

75.06 Organización de Datos.  
Calendario 1er Cuatrimestre 2020

Día	Temas	Día	Temas
13/4	Administrivia, Introducción a Data Science, Introducción a Big Data, Tipos de Datos.	16/4	DataFrames. Análisis Exploratorio de Datos, Introducción a Pandas y Visualización de datos.
20/4	DataFrames. Análisis Exploratorio de Datos, Pandas Avanzado. El paradigma split-apply-combine. Mas visualización de datos.	23/4	DataFrames. Análisis Exploratorio de Datos, Pandas muy avanzado. <b>Vencimiento para la formación de grupos.</b> .
27/4	Consultas sobre Pandas	30/4	Spark I. Hadoop. HDFS. Teoría de Map Reduce. Introducción a Spark. RDDs, modelo de ejecución, transformaciones y acciones en pyspark.
4/5	Spark II. Transformaciones y Acciones II. Ejemplos con pyspark. Joins. Broadcast join.	7/5	Spark III. Estructuras de alto nivel.
11/5	Consultas sobre Spark	14/5	Hashing. Construcción de Funciones de Hashing, Hashing Universal, Hashing Perfecto, Cuckoo Hashing, Teorema de Johnson-Lindenstrauss. Feature Hashing.
18/5	LSH. Minhashes, amplificación de minhashes. LSH para la distancia de Jaccard, LSH para la distancia angular, LSH para la distancia euclideana.	21/5	LSH 2. Minhashes, amplificación de minhashes. LSH para la distancia de Jaccard, LSH para la distancia angular, LSH para la distancia euclideana. <b>Entrega de TP1.</b>
25/5	FERIADO	28/5	Information Retrieval I. Introducción al problema de Search. Indices Invertidos, construcción de índices, almacenamiento del léxico y punteros. Resolución de consultas. N-gramas.[9].
1/6	Information Retrieval II. Construcción de índices invertidos y aplicaciones.	4/6	Information Retrieval III Resolución de consultas ranqueadas, método del coseno, BM25, evaluación de consultas. LSI.
8/6	Reducción de Dimensiones. SVD y PCA. Multidimensional Scaling, UMAP, Laplacian Eigenmaps, TSNE. [G5][8]	11/6	CLASE DE EJERCICIOS
15/6	FERIADO	18/6	PARCIAL
22/6	Machine Learning I. Teoría de Machine Learning. Teorema NFL. Underfitting y Overfitting. Naive Bayes. KNN	25/6	Machine Learning II. Clasificación. Modelos Lineales: Logistic Regression, Perceptron, Redes Neuronales. Deep Learning. Tensorflow. Keras.
29/6	Machine Learning III. Clasificacion, Modelos basados en Arboles: Arboles de decisión, Random Forest, Boosting, XGBoost. Ensamblés.	2/7	Machine Learning IV. Feature Engineering y casos de estudio.
6/7	NLP. Consultas de TP. <b>Entrega de notas del Parcial.</b>	9/7	FERIADO
13/7	Clustering. Clustering jerárquico. K-Means, K-Means online, K-Means++, Clustering espectral. DBScan y HDBScan.	16/7	Consultas
20/7	RECUPERATORIO	23/7	Recomendaciones. Sistemas basados en contenido. Collaborative Filtering. Matrix Factorization. Factorization Machines.
27/7	Streaming. Algoritmos de Streaming.	30/7	Algoritmos sobre redes: Page Rank y derivados. Pagerank, topicrank, trustrank, simrank, textrank, visualrank. Detección de comunidades en redes.
3/8	Consultas de TP. <b>Entrega de notas del Recuperatorio.</b>	6/8	CORRECCIÓN FINAL DE TPS
10/8	CORRECCIÓN FINAL DE TPS	13/8	CORRECCIÓN FINAL DE TPS
17/8	FERIADO	20/8	CLASE DE EJERCICIOS
24/8	Segundo Recuperatorio, Coloquio y Examen por Promoción.		