Modelos para la identificación de fraude en el mercado eléctrico

FuzzyMAD

Daniel Vélez Serrano

- El **fraude** es uno de los **problemas más complejos** a los que se enfrenta una compañía: **elevados costes originados por actividades difíciles de detectar.**

- **Tradicionalmente**, la identificación de estas actividades fraudulentas se basa en **reglas** definidas por la propia compañía de acuerdo a su **experiencia** e **intuición.**

- Sin embargo, los **perfiles del defraudador cambian constantemente** volviéndose **más sofisticados** y en dichos perfiles participan cada vez **más factores** que resultan **difíciles** de **encontrar** de una manera meramente artesanal.

- Dentro de este este contexto, los **modelos analíticos** ofrecen **ventajas competitivas** que buscan **optimizar los beneficios** de la compañía, **aumentando los ahorros generados por fraudes capturados y reduciendo los costes derivados de las inspecciones.**

Tipología: Fraude en comercialización de electricidad

Descripción: Se produce cuando un cliente manipula su contador para que marque (o directamente no marque) un volumen energético inferior al que ha consumido.

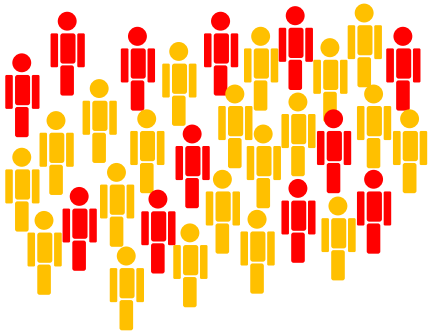
Identificación del fraude:
Un inspector debe comprobar las instalaciones y revisar los datos de consumo para dictaminar si existe fraude.

Se diferencian los conceptos de fraude y anomalía técnica.

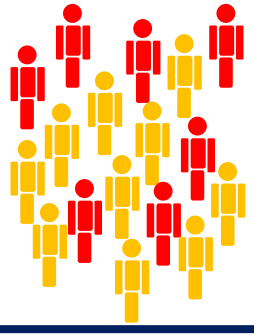
Objetivo: El coste de revisar una instalación sospechosa es alto y por ello resulta importante determinar con precisión las viviendas que deben ser investigadas.

REGLAS DEL JUEGO

CAMPAÑAS

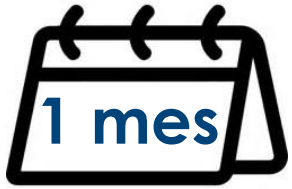


TRAINING



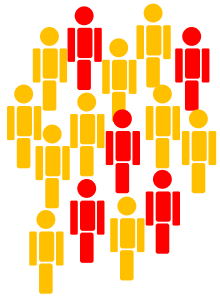
10.000 CUPS

Variables de localización, marca/modelo contador, consumo, inspecciones anteriores, reclamaciones, sociodemográficas, etc.



Ajustar un modelo analítico

TEST

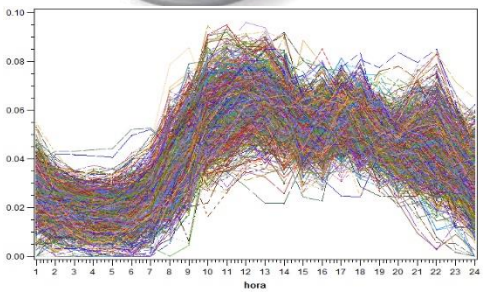


<u>Objetivo</u>

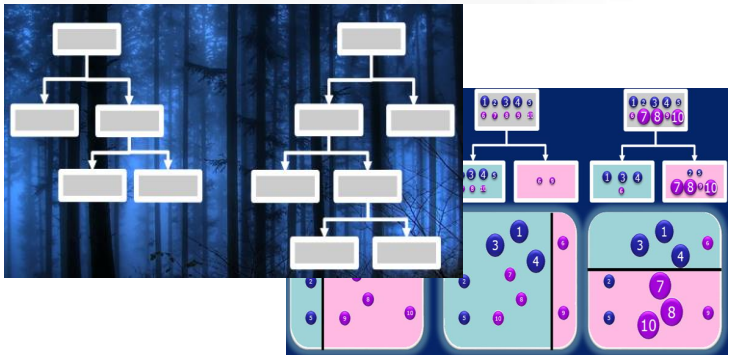
Seleccionar los 100 clientes más propensos a ser fraude

Visión general de la metodología

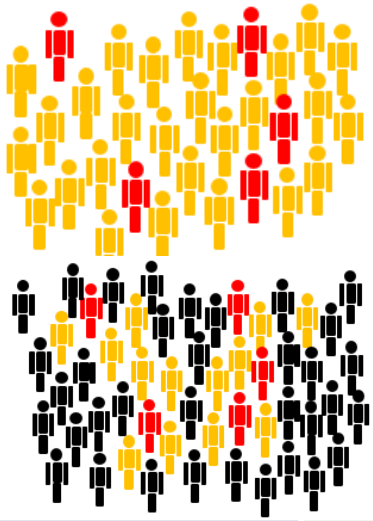
TRATAMIENTO DE DATOS

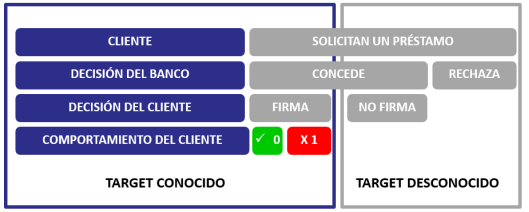


MODELIZACIÓN (I)



MODELIZACIÓN (II)





El grado de detalle que aporta la información de los *smart meter* permite mejorar la tasa de acierto de los modelos.
Por volumetría puede ser necesario recurrir a tecnologías de **Big Data** para su procesamiento

Los tradicionales modelos "estadísticos" (árboles, regresión), buscan la interpretabilidad pero su capacidad de acierto es sensiblemente inferior a los modelos de **Machine Learning: *random forest, gradient boosting*, etc.**

En la modelización de fraude surgen además problemas derivados de la existencia de sesgo en las muestras de datos disponibles: criterios que guían las campañas que hace la compañía.
Por ello, es preciso aplicar técnicas específicas: **Inferencia de rechazo.**

- Grandes volúmenes de datos:

$$2 \times 365 \times 24 \times 10^{7} = 175.200 \times 10^{6}$$

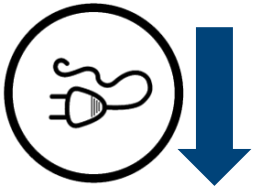
años **días** **horas** **CUPS** **valores**

- Alrededor de 1.000 variables para alrededor de 10.000.000 de CUPS.

