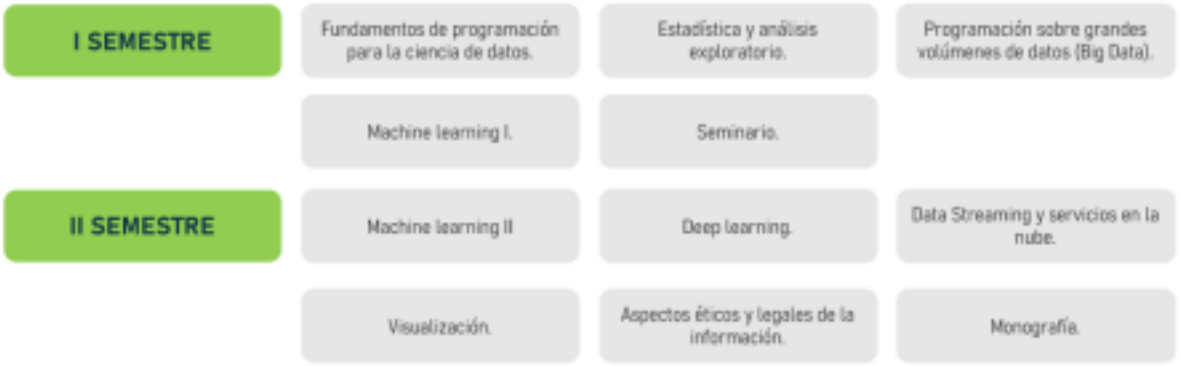
UNIVERSIDAD DE ANTIOQUIA
Facultad de Ingeniería

Especialización en Analítica y Ciencia de Datos

Universidad de Antioquia. Facultad de Ingeniería.

Plan de estudios



I Semestre

01. Fundamentos de programación para la ciencia de datos

Unidad 1. Introducción a Python.

1. Jupyter Notebook System sobre SPARK.
2. Conceptos básicos de Python.
3. Tipos de datos.
4. Lectura de archivos en diversos formatos.
5. Funciones en Python.

Unidad 2. Librerías.

6. ¿Qué son las librerías en python? y su importancia en la gestión de datos.
7. Librerías numéricas de python.

Unidad 3. Manipulación de datos.

8. Dataframe. Creación, uso y transformación.
9. El SQL y la manipulación de dataframe.
10. Generación de números aleatorios, muestras y distribuciones.

Unidad 4. Visualización de información.

11. Importancia de la visualización de datos.
12. Visualización de datos con matplotlib.

02. Estadística y análisis exploratorio

Unidad 1. Introducción a variables aleatorias y momentos.

1. Variables aleatorias, función de distribución, distribuciones bivariadas, marginales, condicionales. Teorema de Bayes.
2. Esperanza de una variable aleatoria, momentos, media y mediana, covarianza y correlación, esperanza condicional. Distribuciones especiales.
3. Tipos de variables y muestreo. Muestreo de funciones de distribución. Taller sobre muestreo, gráficas: histograma y estimador de densidad kernel.

Unidad 2. Función de distribución Gausiana y teorema de Bayes.

4. Distribuciones a priori y a posteriori, distribuciones conjugadas, estimador de Bayes. 5. Función de Distribución Gaussiana Multivariada, distribución Gausiana Condicional, distribución Gausiana Marginal, Inferencia, Máxima verosimilitud para la Gaussiana. 6. Taller sobre estimadores e inferencia.

Unidad 3. Intervalos de confianza y test de diferenciación de medias.

7. Distribución chi-cuadrada, distribución *t*, Intervalos de confianza,
8. Pruebas de hipótesis, el test *t*, test de bondad de ajuste, KS-test, ANOVA y MANOVA.
9. Taller práctico sobre estimación de intervalos de confianza y test de hipótesis. Gráfica Volcano.

Unidad 4. Preparación de datos.

10. Preparación de datos, atributos redundantes, limpieza de datos y normalización.
11. Detección de datos atípicos, imputación de variables y codificación.
12. Taller de aplicación preparación de datos y visualización.

03. Programación sobre grandes volúmenes de datos

1. Sistema de archivos distribuidos (HDFS).
2. Map-reduce.
3. Eficiencia de los procesos paralelos y distribuidos.
4. SPARK y RDD.
5. Dataframe y SQL.
6. Librerías básicas para ML

04. Machine Learning I

Unidad 1. Introducción y fundamentos del aprendizaje automático.

1. Introducción, Definiciones, Sklearn Script básico de una simulación en ML.
2. Regresión lineal y regresión logística + Taller.
3. Taller con dataset grande limpieza de datos + train/test. con métrica de score básica para regresión y para clasificación.

Unidad 2. Clasificación y selección de modelos.

