IA Agil = aprende por si sola (skynet)

Machine Learning= Base/Nucleo de la IA

Modelos supervisados (tienen etiquetas) y los no supervisados (no necesitan y),existen también los de refuerzos

Etiquetas = Y

Parámetros = los datos a trabajar

Hyperparametros = los algoritmos

Continua (0,1,2,3,4,5…),Discreta (0/1)

SUPERVISADOS

Modelos Clasificación (Discreta),

* Modelos Clasificacion: Regresion Logistica, RandomForest,KNN,SuperVectorMachine,Naive Bayes,...
* Modelos Regresion: DecisionTree,etc

NO SUPERVISADOS

Modelos Regresion (continua),(Modelos: Reducción de dimensionalidad,PCA,Deteccion de anomalias,Asociacion,...)

Modelos Clasificacion: Accuary

R2 AJUSTADO (regresión lineal múltiple)

R2 SIMPLE

RSS,MAE,EMAX

NO SUPERVISADO

HOMOGENEIDAD:VER CUANTO SE PARECEN ENTRE UNO CON OTROS Y TAMBIEN SUS DIFERENCIAS (CLUSTER)

METODOLOGIAS DE PROYECTOS DE MACHINE LEARNING (también MINERIA DATOS, BI):SEMMA, KDD, CRISP-DM

\*Existen las bases de datos estructuradas, semi estructuradas y no estructuradas

La Ciencia de Datos, requiere equipos

multidisciplinarios. Estos incluyen:

* Ingenieros de Datos, que trabajan con

big data y tecnologías en la nube, saben

cómo construir canalizaciones de datos,

diseñar bases de datos y extraer datos

de ellas.

* Analistas de Datos, que son personas

que puedan tomar esos datos,

limpiarlos, analizarlos, ejecutar

experimentos con ellos y comunicar los

resultados.

* Ingenieros de Machine Learning,

desarrolladores de software que están

familiarizados con varias bibliotecas de

ciencia de datos y saben cómo

implementar modelos matemáticos a los

datos.

* Especialista en visualización,

centrado en el diseño para crear gráficos

e informes altamente refinados.

* Científico de Datos, quien es un

experto en el dominio del problema

—---------------------------------------------------------

ORDENES DE MAGNITUD

HAY QUE ESTANDARIZAR LOS DATOS (NORMALIZAR,ESCALAR,ETC) EJ: 1KM, 1000 MT , 10000 CM

————

Estadística descriptiva, inferencial y prescriptiva

* Descriptiva: medida de posicion central (Media, moda,mediana), medidas de posición no central ( percentiles, decintiles, cuartieles), dispersión( desviación estándar,la varianza, coeficiente de correlación, covarianza)

CORRELACIÓN NO IMPLICA QUE SEA CAUSAL

* Inferencial: ( probabilidad teoría [todos los datos tienen la misma probabilidad, es un porcentaje], probabilidad empírica [no todos los datos tienen la misma probabilidad, es margen de porcentaje como resultado],
* Prescriptiva: programación lineal, optimizar

Heterocedasticidad: No existe correlación.

EDA: Exploracion de Datos Analitico, Análisis descriptivo, análisis tipo variable, análisis de datos faltantes,análisis de datos anomalos y análisis de dispersion.

Datos faltantes: Patrones, aleatorio, que hacer? eliminar o imputar datos (constante , estadistico[por la moda,mediana,promedio,etc],o algortimico[regresion,etc]).