

CICLO DE SEMINARIOS

CONCURSO DE POLÍTICAS PÚBLICAS

2018

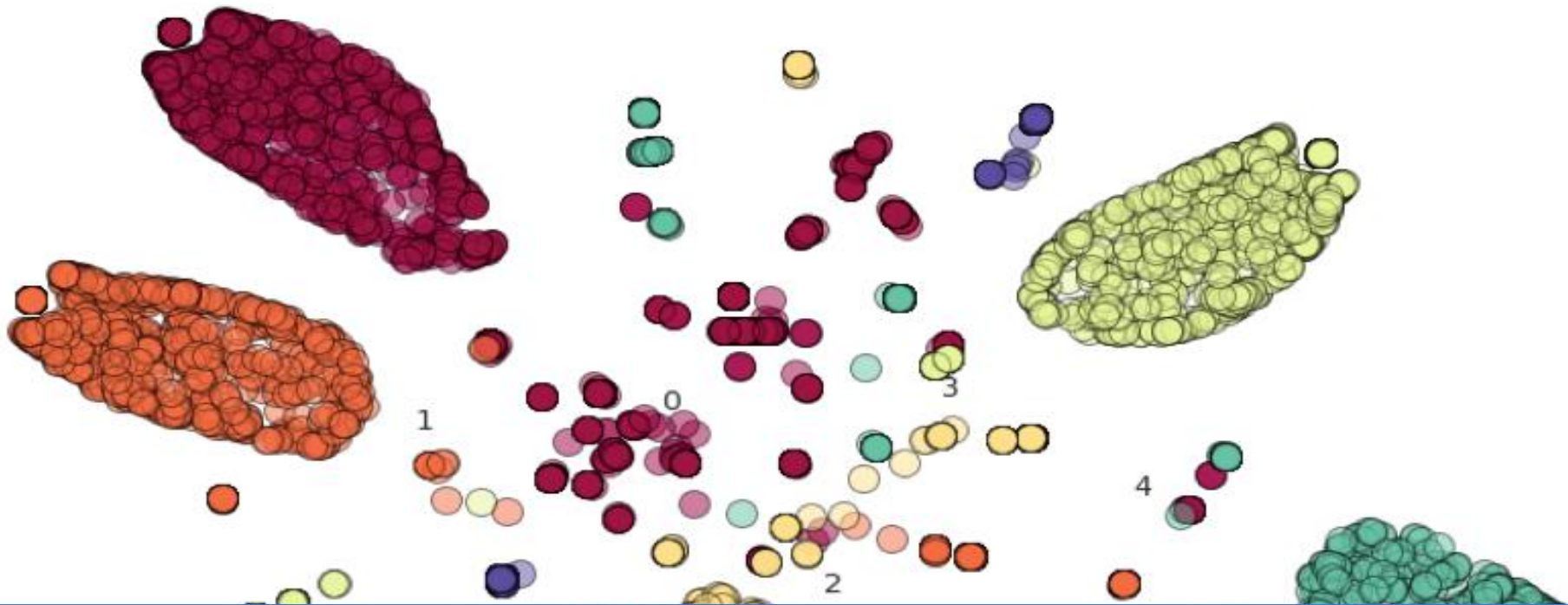
Centro UC
Políticas Públicas

MES DEL
Compromiso
Público UC

BIG DATA PARA EL ESTUDIO DE LA REINCIDENCIA CRIMINAL

#SeminarioReincidenciaCriminal
@cppublicasuc

www.politicaspUBLICAS.uc.cl



Seminario de Políticas Públicas, 19 de noviembre de 2018

Big Data para la Elaboración de Políticas Públicas en la Identificación de Comportamiento Criminal

Elwin van 't Wout | Eduardo Valenzuela | Kenzo Asahi
Christian Pieringer | David Torres | Pilar Larroulet



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE

PDI
POLICÍA DE INVESTIGACIONES
DE CHILE

Equipo de Trabajo

Elwin van 't Wout
Christian Pieringer

- Instituto de Ingeniería Matemática y Computacional
- Experiencia en modelamiento matemático y análisis de datos

Eduardo Valenzuela
Pilar Larroulet

- Facultad de Ciencias Sociales
- Experiencia en la sociología del crimen

Kenzo Asahi

- Escuela de Gobierno
- Experiencia en economía urbana y econometría

David Torres Iribarra

- Escuela de Psicología
- Experiencia en análisis de datos longitudinales

Harmin Cottenie
María José Rojas
Roberto Zulantay
Carlos Saez Larraide

- Centro Nacional de Análisis Criminal, PDI
- Experiencia en la investigación policial

Un Trabajo Junto a la PDI para el Estado

Policía de Investigaciones de Chile (PDI)

- Su misión consiste en investigar los delitos en conformidad a las instrucciones que al efecto dicte el Ministerio Público.
- Emplea a 12 mil de detectives.

