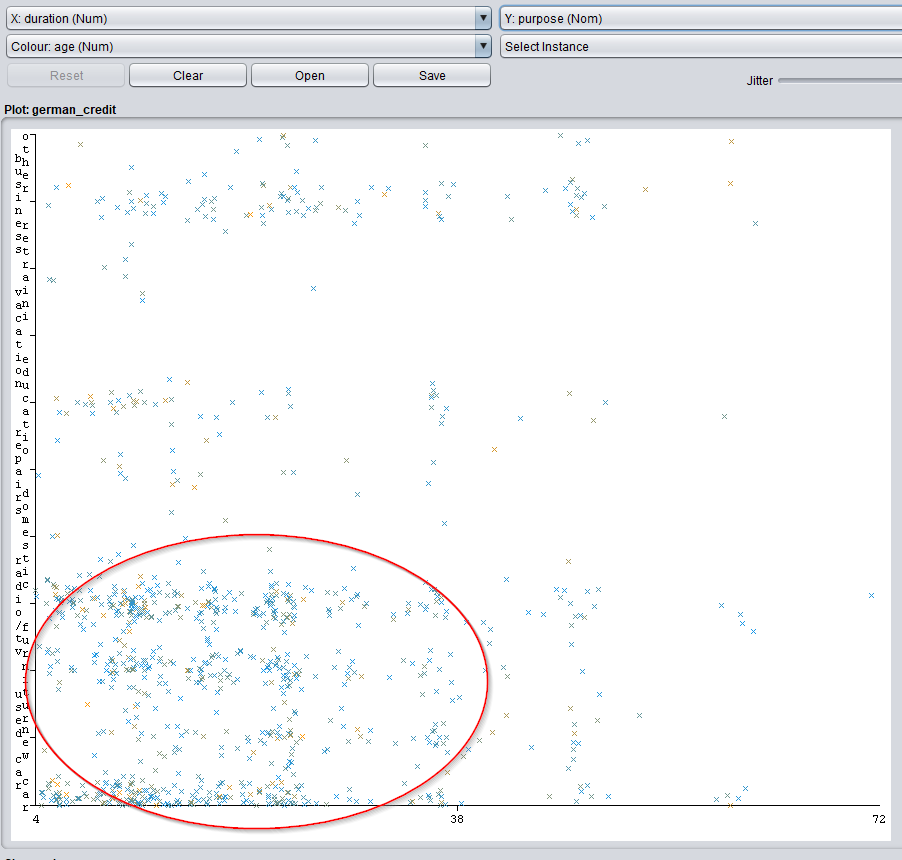
**Autor**: José Romualdo Villalobos Pérez

**ID**: 000294087

# El problema:

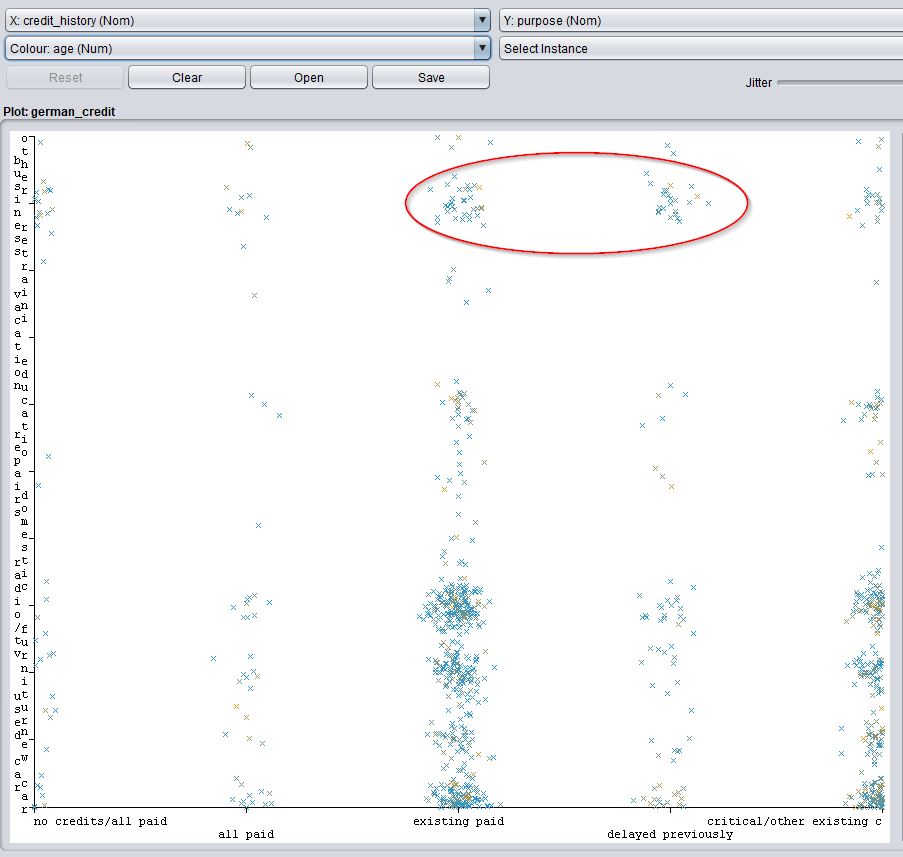
Un banco alemán desea establecer políticas de mercadeo según el tipo de cliente. Para esto se desean identificar tipos de clientes de una base de datos de 1000 registros con 20 atributos cada uno.

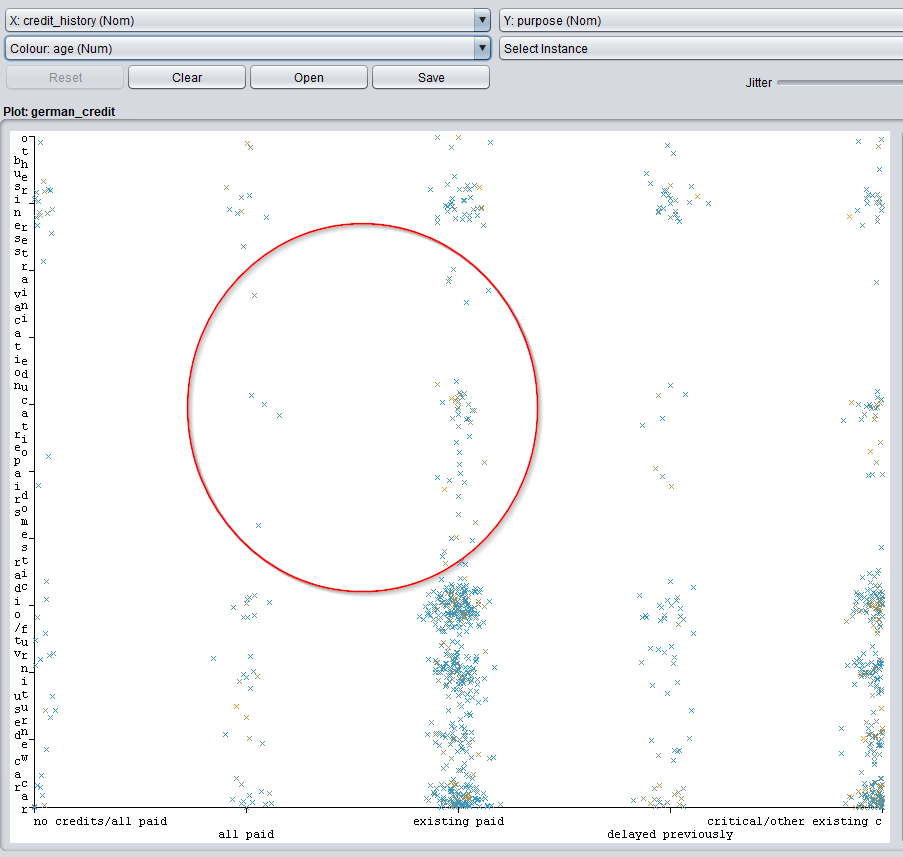
# Definiendo el número de clusters a detectar



