****

**Inteligencia Artificial Aplicada a la Ingeniería**

**Trabajo Práctico Individual**

Noviembre | 2023

Cuerpo docente:

Dr. Ing. Enrique Gabriel Baquela

Ing. José Ernesto Valentini

Autores:

Policrite, Franco

Índice

[Introducción 2](#_Toc149926590)

[Desarrollo 2](#_Toc149926591)

[Primera aproximación a los datos: Orange 2](#_Toc149926592)

[Conclusión 2](#_Toc149926593)

# Introducción

En el ámbito de la inteligencia artificial, los árboles de decisión son una herramienta esencial que desempeña un papel importante en la toma de decisiones en diversos campos, incluida la ingeniería industrial. Estos árboles son modelos predictivos que utilizan estructuras similares a árboles para describir y clasificar decisiones y sus posibles resultados.

El poder de los árboles de decisión para manejar conjuntos de datos heterogéneos y variables múltiples es uno de sus atractivos principales. Además, el uso de los árboles de decisión es super frecuente en la inteligencia artificial debido a su facilidad de interpretación y explicabilidad.

Son modelos predictivos que ayudan a las personas a tomar decisiones basándose en condiciones y reglas lógicas. Los puedes representar como un conjunto de decisiones organizadas en una estructura similar en un árbol.

Cada nodo interno en un árbol de decisión representa una prueba sobre un atributo, cada rama representa el resultado de la prueba, y cada hoja representa la etiqueta de clase o el valor de salida. En otras palabras, tomando decisiones a medida que avanzas de la raíz hacia las hojas, se basa en las características de los datos.

Dividir el conjunto de datos en subconjuntos más pequeños en función de las características elegidas es un paso en el proceso de construcción de un árbol de decisión. Hasta que se cumple un criterio, como la pureza de las hojas o la profundidad máxima del árbol, este proceso se repite.

Los árboles de decisión son populares porque pueden manejar conjuntos de datos complejos y son fáciles de entender. Además, se pueden usar para problemas de regresión y clasificación.

En este trabajo práctico, abordaremos el tema de árboles de decisión desde un simple ejemplo obtenido , en el cual mediante un Script que se le solicito a ChatGPT(IA, s.f.), se puede observar e interpretar de mejor forma como es que el algoritmo de árboles de decisión funcionan.

# Desarrollo

En primera parte, se investigo que son los arboles de decisión dentro de la inteligencia artificial. Se visualizaron videos para poder comprender mejor este algoritmo , en el cual todos trabajaban sobre **Scikit-learn** (Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort and Vincent Michel, s.f.),siendo esta una biblioteca de machine learning que trabaja con el lenguaje de programación Python,tambien es una herramienta introductoria y simple para el análisis de datos y modelado predictivo. Scikit-learn proporciona algoritmos fáciles de usar para usuarios principiantes.

Se paso a la segunda parte,en la cual le pedimos a ChatGPT (IA, s.f.) un script de:

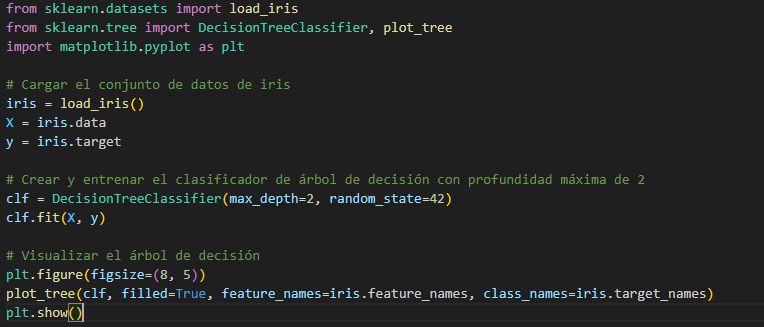
* <= 3 niveles (la raíz (nivel 0), los nodos de decisión en el segundo nivel (nivel 1), los nodos de hoja en el tercer nivel (nivel 2))
* <=3 nodos de decisión.
* Visualizacion Grafica.

Chat Gpt , previo a esto nos solicito tener instalado Python , se procedio a instalar Python.

