



ANÁLISIS INTELIGENTE DE DATOS EN ENTORNOS BIG DATA

Carrera:

Maestría y Especialización
en Inteligencia de Datos
orientada a Big Data
Doctorado en Ciencias Informáticas

Carga Horaria: 70 Hs.

Profesor a Cargo:

Dr. José Angel Olivas Varela
Dr. Waldo Hasperué

Créditos: 4

OBJETIVO

Introducir una mirada diferente al análisis inteligente de datos basada en la Inteligencia de Negocios o BI (Business Intelligence). El énfasis estará puesto en el procesamiento de Datos Masivos o Big Data.

MODALIDAD DE EVALUACION

Para aprobar el curso se requiere un 80% de asistencia y la realización de un trabajo final que se definirá una vez completada la exposición de los contenidos teóricos. Dicho proyecto podrá ser realizado en forma individual o grupal y tendrá por objetivo profundizar uno o varios de los conceptos vistos en clase. La entrega consistirá en un reporte técnico de calidad científica producto de un estudio teórico o experimental específico.

PROGRAMA

Parte I: Conceptos

- Inteligencia de datos: objetivos, retos, límites, ejemplos.

- Diferencias entre Minería de Datos, Machine Learning, Business Intelligence e Inteligencia Artificial, Ciencia de Datos, Almacenes de Datos.
- Inteligencia de datos en distintos ámbitos: empresas, gobiernos, economía, leyes, finanzas, contadurías, recursos humanos, manejo de calidad, organización de proyectos, manejo de recursos, industria, deportes, redes sociales.
- Técnicas inteligentes para procesar datos: modelos supervisados y no supervisados. Problemas de clasificación, predicción, clustering y asociación. Modelos basados en reglas, árboles, redes neuronales, similitud, algoritmos evolutivos y genéticos. Conjuntos de entrenamiento, prueba y validación, generalización, sobreajuste.
- Presentación de información y toma de decisiones a partir de datos.
- Big Data: Definición, características principales, distinción con datos normales. Problemas que plantea, soluciones con técnicas inteligentes. Adaptación de técnicas inteligentes a entornos Big Data.

Parte II: Aplicaciones

- Frameworks para realizar análisis inteligente de datos con Big Data: Spark, Spark Streaming, MLib.
- Spark: Paradigma de programación, clusters, flujo de datos, almacenamiento distribuido. Operaciones: map, filter, joins, inmutabilidad, ejecución *lazy*, tolerancia a fallos.
- MLLib, modelos de aprendizaje automático disponibles, paradigma de carga de datos, entrenamiento y ejecución de modelos, integración con Spark.
- Aplicación a problemas puntuales:
 - Segmentación, predicción/pronóstico/estimación.
 - Análisis de Series Temporales.
 - Sistemas recomendadores.
 - Redes Sociales: Análisis de sentimientos y opiniones.
 - Acceso y búsqueda Web (Web Search and Information Mining)

BIBLIOGRAFIA

- Dean, Jared. *Big data, data mining, and machine learning: value creation for business leaders and practitioners*. John Wiley & Sons, 2014.



- Provost, Foster, and Tom Fawcett. *Data Science for Business: What you need to know about data mining and data-analytic thinking*. " O'Reilly Media, Inc.", 2013.
- Minelli, Michael, Michele Chambers, and Ambiga Dhiraj. *Big data, big analytics: emerging business intelligence and analytic trends for today's businesses*. John Wiley & Sons, 2012.
- Soumendra Mohanty et al. (2013). *Big Data Imperatives: Enterprise Big Data Warehouse, BI Implementations and Analytics*. Apress. ISBN 978-1430248729.
- James A. Scott. *Getting Started with Apache Spark*. MapR Technologies, Inc., 2015.
- O'Reilly Media, Inc.(2016) *Big Data Now*. Inc. ISBN 978-1-449-35671-2.
- Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman (2011). *Mining of Massive Datasets*. Cambridge University Press. ISBN 978-1107015357.
- Kord Davis, Doug Patterson (2012). *Ethics of Big Data*. O'Reilly Media, Inc. ISBN 978-1-449-31179-7.
- Xindong Wu, Xingquan Zhu, Gong-Qing Wu and Wei Ding. (2014). *Data Mining with Big Data*. IEEE transactions on knowledge and data engineering, Vol. 26 (1): 97-107.
- Bifet, A. (2013). *Mining big data in real time*. Informatica, 37(1).
- Fan, W., & Bifet, A. (2013). *Mining big data: current status, and forecast to the future*. ACM SIGKDD Explorations Newsletter, 14(2), 1-5.
- Amatriain, X. (2013). *Mining large streams of user data for personalized recommendations*. ACM SIGKDD Explorations Newsletter, 14(2), 37-48.
- Leskovec, J. et al. (2014). *Mining of massive datasets*. Cambridge University Press.
- Shanmuganathan, S. (2014). *From data mining and knowledge discovery to big data analytics and knowledge extraction for applications in science*. Journal of Computer Science, vol.10(12), pp.1 - 8 (8). Science Publications
- Jain, Mugdha, and Chakradhar Verma. "Adapting k-means for Clustering in Big Data." *International Journal of Computer Applications* 101.1 (2014): 19-24.
- Cui, Xiaoli, et al. "Optimized big data K-means clustering using MapReduce." *The Journal of Supercomputing* 70.3 (2014): 1249-1259.
- Singh, Dilpreet, and Chandan K. Reddy. "A survey on platforms for big data analytics." *Journal of Big Data* 2.1 (2015): 8.
- Havens, Timothy C., et al. "Fuzzy c-means algorithms for very large data." *IEEE Transactions on Fuzzy Systems* 20.6 (2012): 1130-1146.