

Machine Learning:

“Una Máquina de Oportunidades”



WIQONN Technologies

“Datos e ingeniería para Humanos”

Ing. Wayner Barrios Bustamante

www.wiqonn.com



[@wiqonn](https://twitter.com/wiqonn)



contact@wiqonn.com



Acerca de mí... **Wayner Barrios Bustamante**

- ▶ Ingeniero de Sistemas, Universidad del Norte 1991
- ▶ Especialista en Redes de Computadoras, Universidad del Norte 2002
- ▶ Experiencia: 27 años en el sector de la Tecnología de la Información y las Telecomunicaciones (TIC)
- ▶ Historia Laboral: Alcaldía Distrital de Barranquilla, Dirección de TIC de la Universidad del Norte, Unidad de Outsourcing en IBM, Gerente de Sistemas y Técnico en Metrotel.
- ▶ Miembro del Comité Directivo de Metrotel.
- ▶ Presidente de la Junta de Directiva de Optecom S.A.S.
- ▶ Docente Universitario, UN – UAC – USB
- ▶ Gerente General y co-fundador de WIQONN Technologies



WIQONN Tech: Una empresa de servicios TICs



Análisis de Datos

Nuestro objetivo es descubrir información útil, informar conclusiones y apoyar la toma de decisiones



Inteligencia Artificial

Agentes inteligentes que toman acciones que maximizan sus posibilidades de lograr su objetivo con éxito. Con aplicaciones en las áreas de:
Machine Learning Computer Vision Natural Language Processing



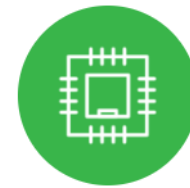
Apps Web & Móviles

Nos enfocamos en crear soluciones adaptadas a los requisitos de nuestros clientes que utilizan tecnología de punta.



Networking

Apoyamos en proyectos de infraestructura en Redes de Acceso, de Sincronismo, IoT, Servicios de Red, IaaS, Cyberseguridad.



Proyectos de Electrónica

Soluciones de sistemas de sensores embebidos y remotos para múltiples aplicaciones.



A signpost with two directional signs. The top sign is green and points right, labeled 'SOLUTION'. The bottom sign is red and points left, labeled 'PROBLEM'.

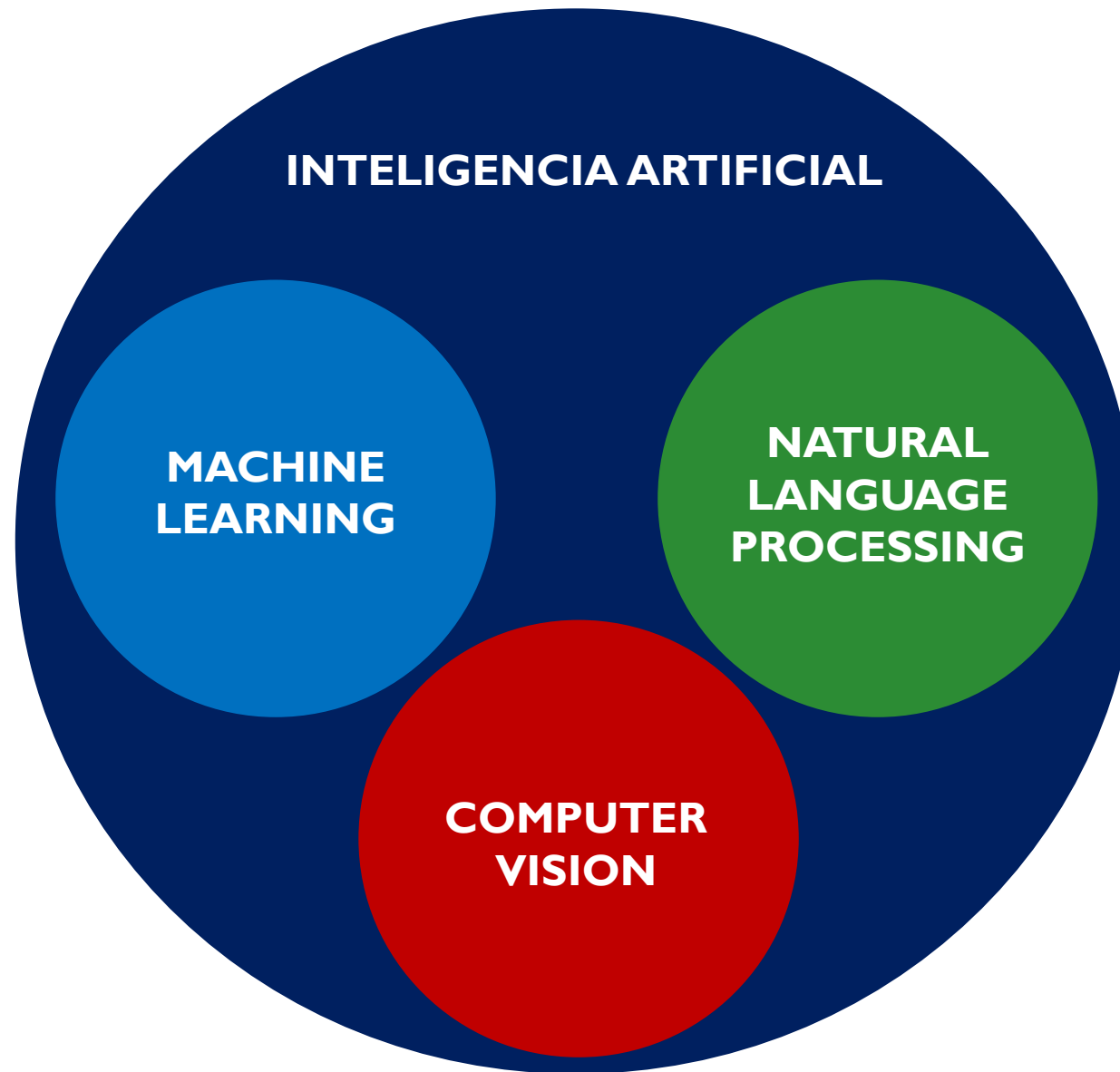


PROBLEM

The word cloud features the following terms:

- ageing population**
- communication**
- education for all**
- fear**
- identity**
- literacy**
- population distribution**
- emancipation of women globally**
- Religion**
- atomic engineering**
- biotech society**
- connected communities**
- energy infrastructure**
- preventative care**
- radio frequency identification (RFID)**
- smart dust**
- wearable computing**
- Convergence**
- Peak Oil**
- disposable quality goods**
- ecological footprint**
- endangered species**
- energy use**
- travel**
- urbanization**
- waste**
- water**
- Climate Change**
- sustainability**
- airport shopping**
- containerized cargo**
- consumer debt**
- democratization of luxury**
- digital currency**
- global trade**
- migration**
- outsourcing**
- wealth gap**
- Cost of Health Care**
- asianization**
- compensation culture**
- ethical investment**
- global governance**
- food legislation**
- pensions**
- strife**
- surveillance society**
- trading blocs**
- the vote**





INTELIGENCIA ARTIFICIAL

Un programa que puede sentir,
razonar, actuar y adaptarse.

