
Solucionario de

ejercicios de autoevaluación

Ejercicios de autoevaluación

Unidad de Aprendizaje 1

1. ¿Quién definió por primera vez el término *business intelligence*?

- a. Richard Millar Devens, en 1958.
- b. Hans Peter Luhn, en 1958.**
- c. Richard Millar Devens, en 1865.
- d. Hans Peter Luhn, en 1865.

2. La implantación del *business intelligence* en una empresa dota a esta de conocimiento útil para una mejor toma de decisiones. Pero, ¿qué tipo de análisis hace de los datos?

- a. Descriptivo.**
- b. Predictivo.
- c. Predictivo y descriptivo.
- d. El análisis debe realizarlo el científico de datos.

3. Determina si las siguientes oraciones son verdaderas o falsas:

a. El *data mining* aporta a la organización un análisis descriptivo de los datos para saber el pasado y el presente de esta.

- Verdadero
- Falso**

b. El *data mining* es una evolución del *business intelligence* tradicional.

- Verdadero**
- Falso

c. El *data mining* aporta a la organización información a través del estudio de patrones y modelos para predecir resultados.

- Verdadero**
- Falso

d. El *data mining* se basa en la inteligencia artificial.

- Verdadero
- Falso

4. El científico de datos:

- a. Es la figura del analizador en el *data mining*.
- b. Se especializa en ordenar y almacenar los datos más relevantes.
- c. Son científicos matemáticos que se especializan en el tratamiento de datos.
- d. Se especializa en el tratamiento de datos.**

5. ¿Cuáles son las 3 V del *big data* más importantes?

- a. Volumen, variación y variedad.
- b. Volumen, velocidad y variedad.**
- c. Volumen, valor y veracidad.
- d. Volumen, valor y variedad.

6. El modelo *MapReduce*:

- a. Se basa en dos elementos: un sistema de ficheros distribuidos y un *software* que tiene implementadas las tareas de cada máquina.**
- b. Se basa en dos elementos: un sistema de ficheros centralizado y un *software* que tiene implementadas las tareas de cada máquina.
- c. Es un sistema de ficheros distribuidos en varias máquinas.
- d. Es un *software* que tiene implementadas las tareas de cada máquina para analizar datos.

7. *Apache Hadoop*:

- a. Pertenece a *Google*.
- b. Es de pago.
- c. Es un *software* libre.**
- d. Es un *software* libre implementado por *Google*.

8. ¿Qué es el *machine learning*?

- a. Análisis automático de las máquinas.
- b. Aprendizaje automático de las máquinas.**
- c. Inteligencia artificial.
- d. Una evolución del *data mining*.

9. ¿Qué es la Web 2.0?

- a. La evolución de las primeras redes de datos.
- b. Una red centralizada donde se crea y comparte información personal.
- c. Una red global donde todos compartimos información mediante la creación de contenido.**
- d. Es la red de internet dotada de mayor seguridad.

10. Determina si las siguientes oraciones son verdaderas o falsas:

- a. El *big data* hace referencia a una cantidad masiva de datos que supera la capacidad de análisis del *software*.

- Verdadero
- Falso

- b. El *big data* hace referencia a una cantidad masiva de datos que puede ser analizada en tiempo real mediante inteligencia artificial.

- Verdadero
- Falso

- c. Al *big data* también se le conoce como macrodatos.

- Verdadero
- Falso

- d. El *big data* también se le llama internet de las cosas.

- Verdadero
- Falso

Ejercicios de autoevaluación

Unidad de Aprendizaje 2

1. Indica si las siguientes afirmaciones sobre la cuantificación de datos en las organizaciones son verdaderas o falsas.

a. Gracias a los sensores instalados en las máquinas se pueden adquirir datos cualitativos que se pueden usar en la estrategia comercial de la organización.

- Verdadero
- **Falso**

b. Todos los datos almacenados en el *data warehouse* de las organizaciones son importantes para su análisis en busca de información.

- **Verdadero**
- Falso

c. A los datos guardados obligatoriamente por las organizaciones solo tienen permitido el acceso los inspectores y las autoridades.

- Verdadero
- **Falso**

d. Las organizaciones deben extraer información tanto de los datos cuantitativos como de los datos cualitativos.

