

Guía para trabajo
Aprendizaje de máquinas supervisado
Inteligencia Artificial 2025-II

La actividad consiste en entrenar modelos de aprendizaje supervisado para lo cual cada grupo debe descargar el conjunto de datos asignado. Para entrenar los modelos, se deberá inspeccionar, y preprocesar los datos si es necesario, diseñar y llevar a cabo experimentos para ajustar parámetros, evaluar y comparar el desempeño de los métodos implementados. El entregable de esta actividad es un notebook de Python que debe contener el desarrollo de los siguientes ítems:

- I. Descripción del problema e inspección del conjunto de datos
 1. Una descripción del problema de clasificación/regresión del cual trata el conjunto de datos asignado, incluyendo una descripción de las variables con la que cuenta su conjunto de datos, número de casos, etc. Para esto deberá consultar la documentación que el repositorio contiene junto con el conjunto de datos, y otras fuentes de acuerdo con el tema tratado. La descripción en palabras debe aparecer en el notebook como texto, adicional a las líneas de códigos utilizadas para mostrar el contenido del conjunto de datos.
 2. Inspección o exploración del dataset utilizando herramientas gráficas de visualización tales como histogramas, diagramas de barra, wordclouds, etc; resúmenes estadísticos, matrices de correlación debidamente explicadas y analizadas. La explicación y análisis de esta inspección debe estar contenida en entrada de textos en el notebook de desarrollo.
- II. Diseño de experimentos y recolección de resultados
 3. Una descripción de las estrategias utilizadas para preparar los datos de acuerdo con las necesidades del trabajo a realizar, tales como conversión de variables, relleno de valores faltantes, selección o reducción de características, balanceo del conjunto de datos, vectorización de los datos, etc. Esto debe estar soportado en el análisis realizado durante la inspección del conjunto de datos.
 4. Descripción de los experimentos realizados para ajustar los parámetros cada uno de los métodos vistos en clase. Al ajustar los parámetros tenga en cuenta que debe tratar de alcanzar el mejor desempeño posible sobre el conjunto de entrenamiento, y al final evaluar cada modelo sobre un conjunto de prueba. Los métodos para evaluar son: regresión multivariada, árboles de decisión, random forest, y redes neuronales (MLP, y DNN). Utilice la función GridSearchCV la cual además de automatizar las pruebas de múltiples combinaciones de parámetros, también implementa una estrategia de repetición de cada experimento llamada validación cruzada.
 5. Una explicación o análisis del resultado final obtenido por los modelos con los parámetros ajustados debe incluir métricas de evaluación como accuracy, matrices de confusión, precision, recall, MSE, MAE, coeficiente de determinación etc. Este análisis también debe aparecer descrito y documentado en el notebook de trabajo.
- III. Comparación de los modelos entrenados y conclusiones
 6. Seleccione para cada método utilizado los mejores modelos obtenidos y elabore tablas y gráficas resúmenes para comparar los resultados.
 7. Escriba conclusiones del trabajo realizado con base en los resultados obtenidos de acuerdo con el objetivo del conjunto de datos asignado, incluyendo trabajo futuro que se pueda realizar para mejorar los resultados, o la evaluación del modelo. Estas deben estar de manera explícita, redactadas de manera clara en el notebook de trabajo.

El notebook desarrollado por cada grupo deberá ser subido a la plataforma del curso el **15 de noviembre**. Cada grupo después tendrá un espacio de 15 minutos asignado para sustentar su trabajo. Consulte la tabla de los grupos por su horario de sustentación que se llevarán a cabo los días 18 y 20 de noviembre para grupos 1 y 2; y el 19 y 21 de noviembre para el grupo 3.

Nota: tal como se indicó al principio del curso la calificación de este trabajo, cuyo total es por 150 puntos, se dividirá en 90 puntos para el Notebook y será una nota grupal, y 60 puntos para la sustentación final del trabajo que serán evaluado de forma individual.

Rúbrica de evaluación:

