

# Aplicando KNIME para Analizar Créditos Bancarios

Este trabajo práctico te guiará a través del uso de KNIME, una potente herramienta de análisis de datos y la extracción de información, para analizar datos históricos de créditos personales otorgados por un banco. Aprenderás a identificar los criterios clave de otorgamiento de créditos, caracterizar grupos de clientes, y determinar los factores que influyen en el comportamiento de aquellos con ingresos más altos. También explorarás el uso de Redes Bayesianas para predecir si se otorgará un crédito a nuevos solicitantes. Serás capaz de efectuar un análisis detallado que te ayudará a desarrollar habilidades fundamentales en el campo de la Minería de Datos.



# Objetivo 1: Identificación de Criterios de Otorgamiento de Créditos

1

## Carga de Datos

Comenzaremos cargando el conjunto de datos "Otorgamiento-Créditos - Categóricos.csv" utilizando el nodo **CSV Reader** de KNIME. Este paso nos permitirá acceder a la información histórica de los clientes a quienes se les otorgó un crédito personal.

2

## Análisis Exploratorio

Realizaremos un análisis exploratorio de los datos utilizando una variedad de nodos de KNIME, como **Statistics**, **Histogram**, **Data Explorer**, **Scatter Matrix** y **Correlation Matrix**. Esto nos ayudará a comprender la naturaleza de los datos y las relaciones entre las diferentes variables.

3

## División del Dataset

Dividiremos el dataset en conjuntos de entrenamiento y prueba usando el nodo **Partitioning: Relative [%]: 70 y Draw Randomly**. Esto nos permitirá evaluar el desempeño de los modelos de clasificación que aplicaremos posteriormente.

4

## Modelado y Evaluación

Aplicaremos diversos algoritmos de clasificación, como **JRip**, **ID3**, **OneR**, **Decision Tree**, **Random Forest** y **Naive Bayes**, para identificar los factores clave que influyen en el otorgamiento de créditos. Evaluaremos el rendimiento de estos modelos utilizando métricas como **Scorer** y **ROC Curve**.



# Clasificación Avanzada

## Modelos Adicionales

Además de los algoritmos de clasificación iniciales, exploraremos modelos más avanzados como **SVM (Support Vector Machines)** y **KNearest Neighbor**. Estos métodos nos permitirán obtener una comprensión más profunda de los patrones y relaciones presentes en los datos.

## Preprocesamiento

Antes de aplicar estos algoritmos más complejos, realizaremos un preprocesamiento de los datos utilizando técnicas como **One-Hot-Encoding**. Esto nos ayudará a transformar las variables categóricas en un formato adecuado para el entrenamiento de los modelos.

## Mejora Continua

A medida que exploramos nuevos enfoques y refinamos nuestros modelos, podremos mejorar continuamente la precisión y el rendimiento de nuestras predicciones, lo que nos permitirá optimizar aún más los procesos de otorgamiento de créditos.

# Validación Cruzada

1

## División del Conjunto de Datos

Inicialmente, dividimos nuestro conjunto de datos en conjuntos de entrenamiento y prueba utilizando el nodo Partitioning de KNIME.

2

## Validación Cruzada

Para obtener una evaluación más robusta de nuestros modelos, agregaremos al flujo el nodo **X-Partitioner**, el cual nos permitirá aplicar validación cruzada con **10 iteraciones**.

3

## Comparación de Resultados

Evaluaremos y compararemos el rendimiento de nuestros modelos utilizando métricas como Scorer, lo que nos ayudará a identificar la mejor aproximación para nuestro caso de estudio.



# Objetivo 2: Segmentación de Clientes

## Algoritmos de Segmentación

Utilizaremos algoritmos de agrupamiento, como **EM**, **FarthestFirst**, **SimpleKMeans**, **K-Means** y **Hierarchical Clustering**, para identificar grupos de clientes con características similares. Esto nos permitirá estudiar líneas de crédito diferenciales y personalizar mejor nuestras ofertas.

## Visualización y Evaluación

Emplearemos nodos de visualización, como **Cluster Assignment**, **Color Manager**, **Shape Manager** y **Scatter Plot**, para representar gráficamente los clústeres identificados. Además, evaluaremos la calidad de los agrupamientos utilizando métricas como **Silhouette Coefficient** y **Entropy Scorer**.

