



UNIVERSIDAD
BOLIVARIANA
DEL ECUADOR

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



La Universidad para todos



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

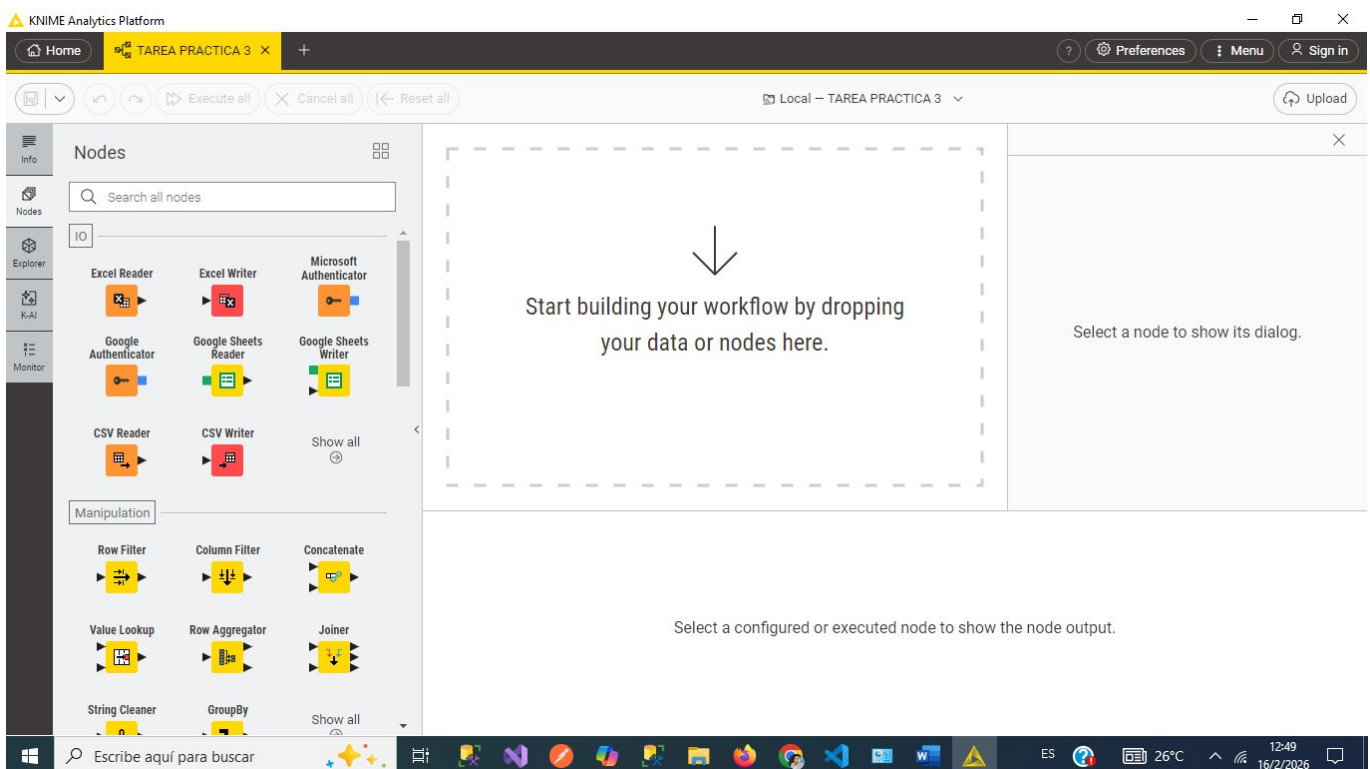
Objetivo: aplicar una técnica de minería de datos utilizando AI Studio (RapidMiner) para resolver un problema específico, comprendiendo tanto los fundamentos teóricos como las aplicaciones prácticas de dicha técnica en la exploración y análisis de datos.

