



UNIVERSIDAD
NACIONAL
DE COLOMBIA

SEGMENTACIÓN SEMÁNTICA DE VEHÍCULOS EN IMÁGENES AÉREAS DE DRONES USANDO APRENDIZAJE PROFUNDO

TALLERISTAS

Prof. John Robert Ballesteros, Ph.D.

Prof. John W. Branch, Ph.D.

Ing. Mateo Cano, M.Sc

Ing. María Camila Durango, M.Sc

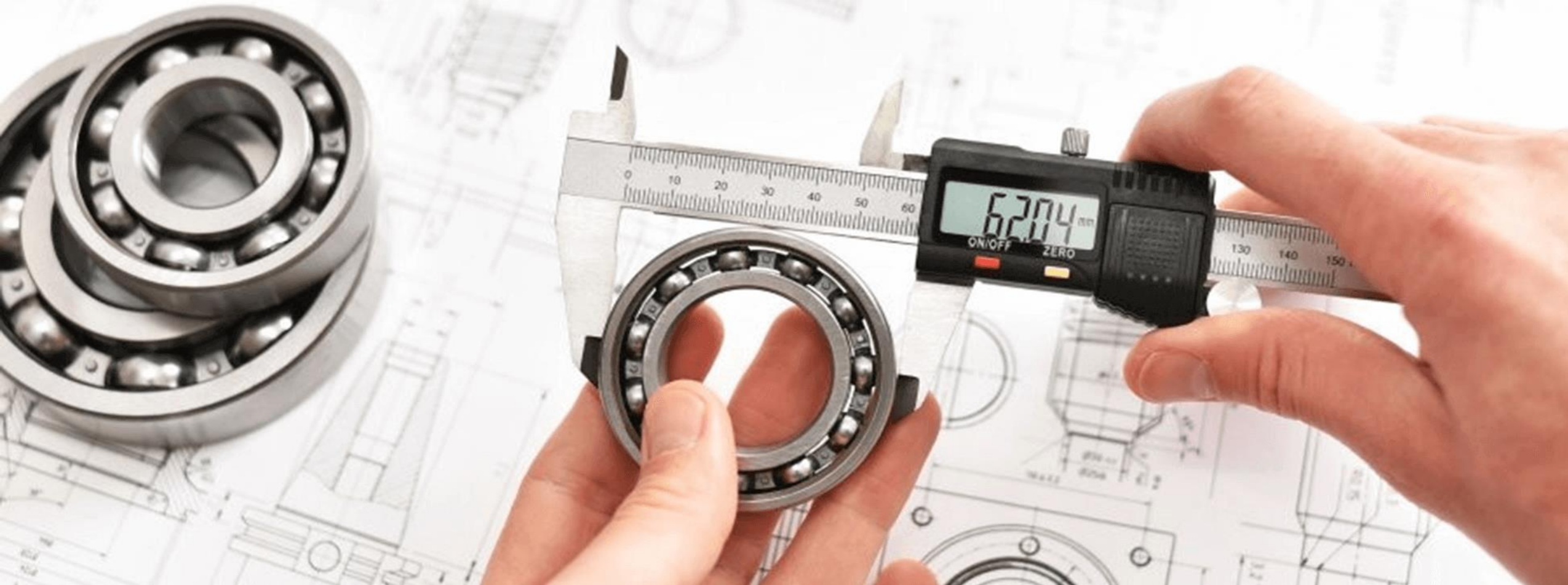
Ing. Camilo Laiton, M.Sc

Grupo de Investigación y Desarrollo en Inteligencia Artificial – GIDIA

Junio 26 de 2023

CICLO DE VIDA DE UN PROYECTO BASADO EN DATOS.





Obtención de datos

DEFINICIÓN DE DATO



Un dato es una representación simbólica (numérica, alfabética, algorítmica, espacial, etc.) de un atributo o variable cuantitativa o cualitativa. Los datos describen hechos empíricos, sucesos y entidades.



