Departamento de Ciencias de la Computación

Carrera de ingeniería en Sistemas e Informática

Implementación de un modelo de aprendizaje automático para el laboratorio de análisis de vulnerabilidades en el CERT Académico de la ESPE

Autores: - John Francisco Ponce Almachi
- Luigi Damián Villarreal Campaña

Director: Ing. Walter Marcelo Fuertes Díaz (PhD)

VERSIÓN: 1.0

AGENDA

1.- FORMULACIÓN DEL PROBLEMA

2.-OBJETIVOS

3.- METODOLOGÍA

4.- EVALUACIÓN DE RESULTADOS

5.- CONCLUSIONES Y RECOMENDACIONES

Amenazas cibernéticas en crecimiento

Marco de gestión

Abril de 2019 diversos medios informáticos, fueron víctimas de un ataque cibernético (alrededor de 40.000.000)

Estándares

ISO

- La Agencia Europea de Seguridad de las Redes de la Información (ENISA)
- La Comisión de la Unión Europea.
- La Unión Internacional de Telecomunicaciones
- OTAN

Objetivo General

Diseñar e implementar un modelo de aprendizaje automático para el análisis de tráfico malicioso, y su implementación en el laboratorio de análisis de vulnerabilidades del ESPE-CERT mediante la creación de scripts ejecutables y sus pruebas funcionales.

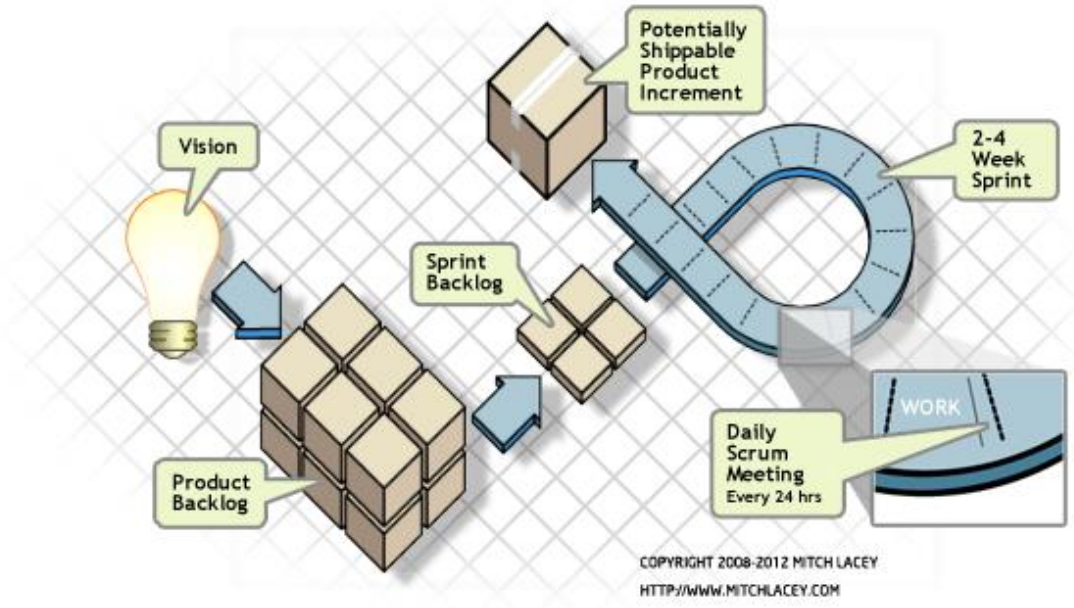
Objetivos Específicos:

a) Evaluar los métodos, técnicas y herramientas de machine learning para el análisis de tráfico malicioso.

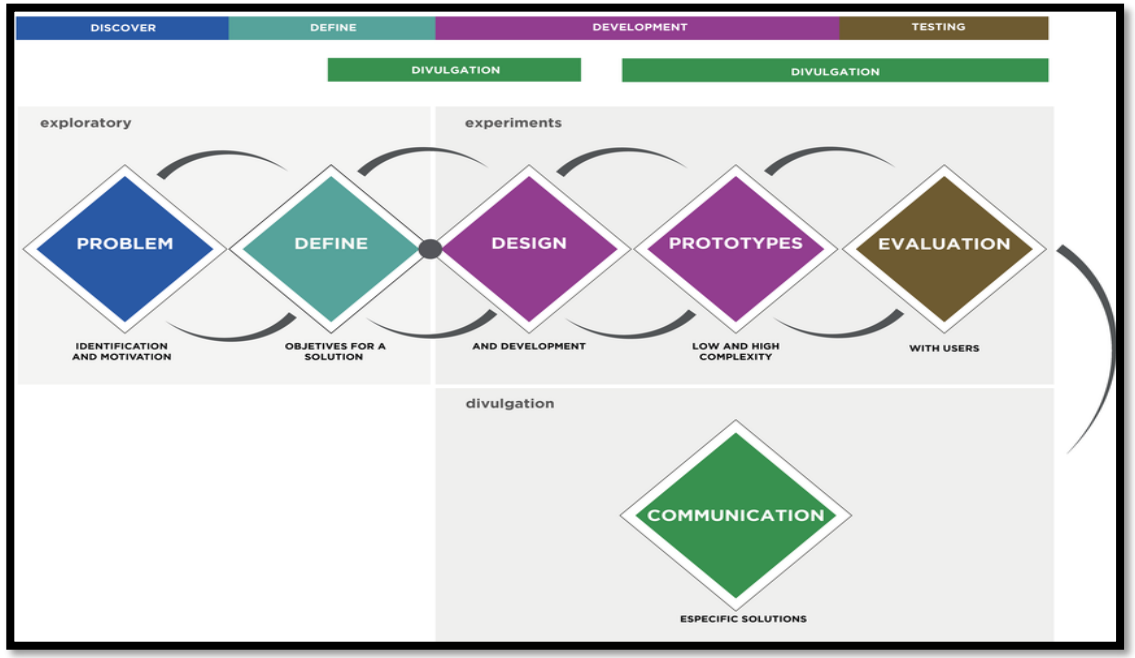
b) Diseñar e implementar un modelo de aprendizaje automático para el análisis de tráfico malicioso, a partir de las fases de la metodología ágil SCRUM.

c) Realizar las pruebas de concepto funcionales y no funcionales del modelo implementado.

d) Documentar los manuales de operación, técnico y de usuario y difundir sus resultados.