MACHINE LEARNING

Algoritmos cuyo rendimiento
mejora a medida que se exponen
a más datos a lo largo del tiempo

DEEP LEARNING

Subconjunto de Machine
Learning en el que las
Redes Neuronales
Multicapas aprenden de
una gran cantidad de
datos



“La **INTELIGENCIA ARTIFICIAL**
Es el intento de que un dispositivo o una
aplicación sea tan o más inteligente que un
Humano”



“**MACHINE LEARNING** es una serie de
algoritmos que hacen que tu dispositivo o
aplicación sea artificialmente inteligente”



“ ... el 90% de los datos existentes en el mundo se han creado en los últimos 2 años a un ritmo de $2,5 \times 10^{18}$ bytes diarios...”

The New York Times



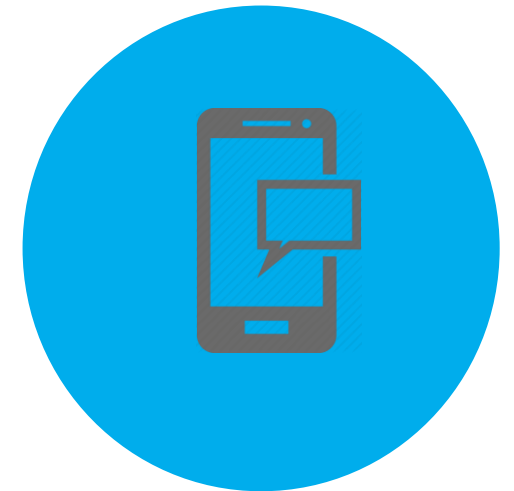
**4.0 Billones
personas usan
Internet**



**3.2 Billones
personas usan
Redes Sociales**



**5.1 Billones
personas usan
Teléfono Móvil**



**3.0 Billones
personas usan
Redes Sociales a
través de su Móvil**







“El conjunto de técnicas y tecnologías para el **tratamiento de datos**, en entornos de gran **VOLUMEN, VARIEDAD** de orígenes y en los que la **VELOCIDAD** de respuesta es crítica” (1)

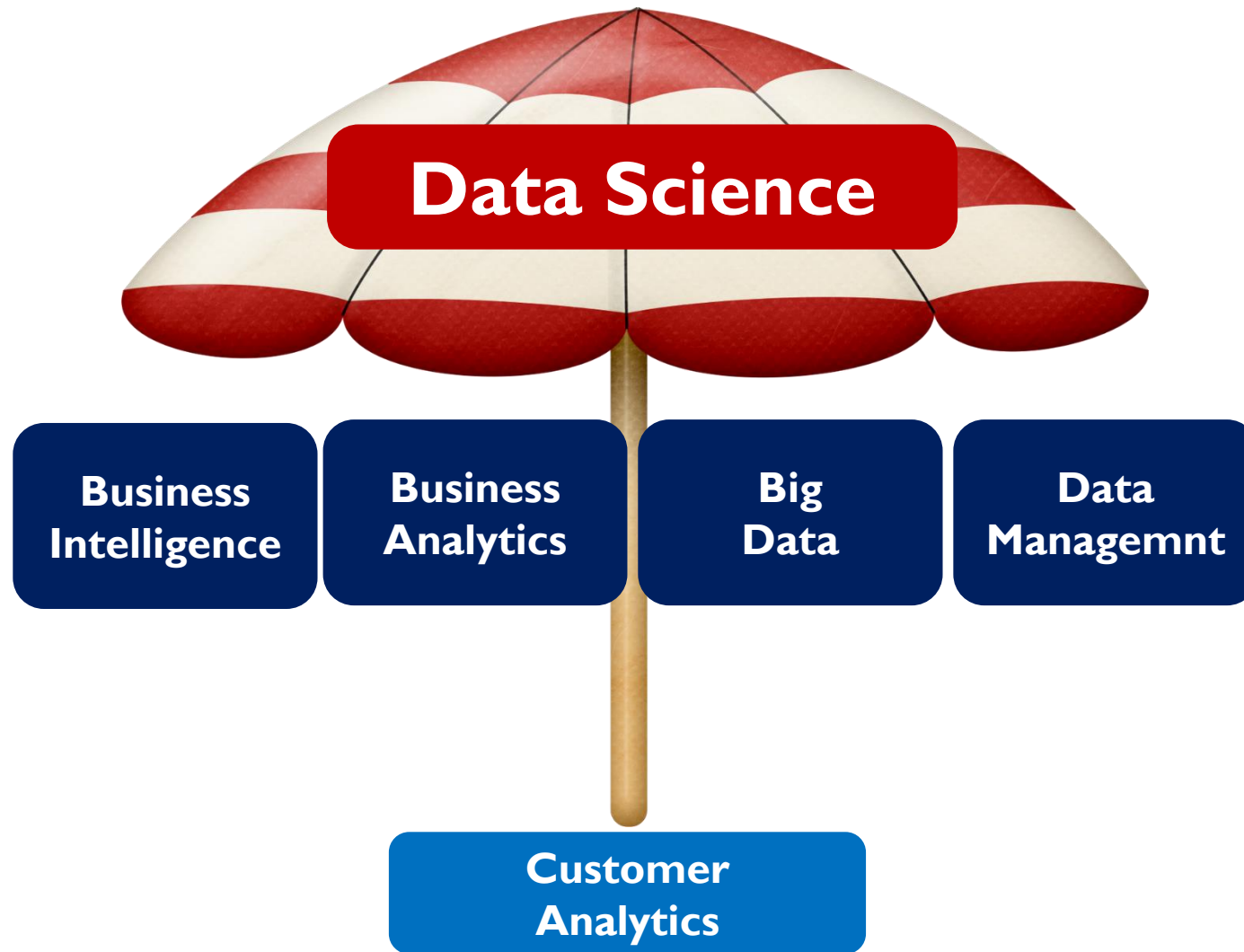
“MODELO DE LAS 3Vs DEL BIG DATA”

(1) Doug Laney, Gartner Group - 2001



Big Data en Colombia: CONPES 3920

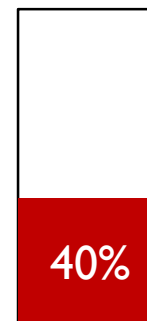
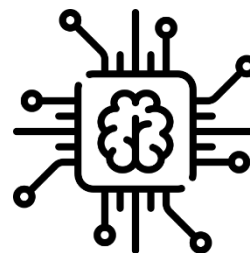




Machine Learning: Identificando las oportunidades

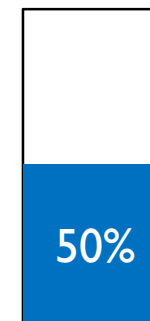
ZooApp

Mi mascota



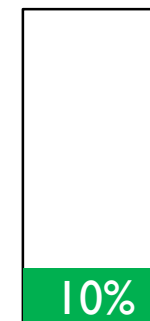
40%

Gato



50%

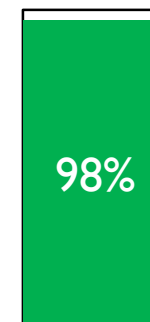
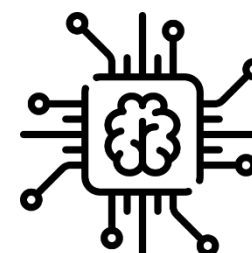
Caballo



