****

**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 Ingeniería Industrial, la toma de decisiones basada en datos se ha convertido en un factor clave para la mejora continua de los procesos y el rendimiento. En este contexto, la inteligencia artificial se destaca como una herramienta poderosa para analizar, procesar y extraer información valiosa a partir de conjuntos de datos complejos. Este trabajo práctico se enfoca en la aplicación de técnicas de IA para analizar datos académicos proporcionados por una universidad, con el propósito de identificar patrones y tendencias que puedan ser fundamentales para la toma de decisiones futuras. A través de la combinación de herramientas, como Orange y el lenguaje de programación Julia, exploraremos cómo la IA puede utilizarse de manera efectiva para transformar datos en conocimiento útil para la comunidad académica.

En el siguiente trabajo, se presentarán los detalles de nuestra aproximación al problema, incluyendo el uso de Orange como herramienta inicial, y el desarrollo de un modelo de IA en un notebook de Julia. Además, se expondrán las ventajas y limitaciones de cada enfoque y se esbozarán recomendaciones para futuras investigaciones y aplicaciones prácticas.

# Desarrollo

Ante la gran cantidad de datos disponibles, optamos por realizar una lluvia de ideas para identificar qué aspectos consideramos particularmente relevantes para nuestro análisis. Llegamos a la conclusión de que sería altamente beneficioso enfocarnos en el análisis de la influencia del historial académico de los primeros tres años en el rendimiento de las materias del cuarto año. Tomando como ejemplo la materia Investigación Operativa, nos propusimos explorar cómo los resultados y el desempeño previo en cursos anteriores podrían afectar el éxito en esta materia específica. Este análisis nos permitirá obtener conocimientos valiosos sobre la correlación entre el pasado académico y el desempeño futuro, lo que podría servir como base para la toma de decisiones educativas estratégicas.

Para alimentar adecuadamente las herramientas de inteligencia artificial que utilizamos en este trabajo práctico, fue necesario realizar modificaciones en las bases de datos originales. Optamos por realizar estas modificaciones en Excel debido a la familiaridad y experiencia que cada uno de nosotros tenía con esta aplicación. Excel nos permitió manipular y ajustar los datos de manera eficiente, lo que resultó en una base de datos más adecuada para su procesamiento en las herramientas de IA.

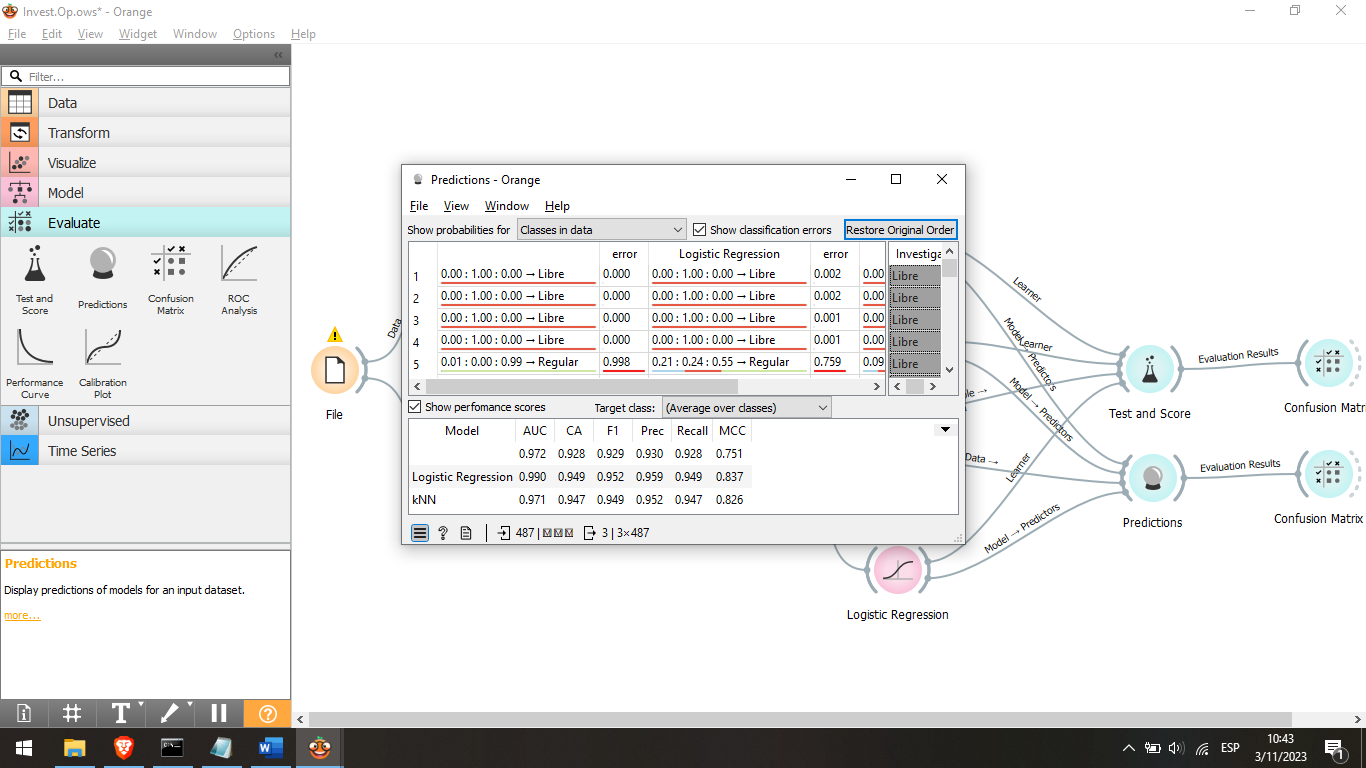
### Primera aproximación a los datos: Orange