Metodología Ágil SCRUM – *Jesús Ramírez*

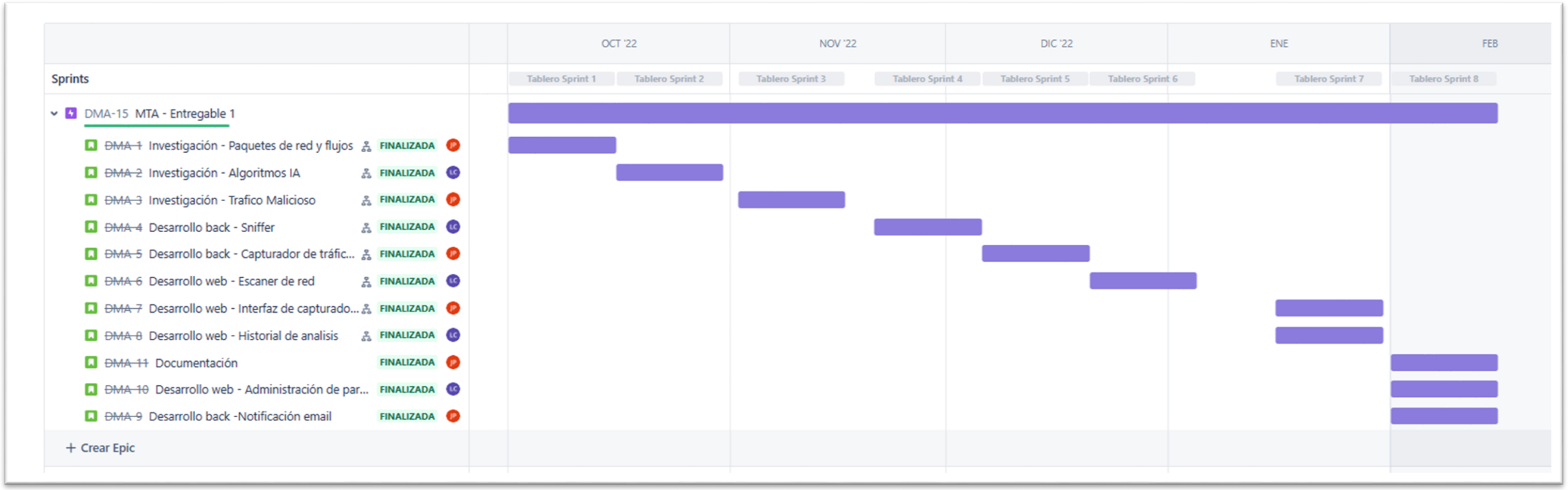


Fases de metodología Ciencia del diseño – *Oscar Díaz*

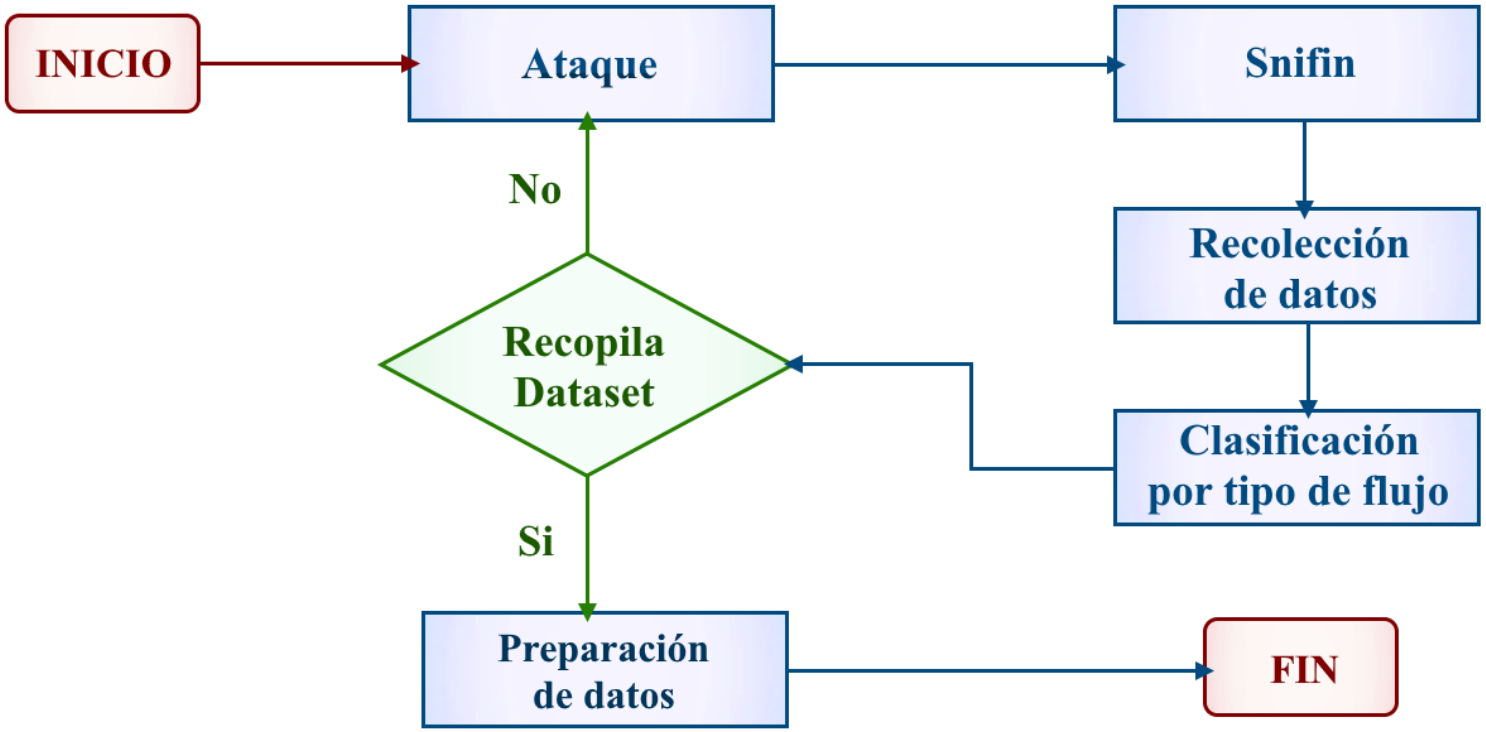
Historias de usuario

Orden de trabajo / Módulo	Actividad	Puntos Historia	Estimación Utilizada
CERT - OT001-R1	❖ Yo como investigador del ESPE-CERT requiero que se haga una análisis de características de paquetes de red que se pueden emplear para alimentar algoritmos de inteligencia artificial	21,00	21,00
CERT - OT001-R2	❖ Yo como investigador del ESPE-CERT requiero que se haga una análisis comparativo de algoritmos de inteligencia artificial basado en los paquetes de redes para aplicar el mas optimo en un análisis de red	21,00	21,00
CERT - OT001-R3	❖ Yo como investigador del ESPE-CERT requiero que se analice y aplique ataques de prueba para recolectar generar datasets	21,00	21,00
CERT - OT001-R4	❖ Yo como operador requiero que se capturen los paquetes de red para aplicar algoritmos de machine learning sobre los mismos	21,00	21,00

