

## CUALIFICACIÓN PROFESIONAL:

Gestión de datos y entrenamiento en sistemas de Inteligencia Artificial basados en aprendizaje automático.

|                             |                                     |
|-----------------------------|-------------------------------------|
| <i>Familia Profesional:</i> | <b>Informática y Comunicaciones</b> |
| <i>Nivel:</i>               | <b>3</b>                            |
| <i>Código:</i>              | <b>IFC749_3</b>                     |
| <i>Estado:</i>              | <b>Tramitación BOE</b>              |
| <i>Publicación:</i>         | <b>RD 46/2022</b>                   |

### Competencia general

Extraer, procesar y aplicar analíticas de datos para el entrenamiento en sistemas de Inteligencia Artificial, basándose en técnicas de aprendizaje automático para la predicción, clasificación o cualquier otro tratamiento inteligente de datos, imágenes, vídeos o lenguaje natural, preprocesando, depurando y particionando los datos en subconjuntos, identificando las variables significativas, verificando la ausencia de sesgos y cumpliendo la normativa aplicable en materia de protección de datos y propiedad intelectual e industrial.

### Unidades de competencia

- UC0966\_3:** CONSULTAR Y EXTRAER INFORMACIÓN DE DISTINTAS PLATAFORMAS DE ALMACENAMIENTO DE DATOS
- UC2492\_3:** Procesar los datos para su uso en sistemas de Inteligencia Artificial basados en aprendizaje automático
- UC2493\_3:** Entrenar modelos en sistemas de Inteligencia Artificial basados en aprendizaje automático

### Entorno Profesional

#### Ámbito Profesional

Desarrolla su actividad profesional en las áreas de sistemas y desarrollo dedicada/o a la gestión y proceso de datos y desarrollo de aplicaciones inteligentes, en entidades de naturaleza pública o privada, empresas de tamaño pequeño, mediano, grande o bien empresas de cualquier tamaño que gestionan sistemas de información para otras organizaciones, tanto por cuenta propia como ajena, con independencia de su forma jurídica. Desarrolla su actividad dependiendo, en su caso, funcional y/o jerárquicamente de un superior. Puede tener personal a su cargo en ocasiones, por temporadas o de forma estable. En el desarrollo de la actividad profesional se aplican los principios de accesibilidad universal y diseño universal o diseño para todas las personas de acuerdo con la normativa aplicable.

#### Sectores Productivos

Se ubica en el sector servicios, en el subsector de la gestión de sistemas de información para otras organizaciones, en el de desarrollo de software, en el desarrollo de sistemas de Inteligencia Artificial basados en aprendizaje automático, en el de consultoría técnica en sistemas de información inteligentes o en cualquier sector productivo con empresas que dispongan de sistemas de información inteligentes para la gestión de sus procesos de negocio.

### Ocupaciones y puestos de trabajo relevantes

Los términos de la siguiente relación de ocupaciones y puestos de trabajo se utilizan con carácter genérico y omnicomprendivo de mujeres y hombres.

- Integradores de sistemas de información
- Técnicos en proceso y analítica de datos

### Formación Asociada (510 horas)

#### Módulos Formativos

- MF0966\_3:** CONSULTA Y MANIPULACIÓN DE INFORMACIÓN CONTENIDA EN GESTORES DE DATOS (210 horas)
- MF2492\_3:** Procesamiento de datos para su uso en sistemas de Inteligencia Artificial basados en aprendizaje automático (150 horas)
- MF2493\_3:** Entrenamiento de modelos en sistemas de Inteligencia Artificial basados en aprendizaje automático (150 horas)

## UNIDAD DE COMPETENCIA 1

### CONSULTAR Y EXTRAER INFORMACIÓN DE DISTINTAS PLATAFORMAS DE ALMACENAMIENTO DE DATOS

Nivel: 3  
Código: UC0966\_3  
Estado: BOE

#### Realizaciones profesionales y criterios de realización

**RP1:** Interpretar las especificaciones de diseño correspondientes a los sistemas gestores de datos, utilizados por la organización para el almacenamiento de la información.

**CR1.1** Los modelos de datos contenidos en los sistemas gestores de datos se reconocen, identificando las relaciones y dependencias de los elementos que los configuran.

**CR1.2** La estructura global de la información del gestor de datos y su almacenamiento se estudian utilizando las especificaciones del diseño, con objeto de localizar las informaciones contenidas en el mismo.

**CR1.3** La arquitectura, componentes y tecnología que intervienen en el diseño se reconocen con el fin de identificar el o los sistemas gestores de datos que contienen la información de la organización.

**CR1.4** Las herramientas y lenguajes de consulta y manipulación para extraer la información contenida en los gestores de datos, se seleccionan según especificaciones del modelo y la tecnología del gestor de datos.

**RP2:** Extraer, transformar y cargar informaciones contenidas en gestores de datos de distinta tipología utilizando herramientas y lenguajes de procesamiento de información, de acuerdo con especificaciones técnicas y funcionales y cumpliendo la normativa de protección de datos.

**CR2.1** La codificación, las herramientas y lenguajes seleccionados de los procedimientos de consulta, manipulación y presentación de información se utilizan según las especificaciones técnicas y funcionales, las necesidades de la organización y normativa aplicable.

**CR2.2** La documentación inherente a las herramientas y lenguajes de consulta y manipulación de la información, se interpreta para obtener la funcionalidad deseada en el proceso de extracción de información, teniendo en cuenta las especificaciones técnicas de las mismas.

**CR2.3** Los procedimientos de extracción y consulta realizados se documentan para su registro según normas de la organización.

**RP3:** Configurar las pasarelas y medios de conexión entre las herramientas cliente y los sistemas gestores de datos, para establecer la comunicación entre ambos según necesidades de la organización.

**CR3.1** La documentación inherente a la herramienta cliente y al sistema gestor de datos, se interpreta para establecer los parámetros de conexión entre los componentes software que intervienen en la conexión según especificaciones técnicas y funcionales.

**CR3.2** Los parámetros de conexión entre las herramientas cliente y los sistemas gestores de datos, se implantan para establecer la conexión entre ambos y permitir la extracción de datos de los sistemas gestores según necesidades operativas.

**CR3.3** Las configuraciones de las conexiones entre las pasarelas y medios de conexión y los sistemas gestores de datos, se prueban para determinar la funcionalidad de extracción de datos según las necesidades de la organización.

## Contexto profesional

### Medios de producción

Equipos informáticos y periféricos. Sistemas operativos y parámetros de configuración. Herramientas ofimáticas. Herramientas cliente específicas para acceder a los sistemas gestores de bases de datos. Lenguajes específicos de acceso a sistemas gestores de bases de datos. Conectores o pasarelas estándares de accesos a sistemas gestores de bases de datos. Protocolos de comunicación. Herramientas de conectividad. Servidores Web. Sistemas de seguridad. Motores de gestores de datos para utilizar, en entorno de pruebas. Servicios distribuidos de información.

