Department of Artificial Intelligence & Machine Learning

Subject : DATA ANALYTICS[23AM602]

UNIT-I

S.No	Questions	BT	CO
1	Define Data Architecture. List its essential components.	L1	CO1
2	What are sensor data and give two examples of sensors used in IoT?	L2	CO1
3	Differentiate between raw signal data and GPS data.	L2	CO1
4	What is Data Quality? Mention any three dimensions of data quality.	L1	CO2
5	Define noise in data. State any two noise reduction techniques	L1	CO1
6	What are missing values? Write any two techniques to handle them	L3	CO2
7	Define outliers. Why do they occur in datasets?	L2	CO3
8	What is duplicate data? How does it impact the accuracy of analysis?	L2	CO2
9	Write the importance of data pre-processing in data analysis.	L3	CO1
10	List any four data cleaning operations performed before model building.	L1	CO2
	Part–B (Long Answer Questions)		
11 a)	Explain in detail the design of Data Architecture with a neat block diagram. Describe ingestion, storage, processing, and governance layers.	L2	CO2
b)	Discuss various data sources such as Sensors, Signals, GPS, Web logs, and Mobile data. Explain their characteristics and challenges in managing them.	L2	CO2
12 a)	Describe the Data Management life cycle. Explain each stage including acquisition, storage, integration, cleaning, processing, and archival.	L4	CO1
b)	What are the major Data Quality issues? Explain noise, outliers, missing values, and duplicate data with examples and effects on analytics.	L4	CO2
13 a)	Explain different noise reduction techniques such as smoothing, filtering, and transformation methods with examples.	L3	CO3
b)	Discuss various outlier detection techniques such as Z-score, IQR, clustering-based detection, and boxplot analysis. Give examples.	L4	CO2
14 a)	Explain missing value handling approaches such as mean/mode imputation, regression imputation, k-NN imputation, and deletion methods.	L3	CO3
b)	Discuss the methods for detecting and removing duplicate data. Explain rule-based, key-based, and ML-based deduplication techniques.	L5	CO1
15 a)	Explain Data Pre-processing in detail. Discuss steps such as cleaning, transformation, normalization, integration, reduction, and discretization with examples.	L2	CO2
b)	Write in detail about Data Processing .	L4	CO1

UNIT-II

S.No		Questions	BT	CO
1		Define Data Analytics. What are its main components?	L1	CO1
2		What is descriptive analytics? Give an example.	L2	CO2
3		List any four commonly used data analytics tools.	L1	CO3
4		What is a business model? State two reasons why it is needed.	L2	CO1
5		Define a database. Give any two types of databases used in analytics.	L1	CO2
6		Differentiate between continuous and categorical variables.	L2	CO2
7		What is missing data? List two reasons why missing values occur.	L2	CO3
8		Define data modeling. Mention two data modeling techniques.	L1	CO1
9		What is imputation? Write two simple imputation methods.	L3	CO1
10		Write any four applications of modeling in business analytics.	L1	CO2
		Part–B(LongAnswerQuestions)		
11	a)	Explain the different types of analytics (descriptive, diagnostic, predictive, prescriptive) with examples.	L2	CO1
	b)	Describe the analytics environment. Discuss tools such as Python, R, SQL, Tableau, Power BI, and Hadoop ecosystems.	L4	CO2
12	a)	Explain the role and importance of business modeling in data analytics. Discuss how business models support decision making.	L2	CO1
	b)	Discuss the different types of databases (relational, NoSQL, distributed, cloud databases) and their relevance in analytics.	L2	CO2
13	a)	Describe types of data: structured, semi-structured, unstructured, and various types of variables with examples.	L2	CO1
	b)	Explain in detail various data modeling techniques such as ER modeling, dimensional modeling, conceptual/logical/physical data models.	L4	CO3
14	a)	What are missing values? Explain different imputation techniques such as mean, mode, median, regression, k-NN, and advanced ML imputation.	L3	CO2
	b)	Discuss applications of modeling in business domains such as finance, marketing, supply chain, HR, and risk management.	L2	CO1
15	a)	Explain the phases of data analytics: data collection, data cleaning, data exploration, data modeling, model evaluation, and deployment.	L2	CO2
	b)	Write in detail about variable classification, measurement scales, and their importance in selecting appropriate analytical methods.	L4	CO3

UNIT-III

S.No		Questions	BT	CO
1		Define regression. What is its purpose in statistical modeling?	L1	CO1
2		What are the BLUE properties in linear regression?	L1	CO2
3		State any four assumptions of the Ordinary Least Squares (OLS) method.	L2	CO3
4		What is Least Square Estimation?	L1	CO3
5		Define multicollinearity. Why is variable rationalization needed?	L2	CO1
6		What is meant by model building in regression?	L2	CO2
7		Define logistic regression. When is it preferred over linear regression?	L1	CO4
8		What are model-fit statistics? List any two.	L2	CO5
9		What is an odds ratio in logistic regression?	L3	CO3
10		Write any four business applications of regression modeling.	L1	CO2
		Part–B(LongAnswerQuestions)		
11	a)	Explain the concept of linear regression. Discuss dependent and independent variables, error terms, and the regression equation with examples.	L2	CO3
	b)	Describe the BLUE (Best Linear Unbiased Estimator) properties. Explain in detail the assumptions required for OLS estimators to be BLUE.	L4	CO1
12	a)	Discuss Least Square Estimation in detail. Explain how OLS estimates regression coefficients using minimization of error terms.	L3	CO2
	b)	Explain variable rationalization techniques such as correlation analysis, variance inflation factor (VIF), and stepwise regression.	L4	CO3
13	a)	Describe the process of model building. Explain model selection criteria such as AIC, BIC, R^2, adjusted R^2, and cross-validation.	L5	CO1
	b)	Explain the theory of logistic regression. Derive the logistic function and discuss sigmoid curve characteristics.	L2	CO3
14	a)	Discuss model fit statistics in logistic regression such as likelihood ratio test, Wald test, Hosmer–Lemeshow test, ROC curve, and AUC.	L4	CO4
	b)	Explain the steps in constructing a logistic regression model. Include data preprocessing, variable selection, training, validation, and interpretation.	L3	CO3
15	a)	Write about applications of regression and logistic regression in business domains such as marketing, finance, HR, risk analytics, and operations.	L2	CO1
	b)	Compare linear and logistic regression. Discuss differences in assumptions, output types, interpretation methods, and use cases.	L4	CO2