CONOCIMIENTO DE NEGOCIO

Variables con poder predictivo

REGLAS TRADICIONALES

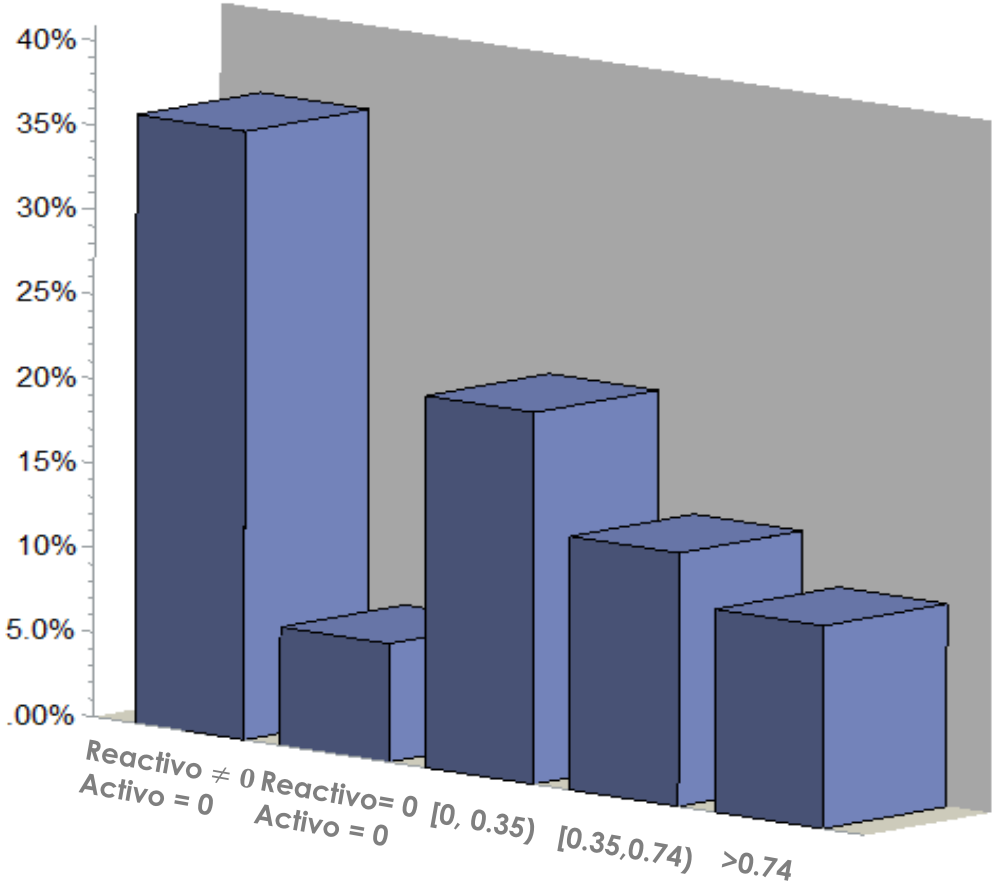


Búsqueda de caídas de consumo

TECNOLOGÍA SMART METER

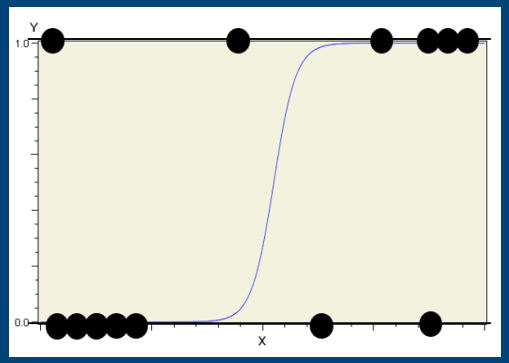
Consumo horario

Energía activa y reactiva



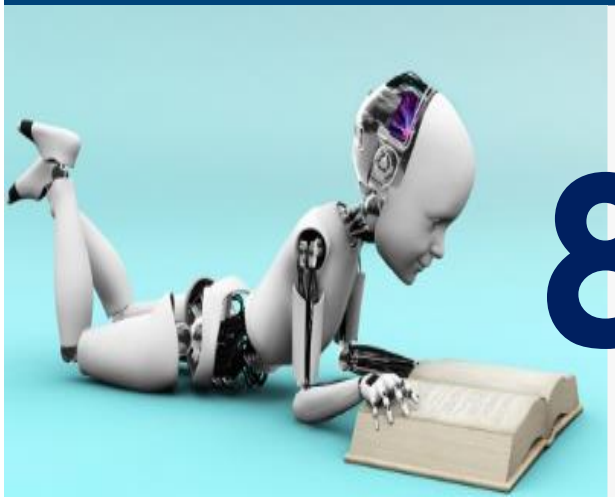
MODELOS DE MACHINE LEARNING

Alta capacidad de interpretación

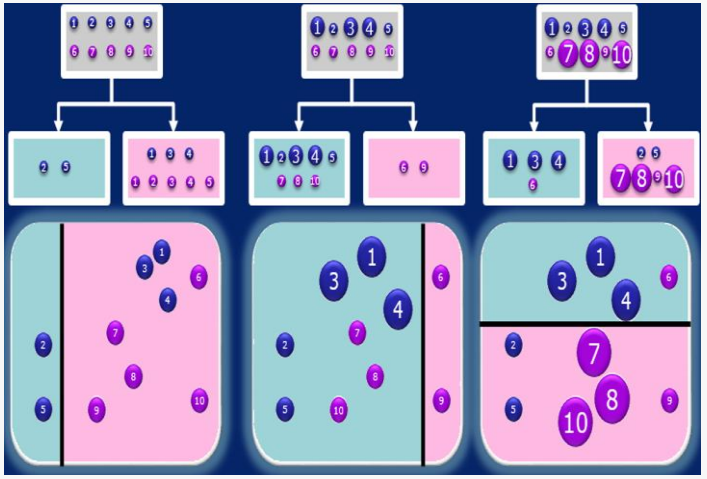


60%

Alta capacidad predictiva

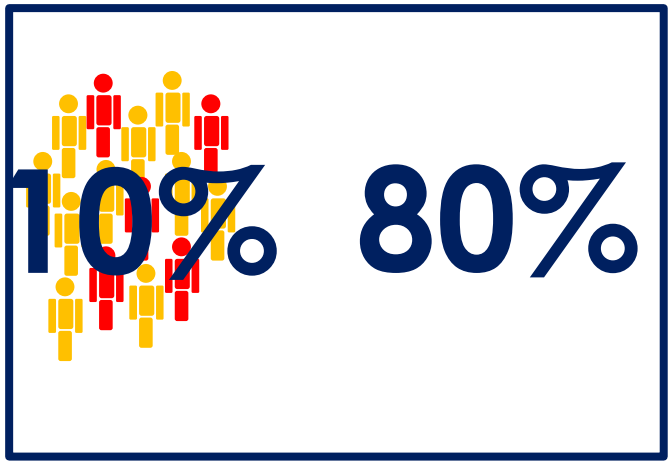