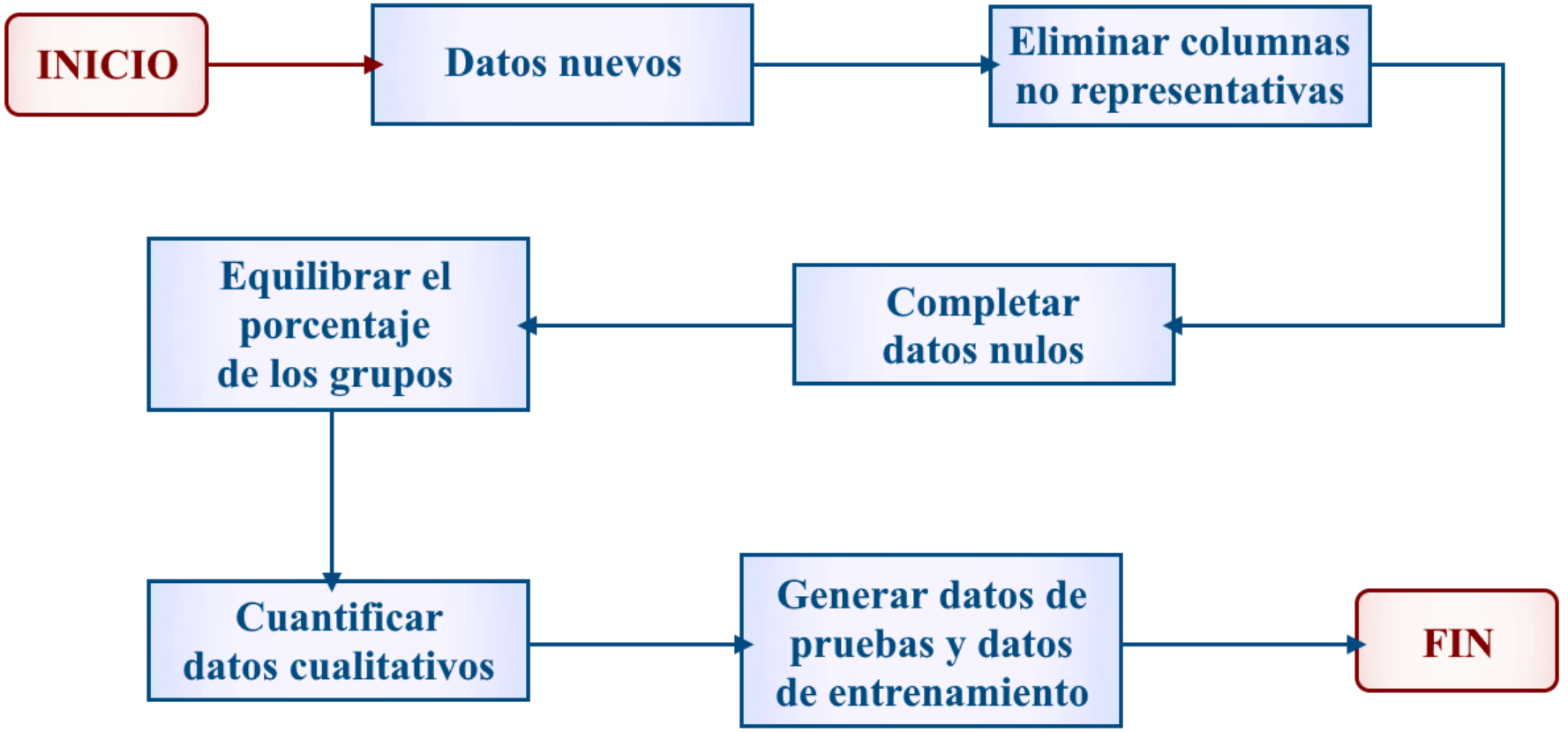
Planificación de los Sprint



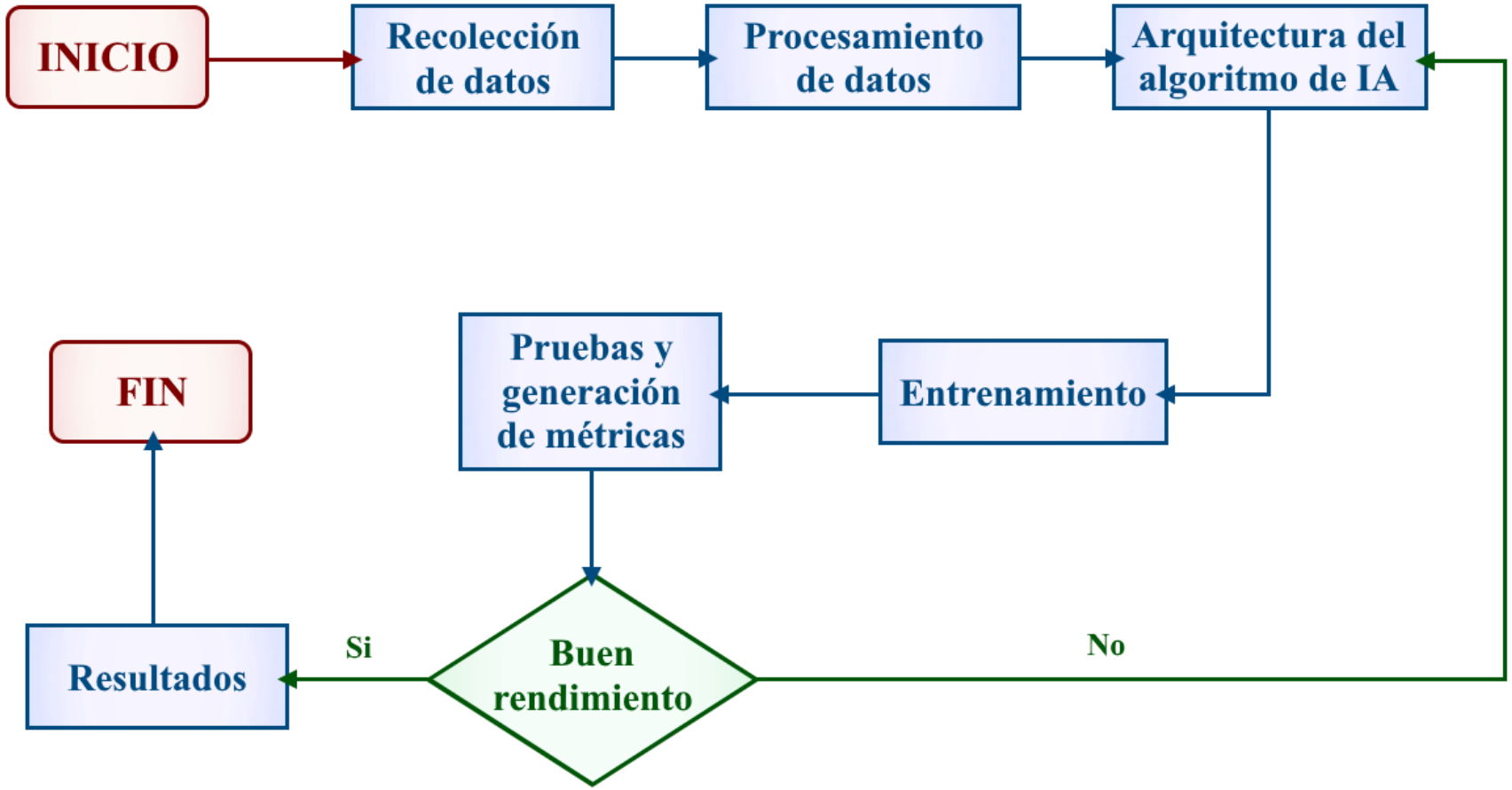
Hoja de seguimiento por Sprint – *Autoría propia*



