



El futuro
es de todos

DNP
Departamento
Nacional de Planeación

Aplicación de analítica de datos para la atención de PQRSD

Unidad de Científicos de Datos
Dirección de Desarrollo Digital

Junio, 2019



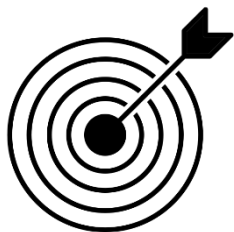
- 1. Trabajo realizado en DNP**
- 2. Oportunidades**
- 3. Necesidades de información y limitaciones**

1. Trabajo realizado en DNP



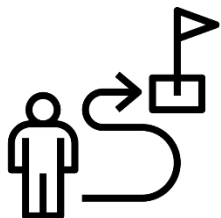
Explotación de datos de PQRSD para entidades públicas

El uso de técnicas de minería de texto y analítica de datos puede facilitar y agilizar los procesos de respuesta de PQRSD en las entidades públicas.



Objetivo

Identificar patrones en las PQRSD que llegan al DNP para definir temas/categorías comunes, con sus posibles responsables.



Metodología

1. Técnicas de procesamiento y análisis de texto.
2. Modelos supervisados y no supervisados para clasificar textos con características similares.

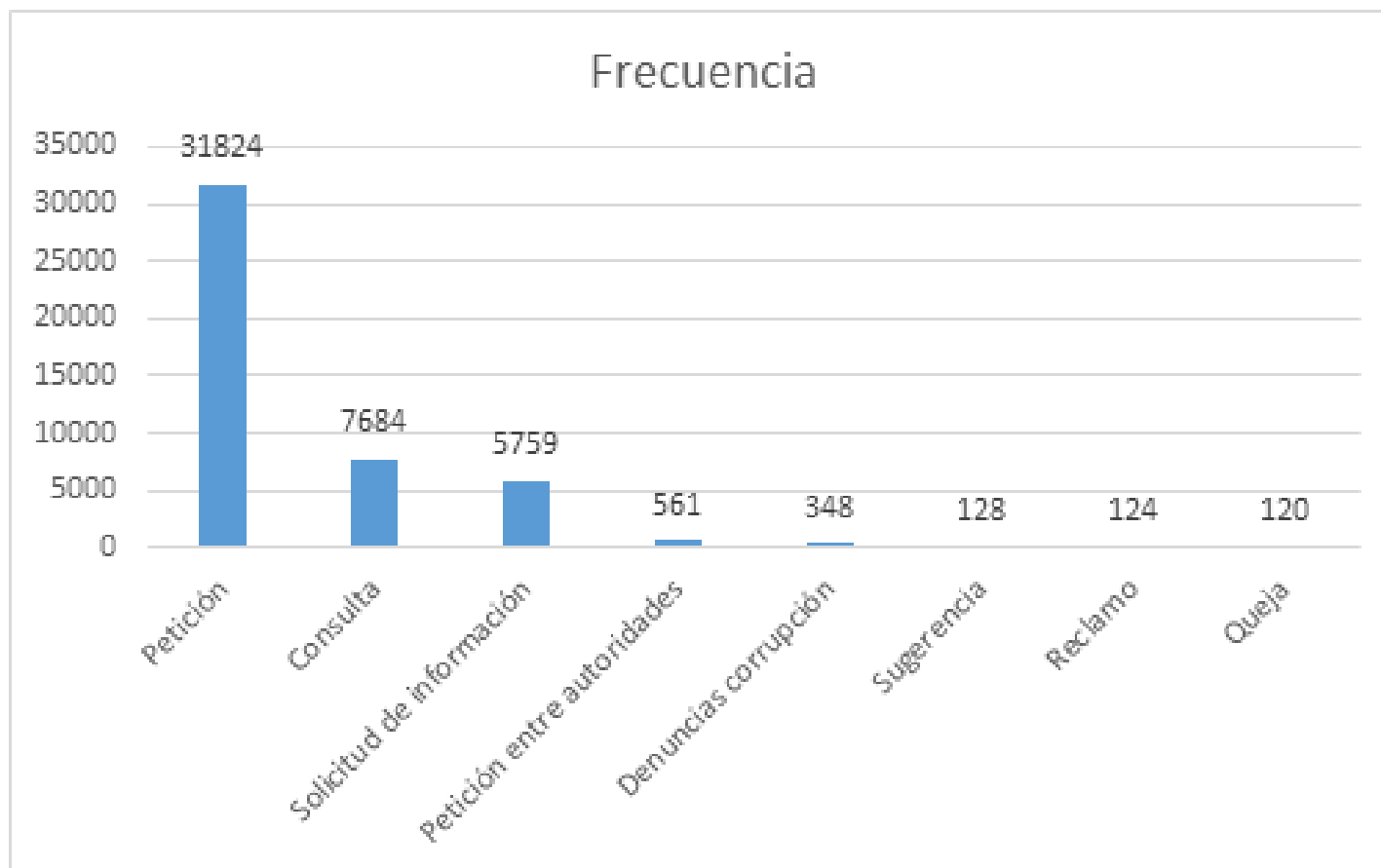


Insumos

48.234 PQRSD radicadas en el DNP entre el 2016 y el 2018.

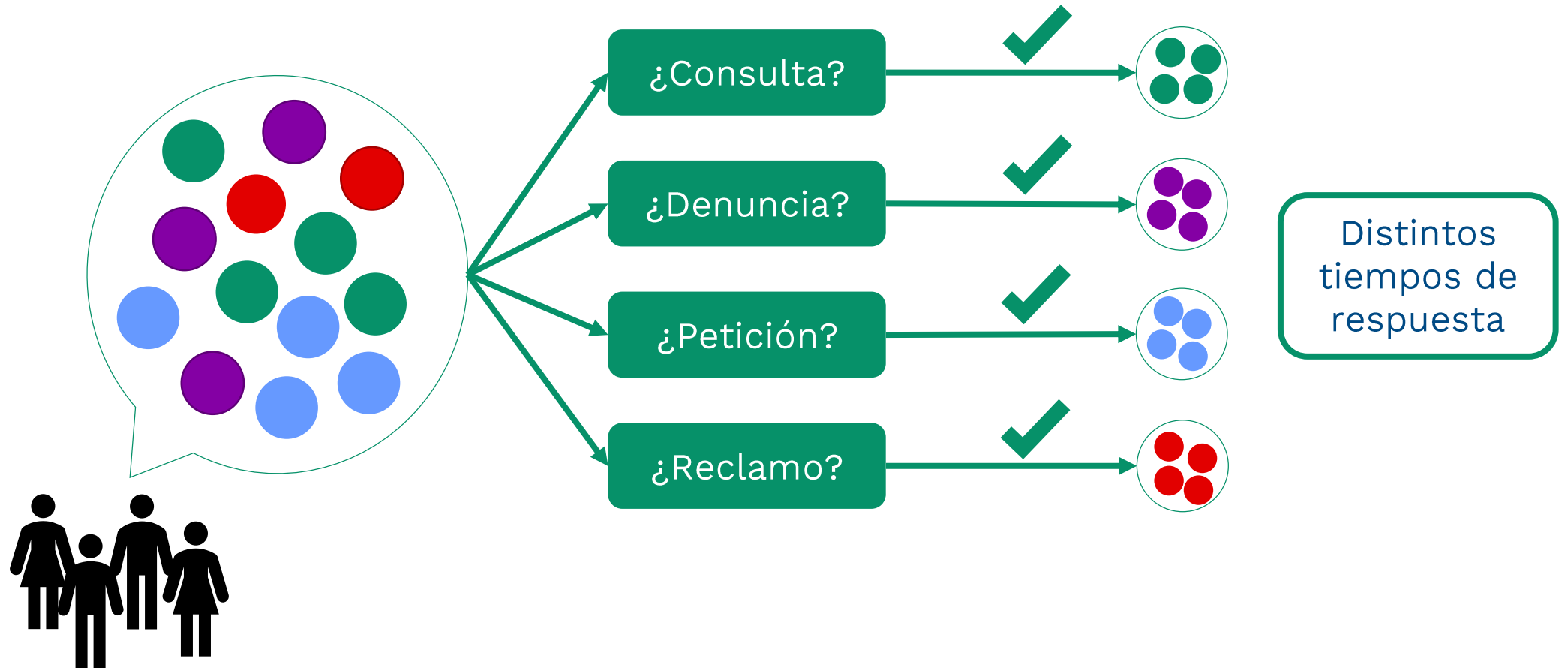
Distribución de las PQRSD por tipo documental

Las peticiones y consultas representan el 84.7% del total de PQRSD.



