تطوير تطبيقات نظم المعلومات ((خاص

12:00-2:00

20/6/2022

أ.د/تيسير حسن

Faculty of Computers & Information, Assiut University
Course code: IS442
4th Level
Midterm Exam
Duration: 2 hours
Total marker: 50
Reviewer: (Add the name)
This exam for the following program(s) : (Add the program's name(s))

* Required

* This form will record your name, please fill your name.

1. الإسم الرباعي (بالعربي فقط) *

2. رقم الجلوس *

3. المستوي *

- الاول
- الثاني
- الثالث
- الرابع 2013
- الرابع 2014
- الرابع 2015
- الرابع 2016
- الرابع 2017
- الرابع 2018

4. البرنامج *

- بايو
- هندسة

5. رقم المعمل *

- ج٠
- د٠
- هـ٠
- اب
- اج
- اد
- اهـ
- أ٢
- ٢ب
- ٢ج
- ٢د
- ٢هـ
- أ٣
- ٣ب
- ٣ج
- ٣د
- ٣هـ
- أ٤
- ٤ب
- ٤ج
- ٤د
- ٤هـ

6. رقم الكمبيوتر *

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

8. A methodology in data science is (1 Point)

- a. A problem definition
- b. Visualizing data
- c. General strategy that guides processes and activities within a domain
- d. Drawing a boxplot

9. In data science, we are concerned with: (1 Point)

- a. Planning a system
- b. Producing a timeline for the project
- c. Solving a problem or analyzing data
- d. Assigning roles and responsibilities to the team

10. In data science methodology, you need feedback in the following phases: (1 Point)

- a. Data requirements, data collection and data understanding
- b. Analytic approach
- c. Business understanding
- d. Deployment

11. Viability means: (1 Point)

- How much data is there?
- b. is the data relevant to the use case at hand?
- c. How quickly is the data being created, moved, or accessed?
- d. Can we trust data ?

12. Volatility means: (1 Point)

- can we keep the data secure?
- b. how can the data be presented to the user?
- c. can this data produce a meaningful return on investment?
- d. how often does the data change?

13. In the analytics life cycle, to understand the data, you need to: (1 Point)

- a. Select modeling technique
- b. Generate test design
- c. Collect initial data
- d. Assess the model

14. To perform the deployment phase in analytics life cycle, you need to: (1 Point)

- a. Select data
- b. Preprocess data
- c. Publish report
- d. Describe data

15. A data engineer builds: (1 Point)

- a. systems that consolidate, store, and retrieve data from the various applications
- b. systems created by software engineers
- c. analysis on top of models
- d. a & b

16. In Big data, we use: (1 Point)

- a. ETL
- b. ELT

17. Real-time in big data refers to: (1 Point)

- a. Sub-second response time
- b. Schedule time
- c. Event-driven
- d. Batch-processing

18. Big data platform helps in Traffic congestion by: (1 Point)

- a. Access social media to gain insight
- b. Federate data between Big Data and RDBMs
- c. Real time analysis to weather and traffic congestion data streams
- d. Work to understand demand and engage customers

19. The Characteristic for industrial production in an Industry 4.0 environment are: (1 Point)

- a. The strong customization of products under the conditions of highly flexibilized (mass-) production.
- b. The customization of products under the conditions of highly flexibilized (mass-) production.
- c. Production of Internet of Things

20. **A data scientist builds:** (1 Point)

- a. Advanced data structures
- b. Distributed computing
- c. Concurrent programming
- d. Analysis on top of data

21. **Modularity is defined as:** (1 Point)

- The ability of cyber-physical systems within Smart Factories to make decisions on their own
- b. Flexible adaptation of Smart Factories to changing requirements by replacing or expanding individual modules
- c. A virtual copy of the Smart Factory which is created by linking sensor data
- d. A virtual copy of the Smart Factory

22. Green Computing is defined as: (1 Point)

- a. adaptive systems consisting of networks of sensors and smart objects
- b. engaging and interacting with local inhabitants to increase awareness
- c. sustainability when green metrics are effectively coupled with its positive socio-economic impacts.
- d. dedicated to advancing technologies that capture digital senses

23. Big data platform capabilities include: (1 Point)

- a. Information Ingest
- b. Analytic appliances
- c. a & b
- d. prescriptive and predictive analytics

24. Hadoop is defined as: (1 Point)

- a. Apache open source software framework for reliable, scalable, distributed computing of massive amount of data
- b. Software for computing analytics
- c. A framework for big data
- d. A framework for distributed databases

25. Hadoop is not good for: (1 Point)

- a. Massive amounts of data through parallelism
- b. When work cannot be parallelized
- c. A variety of data
- d. Inexpensive commodity hardware

26. Sqoop is: (1 Point)

- a. A data access tool
- b. Data Aggregation from many sources
- c. A fast, scalable, durable, and fault-tolerant
- d. A Tool to easily import information from structured databases