Si la duración del crédito esta entre 4 y 38 meses lo mas posible es que el propósito del crédito sea uno de los siguientes:

1. Radio/TV
2. Furniture
3. Used Car
4. New Car

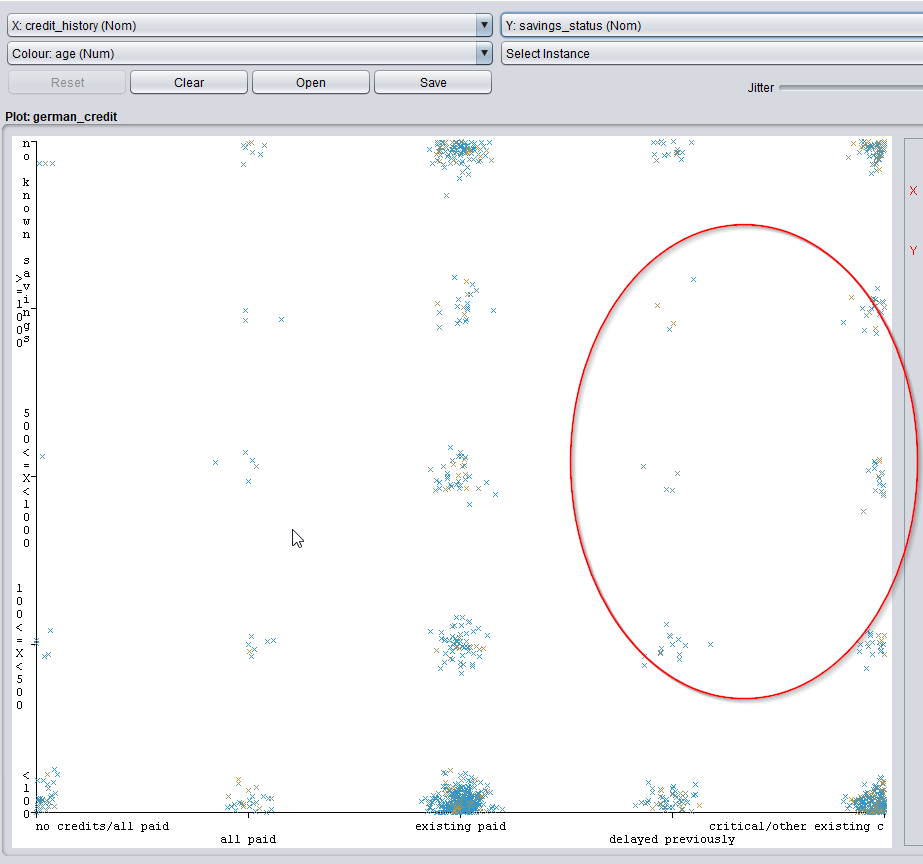
  
Si el propósito del crédito es para business lo mas probable es que el cliente page bien el crédito o que se haya atrasado en algún pago anteriormente.

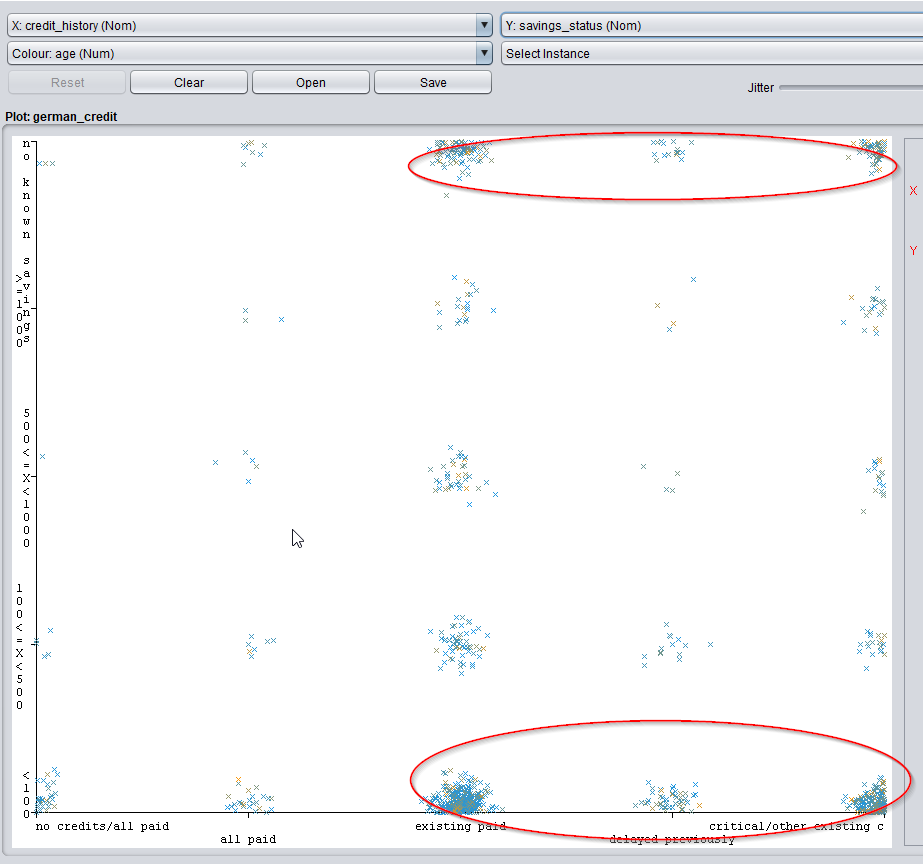


Las personas que sacan crédito para:

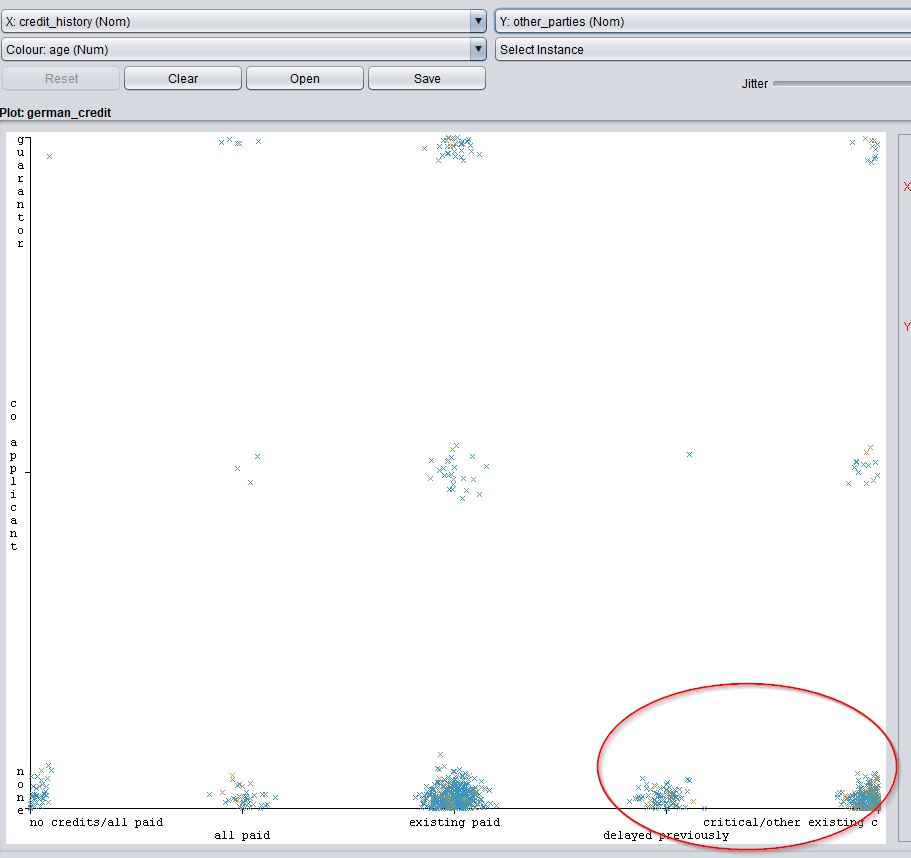
1. Retraining
2. Education
3. Repairs
4. Domestic

Por lo general no se atrasan en los pagos, son muy cumplidas.

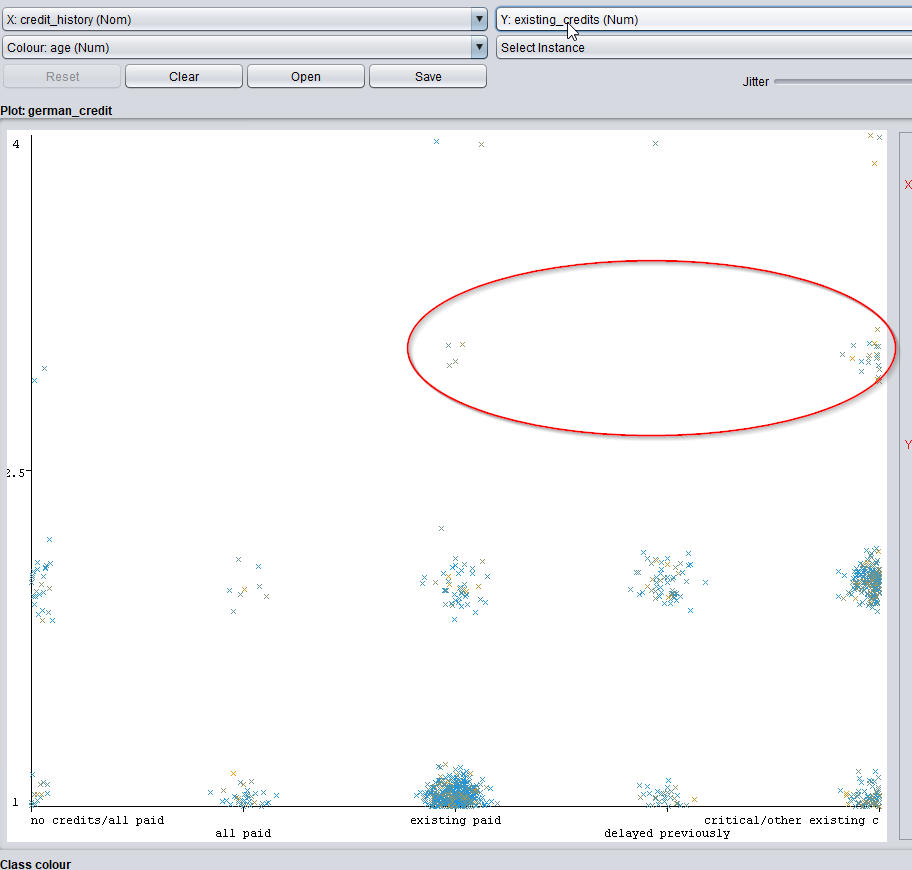


Las personas con ahorros de entre 100 y más de 1000 se retrasan menos en los pagos. 