10%

Perro

Score 5%

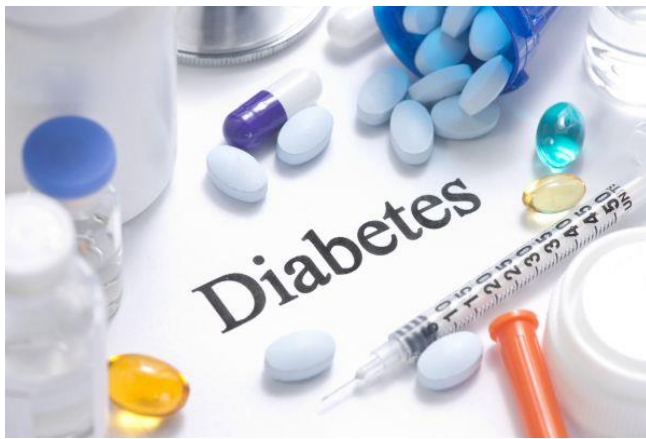


98%

Perro

Score 98%





Determinar si un **PACIENTE** debe ser diagnosticado con **DIABETES** a partir de los datos obtenidos de sus Historias Clínicas.

ATRIBUTOS DE LA INFORMACIÓN

- # Embarazos
- Glucosa
- Presión Arterial
- Pliegue Cutáneo
- Insulina
- BMI
- Pedigree Diabetes
- Edad
- Diagnóstico

CONJUNTO DE DATOS

2, 148, 72, 35, 0, 33.6, 0.627, 50, 1
1, 85, 66, 29, 0, 26.6, 0.351, 31, 0
3, 183, 64, 0, 0, 23.3, 0.672, 32, 1
1, 89, 66, 23, 94, 28.1, 0.167, 21, 0
0, 137, 40, 35, 168, 43.1, 2.288, 33, 1
5, 116, 74, 0, 0, 25.6, 0.201, 30, 0
3, 78, 50, 32, 88, 31, 0.248, 26, 1
1, 115, 0, 0, 0, 35.3, 0.134, 29, 0
2, 197, 70, 45, 543, 30.5, 0.158, 53, 1
2, 125, 96, 0, 0, 0, 0.232, 54, 1



760 instancias
9 atributos



Determina la edad de una Mujer



10 años



21 años



38,5 años

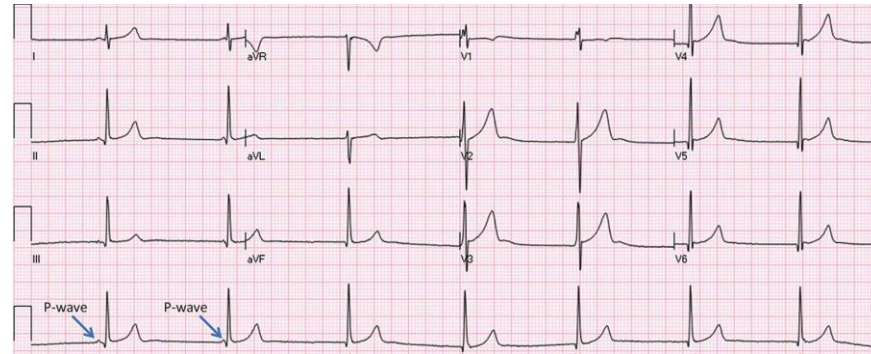


47,2 años

Desarrolla un programa que te permita determinar la edad a partir del rostro de una mujer

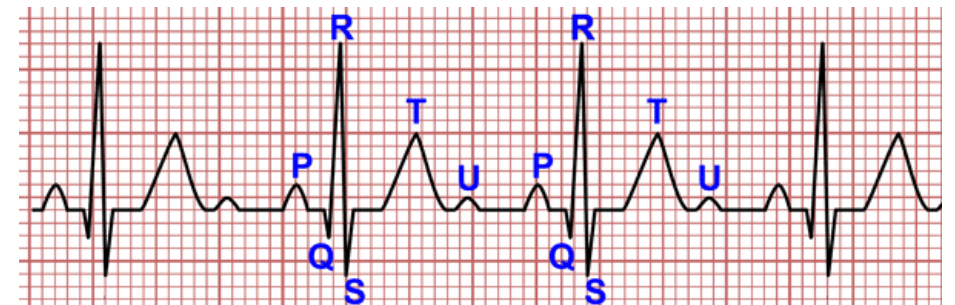


Desarrolle una herramienta de tecnología de la información que brinde apoyo en el análisis de ECG para diagnosticar enfermedades cardíacas de manera oportuna ⁽¹⁾



Un **ECG** detecta:

- Las irregularidades en el ritmo cardíaco (arritmias)
- Defectos del corazón
- Problemas con las válvulas de su corazón
- Arterias bloqueadas o estrechas de su corazón (enfermedad de la arteria coronaria)
- Un ataque al corazón, en situaciones de emergencia
- Un ataque al corazón previo



⁽¹⁾ A Real-Time QRS Detection Algorithm – 1985, Jiapu Pan & Willis J. Tompkins





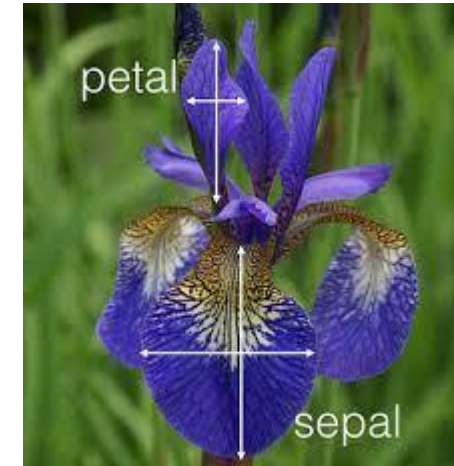
Determinar el tipo de planta de flor de Iris a partir de las dimensiones de sus sépalos y pétalos.

ATRIBUTOS DE LA INFORMACIÓN

1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm
5. class:
 - Iris Setosa
 - Iris Versicolour
 - Iris Virginica

