



Instrumento de Planeación para un Curso

DATOS GENERALES

Carrera:	INGENIERIA DE SOFTWARE Y SISTEMAS COMPUTACIONALES	Grupo:	711		
Materia:	MINERIA DE DATOS (LIS752)	Período:	AGO-DIC 2018		
Maestro:	RODOLFO ROCHA URRUTIA	Fechas en que se cursará:	06/08/2018 – 08/12/2018		
Tipo de curso:	Tradicional (X)	Taller ()	Seminario ()	Laboratorio (X)	Otro ()

Objetivo General:

Al término del curso el alumno aplicará las técnicas de minería de datos en almacenes de datos orientados a la explotación de información para generar patrones útiles para las organizaciones

Enfoque y relación con el plan de estudios:

La minería de datos es un área emergente de los años 90's en la Computación, sus repercusiones han logrado trascender prácticamente todas las áreas del conocimiento y han transformado completamente la forma en que percibimos nuestro entorno cotidiano. Como parte de su evolución, ahora forma parte de la llamada "Ciencia de Datos" (Data Science, en inglés). Su entendimiento es imprescindible hoy día, sobre todo por ser considerada según la mayoría de los expertos como la disciplina de mayor demanda en los próximos 10 años. La toma de decisiones en los entornos organizacionales ha sido fortalecida por técnicas que permiten predecir no solamente ventas futuras o el perfil mercadológico de sus clientes, sino además identifican asociaciones ocultas (patrones) en los productos más comprados o la navegación menos utilizada del sitio web de la compañía, etc.

Materias que apoyan al curso: Programación Estructurada y Orientada a Objetos, Diseño de Almacenes de Datos, Modelos Probabilísticos y Estadísticos.

Materias a las que apoya el curso: Desarrollo de Inteligencia de Negocios

Políticas del curso

- Pase de asistencia a los 10 min. de la hora de entrada. La acumulación de 2 retardos originará una falta.
- Las faltas no son justificables (dudas, acudir con el coordinador de la carrera)
- No se permite la entrada con alimentos y bebidas (sin protección) al aula.
- Configurar el celular en forma pasiva (whatsapp o mensajes) y solo atender casos verdaderamente urgentes fuera del aula.
- Se puede salir al baño o fuera del aula las veces que sea necesario sin avisar al profesor, a menos que se tenga que retirar definitivamente de la clase.

Normas para la entrega de trabajos

- Entregarlos en el tiempo y la forma pactados (al inicio del curso).
- Todos los trabajos serán enviados a través del portal de la Institución y salvo casos donde no se encuentre disponible dicho portal, se indicará algún repositorio especial en la nube. Por lo que la retroalimentación será a través del mismo ó directamente con el profesor.
- Cualquier otra duda acerca del tema, enviar un correo personal a la siguiente dirección: ekrodolfo@gmail.com.

Primer Período

Fechas: de 06/08/2018 a 14/09/2018

No. de sesiones: 12

Contenido temático: Temas y Subtemas:	Fechas:
I LOS CONCEPTOS BÁSICOS	
1.1 Conceptos Básicos de Minería de Datos (Ciencia de Datos - DS)	06/08/2018
1.2 Áreas que apoyan la DS	06/08/2018
1.3 Ética y Privacidad en la DS	09/08/2018
II LA PREPARACIÓN Y GESTIÓN DE DATOS	
2.1 La Pirámide del Conocimiento (DIKW)	09/08/2018
2.2 La Inteligencia de Negocios (BI – Business Intelligence)	13/08/2018
2.3 El Almacén de Datos Central (Datawarehouse & Data Mart)	13/08/2018
2.4 Construcción del Data Mart (MOLAP vs ROLAP)	13/08/2018
2.5 El proceso ETL (Extract – Transform – Load)	16/08/2018
2.6 Visualización de Datos (Dashboard)	20/08/2018
III FASES DE LA MINERÍA DE DATOS	
3.1 Metodologías en la DS (KDD, CRISP-DM y SEMMA)	23/08/2018
3.2 Fases de un proyecto DS:	23/08/2018
3.2.1 Comandos Básicos del Lenguaje R.	27-30/08/2018
3.2.2 Recolección de Datos	3-6/09/2018

Actividades de aprendizaje en clase	Material de apoyo:	Fechas:
P1.- Análisis de Artículos sobre DS (Ética y Privacidad) (Grupal)	- Cañón, Laboratorio	09/08/2018
P2.- Práctica de Construcción de Data Mart (OLAP)	- Suite Microsoft (SQL Server 2016, PowerBI Desktop).	13/08/2018
P3.- Práctica de un Proceso ETL	- Suite Pentaho 8.0 (PCE, PDI, Kettle)	16/08/2018
P4.- Práctica de Visualización de Datos (Dashboard)	- VM (Linux Ubuntu Server 16.04 > ó Windows Server 2012 >)	20/08/2018
P5.- Práctica de Comandos Básicos R		27-30/08/2018

