

COORDINACIÓN DE POSGRADOS

CARRERA: **MAESTRÍA GENERACIÓN Y ANÁLISIS DE INFORMACIÓN ESTADÍSTICA**

CICLO DE: **ORIENTACION EN ESTADISTICAS DE OPINION Y MERCADOS**

MATERIA: **SEMINARIO DE APLICACIÓN DE BIG DATA Y DATA SCIENCE**

CARGA HORARIA: **32 HS**

AÑO: **2021** CUATRIMESTRE: **SEGUNDO**

NOMBRE Y APELLIDO DE DOCENTE TITULAR: **Juan José LLORET**

NOMBRE Y APELLIDO DOCENTE ADJUNTO: **Santiago MARCO**

OBJETIVO GENERAL

El fenómeno de Big Data en un dominio, es un proceso en donde se conjugan los datos disponibles, las habilidades del científico de datos y la utilización de recursos metodológicos/tecnológicos. El objetivo es alcanzar una solución apropiada al problema que permita explorar, analizar, explicar, interpretar o comunicar.

Esta particularidad para abordar el conocimiento del dominio, se ha convertido en una modalidad cotidiana en los ámbitos de los negocios, las instituciones y la sociedad.

Es por eso que el presente curso, preponderantemente práctico, apunta a ganar por parte del cursante, familiaridad en los conceptos que subyacen en el fenómeno Big Data y a la vez poder desarrollar en su ámbito de actividad, los primeros pasos con esta disciplina.

OBJETIVOS ESPECIFICOS

Familiarizar a los asistentes con los conceptos básicos de Big Data y Data Science

Desarrollar una visión panorámica de las distintas herramientas de Data Science para planificación, control y análisis de la información.

Mostrar a través de casos, la utilización de Data Science en diferentes dominios

Explorar el impacto que tiene en la sociedad a través del cambio de paradigmas económicos y sociales.

CONTENIDOS

MODULO 1 - Concepto de Big Data - 4 horas

- Los conceptos generales de Big data y el encuadre en la Ciencia de datos.
- Las etapas que se deben considerar en un proyecto de Big Data. Su formulación y los protocolos a seguir. El Método CRISP como protocolo a seguir en la formulación de un proyecto.
- Ejemplos en diferentes dominios.

MODULO 2 - La Obtención de los Datos y la generación de la información – 4 horas

- Características de los datos. Las fuentes originarias.
- Conocimiento del dominio

- Las métricas.
- Extracción de los datos. Diversos métodos
- La complejidad de los datos
- La preparación de los datos para los distintos procesos de Big Data.
- El Internet de las cosas
- Los sensores remotos
- Los datos sistemáticos. Los paneles

MODULO 3 - El aprendizaje Automático (Machine Learning) y la inteligencia Artificial – 12 horas

- Concepto de aprendizaje automático. Definición del proyecto de Machine Learning
- Enfoque sobre el tipo de datos disponibles y su relación con el propósito del proyecto
- Rol de los atributos y formulación de acuerdo al dominio. Selección de atributos
- Algoritmos candidatos, el entrenamiento y el testeo.
- Elegibilidad de los algoritmos
- Clasificación, agrupamientos y reglas de asociación
- Uso de text mining en definición de campos categóricos.
- Aprendizaje supervisado (Bayes, Regresiones, Arboles, etc.)
- Aprendizaje no supervisado (Análisis de componentes Principales, Clustering)
- Plataforma de minería y modelado.
- Criterios de elección del Algoritmo ganador.
- Inteligencia Artificial aplicados en casos de clasificación y reconocimiento de patrones.

MODULO 4 - Visualización del Conocimiento – 8 horas

- Visualización de los datos.
- Concepto de visualización de la información. *Information literacy*
- Características y tabla de clasificación de la visualización
- Sistemas de información geográficos/temporales.
- Visualizaciones para representar la información. Uso en investigación
- Visualizaciones para presentar la información.
- Visualizadores.
- Tableros de control y visualizaciones dinámicas.
- *StoryTelling*
- Esquema de formalización y optimización de las visualizaciones
- Visualizadores: pro y contra de los visualizadores. Uso en los diferentes dominios.
- Se recrearán casos públicos y se utilizarán diferentes galerías. Se mostrara las fortalezas y debilidades de una visualización, tratando de poner énfasis en la obtención de información útil para la toma decisión y como herramienta de seguimiento en los procesos de control de gestión

MODULO 5 - Las aplicaciones de Big Data – Conclusiones - 4 horas

- Análisis del problema y la planificación
- Formulación de las etapas. Viaje por las etapas
- Utilización de métodos de pre-proceso, modelado y visualización de ciencia de datos
- Visualización de información en Salud, Educación, Medio Ambiente, Demografía, *Business Intelligence*, Marketing, etc.

METODOLOGÍA y MODALIDAD DE CURSADA

El formato del curso será no presencial híbrido, con encuentros sincrónicas y asincrónicas.

Cada clase tendrá dos partes:

- 1) el desarrollo los distintos conceptos temáticas conceptuales
- 2) practica de los ejercicios y consultas sobre la tarea grupal.

El soporte de los docentes se realizara también a través de un foro, en donde cada alumno hará las preguntas o consultas que considere necesarias.

Se realizarán prácticas utilizando aplicaciones de estadísticas, librerías Python y R, así como aplicaciones de *dataminig* y visualización.

REQUISITOS PARA LA APROBACIÓN DE LA MATERIA

La aprobación del curso requiere:

a- ASISTENCIA A CLASES

La asistencia al menos del 80% de las clases sincrónica.

b- EVALUACIÓN

- la realización domiciliaria de una serie de ejercicio de las prácticas.
- la elaboración de un trabajo grupal de Big Data específico, la presentación grupal del mismo y la entrega de un informe con detalle de los objetivos del proyecto, las características del dominio, las etapas realizadas y los resultados obtenidos.

BIBLIOGRAFIA

- ✓ Material de la cátedra, se provee a lo largo del curso
- ✓ Hechos y Estadísticas - M. J. Moroney - EUDEBA
- ✓ Herramientas y Técnicas de Aprendizaje Automático en Data Mining 3era edición - Witten IH, Frank E, Mark A. Hall.
https://www.academia.edu/23331284/Data_Mining_Practical_Machine_Learning_Tools_and_Techniques_3rd_Edition
- ✓ Instalación de Weka: <https://www.cs.waikato.ac.nz/ml/weka/>
- ✓ Manual de Weka para la Versión 3-8-5 R. Bouckaert, Eibe Frank, Mark Hall, Richard Kirkby, Peter Reutemann, Alex Seewald, David Scuse December 21, 2020
- ✓ Una introducción al aprendizaje Automático Miroslav Kubat
http://ndl.ethernet.edu.et/bitstream/123456789/88653/1/2017_Book_AnIntroductionToMachineLearning.pdf
- ✓ Big Data Multimedia para Internet de las Cosas Aplicaciones, Conceptos, Paradigmas y Soluciones Sudeep Tanwar, Sudhanshu Tyagi, Neeraj Kumar
https://www.researchgate.net/publication/328032193_Multimedia_big_data_computing_and_Internet_of_Things_applications_A_taxonomy_and_process_model
- ✓ Galería de visualizaciones Flourish: <https://flourish.studio/examples/>
- ✓ Galería de Visualizaciones Tableau: <https://www.tableau.com/solutions/gallery>