

# Análisis y clasificación de las denuncias remitidas a la Secretaría de Transparencia



Unidad de Científicos de Datos  
2017



**DNP** Departamento  
Nacional  
de Planeación

## Contenido

1.	Metodología de modelo de regresión logística para denuncias .....	3
1.1.	Introducción .....	3
1.2.	Fuente de datos.....	7
1.3.	Metodología .....	8
1.4.	Resultados .....	9
1.5.	Conclusiones.....	12
1.6.	Glosario.....	12

# 1. Metodología de modelo de regresión logística para denuncias

En este proyecto dado el crecimiento sostenido en las denuncias recibidas por la ST por casos de corrupción se generó la necesidad de desarrollar un proyecto de analítica de datos para proveer herramientas de prevención e intervención.

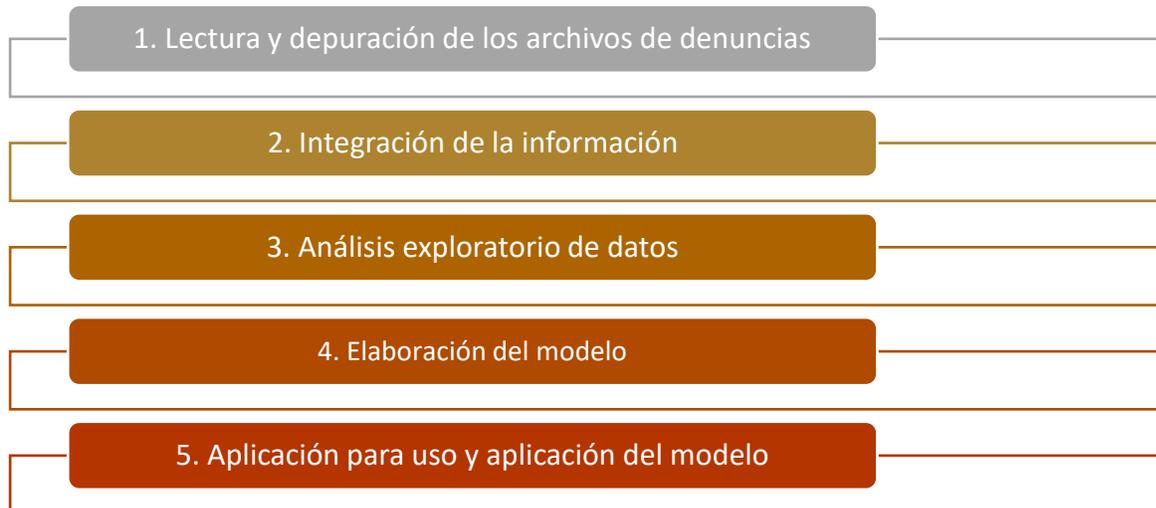
## 1.1. Introducción

El objetivo general de este proyecto es desarrollar una herramienta predictiva que sugiera a qué ente de control deben ser remitidas las denuncias.

Los objetivos específicos de los proyectos son:

- ✓ Construir una base de datos con las denuncias enviadas a la ST (2015 – 2017) y sus remisiones a los entes de control cuando es el caso.
- ✓ Visualizar patrones de ocurrencia de las denuncias en el territorio nacional, por diferentes variables y detectar las relaciones existentes entre las denuncias con las entidades de control a las que son enviadas
- ✓ Desarrollar un modelo predictivo que, con la información estructurada de las denuncias, permita predecir a qué entes de control deben ser estas remitidas.

