

PLAN DE ESTUDIOS

# Data Science



# Temario

## Clase por clase

### CLASE 1

- Introducción a Data Science
- Fundamentos de programación con Python

### CLASE 2

- Python - Tipos de datos.
- Funciones

### CLASE 3

- Clases y Objetos

### CLASE 4

- Consultas básicas SQL
- Filtrado de información

### CLASE 5

- Condicionales
- Subconsultas
- Tipo de datos y conversiones

### CLASE 6

- Tratamiento de tablas
  - Crear
  - Actualizar
  - Eliminar
  - Cargar
- Window Functions en SQL

## CLASE 7

- Vistas
- Uniones
- Joins

## CLASE 8

- Variables discretas y continuas
- Estadísticos: tendencia central y dispersión
- La forma de una distribución
- Población y muestra

## CLASE 9

- Correlación y causalidad
- Estadísticos con Numpy
- Probabilidad y estadística con Pandas

## CLASE 10

- Exploración de datos
- Matplotlib
- Seaborn

## CLASE 11

- Definición de Machine Learning
- Introducción al aprendizaje supervisado
- Modelos de regresión: definición y enfoque

## CLASE 12

- Modelos de clasificación: definición y enfoque
- Pipeline del machine learning

## CLASE 13

- Preprocesamiento de los datos
  - Manejo de los valores faltantes
  - Tratamiento de variables categóricas

## CLASE 14

- Escalado de datos
- Manejo de valores atípicos
- La maldición de la dimensionalidad: introducción

## CLASE 15

- Introducción a sklearn
  - Sklearn en el preprocesamiento de datos
  - Pipelines de machine learning

## CLASE 16

- Introducción:
  - Definiciones básicas
  - Diagramas de voronoi
  - Descenso del gradiente
  - Métricas de evaluación

## CLASE 17

- Algoritmos de Regresión
  - Regresión simple
  - Árboles regresores

## CLASE 18

- Regresión por descenso del gradiente
- Regresión avanzada manejo polinomial

## CLASE 19

- Algoritmos de Clasificación
  - Árboles de decisión
  - K-vecinos más cercanos
- Regresión logística
- Máquinas de Vectores de Soporte

## CLASE 20

- Modelos de ensamble
  - Bagging
  - Random Forest

## CLASE 21

- Boosting y XGBoost
- Staking

## CLASE 22

- Interpretación de las métricas en modelos de regresión
  - Error absoluto
  - Error cuadrático medio

## CLASE 23

- Interpretación de las métricas en modelos de clasificación
  - Matriz de confusión
  - Accuracy
  - Precisión
  - Recall
  - F-Score

## CLASE 24

- Resultados en términos de probabilidades
- La curva ROC

## CLASE 25

- Validación y optimización de modelos de Machine Learning
  - Sesgo y varianza
  - Overfitting y Underfitting
  - Curvas de validación

## CLASE 26

- Bootstrapping
- Validación cruzada y su interpretación estadística

## CLASE 27

- Definición y ejemplos de hiperparámetros
- Espacios combinatorios
- Gridsearch y RandomSearch

## CLASE 28

- Introducción:
  - Definiciones básicas
  - Algoritmos no supervisados vs algoritmos supervisados

## CLASE 29

- Algoritmos de agrupamiento
  - k-means
  - dbscan
  - Clustering-jerárquico
  - Aplicaciones

## CLASE 30

- Algoritmos de reducción de la dimensionalidad
  - Descomposición por valores singulares - SVD

## CLASE 31

- Análisis de componentes principales - PCA
- Aplicaciones

## CLASE 32

- Aplicaciones de aprendizaje no supervisados sobre la ingeniería de Features
  - Análisis descriptivo de variables
  - Aplicaciones del PCA
  - Reducción de la dimensionalidad

## CLASE 33

- Introducción
- Perceptrón
- Perceptrón multicapa

## CLASE 34

- Funciones de pérdida
- El descenso del gradiente en el ajuste de pesos
- Interpretación, validación y optimización en modelos de redes neuronales

## CLASE 35

- Tensorflow y keras

## CLASE 36

- Introducción al manejo de imágenes
  - Representación
  - Imágenes a color
  - Imágenes en escala de grises

## CLASE 37

- Introducción a la convolución
  - Operación de convolución
  - Detectores de características

## CLASE 38

- Redes neuronales convolucionales
  - Etapa de convolución
  - Pooling
  - Flattening
  - Full connection

## CLASE 39

- Introducción
  - Flujo de trabajo
  - Normalización
  - Vectorización
  - Term Frequency - Inverse Document Frequency - TF-IDF

## CLASE 40

- Bag of Words
  - Vectorizar con N-gramas
  - Vectorizar TF-IDF

## CLASE 41

- GPT2
  - Introducción
  - Transformers
  - Recopilación de un conjunto de datos
  - Entrenamiento y puesta a punto de nuestro modelo
  - Casos de uso y el futuro