Centro Nacional de Análisis Criminal (CENACRIM)

- Creado en 2015 como una unidad innovadora en la PDI.
- Tiene como objetivo principal la recolección, evaluación y análisis de información que sea de utilidad en el trabajo policial y la estrategias que de él se desprendan, promoviendo visiones globales y sistemáticas sobre los fenómenos delictuales.

Contexto del Estudio

- En la encuesta del Centro de Estudios Públicos de 2017, “la delincuencia” concentra un 27% de las respuestas situándose como primera prioridad a nivel del país.
- Acciones gubernamentales para enfrentar la delincuencia
 - Operaciones policiales
 - Programas de prevención
- Generación de políticas públicas sobre delincuencia basadas en
 - Experiencia de los actores involucrados
 - Estudios teóricos de criminología
 - Encuestas de victimización
 - Análisis de los datos administrativos

Marco Teórico de Delincuencia

- La literatura criminológica apunta a diversos caminos al involucramiento en el delito que no son asociados a causas únicas si no a un conjunto de factores de riesgo
- Hay hechos empíricos que apuntan a patrones de comportamiento delictual que principalmente dependen de la edad y género de la persona involucrada
- Si bien la literatura reconoce que el comportamiento pasado es el mejor predictor del comportamiento futuro, se apunta a la dificultad de predecir el comportamiento en períodos más largos de tiempo
- Si bien los patrones delictuales son posibles de identificar retrospectivamente, no se pueden predecir prospectivamente

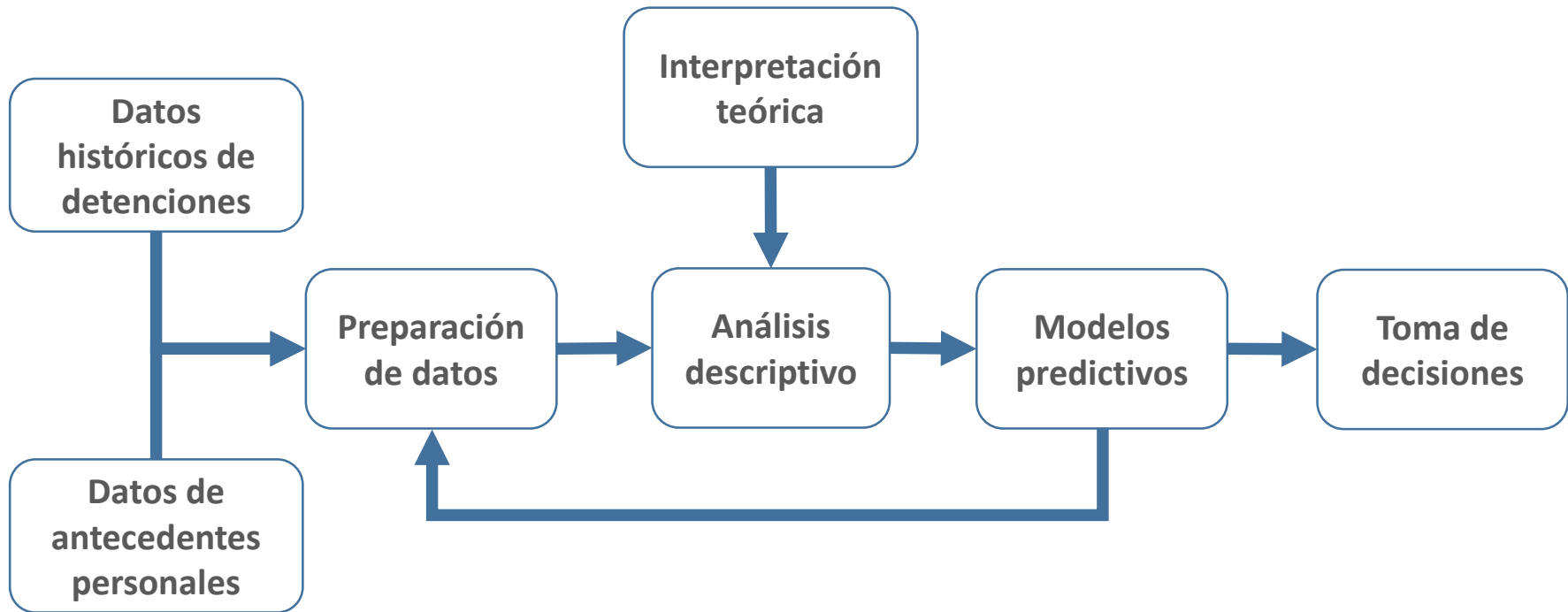
Objetivo General del Estudio

Mejoramiento en la priorización de recursos gubernamentales en las acciones para la reducción de la delincuencia a través de una tipología de comportamiento delictual basada en el análisis de los datos administrativos.