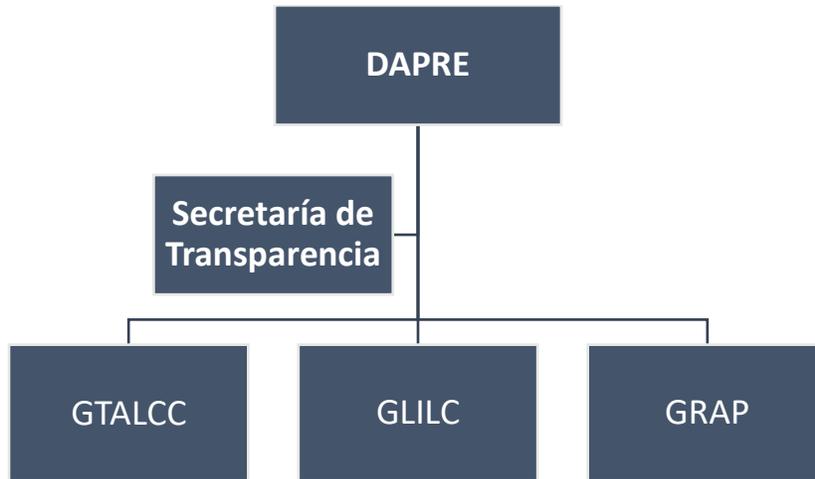
El proyecto se desarrolla en cinco diferentes fases:



La Secretaría de Transparencia es una dependencia misional del Departamento Administrativo de la Presidencia de la República. Su objeto es el de asesorar y asistir al director de esta entidad en la formulación y diseño de políticas públicas en materia de transparencia y lucha contra la corrupción. Asimismo, tiene a su cargo la Secretaría Técnica de la Comisión Nacional de Moralización. (ST, 2014).

El grupo de revisión de análisis de peticiones, denuncias y reclamos de corrupción se encarga de acopiar y analizar las denuncias para su posterior remisión a las entidades de control y otras instituciones.

Otros grupos de la Secretaría de Transparencia tienen un rol importante en el análisis de las denuncias, como son el Grupo de Política de Transparencia, Acceso a la Información y Lucha contra la Corrupción (GTALCC) y el Grupo Laboratorio de Innovación en Lucha contra la Corrupción (GLILC)



Desde el año 2014 hasta el mes de septiembre se han atendido cerca de 12774 denuncias, de las cuales la mayoría de estas se han presentado en lo corrido del 2017.

