RAAK

COLLEGE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, New Delhi & Affiliated to Pondicherry University)
An ISO 9001:2015 Certified Institution

VALUE ADDED COURSES
2021-2022
Department of Information Technology
21IT01- Programming for Data Science
MARK SHEET

Sl. No	Register Number	Student Name	Marks
1	18TH1002	S.BALASUBRAMANIAN	96
2	18TH1003	R.DEEPIKA	92
3	18TH1004	R.DIVYA	92
4	18TH1005	J. KALAIVANI	96
5	18TH1006	K. KANIMOZHI	96
6	18TH1008	A. MANIKANDAN	84
7	18TH1009	S.MANJU	96
8	18TH1010	S.MOHAMED SHAFEE	88
9	18TH1012	K.MUGUNDHAN	88
10	18TH1014	R.RANJITH	88
11	18TH1015	S. SANTHIYA	84
12	18TH1017	P. THAMIZHARASAN	96
13	18TH1019	A. VISHNUPRIYA	88

HOD

PRINCIPAL

Dr. S. SEENUVASAMURTHI, M.E., Ph.D.
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road,
Sulthanpet Post,
Puducherry - 605 110

VALUE ADDED COURSES

2021-2022

Department of Information Technology

21IT01 - Programming For Data Science

NAME :

CLASS :

DATE :

1. Which of the following is a core data type in Python?

    a) Class

    b) Dictionary

    c) List

    d) Tuple

    Answer: b) Dictionary

2. In Python, what is the output of time.time()?

    a) Current time in milliseconds

    b) Past 1 hour time

    c) Current time in milliseconds since January 1, 1970, GMT

    d) Current time in seconds since January 1, 1970, GMT

    Answer: d) Current time in seconds since January 1, 1970, GMT

3. Which of these actions identifies data properties?

    a) Data wrangling

    b) Data mining

    c) Data cleaning

d) Both a and b

Answer: b) Data mining

4. Why do you sample data for data analysis?

   a) To increase the dataset size

   b) To decrease the dataset size

   c) To decrease dimensionality

   d) To select a representative subset of data

   Answer: d) To select a representative subset of data

5. Which supervised learning algorithm is preferable for data classification?

   a) Random Forest

   b) k-Means

   c) Principal Component Analysis

   d) Hierarchical Clustering

   Answer: a) Random Forest

6. What is the main disadvantage of the K-means algorithm?

   a) It is computationally expensive

   b) It can get stuck in local minima

   c) It requires a large amount of labeled data

   d) It can only handle numerical data

   Answer: b) It can get stuck in local minima

7. What is the purpose of data munging?

   a) Evaluation of model performance

   b) Data visualizations

   c) Preparing data for analysis

d) Feature selection

Answer: c) Preparing data for analysis

8. Which of the following is used for dimensionality reduction?

   a) Naive Bayes

   b) k-Means

   c) Logistic Regression

   d) Principal Component Analysis (PCA)

   Answer: d) Principal Component Analysis (PCA)

9. What is a confusion matrix used for?

   a) A matrix that confuses the model

   b) A complex data structure

   c) A table used to describe the performance of a classification model

   d) A tool for data transformation

   Answer: c) A table used to describe the performance of a classification model

10. What is the critical factor in choosing an appropriate node during tree construction in decision trees?

    a) Attribute with the highest information gain

    b) Attribute with high information gain and entropy

    c) Attribute with the lowest information gain

    d) Attribute with high entropy

    Answer: a) Attribute with the highest information gain

11. What is 'Natural Language Processing' (NLP)?

    a) Processing and analyzing human language

    b) A method to process numeric data

Dr. S. SEENUVASAMURTHI, M.E., Ph.C.
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road,
Sulthanpet Post,
Puducherry - 605 110

c) The process of converting voice to text

d) A type of data visualization technique

Answer: a) Processing and analyzing human language

12. What is 'Cross-Validation' in machine learning?

a) A method for combining models

b) A way to validate the model on training data

c) A technique for assessing how the results of a statistical analysis will generalize to an independent data set

d) A method to visualize data

Answer: c) A technique for assessing how the results of a statistical analysis will generalize to an independent data set

13 Select the commonly used algorithm in data science regression:

A. Naive Bayes

B. k-Means

C. Logistic Regression

D. Decision Tree

Answer: C

14 . What does CLI stand for?

