

## Test Tema 71 #1

Actualizado el 13/04/2025

**1. Existen distintos tipos de algoritmos que se aplican en DataMining. ¿A qué categoría pertenece el algoritmo de Bayes Naive?**

- a) Algoritmos de clasificación.
- b) Algoritmos de regresión.
- c) Algoritmos de segmentación.
- d) Algoritmos de asociación.

**2. ¿Cuál no es un modelo de extracción de datos en minería de datos?**

- a) De verificación.
- b) De descubrimiento.
- c) De eliminación
- d) Predictivo.

**3. El algoritmo de minería de datos es el mecanismo que crea un modelo de minería de datos. Para crear un modelo, un algoritmo analiza primero un conjunto de datos y luego busca patrones y tendencias específico. ¿Cuál de los siguientes NO es un algoritmo de minería de datos?**

- a) Algoritmos de clasificación.
- b) Algoritmos de regresión.
- c) Algoritmos de segmentación.
- d) Algoritmos de predestinación.

**4. Indique cuál de los siguientes es SW libre para minería de datos:**

- a) SAS
- b) Clementine
- c) Weka
- d) SPSS

**5. ¿Con cuál de las siguientes categorías de software está más relacionado el Business Intelligence?**

- a) ERP.
- b) Datawarehouse y OLAP.
- c) Document Management y Content Management.
- d) SCM.

**6. Seleccione qué estrategias son de validación de modelo de datos en ciencia de datos.**

- a) Validación simple.
- b) LOOCV
- c) K-Fold Cross-Validation.
- d) Todas las respuestas anteriores son correctas.

**7. ¿Cuál de los siguientes componentes del ecosistema Hadoop no está relacionado con el almacenamiento?**

- a) Hive.
- b) Impala.
- c) Hbase.
- d) Storm.

8. Indique cuál de los siguientes términos corresponde al nombre de una plataforma de software libre compatible con Apache Kafka para la captura de cambios en los datos (CDC - Change Data Capture):

- a) Midjourney
- b) Debezium
- c) Murf
- d) Google Bard

9. ¿En qué tipo de técnica de minería de datos se utiliza el algoritmo IK-medoids?

- a) Predicción.
- b) Categorización.
- c) Clustering.
- d) Estadísticas.

10. Lenguaje de programación usado en Ciencia de Datos.

- a) Julia.
- b) Laura.
- c) Clara.
- d) Jorge.

11. Indicar cuál de los siguientes no es un modelo de minería de datos:

- a) De verificación.
- b) De descubrimiento.
- c) De predicción.
- d) De validación.

12. ¿Qué entiende por ETL? Señale la correcta.

- a) Extract, Transform, Load.
- b) Extract, Transfer, Load.
- c) Extract, Translate, Load.
- d) Extract, Train, Load.

13. Si los datos contienen valores 'NULL' (valores desconocidos) antes de realizar una consulta de data mining:

- a) Debo eliminarlos.
- b) Sólo se eliminan si son pocos. Si son muchos se utiliza alguna técnica para predecir el valor de esos nulos y trabajar con las aproximaciones obtenidas.
- c) En la fase de preproceso se les asignará un valor por defecto.
- d) Se asigna el valor de la media multiplicado por 2.

14. ¿Cual de las siguientes herramientas de minería de datos o Datamining NO es un 'producto open source'?

- a) Knime.
- b) RapidMiner.
- c) ELKI.
- d) Teradata.

15. ¿Qué lenguaje de programación no es usado en ciencia de datos?

- a) R.
- b) Julia.
- c) Mandorian.
- d) SCALA.

**16. Experto encargado de crear informes operativos y visuales**

- a) Data Analyst.
- b) Data Translator.
- c) Data Visualization Developer.
- d) Ninguna de las anteriores.

**17. ¿Cómo se denomina el proceso de limpieza de datos en Data Science?**

- a) Data hygienist.
- b) Data deletion.
- c) Data scrubbing.
- d) Data crawling.

**18. Indique cuál de los siguientes es un software de data warehouse que se apoya en Apache Hadoop y que permite la lectura, escritura y gestión de grandes conjuntos de datos, así como su consulta empleando sintaxis SQL:**

- a) Hive.
- b) Mahout.
- c) ZooKeeper.
- d) -

**19. ¿Qué es la técnica resampling?**

- a) Técnica para estimar la capacidad predictiva de los modelos de data.
- b) Técnica de búsqueda.
- c) Técnica de transformación en calidad de datos.
- d) Ninguna de las anteriores.

