Preprocesamiento y limpieza de datos problema 2:

Problema2v1: contiene el extracto con todos los registros obtenidos desde la plataforma del gss explorer.

Problema2v2: contiene la primera limpieza del extracto, donde se elimina un campo generado automáticamente para separar los registros de cada año

Problema2v3: se eliminan los registros que están asociados a la balota A y de la balota D, debido a que generan una gran distorsión, ya que casi todo el contenido está asociado según la encuesta al valor “No Aplica”.

Problema2v4: se establece una discretización de la variable year\_encuesta, para que se visualicen como rangos de tiempo para poder utilizar reglas de asociación y determinar el grado de efectividad o si por el contrario no se obtienen resultados que ayuden a la predicción del voto.

Problema2v5v1: se llevó a cabo una discretización de la variable afiliación\_politica, para disminuir la cantidad de campos contenidos en la misma, disminuyendo la dispersión de registros, y así intentar la mejora de los modelos de predicción, se hizo de tal manera en que quedaron: los valores ind,near rep e ind,near dem se unieron al valor independiente.

Problema2v5v2: se llevó a cabo una discretización de la variable afiliación\_politica, para disminuir la cantidad de campos contenidos en la misma, disminuyendo la dispersión de registros, y así intentar la mejora de los modelos de predicción, se hizo de tal manera en que quedaron: los valores ind,near rep e ind,near dem se unieron a republicano y demócrata respectivamente.

Problema2v6v1: se eliminan registros que aparecen como “no aplica” en el rango de años 2002 y 2004

Problema2v6v2: se eliminan registros que aparecen como “no aplica” en el rango de años 2002 y 2004