



UNIVERSIDAD SANTO TOMÁS

PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA

SECCIONAL TUNJA

VIGILADA MINEDUCACIÓN - SNIES 1732



Acreditación Institucional
Internacional

OTORGADA POR EL IAC CINDA ACUERDO 55 DEL 9 DE MAYO-VIGENCIA 5 AÑOS



Vigencia por seis años





UNIVERSIDAD SANTO TOMÁS
PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA
SECCIONAL TUNJA

VIGILADA MINEDUCACIÓN - SNIES 1732

Faculty: Systems engineer

Course: Deep Learning

Topic: Pasos y normas del Machine Learning

Professor: Luis Fernando Castellanos Guarín

Email: Luis.castellanosg@usantoto.edu.co

Phone: 3214582098

Pasos “universales” que se deben seguir para construir una Inteligencia Artificial (I.A)



CONTENIDO

- Paso 1: Recolectar Datos
- Paso 2: Preparar los datos
- Paso 3: Elegir el modelo (“experiencia”)
- Paso 4 Entrenar nuestra máquina
- Paso 5: Evaluación (inferencia >80%)
- Paso 6 : Predicción o Inferencia
- Paso 7 : Empaquetar y desplegar la I.A

¡Siempre
hacia lo alto!





Paso 1: Recolectar Datos

Para cada problemática que se desea resolver, es necesario:

- investigar alcances (éticos y legales).
- Obtener datos, ojala reales (los simulados pueden ser muy ideales).

La calidad y cantidad de información que se pueda obtener impactará directamente en lo bien o mal que luego funcione la I.A.

Si se tiene información en una base de datos ya existente (evaluar la autenticidad y calidad en como fue construida) o crearla desde cero.

Si es un pequeño proyecto se pueden manejar planilla de cálculos que luego se exportará fácilmente como archivo csv.

Una buena técnica es usar el “**web scraping**” para recopilar información de manera automática de diversas fuentes (y/o servicios rest/ APIs).



Paso 2: Preparar los datos

Es importante mezclar de forma aleatoria la base de datos para no hacer OVERFITTING que obtengas ya que el orden en que se procesen los datos dentro de tu máquina no debe de ser determinante.

También es un bueno hacer gráficos estadísticos de los datos y revisar si hay correlaciones entre las distintas características (“features”, suelen ser las columnas de la base datos o archivo).

Los gráficos y matrices permitirán seleccionar “Características ideales”, las que se elijan impactarán directamente en los tiempos de ejecución y en los resultados.

En las IA de clasificación garantizar que la cantidad de datos que se tienen para cada resultado(clase), sea representativo, ya que si no, el aprendizaje podrá ser tendencioso hacia un tipo de respuesta y cuando nuestro modelo intente generalizar el conocimiento fallará.

Al separar los datos en dos grupos (entrenamiento y evaluación) del modelo la proporción depende de la cantidad de datos :

- < 10.0000 70/30
- < 1.000.000 75/25
- >1.000.000 80/20

Es valido preprocesar, normalizar, eliminar duplicados y hacer corrección de errores (**que hacer con los nulos????**).



Paso 3: Elegir el modelo (“experiencia”)

Modelo	Aplicaciones (Ejemplo de uso)
Logistic Regression	Predicción de precios
Fully connected networks	Clasificación
Convolutional Neural Networks	Procesamiento de imágenes
Recurrent Neural Networks	Reconocimiento de Voz
Random Forest	Detección de Fraude
GENETIC ALgorithm	Enseñarle a la máquina a jugar videojuegos y vencer!
Generative Models (GAN)	Creación de imágenes
K-means	Crear Clusters a partir de datos sin etiquetar.
k-Nearest Neighbors	motores de recomendación (por similitud/cercanía)
Bayesian Clasifiers	Clasificación
¿CUALES FALTARON Y CUALES NO HEMOS VISTO?	



Paso 4 Entrenar nuestra máquina

Utilizando el set de datos de entrenamiento para ejecutar el (FIT) del modelo seleccionado.



¡Siempre
hacia lo alto!



Paso 5: Evaluación (inferencia >80%)

Luego de entrenar solo queda comprobar la I.A con el set de datos de Evaluación que contiene datos que el modelo desconoce.

Verificar la precisión de nuestro modelo ya entrenado.

Si la exactitud es menor o igual al 50% ese modelo no será útil ya que sería como lanzar una moneda al aire para tomar decisiones.

Se es superior a 50% pero inferior a 80%, puede que debamos revisar:

- los datos que se están usando para entrenar (descartar algunos o agregar otros).
- Los hyperparametros usados para el entrenamiento (aumentar los ciclos de entrenamiento, ampliar la taza de aprendizaje, minimizar el costo de la función, etc).

Si alcanzamos un 80% o más podremos tener una buena confianza en los resultados que nos otorga el modelo.

¡Siempre
hacia lo alto!



Paso 6 : Predicción o Inferencia

Con el modelo de Aprendizaje Automático (I.A) le suministramos nueva información y comenzar a predecir o inferir resultados “en la vida real”...

¡Siempre
hacia lo alto!



Paso 7 : Empaquetar y desplegar la I.A

- Serializar el modelo de I.A usando un “serializador de objetos” en Python (**Pickle , h5py**)
- Crear una API (dganjo, flask) usando servidores en PYTHON
- Instalar librerías necesarias.
- Desplegar la API



UNIVERSIDAD SANTO TOMÁS

PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA

SECCIONAL TUNJA

VIGILADA MINEDUCACIÓN - SNIES 1732

¡Siempre hacia lo alto!

USTATUNJA.EDU.CO



@santotomastunja