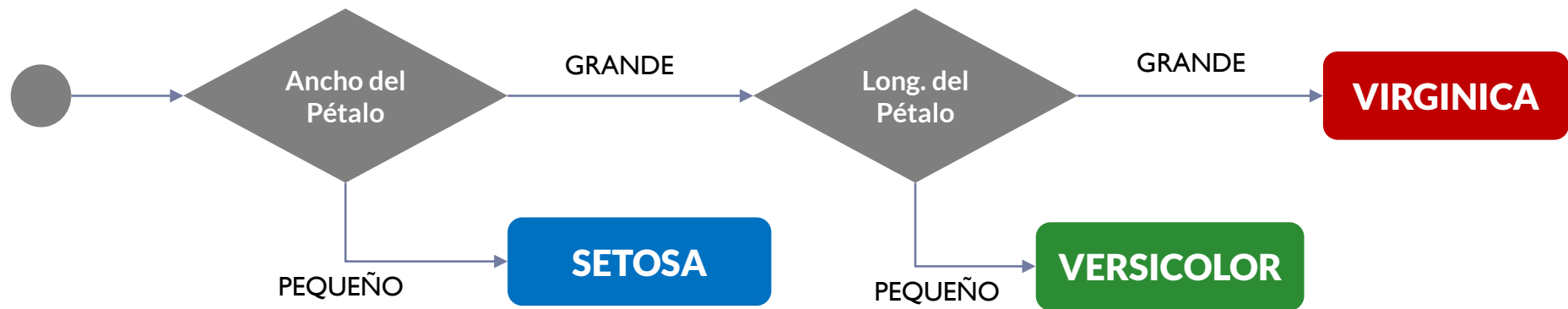
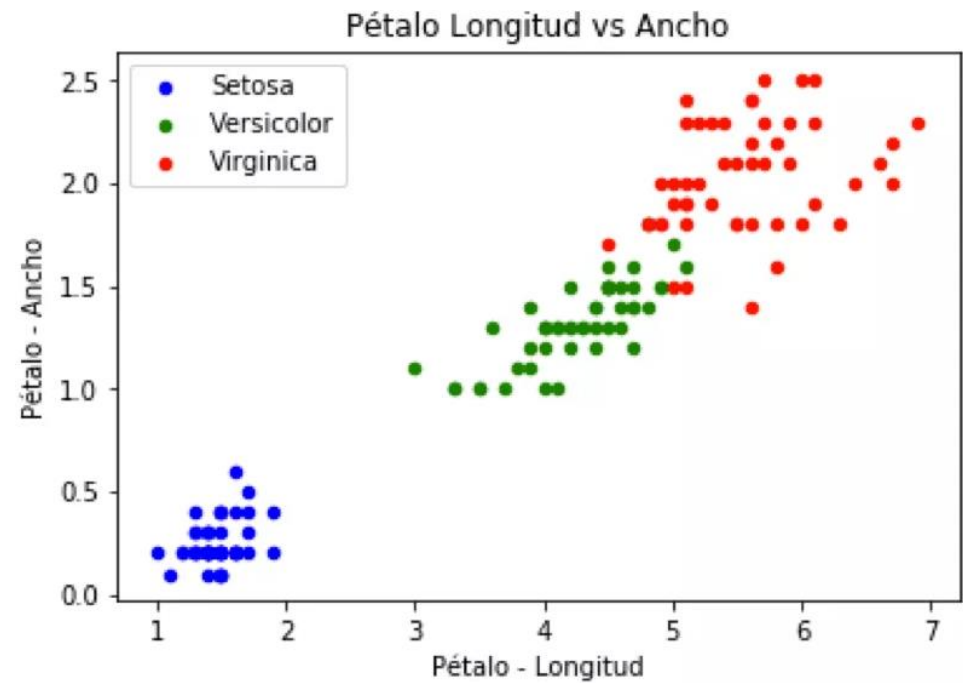
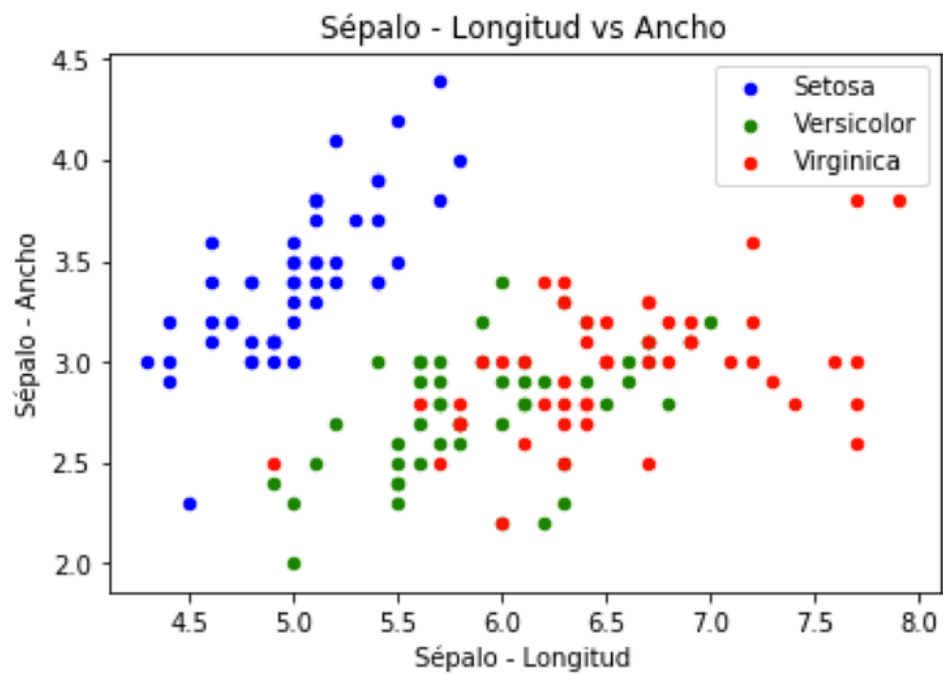
CONJUNTO DE DATOS

4.6, 3.2, 1.4, 0.2, Iris-setosa
5.3, 3.7, 1.5, 0.2, Iris-setosa
5.0, 3.3, 1.4, 0.2, Iris-setosa
7.0, 3.2, 4.7, 1.4, Iris-versicolor
6.4, 3.2, 4.5, 1.5, Iris-versicolor
5.7, 2.8, 4.1, 1.3, Iris-versicolor
6.3, 3.3, 6.0, 2.5, Iris-virginica
5.8, 2.7, 5.1, 1.9, Iris-virginica
7.1, 3.0, 5.9, 2.1, Iris-virginica



150 instancias
5 atributos





A partir de las características de un artículo de Inteligencia Artificial intenta predecir, cuántas veces será compartido en Redes Sociales.

Title: Titulo del Artículo

URL: Ruta al artículo

Word count: La cantidad de palabras del artículo,

of Links: Los enlaces externos que contiene,

of comments: Cantidad de comentarios,

Images video: Suma de imágenes (o videos),

Elapsed days: La cantidad de días transcurridos (al momento de crear el artículo)

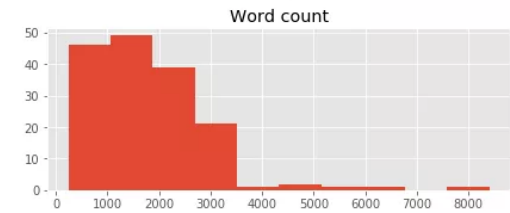
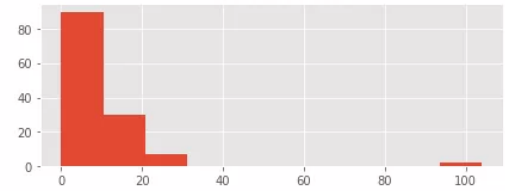
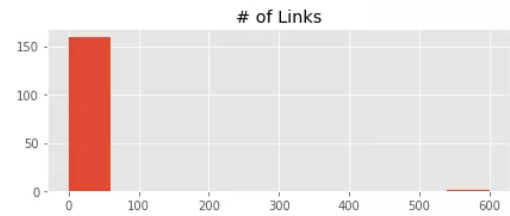
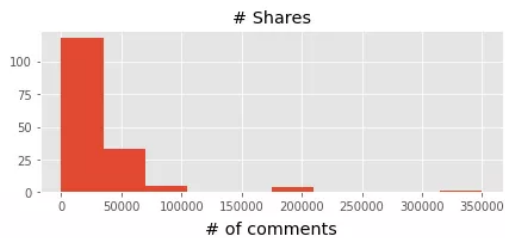
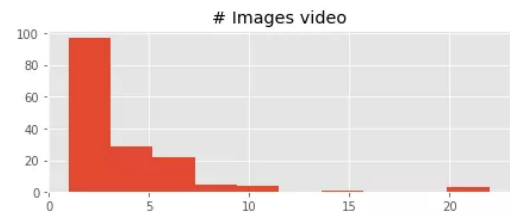
Shares: Cantidad de veces que se compartió el artículo.