Luego tambien nos solicito instalar:

* scikit-learn: pip install scikit-learn
* matplotlib (Se utiliza para crear visualizaciones del árbol de decisión):   
  pip install matplotlib

Habiendo instalado todo lo solicitado , ChatGPT (IA, s.f.) nos arrojo el siguiente Script:



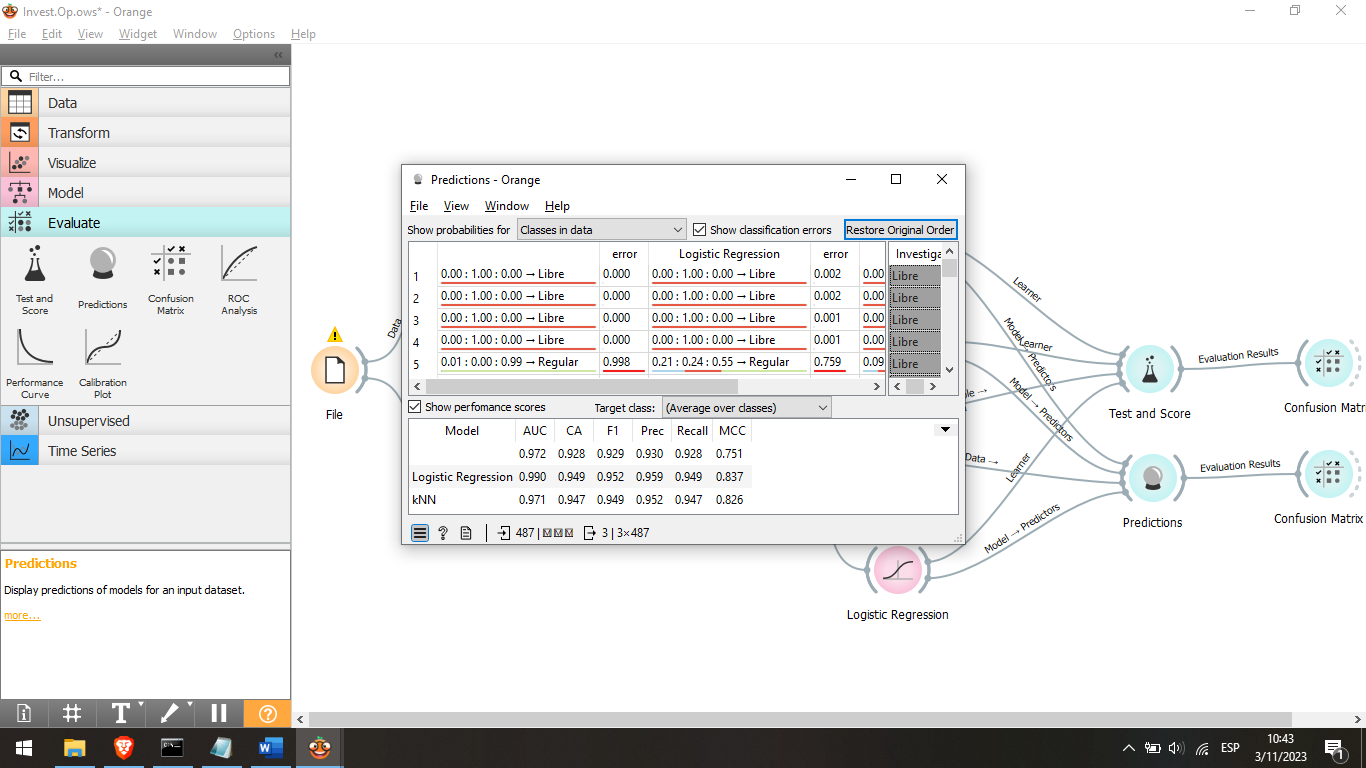
### Primera aproximación a los datos: Orange

En primera instancia, Orange fue de utilidad para intentar realizar los primeros modelos, y a partir de los distintos errores y problemas que fueron surgiendo pudimos modificar y adaptar la base de datos funcionalmente de acuerdo a los requerimientos de la herramienta.

Una vez modelada la base de datos con todos los predictores que consideramos que son de influencia en el desarrollo académico de los alumnos durante el cuarto año de la carrera, decidimos que la variable a predecir sea categórica, teniendo como posibles estados: Ap. Directa, Regular y Libre.

