



# Introducción a la Ciencia de Datos para la Toma de Decisiones en el Sector Energético

Oscar A. Esquivel-Flores

Escuela de Gobierno y transformación pública

## RESUMEN

En el marco de las actividades científico-tecnológicas que se desarrollan para la “Red de Conocimientos Cero Incidentes en la Red de Ductos de México” se presenta el seminario titulado **Introducción a la Ciencia de Datos para la Toma de Decisiones en el Sector Energético** en el que se exponen diversas herramientas computacionales y metodológicas para el análisis de datos con aplicación en el área de hidrocarburos. A continuación se presentan los contenidos, aprendizajes y bibliografía de los módulos I-IV de este seminario.

## Índice

<b>1. Introducción</b>	<b>1</b>
1.1. Objetivo . . . . .	1
1.2. Requisitos . . . . .	1
1.3. Enfoques . . . . .	1
<b>2. Contenidos</b>	<b>2</b>
2.1. Python . . . . .	3
2.2. Python Científico . . . . .	4
2.3. Aprendizaje Automático . . . . .	5
2.4. Caso De Estudio . . . . .	6

# 1. Introducción

## 1.1. Objetivo

El objetivo del seminario consiste en que el participante afine sus habilidades computacionales y profundice en el análisis de datos auxiliado de las herramientas que ofrece el lenguaje de programación Python.

El diplomado abarcará los siguientes temas:

- Conocimiento del lenguaje de programación Python
- Manejo de las librerías del entorno científico de Python
- Introducción al análisis de datos por medio de algoritmos de aprendizaje automático
- Desarrollo de soluciones tecnológicas a un caso de estudio

Estos temas de los módulos I-IV se presentarán en 4 sesiones de 5 horas cada una. Se ofrecerá al participante el material didáctico necesario para la debida comprensión de los conceptos y se utilizará software libre.

## 1.2. Requisitos

Se asume que el participante tiene conocimientos básicos sobre algún lenguaje de programación y los conocimientos de estadística descriptiva. Cada participante puede utilizar su computadora personal (se sugiere sistema operativo tipo unix, *no indispensable*), no obstante se cuenta con el equipo necesario en la sala de exposición.

## 1.3. Enfoques

El **enfoque didáctico** del diplomado está orientado al logro de los siguientes aspectos:

- *Teórico*: Conocer los conceptos relativos a las metodologías computacionales utilizadas actualmente para el manejo de datos.
- *Práctico*: Desarrollar las habilidades en el manejo de herramientas computacionales que permitan extraer, organizar y analizar información proveniente de situaciones prácticas en el sector de hidrocarburos.
- *Analítico*: Analizar información y aplicar algoritmos básicos de aprendizaje automático a los datos disponibles.

El uso de herramientas computacionales actuales bajo los enfoques planteados anteriormente, coadyuvará al participante a:

- Valorar del conocimiento científico y tecnológico actuales como herramientas fundamentales para agregar valor al sector energético.
- Entender al conocimiento como un proceso multidisciplinario.
- Resolver problemas por medio de las tecnologías computacionales.
- Plantear soluciones innovadoras.
- Adquirir habilidades de trabajo intelectual y conocimientos específicos a través de trabajo colaborativo y participativo.
- Fundamentar con racionalidad y rigor científico las soluciones propuestas a problemas específicos.
- Desarrollar un pensamiento lógico, reflexivo, crítico auxiliado del uso de nuevas tecnologías computacionales.
- Ampliar su experiencia profesional.

## 2. Contenidos

Los módulos a impartirse son:

- La **Módulo I** proporciona una visión general del lenguaje de programación Python.
- La **Módulo II** profundiza en el entorno científico de Python. Se revisan las herramientas mas representativas de este ecosistema.
- La **Módulo III** introduce al análisis de datos por medio de técnicas de aprendizaje automático. Se implementan los algoritmos y se utiliza la librería Scikit-Learn
- La **Módulo IV** recopila las habilidades adquiridas y elabora estrategias de análisis de datos aplicadas a un caso de estudio particular.

Los tiempos para cada sección son los siguientes:

- Conceptos básicos de Python (5 horas)
- Herramientas de obtención, manipulación de datos (5 horas)
- Visualización de datos (5 horas)
- Análisis de datos y caso de estudio (5 horas)

La propuesta de contenidos organizados por temática, aprendizajes y estrategias didácticas es la siguiente:

## 2.1. Python

Módulo I		
<b>Propósito:</b> al finalizar la sección el participante tendrá una visión general del lenguaje de programación Python y las ventajas de su uso.		
Temática	Aprendizajes	Estrategias
Conceptos básicos de la programación en Python: <ul style="list-style-type: none"><li>■ Instalación de Python</li><li>■ Variables y operaciones</li><li>■ Operadores</li><li>■ Funciones</li></ul> “Codeando” con Python: <ul style="list-style-type: none"><li>■ Tipos de datos</li><li>■ Listas, cadenas</li><li>■ Control de flujo</li><li>■ Módulos, atributos y métodos</li><li>■ Archivos</li><li>■ Clases y objetos</li></ul>	El participante: <ul style="list-style-type: none"><li>■ Conoce las distribuciones e instalación del ecosistema Python</li><li>■ Compara los tipos de ambientes de trabajo de Python</li><li>■ Comprende la sintaxis básica de Python</li><li>■ Usa, define e implementa funciones <i>Pythonicas</i></li><li>■ Diseña e implementa algoritmos utilizando los tipos de datos estudiados</li><li>■ Conoce los módulos, propiedades y métodos de objetos</li><li>■ Conoce la manera de abrir, leer y escribir en archivos</li></ul>	<ul style="list-style-type: none"><li>■ Exposición de las distribuciones de Python y la correspondiente instalación.</li><li>■ Propiciar una discusión colectiva sobre las ventajas del lenguaje después de explorar su sintaxis básica.</li><li>■ Elaborar distintos ejercicios de programación en los que se utilicen los elementos básicos del lenguaje y ejercitar el diseño de funciones para resolver y automatizar acciones variadas.</li></ul>

La bibliografía y referencias sugeridas para esta sección son:

- *Python*, recuperado el 7 de enero de 2018 de <https://www.python.org/>
- *Python Docs*, recuperado el 10 de enero de 2018 de <https://docs.python.org/3/>
- *The Python Wiki*, recuperado el 7 de enero de 2018 de <https://wiki.python.org/moin/BeginnersGuide/Overview>
- BEAZLEY, D. y Jones, B., *Python Cookbook*. 3a ed. New York, O'Reilly, 2013.
- BAHIT, Eugenia, *Curso: Python para Principiantes*, Buenos Aires, Safe Creative, 2012.
- BARRY, Paul, *Head First Python*. Boston, O'Reilly, 2011.
- DOWNEY, Allen, *Think Python. How to Think Like a Computer Scientist*, Green Tea Press, 2013.
- GONZALEZ, Raúl, *Python para todos*, Barcelona, Creative Commons.
- GRIES, P., et. al., *Practical Programming. An Introduction to Computer Science Using Python 3*. 2a ed. The Pragmatic Bookshelf, 2013.
- LUTZ, Mark, *Python Pocket Reference. Python in your Pocket*, O'Reilly, 2014.

## 2.2. Python Científico

Módulo II		
<b>Propósito:</b> al finalizar la unidad el alumno conocerá y utilizará las librerías científicas del ecosistema <i>SciPy</i> mediante la codificación de segmentos de código con Ipython notebooks.		
Temática	Aprendizajes	Estrategias
Python en la ciencia: <ul style="list-style-type: none"> <li>■ Jupyter-Notebooks</li> <li>■ Python Científico</li> <li>■ NumPy</li> <li>■ Matplotlib</li> <li>■ Plotly</li> </ul>	El participante: <ul style="list-style-type: none"> <li>■ Conoce Python interactivo con Jupyter-Notebooks <i>notebooks</i></li> <li>■ Conoce los paquetes y librerías diseñados para el cómputo científico en Python</li> <li>■ Utiliza los elementos más importantes de la librería <i>NumPy</i></li> <li>■ Estudia los elementos básicos de <i>Matplotlib</i></li> <li>■ Diseña e implementa código para obtener gráficas en 2D y 3D</li> <li>■ Conoce los elementos sustanciales de la librería <i>Plotly</i></li> <li>■ Elabora gráficos dinámicos con <i>Plotly</i></li> </ul>	<ul style="list-style-type: none"> <li>■ Realizar ejercicios utilizando el entorno Jupyter-Notebooks</li> <li>■ Elaborar un mapa de librerías dentro del entorno científico de Python (<i>SciPy</i>)</li> <li>■ Describir las características principales de diversas librerías para el cómputo científico</li> <li>■ Practicar el cálculo de operaciones matriciales utilizando <i>Numpy</i></li> <li>■ Ejemplificar con segmentos de código la elaboración de gráficos con <i>Matplotlib</i></li> <li>■ Instalar los paquetes necesarios para utilizar <i>Plotly</i></li> <li>■ Exposición de las funcionalidades de librería <i>Plotly</i></li> </ul>

La bibliografía y referencias sugeridas para esta unidad son:

- *IPython. Interactive Computing*, consultado el 7 de enero de 2018 de <https://www.python.org/>
- *Jupyter. The Jupyter Notebook*, consultado el 10 de enero de 2018 de <http://jupyter.org/>
- *Matplotlib*, consultado el 10 de enero de 2017 de <https://matplotlib.org>
- *Scipy*, consultado el 10 de enero de 2018 de <http://www.scipy.org/>
- *Numpy*, consultado el 10 de enero de 2018 de <http://www.numpy.org/>
- *Python Scientific Lecture Notes (Spanish translation)*, consultado el 7 de septiembre de 2018 de <https://claudiovz.github.io/scipy-lecture-notes-ES/index.html>
- *Plotly. Modern Visualization for the Data Era*, consultado el 7 de septiembre de 2018 de <https://plot.ly/>
- FANGOHR, H., *Introduction to Python for Computational Science and Engineering*, University of Southampton, 2015.