### Productos y resultados

Datos extraídos en diferentes formatos del sistema de gestión de información. Plantillas tipo para recuperar la información del sistema de gestión de información. Consultas para la manipulación de informaciones contenidas en gestores de datos. Conexiones lógicas entre gestores de datos y clientes.

### Información utilizada o generada

Diseño y especificaciones del tratamiento y almacenamiento de la información. Visión global del sistema de información a realizar, entregar y explotar. Modelos de datos contenidos en los sistemas gestores de datos. Manuales de uso y funcionamiento de los sistemas informáticos. Manuales de funcionamiento del software. Manuales de las herramientas de desarrollo y consulta utilizadas. Documentación de cursos de formación. Documentación de explotación del sistema de información. Soportes técnicos de asistencia. Normativa aplicable acerca de la propiedad intelectual y los derechos de autor. Manuales de "usabilidad".

## UNIDAD DE COMPETENCIA 2

### Procesar los datos para su uso en sistemas de Inteligencia Artificial basados en aprendizaje automático

Nivel: 3

Código: UC2492\_3

Estado: Tramitación BOE

### Realizaciones profesionales y criterios de realización

**RP1:** Extraer datos de las fuentes, previa identificación y clasificación de las mismas, estableciendo el filtrado inicial y configurando las pasarelas que permitan su acceso para su posterior carga y análisis.

**CR1.1** Las fuentes de datos se conectan, configurando las pasarelas que permiten su acceso y teniendo en cuenta:

- Su origen: sistemas de almacenamiento, sistemas IoT, plataformas de datos en 'streaming', integración con APIs u otro,
- Su naturaleza: estructurados, tales como datos cuantitativos u otros o no estructurados,
- Si se trata de datos formales, tales como los provenientes de fuentes jurídicas o bien datos no formales tales como audios, imágenes o textos provenientes de fuentes como las redes sociales.

**CR1.2** Los filtros y consultas para la extracción de datos desde las fuentes se establecen, previa selección, configurándolos, considerando la posibilidad de distribución de datos y de procesamiento cercano a las fuentes para maximizar tiempos de proceso, conforme a las especificaciones de análisis y según la naturaleza de los datos.

**CR1.3** Los datos se extraen desde sus orígenes, conforme a las especificaciones de análisis y según la naturaleza de los datos.

**RP2:** Determinar la estructura y relaciones correspondientes al dominio de aplicación en los datos disponibles, mediante exploración para el análisis según objetivos de negocio.

**CR2.1** Los datos recogidos de fuentes diversas se almacenan para componer conjuntos de datos, identificando su naturaleza e integrándolos en el destino.

**CR2.2** Los conjuntos de datos recogidos se exploran por separado, identificando su estructura y relaciones para el análisis de variables por parte de la persona responsable.

**CR2.3** La descripción de cada conjunto de datos se documenta, incluyendo entidades, volumetrías, relaciones y descripción de atributos, entre otros, para trabajar en las especificaciones de análisis de los datos, según los objetivos de negocio.

**RP3:** Verificar los datos mediante técnicas estadísticas, escribiendo código para explorar sus atributos con el fin de evaluar la cobertura y calidad del conjunto, de modo que permita la selección de los atributos clave para el entrenamiento de modelos, siguiendo las especificaciones de análisis.

**CR3.1** La calidad de los datos se valora, aplicando técnicas estadísticas, conforme a las especificaciones de análisis vinculadas al cumplimiento de las necesidades de negocio.

**CR3.2** El conjunto de datos se evalúa, mediante estadísticas sobre cobertura y sesgo tales como frecuencia y distribución, valorando el cumplimiento de los objetivos de negocio conforme a las especificaciones de análisis.

**CR3.3** La descripción del resultado de las verificaciones sobre el conjunto de datos se documenta, indicando cobertura, frecuencia y sesgo entre otros, para que se tomen decisiones de diseño sobre la selección y uso del mismo en el modelado.

**RP4:** Seleccionar los datos mediante la depuración, limpieza, construcción y posterior clasificación en conjuntos, para su uso posterior en el modelado analítico, según las especificaciones del diseño.

**CR4.1** El conjunto de datos y sus atributos se seleccionan, atendiendo a los criterios y especificaciones del diseño, como resultado de las verificaciones realizadas.

**CR4.2** El conjunto de datos de entrenamiento se limpia, depurándolo para su utilización en el modelado, normalizándolo, categorizándolo o imputando valores ausentes, entre otros.

**CR4.3** El conjunto de datos depurado se construye, atendiendo a las especificaciones de diseño de los modelos a implementar y según las técnicas de Inteligencia Artificial en las que se basen dichos modelos, añadiendo variables sintéticas fruto de cálculos agregados tales como:

- Normalización de atributos a media cero, desviación típica uno,
- Reescalado de atributos dentro de un rango predefinido,
- Media móvil calculada con respecto a una ventana deslizante temporal.

**CR4.4** El conjunto de datos depurado se divide en particiones, según indique el científico de datos, habitualmente entrenamiento, validación y test, para entrenar y evaluar el modelo.

## Contexto profesional

### Medios de producción

Equipos y sistemas cliente y servidor. Sistemas operativos, firmware y software de base. Soporte hardware para almacenamiento de información y/o almacenamiento de la misma en la nube. Software y herramientas de exploración y análisis estadístico de datos, lenguajes de programación específicos. Entornos de desarrollo y programación.

### Productos y resultados

Datos analizados y conjuntos de datos y variables y relaciones identificados. Conjuntos de datos analizados y conformes a las necesidades de negocio. Verificaciones de conjuntos de datos documentados. Conjuntos de datos seleccionados y depurados. Variables sintéticas agregadas añadidas a los conjuntos de datos. Conjuntos de datos divididos en particiones.

### Información utilizada o generada

Normas externas de trabajo: (normativa aplicable de protección de datos y propiedad intelectual e industrial) Normas internas de trabajo: (modelo o plan de negocio, proyecto software, diseño y especificaciones de los datos utilizados, modelos de datos contenidos en los sistemas gestores de datos, normativa interna para elaboración de documentación) Documentación técnica: (manuales de las herramientas de análisis estadístico, desarrollo y consulta utilizadas, documentación de cursos de formación, soportes técnicos de asistencia).

