

UNIVERSIDAD DEL VALLE – SEDE TULUÁ
INGENIERÍA DE SISTEMAS
INTRODUCCIÓN A LA INTELIGENCIA ARTIFICIAL
INFORME MINIPROYECTO

Nombres: Jairo Alonso Vera

1. **Objetivo:** Aplicar una técnica y conceptos de minería de datos vistas a lo largo del curso para la resolución de un problema.
2. **Marco Teórico:** La minería de datos es un campo de las ciencias de la computación referido al proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos. El objetivo general del proceso de minería de datos consiste en extraer información de un conjunto de datos y transformarla en una estructura comprensible para su uso posterior.

Las técnicas más representativas de la minería de datos son:

- **Redes neuronales:** Son un paradigma de aprendizaje y procesamiento automático inspirado en la forma en que funciona el sistema nervioso de los animales. Se trata de un sistema de interconexión de neuronas en una red que colabora para producir un estímulo de salida. Algunos ejemplos de red neuronal son: El Perceptrón, El Perceptrón multicapa, Los Mapas Auto organizados, también conocidos como redes de Kohonen..

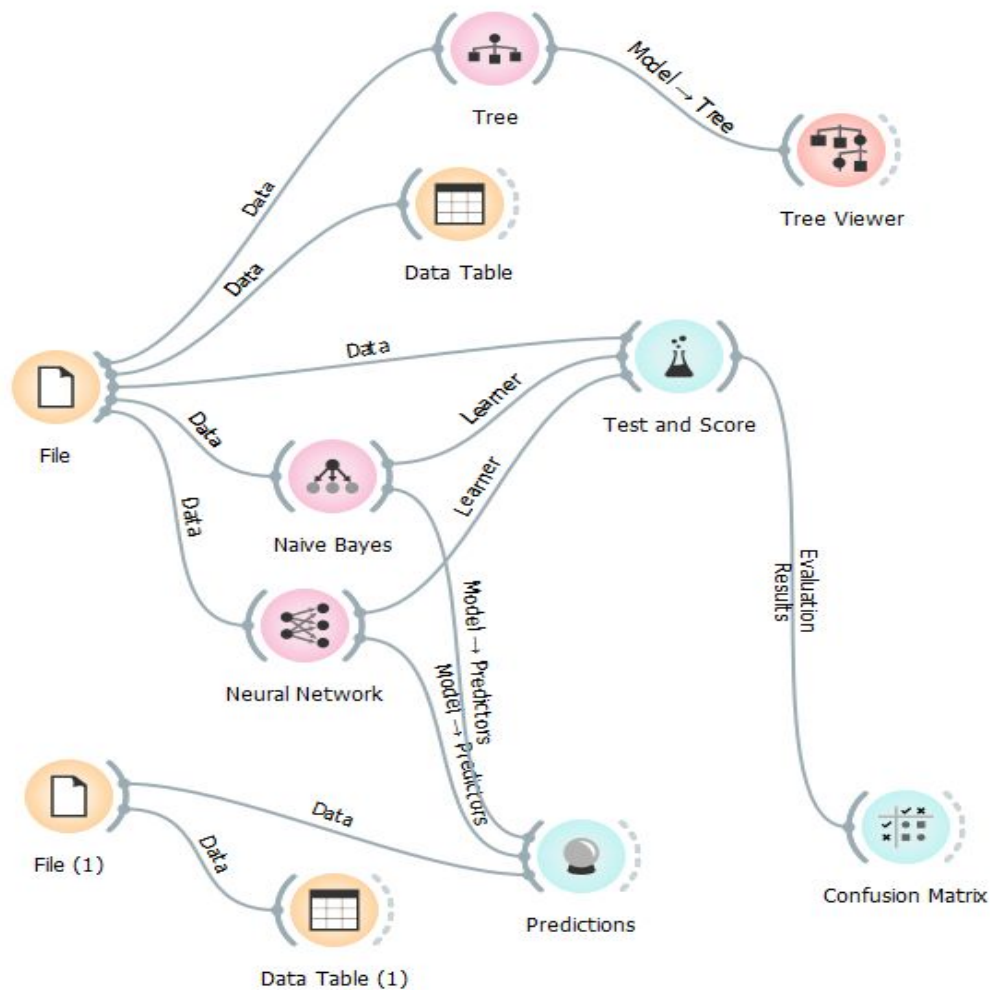
- **Regresión lineal:** Es la más utilizada para formar relaciones entre datos. Rápida y eficaz pero insuficiente en espacios multidimensionales donde puedan relacionarse más de 2 variables.
- **Árboles de decisión:** Un árbol de decisión es un modelo de predicción utilizado en el ámbito de la inteligencia artificial, dada una base de datos se construyen estos diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que suceden de forma sucesiva, para la resolución de un problema. Ejemplos: Algoritmo ID3, Algoritmo C4.5.
- **Modelos estadísticos:** Es una expresión simbólica en forma de igualdad o ecuación que se emplea en todos los diseños experimentales y en la regresión para indicar los diferentes factores que modifican la variable de respuesta.
- **Agrupamiento o Clustering:** Es un procedimiento de agrupación de una serie de vectores según criterios habitualmente de distancia; se tratará de disponer los vectores de entrada de forma que estén más cercanos aquellos que tengan características comunes. Ejemplos: Algoritmo K-means, Algoritmo K-medoids.
- **Reglas de asociación:** Se utilizan para descubrir hechos que ocurren en común dentro de un determinado conjunto de datos.

