

***Máster de Formación Permanente en
BigData e Inteligencia Artificial***



**Universidad Autónoma
de Madrid**

Título Propio de la Universidad Autónoma de Madrid

Escuela Politécnica Superior

Universidad Autónoma de Madrid

Estructura en créditos ECTS

Fundamentos: sistemas y arquitecturas (Online)	3 ECTS
Fundamentos: lenguajes (Online)	3 ECTS
Bases de datos NoSQL (Online)	3 ECTS
Estadística	3 ECTS
Fundamentos de análisis de datos y del aprendizaje automático	5 ECTS
Infraestructura para Big Data e Inteligencia Artificial	5 ECTS
Ciclo de vida analítico, fuentes de datos y aprovisionamiento	5 ECTS
Procesamiento de datos con Spark	3 ECTS
Explotación y visualización de datos	3 ECTS
Seguridad, privacidad y protección de datos	3 ECTS
Tecnologías en Inteligencia Artificial	4 ECTS
Aplicaciones de Inteligencia Artificial en contenidos multimedia y biometría	5 ECTS
Inteligencia Artificial Generativa y Explicativa	3 ECTS
Trabajo Fin de Máster	12 ECTS
	60 ECTS

Programa

Fundamentos: sistemas y arquitecturas (ONLINE plataforma UAMx)

1. Sistema Operativo Linux
 - 1.4. Conceptos generales de Linux
 - 1.5. Comandos, variables de entorno, scripts
 - 1.6. Control y planificación de procesos.
 - 1.7. Sistemas de almacenamiento y sistemas de ficheros
 - 1.8. Administración básica de Linux
2. Redes de comunicación
 - 2.1. Componentes y tipos de redes.
 - 2.2. Conceptos básicos: Direccionamiento IP, máscara de red, puerta de enlace, servidor de nombres (DNS), direccionamiento dinámico DHCP.
 - 2.3. Configuración de una red TCP/IP en Linux.
 - 2.4. Acceso remoto a equipos y ficheros: ssh, ftp
 - 2.5. Conceptos básicos de seguridad: Claves pública y privada, VPN.
3. Máquinas Virtuales
 - 3.1. Concepto de virtualización
 - 3.2. Tipos de virtualización de plataforma
 - 3.3. Instalación y gestión de una Máquina Virtual
 - 3.4. Creación de máquinas virtuales con Vagrant
 - 3.5. Infraestructura como Servicio (IaaS): máquinas virtuales bajo demanda y con capacidades actualizables en tiempo real
4. Cluster de ordenadores
 - 4.1. Multicomputador
 - 4.2. Clusters de ordenadores
 - 4.3. Construir, desplegar y gestionar un cluster
 - 4.4. Planificación y balanceo de tareas
 - 4.5. Tipos de Cluster y aplicaciones



Fundamentos: lenguajes (ONLINE plataforma UAMx)

1. Python
 - 1.1. Introducción a python
 - 1.1.1. Instalación
 - 1.1.2. Intérpretes: python, ipython, notebooks
 - 1.1.3. Diferencias entre versiones
 - 1.2. Tipos básicos: cadenas, listas, diccionarios, tuplas, etc.
 - 1.3. Funciones, funciones lambda e imports.
 - 1.4. Sentencias de control e iteración
 - 1.4.1. Loops e ifs
 - 1.4.2. Algunas formas de utilizar programación funcional: map, reduce.
 - 1.5. Entrada y salida de ficheros
 - 1.6. Programación orientada a objetos en Python
2. Librerías:
 - 2.1. numpy,
 - 2.2. matplotlib,
 - 2.3. pandas.

