Catálogo de Especialidades Formativas

PROGRAMA FORMATIVO

DATA SCIENCE

Junio 2025

GOBIERNO DE ESPAÑA
MINISTERIO DE TRABAJO Y ECONOMÍA SOCIAL
SERVICIO PÚBLICO DE EMPLEO ESTATAL
SEPE

SISTEMA NACIONAL DE EMPLEO

IDENTIFICACIÓN DE LA ESPECIALIDAD Y PARÁMETROS DEL CONTEXTO FORMATIVO

Denominación de la especialidad:	DATA SCIENCE
Familia Profesional:	INFORMÁTICA Y COMUNICACIONES
Área Profesional:	SISTEMAS Y TELEMÁTICA
Código:	IFCT0144
Nivel de cualificación profesional:	3

Objetivo general

Utilizar las principales arquitecturas y herramientas tecnológicas, así como métodos matemáticos y estadísticos utilizados en ciencia de datos, integrando los conocimientos teóricos y prácticos necesarios para la práctica de la ciencia de datos, tanto en la dimensión de ingeniería como en la de análisis de datos.

Relación de módulos de formación

Módulo 1	Fundamentos de Ciencia de Datos	60 horas
Módulo 2	Métodos Estadísticos y Minería de Datos	100 horas
Módulo 3	Captura, Almacenamiento y Gestión de Datos	120 horas
Módulo 4	Procesamiento Distribuido y Arquitecturas Big Data	140 horas
Módulo 5	Análisis Avanzado y Machine Learning	100 horas
Módulo 6	Visualización de Datos y Presentación de Resultados	80 horas

Modalidades de impartición

Presencial
Mixta

Duración de la formación

Duración total en cualquier modalidad de impartición	600 horas
Mixta	Duración total de la formación presencial: 410 horas

Requisitos de acceso del alumnado

Acreditaciones / titulaciones	Cumplir como mínimo alguno de los siguientes requisitos: -Título de Bachiller o equivalente -Título de Técnico Superior (FP Grado superior) o equivalente -Haber superado la prueba de acceso a Ciclos Formativos de Grado Superior -Certificado de profesionalidad de nivel 3
Experiencia profesional	Se requiere un mínimo de 6 meses de experiencia laboral desarrollando funciones en el sector de TI, análisis de datos o áreas afines, en el caso de no disponer de la titulación requerida.
Otros	El alumnado aspirante deberá demostrar conocimientos básicos

Otros	de informática (al menos sistemas operativos y ofimática básica) y matemáticas o estadística, en caso de no disponer de la titulación o la experiencia requerida.
Modalidad mixta	Además de lo indicado anteriormente, los participantes han de tener las destrezas suficientes para ser usuarios de la plataforma virtual en la que se apoya la acción formativa.

Justificación de los requisitos del alumnado
La acreditación de la titulación se justificará mediante la presentación del título o certificado académico correspondiente. La acreditación de la experiencia profesional, mediante la vida laboral. La acreditación de otros requisitos se realizará mediante una prueba o mediante aportación de cualquier evidencia documental de haber cursado una formación en la materia exigida.

Prescripciones de formadores y tutores

Acreditación requerida	Cumplir como mínimo alguno de los siguientes requisitos: • Licenciado/a, Ingeniero/a, Arquitecto/a o el Título de Grado correspondiente u otros títulos equivalentes en los ámbitos de conocimiento de ingeniería informática y de sistemas, ingeniería eléctrica, ingeniería electrónica, ingeniería de la telecomunicación, ingeniería industrial, matemáticas o estadística. • Diplomado/a, Ingeniero/a Técnico/a, Arquitecto/a Técnico/a o el Título de Grado correspondiente y otros títulos equivalentes en los ámbitos de conocmiento de ingeniería informática y de sistemas, ingeniería eléctrica, ingeniería electrónica, ingeniería de la telecomunicación, ingeniería industrial, matemáticas o estadística.
Experiencia profesional mínima requerida	La experiencia se acreditará en el ámbito de la especialidad (análisis de datos, big data, data sciencie o similar) y será de al menos 3 años, en caso de no disponer de la titulación requerida.
Competencia docente	Se requiere acreditar un mínimo de 500 horas de experiencia docente o 250 horas si la experiencia está relacionada con especialidades o certificados de profesionalidad de la misma familia profesional en la que está encuadrada la presente especialidad.O estar en posesión de alguna de las siguientes titulaciones:-Certificado de Profesionalidad de Docencia de la Formación Profesional para el empleo (SSCE0110) o equivalente.-Certificado de Aptitud Pedagógica (CAP).-Título de licenciado en Pedagogía, Psicopedagogía o de Maestro en cualquiera de sus especialidades, graduado universitario en el ámbito de la Psicología o de la Pedagogía, o un título universitario oficial de posgrado en los citados ámbitos.-Título profesional de Especialización Didáctica, el Certificado de Cualificación Pedagógica, el Máster Universitario habilitante para el ejercicio de las Profesiones Reguladas de Profesor de Educación Secundaria Obligatoria y Bachillerato, Formación Profesional y Escuelas Oficiales de idiomas.