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.

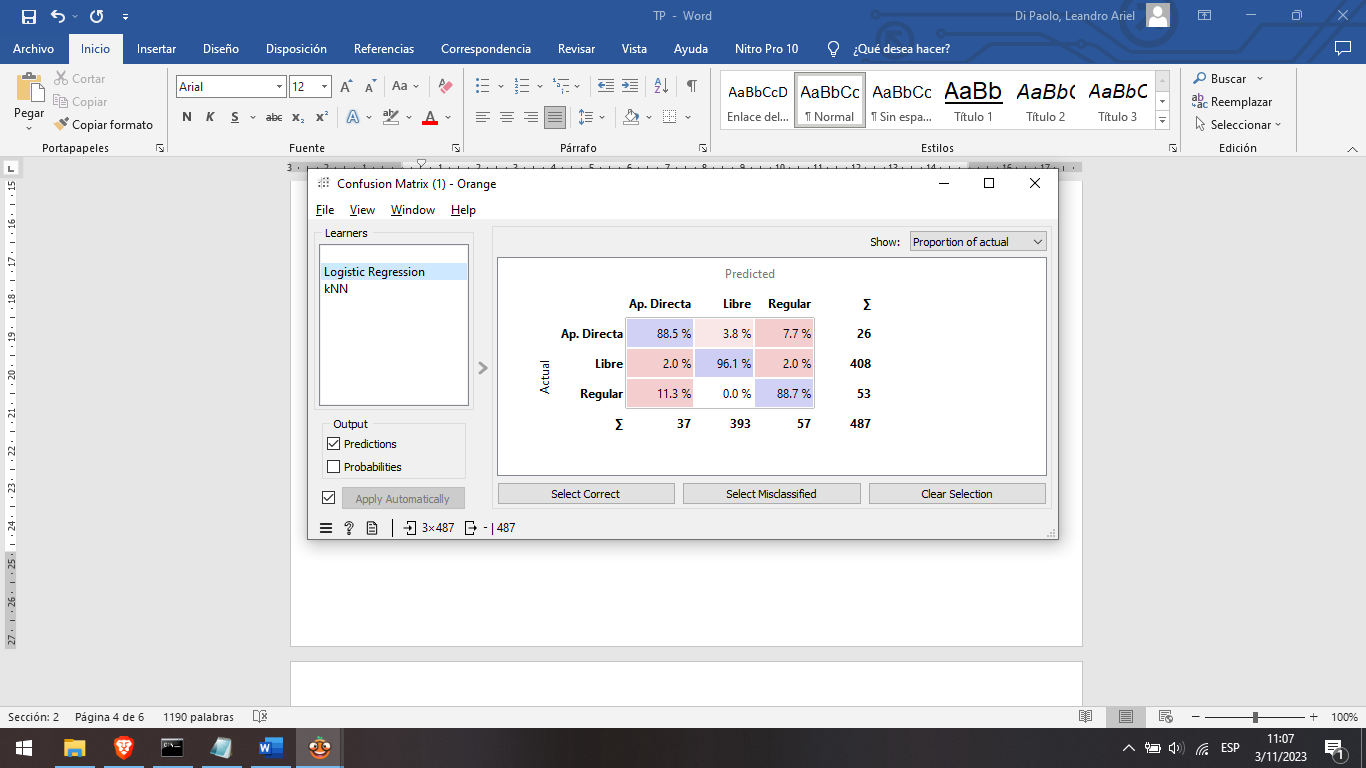
Elegimos utilizar tres modelos distintos, el kNN, el modelo de regresión logística y una red neuronal, para predecir el resultado de los estudiantes en la materia en cuestión. El modelo kNN se basa en la proximidad de vecinos, lo que lo hace efectivo para identificar patrones de similitud entre estudiantes. 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:

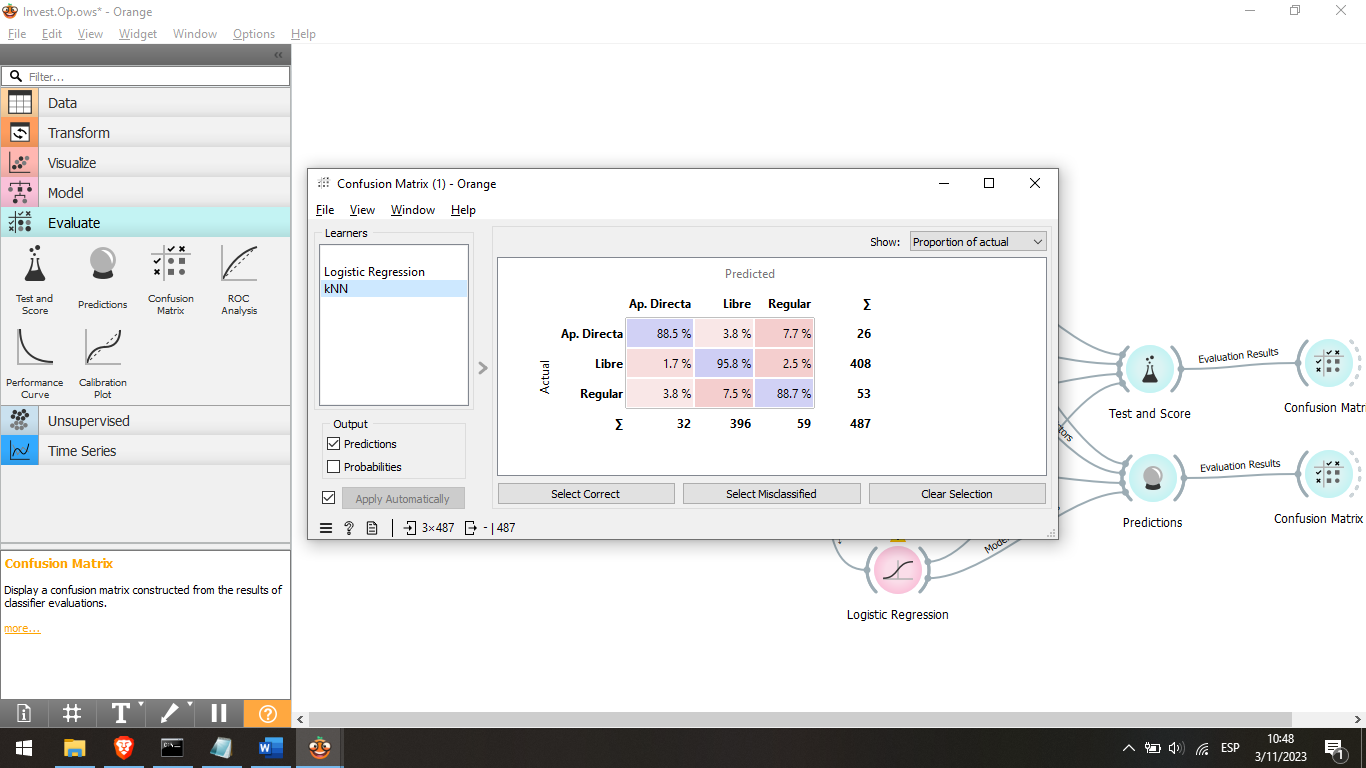


Donde el mejor modelo sería la Regresión Logística con 94,9% de exactitud, seguido por la Regresión logística con 94,7% y por último la Red Neuronal con 92,8%.

