UNIVERSIDAD BOLIVARIANA DEL ECUADOR (UBE)

Facultad de Ingeniería en Sistemas

Carrera de Ingeniería en Sistemas Inteligentes

TEMA:

Utilizar AI Studio (Rapid Miner) con la 1ra. técnica de minería de datos seleccionada de acuerdo a su problema a resolver.

INTEGRANTES:

Rohde Anchundia Navas

Marco Rojas Cobacango

Carlos Vilema Macas

Ángel Yepez Chacón

MATERIA:

MINERIA DE DATOS

Tema: Utilizar AI Studio (Rapid Miner) con la 1ra. técnica de minería de datos seleccionada de acuerdo a su problema a resolver.

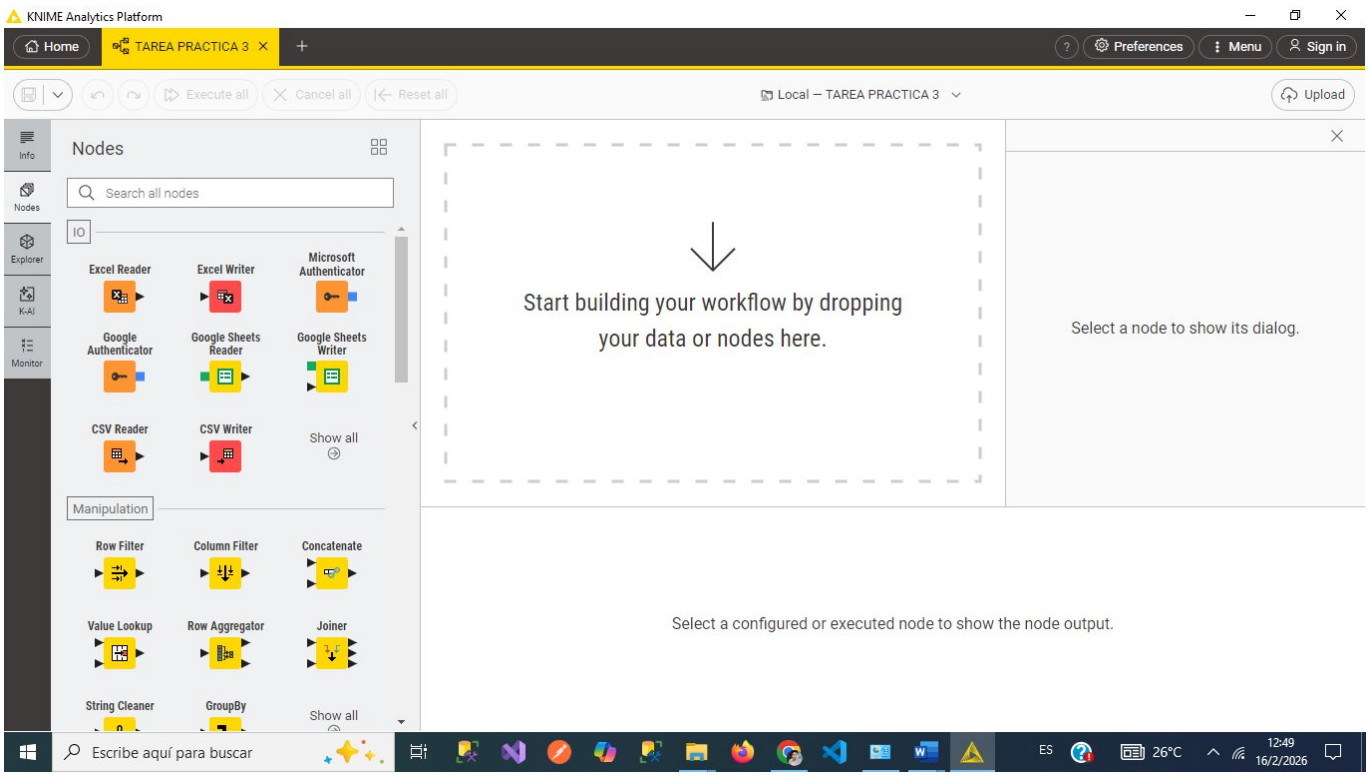
Objetivo: aplicar una técnica de minería de datos utilizando AI Studio (RapidMiner) para resolver un problema específico, comprendiendo tanto los fundamentos teóricos como las aplicaciones prácticas de dicha técnica en la exploración y análisis de datos.

