

# Violencia de género

*Usando procesamiento natural de lenguaje para  
identificar la presencia y el tipo de violencia de género  
en denuncias frente a un juzgado*

**Yabra Muvdi**

# Contexto



## Objetivo

El Juzgado PCyF 10 de la Ciudad Autónoma de Buenos Aires busca entender de qué forma se puede utilizar el texto de las denuncias que recibe para identificar la ocurrencia de diversos tipos de violencia de género.

## Insumos

El Juzgado ha consolidado una base de datos con información sobre cada una de las denuncias y los archivos donde se encuentra el texto de las denuncias.

# Metodología propuesta

1.

## **Aprendiendo del texto**

Utilizando el texto de las denuncias es posible construir una serie de modelos que aprendan a relacionar las palabras presentes en las denuncias con la ocurrencia de violencia de género y su tipo.

2.

## **Estimando probabilidades**

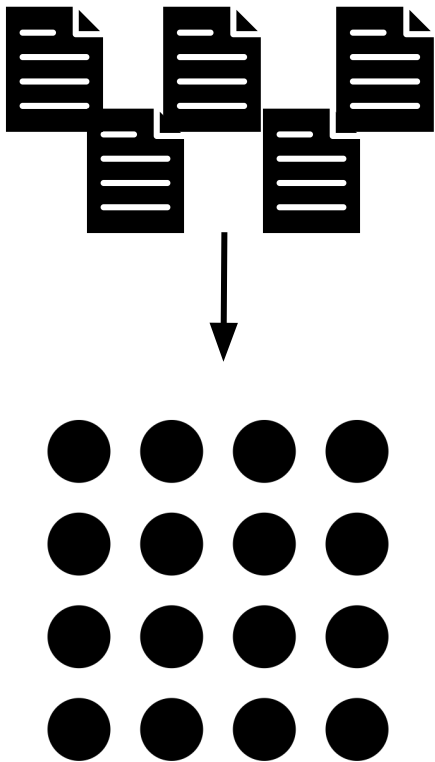
Los modelos construidos en la etapa 1 permiten estimar las probabilidades de que haya ocurrido cada tipo de violencia de género a partir del texto de la denuncia.

3.

## **Combinando probabilidades con información de las denuncias**

La base de datos del Juzgado contiene información potencialmente valiosa para entender la presencia de violencia de género y su tipo. En esta última etapa se construirá un modelo que combine las probabilidades estimadas en la etapa 2 con el resto de la información para predecir los tipos de violencia de género.

# Texto como datos



Antes de construir cualquier modelo, es necesario definir la forma en la que se va a transformar la información desestructura contenida en el texto en variables numéricas. Recientemente, algunas aproximaciones como los **“word embeddings”** o los **“transformers”** han mostrado excelentes resultados en diversas tareas.

Sin embargo, la metodología propuesta hará uso de una aproximación más tradicional: la **bolsa de palabras** (“bag of word”). Esto con el fin de facilitar la interpretabilidad y comunicación tanto del modelo como de los resultados.

# Texto como datos



## Bolsa de palabras

La bolsa de palabras es una representación de un texto en la que se cuenta el número de veces que cada palabra aparece en dicho texto. Este conteo permite construir una representación numérica del texto.