## UNIDAD DE COMPETENCIA 3

### Entrenar modelos en sistemas de Inteligencia Artificial basados en aprendizaje automático

Nivel: 3

Código: UC2493\_3

Estado: Tramitación BOE

#### Realizaciones profesionales y criterios de realización

**RP1:** Buscar correlaciones entre las variables utilizando herramientas que incorporan técnicas de la estadística y el aprendizaje automático, con anterioridad al diseño y entrenamiento de modelos y siguiendo las indicaciones proporcionadas por el científico de datos responsable.

**CR1.1** Las correlaciones entre variables, tanto lineales como no lineales se buscan, aplicando técnicas de la estadística y del aprendizaje automático ('machine learning'- ML-).

**CR1.2** Los procedimientos de categorización y codificación de las variables tales como mapas de bits ('one-hot vectors') u otros se aplican mediante software o herramientas.

**CR1.3** Las transformaciones matemáticas tales como logaritmos, exponenciales, funciones trigonométricas u otras se aplican antes del entrenamiento de los modelos, teniendo en cuenta la distribución de sus valores dentro del conjunto de datos.

**CR1.4** Las variables de entrada más explicativas se seleccionan para tenerlas en cuenta durante el diseño de los modelos, considerando aquellas que tienen mayor correlación con las variables de salida o variables objetivo.

**CR1.5** El documento que describe las variables se redacta, incluyendo todos los detalles analizados en relación a las variables tanto de entrada como de salida y a las correlaciones encontradas, con el fin de servir para tomar decisiones durante el diseño de los modelos y para ser integrado en un informe final.

**RP2:** Reducir la dimensión de las muestras de los conjuntos de datos, mediante programación o herramientas software, para obtener una representación de los mismos mediante variables latentes siguiendo instrucciones de la persona responsable.

**CR2.1** Las técnicas de reducción de la dimensión, tales como análisis de componentes principales -PCA-, t-SNE, autocodificadores ('autoencoders') basados en redes neuronales y otras, disponibles como funcionalidades en herramientas software existentes se utilizan, eligiendo aquellas que realizan una mayor reducción de la dimensión mientras conservan la mayor variabilidad (capacidad explicativa) de las muestras.

**CR2.2** Los métodos y procedimientos se programan, usando el lenguaje definido para el proyecto, realizando la reducción de la dimensión de las muestras con objeto de crear una copia del conjunto de datos con las transformaciones ya aplicadas, o bien aplicando las transformaciones durante el entrenamiento, antes de que las muestras se utilicen como entrada a los modelos para su entrenamiento y evaluación.

**CR2.3** El trabajo realizado se documenta, indicando:

- Las técnicas de reducción de la dimensión aplicadas al conjunto de datos con el que se esté trabajando,
  - Los valores concretos de los parámetros de configuración de cada técnica,
  - Los resultados obtenidos en cuanto a reducción de la dimensión y pérdida de explicación de la variabilidad,
- enfatisando el balance entre la proporción de reducción y la pérdida de variabilidad en función de uno o más parámetros.

**RP3:** Visualizar los datos, representándolos gráficamente, con la finalidad de corroborar las correlaciones encontradas, verificando la reducción de la dimensión aplicada sobre el conjunto de datos con los que se esté trabajando.

**CR3.1** Las gráficas se generan, escribiendo programas al efecto, utilizando varias representaciones, con el fin de verificar tanto las correlaciones encontradas como las transformaciones aplicadas para reducir la dimensión de las muestras, bajo la supervisión de la persona responsable.

**CR3.2** La información recogida mediante gráficas se elabora, usando medios didácticos tales como diapositivas o infografías, organizándola para presentarla al equipo de trabajo, con el fin de sacar conclusiones de manera conjunta y así decidir colectivamente qué subconjunto de todas las transformaciones probadas se utilizarán para probar su efecto con los distintos modelos a entrenar y evaluar.

**CR3.3** Las gráficas se recogen en un informe, presentándolas en formatos tales como histogramas, mapas de dispersión y otras técnicas de representación, explicando cómo se han obtenido y aportando una valoración de cada gráfica, destacando aspectos cualitativos como la correlación entre las variables latentes fruto de las transformaciones y/o reducciones y las variables salida.

**RP4:** Diseñar modelos basados en aprendizaje automático ('machine learning' -ML-) para aplicarlos sobre el conjunto de datos con el objeto de abordar el problema planteado según su tipo -de regresión o de clasificación-, siguiendo las indicaciones proporcionadas por el científico de datos responsable.

**CR4.1** Las técnicas de aprendizaje automático tales como redes neuronales, máquinas de soporte vectorial, modelos de mixturas de gaussianas, árboles de decisión u otras se eligen para diseñar los modelos en base al análisis exploratorio de los datos y el estudio visual realizados previamente.

**CR4.2** Los modelos diseñados se ensayan, programando código para entrenarlos, realizando experimentos que permitan ver qué combinaciones de parámetros de configuración conducen a mejores resultados, tales como:

- El coeficiente de aprendizaje para las redes neuronales ('learning rate'), el número de capas, el número de neuronas en cada capa, los tipos de activación a utilizar en las capas ocultas,
- El número de grupos ('clusters') para el algoritmo 'k means', y
- El número de árboles a generar en el caso de 'random forests'.

**CR4.3** Las métricas a utilizar para evaluar modelos tales como:

- La desviación porcentual entre valores predichos y reales en problemas de regresión como la predicción del volumen de ventas semanal por producto,
- La ratio de falsos positivos y la de falsos negativos para un problema de clasificación,
- Otras métricas relacionadas con las ratios anteriores como 'accuracy', 'precision', 'recall' o 'sensitivity', 'specificity', y área bajo la curva ROC ('Receiver Operating Characteristic'), se seleccionan, teniendo en cuenta las técnicas a aplicar y el problema a resolver.



**CR4.4** Los modelos diseñados en base a las métricas seleccionadas se evalúan, con el fin de decidir, bien la creación de nuevos modelos desde cero, bien el rediseño de modelos anteriores, en función de que los resultados obtenidos se ajusten a los esperados, siguiendo indicaciones de la persona responsable.

**CR4.5** Los comentarios en el código escrito se incluyen, detallando el papel que juega una variable, describiendo las funciones y algoritmos con el fin de ser reutilizado en el desarrollo de soluciones a otros problemas similares y siguiendo las normas establecidas por la organización.

**CR4.6** La documentación se genera, detallando los modelos diseñados con todas sus variantes, con el fin de que sirva como punto de partida para el informe que incluirá los resultados de evaluar el rendimiento de cada uno de los modelos.

**RP5:** Evaluar los modelos diseñados, definiendo un subconjunto de test, cuando el conjunto de datos no lo tuviera previamente definido, y creando nuevas particiones de validación y entrenamiento una vez separado el subconjunto de test.