UNIT-IV

S.No		Questions	BT	CO
1		Define object segmentation. Why is it used in analytics?	L1	CO2
2		Differentiate between regression and segmentation.	L2	CO3
3		What is supervised learning? Give one example.	L1	CO2
4		What is unsupervised learning? Mention one technique.	L2	CO1
5		Define overfitting in decision trees.	L1	CO2
6		What is tree pruning? Why is it needed?	L2	CO1
7		Write two differences between regression trees and classification trees	L2	CO3
8		What are ARIMA models used for?	L1	CO1
9		List any two measures of forecast accuracy.	L3	CO2
10		What is STL decomposition in time series analysis?	L2	CO3
		Part–B(LongAnswerQuestions)		
11	a)	Explain object segmentation in detail. Discuss how segmentation differs from regression, and explain its importance in analytics tasks.	L2	CO1
	b)	Describe supervised and unsupervised learning approaches. Explain their use in segmentation, clustering, and classification tasks.	L2	CO2
12	a)	Explain the process of decision tree building for regression and classification. Discuss node splitting, impurity measures, and stopping rules.	L4	CO2
	b)	Discuss overfitting in decision trees. Explain how pruning and complexity parameters control overfitting with examples.	L5	CO1
13	a)	Explain multiple decision tree methods such as Random Forests, Bagging, and Boosting. Compare their advantages.	L4	CO2
	b)	Write in detail about ARIMA models. Explain identification, estimation, and diagnostic checking with relevant examples.	L2	CO3
14	a)	Discuss various measures of forecast accuracy including MAE, MAPE, RMSE, SMAPE, and MASE. Explain when each measure is appropriate	L4	CO1
	b)	Explain the STL (Seasonal-Trend-Loess) decomposition approach. Discuss how seasonality, trend, and remainder are extracted.	L2	CO2
15	a)	Describe how features like height, average energy, and other statistical features are extracted from a time series model. Explain how these features aid in prediction.	L3	CO2
	b)	Explain the complete workflow of using segmentation, tree-based models, and time series methods for predictive analytics in business domains. Include data preparation, model development, validation, and interpretation.	L4	CO1

UNIT-V

S.No		Questions	BT	CO
1		Define data visualization. Why is it important in analytics?	L1	CO1
2		What are pixel-oriented visualization techniques? Give one example.	L2	CO2
3		List two advantages of geometric projection-based visualization techniques	L2	CO3
4		What is the main idea behind icon-based visualization?	L1	CO1
5		Define hierarchical visualization. Give any one hierarchical visualization method.	L1	CO2
6		What is multidimensional data visualization?	L2	CO1
7		Write any two challenges in visualizing complex data.	L2	CO3
8		What are projection techniques in visualization? Provide one example.	L1	CO2
9		What is the role of interaction in visual analytics?	L3	CO3
10		Define relation visualization and give one example of relation graphs.	L2	CO1
Part–B(LongAnswerQuestions)				
11	a)	Explain pixel-oriented visualization techniques. Discuss their types, advantages, and limitations with examples.	L2	CO2
	b)	Describe geometric projection visualization techniques such as scatterplots, PCA projections, parallel coordinates, and multidimensional scaling.	L2	CO3
12	a)	Explain icon-based visualization techniques. Discuss Chernoff faces, star plots, and stick-figure icons with suitable illustrations.	L4	CO1
	b)	Discuss hierarchical visualization techniques such as dendrograms, tree-maps, sunburst charts, and radial tree visualization.	L4	CO3
13	a)	Explain how complex data relations such as social networks, association graphs, and entity-relationship visualizations are represented.	L4	CO3
	b)	Compare pixel-oriented, geometric-projection, icon-based, and hierarchical visualization techniques. Explain scenarios for using each.	L5	CO4
14	a)	Discuss visualization challenges in handling large-scale, high-dimensional, and heterogeneous data. Provide solutions using advanced visualization frameworks.	L4	CO2
	b)	Explain methods for visualizing multidimensional data: heatmaps, 3D plots, parallel coordinate plots, and dimensionality reduction-based projections.	L2	CO3
15	a)	Describe the process of designing an effective visualization: data selection, encoding, layout, color scheme selection, and interaction design.	L3	CO1
	b)	Explain techniques for visualizing complex relations using graph-based visualization: node-link diagrams, adjacency matrices, force-directed layouts, and clustering.	L4	CO3