

PM_BIG DATA & DATA

OBJETIVO(S) GENERAL(ES) DE LA ASIGNATURA

AL TÉRMINO DE LA MATERIA EL ESTUDIANTE:

- Identificará y resolverá problemas relacionados con el paradigma Big Data
- Clasificará y almacenará los diferentes tipos de datos de este paradigma
- Realizará analítica de datos de diferentes fuentes (Web, redes sociales, M2M, Cloud, entre otras)
- Obtendrá las habilidades metodológicas de investigación para sentar las bases de conocimiento para futuros Científicos de Datos

TEMAS Y SUBTEMAS

Introducción a Big Data y la ciencia de datos.

1.1 Big Data: Datos estructurados y datos no estructurados

1.2 Data Analysis, Data Analytics, Data Engineer, Data Scientist

1.3 De la hoja de cálculo al trabajo en la nube

1.4 El arte de la visualización de los datos: Flourish – Gapminder - Gephi:

Estadísticas animadas

1.5 Casos prácticos

2 Descubrimientos en los datos a través de Cognos analytics de IBM.

2.1 Creación de cuenta en Cognos Analytics y exploración de la plataforma de IBM

2.2 Preparación de archivos en Excel y CSV

2.3 Algunas fuentes de información para macro análisis

2.4 Visualización de datos y preparación de reportes profesionales:

2.4.1 Construir tableros dinámicos (dashboards)

2.4.2 Descubrir relaciones e identificar patrones en los datos

2.4.3 Crear presentaciones siguiendo una narrativa temporal (stories)

2.4.4 Casos prácticos.

3 “R” para ciencia de datos.

3.1 Análisis de sentimientos en “R”

3.1.1 Un paseo por el software “R”

3.1.2 Paquetería para Ciencia de Datos

3.1.3 Caso práctico: El estado de ánimo de mi libro o la noticia del día

3.1.4 Codificación en “R” v. Corpus del documento

3.1.5 Palabras de opinión negativa

3.2 Minería de datos en Twitter para sentiment análisis.

3.2.1 Creación de cuenta para desarrollador, Application Programming Interface (API)

3.2.2 Paquetería para Twitter Mining

3.2.3 Caso práctico: El estado de ánimo de los tuiteros respecto al “NAFTA”, “Trump” u.

4 Análisis de datos en Python.

4.1 Conceptos básicos de Python.



PROGRAMA DE MATERIA

MAESTRÍA EN CIENCIA DE DATOS (RVOE 2727)

- 4.1.1 ¿Qué es y cómo instalar Python?
- 4.1.2 Ejecución de código en Jupyter
- 4.1.3 Sintaxis
- 4.2 Análisis de datos en Python.
 - 4.2.1 Fuentes de datos
 - 4.2.2 Tipos de datos que se pueden recopilar
 - 4.2.3 Gestión, exploración y visualización efectiva de los datos
- 4.3 Interpretación de datos inferenciales y modelación estadística.
 - 4.3.1 Análisis de datos con pandas, numpy y matplotlib
 - 4.3.2 Importar y exportar datos desde Excel
 - 4.3.3 Arrays & Dataframes
- 4.4 Casos prácticos de análisis y visualización de datos.
 - 4.4.1 Visualización dinámica
- 5 Análisis de redes sociales. (social network analysis)
 - 5.1 ¿Qué es Social Analytics y cuáles son sus usos?
 - 5.2 Análisis de redes sociales con NodeXL:
 - 5.2.1 Conceptos de Social Network
 - 5.2.2 Importar y preparar datos de Twitter
 - 5.2.3 Agrupar por clusters (conglomerados)
 - 5.2.4 Calcular métricas v. Análisis de series de tiempo
 - 5.2.5 Análisis de sentimientos en texto
 - 5.2.6 Network
 - 5.3 Consideraciones éticas sobre el análisis de redes sociales

MODALIDADES TECNOLÓGICAS E INFORMÁTICAS

Las modalidades y herramientas tecnológicas e informáticas para el desarrollo de las

actividades académicas son:

- Comunicación digital colaborativa estudiante-docente a través de Plataforma Moodle
- Gestión de actividades académicas mediante el uso de las herramientas de la Plataforma Moodle
- Uso de recursos para el aprendizaje como documentos, enlaces web, videos.
- Desarrollo de proyectos de investigación y procesamiento de información a través de Internet
- Uso de correo electrónico institucional.

