



Machine Learning con Python

César Quezada
Especialista en Analítica & BI – DMC

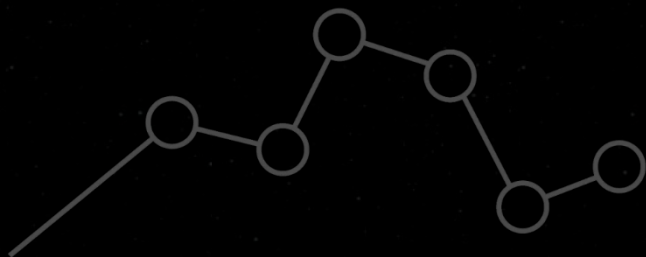


Hola!
Soy [César Quezada]

Ingeniero Estadístico UNI

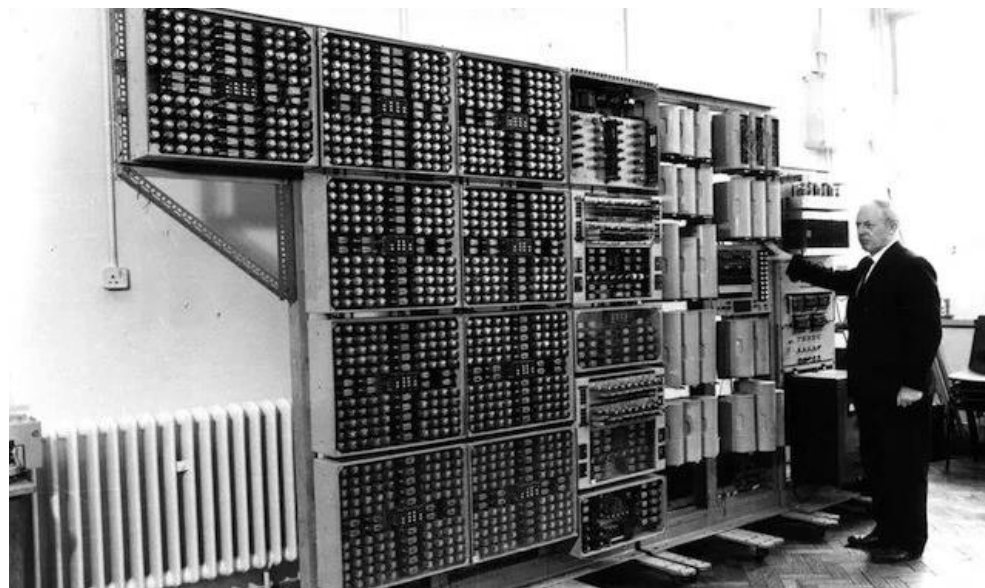
**Apasionado del mundo de la analítica
Especialista en Analítica & BI**

Di!C
Perú



¿Qué es Big Data?

¿Donde se generan todos los tipos de datos?



En 1944 se crea la primera computadora – ENIAC. Fue elaborada por científicos e ingenieros de la universidad de Pensilvania - EEUU



¿Qué es Big Data?

¿Donde se generan todos los tipos de datos?



Redes sociales y multimedia
(todos generamos datos)



Instrumentos científicos
(colección de toda clase de datos)



Dispositivos móviles
(seguimiento de objetos)