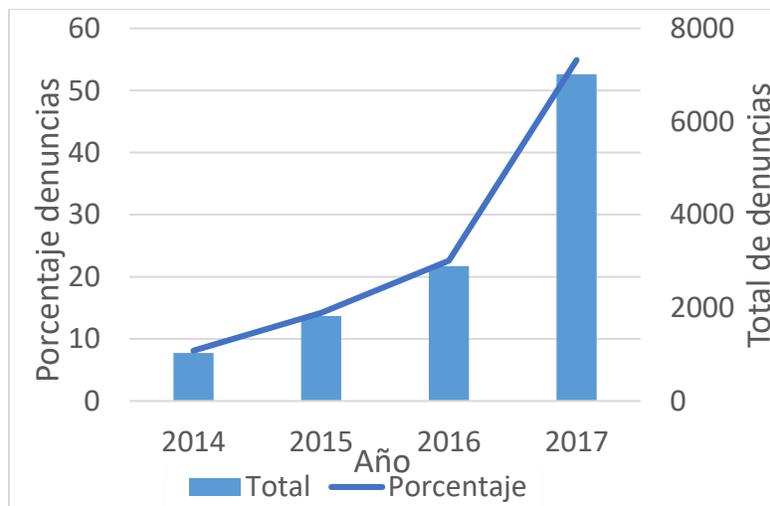
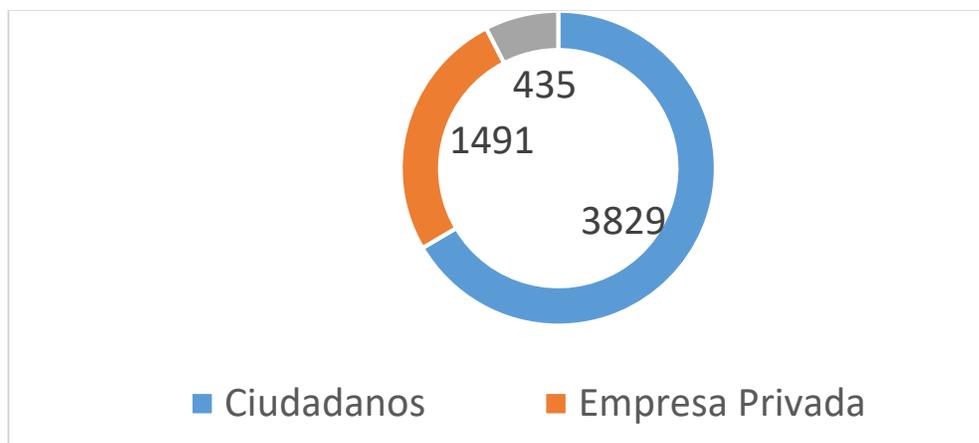
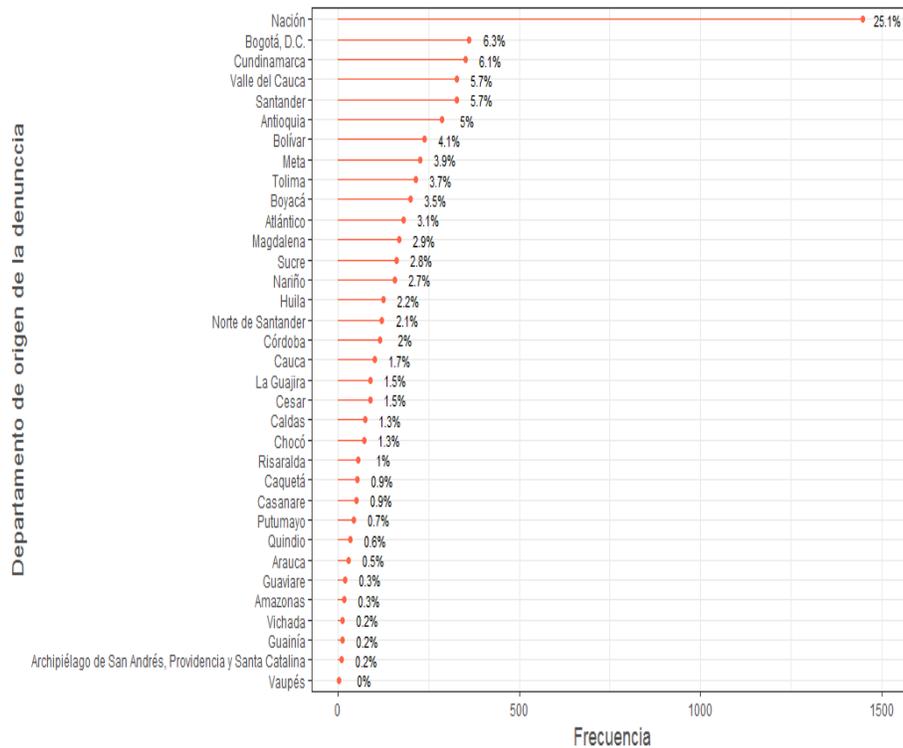
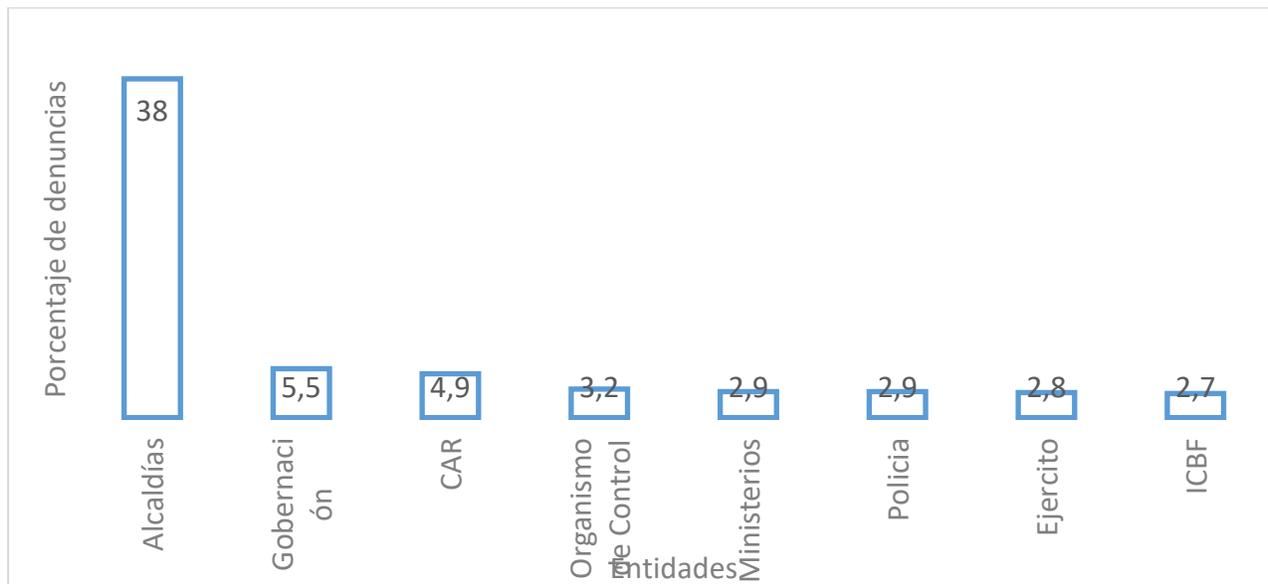