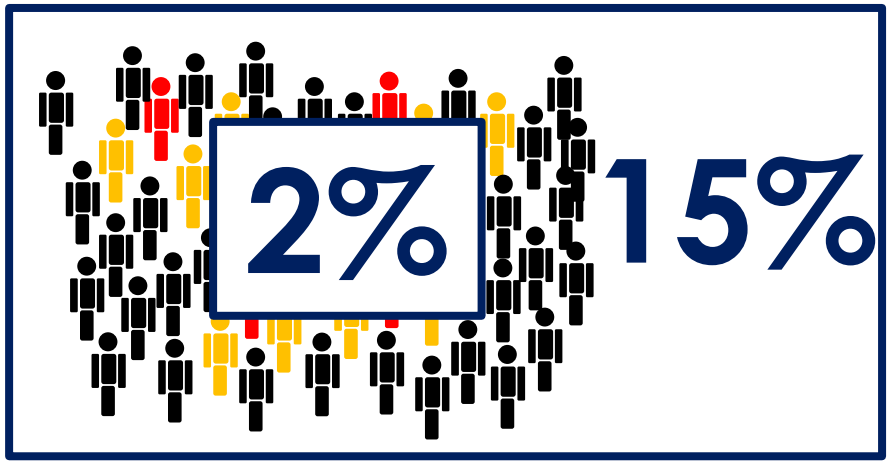
80%



- Al realizar inspecciones ("**salir a campo"),** donde la tasa real de fraude es mucho menor, **las tasas de acierto caen drásticamente.**

- Los modelos se entrenan con clientes que han sido inspeccionados (sospechosos) y por tanto, **las tasas de acierto obtenidas son verosímiles sobre una población sospechosa.**



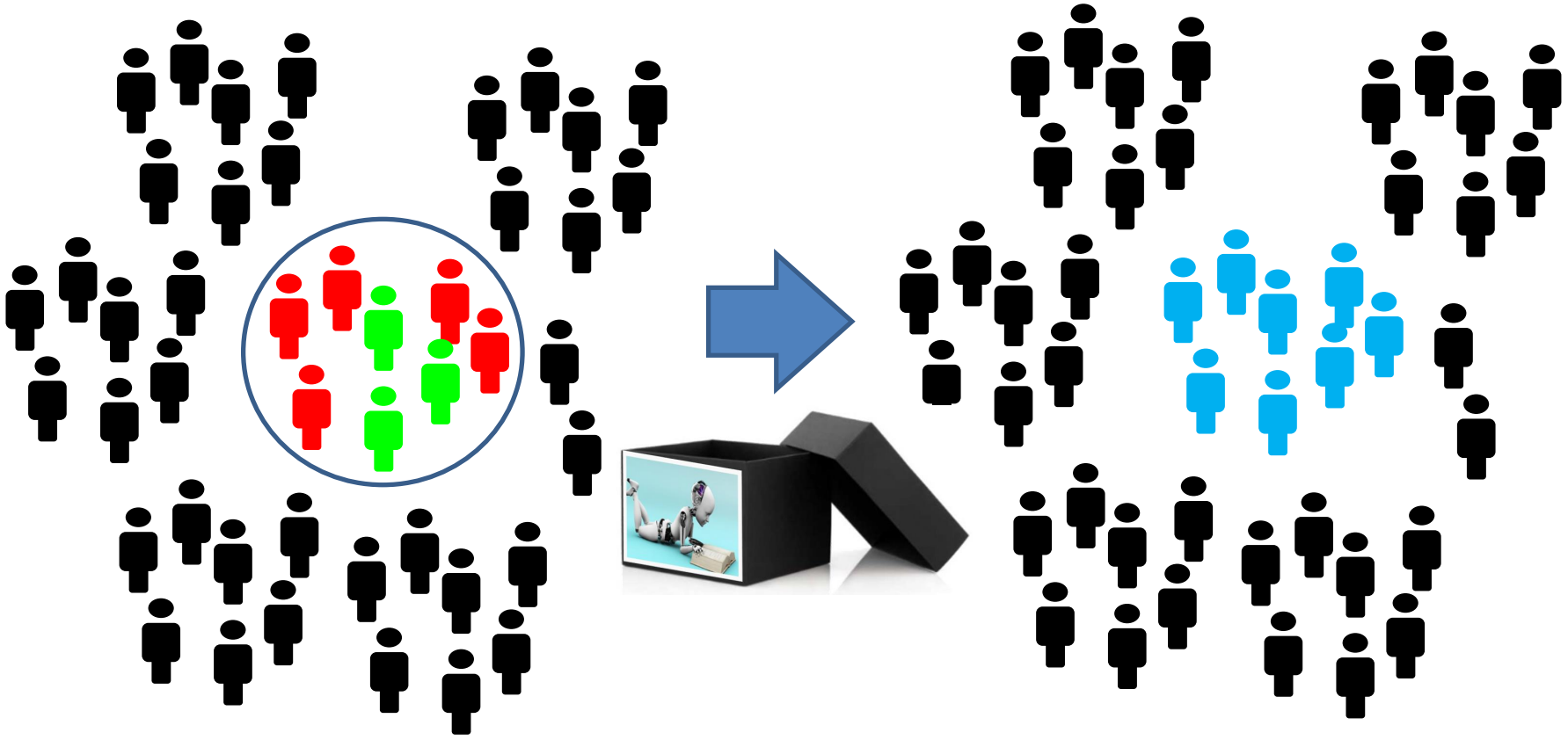


¿Cuáles son los criterios con los que la compañía realiza las inspecciones?