Proceso de recolección de datos – *Autoría propia*

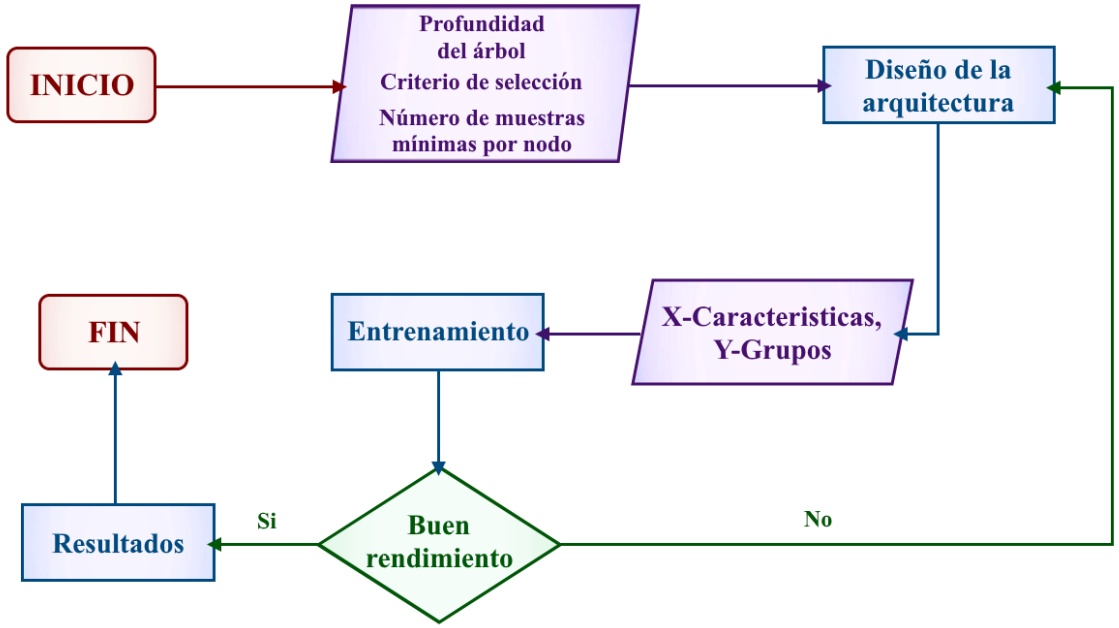


Preparación de datos – *Autoría propia*

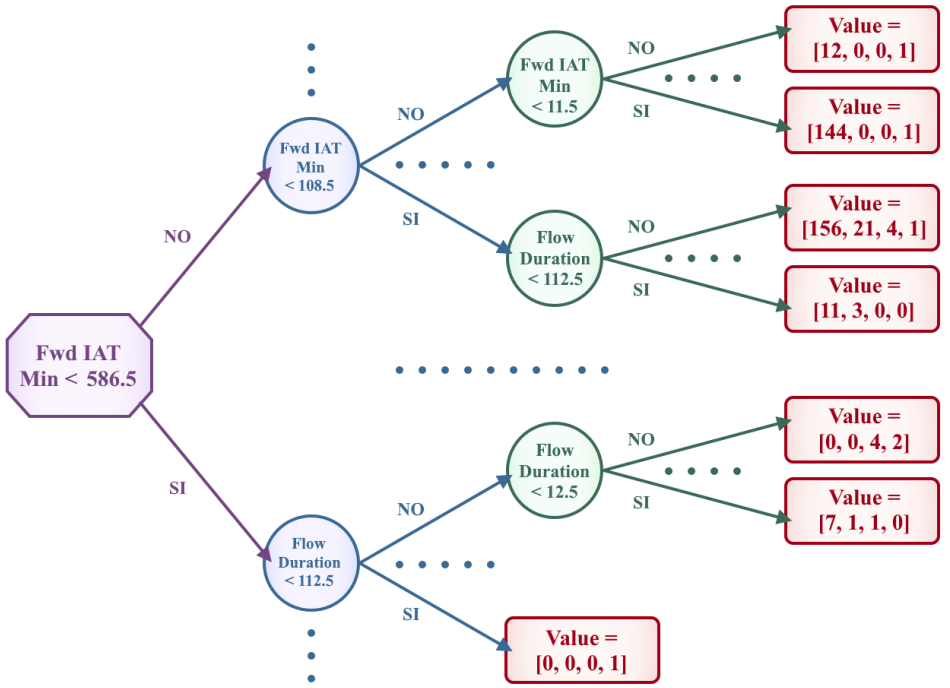


Proceso de desarrollo de algoritmos de inteligencia artificial – *Autoría propia*

