



WEKA

confit-g

Integrantes:

- Dennis David Sejas
- Kasandra Mamani Rodriguez
- Brayan Sabinas

¿OBTENCIÓN DE LOS DATOS?



El proceso de Descubrimiento de Conocimiento en Bases de Datos (KDD) implica varios pasos para transformar datos sin procesar en información útil. En el contexto del archivo credit-g, que contiene datos de clientes de una entidad bancaria que solicitaron crédito, estos pasos incluyen:

PASOS:

1. Entendimiento del Problema: En este caso, el objetivo es comprender los factores que determinan si un cliente es confiable para recibir un crédito

2. Preparación de los Datos: Esto implica cargar el archivo credit-g.arff en Weka, inspeccionando los 20 atributos (7 numéricos y 13 nominales) que describen a cada cliente y al crédito solicitado. Se debe analizar la distribución de las clases ("fiable" o "no fiable") y considerar la creación de atributos derivados como el cociente entre variables si es necesario.

3. Filtrado de Atributos: Identificar atributos irrelevantes que pueden afectar negativamente la precisión del modelo. Weka ofrece técnicas de "envoltura" (wrappers) que utilizan un algoritmo de aprendizaje (como Naive Bayes) para evaluar subconjuntos de atributos y seleccionar los más relevantes.

4. Modelado: Se pueden aplicar algoritmos de clasificación como árboles de decisión (J48), Naive Bayes, o incluso la combinación de modelos (Bagging, Boosting) para construir modelos predictivos.

5. Evaluación: Se evalúa la precisión del modelo utilizando técnicas como validación cruzada, considerando la matriz de costes que cuantifica el impacto de clasificaciones incorrectas

6. Interpretación y Uso: Se interpretan los resultados y se traduce el conocimiento descubierto en acciones, como políticas de concesión de crédito.



2.¿QUE SIGNIFICA CADA DATO?



El archivo credit-g contiene información sobre 1000 clientes de una entidad bancaria que han solicitado un crédito. Cada cliente está representado por 20 atributos, 7 numéricos y 13 categóricos. Además, el archivo incluye información sobre la clase a la que pertenece cada cliente, que indica si es "fiable" o "no fiable" para la concesión del crédito.

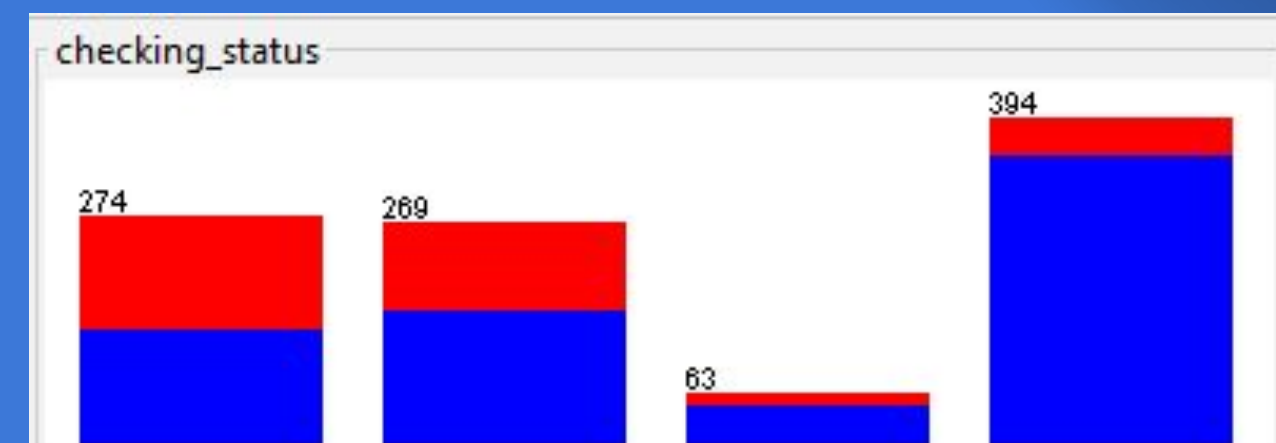
¿QUE SIGNIFICA CADA DATO?