- **Verdadero**
- Falso

e. Los datos cualitativos normalmente se adquieren mediante preguntas tipo test en encuestas.

- Verdadero
- **Falso**

2. ¿Qué son los KPI?

- a. Son una serie de empleados que se encargan de medir el rendimiento de procesos y, además, relacionarlos con los objetivos fijados para dichos procesos.
- b. Son indicadores que se utilizan para medir el rendimiento de procesos y, además relacionarlos con los objetivos fijados para dichos procesos.**
- c. Son un conjunto de indicadores que recogen información cualitativa de los procesos de la organización.
- d. Son un conjunto de indicadores que sirven para medir el rendimiento de los empleados.

3. ¿Cuál de las siguientes opciones describe mejor la relación entre innovación y *big data*?

- a. Proceso de mantener el *statu quo* y evitar cambios en el análisis de grandes volúmenes de datos.
- b. Acción de copiar exactamente las estrategias de la competencia en el contexto del procesamiento de información masiva.
- c. Acto de introducir novedades significativas para mejorar o transformar productos, servicios o procesos mediante el uso inteligente de grandes conjuntos de datos.**
- d. Rutina de seguir métodos tradicionales sin buscar mejoras en la gestión de información a gran escala.

4. ¿Cuál de estas no es una función del *data governance*?

- a. Crear los estándares y las políticas que guiarán el uso de los datos.
- b. Establecer los roles y las responsabilidades internas.
- c. Decidir la tecnología más adecuada para la gestión de datos.
- d. Definir y desarrollar la arquitectura de los datos.**

5. Indica si las siguientes afirmaciones sobre el *data modeling and design* son verdaderas o falsas.

- a. Transforma el modelo lógico en modelo físico y funcional.

- Verdadero
- Falso

b. Trabaja directamente con las bases de datos.

- Verdadero
- Falso

c. Se dedica a estructurar los datos y a almacenarlos en las bases de datos.

- Verdadero
- Falso

d. Cuenta con la figura del ingeniero o arquitecto de datos.

- Verdadero
- Falso

6. ¿Cuáles de estas se consideran funciones del *data security*?

- a. Monitorizar los datos para que se asegure la privacidad.
- b. Utilizar el enmascaramiento de datos como técnica de protección de datos.
- c. Controlar el acceso a los datos mediante la autenticación y la autorización a los interesados.
- d. Ejecutar las políticas y actualizaciones propuestas por el *data governance*.

7. ¿Qué características no son de aplicación para asegurar la calidad de los datos en el *data quality*?

- a. Universalidad
- b. Precisión
- c. Integridad
- d. Publicidad
- e. Estandaridad

8. ¿Cuál de las siguientes afirmaciones sobre el problema de saber entender los datos que se presentan no es correcta?

- a. Para saber entender los datos son de ayuda las herramientas visuales.
- b. Para saber entender los datos es de ayuda la contratación de personal especializado.
- c. Solo la dirección de una organización tiene obligación de adquirir conocimientos específicos sobre el tratamiento de datos.**
- d. Las herramientas visuales no son suficientes para asegurar un correcto entendimiento de lo que en ellas se representa.

9. ¿Cuál es el nombre de la norma europea de protección de datos?

- a. Ley Orgánica de Protección de Datos.
- b. Reglamento Nacional de Protección de Datos.
- c. Reglamento General de Protección de Datos.**
- d. Ley Europea de Protección de Datos.

10. La legislación española sobre protección de datos será de aplicación cuando:

- a. Alguno de sus puntos no contradiga al Reglamento Europeo.**
- b. En todo caso por tener más validez que el Reglamento Europeo.
- c. En ningún caso por el principio de primacía de los reglamentos europeos.
- d. Solo en los casos en que el presidente del gobierno no desee previa votación de las Cortes Generales.

Ejercicios de autoevaluación

Unidad de Aprendizaje 3

1. Selecciona "A" si los procedimientos que se muestran a continuación pertenecen al *business intelligence* y "B" si pertenecen al *big data*.

1. (A/B) Modelado predictivo.
2. (A/B) Informes ad hoc.
3. (A/B) Informes estándar.
4. (A/B) Monitorización de datos.