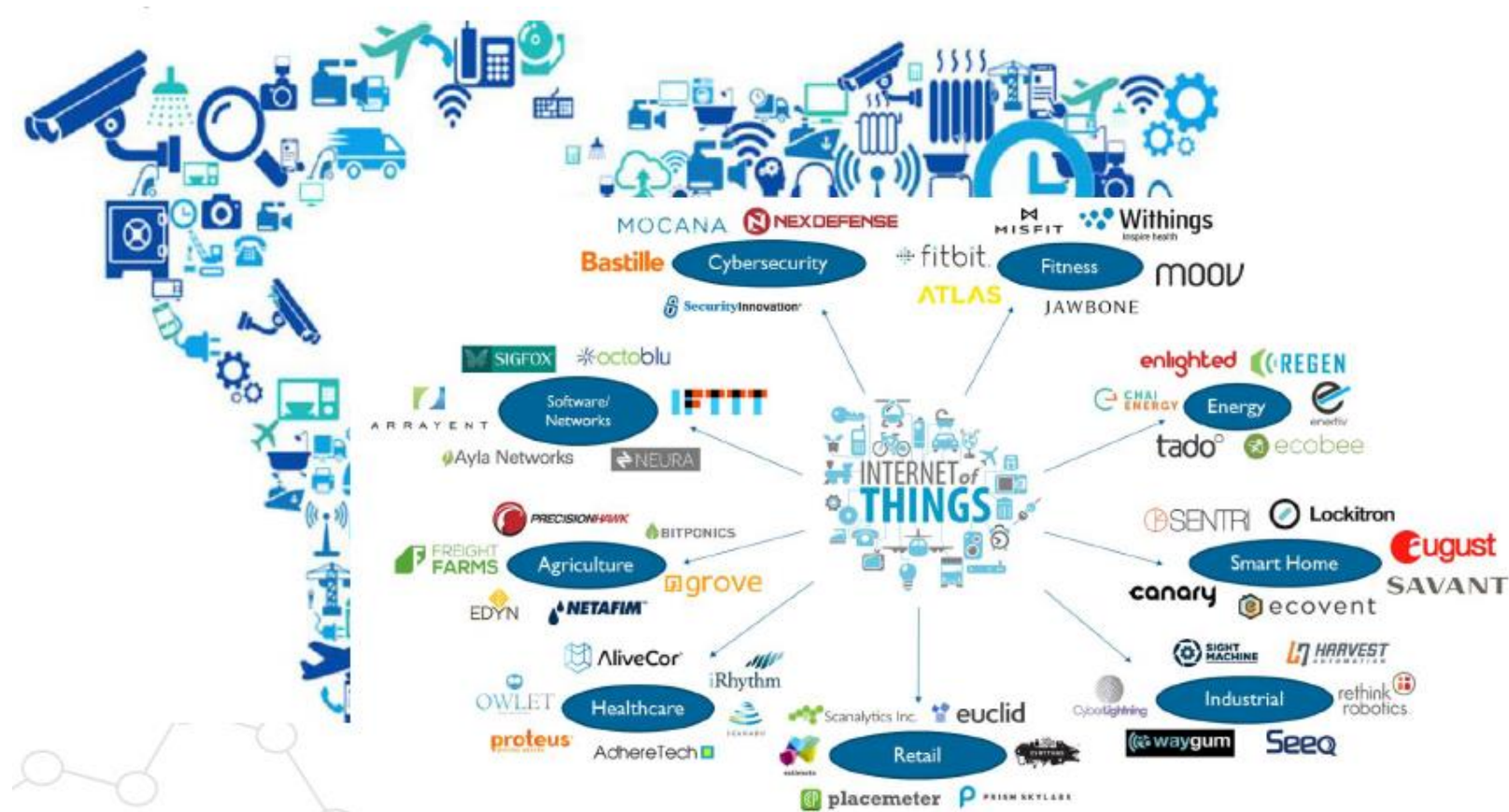
Redes de sensores
(se miden toda clase de datos)

¿Qué es Big Data?

¿Donde se almacenan todos los datos que se generan en nuestra vida diaria?



¿Qué es Big Data?



Big data es la tecnología clave para procesar grandes cantidades de datos

¿Qué es un Data Science?

Datos de distintas fuentes



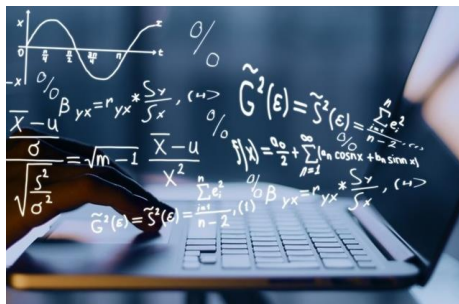
La ciencia que nos permite sacar conclusiones a partir de ello es:



Dominio de Matemática, Estadística, Sistemas, Informática, Marketing, optimización, visualización, entre otros.

¿Por qué la necesidad de un Data Science?