Método de *Augmentation*

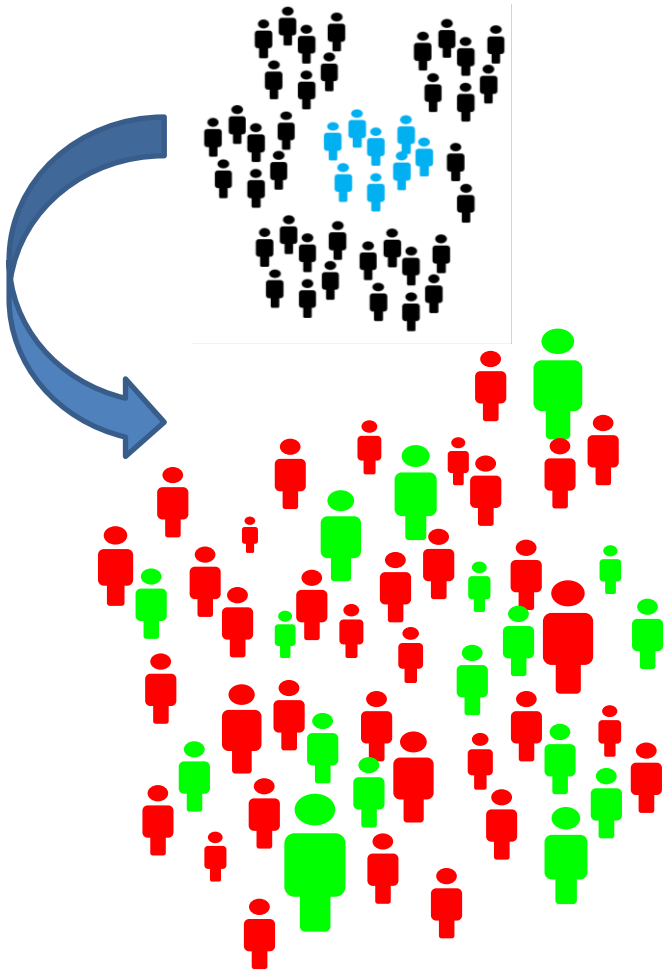
1. Se ajusta un modelo para predecir los criterios de inspección

2. El modelo calcula la probabilidad de ser inspeccionado para cada cliente



Método de *Augmentation*

3. Se da más peso a aquellos clientes que tienen menor probabilidad de ser inspeccionados, y para los cuales se conoce el target



4. Los CUPS a los que el modelo da más peso son más similares a los que se inspeccionarían a través de una muestra aleatoria (útil para detectar nuevas reglas de fraude)

5. Se aplica un nuevo modelo sobre la muestra con las observaciones ponderadas

FASE III: Inferencia de rechazo

Las técnicas de **INFERENCIA DE RECHAZO** surgen en el ámbito del ***CREDIT SCORING:*** Las entidades bancarias disponen de una relación de reglas en función de las cuales deciden conceder o no préstamos.

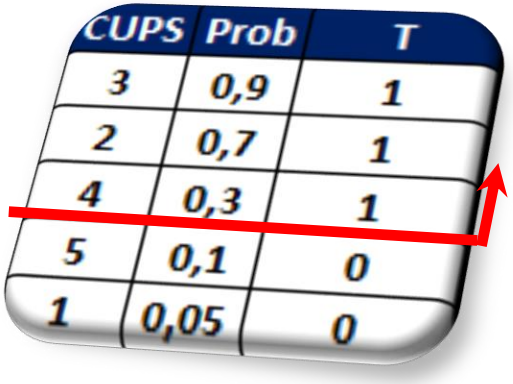
En consecuencia, la **INFORMACIÓN** que queda almacenada **NO ESTÁ SIEMPRE COMPLETA**: para solicitudes rechazadas solo se dispone de las explicativas.

Se busca **inferir** el target para solicitudes que han sido **rechazadas** (clientes no inspeccionados)

IMPORTANTE: En problemas de fraude, el concepto de "no inspeccionado" ("rechazado") no implica necesariamente que el cliente no sea sospechoso.

Método Hard cutoff.-

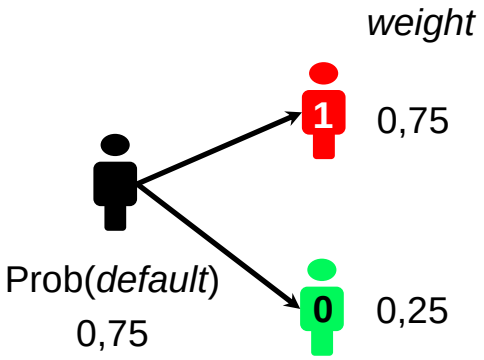
- Se **puntúa con el modelo ajustado a toda la población rechazada.**
- Se **establece un valor probabilístico** por encima del cual el target será considerado 1 y por debajo del cual será considerado 0.
- Se ajusta un **modelo que se entrene tanto con la muestra de clientes aceptados como con los rechazados con target inferido.**