Modalidad mixta	Además de cumplir con las prescripciones establecidas anteriormente, los tutores-formadores deben acreditar una formación, de al menos 30 horas, o experiencia, de al menos 60 horas, en esta modalidad y en la utilización de las tecnologías de la información y comunicación.

Justificación de las prescripciones de formadores y tutores

La acreditación de la titulación se justificará mediante la presentación del título o certificación académica correspondiente. La experiencia profesional mediante vida laboral actualizada o certificado de funciones expedido por organización en la que prestó sus servicios. La competencia docente a través de vida laboral actualizada y/o alguna de las titulaciones acreditativas contempladas.

Requisitos mínimos de espacios, instalaciones y equipamientos

Espacios formativos	Superficie m² para 15 participantes	Incremento Superficie/ participante (Máximo 30 participantes)
Aula de informática	45.0 m²	2.4 m² / participante

Espacio formativo	Equipamiento
Aula de informática	-Mesa y silla para el/la formador/a-Mesas y sillas para el alumnado-Material de aula-Pizarra-PC instalado en red con posibilidad de impresión de documentos, cañón con proyector e internet para el personal formador-PCs instalados en red e internet con posibilidad de impresión para el alumnadoTodos los ordenadores han de tener las siguientes características mínimas:- Procesador i5- Memoria RAM 16 GB - Disco duro 500 GB - Tarjeta de red 10/100/1000 Mbps - Tarjeta gráfica- Tarjeta de sonido - Periféricos: Teclado, Ratón y Monitor colorSoftware específico para el aprendizaje de cada acción formativa:- Sistema operativo (versión/es con soporte oficial)-Navegador-Hadoop-Spark-Hive-Flume-Python o R-SQL y NoSQL-Entornos de desarrollo (Jupyter Notebooks o similar).

La superficie de los espacios e instalaciones estarán en función de su tipología y del número de participantes. Tendrán como mínimo los metros cuadrados que se indican para 15 participantes y el equipamiento suficiente para los mismos.

En el caso de que aumente el número de participantes, hasta un máximo de 30, la superficie de las aulas se incrementará proporcionalmente (según se indica en la tabla en lo relativo a m²/participante) y el equipamiento estará en consonancia con dicho aumento.

No debe interpretarse que los diversos espacios formativos identificados deban diferenciarse necesariamente mediante cerramientos.

Las instalaciones y equipamientos deberán cumplir con la normativa industrial e higiénico-sanitaria correspondiente y responderán a medidas de accesibilidad y seguridad de los participantes.

En el caso de que la formación se dirija a personas con discapacidad se realizarán las adaptaciones y los ajustes razonables para asegurar su participación en condiciones de igualdad.

Aula virtual

Características
• La impartición de la formación mediante aula virtual se ha de estructurar y organizar de forma que se garantice en todo momento que exista conectividad sincronizada entre las personas formadoras y el alumnado participante así como bidireccionalidad en las comunicaciones. • Se deberá contar con un registro de conexiones generado por la aplicación del aula virtual en que se identifique, para cada acción formativa desarrollada a través de este medio, las personas participantes en el aula, así como sus fechas y tiempos de conexión.

