Registration No.: | | | | | | | | | | |

Total Number of Pages: 02

Course: M.Tech
Sub_Code: 23PC1008

2nd Semester Regular Examination: 2024-25
SUBJECT: Advanced Data Visualization
BRANCH(S): CSEDS
Time: 3 Hours
Max Marks: 100
Q.Code: S291

Answer Question No.1 (Part-1) which is compulsory, any eight from Part-II and any two from Part-III.
The figures in the right hand margin indicate marks.

Part-I

Q1 **Answer the following questions:** **(2 x 10)**

a) What are the characteristics of effective data visualization?
b) Define Hierarchical Visualization.
c) Write name of some common tools used for data visualization.
d) How do you decide whether to use a bar chart or a line chart?
e) What do you mean by data annotation?
f) List the benefits of data visualization.
g) What is contour plot?
h) What are subplots?
i) Name the pixel-oriented visualization techniques.
j) Differentiate between interactive data visualizations and static charts.

Part-II

Q2 **Only Focused-Short Answer Type Questions- (Answer Any Eight out of Twelve)** **(6 x 8)**

a) Explain different types of data used in design process.
b) How do you deal with outliers in data visualization?
c) What is the significance of data cleansing in data visualization?
d) What is the difference between data and data visualization?
e) Explain Icon-Based Visualization Techniques with examples.
f) What are the differences between graphs and charts?
g) How can more than 3D be represented in a single chart?
h) Describe Confidence and Prediction intervals.
i) What do you mean by data reduction? Explain with example.
j) Explain Rank Analysis Tools in brief.
k) Describe data transformation techniques used in data processing.
l) What is a box plot? Explain how outliers are represented in a box plot.

Part-III

Only Long Answer Type Questions (Answer Any Two out of Four) **(16 x 2)**

Q3	Describe Geographical Analysis Tools for visualizing the complex data.	**(16)**
Q4	Discuss the techniques such as parallel coordinates, radar charts, and scatterplot matrices in detail.	**(16)**
Q5	Explain Multivariate Data Visualization with neat diagram.	**(16)**
Q6	Describe Residual analysis in visualization model building.	**(16)**

Registration No.: | 2 | 4 | 0 | 7 | 1 | 0 | 2 | 0 | 1 | 6 |

Total Number of Pages: 02

Course: M.Tech
Sub_Code: 23PC1007

2nd Semester Regular Examination: 2024-25
SUBJECT: Big Data Analysis
BRANCH(S): CSEDS
Time: 3 Hours
Max Marks: 100
Q.Code: S332

Answer Question No.1 (Part-I) which is compulsory, any eight from Part-II and any two from Part-III.
The figures in the right hand margin indicate marks.

Part-I

Q1 **Answer the following questions:** **(2 x 10)**

a) Define the term "Big Data".
b) What are the key characteristics of Big Data (5Vs)?
c) What is schema-less modeling in NoSQL?
d) Differentiate between key-value store and document store.
e) Write any two challenges in Big Data Analytics.
f) Mention the advantages of HDFS.
g) What is Pig Latin?
h) Define HiveQL.
i) Write two real-time applications of Spark Streaming.
j) What is sentiment analysis in Big Data context?

Part-II

Q2 **Only Focused-Short Answer Type Questions- (Answer Any Eight out of Twelve)** **(6 x 8)**

a) Explain the architecture requirements of Big Data Analytics systems.
b) Describe the types of NoSQL databases with examples.
c) Discuss the need for Big Data frameworks.
d) Compare and contrast traditional RDBMS with NoSQL databases.
e) What is MapReduce? Explain its working principle.
f) Illustrate the key features of Hadoop ecosystem.
g) Explain the basic components and commands used in Pig.
h) How are tables created and queried in Hive?
i) Write a short note on HBase and its usage.
j) Explain the concept of Stream Sampling.
k) Describe any one case study where Big Data analytics is used in E-Commerce.
l) What are the performance tuning techniques in HDFS?