## Insights Clave

El análisis de los grupos de clientes nos brindará valiosos insights sobre las características y necesidades de nuestros usuarios, lo que nos ayudará a diseñar estrategias de crédito más efectivas y personalizadas.





# Objetivo 3: Factores Relevantes para para Clientes con Altos Ingresos

1

## Filtrado de Datos

Para este objetivo, aplicaremos un filtro a nuestro conjunto de datos para considerar únicamente a los clientes **con ingresos superiores a \$15,000**. Esto nos permitirá enfocar nuestro análisis en este segmento específico de la población.

2

## Aplicación de Naive Bayes

Utilizaremos el algoritmo **de Naive Bayes** para determinar los factores que más inciden en el otorgamiento de créditos a los clientes con altos ingresos. Compararemos estos resultados con los obtenidos anteriormente para todo el conjunto de datos.

3

## Insights Relevantes

El análisis de los factores clave para los clientes de altos ingresos nos ayudará a diseñar ofertas de diseñar ofertas de crédito más atractivas y acordes a las necesidades de este segmento, segmento, optimizando así nuestros procesos de negocio.

# Objetivo 4: Predicción de Otorgamiento de Créditos con Redes Bayesianas

## Cálculo de Probabilidades

Aplicaremos el enfoque de Redes Bayesianas para calcular las **probabilidades A priori y condicionales** de cada clase de otorgamiento de crédito. Esto nos permitirá predecir si se otorgará o no un crédito para nuevos casos de manera más precisa.

## Predicciones Específicas

Utilizaremos el **modelo de Redes Bayesianas** para realizar predicciones sobre **casos específicos**, evaluando la probabilidad de otorgamiento de crédito. **Normalizaremos** los resultados para obtener una mejor interpretación de los mismos.

## Toma de Decisiones

Los insights obtenidos a través de las Redes Bayesianas serán fundamentales para desarrollar un mejor proceso de toma de decisiones en el otorgamiento de créditos, optimizando así los resultados para el banco y brindando una experiencia más satisfactoria a los clientes.

## Mejora Continua

A medida que se incorporen nuevos datos y se refinan los modelos, podremos mejorar continuamente la precisión de nuestras predicciones, lo que nos permitirá adaptarnos mejor a los cambios en el mercado y las necesidades de los clientes.

# Objetivo 5: Selección de Atributos Relevantes



## Algoritmo Apriori

Utilizaremos el nodo **Apriori** de la de la extensión de WEKA para KNIME, con el objetivo de identificar los atributos más significativos del conjunto de datos.



## Algoritmo CfsSubsetEval

Aplicaremos el nodo **AttributeSelectedClassifier** de la de la extensión de WEKA para KNIME, que emplea el algoritmo **CfsSubsetEval** y **J48** para seleccionar las características más relevantes.

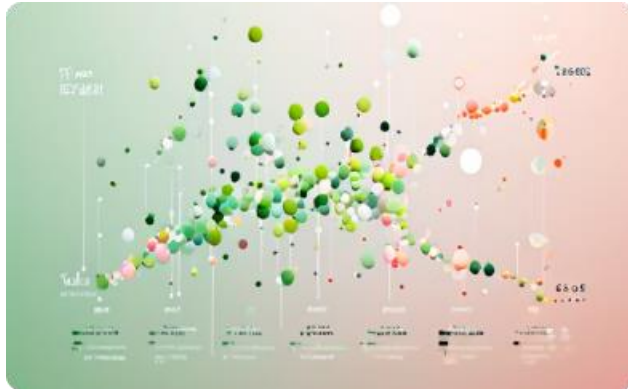


## Selección de Características Características en KNIME

Exploraremos también las capacidades nativas de KNIME, utilizando el nodo **Feature Selection Loop Start** para identificar los atributos más significativos del caso de estudio.