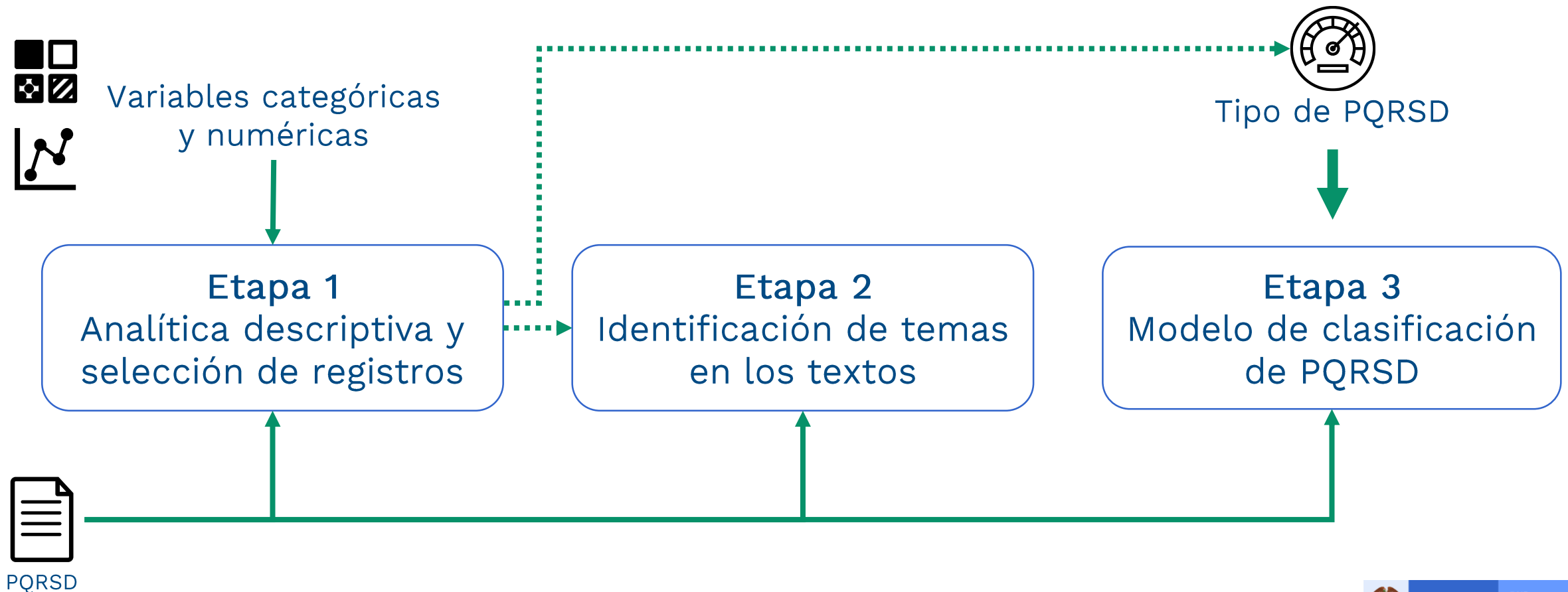
Modelo de clasificación

Se entrenó un modelo para clasificar PQRSD por tipo de documento.