3. Lenguaje R
 - 3.1. Introducción a R
 - 3.2. Objetos y atributos en R
 - 3.3. Vectores
 - 3.4. Arrays
 - 3.5. Listas
 - 3.6. Data frames
 - 3.7. Lecturas de ficheros
 - 3.8. Funciones y sentencias de control
 - 3.9. Gráficas
 - 3.10. Depuración y medición de tiempos
 - 3.11. Notebooks

Bases de datos NoSQL (ONLINE plataforma UAMx)

1. Introducción a las bases de datos NoSQL
 - 1.1. ¿Qué son?
 - 1.2. Tipos de BBDD NoSQL
 - 1.3. Ventajas y desventajas
2. Base de Datos MongoDB
 - 2.1. Introducción
 - 2.2. Organización de los datos
 - 2.3. Manejo básico de los datos
 - 2.4. Métodos básicos de agregación
 - 2.5. MapReduce
 - 2.6. Aggregation Framework
 - 2.7. Uso de índices
3. Base de Datos Redis
 - 3.1. Estructuras de datos
 - 3.2. Programación en Lua
 - 3.3. Bibliotecas de Lua y depuración de Scripts Lua
4. Base de Datos Cassandra
 - 4.1. Introducción
 - 4.2. Cassandra Query Language
5. Base de datos Neo4j
 - 5.1. Introducción
 - 5.2. Lenguaje de consulta Cypher



Universidad Autónoma
de Madrid

Estadística

1. Introducción
 - 1.1. ¿Qué es la estadística?
 - 1.2. Modelo estadístico
 - 1.3. Método estadístico
 - 1.4. Algunas herramientas de análisis de datos mediante estadística
2. Descripción de los datos
 - 2.1. Descripción de una variable
 - 2.2. Descripción multivariante
3. Modelos en estadística
 - 3.1. Probabilidad y variables aleatorias
 - 3.2. Modelos univariantes de distribución de probabilidad
 - 3.3. Modelos multivariantes de distribución de probabilidad

4. Inferencia Estadística
 - 4.1. Estimación puntual
 - 4.2. Estimación por intervalos
 - 4.3. Estimación bayesiana
 - 4.4. Contraste de hipótesis

Fundamentos de análisis de datos y del aprendizaje automático

1. Introducción al aprendizaje automático
 - 1.1. Tipos de aprendizaje automático, conceptos básicos, tipos de atributos
 - 1.2. Flujo de un proyecto de aprendizaje automático
 - 1.3. Validación de modelos: tasas de error, matriz de confusión, curvas ROC y validación cruzada
 - 1.4. Regresión lineal, regresión logística
 - 1.5. Vecinos próximos en clasificación y regresión
 - 1.6. Sesgo y varianza. Maldición de la dimensionalidad
2. Preprocesado de datos
 - 2.1. Construcción de la base de datos; tratamiento de múltiples fuentes
 - 2.2. Preparación y auditoría de la base de datos
 - 2.3. Distribución de las variables.
 - 2.4. Reducción de la dimensionalidad
 - 2.5. Información no estructurada; casos prácticos
3. Aprendizaje automático
 - 3.1. Modelos predictivos básicos: Regresión lineal y logística. Vecinos Próximos en clasificación y regresión. Sesgo y varianza
 - 3.2. Clustering: K-means, Clúster jerárquico, basado en mezclas de Gaussianas EM, basado en densidades DBSCAN
 - 3.3. Conjuntos de clasificadores y árboles de decisión: Algoritmos de construcción de árboles de decisión, poda, CART, C4.5.
 - 3.4. Conjuntos de clasificadores (aleatorización): bagging, Random Forests y variantes.
 - 3.5. Conjuntos de clasificadores (optimización): boosting y Gradiente Boosting
 - 3.6. Clasificación y regresión con máquinas de vectores soporte SVM
 - 3.7. Clasificación de margen máximo. Regresión de margen máximo. SVM sobre núcleos lineales