161 instancias
8 atributos

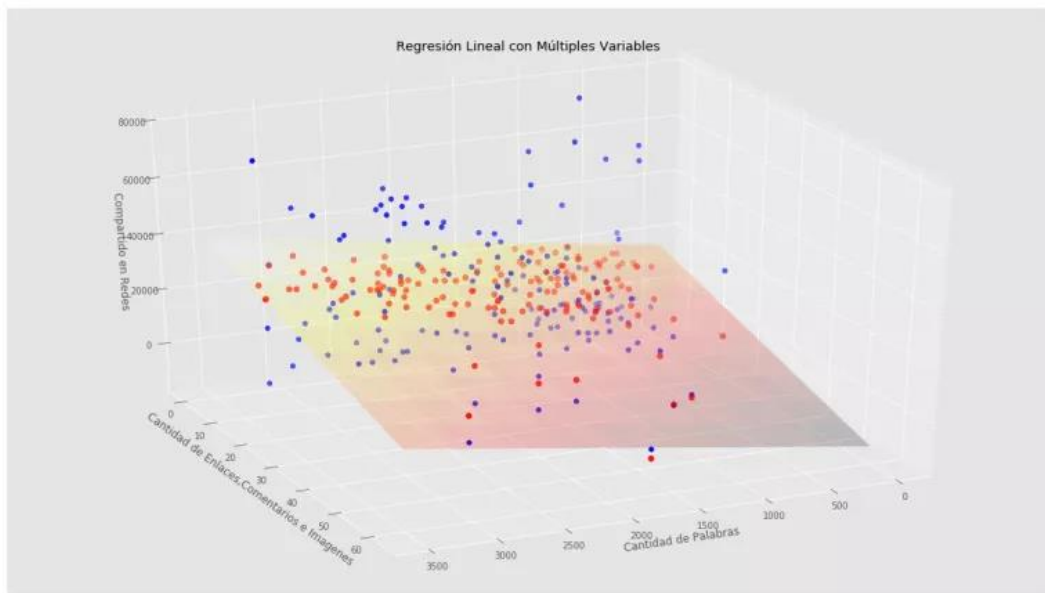
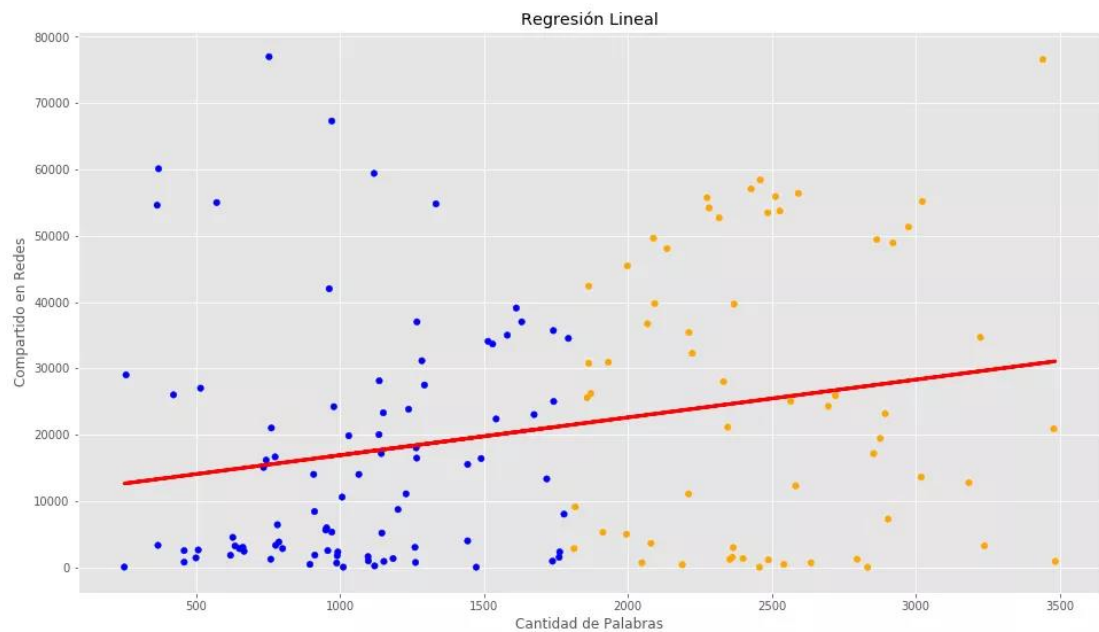
| | Title | url | Word count | # of Links | # of comments | # Images video | Elapsed days | # Shares |
|---|---|---|------------|------------|---------------|----------------|--------------|----------|
| 0 | What is Machine Learning and how do we use it ... | https://blog.signals.network/what-is-machine-l... | 1888 | 1 | 2.0 | 2 | 34 | 200000 |
| 1 | 10 Companies Using Machine Learning in Cool Ways | NaN | 1742 | 9 | NaN | 9 | 5 | 25000 |
| 2 | How Artificial Intelligence Is Revolutionizing... | NaN | 962 | 6 | 0.0 | 1 | 10 | 42000 |
| 3 | Dbrain and the Blockchain of Artificial Intell... | NaN | 1221 | 3 | NaN | 2 | 68 | 200000 |
| 4 | Nasa finds entire solar system filled with eig... | NaN | 2039 | 1 | 104.0 | 4 | 131 | 200000 |



| | Word count | # of Links | # of comments | # Images video | Elapsed days | # Shares |
|--------------|-------------|------------|---------------|----------------|--------------|---------------|
| count | 161.000000 | 161.000000 | 129.000000 | 161.000000 | 161.000000 | 161.000000 |
| mean | 1808.260870 | 9.739130 | 8.782946 | 3.670807 | 98.124224 | 27948.347826 |
| std | 1141.919385 | 47.271625 | 13.142822 | 3.418290 | 114.337535 | 43408.006839 |
| min | 250.000000 | 0.000000 | 0.000000 | 1.000000 | 1.000000 | 0.000000 |
| 25% | 990.000000 | 3.000000 | 2.000000 | 1.000000 | 31.000000 | 2800.000000 |
| 50% | 1674.000000 | 5.000000 | 6.000000 | 3.000000 | 62.000000 | 16458.000000 |
| 75% | 2369.000000 | 7.000000 | 12.000000 | 5.000000 | 124.000000 | 35691.000000 |
| max | 8401.000000 | 600.000000 | 104.000000 | 22.000000 | 1002.000000 | 350000.000000 |



REGRESIONES LINEALES



Terminología de Machine Learning

Los sistemas de **Machine Learning** aprenden cómo combinar entradas para producir predicciones útiles sobre datos nunca antes vistos.

