



**TESIS DE GRADO  
INGENIERÍA INDUSTRIAL**

**CARACTERIZACIÓN DE LA POBLACIÓN  
CARCELARIA EN ARGENTINA MEDIANTE LA  
APLICACIÓN DE MINERÍA DE DATOS PARA LA  
PREVENCIÓN DE HECHOS DELICTIVOS.**

**AUTOR:**

**PATRICIO GUTIÉRREZ RÜEGG**

**DIRECTOR DE TESIS:**

**M. ING. PAOLA BRITOS**

**DR. RAMÓN GARCÍA-MARTÍNEZ**

**2008**



## **RESUMEN**

Desde hace ya varios años la inseguridad pasó a ser un tema recurrente en la vida de los argentinos. Reducir el índice de delitos es un tema que tiene muy presente cada nuevo gobierno. La toma de decisiones eficientes y la implementación de medidas preventivas en lo que respecta a política criminal y social, es el primer paso que se debería analizar a fin de lograr reducir la cantidad de hechos delictivos.

El análisis de la información criminal que los diferentes entes del Estado generan es un punto de vital importancia a los fines de prevención del delito. Se cree que no basta solo con analizar los registros sobre los distintos delitos cometidos en el pasado, sino que es necesario poder llevar a cabo estudios que involucren a los autores materiales de dichos hechos. Generalmente se analizan estadísticas descriptivas clásicas relacionadas con los hechos propiamente dichos, pero nunca se hace verdadero foco en la persona que ejecuta dicho delito, el delincuente. Se entiende que es necesario un tratamiento más complejo sobre la información criminal.

El presente proyecto se plantea en el plano de la información criminal, enfocándose específicamente en los datos de la población carcelaria en Argentina registrados por la Dirección Nacional de Política Criminal dependiente del Ministerio de Justicia y Derechos Humanos de la Nación Argentina.

Se busca desarrollar y proponer un modelo metodológico basado en la minería de datos que ayude a caracterizar a la sociedad carcelaria a fines de obtener patrones de comportamiento y conducta de los delincuentes. Si se lograra comprender los posibles motivos que llevaron a delinquir a esta porción de la sociedad, se facilitaría la generación de estrategias de prevención efectivas.



## **ABSTRACT**

For several years security has been a recurrent theme in the lives of Argentines. Reducing the crime index is a very pressing premise for every new administration. To reduce the amount of criminal acts the first step should be to analyze the politics of efficient decision-making and the implementation of preventive measures regarding criminal and social politics.

The analysis of criminal data generated by different federal agencies is of vital importance for crime prevention. It is believed that it is not enough to analyze the reports on different crimes committed in the past, but it is also necessary to study the actual perpetrators of such acts. Generally, studies focus on classic descriptive statistics related to the events in question, but never truly focus on the person who committed the crime, the criminal. It is therefore understood that a more complex management of criminal data is needed.

The current project is based on the criminal information focusing specifically on the data from the incarcerated population in Argentina registered by the *Dirección Nacional de Política Criminal* (National Department of Criminal Law) depending on the *Ministerio de Justicia y Derechos Humanos de la Nación Argentina* (The Argentinean National Ministry of Justice and Human Rights.)

The objective is to develop and propose a methodological model based on data mining that helps characterize the inmate population thus obtaining behavioral and personality traits of criminals. If an understanding can be achieved of the possible motives that drove this sector of our society to crime, then development of strategies for effective prevention would be facilitated.