**20. ¿Qué es la minería de datos?**

- a) Es el análisis de datos que permite extraer información de bases de datos mediante la búsqueda manual de patrones y relaciones.
- b) Es el análisis de datos que permite extraer información de bases de datos mediante la búsqueda automatizada de patrones y relaciones.
- c) Es el análisis de datos que permite extraer información de bases de datos mediante la búsqueda aleatoria de patrones y relaciones.
- d) Nada de lo anterior es correcto.

**21. ¿Qué componente es introducido en Hadoop 2.0?**

- a) Yarn.
- b) Yern.
- c) Yurn.
- d) Todas las respuestas anteriores son falsas.

**22. En el contexto de "Big Data", la siguiente definición: "Conjunto de principios y fundamentos, tanto científicos como aplicados, que guían la extracción de conocimiento de los datos, y en los que se basan las herramientas, técnicas y procedimientos de Data Mining" se corresponde, de entre los siguientes, con el término:**

- a) Business Intelligence (BI).
- b) Business Analytics (BA).
- c) Machine Learning.
- d) Data Science.

**23. ¿Cómo se denomina el proceso de creación de set de datos más pequeños a partir de uno más grande en ciencia de datos?**

- a) Data hygienist.
- b) Data squashing.
- c) Data scrubbing.
- d) Data crawling.

**24. De las afirmaciones que se indican a continuación, ¿cuál de ellas NO responde a una definición de Inteligencia de Negocio o Business Intelligence (BI)?**

- a) Las estrategias, acciones y herramientas para la creación de conocimiento mediante el análisis de la información de la organización.
- b) Dispositivo software que proporciona servicios de aplicación a computadoras clientes.
- c) Metodología y uso de tecnologías que transforman los datos en información para mejorar la toma de decisiones.
- d) Reunir, depurar y transformar datos que se puedan analizar con la finalidad de mejorar la toma de decisiones internas, para crear ventajas competitivas.

**25. Dentro de un cluster HDFS (Hadoop Distributed File System), una de las funcionalidades del elemento software llamado "NameNode" es:**

- a) Regular el acceso a los ficheros por los clientes.
- b) Servir las peticiones de lectura del sistema de fichero de los clientes.
- c) Servir las peticiones de escritura del sistema de fichero de los clientes.
- d) Este elemento software no existe dentro de la arquitectura HDFS.

**26. De acuerdo a los ejes en los que se estructura el Plan España Digital 2026, seleccione en cual encuadraría mejor el Big Data:**

- a) Eje 4.
- b) Eje 5.
- c) Eje 6.
- d) Eje 7.

**27. En el marco de un proyecto de "Big Data", de entre las siguientes opciones, elija la que indique la fase de la metodología "KDD" que consiste en el tratamiento preliminar de los datos, transformación y generación de nuevas variables a partir de las ya existentes para obtener una estructura de datos apropiada:**

- a) Transformación.
- b) Preprocesamiento.
- c) Minería de datos.
- d) Selección.

**28. ¿Para qué se utiliza principalmente el algoritmo k-means?**

- a) Clustering.
- b) Categorización.
- c) Predicción.
- d) Detección de fraude.

**29. Técnica usada en Ciencia de Datos**

- a) Clasificación.
- b) Regresión.
- c) Agrupación.
- d) Todas las respuestas anteriores son correctas.

**30. A la hora de aplicar métodos de data mining:**

- a) Lo único importante es que se disponga de muchos datos.
- b) Debe disponerse de muchos datos residentes en una base de datos relacional.
- c) Lo más importante es disponer de una buena herramienta comercial.
- d) Los datos deben haber sido preprocesados.

**31. ¿Cuál de las siguientes es una librería de Python para ciencia de datos?**

- a) Numpy.
- b) Pandas.
- c) Scikit-learn.
- d) Todas las respuestas anteriores son correctas.

**32. ¿Qué significa MLOPS?**

- a) Mega Load Operations.
- b) Multiple-loads Operations.
- c) Machine Learning Operations.
- d) Ninguna de las anteriores.

**33. ¿Cuál de los siguientes términos NO es una técnica empleada en Minería de Datos?**

- a) Las redes neuronales.
- b) Look and Find.
- c) Los árboles de decisión.
- d) Las redes bayesianas.

**34. Una tabla en el modelo relacional equivale en una base de datos NoSQL de Apache Cassandra a:**

- a) Familia de columnas.
- b) Colección.
- c) Keyspace.
- d) HyperTable.