CUPS	Prob	T
3	0,9	1
2	0,7	1
4	0,3	1
5	0,1	0
1	0,05	0

Método Fuzzy.-

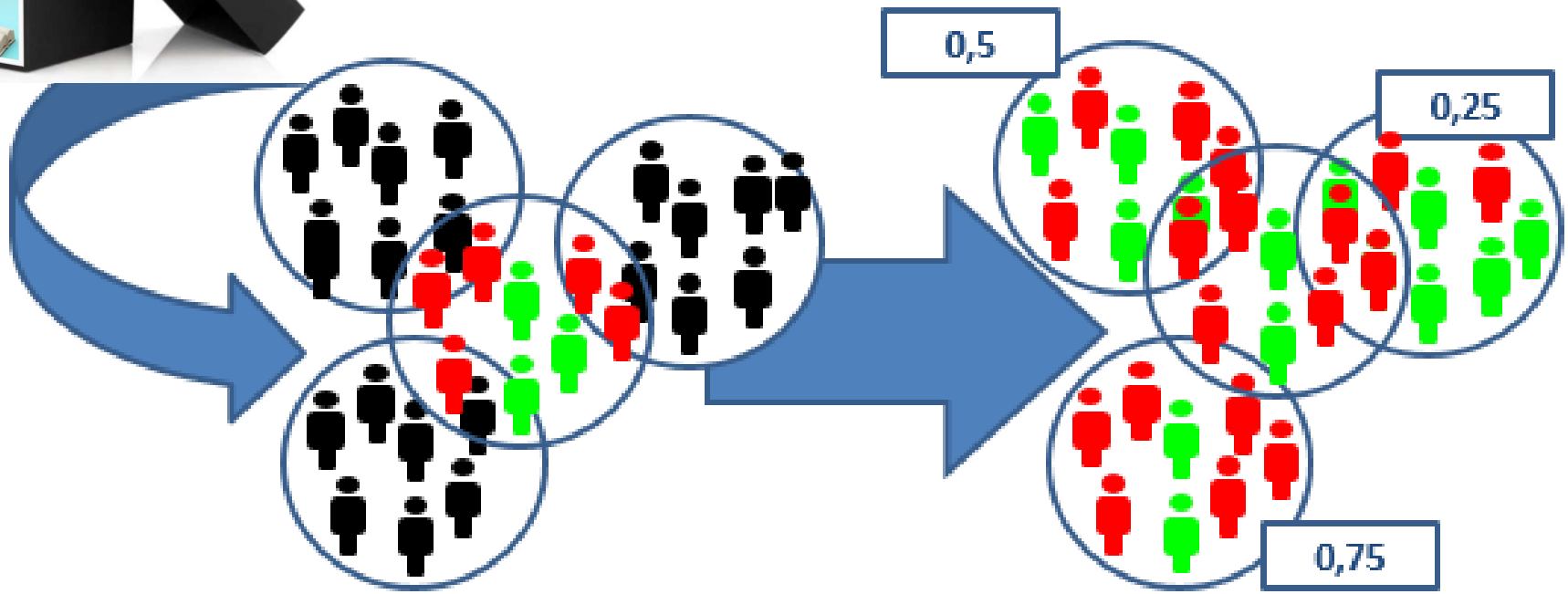
- Se **puntúa con el modelo ajustado a toda la población rechazada.**
- **Por cada observación de dicha población se generan dos observaciones: una a la que se asigna target 1 (morosidad) y otra a la que se asigna target 0** (no morosidad).
- A cada una de dichas observaciones se asigna **un peso proporcional a la probabilidad** de la clase en cuestión.
- Se ajusta un **modelo que se entrene tanto con la muestra de clientes/solicitudes aceptadas como con la muestra rechazada ponderada.**



- ***Método de Parcelling.-***

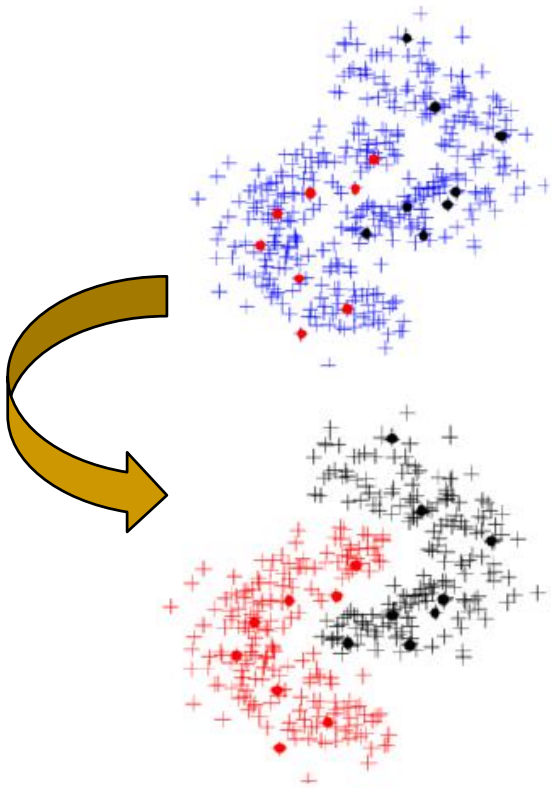
1. Se aplica el modelo sobre los **clientes NO inspeccionados**

2. Se generan parcelas de clientes (NO inspeccionados) con probabilidad parecida de fraude y **se asignan 1's y 0's aleatoriamente** en dichos grupos **de acuerdo a la probabilidad** proporcionada por el modelo



3. Se hace un último modelo con Inspeccionados + No inspeccionados (con target inferido)

En la línea del método de Parceling se podría utilizar una **técnica de modelización No Supervisada:**

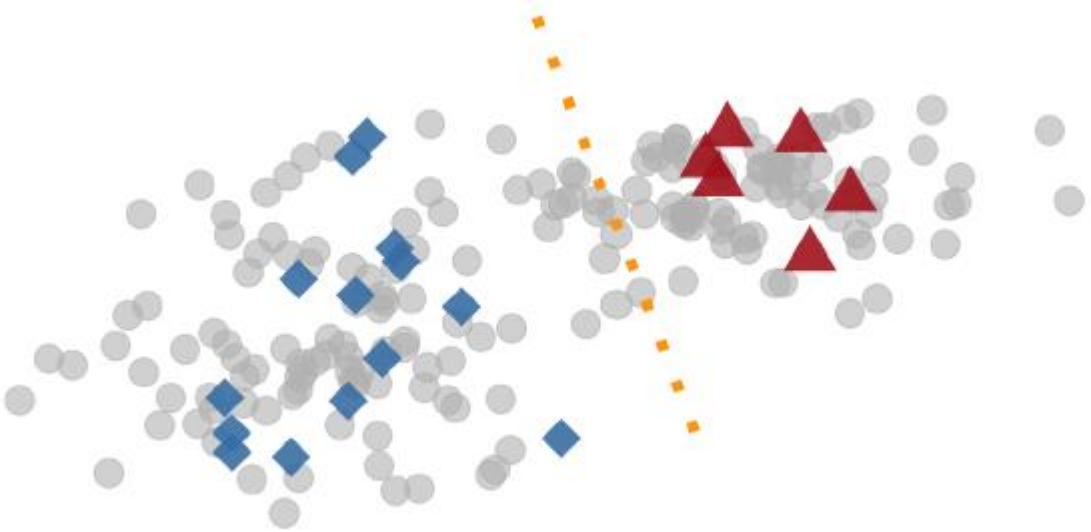


- **Agrupar clientes inspeccionados y no inspeccionados (k-means, Kohonen, etc.).**
- **Etiquetar a los no inspeccionados de cada cluster** de forma aleatoria pero **respondiendo a la tasa de fraude observada** en los inspeccionados.
- **Ajustar un modelo que se entrene con clientes aceptados y rechazados con target inferido.**