Hipótesis del Estudio

Los modelos matemáticos-estadísticos permiten describir y predecir una tipología de comportamiento de los actos delictuales esperados, basada en la historia delictual y características personales, con una precisión que sea de ayuda en la toma de decisiones.

Metodología de Big Data



Bases de Datos del Proyecto

Detenciones

- Datos históricos de detenciones realizadas en la Región Metropolitana desde el año 2009
- Un total de 750.000 registros y 300.000 personas

Armas

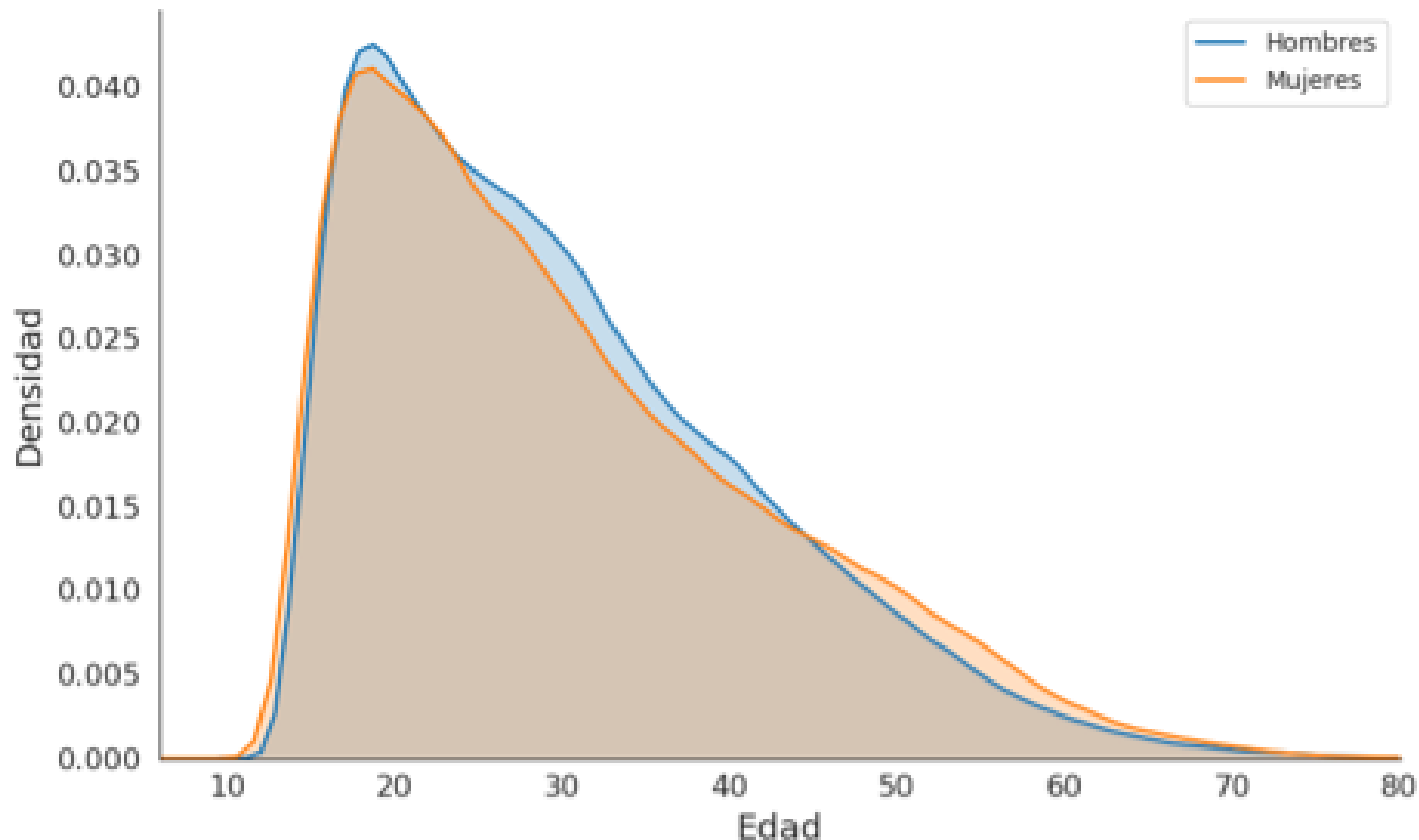
- Registros de armas inscritas en Chile