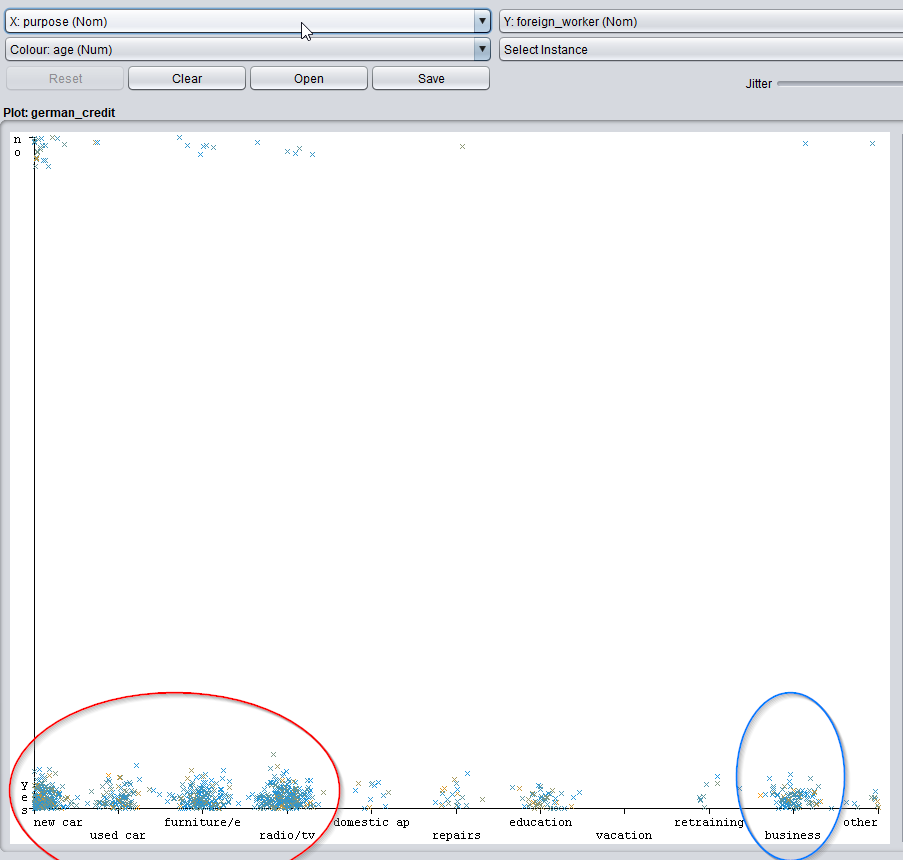
Las personas con ahorros de menos de 100 o ahorros desconocidos tienen muchas posibilidades de atrasarse en los pagos del crédito.



Si la persona no es co-aplicante o garante tiene muchas posibilidades de atrasarse en los pagos.

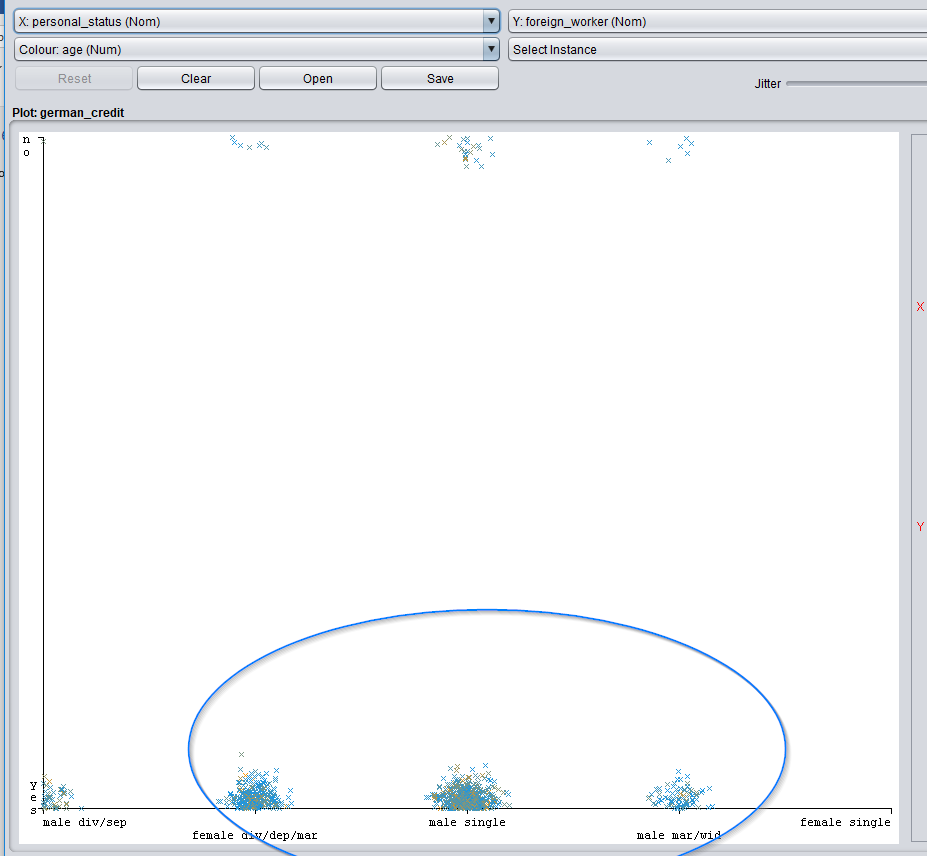


Las personas con tres o más créditos tienen muchas posibilidades de atrasarse en los pagos

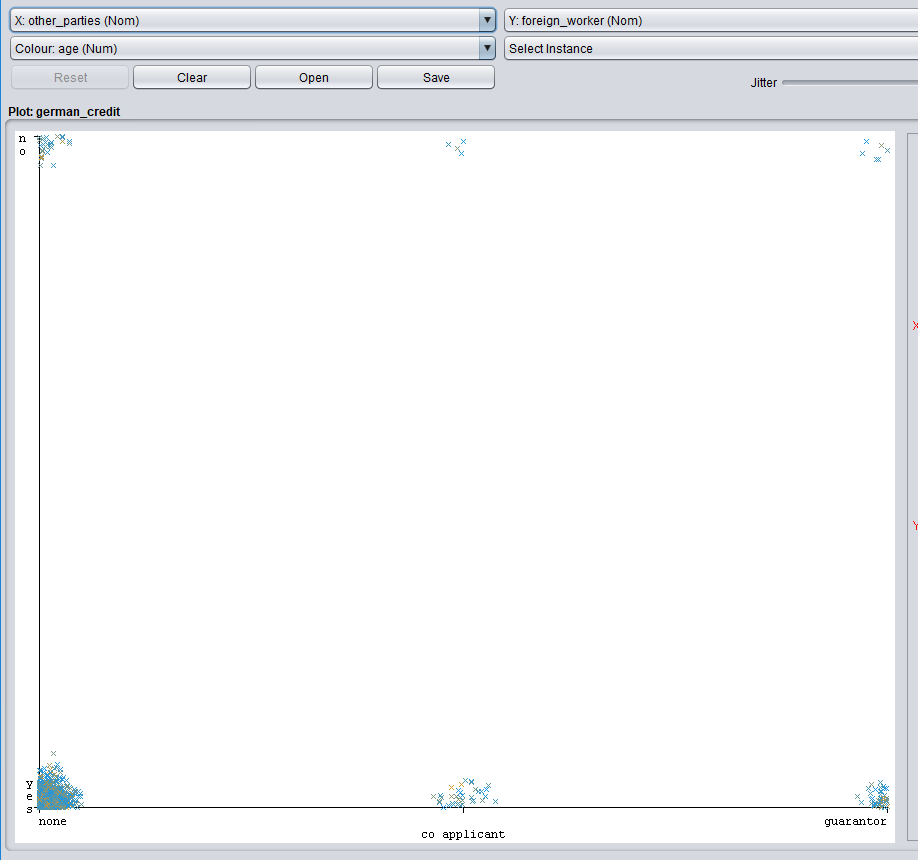


La mayoría de los créditos para negocios son pedidos por trabajadores extranjeros.

