PRESIDENCY UNIVERSITY
BENGALURU

SCHOOL OF ENGINEERING

TEST 1

Winter Semester: 2021-22

Course Code: PET 1003

Course Name: Data Analytics for Oil and Gas Exploration

Program & Sem: B.Tech and IV sem

Branch & Sem: PET and IV sem

Date: 27-04-2022

Time: 11:30 AM to 12:30 PM

Max Marks: 30

Weightage: 15%

Instructions:

(i) *Question Paper has THREE Parts, i.e. Part A, Part B, and Part C.*
(ii) *All the Questions should be answered.*
(iii) *Extra-time will not be given in any circumstances.*
(iv) *Read all the questions properly and answer accordingly.*

Part A [Memory Recall Questions]

Answer all the NINE Questions. Each question carries ONE mark. (9Qx1M=9M)

1. SEMMA process involves _____ of data.

(C.O. No. 1   Knowledge)

2. The 3 key ingredients to the Big Data Amalgam in Upstream Data Analysis are:

(C.O. No. 2   Knowledge)

3. Green field projects in Oil and Gas Industries are generally implemented by _____.

(C.O. No. 1   Knowledge)

4. Descriptive Modeling solves which of the question of Data Analytics?
A. What happened?   B. Why did it happen?
C. What will happen?   D. What's the best that could happen?

(C.O. No. 1   Knowledge)

5. The 3 Tenets of Upstream Data are:

(C.O. No. 1   Knowledge)

6. Logistic Regression in Data Analytics is used for _____ data.

(C.O. No. 1   Knowledge)

7. Predictive Modeling solves which of the question of Data Analytics?
A. What happened?   B. Why did it happen?
C. What will happen?   D. What's the best that could happen?

(C.O. No. 1   Knowledge)

8. DOFFs in oil and gas industry refers to _____.

(C.O. No. 1   Knowledge)

9. In exploration and production industry, the most common activation function used in the Artificial Neural Network is _____.

(C.O. No. 1   Knowledge)

Part B [Thought Provoking Questions]

Answer both the Questions. Each question carries THREE marks.

(3Q x 3M = 9M)

10. SEMMA process deals with inordinate amounts of data in a multivariate system.It is applicable across a variety of industries and provides methodologies for diverse business problems in all fields. Explain the significance of SEMMA process in Upstream Industry.   (C.O. No. 1 Comprehension)

11. Suppose an ANN model is developed for Positive Displacement Pump.Considering RPM,Flow Rate and Discharge Pressure as the input parameters to the model, what can be the output of the model and discuss the objective after solving the ANN problem?   (C.O. No. 1 Comprehension)

12. The term Greenfield vs Brownfield is not only related to oil & gas industries but is widely used in many industries like IT, Construction, Manufacturing, Building Services, etc. Seismic data are now becoming pivotal as 3D and 4D surveys are accelerated across green and brown fields. Compare Green Field Projects with respect to Brown Field Projects.   (C.O. No. 1   Comprehension)

Part C [Problem Solving Questions]

Answer both the Questions. Each question carries SIX marks.   (2Qx6M=12M)

13. Use the basic regression equation and comment on the parameters which could be taken to draw a basic multiple regression equation for viscosity of crude. Choose the parameter which is most sensitive with higher weighting coefficient.

(C.O. No. 1   Application)

14. Over 70 to 80 percent of time is consumed by managing and organizing the upstream data. Bad and duplicate data reduce your efficiency, increase your maintenance costs, and contribute to higher development costs for every Oil and Gas Project. Determine the sources of Bad Data. Determine the significant source which can be highly detrimental for working of an Oil and Gas project.

(C.O. No. 2   Application)

PRESIDENCY UNIVERSITY
BENGALURU

SCHOOL OF ENGINEERING

TEST 2

Winter Semester: 2021-22

Course Code: PET 1003

Course Name: Data Analytics for Oil and Gas Exploration

Program & Sem: B.Tech and IV Sem

Date: 2nd June 2022

Time: 11:30 AM to 12:30 PM

Max Marks: 30

Weightage: 15%

Instructions:
- *(v)   Question Paper has THREE Parts, i.e. Part A, Part B, and Part C.*
- *(vi) All the Questions should be answered.*
- *(vii) Extra-time will not be given in any circumstances.*
- *(viii) Read all the questions properly and answer accordingly.*

Part A [Memory Recall Questions]

Answer all the NINE Questions. Each question carries ONE mark.   (9Qx1M=9M)

Q1. Data that provides indirect measures of the property of interest with a significant degree of uncertainty is called _____.   (C.O. No. 2 Knowledge)

Q2. Sampling in which every item in the population has an equal chance of being chosen is _____.   (C.O. No. 2 Knowledge)

Q3. Sampling in which samples are equally spaced is known as ___.(C.O. No. 2 Knowledge)

Q4. Geologic age of rocks belongs to _____ data measurement type. (C.O. No. 2 Knowledge)
A. Nominal   B. Interval
C. Ordinal   D. Ratio

Q5. _____ party data involves collecting data from the website visits of a particular company.