Part-III

Only Long Answer Type Questions (Answer Any Two out of Four) **(16 x 2)**

Q3 Describe the evolution of Big Data and elaborate on the best practices used in Big Data Analytics. Also, explain different use cases highlighting the applications of Big Data. **(16)**

Q4 Discuss in detail the architecture and data model of a document store and a graph database. How are these databases useful in analyzing social media platforms like Twitter? **(16)**

Q5 Explain the working of Pig and Hive in detail. Compare their features, query language, and execution modes. Provide relevant examples of data operations. **(16)**

Q6 What is real-time analytics? Describe Spark's architecture for mining data streams. Discuss a case study on stock market prediction using real-time sentiment analysis. **(16)**

Registration No.: | 2 | 4 | 0 | 7 | 1 | 0 | 2 | 0 | 1 | 6 |

Total Number of Pages: 02

Course: M.Tech
Sub_Code: 23PC1005

2nd Semester Regular Examination: 2024-25
SUBJECT: Machine Learning
BRANCH(S): CSEDS
Time: 3 Hours
Max Marks: 100
Q.Code: S396

Answer Question No.1 (Part-1) which is compulsory, any eight from Part-II and any two from Part-III.
The figures in the right hand margin indicate marks.

Part-I

Q1 **Answer the following questions:** **(2 x 10)**

a) What are the objectives of machine learning.
b) Identify the challenges of the clustering algorithm.
c) Distinguish between Bagging and Boosting.
d) What is meant by cross-validation and resampling?
e) What is the role of a loss function?
f) Define support vectors in SVM.
g) Describe multilayer networks.
h) Give the formula for accuracy in classification problems.
i) What is the k in k-NN algorithm?
j) Differentiate between L1 and L2 regularization.

Part-II

Q2 **Only Focused-Short Answer Type Questions- (Answer Any Eight out of Twelve)** **(6 × 8)**

a) Explain the difference between Supervised, Unsupervised, and Reinforcement Learning with examples.
b) What is overfitting and underfitting? How can these issues be addressed?
c) What do you mean by Training, Validation, and Testing? Explain k-fold cross validation techniques.
d) What is Gradient Descent? Explain its types with advantages and disadvantages.
e) What do you mean by feature extraction? Explain the PCA for feature extraction.
f) Explain how Support Vector Machine can be used for classification of linearly separable data.
g) What are the steps involved in the k-means clustering algorithm? What are its limitations?
h) Why the estimate accuracy of Hypothesis is necessary? Explain the process to estimate the error between any two learning methods.

i) What is a ROC curve? Explain how it helps in evaluating classifier performance.

j) Explain the terms Gain And Entropy? How it is used to build the Decision tree in algorithm? Illustrate using an example.

k) What is kernel in SVM? Explain popular kernels used in SVM along with a scenario of their application.

l) Explain the principle of Maximum Likelihood Estimation (MLE) in the context of machine learning.

Part-III

Only Long Answer Type Questions (Answer Any Two out of Four) **(16 x 2)**

Q3 Discuss the differences, advantages, and disadvantages of various machine learning algorithms such as Linear Regression, Decision Trees, SVM, and Neural Networks. Provide suitable use cases. **(16)**

Q4 Describe the architecture and working of a multilayer perceptron (MLP). Explain forward propagation, backpropagation, and how neural networks learn using gradient descent. **(16)**

Q5 What is ensemble learning? Explain different ensemble techniques like bagging, boosting, and stacking. Discuss how ensemble methods improve predictive performance with examples. **(16)**

Q6 Explain Bayes' Theorem with an example and describe its application in pattern classification. Demonstrate why the Bayes decision rule yields the minimum error probability in multiclass classification problems. Also, explain the importance of the Radial Basis Function in machine learning. **(16)**

Registration No.: | 2 | 4 | 0 | 7 | 1 | 0 | 2 | 0 | 1 | 6 |

Total Number of Pages: 02

Course: M.Tech
Sub_Code: 23PC1006

2nd Semester Regular Examination: 2024-25
SUBJECT: RESEARCH METHODOLOGY, ETHICS AND IPR
BRANCH(S): CSEDS
Time: 3 Hours
Max Marks: 100
Q.Code: S491

Answer Question No.1 (Part-I) which is compulsory, any eight from Part-II and any two from Part-III.
The figures in the right hand margin indicate marks.