Elementos	5	4	3	2	1	0
Descripción del problema y objetivo	Se explica la problemática asociada a los datos de manera clara y documentado, utilizando fuentes tanto del repositorio de los datos, como de otras fuentes distintas y se propone un objetivo del trabajo acorde con el problema presentado.	Se explica la problemática asociada a los datos de manera poco elaborada, solo se hace referencia al contenido encontrado en el repositorio del cual se descargó los datos y se propone un objetivo del trabajo acorde con el problema presentado.	El problema se presenta de forma muy poco elaborada; pero el objetivo propuesto está acorde con el conjunto de datos asignado	Se planteó un objetivo del trabajo; pero no hay una descripción del problema asociado con los datos asignados, o esta descripción no es clara.	Ni el problema, ni el objetivo del trabajo están bien redactados y acorde con el conjunto de datos asignado.	El problema no se presentó, ni el objetivo del trabajo.
Inspección de los datos	El estudiante realizó inspección visual de los datos, la cual está muy bien soportada con gráficas y tablas explicadas, con sus respectivos títulos e interpretación.	El estudiante realizó inspección visual de los datos, la cual está soportada con gráficas y tablas, alguna de ellas no explicadas, y/o faltaron títulos descriptivos de las figuras y tablas utilizadas.	El estudiante realizó inspección visual de los datos; utilizó gráficas y tablas; pero no hizo ninguna explicación acerca de ellas, ó la interpretación que se presenta en el análisis es muy poco elaborada.	Solo se presentó pocas líneas de texto sobre la calidad del conjunto de datos respectivo; pero no hubo un análisis muy elaborado. La inspección realizada no se soportó ni en gráficas, ni en tablas.	No hay gráficas suficientes para observar el comportamiento de las variables y tomar decisiones con respecto a ellas.	No se hizo ninguna inspección de los datos.
Preprocesamiento de los datos	Se aplicaron operaciones de preprocesamiento que eran necesarias, y que estaban bien justificadas por la inspección de los datos.	Se aplicaron operaciones de preprocesamiento que eran necesarias; pero algunas de ellas no están soportadas o justificadas en la inspección o análisis de los datos.	No es claro la aplicación de algunas operaciones de preprocesamiento que no eran necesarias.	No se aplicaron algunas operaciones de preprocesamiento que eran necesarias.	Las operaciones presentadas para preprocesar los datos, ninguna de ellas eran necesarias, ni se encuentran justificadas.	No hay ningún tipo de preprocesamiento de datos, y es claro que había pasos previos que se debían realizar.
Implementaciones de los modelos (generación de resultados)	Todos los modelos pedidos en el literal 4 de esta guía fueron implementados correctamente, se	Todos los modelos pedidos en el literal 4 de esta guía fueron implementados; pero algunos de ellos	Todos los modelos pedidos en el literal 3 de esta guía fueron implementados; pero los experimentos	Algunos modelos no fueron implementados, o no fueron bien implementados. Sin embargo, se generaron	Los modelos implementados presentan errores, no se ajustaron hiperparámetros y	No implementaron ningún modelo

	ajustaron hiperparámetros, se ejecutaron los modelos y se generaron suficientes resultados siguiendo una metodología adecuada para tal fin como los es GridSearchCV.	presentan algún tipo de error, o faltó rigurosidad en el ajuste de hiperparámetros o en la generación de resultados utilizando una metodología adecuada como lo es GridSearchCV.	realizados para ajustar los hiperparámetros fue poco rigurosa, y no se siguió ninguna metodología adecuada tal como GridSearchCV	resultados con ellos ajustando hiperparámetros y utilizando GridSearchCV.	tampoco se utilizó GridSearchCV.	
Resultados generados	Se observa el trabajo y el empeño del estudiante al realizar diferentes pruebas y evaluar los modelos con diferentes hiperparámetros, utilizando GridSearchCV.	El estudiante llevó a cabo algunas pruebas de hiperparámetros con los métodos implementados; pero faltó más experimentación, pudo haber mejorado sus resultados un poco más.	Hay algunas pruebas realizadas sin ninguna secuencia lógica, se cumple con mostrar que evaluó e hizo pruebas con sus modelos; pero faltó rigurosidad y empeño para mejorar la precisión de los modelos.	Las pruebas realizadas fueron muy escasas, solo se evaluaron y probaron algunos modelos y otros no. Faltó trabajo de experimentación y prueba para ajustar los hiperparámetros de los modelos.	Reportaron resultado con cualquier combinación inicial de hiperparámetros; y no se evidencia que hayan explorado otras combinaciones.	No se hizo el trabajo
Análisis de resultados	Los resultados obtenidos están explicados y analizados, de tal forma que se nota una secuencia de decisiones en los experimentos y pruebas realizados. Además, se presenta un análisis claro comparando los métodos. Estos resultados están bien resumidos y tablas y gráficas.	Los resultados obtenidos están explicados y analizados, de tal forma que se nota una secuencia de decisiones en los experimentos y pruebas realizados. Sin embargo, no se presenta un análisis claro de estos resultados. Se presenta un análisis comparativo poco elaborado.	Los resultados obtenidos están explicados y analizados, pero no se nota una secuencia de decisiones en los experimentos y pruebas realizados. Se presenta un análisis comparativo poco elaborado.	Algunos experimentos y resultados obtenidos fueron explicados; pero hay otros resultados que simplemente se mostraron sin explicación, ni análisis. Se presenta un análisis comparativo poco elaborado.	Se presentan muy pocos o ningún resultado, para algunos modelos; no hay ningún análisis, solo se colocaron valores, y matrices de confusión, y/o gráficas, y/o tablas sin explicación. Se presenta un análisis comparativo poco elaborado.	No se presentó ningún análisis de resultados
Conclusiones	Se presenta unas conclusiones claras, bien redactadas, en las que se tiene en cuenta el objetivo del conjunto de datos, y se hace una reflexión de	Se presenta unas conclusiones, en las que se tiene en cuenta el objetivo del conjunto de datos, y se hace una reflexión de posible trabajo	Se presenta conclusiones con respecto al objetivo del conjunto de datos; pero no una reflexión sobre trabajo futuro que pueda conllevar a un	Se presenta conclusiones con respecto al objetivo del conjunto de datos; pero no una reflexión sobre trabajo futuro que pueda conllevar a un	Las conclusiones presentadas no tienen nada que ver con lo solicitado en la guía para este ítem.	No se presentaron conclusiones