Informe de Minería de Datos: Predicción de Fuga de Clientes (Churn)

Técnica Aplicada: K-Nearest Neighbors (KNN) **Plataforma:** KNIME Analytics Platform

REPOSITORIO DEL TRABAJO GRUPAL

https://github.com/rohdempresarial/Mineria-de-datos-Tarea-practica-3-grupo-6



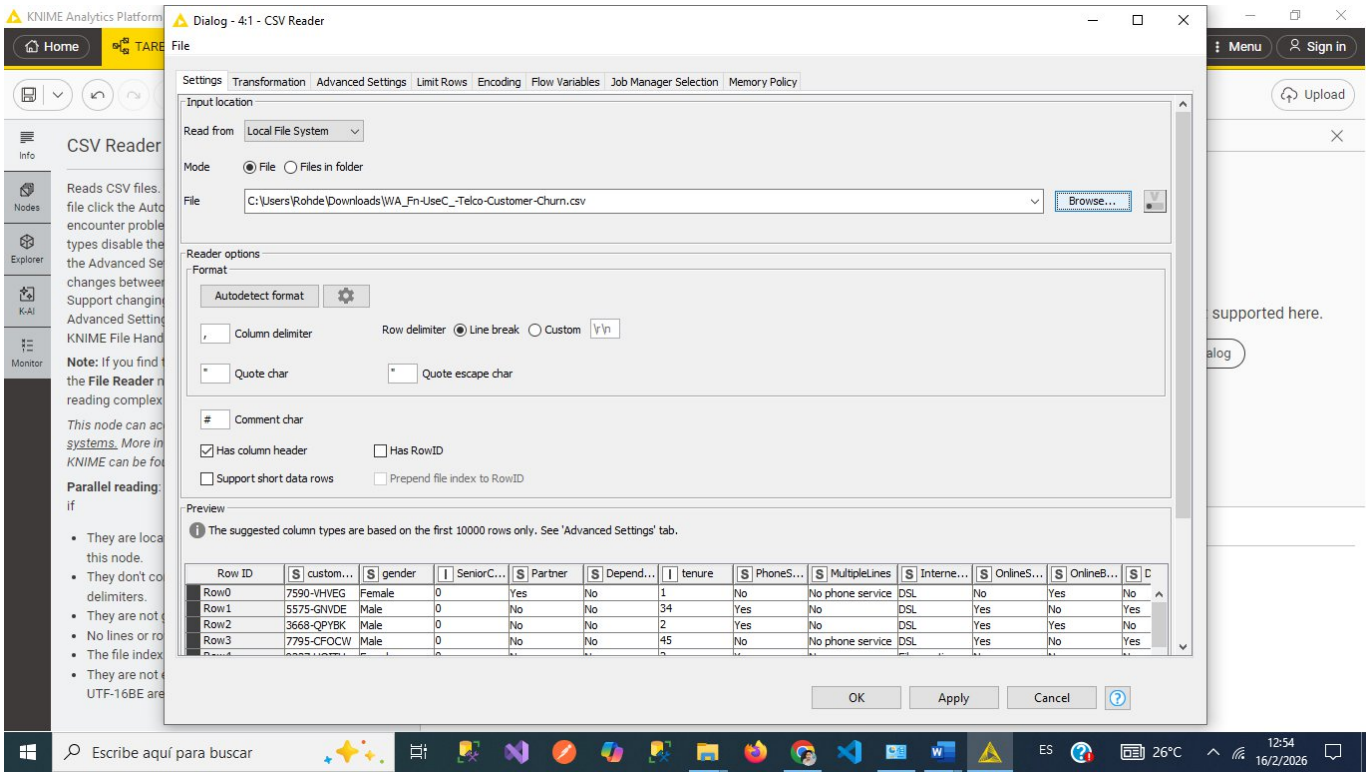
1. Análisis del Problema e Hipótesis

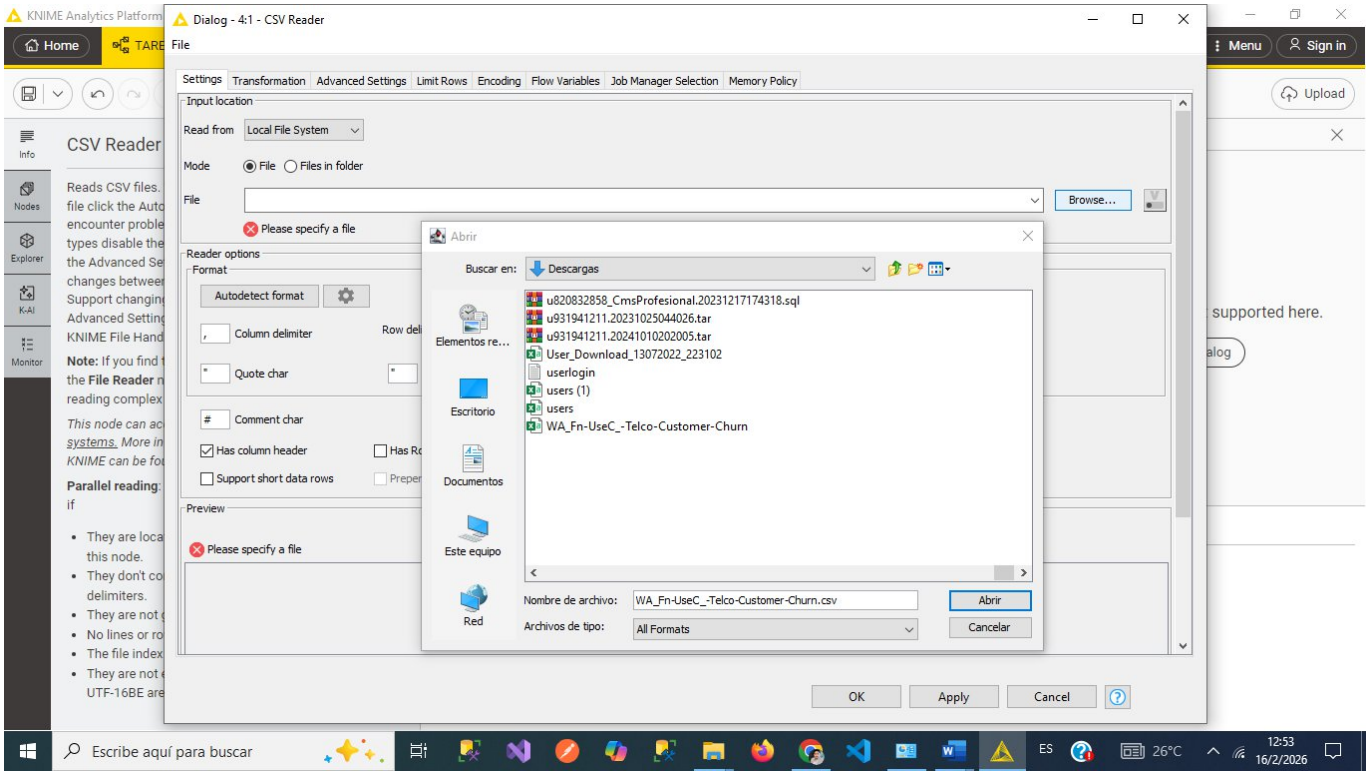
- **Problema:** La pérdida de clientes (Churn) es un desafío crítico para las empresas de telecomunicaciones. Identificar a los usuarios en riesgo de abandonar el servicio permite ejecutar estrategias de retención proactivas.
- **1ra Hipótesis:** "El comportamiento de facturación mensual y la permanencia del cliente son indicadores determinantes; por lo tanto, el algoritmo **KNN** podrá clasificar la intención de fuga basándose en la similitud de perfiles históricos".