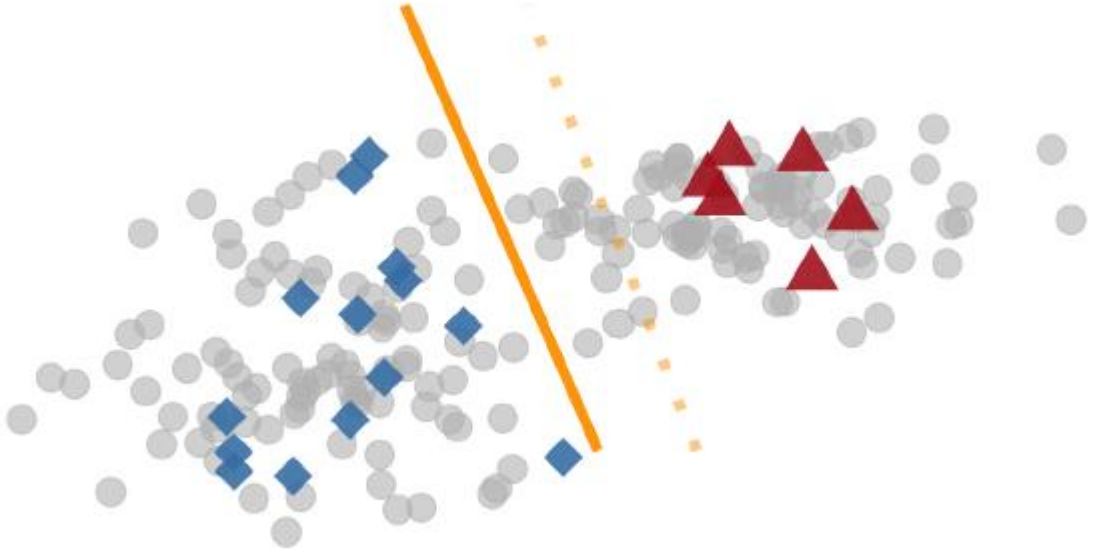
Las **técnicas de inferencia de rechazo pueden arrastrar el sesgo de los criterios de inspección** de la compañía. En este contexto, **también las técnicas de clustering pueden ser una alternativa interesante para identificar casos raros** (outliers multivariante) sospechosos de ser fraude.

Técnicas alternativas para la identificación de fraude: Modelización semi - supervisada

- La combinación de modelos supervisados y no supervisados da lugar a la **modelización semi – supervisada.**



Separación construida de acuerdo a las observaciones etiquetadas (supervisado)

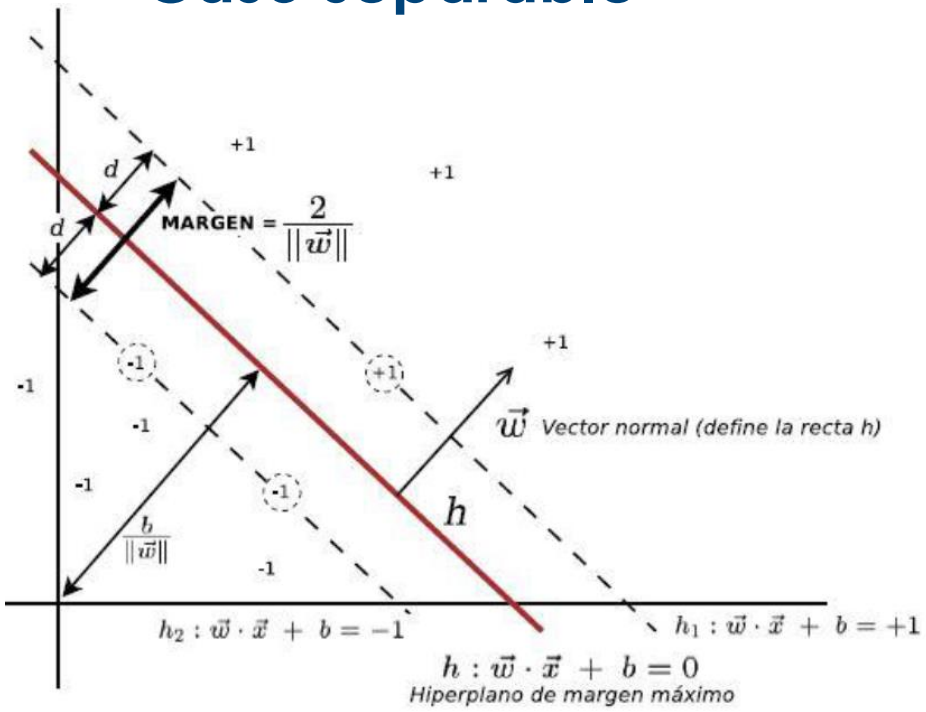


Separación construida de acuerdo a las observaciones etiquetadas y no etiquetadas (semi - supervisado)

- La modelización mediante SVM busca determinar el hiperplano de máximo margen ($\max_w \frac{2}{\|\vec{w}\|}$) que permite separar las clases.

