

**Guía para Taller final de IA**  
**Aprendizaje de máquinas supervisado**  
**Inteligencia Artificial 2022-I**

La actividad consiste en entrenar modelos de aprendizaje supervisado para lo cual cada grupo debe descargar el conjunto de datos asignado. Para cada modelo, se deberán diseñar y llevar a cabo experimentos para ajustar sus parámetros. El resultado de esta actividad es un informe que debe contener el desarrollo de los siguientes ítems:

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. 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.
2. Una descripción de una inspección de los datos incluyendo gráficas y tablas, como matrices de correlación, histogramas, etc.; y de las operaciones de preprocesamiento que sean necesarias para mejorar la calidad de los datos antes de entrenar los distintos modelos.
3. Descripción de los experimentos realizados para ajustar los parámetros cada uno de los métodos vistos en clase. Debe utilizar la función GridSearchCV, la cual además de automatizar las pruebas de múltiples combinaciones de parámetros, también implementa validación cruzada. 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: árboles de decisión, random forest y redes neuronales.
4. En el caso de los modelos de redes neuronales debe implementar modelos secuenciales utilizando las librerías TensorFlow y Keras.
5. Una explicación o análisis del resultado final obtenido por los modelos con los parámetros ajustados debe incluir matrices de confusión en caso de clasificación, o gráficas que ilustren el error de ajuste en caso de un problema de regresión.
6. Conclusiones de 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.

Cada grupo deberá subir a la plataforma un archivo adjunto que contenga, las implementaciones de los distintos modelos entrenados, donde se evidencie los experimentos efectuados para ajuste de parámetros, el conjunto de datos asignado (en caso de que hagan alguna manipulación de los datos originales), y un informe en Word que contenga el desarrollo de los ítems indicados arriba.

Nota: como ejemplo de comparación de modelos, y de algunos elementos de lo solicitado para el informe de este trabajo adjunto el artículo: *A Comparison between Decision Tree and Random Forest in Determining the Risk Factors Associated with Type 2 Diabetes*. Se hará una sesión de asesoría para este trabajo los días 2 y 3 de junio en el horario respectivo a la clase.

Rúbrica de evaluación:

Elementos	Sobresaliente	Destacado	Satisfactorio	Básico	No cumplimiento
Descripción del problema	Se observa que el estudiante documento la problemática utilizando fuentes tanto del repositorio de los datos, como de otras fuentes distintas y hay una buena descripción de las variables.	Se observa que el estudiante documento la problemática utilizando fuentes tanto del repositorio de los datos, como de otras fuentes; sin embargo, pudo ser más elaborado incluyendo la descripción de las variables.	Se hizo una buena descripción del problema y de las variables con base en la información que aporta el repositorio de donde descargo los datos.	La descripción de los datos es pobre, se lee como una traducción literal de la información contenida en la descripción del repositorio.	La descripción es pobre, y no incluye en ella una de los dos el problema en sí, o la descripción de las variables, o ninguna de las dos.
Inspección 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 y descripción.	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 descripción 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 se aplicaron operaciones de preprocesamiento que eran necesarias.	No hay gráficas suficientes para observar el comportamiento de las variables y tomar decisiones con respecto a ellas.
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 en la inspección 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.	No se aplicó ninguna operación de preprocesamiento necesarias.
Implementaciones de los modelos	Todos los modelos fueron implementados, incluyendo los modelos de redes neuronales secuenciales con TensorFlow y Keras. El Notebook desarrollado es claro,	Todos los modelos fueron implementados, incluyendo los modelos de redes neuronales secuenciales con TensorFlow y Keras. El Notebook desarrollado contiene todo lo pedido; pero le faltó documentación para	Todos los modelos fueron implementados, incluyendo los modelos de redes neuronales secuenciales con TensorFlow y Keras. El Notebook es difícil de seguir y le faltó documentación.	Algunos modelos no fueron implementados.	No se entregaron las implementaciones de los modelos.

	y fácil de seguir, tiene documentación.	poder seguirlo fácilmente.			
Experimentación y prueba	Se observa el trabajo y el empeño del estudiante al realizar diferentes pruebas y evaluar los modelos con diferentes parámetros, utilizando validación cruzada.	El estudiante llevó a cabo algunas pruebas de parámetros 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 métodos y otros no. Faltó trabajo de experimentación y prueba para ajustar los parámetros de los modelos.	No hicieron ninguna experimentación, ni ajuste de parámetros de los modelos implementados.
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.	Los resultados obtenidos están descritos; pero falta análisis de lo observado, no se observa una secuencia en las decisiones que se tomaron al realizar los experimentos producto del análisis de los resultados.	Algunos experimentos y resultados obtenidos fueron explicados; pero hay otros resultados que simplemente se mostraron sin explicación, ni análisis.	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.	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 posible trabajo futuro para mejorar el desempeño de los modelos obtenidos, y del trabajo realizado. Se utilizaron términos técnicos apropiados.	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 futuro para mejorar el desempeño de los modelos obtenidos, y del trabajo realizado. Sin embargo, hay algunas fallas en los términos técnicos apropiados.	Se presenta conclusiones con respecto al objetivo del conjunto de datos; pero no una reflexión sobre trabajo futuro que pueda conllevar a un mejor desempeño de los modelos. Los términos técnicos utilizados fueron apropiados.	Se presenta conclusiones con respecto al objetivo del conjunto de datos; pero no una reflexión sobre trabajo futuro que pueda conllevar a un mejor desempeño de los modelos. También, hay algunas fallas en los términos técnicos apropiados.	Las conclusiones presentadas no tienen nada que ver con lo solicitado en la guía para este ítem. O no se presentaron conclusiones.