

APLICACIONES DE ANÁLISIS

Nº de Créditos: 5 ECTS
Segundo Semestre
Segundo Curso

EQUIPO DOCENTE

Pulido Cañabate, Estrella
Escuela Politécnica Superior
UAM
Coordinadora

Camacho Fernández, David
Escuela Politécnica Superior
UAM

Fernández Pascual, Angela
Escuela Politécnica Superior
UAM

Antonio González Pardo
Escuela Politécnica Superior
UAM

Haya Coll, Pablo
Escuela Politécnica Superior
UAM

Hernández Bravo, Ángel
IBM

Moreno Ruíz, María José
IBM

Rodríguez Luján, Irene
Escuela Politécnica Superior
UAM

Sanz Segurrola, Iñigo
Indizen

Varona Martínez, Pablo
Escuela Politécnica Superior
UAM

OBJETIVOS

- Utilizar herramientas avanzadas para modelar y resolver problemas relacionados con la extracción de datos de redes sociales.
- Analizar y desarrollar diferentes modelos para el procesado y análisis de conocimiento de redes sociales.
- Utilizar técnicas avanzadas de minería de datos y computación avanzada para resolver problemas en los que se dispone de un volumen enorme de datos.
- Análisis y predicción de series temporales dentro del contexto de las energías renovables.
- Enfrentarse a problemas de negocio realizando un análisis del mismo desglosando probabilidades y beneficios/costes estimados para cada caso,

identificando las técnicas de Ciencia de Datos aplicables y comunicándose con otros componentes del equipo a través del uso de gráficas para la visualización del beneficio esperado de las posibles soluciones.

- Identificación y diseño de soluciones TIC en la integración y tratamiento de datos biomédicos masivos.
- Recabar y hacer uso de datos biomédicos a partir de dispositivos móviles.
- Entender, analizar e interpretar datos de genómica de *Next Generation Sequencing (NGS)*.
- Utilizar algunas de las herramientas más comunes para el análisis de datos NGS.
- Identificar escenarios de Big Data en el entorno de Ciudades Inteligentes. Diferentes casos de uso y la problemática de las diferentes fuentes de datos así como su manipulación.

PROGRAMA DETALLADO

- Business Analytics
 - Introducción
 - Soluciones de data science para problemas de negocio
 - Evaluación de modelos para negocio
 - Visualización del rendimiento de modelos
 - Ingeniería analítica
 - Evaluación de propuestas
- Biomedicina y genómica
 - Introducción al Big Data en Biomedicina
 - Diversidad e integración de información biomédica
 - Disponibilidad y acceso a bases de datos biomédicas
 - Herramientas básicas de tratamiento masivo de datos para el diagnóstico y prevención de enfermedades
 - Dispositivos móviles y big-data biomédico
 - Genómica
 - Introducción a la genómica y el Big Data
 - Algoritmos para el análisis de datos NGS
 - Análisis de expresión diferencial en datos de NGS
 - Herramientas
- Predicción de energías renovables con series temporales
 - Introducción:
 - Qué es una serie temporal.
 - Las energías renovables como series temporales.

- Análisis de series temporales: visualización, estacionaridad y estacionalidad.
- Predicción de energías renovables usando modelos autoregresivos.
- Predicción de energías renovables usando modelos de series temporales con entradas exógenas.
- Real-life analytics
- Impacto social: análisis de redes sociales
 - Los medios sociales como fuente de datos.
 - Introducción al social mining
 - Minería de redes sociales y de opiniones
 - Métricas
 - Herramientas
 - Casos de estudio:
 - Cómo medir la concienciación de los lectores.
 - Grupos de conversación en Twitter: Análisis políticos.
 - Social Mining: Analizando los datos de las redes sociales.
 - Conceptos básicos sobre grafos.
 - Métricas para análisis de grafos o redes.
 - Detección de topics en los textos.
 - Análisis de sentimiento.
 - Visualización de datos (información de redes, geoespacial y textual)
 - Análisis de comunidades.
 - Algoritmos de búsquedas de comunidades.
 - Métricas de evaluación de comunidades.
 - Casos prácticos de detección de comunidades.
 - Casos prácticos sobre Twitter
 - Extracción y pre-procesado de datos.
 - Almacenamiento de datos en BD no relacionales (MongoDB).
 - Filtrado, ordenación y agrupamientos de documentos (tweets).
 - Generar y analizar grafos de retweets, favoritos, amistad...
 - Analizando y visualizando los datos de Twitter.
- Impacto social: computación cognitiva
 - Bases de la Computación Cognitiva. IBM Watson.
 - Estrategia de Soluciones Cognitivas de IBM.
 - Infraestructura para Soluciones Cognitivas.
 - Servicios cognitivos a través de IBM Bluemix.
- Smarter cities
 - Introducción: Por qué Smarter Cities
 - Plataformas, arquitecturas y estándares en Smarter Cities
 - La conexión con Internet of Things: plataformas y arquitecturas
 - La potencia de Open Data en Smarter Cities