2. ¿Cuál de las siguientes fuentes de información no forma parte del *business intelligence*?

- a. Sistemas operacionales
- b. Fuentes externas
- c. **Data mining**
- d. Sistemas departamentales

3. Indica si las siguientes afirmaciones son verdaderas o falsas.

a. El proceso ETL comprende la extracción, transformación y análisis de los datos.

- Verdadero
- **Falso**

b. Durante el proceso de extracción, el proceso guarda los datos en el *data staging* antes de guardarlos en el *data warehouse*.

- **Verdadero**
- Falso

c. En el proceso de integración se procede al volcado de datos en el *data warehouse* y se comprueba que estos coincidan con los valores reales.

- **Verdadero**
- Falso

d. En el proceso de limpieza de datos debe haber intervención humana para seleccionar los datos correctos.

- Verdadero
- **Falso**

4. ¿Qué elemento de los siguientes no pertenece al *data warehouse*?

- a. *Data mart*
- b. *Meta data*
- c. *Data staging*
- d. ***Data storing***

5. Indica si las siguientes afirmaciones sobre las herramientas *ad hoc* son verdaderas o falsas.

a. Estas herramientas se basan en la tecnología del internet de las cosas.

- Verdadero
- **Falso**

b. Estas herramientas están diseñadas para fines concretos y no son replicables.

- **Verdadero**
- Falso

c. Para generar informes es necesario estar dentro del entorno de la empresa.

- Verdadero
- **Falso**

d. Las herramientas *ad hoc* solo son manejadas por el personal técnico de la empresa.

- Verdadero
- **Falso**

6. Indica qué tipo de *dashboard* es el indicado para las operaciones de monitorización.

- a. Cuadro de mando operativo.
- b. Cuadro de mando estratégico.
- c. Cuadro de mando táctico.
- d. Todas las opciones son correctas.

7. Indica si las siguientes afirmaciones sobre el cuadro de mando son verdaderas o falsas.

a. El cuadro de mando operativo necesita actualizaciones diariamente.

- Verdadero
- Falso

b. Los cuadros de mando son herramientas para monitorizar y representar información relativa a la empresa.

- Verdadero
- Falso

c. Los cuadros de mando solo cuentan con elementos de analítica tradicional.

- Verdadero
- Falso

d. El cuadro de mando táctico monitoriza los procesos en tiempo real.

- Verdadero
- Falso

8. Señala cuál de las siguientes opciones no se considera una transformación de un cubo OLAP.

- a. *Slicing*
- b. ***Drill-up***
- c. *Drill-down*
- d. *Roll-up*

9. Indica cuál de los siguientes no es un tipo de herramienta OLAP.

- a. ROLAP (*Relational OLAP*).
- b. ROLAP (*Reciprocal OLAP*).**
- c. MOLAP (*Multidimensional OLAP*).
- d. HOLAP (*Hybrid OLAP*).

10. Determina cuáles de las siguientes afirmaciones son verdaderas y cuáles son falsas.

- a. La tecnología *big data* es incompatible con el *business intelligence*.
 - Verdadero
 - **Falso**

- b. En el *business intelligence* existen herramientas de análisis de datos.
 - **Verdadero**
 - Falso

- c. Las analíticas tradicional y avanzada pueden ser procedimientos compatibles.
 - **Verdadero**
 - Falso

- d. Los *data warehouses* solo pueden ser bases de datos relacionales.
 - **Verdadero**
 - Falso

Ejercicios de autoevaluación

Unidad de Aprendizaje 4

1. Selecciona "A" si los siguientes métodos de recogida de datos pertenecen a datos estructurados y "B" si pertenecen a datos no estructurados.

1. (A/B) Generados por los usuarios.
2. (A/B) Creados.
3. (A/B) Experimentación.
4. (A/B) Compilados.

2. ¿Cuál de las afirmaciones que se presentan a continuación es correcta?

- a. Los datos semiestructurados no se pueden almacenar en bases de datos.
- b. Los archivos PDF son datos semiestructurados.
- c. Los datos no estructurados son más fáciles de analizar.
- d. Los datos estructurados son más fáciles de analizar.**

3. Indica si las siguientes afirmaciones son verdaderas o falsas.

a. Los datos RFID son de tipo estructurado.