A. Common Language Interface

B. Command-line Interface

C. Computer Language Interaction

D. Command Language Instruction

Answer: B

Dr. S. SEENUVASAMURTHI, M.E., Ph.D.
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road,
Sulthanpet Post,
Puducherry - 605 110

15. Which Python data structure is responsible for the storage and manipulation of tabular data in Data Science?

A. Array

B. List

C. Dictionary

D. DataFrame

Answer: D

16. Which of these is a key characteristic of a hacker?

A. Afraid to say they don't know the answer

B. Willing to find answers on their own

C. Not willing to find answers on their own

D. All of the above

Answer: B

17. Which of the following focuses on the discovery of (previously) unknown properties on the data?

A. Data mining

B. BigData

C. Data wrangling

D. Machine Learning

Answer: A

18. Which of the following is a revision control system?

A. Git

B. Numpy

C. Scipy

D. Slidify

Dr. S. SEENUVASAMURTHI, M.E., Ph.D.
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road,
Sulthanpet Post,
Puducherry - 605 110

Answer: A

19. Which supervised learning algorithm is used for classification problems?

A. K-Means

B. Decision Tree

C. Principal Component Analysis

D. Gaussian Mixture Model

Answer: B

20. Which step in data analysis involves combining datasets to create a larger dataset?

A. Merging

B. Cleaning

C. Wrangling

D. Transforming

Answer: A

21. What does the "K" in K-Means algorithm represent?

A. Number of iterations

B. Number of attributes

C. Number of clusters

D. Number of data points

Answer: C

22. Which of the following is an unsupervised learning algorithm?

A. Linear Regression

B. Decision Tree

C. K-Means Clustering

D. Logistic Regression

Dr. S. SEENUVASAMURTHI, M.E., Ph.D.
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road,
Sulthanpet Post,
Puducherry - 605 110

Answer: C

23. Which type of analysis is commonly applied to census data?

A. Inferential analysis

B. Descriptive analysis

C. Causal analysis

D. Predictive analysis

Answer: B

24. What is the purpose of data munging?

A. Evaluation of model performance

B. Data visualizations

C. Preparing data for analysis

D. Feature selection

Answer: C

25. What does PCA stand for in data science?

A. Principal Component Analysis

B. Principal Coordinate Analysis

C. Principal Cluster Analysis

D. Principal Computation Analysis

Answer: A

Dr. S. SEENUVASAMURTHI, M.E., Ph.D.,
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road,
Sulthanpet Post,
Puducherry - 605 110

VALUE ADDED COURSES

2021-2022

Department of Information Technology

21IT01 - Programming For Data Science

NAME : S. MANJU

CLASS : IV /IT

DATE : 20/08/2021

96 %

1. Which of the following is a core data type in Python?

   a) Class

   b) Dictionary

   c) List

   d) Tuple

2. In Python, what is the output of time.time()?

   a) Current time in milliseconds

   b) Past 1 hour time

   c) Current time in milliseconds since January 1, 1970, GMT

   d) Current time in seconds since January 1, 1970, GMT

3. Which of these actions identifies data properties?

   a) Data wrangling

   b) Data mining

   c) Data cleaning

   d) Both a and b

Dr. S. SEENUVASAMURTHI, M.E., Ph.D.
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road.
Sulthanpet Post,
Puducherry - 605 110

4. Why do you sample data for data analysis?

    a) To increase the dataset size

    b) To decrease the dataset size

    c) To decrease dimensionality

    d) To select a representative subset of data

5. Which supervised learning algorithm is preferable for data classification?

    a) Random Forest

    b) k-Means

    c) Principal Component Analysis

    d) Hierarchical Clustering

6. What is the main disadvantage of the K-means algorithm?

    a) It is computationally expensive

    b) It can get stuck in local minima

    c) It requires a large amount of labeled data

    d) It can only handle numerical data

7. What is the purpose of data munging?

    a) Evaluation of model performance

    b) Data visualizations

    c) Preparing data for analysis

    d) Feature selection

Dr. S. SEENUVASAMURTHI, M.E., Ph.D.
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road.
Sulthanpet Post,
Puducherry - 605 110

8. Which of the following is used for dimensionality reduction?

   a) Naive Bayes

   b) k-Means

   c) Logistic Regression

   d) Principal Component Analysis (PCA)

9. What is a confusion matrix used for?

   a) A matrix that confuses the model

   b) A complex data structure

   c) A table used to describe the performance of a classification model

   d) A tool for data transformation

10. What is the critical factor in choosing an appropriate node during tree construction in decision trees?

    a) Attribute with the highest information gain

    b) Attribute with high information gain and entropy

    c) Attribute with the lowest information gain

    d) Attribute with high entropy

11. What is 'Natural Language Processing' (NLP)?

    a) Processing and analyzing human language

    b) A method to process numeric data

c) The process of converting voice to text

d) A type of data visualization technique

12. What is 'Cross-Validation' in machine learning?

   a) A method for combining models

   b) A way to validate the model on training data

   c) A technique for assessing how the results of a statistical analysis will generalize to an independent data set

   d) A method to visualize data

13 Select the commonly used algorithm in data science regression:

   A. Naive Bayes

   B. k-Means

   C. Logistic Regression

   D. Decision Tree

14 . What does CLI stand for?

