

75.06 Organización de Datos.
Calendario 2do Cuatrimestre 2020

| Día | Temas | Día | Temas |
|--------|--|-------|--|
| 28/9 | Administrivia, Introducción a Data Science, Introducción a Big Data, Tipos de Datos. | 1/10 | DataFrames. Análisis Exploratorio de Datos, Introducción a Pandas y Visualización de datos. |
| 5/10 | DataFrames. Análisis Exploratorio de Datos, Pandas Avanzado. El paradigma split-apply-combine. Mas visualización de datos. | 8/10 | DataFrames. Análisis Exploratorio de Datos, Pandas muy avanzado. Vencimiento para la formación de grupos. . |
| 12/10 | FERIADO | 15/10 | Parcialito: Pandas |
| | | | Spark I. Hadoop. HDFS. Teoría de Map Reduce. Introducción a Spark. RDDs, modelo de ejecución, transformaciones y acciones en pyspark. |
| 19/10 | Spark II. Transformaciones y Acciones II. Ejemplos con pyspark. Joins. Broadcast join. | 22/10 | Spark III. Estructuras de alto nivel. |
| 26/10 | Parcialito: Spark | 29/10 | Minhashing. Función de minhash, amplificación. Minhash para la distancia de Jaccard, distancia angular y distancia euclídea. |
| | Hashing. Construcción de Funciones de Hashing, Hashing Universal, Hashing Perfecto, Cuckoo Hashing, Teorema de Johnson-Lindenstrauss. Feature Hashing. | | |
| 2/11 | LSH. Objetivo. Uso de funciones de minhash. Estructura de tabla LSH. Resolución de consultas. Entrega de TP1. | 5/11 | Information Retrieval I. Introducción al problema de Search. Índices Invertidos, construcción de índices, almacenamiento del léxico y punteros. Resolución de consultas. N-gramas.[9]. |
| 9/11 | Parcialito: Hashing y LSH | 12/11 | Information Retrieval III Resolución de consultas ranqueadas, método del coseno, BM25, evaluación de consultas. LSI. |
| | Information Retrieval II. Construcción de índices invertidos y aplicaciones. | | |
| 16/11 | Reducción de Dimensiones. SVD y PCA. Multidimensional Scaling, UMAP, Laplacian Eigenmaps, TSNE. [G5][8] | 19/11 | Machine Learning I. Teoría de Machine Learning. Teorema NFL. Underfitting y Overfitting. Naive Bayes. KNN |
| 23/11 | FERIADO | 26/11 | Parcialito: Information Retrieval |
| | | | Machine Learning II. Clasificación. Modelos Lineales: Logistic Regression, Perceptron, Redes Neuronales. Deep Learning. Tensorflow. Keras. |
| 30/11 | Machine Learning III. Clasificación, Modelos basados en Arboles: Arboles de decisión, Random Forest, Boosting, XGBoost. Ensamblés. | 3/12 | Machine Learning IV. Feature Engineering y casos de estudio |
| 7/12 | FERIADO | 10/12 | Clase de consultas |
| 14/12 | RECUPERATORIO | 17/12 | Consultas de TP |
| Receso | | | |
| 8/2 | NLP. Consultas de TP. | 11/2 | Clustering. Clustering jerárquico. K-Means, K-Means online, K-Means++, Clustering espectral. DBScan y HDBScan. |
| 15/2 | FERIADO | 18/2 | Recomendaciones. Sistemas basados en contenido. Collaborative Filtering. Matrix Factorization. Factorization Machines. |
| 22/2 | Streaming. Algoritmos de Streaming. | 25/2 | Algoritmos sobre redes: Page Rank y derivados. Pagerank, topicrank, trustrank, simrank, textrank, visualrank. Detección de comunidades en redes. |
| 1/3 | ENTREGA FINAL DE TPS | 4/3 | ENTREGA FINAL DE TPS |
| 8/3 | Segundo Recuperatorio, Coloquio y Examen por Promoción. | | |