27. Hive: (1 Point)

- a. Facilitates easy data summarization
- b. Is a platform for analyzing large data sets
- c. is a distributed, scalable, big data store
- d. modeled after Google's BigTable

28. Data center virtualization: (1 Point)

- a. is a convenient, on-demand network access
- b. uses resource pooling
- c. Abstracts most of a data center's hardware into software
- d. is a measured service

29. We use virtualization for : (1 Point)

- a. Resource efficiency and faster provisioning
- b. Resource efficiency
- c. price cut-down
- d. Services

30. IaaS is defined as: (1 Point)

- a. The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources.
- b. Input as a service
- c. Data as a service
- d. Implementation as a service

31. Community cloud is defined as: (1 Point)

- a. The cloud is operated solely for an organization
- b. The cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns.
- c. The cloud infrastructure is made available to the general public
- d. The cloud infrastructure is a composition of two or more clouds

32. An Advantage of cloud computing includes: (1 Point)

- a. The data storage is independent
- b. scalability is static
- c. Reducing run time and response time.
- d. The need of high quality equipment

33. The Hadoop computing model depends on: (1 Point)

- a. Notion of a transaction
- b. Distributed transaction
- c. Notion of jobs divided into tasks
- d. Using transaction properties

34. HDFS is designed to be: (1 Point)

- a. Resilient to node failure
- b. Data virtualization
- c. Batch processing
- d. Not replicating the data

35. A Task tracker monitors: (1 Point)

- a. the job tracker
- b. both job tracker and task tracker
- c. job division
- d. the execution of each task

36. Name Node stores: (1 Point)

- a. Block operations
- b. the location of each block in a file
- c. a framework for job scheduling
- d. A YARN-based system for parallel processing

37. If a data node fails, then: (1 Point)

- a. Both name nodes and job tracker detect the failure
- b. Only name nodes detect the failure
- c. Only job tracker detect the failure
- d. Name node re-schedule the task

38. Data acquisition means: (1 Point)

- a. Analyzing data
- b. Understanding data
- c. Obtaining data from a variety of sources, including RDBMS systems, NoSQL, and others
- d. Exploring statistical relations between data

39. **In Data Analysis and Modeling phase in the data science pipeline, the data scientist explores the statistical relationship between the variables in the data and uses Machine Learning** (1 Point)

- True
- False

40. **Data crawling is defined as popular technology used for systematically browsing the web pages** (1 Point)

- ◯ True
- ◯ False

41. **The most time-consuming step in the data science pipeline is data preparation** (1 Point)

- ◯ True
- ◯ False

42. **Data Science pipeline, communicate & operationalize mean give the data back in a compelling form and structure - one-off report, scalable web product** (1 Point)

- ◯ True
- ◯ False

43. **Vulnerability means can we keep the data up-to-date** (1 Point)

- ◯ True
- ◯ False

44. **Event-driven data processing happens when a certain action or condition triggers it.** (1 Point)

- ◯ True
- ◯ False

45. **The objective of the Reduce function is to combine the input from HDFS** (1 Point)

- ◯ True
- ◯ False

46. Sentiment analysis is used to identify Positive/Negative opinions on product reviews (1 Point)

- ◯ True
- ◯ False

47. Major challenges in the problem of text mining are very low dimensions of text (1 Point)

- ◯ True
- ◯ False

48. **Value means can this data produce a meaningful return on investment?** (1 Point)

- ◯ True
- ◯ False

49. When performing Arabic sentiment analysis, one must Preprocess the statements and build your own corpus. (1 Point)

- ◯ True
- ◯ False

50. Examples of Arabic sentiment analysis include (1 Point)

- ◯ opinions on cars, books, and writers
- ◯ Micrarray gene expression
- ◯ Genomic scale data

51. Regular expression is a technique for finding words, strings or a particular patterns in the text (1 Point)

- ◯ True
- ◯ False

52. Wireless monitoring devices that postoperative patients and those with chronic diseases are wearing at home and in their daily lives are examples of big data. (1 Point)

- ◯ True
- ◯ False

53. A personalized medicine is a medical model that proposes the customization of healthcare, with medical
decisions, practices, and/or products being tailored to the individual patient (1 Point)

- ◯ True
- ◯ False

54. Diagnostic testing is often employed for selecting appropriate and optimal therapies based on the context of a patient's genetic content or other molecular or cellular analysis (1 Point)

- ◯ True
- ◯ False

55. Real time analysis to weather and traffic congestion data streams to identify traffic patterns reducing
transportation costs. (1 Point)

- ◯ True
- ◯ False

56. PubMed is a typical example of applying text mining in the bioinformatics area (1 Point)

- ◯ True
- ◯ False

57. Frequent patterns can always be applied for discovering new medications for certain diseases (1 Point)

- ◯ True
- ◯ False