   A. Common Language Interface

   B. Command-line Interface

   C. Computer Language Interaction

   D. Command Language Instruction

15. Which Python data structure is responsible for the storage and manipulation of tabular data in Data Science?

   A. Array

   B. List

   C. Dictionary

   D. DataFrame

16. Which of these is a key characteristic of a hacker?

   A. Afraid to say they don't know the answer

   B. Willing to find answers on their own

   C. Not willing to find answers on their own

   D. All of the above

17. Which of the following focuses on the discovery of (previously) unknown properties on the data?

   A. Data mining

   B. BigData

   C. Data wrangling

   D. Machine Learning

18. Which of the following is a revision control system?

   A. Git

   B. Numpy

   C. Scipy

   D. Slidify

Dr. S. SEENUVASAMURTHI, M.E., Ph.D.
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road
Sulthanpet Post

19. Which supervised learning algorithm is used for classification problems?

    A. K-Means

    B. Decision Tree

    C. Principal Component Analysis

    D. Gaussian Mixture Model

20. Which step in data analysis involves combining datasets to create a larger dataset?

    A. Merging

    B. Cleaning

    C. Wrangling

    D. Transforming

21. What does the "K" in K-Means algorithm represent?

    A. Number of iterations

    B. Number of attributes

    C. Number of clusters

    D. Number of data points

22. Which of the following is an unsupervised learning algorithm?

    A. Linear Regression

    B. Decision Tree

    C. K-Means Clustering

    D. Logistic Regression

Dr. S. SEENUVASAMURTHI, M.E., Ph.D..
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road.
Sulthanpet Post,
Puducherry - 605 11°

23. Which type of analysis is commonly applied to census data?

    A. Inferential analysis

    B. Descriptive analysis

    C. Causal analysis

    D. Predictive analysis

24. What is the purpose of data munging?

    A. Evaluation of model performance

    B. Data visualizations

    C. Preparing data for analysis

    D. Feature selection

25. What does PCA stand for in data science?

    A. Principal Component Analysis

    B. Principal Coordinate Analysis

    C. Principal Cluster Analysis

    D. Principal Computation Analysis

Dr. S. SEENUVASAMURTHI, M.E., Ph.C.
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road,
Sulthanpet Post,
Puducherry - 605 110

RAAK

COLLEGE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, New Delhi & Affiliated to Pondicherry University)
An ISO 9001:2015 Certified Institution

Certificate of Completion

2021-2022

This is to certify that Mr/Ms .... MANJU S ....

Year...... IV ...... Department...... IT .... has successfully Completed the Value added course.

SCORE: .... 96 ....

COURSE

TITLE: ...... PROGRAMMING FOR DATA SCIENCE ......

HOD

COURSE

DURATION: ...... 19-8-21 to 13-8-21 ......

Dr. S. SEENUVASAMURTHI, M.E., Ph.L.
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road,
Sulthanpet Post,
Puducherry - 605 110

PRINCIPAL

RAAK

COLLEGE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, New Delhi & Affiliated to Pondicherry University)
An ISO 9001:2015 Certified Institution

VALUE ADDED COURSES
2021-2022
Department of Information Technology
21IT02- Predictive Analytics
MARK SHEET

No	Register Number	Student Name	Marks
1	19TH0502	AGATHIAN M	88
2	19TH0503	AJEETH A	84
3	19TH0504	ARAVINDKRISHNA S	96
4	19TH0505	ARUN T	88
5	19TH0508	CHANDRAKUMAR. E	82
6	19TH0509	DEVIBALA E	88
7	19TH0510	DHIYA B	96
8	19TH0511	DIIIYA S	88
9	19TH0512	DIIIYASRI R	96
10	19TH0513	ELAMATHI M	92
11	19TH0514	FATHIMA AASMIN C	92
12	19TH0515	HARIHARAN A	96
13	19TH0516	HARIHARAN R	96
14	19TH0517	JOHNSON VASANTHARAJ. S	84
15	19TH0518	KALIMUTHU K	96

Dr. S. SEENUVASAMURTHI, M.E., Ph.D.
PRINCIPAL
RAAK College of Engineering & Technology
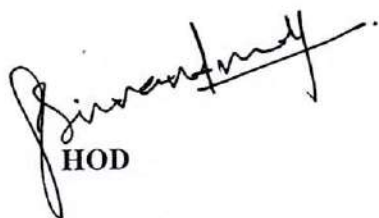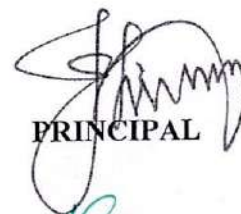No.1, Muthupillai Palayam Road,
Sulthanpet Post,
Puducherry - 605 110

16	19TH0519	KANNADASAN K	88
17	19TH0521	KIRUBA. B	88
18	19TH0522	MAHESH H	88
19	19TH0523	MUGILAN M	84
20	19TH0524	MURALIDHARAN V	96
21	19TH0525	NAVANEETHA KRISHNAN S	88
22	19TH0526	NITHYA R	96
23	19TH0527	NOORUDEEN M	92
24	19TH0528	PASUPATHI N	92
25	19TH0529	PUGAZHENDHI S	96
26	19TH0530	REVATHI B	96
27	19TH0531	SAMPATHKUMAR V	84
28	19TH0532	SANDHIYA P	96
29	19TH0533	SARANYA T	88
30	19TH0534	SARIGA T	88
31	19TH0535	SASIREKA M	88
32	19TH0536	SATHIYAVANI S	84
33	19TH0537	SELVA SUNDARAM K	96
34	19TH0538	SHEKANAS. K	88