03.5 **Flujograma algoritmos**

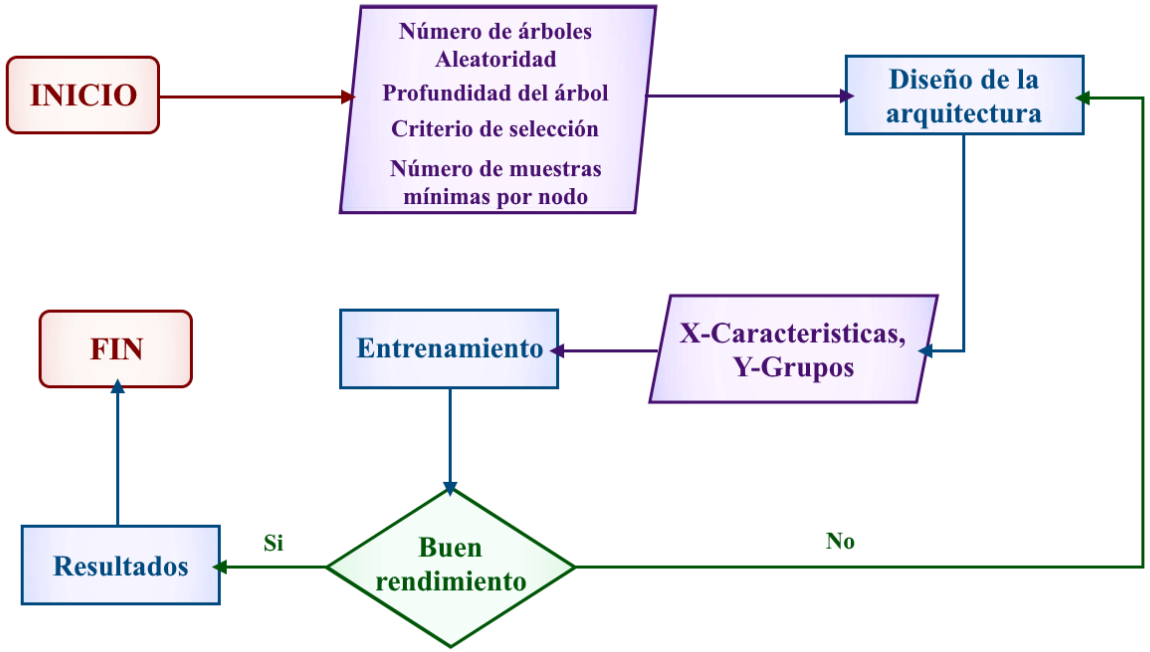


Proceso de diseño de Árbol de decisiones – *Autoría propia*

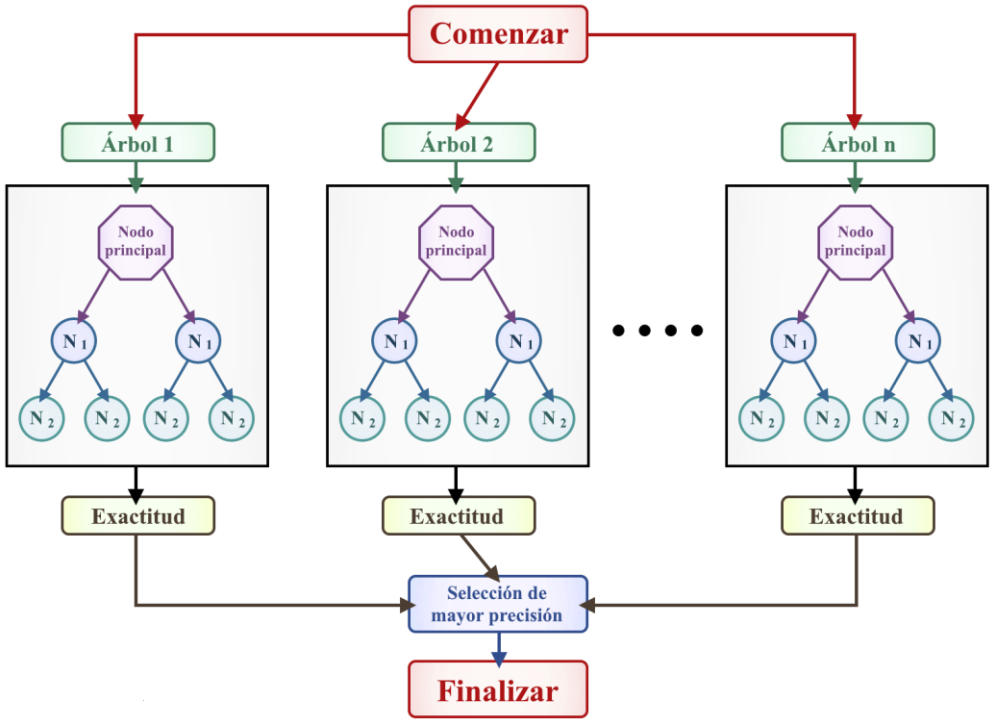


Diseño de Árbol de decisiones – *Autoría propia*

04.6 **Flujograma algoritmos**

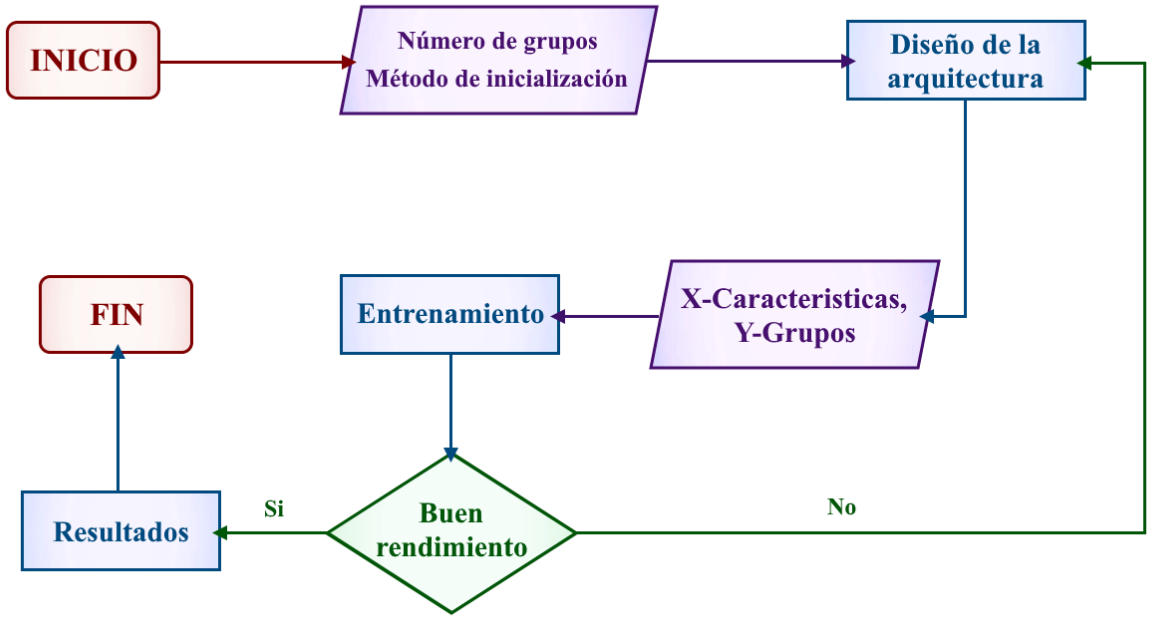


Proceso de diseño de Random Forest – *Autoría propia*

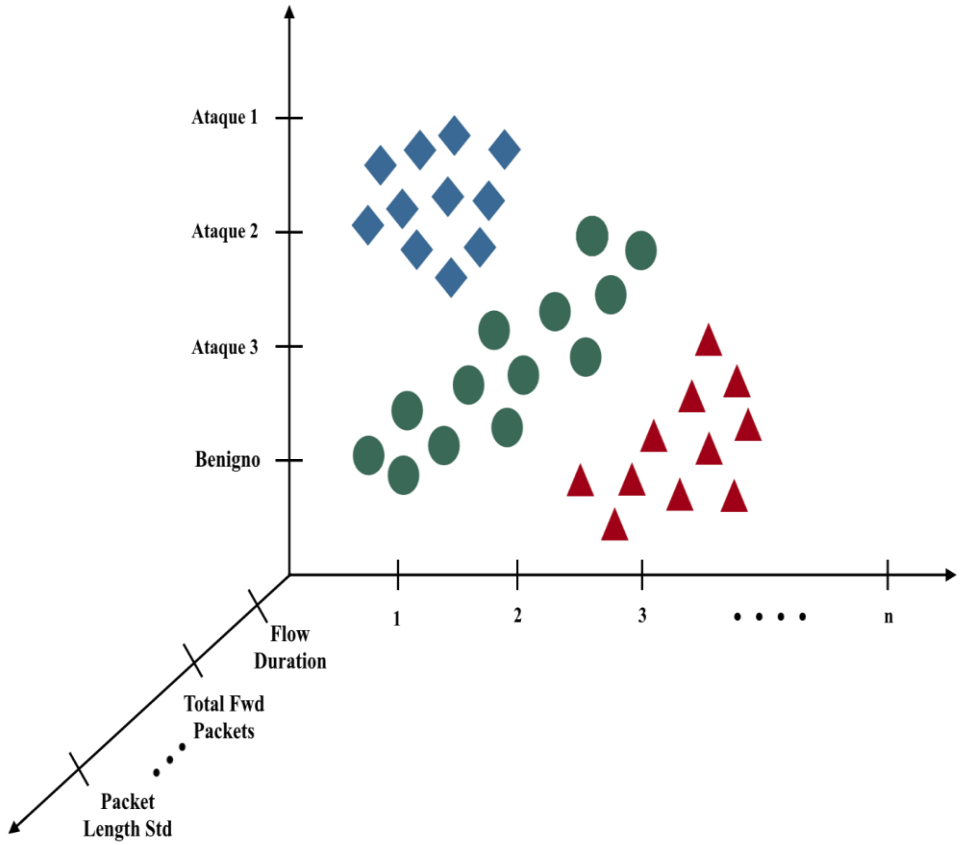