Figura : Total de denuncias

La distribución de las denuncias en el territorio nacional muestra que la mayoría de estas son realizadas a entidades de orden nacional (25,1%), Bogotá, Cundinamarca y el Valle del Cauca tienen (18,1%) de estas.



El 66,5% de las denuncias provienen de la ciudadanía, el 25,9% de la empresa privada y sólo el 7,6% de las denuncias son realizadas por el sector público.



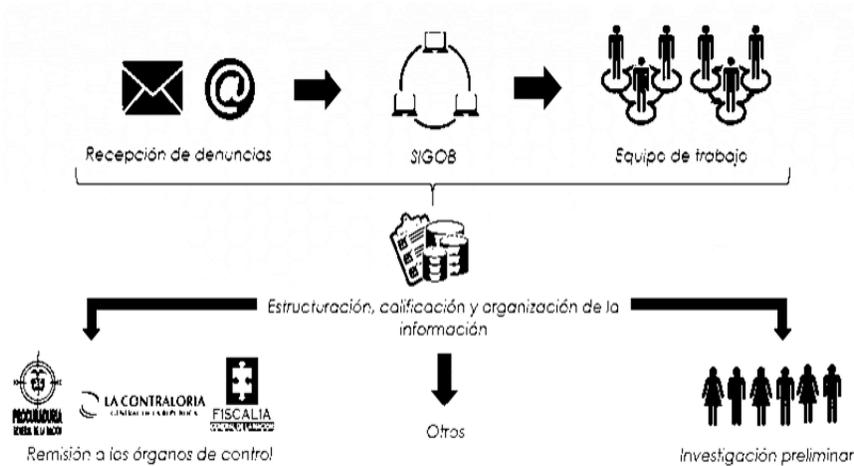
Las entidades más denunciadas son las alcaldías, las gobernaciones con cerca de la mitad de las denuncias (48,4%). Las causas más frecuentes de las denuncias son:

Remitido a Entres de Control	%
Incumplimiento ley transparencia	65,7
Mala administración y malversación de Fondos	56
Peculado por apropiación - enriquecimiento ilícito	55,8
Tráfico de influencias	53,5
Deficiencia en la planeación	53,2
Falsedad doc. público.	50

El proceso de análisis de las denuncias comienza con la recepción de las denuncias por medio físico o magnético (vía correo electrónico), estas denuncias son almacenadas en un sistema de información

(SIGOB), posteriormente un equipo de trabajo analiza las denuncias y se codifican las denuncias detallando diferentes aspectos de las denuncias.

Las denuncias son remitidas a los entes de control (contraloría, procuraduría y fiscalía) departamentales o nacionales o a otras entidades como personerías locales.



## 1.2. Fuente de datos

Se utilizan dos fuentes de información: hojas de cálculo donde se especifican las características de las denuncias y una segunda fuente de remisión de las denuncias en donde se detalla la entidad a donde se remiten la denuncia.