**35. En el área de la ciencia de datos, la regresión lineal se usa fundamentalmente en los procesos de:**

- a) Preprocesamiento de la información.
- b) Validación.
- c) Modelado.
- d) ETL.

**36. Herramienta de preprocesamiento de datos en proyectos de ciencia de datos.**

- a) Weka.
- b) Pandas.
- c) RapidMiner.
- d) Todas las respuestas anteriores son ciertas.

**37. Las consultas de data minig se realizan:**

- a) Sobre las bases de datos de la organización.
- b) Sobre las bases de datos cuyos datos se han obtenido de las bases de datos operacionales de la organización.
- c) Sobre cualquier tipo de dato sin necesidad de un preproceso previo.
- d) No se realizan sobre bases de datos.

**38. El objetivo principal de la minería de datos es:**

- a) Descubrir el conocimiento oculto en grandes volúmenes de datos y expresarlo de forma inteligible.
- b) Reorganizar adecuadamente los grandes volúmenes de datos.
- c) Tener copias de seguridad fiables y rápidas de todos y cada uno de los datos de que se dispone.
- d) Ninguna de las anteriores respuestas es correcta.

**39. En el modelo de minería de datos de árboles de decisión (ID3) de Quinlan:**

- a) Se obtiene un nuevo conjunto de reglas a partir de uno previo, utilizando mecanismos de deducción basados en la entropía
- b) Se obtiene un conjunto de reglas a partir de un conjunto de ejemplos, clasificados a priori, utilizando mecanismos de inducción
- c) Se obtiene un nuevo conjunto de reglas que agrupan, por similitud, un conjunto de ejemplos, utilizando aprendizaje no supervisado
- d) Se obtiene un nuevo conjunto de reglas a partir de uno previo, mediante eliminación de reglas incorrectas e inclusión de reglas necesarias, utilizando un conjunto de ejemplos para dicha depuración

**40. ¿Qué técnica es usada para el modelado de datos en un proyecto de data science?**

- a) Análisis Bayesiano.
- b) Regresión Lineal.
- c) SVM.
- d) Todas las respuestas anteriores son ciertas.

**41. En el ámbito de la minería de datos ¿cómo se llama la herramienta con la que se establecen relaciones entre elementos?**

- a) Clasificación.
- b) Asociación.
- c) Agrupación.
- d) Secuenciación.

**42. Seleccione una herramienta Open source usada para procesos ETL**

- a) Azure Data Factory.
- b) SSIS.
- c) Pentaho.
- d) SAS Data Integrator.

**43. Indique cuál de los siguientes NO es un framework de Deep Learning:**

- a) PyTorch.
- b) DeepMind.
- c) Caffe.
- d) MXNet.

**44. CRISP-DM. Señale la incorrecta**

- a) Cross-Industry Standard Process for Data Mining.
- b) Se considera metodología.
- c) Se usa en proyectos de ciencia de datos.
- d) Se considera un estandar para creación de datamarts.

**45. ¿Qué herramienta es usada para MLOPS?**

- a) Jupyter.
- b) ONNX.
- c) Metaflow.
- d) Todas las respuestas anteriores son correctas.

**46. ¿Cómo se denomina el proceso de indexación en la técnica de Web Scraping?**

- a) Crawling.
- b) Rasping.
- c) Raspering.
- d) Web-Rasping.

**47. Dentro de los roles que se pueden desempeñar en un proyecto de "Big Data", indique cuál, de entre los siguientes, se corresponde con la persona que "junta los datos descubiertos, los organiza de forma que estén listos para analizar y se asegura de que se pueden consultar de forma útil, con tiempos de respuesta adecuados por todos los usuarios":**

- a) Explorador de Datos.
- b) Arquitecto de Datos.
- c) Higienista de Datos.
- d) Científico de Datos.

**48. En el ámbito de los algoritmos de aprendizaje:**

- a) Los algoritmos de agrupamiento son métodos de aprendizaje supervisado y los de clasificación son métodos de aprendizaje no supervisado
- b) Los algoritmos de agrupamiento son métodos de aprendizaje no supervisado y los de clasificación son métodos de aprendizaje supervisado
- c) Tanto los algoritmos de agrupamiento como de clasificación son métodos aprendizaje supervisado
- d) Tanto los algoritmos de agrupamiento como de clasificación son métodos aprendizaje no supervisado

**49. Entre las técnicas usadas en minería de datos se encuentran las redes neuronales artificiales, ¿cuál de los siguientes es uno de sus modelos?**

- a) Redes de Shelman.
- b) Redes de Coperfield.
- c) Redes ART.
- d) Máquina de Gauss.