Matemática y Estadística



Conocimiento y dominio de habilidades blandas



Programación y base de datos



Comunicación y visualización



¿Ejemplos de predicción?

facebook

Perfiles de tus amigos

Linked in

Habilidades laborales

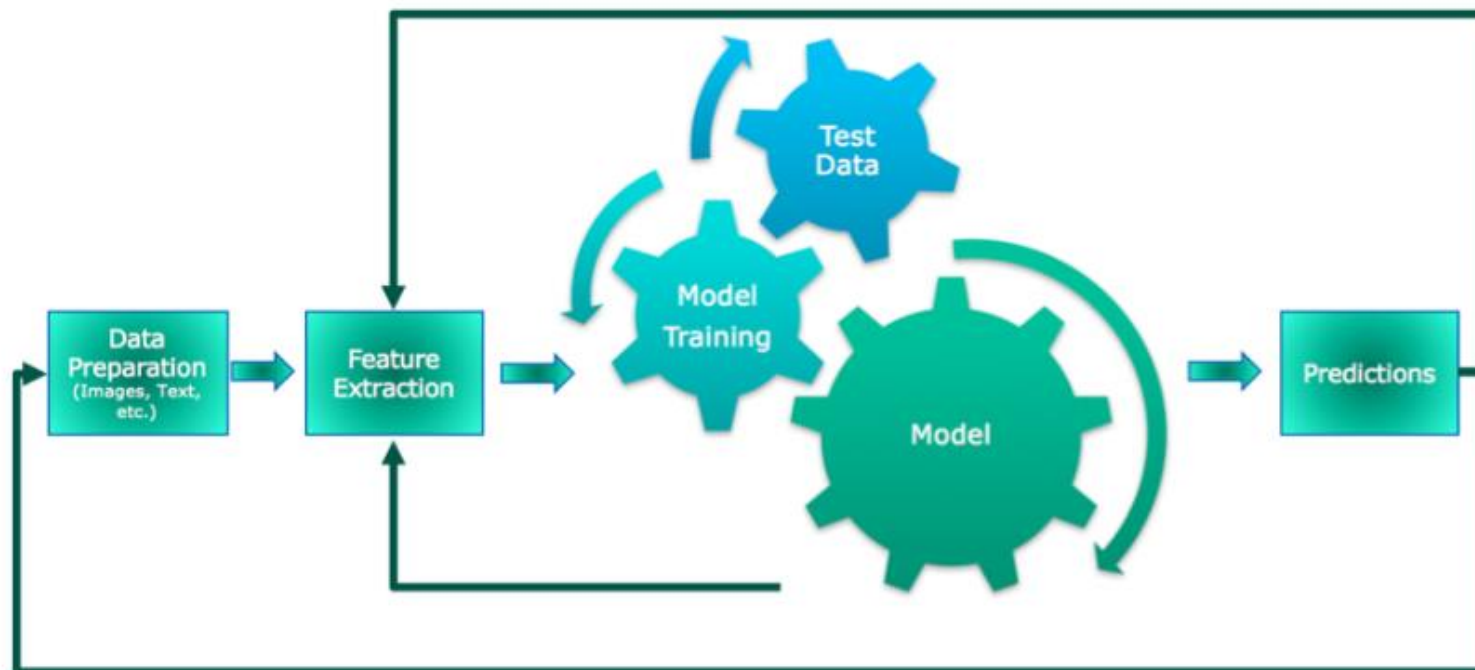
NETFLIX