4. Paramétrico vs No paramétrico: K-nn vs Gaussian. Taller sobre los modelos, fronteras de decisión.
5. Selección de modelos, overfitting y regularización.
6. Taller con dataset real selección de modelos: k-fold, k-folds estratificado, k-fold por grupos, Boostrapping.

Unidad 3. Árboles de decisión y máquinas de vectores de soporte.

7. Árboles, Bagging + Random Forest.
8. Máquinas de Vectores de Soporte, One vs All, All vs All.
9. Taller práctico comparación de modelos de la semana.

Unidad 4. Boosting y selección de características.

10. Boosting: Adaboost y Gradient Boosting.
11. Selección de características e importancia de variables.
12. Taller de aplicación de las técnicas de la semana.

05. Seminario

1. Los fundamentos metodológicos y conceptuales para la identificación y formulación de proyectos de investigación aplicada, a nivel de analítica y ciencia de datos.
2. Consideraciones prácticas de conceptos de analítica de datos en proyectos aplicados.
3. Elaboración y presentación de propuesta monográfica.

II Semestre

06. Machine Learning II

Unidad 1. Fundamentos de clustering y reducción de dimensionalidad.

1. Clustering k-means, agglomerative clustering, silhouette
2. PCA, LDA, t-SNE.
3. Taller uso de técnicas y visualización de cluster en baja dimensión.

Unidad 2. Técnicas avanzadas e integración en flujos de trabajo.

4. Spectral clustering.
5. Pipelines.

Unidad 3. Aprendizaje por refuerzo.

6. Reinforcement Learning.

Unidad 4. Taller.

7. Taller de aplicación integral con datasets propios del estudiante.

07. Deep Learning

Unidad 1: Introducción a las redes neuronales artificiales.

1. ¿Qué son las Redes neuronales?
2. Practice.
3. Metric.

Unidad 2: Redes neuronales profundas.

4. Técnicas.
5. Algoritmos de Entrenamiento.

Unidad 3: Aplicaciones de las redes neurales.

6. Analítica de imágenes con redes neuronales convolucionales.
7. Analítica de series temporales con redes neuronales recurrentes.

08. Data streaming y servicios en la nube

Unidad 1. Introducción al Streaming de Datos.

(Clasificación native stream processing. Micro-batch processing , Tipos de Datos - JSON, XML, YAML, Protocol Buffers, Apache Thrift - Data lakes vs Data Stream).

Unidad 2. Tipos de procesamiento.

(Procesamiento basado en eventos,Batch File- Based Processing, Continuous Operator Stream Processing, Stream Processing Services - Fuentes de Datos VS Almacenamiento de datos).

Unidad 3. Servicios y plataformas en la Nube.

09. Visualización

Unidad 1. Introducción.

1. Conceptos sobre técnicas de visualización.
2. Teoría de la Percepción.
3. Teoría del Color.
4. Teoría de análisis gráfico estadístico multivariado.

Unidad 2. Los datos en los gráficos.

5. Gramática de los Gráficos.
    a. Ejemplos prácticos.
6. Análisis Exploratorio de Datos.
    a. Ejemplos prácticos.

Unidad 3. Los Gráficos.

7. Taxonomía de los gráficos (Tabla periódica).
    a. Distribución
    b. Correlación
    c. Barras
    d. Jerárquicos
    e. Redes
    f. Evolución (Series)
    g. Circulares
    h. Ejemplos prácticos

8. Técnicas modernas de visualización de información, enfoque hacia altos volúmenes de datos, con reducción de dimensión.
9. Caso de estudio (*uso práctico de las técnicas y tecnologías presentadas en el curso*).
10. Análisis de ventajas y desventajas del uso de la Visualización.
11. Conclusiones.

10. Aspectos éticos y legales de la información

Unidad 1. Normatividad

1. El Habeas Data. La ley 1581 del 2012 y el decreto 1377 del 2013.
2. La normatividad internacional. El Reglamento General de Protección de Datos de la UE.

Unidad 2. La regulación en las TIC

3. La industria 4.0 y el impacto de las tecnologías digitales.
4. La privacidad y el uso de las tecnologías.

11. Monografía

Unidad 1. Asignación de asesor y definición de metodología de trabajo.

Selección y asignación de asesor(a) integral. Definición de la metodología de trabajo.

Unidad 2. Herramientas para la redacción de la monografía.

Elementos básicos para la escritura y presentación de trabajos técnicos. Recomendaciones para exposición pública de los avances.

Unidad 3. Seguimiento a los objetivos y resultados esperados.

Evaluación intermedia del avance en el cumplimiento de los objetivos, y cumplimiento de resultados esperados. Evaluación final del cumplimiento de resultados esperados y alcance de los objetivos propuestos