## Ponderando las palabras

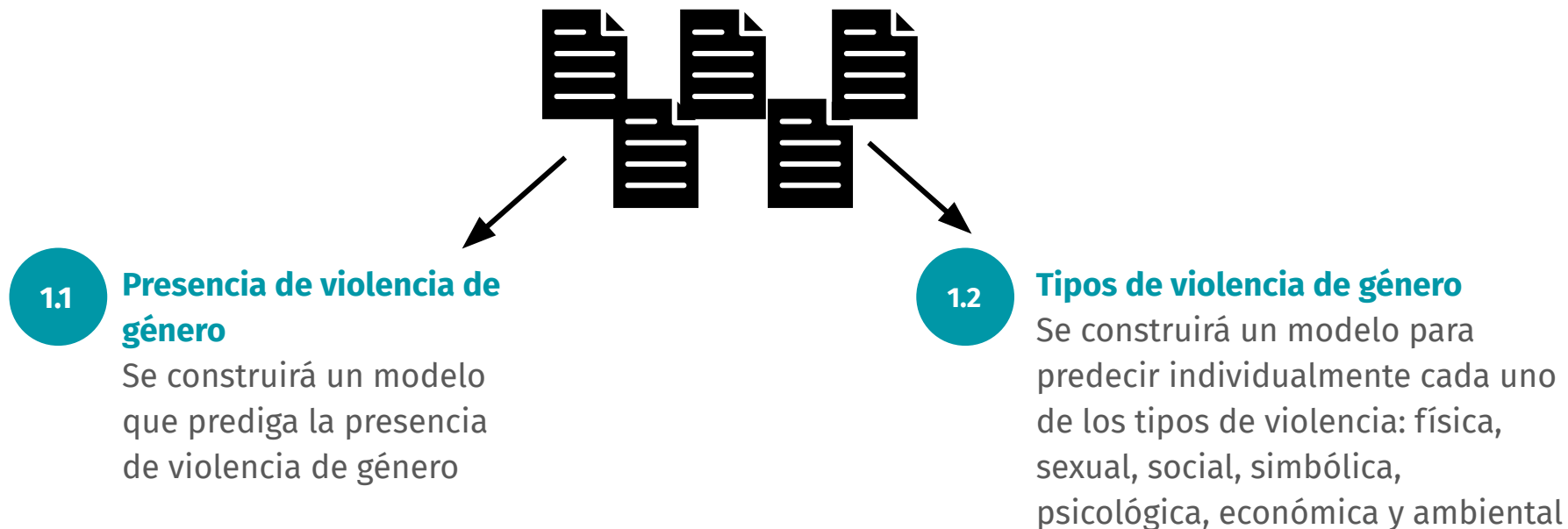
Además de contar la aparición de cada palabra, es posible asignar distintas ponderaciones a cada una de ellas. Esto puede hacerse, por ejemplo, para asignar mayor peso a aquellas palabras que tan solo aparecen en pocos documentos. La aproximación elegida será la de ponderar cada palabra con su **“inverse document frequency (idf)”** definido como:

$$\text{idf}(t) = \log \frac{1+n}{1+\text{df}(t)} + 1$$

# **1. Aprendiendo del texto**

# Modelos

Teniendo una representación numérica del texto, es posible construir modelos que permitan predecir la presencia de violencia de género y su tipo.



# Modelos

Para cada una de las dos tareas descritas (1.1 y 1.2) se entrenarán y evaluarán **4 modelos** distintos:

- Regresión logística
- Clasificador bayesiano ingenuo
- K vecinos más próximos
- Bosques aleatorios

Se considerarán diversos parámetros para cada modelo y se escogerán los mejores mediante **validación cruzada**