ACTIVIDADES DE APRENDIZAJE

Actividades con docente

- Seminario

o La clase procurará utilizar el esquema de seminario, de tal manera que cada semana se presenten los diferentes temas que componen esta



PROGRAMA DE MATERIA

MAESTRÍA EN CIENCIA DE DATOS (RVOE 2727)

asignatura, analizar y clasificar datos estructurados, no estructurados y semi estructurados, acompañados por parte del titular de la materia, y de expertos en los diferentes temas.

- Clases magistrales

o El docente preparará y presentará una exposición a los estudiantes, describiendo a detalle las principales características de los temas vistos en clase. Se espera que los alumnos participen en esta exposición, a través de preguntas o argumentos relacionados con la generación de

Cubos OLAP.

- Asesorías sobre temas planteados

o El estudiante, con el apoyo y seguimiento del docente, realizará análisis de conceptos y ejercicios relacionados con los métodos de redes neuronales, con el objetivo de conocerlas y de formar un criterio al momento de resolver problemas vinculados con esta disciplina.

- Práctica:

o Instalar un software libre que permita realizar la analítica de Big Data.

o Realizar una consulta, para obtener un reporte, query y visualización en alguna de las siguientes plataformas: SAS, IBM (Cognos), SAP(Business Object), Pentaho, o alguna otra similar.

o Desarrollar un proyecto donde el estudiante aplique técnicas de analítica de Big Data.

ACTIVIDADES DE APRENDIZAJE INDEPENDIENTES

- Ejercicio práctico

o El estudiante analizará el caso práctico descrito por el docente, a efecto de

destacar las principales características, la justificación de las decisiones

tomadas, y en general, identificar potenciales recomendaciones con respecto a

“R” para ciencia de datos.

- Clase magistral

o El docente preparará y presentará una exposición a los estudiantes, describiendo

a detalle las principales características de los temas vistos en clase. Se espera

que los alumnos participen en esta exposición, a través de preguntas o

argumentos relacionados con los casos prácticos de análisis y visualización de



PROGRAMA DE MATERIA
MAESTRÍA EN CIENCIA DE DATOS (RVOE 2727)

datos.

CRITERIOS Y PROCEDIMIENTOS DE EVALUACIÓN Y ACREDITACIÓN

La calificación final del curso se integrará de acuerdo con los siguientes porcentajes obtenidos por parcial:

Primer Parcial	25%
Segundo Parcial	25%
Tercer Parcial	50%

La calificación de cada uno de los parciales se calculará con base en las ponderaciones a actividades, exámenes, proyectos, etc. definidas al inicio del curso por cada docente, esta ponderación deberá ser comunicada a los estudiantes al inicio del curso. La calificación parcial deberá expresarse numéricamente en una escala de 0 a 10 con números enteros, siendo aprobatoria a partir de seis. En caso de calificaciones fraccionarias aprobatorias, si la fracción es menor a 0.50 se pasará al entero inmediato inferior, si es igual o mayor se pasará al entero inmediato superior. Las calificaciones reprobatorias se redondean al entero inmediato anterior.

Son las técnicas, instrumentos y herramientas sugeridas para constatar los desempeños académicos de las actividades de aprendizaje. Se recomienda evaluar los siguientes aspectos:

- Investigación.
- Participación en el aula.
- Cuestionarios.
- Prácticas de laboratorio.
- Trabajo en equipo.
- Proyecto de aplicación.