UNIVERSIDAD NACIONAL DE LA MATANZA

INTELIGENCIA DE NEGOCIOS

**Trabajo Práctico de Aplicación 2:**

**Introducción a KNIME – EI Créditos Bancarios**

**Objetivo del Trabajo**

Familiarizarse con la funcionalidad que proporciona KNIME para desarrollar los Procesos de Explotación de Información (EI) identificados, que darán soporte a los Objetivos de Negocio del Caso de Estudio en un banco que desea lanzar una nueva línea de créditos personales.

**Enunciado**

El trabajo práctico se basa en analizar una muestra de datos históricos de 144 clientes a los que se les otorgó un crédito personal, utilizando KNIME para cumplir los siguientes objetivos:

1. Identificar los criterios de otorgamiento de créditos.
2. Identificar y caracterizar grupos de clientes para estudiar líneas de crédito diferenciales.
3. Identificar los factores de incidencia en clientes con ingresos superiores a $15.000.
4. Determinar si se otorga o no un crédito para nuevos casos usando Redes Bayesianas.
5. Seleccionar los atributos más significativos del caso de estudio.

**Actividades Preliminares**

1. **Instalación de KNIME:**
   * Descargar e instalar la versión 5.2 de KNIME desde su sitio web.
   * Consultar la documentación oficial y tutoriales disponibles en los recursos proporcionados.
2. **Revisión de Materiales:**
   * “Introducción al Análisis de Datos - Prácticas con Power BI, R y KNIME”.
   * “Practicing Data Science – The Data Science Case Study Collection”.
   * Tutoriales en el KNIME Learning Center y YouTube.

**Desarrollo del Trabajo**

**Objetivo 1: Identificar los Criterios de Otorgamiento de Créditos**

1. **Cargar el Dataset:**
   * Utilizar el nodo CSV Reader para cargar el archivo “Otorgamiento-Creditos - Categoricos.csv”.
2. **Análisis Exploratorio de Datos (EDA):**
   * Utilizar nodos como Statistics, Histogram, Data Explorer, Scatter Matrix y Correlation Matrix.
3. **División del Dataset:**
   * Dividir los datos en entrenamiento y prueba usando Partitioning con opciones Relative [%]:70 y Draw Randomly.
4. **Modelos de Clasificación:**
   * Aplicar algoritmos como JRip, ID3, OneR, Decision Tree, Random Forest, y Naive Bayes.
   * Evaluar los modelos con Scorer y ROC Curve.
5. **Modelos Adicionales:**
   * Aplicar SVM y KNearest Neighbor, realizando preprocesamiento con One-Hot-Encoding.

**Tarea Adicional: Validación Cruzada con X-Partitioner**

1. **Agregar al flujo de trabajo el nodo X-Partitioner:**
   * Conectar el archivo leído con el nodo X-Partitioner, configurando Number of Validations: 10 y Random Sampling.
   * Conectar X-Partitioner con un clasificador de preferencia (por ejemplo, Decision Tree Learner y Decision Tree Predictor).
   * Evaluar el rendimiento del modelo con Scorer y comparar con los resultados obtenidos usando el nodo Partitioning.

**Objetivo 2: Identificar y Caracterizar Grupos de Clientes**

1. **Algoritmos de Segmentación:**
   * Utilizar algoritmos como EM, FarthestFirst, SimpleKMeans, K-Means y Hierarchical Clustering.
2. **Visualización y Evaluación de Clústeres:**
   * Utilizar nodos de visualización como Cluster Assignment, Color Manager, Shape Manager y Scatter Plot.
   * Evaluar la calidad de los clústeres con Silhouette Coefficient y Entropy Scorer.

**Objetivo 3: Factores de Incidencia en Clientes con Ingresos Superiores a $15.000**

1. **Filtro de Datos:**
   * Filtrar clientes con ingresos superiores a $15.000 utilizando Row Filter.
2. **Aplicación de Naive Bayes:**
   * Aplicar Naive Bayes al dataset filtrado y comparar con los resultados anteriores.

**Objetivo 4: Predicción de Créditos para Nuevos Casos usando Redes Bayesianas**

1. **Cálculo de Probabilidades:**
   * Calcular probabilidades a priori y condicionales para cada clase.
   * Predicciones para casos específicos y normalización de resultados.

**Objetivo 5: Seleccionar los Atributos más Significativos del Caso de Estudio**

1. **Algoritmos de Selección de Características:**
   * **Algoritmo Apriori:**
     + Utilizar el nodo Apriori (3.7) de la extensión de WEKA para KNIME.
   * **Algoritmo CfsSubsetEval:**
     + Utilizar el nodo AttributeSelectedClassifier (3.7) de la extensión de WEKA para KNIME.
   * **Selección de Características con KNIME:**
     + Utilizar el nodo Feature Selection Loop Start (1:1) de los originales de KNIME.

**Entregables**

1. **Workflow KNIME:**
   * Archivo GrupoNN\_TPApp2\_MineriaDatos.knwf.
2. **Informe Final en Word:**
   * Archivo GrupoNN\_TPApp2\_MineriaDatos.docx.
3. **Planillas Excel:**
   * Incluir Anexo Solapa A con tabla comparativa de métodos de clasificación.
   * Incluir Anexo Solapa B con tablas de resultados de Redes Bayesianas.
   * Archivo GrupoNN\_TPApp2\_MineriaDatos.xlsx.

**Fecha Límite de Entrega**

* **Fecha Límite:** 18/06/2025