

PROGRAMA DE CURSO

Código	Nombre			
IN7J0	Introducción a la Ciencia de Datos			
Nombre en Inglés				
Data Science 101				
SCT	Unidades Docentes	Horas de Cátedra	Horas Docencia Auxiliar	Horas de Trabajo Personal
6	10	3		7
Requisitos			Carácter del Curso	
			Electivo del MBE	
Resultados de Aprendizaje				
<p>Este curso proporciona una amplia introducción sobre la ciencia de datos y sus aplicaciones en la Industria, buscando entregar herramientas y conocimientos prácticos que permitan a los alumnos identificar y entender el problema, modelarlo, evaluar sus resultados y comunicarlos de forma efectiva.</p> <p>El alumno aprenderá a buscar en los datos relaciones ocultas y patrones. Este curso presenta algunos temas avanzados de la minería de datos desde el punto de vista de sus aplicaciones más recientes, tales como la minería de textos y la minería de redes sociales, sus aplicaciones al diseño de sistemas de recomendación y a la minería de opiniones, y algunas técnicas avanzadas no cubiertas tradicionalmente.</p>				

Semestre	
Primavera 2017	
Profesores Encargados	Profesor Auxiliar
Sebastián A. Ríos, Constanza Contreras, Jorge Retamales	Sebastián Gutierrez
Horario	
Cátedras: 1 vez a la semana en horario por definir.	
Metodología Docente	Evaluación General
<p>La metodología se basa en:</p> <ul style="list-style-type: none"> Talleres Prácticos. Proyecto Final 	<ul style="list-style-type: none"> Talleres Prácticos (60%) Los alumnos desarrollaran en clases tareas aplicando las técnicas entregadas en el curso. Proyecto Final (40%) Se realizará un proyecto final de inteligencia de negocios en el cual los alumnos deberán aplicar una o más técnicas entregadas en el curso sobre datos propios.

Contenidos del Curso	
<ul style="list-style-type: none"> Parte 0: Introducción (1 semana) 	<ul style="list-style-type: none"> Definiciones básicas (ciencia de datos y Data Scientists) Evolución de la minería de datos Actualidad en Inteligencia de Negocios Ejemplos, contexto, aplicaciones Contenidos/evaluaciones del curso
<ul style="list-style-type: none"> Parte 1: Conceptos claves (1 semana) 	<ul style="list-style-type: none"> Datos e información Metodologías KDD – CRISP DM Data Manipulation – Analytics – Visualization Aprendizaje supervisado/no supervisado Clasificadores y clustering
<ul style="list-style-type: none"> Parte 2: Exploración de datos (2 semanas) 	<ul style="list-style-type: none"> Identificación de problemas Planteamiento de problema/solución Entendimiento del problema, datos y tipos de soluciones Recopilación/exploración de datos e información Definición y selección de variables
<ul style="list-style-type: none"> Parte 3: Pre-procesamiento de datos (2 semanas) 	<ul style="list-style-type: none"> Trabajar con missing values Balanceo de clases Imputación de datos Outliers Transformación de datos Reducción de dimensionalidad
<ul style="list-style-type: none"> Parte 4: Técnicas básicas de minería de datos (1 semana) 	<ul style="list-style-type: none"> Muestreo Training – validation – testing Sobreajuste Hold out y k-fold cross validation Medidas de evaluación de modelos (accuracy, precision, recall) Regresión lineal y logística
<ul style="list-style-type: none"> Parte 5: Modelos de minería de datos (8 semanas) 	<ul style="list-style-type: none"> Técnicas avanzadas de clasificación (SVM, NN, DT, RF, Naive-Bayes) Minería de textos y de opiniones Clustering (Jerárquicos, K-Means, Fuzzy C-means, dbscan, optics) Sistemas de recomendación Minería de redes sociales

<ul style="list-style-type: none"> • Parte 6: Otros temas en ciencia de datos (1 semana) 	<ul style="list-style-type: none"> • Visualización, data products y análisis de datos visuales. • Provenance, privacy, ethics and governance • Tendencia de la minería de datos
---	--

Reglas del curso
<ul style="list-style-type: none"> • Se recibirán las tareas en U-cursos, donde se explique el trabajo realizado y sus conclusiones. Este resumen debe contener: Introducción, Problema, Técnica aplicada, Resultados y conclusiones. Los Anexos no se consideran parte del Resumen. • Toda copia literal del material del expositor, Internet o trabajos de otros alumnos, será calificada con la nota mínima 1.0. Si se detecta la copia a otro alumno, lo que incurre en una falta ética, el alumno corre el riesgo de reprobado automáticamente el curso.

Semana	Actividad	Tema
1	Presentación general curso (Parte 0)	Explicación de objetivos y metodología del curso. Introducción general de los conceptos y los contenidos que se abordarán durante el semestre. Muestra de sus aplicaciones y utilidad en la Industria.
2	Parte 1	Conceptos claves
3	Parte 2	Identificación y planteamiento de problemas
4	Parte 2	Definición de variables y selección de atributos
5	Parte 3	Missing values, outliers, balanceo de datos
6	Parte 3	Transformación de atributos, reducción de dimensionalidad
7	Parte 4	Técnicas básicas de minería de datos
8	Parte 5: Clasificación	Naive-Bayes, Decision Trees, Random Forest
9	Parte 5: Clasificación	Neural net, Support Vector Machine
10	Parte 5: Clustering	Jerárquico, K-Means, K-Medoids
V	Vacaciones mitad semestre	
11	Parte 5: Clustering	Fuzzy C-Means, dbscan, optics
12	Parte 5: Minería de Texto	Pre-procesamiento y clasificación
13	Parte 5: Minería de Texto	Semántica latente y análisis de sentimientos
14	Parte 5: Análisis de Redes Sociales	Grafos, aplicaciones, métricas de las redes sociales

15	Parte 5: Sistemas de recomendación	Tipos de recomendación
16	Parte 6: Otros Temas	Privacidad, visualización de datos

Bibliografía General

Bibliografía obligatoria

- [1] J. Han & M. Kamber, Data Mining: Concepts and Techniques, 3rd ed. Morgan Kaufmann Publishers, Enero 2011
- [2] I. Witten y E. Frank, Data Mining: Practical Machine Learning Tools and Techniques, 2d ed. Morgan Kaufmann, 2005
- [3] James, Gareth, et al. An introduction to statistical learning. Vol. 6. New York: Springer, 2013.

Bibliografía complementaria

- [4] J. Hernández Orallo, M. J. Ramírez Quintana, C. Ferri, Introducción a la minería de datos, Pearson Prentice Hall, 2004
- [5] Murphy, Kevin P. Machine Learning: a probabilistic perspective. MIT press, 2012.

Vigencia desde:	Junio de 2017
Elaborado por:	Sebastián Ríos, Constanza Contreras, Jorge Retamales