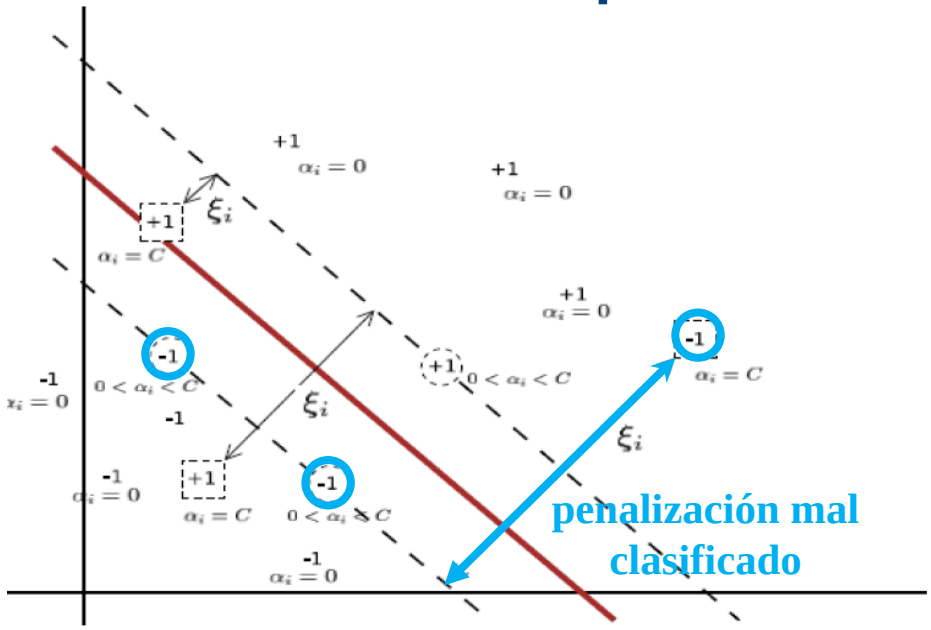
SVM

Caso separable



$$\min_w \frac{1}{2} \|\vec{w}\|^2$$

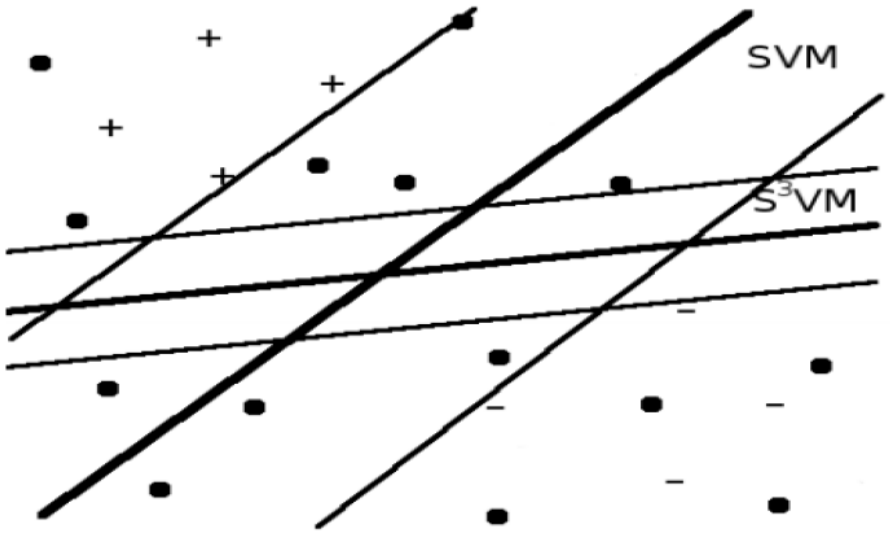
Caso cuasi - separable



$$\min_w \frac{1}{2} \|\vec{w}\|^2 + C \sum_{i=1}^{l} \xi_i$$

- Las **etiquetas asociadas a los registros para los que se desconoce la variable target (no inspeccionados)** se convierten en **variables de decisión** a inferir **(inferencia de rechazo)**, imponiéndose una restricción adicional que hace referencia al máximo número de 1's (tasa de fraude) a asignar.

S^3VM = Semi Supervised SVM



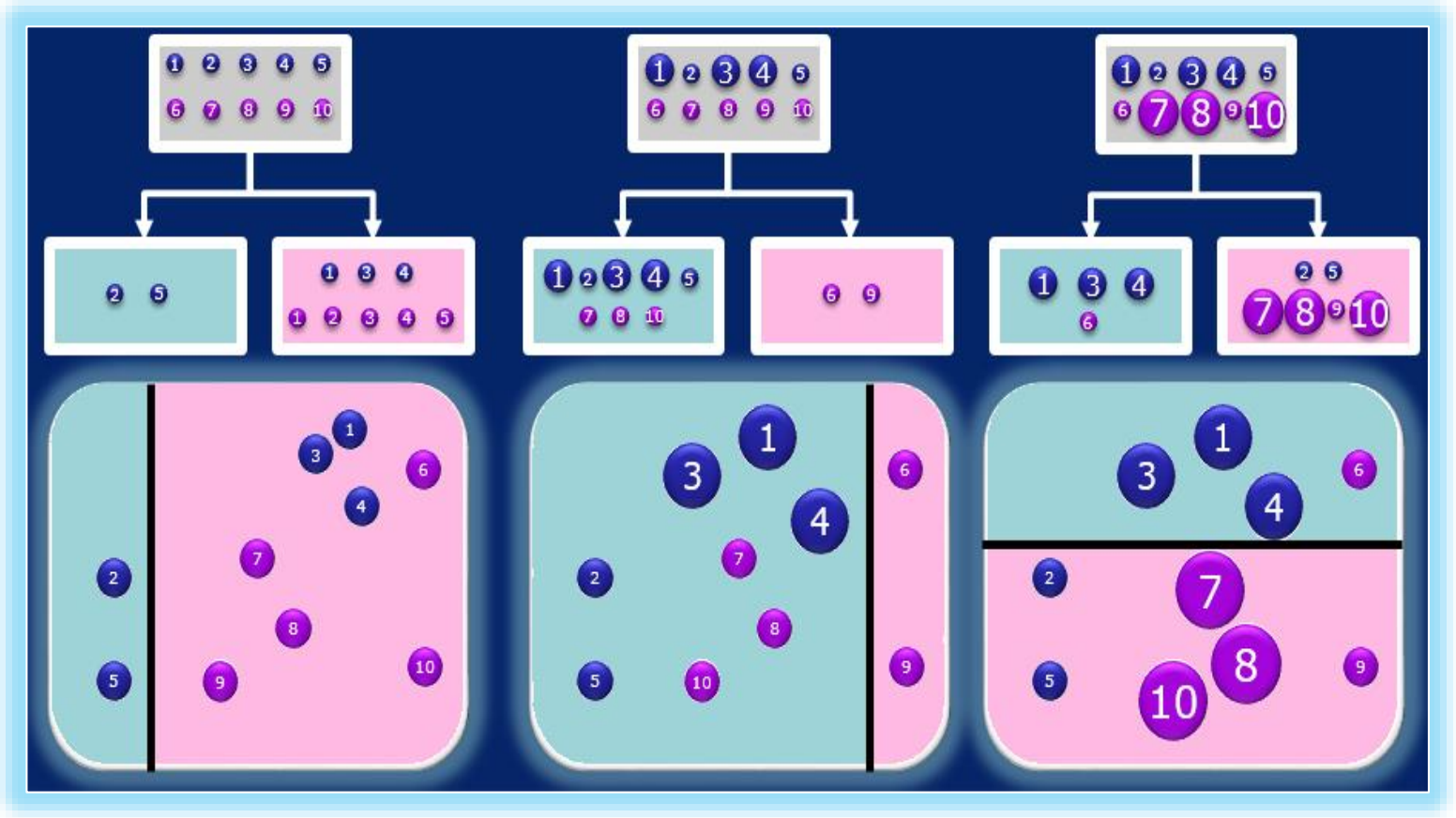
ID	Y	X1	X2	…	Xp
1	0	-3	Mad	…	2/3
2	1	2	Sev	…	1
…	…	…	…	…	…
L	0	2	Bar	…	-3/4
L+1	NA	5	Mad	…	1/2
…		…	…	…	…
n	NA	1	Val	…	0