Diseño de Random Forest – *Autoría propia*

03.7 **Flujograma algoritmos**

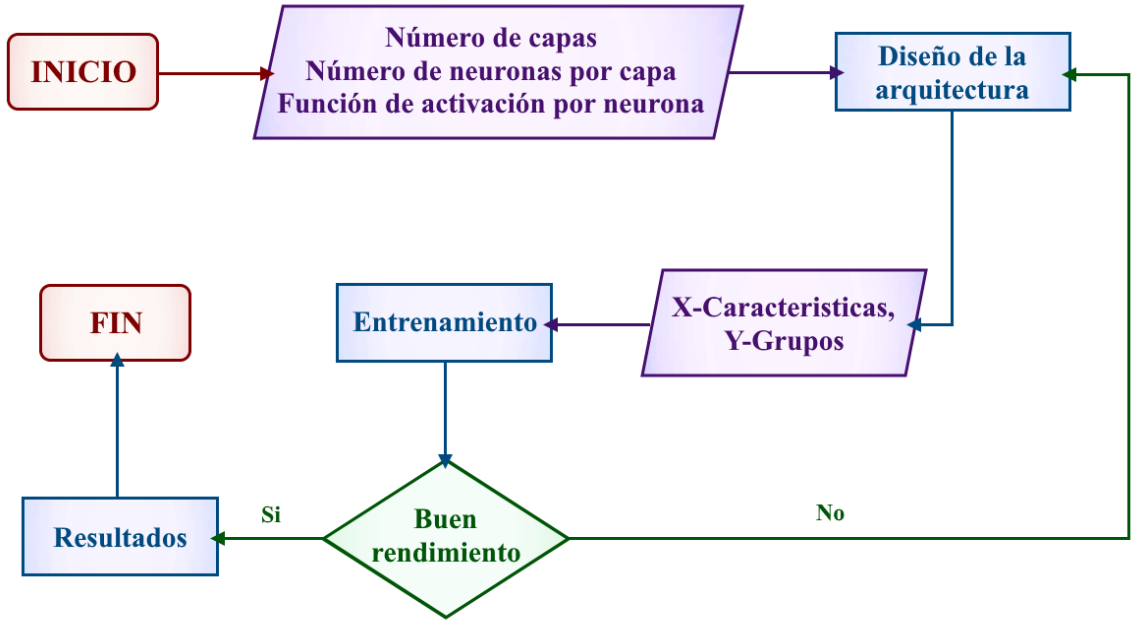


Proceso de diseño de red K-means – *Autoría propia*

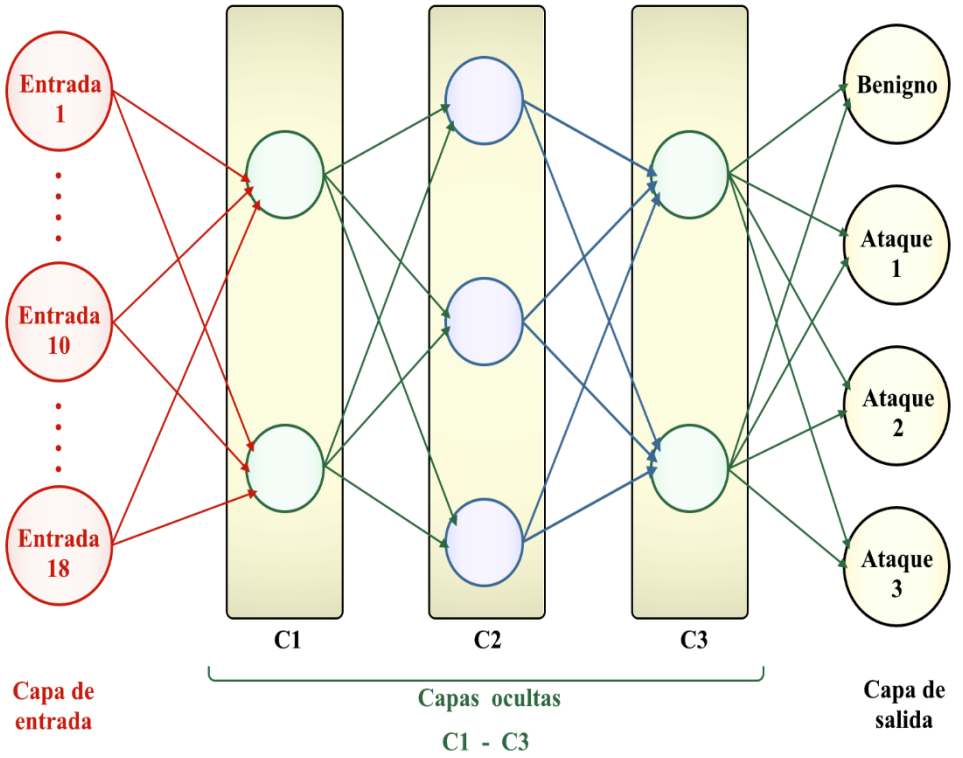


Diseño de K-means – *Autoría propia*

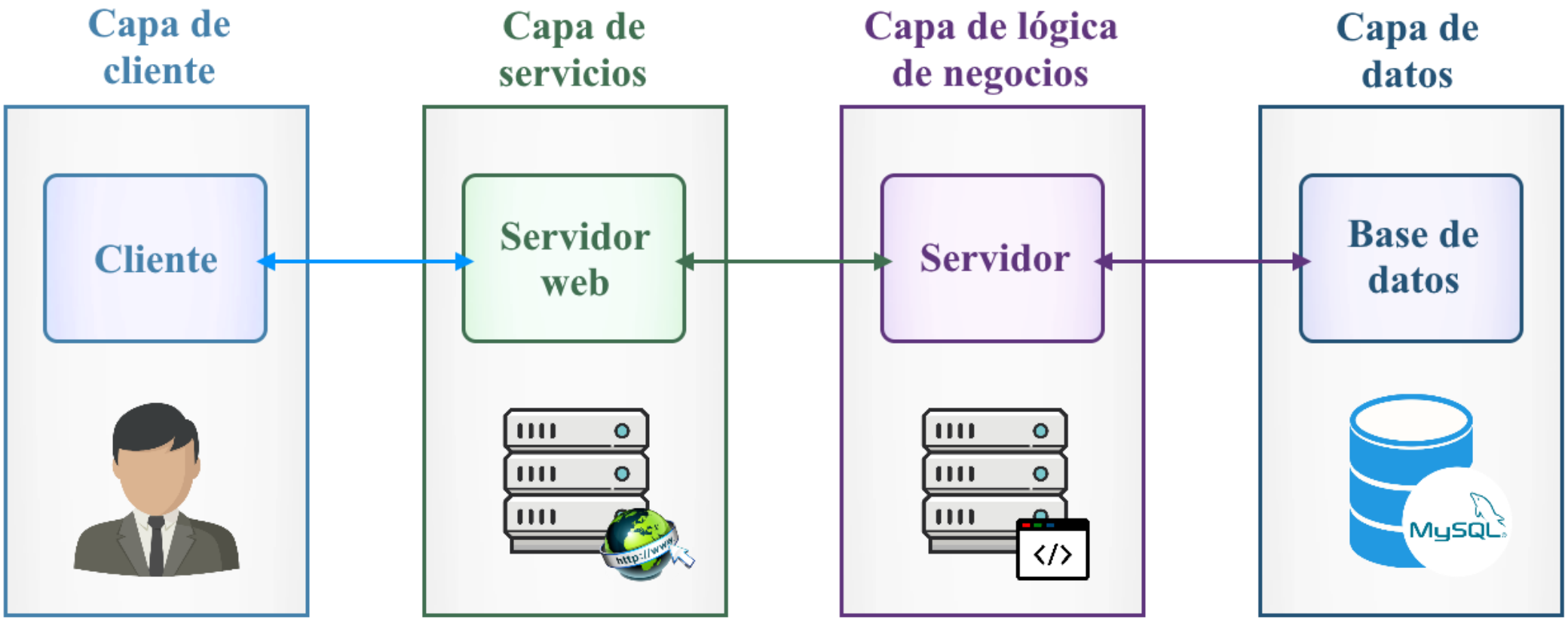
03.8 **Flujograma algoritmos**



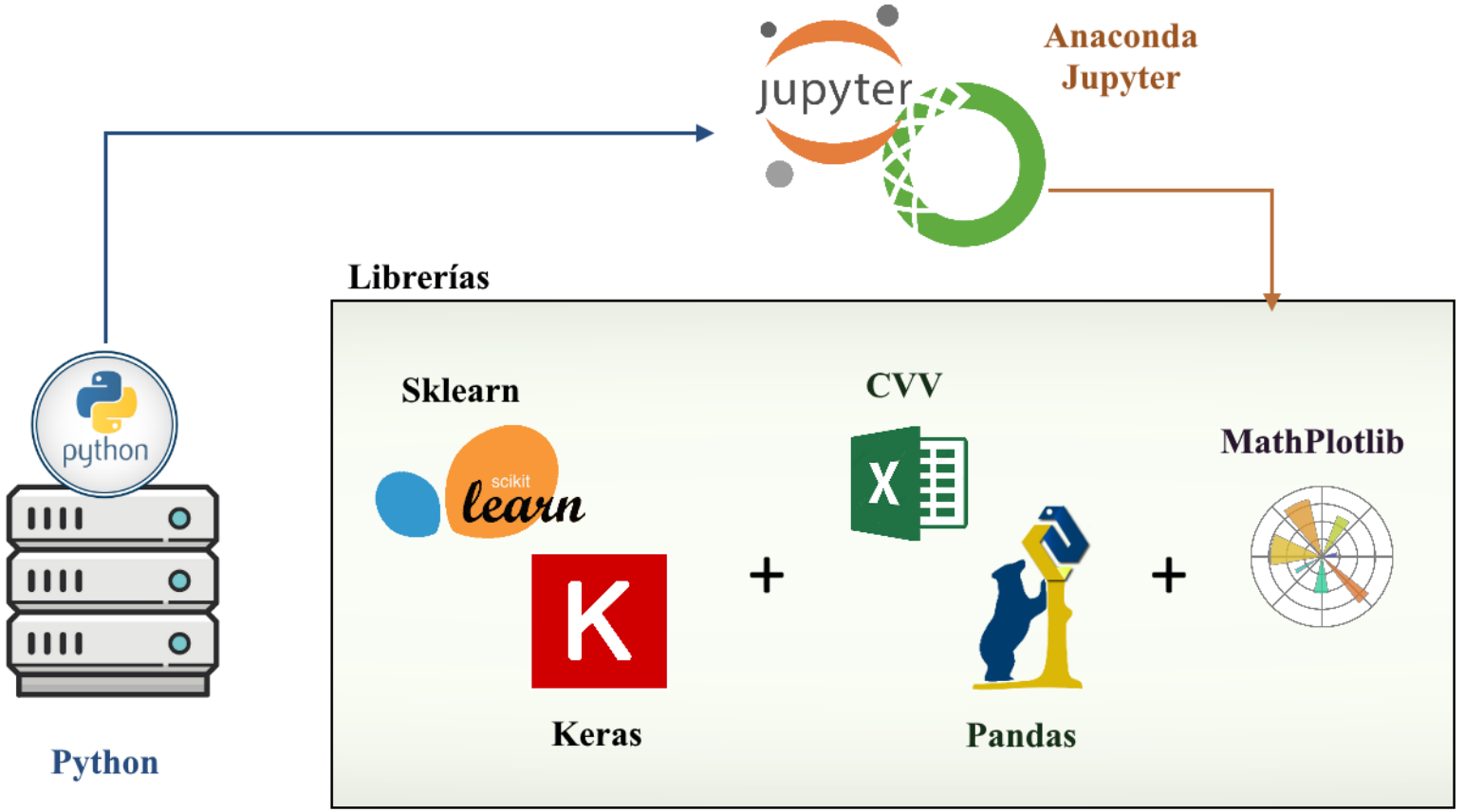
Proceso de diseño de red neuronal – *Autoría propia*