                                                        (C.O. No. 2 Knowledge)

Q6. _____ data is well organized, digitalized, labeled and easily searchable along with defined data types.   (C.O. No. 2 Knowledge)

Q7. Which of the following data type measurement involves porosity , permeability and saturation?   (C.O. No. 2 Knowledge)

A.Nominal   B. Interval
C.Ordinal   D. Ratio

Q8. PODS in oil and gas industry refers to _____.   (C.O. No. 2
Knowledge)

Q9. PPDM in upstream data analytics refers to _____.   (C.O. No. 2
Knowledge)

Part B [Thought Provoking Questions]

Answer all the Questions. Each question carries THREE marks.   (3Q x 3M = 9M)

Q10. Major challenge for oil and gas digitalization is handling unstructured data. Unstructured data is not easily searchable and lack predefined labels. Explain the reasons which could be holding the upstream industry back in researching the unstructured data?
(C.O. No.2Comprehension)

Q11. Sample refers to the set of values and locations that have been measured. Sampling involves Random Sampling and Regular Sampling where the data is unbiased for the representation of the population. In Oil and Gas Industry , data is generally biased and prone to opportunity sampling. Describe the sources of sampling bias in upstream data which makes the sample less reliable and representative for explaining the entire population.
(C.O. No. 2   Comprehension)

Q12. The term Hard Data vs Soft Data is not only related to oil & gas industries but is widely used in many industries like IT, Construction, Manufacturing, Building Services, etc. Compare Hard data with respect to Soft data in upstream industry.
(C.O. No. 2   Comprehension)

Part C [Problem Solving Questions]

Answer both the Questions. Each question carries SIX marks.   (2Q x 6M = 12M)

Q13. The data for Facies F1, F2 and F3 in 10 wells is represented below:

Position	Well 1	Well 2	Well 3	Well 4	Well 5	Well 6	Well 7	Well 8	Well 9	Well 10
Top	F3	F2	F2	F1	F1	F1	F2	F2	F1	F1
Middle	F1	F1	F1	F2	F1	F2	F1	F2	F2	F2
Bottom	F2	F2	F2	F3	F3	F3	F3	F3	F3	F2

Event A represents that F1 is the middle facies. Similarly the other Event B represents that F3 is the bottom facies. Calculate and determine if the events are independent. What can be done in data cleansing step to mitigate if the problem of dependency exists in such data.
(C.O. No. 2   Application)

Q14. One in every thousand BOPs has a serious crack. X-ray analysis has a 99 % chance of detecting the crack correctly. If the BOP does not have a crack, there is a 2% chance that the X-ray detects a crack. The rate of BOP cracks is 0.1%. A BOP has been X-rayed and the result is positive. Determine the chance that the BOP actually does have a crack.

PRESIDENCY UNIVERSITY
BENGALURU

SCHOOL OF ENGINEERING

END TERM EXAMINATION

Winter Semester: 2021 - 22

Course Code: PET 1003

Course Name: Data Analytics for Oil and Gas Exploration

Program & Sem: B.Tech and IV Sem

Date: 4th July 2022

Time: 09:30 AM to 12:30 PM

Max Marks: 100

Weightage: 50%

Instructions:

(ix) Read the all questions carefully and answer accordingly.
(x) Question Paper has THREE Parts, i.e. Part A, Part B, and Part C.
(xi) All the Questions should be answered.
(xii) Extra-time will not be given in any circumstances.