Las películas que te gustan



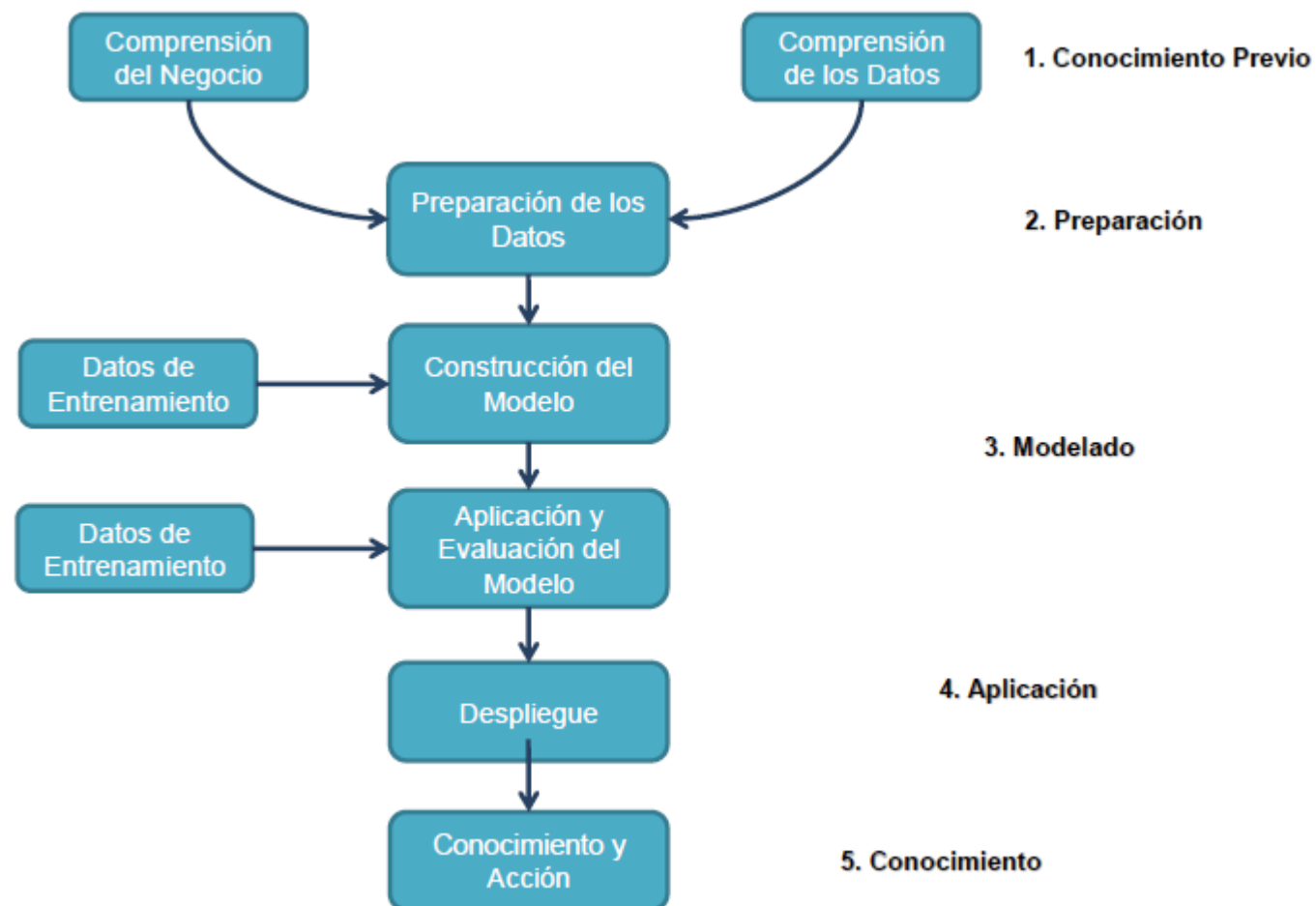
Reconocimiento facial

Introducción al Machine Learning



Machine learning es la técnica que permite entrenar a las computadoras para que aprendan a descubrir por si solos los patrones que se encuentran dentro de los datos