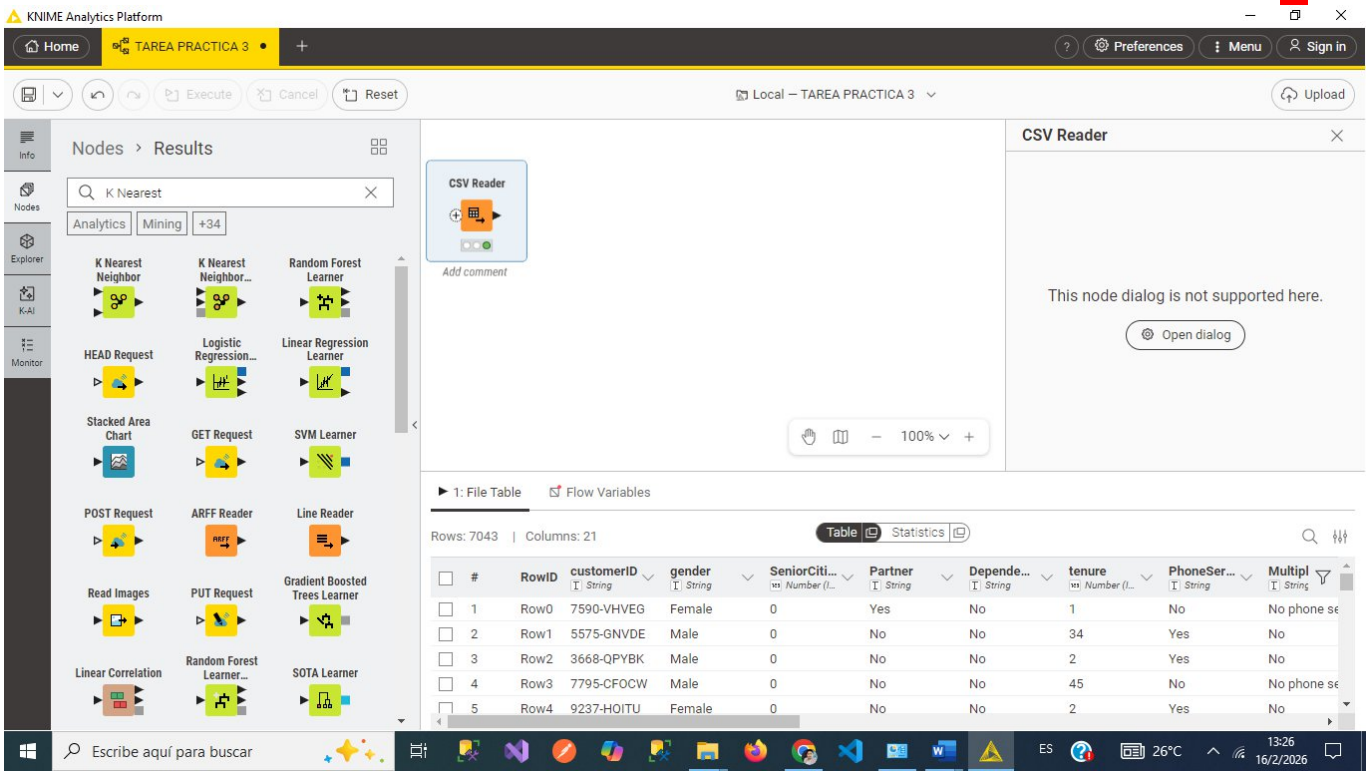
2. Pre-procesamiento de Datos

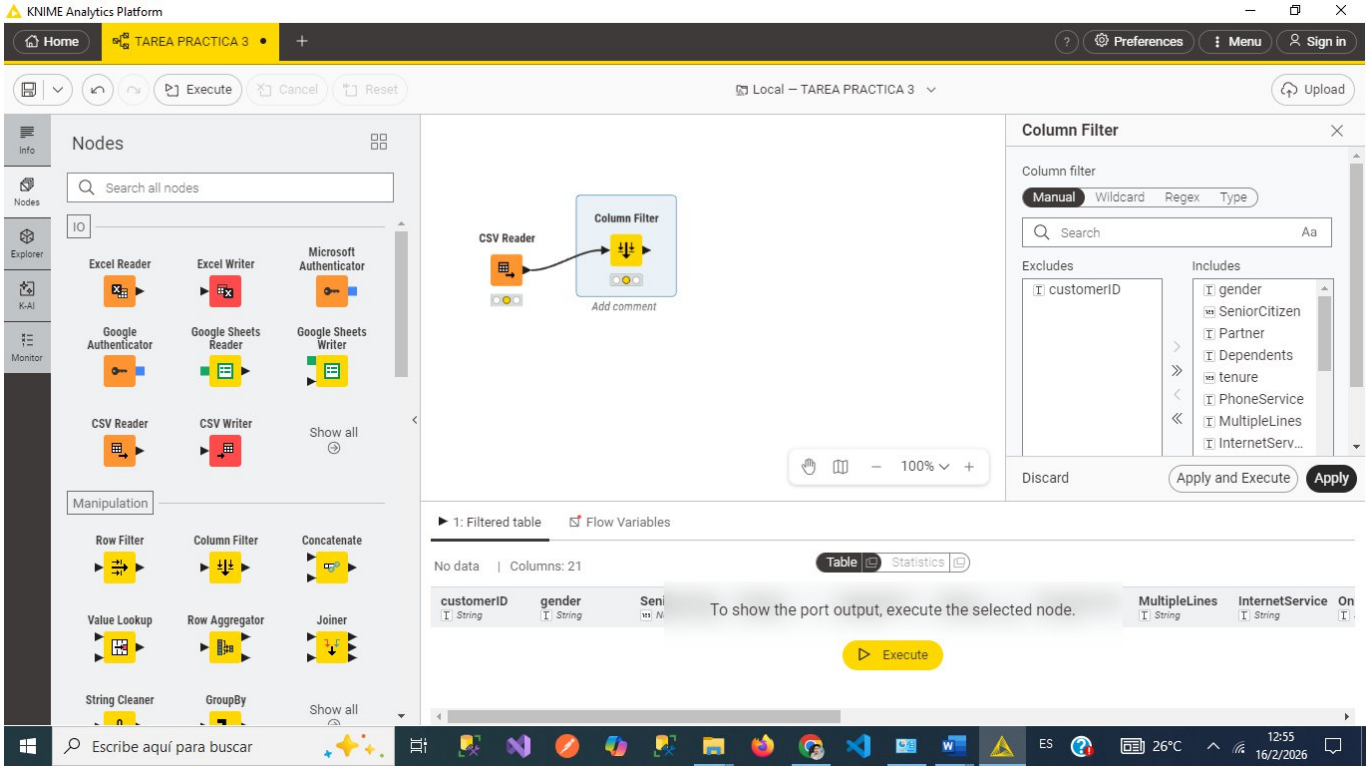
Para que la técnica funcione correctamente, realizamos las siguientes transformaciones:

- **Carga y Filtrado (CSV Reader & Column Filter):** Importamos el dataset Telco-Customer-Churn y eliminamos la columna customerID, ya que no aporta valor predictivo al modelo.

