3711

B. Tech. (Computer Science & Engineering)

8th Semester (G-Scheme) Examination, May–2024

QUALITY ENGINEERING

Paper–PEC-ME-410-G/OEC-ME-410-G

Time allowed : 3 hours] *[Maximum marks : 75*

*Note : **Question No. 1 is compulsory.** Attempt total **five** questions selecting **one** question from each Unit. All questions carry equal marks.*

1. Explain the following terms: 6×2.5

   (a) Quality function
   (b) Hypergeometric distribution
   (c) ISO : 9000
   (d) JIT
   (e) Frequency Distribution
   (f) 5's approach

Unit-I

2. What do you meant by Quality? Explain the importance of Quality in Industry. Also explain quality characteristics. 15

3. Explain the concept of TQC and its implementation in details. 15

Unit-II

4. Explain the Statistical tools in Quality Control. 15

5. Explain the Binomial, Poisson and Normal Distribution with practical examples on the basics of sampling distribution. 15

Unit-III

6. What is Inspection system? Explain different forms of Inspection. 15

7. What is Standardization? Explain the need for Standardization. Also explain the history of ISO: 2000 series standard. 15

Unit-IV

8. What are basic concepts of TQM? Mention and explain the application of TQM in Industry. Also explain the role of senior management in TQM. 15

9. Explain the following terms: 2×7.5=15

   (a) Kaizen
   (b) Relevance of TQM
   (c) Kaizen

B.Tech. (Computer Science & Engineering),

8th Semester

(G-Scheme) Examination, May-2024

MACHINE LEARNING

Paper: PCC-CSE-402G

Time allowed : 3hours] *[Maximum marks : 75*

Note: *All **Questions** carry equal marks. **Question No. 1 is compulsory.** In addition to the compulsory question, students will have to attempt four more questions, selecting **one** question from each unit.*

1. Compulsory Question :

   (a) What are the three main types of Machine Learning classification ?

   (b) Define Dimensionality Reduction and why it is important in Machine Learning ?

   (c) Explain column vector.

   (d) How do you represent a dataset as a Matrix ?

   (e) Name two techniques used in Data preprocessing in Machine Learning.

(f) What is PCA, and how does it help with dimensionality reduction ?

Unit–I

2. What is machine learning ? Describe the features of Machine Learning algorithms and their importance in building predictive models.

3. Create a detailed block diagrammatic representation of learning machines, highlighting the key components and their roles in the Machine Learning process.

Unit–II

4. What is a dataset ? Explain with a suitable example. Demonstrate how to represent a dataset as a Matrix. Discuss the advantages of matrix representation in Machine Learning.

5. Describe the process of Data preprocessing in Machine Learning, focusing on Feature Normalization, Mean calculation, Column standardization, and Covariance estimation.

Unit–III

6. Define Supervised Learning and explain its working principle. Provide a step-by-step example of how a Supervised Learning algorithm processes training data and makes predictions.

7. Discuss the importance of labelled data in Supervised Learning and its role in training predictive models. Also, explain Decision Trees by taking a suitable example.

Unit–IV

8. Compare and contrast Boosting, Bagging, and Random Forests as Ensemble Methods in Unsupervised Learning. Discuss how these methods combine multiple models and improve the overall performance of unsupervised learning tasks ? Provide insights into scenarios where each ensemble method is most effective.

9. Describe the Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC) as evaluation metrics for binary classification models. How does the ROC curve visualize the trade-off between a true positive rate and a false positive rate ? What insights can be derived from the AUC value ?

3718

Tech. (Computer Science & Information Technology)

8th Semester, (G-Scheme) Examination, May–2024

BIG DATA ANALYSIS

Paper-PCC-CSE-404G/PCC-IT403G

me allowed : 3 hours] *[Maximum marks : 75*

*ote : **Question No. 1 is compulsory**. Attempt **five** questions in total selecting **one** question from each unit.*

Write short note on : 6×2.5=15

(a) Enlist 5 challenges associated with managing and analyzing large volumes of data.

(b) Compare and contrast RDBMS and NoSQL.

(c) What are different types of data models and their applications in organizing and structuring large datasets?

(d) What do you mean by Big Data Processing Pipeline?

(e) Discuss the challenges and considerations involved in designing and implementing distributed file systems for Big Data applications.

(e) Briefly elaborate the six V of big data.

Unit–I

2. Explain the characteristics of Big Data and discus they contribute to the challenges and opportuni managing large volumes of data.

3. (a) Describe the steps involved in the Data Sc process. How does each step contribu extracting value from Big Data?

   (b) Illustrate with a real-world scenario for involving Data Science process.

Unit–II

4. Compare and contrast Relational Database Manage System (RDBMS) and NoSQL database in the co of Big Data storage and Management. Discuss advantages and disadvantages of each approach.

5. (a) Explain the concept of Data lakes and how are different from data marts.

   (b) Discuss the role of ETL (Extract, Transform, L processes and data pipelines in building maintaining Data Lakes.

Unit–III

6. Explain the concept of scalability in the context of Data storage and Management systems. Discuss scalability challenges associated with traditional DB and how they differ those in Big Data Manage Systems.

7. Discuss the importance of data quality in Big Data Management. What are the key Challenges in ensuring data quality at scale? How can organizations address these challenges effectively? 15

Unit–IV

8. Describe the key components of Big Data processing pipelines. How do these components work together to ingest, process, and analyze large volumes of data? 15

9. Write short notes on :

(a) Hadoop 7

(b) Cassandra 8