Fases de un problema de Machine Learning



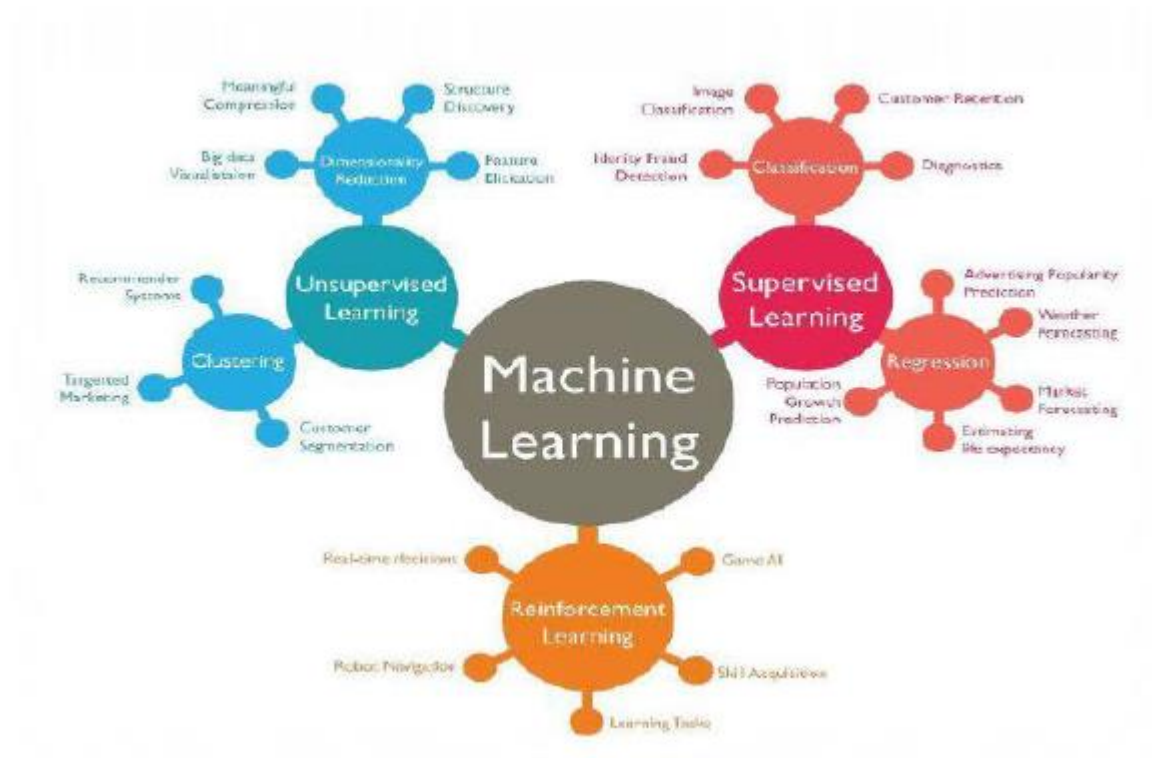
Definiciones básicas

- **Conjunto de Datos (Data Set):** El total del conjunto de datos sobre los que queremos desarrollar un algoritmo de Machine Learning con el fin de obtener un modelo que lo represente lo mejor posible. Contendrá variables independientes y dependientes.
- **Variables Independientes (Features), (VI):** Aquellas columnas del Data Set que serán usadas por el algoritmo para generar un modelo que prediga lo mejor posible las variables dependientes.

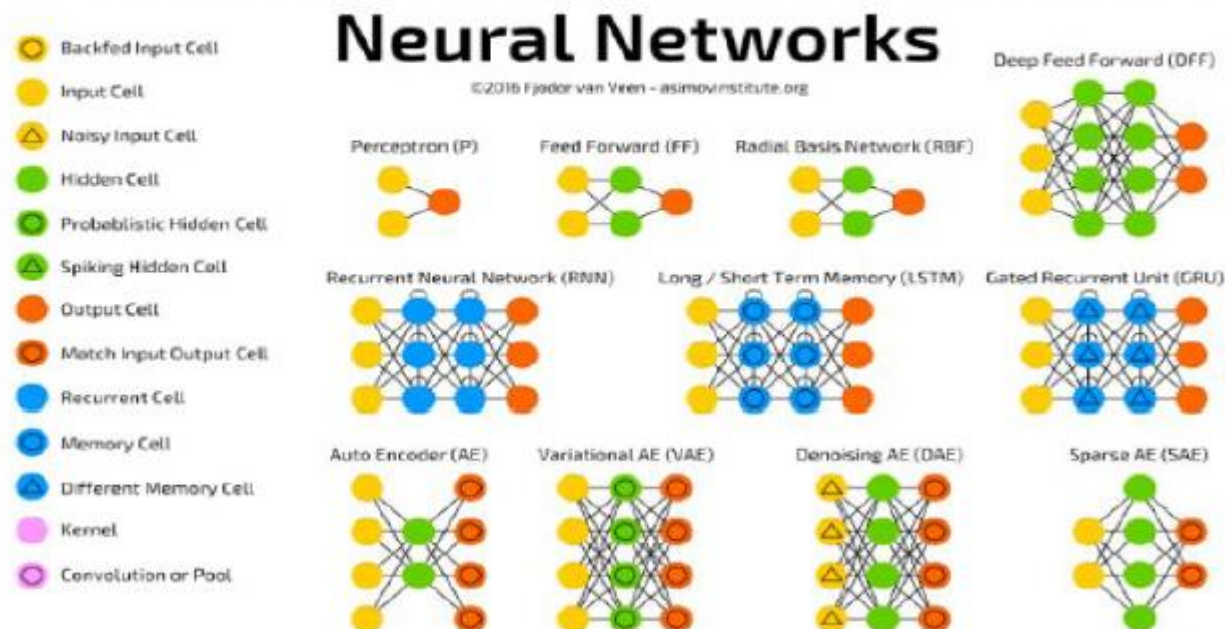
Definiciones básicas

- **Variables dependientes (Labels,Target), (VD):** Columna del data set que responde a una correlación de VI y que debe ser predicha por el futuro modelo
- **Conjunto de Datos de Entrenamiento (Training Set):** Subconjunto del Data Set que será utilizado para entrenar el modelo que se pretende generar.
- **Conjunto de Datos de Test (Test Set):** Subconjunto del data set que se le pasará al modelo una vez haya sido entrenado para comprobar, mediante el uso de diferentes métricas, sus indicadores más importantes de calidad.