Infraestructura para Big Data e Inteligencia Artificial

1. Arquitecturas para tratar grandes volúmenes de información
 - 1.1. Evolución de las tecnologías: Multicore, GPUs, FPGA.
 - 1.2. Arquitecturas de referencia para Bigdata e Inteligencia Artificial.
2. Supervisión y mantenimiento de un clúster para grandes cantidades de datos.
 - 2.1. Configuración inicial del clúster
 - 2.2. Gestión de logs y diagnóstico
 - 2.3. Administración y mantenimiento
 - 2.4. Seguridad, autenticación y autorización
 - 2.5. Planificación de trabajos y configuración avanzada
 - 2.6. Monitorización del cluster
3. Principios de programación paralela y distribuida en Python
 - 3.1. Paralelismo en CPU
 - 3.2. Aceleración explotando el paralelismo de GPU
 - 3.3. Computación distribuida en clúster
4. Evaluación de prestaciones y optimización
 - 4.1. Benchmarking y tuneado de parámetros en un caso práctico
 - 4.2. Medidas de rendimiento en un caso real: análisis de tráfico DNS y HTTP en una red de comunicaciones

5. Infraestructura para el procesamiento de datos con Spark
 - 5.1. Arquitectura de un sistema Spark. Estructura interna y flujos de datos
 - 5.2. Modos de ejecución en Spark: local vs cluster. Ámbitos y contextos: driver, ejecutores
 - 5.3. Gestión de memoria
 - 5.4. Modos de ejecución en cluster. Comunicación entre nodos
 - 5.5. Ciclo de vida de un programa Spark
 - 5.6. Configuración
 - 5.7. Ejecución de tareas: spark-submit dependencias, REPL (spark-shell, pyspark), notebooks
 - 5.8. Acceso a datos: E/S (local, HDFS, etc).
 - 5.9. Interfaces de monitorización y análisis: Spark UI, Spark Master, Spark HistoryServer
6. Virtualización de infraestructura
 - 6.1. Infraestructura local vs Cloud
 - 6.2. Infraestructura como Servicio (IaaS)
 - 6.3. Cloud privado: Propuestas Openstack y OpenNebula
 - 6.4. Cloud público: Propuestas Amazon AWS, Google Cloud y Microsoft Azure
 - 6.5. Cloud público vs Cloud privado
7. Virtualización basada en contenedores
 - 7.1. Diseño de aplicaciones en contenedores
 - 7.2. Gestión de imágenes y versiones
 - 7.3. Orquestación y Comunicación
 - 7.4. Seguridad
8. Nuevas plataformas: computación cuántica
 - 8.1. Principios de la computación cuántica
 - 8.2. Qubits, puertas y circuitos cuánticos
 - 8.3. Paralelismo de procesamiento. Entrelazamiento. Reversibilidad.
 - 8.4. Computación cuántica basada en circuitos vs optimizadores cuánticos (Quantum Annealing)
 - 8.5. Algoritmos cuánticos


Universidad Autónoma
de Madrid

Ciclo de vida analítico, fuentes de datos y aprovisionamiento

1. Proyecto Apache Hadoop
 - 1.1. Sistemas de ficheros distribuido HDFS
 - 1.2. Modelo de Programación MapReduce
 - 1.3. Desarrollo con Pig
2. Persistencia de datos
 - 2.1. Hive: data warehouse sobre Hadoop, HiveQL
 - 2.2. Hbase: componentes y lenguaje de consulta
 - 2.3. Beneficios y compatibilidad con otros formatos de persistencia: Parquet, ORC
3. Análisis de datos en movimiento
 - 3.1. Apache Storm: componentes, despliegue y topologías
 - 3.2. Apache Flink
4. Gobierno del Dato
 - 4.1. Necesidad y objetivos de la gobernanza de datos
 - 4.2. Metadatos, linaje del dato y terminología
 - 4.3. Plataformas de Gobierno y calidad del dato: Apache Atlas y Apache Griffin
5. Fuentes de datos y descubrimiento
 - 5.1. Internet de las cosas
 - 5.2. Conceptos y escenarios
 - 5.3. Plataformas IoT
 - 5.4. Casos de éxito IoT
 - 5.5. Industria 4.0