Si la especialidad se imparte en **modalidad mixta**, para realizar la parte presencial de la formación, se utilizarán los espacios formativos y equipamientos necesarios indicados anteriormente.

Para impartir la formación en **modalidad mixta**, se ha de disponer del siguiente equipamiento.

Plataforma de teleformación

La plataforma de teleformación que se utilice para impartir acciones formativas deberá alojar el material virtual de aprendizaje correspondiente, poseer capacidad suficiente para desarrollar el proceso de aprendizaje y gestionar y garantizar la formación del alumnado, permitiendo la interactividad y el trabajo cooperativo, y reunir los siguientes requisitos técnicos de infraestructura, software y servicios:

- **Infraestructura:**

  Tener un rendimiento, entendido como número de alumnos que soporte la plataforma, velocidad de respuesta del servidor a los usuarios, y tiempo de carga de las páginas Web o de descarga de archivos, que permita:

  a) Soportar un número de alumnos equivalente al número total de participantes en las acciones formativas de formación profesional para el empleo que esté impartiendo el centro o entidad de formación, garantizando un hospedaje mínimo igual al total del alumnado de dichas acciones, considerando que el número máximo de alumnos por tutor es de 80 y un número de usuarios concurrentes del 40% de ese alumnado.

  b) Disponer de la capacidad de transferencia necesaria para que no se produzca efecto retardo en la comunicación audiovisual en tiempo real, debiendo tener el servidor en el que se aloja la plataforma un ancho de banda mínimo de 300 Mbs, suficiente en bajada y subida.

  Estar en funcionamiento 24 horas al día, los 7 días de la semana.

- **Software:**

- Compatibilidad con el estándar SCORM y paquetes de contenidos IMS.
- Niveles de accesibilidad e interactividad de los contenidos disponibles mediante tecnologías web que como mínimo cumplan las prioridades 1 y 2 de la Norma UNE 139803:2012 o posteriores actualizaciones, según lo estipulado en el capítulo III del Real Decreto 1494/2007, de 12 de noviembre.
- El servidor de la plataforma de teleformación ha de cumplir con los requisitos establecidos en la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales, por lo que el responsable de dicha plataforma ha de identificar la localización física del servidor y el cumplimento de lo establecido sobre transferencias internacionales de datos en los artículos 40 a 43 de la citada Ley Orgánica 3/2018, de 5 de diciembre, así como, en lo que resulte de aplicación, en el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas respecto del tratamiento de datos personales y la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE.
- Compatibilidad tecnológica y posibilidades de integración con cualquier sistema operativo, base de datos, navegador de Internet de los más usuales o servidor web, debiendo ser posible utilizar las funciones de la plataforma con complementos (plug-in) y visualizadores compatibles. Si se requiriese la instalación adicional de algún soporte para funcionalidades avanzadas, la plataforma debe facilitar el acceso al mismo sin coste.
- Disponibilidad del servicio web de seguimiento (operativo y en funcionamiento) de las acciones formativas impartidas, conforme al modelo de datos y protocolo de transmisión establecidos en el anexo V de la Orden/TMS/369/2019, de 28 de marzo.

- **Servicios y soporte:**

- Sustentar el material virtual de aprendizaje de la especialidad formativa que a través de ella se imparta.
- Disponibilidad de un servicio de atención a usuarios que de soporte técnico y mantenga la infraestructura tecnológica y que, de forma estructurada y centralizada, atienda y resuelva las consultas e incidencias técnicas del alumnado. Las formas de establecer contacto con este servicio, que serán mediante teléfono y mensajería electrónica, tienen que estar disponibles para el alumnado desde el inicio hasta la finalización de la acción formativa, manteniendo un horario de funcionamiento de mañana y de tarde y un tiempo de demora en la respuesta no superior a 48 horas laborables.
- Personalización con la imagen institucional de la administración laboral correspondiente, con las pautas de imagen corporativa que se establezcan.