# Presencia de violencia de género - Resultados

## Modelo seleccionado

Bosques aleatorios

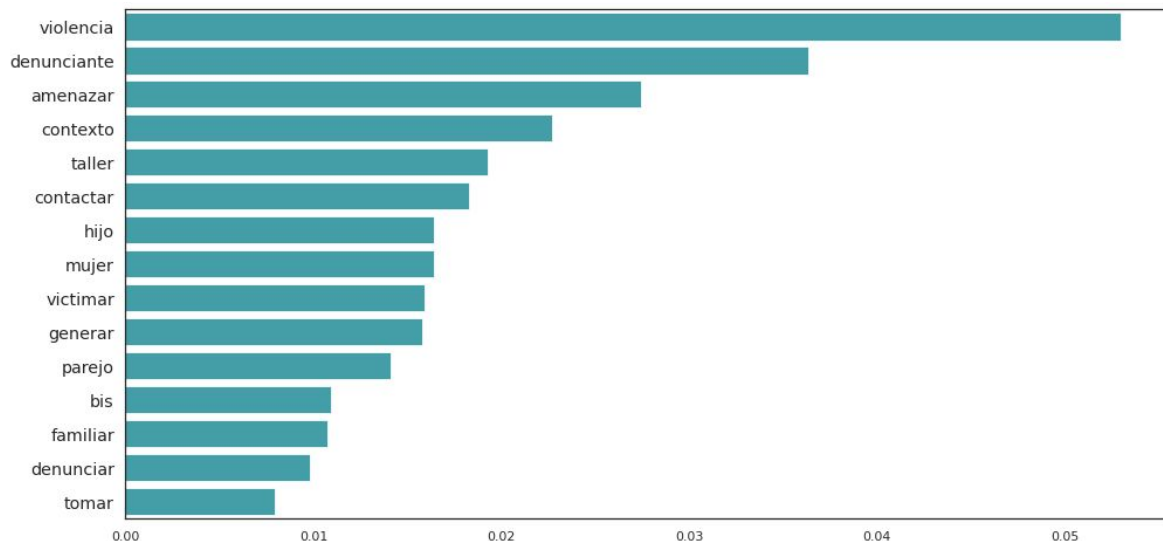
## Puntajes en datos de prueba validación cruzada

Accuracy: 93%

F1: 85%

Precisión: 95%

Recall: 76%



### Importancia de cada palabra en el modelo (“feature importance”)

En términos generales, esta medida recoge qué tanto cada una de las variables del modelo - en este caso palabras - contribuye a aumentar la pureza de los árboles de decisión. Entre más alto sea este número mayor será la importancia de la palabra.

# Tipos de violencia - Física

## Modelo seleccionado

Bosques aleatorios

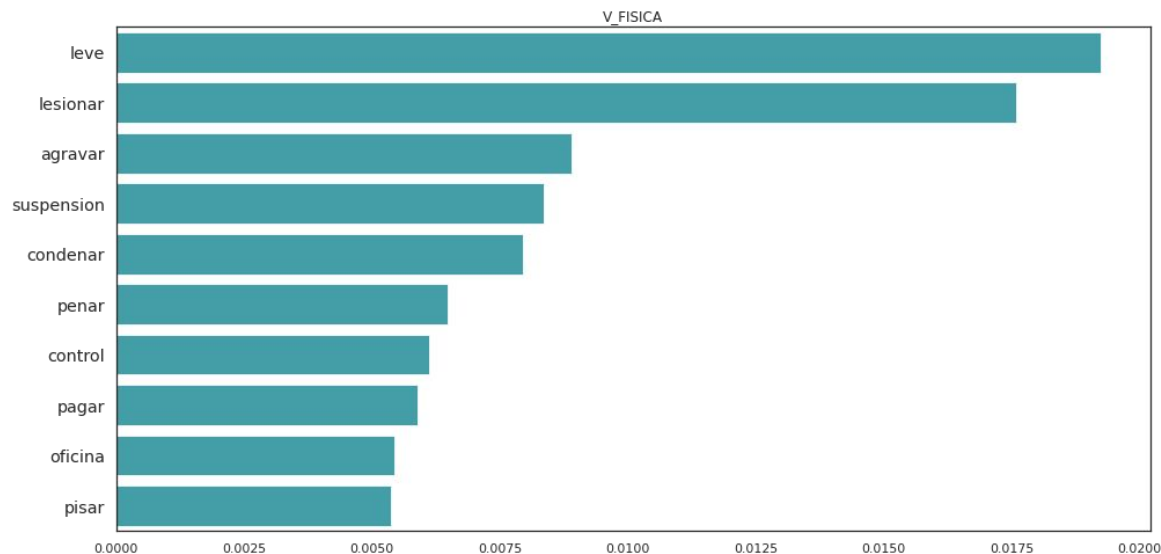
## Puntajes en datos de prueba validación cruzada