## **CLASE 42**

- Web Scraping con BeautifulSoup
  - Ejemplos

## **CLASE 43**

- Introducción
- La matriz de utilidad
- Filtrado basado en contenido
- Filtrado colaborativo

## **CLASE 44**

- Distancia y similitud
- Descomposición UV
  - Evaluación e interpretación del rendimiento

## **CLASE 45**

- Sistemas recomendadores híbridos

## **CLASE 46**

- Introducción a MLOps
- Ciclo de vida de un modelo de machine learning

## **CLASE 47**

- MPIs

## **CLASE 48**

- MLFlow

## MÓDULOS

### Módulo I

#### ELEMENTOS DE DATA SCIENCE

##### Competencia

- Analiza datos a pequeña y gran escala, utilizando herramientas matemáticas, estadísticas y computacionales para resolver problemas complejos.
- Implementa un modelo baseline de machine learning utilizando aprendizaje supervisado con el fin de comprender el proceso de extracción e interpretación de métricas de negocio.

##### Unidad Didáctica I

###### Introducción al Data Scientist

###### Temario

- Introducción a Data Science
- Fundamentos de programación con Python
- Python - Tipos de datos.
- Funciones
- Clases y Objetos

##### Unidad Didáctica II

###### Nombre: Fundamentos de Bases de Datos

###### Temario

- Consultas básicas SQL
- Filtrado de información
- Funciones de agregado SQL
- Condicionales
- Subconsultas
- Tipo de datos y conversiones
- Tratamiento de tablas
  - Crear
  - Actualizar
  - Eliminar
  - Cargar

- Window Functions en SQL
- Vistas
- Uniones
- Joins

### **Unidad Didáctica III**

**Nombre: Estadística Básica para el análisis exploratorio de datos**

#### **Temario**

- Variables discretas y continuas
- Estadísticos: tendencia central y dispersión
- La forma de una distribución
- Población y muestra
- Correlación y causalidad
- Estadísticos con Numpy
- Probabilidad y estadística con Pandas
- Exploración de datos
  - Matplotlib
  - Seaborn

### **Unidad Didáctica IV**

**Nombre: Introducción al Machine Learning**

#### **Temario**

- Definición de Machine Learning
- Introducción al aprendizaje supervisado
- Modelos de regresión: definición y enfoque
- Modelos de clasificación: definición y enfoque
- Pipeline del machine learning

### **Actividad Integradora**

Práctica: Estrategias de casos de usos

Teoría: Cuestionarios

## MÓDULO II

# MACHINE LEARNING AVANZADO

### Competencia

Desarrolla modelos de machine learning utilizando aprendizaje supervisado y no supervisado para caracterizar procesos reales y medir el impacto usando indicadores clave de dominio.

### Unidad Didáctica I

**Nombre:** Ingeniería de Features

#### Temario

- Introducción a Data Science
- Fundamentos de programación con Python
- Python - Tipos de datos.
- Funciones
- Clases y Objetos

### Unidad Didáctica II

**Nombre:** Fundamentos de Bases de Datos

#### Temario

- Consultas básicas SQL
- Filtrado de información
- Funciones de agregado SQL
- Condicionales
- Subconsultas
- Tipo de datos y conversiones
- Tratamiento de tablas
  - Crear
  - Actualizar
  - Eliminar
  - Cargar
- Window Functions en SQL
- Vistas
- Uniones
- Joins

### **Unidad Didáctica III**

**Nombre:** Estadística Básica para el análisis exploratorio de datos

#### **Temario**

- Variables discretas y continuas
- Estadísticos: tendencia central y dispersión
- La forma de una distribución
- Población y muestra
- Correlación y causalidad
- Estadísticos con Numpy
- Probabilidad y estadística con Pandas
- Exploración de datos
  - Matplotlib
  - Seaborn

### **Unidad Didáctica IV**

**Nombre:** Introducción al Machine Learning

#### **Temario**

- Definición de Machine Learning
- Introducción al aprendizaje supervisado
- Modelos de regresión: definición y enfoque
- Modelos de clasificación: definición y enfoque
- Pipeline del machine learning

#### **Actividad Integradora**

**Práctica:** Estrategias de casos de usos

**Teoría:** Cuestionarios

## **MÓDULO II**

### **MACHINE LEARNING AVANZADO**

#### **Competencia**

Desarrolla modelos de machine learning utilizando aprendizaje supervisado y no supervisado para caracterizar procesos reales y medir el impacto usando indicadores clave de dominio.