## **TABLA DE CONTENIDOS**

1. INTRODUCCIÓN.....	1
1.1 INTRODUCCIÓN AL PROBLEMA .....	1
1.2 MOTIVACIONES Y DESTINATARIOS .....	3
1.3 COMPOSICIÓN DEL TRABAJO.....	3
2. ESTADO DE LA CUESTIÓN .....	5
2.1 INTRODUCCIÓN A LA MINERÍA DE DATOS.....	5
2.1.1 Algoritmos de Minería de Datos.....	6
2.1.1.1 Algoritmos de Clustering o Agrupamiento.....	7
2.1.1.1.1 SOM.....	7
2.1.1.1.2 K-Means .....	9
2.1.1.2 Algoritmos de Clasificación: Inducción .....	10
2.1.1.2.1 ID3.....	10
2.1.1.2.2 C4.5 .....	11
2.1.1.2.3 CHAID .....	12
2.1.1.3 Selección de atributos.....	13
2.2 CRIPS-DM: METODOLOGÍA DE PROYECTOS DE MINERÍA DE DATOS.....	14
2.2.1 Objetivos de cada fase en la metodología CRISP-DM. ....	16
2.3 MINERÍA DE DATOS APLICADA A LA INFORMACIÓN CRIMINAL.....	17
2.3.1 Proyecto <i>Able Danger</i> .....	18
2.3.2 Proyecto <i>COPLINK</i> – Universidad de Arizona .....	19
2.4 INFORMACIÓN CRIMINAL EN ARGENTINA .....	23
2.4.1 Dirección Nacional de Política Criminal .....	23
2.4.1.1 Fuentes Oficiales de Información y Estudios de Victimización.....	23
2.4.1.1.1 Sistema Nacional de Información Criminal (SNIC) .....	24
2.4.1.1.2 Sistema Nacional de Estadísticas Judiciales (SNEJ).....	26
2.4.1.1.3 Encuestas de Victimización.....	27
2.4.1.1.4 Sistema Nacional de Estadísticas sobre Ejecución de la Pena (SNEEP) ...	28
2.4.1.1.5 Sistema Unificado de Registros Criminales (SURC).....	31
2.4.1.1.6 Mapa del delito de la Ciudad Autónoma de Buenos Aires .....	33
2.5 DIAGNÓSTICO DEL SERVICIO PENITENCIARIO EN ARGENTINA.....	34
2.5.1 Evolución de la Población Penal en Argentina.....	34
2.5.2 Situación a nivel internacional.....	37
2.5.3 Sobrepoblación y hacinamiento en Argentina .....	39
2.5.3.1 Sobrepoblación. Dimensión del problema.....	40
2.5.4 Situación penal de los detenidos .....	41
2.5.5 Costo promedio de cada preso. ....	42
2.5.6 Trabajo en las cárceles .....	43
2.5.7 Reflexiones acerca del diagnóstico del Servicio Penitenciario.....	43

3. DESCRIPCIÓN DEL PROBLEMA .....	45
3.1. TRATAMIENTO DE LA INFORMACIÓN .....	45
3.2. EL PROBLEMA ESPECÍFICO .....	45
4. SOLUCIÓN.....	47
4.1 PROPUESTA DE SOLUCIÓN .....	47
4.2 APLICACIÓN DE CRISP-DM .....	47
4.2.1 Fase I: Comprensión del negocio .....	48
4.2.1.1 Determinar los objetivos del negocio.....	48
4.2.1.2 Evaluación de la Situación .....	48
4.2.1.3 Plan de Proyecto.....	50
4.2.2 Fase II: Comprensión de los datos.....	51
4.2.2.1 Recolección de los datos iniciales .....	51
4.2.2.2 Descripción de los datos iniciales .....	52
4.2.2.2.1 Dataset seleccionado .....	54
4.2.2.3 Exploración de los datos .....	58
4.2.2.4 Verificación de Calidad de los Datos.....	64
4.2.3 Fase III: Preparación de los datos.....	65
4.2.3.1 Preparación del Dataset.....	65
4.2.3.2 Selección de los Datos .....	66
4.2.3.3 Limpieza de los Datos .....	68
4.2.3.4 Construcción e integración de los Datos .....	68
4.2.4 Fase IV: Modelado .....	70
4.2.4.1 Selección de la técnica de modelado.....	71
4.2.4.2 Construcción del modelo y resultados experimentales .....	71
4.2.4.2.1 Clustering.....	71
4.2.4.2.2 Resultados del Clustering .....	72
4.2.4.2.3 Selección de Atributos .....	86
4.2.4.2.4 Clasificación .....	87
4.2.4.3 Evaluar el modelo. ....	96
4.2.5 Fase V: Evaluación.....	97
5. CONCLUSIONES Y REFLEXIONES.....	99
5.1 CONCLUSIONES DEL PROYECTO .....	99
5.2 REFLEXIONES.....	100
6. FUTURAS LÍNEAS DE INVESTIGACIÓN .....	103
7. APORTES REALIZADOS .....	105
8. REFERENCIAS .....	107
9. ANEXOS.....	111