6. Aproveccionamiento, tecnologías para captura y modificación de datos
 - 6.1. Apache Nifi, componentes y desarrollo
 - 6.2. Apache Kafka, escenarios de aprovisionamiento, despliegue y configuración

Procesamiento de datos con Spark

1. Fundamentos de Spark
 - 1.1. Introducción: arquitectura y organización
 - 1.2. Datos en Spark: Resilient Distributed Datasets (RDDs)
 - 1.3. Flujo de un programa spark
 - 1.4. Entrada y salida de datos
 - 1.5. Transformaciones
 - 1.6. Persistencia
 - 1.7. Acciones
 - 1.8. Variables compartidas: broadcast y acumuladores
2. Tuning en Spark
 - 2.1. Vista general de las APIs ofrecidas por Spark: Scala, Java, Python, R
 - 2.2. SparkR: paralelización de DataFrames de R
3. Spark SQL
 - 3.1. Introducción a DataFrames
 - 3.2. Fuentes de datos: Hive, JDBC/ODBC, etc
 - 3.3. API de DataFrames
4. Procesamiento de grafos vía Spark
 - 4.1. Introducción general a los operadores sobre grafos
 - 4.2. Grafos en Spark: GraphX
 - 4.3. Paquetes adicionales para Spark: GraphFrames
 - 4.4. Algoritmos de grafos sobre GraphFrames
5. Procesado en tiempo real: Spark Streaming
 - 5.1. Spark Streaming clásico: Discretized Streams (DStreams)
 - 5.2. Operaciones con DStreams: estado, robustez, ventanas
 - 5.3. Streaming estructurado
 - 5.4. Operaciones sobre streaming estructurado: flujos, ventanas, entregas
 - 5.5. Fuentes de datos para streaming: Kafka
 - 5.6. Machine Learning sobre datos en streaming
6. Machine Learning: Spark ML
 - 6.1. Aprendizaje supervisado: clasificación y regresión
 - 6.2. Aprendizaje no supervisado
 - 6.3. Creación de pipelines de aprendizaje automático

Explotación y visualización de datos

1. Introducción, importancia de la visualización
 - 1.1 Visualización gráfica eficiente de datos
 - 1.2 Sistemas de coordenadas
 - 1.3 Color
 - 1.4 Finalidad y alcance de la visualización
2. Tipos de Visualizaciones
 - 2.1 Cantidades
 - 2.2 Distribuciones
 - 2.3 Proporciones
 - 2.4 Asociaciones variables cuantitativas
 - 2.5 Series temporales y tendencias

3. IBM Cognos Analytics
 - 3.1 Acceso a datos desde Cognos Analytics
 - 3.2 Creación de consolas gráficas
 - 3.3 Análisis de datos integrado
 - 3.4 Funciones de IA en Cognos Analytics
4. Tableau
 - 4.1 Acceso a datos desde Tableau
 - 4.2 Hojas y Dashboards
 - 4.3 Patrones temporales
 - 4.4 Información espacial y geográfica
 - 4.5 Gráficos interactivos, filtros
5. Grafana
 - 5.1 Conceptos básicos: roles, datasources y dashboards
 - 5.2 Templating y automatización
 - 5.3 Alerting y plugins
6. Explotación de datos
 - 6.1 Indexación y almacenamiento no estructurados con Elasticsearch
 - 6.2 casos de uso: ElasticSearch y Grafana