  Con el objeto de gestionar, administrar, organizar, diseñar, impartir y evaluar acciones formativas a través de Internet, la plataforma de teleformación integrará las herramientas y recursos necesarios a tal fin, disponiendo, específicamente, de herramientas de:

  - Comunicación, que permitan que cada alumno pueda interaccionar a través del navegador con el tutor-formador, el sistema y con los demás alumnos. Esta comunicación electrónica ha de llevarse a cabo mediante herramientas de comunicación síncronas (aula virtual, chat, pizarra electrónica) y asíncronas (correo electrónico, foro, calendario, tablón de anuncios, avisos). Será obligatorio que cada acción formativa en modalidad de teleformación disponga, como mínimo, de un servicio de mensajería, un foro y un chat.

- Colaboración, que permitan tanto el trabajo cooperativo entre los miembros de un grupo, como la gestión de grupos. Mediante tales herramientas ha de ser posible realizar operaciones de alta, modificación o borrado de grupos de alumnos, así como creación de «escenarios virtuales» para el trabajo cooperativo de los miembros de un grupo (directorios o «carpetas» para el intercambio de archivos, herramientas para la publicación de los contenidos, y foros o chats privados para los miembros de cada grupo).
- Administración, que permitan la gestión de usuarios (altas, modificaciones, borrado, gestión de la lista de clase, definición, asignación y gestión de permisos, perfiles y roles, autenticación y asignación de niveles de seguridad) y la gestión de acciones formativas.
- Gestión de contenidos, que posibiliten el almacenamiento y la gestión de archivos (visualizar archivos, organizarlos en carpetas –directorios- y subcarpetas, copiar, pegar, eliminar, comprimir, descargar o cargar archivos), la publicación organizada y selectiva de los contenidos de dichos archivos, y la creación de contenidos.
- Evaluación y control del progreso del alumnado, que permitan la creación, edición y realización de pruebas de evaluación y autoevaluación y de actividades y trabajos evaluables, su autocorrección o su corrección (con retroalimentación), su calificación, la asignación de puntuaciones y la ponderación de las mismas, el registro personalizado y la publicación de calificaciones, la visualización de información estadística sobre los resultados y el progreso de cada alumno y la obtención de informes de seguimiento.

Material virtual de aprendizaje:

El material virtual de aprendizaje para el alumnado mediante el que se imparta la formación se concretará en el curso completo en formato multimedia (que mantenga una estructura y funcionalidad homogénea), debiendo ajustarse a todos los elementos de la programación (objetivos y resultados de aprendizaje) de este programa formativo que figura en el Catálogo de Especialidades Formativas y cuyo contenido cumpla estos requisitos:

- Como mínimo, ser el establecido en el citado programa formativo del Catálogo de Especialidades Formativas.
- Estar referido tanto a los objetivos como a los conocimientos/ capacidades cognitivas y prácticas, y habilidades de gestión, personales y sociales, de manera que en su conjunto permitan conseguir los resultados de aprendizaje previstos.
- Organizarse a través de índices, mapas, tablas de contenido, esquemas, epígrafes o titulares de fácil discriminación y secuenciase pedagógicamente de tal manera que permiten su comprensión y retención.
- No ser meramente informativos, promoviendo su aplicación práctica a través de actividades de aprendizaje (autoevaluables o valoradas por el tutor-formador) relevantes para la adquisición de competencias, que sirvan para verificar el progreso del aprendizaje del alumnado, hacer un seguimiento de sus dificultades de aprendizaje y prestarle el apoyo adecuado.
- No ser exclusivamente textuales, incluyendo variados recursos (necesarios y relevantes), tanto estáticos como interactivos (imágenes, gráficos, audio, video, animaciones, enlaces, simulaciones, artículos, foro, chat, etc.). de forma periódica.
- Poder ser ampliados o complementados mediante diferentes recursos adicionales a los que el alumnado pueda acceder y consultar a voluntad.

- Dar lugar a resúmenes o síntesis y a glosarios que identifiquen y definan los términos o vocablos básicos, relevantes o claves para la comprensión de los aprendizajes.
- Evaluar su adquisición durante y a la finalización de la acción formativa a través de actividades de evaluación (ejercicios, preguntas, trabajos, problemas, casos, pruebas, etc.), que permitan medir el rendimiento o desempeño del alumnado.