1. Características de las denuncias: esta información es copiada por los asesores del GRAP, en una hoja de cálculo. Se dispone información desde el año 2014 al 2017. Las variables que contiene esa hoja de cálculo son:
  - ✓ Entidad denunciante
  - ✓ Departamento de origen de la denuncia
  - ✓ Municipio de origen de la denuncia
  - ✓ Resumen de la denuncia
  - ✓ Presunto acto de corrupción o irregularidad
  - ✓ Tipo de forma de contratación
  - ✓ Entidad mencionada en la denuncia
  - ✓ Orden de la entidad denunciada
  - ✓ Sector de la entidad denunciada

Remisión de las denuncias: esta información recoge las remisiones de las denuncias a los entes de control y a otras entidades, la información es copiada en diferentes hojas de cálculo que tienen las denuncias enviadas en cada año. Las variables de esta base de datos son:

- ✓ Fecha de envío a destinatario (entidad de control)
- ✓ Destinatario: destinatario de la denuncia.
- ✓ Dependencia: entidad o actor al que se le remite la denuncia
- ✓ Ciudad de envío: ciudad de la dependencia a la cual se envía la denuncia
- ✓ Clase de envío: denuncias, administrativa.
- ✓ Número de folios: número de folios de la denuncia remitida.

### 1.3. Metodología

- ✓ El proceso para el análisis de las denuncias es el siguiente:
- ✓ Se realiza la lectura de las dos bases de datos (denuncias y remisiones), y se realizan procesos de recodificación y de edición de datos:
- ✓ Se renombran de forma adecuada los nombres de las variables de las bases de datos.
- ✓ Se unifican las diferentes grafías de las diferentes categorías de cada una de las variables.
- ✓ Se arreglan los formatos de las fechas.
- ✓ En la variable forma de contratación, los valores faltantes se recodifican a “contractual”.
- ✓ En lo sobres, arreglar las clases para que existan unas únicas categorías:

Se lleva a cabo la integración de dos bases de datos la de remisiones por medio de una matriz que contiene el enlace entre cada denuncia con cada remisión. Esta relación es de unos a muchos, puesto que cada denuncia puede ser remitida a varias entidades (por ejemplo, una procuraduría regional, a la fiscalía y u a una tesorería).

ID Denuncia	ID Remisión
Denuncia 1	1
Denuncia 1	2
Denuncia 1	3
Denuncia 1	4
Denuncia 1	5
Denuncia 2	1
Denuncia 2	2
Denuncia 2	3
.	.
.	.
.	.
Denuncia n	1
Denuncia n	2

Estructura de las denuncias

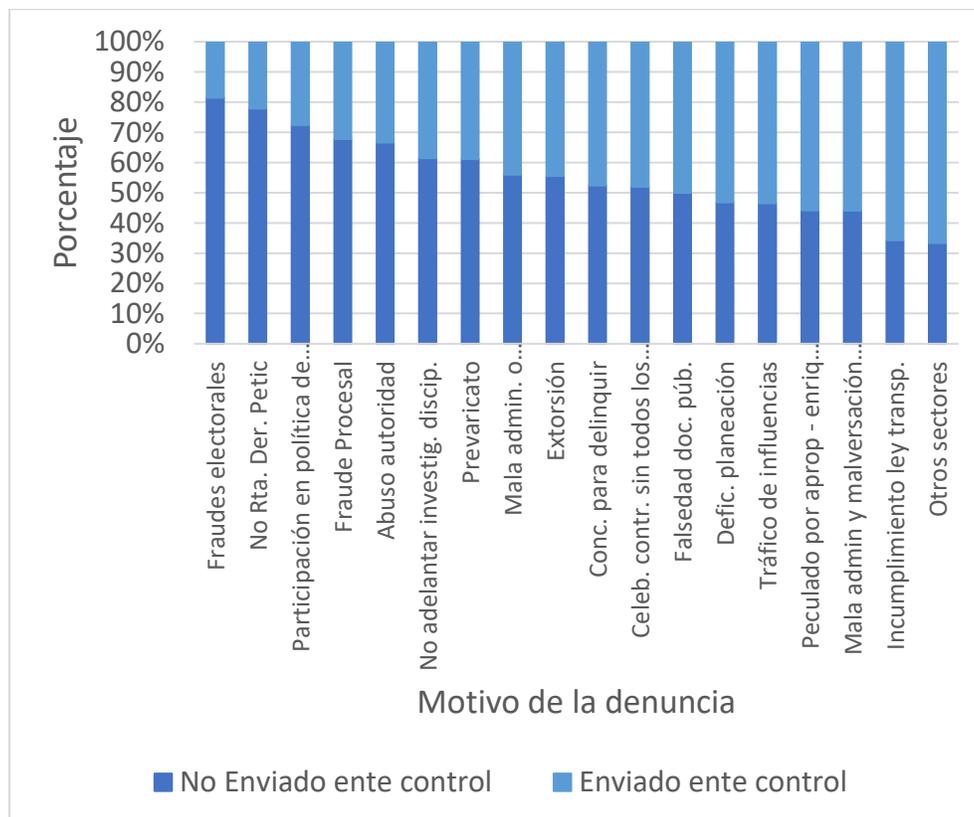
Se integra la base de datos de denuncias con la base de datos de remisiones, se realiza en primer lugar un *inner join* entre la base de denuncias y la matriz mencionada anteriormente.