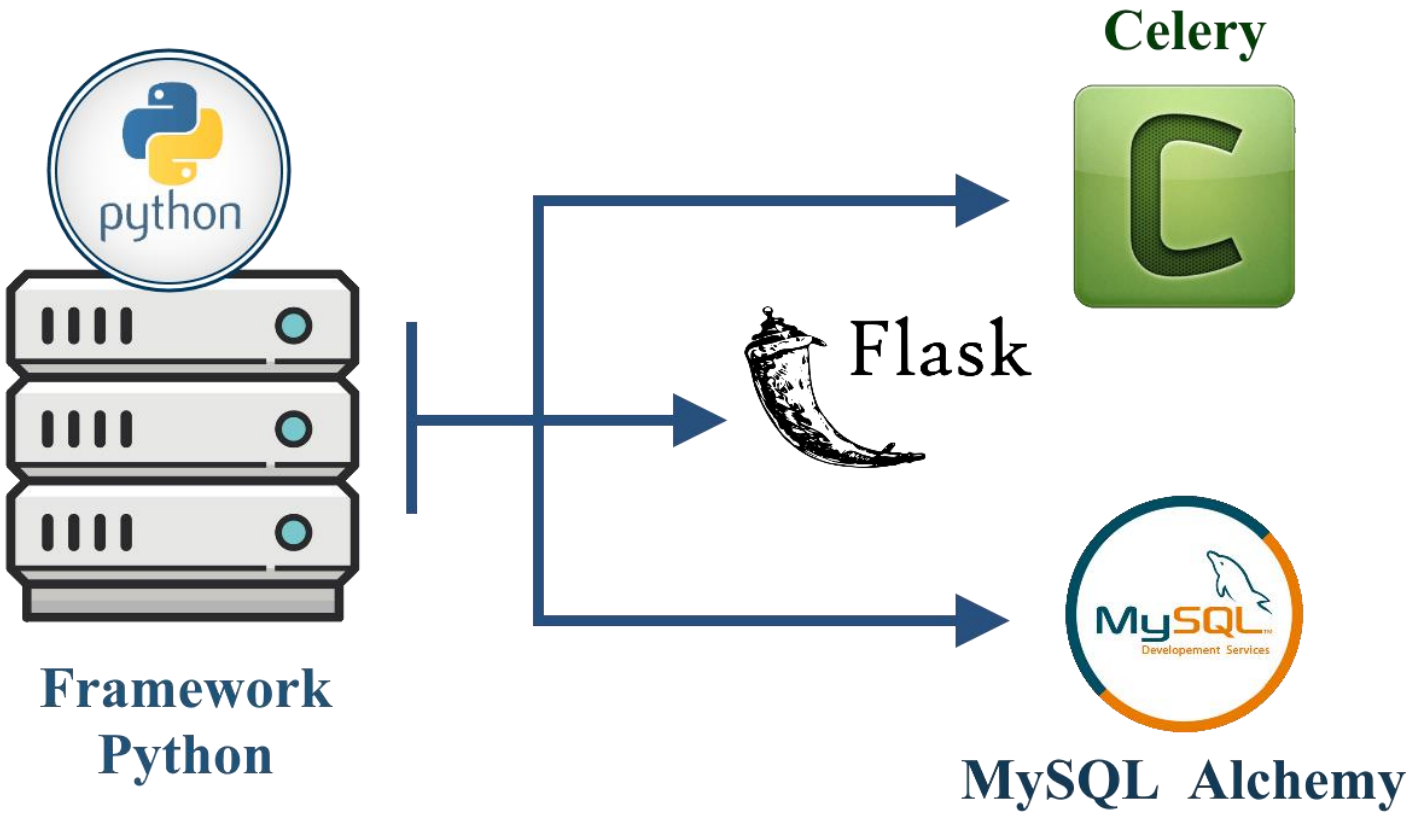
Diseño de red neuronal– *Autoría propia*



Diseño de la arquitecturas por capas– *Autoría propia*



Herramientas de desarrollo de algoritmos de IA– *Autoría propia*



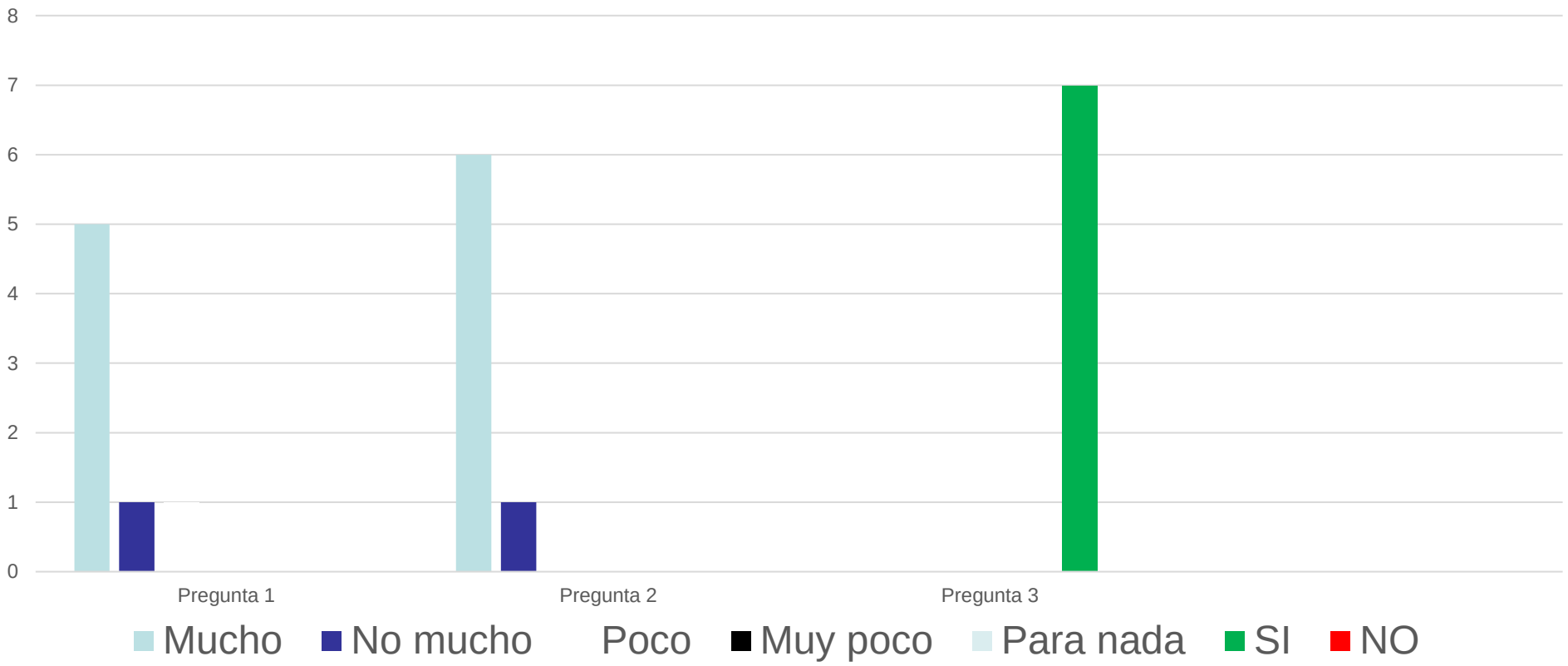
Herramientas de desarrollo de aplicación web – *Autoría propia*