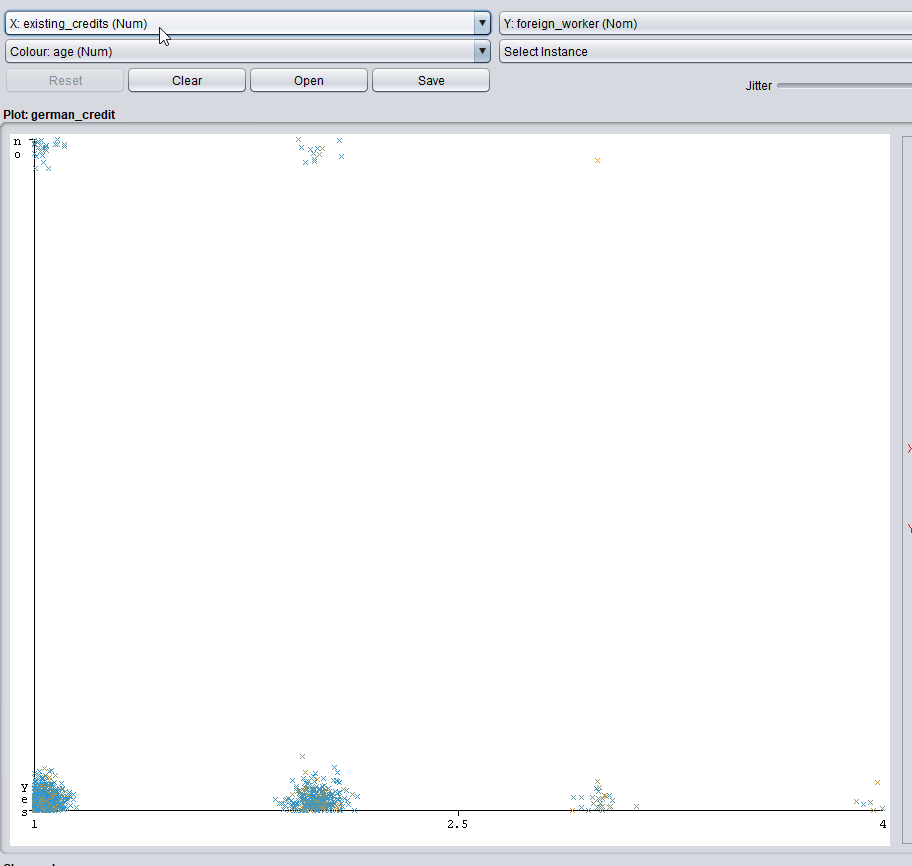
En general la mayoría de los créditos son solicitados por trabajadores extranjeros.



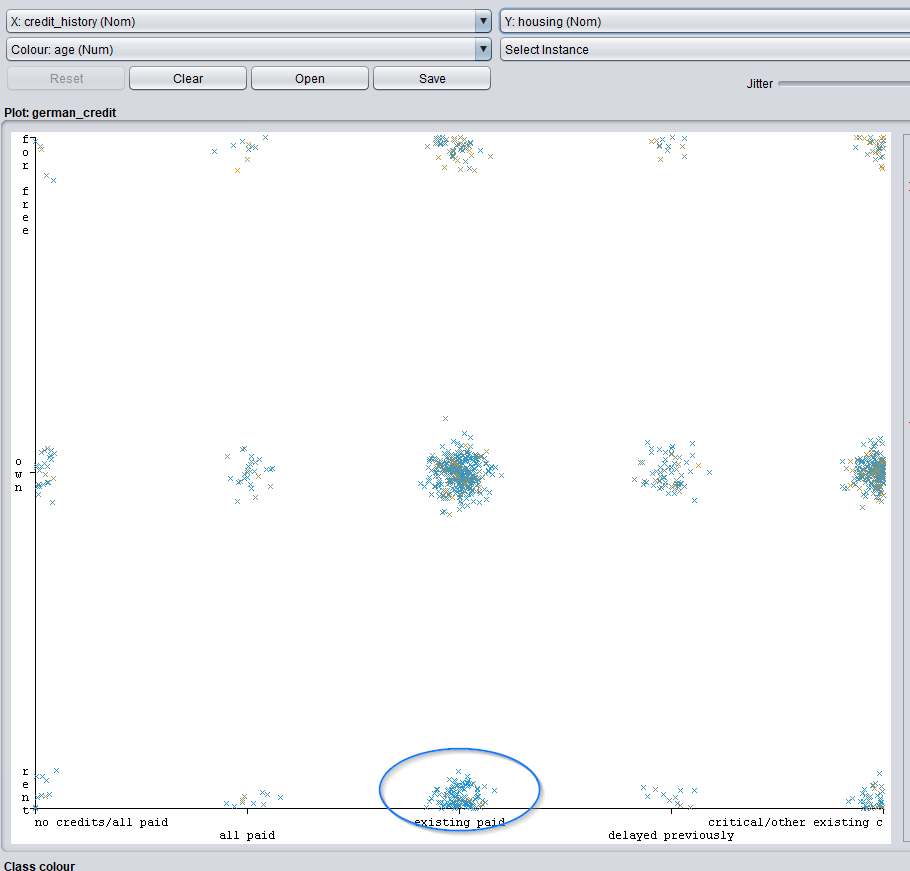
La mayoría de los trabajadores extranjeros se encuentran en una relación o son hombres solteros.