La base de datos de datos de remisiones se unifica con la matriz anterior y se realiza un proceso de reestructuración de tal forma que cada fila represente una denuncia. A cada entidad a la que se remite una denuncia se le asigna una variable, se calcula el máximo de folios por cada remisión y se calcula la moda del tipo de denuncia.

Las dos bases de datos con la información que permita unir las se unen resultando un archivo de denuncias y variables indicadoras que permiten identificar a que entidad fueron remitidas las denuncias.

#### 1.4. Resultados

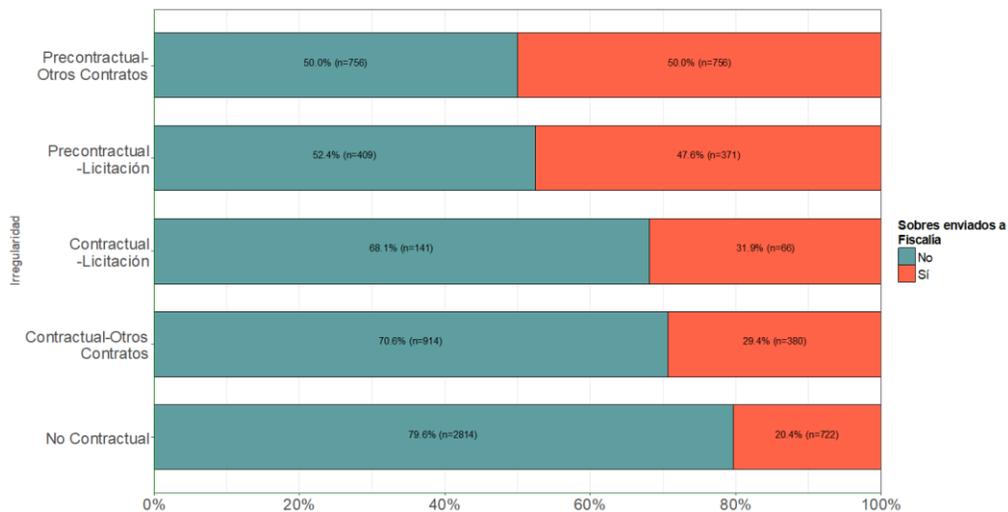
Una vez conformada las bases de datos, se realiza un análisis exploratorio para ver las variables que más relacionadas están con las remisiones a entidades de control.



