

Profesor: Yasmany Prieto

Tarea 1: Aprendizaje supervisado. Diseño y evaluación de clasificadores.

Objetivos:

- Diseñar y evaluar clasificadores para problemas reales.
- Analizar y presentar un artículo científico del tema.

Tiempo de desarrollo de la tarea:

20 días

Equipos:

De hasta 3 estudiantes

Enunciado:

Problema 1.

En redes de computadores un Firewall o cortafuegos es un sistema diseñado para bloquear accesos no autorizados a la red local. Los cortafuegos verifican las conexiones entre un host fuera de la red y un host local, bloqueando aquellas que no cumplen con los requisitos de seguridad definidos. Las políticas de seguridad definidas en los cortafuegos abarcan desde la inspección de direcciones IP, puertos, y protocolos hasta el análisis del comportamiento de las aplicaciones de red en los paquetes de entrada y salida.

La base de datos reportada en (<https://archive.ics.uci.edu/ml/datasets/Internet+Firewall+Data>) contiene *logs* del comportamiento de un Firewall ante varios intentos de comunicación. La base de datos cuenta con 11 características que describen cada intento de conexión y una variable que describe la acción que tomó el Firewall. Se desea construir un clasificador capaz de emular el comportamiento de este Firewall, de manera que ante un intento de conexión el clasificador reconozca que acción debe tomar, de acuerdo con el comportamiento aprendido del Firewall original.

- i) Describa el dataset, identificando las características y la variable de clasificación, que tipo de datos es cada una, y que valores toma la variable de clasificación. ¿El dataset presenta datos faltantes?
- ii) Para el clasificador que propone como solución modifique los parámetros que tiene disponibles en la implementación de *sklearn*. Construya una tabla como la que aparece a continuación y escoja cuáles son los valores óptimos de los parámetros (pruebe al menos con 5 valores de los parámetros). Tome uno de los parámetros y grafique su relación con la métrica de desempeño. Interprete esta gráfica de acuerdo con el posible *underfitting* u *overfitting* del modelo.

Parámetro 1	...	Parámetro n	<i>accuracy promedio por clase</i>
...			
...			

- iii) Del clasificador que propone como solución del problema entregue una tabla que muestre los resultados de la matriz de confusión, así como los valores de las métricas de *recall* y *precision* de cada clase. Tome como referencia la que aparece a continuación. Interprete un valor de *recall* y *precision*. Obtenga las métricas de *accuracy* y *accuracy promedio por clase*, comente la diferencia entre estos valores.

		Prediction				
		<i>durionis</i>	<i>ficulneus</i>	<i>fructosus</i>	<i>pseudo.</i>	Recall
Target	<i>durionis</i>	5	0	2	0	0.714
	<i>ficulneus</i>	0	6	1	0	0.857
	<i>fructosus</i>	0	1	10	0	0.909
	<i>pseudo.</i>	0	0	2	3	0.600
Precision		1.000	0.857	0.667	1.000	

- iv) Ocupe alguno de los métodos vistos en clase para reducir la dimensión de las características del problema y evalúe su impacto en las soluciones obtenidas (compare las métricas de desempeño). Pruebe al menos con la mitad de las características y con dos características.
- v) Ocupe algún método de balanceo de clases de ser necesario y evalúe su impacto en las soluciones obtenidas (compare las métricas de desempeño y la matriz de confusión).

Problema 2.

Escoja un artículo científico donde se emplee al menos un método de reconocimiento de patrones con aprendizaje supervisado, para resolver algún problema práctico. De este artículo extraiga la siguiente información:

- ¿Cuál es el problema práctico que se busca resolver? ¿Por qué es relevante? ¿Por qué se propone resolverlo mediante métodos de reconocimiento de patrones?
- ¿Qué métodos de clasificación se emplean en el trabajo? ¿Por qué éstos y no otros?
- ¿Cuál es la metodología empleada en el trabajo? ¿Cómo se diseñan los experimentos? ¿Qué tipos de datos se emplean? ¿Qué métricas se usan para medir los resultados?, etc.
- ¿Cuáles fueron los resultados del trabajo y que se concluyó? Discuta gráficos, comparaciones, etc.
- ¿Qué hubiese hecho usted diferente? ¿Por qué?

Incluya en el informe la referencia al artículo.

Entrega:

- i) Se debe entregar un informe que contemple las respuestas a los Problemas 1 y 2.
- ii) El análisis del artículo incluye una presentación donde se comunique la información extraída. La fecha de presentación es el **1 de junio de 2023**.

Adicionalmente, se puede incluir de manera opcional un resumen sobre que se aprendió de la experiencia o si se experimentó alguna dificultad específica.

¿Cómo se evaluará mi trabajo?

Aspectos para evaluar	Ponderación
Descripción del dataset	0.2
Selección del clasificador óptimo. Ajuste de parámetros de clasificador.	1.2
Interpretación de la relación entre el valor de los parámetros del clasificador y el valor de la métrica de desempeño.	0.4
Obtención de matriz de confusión e interpretación de <i>recall</i> y <i>precision</i> .	0.3
Interpretación de diferencia entre <i>accuracy</i> y <i>accuracy promedio por clase</i> .	0.3
Uso de métodos para reducción de dimensionalidad. Interpretación de los resultados obtenidos.	0.4
Uso de métodos de balanceo de clases o justifique que no es necesario su uso. Interpretación de los resultados obtenidos.	0.4
El artículo científico escogido está relacionado con el uso de clasificadores y el aprendizaje supervisado.	0.5
Sintetiza correctamente toda la información que se solicita.	1.0
Propone alguna mejora a los resultados del artículo a partir de los conocimientos del curso.	0.3
Presentación del artículo en clase.	0.5
El informe es claro y organizado. Se usa vocabulario de la asignatura.	0.5
Total	6.0