Datos biográficos

- Registros familiares

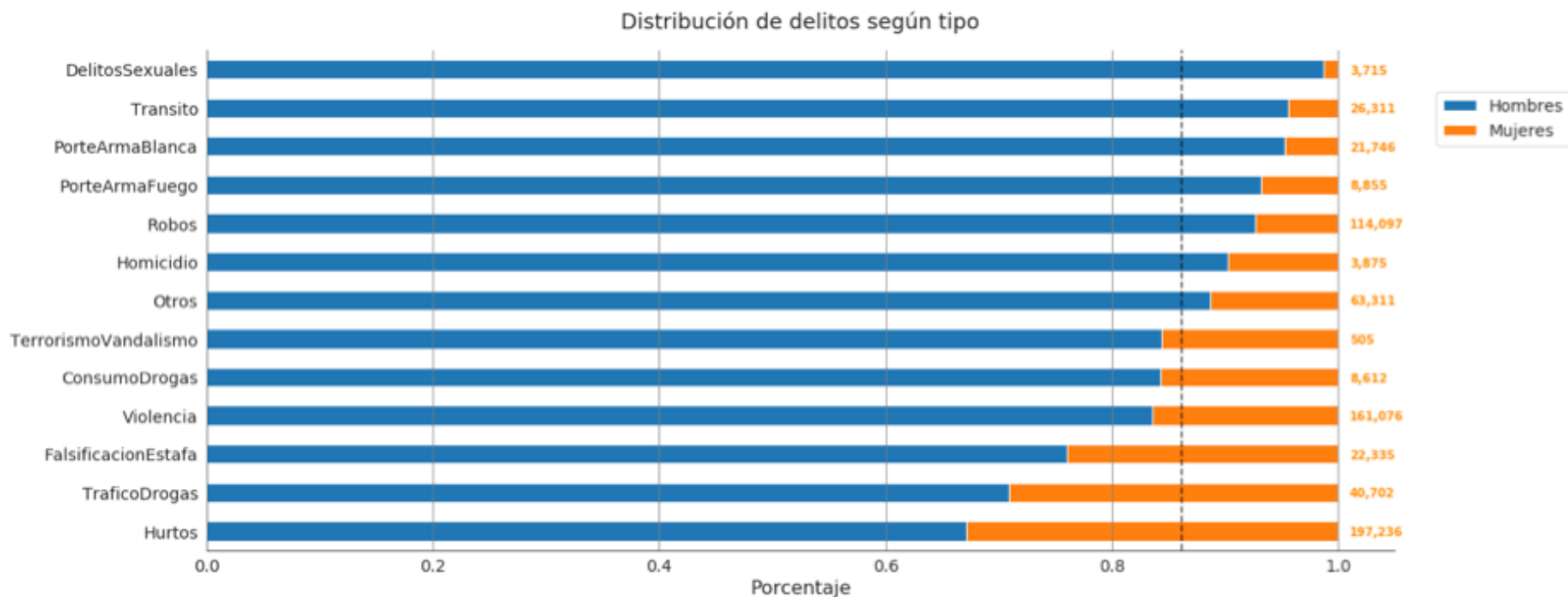
La edad tiene un impacto en la perpetración de delitos.

La delincuencia comienza en la juventud y baja rápidamente en la adultez, principalmente cuando una persona tiene un vínculo estable, por ejemplo matrimonio y trabajo supervisado.



El género tiene un impacto en la frecuencia y tipo de delito.

Los hombres están relativamente más involucrados en delitos más violentos como robo y homicidio y las mujeres en delitos que involucran tráfico de drogas.