Actividades de aprendizaje extra-clase:	Fecha de entrega:
A1.- Análisis de la Serie “Numb3rs” y su relación con DS	19/08/2018
A2.- Diagrama sobre Terminología clave en DS	09/09/2018
A3.- Línea de Tiempo sobre DS.	09/09/2018

Evaluación Parcial (Criterios de la calificación parcial)	Fechas:
<ul style="list-style-type: none"> - Prácticas (Individual/Grupal) 50 % - Investigaciones (Individual/Grupal) 30 % - Examen Teórico/Práctico 20 % <p>-----</p> <p>Total 100 %</p>	<p>Examen: 10/09/2018</p> <p>Revisión: 11/09/2018</p> <p>Entrega: 13/09/2018</p>

Segundo Período

Fechas: de 17/09/2018 a 26/10/2018

No. de sesiones: 12

Contenido temático: Temas y Subtemas:	Fechas:
III FASES DE LA MINERÍA DE DATOS 3.2.3 Limpieza de Datos 3.2.4 Normalización de Datos 3.2.5 Transformación de Datos 3.2.6 Exploración de Datos	17-20/09/2018 24/09/2018 27/09/2018 01/10/2018
IV LOS ELEMENTOS DE ENTRADA Y SALIDA EN LA MINERÍA DE DATOS 4.1 Métodos de Reducción de Dimensiones 4.1.1 PCA (Principal Components Analysis)	04/10/2018 04/10/2018
V TÉCNICAS DE MINERÍA DE DATOS 5.1 Métodos Supervisados y No Supervisados de DS 5.2 Métodos de Predicción Estadísticos: 5.2.1 Regresión Lineal 5.2.2 Regresión Múltiple	15/10/2018 15/10/2018 15/10/2018 18/10/2018

Actividades de aprendizaje en clase	Material de apoyo:	Fechas:
P6.- Práctica de Recolección – Limpieza de Datos P7.- Práctica de Normalización - Transformación de Datos P8.- Práctica de Exploración - PCA P9.- Práctica de Métodos de Predicción Estadísticos	- Cañón, Laboratorio - Pack R, R Studio, Anaconda Studio	3-20/09/2018 24-27/09/2018 01-04/10/2018 15-18/10/2018

Actividades de aprendizaje extra-clase:	Fecha de entrega:
A4.- Cuadro Comparativo entre metodologías SEMMA, CRISP-DM - DS A5.- Línea de Tiempo sobre la Historia de ML (Machine Learning) A6.- Diagrama sobre Métodos de Predicción Estadísticos (Lineal, Múltiple)	30/09/2018 30/09/2018 21/10/2018

Evaluación Parcial (Criterios de la calificación parcial)	Fechas:												
<table> <tr> <td>- Prácticas (Individual/Grupal)</td><td>30 %</td></tr> <tr> <td>- Investigaciones (Individual/Grupal)</td><td>20 %</td></tr> <tr> <td>- Examen Teórico/Práctico</td><td>30 %</td></tr> <tr> <td>- Proyecto (Parcial)</td><td>20 %</td></tr> <tr> <td colspan="2">-----</td></tr> <tr> <td>Total</td><td>100 %</td></tr> </table>	- Prácticas (Individual/Grupal)	30 %	- Investigaciones (Individual/Grupal)	20 %	- Examen Teórico/Práctico	30 %	- Proyecto (Parcial)	20 %	-----		Total	100 %	Examen: 22/10/2018 Revisión: 23/10/2018 Entrega: 25/10/2018
- Prácticas (Individual/Grupal)	30 %												
- Investigaciones (Individual/Grupal)	20 %												
- Examen Teórico/Práctico	30 %												
- Proyecto (Parcial)	20 %												

Total	100 %												

Tercer Período

Fechas: de 29/10/2018 a 07/12/2018

No. de sesiones: 8

Contenido temático: Temas y Subtemas:	Fechas:
V TÉCNICAS DE MINERÍA DE DATOS	
5.3 Métodos de Clasificación:	
5.3.1 Regresión Logística	29/10/2018
5.3.2 Árboles de Decisión (CART, Random Forest)	01/11/2018
5.4 Reglas de Asociación (Apriori)	05/11/2018
5.5 Agrupamiento (Clustering) con K Means	08/11/2018
VI LA GESTIÓN DE MODELOS DE MINERÍA DE DATOS	
6.1 Modelos de Entrenamiento - Prueba	12/11/2018
6.2 Validación Cruzada (Matriz de Confusión)	12/11/2018
6.3 Gráfico ROC (Receiver Operating Characteristic)	12/11/2018
6.4 Evaluación de modelos (Matriz de Comparación)	12/11/2018
VII MINERÍA WEB y OTRAS TENDENCIAS	
7.1 Técnicas de extracción Web - Texto	15/11/2018
7.1.1 Técnicas de extracción en Redes Sociales, Mapas, etc.	15/11/2018
7.1.2 Manejo de Datos No-SQL con MongoDB	15/11/2018
7.1.3 Práctica de Extracción de Texto	22/11/2018
7.2 Otras técnicas avanzadas:	
7.2.1 Redes Neuronales (NN) & Deep Learning	22/11/2018
7.2.2 Máquinas de Soporte Vectorial (SVM)	22/11/2018
7.2.3 Métodos Bayesianos	22/11/2018
7.2.4 Lógica difusa	22/11/2018
7.2.5 Algoritmos genéticos	22/11/2018

