**Dr.SNS RAJALAKSHMI COLLEGE OF ARTS & SCIENCE (AUTONOMOUS)**

**COIMBATORE-49.**

**DEPARTMENT OF DATA ANALYTICS**

**INTRODUCTION TO DATA ANALYTICS - 21UCU402**

I What is data science ? (CO:I) (BL:I) 3

I When to use Big data concept ? (CO:I) (BL:I) 3

I Define KDD Process (CO:I) (BL:I) 3

I Classify Examples of data use (CO:I) (BL:II) 3

I List any two algorithm used in data analytics and explain in breif ?(CO:I) (BL:II) 3

I Summarize Univariate data visualization (CO:I) (BL:II) 3

I Identify the usages of data (CO:I) (BL:III) 3

I Experiment What is Data?(CO:I) (BL:III) 3

I Utilize Big Data and Data Science(CO:I) (BL:III) 3

I Analyze Small Data (CO:I) (BL:IV) 3

I Simplify Polish Company Insolvency Data (CO:I) (BL:IV) 3

I A survey on What is Data? (CO:I) (BL:IV) 3

II What is attribute selection?(CO:II) (BL:I) 3

II When and where to use infographics and word clouds ?(CO:II) (BL:I) 3

II Recall the process behind multivariate statistics(CO:II) (BL:I) 3

II Classify between data and noisy data(CO:II) (BL:II) 3

II Demonstrate the ability of filters with examples?(CO:II) (BL:II) 3

II Illustarte the meaning of embedded data(CO:II) (BL:II) 3

II Apply word clouds with filters (CO:II) (BL:III) 3

II Model the concept of Converting Nominal to Relative(CO:II) (BL:III) 3

II Apply a logic to avoid repeated data (CO:II) (BL:III) 3

II Examine the meaning of inconsistent data (CO:II) (BL:IV) 3

II Compare data and filtered data and pre processed data(CO:II) (BL:IV) 3

II Analyze reduction techniques and the need of implementing the concept of reduction (CO:II) (BL:IV) 3

III What is the concept of clustering in brief ? (CO:III) (BL:I) 3

III Is Apriori an algorithm if so, when to use it and where to use it ?(CO: III) (BL:I) 3

III Define the word Dendrograms?(CO: III) (BL:I) 3

III Demonstrate How K-means concept is working (CO: III) (BL:II) 3

III Explain about clustering techniques (CO: III) (BL:II) 3

III Illustrate the implementation of Linkage Criterion(CO: III) (BL:II) 3

III Experiment about the concept of DBSCAN(CO: III) (BL:III) 3

III Identify cross support patterns (CO: III) (BL:III) 3

III Utilize the usage of mining (CO: III) (BL:III) 3

III Distinguish between Eclat and Fp-growth (CO: III) (BL:IV) 3

III Define the various validations on clustering techniques (CO: III) (BL:IV) 3

III Examine the concept of association rules (CO: III) (BL:IV) 3

IV How to apply regression on any data ?(CO: IV) (BL:I) 3

IV Define the meaning of classification ?(CO: IV) (BL:I) 3

IV Find the difference between regression and linear regression?(CO: IV) (BL:I) 3

IV Classify various types of regression concepts(CO: IV) (BL:II) 3

IV Explain in brief about the root cause of emprical error(CO: IV) (BL:II) 3

IV Outline about generalization concepts (CO: IV) (BL:II) 3

IV Experiment about lasso regression algorithm (CO: IV) (BL:III) 3

IV Identify the implementation of KNN algorithm (CO: IV) (BL:III) 3

IV Organize the results of model validation technique(CO: IV) (BL:III) 3

IV Categorize the list of algorithms used in classification and regression (CO: IV) (BL:IV) 3

IV Examine the concept of ridge regression (CO: IV) (BL:IV) 3

IV Inference on Naive Bayes Algorithm (CO: IV) (BL:IV) 3

V What is active learning ? (CO: V) (BL:I) 3

V How to use algorithm Bias technique ? (CO: V) (BL:I) 3

V Where are AdaBoost concept got applied in the real world ? (CO: V) (BL:I) 3

V Classify support vector machine (CO: V) (BL:II) 3

V Compare Neural network and Artificial nueral network (CO: V) (BL:II) 3

V Demonstrate model preparation in machine learning concept (CO: V) (BL:II) 3

V Construct a diagram to represent deep neural network (CO: V) (BL:III) 3

V Experiment backpropagation(CO: V) (BL:III) 3

V Organize the usage of random forest (CO: V) (BL:III) 3

V Discover the concept of the search based algorithm (CO: V) (BL:IV) 3

V Examine the utilities of ensemble learning using machine learning (CO: V) (BL:IV) 3

