# CIENCIA **DE DATOS**

#### **MATEMÁTICAS**

- Álgebra Lineal
- Matrices
- Vectores
- Optimización
- Regresión

- Reducción de
- dimensionalidad
- Estimación
- de densidad
- Clasificación

#### **PROBABILIDAD**

- Distribuciones discretas
- Distribuciones continuas
- Distribución normal
- Probabilidad
- Variables aleatorias
- Función de una cariable aleatoria

# **ESTADÍSTICAS**

- Descripción de datos
- Muestras aleatorias
- Distribución muestral
- Estimación de parámetros
- Prueba de hipótesis
- ANOVA

- Procesos estocasticos
- Diseño de experimentos
- Regresión lineal simple
- Correlación
- Regresión múltiple
- Gráficos básicos

#### **PROGRAMACIÓN**

Básico

-dplyr

Tidyr

-ggplot2

·shiny, etc.

# Python

- Básico
- Numpy
- Pandas

· SQL

- Matplotlib
- Seaborn, etc.

Base de Datos

MongoDB

#### **Otros**

- - Estructura de datos,
  - Web scraping
  - Git

#### APRENDIZAJE AUTOMÁTICO

#### Introducción

- Funcionamiento de modelos
- Exploración de datos
- Validación de modelos
- Sobreajuste y subajuste
- Bosque aleatorio
- Scikit-learn

#### Intermedio

- Valores vacíos
- Variables categoricas
- Pipelines/tuberías
- Validación cruzada
- Xgboost
- Derrame de datos

#### APRENDIZAJE PROFUNDO

- Redes neuronales artificiales
- Redes neuronales convolucionales
- Redes recurrentes
- Keras
- Pytorch
- Tensorflow
- Una neurona
- Redes neuronales profundas
- Descenso de gradiente estocastico
- Clasificación binaria

## INGENIERIA DE VARIABLES

- Modelo base
- Codificación categórica
- Generación de variables
- Selección de variables

# PROCESAMIENTO DE **LENGUAJE NATURAL**

- Clasificación de texto
- Vectorizacion de palabras

# **HERRAMIENTA DE VISUALIZACIÓN**

- PowerBI
- Tableau
- Qliksense
- Qlikview

# DESPLIEGUE

- Azure
- Heroku
- Google cloud
- Flask
- Django

# **OTROS PUNTOS**

- Conocimiento del área
- Habilidades de comunicación
- Aprendizaje reforzado
- Estudio de casos

**iSEGUIR PRACTICANDO!**