Dr. S. SEENUVASAMURTHI, M.E., Ph.D.
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road,
Sulthanpet Post,
Puducherry - 605 110

35	19TH0539	SIIIABALAN J	84
36	19TH0540	SUJITH. K	88
37	19TH0541	SURIYA PRAKASH J	96
38	19TH0542	SWETHA K	92
39	19TH0543	THANNARASI. V	96
40	19TH0544	THENADAYALAN V	92
41	19TH0545	U. GEORRGE	96
42	19TH0546	UMAR SHARIFS M	92
43	19TH0547	VIJAYWILLSONDASS A	96

HOD

PRINCIPAL

Dr. S. SEENUVASAMURTHI, M.E., Ph.D.
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road,
Sulthanpet Post,
Puducherry - 605 110

VALUE ADDED COURSES

2021-2022

Department of Information Technology

21IT02 - Predictive Analytics

NAME :

CLASS :

DATE :

1. Predictive analytics uses statistics and _____ to determine future performance.

    A. Algorithmic techniques

    B. Modeling techniques

    C. System development and design techniques

    D. None of the mentioned above

Answer: B

2. Amongst which of the following is / are the applications of Predictive Analytics,

    A. Translating voice to text for mobile phone messaging

    B. Investment portfolio development

    C. Weather forecasts

    D. All of the mentioned above

Answer: D

3. Organizations are turning to predictive analytics to increase their bottom line and competitive advantage.

    A. True

    B. False

Answer: A

4. Business intelligence systems do not help businesses make better decisions by collecting, storing, analyzing and reporting on past data,

A. True

B. False

Answer: B

5. Companies use predictive analytics models to forecast inventory, manage resources, and operate more efficiently.

A. True

B. False

Answer: A

6. Organizations use data to predict when routine equipment maintenance will be required and can then schedule it before a problem or malfunction arises.

A. True

B. False

Answer: A

7. Predictive analytics is a process harnesses _____, often massive, data sets into models.

A. Heterogeneous

B. Storage

C. Network

D. None of the mentioned above

Answer: A

8. Amongst which of the following is / are the correct workflow of predictive analytics,

A. Import data → Clean the data → Develop a predictive model → Integrate the model

B. Clean the data → Develop a predictive model → Import data → Integrate the model

C. Clean the data → Develop a predictive model → Import data → Integrate the model

Dr. S. SEENUVASAMURTHI, M.E., Ph.C.
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road,
Sulthanpet Post,
Puducherry - 605 110

D. None of the mentioned above

Answer: A

9. Predictive analytics relies on capturing relationships between explanatory variables and the
_____.

    A. Predicted variables

    B. Descriptive variables

    C. Prescriptive variables

    D. All of the mentioned above

Answer: A

10. Amongst which of the following is / are the types of predictive analytics techniques,

    A. Predictive models

    B. Descriptive models

    C. Decision models

    D. All of the mentioned above

Answer: D

11. Machine learning predictive analytics is a category of algorithm that can receive input data and use statistical analysis to predict outputs,

    A. True

    B. False

Answer: A

12. Amongst which of the following is / are true with reference to predictive modeling,

    A. A tool used in predictive analytics

    B. A process that uses data mining and statistics to develop models

    C. Examine current and historical datasets for underlying patterns

Dr. S. SEENUVASAMURTHI, M.E., Ph.C.
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road,
Sulthanpet Post,
Puducherry - 605 110

D. All of the mentioned above

Answer: D

13. Predictive modeling is used throughout a range of industries, including meteorology, archaeology, automobile insurance, and algorithmic trading,

A. True

B. False

Answer: A

14. Machine learning in predictive analytics uses to enable computers to learn without being explicitly programmed by building algorithms that can receive _____ and use statistics to predict an output.

A. Input data

B. Output data

C. Process data

D. All of the mentioned above

Answer: A

15. Amongst which of the following is / are true with reference to Decision trees,

A. A learning model

B. Uses observations

C. Develop conclusions

D. All of the mentioned above

Answer: D

16. The IBM _____ analytics appliances combine high-capacity storage for Big Data with a massively-parallel processing platform for high-performance computing.

A) Watson

Dr. S. SEENUVASAMURTHI, M.E., Ph.C.
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road,
Sulthanpet Post,
Puducherry - 605 110

B) Netezza

C) InfoSight

D) LityxEQ

Answer: A

17. _____ is an integrated hosted analytics platform for marketing insights, predictive models, and marketing optimization"
A) LityxEQ

B) WatSon

C) LityxIQ

D) InfoSight

Answer: C

18. _____ is rapidly being adopted for computing descriptive and query types of analytics on Big data.

A) EDR

B) Hadoop

C) Azure

D) InfoSight

Answer: B

19. _____ involves predicting a response with meaningful magnitude, such as quantity sold, stock price, or return on investment.

A) Regression

B) Summarization

C) Clustering

D) Classification

Answer: A

20. Which of the following involves predicting a categorical response?

A) Regression

B) Summarization

C) Clustering

D) Classification

Answer: D

21. Which of the following contains pre-built predictive tools?

A) alteryx

B) fossil

C) paleoTS

D) ssas

Answer: A

22. _____ is proprietary tool for predictive analytics.