Part-I

Q1 **Answer the following questions:** **(2 x 10)**

a) What is the meaning of research?
b) Write about identification of research problem.
c) Write the condition for rejection of null hypothesis in hypothesis testing.
d) What is the conflicts of interest and how to avoid it in research?
e) How patent law helps the researcher?
f) What is non-literal infringement?
g) How long does copyright last? How copyright helps the researchers and scientists?
h) Write about role of claims.
i) What is criminal liability of copyright?
j) Write about the equivalent doctrine.

Part-II

Q2 **Only Focused-Short Answer Type Questions- (Answer Any Eight out of Twelve)** **(6 x 8)**

a) Explain about the Data Processing.
b) Discuss about Z-test and chi-square test.
c) Explain the predatory publishers and journals in research.
d) Write about design of questionnaire with suitable example.
e) Explain about Identification of publication misconduct, complaints, and appeals?
f) Discuss about the IPR in India genesis and development.
g) Write about the rights of patentee and working of patent.
h) Write transfer and termination of transfers of copyright with suitable examples.
i) What is patents granted with conditions?
j) Write about the rights covered by copyright,
k) Discuss about the right of publicity, federal preemption.
l) Write the use of intellectual property in research.

Part-III

Only Long Answer Type Questions (Answer Any Two out of Four) **(16 x 2)**

Q3	Explain in details about measurement and data collection: primary data, secondary data, also explain about the Discriminant Analysis	**(16)**
Q4	Write the best practices/ standards setting initiatives and guidelines: COPE, WAME, etc. Also discuss about violation of publication ethics.	**(16)**
Q5	Explain the restoration, surrender, and revocations of patents. Also write about layout of the patents	**(16)**
Q6	Explain in details the equivalent doctrine, pith and marrow doctrine. Also discuss the trade secret, contract, and misappropriation.	**(16)**

Registration No.: | 2 | 4 | 0 | 7 | 1 | 0 | 2 | 0 | 1 | 6 |

Total Number of Pages: 02

Course: M.Tech
Sub_Code: 23PE1005

2nd Semester Regular Examination: 2024-25
SUBJECT: Natural Language Processing
BRANCH(S): CSEDS
Time: 3 Hours
Max Marks: 100
Q.Code: S558

Answer Q1 (Part-1) which is compulsory, any eight from Part-II and any two from Part-III.
The figures in the right hand margin indicate marks.

Part-I

Q1 **Answer the following questions :** **(2 x 10)**

a) Differentiate between bigram and trigram.
b) Describe why production rules with zero probability are problematic
c) 3-grams are better than bigrams for part-of-speech tagging. Is it true or false? Explain your answer.
d) Information extraction is harder than text categorization. Is it true or false? Explain your answer
e) What do you understand by ambiguity?
f) List the various issues of machine translation system.
g) What are the problems with PCFG?
h) What do you mean by lemmatization?
i) What do you mean by stemmer?
j) What is the difference between phrase-based and feature-based NLP?

Part-II

Q2 **Only Focused-Short Answer Type Questions- (Answer Any Eight out of Twelve)** **(6 × 8)**

a) Discuss various relations among the word senses.
b) Discuss the relative advantages and disadvantages of partial versus full parsing.
c) What do you mean by anaphora resolution? Discuss its types and applications.
d) How to do a discourse analysis? Explain with example.
e) Write short note on WordNet.
f) Discuss stochastic part-of-speech tagging.
g) Discuss any two advanced applications of NLP.
h) How to deal with spelling error detection and correction? Explain.
i) Why is word sense disambiguation important for language technology? Explain with examples.
j) Explain how unification is implemented.
k) What is text summarization? Explain with an example.
l) Explain the application of semantics.

Part-III

Only Long Answer Type Questions (Answer Any Two out of Four) **(16 x 2)**

Q3	Explain the issues in computational morphology with suitable example.	**(16)**
Q4	Explain Earley parser and CYK parser with suitable examples.	**(16)**
Q5	Explain the natural language understanding (NLU) system architecture with neat sketch diagram.	**(16)**
Q6	Elaborate on the impact of Natural language processing in online search.	**(16)**