

Name:

Enrolment No:



UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
END Semester Examination, December 2021

Programme Name: B.Tech. Non CIT
Course Name : Data Science Tools
Course Code : MRDA0203
Nos. of page(s) : 08

Semester : V
Time : 03 hrs
Max. Marks: 100

SECTION A [20 marks]

S. No.		Marks	CO
Q 1.	What is the best process for contributing a Bugfix to a foreign repository?		
(A)	(a) Ask the repository owner for write access to the repository. (b) Send the fix via email to the author. (c) Fork the repository, update the fork and create a pull request. (d) All of them		
(B)	Which of the following statements about repositories are correct? (Select all that apply.) (a) The remote repository is only accessible by myself. (b) The local repository is only accessible by myself. (c) The staging is only accessible by myself. (d) The remote repository is accessible by all contributors. (e) The local repository is accessible by all contributors.	2+2	CO4
Q 2.	GitHub in simple words is a collection of folders and files. It is a Git repository hosting service, but it adds many of its own features. Enlist at least 04 such features supported by GitHub.	4	CO1
Q 3.	As the merge command is used the new create branch such that it will be merged to the master branch and the file will be inserted to it. Write a code in GitHub to execute the aforementioned statement.	4	CO2

Q 4.			
(a)	Write the Syntax of opening a “Pull Request” in GitHub.	2+2	CO2
(b)	Describe the Pre-requisites for working with GitHub on local machine:		
Q5.			
(A)	Fill in the blank: It's a best practice to remove or replace _____ before publishing to GitHub. (a) code cells (b) credentials (c) markdown text (d) charts	2+2	CO3
(B)	Fill in the blank: When working in a Jupyter Notebook, before returning to a project, it's important to _____. (a) Insert cells. (b) Insert to code. (c) Run cells. (d) Save your notebook.		
SECTION B [40 marks]			
Q 6.	Data Refinery provides which of the following services?		
(A)	(a) Catalog the data assets. (b) Monitor for bias and model drift. (c) Visualize and prepare data. (d) Automatically build models. (e) All of the above.		
(B)	IBM SPSS Modeler includes what kind of models? (a) Classification models (for data with a categorical target). (b) Regression models (for data with a continuous target). (c) Clustering models (for data with no target variables). (d) Other kinds of models. (e) All of the above.	10	CO2
(C)	Open Neural Network eXchange (ONNX) was originally created for what models?		

	<p>(a) Clustering models (b) Support Vector Machines (SVM). (c) Regression models (d) Decision trees (e) Deep learning models.</p>		
(D)	<p>Fill in the blank: If you'd like to schedule a notebook in Watson Studio to run at a different time, you can create a(n) _____.</p> <p>(a) markdown cell (b) API (c) Job (d) asset</p>		
(E)	<p>Fill in the blank: In the _____ tab you can define the hardware size and software configuration for the runtime associated with Watson Studio tools such as Notebook.</p> <p>(a) Assets (b) Overview (c) Settings (d) environments</p>		
Q 7.	<p>Discuss the benefits of IBM Watson Studio and justify your answer via correlating its features w.r.t. Extensive open source frameworks.</p>	10	CO4
Q8.	<p>When learning data science what open source tools are the mostly used? Justify your response with suitable example.</p>	10	CO3
Q 9.	<p>(A) Which products (of those we covered) allow you to build data pipelines using graphical user interface and no coding?</p> <p>(a) Only IBM SPSS Statistics. (b) Only IBM SPSS Modeler. (c) OpenScale (d) IBM SPSS Modeler and Modeler Flows in Watson Studio. (e) All of the above.</p> <p>(B) Which features of Data Refinery help save hours and days of data preparation?</p>	10	CO4