Descripción de los Atributos:

1. **Checking_status**: Estado de la cuenta corriente del cliente:

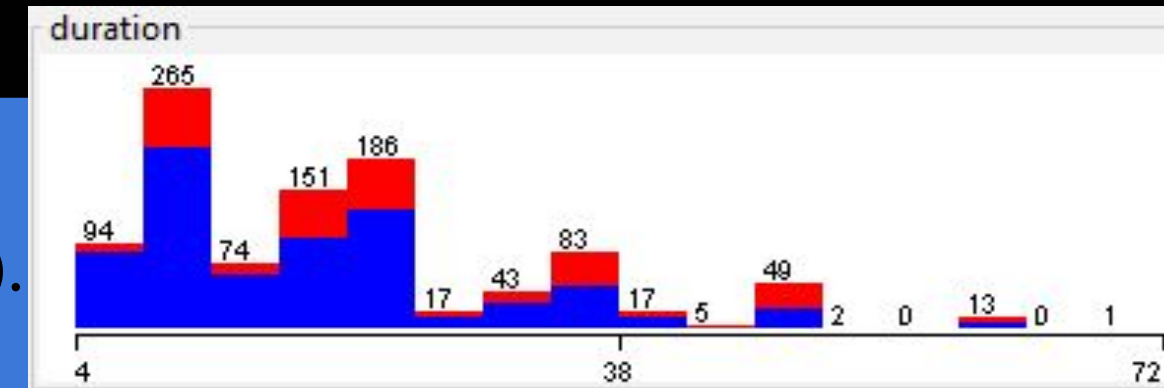
- * <0 : Saldo menor a 0 marcos alemanes.
- * $0 \leq X < 200$: Saldo entre 0 y 200 marcos alemanes.
- * ≥ 200 : Saldo mayor o igual a 200 marcos alemanes.
- * no checking: No posee cuenta corriente.



¿QUE SIGNIFICA CADA DATO?

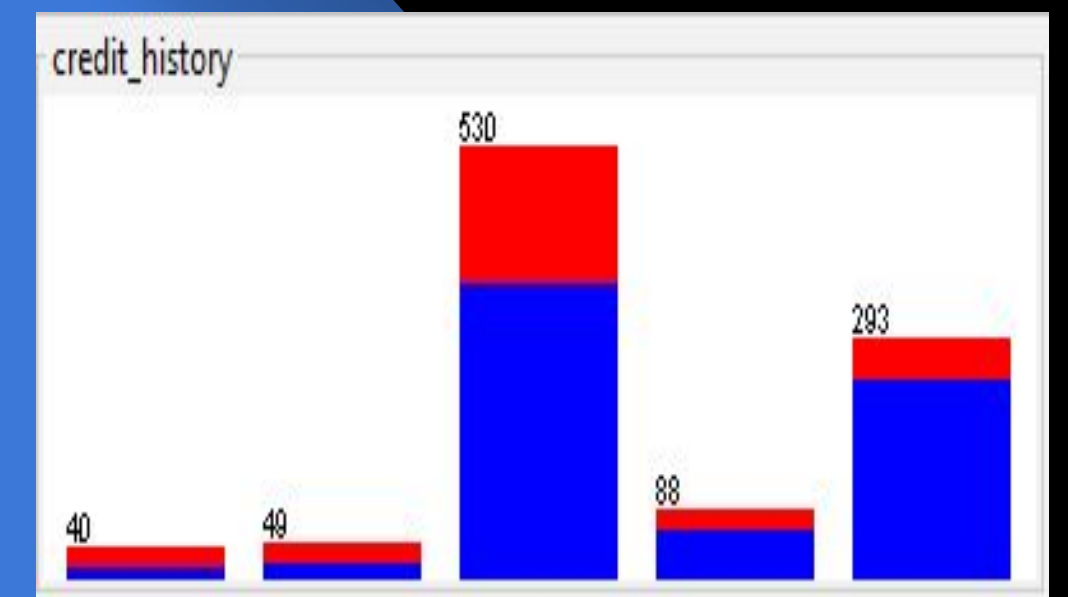


2.duration: Duración del crédito en meses (numérico).



3.credit_history: Historial crediticio del cliente:

- * no credits/all paid: No ha solicitado créditos o todos los créditos han sido pagados.
- * all paid: Todos los créditos en este banco han sido pagados.
- * existing paid: Créditos existentes pagados debidamente hasta la fecha.
- * delayed previously: Retrasos en pagos en el pasado.
- * critical/other existing credit: Cuenta crítica/ existen otros créditos (no en este banco).