3. Aplicación: el problema propuesto y a resolver es la detección de fraude de tarjetas de crédito. Utilizamos las técnicas de modelación Bayesiana y redes neuronales para realizar una tarea de clasificación implementando programa Orange. Comparamos los resultados de Naive Bayes con el modelo Neural Network. Conectamos los

datos del dataset de File a Test & Score, también Naive Bayes y Neural Network con Test & Score y observamos sus puntajes de predicción.

Implementación en orange

se uso un dataset descargado de la pagina kaggle.com, el cual contenía información sobre transacciones bancarias de un grupo de personas, de la cual basada en ellas y en una columna llamada clase la cual fue binarizada y sobre la cual se hizo el target para clasificación y predicciones

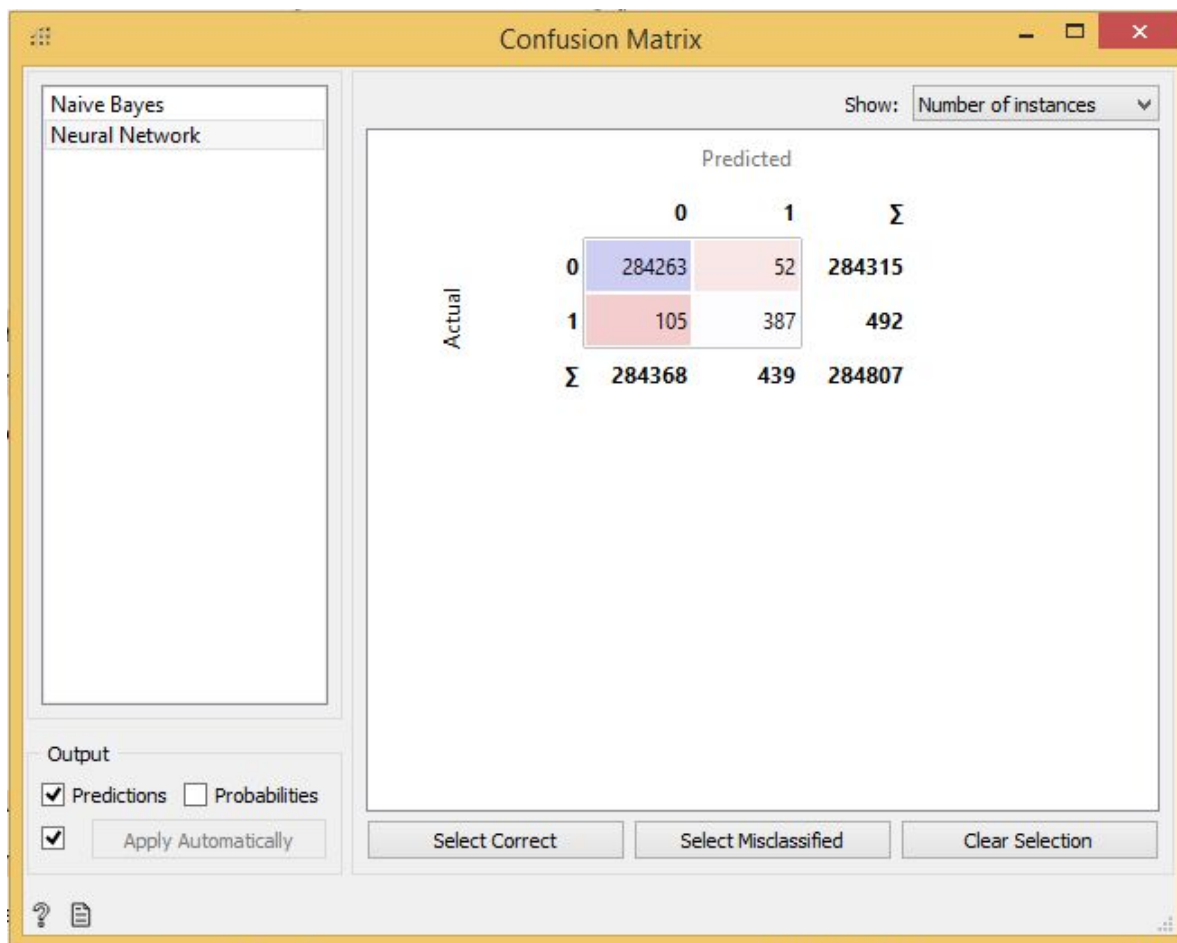


el primer “file” tiene cargado el dataset binarizado para la clasificación

y el segundo “file” tiene un dataset mas pequeño al cual se le suprimió la columna binarizada para así poder hacer una predicción basada en los modelos de redes neuronales y bayesiano

