple, G.C. (2022). Vital Statistics: An Introduction to the Science of Demography, Legare Street Press.
7. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2017). Data Mining: Practical machine learning tools and techniques, Elsevier Inc.