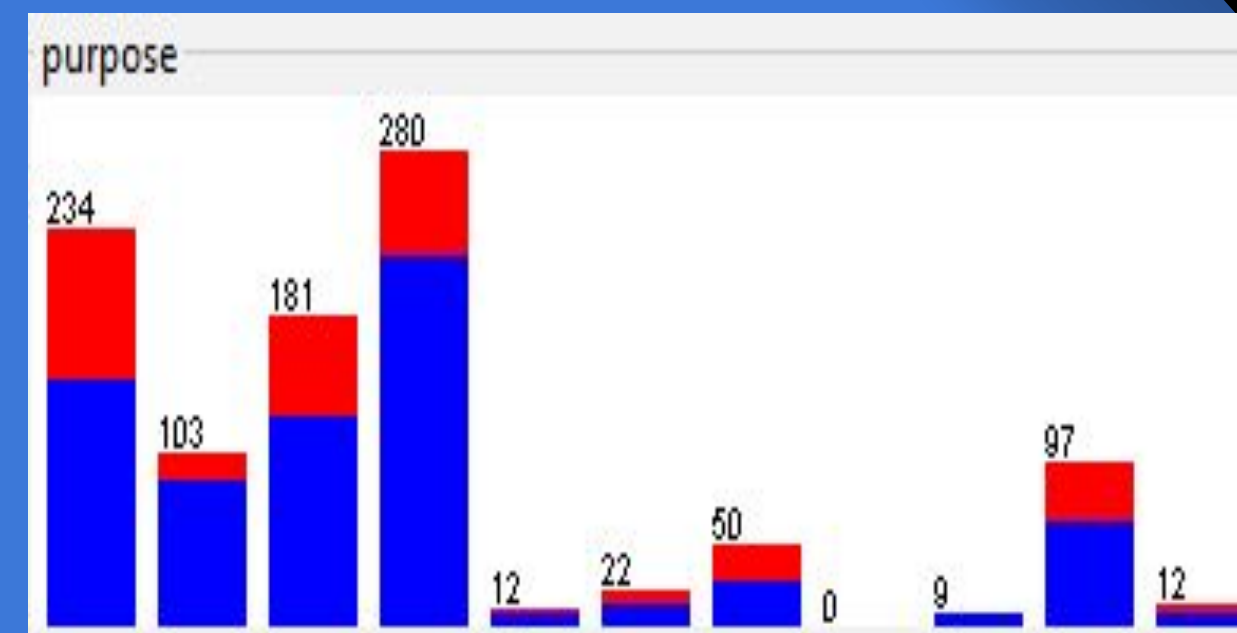


¿QUE SIGNIFICA CADA DATO?



4.Purpose: Propósito del crédito:

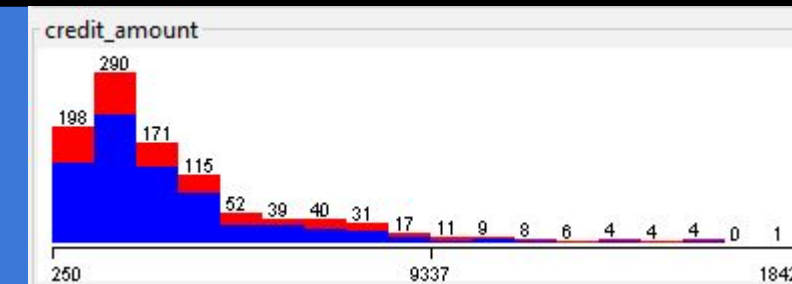
- * new car: Auto nuevo.
- * used car: Auto usado.
- * furniture/equipment: Muebles/ equipamiento.
- * radio/tv: Radio/ televisión.
- * domestic appliance: Electrodomésticos.
- * repairs: Reparaciones.
- * education: Educación.
- * vacation: Vacaciones.
- * retraining: Reciclaje profesional.
- * business: Negocios.
- * other: Otros.



¿QUE SIGNIFICA CADA DATO?