Análisis Descriptivo

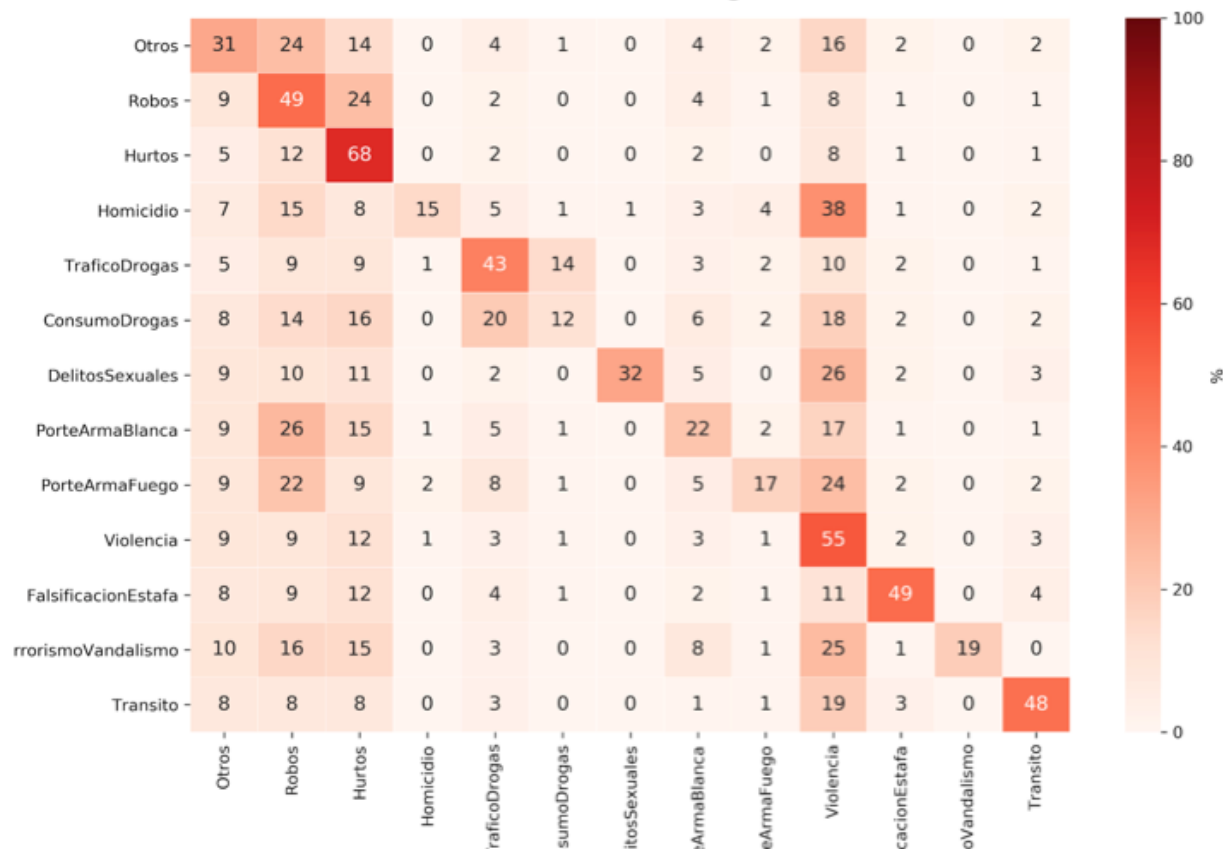
Objetivo específico

Describir el comportamiento delictual de las personas detenidas con el fin de mejorar la gestión y organización de los actores policiales.

Existe una asociación intertemporal en el tipo de delito cometido.

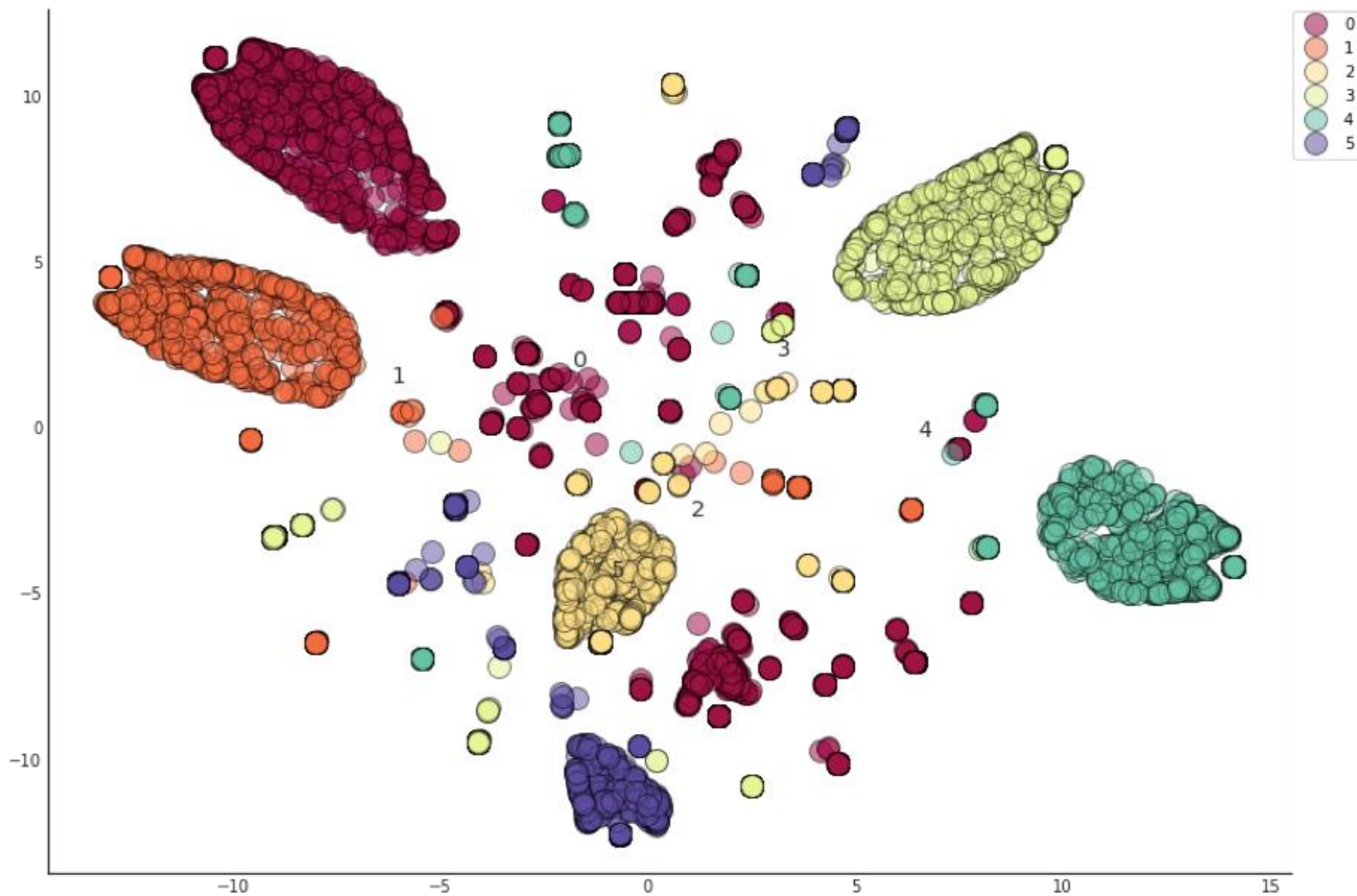
Las personas se especializan en un tipo de delitos, aumentando la probabilidad de reincidir en el mismo tipo de delito.

