تطويرتطبيقات نظم المعلومات

10:12

6/9/2021

أ.د/تيسيرحسن

1. الإسم الرباعي (بالعربي فقط) *

> ابرام صفوت برتي نظير

2. رقم الجلوس *

> 162016003

3. المستوي *

- ⭘ الاول

- ⭘ الثاني

- ⭘ الثالث

○ الرابع 2013

○ الرابع 2014

○ الرابع 2015

◉ الرابع 2016

○ الرابع 2017

4. البرنامج *

◉ عام

○ بايو

○ هندسة

5. رقم المعمل *

٣ج	⌄

6. رقم الكمبيوتر *

19

7. الكود (قد تمت مراجعة بيانات الطالب ورقم الجلوس) *

Farouk

8. A data scientist:
   (1 Point)

○ a. Has strong analytical skills,

○ b. Has a combination of skills that enable them to handle raw, unstructured data

○ c. Applies complex analytical techniques at scale

◉ d. All of the above

9. A corpus is:
(1 Point)

○ a. Tweets

◉ b. A large collection of text documents.

○ c. A dictionary of terms

○ d. A collection of images

10. Virtualization uses:
(1 Point)

◉ a. Software to create an abstraction layer over computer hardware that allows the hardware elements of a single computer—processors, memory, storage and more—to be divided into multiple virtual computers, commonly called virtual machines (VMs).

○ b. Frequent electronic copying data from a database in one computer or server to a database in another so that all users share the same level of information.

○ c. Electronic copying of the data

○ d. Distributing the data

11. In Model planning phase, you perform the following steps **EXCEPT:**
(1 Point)

○ a. Data Exploration

○ b. Inputs from stakeholders

c. Model selection

d. List Data resources

12. Regular expression is a technique for:
(1 Point)

a. Clustering

b. Classification

c. Finding words, strings or particular patterns in the text

d. Association rule mining

13. A typical example of sentiment analysis is:
(1 Point)

a. Buying a specific product

b. Buying certain items together in a supermarket

c. Buying certain items together in a specific region

d. Determining a geographical region where a specific writer lives

14. You formulate your initial hypothesis in the following phase:
(1 Point)

a. Data preparation

b. Model Planning

c. Discovery

d. Model building

15. Possible tools for the data preparation phase are:
(1 Point)

- ⦿ a. Hadoop, R, and Mining tools

- ○ b. DBMS

- ○ c. Weka

- ○ d. Octave

16. Role of the data engineer is responsible for:
(1 Point)

- ○ a. Core business of the project

- ⦿ b. Deep technical skills to assist tuning technical skills for data management

- ○ c. Configures database management

- ○ d. Analytical techniques

17. Quasi-structured data are:
(1 Point)

- ○ a. Text data

- ○ b. Relational tables

- ⦿ c. Textual data with erratic data formats, such as Web clickstream data

- ○ d. None of the above

18. Focus of data analytics life cycle is:
(1 Point)

- ○ a. On business intelligence

b. On data science projects

c. On stakeholders

d. Communicating results

19. Hadoop has the following advantages **EXCEPT:**
(1 Point)

a. Expensive machines

b. Notion of jobs

c. Any data will fit in any format

d. Scalable and flexible

20. The JobTracker node is responsible for:
(1 Point)

a. Operating the Mapper and Reducer functions to available TaskTrackers and monitoring the results, while TaskTracker nodes actually run the jobs and communicate results back to the JobTracker.

b. Distributing the Mapper and Reducer functions to available TaskTrackers and monitoring the results, while TaskTracker nodes actually run the jobs and communicate results back to the JobTracker.

c. Distributing the Mapper function to available TaskTrackers and monitoring the results, while TaskTracker nodes actually run the jobs and communicate results back to the JobTracker.

d. Distributing the Mapper and Reducer functions to available TaskTrackers and monitoring the results, while TaskTracker nodes actually run the jobs and communicate results back to the NameNode.