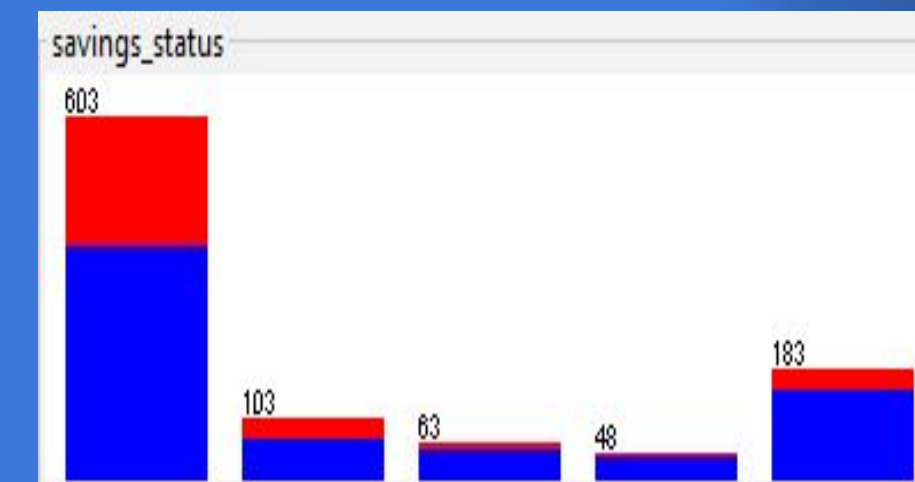


5.credit_amount: Monto del crédito (numérico).



6.savings_status: Estado de la cuenta de ahorros/bonos:

- * <100 : Menos de 100 marcos alemanes.
- * $100 \leq X < 500$: Entre 100 y 500 marcos alemanes.
- * $500 \leq X < 1000$: Entre 500 y 1000 marcos alemanes.
- * ≥ 1000 : Mayor o igual a 1000 marcos alemanes.
- * no known savings: Sin información sobre ahorros.

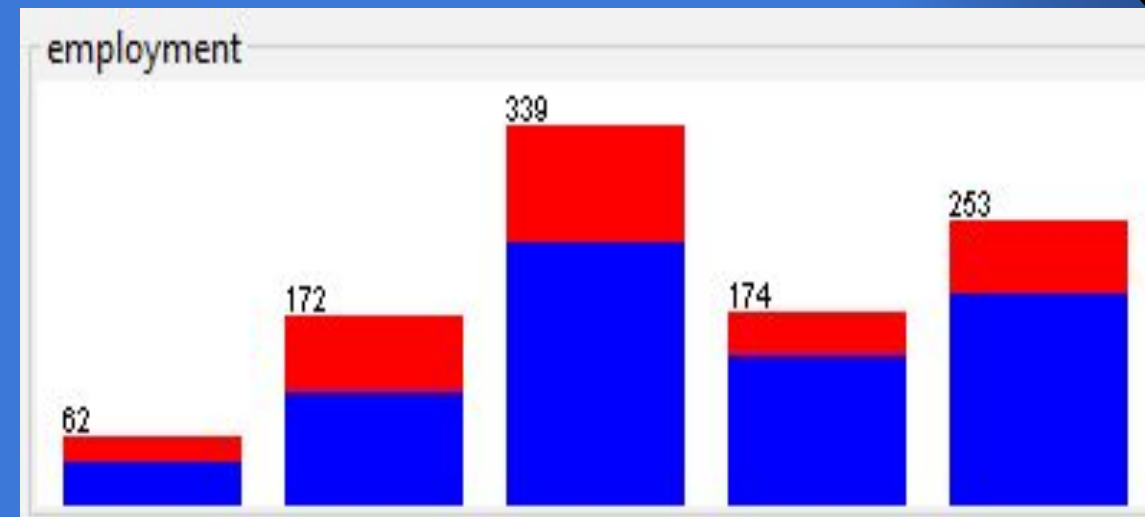


¿QUE SIGNIFICA CADA DATO?