Ocupaciones y puestos de trabajo relacionados

- 21121032 MÉDICOS ESPECIALISTAS EN ANÁLISIS CLÍNICOS

- 21401015 FARMACÉUTICOS DE ANÁLISIS CLÍNICOS

- 22201281 PROFESORES TÉCNICOS DE FORMACIÓN PROFESIONAL (INFORMÁTICA Y COMUNICACIONES)
- 75311119 INSTALADORES DE EQUIPOS Y SISTEMAS ELECTRÓNICOS

- 75311128 INSTALADORES ELECTRÓNICOS DE MICROSISTEMAS

- 75331058 INSTALADORES DE EQUIPOS Y SISTEMAS DE COMUNICACIÓN

- 75331067 INSTALADORES DE EQUIPOS Y SISTEMAS DE TELECOMUNICACIÓN

Requisitos oficiales de las entidades o centros de formación

Estar inscrito en el Registro de entidades de formación (Servicios Públicos de Empleo).

DESARROLLO MODULAR

MÓDULO DE FORMACIÓN 1: Fundamentos de Ciencia de Datos

OBJETIVO

Aprender los conceptos fundamentales de la ciencia de datos y su importancia en el entorno actual, los principios de la recolección, manipulación y análisis de datos y conocer las principales herramientas y entornos utilizados en la ciencia de datos.

DURACIÓN EN CUALQUIER MODALIDAD DE IMPARTICIÓN: 60 horas

Mixta: Duración de la formación presencial: 40 horas

RESULTADOS DE APRENDIZAJE

Conocimientos / Capacidades cognitivas y prácticas

- Comprensión de conceptos clave de la Ciencia de Datos
  - Historia y evolución de la Ciencia de Datos.
  - Conceptos clave: datos, información y conocimiento.
  - Rol del Científico de Datos en la organización.
- Identificación de Tipos y Fuentes de Datos
  - Datos estructurados, no estructurados y semi-estructurados.
  - Fuentes de datos: bases de datos, APIs, sensores, redes sociales.
  - Importancia de la calidad de los datos.
- Uso de Herramientas y Entornos de Ciencia de Datos

- Introducción a lenguajes de programación: Python, R.
- Herramientas: Jupyter, RStudio.
- Integración con entornos colaborativos (Git, notebooks compartidos).

- Comprensión del proceso general de trabajo en ciencia de datos

- Definición de objetivos.
- Preparación y limpieza de datos (Data Wrangling).
- Exploración y visualización inicial.
- Introducción a CRISP-DM y metodologías similares.

- Aplicación de medidas de eficiencia energética y sostenibilidad ambiental en el trabajo con grandes volúmenes de datos y presentación de resultados.

Habilidades de gestión, personales y sociales

- Coordinación en actividades prácticas de trabajo colaborativo para resolver problemas simples de análisis de datos.
- Preocupación por el seguimiento y cumplimiento de políticas y regulaciones de seguridad.
- Preocupación por el seguimiento y cumplimiento de políticas de protección de datos.
- Trabajo buscando la eficiencia energética (reducción del uso de memoria volátil y persistente, uso de entornos de desarrollo con impacto reducido en el consumo de recursos, etc.).

Resultados que tienen que adquirirse en presencial

Deberán realizarse de forma presencial las siguientes actividades:

--Uso de Herramientas y Entornos de Ciencia de Datos-Introducción a lenguajes de programación: Python, R.-Herramientas: Jupyter, RStudio.-Integración con entornos colaborativos (Git, notebooks compartidos).-Comprensión del proceso general de trabajo en ciencia de datos -Definición de objetivos.-Preparación y limpieza de datos (Data Wrangling).-Exploración y visualización inicial.-Introducción a CRISP-DM y metodologías similares.

MÓDULO DE FORMACIÓN 2: Métodos Estadísticos y Minería de Datos

OBJETIVO

Aplicar métodos estadísticos aplicados al análisis de datos, comprendiendo las técnicas de minería de datos y su aplicación en la extracción de información útil y utilizar modelos estadísticos para resolver problemas prácticos en ciencia de datos.