21. Obtaining the objective of the analytics determines:
(1 Point)

a. The preprocessing phase

○ b. The model planning phase

○ c. The model building phase

○ d. The data attributes you need

22. Hadoop deals with:
(1 Point)

○ a. Data schema

○ b. Tables and relations

● c. Flat files in any format

○ d. Data fragmentation

23. HDFS works by:
(1 Point)

● a. A Name Node and a Data Node

○ b. A Name Node

○ c. A Data Node

○ d. Block mappings

24. Velocity of the data means:
(1 Point)

○ a. Uncertainty of the data

○ b. Real-time data

● c. Data is being generated fast and need to be processed fast

○ d. Online transaction processing

25. In the model building phase, you perform the following:
(1 Point)

- ○ a. Interpret the results

- ● b. Answer the question is model accurate enough

- ○ c. Identify key findings

- ○ d. Quantify business values

26. Heterogeneous data, such as clinical and x-rays, is required in Big data analytics in healthcare in order to:
(1 Point)

- ○ a. Provide better decision making for the patient

- ○ b. Provide visualization of each part of the data

- ○ c. Apply big data analytics life cycle

- ● d. All of the above

27. HDFS does the following to solve the problem when the end machines are too expensive:
(1 Point)

- ○ a. Run on brands hardware

- ● b. Run on commodity hardware

- ○ c. Replicate the data

- ○ d. Tune the data

28. Presentation for Project Sponsors contains:
(1 Point)

- ● a. Easy visuals for the presenter to explain and for the viewer to grasp.

b. Changes to business processes and reports.

c. Technical graphs (such as Receiver Operating Characteristic [ROC] curves, density plots, and histograms)

d. Technical reports

29. The following statements about the operationalize phase of the big data analytics life cycle are true
EXCEPT:
(1 Point)

a. Assess the benefits

b. Run a pilot

c. Model execution in production environment

d. Quantify the business value ●

30. An example of Unstructured data is:
(1 Point)

a. Click-Streams

b. Text data

c. XML data ●

d. Tables

31. MapReduce is responsible for:
(1 Point)

a. Storage

b. SQL

c. Computation ●

    ○ d. Columnar database

32. Major challenges of text analysis are:
    (1 Point)

    ● a. High dimensionality and unstructured text

    ○ b. High dimensionality and structured text

    ○ c. Unstructured text and volume

    ○ d. Uncertainty and velocity

33. Big data characteristics include:
    (1 Point)

    ○ a. Volume

    ○ b. Variety

    ○ c. Veracity

    ● d. All of the above

34. The following resources are required to complete the discovery phase:
    (1 Point)

    ○ a. Available technology

    ○ b. Available data

    ○ c. People for the working team

    ● d. All of the above

35. Precision is defined as:
    (1 Point)

a. The percentage of returned documents among all relevant documents in the corpus.

b. The percentage of documents in the results that are relevant.

c. The accuracy

d. Support

36. The objective of the Reduce function is:
(1 Point)

a. To apply a recommender system on the mapper data

b. To apply association rule mining on the mapper data

c. To combine the input from the Mapper

d. To visualize the mapper data

37. The objective of the Reduce function is:
(1 Point)

a. To combine the input from the Mapper

b. To apply a recommender system on the mapper data

c. To apply association rule mining on the mapper data

d. To visualize the mapper data

38. The goals of model building phase in big data life cycle include:
(1 Point)

a. Develop datasets for training and testing and production purposes.

b. Get the environment you can for building models and workflows

c. a &b

d. Visualize results

39. The goal of discovery phase in big data analytics life cycle is to determine:
   (1 Point)

   ⦿ a. If we have enough information to draft an analytic plan and share for peer review

   ○ b. Data quality issues

   ○ c. Verify the results

   ○ d. Assure model accuracy

40. Business Intelligence focuses on:
   (1 Point)

   ⦿ a. Using a consistent set of metrics to measure past performance and inform business planning

