



Instituto de
Matemática
Interdisciplinar

IMI Data Science Club

**Ciencia de Datos:
gestionar información para decidir mejor**

Begoña Vitoriano (IMI, UCM)

Datos e Información

- **Dato:** representación simbólica (numérica, alfabética, espacial...) de un atributo o variable cuantitativa o cualitativa.
 - Se refieren a información existente o conocimiento representado en alguna forma adecuada para su mejor uso
- **Información:** conjunto organizado de datos procesados que constituyen un mensaje que cambia el estado de conocimiento de quien recibe dicho mensaje
 - Es conocimiento explícito extraído como resultado de interacción o percepciones sensibles del entorno.
 - A diferencia de los datos, tienen estructura útil que modificará las interacciones con el entorno



Ciencia e Ingeniería de Datos

- **Ciencia de Datos:** campo interdisciplinar que usa métodos matemáticos, procesos y algoritmos para extraer conocimiento e ideas de los datos
 - Continuación de algunos campos de análisis de datos como la estadística, la minería de datos, el aprendizaje automático y la analítica predictiva
- **Ingeniería / Arquitectura de datos:** campo de la ingeniería de software que recoge y recopila los datos, los almacena, los procesa por lotes o en tiempo real, y los prepara y entrega para ser utilizados en ciencia de datos.
 - Diseño y desarrollo de sistemas de información



Problemas de Ingeniería de Datos

- **Ingeniería de datos:** Almacenamiento de datos, bases de datos, comunicaciones, hardware (**Computación**)
 - **Big data / Datos masivos:** millones de petabytes generados dinámicamente (Redes sociales, observatorios astronómicos...). Almacenar, buscar y tratar (paralelizar)
 - **Blockchain:** estructura de datos en la que la información contenida se agrupa en conjuntos (bloques) a los que se les añade metainformaciones relativas a otro bloque de la cadena anterior en una línea temporal. Gracias a la criptografía, la información contenida en un bloque solo puede ser repudiada o editada modificando todos los bloques posteriores



Problemas de Ciencia de Datos

- **Problemas/Análisis de datos:** estadística (regresión, predicción, clasificación supervisada y no supervisada...) y nuevos (reconocer texto, identificar líderes sociales, reconocimiento de imágenes...)
- **Métodos:**
 - **Estadística clásica y computacional** (minería de datos...)
 - **Aprendizaje automático:** técnicas para que los ordenadores *aprendan*, generalizando comportamientos a partir de información suministrada en forma de ejemplos (inducción). Se solapa con la estadística computacional
 - **Aprendizaje profundo:** modelar abstracciones alto nivel en datos usando arquitecturas de transformaciones no lineales múltiples
 - **Optimización** (minimización errores, svm...)
 - Alta carga **computacional:** R fundamental, Python

Toma de Decisiones

- **Fin del conocimiento: Toma de Decisiones**
- Proceso de identificar y escoger entre alternativas basado en los valores, preferencias y creencias del decisor, siendo la elección óptima o al menos satisfactoria
 - **Sistemas de ayuda a la decisión:** software de toma de decisiones cuando se enfrentan a decisiones muy complejas o al considerar a muchas partes interesadas, categorías u otros factores que afectan las decisiones
 - Apoyo a la decisión **automatizado**
 - **Optimización / Investigación Operativa**
 - Input: **conocimiento** del entorno/sistema
 - Crucial: **información** suficiente, estructurada y fiable

IMI Data Science Club

Agrupar actividades de **Data Science** en el IMI

- **Programas IMI:**
 - DEC-HUMLOG (Decision Aid Models for Humanitarian Logistics and Development)
 - Development and Implementation of Bayesian Methods and Design of a User Interface for the Analysis of Contingency Tables and Expression Data
- **Grupos investigación:**
 - Data Science and Soft Computing for Social Analytics and Decision Aid
 - Modelos de Decisión en Logística y Gestión de Desastres
 - Métodos Bayesianos
 - Teoría de Juegos y Aplicaciones
- **Proyectos:**
 - Conceptos y Aplicaciones de los Sistemas Inteligentes
 - ¿Nos dirigimos hacia una sociedad desintermediada? Un estudio de la participación social mediante análisis de redes, big data y aprendizaje automático
 - Técnicas de obtención, procesamiento y representación de información difusa para la toma de decisiones
 - Geospatial based Environment for Optimisation Systems Addressing Fire Emergencies
 - Análisis Juego-Teórico de las Redes Sociales
 - Modelos de Decisión y Ciencia de Datos en Sector de Restauración
 - Modelos de Ayuda a la Decisión para Logística Humanitaria en Gestión de Desastres



Instituto de
Matemática
Interdisciplinar

IMI Data Science Club Plan 2019

- **Abril.** Ciencia de Datos: gestionar información para decidir mejor
- **Mayo.** Conferencia Daniel Vélez (F. Matemáticas, UCM)
- **Mayo.** Conferencia Luis Magdalena (UPM) (por confirmar)
- **Junio.** Curso doctorado “Una introducción al análisis de datos y la cuantificación de incertidumbre en problemas inversos”. Marcos A. Capistrán (CIMAT, México)
- **Junio.** Presentaciones Modelling Week
- **Octubre.** Presentaciones TFM TECI + MIM
- **Noviembre.** Curso doctorado “Máquinas de Vector Soporte: Aplicación a Problemas de Clasificación Supervisada”. E. Carrizosa (F. Matemáticas, U. Sevilla)
- **Diciembre.** Curso doctorado “Modelos de Localización, análisis de datos y problemas afines”. J. Puerto (F. Matemáticas, U. Sevilla)
- **Diciembre.** Jornada +FuzzyMAD (curso doctorado, presentaciones)