



PROGRAMA DE ESTUDIOS

NOMBRE DE LA ASIGNATURA

Datos masivos

SEMESTRE	CLAVE DE LA ASIGNATURA	TOTAL DE HORAS
Décimo semestre	075106	80

OBJETIVO(S) GENERAL(ES) DE LA ASIGNATURA

Al finalizar el curso, el alumno será capaz de identificar las bases de datos relacionales y no relacionales y los lenguajes utilizados para gestionarlos, los algoritmos de aprendizaje máquina utilizados en la manipulación de grandes cúmulos de datos, comprenderá las técnicas para almacenarlos, filtrarlos y muestrearlos.

TEMAS Y SUBTEMAS

1. **Introducción a SQL y NoSQL**
 - 1.1. Bases de datos relacionales.
 - 1.2. Lenguaje SQL.
 - 1.3. Bases de datos no relacionales.
 - 1.4. MongoDB.
2. **Incremento de complejidad**
 - 2.1. Transformaciones no lineales.
 - 2.2. Regularización de modelos lineales.
 - 2.2.1. Regresión Ridge.
 - 2.2.2. Regresión LASSO.
 - 2.3. Descenso de gradiente estocástico.
3. **Aprendizaje máquina para datos masivos**
 - 3.1. Naive Bayes.
 - 3.2. Clasificadores de árboles de decisión.
 - 3.3. Impulso y aprendizaje conjunto.
 - 3.4. Máquinas de soporte vectorial.
 - 3.5. Aprendizaje supervisado.
 - 3.6. Aprendizaje no supervisado.
 - 3.7. Aprendizaje semi-supervisado.
 - 3.8. Ingeniería de funciones.
 - 3.9. Aprendizaje profundo.
4. **Manejos de datos masivos**
 - 4.1. Conjuntos de datos masivos.
 - 4.2. Volumen, variedad y velocidad.
 - 4.3. Algoritmos para Big Data.
 - 4.3.1. Análisis de Big Oh
 - 4.3.2. Hashing
 - 4.3.3. Almacenamiento jerárquico
 - 4.3.4. Algoritmos de transmisión y de un solo pase
 - 4.4. Filtrado y muestreo
 - 4.5. Paralelismo
 - 4.6. Arquitectura MapReduce, Hadoop y Spark
 - 4.7. Implicaciones éticas y sociales





PROGRAMA DE ESTUDIOS

5. Programación para almacenaje de datos masivos

- 5.1. *Hadoop Distributed File System (HDFS)*
- 5.2. *MapReduce*
- 5.3. *Pig*
- 5.4. *Spark*
- 5.5. Gestión de flujo de trabajo (*Luigi*)

ACTIVIDADES DE APRENDIZAJE

Sesiones dirigidas por el profesor en las que presenta los conceptos y, al mismo tiempo, se realizarán programas que los ilustrarán, se sugiere utilizar algún Notebook como Collaboratory o Jupyter para realizar programas con el lenguaje Python, así como GitHub para compartir y descargar código. Se recomienda ampliamente impartir el curso en un laboratorio con equipo de cómputo disponible para cada estudiante. Se iniciará con un proyecto desde el inicio del curso con fines didácticos y de práctica de cada uno de los temas.

CRITERIOS Y PROCEDIMIENTOS DE EVALUACIÓN Y ACREDITACIÓN

En términos de los artículos 25 incisos (b), (e), (f) y (g); del 48 al 62, del Reglamento de alumnos de licenciatura aprobado por el H. Consejo Académico el 19 de mayo del 2016, los lineamientos que habrán de observarse en lo relativo a los criterios y procedimientos de evaluación y acreditación, entre lo más importante:

Al inicio del curso el profesor deberá indicar el procedimiento de evaluación que deberá comprender, al menos tres evaluaciones parciales que tendrán una equivalencia del 50% de la calificación final y un examen ordinario que equivaldrá al restante 50%.

Las evaluaciones podrán ser escritas y/o prácticas y cada una consta de un examen teórico-práctico, tareas y proyectos. La parte práctica de cada evaluación deberá estar relacionada con la ejecución exitosa y la documentación de la solución de problemas sobre temas del curso.

Además, pueden ser consideradas otras actividades como: el trabajo extra-clase, la participación durante las sesiones del curso y la asistencia a las asesorías.

El examen tendrá un valor mínimo de 50%; las tareas, proyectos y otras actividades, un valor máximo de 50%.

BIBLIOGRAFÍA (TIPO, TÍTULO, AUTOR, EDITORIAL Y AÑO)

Libros Básicos:

1. **Data science design manual**, Steven S. Skiena, Text in computer science, Springer 2017.
2. **Hadoop with Python**, Zachary Radtka and Donald Miner, O'Reilly Media, Inc., 2016.
3. **Introducing data science, big data, machine learning and more using Python tools**, Davy Cielen, Arno D. B. Meysman and Mohamed Ali, Manning Publications Co., 2016.

Libros de Consulta:

1. **Python for data Science**, John Paul Mueller and Luca Massaron, John Wiley & Sons, 2015.
2. **Statistical and Machine-Learning Data Mining, Techniques for Better Predictive Modeling and Analysis of Big Data**, Bruce Ratner, second edition, CRC Press, Taylor & Francis Group, 2011.
3. **Big Data Analytics with R and Hadoop**, Vignesh Prajapati, Packt Publishing, 2013.

PERFIL PROFESIONAL DEL DOCENTE

Estudios de maestría o doctorado en matemáticas, matemáticas aplicadas o computación con conocimientos en ciencia de datos, *machine learning* y *deep learning*.

Vo.Bo.



JEFE DE CARRERA
LICENCIATURA EN
MATEMÁTICAS APLICADAS
DR. FRANCO BARRAGÁN MENDOZA
JEFE DE CARRERA

AUTORIZÓ



DR. AGUSTÍN SANTIAGO ALVARADO
VICE-RECTOR ACADÉMICO