Actividades de aprendizaje en clase	Material de apoyo:	Fechas:
P10.- Práctica de Clasificación (RL, Árboles)	- Cañón, Laboratorio	29-01/11/2018
P11.- Práctica de Reglas de Asociación (A priori)	Pack R, R Studio,	05/11/2018
P12.- Práctica de Agrupamiento (KMeans)	Anaconda Studio,	08/11/2018
P13.- Práctica de Validación de Modelos	MongoDB	12/11/2018

Actividades de aprendizaje extra-clase:	Fecha de entrega:
A7.- Investigación sobre Otras técnicas avanzadas (NN, SVM, Redes Bayesianas, etc).	11/11/2018

Evaluación Parcial (Criterios de la calificación parcial)	Fechas:										
<table><tr><td>- Prácticas (Individual/Grupal)</td><td>30 %</td></tr><tr><td>- Investigaciones (Individual/Grupal)</td><td>20 %</td></tr><tr><td>- Proyecto Final</td><td>50 %</td></tr><tr><td colspan="2">-----</td></tr><tr><td>Total</td><td>100 %</td></tr></table>	- Prácticas (Individual/Grupal)	30 %	- Investigaciones (Individual/Grupal)	20 %	- Proyecto Final	50 %	-----		Total	100 %	Examen: 26/11/2018
- Prácticas (Individual/Grupal)	30 %										
- Investigaciones (Individual/Grupal)	20 %										
- Proyecto Final	50 %										

Total	100 %										
	Revisión: 03/12/2018										
	Entrega: 06/12/2018										

Evaluación Final

OBJETIVO: *Aplicar de forma integral los conocimientos teórico-prácticos sobre la materia de Minería de Datos (DS)*

1.- Elegir una empresa (real o ficticia) de cualquier giro donde sea posible reunir 2 años de datos (ó 1000 registros como mínimo) en algún formato origen (CSV, DB, JSON, XML, etc) (Ponderación: 6.0)

Se deberá seguir la metodología CRISP-DM:

- ✓ Entendimiento del Problema (Análisis de Requerimientos, Diagrama AS IS – TO BE).
- ✓ Preparación y Transformación de los Datos (Recolección, Limpieza, Transformación, Normalización, etc)
- ✓ Modelación de al menos 2 algoritmos de ML (Supervisados o No Supervisados). Justificar su uso.
- ✓ Validación de los Modelos de Prueba y Evaluación (ROC, Matriz de Confusión).
- ✓ Presentación de resultados finales (Gráficos y Tablas)

Duración de 15 a 30 minutos máximo por equipo (3 integrantes máximo, calificación individual).

En la exposición oral los criterios a utilizar son:

1. Dominio del tema (2.0)
2. Fluidez (0.5)
3. Manejo de escenario (0.5)
4. Demostración (3.0)

2.- Generar un reporte o video donde se describa paso a paso la forma en que se desarrolló cada punto del proyecto (Reporte en formato PDF o video MP4-MKV-AVI) que contenga imágenes propias (no de terceros o internet). (Ponderación: 4.0)

Bibliografía básica:

1. EMC Education Services. Data Science and Big Data Analytics. John Wiley and Sons. USA 2015.
2. Gendron Jay. Introduction to R for Business Intelligence. Packt Publishing. UK 2014.
3. Lanz Brett. Machine Learning with R Second Edition. Packt Publishing. UK 2015.
4. Larose Daniel T. & Chantal D. Larose. Discovering Knowledge in Data (an Introduction to Data Mining). Second Edition. John Willey and Sons. USA 2014.
5. Larose Daniel T. & Chantal D. Larose. Data Mining and Predictive Analytics. Second Edition. John Willey and Sons. USA 2014.
6. Ojeda Tony and others. Practical Data Science Cookbook. Packt Publishing. UK 2014.
7. Pierson Lillian. Data Science for Dummies. John Willey and Sons. USA 2015.
8. Squire Megan. Mastering Data Mining with Python. Packt Publishing. UK 2016.
9. Viswanathan Viswa & Shanti. R Data Analysis Cookbook. Packt Publishing. UK 2015.
10. Zumel Nina & Mount John. Practical Data Science with R. Manning Publications Co. USA 2014.

Bibliografía complementaria:

11. De Vries Andre & Meys Joris. R for Dummies. John Willey and Sons. England 2012.
12. Meadows Alex & others. Pentaho Data Integration Cookbook Second Edition. Packt Publishing. UK 2013.
13. Rainardi Vincent. Building a Data Warehouse With Examples in SQL Server. Apress. USA 2008.
14. Watt Jeremy, Borhani Reza, Katsaggelos Aggelos K. Machine Learning Refined. Foundations, Algorithms, and Applications. Cambridge University Press. UK 2016.

Observaciones, aciertos, deficiencias y sugerencias para otros cursos (se llena durante el curso)