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>



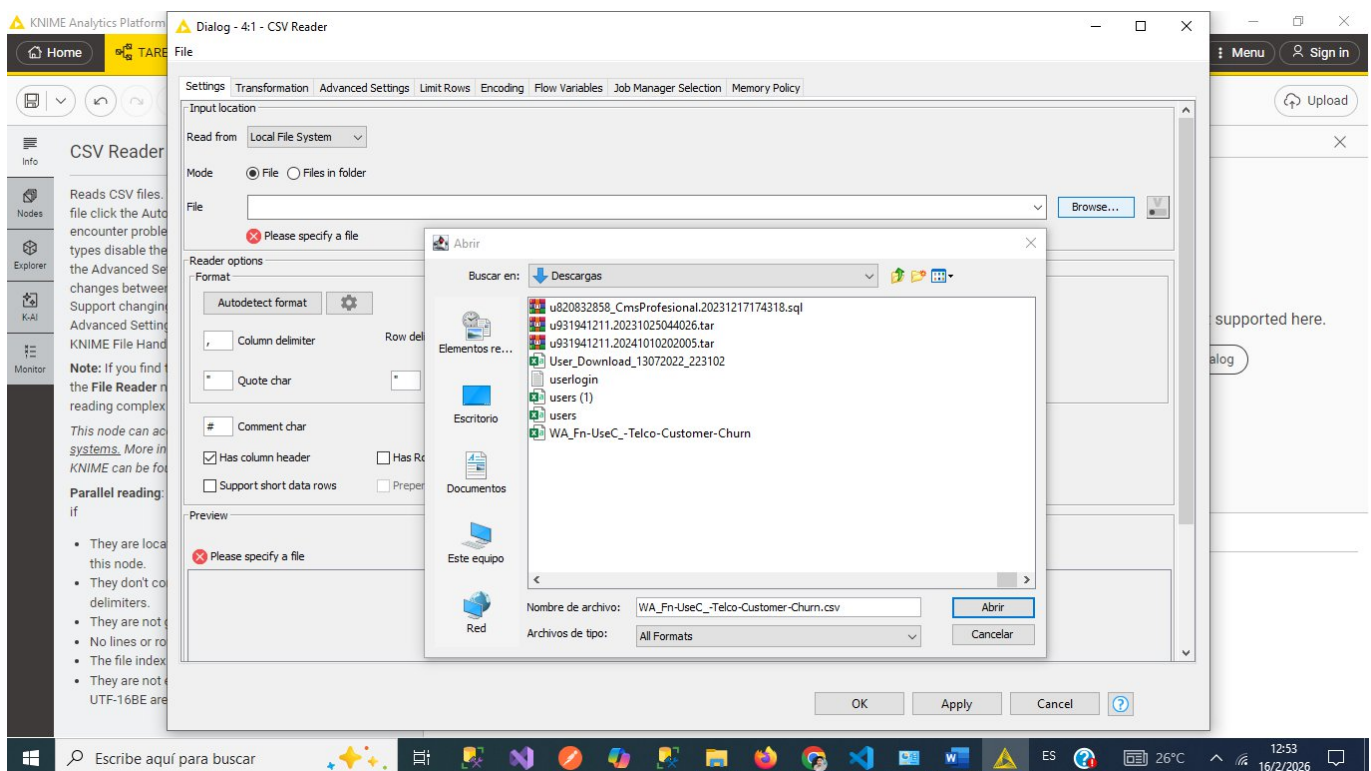