Matriz de Transición entre Categorías de Delitos



Existe una tipología de patrones delictuales.

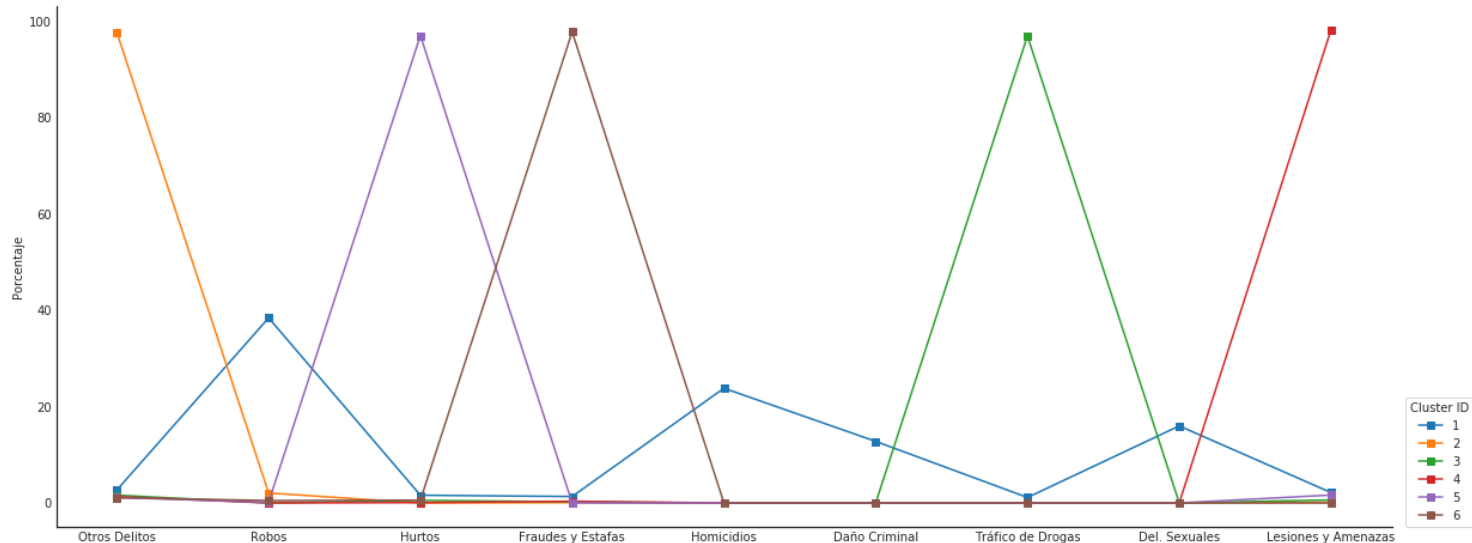
El método de conglomeración puede identificar distintos grupos de personas detenidas con características similares.



Existe una tipología de patrones delictuales.

La historia de detenciones y los antecedentes personales permiten identificar algunos grupos de personas. Cada conglomerado agrupa personas con características similares.

conglomerado	1	2	3	4	5	6
% de Individuos	32,14 %	14,81 %	12,77 %	14,71 %	13,83 %	11,75 %
% Armas Inscritas	8,76 %	0,0 %	0,0 %	0,0 %	0,0 %	0,0 %
% de Mujeres	8,92 %	15,97 %	29,47 %	15,43 %	30,55 %	29,87 %
% de estado civil soltero	19,42 %	25,27 %	19,18 %	23,43 %	14,21 %	28,44 %



Análisis Descriptivo

Conclusiones

1. La edad y género tiene un impacto en la frecuencia de detenciones y el tipo de delito.
2. Hay una habitualidad en especializarse en un tipo de delito.

Extensión

1. Uso de las herramientas como un apoyo adicional en la gestión interna de la PDI.
2. Usar el conocimiento en la habitualidad delictual como un apoyo adicional en la especialización de los actores gubernamentales.