Accuracy: 83%

F1: 84%

Precisión: 87%

Recall: 80%



Importancia de cada palabra en el modelo ("feature importance")

# Tipos de violencia - Psicológica

## Modelo seleccionado

Bosques aleatorios

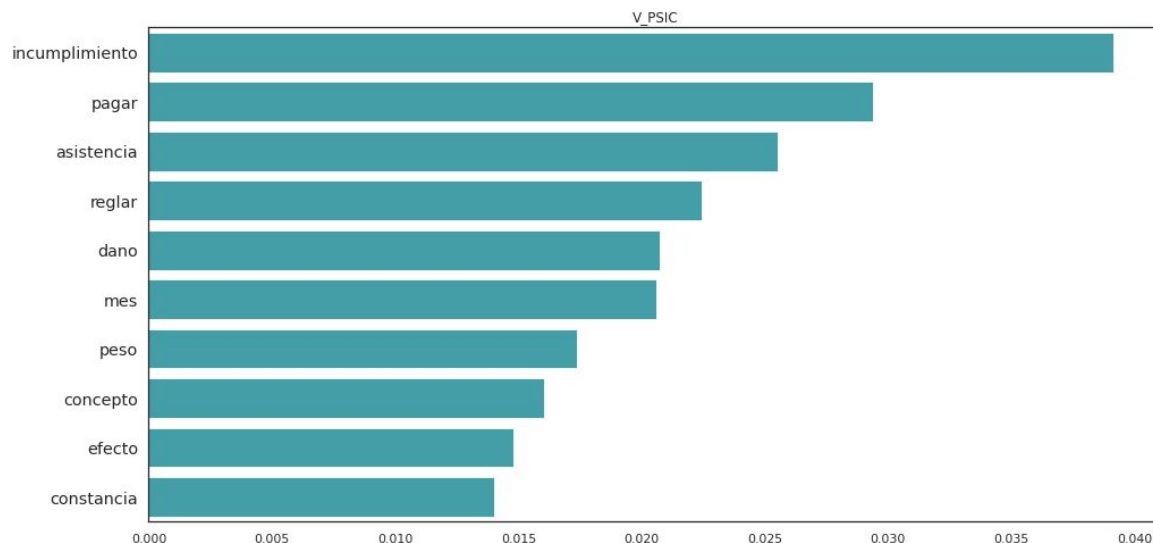
## Puntajes en datos de prueba validación cruzada

Accuracy: 87%

F1: 93%

Precisión: 86%

Recall: 100%



Importancia de cada palabra en el modelo ("feature importance")

# Tipos de violencia - Económica

## Modelo seleccionado

Bosques aleatorios

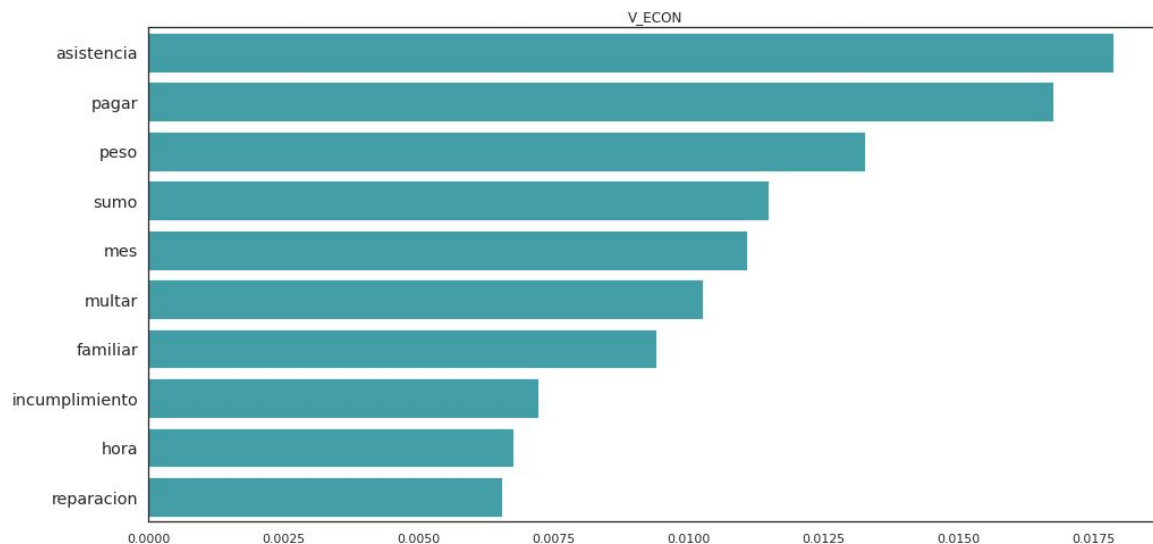
## Puntajes en datos de prueba validación cruzada