	<p>(a) Flexibility of using Intuitive user interface and coding templates enabled with powerful operations to shape and clean data.</p> <p>(b) Data visualization and profiles to spot the difference and guide data preparation steps.</p> <p>(c) Incremental snapshots of the results allowing the user to gauge success with each iterative change.</p> <p>(d) Saving, editing and fixing the steps provide ability to iteratively fix the steps in the flow.</p> <p>(e) All of the above.</p>		
(C)	<p>Watson Knowledge Catalog provides what functionality?</p> <p>(a) Build data and water pipelines.</p> <p>(b) Create data and deploy models into production.</p> <p>(c) Catalog all books mentioning Doctor Watson and Sherlock Holmes.</p> <p>(d) Catalog data and ML assets, help to find relevant assets, keep track of asset lineage, enforce data governance.</p> <p>(e) Process data, build and deploy models.</p>		
(D)	<p>Fill in the blank: PMML, PFA, and ONNX are _____.</p> <p>(a) Passwords for some super-secret system.</p> <p>(b) Abbreviations for machine learning algorithm names.</p> <p>(c) Robots that are plotting to take over the planet.</p> <p>(d) Codes for getting rid of undesired data or models.</p> <p>(e) Open standards for predictive model serialization, exchange, and deployment.</p>		
(E)	<p>Which node must be used in Modeler flows before any modeling node?</p> <p>(a) Output node</p> <p>(b) Type node</p> <p>(c) Derive node</p> <p>(d) Auto Numeric node</p> <p>(e) All of the above</p>		

SECTION-C [40 marks]

<p>Q 10.</p> <p>(I)</p>			
<p>(A)</p>	<p>Fill in the blank: Auto Classification node can be used for data with _____.</p> <p>(a) No target variables. (b) A categorical target variable. (c) A continuous target variable. (d) Any target variable. (e) All of the above</p>		
<p>(B)</p>	<p>Fill in the blank: Auto Numeric node can be used for data with _____.</p> <p>(a) No target variables. (b) A categorical target variable. (c) A continuous target variable. (d) Any target variable. (e) All of the above.</p>		
<p>(C)</p>	<p>IBM SPSS Modeler evolved from which product?</p> <p>(a) Clementine (b) Oracle (c) IBM DB2 (d) SPSS (e) Netezza</p>	<p>10</p>	<p>CO2</p>
<p>(D)</p>	<p>Fill in the blank: IBM SPSS Statistics syntax can be created using _____.</p> <p>(a) Graphical user interface of IBM SPSS Statistics product or syntax editor. (b) IBM SPSS Modeler streams. (c) OpenScale (d) AutoAI (e) Watson Studio Modeler flows.</p>		

<p>(E)</p>	<p>AutoAI provides which of the following services?</p> <ul style="list-style-type: none"> (a) Monitoring for fairness, bias, and model drift. (b) Automatic finding of optimal data preparation steps, model selection, and hyper-parameter optimization. (c) Cataloging data and model assets. (d) Creating SPSS syntax. (e) All of the above. 		
<p>(II)</p> <p>(A)</p> <p>(B)</p> <p>(C)</p> <p>(D)</p>	<p>SQL is what type of query language?</p> <ul style="list-style-type: none"> (a) Cypher (b) Database (c) Contextual (d) Access <p>Is Keras a machine learning or deep learning library for Python?</p> <ul style="list-style-type: none"> (a) Yes (b) No <p>What type of model would you use if you wanted to find the relationship between a set of variables?</p> <ul style="list-style-type: none"> (a) Regression model (b) Classification model (c) Clustering model <p>What is data governance?</p> <ul style="list-style-type: none"> (a) An implementation of data processes and controls (b) Deleting data stored within a central repository (c) Storing data in a central repository (d) Creating processes and controls around the access of data 	<p>10</p>	<p>CO1</p>

(E)	<p>Jupyter Notebooks is the tool most R developers use.</p> <p>(a) Yes (b) No</p>		
11	<p>Brief the following:</p> <p>(I) (i) Cloud based Tools for Data Science (ii) Data Refinery (iii) IBM SPSS Modeler (iv) Packages (v) AutoAI</p>	10	CO3
(II)	<p>(A) Which of the following functions does R Studio provide?</p> <p>(a) Creating relationships between data tables. (b) Storing data in tables. (c) Documenting R code applications. (d) Editing and execution of R code.</p> <p>(B) What type of file format is a Jupyter Notebook?</p> <p>(a) .xlsx (b) .doc (c) .ipynb (d) .ppt</p> <p>(C) What capability monitors and manages models to operate trusted AI?</p> <p>(a) Watson Knowledge Catalog (b) Modeler Flows (c) AutoAI (d) OpenScale</p> <p>(D) Creating a scheduled Job and using a custom environment to run a data flow/pipeline on different workloads can help build what?</p> <p>(a) APIs with automation (b) Fixed workloads (c) Not supported (d) Repeatable Data Pipelines for workloads of almost any size</p>	10	CO4

(E)	What type of node is used to prepare the data for modeling in Modeler flows? (a) Auto data prep node (b) A modeling node (c) All of the above (d) A data source node (e) A type node		
------------	---	--	--