**CR5.1** El conjunto de datos se divide en dos subconjuntos, entrenamiento y test, si no estaban previamente definidos, usando las muestras del subconjunto de test solamente para evaluar el modelo y respetando que el subconjunto de test represente un porcentaje del conjunto de datos completo entre el 10% y el 30%.

**CR5.2** El subconjunto de entrenamiento se vuelve a dividir en dos nuevos subconjuntos, entrenamiento propiamente dicho y validación, escogiendo las muestras de manera aleatoria, pero conservando la misma distribución de clases.

**CR5.3** La validación cruzada de los modelos se lleva a cabo, generando distintas particiones de entrenamiento y validación, donde el número de distintas particiones lo indicará la persona responsable, repitiendo el proceso completo de entrenamiento y evaluación del modelo por cada una de las particiones, decidiendo en qué iteración del entrenamiento se guarda la versión del modelo para ser evaluado con el subconjunto de test.

**CR5.4** Cada modelo entrenado y validado con cada partición se evalúa con el subconjunto de test, mostrando la media aritmética, la varianza u otra medida de las métricas elegidas para evaluar los modelos a partir de las particiones creadas para la validación cruzada.

**CR5.5** El informe de la evaluación se crea, tomando como punto de partida el informe con todos los detalles de los modelos diseñados e incluyendo gráficas y tablas que muestren las métricas elegidas, para evaluar los modelos y su evolución durante el entrenamiento con respecto a los subconjuntos de validación y que muestren los resultados de los distintos modelos diseñados y sus variantes, siempre con respecto al mismo subconjunto de test con el fin de poder comparar rápidamente el rendimiento y la robustez de cada modelo y sus variantes.

## Contexto profesional

### Medios de producción

Lenguajes de programación, herramientas software y plataformas de desarrollo de Inteligencia Artificial para el tratamiento, análisis, visualización y transformación de los datos; Implementaciones de técnicas de aprendizaje automático.

### Productos y resultados

Datos analizados. Dimensión de muestras de conjuntos de datos reducidas. Datos visualizados y representados. Subconjuntos de test definidos. Particiones de validación y entrenamiento definidas. Modelos diseñados, entrenados y listos para ser desplegados. Informe de evaluación realizado. Modelos

evaluados. Informe de comparación con modelos basados en otras técnicas de aprendizaje automático distintas a la técnica en la que se basa el modelo seleccionado realizado.

### Información utilizada o generada

Normas externas de trabajo (normativa aplicable de protección de datos, propiedad intelectual e industrial). Normas internas de trabajo (planificación del trabajo en base al proyecto, plan de seguridad) Documentación técnica (Guía de cómo se ha utilizado el sistema de Inteligencia Artificial elegido para realizar los experimentos, referencias a artículos de investigación y libros relacionados con el área del aprendizaje automático, manuales de las herramientas software utilizadas para entrenar los modelos, documentación de requisitos hardware tanto de la plataforma como de las herramientas software o 'toolkits'.

## MÓDULO FORMATIVO 1

### CONSULTA Y MANIPULACIÓN DE INFORMACIÓN CONTENIDA EN GESTORES DE DATOS

|                   |  |
|-------------------|--|
| Nivel:            | 3  |
| Código:           | MF0966_3   |
| Asociado a la UC: | UC0966_3 - CONSULTAR Y EXTRAER INFORMACIÓN DE DISTINTAS PLATAFORMAS DE ALMACENAMIENTO DE DATOS |
| Duración (horas): | 210  |
| Estado:           | BOE  |

#### Capacidades y criterios de evaluación

- C1:** Identificar las tipologías de gestores de datos y sus correspondientes modelos de datos según unas especificaciones de diseño dadas.
- CE1.1** Describir los atributos y propiedades de los elementos que componen los modelos de datos utilizados por los gestores de datos, para identificar su función dentro del sistema según las tecnologías utilizadas.
  - CE1.2** Identificar los posibles cambios que se pueden producir en el modelo de datos como consecuencia de la evolución tecnológica y funcional del sistema gestor de base de datos asociado.
  - CE1.3** Identificar las características generales y específicas de los sistemas gestores de datos operacionales, para optimizar la consulta y extracción de información según especificaciones técnicas y estándares de mercado.
  - CE1.4** Distinguir las características de los sistemas gestores de datos orientados a datos específicos (imágenes, sonidos, documentos, planos geográficos y videos, entre otros), para mejorar los procedimientos de consulta utilizados según especificaciones técnicas y estándares de mercado.
  - CE1.5** Formular, reconocer y especificar técnicas de recuperación de la información, para ser utilizadas en el sistema de información según el modelo de datos y tipología del gestor de datos implementado en el sistema de información.
- C2:** Distinguir las características y funciones de los lenguajes y herramientas de consulta y extracción de información de los sistemas gestores de datos, teniendo en cuenta sus especificaciones técnicas y funcionales.
- CE2.1** Determinar las funcionalidades y características de los lenguajes de consulta utilizados en los gestores de datos, para su utilización en los procedimientos de extracción y consulta de información según las especificaciones técnicas propias del lenguaje.
  - CE2.2** Determinar las funcionalidades y características de las herramientas utilizadas en los gestores de datos, para su utilización en los procedimientos de extracción y consulta de información según especificaciones técnicas de las herramientas.

**C3:** Aplicar procedimientos de extracción y consulta de información en el sistema gestor de datos, utilizando lenguajes de consulta específicos según necesidades dadas.

**CE3.1** Utilizar la gramática, sintaxis y semántica del lenguaje de consulta utilizado, para la creación de procedimientos de extracción de la información según las normas de codificación y desarrollo dadas.

**CE3.2** En un supuesto práctico, de extracción y consulta de la información utilizando lenguajes de consulta específicos del sistema gestor de datos, según un diseño especificado:

- Identificar las fuentes de información para la creación de los procedimientos de consulta y extracción.
- Crear procedimientos que permitan formular consultas para la recuperación de la información.
- Recuperar y archivar procedimientos de consultas y extracción previamente establecidos para explotar la información del sistema.
- Depurar y verificar los procedimientos creados utilizando las especificaciones implementadas en el lenguaje para asegurar la calidad de los resultados.
- Establecer formatos de presentación y publicación de la información extraída del sistema gestor de datos utilizado para la posterior distribución de la misma.
- Documentar los procedimientos creados.

**C4:** Aplicar procedimientos de extracción y consulta de información en el sistema gestor de datos, utilizando herramientas específicas.

**CE4.1** Clasificar las diferentes formas de trabajo ofrecidas por la herramienta de aplicación, para su selección en la creación de procedimientos de extracción y consulta de información, según necesidades funcionales y especificaciones técnicas.