DURACIÓN EN CUALQUIER MODALIDAD DE IMPARTICIÓN: 100 horas

Mixta: Duración de la formación presencial: 70 horas

RESULTADOS DE APRENDIZAJE

Conocimientos / Capacidades cognitivas y prácticas

- Conocimiento de fundamentos Estadísticos para Ciencia de Datos
  - Estadística descriptiva: medidas de centralidad y dispersión.
  - Inferencia estadística: estimación, intervalos de confianza.
  - Pruebas de hipótesis: t-test, chi-cuadrado, ANOVA.
- Utilización de métodos de Minería de Datos
  - Clasificación: regresión logística, árboles de decisión.
  - Clustering: k-means, DBSCAN, jerárquico.
  - Reglas de asociación y análisis de cesta de mercado.
- Uso de técnicas de Reducción de Dimensionalidad
  - Análisis de componentes principales (PCA).
  - Análisis factorial.
  - Selección de características.
- Visualización de Datos
  - Herramientas y librerías: Matplotlib, Seaborn, ggplot2.
  - Creación de dashboards interactivos.
  - Mejores prácticas en visualización.

Habilidades de gestión, personales y sociales

- Trabajo en equipo para desarrollar modelos de análisis en conjunto y evaluar resultados.
- Pensamiento analítico.

Resultados que tienen que adquirirse en presencial

Deberán realizarse de forma presencial las siguientes actividades:

--Utilización de métodos de Minería de Datos -Clasificación: regresión logística, árboles de decisión.-Clustering: k-means, DBSCAN, jerárquico.-Reglas de asociación y análisis de cesta de mercado.-Uso de técnicas de Reducción de Dimensionalidad- Análisis de componentes principales (PCA).-Análisis factorial.-Selección de características.-Visualización de Datos -Herramientas y librerías: Matplotlib, Seaborn, ggplot2.-Creación de dashboards interactivos.-Mejores prácticas en visualización.

MÓDULO DE FORMACIÓN 3: Captura, Almacenamiento y Gestión de Datos

OBJETIVO

Obtener y almacenar datos desde múltiples fuentes, comprendiendo las arquitecturas y tecnologías de bases de datos convencionales y no convencionales e introducir conceptos de privacidad y protección de datos.

DURACIÓN EN CUALQUIER MODALIDAD DE IMPARTICIÓN: 120 horas

Mixta: Duración de la formación presencial: 80 horas

RESULTADOS DE APRENDIZAJE

Conocimientos / Capacidades cognitivas y prácticas

- Conocimiento de los fundamentos de Bases de Datos
  - Modelos de bases de datos: relacionales y NoSQL.
  - Lenguaje SQL para análisis de datos.
  - Introducción a bases de datos NoSQL: MongoDB, Cassandra.
- Extracción y Captura de Datos
  - Web scraping con Python: Beautiful Soup, Selenium.
  - APIs REST y SOAP.
  - Datos en tiempo real: Flume, Kafka.
- Gestión y Transformación de Datos
  - ETL (Extract, Transform, Load): concepto y herramientas.
  - Limpieza y normalización de datos.
  - Tratamiento de datos faltantes y valores atípicos.
- Gobierno y Seguridad de los Datos
  - Gestión de la privacidad: GDPR, CCPA.
  - Calidad y linaje de datos.
  - Herramientas para la gobernanza de datos.

Habilidades de gestión, personales y sociales

- Capacidad para gestionar proyectos de recopilación de datos en equipos colaborativos.
- Preocupación por el seguimiento y cumplimiento de políticas y regulaciones de seguridad.
- Preocupación por el seguimiento y cumplimiento de políticas de protección de datos.

Resultados que tienen que adquirirse en presencial

Deberán realizarse de forma presencial las siguientes actividades:

--Extracción y Captura de Datos-Web scraping con Python: Beautiful Soup, Selenium.-APIs REST y SOAP.-Datos en tiempo real: Flume, Kafka.-Gestión y Transformación de Datos -ETL (Extract, Transform, Load): concepto y herramientas.-Limpieza y normalización de datos.-Tratamiento de datos faltantes y valores atípicos.-Gobierno y Seguridad de los Datos-Gestión de la privacidad: GDPR, CCPA.-Calidad y linaje de datos.-Herramientas para la gobernanza de datos.