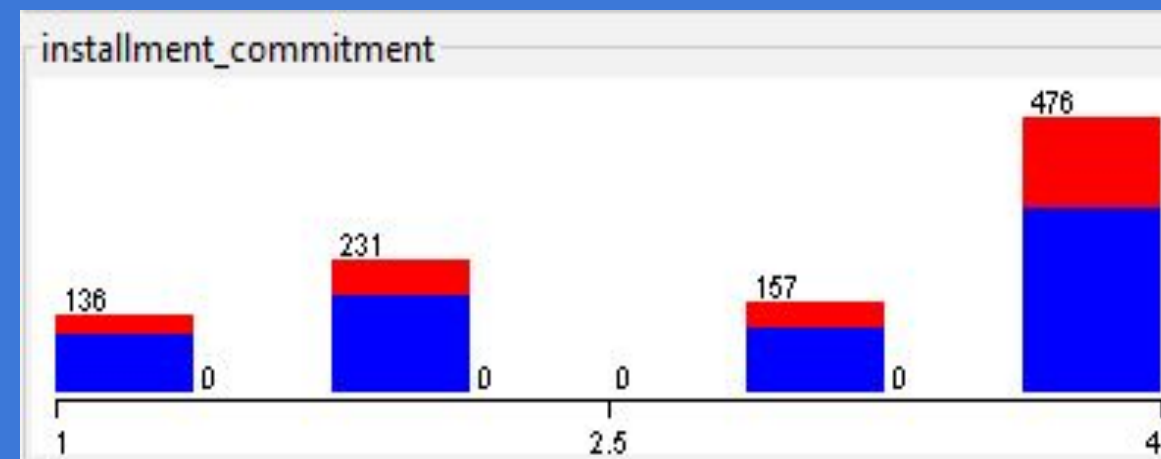


7. employment: Tiempo de empleo:

- * unemployed: Desempleado.
- * <1 : Menos de 1 año.
- * $1 \leq X < 4$: Entre 1 y 4 años.
- * $4 \leq X < 7$: Entre 4 y 7 años.
- * ≥ 7 : Mayor o igual a 7 años.



8. installment_commitment: Porcentaje de ingresos comprometido para el pago de cuotas (numérico).

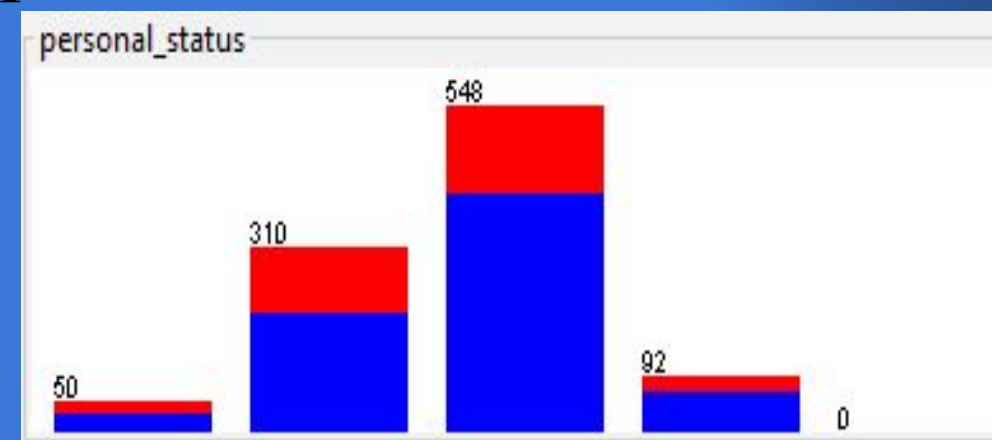


¿QUE SIGNIFICA CADA DATO?



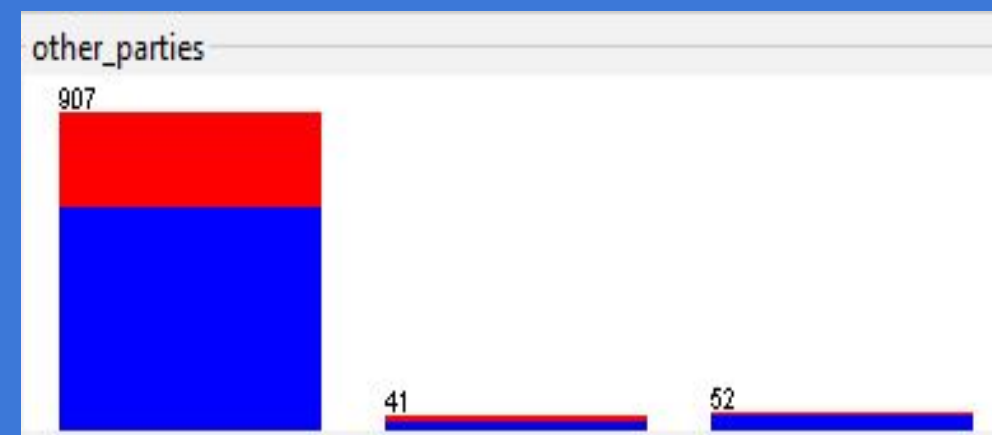
9. **personal_status**: Estado civil y sexo:

- * male div/sep: Hombre divorciado/separado.
- * female div/dep/mar: Mujer divorciada/dependiente/casada.
- * male single: Hombre soltero.
- * male mar/wid: Hombre casado/viudo.
- * female single: Mujer soltera.



10. **other_parties**: Otras partes involucradas en el crédito:

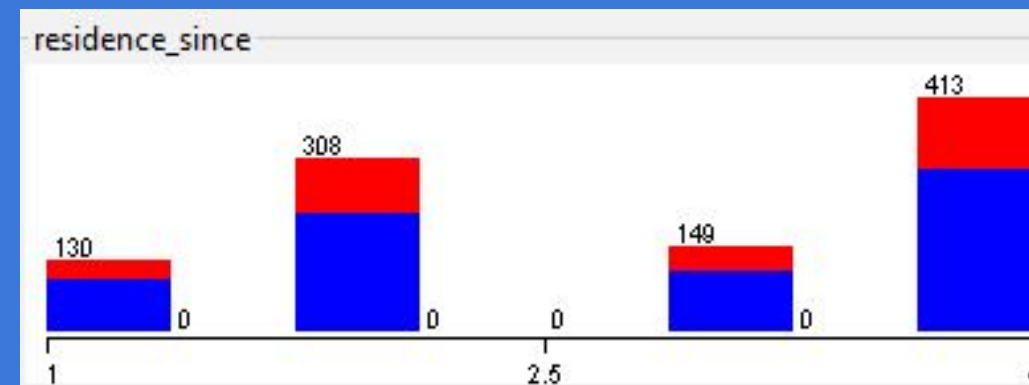
- * none: Ninguna.
- * co applicant: Co-solicitante.
- * guarantor: Garante.



¿QUE SIGNIFICA CADA DATO?

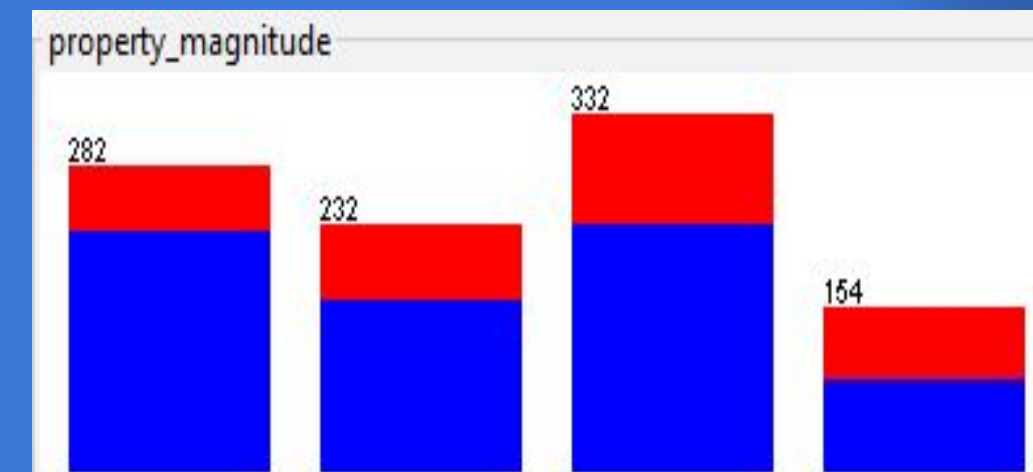


11. residence_since: Tiempo de residencia en años (numérico).