Los datos aisladamente pueden no contener información humanamente relevante. Solo cuando un conjunto de datos se examina conjuntamente a la luz de un enfoque, hipótesis o teoría se puede apreciar la información contenida en dichos datos.

Los datos pueden consistir en **números, estadísticas o proposiciones descriptivas.**

Definición aceptable para datos estructurados. Son los primeros que se nos vienen a la mente cuando nos hablan de datos.

ESTÁNDARES JERÁRQUICOS DE CALIDAD PARA LOS DATOS

1. Confiabilidad

- ✓ **Exactitud:** Los datos proporcionados son precisos.
- ✓ **Consistencia:** Todos los datos son consistentes o verificables.
- ✓ **Integridad:** El formato de los datos es claro y cumple los criterios.
- ✓ **Compleitud:** Una deficiencia de un componente afectará la precisión y la integridad de los
- ✓ **Credibilidad:** Cuando los datos provienen de organizaciones especializadas, gubernamentales o industriales, son auditados regularmente, verificados.

2. Disponibilidad.

- ✓ **Accesibilidad:** Los datos son fáciles de adquirir.
- ✓ **Oportunidad:** Los datos llegan a tiempo. Los datos se actualizan regularmente.

3. Usabilidad.

- ✓ **Estandarización:** Los datos están en formatos estándar, fáciles de usar o transformar.

4. Pertinencia

- ✓ **Conveniencia:** Los datos recogidos exponen completamente el tema de interés o parte de él.

5. Presentación

- ✓ **Legibilidad:** Los datos (contenido, formato, etc.) son claros y comprensibles.

CLASIFICACIÓN DE TIPOS DE DATOS

Datos estructurados:

Los datos estructurados tienen perfectamente definido la longitud, el formato y el tamaño de sus datos. Se almacenan en formato tabla, hojas de cálculo o en bases de datos relacionales.

Datos no estructurados:

Los datos no estructurados se caracterizan por no tener un formato específico. Se almacenan en múltiples formatos como documentos PDF o Word, correos electrónicos, ficheros multimedia de imagen, audio o video, etc.

Datos semiestructurados:

Los datos semiestructurados son una mezcla de los dos anteriores, no presentan una estructura perfectamente definida como los datos estructurados pero si presentan una organización definida en sus metadatos donde describen los objetos y sus relaciones, y que en algunos casos están aceptados por convención, como por ejemplo los formatos HTML, XML o JSON.

Fuente: <https://rpubs.com/>

CLASIFICACIÓN DE TIPOS DE DATOS

Los datos pueden ser clasificados en 4 dominios, dependiendo de su origen:

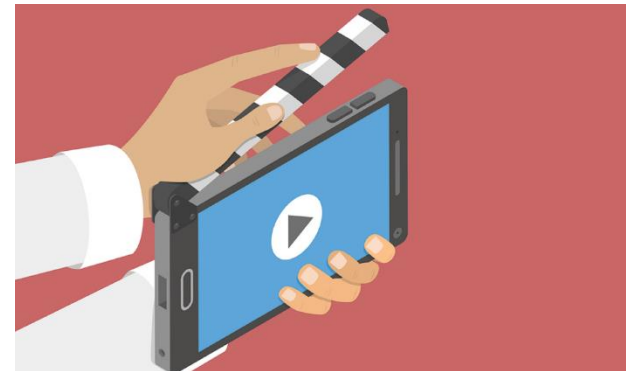
TEXTO

CUANDO **EL** OJO **VE** UN **COLOR** SE **EXCITA** INMEDIATAMENTE, Y ÉSTA **ES** SU **NATURALEZA**, ESPONTÁNEA Y DE **NECESIDAD**, PRODUCIR **OTRA** EN LA QUE **EL** COLOR **ORIGINAL** COMPRENDE LA ESCALA **CROMÁTICA** ENTERA. **UN** ÚNICO **COLOR** EXCITA, **MEDIANTE** UNA **SENSACIÓN** ESPECÍFICA, LA TENDENCIA A LA **UNIVERSALIDAD**. EN **ESTO** RESIDE LA LEY **FUNDAMENTAL** DE **TODA** ARMONÍA **DE** LOS **COLORES...**

