1. ETL Extract Transform Load

* Obtener los datos en un Excel y analizar cuáles sirven, depurar valores, ver si hay algo extra para calcular a priori.
  + Preguntas sobre los valores de los datos:
    - Se cambiaron los -- y vacíos por “no”, donde correspondía
    - Se normalizaron todos los valores de Plan 1988 y 88 a 88, ya que son equivalentes.
    - Los valores de Plan S95 y 1995 non son equivalentes.
    - Se marcaron como “usables = 0” los registros incompletos para no usarlos hasta tener la información necesaria. Faltan datos de Plan y adeuda trab final con ?
    - Una vez realizados estos cambios, se cambiaron los valores de sí y no por si=1 y no=0, para utilizar booleanos en la BD.
  + Se agregó el dato de Alumnos Recibido al final del 1er año de Reválidas (SI/NO)
* Importar los datos a la BD a partir de un Excel
  + Definir DBMS: PostgreSQL- listo
  + Copy con archivo CSV: cambiar permisos de lectura en el disco o mover el archivo a donde el usuario pueda abrirlo.

COPY public.prueba FROM 'D:\prueba.csv' USING DELIMITERS ';' with csv

* Agregar Store Procedures si es necesario calcular datos derivados a partir de las variables originales

1.2 Variables Involucradas

2. Calcular la Clasificación de Base para ver la precisión mínima a alcanzar

Total de alumnos inscriptos=123

Graduados al finalizar R1=50

Clasificación de Base = 40,65% de presición

3. Calcular con los distintos métodos sobre las distintas variables para ver cuáles podrían ser índices viables (según la precisión). Incluyendo el dato “Alumno Recibido” (S/N).

* OneR

**Script**

# query the data from postgreSQL

df\_postgres <- dbGetQuery(con, "SELECT anio\_ingreso,finales\_adeudados,cursadas\_adeudadas,optativas\_adeudadas,adeuda\_trab\_final\_inicio,plan,tutor\_r1, recibido\_fin\_r1 from revalidas WHERE usable=TRUE")

#df\_postgres <- dbGetQuery(con, "SELECT finales\_adeudados,cursadas\_adeudadas,optativas\_adeudadas,adeuda\_trab\_final\_inicio,plan,tutor\_r1, recibido\_fin\_r1 from revalidas WHERE usable=TRUE")

df\_postgres [ , ]

str(df\_postgres)

#install.packages(OneR)

require("OneR")

model <- OneR(df\_postgres, verbose = TRUE)

summary(model)

plot(model)

prediction <- predict(model, df\_postgres)

eval\_model(prediction, df\_postgres)

**Resultados grales:**

Attribute Accuracy

1 \* anio\_ingreso 64.75%

2 plan 62.3%

3 finales\_adeudados 59.84%

4 cursadas\_adeudadas 59.02%

4 optativas\_adeudadas 59.02%

4 adeuda\_trab\_final\_inicio 59.02%

4 tutor\_r1 59.02%

* J48
* …?

Analizar los resultados para cada combinación método-variable/s (anotar).

4. Automatizar el proceso de carga de datos en BD, cálculo del/los indicador/es y presentación de resultados.