- LANGTANGEN, H., *Python Scripting for Computational Science*, 3rd edition, Springer, 2007.
- LANGTANGEN, H., *A primer on Scientific Programming with Python*, 2nd edition, Springer, 2011.
- TOSI, Sandro, *Matplotlib for Python Developers*, Packt, U.K., 2009.

## 2.3. Aprendizaje Automático

Módulo III		
<b>Propósito:</b> al finalizar la sección el participante comprenderá y aplicará algunos algoritmos de aprendizaje automático utilizando Python y las librerías científicas <i>Scikit-Learn</i>		
Temática	Aprendizajes	Estrategias
Python en la ciencia: <ul style="list-style-type: none"> <li>■ Pandas</li> <li>■ Principios de Aprendizaje Automático</li> <li>■ Algoritmos supervisados</li> <li>■ Algoritmos no supervisados</li> <li>■ Limpieza de datos</li> <li>■ Implementación de algoritmos</li> <li>■ Scikit-Learn</li> </ul>	El participante: <ul style="list-style-type: none"> <li>■ Conoce el paquete <i>Pandas</i> y sus funcionalidades para el manejo de datos</li> <li>■ Adquiere habilidades para extraer y organizar datos con <i>Pandas</i></li> <li>■ Comprende los principios básicos del Aprendizaje automático.</li> <li>■ Conoce las propiedades del aprendizaje supervisado y no supervisado</li> <li>■ Implementa algoritmos de aprendizaje automáticos</li> <li>■ Utiliza el paquete <i>Scikit-Learn</i> para implementar algoritmos de aprendizaje automático</li> </ul>	<ul style="list-style-type: none"> <li>■ Realizar lecturas previas sobre aprendizaje automático</li> <li>■ Elaborar un mapa de algoritmos de aprendizaje automático</li> <li>■ Exposición de los fundamentos teóricos de algunos algoritmos básicos de regresión y clasificación</li> <li>■ Implementación de los algoritmos estudiados en trabajo colaborativo.</li> <li>■ Exponer las características básicas del paquete <i>Scikit-Learn</i></li> <li>■ Propiciar una discusión colectiva sobre la elaboración de soluciones a problemáticas por medio de algoritmos de aprendizaje automático</li> </ul>

La bibliografía sugerida para esta unidad es:

- *Pandas. Python Data Analysis Library*, consultado el 10 de enero de 2018 de <http://pandas.pydata.org>
- *Scikit-learn. Machine Learning in Python*, consultado el 10 de enero de 2018 de <http://scikit-learn.org/stable>
- McKINNEY, W., *Python for Data Analysis*, O'Reilly, 2013.
- NELLI, F., *Python Data Analytics*, Apress, 2015.

- MÜLLER, A., Guido, S., *Introduction to Machine Learning with Python*, O'Reilly, 2016.
- HASTIE, T., *et al.*, *The Elements of Statistical Learning*, 2nd. edition, Springer, 2008.
- MURPHY, K., *Machine Learning. A probabilistic perspective*, MIT Press, 2012.
- GÉRON, A., *Hands-On Machine Learning with Scikit-Learn and TensorFlow*, O'Reilly, 2017.

## 2.4. Caso De Estudio

Módulo IV		
<b>Propósito:</b> al finalizar la sección el participante diseñó una estrategia de solución a un caso de estudio particular en las que utiliza las herramientas computacionales del entorno científico de python.		
Temática	Aprendizajes	Estrategias
<ul style="list-style-type: none"> <li>■ Definición de caso de estudio</li> <li>■ Elaboración de estrategias de solución</li> <li>■ Codificación de algoritmos</li> <li>■ Pruebas y afinación de algoritmos</li> <li>■ Conclusiones cualitativas y cuantitativas</li> </ul>	<p>El participante:</p> <ul style="list-style-type: none"> <li>■ Elabora estrategias de solución de problemas utilizando herramientas computacionales</li> <li>■ Adquiere habilidades para analizar datos</li> <li>■ Valora las habilidades adquiridas</li> <li>■ Evalúa la eficacia de python y su ecosistema científico como herramienta para analizar datos</li> <li>■ Logra obtener conclusiones cualitativas del análisis cuantitativo de datos</li> </ul>	<ul style="list-style-type: none"> <li>■ Elaboración de grupos de trabajo para elegir y diseñar una solución a un problema particular.</li> <li>■ Trabajo en conjunto para llevar a cabo el flujo de tareas.</li> <li>■ Evaluar de manera colectiva los resultados obtenidos.</li> <li>■ Fomentar la discusión colectiva para llegar a conclusiones.</li> </ul>

La bibliografía sugerida para esta unidad es:

- *Kaggle: Your Home for Data Science*, consultado el 10 de enero de 2018 de <https://www.kaggle.com/>
- OJEDA, T., *et al.*, *Practical Data Science Cookbook*, Packt Publishing, 2014.
- BOWLES, M., *Machine Learning in Python. Essential Techniques for Predictive Analysis*, Wiley, 2015.
- AVILA, J., Hauck, T., *Scikit-learn Cookbook*, 2nd. editon, Packt, 2017.