Dataset

Conjunto de datos del modelo, que podemos expresar como $\{x_1, x_2, x_3, \dots, x_N, y_1, y_2, \dots\}$

Atributo

Es una variable de entrada, y corresponde a un campo (columna) del conjunto de datos. Un proyecto de aprendizaje automático simple podría usar un solo atributo, mientras que otro más sofisticado podría usar millones de atributos, especificados como $\{x_1, x_2, x_3, \dots, x_N\}$

Etiqueta

Es el valor que estamos prediciendo, expresado como $\{y_1, y_2, \dots\}$. La etiqueta podría ser el precio futuro del café, el tipo de animal que se muestra en una imagen, el significado de un clip de audio o simplemente cualquier cosa.



Modelo

Un modelo define la relación entre los atributos y la etiqueta. Por ejemplo, un modelo para determinar el tipo de planta podría asociar de manera muy definida determinados atributos con “la longitud y el ancho” del pétalo o sépalo.

Definir como desde X podemos llegar a Y y encontrar la siguiente relación $Y \approx F(X, W)$

Identificamos las siguiente fases de un Modelo:

Entrenamiento significa crear o aprender el modelo. Es decir, le muestras ejemplos etiquetados al modelo y permites que este aprenda gradualmente las relaciones entre los atributos y la etiqueta:

$(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), (x_5, y_5), \dots$ y encontrar W ideal.

Inferencia significa aplicar el modelo entrenado a ejemplos sin etiqueta. Es decir, usas el modelo entrenado para realizar predicciones útiles Y' . Es decir $Y' = F(X, W) = w_N x_N + w_{N-1} x_{N-1} + w_{N-2} x_{N-2} + \dots w_1 x_1 + w_0$

Parámetros de entrenamiento

Variables que permiten controlar la calidad del modelo resultante. $\{w_0, w_1, w_2, \dots, w_N\}$



Pérdida (L2)

También llamada error cuadrático (EC) se define como el cuadrado de la diferencia entre la predicción y la etiqueta. n modelo define la relación entre los atributos y la etiqueta.

$(F(x_i, W) - y_i)^2$ de tal forma que $F(x_i, W)$ es llamada y'_i la predicción

Costo de la función L(W)

$$\frac{1}{2n} \sum_{i=1}^n (F(x_i, W) - y_i)^2$$

Objetivo

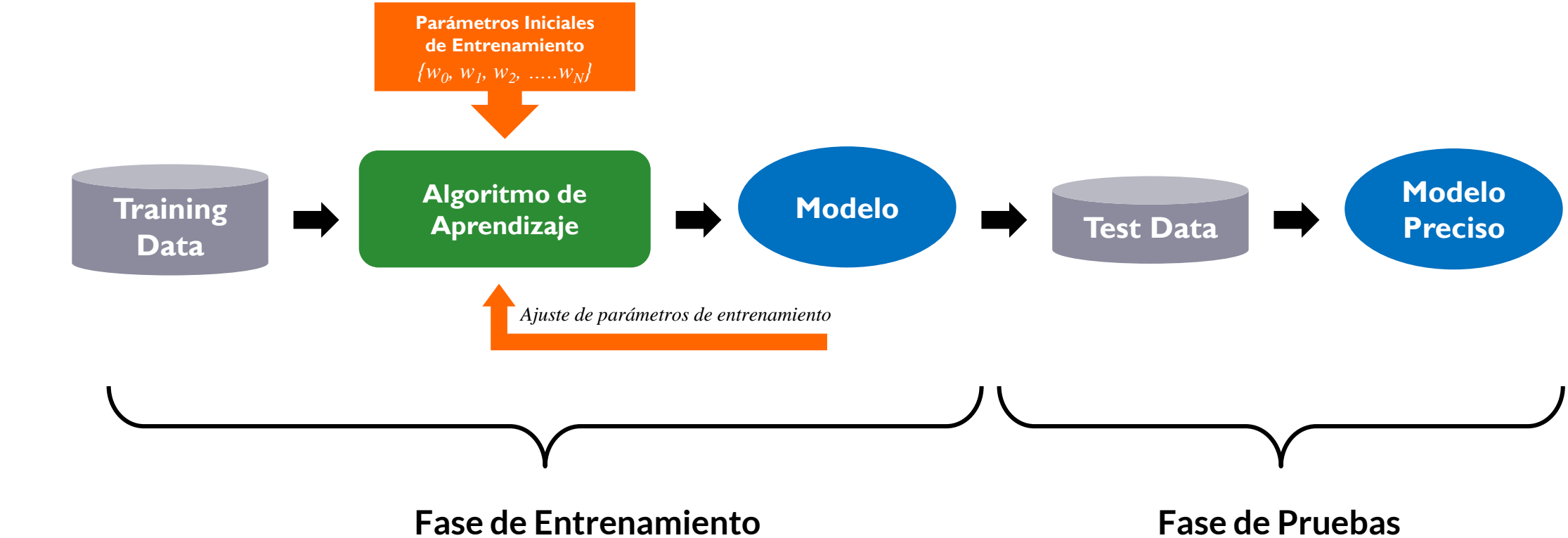
$$\text{Min}_w L(W)$$


Gradiente Descendente

$w_0 = w_0 - \alpha \frac{dL(W)}{dw_0}$ y $w_1 = w_1 - \alpha \frac{dL(W)}{dw_1}$, donde α es la tasa de aprendizaje