Accuracy: 87%

F1: 64%

Precisión: 100%

Recall: 47%



Importancia de cada palabra en el modelo ("feature importance")

# Tipos de violencia - Sexual

## Modelo seleccionado

Bosques aleatorios

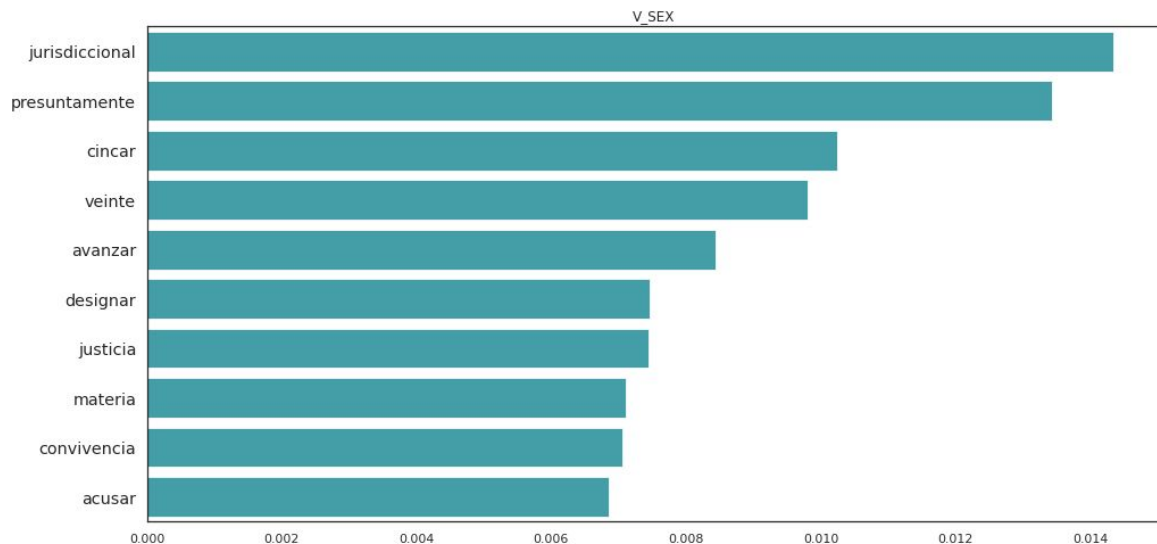
## Puntajes en datos de prueba validación cruzada

Accuracy: 97%

F1: 75%

Precisión: 100%

Recall: 60%



Importancia de cada palabra en el modelo ("feature importance")

