

Universidad del Valle  
Escuela de Ingeniería de Sistemas y Computación  
Inteligencia artificial  
Informe sobre *Machine learning*

Para la experimentación con técnicas de *machine learning* se utilizará un conjunto de datos de 27901 estudiantes quienes participaron en un estudio sobre depresión. Dentro de la información recolectada se tienen datos sobre el estilo de vida, hábitos académicos o laborales y salud mental. Cada persona está descrita mediante 9 atributos, entre los que se incluyen la edad, el promedio acumulado (CGPA), la duración del sueño, el tipo de estudios que realiza, la carga horaria de estudio o trabajo, y si ha tenido pensamientos suicidas o antecedentes familiares de enfermedades mentales. La variable dependiente que se quiere predecir es *Depression*, la cual toma los valores 1 o 0 para indicar la presencia o ausencia de síntomas depresivos, respectivamente. En este informe se desarrollarán modelos para predecir dicha variable a partir de los demás atributos.

#	Atributo	Descripción	Tipo de variable	Posibles valores
1	Gender	Sexo de la persona	Catógorica	Female, Male
2	Age	Edad de la persona en años	Numérica	18.0 - 59.0
3	CGPA	Promedio acumulado (Cumulative Grade Point Average)	Numérica	0.0 - 10.0
4	Sleep Duration	Duración habitual del sueño por noche	Catógorica	Less than 5 hours 5-6 hours, 7-8 hours, More than 8 hours Others
5	Degree	Indica el nivel educativo en el que se encuentra actualmente	Catógorica	High school Undergraduate Postgraduate Doctorate Others
6	Suicidal Thoughts	Indica si ha tenido pensamientos suicidas	Catógorica	No, Yes
7	Work/Study Hours	Horas promedio dedicadas a estudio o trabajo por día	Numérica	0.0 - 12.0
8	Family History of Mental Illness	Antecedentes familiares de enfermedades mentales	Catógorica	No, Yes
9	Depression	Etiqueta binaria que indica presencia de síntomas depresivos	Variable dependiente a predecir	0 = No tiene depresión 1 = Sí tiene depresión

El objetivo de este informe es crear dos notebooks. En el primer notebook se usará la técnica de redes neuronales probando diferentes topologías y modificando los hiperparámetros. Para esto, debe entregar un notebook donde se realicen las siguientes tareas:

1. Leer el archivo depression.csv
2. Seleccionar aleatoriamente el 80% del conjunto de datos para entrenar y el 20% restante para las pruebas
3. Utilizar una estrategia para normalizar los datos numéricos y una forma de codificar los atributos categóricos
4. Construir 5 redes neuronales variando en la topología de la red la cantidad de capas ocultas y de neuronas por cada capa oculta. Puede también variar los hiperparámetros solver y la función de activación. En todas las pruebas debe usar un random\_state=123.
5. Incluya en el notebook una tabla con el *accuracy* obtenido sobre el conjunto de prueba para las 5 redes neuronales del punto anterior
6. Indique en el notebook, usando una celda de tipo texto, los hiperparámetros que por el momento le permiten obtener la red con mayor *accuracy*
7. Seleccione uno de los hiperparámetros disponibles en la documentación ([https://scikit-learn.org/stable/modules/generated/sklearn.neural\\_network.MLPClassifier.html](https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html)) que sea diferente al solver, a la función de activación, y al random\_state. Realice dos variaciones en el hiperparámetro seleccionado manteniendo los otros hiperparámetros del punto anterior. Indique el *accuracy* obtenido al modificar el hiperparámetro seleccionado y analice si la red mejora, empeora, o mantiene su exactitud. Incluya en el notebook dicho análisis

En el segundo notebook se deben realizar las siguientes tareas:

1. Leer el archivo depression.csv
2. Seleccionar aleatoriamente el 80% del conjunto de datos para entrenar y el 20% restante para las pruebas
3. Utilizar una estrategia para normalizar los datos numéricos y una forma de codificar los atributos categóricos
4. Configurar los hiperparámetros del árbol de decisión de la siguiente manera: criterion=gini, splitter=best, y random\_state=123. Obtener 10 árboles de decisión que resultan de modificar el hiperparámetro max\_depth desde 1 hasta 10 con incrementos de 1
5. Incluya en el notebook una tabla con el *accuracy* obtenido sobre el conjunto de prueba para los 10 árboles del punto anterior
6. Repita el mismo procedimiento del punto 4 usando como hiperparámetros criterion=entropy, splitter=best, y random\_state=123. Nuevamente debe modificar el hiperparámetro max\_depth desde 1 hasta 10 con incrementos de 1
7. Incluya en el notebook una tabla con el *accuracy* obtenido sobre el conjunto de prueba para los 10 árboles del punto anterior
8. Indique en el notebook los hiperparámetros que por el momento le permiten obtener el árbol con mayor *accuracy*
9. Seleccione uno de los hiperparámetros disponibles en la documentación (<https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html>) que sea diferente al criterion, splitter, max\_depth, y random\_state. Realice dos variaciones en el hiperparámetro seleccionado manteniendo los otros hiperparámetros del punto anterior. Indique el *accuracy* obtenido al modificar el hiperparámetro seleccionado y analice si el árbol de decisión mejora, empeora, o mantiene su exactitud.