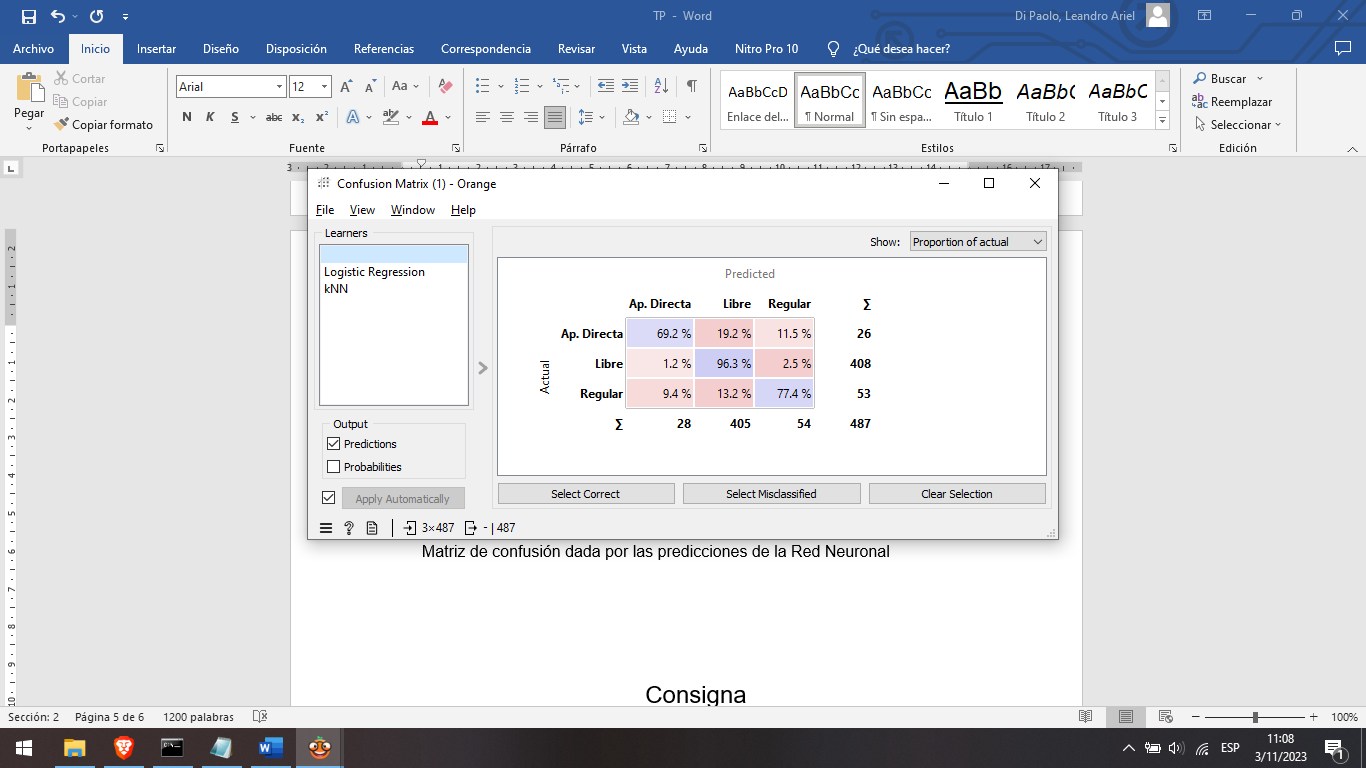
Matriz de confusión dada por las predicciones del modelo de Regresión Logística.



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.