A) R

B) SAS

C) SSAS

D) EDR

Answer: B

23. Which of the following is preferred for text analytics?

A) R

B) Python

C) S

D) EDR

Answer: B

24. _____ is simplest class of analytics.

A) Descriptive

B) Predictive

C) Prescriptive

D) Summarization

Answer: A

25. _____ is a JavaScript charting library and feature-rich API set that lets you build interactive Flash or HTML5 charts.

A) InstantAtlas

B) Alterian

C) ZingChart

D) paleoTS

Answer: C

VALUE ADDED COURSES

2021-2022

Department of Information Technology

21IT02 - Predictive Analytics

NAME : ARAVINDKRISHNA.S

CLASS : IIP/IT

DATE : 20-8-2021

1. Predictive analytics uses statistics and ____ to determine future performance.

    A. Algorithmic techniques

    B. Modeling techniques

    C. System development and design techniques

    D. None of the mentioned above

$\frac{24}{25}$ 96%

2. Amongst which of the following is / are the applications of Predictive Analytics,

    A. Translating voice to text for mobile phone messaging

    B. Investment portfolio development

    C. Weather forecasts

    D. All of the mentioned above

3. Organizations are turning to predictive analytics to increase their bottom line and competitive advantage.

    A. True

    B. False

Dr. S. SEENUVASAMURTHI, M.E., Ph.D.
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road,
Sulthanpet Post,
Puducherry - 605 110

4. Business intelligence systems do not help businesses make better decisions by collecting, storing, analyzing and reporting on past data,

    A. True

    B. False

5. Companies use predictive analytics models to forecast inventory, manage resources, and operate more efficiently.

    A. True

    B. False

6. Organizations use data to predict when routine equipment maintenance will be required and can then schedule it before a problem or malfunction arises.

    A. True

    B. False

7. Predictive analytics is a process harnesses _____, often massive, data sets into models.

    A. Heterogeneous

    B. Storage

    C. Network

    D. None of the mentioned above

8. Amongst which of the following is / are the correct workflow of predictive analytics,

    A. Import data → Clean the data → Develop a predictive model → Integrate the model

    B. Clean the data → Develop a predictive model → Import data → Integrate the model

    C. Clean the data → Develop a predictive model → Import data → Integrate the model

    D. None of the mentioned above

Dr. S. SEENUVASAMURTHI, M.E., Ph.D.
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road,
Sulthanpet Post,
Puducherry - 605 110

9. Predictive analytics relies on capturing relationships between explanatory variables and the
_____.

A. Predicted variables

B. Descriptive variables

C. Prescriptive variables

D. All of the mentioned above

10. Amongst which of the following is / are the types of predictive analytics techniques,

A. Predictive models

B. Descriptive models

C. Decision models

D. All of the mentioned above

11. Machine learning predictive analytics is a category of algorithm that can receive input data and use statistical analysis to predict outputs,
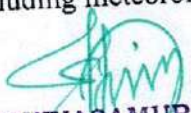
A. True

B. False

Answer: A

12. Amongst which of the following is / are true with reference to predictive modeling,

A. A tool used in predictive analytics

B. A process that uses data mining and statistics to develop models

C. Examine current and historical datasets for underlying patterns

D. All of the mentioned above

13. Predictive modeling is used throughout a range of industries, including meteorology, archaeology, automobile insurance, and algorithmic trading,

A. True

B. False

14. Machine learning in predictive analytics uses to enable computers to learn without being explicitly programmed by building algorithms that can receive _____ and use statistics to predict an output.

    A. Input data

    B. Output data

    C. Process data

    D. All of the mentioned above

15. Amongst which of the following is / are true with reference to Decision trees,

    A. A learning model

    B. Uses observations

    C. Develop conclusions

    D. All of the mentioned above

16. The IBM _____ analytics appliances combine high-capacity storage for Big Data with a massively-parallel processing platform for high-performance computing.

    A) Watson

    B) Netezza

    C) InfoSight

    D) LityxEQ

17. _____ is an integrated hosted analytics platform for marketing insights, predictive models, and marketing optimization"

    A) LityxEQ

    B) WatSon

    C) LityxIQ

    D) InfoSight

Dr. S. SEENUVASAMURTHI, M.E., Ph.C.
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road,
Sulthanpet Post,
Puducherry - 605 110

18. _____ is rapidly being adopted for computing descriptive and query types of analytics on Big data.

A) EDR

B) Hadoop

C) Azure

D) InfoSight

19. _____ involves predicting a response with meaningful magnitude, such as quantity sold, stock price, or return on investment.

A) Regression

B) Summarization

C) Clustering

D) Classification

20. Which of the following involves predicting a categorical response?

A) Regression

B) Summarization

C) Clustering

D) Classification

21. Which of the following contains pre-built predictive tools?

A) alteryx

B) fossil

C) paleoTS

D) ssas

Dr. S. SEENUVASAMURTHI, M.E., Ph.C.
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road,
Sulthanpet Post,
Puducherry - 605 110

22. _____ is proprietary tool for predictive analytics.