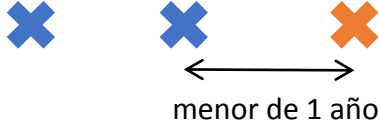
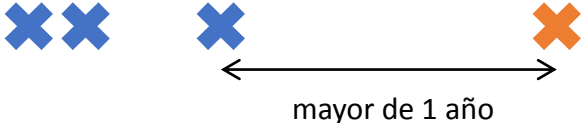

Análisis Predictivo

Objetivo específico

Predecir el comportamiento delictual futuro de las personas detenidas con el fin de mejorar la priorización de investigaciones policiales.

Análisis Predictivo

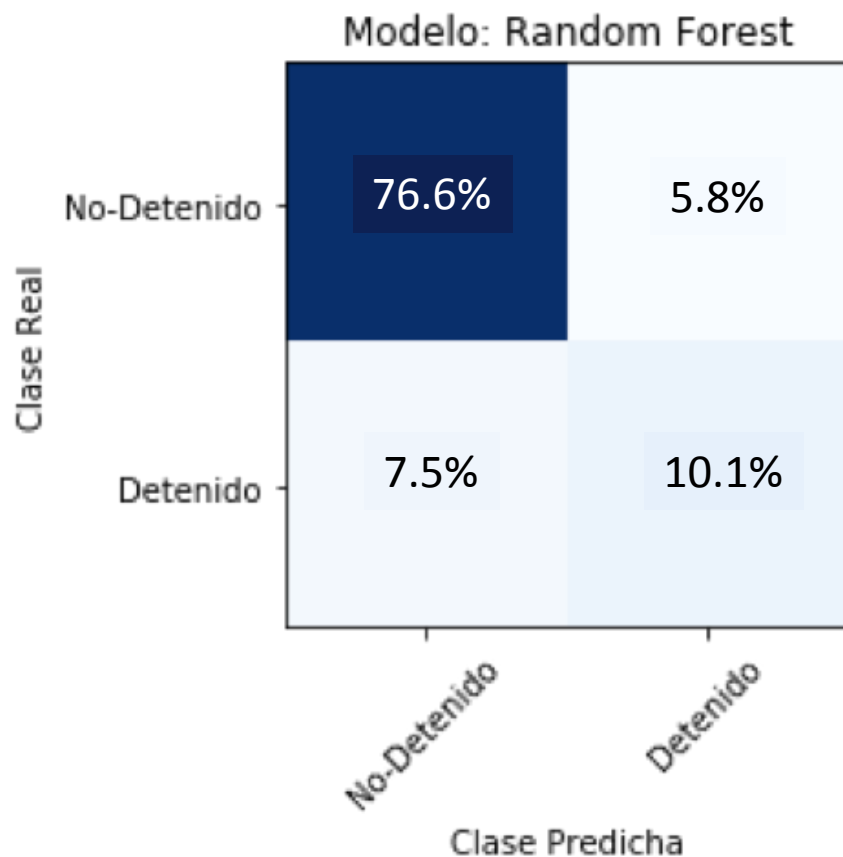
- Etiquetamos las personas en la base de datos que fueron detenidas por última vez en un tiempo menor a un año después la penúltima detención como 'detenida'.
- Hacemos un algoritmo de predicción en base de la historia excluyendo la última detención.
 - La predicción es en base de la historia de detenciones y datos biográficos.

	historia de detenciones	etiqueta
persona 1		detenida
persona 2		no detenida
persona 3		no detenida

Predicción si una persona será detenida en un año después su última detención.

En un 63% de los casos la predicción de ser detenida es correcta (precisión).

En un 57% de los casos de ser detenida la predicción es correcta (*recall*).



Los casos de falsos negativos y falsos positivos son considerables.

Los métodos de predicción tienen una precisión similar. El F-score es un promedio de la precisión y recall. Tener disponible más datos podría mejorar el modelo.

Modelo: Random Forest	Porcentaje
% de falsos entre las predicciones de ser detenida	37%
% de las personas detenidas no detectadas por la predicción	43%

Modelo	F-Score
Regresión Logística	0,51 ± 0,12
Naive Bayes	0,56 ± 0,07
Random Forest	0,56 ± 0,18
Árbol de Decisión	0,52 ± 0,23
MLP	0,56 ± 0,10

El modelo predictivo entrega conocimiento sobre los atributos claves.
En el modelo de Random Forest, los atributos más importantes en la predicción son relacionados a la historia de detenciones.

Indicador	Importancia relativa
Tiempo promedio entre detenciones	0.195
Número de detenciones	0.100
Familiares con detenciones	0.005
Género	0.001
Armas inscritas	0.001

El modelo usa más indicadores para la predicción que los mostrados en la tabla.

Estos resultados dependen intrínsecamente al método de predicción, la variable a predecir y la base de datos utilizada.