- Integración de Data Streams
- Modelos de Big Data y Big Analytics en SmarterCities
- Provisión de Servicios Digitales. El ecosistema: API Management, SmarterCities/IoT PaaS
- Gobernanza de Sistemas de Smarter Cities
- Modelado de ciudades: indicadores

BIBLIOGRAFÍA

- Mining the Social Web, Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More. Matthew A. Russell. O'Reilly Media. 2013.
- Introduction to time series and forecasting, P.J. Brockwell, R.A. Davis, Springer Texts in Statistics (2002)
- Time series analysis, J. D. Hamilton, Princeton University Press, Princeton, NJ (1994)
- Time series prediction, A.S. Weigend, New York: Routledge (2018)
- Weiser, Mark (1991). "The Computer for the 21st Century". <http://www.ubiq.com/hypertext/weiser/SciAmDraft3.html>
- Pentland, Alex (2014) "Social Physics: How Social Networks Can Make Us Smarter". Penguin Publishing Group
- [MIT Human Dynamics Lab](#)
- Social media mining: an introduction. Zafarani, Reza and Abbasi, Mohammad Ali and Liu, Huan. Cambridge University Press. 2014.
- Sentiment Analysis and Opinion Mining, Bing Liu. Morgan&Claypool Publishers. 2012.
- Data Science for Business, Foster Provost, Tom Fawcett. O'Reilly Media. 2013.
- Springer handbook of speech processing. Benesty, J. Springer Science & Business Media. 2008.
- Introduction to biometrics. Jain, A. K., Ross, A. A., & Nandakumar, K. Springer Science & Business Media. 2011.
- Handbook of natural language processing (Vol. 2). Indurkha, N., & Damerau, F. J. (Eds.). CRC Press. 2010.
- Natural language processing with Python. Bird, S., Klein, E., & Loper, E. O'Reilly Media, Inc. 2009.
- Connected: The Surprising Power of Our Social Networks. N Christakis and J Fowler. 2011.
- Six degrees: The Science of a Connected Age. DJ Watts. 2004.
- Linked. AL Barabasi. 2002.
- Networks. M. Newmann. Oxford University Press. April 2010.
- Networks, Crowds, and Markets. D. Easley and J. Kleinberg. Cambridge University Press, Cambridge. 2010.

- Social Network Analysis. S. Wasserman and K. Faust. Cambridge University Press, Cambridge. 1994.
 - Coursera – Genomic Data Science Specialization: <https://www.coursera.org/specializations/genomic-data-science>
 - Materiales docentes del profesor Ben Langmead: <http://www.langmead-lab.org/teaching-materials/>
 - Claverie, Jean-Michel, and Cedric Notredame. Bioinformatics for dummies. John Wiley & Sons, 2011.
 - Fonseca, Nuno A., et al. "Tools for mapping high-throughput sequencing data." Bioinformatics (2012): bts605.
 - Anders, Simon, et al. "Count-based differential expression analysis of RNA sequencing data using R and Bioconductor." Nature protocols 8.9 (2013): 1765-1786.
 - Datta, Somnath, and Dan Nettleton. Statistical Analysis of Next Generation Sequencing Data. Springer, 2014.
 - Gentleman, Robert. R programming for bioinformatics. CRC Press, 2008.
 - Bassi, Sebastian. Python for bioinformatics. CRC Press, 2016.
 - Korpelainen, Eija, et al. RNA-seq Data Analysis: A Practical Approach. CRC Press, 2014.

MÉTODOS DOCENTES

- Lección magistral
- Resolución de problemas
- Prácticas de laboratorio
- Estudio de casos

MÉTODOS DE EVALUACIÓN

- Examen: **50%**
- Evaluación continua: **50%**

Para superar la asignatura es necesario

- Obtener una calificación mínima de 5 en el examen, y
- Obtener una calificación mínima de 3,5 en cada una de las prácticas de la asignatura