V List advanced predictive topics (CO: V) (BL:IV) 3

I Choose Univariate Data Visualization and its usage ?(CO:I) (BL:I) 6

I Find the difference between small data and big data ?(CO:I) (BL:I) 6

I Relate Univariate statistics concept(CO:I) (BL:I) 6

I Classify two ordinal attribute mechanism(CO:I) (BL:II) 6

I Demonstrate the usage of CRISP-DM Methodology (CO:I) (BL:II) 6

I Summarize the concept of Big data (CO:I) (BL:II) 6

I Analyze the real world usage of data analytics with real world applications (CO:I) (BL:III) 6

I Categorize the problems in preparing a knowledge data discovery ?(CO:I) (BL:III) 6

I Discover decriptive bivariate analysis(CO:I) (BL:III) 6

I Examine What we can do with data ?(CO:I) (BL:IV) 6

I Categorize Examples of Data Use(CO:I) (BL:IV) 6

I Compare Descriptive Bivariate Analysis(CO:I) (BL:IV) 6

II How to aggregate various types of attributed data ?(CO:II) (BL:I) 6

II Define about multivariate frequencies ?(CO:II) (BL:I) 6

II Relate the main logic of wrappers ?(CO:II) (BL:I) 6

II Compare in detail about the step by step process of data preprocessing (CO:II) (BL:II) 6

II Explain about missing values in data and how it is different from redundant data in data analytics ? (CO:II) (BL:II) 6

II Interpret the procedure about principal component analysis (CO:II) (BL:II) 6

II Distinguish between attribute aggregation and attribute selection (CO:II) (BL:III) 6

II Make use of preprocessing techniques in data analytics (CO:II) (BL:III) 6

II Utilize various convering concepts (CO:II) (BL:III) 6

II Dissect Converting Relative or Absolute to Ordinal or Nominal(CO:II) (BL:IV) 6

II Examine the frequencies, statistics and data visualization association with multivariate logic (CO:II) (BL:IV) 6

II simplify Infographics and Word Clouds(CO:II) (BL:IV) 6

III How to use association rules to resolve related issue (CO:III) (BL:I) 6

III Classify Simpson's paradox(CO:III) (BL:I) 6

III Tell about distance measures(CO:III) (BL:I) 6

III Ouline the meaning of k-means and how it works (CO:III) (BL:II) 6

III Summarize Distance Measures for Non-conventional Attributes(CO:III) (BL:II) 6

III Illustrate Distance Measures for Objects with Quantitative Attributes(CO:III) (BL:II) 6

III Identify sequential pattern idea and pattern logic on sequential data(CO:III) (BL:III) 6

III Make use of Other types of patterns (CO:III) (BL:III) 6

III Organize the outcomes of LIFT concept in data analytics(CO:III) (BL:III) 6

III Analyze clustering with its techniques to provide an output to a real world problem(CO:III) (BL:IV) 6

III Categorize Setting up the min\_sup Threshold concept(CO:III) (BL:IV) 6

III Contrast the general usage of association and clustering(CO:III) (BL:IV) 6

IV A case study about comparison of classification and regression(CO:IV) (BL:I) 6

IV How the Finding the Parameters of the Model is working in data analytics(CO:IV) (BL:I) 6

IV Define probablistic classification algorithm(CO:IV) (BL:I) 6

IV Explain about Partial Least Squares Regression(CO:IV) (BL:II) 6

IV Demonstrate the concept of Predictive Performance Estimation(CO:IV) (BL:II) 6

IV Outline about bias-variance trade-off concept(CO:IV) (BL:II) 6

IV Build an example on Logistic Regression Algorithm(CO:IV) (BL:III) 6

IV Construct Shrinkage Methods and explain them in detail(CO:IV) (BL:III) 6

IV Build Predictive Performance Measures for Regression(CO:IV) (BL:III) 6

IV Divide lasso , linear and logistice regression algorithms(CO:IV) (BL:IV) 6

IV Distinguish about Distance-based Learning Algorithms(CO:IV) (BL:IV) 6

IV Dissect Partial Least Squares Regression(CO:IV) (BL:IV) 6

V Choose Advanced Data Preparation Techniques for Prediction(CO:V) (BL:I) 6

V Find Decision Tree Induction Algorithms(CO:V) (BL:I) 6

V Define Semi-supervised Learning(CO:V) (BL:I) 6

V Compare Description and Prediction with Supervised Interpretable(CO:V) (BL:II) 6