$$\min_{w} \frac{1}{2}\|\vec{w}\|^2 + C\sum_{i=1}^{l}\xi_i + C^*\sum_{i=l+1}^{n} y_i$$

$$\sum_{i=l+1}^{n} y_i = \mathrm{r}$$

- Los modelos tipo Boosting se construyen por **ensamblaje iterativo de clasificadores débiles** ajustados a muestras sucesivas en las que **las observaciones peor clasificadas tienen más peso.**

Boosting



- Formalmente, el algoritmo es el siguiente:

  - Partir de un clasificador inicial $\widehat{h_0}(x) = argmin_\alpha \sum_{i=1}^{N} L(target_i, \alpha)$
  - Iterar desde j = 1,…J (tantas como clasificadores débiles se pretenda utilizar):
    - Para i=1,2,…,N, calcular el residual $residual_{ij} = -[\frac{\partial L(target_i, \widehat{h_{j-1}}(x_i))}{\partial \widehat{h_{j-1}}(x_i)}]$ y ajustar un árbol de decisión a los $residual_{ij}$, de modo que se obtengan regiones terminales $R_{kj}, k = 1,2,\dots,J_j$
    - Para $k = 1,2,\dots,J_j$ calcular $\alpha_{kj} = argmin_\alpha \sum_{x_i \epsilon R_{kj}} L(target_i, \widehat{h_{j-1}}(x_i) + \alpha)$
    - Calcular $\widehat{h_j}(x) = \widehat{h_{j-1}}(x) + \sum_{k=1}^{J_j} \alpha_{kj} I_{(x \epsilon R_{kj})}$
  - Clasificador fuerte: $\widehat{H}(x) = \widehat{h_J}(x)$

- El **algoritmo SemiBoost** es una versión no supervisada del algoritmo que busca **satisfacer dos condiciones:**

  - **Observaciones no etiquetadas (u = unlabeled) parecidas a otras etiquetadas (l =labeled) deben tener la misma etiqueta.**

    La inconsistencia entre unas y otras etiquetas se mide mediante:

$$F_l(y,S)=\sum_{i\in etiquetado}\sum_{j\in noEtiquetado} S_{ij}\exp(-2y_i^l y_j^u)$$

  - **Observaciones no etiquetadas parecidas deben tener una misma etiqueta.**

$$F_u(y_u\ ,S)=\sum_{i,j\in noEtiquetado} S_{ij}\exp(y_i^u - y_j^u)$$

    siendo S la matriz de similaridad (parentesco) entre las observaciones.

- Asignar valores {+1,-1} a las etiquetas desconocidas para que la **inconsistencia total,** $F(y,S)= F_l(y,S) + C^* F_u(y_u\ ,S)$, **sea mínima** donde C* es un parámetro que pondera la importancia que tienen los datos no etiquetados.

- El **algoritmo SemiBoost** incorpora estas condiciones sobre el boosting original:
  - Calcular la matriz de similaridad S_{ij} entre todos los puntos e inicializar $\widehat{\mathrm{H}}(\mathrm{x})$
  - Iterar desde j = 1,…J (tantas como clasificadores débiles se pretenda utilizar):
    - Calcular a partir de S, las confianzas que se tiene de que cada observación no etiquetada "i" pertenezca a cada clase (p_i y 1- p_i en el caso binario).
    - Asignar las observaciones no etiquetadas a la clase con mayor confianza.
    - Ponderar las observaciones no etiquetadas de acuerdo a su confianza (cuando p_i = 0,5 la confianza es mínima y se da peso 0).
    - Aplicar el algoritmo de boosting sobre las observaciones etiquetadas y las no etiquetadas ponderadas, para obtener el clasificador débil $\widehat{h_j}(\mathrm{x})$ y los pesos α.
  - La predicción final la proporciona el clasificador fuerte $\widehat{\mathrm{H}}(\mathrm{x}) = \widehat{h_J}(\mathrm{x})$.

- Se trata de resolver $\underset{h(x),\alpha}{\arg\min} \sum_{i \in etiquetado} \sum_{j \in noEtiquetado} S_{ij} \exp(-2y_i^l(H_j + \alpha h_j)) +$

  $+ C^* \sum_{i,j \in noEtiquetado} S_{ij} \exp(H_i - H_j) \exp(\alpha(h_i - h_j))$ sujeto a $h(x_i) = y_i^l \ \forall i \ etiquetado$

El **FRAUDE** es un problema analítico **COMPLEJO**

Tradicionalmente, la identificación se ha hecho en base a **EXPERIENCIA** e **INTUICIÓN**

Los **PERFILES del defraudador cambian constantemente: MÁS SOFISTICADOS**

Dentro de este este contexto, los **modelos analíticos** ofrecen **VENTAJAS COMPETITIVAS**

Preguntas