La siguiente tipología de denuncias son las más propensas para enviar a los entes de control:

Remitido a Entes de Control	%
Incumplimiento ley transparencia.	65,7
Mala administración y malversación de fondos	56
Peculado por apropiación - enriquecimiento ilícito	55,8
Tráfico de influencias	53,5
Deficiencia planeación	53,2
Falsedad doc. Público	50

Por otro lado, los procesos precontractuales son los más procesos para enviar a los entes de control:



Las variables más asociadas con las remisiones se presentan a continuación:

- Tipo de contratación ( $X_1$ )
- Tiempo de remisión ( $X_2$ )
- Orden entidad ( $X_3$ )
- Sector entidad ( $X_4$ )

- Núm. Folios ( $X_5$ )
- Dpto. origen denuncia ( $X_6$ )
- Tipo de irregularidad ( $X_7$ )

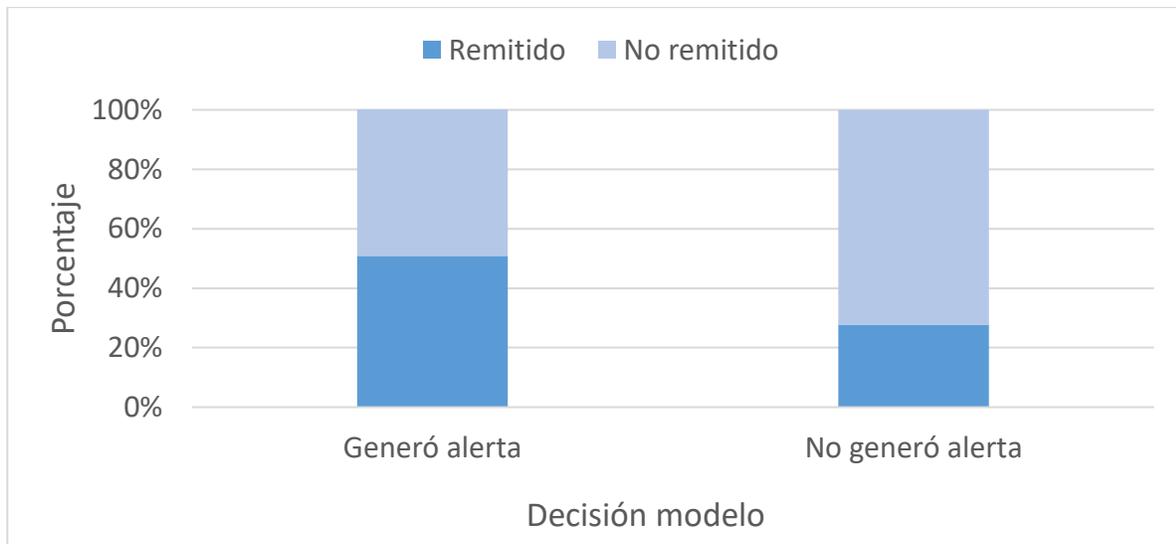
Una vez definidas las variables más relevantes se ajusta un modelo de regresión logística

$$\text{Pr(Probab. remisión a ente de control)} = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_6 X_6)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_6 X_6)}$$

Este modelo permite identificar las denuncias que deben ser remitidas a los entes de control:

El modelo se calibra de forma que se minimice la tasa de falsos negativos (denuncias que deben ser remitidas a entes de control, pero no se emitieron alertas sobre ellas).

Las denuncias sobre las cuales se emite alerta se deben analizar al detalle ya que estas denuncias tienen características que las hacen muy propensas a revisión por parte de los entes de control. (Los falsos positivos pueden deberse a rezagos en los tiempos).



## **1.5. Conclusiones**

- ✓ Se elaboró una herramienta predictiva para clasificar las denuncias según la entidad de control a la cual deba remitirse.
- ✓ No se disponía de herramientas que analizarán la información consignada en las denuncias y convirtiera esta información en un insumo de política pública.
- ✓ Se elaboró una herramienta que permite cuantificar cuáles son los factores que hacen de una denuncia más susceptible de análisis y revisión por parte de los entes de control

## **1.6. Glosario**

- DAPRE: Departamento Administrativo de Presidencia de la República
- Grupo Laboratorio de Innovación en Lucha contra la Corrupción (GLILC)
- Grupo de Política de Transparencia, Acceso a la Información y Lucha contra la Corrupción (GTALCC)
- Grupo de Revisión y Análisis de Peticiones, Denuncias y Reclamos de Corrupción (GRAP)