



ANÁLISIS INTELIGENTE DE DATOS EN ENTORNOS BIG DATA

Año 2019

Carrera: ***Maestría en Inteligencia de
Datos orientada a Big Data***

Carga Horaria: 64 Hs.

Profesor a Cargo:

***Dr. José Angel Olivas Varela, Dr. Waldo
Hasperué***

OBJETIVO

Presentar el análisis inteligente de datos, principalmente en entornos masivos o Big Data, desde de punto de vista del estudio detallado de los datos disponibles, las técnicas inteligentes a aplicar para su análisis y las diferentes salidas esperables como resultado del análisis.

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.
- Ciencia e Ingeniería de Datos, Inteligencia Artificial y Aprendizaje Automático (Machine Learning), Técnicas estadísticas, Business Intelligence Minería de datos, textos, opiniones y grafos.



- 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, Mllib.
- 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

- Serrano-Guerrero, J., Olivas, J. A., Romero, F. P., Herrera-Viedma, E.: "Sentiment analysis: A review and comparative analysis of web services". Information Sciences, 311 (2015), 18–38.



- 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.
- James A. Scott. *Getting Started with Apache Spark*. MapR Technologies, Inc., 2015.
- O'Reilly Media, Inc.(2016) *Big Data Now*. Inc.
- Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman (2011). *Mining of Massive Datasets*. Cambridge University Press.
- Kord Davis, Doug Patterson (2012). *Ethics of Big Data*. O'Reilly Media, Inc. 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.
- 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
- Singh, Dilpreet, and Chandan K. Reddy. "A survey on platforms for big data analytics." *Journal of Big Data* 2.1 (2015): 8.