

75.06 Organización de Datos.
Calendario 2do Cuatrimestre 2019

Día	Temas	Día	Temas
19/8	FERIADO	22/8	Administrivia, Introducción a Data Science, Introducción a Big Data,, Tipos de Datos.
26/8	DataFrames. Análisis Exploratorio de Datos, Introducción a Pandas y Visualización de datos.	29/8	DataFrames. Análisis Exploratorio de Datos, Pandas Avanzado. El paradigma split-apply-combine. Mas visualización de datos.
2/9	DataFrames. Análisis Exploratorio de Datos, Pandas muy avanzado. Vencimiento para la formación de grupos.	5/9	Spark I. Hadoop. HDFS. Teoría de Map Reduce. Introducción a Spark. RDDs, modelo de ejecución, transformaciones y acciones en pyspark.
9/9	Spark II. Transformaciones y Acciones II. Ejemplos con pyspark. Joins. Broadcast join.	12/9	Spark III. Estructuras de alto nivel.
16/9	ITI I. Complejidad de Kolmogorov. NCD. Teoría de Compresión de Datos. Huffman estático y dinámico. La familia LZ.	19/9	Hashing. Construcción de Funciones de Hashing, Hashing Universal, Hashing Perfecto, Cuckoo Hashing, Teorema de Johnson-Lindenstrauss. Feature Hashing.
23/9	LSH. Minhashes, amplificación de minhashes. LSH para la distancia de Jaccard, LSH para la distancia angular, LSH para la distancia euclideana.	26/9	LSH 2. Minhashes, amplificación de minhashes. LSH para la distancia de Jaccard, LSH para la distancia angular, LSH para la distancia euclideana.
30/9	Reducción de Dimensiones. SVD y PCA. Multidimensional Scaling, ISOMAP, Laplacian Eigenmaps, TSNE. [G5][8] Entrega de TP1.	3/10	Information Retrieval I. Introducción al problema de Search. Índices Invertidos, construcción de índices, almacenamiento del léxico y punteros. Resolución de consultas. N-gramas.[9]
7/10	Information Retrieval II. Construcción de índices invertidos y aplicaciones.	10/10	Information Retrieval III Resolución de consultas ranqueadas, método del coseno, BM25, evaluación de consultas. LSI..
14/10	FERIADO	17/10	CLASE DE EJERCICIOS
21/10	PARCIAL	24/10	Machine Learning I. Teoría de Machine Learning. Teorema NFL. Underfitting y Overfitting. Naive Bayes. KNN
28/10	Machine Learning II. Clasificación. Modelos Lineales: Logistic Regression, Perceptron, Redes Neuronales. Deep Learning. Tensorflow. Keras.	31/10	Machine Learning III. Clasificación, Modelos basados en Árboles: Árboles de decisión, Random Forest, Boosting, XGBoost. Ensamblados.
4/11	Machine Learning IV. Feature Engineering y casos de estudio.	7/11	Clustering. Clustering jerárquico. K-Means, K-Means online, K-Means++, Clustering espectral. DBScan y HDBScan.
11/11	Recomendaciones. Sistemas basados en contenido. Collaborative Filtering. Matrix Factorization. Factorization Machines. Entrega de notas del Parcial.	14/11	Streaming. Algoritmos de Streaming.
18/11	FERIADO	21/11	RECUPERATORIO
25/11	Algoritmos sobre redes: Page Rank y derivados. Pagerank, topicrank, trustrank, simrank, textrank, visualrank. Detección de comunidades en redes.	28/11	CLASE DE EJERCICIOS. Entrega de notas del recuperatorio.
2/12	CORRECCIÓN FINAL DE TPS	5/12	CORRECCIÓN FINAL DE TPS
9/12		12/12	Segundo Recuperatorio, Coloquio y Examen por Promoción.