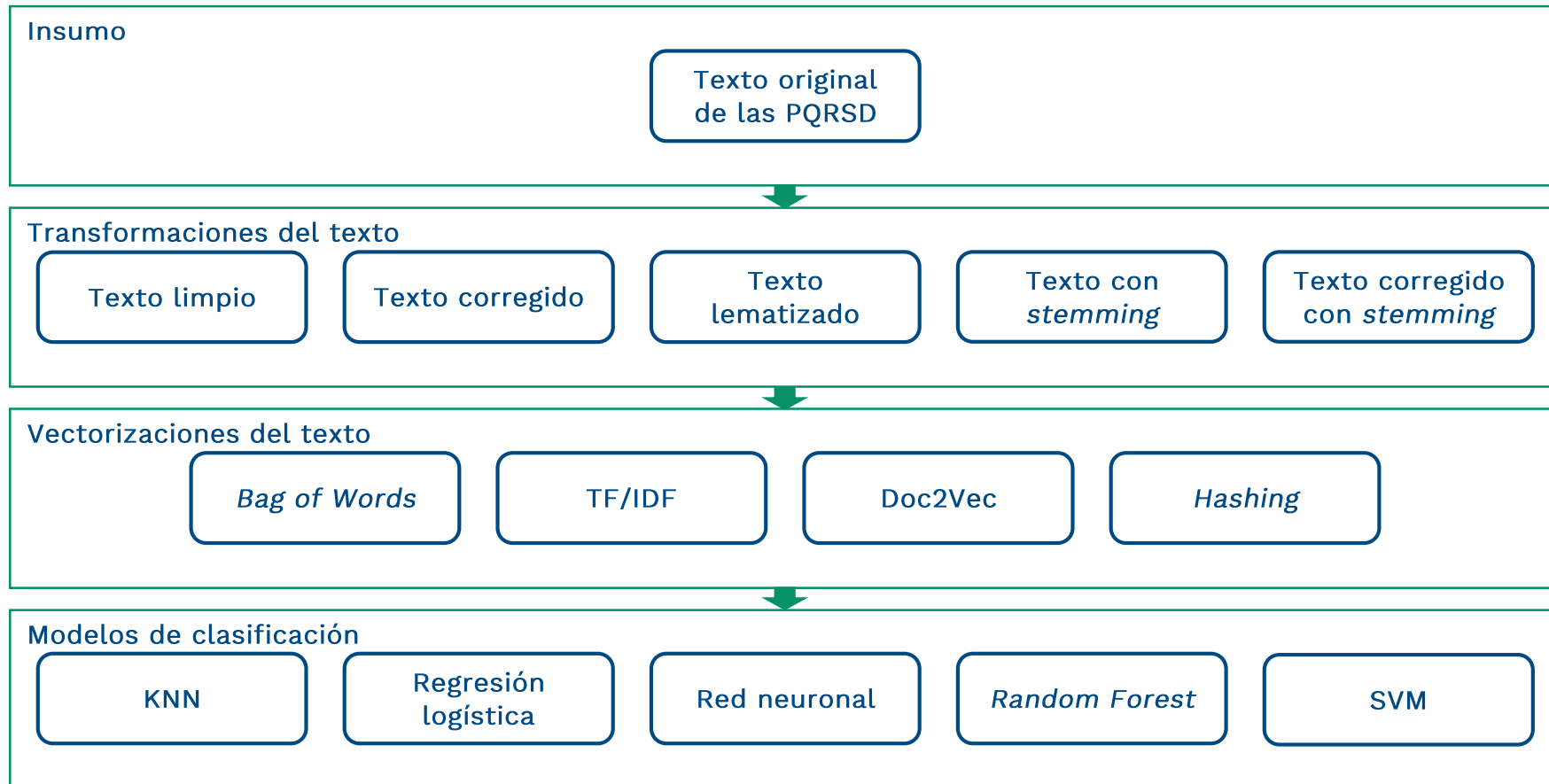
Flujo de datos para crear modelos de clasificación

Se definen 3 etapas distintas; una de exploración y selección de datos y otras dos de desarrollo de modelos de aprendizaje de máquina.



Enfoques de modelado

Se evaluaron más de 40 enfoques de modelado para predecir el tipo de documento.