V Classify SVM for Regression (CO:V) (BL:II) 6

V Summarize the concept of hierarchical classification (CO:V) (BL:II) 6

V Construct the cumalative works of non-binary classification tasks (CO:V) (BL:III) 6

V Experiment the usage of ranking classificaiton with examples (CO:V) (BL:III) 6

V Make use of Bagging technologies to enhance the data redability (CO:V) (BL:III) 6

V Categorize imbalanced data classificaion (CO:V) (BL:IV) 6

V Contrast one - class classification implementations (CO:V) (BL:IV) 6

V Dissect the usage of adaBoost algorithm with different real world operations (CO:V) (BL:IV) 6

I Define The KDD Process (CO:I)(BL:I) 10

I Explain Big Data Architectures(CO:I)(BL:II) 10

I Identify An online e-commerce company, XYZMart, is facing challenges with declining sales and wants to use data analytics to identify the factors contributing to this decline and develop strategies to boost revenue. Your task is to analyze their customer data and provide actionable insights to help XYZMart increase sales and improve customer satisfaction.(CO:I)(BL:III) 10

I Experiment The CRISP-DM Methodology(CO:I)(BL:III) 10

I Compare Quantitative and Qualitative Attributes(CO:I)(BL:IV) 10

I Inspect Univariate Statistics(CO:I)(BL:IV) 10

II Define data quality , preprocessing and data transformation(CO:II)(BL:I) 10

II Explain about descriptive Multivariate Analysis(CO:II)(BL:II) 10

II Interpret on dimensionality reduction(CO:II)(BL:III) 10

II Experiment on converting to different scale and scale type(CO:II)(BL:III) 10

II Assume A large regional healthcare provider, HealthSynergy, is grappling with data quality issues and an increasing problem of redundant data in their information system. Your task is to conduct an in-depth analysis to identify the root causes of data quality problems and recommend strategies to eliminate redundant data and enhance data accuracy in their system.(CO:II)(BL:IV) 10

II Simplify what is data and it's quality and filters and outliers(CO:II)(BL:IV) 10

III Find the difference between frequent pattern mining and frequent sequence mining (CO:III)(BL:I) 10

III Explain the real world usages of FP-Growth algorithm (CO:III)(BL:II) 10

III Devlop a solution using apriori - join based method (CO:III)(BL:III) 10

III Make use of behind support and confidence logic (CO:III)(BL:III) 10

III Compare Centroids and Distance Measures(CO:III)(BL:IV) 10

III Analyze the Customer Segmentation Using Agglomerative Hierarchical Clustering, A regional retail chain, MarketXpress, is looking to enhance its marketing strategy by segmenting its customer base. Your task is to use agglomerative hierarchical clustering to group customers based on their purchase behavior and demographics, allowing MarketXpress to tailor their marketing efforts more effectively.(CO:III)(BL:IV) 10

IV Define K nearest neighbour algorithm in detail with examples (CO:IV)(BL:I) 10

IV Explain about technique and model selection (CO:IV)(BL:II) 10

IV Experiment methods that use linear combinations of attributes (CO:IV)(BL:III) 10

IV Discover the Predicting House Prices: A Comparative Study of Regression Algorithms,A real estate agency, HomeValue Pro, wants to develop an accurate predictive model for estimating house prices in a competitive housing market. Your task is to perform a comparative study of various regression algorithms to determine which one yields the best predictive performance for their specific dataset.(CO:IV)(BL:III) 10

IV Analyze classification and predictive performance measures for classification (CO:IV)(BL:IV) 10

IV Discover the usage of case-based reasoning (CO:IV)(BL:IV) 10

V List the usage of Multivariate Adaptive Regression Splines(CO:V)(BL:I) 10

V Demonstrate Decision tree induction algorithm in detail (CO:V)(BL:II) 10

V Experiment deep networks and deep learning algorithms (CO:V)(BL:III) 10

V Make use of Enhancing Data Summarization Using Artificial Neural Networks,A leading news agency, InfoDigest, is seeking to automate the summarization of complex articles and reports to improve content accessibility for its readers. Your task is to implement an artificial neural network (ANN) to create concise and readable summaries from lengthy news articles and evaluate its effectiveness in providing informative yet concise content (CO:V)(BL:III) 10

V A detailed survey on AdaBoost algorithm (CO:V)(BL:IV) 10

V Simplify the usage of ensemble learning concept(CO:V)(BL:IV) 10