## Unidad Didáctica I

**Nombre:** Ingeniería de Features

### Temario

- Preprocesamiento de los datos
  - Manejo de los valores faltantes
  - Tratamiento de variables categóricas
  - Escalado de datos
  - Manejo de valores atípicos
  - La maldición de la dimensionalidad: introducción
- Introducción a sklearn
  - Sklearn en el preprocesamiento de datos
  - Pipelines de machine learning

## Unidad Didáctica II

**Nombre:** Algoritmos de aprendizaje supervisado

### Temario

- Introducción:
  - Definiciones básicas
  - Diagramas de voronoi
  - Descenso del gradiente
  - Métricas de evaluación
- Algoritmos de Regresión
  - Regresión simple
  - Árboles regresores
  - Regresión por descenso del gradiente
  - Regresión avanzada manejo polinomial
- Algoritmos de Clasificación
  - Árboles de decisión
  - K-vecinos más cercanos
  - Regresión logística
  - Máquinas de Vectores de Soporte
- Modelos de ensamble
  - Bagging
  - Random Forest
  - Boosting y XGBoost
  - Staking

## Unidad Didáctica III

**Nombre: Interpretación, validación y optimización de modelos**

### Temario

- Interpretación de las métricas en modelos de regresión
  - Error absoluto
  - Error cuadrático medio
- Interpretación de las métricas en modelos de clasificación
  - Matriz de confusión
  - Accuracy
  - Precisión
  - Recall
  - F-Score
  - Resultados en términos de probabilidades
  - La curva ROC
- Validación y optimización de modelos de Machine Learning
  - Sesgo y varianza
  - Overfitting y Underfitting
  - Curvas de validación
  - Bootstrapping
  - Validación cruzada y su interpretación estadística
  - Definición y ejemplos de hiperparámetros
  - Espacios combinatorios
  - Gridsearch y RandomSearch

## Unidad Didáctica IV

**Nombre: Algoritmos de aprendizajes no supervisados**

### Temario

- Introducción:
  - Definiciones básicas
  - Algoritmos no supervisados vs algoritmos supervisados
- Algoritmos de agrupamiento
  - k-means
  - dbscan
  - Clustering-jerárquico
  - Aplicaciones
- Algoritmos de reducción de la dimensionalidad
  - Descomposición por valores singulares - SVD
  - Análisis de componentes principales - PCA
  - Aplicaciones

- Aplicaciones de aprendizaje no supervisados sobre la ingeniería de Features
  - Análisis descriptivo de variables
  - Aplicaciones del PCA
  - Reducción de la dimensionalidad

### **Actividad Integradora**

**Práctica:** Estrategias de casos de usos

**Teoría:** Cuestionarios

## **MÓDULO III**

### **ALGORITMOS AVANZADOS Y APLICACIONES**

#### **Competencia**

Aplica algoritmos de machine learning a diferentes campos especializados de la inteligencia artificial para el desarrollo de aplicaciones reales de ciencia de datos

#### **Unidad Didáctica I**

**Nombre:** Redes neuronales artificiales

##### **Temario**

- Introducción
- Perceptrón
- Perceptrón multicapa
- Funciones de pérdida
- El descenso del gradiente en el ajuste de pesos
- Interpretación, validación y optimización en modelos de redes neuronales
- Tensorflow y keras

#### **Unidad Didáctica II**

**Nombre:** Procesamiento de imágenes

##### **Temario**

- Introducción al manejo de imágenes
  - Representación
  - Imágenes a color
  - Imágenes en escala de grises
- Introducción a la convolución
  - Operación de convolución
  - Detectores de características

- Redes neuronales convolucionales
  - Etapa de convolución
  - Pooling
  - Flattening
  - Full connection

### Unidad Didáctica III

**Nombre:** Procesamiento del lenguaje natural

#### Temario

- Introducción
  - Flujo de trabajo
  - Normalización
  - Vectorización
  - Term Frequency - Inverse Document Frequency - TF-IDF
- Bag of Words
  - Vectorizar con N-gramas
  - Vectorizar TF-IDF
- GPT2
  - Introducción
  - Transformers
  - Recopilación de un conjunto de datos
  - Entrenamiento y puesta a punto de nuestro modelo
  - Casos de uso y el futuro
- Web Scraping con BeautifulSoup
  - Ejemplos

### Unidad Didáctica IV

**Nombre:** Sistemas de Recomendación

#### Temario

- Introducción
- La matriz de utilidad
- Filtrado basado en contenido
- Filtrado colaborativo
- Distancia y similitud
- Descomposición UV
- Evaluación e interpretación del rendimiento
- Sistemas recomendadores híbridos

## Unidad Didáctica V

**Nombre:** Estrategias de despliegue de Machine Learning

### Temario

- Introducción a MLOps
- Ciclo de vida de un modelo de machine learning
- MPIS
- MLFlow

### Actividad Integradora

**Práctica:** Estrategias de casos de usos

**Teoría:** Cuestionarios



Contáctanos:

---

[aplicantes@devplace.com.ar](mailto:aplicantes@devplace.com.ar)