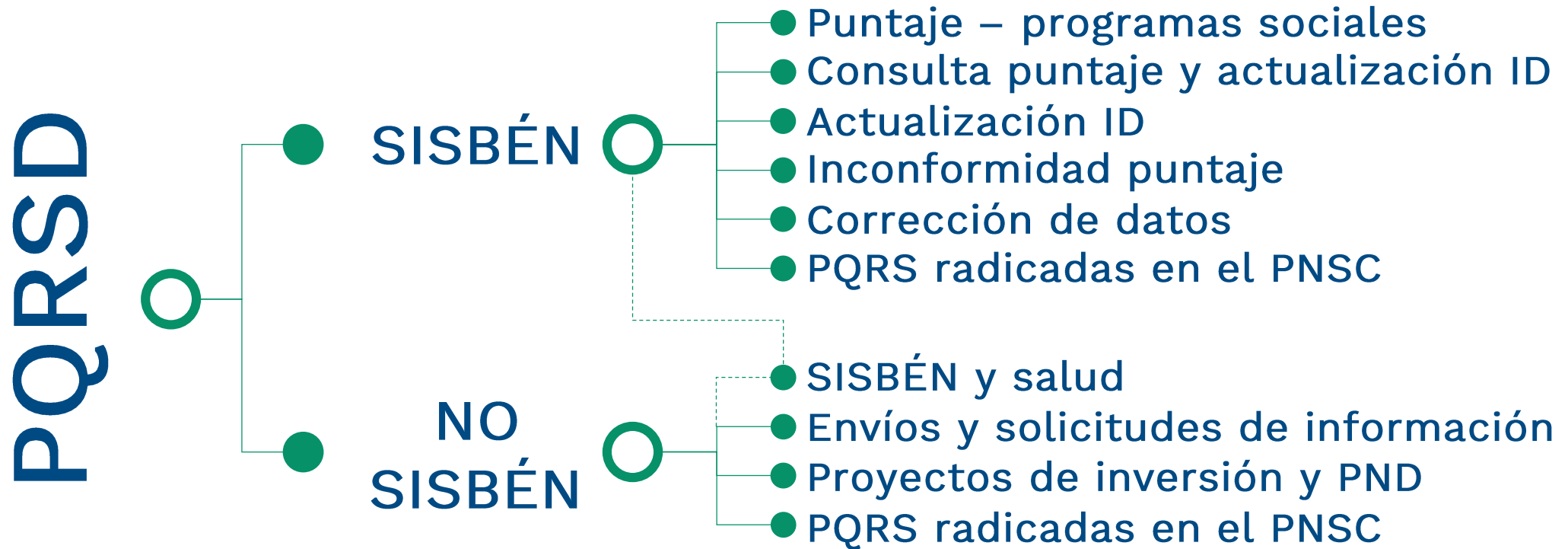
Resultados modelos de clasificación

Se obtuvo una baja precisión para todos los modelos entrenados, lo que se consideró insuficiente aun teniendo en cuenta la complejidad del problema multiclase (8 clases).

Tipo de modelo	Tipo de vectorización	Tipo de texto	Precisión (%)
KNN	Doc2Vec	Lematizado	43,56%
Multinomial	Doc2Vec	Lematizado	44,67%
SVM	Doc2Vec	Lematizado	26,89%
Red Neuronal	Doc2Vec	Lematizado	43,56%
Random Forest	Doc2Vec	Lematizado	45,89%

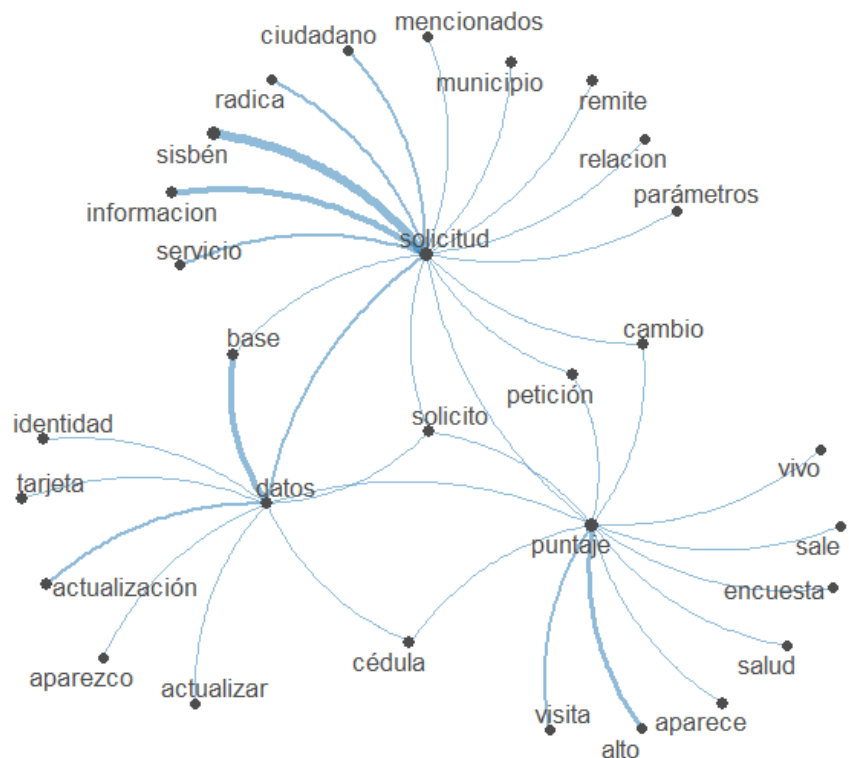
Agrupamiento de PQRSD

Se realizó un *clustering* para identificar algunos de los ejes temáticos más relevantes.