**CE4.2** En un supuesto práctico, de extracción y consulta de la información utilizando herramientas específicas en el sistema gestor de datos y según unas necesidades dadas:

- Identificar las fuentes de información para la creación de los procedimientos de consulta y extracción.
- Formular consultas que permitan la recuperación de la información según un diseño dado.
- Recuperar y archivar consultas previamente establecidas para optimizar los procesos de explotación de la información.
- Establecer formatos de presentación y publicación de la información extraída del sistema gestor de datos utilizado para realizar la distribución de la misma según un diseño dado.

**C5:** Aplicar los procedimientos de configuración de pasarelas y medios de conexión para extraer y consultar información de sistemas gestores de datos, según necesidades funcionales.

**CE5.1** Enunciar características generales de pasarelas y medios de conexión para permitir su selección y utilización en los procedimientos de extracción y consulta de información, según las especificaciones técnicas de los sistemas gestores de datos.

**CE5.2** Identificar los parámetros a configurar en una pasarela o medio de conexión para poder extraer y consultar información de sistemas gestores de datos, según especificaciones técnicas de la infraestructura de comunicaciones y del sistema gestor de datos.

**CE5.3** Usar medios de conexión y pasarelas para extraer y consultar información del sistema gestor de datos implementado, según especificaciones funcionales.

## Capacidades cuya adquisición debe ser completada en un entorno real de trabajo

C3 respecto a CE3.2; C4 respecto a CE4.2.

### Otras Capacidades:

Adaptarse a la organización, a sus cambios organizativos y tecnológicos así como a situaciones o contextos nuevos.

Responsabilizarse del trabajo que desarrolla y del cumplimiento de objetivos.

Transmitir información con claridad, de manera ordenada, estructurada, clara y precisa respetando los canales establecidos en la organización.

Mantener una actitud asertiva, empática y conciliadora con las personas demostrando cordialidad y amabilidad en el trato.

Mostrar iniciativa en la búsqueda de soluciones y en la resolución de problemas.

Adoptar actitudes posturales saludables en el entorno de trabajo.

## Contenidos

### 1 Modelos de datos

Descripción. Características generales. Tipos de modelos de datos. Compatibilidades entre modelo de datos.

### 2 Sistemas gestores de datos

Características. Componentes de un sistema gestor de datos. Funcionamiento de un sistema gestor de datos. Relación entre los sistemas gestores de datos y los modelos de datos. Características de los sistemas gestores de datos operacionales (OLAT). Características de los sistemas gestores de datos orientados a las tomas de decisiones o análisis de datos (OLAP). Sistemas gestores de datos avanzados.

### 3 Procedimientos de extracción y consulta de información en los sistemas gestores de datos

Búsquedas en sistemas gestores. Optimización de la recuperación. Utilización de lenguajes de consultas. Ventajas e inconvenientes. Entornos de trabajo. Utilización de herramientas. Ventajas e inconvenientes. Funcionalidades. Presentación de la información. Normativa aplicable en materia de protección de datos.

### 4 Lenguajes de consulta y extracción de datos

Características. Funcionalidades. Lenguajes utilizados en sistemas gestores de datos relacionales (SQL). Lenguajes utilizados en sistemas gestores orientados a objetos (OQL). Lenguajes de consulta utilizados en sistemas gestores de datos relacionales orientados a objetos. Lenguajes de consulta utilizados en sistemas gestores de datos relacionales con extensiones de formatos de documento de marcas extendidas (XML-QL). Otros lenguajes de consulta y extracción de datos. Procedimientos de pruebas.

### 5 Herramientas de consulta y extracción de información de los sistemas gestores de datos

Características. Funcionalidades de las herramientas. Herramientas utilizadas en los sistemas gestores de datos operacionales. Herramientas utilizadas en los sistemas gestores de datos orientados a los análisis de datos. Otras herramientas utilizadas en los sistemas gestores de datos.

## 6 Pasarelas y medios de conexión

Características. Parámetros de configuración. Procedimientos de verificación. Guía de uso y utilización de pasarelas y medios de conexión.

### Parámetros de contexto de la formación

#### Espacios e instalaciones

Los espacios e instalaciones darán respuesta, en forma de aula, aula-taller, taller de prácticas, laboratorio o espacio singular, a las necesidades formativas, de acuerdo con el Contexto Profesional establecido en la Unidad de Competencia asociada, teniendo en cuenta la normativa aplicable del sector productivo, prevención de riesgos laborales, salud laboral, accesibilidad universal, diseño universal o diseño para todas las personas y protección medioambiental

#### Perfil profesional del formador o formadora:

1. Dominio de los conocimientos y las técnicas relacionados con la consulta y extracción de información de distintas plataformas de almacenamiento de datos, que se acreditará mediante una de las dos formas siguientes:

- Formación académica de nivel 2 (Marco Español de Cualificaciones para la Educación Superior) o de otras de superior nivel relacionadas con el campo profesional.

- Experiencia profesional de un mínimo de 2 años en el campo de las competencias relacionadas con este módulo formativo.

2. Competencia pedagógica acreditada de acuerdo con lo que establezcan las Administraciones competentes.

## MÓDULO FORMATIVO 2

### Procesamiento de datos para su uso en sistemas de Inteligencia Artificial basados en aprendizaje automático

|                   |  |
|-------------------|--|
| Nivel:            | 3  |
| Código:           | MF2492_3   |
| Asociado a la UC: | UC2492_3 - Procesar los datos para su uso en sistemas de Inteligencia Artificial basados en aprendizaje automático |
| Duración (horas): | 150  |
| Estado:           | Tramitación BOE  |

### Capacidades y criterios de evaluación

**C1:** Aplicar técnicas de extracción de datos de las fuentes, previa identificación y clasificación de las mismas, estableciendo el filtrado inicial y configurando pasarelas que permitan su acceso para su posterior carga y análisis.

**CE1.1** Identificar las fuentes de datos y conocimiento, clasificándolas:

- Según su origen: sistemas gestores de datos, sistemas IoT, plataformas de datos en 'streaming', integración con APIs u otro,
- Según su naturaleza: estructuradas o no estructuradas,
- Según sean formales, tales como las provenientes de fuentes jurídicas o bien no formales, como audios, imágenes o textos provenientes de redes sociales.

**CE1.2** Describir tipos de filtros y consultas para la extracción de datos, explicando su funcionalidad.

**CE1.3** Describir procedimientos de manejo de datos masivos y de mejora de tiempos de proceso de los mismos, tales como el proceso cercano a las fuentes en el caso de fuentes de datos distribuidas, explicando sus mecanismos de funcionamiento.