Secuencia típica de un Aprendizaje



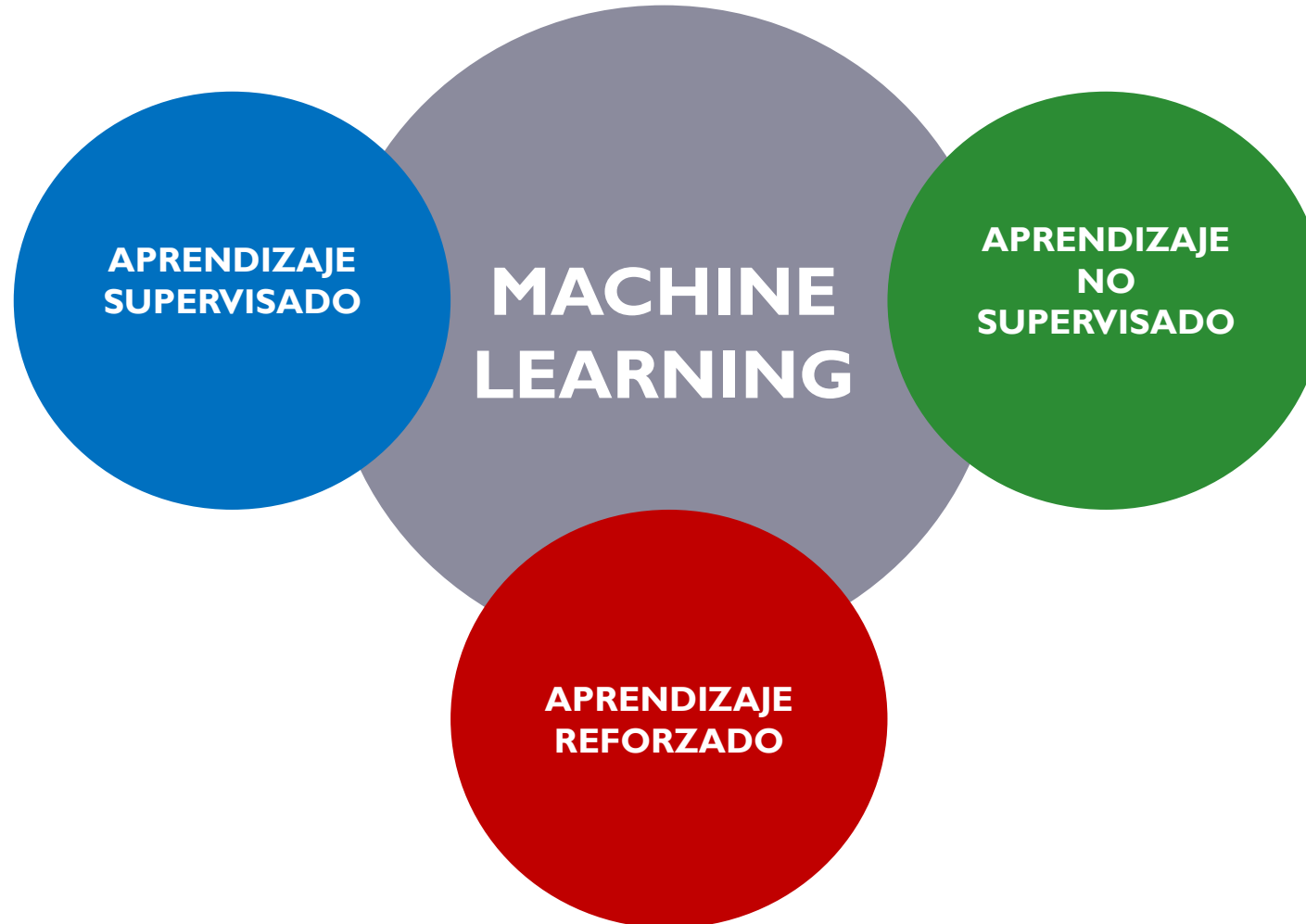
Dataset

- Training_data 80%
- Test_data 20%

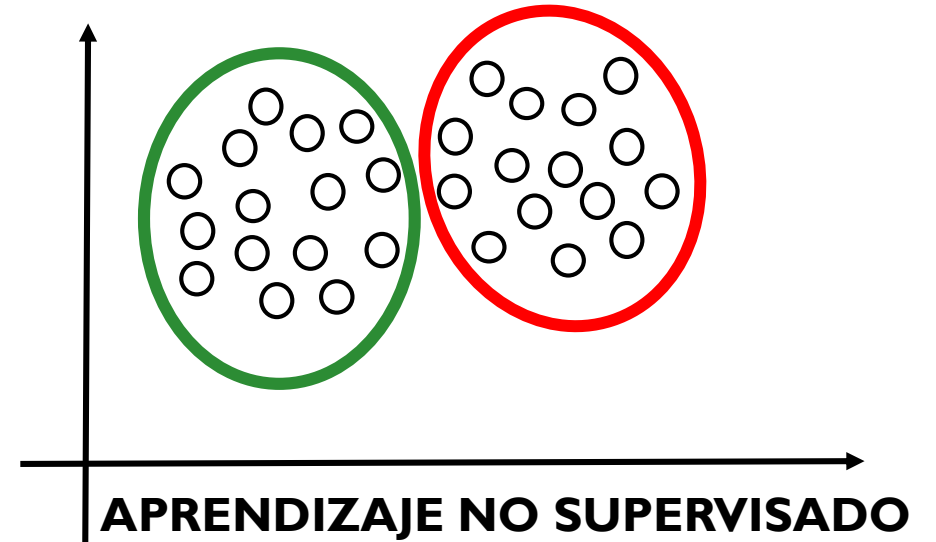
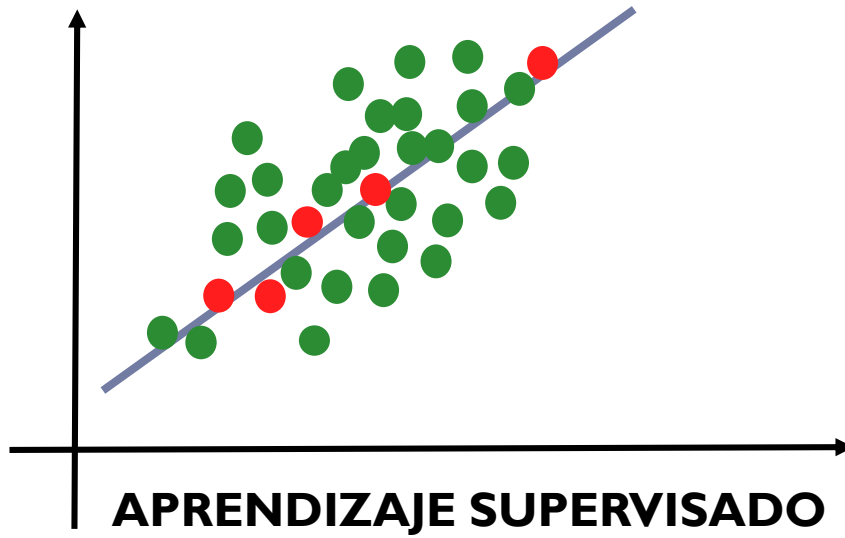