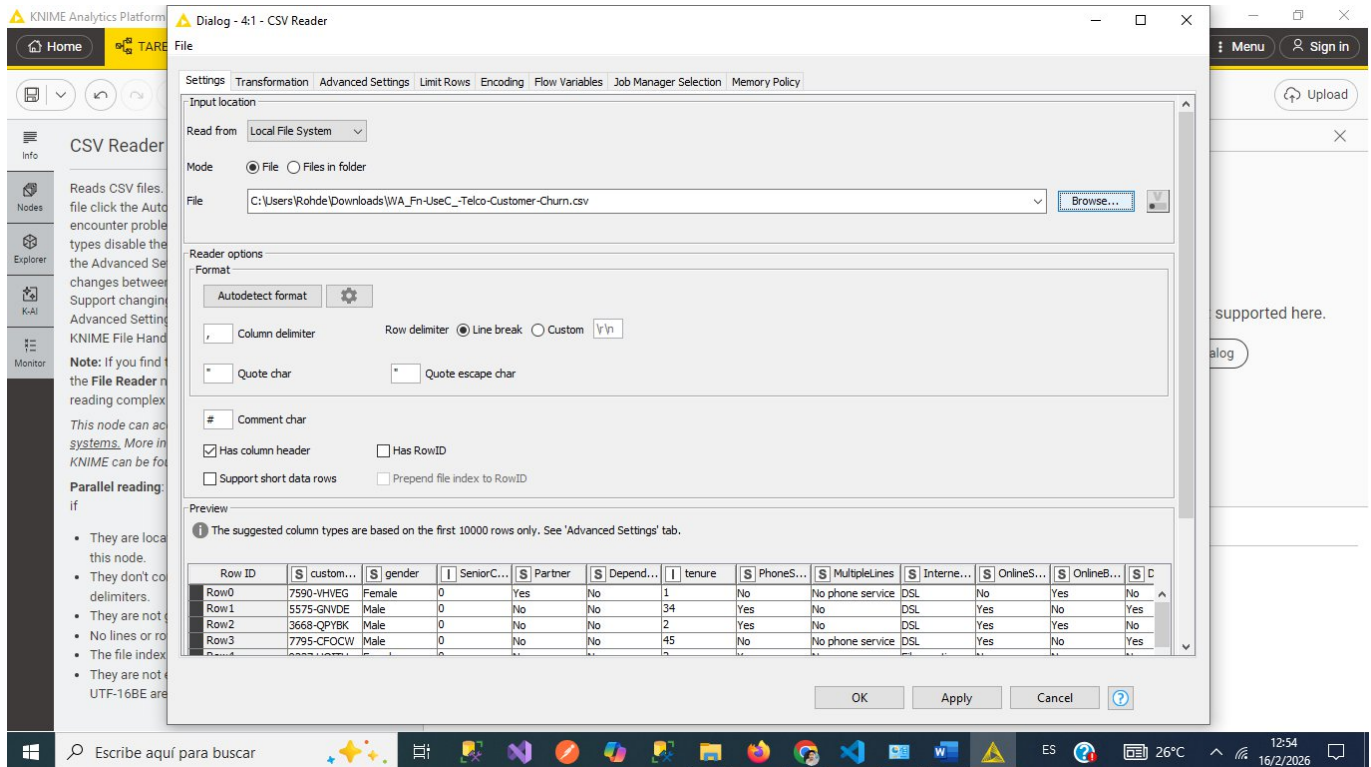
1. Análisis del Problema e Hipótesis