**CE1.4** Describir tipos de pasarelas de conexión, identificando mecanismos de configuración y conexión.

**CE1.5** En un supuesto práctico de extracción de datos de las fuentes:

- Conectar las fuentes de datos, configurando las pasarelas que permiten su acceso y teniendo en cuenta el tipo: estructurados, no estructurados; formales o no formales.
- Establecer los filtros y consultas para la extracción de datos desde las fuentes, previa selección, configurándolos, considerando la posibilidad de distribución de datos y de procesamiento cercano a las fuentes para maximizar tiempos de proceso, conforme a las especificaciones de análisis y según la naturaleza de los datos.
- Extraer los datos desde sus orígenes, conforme a las especificaciones de análisis y según la naturaleza de los datos

**C2:** Aplicar técnicas exploratorias para identificar y describir la estructura y variables en los datos en el dominio de aplicación.

**CE2.1** Describir técnicas y posibilidades de almacenaje de los datos para componer conjuntos, identificando su naturaleza.

**CE2.2** Identificar la estructura, variables y relaciones de los conjuntos de datos aplicando técnicas exploratorias.

**CE2.3** Identificar entidades, volumetrías, relaciones y atributos, para documentar cada conjunto de datos, describiendo sus características.

**CE2.4** En un supuesto práctico de aplicación de técnicas exploratorias para identificar y describir la estructura y variables en los datos en el dominio de aplicación:

- Almacenar los datos recogidos de fuentes diversas para componer conjuntos de datos, identificando su naturaleza.
- Explorar los conjuntos de datos recogidos por separado, identificando su estructura y relaciones para el análisis de variables.
- Documentar cada conjunto de datos, identificando entidades, volumetrías, relaciones y descripción de atributos, entre otros, para trabajar en las especificaciones de análisis de los datos, según los objetivos de negocio.

**C3:** Aplicar técnicas de verificación de datos mediante cálculos estadísticos, escribiendo código para explorar atributos de los datos con el fin de evaluar la cobertura y calidad del conjunto, de modo que permita la selección de los atributos clave para el entrenamiento de modelos, siguiendo unas especificaciones de análisis proporcionadas.

**CE3.1** Describir las técnicas estadísticas de análisis de la calidad técnica de los datos, que se aplican para su verificación, explicando sus características y objetivos.

**CE3.2** Identificar técnicas de evaluación de la calidad de los datos basadas en cálculos estadísticos tales como frecuencia y distribución, para detectar la cobertura y el sesgo, describiendo sus características.

**CE3.3** En un supuesto práctico de verificación de datos:

- Aplicar técnicas estadísticas sobre los datos, identificando y describiendo la calidad técnica de los mismos, escribiendo código o usando herramientas software.
- Aplicar técnicas estadísticas sobre los datos, tales como frecuencia y distribución, identificando y describiendo la cobertura y sesgo de los mismos, escribiendo código o usando herramientas software.
- Documentar los resultados de la evaluación, detallando cualitativa y cuantitativamente el sesgo, la frecuencia, la cobertura y la calidad encontrados.

**C4:** Aplicar técnicas de selección de datos mediante la depuración, limpieza, construcción y posterior clasificación en conjuntos, para su uso posterior en el modelado analítico, según unas especificaciones del diseño.

**CE4.1** Identificar los atributos de los datos, explicando sus características, para su selección según especificaciones del diseño y tomando como base los resultados de las verificaciones realizadas.

**CE4.2** Identificar técnicas de limpieza y depurado de datos de entrenamiento, tales como la normalización, categorización o imputación de valores ausentes, entre otros, describiendo sus características y operativa.

**CE4.3** Identificar técnicas de construcción de un conjunto de datos, describiendo sus características y las posibles variables sintéticas a añadir, atendiendo a las especificaciones de diseño de los modelos a implementar y según las técnicas de Inteligencia Artificial en las que se basen dichos modelos.

**CE4.4** En un supuesto práctico de selección de datos:

- Depurar los datos de entrenamiento, siguiendo especificaciones y criterios dados, aplicando normalización y teniendo en cuenta las evaluaciones y verificaciones detectadas previamente.



- Depurar los datos de entrenamiento, siguiendo especificaciones y criterios dados, aplicando categorización y teniendo en cuenta las evaluaciones y verificaciones detectadas previamente.
- Depurar los datos de entrenamiento, siguiendo especificaciones y criterios dados, aplicando imputación de valores ausentes y teniendo en cuenta las evaluaciones y verificaciones detectadas previamente.
- Construir un conjunto de datos, añadiendo variables sintéticas fruto de cálculos agregados tales como normalización, reescalado y media móvil respecto a una ventana deslizando temporal.

## Capacidades cuya adquisición debe ser completada en un entorno real de trabajo

C1 respecto a CE1.5; C2 respecto a CE2.4; C3 respecto a CE3.3; C4 respecto a CE4.4.

### Otras Capacidades:

Responsabilizarse del trabajo que desarrolla y del cumplimiento de los objetivos.

Demostrar cierto grado de autonomía en la resolución de contingencias relacionadas con su actividad.

Comunicarse eficazmente con las personas adecuadas en cada momento, respetando los canales establecidos en la organización.

Adaptarse a la organización, a sus cambios organizativos y tecnológicos, así como a situaciones o contextos nuevos.

Adoptar actitudes posturales adecuadas en el entorno de trabajo.

Mostrar una actitud de respeto hacia los compañeros, procedimientos y normas de la empresa.

Cumplir las medidas que favorezcan el principio de igualdad de trato y de oportunidades entre hombres y mujeres.

Valorar el talento y el rendimiento profesional con independencia del sexo.

Aplicar de forma efectiva el principio de igualdad de trato y no discriminación en las condiciones de trabajo entre mujeres y hombres.

## Contenidos

### 1 Fuentes de datos para un sistema de Inteligencia Artificial basado en aprendizaje automático

Fuentes de datos y de conocimiento. Definición.

Clasificaciones para descripción de datos. Orígenes y tipos (estructuradas y no estructuradas; formales y no formales).

Tecnologías y herramientas de extracción y conexión para conjuntos masivos de datos según su naturaleza y uso, tales como: conectores IoT, conectores para eventos en tiempo real, conectores SQL u otros.

Tecnologías y herramientas de almacenamiento para conjuntos de datos según su naturaleza y uso tales como gestores de datos SQL o noSQL, centralizados o distribuidos, basados en grafos, clave-valor u otros.

El ciclo de preparación de los datos para un sistema de Inteligencia Artificial.

### 2 Perfilado de datos

Conceptos y dimensiones de perfilado de datos: entidades, volumetrías, relaciones y atributos de un conjunto de datos.

Técnicas exploratorias y herramientas para visualización y manipulación de datos.

### 3 Verificación de la calidad de los datos

Calidad de datos.