Ramas de Machine Learning



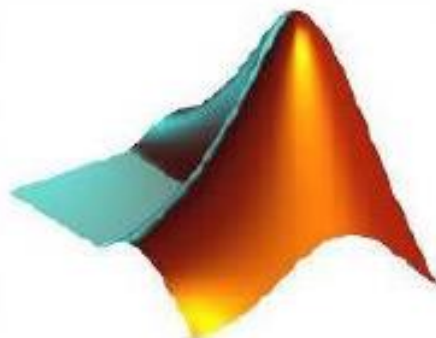
DEEP LEARNING (APRENDIZAJE PROFUNDO)



Panorama tecnológico: Software Machine Learning



SPSS Modeler

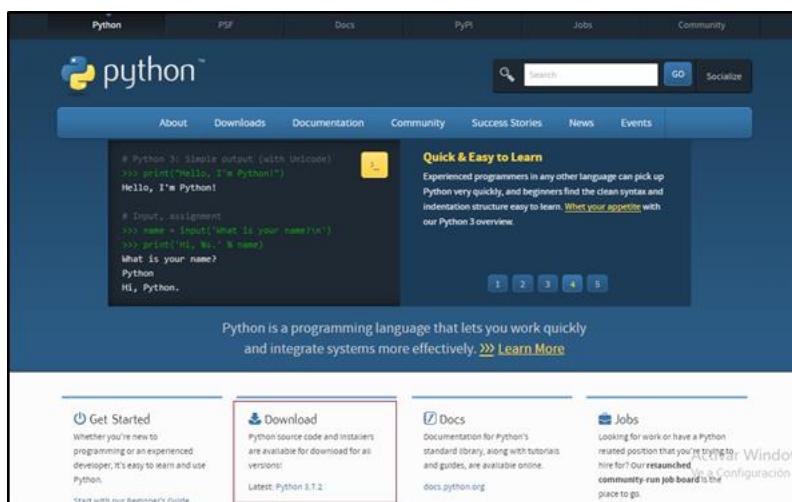




Python es un entorno de trabajo útil para cálculos científicos, trabajo con base de datos, operaciones financieras y gestión de páginas web

Instalación Python

<https://www.python.org/>





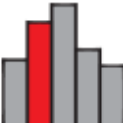





Instalación Anaconda

<https://www.anaconda.com/download/>



Manejo de entornos de trabajo

 <p>JupyterLab 0.35.3</p> <p>An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.</p> <p>Launch</p>	 <p>Notebook 5.7.4</p> <p>Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.</p> <p>Launch</p>	 <p>Qt Console 4.4.3</p> <p>PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips, and more.</p> <p>Launch</p>	 <p>Spyder 3.3.2</p> <p>Scientific PYTHON Development EnviRonment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection features</p> <p>Launch</p>
 <p>Glueviz 0.13.3</p> <p>Multidimensional data visualization across files. Explore relationships within and among related datasets.</p>	 <p>Orange 3 3.17.0</p> <p>Component based data mining framework. Data visualization and data analysis for novice and expert. Interactive workflows</p>	 <p>RStudio 1.1.456</p> <p>A set of integrated tools designed to help you be more productive with R. Includes R essentials and notebooks.</p>	 <p>VS Code 1.30.2</p> <p>Streamlined code editor with support for development operations like debugging, task running and version control.</p>

Estudio de caso

Caso aplicativo Machine Learning:

REINGRESO HOSPITALARIO DE PACIENTES DIAGNOSTICADOS CON DIABETES

Se desea analizar la propensión que un paciente ya antes atendido en el centro de salud vuelva a ingresar a dicho centro para solicitar algún servicio ambulatorio, medicina, atención de emergencia, etc.

Importante:

Se dice que un paciente es readmitido si vuelve al centro de salud por lo menos una vez después de haber sido dado de alta.

¡Gracias!

César Quezada

Especialista en Analítica de Data & BI en DMC Perú



Me puedes encontrar como:



[cesar.quezada@dmc.pe]



[www.linkedin.com/in/cquezadab]



[cesar.quezada19]



[Data Scientist's Journal]