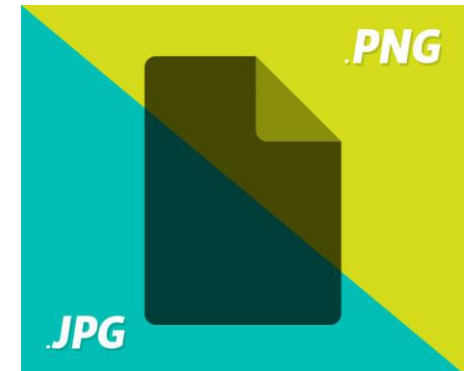
AUDIO



VIDEO



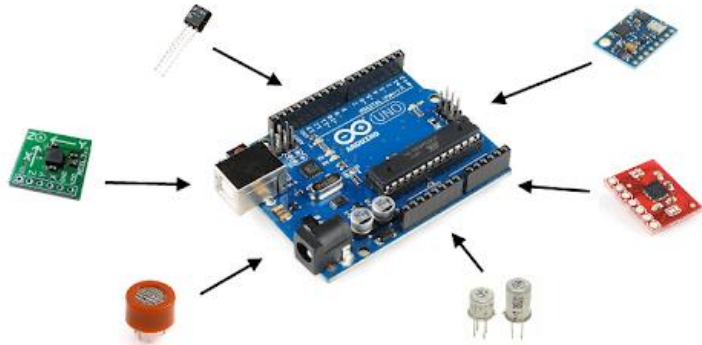
IMAGEN



CAPTURA MANUAL & AUTOMÁTICA DE DATOS



CAPTURA MANUAL

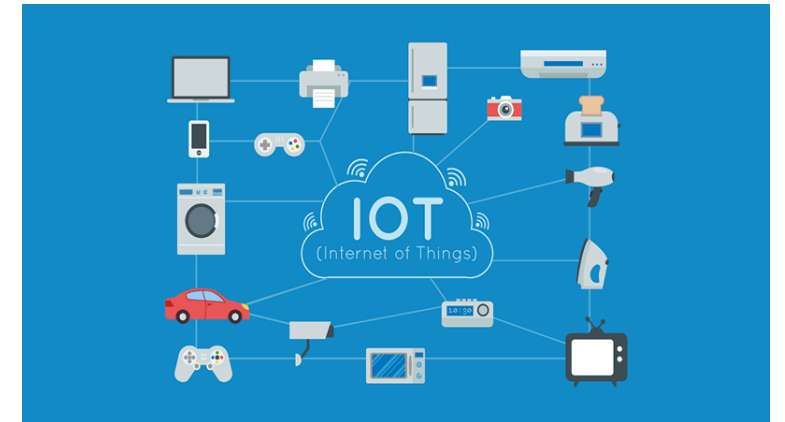


Evolución de la sensórica (menos costo)

CAPTURA AUTOMÁTICA



Generación de datos en la web



Generación de datos en con IoT

DIVERSIDAD EN LOS FORMATOS DE DATOS.



Texto, numérico.