   ○ b. Metrics

   ○ c. Monitoring plan

   ○ d. Business plan

41. Mapper and Reducers consume and produce:
   (1 Point)

   ○ a. Relational tables only

   ⦿ b. pairs

   ○ c. Identifier

   ○ d. value

42. Examples of non-Big Data include:
   (1 Point)

a. Communications companies

b. Egyptian Million Health Initiative

● c. Clinical data

d. Social Media

43. The goals of communicate results phase in big data life cycle include:
(1 Point)

a. Interpret the results

b. Compare IH from phase 1

c. Summarize findings depending on audience

● d. All of the above

44. HDFS does the following to solve the problem when the data is too big to store on one machine:
(1 Point)

a. Stores the data on a server

● b. Stores the data on multiple machines

c. Stores the data on clients only

d. All of the above

45. Benefit of TFIDF is:
(1 Point)

● a. It is a measure that considers the prevalence of a term within a document (TF) and the scarcity of the term over the entire corpus (IDF).

b. TFIDF is very hard to calculate

c. TFIDF requires understanding of the meaning of the text

d. TFIDF is sentiment analysis

46. Virtualization is performed because of its:
(1 Point)

a. Tough management

b. Downtime

c. Resource efficiency

d. Provisioning

47. In the Operationalize phase, you perform the following:
(1 Point)

a. Run a pilot

b. Assess the stakeholders

c. Provide pre-final deliverables

d. None of the above

48. A data scientist key activities are:
(1 Point)

a. Reframe business challenges as well as data analytics

b. Replicate the data

c. Analyze business drivers

d. Draw different scenarios of the data

49. Major challenges in the problem of text mining are:
(1 Point)

○ a. Very limited structures and formats in which the text may be provided

○ b. It is very hard to determine the meaning of words

○ c. Very limited variations of words' synonyms

◉ d. High dimensionality due to the number of possible words

50. You can apply the following algorithms for sentiment analysis:
(1 Point)

○ a. Clustering

○ b. Association rule mining

○ c. Classification

◉ d. All of the above

51. Fault-tolerance is:
(1 Point)

◉ a. The way in which an operating system (OS) responds to a hardware or software failure.

○ b. The way to recover failures

○ c. The way to tune databases

○ d. The way to secure databases

52. In the model planning phase, you perform the following:
(1 Point)

◉ a. Select methods based on hypotheses.

○ b. Make sure the data model is robust

○ c. Make the data model work

d. Visualize the data

53. Parsing is defined as:
    (1 Point)

    ○ a. The process step that takes the un-structured or a semi-structured document and imposes a structure for the downstream analysis.

    ○ b. A semistructured data

    ○ c. A quasi-structured data

    ○ d. A well-defined dataset

54. Big data applications include the following:
    (1 Point)

    ◉ a. Public health in Egypt

    ○ b. Friends' data

    ○ c. Wedding invitations

    ○ d. Doctor private clinics

55. DaaS is defined as:
    (1 Point)

    ○ a. Distributed as a service

    ◉ b. Data Science as a service

    ○ c. Done as a service

    ○ d. Building on the concept that its data product can be provided to the user on demand, regardless of geographic or organizational separation between provider and consumer

56. The main characteristics of Hadoop include:
   (1 Point)

   ○ a. Sequential computation framework

   ● b. Fault-tolerant data storage, parallel computation framework, and job coordination

   ○ c. Fault-tolerant

   ○ d. One master and no slaves

57. Sentiment analysis refers to :
   (1 Point)

   ○ a. A group of tasks that use statistics and natural language processing to mine opinions to identify and extract subjective information from texts is used to identify:

   ○ b. Positive/Negative opinions on product reviews

   ● c. Positive/Negative/Neutral of specific topic people are talking about

   ○ d. Predicting opinions during the next five years on Education