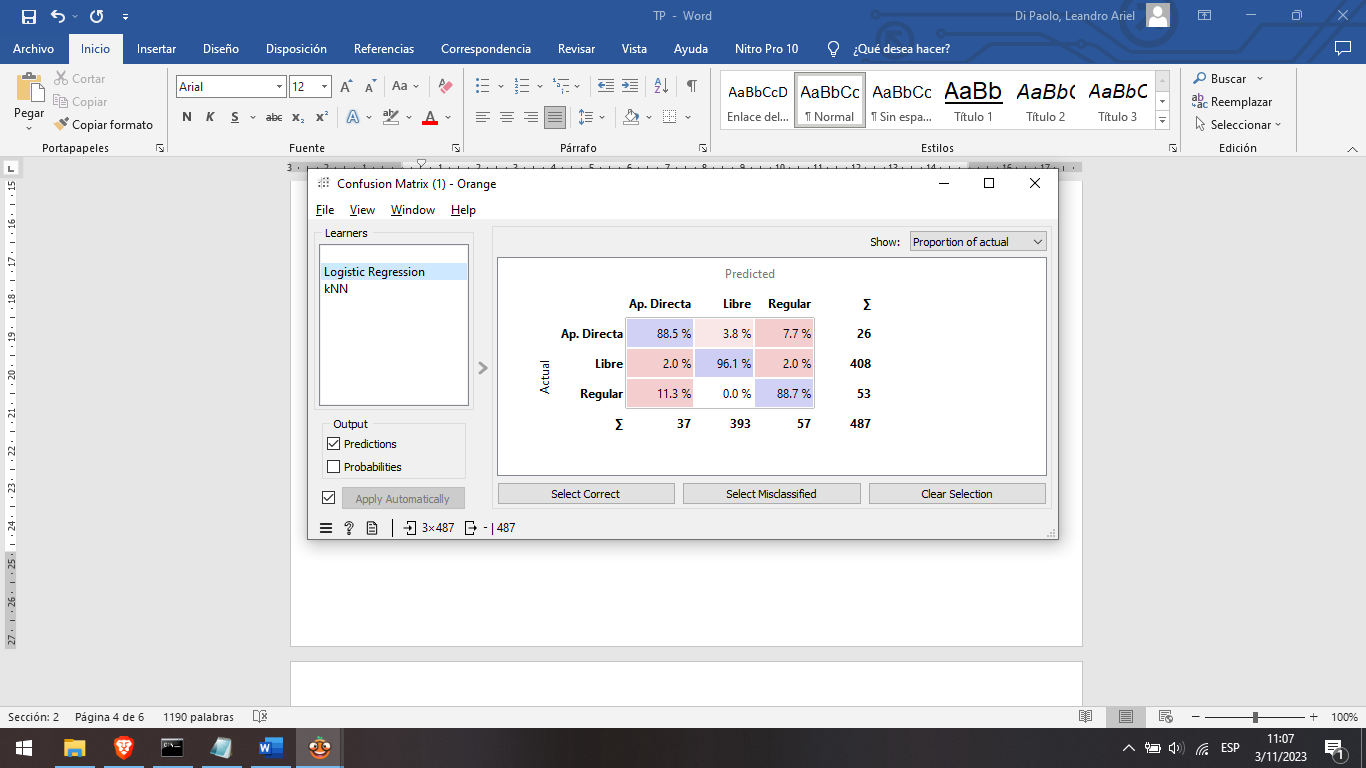
Elegimos utilizar tres modelos distintos, el kNN, el modelo de regresión logística y una red neuronal, para predecir el resultado de los estudiantes en la materia en cuestión. El modelo kNN se basa en la proximidad de vecinos, lo que lo hace efectivo para identificar patrones de similitud entre estudiantes. Por otro lado, la regresión logística es una elección sólida para problemas de clasificación. La red neuronal, en cambio, es capaz de capturar relaciones más complejas y no lineales en los datos, lo que es fundamental para entender el rendimiento de los estudiantes.

Los resultados de Classification Accuracy obtenidos con cada uno de los modelos fueron los siguientes:

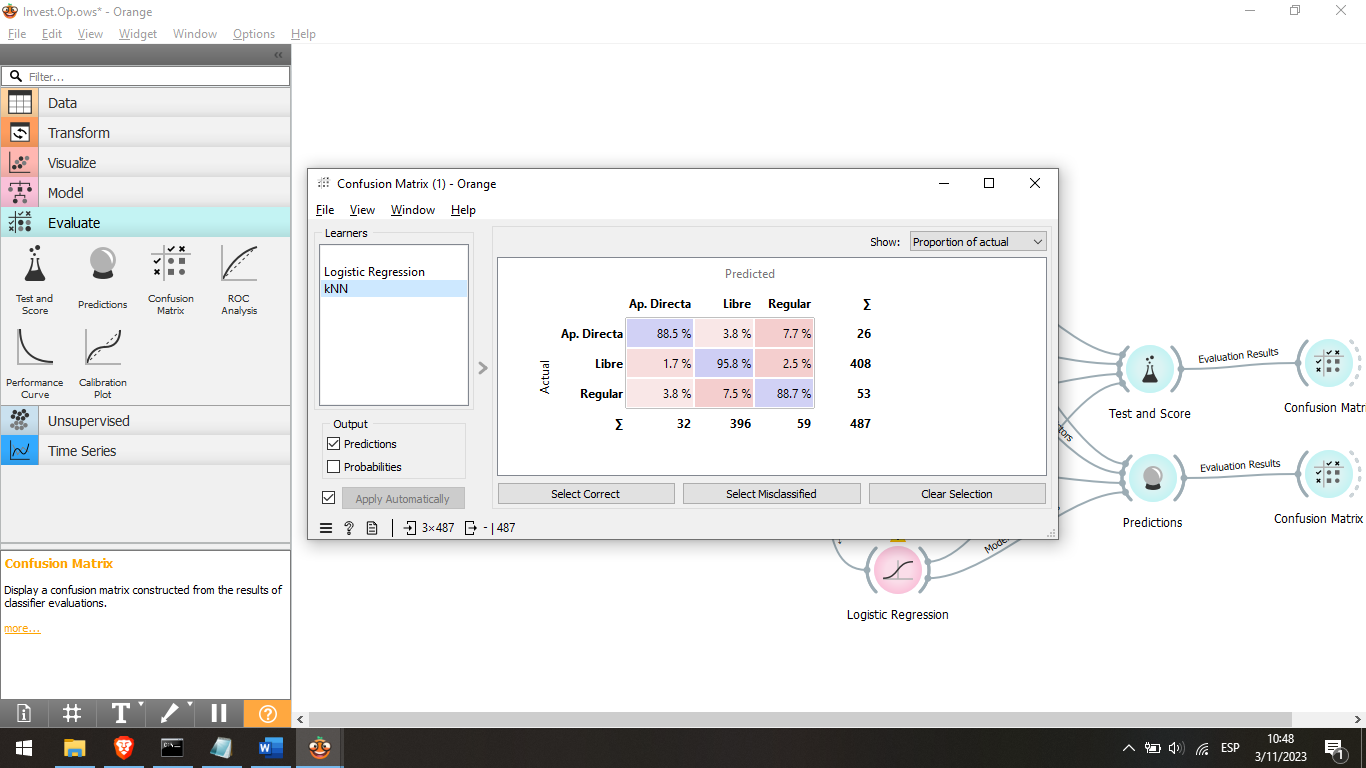


Donde el mejor modelo sería la Regresión Logística con 94,9% de exactitud, seguido por la Regresión logística con 94,7% y por último la Red Neuronal con 92,8%.