12. property_magnitude: Tipo de propiedad:

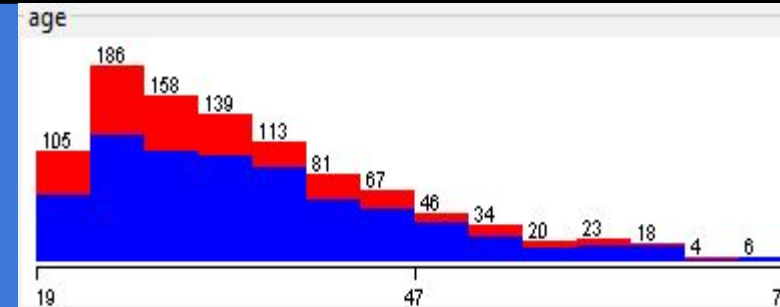
- * real estate: Bienes raíces.
- * life insurance: Seguro de vida.
- * car: Auto.
- * no known property: Sin propiedad conocida.



¿QUE SIGNIFICA CADA DATO?

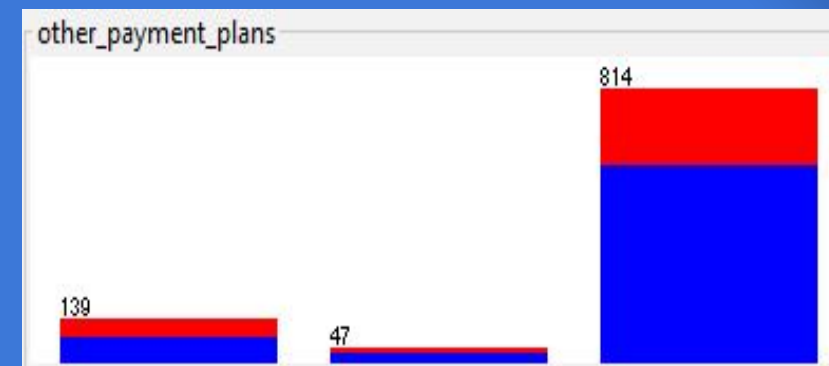


13. age: Edad del cliente en años (numérico).



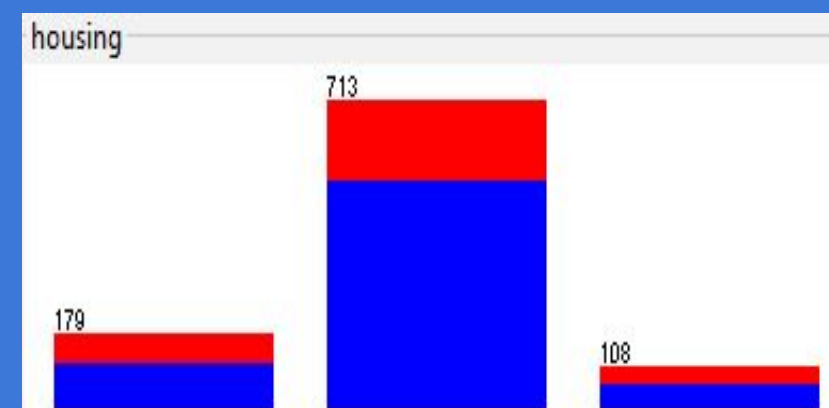
14.other_payment_plans: Otros planes de pago:

- * bank: Banco.
- * stores: Tiendas.
- * none: Ninguno.



15.housing: Tipo de vivienda:

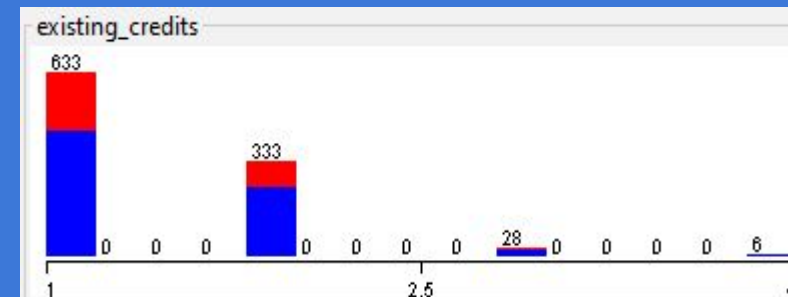
- * rent: Alquilada.
- * own: Propia.
- * for free: Gratuita.



¿QUE SIGNIFICA CADA DATO?



16.existing_credits: Número de créditos existentes en este banco (numérico).



