

75.06 Organización de Datos.
Calendario 2do Cuatrimestre 2018
 [G##] = Gradiance /// [Num] = capítulo apunte

Día	Temas	Día	Temas
20/8	FERIADO	23/8	Administrivia, Introducción a Data Science, Introducción a Big Data, Jerarquía de Almacenamiento, Tipos de Datos, Arquitecturas Lambda. [G0][G0.5][1]
27/8	DataFrames. Análisis Exploratorio de Datos, Introducción a Pandas y Visualización de datos. [2,3]	30/8	DataFrames. Análisis Exploratorio de Datos, Pandas Avanzado. El paradigma split-apply-combine. Mas visualización de datos. [2,3]
3/9	DataFrames. Análisis Exploratorio de Datos, Pandas muy avanzado. Vencimiento para la formación de grupos. [G1][2,3]	6/9	Spark I. Hadoop. HDFS. Teoría de Map Reduce. Introducción a Spark. RDDs, modelo de ejecución, transformaciones y acciones en pyspark. [4]
10/9	Spark II. Transformaciones y Acciones II. Ejemplos con pyspark. Joins. Broadcast join.[4]	13/9	Spark III. Estructuras de alto nivel. Dataframes, SparkSQL. [G2][4]
17/9	ITI I. Complejidad de Kolmogorov. NCD. Teoría de Compresión de Datos. Huffman estático y dinámico. Compresión Aritmética. [5]	20/9	ITI II. La Familia LZ de compresores. Block Sorting. PAQ. [G3][5]
24/9	Hashing. Construcción de Funciones de Hashing, Hashing Universal, Hashing Perfecto, Cuckoo Hashing, Teorema de Johnson-Lindenstrauss. Feature Hashing. [6]. Entrega de TP1.	27/9	LSH. Minhashes, amplificación de minhashes. LSH para la distancia de Jaccard, LSH para la distancia angular, LSH para la distancia euclídeana. [G4][7]
1/10	Reducción de Dimensiones. SVD y PCA. Multidimensional Scaling, ISOMAP, Laplacian Eigenmaps, TSNE. [G5][8]	4/10	Information Retrieval I. Introducción al problema de Search. Índices Invertidos, construcción de índices, almacenamiento del léxico y punteros. Resolución de consultas. N-gramas.[9]
8/10	Information Retrieval II. Resolución de consultas ranqueadas. Método del coseno. BM25. LSI. Evaluación de consultas. [G6][9]	11/10	CLASE DE EJERCICIOS
15/10	FERIADO	18/10	PARCIAL
22/10	Machine Learning I. Teoría de Machine Learning. Teorema NFL. Underfitting y Overfitting. Naive Bayes. KNN [G7][10,11]	25/10	Machine Learning II. Clasificación. Modelos Lineales: Logistic Regression, Perceptron, Redes Neuronales. Deep Learning. Tensorflow. Keras. [G8][12]
29/10	Machine Learning III. Clasificación, Modelos basados en Árboles: Árboles de decisión, Random Forest, Boosting, XGBoost. Ensembles.[12]	1/11	Machine Learning IV. Feature Engineering y casos de estudio.
5/11	Clustering. Clustering jerárquico. K-Means, K-Means online, K-Means++, Clustering espectral. DBScan y HDBScan. [G9][13]. Entrega de notas del Parcial.	8/11	Recomendaciones. Sistemas basados en contenido. Collaborative Filtering. Matrix Factorization. Factorization Machines. [G10][14]
12/11	RECUPERATORIO	15/11	Streaming. Algoritmos de Streaming. [G11][15].
19/11	FERIADO	22/11	Page Rank y derivados. Pagerank, topicrank, trustrank, simrank, textrank, visualrank. [G12][16]
26/11	Redes Sociales. Análisis de Redes Sociales, características y algoritmos. Entrega de Notas del Recuperatorio. [G13][17]	29/11	“Usando de todo un poco para armar un Data Lake” <i>Maximiliano Mendez, Hugo Bellomosto. Despegar.</i>
3/12	CORRECCION FINAL DE TPS	6/12	CORRECCION FINAL TPS
10/12		13/12	Segundo Recuperatorio, Coloquio y Examen por Promoción.