A) R

B) SAS

C) SSAS

D) EDR

23. Which of the following is preferred for text analytics?

A) R

B) Python

C) S

D) EDR

24. _____ is simplest class of analytics.

A) Descriptive

B) Predictive

C) Prescriptive

D) Summarization

25. _____ is a JavaScript charting library and feature-rich API set that lets you build interactive Flash or HTML5 charts.

A) InstantAtlas

B) Alterian

C) ZingChart

D) paleoTS

Dr. S. SEENUVASAMURTHI, M.E., Ph.D.
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road,
Sulthanpet Post,
Puducherry - 605 110

RAAK

COLLEGE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, New Delhi & Affiliated to Pondicherry University)

An ISO 9001:2015 Certified Institution

Certificate of Completion

2021-2022

This is to certify that Mr/Ms .... AJEETH A ....

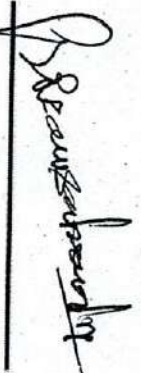Year..... III ..... Department...... II ...... has successfully Completed the Value added course.

SCORE: ...... 84 ......

COURSE
TITLE: ..... PREDICTIVE ANALYTICS .....

COURSE
DURATION: (9-8-21 to 13-8-21)

HOD

Dr. S. SEENUVASAMURTHI, M.E., Ph.E.
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road,
Sulthanpet Post,
Puducherry - 605 110

PRINCIPAL

RAAK

COLLEGE OF ENGINEERING AND TECHNOLOGY

VALUE ADDED COURSES
2021-2022
Department of Information Technology
21IT03 - Data Visualization and Presentation
MARK SHEET

Sl. No	Register Number	Student Name	Marks
1	20TH0251	ARUNADEVI K	88
2	20TH0252	ASHOKKUMAR R	84
3	20TH0253	DHARANIRASAN R	96
4	20TH0254	JESINTHA MARY D	88
5	20TH0255	LOGESH M	82
6	20TH0256	MOHAMED DHOWFIQ A	88
7	20TH0257	MOHANDASS T	96
8	20TH0258	MONIKA M	88
9	20TH0259	MUTHAZHAGAN S	96
10	20TH0260	PRETHESHWARAN S	92
11	20TH0261	PRIYANKA M	92
12	20TH0262	SAKTHI YUVARAJ V	96
13	20TH0263	SOWMIYA KK	96
14	20TH0266	YAZHINI K	84
15	20TD0921	RAAGA RAMYA DR	96

Dr. S. SEENUVASAMURTHI, M.E., Ph.D.,
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road,
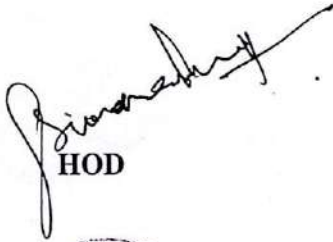Sulthanpet Post,
Puducherry - 605 110

16	20THL007	EZHILARASI R	88
17	20THL008	GOWSIKA A	88
18	20THL010	MALIKBASHA A	88
19	20THL011	THALAPATHIGUGAN G	84

HOD

PRINCIPAL

Dr. S. SEENUVASAMURTHI, M.E., Ph.D.
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road,
Sulthanpet Post,
Puducherry - 605 110

VALUE ADDED COURSES

2021-2022

Department of Information Technology

21IT03 - Data Visualization and Presentation

NAME :

CLASS :

DATE :

1.What is the primary purpose of data visualization in data science?

A. Summarizing data

B. Communicating insights

C. Cleaning data

D. Storing data

Answer: B

2. Which data visualization type is best suited for showing the distribution of a single continuous variable?

A. Bar chart
B. Scatter plot
C. Histogram
D. Pie chart

Answer: C

3. What is the term for a data visualization that shows the relationship between two continuous variables?

A. Heatmap

B. Scatter plot

C. Bar chart

D. Pie chart

Answer: B

4.Which color scheme is often used for sequential data in data visualizations to represent a progression from low to high values?

A. Rainbow

B. Viridis

C. Red-Blue

D. Random

Answer: B

5. What does the term "chartjunk" refer to in data visualization?

A. Non-essential or distracting elements in a chart

B. Accurate data representation

C. Data cleaning

D. Data analysis

Answer: A

6. Which data visualization type is suitable for comparing proportions of different categories in a dataset?

A. Scatter plot

B. Bar chart

C. Line chart

D. Pie chart

Answer: D

7. What is the benefit of using line charts in data visualization?

A. They show the distribution of a single numerical variable

B. They are useful for comparing multiple categories at once

Dr. S. SEENUVASAMURTHI, M.E., Ph.C.
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road,
Sulthanpet Post,
Puducherry - 605 110