Tablero de visualización y análisis

Se desarrolló una aplicación para realizar análisis exploratorios, conteos de frecuencias y redes de términos. También contenía el modelo de clasificación desarrollado.



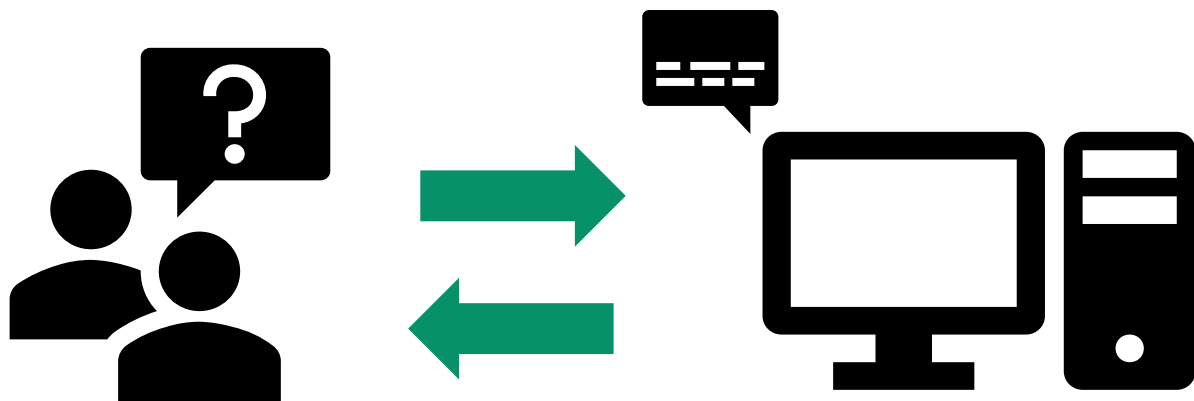
The screenshot shows the 'Análisis de PQRSD' application interface. The left sidebar contains a menu with the following items: 'PQRS D', 'Gráficos de frecuencias', 'Nubes de palabras', 'Red de términos principales', 'Palabras clave por clúster', and 'Clasificación de PQRS'. The main content area has a header 'Ingrese una PQRS' and a text input field containing 'Envía denuncia de robo de recursos'. Below the input field is a 'Clasificar' button. At the bottom of the main area, the text 'DENUNCIAS CORRUPCION' is displayed.

2. Oportunidades



Opciones de trabajo futuro

Se recomendaron dos alternativas para explotar información de PQRSD.



Sistema de respuestas automáticas a temas definidos



Chatbot para interacción con el usuario

Algunos temas que podrían abordarse para respuestas automáticas

Consulta puntaje

Inconformidad puntaje

Solicitud encuesta

Corrección de datos

Actualización de documento

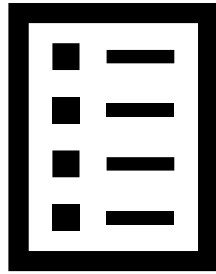
¿Estoy en SISBÉN?

3. Necesidades de información y limitaciones



Limitaciones en la formulación del modelo

Debido al método de clasificación de las PQRSD el modelo presenta limitaciones predictivas



Modificación del tipo de documento en el proceso



Desbalanceo de clases por tipo documental



Baja capacidad predictiva del modelo

Lecciones aprendidas

Del piloto realizado, se identificaron aspectos para tener en cuenta en futuros proyectos.

El problema o necesidad a tratar debe definirse con los actores directamente implicados

Debe evaluarse la calidad de los datos considerando su metodología de recolección

Los textos provenientes de distintas fuentes pueden tener estructuras muy diferentes

La estandarización de procesos y registros es muy deseable para la obtención de buenos insumos



Conclusiones



1. Automatizar la clasificación de PQRSD por tipo de documento no es viable en el momento.
2. Para desarrollar modelos de clasificación en un futuro, es necesario estandarizar el registro de las PQRSD en el sistema de información.
3. Los clústeres permiten identificar temas sobre los cuales llegan numerosas PQRSD, como inconformidad con el puntaje de SISBÉN.

El análisis establece un precedente sobre la explotación de datos de PQRSD para hacer más eficiente su recepción y respuesta.