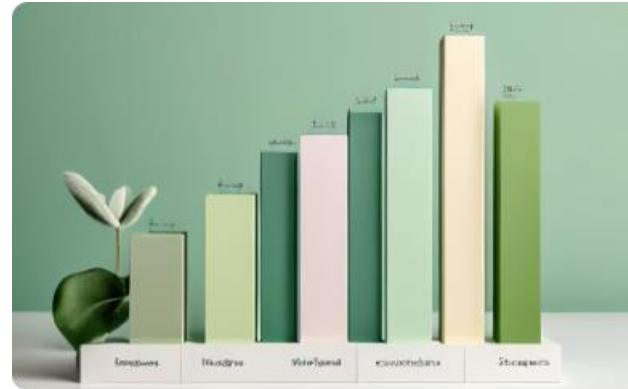


# Visualización de Resultados



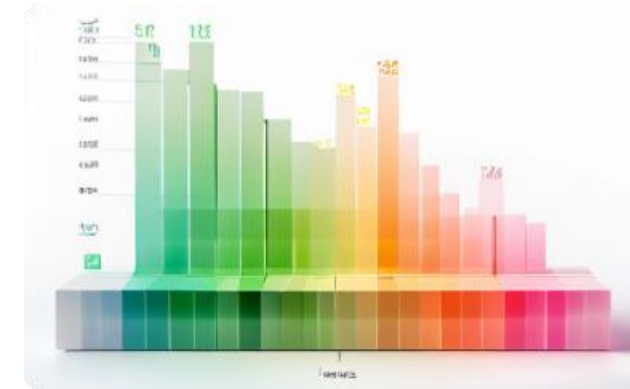
## Visualización de Clústeres

Utilizaremos nodos de visualización como Scatter Plot para representar gráficamente los grupos de clientes identificados mediante los algoritmos de agrupamiento. Esto nos permitirá comprender mejor las características y las diferencias entre los segmentos.



## Ranking de Atributos

Crearemos visualizaciones como gráficos de barras para mostrar el ranking de importancia de los atributos, lo que facilitará la interpretación de los resultados de los algoritmos de selección de características.



## Métricas de Evaluación

Emplearemos matrices de confusión y otras visualizaciones para evaluar el desempeño de nuestros modelos de clasificación. Esto nos permitirá entender mejor los aciertos y errores de las predicciones.

# Entregables del Trabajo Práctico

Workflow KNIME	Archivo GrupoNN_TPApp2_MineriaDatos.knwf
Informe Final	Archivo GrupoNN_TPApp2_MineriaDatos.docx
Planillas Excel	Archivo GrupoNN_TPApp2_MineriaDatos.xlsx

Estos tres entregables conforman los resultados del trabajo práctico.

El workflow KNIME contiene el flujo de trabajo completo con los nodos y configuraciones utilizadas.

El informe final detalla el proceso y los hallazgos clave.

Y las planillas Excel incluyen tablas comparativas y resultados de las Redes Bayesianas.



# Fecha Límite de Entrega

1

## Fecha Límite

La fecha límite para la entrega de este trabajo práctico se encuentra en MleL y en el enunciado del mismo.

2

## Puntualidad

Es importante respetar la fecha límite establecida para asegurar que el trabajo sea evaluado a tiempo y poder recibir la retroalimentación correspondiente.

3

## Recomendación

Se recomienda comenzar a trabajar en el proyecto con anticipación, de modo que puedan lograr los mejores resultados posibles.



# Conclusiones y Próximos Pasos

A lo largo de este trabajo práctico, hemos explorado cómo KNIME puede ser una herramienta poderosa para el análisis de datos y la extracción de información valiosa en el contexto de la Inteligencia de Negocios. Hemos aplicado diversos algoritmos y técnicas para identificar los criterios de otorgamiento de créditos, segmentar a los clientes, analizar los factores clave para clientes con altos ingresos, realizar predicciones utilizando Redes Bayesianas y seleccionar los atributos más relevantes.

Los insights y resultados obtenidos serán fundamentales para que el banco pueda banco pueda tomar decisiones más informadas, diseñar ofertas de crédito más crédito más atractivas y personalizadas, y mejorar continuamente sus procesos de procesos de negocio. A medida que sigamos aplicando estos enfoques y refinando refinando nuestros modelos, podremos consolidar aún más nuestra posición en el posición en el mercado y brindar una experiencia excepcional a nuestros clientes. clientes.