- Verdadero
- Falso

b. Los datos de las redes sociales son de poca utilidad, ya que están fuera del ámbito de *big data*.

- Verdadero
- Falso

c. Las ciudades inteligentes son uno de los principales productos surgidos de *big data*.

- Verdadero
- Falso

d. Los datos gubernamentales siempre son de acceso prohibido para el *big data*.

- Verdadero
- **Falso**

4. ¿Cuál de los siguientes métodos de representación de datos no está orientado a seres humanos?

- a. **Sistema binario**
- b. Sistema decimal
- c. Mapas
- d. Infografías

5. Indica si las siguientes afirmaciones sobre los cuadros comparativos y las infografías son verdaderas o falsas.

a. Los cuadros comparativos pueden contener imágenes.

- **Verdadero**
- Falso

b. Las infografías representan información resumida y gráfica.

- **Verdadero**
- Falso

c. Las infografías pueden mostrar información en tiempo real y *online*.

- **Verdadero**
- Falso

d. Los cuadros comparativos son datos estructurados.

- Verdadero
- **Falso**

6. ¿Cuál de las siguientes variables son de carácter cuantitativo?

- a. Variables acumuladas
- b. Variables ordinales
- c. Variables discretas**
- d. Variables dicotómicas

7. Indica si las siguientes afirmaciones sobre la representación gráfica de datos son verdaderas o falsas.

a. Las gráficas de barras compuestas solo pueden representar variables cualitativas.

- Verdadero
- **Falso**

b. Los pictogramas no pueden contener imágenes, ya que dificultan su comprensión.

- Verdadero
- **Falso**

c. El polígono de frecuencias está orientado a variables cuantitativas.

- **Verdadero**
- Falso

d. Los gráficos que representan variables cuantitativas integrales tienen forma de escalera.

- **Verdadero**
- Falso

8. ¿Cuál de las siguientes son variables cualitativas?

- a. Medidas de voltaje de una pila.
- b. Cantidad de ventas según el tipo de clientes (numerados).
- c. Compras de vehículos por países (por nombre).**
- d. Ventas en función de compras.

9. Señala cuál de los siguientes no es un tipo de base de datos.

- a. En estrella
- b. NoSQL
- c. En red
- d. Semiestructurada**

10. Indica cuál de las siguientes afirmaciones sobre la representación de datos digital son verdaderas y falsas.

a. El código binario está formado por ceros y unos.

- **Verdadero**
- Falso

b. Con el código binario no se pueden representar caracteres.

- Verdadero
- **Falso**

c. La codificación binaria es de base 2.

- **Verdadero**
- Falso

d. Un *byte* equivale a 1024 bits.

- Verdadero
- **Falso**

Ejercicios de autoevaluación

Unidad de Aprendizaje 5

1. Selecciona si las siguientes afirmaciones son verdaderas o falsas.

a. El *big data* no contempla la necesidad de hacer un uso eficaz de los datos y la viabilidad de estos.

- Verdadero
- **Falso**

b. El volumen de datos es fundamental para hablar del *big data*.

- **Verdadero**
- Falso

c. El *big data* está orientado a una gestión de datos de forma distribuida y escalable.

- **Verdadero**
- Falso

d. La analítica predictiva en tiempo real es algo que surgió antes del data 2.0.

- Verdadero
- **Falso**

2. ¿Cuál de las siguientes afirmaciones es correcta?

a. *Apache* fue el primero en crear el motor *MapReduce*.

b. El sistema de ficheros *Google File System* es, en esencia, de tipo SQL.

c. ***Apache Hadoop* es una plataforma *open source*.**

d. Algunas de las aplicaciones de *Hadoop* son *open source*.

3. Indica si las siguientes afirmaciones sobre HDFS son verdaderas o falsas.

a. Los *namenodes* son cada uno de los nodos de un clúster que realiza los trabajos con los datos.

- Verdadero
- **Falso**

b. HDFS divide los datos en trozos y, además, hace replicaciones con ellos.