MÓDULO DE FORMACIÓN 4: Procesamiento Distribuido y Arquitecturas Big Data

OBJETIVO

Conocer los sistemas distribuidos y su importancia en el procesamiento de grandes volúmenes de datos, utilizar plataformas de procesamiento como Hadoop y Spark y aplicar arquitecturas Big Data para resolver problemas empresariales y técnicos.

DURACIÓN EN CUALQUIER MODALIDAD DE IMPARTICIÓN: 140 horas

Mixta: Duración de la formación presencial: 100 horas

RESULTADOS DE APRENDIZAJE

Conocimientos / Capacidades cognitivas y prácticas

- Conocimiento de los conceptos fundamentales y el ecosistema Big Data
    - Conceptos fundamentales: 5 V's (Volumen, Velocidad, Variedad, Veracidad, Valor).
    - Ecosistema Hadoop: HDFS, MapReduce.
    - Comparativa entre Hadoop y Spark.
- Identificación de sistemas de Procesamiento Distribuido
    - Apache Spark: RDDs, DataFrames y Spark SQL.
    - Procesamiento por lotes vs procesamiento en tiempo real.
    - Apache Flink y Storm.
- Uso de herramientas y tecnologías Big Data
    - Sistemas de almacenamiento: HBase, Cassandra.
    - Sistemas de mensajería: Kafka, RabbitMQ.
    - Data Lakes y Warehousing.
- Implementación de Arquitecturas
    - Diseño de pipelines de datos.
    - Arquitecturas Lambda y Kappa.
    - Casos prácticos: desde la captura hasta la visualización.

Habilidades de gestión, personales y sociales

- Coordinación en la gestión de proyectos distribuidos donde cada miembro asume una parte del procesamiento de datos.
- Gestión eficaz del tiempo.

Resultados que tienen que adquirirse en presencial

Deberán realizarse de forma presencial las siguientes actividades:

--Uso de herramientas y tecnologías Big Data-Sistemas de almacenamiento: HBase, Cassandra.-Sistemas de mensajería: Kafka, RabbitMQ.-Data Lakes y Warehousing.-Implementación de Arquitecturas-Diseño de pipelines de datos.-Arquitecturas Lambda y Kappa.-Casos prácticos: desde la captura hasta la visualización.

MÓDULO DE FORMACIÓN 5: Análisis Avanzado y Machine Learning

OBJETIVO

Aplicar técnicas avanzadas de análisis de datos para la creación de modelos predictivos y descriptivos, concoer los conceptos de Machine Learning supervisado y no supervisado e implementar modelos de Machine Learning y evaluarlos en contextos empresariales.

DURACIÓN EN CUALQUIER MODALIDAD DE IMPARTICIÓN: 100 horas

Mixta: Duración de la formación presencial: 70 horas

RESULTADOS DE APRENDIZAJE

Conocimientos / Capacidades cognitivas y prácticas

- Conocimiento de los elementos básicos de Machine Learning
  - Tipos de aprendizaje: supervisado, no supervisado y por refuerzo.
  - Métricas de evaluación: precisión, recall, F1-score.
  - Overfitting y regularización.
- Creación de modelos de Machine Learning
  - Modelos supervisados: regresión lineal, SVM, random forests.
  - Redes neuronales básicas.
  - Modelos no supervisados: clustering, reducción de dimensionalidad.
- Deep Learning y Redes Neuronales Avanzadas
  - Introducción a TensorFlow y PyTorch.
  - Redes convolucionales (CNN) y recurrentes (RNN).
  - Transfer Learning.
- Optimización y Deployment
  - Técnicas de optimización de modelos: Grid Search, Random Search.
  - Implementación de modelos en producción.
  - Herramientas para la gestión de modelos: MLflow, Docker.

Habilidades de gestión, personales y sociales

- Capacidad de gestión de proyectos de análisis y evaluación colaborativa de modelos predictivos.