- **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.



KNIME Analytics Platform

Home TAREA PRACTICA 3

Local - TAREA PRACTICA 3

Nodes > Results

Search: K Nearest

Analytics Mining +34

K Nearest Neighbor K Nearest Neighbor... Random Forest Learner

HEAD Request Logistic Regression... Linear Regression Learner

Stacked Area Chart GET Request SVM Learner

POST Request ARFF Reader Line Reader

Read Images PUT Request Gradient Boosted Trees Learner

Linear Correlation Random Forest Learner... SOTA Learner

CSV Reader

This node dialog is not supported here.

Open dialog

1: File Table Flow Variables

Rows: 7043 Columns: 21

#	RowID	customerID	gender	SeniorCiti...	Partner	Depende...	tenure	PhoneSer...	Multipl
1	Row0	7590-VHVEG	Female	0	Yes	No	1	No	No phone se
2	Row1	5575-GNVDE	Male	0	No	No	34	Yes	No
3	Row2	3668-QPYBK	Male	0	No	No	2	Yes	No
4	Row3	7795-CFOCW	Male	0	No	No	45	No	No phone se
5	Row4	9237-HOITU	Female	0	No	No	2	Yes	No

KNIME Analytics Platform

Home TAREA PRACTICA 3

Local - TAREA PRACTICA 3

Nodes

Search: Search all nodes

Excel Reader Excel Writer Microsoft Authenticator

Google Authenticator Google Sheets Reader Google Sheets Writer

CSV Reader CSV Writer Show all

Manipulation

Row Filter Column Filter Concatenate

Value Lookup Row Aggregator Joiner

String Cleaner GroupBy Show all

CSV Reader

Column Filter

Column filter

Manual Wildcard Regex Type

Search

Excludes Includes

customerID gender SeniorCitizen Partner Dependents tenure PhoneService MultipleLines InternetServ...

Discard Apply and Execute Apply

1: Filtered table Flow Variables

No data Columns: 21

customerID	gender	Seni
String	String	N

To show the port output, execute the selected node.

Execute

- **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.

KNIME Analytics Platform

Home TAREA PRACTICA 3

Nodes > Results

Search: normal

Manipulation Column +17

Workflow: CSV Reader → Column Filter → Normalizer

Normalizer Configuration:

- Normalization method: Min-max
- Minimum: 0
- Maximum: 1

Data Table (Rows: 7043 | Columns: 20):

#	RowID	gender	SeniorCiti...	Partner	Depende...	tenure	PhoneSer...	MultipleLi...	Interne
1	Row0	Female	0	Yes	No	0.014	No	No phone serv	DSL
2	Row1	Male	0	No	No	0.472	Yes	No	DSL
3	Row2	Male	0	No	No	0.028	Yes	No	DSL

- **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.

KNIME Analytics Platform

Home TAREA PRACTICA 3

Nodes > Results

Search: Partitioning

Manipulation Column Row Transform +34

Workflow: CSV Reader → Column Filter → Normalizer → Table Partitioner

Table Partitioner Configuration:

- Relative size: 80
- Sampling strategy: Random
- Fixed random seed: ☒ 1678807467440

Data Table (Rows: 5634 | Columns: 20):

#	RowID	gender	SeniorCiti...	Partner	Depende...	tenure	PhoneSer...	MultipleLi...	Interne
1	Row0	Female	0	Yes	No	0.014	No	No phone serv	DSL
2	Row1	Male	0	No	No	0.472	Yes	No	DSL
3	Row2	Male	0	No	No	0.028	Yes	No	DSL

3. Configuración y Aplicación de la Técnica (Parametrización)

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



La Universidad para todos

- **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.

1: Classified Data

#	RowID	gender	SeniorCit...	Partner	Depende...	tenure	PhoneSer...	MultipleLi...	Interne
1	Row9	Male	0	No	Yes	0.861	Yes	No	DSL
2	Row22	Male	0	No	No	0.014	Yes	No	No
3	Row33	Male	0	No	No	0.014	Yes	No	No

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.

KNIME Analytics Platform

Home TAREA PRACTICA 3 +

Execute Cancel Reset

Local - TAREA PRACTICA 3 Upload

Nodes > Results

Search: K Nearest

Analytics Mining +34

Nodes Explorer Monitor

K Nearest Neighbor K Nearest Neighbor... Random Forest Learner

HEAD Request Logistic Regression... Linear Regression Learner

Stacked Area Chart GET Request SVM Learner

POST Request ARFF Reader Line Reader

Read Images PUT Request Gradient Boosted Trees Learner

Linear Correlation Random Forest Learner... SOTA Learner

CSV Reader Normalizer K Nearest Neighbor

Column Filter Table Partitioner Scorer

Scorer

First column: Churn

Second column: Class [kNN]

Sorting strategy: Insertion order

Reverse order: ☐

Discard Apply and Execute Apply

1: Confusion matrix 2: Accuracy statistics Flow Variables

Rows: 2 Columns: 2

#	RowID	No	Yes
		Number (Integer)	Number (Integer)
1	No	895	135
2	Yes	193	183

Escribe aquí para buscar

ES 26°C 13:20 16/2/2026

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