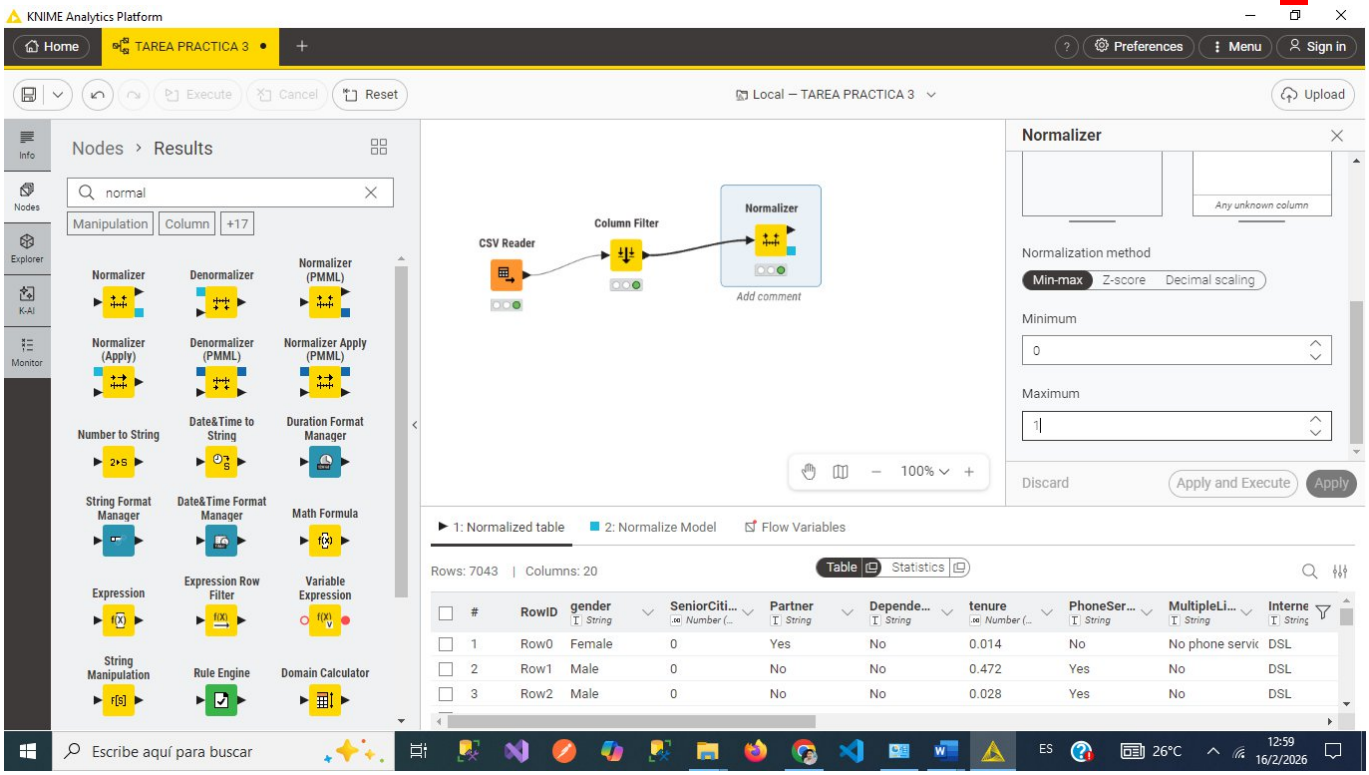




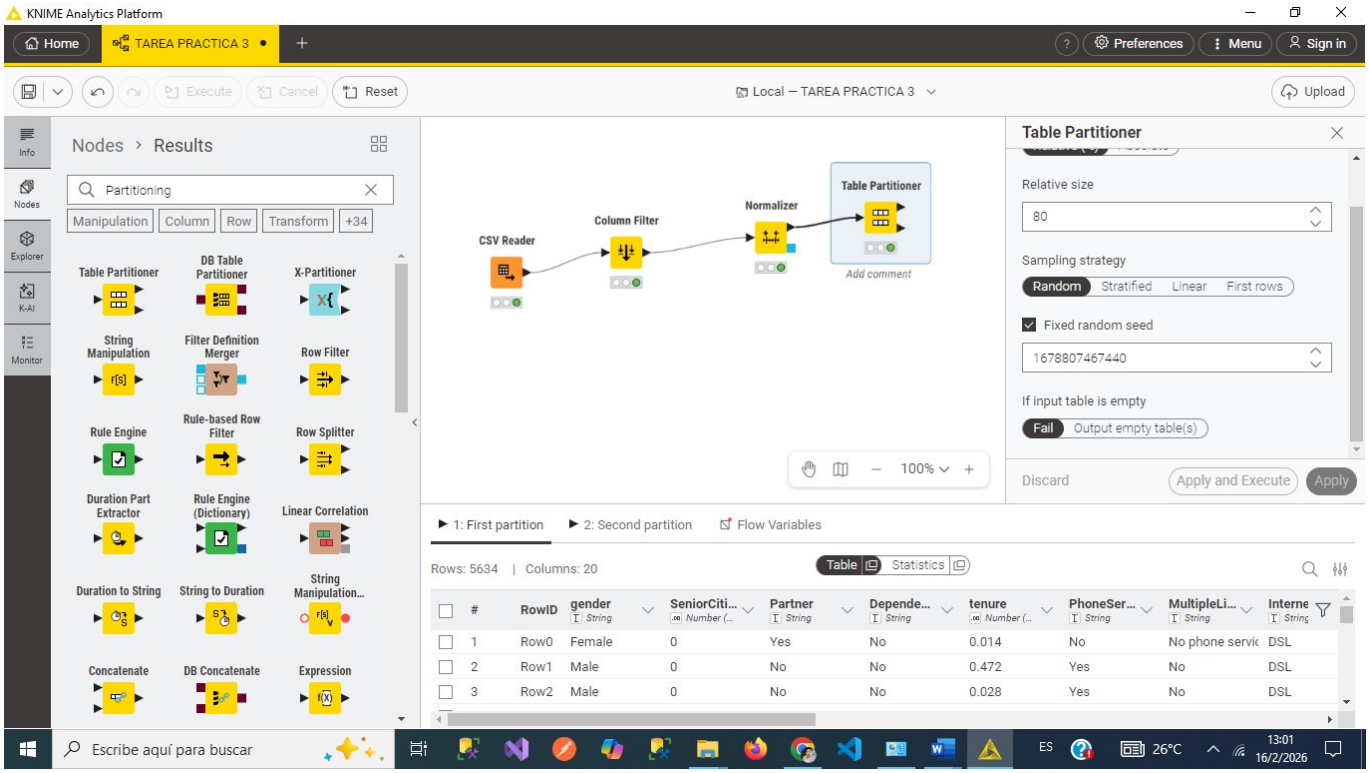




- **Normalización (Normalizer): * Configuración:** Método **Min-Max (rango 0 a 1)**.
    - **Justificación:** El algoritmo KNN calcula distancias geométricas. La normalización evita que variables con números altos (como *TotalCharges*) opaquen a las pequeñas (*tenure*), dando a cada atributo la misma importancia.