	posible trabajo futuro para mejorar el desempeño de los modelos obtenidos, y del trabajo realizado. Se utilizaron términos técnicos apropiados.	futuro para mejorar el desempeño de los modelos obtenidos, y del trabajo realizado. Sin embargo, hay algunas fallas en los términos técnicos apropiados.	mejor desempeño de los modelos. Los términos técnicos utilizados fueron apropiados.	mejor desempeño de los modelos. También, hay algunas fallas en los términos técnicos apropiados.		
Presentación final del Notebook	El Notebook está bien organizado en secciones, con suficiente texto explicando el problema, y el tratamiento que se le dio al conjunto de datos asignado, así como también las conclusiones.	El Notebook está bien organizado en secciones, pero algunas secciones no contienen suficiente texto explicando el tratamiento que se le dio al conjunto de datos asignado.	El Notebook está bien organizado en secciones, pero algunas secciones no están documentadas.	El Notebook está parcialmente organizado en secciones, hay algunas partes donde se pierde la secuencia lógica del trabajo y falta documentación.	El Notebook no está bien organizado en secciones, es difícil de seguir una secuencia clara y lógica del trabajo realizado.	No entregaron el Notebook

Rubrica ABET para el grupo 3:

FACULTY OF ENGINEERING						
SYSTEM ENGINEERING PROGRAM						
STUDENT OUTCOMES PERFORMANCE INDICATORS RUBRIC						
SO6: Ability to develop and perform appropriate experiments, analyze and interpret data, and use engineering judgment to draw conclusions.	QUALITATIVE ASSESSMENT					
	NON-COMPLIANCE	BASIC	SATISFACTORY	OUTSTANDING		
PERFORMANCE INDICATOR		0% - 59%	60% - 69%	70% - 89%	90% - 100%	EVIDENCES
PI - 61	Develop and carry out experiments with rigor and order in execution, precision and security in data collection.	The student does not organize activities, make accurate measurements, or report reliable data on results.	The student organizes activities, takes measurements and reports the minimum data required.	The student documents, structures an advance test plan with the required elements, organizes the activities, completes the planning, collects, and reports reliable data.	The student identifies the needs for the test plan in advance with the required elements, establishes a method to define the metrics, organizes activities by establishing an effective sequencing, complies with the planning, collects the results of interest and reports reliable data.	Puntos 3 y 4 de la guía general del curso
PI - 62	Analyze the results obtained and interpret them taking into account theoretical and technical elements.	The student does not analyze or interpret the results obtained.	The student analyzes data and interprets minimum required results.	The student analyzes and interprets the results with technical and theoretical criteria in accordance with the test plan, integrating methodologies for data evaluation.	The student refines, analyzes and interprets the results based on technical and theoretical criteria in accordance with the test plan, integrating methodologies for data evaluation.	Puntos 5 y 6 de la guía general del curso

PI - 63	Integrates data and draws conclusions based on engineering judgment.	The student draws conclusions in an ambiguous and imprecise manner; and they are not consistent with the data obtained through the experimentation process.	The conclusions are written in a precise manner with minimal integration of the data obtained.	The student systematically integrates data and makes value judgments correlated with the data analysis; excluding atypical data from the experiment, which is effectively documented.	The student draws concrete and precise conclusions; argues the limitations of the experiment and their impact on the results obtained. He/she also raises discussions based on the resulting judgments.	Puntos 6 y 7 de la guía general del curso
---------	--	---	--	---	---	---