Evaluación de calidad de datos para la inteligencia artificial.

Técnicas de evaluación de calidad técnica de datos y no conformidades.

Definición de cobertura y sesgo para Inteligencia Artificial.

Técnicas de evaluación de sesgo y cobertura basadas en cálculos estadísticos.

#### 4 Limpieza y depurado de datos

Perfilado de datos para el caso de uso con la perspectiva de negocio: atributos, características, dimensionalidad y variables sintéticas. Técnicas de limpieza y depurado de datos de entrenamiento.

Normalización, categorización, imputación de valores ausentes y otras.

Técnicas de generación de un conjunto de datos para el entrenamiento de un sistema de Inteligencia Artificial.

### Parámetros de contexto de la formación

#### Espacios e instalaciones

Los talleres e instalaciones darán respuesta a las necesidades formativas de acuerdo con el contexto profesional establecido en la unidad de competencia asociada, teniendo en cuenta la normativa aplicable del sector productivo, prevención de riesgos laborales, accesibilidad universal y protección medioambiental. Se considerará con carácter orientativo como espacios de uso:

- Instalación de 2 m<sup>2</sup> por alumno o alumna.

#### Perfil profesional del formador o formadora:

1. Dominio de los conocimientos y las técnicas relacionados con el procesamiento de datos para su uso en sistemas de Inteligencia Artificial basados en aprendizaje automático, que se acreditará simultáneamente mediante las dos formas siguientes:

- Formación académica de nivel 2 (Marco Español de Cualificaciones para la Educación Superior) o de otras de superior nivel relacionadas con el campo profesional.

- Experiencia profesional de un mínimo de 2 años en el campo de las competencias relacionadas con este módulo formativo.

2. Competencia pedagógica acreditada de acuerdo con lo que establezcan las Administraciones competentes.

## MÓDULO FORMATIVO 3

### Entrenamiento de modelos en sistemas de Inteligencia Artificial basados en aprendizaje automático

|                   |  |
|-------------------|--|
| Nivel:            | 3  |
| Código:           | MF2493_3   |
| Asociado a la UC: | UC2493_3 - Entrenar modelos en sistemas de Inteligencia Artificial basados en aprendizaje automático |
| Duración (horas): | 150  |
| Estado:           | Tramitación BOE  |

### Capacidades y criterios de evaluación

**C1:** Utilizar herramientas estadísticas y de aprendizaje automático para encontrar correlaciones entre variables con anterioridad al diseño y entrenamiento de modelos.

**CE1.1** Describir técnicas de la estadística y el aprendizaje automático ('machine learning') para la detección de correlaciones entre variables, tanto lineales como no lineales.

**CE1.2** Describir procedimientos de categorización de variables tales como mapas de bits ('one-hot vectors') u otros, explicando sus características y objetivos.

**CE1.3** Describir técnicas de transformación matemática tales como logaritmos, exponenciales, funciones trigonométricas u otras, explicando sus características y objetivos.

**CE1.4** Enumerar herramientas que incorporan técnicas estadísticas y/o de aprendizaje automático, describiendo sus características.

**CE1.5** En un supuesto práctico de aplicación de técnicas de análisis de datos:

- Encontrar las correlaciones entre variables, tanto lineales como no lineales, aplicando técnicas de la estadística y del aprendizaje automático a través de herramientas o código.
- Aplicar procedimientos de categorización y codificación de las variables tales como mapas de bits ('one-hot vectors') u otros, mediante software o herramientas.
- Aplicar transformaciones matemáticas a las variables, teniendo en cuenta su distribución.
- Identificar las variables de entrada más explicativas, tomando las que tienen mayor correlación con las variables de salida, para tenerlas en cuenta durante el diseño de los modelos
- Documentar las variables de entrada y de salida, incluyendo los detalles analizados y las correlaciones encontradas.

**C2:** Aplicar técnicas de reducción sobre muestras de conjuntos de datos, mediante programación o herramientas software, para obtener una representación de los mismos mediante variables latentes.

**CE2.1** Identificar técnicas de reducción de la dimensión, tales como análisis de componentes principales -PCA-, t-SNE, autocodificadores ('autoencoders') basados en redes neuronales y otras, explicando sus características.

**CE2.2** Describir lenguajes de programación y herramientas aplicables a la codificación de técnicas de reducción, explicando sus características y funcionalidades.

**CE2.3** En un supuesto práctico de aplicación de técnicas de reducción sobre las muestras de un conjunto de datos:

- Elegir las técnicas de reducción a aplicar, considerando aquellas que realizan una mayor reducción de la dimensión mientras conservan la mayor variabilidad de las muestras.
- Aplicar la técnica de reducción elegida usando el lenguaje indicado en el supuesto y aplicando las transformaciones.
- Aplicar la técnica de reducción elegida usando herramientas software y aplicando las transformaciones.
- Documentar el trabajo realizado, indicando las técnicas de reducción de la dimensión aplicadas al conjunto de datos con el que se esté trabajando, los valores concretos de los parámetros de configuración de cada técnica y los resultados obtenidos en cuanto a reducción de la dimensión y pérdida de explicación de la variabilidad, enfatizando el balance entre la proporción de reducción y la pérdida de variabilidad en función de uno o más parámetros.

**C3:** Aplicar técnicas de representación gráfica, visualizando los datos para corroborar las correlaciones encontradas y verificar la reducción aplicada anteriormente.

**CE3.1** Enumerar técnicas de representación gráfica, identificando su tipo, elementos y utilidad práctica.

**CE3.2** Describir herramientas de utilidad para la representación gráfica, explicando sus características.

**CE3.3** En un supuesto práctico de visualización gráfica de datos:

- Escribir programas utilizando técnicas de representación para corroborar las correlaciones y verificar la reducción de la dimensión sobre el conjunto de datos.
- Elaborar las gráficas usando infografías o diapositivas, organizándolas para su presentación al equipo de trabajo y elegir las transformaciones a usar en los modelos a entrenar.
- Redactar un informe con los resultados, explicando cómo se ha obtenido cada gráfica y aportando una valoración en base a aspectos cualitativos y cuantitativos y la correlación entre variables latentes.

**C4:** Aplicar técnicas de diseño de modelos basados en aprendizaje automático ('machine learning' -ML-) para aplicarlos sobre un conjunto de datos con el objeto de abordar el problema planteado según su tipo -de regresión o de clasificación.

**CE4.1** Enumerar técnicas de aprendizaje automático, tales como redes neuronales, máquinas de soporte vectorial, modelos de mixturas gaussianas, árboles de decisión u otras, describiendo sus características.