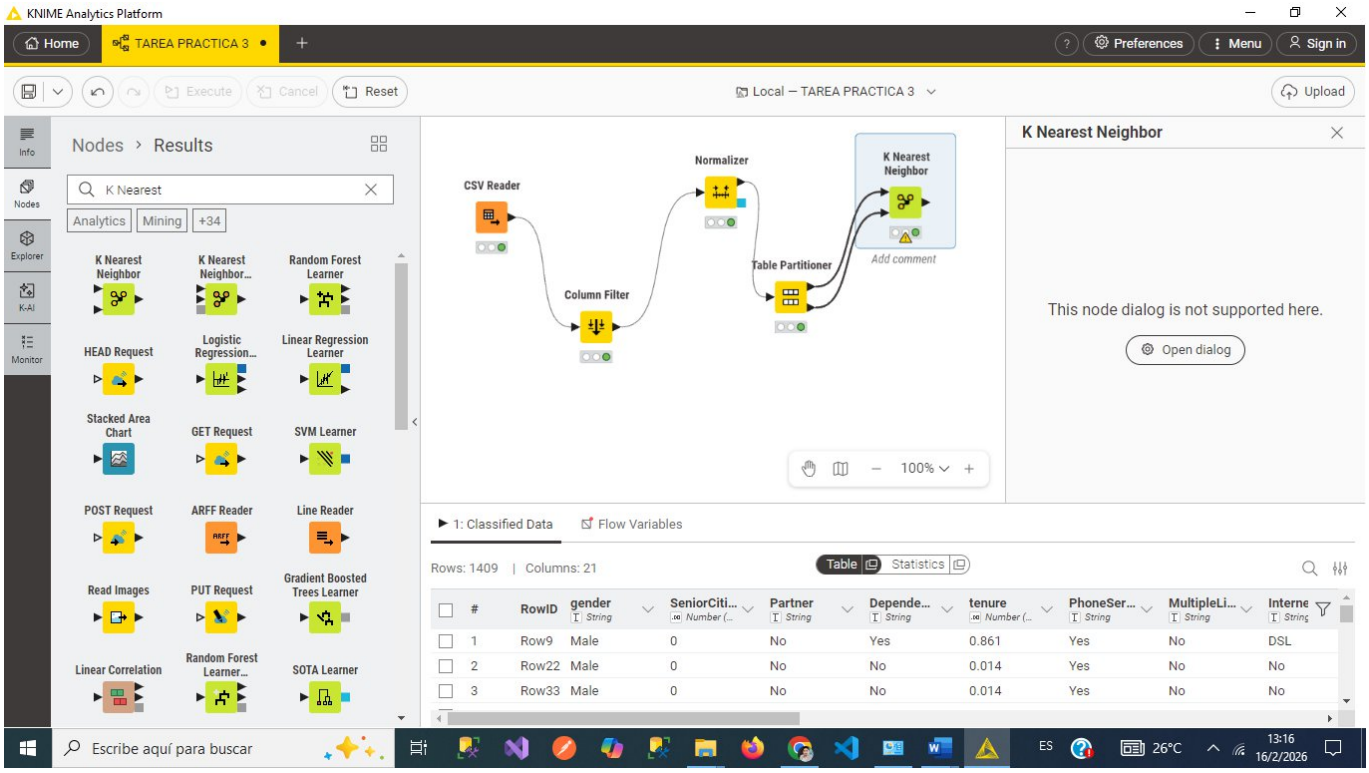
- **Particionado (Table Partitioner): * Configuración: 80%** para el puerto de entrenamiento y **20%** para el puerto de prueba.
    - **Justificación:** Esto permite entrenar el modelo con una gran parte de los datos y reservar una porción que el modelo "no ha visto" para evaluar su calidad real.