C. They are effective for showing trends and changes over time

D. They display the parts of a whole in a circular format

Answer: C

8. Which data visualization is commonly used to show the relationship between two numerical variables?

    A. Bar chart

    B. Line chart

    C. Scatter plot

    D. Pie chart

    Answer: C

9. What is the benefit of using color in data visualization?

    A. To make the visualizations more monotonous

    B. To hide data points

    C. To highlight important data points and patterns

    D. To decrease the visibility of the visualizations

    Answer: C

10. Which of the following is not a limitation of parallel coordinate technique?

    A. Visual clutter

    B. Overlap

    C. Cannot visualize many data records

    D. Cannot visualize high dimension data

    Answer: D

11. Which method shows hierarchical data in a nested format?

    A. Treemaps

B. Scatter plots

C. Population pyramids

D. Area charts

Answer: A

12. How does interactive data visualization benefit data analysis?

A. It allows users to hide data points they don't like

B. It slows down data analysis

C. It enables users to explore and analyze data in real-time

D. It decreases the visibility of data visualizations

Answer: C

13. What is the purpose of providing data context in data visualization?

A. To make the data more confusing

B. To present the data in multiple formats

C. To allow users to interact with the data visualizations

D. To help users understand the significance and implications of the data

Answer: D

14. Which data visualization is suitable for showing the change in values over time?

A. Bar chart

B. Line chart

C. Scatter plot

D. Pie chart

Answer: B

Dr. S. SEENUVASAMURTHI, M.E., Ph.C.,
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road,
Sulthanpet Post,
Puducherry - 605 110

15. Which of the following lists names of variables in a data.frame in R?

A. par()

B. names()

C. barchart()

D. quantile()

Answer: B

16. What is the primary purpose of data visualization in data science?

A. Summarizing data

B. Communicating insights

C. Cleaning data

D. Storing data

Answer: B

17. Which data visualization type is best suited for showing the distribution of a single continuous variable?

A. Bar chart

B. Scatter plot

C. Histogram

D. Pie chart

Answer: C

18. What is the term for a data visualization that shows the relationship between two continuous variables?

A. Heatmap

B. Scatter plot

C. Bar chart

Dr. S. SEENUVASAMURTHI, M.E., Ph.C.
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road,
Sulthanpet Post,
Puducherry - 605 110

D. Pie chart

Answer: B

19. Which color scheme is often used for sequential data in data visualizations to represent a progression from low to high values?

   A. Rainbow

   B. Viridis

   C. Red-Blue

   D. Random

   Answer: B

20. What does the term "chartjunk" refer to in data visualization?

   A. Non-essential or distracting elements in a chart

   B. Accurate data representation

   C. Data cleaning

   D. Data analysis

   Answer: A

21. Which data visualization type is suitable for comparing proportions of different categories in a dataset?

   A. Scatter plot

   B. Line chart

   C. Pie chart

   D. Histogram

   Answer: C

22. Which of the following is a common tool for checking normality in data visualization?

   A. qqline()

B. qline()

C. anova()

D. lm()

Answer: A

23. Which data visualization method is best for showing hierarchical data in a nested format?

A. Treemaps

B. Scatter plots

C. Population pyramids

D. Area charts

Answer: A

24. What is the main advantage of using a line chart in data visualization?

A. Comparing multiple categories at once

B. Showing trends and changes over time

C. Displaying parts of a whole

D. Showing the distribution of a single variable

Answer: B

25. Which type of graph is used to show the relationship between three numerical variables?

A. Bar chart

B. Line chart

C. Scatter plot

D. Bubble chart

Answer: D

Dr. S. SEENUVASAMURTHI, M.E., Ph.C.
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road.
Sulthanpet Post.
Puducherry - 605 110

25. When the two sequences have substantial regions of similarity, many dots line up to form contiguous _____ lines.

a) crossings on

b) horizontal

c) diagonal

d) vertical

Answer: b

VALUE ADDED COURSES

2021-2022

Department of Information Technology

21IT03 - Data Visualization and Presentation

NAME : MOHANDASS. T

CLASS : II/IT

DATE : 20-8-2021

1. What is the primary purpose of data visualization in data science?

    A. Summarizing data

    B. Communicating insights

    C. Cleaning data

    D. Storing data

$\frac{24}{25}$ 96 /.

2. Which data visualization type is best suited for showing the distribution of a single continuous variable?

    A. Bar chart
    B. Scatter plot
    C. Histogram
    D. Pie chart

3. What is the term for a data visualization that shows the relationship between two continuous variables?

    A. Heatmap

    B. Scatter plot

    C. Bar chart

    D. Pie chart

Dr. S. SEENUVASAMURTHI, M.E., Ph.D.
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road,
Sulthanpet Post,
Puducherry - 605 110

4. Which color scheme is often used for sequential data in data visualizations to represent a progression from low to high values?