Además, no hay información sobre las condenas, sólo las detenciones.

Análisis Predictivo

Conclusiones

1. El modelo predictivo tiene una cantidad sustancial de 'falsos positivos', es decir, personas que son falsamente detectadas como un riesgo.
2. El modelo predictivo no detecta gran parte de las personas que son detenidas nuevamente.

Extensión

1. La situación actual del análisis predictivo no permite el uso de los modelos como herramienta en la priorización de recursos policiales.
2. Se espera obtener una precisión mayor con la disponibilidad de datos adicionales.

Desafíos de Herramientas Computacionales

Volumen, variedad, y valor de los datos

Agilizar el intercambio de la información administrativa en temas ligados a antecedentes de criminalidad que manejan actores gubernamentales.

Discusión ética y legal

Estudiar el potencial sesgo de las herramientas automatizadas y la priorización entre la seguridad ciudadana y privacidad en el sistema legal.

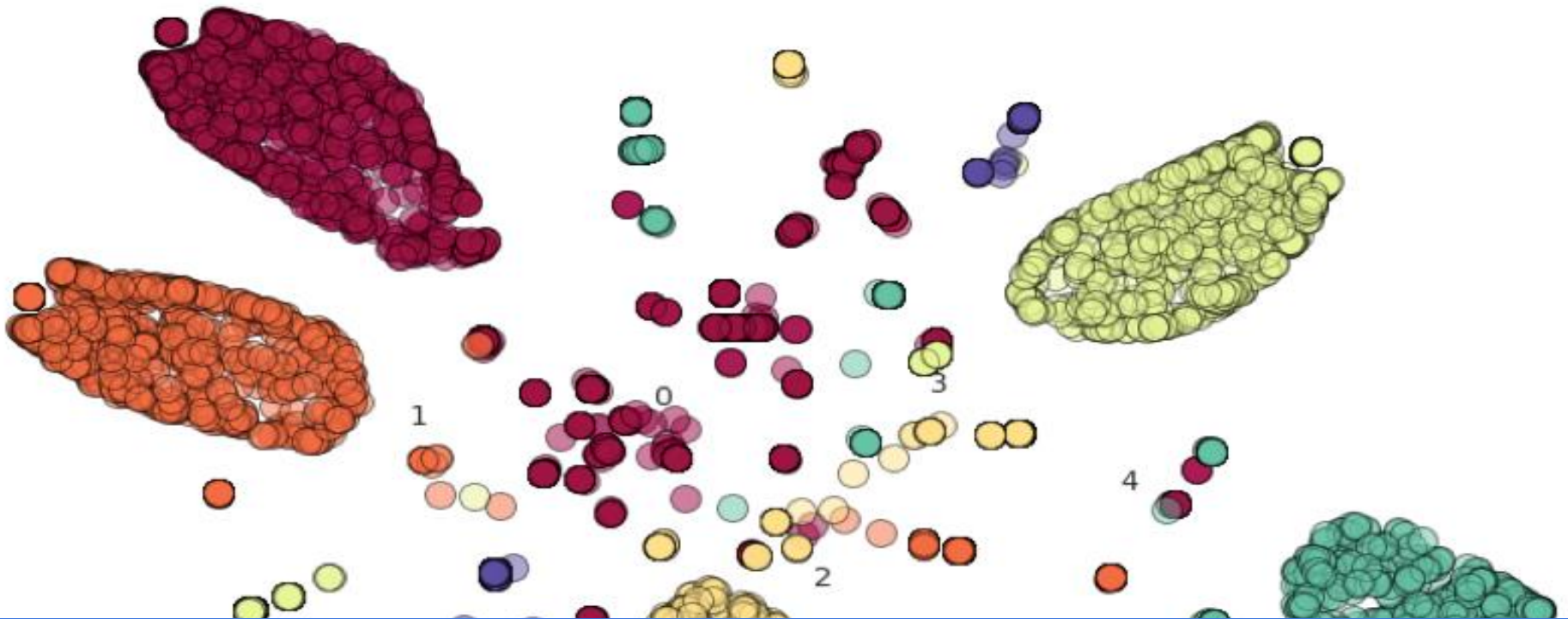
Síntesis

Propuesta en base del modelo descriptivo

Comenzar con el uso de las herramientas computacionales como un apoyo adicional en la gestión interna de la PDI y la especialización de los actores gubernamentales.

Propuesta en base del modelo predictivo

La baja precisión y los costos sociales que implica un falso positivo nos lleva a no recomendar aún el uso de modelos para la predicción del delito a nivel individual o para la asignación de recursos policiales.



Seminario de Políticas Públicas, 19 de noviembre de 2018

Big Data para la Elaboración de Políticas Públicas en la Identificación de Comportamiento Criminal

Elwin van 't Wout | Eduardo Valenzuela | Kenzo Asahi
Christian Pieringer | David Torres | Pilar Larroulet