**50. Herramienta propia usada en la técnica de Web scraping**

- a) Octoparse.
- b) Visual Scraper.
- c) Web Harvy.
- d) Todas las respuestas anteriores son ciertas.

**51. ¿Cuál de las siguientes sería una herramienta apropiada para visualizar datos en un proyecto de BI (Business Intelligence)?**

- a) Filezilla.
- b) Jmeter.
- c) Artifactory.
- d) Microsoft Power BI.

**52. ¿Cuáles de los siguientes roles está relacionado con Big Data?**

- a) Data Explorer.
- b) Data Steward.
- c) Data Hygienist.
- d) Todas las respuestas anteriores son ciertas.

**53. Herramienta de visualización de datos**

- a) Power BI
- b) Qlik
- c) Tableau
- d) Todas las respuestas anteriores son correctas.

**54. Las herramientas de minería de datos:**

- a) Se utilizan en entornos altamente transaccionales.
- b) Emplean métodos de búsqueda por asociaciones, secuencias, clasificaciones, pronósticos, y/o agrupamientos.
- c) Trabajan exclusivamente con bases de datos Oracle y Adabas.
- d) Se denominan también sistemas OLTP.

**55. Seleccione una herramienta usada para procesos ETL**

- a) Azure Data Factory.
- b) ODI.
- c) Pentaho.
- d) Todas las anteriores son correctas.

**56. Entre las extensiones del Data Mining se encuentra:**

- a) Web Mining y Text Mining.
- b) Web Mining, Text Mining e Hipertext Mining.
- c) Text Mining e Hipertext Mining.
- d) Web Mining e Hipertext Mining.

**57. Entre las técnicas más comúnmente usadas en Datamining, ¿cuál de las siguientes clasifica cada registro en un conjunto de datos basado en una combinación de las clases de los k registros más similares a él en un conjunto de datos históricos (donde  $k \geq 1$ )?**

- a) Redes neuronales artificiales.
- b) Árboles de decisión.
- c) Método del vecino más cercano.
- d) Regla de Inducción.

**58. SCALA**

- a) Son las siglas de Scalable Language.
- b) Es usado como lenguaje en Ciencia de Datos.
- c) La compilación del código fuente de Scala produce bytecode Java para su ejecución en una JVM.
- d) Todas las anteriores son correctas.



**59. Los ETL son herramientas utilizadas habitualmente en la minería de datos. ¿En qué consisten dichas herramientas?**

- a) Son bases de datos multidimensionales que permiten tener la información correctamente estructurada siguiendo multitud de criterios o dimensiones.
- b) Son herramientas que realizan informes destinados a la alta dirección de la organización.
- c) Son aplicaciones que permiten el acceso a la base de datos a través de páginas HTML.
- d) Son herramientas que permiten extraer información de diversos orígenes, realizar modificaciones en dicha información y cargarla en un sistema destino.

**60. Selenium es una librería de Python que se usa en ciencia de datos para:**

- a) Web Scraping.
- b) Minería de Datos.
- c) Es un algoritmo de IA.
- d) Es una librería Data Translator.

**61. ¿Cuál de las siguientes es una librería de Python para la visualización de datos?**

- a) Numpy.
- b) Plotly.
- c) ONNX.
- d) Todas las respuestas anteriores son correctas.

**62. Rol de equipo de ciencia de datos experto en limpieza de datos**

- a) Data Artist.
- b) Data Hygienist.
- c) Data Analyst.
- d) Data Steward.

**63. ¿Cuál de las siguientes respuestas incluye alguna tecnología que NO forma parte de la Inteligencia de Negocio o Business Intelligence (BI)?**

- a) Data warehouse y análisis OLAP.
- b) Cuadro de mando y minería de datos.
- c) OLTP y DNS.
- d) Análisis OLAP y Dashboards de visualización de datos.

**64. Qué tipo de análisis en Ciencia de datos corresponde con un examen profundo o detallado de datos para entender por qué ha ocurrido algo.**

- a) Descriptivo.
- b) De diagnóstico.
- c) Predictivo.
- d) Prescriptivo.

**65. ¿Qué lenguaje de programación es usado en ciencia de datos?**

- a) R.
- b) Julia.
- c) Spyder.
- d) Todas las respuestas anteriores son correctas.