# Tipos de violencia - Social

## Modelo seleccionado

Bosques aleatorios

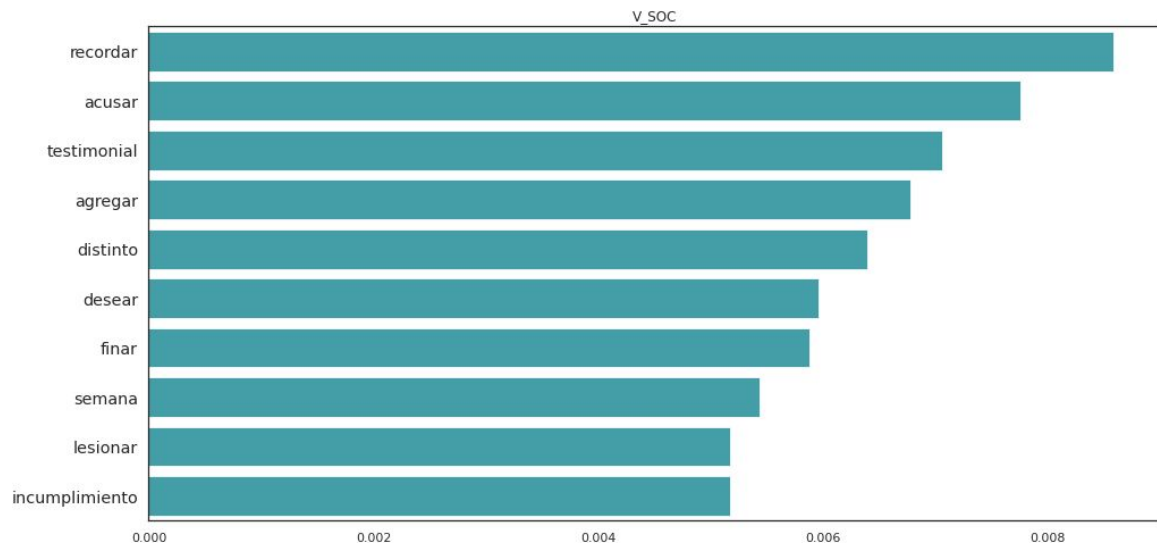
## Puntajes en datos de prueba validación cruzada

Accuracy: 87%

F1: 44%

Precisión: 100%

Recall: 29%



Importancia de cada palabra en el modelo ("feature importance")

# Tipos de violencia - Ambiental

## Modelo seleccionado

Bosques aleatorios

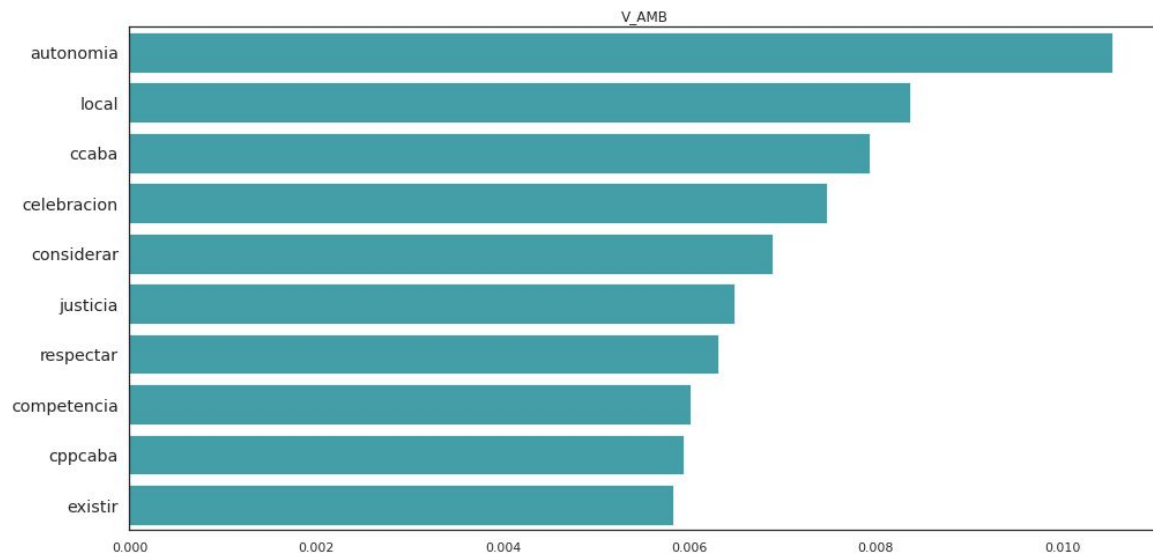
## Puntajes en datos de prueba validación cruzada