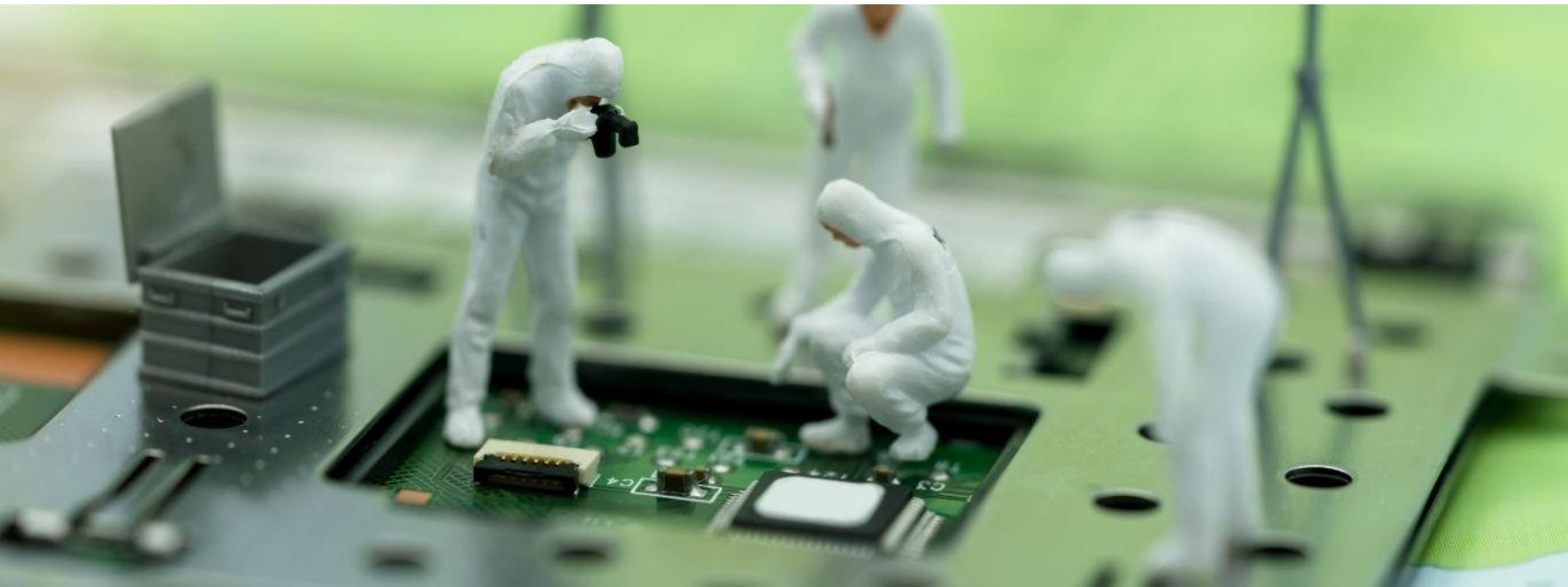
Imagen.



Audio.



Video.



Limpieza de datos

CICLO DE LIMPIEZA DE DATOS.



Importar datos: Realizar transformaciones de formatos de almacenamiento.

Fusionar datasets: Unir datos de diferentes fuentes para conformar, una sola unidad con todos los features.

Corrección de datos: Estrategias para reponer datos perdidos, datos erróneos, eliminar ruido (outlier).

Estandarizar datos: Normalizar o estandarizar la data para tener la misma escala y disminuir la cantidad de operaciones de computo.

Datos duplicados: Eliminar muestras duplicadas.

Verificar y enriquecer los datos: Visualizar la data para una mejor comprensión, incluir otros datos no considerados al principio, combinarlos mediante la estadística para eliminar redundancia.

Exportar datos: Exportar datos a otros formatos para diferentes tipos de análisis. CSV, JSON etc.

PRE-PROCESAMIENTO DE DATOS.

- ✓ Los datos a utilizar deben pasar por un proceso de preprocesamiento. Esto para seguir un estándar en los datos y lograr un mayor desempeño y exactitud a la hora de resolver el problema.
- ✓ Si los datos no pasan por este proceso, los resultados en las futuras etapas no podrán alcanzar los valores reales de precisión posibles.





Preguntas





UNIVERSIDAD
NACIONAL
DE COLOMBIA