Resultados que tienen que adquirirse en presencial

Deberán realizarse de forma presencial las siguientes actividades:

--Creación de modelos de Machine Learning -Modelos supervisados: regresión lineal, SVM, random forests.-Redes neuronales básicas.-Modelos no supervisados: clustering, reducción de dimensionalidad.-Deep Learning y Redes Neuronales Avanzadas -Introducción a TensorFlow y PyTorch.-Redes convolucionales (CNN) y recurrentes (RNN).-Transfer Learning.-Optimización y Deployment-Técnicas de optimización de modelos: Grid Search, Random Search.-Implementación de modelos en producción.-Herramientas para la gestión de modelos: MLflow, Docker.

MÓDULO DE FORMACIÓN 6: **Visualización de Datos y Presentación de Resultados**

OBJETIVO

Crear visualizaciones efectivas que comuniquen los resultados de análisis complejos y elaborar informes comprensibles y visualmente atractivos basados en los resultados obtenidos.

DURACIÓN EN CUALQUIER MODALIDAD DE IMPARTICIÓN: 80 horas

Mixta: Duración de la formación presencial: 50 horas

RESULTADOS DE APRENDIZAJE

Conocimientos / Capacidades cognitivas y prácticas

- Visualización de Datos
  - Principios básicos de diseño visual: claridad, simplicidad y narrativa.
  - Tipos de gráficos y sus aplicaciones: barras, líneas, áreas, dispersión.
  - Introducción a la percepción visual y cómo influye en la comprensión de datos.
- Uso de herramientas de Visualización de Datos
  - Librerías en Python: Matplotlib, Seaborn, Plotly.
  - Uso de herramientas especializadas: Tableau, Power BI.
  - Creación de gráficos interactivos para la web con D3.js.
- Diseño de Dashboards y Reportes Interactivos
  - Diseño y desarrollo de dashboards efectivos.
  - Integración de múltiples fuentes de datos.
  - Exportación y automatización de reportes en Excel y PDF.
- Creación de Storytelling con Datos
  - Creación de narrativas visuales para comunicar hallazgos.
  - Uso de infografías y diagramas para reforzar mensajes clave.
  - Personalización del mensaje según la audiencia.
- Identificación de mejores Prácticas y Ética en la Presentación de Datos
  - Evitar sesgos en la presentación de datos.
  - Identificación y prevención de malas prácticas (gráficos engañosos).
  - Adaptación del diseño visual a criterios de accesibilidad.

Habilidades de gestión, personales y sociales

- Presentación de resultados y comunicación efectiva de análisis complejos a audiencias no técnicas.
- Vigilancia de ahorro energético y la reducción del impacto ambiental en la elaboración de presentaciones (impresión, papel, etc.).
- Cumplimiento de normas éticas y legales.

Resultados que tienen que adquirirse en presencial

Deberán realizarse de forma presencial las siguientes actividades:

--Uso de herramientas de Visualización de Datos -Librerías en Python: Matplotlib, Seaborn, Plotly.-Uso de herramientas especializadas: Tableau, Power BI.-Creación

de gráficos interactivos para la web con D3.js.-Diseño de Dashboards y Reportes Interactivos-Diseño y desarrollo de dashboards efectivos.-Integración de múltiples fuentes de datos.-Exportación y automatización de reportes en Excel y PDF.- Creación de Storytelling con Datos -Creación de narrativas visuales para comunicar hallazgos.-Uso de infografías y diagramas para reforzar mensajes clave.- Personalización del mensaje según la audiencia.

EVALUACIÓN DEL APRENDIZAJE EN LA ACCIÓN FORMATIVA

- La evaluación tendrá un carácter teórico-práctico y se realizará de forma sistemática y continua, durante el desarrollo de cada módulo y al final del curso.
- Puede incluir una evaluación inicial de carácter diagnóstico para detectar el nivel de partida del alumnado.
- La evaluación se llevará a cabo mediante los métodos e instrumentos más adecuados para comprobar los distintos resultados de aprendizaje, y que garanticen la fiabilidad y validez de la misma.
- Cada instrumento de evaluación se acompañará de su correspondiente sistema de corrección y puntuación en el que se explicite, de forma clara e inequívoca, los criterios de medida para evaluar los resultados alcanzados por los participantes.
- La puntuación final alcanzada se expresará en términos de Apto/ No Apto.