A. Rainbow

B. Viridis

C. Red-Blue

D. Random

5. What does the term "chartjunk" refer to in data visualization?

A. Non-essential or distracting elements in a chart

B. Accurate data representation

C. Data cleaning

D. Data analysis

6. Which data visualization type is suitable for comparing proportions of different categories in a dataset?

A. Scatter plot

B. Bar chart

C. Line chart

D. Pie chart

7. What is the benefit of using line charts in data visualization?

A. They show the distribution of a single numerical variable

B. They are useful for comparing multiple categories at once

C. They are effective for showing trends and changes over time

D. They display the parts of a whole in a circular format

Dr. S. SEENUVASAMURTHI, M.E., Ph.D.,
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road,
Sulthanpet Post,
110

8. Which data visualization is commonly used to show the relationship between two numerical variables?

   A. Bar chart

   B. Line chart

   C. Scatter plot

   D. Pie chart

9. What is the benefit of using color in data visualization?

   A. To make the visualizations more monotonous

   B. To hide data points

   C. To highlight important data points and patterns

   D. To decrease the visibility of the visualizations

10. Which of the following is not a limitation of parallel coordinate technique?

    A. Visual clutter

    B. Overlap

    C. Cannot visualize many data records

    D. Cannot visualize high dimension data

11. Which method shows hierarchical data in a nested format?

    A. Treemaps

    B. Scatter plots

    C. Population pyramids

    D. Area charts

12. How does interactive data visualization benefit data analysis?

    A. It allows users to hide data points they don't like

    B. It slows down data analysis

    C. It enables users to explore and analyze data in real-time

    D. It decreases the visibility of data visualizations

13. What is the purpose of providing data context in data visualization?

    A. To make the data more confusing

    B. To present the data in multiple formats

    C. To allow users to interact with the data visualizations

    D. To help users understand the significance and implications of the data

14. Which data visualization is suitable for showing the change in values over time?

    A. Bar chart

    B. Line chart

    C. Scatter plot

    D. Pie chart

15. Which of the following lists names of variables in a data.frame in R?

    A. par()

    B. names()

    C. barchart()

    D. quantile()

Dr. S. SEENUVASAMURTHI, M.E., Ph.C.
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road,
Sulthanpet Post,
Puducherry - 605 110

16. What is the primary purpose of data visualization in data science?

   A. Summarizing data

   B. Communicating insights

   C. Cleaning data

   D. Storing data

17. Which data visualization type is best suited for showing the distribution of a single continuous variable?

   A. Bar chart

   B. Scatter plot

   C. Histogram

   D. Pie chart

18. What is the term for a data visualization that shows the relationship between two continuous variables?

   A. Heatmap

   B. Scatter plot

   C. Bar chart

   D. Pie chart

19. Which color scheme is often used for sequential data in data visualizations to represent a progression from low to high values?

   A. Rainbow

   B. Viridis

   C. Red-Blue

   D. Random

Dr. S. SEENUVASAMURTHI, M.E., Ph.D.,
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road.
Sulthanpet Post,
Puducherry - 605 110

20. What does the term "chartjunk" refer to in data visualization?

A. Non-essential or distracting elements in a chart

B. Accurate data representation

C. Data cleaning

D. Data analysis

21. Which data visualization type is suitable for comparing proportions of different categories in a dataset?

A. Scatter plot

B. Line chart

C. Pie chart

D. Histogram

22. Which of the following is a common tool for checking normality in data visualization?

A. qqline()

B. qline()

C. anova()

D. lm()

23. Which data visualization method is best for showing hierarchical data in a nested format?

A. Treemaps

B. Scatter plots

C. Population pyramids

D. Area charts

Dr. S. SEENUVASAMURTHI, M.E., Ph.E.
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road,
Sulthanpet Post,
Puducherry - 605 110

24. What is the main advantage of using a line chart in data visualization?

    A. Comparing multiple categories at once

    B. Showing trends and changes over time

    C. Displaying parts of a whole

    D. Showing the distribution of a single variable

25. Which type of graph is used to show the relationship between three numerical variables?

    A. Bar chart

    B. Line chart

    C. Scatter plot

    D. Bubble chart

Dr. S. SEENUVASAMURTHI, M.E., Ph.C.
PRINCIPAL
RAAK College of Engineering & Technology
No.1, Muthupillai Palayam Road,
Sulthanpet Post,
Puducherry - 605 110

RAAK

COLLEGE OF ENGINEERING AND TECHNOLOGY

(Approved by AICTE, New Delhi & Affiliated to Pondicherry University)

An ISO 9001:2015 Certified Institution

Certificate of Completion

2021-2022

This is to certify that Mr/Ms ...... LOGESH M ....

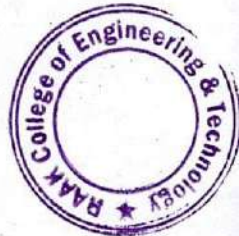Year...... II ...... Department...... II ...... has successfully Completed the Value added course.

SCORE: ...... 82 ......

COURSE

COURSE TITLE: DATA VISUALIZATION & PRESENTATION

DURATION: (9-8-21 to 13-8-21)

Dr. S. SEENUVASAMURTHI, M.E., Ph.C.
PRINCIPAL
RAAK College of Engineering & Technology
No.1. Muthupillai Palayam Road,
Sulthanpet Post,
Puducherry - 605 110

PRINCIPAL

HOD