Seguridad, privacidad y protección de datos

1. Introducción al concepto de privacidad
2. Tecnologías criptográficas para la protección de la privacidad
 - 2.1. Fundamentos criptográficos de la protección de la información
 - 2.2. Problema de la gestión de la identidad digital mediante certificados digitales: el estándar X.509
 - 2.3. Definición de los conceptos de trazabilidad, enlazado, anonimato y pseudo-anonimato: firmas grupales
 - 2.4. Navegación anónima: introducción a las redes de mezcla de tráfico, *onion routing* y ofuscación de tráfico
3. Privacidad como control estadístico del acceso a datos
 - 3.1. Descripción de datos y metadatos que permiten la re-identificación de individuos
 - 3.2. Valor y riesgo asociado a los datos y metadatos accesibles en abierto (Open Data)
4. Marco jurídico de protección de datos y normas de transferencia internacionales
 - 4.1. Introducción al Reglamento General de Protección de Datos (RGPD).
 - 4.2. Privacy Shield Framework (transferencia de datos entre EE.UU. y UE).
 - 4.3. Privacy Directive ('cookie law')
5. Implicaciones ético-legales de la inclusión de Big Data en la toma de decisión: gobernanza y rendición de cuentas
 - 5.1. La gobernanza de los algoritmos como problema epistémico vs. moral o como problema técnico vs. de gestión o legal
 - 5.2. Actual marco legal europeo
 - 5.3. Introducción al debate sobre Algorithmic Fairness, Accountability and Transparency (FAT*) a través de distintos casos de estudio.
 - 5.4. Consideraciones sobre el conjunto de datos y la selección de modelos de aprendizaje automático: riesgos de manipulación, problemas de predicción y path dependency
6. Sesgos en Inteligencia Artificial

Tecnologías en Inteligencia Artificial

1. Aprendizaje con redes neuronales clásicas
 - 1.1. Perceptrones de Rosenblatt
 - 1.2. Perceptrones multicapa
 - 1.3. Estrategias para aprendizaje de perceptrones
2. Aprendizaje con redes neuronales profundas (Deep Learning)
 - 2.1. Innovaciones, función de activación, función de coste
 - 2.2. Autocodificadores
 - 2.3. Convolución
 - 2.4. Técnicas para evitar sobreajuste
3. Marcos de desarrollo para IA: TensorFlow y Keras
 - 3.1. Introducción a Keras
 - 3.2. Arquitectura de TensorFlow: Tensores, grafos, diagrama de flujo de datos
 - 3.3. Diferenciación automática y optimización con TensorFlow
 - 3.4. Introducción a Keras
 - 3.5. Aplicaciones prácticas utilizando TensorFlow - Keras: autoencoders, redes neuronales convolucionales (CNN), redes neuronales recurrentes, GAN, aprendizaje por refuerzo.
4. Predicción con series temporales
 - 4.1. Definición de serie temporal
 - 4.2. Análisis de series temporales: visualización, tendencia, estacionalidad y estacionaridad.
 - 4.3. Predicción usando modelos clásicos autorregresivos: AR, MA, ARMA, ARIMA, SARIMA.
 - 4.4. Modelos de aprendizaje profundo para series temporales
5. Computación cognitiva
 - 5.1. Bases de la Computación Cognitiva. IBM Watson.
 - 5.2. Estrategia de Soluciones Cognitivas de IBM.
 - 5.3. Infraestructura para Soluciones Cognitivas.
 - 5.4. Impacto social: servicios cognitivos a través de IBM Cloud.