**CE4.2** Identificar los posibles experimentos a codificar y aplicar para evaluar modelos diseñados y las combinaciones de parámetros que conducen a mejores resultados tales como:

- El coeficiente de aprendizaje para las redes neuronales ('learning rate'), el número de capas, el número de neuronas en cada capa, los tipos de activación a utilizar en las capas ocultas,
- El número de grupos ('clusters') para el algoritmo k-means, y
- El número de árboles a generar en el caso de 'random-forests'.

**CE4.3** Describir las métricas a utilizar para evaluar modelos, tales como:

- La desviación porcentual entre valores predichos y reales en problemas de regresión como la predicción del volumen de ventas semanal por producto.
- La ratio de falsos positivos y la de falsos negativos para un problema de clasificación.
- Otras métricas relacionadas con las ratios anteriores como 'accuracy', 'precision', 'recall' o 'sensitivity', 'specificity', y área bajo la curva ROC ('Receiver Operating Characteristic'), explicando sus características.

**CE4.4** En un supuesto práctico de diseño de modelos basados en técnicas de aprendizaje automático:

- Elegir las técnicas de aprendizaje automático a aplicar, en base al análisis exploratorio de los datos y el estudio visual y gráfico realizado previamente.
- Evaluar los modelos diseñados, programando código para entrenarlos y realizando experimentos que permitan ver qué combinaciones de parámetros de configuración conducen a mejores resultados.
- Seleccionar métricas en función de las técnicas de aprendizaje automático aplicadas, teniendo en cuenta el problema a resolver.
- Evaluar los modelos diseñados en base a las métricas seleccionadas, decidiendo, bien la creación de nuevos modelos desde cero, bien el rediseño de modelos anteriores en función de que los resultados obtenidos se adecúen a los esperados.
- Incluir comentarios en el código escrito, detallando el papel que juega cada variable, describiendo las funciones y algoritmos.
- Generar la documentación, detallando los modelos diseñados con todas sus variantes.

## Capacidades cuya adquisición debe ser completada en un entorno real de trabajo

C1 respecto a CE1.5; C2 respecto a CE2.3; C3 respecto a CE3.3; C4 respecto a CE4.4.

### Otras Capacidades:

Responsabilizarse del trabajo que desarrolla y del cumplimiento de los objetivos.

Demostrar cierto grado de autonomía en la resolución de contingencias relacionadas con su actividad.

Comunicarse eficazmente con las personas adecuadas en cada momento, respetando los canales establecidos en la organización.

Adaptarse a la organización, a sus cambios organizativos y tecnológicos, así como a situaciones o contextos nuevos.

Adoptar actitudes posturales adecuadas en el entorno de trabajo.

Mostrar una actitud de respeto hacia los compañeros, procedimientos y normas de la empresa.

Cumplir las medidas que favorezcan el principio de igualdad de trato y de oportunidades entre hombres y mujeres.

Valorar el talento y el rendimiento profesional con independencia del sexo.

Aplicar de forma efectiva el principio de igualdad de trato y no discriminación en las condiciones de trabajo entre mujeres y hombres.

## Contenidos

### 1 Técnicas Estadísticas en el ámbito del aprendizaje automático

Técnicas estadísticas para el análisis exploratorio de datos, transformaciones simples y/o de reducción de la dimensión de las muestras. Características

Distribuciones de probabilidad.

Técnicas de representación gráfica aplicables a la representación de variables en muestras: histogramas, mapas de dispersión y otras técnicas

Técnicas de transformación de datos de entrada. Técnicas aplicadas a cada variable de manera singular (Escalado a valores dentro de un rango, normalización a media cero y desviación típica uno). Funciones matemáticas de una o más variables (logaritmo, exponencial, raíz cuadrada,  $x^2$ ,  $x^n$ , seno, coseno, tangente hiperbólica u otros). Técnicas aplicadas a todas las variables de entrada (expansión polinómica para aumento de la dimensión de las muestras en problemas de regresión,

técnicas de reducción de la dimensión de las muestras: PCA, t-SNE, 'autoencoders' basados en redes neuronales. Finalidad y aplicación).

Herramientas software que implementen técnicas estadísticas en el ámbito del análisis de datos.

## 2 Visualización de datos y resultados en el aprendizaje automático

Técnicas de visualización de datos y resultados (Gráficas de dispersión, Histogramas, Mapas de color/densidad, Curva ROC u otros).

Herramientas software aplicables a técnicas de representación de datos y de resultados.

## 3 Técnicas de entrenamiento en el aprendizaje automático

Tipos de aprendizaje automático: supervisado, no supervisado, semi-supervisado, por refuerzo.

Tipos de problemas: clasificación, regresión, segmentación, detección de objetos.

Técnicas paramétricas: Basadas en distribuciones de probabilidad conocidas: 'Gaussian Mixture Models' u otras.

Técnicas no paramétricas: (árboles de decisión: clásicos, 'random forest', 'gradient boosting', 'extremely randomized trees'; Vecinos más próximos (k-NN); 'Kernel density estimation'; 'Support Vector Machines' (SVM); K-Means.

Técnicas Holísticas: redes neuronales: 'fully connected', 'convolucional', 'recurrent', 'transformers', 'autoencoders', 'encoder-decoder', seq2seq u otras.

Técnicas de regularización para evitar el sobreaprendizaje.

Métricas para evaluar la calidad de los modelos diseñados y entrenados: MSE, MAE, 'accuracy', precisión, 'recall' o 'sensitivity', 'specificity', 'F1-score', ratio de falsos positivos y de falsos negativos, área bajo la curva ROC, 'DICE coefficient', 'Intersection over Union' (IoU), y otros específicos del problema a abordar.

Herramientas software con la implementación de las técnicas de aprendizaje automático.

## Parámetros de contexto de la formación

### Espacios e instalaciones

Los talleres e instalaciones darán respuesta a las necesidades formativas de acuerdo con el contexto profesional establecido en la unidad de competencia asociada, teniendo en cuenta la normativa aplicable del sector productivo, prevención de riesgos laborales, accesibilidad universal y protección medioambiental. Se considerará con carácter orientativo como espacios de uso:

- Instalación de 2 m<sup>2</sup> por alumno o alumna.

### Perfil profesional del formador o formadora:

1. Dominio de los conocimientos y las técnicas relacionados con el entrenamiento de modelos en sistemas de Inteligencia Artificial basados en aprendizaje automático, que se acreditará simultáneamente mediante las dos formas siguientes:

- Formación académica de nivel 2 (Marco Español de Cualificaciones para la Educación Superior) o de otras de superior nivel relacionadas con el campo profesional.

- Experiencia profesional de un mínimo de 2 años en el campo de las competencias relacionadas con este módulo formativo.

2. Competencia pedagógica acreditada de acuerdo con lo que establezcan las Administraciones competentes.