Part A [Memory Recall Questions]

Answer all the Questions. Each question carries FIVE marks.
(4Qx 5M= 20M)

1. Select the correct option among the following questions:

A) Sampling in which every item in the population has an equal chance of being chosen is

i) Regular Sampling	ii) Random Sampling
iii) Biased Sampling	iv) Opportunity Sampling

(C.O.No.2) [Knowledge]

B) Which data provides indirect measures of the property of interest with a significant degree of uncertainty ?

i) Unstructured Data	ii) Structured Data
iii) Soft Data	iv) Hard Data

(C.O.No.2) [Knowledge]

C) The data that is well organized, digitalized, labeled and easily searchable along with defined data types is

i) Unstructured Data	ii) Structured Data
iii) Soft Data	iv) Hard Data

(C.O.No.2) [Knowledge]

D) Logistic Regression in Data Analytics is used for _____ data.

i) Numerical	ii) Nominal
ii) Ordinal	iv) Categorical

(C.O.No.1) [Knowledge]

E) Which party data involves collecting data from the website visits of a particular company?

 i) First Party　　　　　　　　　　　　ii) Second Party
 iii) Third Party　　　　　　　　　　　iv) Fourth Party

(C.O.No.2) [Knowledge]

2. Match the following:

A. Descriptive Modeling　　　　　　　　　　　　I) What happened?

B. Predictive Modeling　　　　　　　　　　　　II) What's the best that could happen?

C. Prescriptive Modeling　　　　　　　　　　　III) Why did it happen?

D. Business Intelligence and Operational Intelligence　　IV) What will happen?

E. Diagnostic Modeling　　　　　　　　　　　　V) What is going to happen?

(C.O.No.1) [Knowledge]

3. Fill in the blanks:

A. DOFFs in oil and gas industry refers to _____.
(C.O.No.1) [Knowledge]

B. The 3 Tenets of Upstream Data are:
(C.O.No.1) [Knowledge]

C. The 3 key ingredients to the Big Data Amalgam in Upstream Data Analysis are _____.

(C.O.No.1) [Knowledge]

D. PPDM in data analytics refers to _____.
(C.O.No.1) [Knowledge]

E. In exploration and production industry, the most common activation function used in the Artificial Neural Network is _____.
(C.O.No.1) [Knowledge]

4. Define the following :

A. Correlation
(C.O.No.4) [Knowledge]

B. Relief
(C.O.No.4) [Knowledge]

C. Lorenz Coefficient
(C.O.No.3) [Knowledge]

D. Coefficient of Variation
(C.O.No.3) [Knowledge]

E. Initial Data Analysis
(C.O.No.3) [Knowledge]

Part B [Thought Provoking Questions]

**Answer all the Questions. Each question carries FIVE marks.
(6Qx5M=30M)**

5. The term Greenfield vs Brownfield is not only related to oil & gas industries but is widely used in many industries like IT, Construction, Manufacturing, Building Services, etc. Seismic data are now becoming pivotal as 3D and 4D surveys are accelerated across green and brown fields. Compare Green Field Projects with respect to Brown Field Projects.   (C.O.No.1) [Comprehension]

6. SEMMA process deals with inordinate amounts of data in a multivariate system.It is applicable across a variety of industries and provides methodologies for diverse business problems in all fields. Explain the significance of SEMMA process in Upstream Industry.
                    (C.O.No.1) [Comprehension]

7. Sample refers to the set of values and locations that have been measured. Sampling involves Random Sampling and Regular Sampling where the data is unbiased for the representation of the population. In Oil and Gas Industry , data is generally biased and prone to opportunity sampling. Describe the sources of sampling bias in upstream data that makes the sample less reliable and representative for explaining the entire population.   (C.O.No.2) [Comprehension]

8. The term Hard Data vs Soft Data is not only related to oil & gas industries but is widely used in many industries like IT, Construction, Manufacturing, Building Services, etc. Compare Hard data with respect to Soft data in Oil and Gas industry.
        (C.O.No.2) [Comprehension]

9. Virtually all the subsurface spatial data is sampled in a biased manner. It is almost impossible to use the raw statistics from these datasets to support decision making. There is a need, however, to adjust the histograms and summary statistics to be representative of the entire volume of interest. Classify the techniques that can be used as solutions to biased spatial data.

                                    (C.O.No.3)

[Comprehension]

10) A Pareto chart is a type of chart that contains both bars and a line graph.The Pareto chart is one of the basic tools of quality control. The left vertical axis in it represents the frequency of occurrence and the right vertical axis in it represents the total number of occurrences. Explain the importance of Pareto Charts in Real Time Drilling Optimization.   (C.O.No.4) [Comprehension]

Part C [Problem Solving Questions]

**Answer all the Questions. Each question carries TEN marks.
(5Qx10M=50M)**

11. Use the basic regression equation and comment on the parameters which could be taken to draw a basic multiple regression equation for viscosity of crude. Choose the parameter which is most sensitive with higher weighting coefficient.
            (C.O.No. 1) [Application]

12. When drilling through a specific formation there have been an average of 3 fluid loss incidents. A) Determine the probability of 5 fluid loss incidents for next well. Also, compute the probability of no fluid loss incidents.