Resultados

Clasificación mediante matriz de confusión

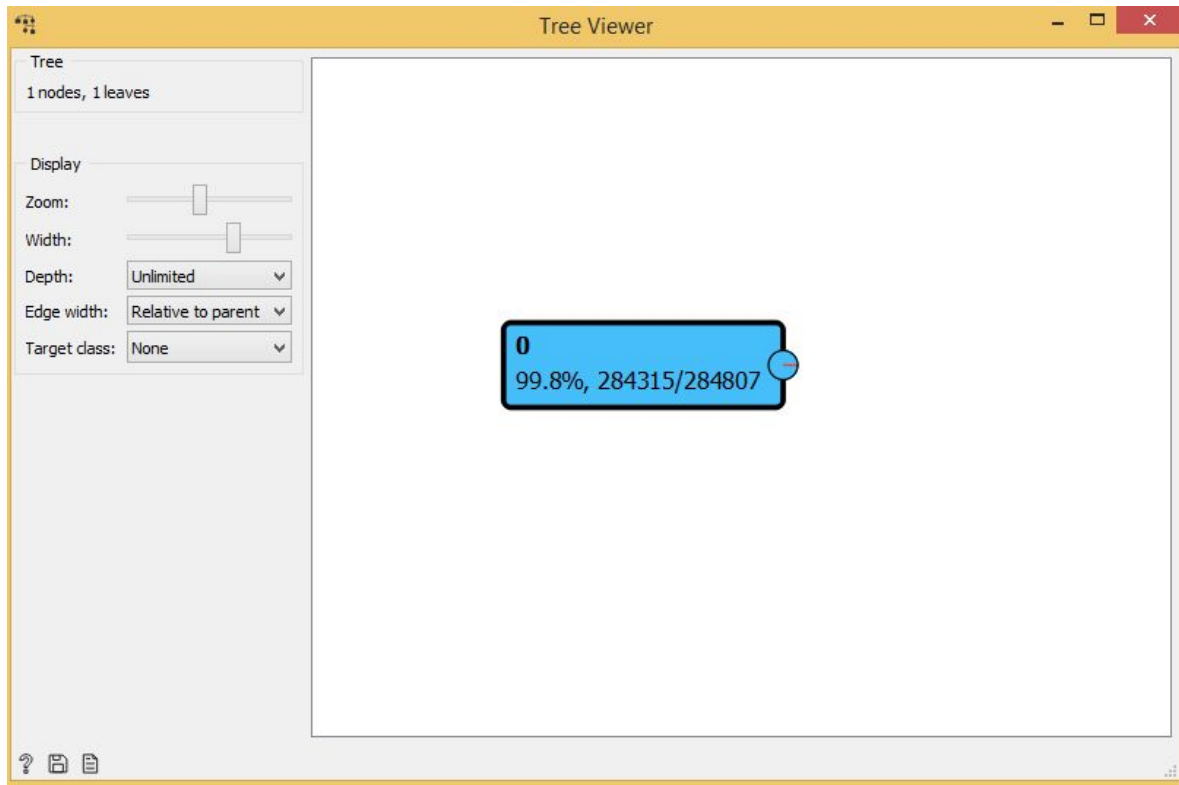


Modelo de predicción

	Neural Network	Naive Bayes	Time	V1	
1	1.00 : 0.00 → 0	1.00 : 0.00 → 0	4	1.22966	0.1410
2	1.00 : 0.00 → 0	1.00 : 0.00 → 0	7	-0.644269	1.4179
3	1.00 : 0.00 → 0	1.00 : 0.00 → 0	7	-0.894286	0.2861
4	1.00 : 0.00 → 0	1.00 : 0.00 → 0	39	-1.33088	0.2675
5	1.00 : 0.00 → 0	1.00 : 0.00 → 0	40	1.11069	0.0819
6	1.00 : 0.00 → 0	1.00 : 0.00 → 0	41	1.15431	0.2654
7	1.00 : 0.00 → 0	1.00 : 0.00 → 0	41	0.986063	-0.202
8	1.00 : 0.00 → 0	1.00 : 0.00 → 0	41	1.13876	-1.192
9	1.00 : 0.00 → 0	1.00 : 0.00 → 0	41	1.14552	0.5750
10	1.00 : 0.00 → 0	1.00 : 0.00 → 0	42	-0.249361	0.3992
11	0.01 : 0.99 → 1	0.00 : 1.00 → 1	7610	0.725646	2.3008
12	0.00 : 1.00 → 1	0.00 : 1.00 → 1	7672	0.70271	2.4264
13	0.00 : 1.00 → 1	0.00 : 1.00 → 1	34256	0.539276	1.5548
14	1.00 : 0.00 → 0	1.00 : 0.00 → 0	34521	1.08123	0.4164
15	0.02 : 0.98 → 1	0.00 : 1.00 → 1	34634	0.333499	1.6998
16	0.81 : 0.19 → 0	1.00 : 0.00 → 0	17230	-0.469327	1.1114
17	0.00 : 1.00 → 1	0.00 : 1.00 → 1	17520	-5.26805	9.0676
18	0.00 : 1.00 → 1	0.00 : 1.00 → 1	17838	-5.18788	6.9677

Donde podemos apreciar que el 1 son los aciertos y 0 los errados y también vemos que ambos modelos usados entregan casi los mismo resultados

Porcentajes de probabilidad



Donde el porcentaje de transacciones validas es de 99.8%