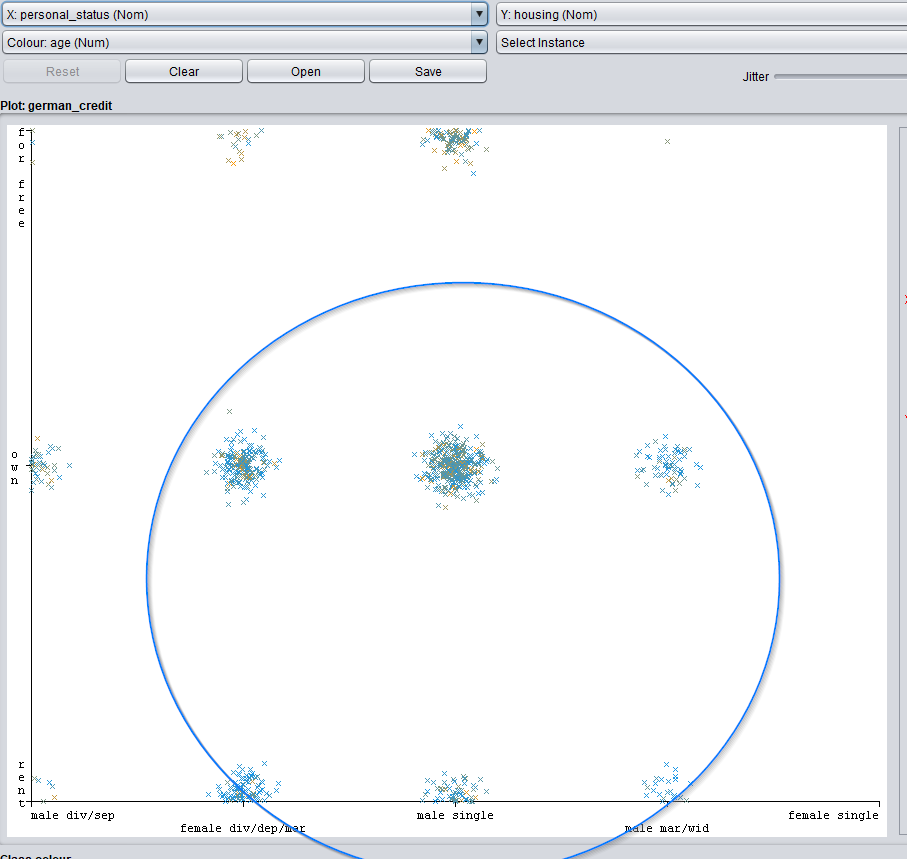
La mayoría de los trabajadores extranjeros no son co-aplicantes ni co-deudores



La mayoría de los trabajadores extranjeros tienen entre 1 y 2 creditos.



Las personas que arriendan su casa tienes menos probabilidades de atrasarse en los pagos, quizás son mas cuidadosos con sus gastos y no los subestiman tanto.



La mayoría de los hombres y mujeres que están en pareja arriendan o tienen casa propia pero no es una casa gratis.

## Definiendo los clústeres

### Clúster 1

Trabajadores extranjeros que piden un crédito para radio/tv, furniture, carro nuevo o carro usado, cabe notar que son artículos que pueden ser considerados esenciales por ellos para su estadía en el país.

### Clúster 2:

Créditos para negocios normalmente hay una posibilidad de que se retrasen en los pagos, tienen ahorros inferiores a los 100 DM (Marco Alemán), hombres solteros o hombres o mujeres casados o una relación, no son co-aplicants ni guarantors casi todos tienen su propia casa, tienen altas habilidades en la labor que desempeñan. Casi todos son trabajadores extranjeros.

### Clúster 3:

Personas que están dispuestas a asumir riesgos de familiares y amigos sirviendo de co-applicants o guarantors por lo general no se atrasan en los pagos.

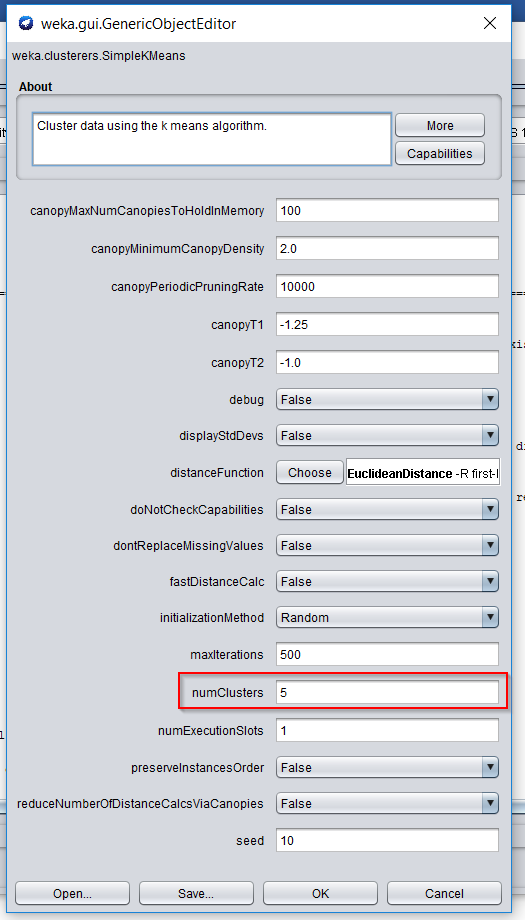
## Clúster 4:

Personas que tienen casa propia o arriendan una y al mismo tiempo son trabajadores extranjeros que tienen entre 1 y 3 créditos principalmente con altas habilidades en su profesión.

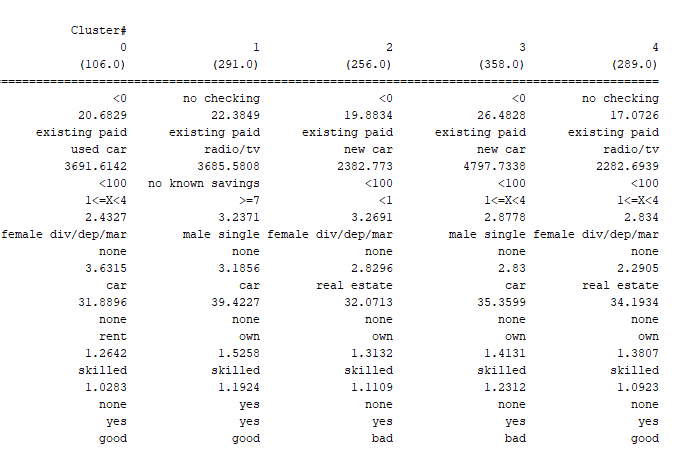
### Clúster 5:

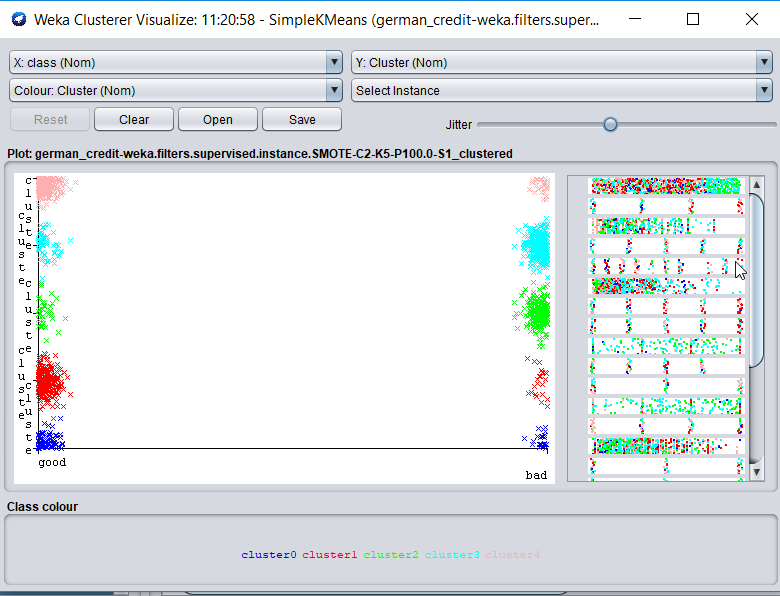
Trabajador extranjero con entre 1 y 2 créditos con personas a cargo.