B) Comment on the type of distribution which follows when such problems are encountered in Drilling.

(C.O.No. 4)

[Application]

13. One in every thousand BOPs has a serious crack. X-ray analysis has a 99 % chance of detecting the crack correctly. If the BOP does not have a crack, there is a 2% chance that the X-ray detects a crack. The rate of BOP cracks is 0.1%. A BOP has been X-rayed and the result is positive. Determine the chance that the BOP actually does have a crack.   (C.O.No. 2) [Application]

14. The data for Facies F1, F2 and F3 in 10 wells is represented below:

Position	Well 1	Well 2	Well 3	Well 4	Well 5	Well 6	Well 7	Well 8	Well 9	Well 10
Top	F3	F2	F2	F1	F1	F1	F2	F2	F1	F1
Middle	F1	F1	F1	F2	F1	F2	F1	F2	F2	F2
Bottom	F2	F2	F2	F3	F3	F3	F3	F3	F3	F2

Event A represents that F1 is the middle facies. Similarly the other Event B represents that F3 is the bottom facies. Calculate and determine if the events are independent. What can be done in data cleansing step to mitigate if the problem of dependency exists in such data.
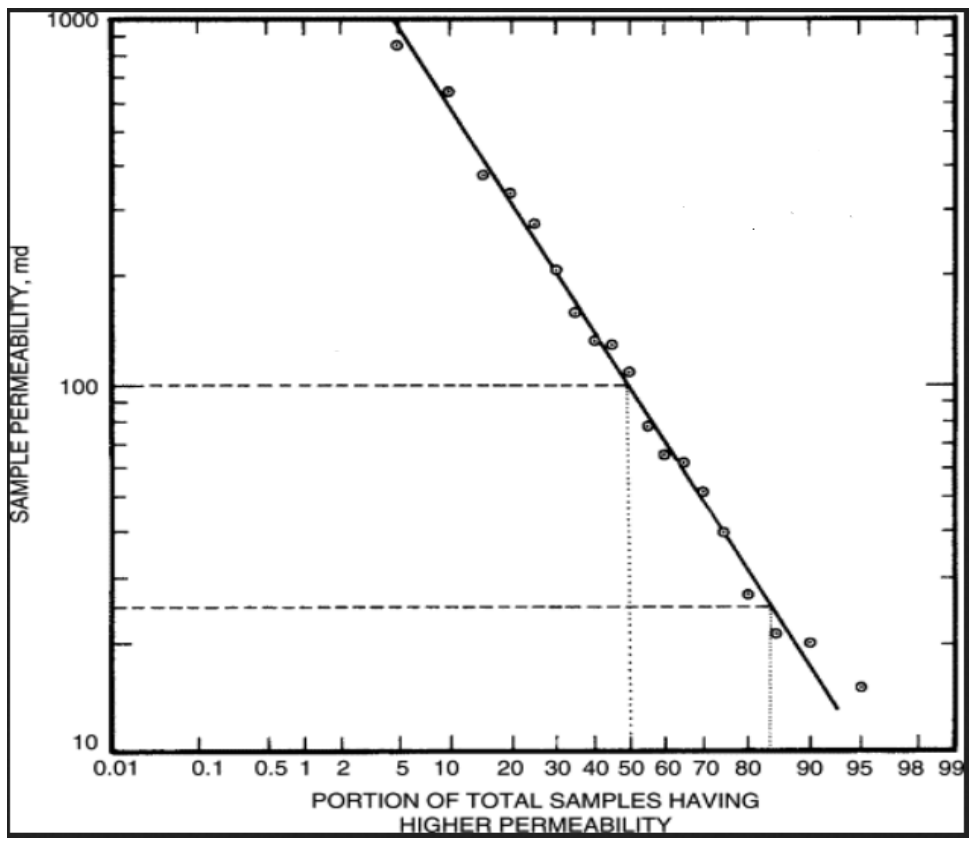
(C.O.No. 2)

[Application]

15. Dykstra Parsons coefficient represents the measure of spread or dispersion of permeability. It provides the subsurface asset team with an initial assessment of heterogeneity encountered by a new well.

A) From the figure shown below, estimate the Dykstra Parsons Coefficient and comment on the value estimated for the heterogeneity of the reservoir.

B) Assume that there are 2 wells . One well is the production well from which the oil is produced. The other well is an injector well , where water is injected to displace oil. After calculating the Dykstra Parsons coefficient from the above, what can happen to such type of reservoir if oil is being swept by water. Justify and comment on the oil production rate.

(C.O.No.

3) [Application]