Accuracy: 92%

F1: 40%

Precisión: 66%

Recall: 29%



Importancia de cada palabra en el modelo ("feature importance")

# Tipos de violencia - Simbólica

## Modelo seleccionado

Bosques aleatorios

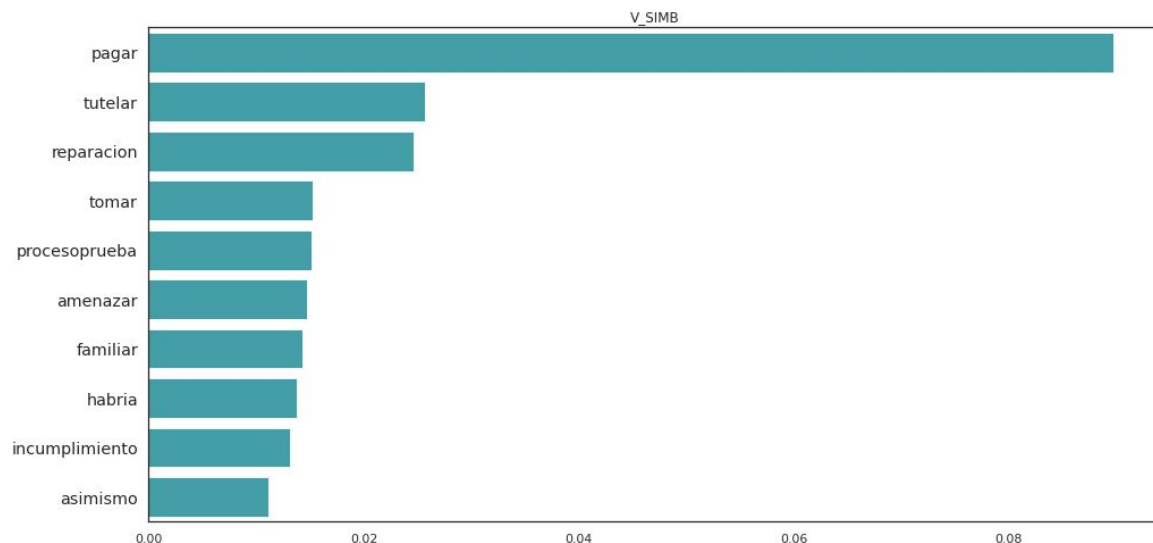
## Puntajes en datos de prueba validación cruzada

Accuracy: 91%

F1: 96%

Precisión: 90%

Recall: 100%



Importancia de cada palabra en el modelo ("feature importance")



## **2. Estimando probabilidades**

# Probabilidades

## Más allá de la clasificación binaria

Una vez entrenados los modelos para predecir la presencia de violencia de género y el tipo de violencia, es posible utilizarlos para - **en vez de generar una predicción binaria - generar probabilidades**. De esta forma, mantenemos más información dado que no estamos forzados a seleccionar exclusivamente entre dos categorías.

## Probabilidades como variables del modelo completo

El objetivo de estimar estas probabilidades es poder utilizarlas como insumo en el modelo final que incluya otras variables relacionadas con la denuncia.