Universidad Autónoma
de Madrid

Aplicaciones de IA en contenidos multimedia y biometría

1. Multimedia (imagen, video). Aplicaciones de IA en el tratamiento de Imagen
 - 1.1. Introducción al tratamiento de imagen y vídeo
 - 1.2. Extracción de características en señales visuales: descriptores en imagen y vídeo
 - 1.3. Descriptores globales: color, puntos de interés. Descriptores a nivel de región segmentada: color, puntos de interés, textura, forma. Descriptores de movimiento: movimiento global, trayectorias
 - 1.4. Estrategias de indexación y búsqueda de contenido visual
 - 1.5. Descripción y análisis de imágenes basada en redes neuronales convolucionales: fundamentos, entrenamiento, arquitecturas y transferencia de conocimiento adquirido
2. Multimedia (audio). Aplicaciones de IA en el tratamiento de sonido
 - 2.1. Extracción de características en señal de voz: detección de voz, tono fundamental, espectro y envolvente espectral, espectrogramas
 - 2.2. Extracción de características en señal musical: detección de notas musicales simultáneas (multipitch), ritmo, armonía, cromagramas, detección de música, localización de fragmentos musicales
 - 2.3. Aplicaciones en voz I: reconocimiento de voz y detección de palabras clave
 - 2.4. Aplicaciones en voz II: detección de hablante, detección de idioma, reconocimiento de emociones
 - 2.5. Aplicaciones sobre audio broadcast: segmentación y separación de hablantes y eventos acústicos, búsquedas, indexación y búsquedas sobre audio.

3. Procesamiento de Lenguaje Natural
 - 3.1. Introducción al PLN y lingüística computacional
 - 3.2. Tokenización, splitter, lematización, Pos Tagging
 - 3.3. Ambigüedad léxica y semántica
 - 3.4. Pipelines de PLN: freeling y Spacy
 - 3.5. Análisis semántico
 - 3.6. Aplicaciones basadas en PLN
 - 3.7. Herramientas de análisis del sentimiento
 - 3.8. Herramientas de detección de entidades
 - 3.9. Herramientas de Chatbot
 - 3.10. Herramientas de Speech2text
4. Aplicaciones de IA en análisis de textos
 - 4.1. Introducción al procesado de texto lingüística computacional, estratos del lenguajes, corpora, estrategias de análisis
 - 4.2. Análisis de caracteres: frecuencias de caracteres, n-gramas de caracteres, complejidad de compresión y riqueza del texto
 - 4.3. Análisis de tokens: tokenización, bag of words, hashing trick, TF-IDF
 - 4.4. Análisis morfosintáctico: lexemas, categorías gramaticales, reglas generativas, árboles de parsing, gramáticas formales.
 - 4.5. Análisis semántico: ontologías, NER, topic modelling
 - 4.6. Modelos estadísticos avanzados: redes neuronales para texto, embeddings de tokens, modelos neuronales convolucionales y recurrentes.
 - 4.7. Modelos grandes de lenguaje LLMs
5. Reconocimiento Biométrico
 - 5.1. Introducción al reconocimiento biométrico de personas
 - 5.2. Métricas de rendimiento
 - 5.3. Fusión de sistemas biométricos
 - 5.4. Reconocimiento facial y firma manuscrita
 - 5.5. Particularidades del reconocimiento biométrico conductual



Inteligencia Artificial Generativa y Explicativa

Universidad Autónoma de Madrid

1. Aplicación de IA generativa al procesamiento de imagen
 - 1.1. Introducción a modelos generativos y prompt engineering
 - 1.2. Edición inteligente de imágenes
 - 1.3. Generación de datos sintéticos para entrenamiento de modelos
2. Aplicación de IA generativa para acceso Inteligente a la Información
 - 2.1. Búsqueda avanzada de información
 - 2.2. Recomendación personalizada de información
 - 2.3. Síntesis y explicación de información recuperada
3. Inteligencia Artificial Explicable (XAI)
 - 3.1. Introducción a la XAI. Motivación, definiciones
 - 3.2. Algoritmos genéricos de explicabilidad (SHAP y LIME)
 - 3.3. Algoritmos específicos de explicabilidad para Deep Learning DL, basados en gradientes y LRP
 - 3.4. Comparación de algoritmos de explicabilidad: Coeficiente de Explicabilidad-Rendimiento EPC
 - 3.5. Introducción a la Interpretabilidad global de los modelos de DL y RNNs
4. Aplicación de IA Generativa en la actualidad
 - 4.1. Casos de uso en investigación
 - 4.2. Casos de uso en la empresa