- **Verdadero**
- Falso

c. Los *datanodes* son cada uno de los nodos de un clúster que realiza los trabajos con los datos.

- **Verdadero**
- Falso

d. Un *namenode* actúa como servidor de un clúster.

- **Verdadero**
- Falso

4. Un *tasktracker*...

a. ... se encarga de dividir cada proceso en subprocesos y distribuirlos en los *jobtrackers*.

b. ... recibe las tareas del *jobtracker* para ejecutarlas.

c. ... se considera un nodo maestro.

d. ... se encarga de comprobar que los *jobtrackers* se encuentran operativos.

5. Indica si las siguientes afirmaciones sobre el ecosistema *Hadoop* son verdaderas o falsas.

a. El proyecto *Spark* tiene una variante *streaming* para trabajar con datos en tiempo real.

- Verdadero
- Falso

b. YARN se implementó ya desde la versión 1.0 de *Hadoop*.

- Verdadero
- Falso

c. HDFS también es un sistema de ficheros orientado al trabajo en tiempo real.

- Verdadero
- Falso

d. *Apache Hive* convierte una instrucción en lenguaje SQL en *Pig* o directamente en lenguaje *MapReduce*.

- Verdadero
- Falso

6. ¿Cuál de las siguientes afirmaciones sobre las distribuciones comerciales de *Hadoop* es correcta?

a. Cloudera ofrece todo su paquete en código abierto, incluso con las mejoras realizadas por otras empresas a las aplicaciones, lo que no permite que sean de pago.

b. Cloudera ha implementado su propio sistema de ficheros.

c. Cloudera puede trabajar tanto en *Linux* como en *Windows*.

d. Cloudera ha implementado su propio sistema de archivos.

7. Indica si las siguientes afirmaciones sobre el almacenamiento NoSQL son verdaderas o falsas.

a. Una ventaja del NoSQL es la de almacenar mayor cantidad de datos a mayor velocidad.

- Verdadero
- Falso

b. Todas las bases de datos NoSQL almacenan sus datos en código binario.

- Verdadero
- Falso

c. En las bases de datos orientadas a objetos aparecen los documentos llamados colecciones. Por ejemplo, la colección «Clientes» sería un documento.

- Verdadero
- Falso

d. Las bases de datos *Key-Value* almacenan la información en forma binaria.

- Verdadero
- Falso

8. MongoDB...

- a. ... es una base de datos orientada a objetos.
- b. ... es una base de datos orientada a documentos.**
- c. ... es una base de datos columnar.
- d. Existen varios tipos de bases de datos en MongoDB.

9. ¿Qué modalidad de servicio no ofrece el *cloud computing*?

- a. Infraestructura como servicio.
- b. Software libre como servicio.**
- c. Plataforma como servicio.
- d. *Software* como servicio.

10. Señala cuál de las siguientes afirmaciones sobre el *cloud computing* son verdaderas o falsas.

a. Uno de los beneficios del *cloud computing* es la adaptabilidad en cuanto al uso de los medios que ofrece.

- Verdadero
- Falso

b. El *cloud computing* solo ofrece el soporte físico como servicio. Los demás elementos, como el *software*, debe ponerlos la empresa consumidora.

- Verdadero
- Falso

c. La computación elástica proporciona un servicio adaptable en base a la necesidad que se tenga de analizar más o menos datos.

- Verdadero
- Falso

d. Una de las soluciones más utilizadas en *big data* es el *cloud computing*.

- Verdadero
- Falso

Ejercicios de autoevaluación

Unidad de Aprendizaje 6

1. Selecciona la afirmación correcta sobre la analítica avanzada.

- a. Para utilizar la analítica avanzada es necesario contar con tecnología *big data*.
- b. La analítica avanzada es capaz de extraer información de grandes cantidades de datos para mejorar la toma de decisiones.**
- c. La analítica avanzada engloba solo las distintas técnicas predictivas.
- d. Gracias a la analítica avanzada se puede representar en tiempo real la situación de un negocio.

