Kanya Maha Vidyalaya, Jalandhar (Autonomous)

SCHEME AND CURRICULUM OF EXAMINATIONS OF THREE YEAR DEGREE PROGRAMME

Bachelor of Science (Information Technology)

Session 2019-20

Additional / Optional paper for Specialization in Data Science

Course Code	Course Name	Course Type	Marks				Examination Time (in Hours)
			Total	**Ext.**		**CA**	
				L	**P**		
BITL-1117	*Data Preprocessing and Data Mining (optional)	O	75	60	-	15	3
	Total		75				

Note:

O - Optional

*One additional/optional paper will be studied by the candidate if she opts for Specialization in Data Science

B.Sc. (Information Technology) Semester- I
(Session 2019-20)
DATA PREPROCESSING AND DATA MINING
COURSE CODE: BITL–1117

Max. Marks: 75

Time: 3 Hrs   **Theory: 60**
CA:15

Instructions for Paper Setter -

Eight questions of equal marks (12 marks each) are to set, two in each of the four sections (A-D). Questions of Sections A-D should be set from Units I-IV of the syllabus respectively. Questions may be divided into parts(not exceeding four). Candidates are required to attempt five questions, selecting at least one question from each section. The fifth question may be attempted from any section

UNIT I
Data Processing: Basic Terminology of Data, Information and Knowledge.
Preprocessing the Data, Data cleaning, Data transformation, Data reduction.
Statistical description of data: Mean, Median and Mode.

UNIT II
Measures of Dispersion: Range, Quartile Deviation, Mean Deviation, Standard Deviation
Data Mining: Introduction, need, Applications, Process and techniques.
Data mining model

UNIT III

Excel Basics: Introduction, Basics of Cell, Modifying columns, rows and cells in excel, cells formatting, create a simple formula in excel, worksheet basics, Charts, printing an excel sheet.

UNIT IV

Creating Complex Formulas in Excel, Working with Basic Functions - to find values for a range of cells
Data analysis tools: Analyze, Detect, Fill from, Forecast, Scenario tool

References/Textbooks:
1. J. Han , M. Kamber and J. Pei , Data Mining: Concepts and Techniques , 3rd. Edition, Morgan Kaufmann Publishers.
2. Handbook of Data Mining by NongYe
3. Fundamentals of Numerical Methods and Statistical Techniques by Anshuman Sharma
4. Excel 2010 Bible by John Walkenbach