**66. ¿Cuál de los siguientes ejemplos de tipos de aprendizaje en minería de datos es no supervisado?**

- a) Modelado y control.
- b) Sistemas de predicción.
- c) Descubrimiento de patrones.
- d) Árboles de predicción.

**67. Indique cuál de las siguientes afirmaciones es cierta respecto a la Minería de Datos:**

- a) Las redes neuronales se engloban dentro de las técnicas descriptivas de minería de datos.
- b) El Análisis de regresión es una técnica empleada en la minería de datos.
- c) Los algoritmos repetitivos son una técnica empleada en la minería de datos.
- d) Los árboles de decisión no son una técnica empleada en la minería de datos.

**68. Seleccione la respuesta INCORRECTA. Algunos de los componentes de Hadoop son:**

- a) HDFS.
- b) Hadoop Common.
- c) Hadoop Yarn.
- d) Hadoop MapExtend.

**69. Los algoritmos “no supervisados” en Minería de Datos:**

- a) Se llaman también 'predictivos'.
- b) Utilizan la técnica de 'árboles de decisión'.
- c) Sirven para descubrir patrones y tendencias en los datos actuales.
- d) Predicen el valor de un atributo de un conjunto de datos, conocidos otros atributos.

**70. Rol de equipo de ciencia de datos experto en almacenamiento y procesado de datos**

- a) Data Engineer.
- b) Data Hygienist.
- c) Data Analyst.
- d) Data Steward.

**71. En el ámbito de la minería de datos se conoce como "científico de datos" al responsable de:**

- a) Realizar el diseño de la estructura de datos y de los procesos ETL.
- b) Diseñar, implementar y mantener la infraestructura.
- c) Crear y coordinar los equipos que realizan la implementación de los proyectos.
- d) Resolver problemas complejos a partir del análisis e interpretación de grandes bases de datos.

**72. En el paradigma MapReduce**

- a) Siempre tiene que haber por lo menos un Mapper, pero puede no haber Reducer
- b) Puede no haber un Mapper, pero siempre tiene que haber al menos un Reducer
- c) Puede no haber ni Mapper ni Reducer
- d) Tiene que haber por lo menos un Mapper y un Reducer

**73. ¿Cuál de las siguientes acciones no se puede considerar del modo de operar del Data Mining?**

- a) Detectar desviaciones.
- b) Crear modelos.
- c) Analizar relaciones.
- d) Todas son correctas.

**74. HDFS (Hadoop Distributed File System). Señale la falsa**

- a) Posee modelo WORM.
- b) Es basado en Java.
- c) Los ficheros que se almacenan están divididos en bloques.
- d) No proporciona redundancia.

**75. ¿En qué lenguaje está escrito el proyecto Apache Kafka?**

- a) Python
- b) Groovy
- c) Scala
- d) JavaScript

**76. ¿Qué utilidad puede tener el Data Mining?**

- a) Predecir valores futuros en base a los datos actuales.
- b) Descubrimiento de fraude.
- c) Descubrimiento de patrones en los datos.
- d) Todas las anteriores.

**77. Los tipos de análisis desarrollados en Data Mining se dividen comúnmente en:**

- a) Gráfico y textual.
- b) Supervisado y no supervisado.
- c) Sintáctico y semántico.
- d) Estadístico y de calidad.

**78. ¿Cuál de los siguientes modelos de minería de datos escogería para detectar un posible fraude?**

- a) El modelo de Verificación.
- b) El modelo de Descubrimiento.
- c) El modelo de Simulación.
- d) El modelo Distribuido.

**79. Rol de equipo de ciencia de datos experto en procesamiento gráfico y textual**

- a) Data Artist.
- b) Data Hygienist.
- c) Data Analyst.
- d) Data Steward.

**80. El objetivo del aprendizaje no supervisado en minería de datos basado en modelado consiste en:**

- a) Asignar nuevas situaciones a clases a partir de ejemplos existentes.
- b) Determinar el comportamiento futuro a partir de datos históricos.
- c) Establecer grupos de parecidos.
- d) Descubren patrones y tendencias en los datos.

**81. ¿Qué significa la técnica de Web scraping?**

- a) Diseñar páginas Web de un modo interoperable.
- b) Extraer información de sitios web a partir de técnicas de software.
- c) Filtrar contenido por sensibilidad en la Web.
- d) Analizar la web en busca de ciberamenazas

**82. Los tipos de análisis desarrollados en Data Mining se dividen comúnmente en:**

- a) Sintáctico y semántico.
- b) Estadístico y de calidad.
- c) Gráfico y textual.
- d) Supervisado y no supervisado.