17.job:Trabajo del cliente:

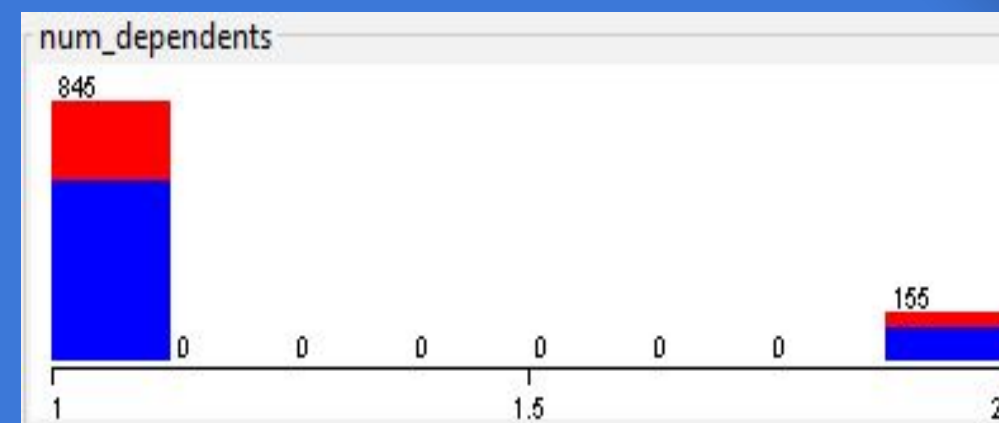
- * unemp/unskilled non res: Desempleado/ no calificado - no residente.
- * unskilled resident: No calificado - residente.
- * skilled: Empleado calificado / oficial.
- *high qualif/self emp/mgmt: Gerencia/ autónomo/ empleado altamente calificado/ oficial.



¿QUE SIGNIFICA CADA DATO?



18.number_of_dependents:* Número de personas que dependen económicamente del cliente (numérico).



19.own_telephone: Posee teléfono propio:

- * none: No.
- * yes: Sí.



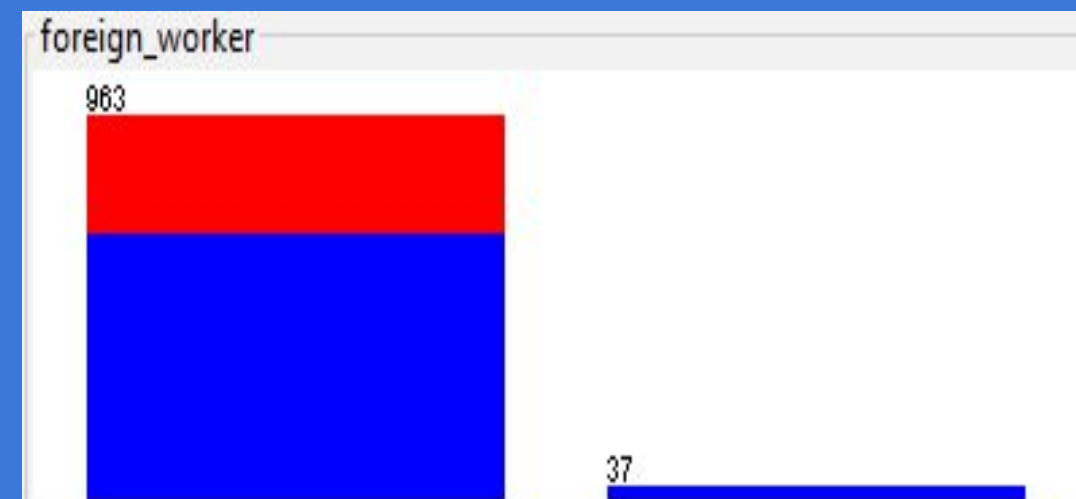
¿QUE SIGNIFICA CADA DATO?



20. **foreign_worker:** Es trabajador extranjero:

* yes: Sí.

* no: No.



3.¿CUÁL ES EL OBJETIVO?



El **objetivo** principal al analizar el archivo credit-g es **desarrollar un modelo predictivo que determine si un nuevo cliente es "fiable" para recibir un crédito**. Este modelo debería minimizar los costes asociados con la concesión de créditos a clientes "no fiables", utilizando la matriz de costes proporcionada.

Información Adicional: La información sobre la matriz de costos y cómo utilizarla en Weka se encuentra en las fuentes. Es importante recordar que la precisión del modelo no es el único factor a considerar; el costo económico de las decisiones erróneas también es crucial en este contexto.

¡MUCHAS GRACIAS!