2. Indica cuál de las siguientes afirmaciones sobre la analítica predictiva son verdaderas o falsas.

- a. En el análisis predictivo se utilizan técnicas de modelación estadística.
 - Verdadero
 - Falso
- b. Un objetivo del análisis predictivo es coger solo datos históricos para realizar predicciones.
 - Verdadero
 - Falso
- c. El análisis predictivo busca patrones y modelos acerca de acciones y acontecimientos futuros.
 - Verdadero
 - Falso

d. El análisis predictivo realiza recomendaciones sobre qué acciones son mejores y peores para la empresa.

- Verdadero
- **Falso**

3. Indica cuáles de las siguientes afirmaciones pertenecen al *data mining*.

- a. **El *data mining* es una etapa dentro del proceso KDD.**
- b. KDD es una etapa dentro del proceso del *data mining*.
- c. **Durante el preprocesamiento, los datos se transforman en un conjunto adecuado para llevar a cabo el análisis.**
- d. **Un tipo de minería de datos es el modelo de segmentación.**

4. Indica cuál de las siguientes afirmaciones sobre el método CRISP-DM son verdaderas o falsas.

a. En la fase de modelado se elige la técnica de *data mining* a utilizar.

- **Verdadero**
- Falso

b. En la fase de preparación de los datos hay que recolectar estos.

- Verdadero
- **Falso**

c. La fase de comprensión de los datos incluye la limpieza de los datos y cambios de formato.

- Verdadero
- **Falso**

d. En la fase de implantación se coge el conocimiento obtenido y se transforma en decisiones.

- **Verdadero**
- Falso

5. El *machine learning*...

- a. ... consiste en crear programas que sean capaces de generalizar comportamientos gracias al procesamiento de información a modo de ejemplo.
- b. ... está basado en el método científico.
- c. ... tiene como objetivo desarrollar técnicas que permitan a las computadoras detectar patrones y modelos.
- d. ... está más orientado al proceso de descubrimiento de patrones de datos desconocidos.

6. Indica cuáles de estas afirmaciones son verdaderas o falsas.

- a. El aprendizaje supervisado se basa en el estudio previo de ejemplos para deducir una función de predicción.

- Verdadero
- Falso

- b. El aprendizaje no supervisado se basa en el estudio previo de ejemplos para deducir una función de predicción.

- Verdadero
- Falso

- c. En el aprendizaje no supervisado los datos en entrada son tratados como variables aleatorias.

- Verdadero
- Falso

- d. En el aprendizaje por refuerzo se utiliza el método de ensayo-error.

- Verdadero
- Falso

7. Señala cuáles de las siguientes afirmaciones son correctas.

- a. En las redes neuronales se utiliza un proceso iterativo de selección que compara cada nuevo modelo con los anteriores y prevalece el mejor.
- b. Los algoritmos de *clustering* clasifican las observaciones en distintos grupos de nubes de datos.**
- c. Los algoritmos genéticos enlazan elementos que simulan neuronas, los cuales interactúan para producir estímulos en la salida.
- d. Los árboles de decisión definen un conjunto de decisiones que generan reglas para clasificar un conjunto de datos.**

8. ¿Cuál de los siguientes algoritmos se considera un método de aprendizaje no supervisado?

- a. Regresión lineal
- b. Árboles de decisión
- c. *Clustering***
- d. Redes neuronales

9. Selecciona las respuestas correctas sobre la analítica prescriptiva.

- a. Evalúa el impacto que tendrán las decisiones que se tomen.**
- b. Se utiliza cuando la toma de decisiones debe ser sistemática.
- c. Realiza recomendaciones sobre las decisiones o acciones.**
- d. Utiliza la optimización matemática para definir las reglas de negocio.

10. Indica cuáles de estas afirmaciones sobre la analítica prescriptiva son verdaderas o falsas.

- a. La analítica prescriptiva proporciona conocimiento sobre las posibles decisiones que se pueden tomar.
 - Verdadero
 - Falso

b. Las reglas de negocio son necesarias para evaluar cada decisión a tomar.

- Verdadero
- Falso

c. La analítica prescriptiva toma modelos y patrones para hacer predicciones y mostrar recomendaciones sobre las decisiones que se deben tomar.

- Verdadero
- Falso

d. La optimización matemática tiene en cuenta distintas restricciones para separar decisiones válidas de decisiones óptimas.

- Verdadero
- Falso