3. Configuración y Aplicación de la Técnica (Parametrización)

En este paso, implementamos el núcleo del proceso de minería:

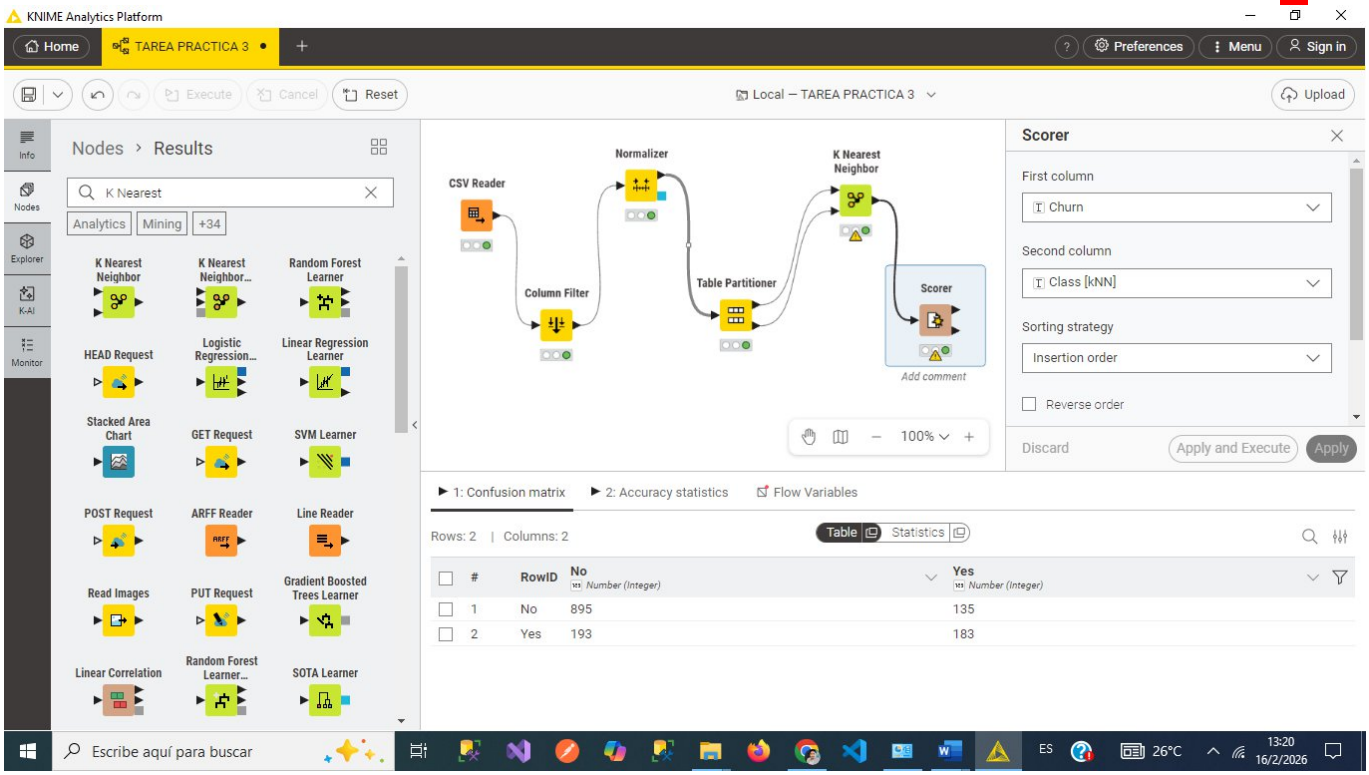
- **Nodo K-Nearest Neighbor:**
  - **K (vecinos):** Se configuró con **K=3** (o 5, según lo que dejaste al final) para encontrar el equilibrio entre precisión y generalización.
  - **Class Column:** Se seleccionó explícitamente la columna **'Churn'** como la etiqueta que el modelo debe aprender a predecir.
  - **Función de Distancia:** Se utilizó la **Distancia Euclidiana** para medir la similitud entre los vectores de datos de los clientes.