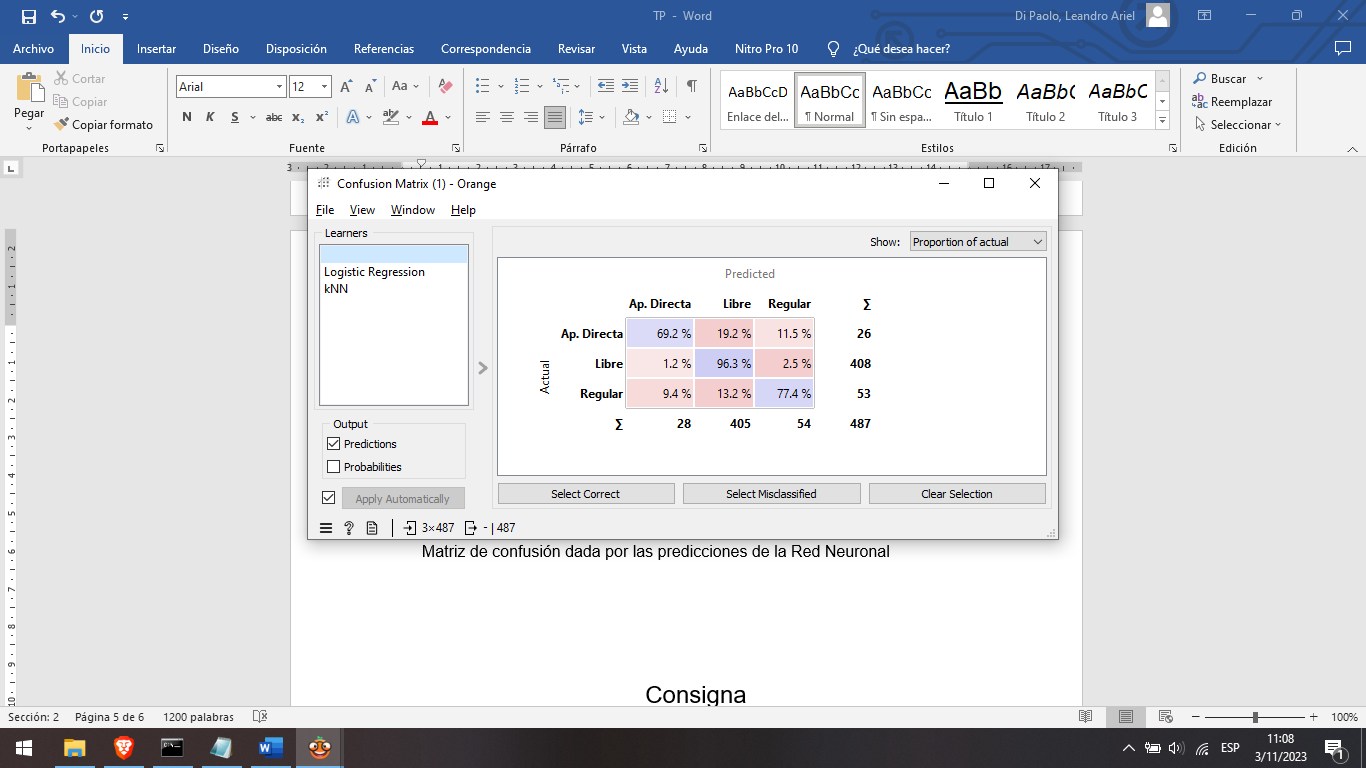
Matriz de confusión dada por las predicciones del modelo de Regresión Logística.



Matriz de confusión dada por las predicciones del modelo kNN



Matriz de confusión dada por las predicciones de la Red Neuronal



Interpretando estos resultados podemos llegar a la conclusión de que la Red Neuronal aplicada fue la menos útil, ya que el tipo de error que tiene en la clasificación de las Ap. Directas es conceptualmente importante, ya que un alumno que tiene altas probabilidades de aprobar directamente la materia puede ser clasificado como si fuera a quedar libre, cuando el error “aceptable” debería ser una clasificación equivocada a la categoría Regular.

Considerando los resultados de los otros dos modelos, se aprecia que no son tan diferentes, sin embargo, es importante resaltar que la Regresión Logística al clasificar los alumnos en estado Regular no va a cometer el error de clasificarlos como Libres.

La ventaja principal que destacamos en Orange es su interfaz gráfica intuitiva, que facilita la creación de flujos de análisis de datos de manera accesible y sin la necesidad de utilizar ningún lenguaje de programación. Además, sobresale en la visualización de resultados, lo que nos ayudó a comprender de manera efectiva el rendimiento de los modelos. Sin embargo, una desventaja relevante que enfrentamos en nuestra experiencia fue su limitación en el rendimiento de conjuntos de datos masivos, ya que, con nuestra base de datos, en muchas ocasiones, experimentamos tiempos de procesamiento considerablemente prolongados.

# Conclusión

En resumen, este trabajo práctico nos llevó a explorar y aplicar herramientas de análisis de datos y aprendizaje automático para entender y predecir el rendimiento de los estudiantes en una materia específica. Este modelado también podría usarse para el resto de las materias de cuarto año cambiando los datos históricos a predecir. Aunque también, adaptando la base de datos serviría para predecir el rendimiento en cualquier materia de cualquier, siempre que se tenga en cuenta el desempeño en las materias anteriormente cursadas.

Sería de gran valor para nosotros poder testear el modelo con los próximos alumnos de Investigación Operativa el próximo año y al finalizar el año lectivo comprobar la eficacia de las predicciones.

Lamentablemente, no pudimos tomarnos el tiempo que era necesario para introducirnos en Julia y poder llevar nuestro análisis a una etapa más profunda. Pero con todo el material brindado queda la puerta abierta para que cada uno investigue y pueda sacar provecho de esta herramienta.