Tabla de procesos de prueba

No	Identificador	Objetivo	Pre condición	Entrada	Pasos	Resultados esperados	Post condición	Estado	Prioridad
1	INICIO_001	Verificar maquetación de la pantalla	N/A	N/A	Iniciar la aplicación web	**Cabecera de la página web con las opciones** a) Logotipo (Imagen) b) Nombre del sistema (Texto) c) Historial (Enlace) d) Acerca de (Enlace) e) Iniciar escaneo (Enlace) f) Configuración (Enlace) g) Inicio (Enlace) **Cuerpo de la página web con las opciones** a) Ilustración representativa (Imagen) b) Información del sistema (Texto) c) Botón iniciar escaneo (Botón) **Pie de página web con las opciones** a) Marquesina (Texto)	N/A	Abierto	Alta

ISO 29119

03.12 **Pruebas usabilidad**



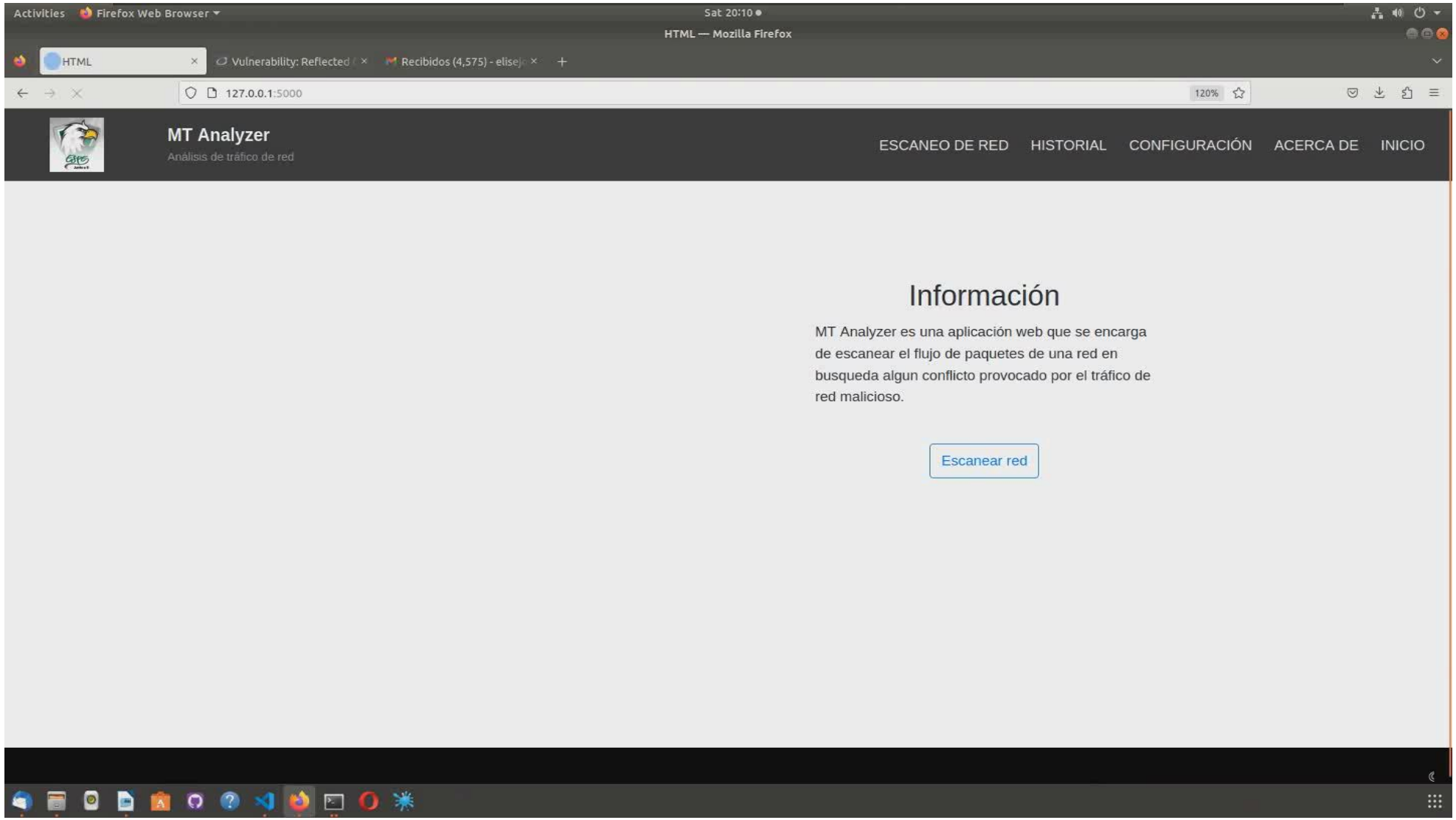
Resultados encuesta de satisfacción al cliente según ISO 9001

Matriz comparativa de Algoritmos vs Métricas

Modelo de aprendizaje	Exactitud	Sensibilidad	Precisión	Puntuación
Decision Tree	0,9803	0,8602	0,6873	0,6873
Random Forest	0,9818	0,8944	0,6990	0,6990
K-means	0,8292	0,0388	0,0156	0,0156
Neuronal networking	0,9450	0,8135	0,3753	0,3753

05.1 Evaluación de resultados

Funcionamiento Aplicación web



a) Al evaluar los algoritmos de inteligencia artificial, se determinó que el más óptimo es Random Forest, con un valor de Exactitud del 98,18% y el menos óptimo fue el K-means con un porcentaje de 82,92%.

b) Python demostró ser un lenguaje de programación completo que permitió desarrollar la funcionalidad para capturar el tráfico de red, algoritmos de inteligencia artificial e implementar la aplicación web con sus servicios.

c) Para realizar las pruebas funcionales y no funcionales, fue necesario aplicar las normativas ISO 9001 para la usabilidad y la ISO 29119 para pruebas de software.

d) El manual técnico y de usuario detallan la estructura y funcionamiento de la aplicación diseñada específicamente para usuarios nuevos.

a) El Dataset debe ser debidamente procesado para generar resultados que no proporcionen incoherencias en la predicción de nuevos datos.

b) Al trabajar con Python, se debe buscar frameworks y librerías. Igualmente, se recomienda el manejo de JIRA u otro software similar.

c) Para las pruebas funcionales y no funcionales, se recomienda buscar una normativa con el fin de proporcionar calidad en el entregable final.

d) Se recomienda revisar los manuales con los clientes, con el fin de comprobar que la información sea clara para los mismos.

Gracias