**3.**

**Combinando probabilidades con  
información de las denuncias**

# Variables denuncias

La base de datos del Juzgado contiene distintas variables que pueden llegar a ser importantes a la hora de determinar la existencia de violencia de género y sus tipos. Algunas de estas son:

- Género del acusado/a
- Nivel de instrucción del acusado/a
- Frecuencia episodios
- Hijos en común

Estas variables se usarán en conjunto con las probabilidades estimadas anteriormente para construir el modelo final

# Modelos

Al igual que para las tareas anteriores, se entrenarán y evaluarán **4 modelos** distintos:

- Regresión logística
- Clasificador bayesiano ingenuo
- K vecinos más próximos
- Bosques aleatorios

Se considerarán diversos parámetros para cada modelo y se escogerán los mejores mediante **validación cruzada**

# Resultados por tipo de violencia (1/2)

## Modelo seleccionado

Bosques aleatorios

## Puntajes en datos de prueba validación cruzada

### Física

Accuracy: 97%

F1: 98%

Precisión: 100%

Recall: 95%

### Psicológica

Accuracy: 99%

F1: 99%

Precisión: 99%

Recall: 100%

### Económica

Accuracy: 99%

F1: 96%

Precisión: 100%

Recall: 92%

### Sexual

Accuracy: 100%

F1: 100%

Precisión: 100%

Recall: 100%

# Resultados por tipo de violencia (2/2)

## Modelo seleccionado

Bosques aleatorios

## Puntajes en datos de prueba validación cruzada

### Social

Accuracy: 99%  
F1: 92%  
Precisión: 100%  
Recall: 86%

### Ambiental

Accuracy: 99%  
F1: 93%  
Precisión: 100%  
Recall: 88%

### Simbólica

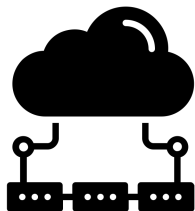
Accuracy: 99%  
F1: 99%  
Precisión: 98%  
Recall: 100%

# Productos

Si bien el análisis y los resultados presentados hasta el momento son parte importante del trabajo hecho, a continuación se especificarán concretamente los productos que se desarrollaron.



# Productos (1/2)



## Extracción automatizada de datos

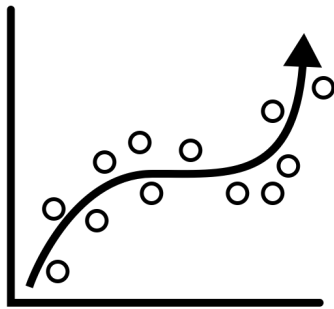
Actualmente, el Juzgado almacena los datos relacionados con cada una de las denuncias en Google Drive. El código entregado interactúa con Google Drive mediante su API de forma que el Juzgado pueda seguir utilizando esta plataforma y la extracción de los datos sigue funcionando.



## Extracción y procesamiento de texto

El texto de las denuncias está actualmente en archivos de diversos formatos (pdf, Word, odt, Docs) almacenados en Google Drive. El código desarrollado permite extraer este texto de acuerdo a su formato procurando que se conserve la mayor cantidad de texto.

# Productos (2/2)



## Modelos predictivos

Se entregan todos los modelos que permiten estimar la probabilidad de cada tipo de violencia. Adicionalmente, re entrenar estos modelos con nuevos datos es muy sencillo y se especifica en el README del repositorio.

## Palabras relacionadas con violencia

Como se pudo ver anteriormente, la aproximación elegida permite entender cuáles son algunas de las palabras que están más asociadas con cada uno de los tipos de violencia

# Consideraciones finales

Esta es solo una primera aproximación a este difícil problema. Todavía quedan muchas ideas por explorar y retos por resolver. Algunos de estos:

- Pocos datos y muchas variables (features)
- Desbalance de clases (e.g. muchos casos negativos de violencias ambiental y muy pocos positivos)
- Otras aproximaciones al texto (word2vec, Bert)
- Explorar más parámetros para los modelos
- Otro tipo de modelos predictivos (redes neuronales)
- Consideraciones éticas del uso de estos sistemas

**Gracias**