Tipos de Machine Learning

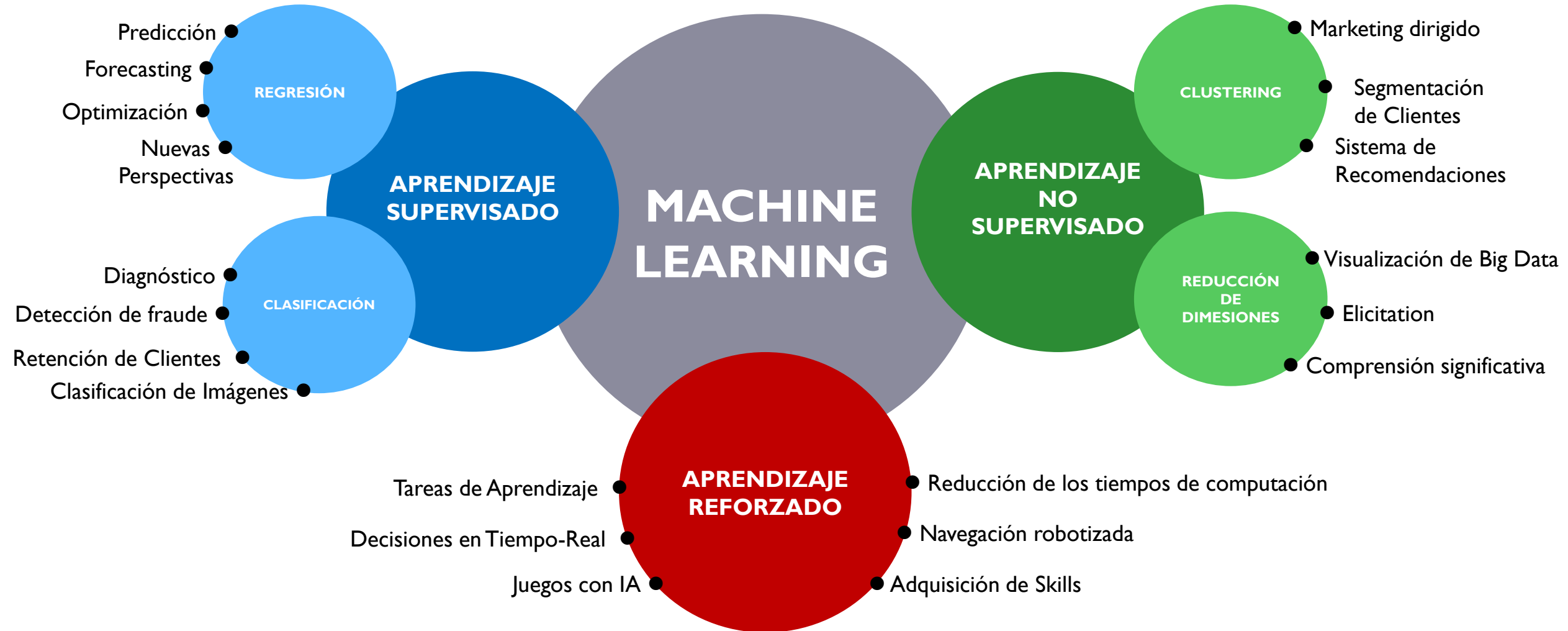
Tipos de Aprendizajes Automáticos



Tipos de Aprendizajes Automáticos



Tipos de Aprendizajes Automáticos



¿Qué algoritmo de Aprendizaje Automático debo “USAR”?

La respuesta a la pregunta varía según muchos factores, entre ellos:

- ▶ El tamaño, la calidad y la naturaleza de los datos
- ▶ El tiempo computacional disponible
- ▶ La urgencia de la tarea
- ▶ ¿Qué quieres hacer con los datos?

Incluso un *científico de datos* experimentado no puede decir qué algoritmo funcionará mejor antes de probarlos.



Consideraciones al elegir un Algoritmo

- ▶ **Precisión:** No siempre es necesario obtener la respuesta más precisa posible. A veces, una aproximación ya es útil, según para lo que se desee usar.
- ▶ **Tiempo de entrenamiento:** La cantidad de minutos u horas necesarios para entrenar un modelo varía mucho según el algoritmo.
- ▶ **Linealidad:** Los algoritmos de clasificación lineal suponen que las clases pueden estar separadas mediante una línea recta (o su análogo de mayores dimensiones).
- ▶ **Cantidad de parámetros:** Los parámetros son los botones que un científico de datos activa al configurar un algoritmo. Son números que afectan al comportamiento del algoritmo, como la tolerancia a errores o la cantidad de iteraciones, o bien opciones de variantes de comportamiento del algoritmo.
- ▶ **Cantidad de características:** Para ciertos tipos de datos, la cantidad de características puede ser muy grande en comparación con la cantidad de puntos de datos.



Algoritmos según el Tipo de Aprendizaje

| Tipo de Aprendizaje | Algoritmo |
|----------------------------|--|
| APRENDIZAJE SUPERVISADO | Linear Regression Logistic Regression k-Nearest Neighbors Support Vector Machines (SVM) Bayesian Classifiers Decision Trees and Random Forest Neural Networks / Deep Learning |
| APRENDIZAJE NO SUPERVISADO | K-Means Principal Component Analysis (PCA) Anomaly Detection |



¿Qué quiero “HACER”?

- ▶ Si lo que necesitas es agrupar o establecer categorías o grupos, puedes utilizar algoritmos de **clasificación**
- ▶ Si quieres predecir un importe, cantidad, ventas, etc... deberás usar algoritmos de **regresión**
- ▶ Si quieres detectar datos incongruentes, incorrectos, fallos, etc... usa **detección de anomalías**
- ▶ Si lo que tienes es un montón de datos, aparentemente inconexos y sin ninguna relación entre ellos, pues utilizar “**clustering**”, de forma que se establezcan estructuras y grupos con patrones comunes



Áreas de Aplicación de Machine Learning

Áreas de aplicación de Machine Learning

- ▶ Seguridad Física e Informática
- ▶ Atención a Clientes
- ▶ Segmentación de Audiencia
- ▶ Transporte Masivo
- ▶ Salud
- ▶ Juegos
- ▶ Economía y Finanzas
- ▶ Conducción autónoma
- ▶ Motores de recomendación



Plataformas de software de Machine Learning

Plataformas de Machine Learning

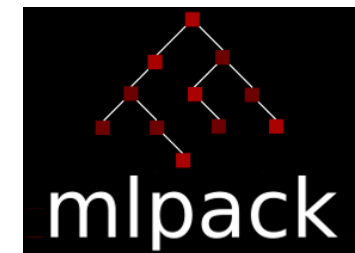


Principales Frameworks para Machine Learning



Caffe2

PYTORCH



Ejemplo Python para aplicación Planta Flor de Iris

```
2 #----- DECLARACION DE LIBRERIAS
3 import numpy as np
4 import pandas as pd
5 import matplotlib.pyplot as plt
6 #
7 #----- CARGUE DEL DATASET - IMPORTAR DATOS DE ENTRADA EN FORMATO CSV
8 iris = pd.read_csv("file-iris.csv")
9 #
10 #----- GRAFICAR Pétalo - Longitud vs Ancho
11 fig = iris[iris.Species == 'Iris-setosa'].plot(kind='scatter', x='PetalLengthCm', y='PetalWidthCm', color='blue', label='Setosa')
12 iris[iris.Species == 'Iris-versicolor'].plot(kind='scatter', x='PetalLengthCm', y='PetalWidthCm', color='green', label='Versicolor', ax=fig)
13 iris[iris.Species == 'Iris-virginica'].plot(kind='scatter', x='PetalLengthCm', y='PetalWidthCm', color='red', label='Virginica', ax=fig)
14 fig.set_xlabel('Pétalo - Longitud')
15 fig.set_ylabel('Pétalo - Ancho')
16 fig.set_title('Pétalo Longitud vs Ancho')
17 plt.show()
18 #
19 #----- APLICACIÓN DE ALGORITMOS DE MACHINE LEARNING
20 from sklearn.model_selection import train_test_split
21 from sklearn.tree import DecisionTreeClassifier
22 #
23 ...
24 #----- SEPARAR EL DATASET EN datos de "train" en entrenamiento y "test" para probar los algoritmos
25 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
26 print('Son {} datos para entrenamiento y {} datos para prueba'.format(X_train.shape[0], X_test.shape[0]))
27 #
28 ....
29 #----- APLICACIÓN DEL ALGORITMO DE CLASIFICADOR DE ARBOL DE DECISIÓN
30 algoritmo = DecisionTreeClassifier()
31 algoritmo.fit(X_train, y_train)
32 Y_pred = algoritmo.predict(X_test)
33 print('Precisión Árboles de Decisión Clasificación: {}'.format(algoritmo.score(X_train, y_train)))
34 |
```





-- FIN DE LA PRESENTACIÓN --

¿Inquietudes y/o Preguntas?

E-mail: wiqonn@gmail.com

Twitter: @wiqonn