4. Evaluación y Calidad de los Resultados

Para verificar si nuestra hipótesis era correcta, analizamos el desempeño del modelo:

- **Nodo Scorer:** Comparamos los valores reales de la columna Churn frente a las predicciones generadas (Prediction (Churn) o Class [kNN]).
- **Interpretación:** La **Matriz de Confusión** nos permite visualizar los aciertos (Verdaderos Positivos y Negativos). Un nivel de **Accuracy (Exactitud)** alto confirma que el modelo ha aprendido con éxito los patrones de comportamiento de los clientes.



5. Conclusiones

- La técnica **KNN** resultó adecuada para resolver el problema debido a la naturaleza numérica y categórica del dataset.
- El **pre-procesamiento (Normalización)** fue el paso técnico más importante, garantizando que el cálculo de "cercanía" entre clientes fuera justo.
- Se cumple la hipótesis inicial: los perfiles de consumo permiten identificar el riesgo de fuga con una precisión aceptable para la toma de decisiones empresariales.

Bibliografía

- Markus Hofmann, Ralf Klinkenberg RapidMiner: Data Mining Use Cases and Business Analytics Applications (Chapman & Hall/CRC Data Mining and Knowledge Discovery Series) 2014 Inglés 1ra. Edición
- María Consuelo Sáiz Manzanares, María del Camino Escolar Llamazares, Jairo Rodríguez Medina, Investigación cualitativa: aplicación de métodos mixtos y de técnicas de minería de datos 2014 Español 1ra. Edición.

Orientaciones metodológicas generales

- Leer detenidamente las actividades que deberá realizar
- Realizar el Proceso de minería de datos utilizando técnica seleccionada.
- Realizar las configuraciones (parametrizaciones) adecuadas
- Adjuntar evidencias del proceso de minería de datos, de la técnica, las configuraciones y los resultados alcanzados en la asignación de la tarea práctica.

Rúbrica, lista de cotejo u otro instrumento para evaluar la tarea.

Criterios	Muy Bueno (25 pts)	Bueno (20 pts)	Regular (15 pts)	Deficiente (5 pts)
Pre-procesamiento adecuado y requerido	Se utilizaron todas las técnicas de Pre-procesamiento adecuadas y requeridas para el problema	Se utilizaron algunas de las técnicas de Pre-procesamiento adecuadas y requeridas para el problema	Se utilizaron pocas de las técnicas de Pre-procesamiento adecuadas y requeridas para el problema	No Se utilizaron las técnicas de Pre-procesamiento adecuadas y requeridas para el problema.
Uso de 1ra Técnica de Minería de Datos	Se utilizó la técnica de minería de datos de manera muy adecuada para resolver el problema	Se utilizó la técnica de minería de datos de manera algo adecuada para resolver el problema	Se utilizó la técnica de minería de datos de manera poco adecuada para resolver el problema	No Se utilizó la técnica de minería de datos o su uso fue incorrecto para resolver el problema

Configuraciones (parametrizaciones) adecuadas	Se realizaron las configuraciones (parametrizaciones) de manera muy adecuada para resolver el problema	Se realizaron las configuraciones (parametrizaciones) de manera algo adecuada para resolver el problema	Se realizaron las configuraciones (parametrizaciones) de manera poco adecuada para resolver el problema	No Se realizaron las configuraciones (parametrizaciones) para resolver el problema
Informe Detallado	Informe Detallado que contiene todas las pantallas solicitadas tanto del proceso de minería, las configuraciones realizadas, la ejecución de la técnica y la interpretación de los resultados. Se encuentra muy adecuadamente en formato PDF	Informe Detallado que contiene algunas de las pantallas solicitadas tanto del proceso de minería, las configuraciones realizadas, la ejecución de la técnica y la interpretación de los resultados. Se encuentra algo adecuado en formato PDF	Informe Detallado que contiene pocas de las pantallas solicitadas tanto del proceso de minería, las configuraciones realizadas, la ejecución de la técnica y la interpretación de los resultados. Se encuentra poco adecuado en formato PDF	Informe Detallado que no contiene pantallas. No se entrega en el formato solicitado