# K-Means



Se obtuvieron los siguientes centroides para los 5 clusters:





Lamentablemente este método de clustering no separó los clientes buenos de los malos.

Ahora, vamos a nombrar los clusters:

**Cluster 0:**

**Nombre:** Independiente

**Descripción Persona:** Mujer responsable y dedicada a su carrera, no le da miedo ir a otros países a trabajar, estas mujeres le huyen a los gastos grandes, prefiere tener un auto usado que uno nuevo, prefiere rentar su vivienda que comprar una no le preocupa lo que pueda pasar en el futuro.

**Cluster 1:**

**Nombre:** Libertad

**Descripción Persona:** Dedicado a su carrera, valora mucho los espacios de ocio, por eso tiene un TV gigante, vive en su propia casa y no tiene pareja.

**Cluster 2:**

**Nombre:** Imperial

**Descripción Persona:** Tiene su propia casa, no le gusta endeudarse a largo plazo, pero a veces no calcula bien sus verdaderas capacidades, le gusta movilizarse en automóvil y prefiere sentir que es la dueña.

**Cluster 3:**

**Nombre:** Orgullo personal

**Descripción Persona:** No tiene pareja, le gusta lucir su automóvil, dedicado a su carrera, se siente cómodo en proyectos de mediano plazo, a veces no calcula bien sus verdaderas capacidades.

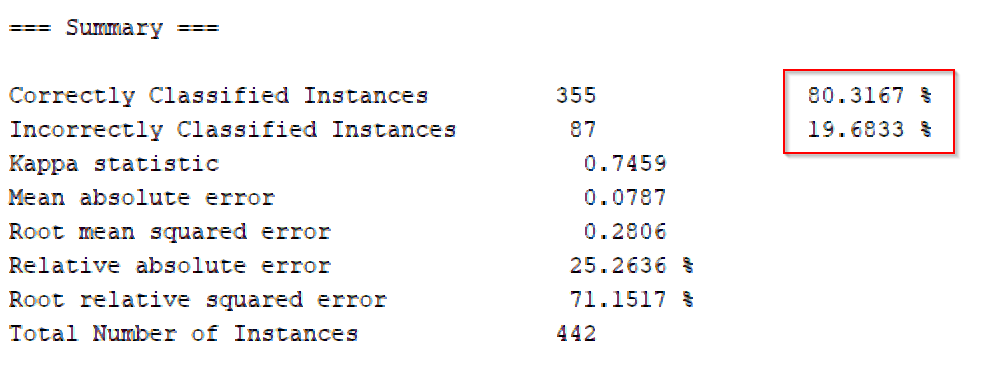
**Cluster 4:**

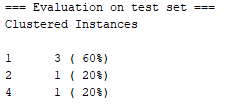
**Nombre:** Vive la vida

**Descripción Persona:** Dedicada a su carrera, valora su comodidad y los momentos de ocio, tiene su propia casa, despreocupada de lo que pueda pasar en el futuro.

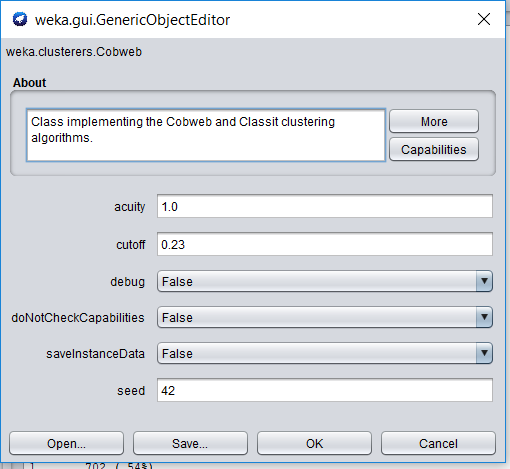
**Probando el modelo con 5 usuarios nuevos:**

Para predecir a que cluster pertenecen 5 usuarios nuevos, vamos a aplicar el clasificador Support Vector Machine donde la clase de cada instancia es el cluster al que pertenece, para este caso entrenamos una SVM con un kernel lineal.

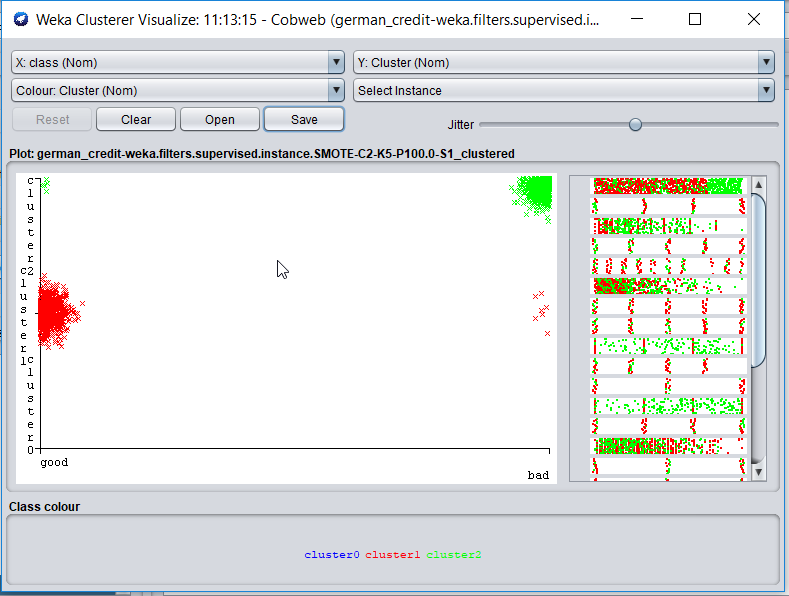




# Cobweb

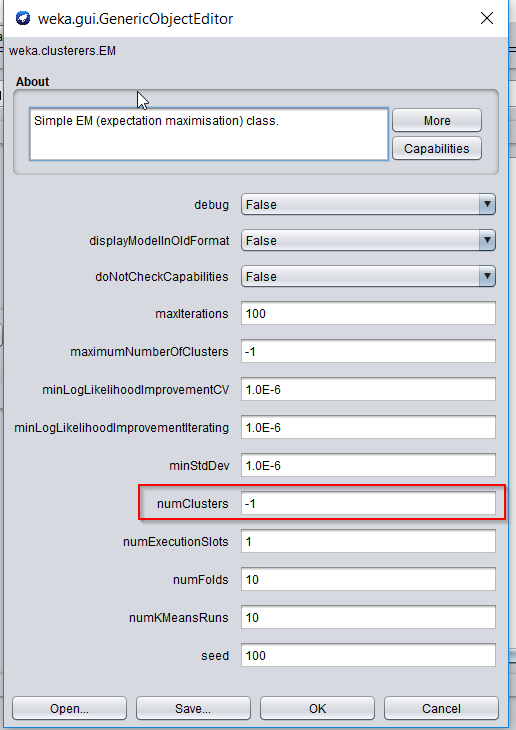


El valor de cutoff es 0.23 para evitar que el árbol crezca mucho, el resultado obtenido se muestra a continuación:

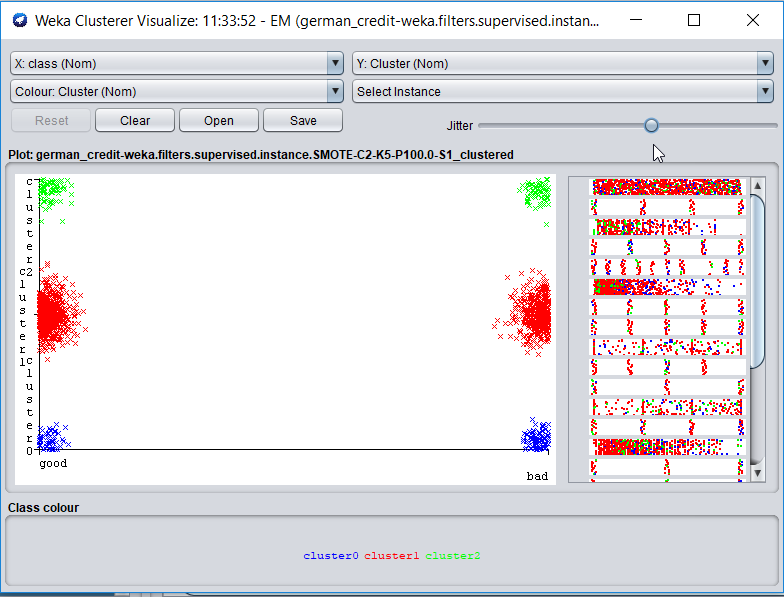


Afortunadamente para el banco, este método de clustering agrupo a los clientes buenos y a los clientes malos en clusters diferentes.

# Expectation Maximization



Cuando numClusters = -1 el algoritmo selecciona un valor automáticamente para K basado en la cross validation.



No separó a los clientes buenos de los clientes malos.

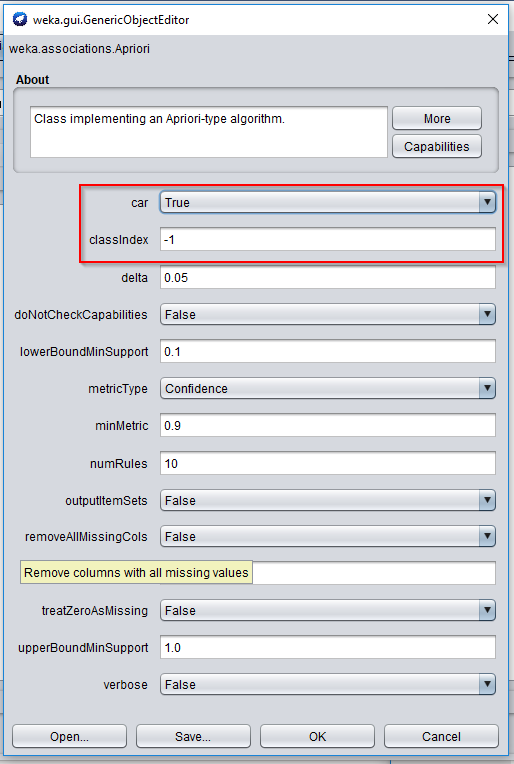
# Self-Organizing Maps

Debigo a que Weka no tiene una clase para aplicar este algoritmo, se implementara en Python y en R únicamente.

*“Self-organizing map is not in weka catalogue the only neural networks do not compress features to viewable 2d map only regular nn's named multilayerperceptron and rbfnetwork which have different (opposite) activation functions”*

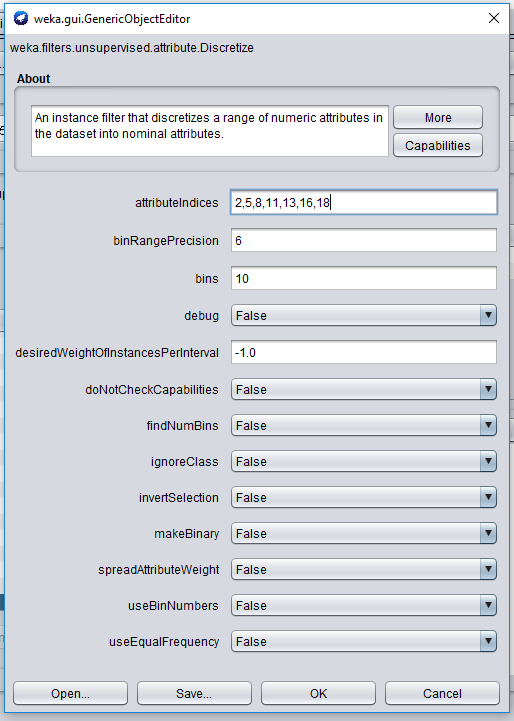
* *Harry M.T. Saarikoski, Weka mail list*

# Association, A priori

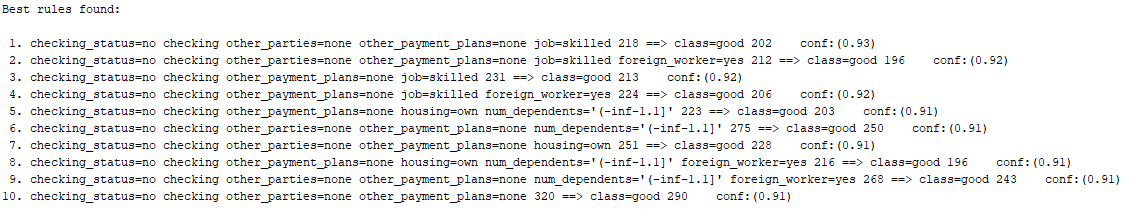


**Car = true** porque las instancias tienen un atributo clase, y **classIndex=-1** porque la clase se encuentra en la última columna.

Debido a que el método apriori solo funciona con variables categóricas, vamos a proceder discretizando los atributos numéricos.



Como resultado, ahora todos los atributos son de tipo nominal, ahora si podemos aplicar las reglas de asociación.



Podemos concluir que este algoritmo brinda al banco información muy valiosa, que podrá permitir sacar mas valor de sus clientes.