## **1. INTRODUCCIÓN**

### **1.1 INTRODUCCIÓN AL PROBLEMA**

La situación del régimen penitenciario en Argentina es un tema que tiene preocupado a los últimos gobiernos desde hace ya varios años. Los incrementos en la población carcelaria hacen suponer que las tasas de delitos deberían haber bajado, sin embargo ocurrió lo contrario. Desde el año 1990 al 2006 el total de hechos delictuosos creció de 560.240 a 1.224.293, lo que representa un aumento de aproximadamente 119%. A su vez, la población alojada en cárceles federales creció un 110% en el mismo período [Dirección Nacional de Política Criminal, 2006]. Aún más, según un estudio elaborado en el Ministerio de Justicia, se prevé que el número de presos en cárceles federales seguirá creciendo en el futuro llegando a unos 22.000 para el año 2025. Esto demuestra que el número de personas presas crece a un ritmo mucho mayor que la población del país, la cual creció solo un 20% desde el año 1990 al 2006 [INDEC, 2007].

En lo que refiere al régimen penitenciario en Argentina, las estadísticas muestran números realmente preocupantes en todos los aspectos a tener en cuenta cuando se habla de cárceles en Latinoamérica, particularmente en Argentina. La sobrepoblación, el hacinamiento, la promiscuidad, el avance del sida, la violencia, son algunos de los temas que se presentan cuando se habla de sociedades carcelarias en Argentina. Por tanto, se cree que sin una política criminal eficiente y reparadora, será imposible bajar en el mediano plazo el alto índice de delitos existente. En el capítulo 2.5 se realiza un diagnóstico sobre el sistema penitenciario en nuestro país, en donde se pueden observar las cuestiones planteadas y sus indicadores.

Debido a estos incrementos y cuestiones alarmantes se cree de suma importancia y relevancia realizar un estudio que involucre a la población carcelaria en Argentina como punto insoslayable dentro de la política criminal.

La problemática planteada lleva a la necesidad de contar con herramientas que permitan desarrollar un diagnóstico válido sobre el estado actual de la cuestión carcelaria, ya que es de vital importancia a los fines de aplicar una política eficaz en la materia. El Sistema Nacional de Estadísticas sobre Ejecución de la Pena (SNEEP) es un aporte en tal sentido. Este sistema de información fue implementado en el año 2002 por la Dirección Nacional de Política Criminal con el fin de contar con información periódica y uniforme acerca de la población penal privada de libertad en la República Argentina. Si bien surgen ciertos límites metodológicos, y hasta ideológicos, al resultar la fuente primaria de información los propios organismos de ejecución de la pena, se cree que se trata de una buena mirada sobre la cuestión que tiene una enorme relevancia [Dirección Nacional de Política Criminal, 2005].

Sin embargo, se cree que con el solo contar con estadísticas vinculadas a los registros criminales o, en el caso del SNEEP, de los criminales propiamente dichos, no es suficiente para poder tomar decisiones completas y correctas en lo que a política criminal respecta. Tradicionalmente se han hecho análisis en intervalos de tiempos acotados sobre alocación de recursos para la prevención de delitos basándose en resultados obtenidos de herramientas estadísticas clásicas. El problema es que estos no reflejan la distribución real de la actividad criminal, volviéndose no tan útiles para el desarrollo de medidas preventivas con el fin de evitar hechos delictivos [Corcoran *et al*, 2004].

Es necesario un tratamiento de la información estadística más complejo, aún más cuando se trabajan con base de datos de más de 50,000 registros, como es el caso de la población carcelaria en Argentina. El objetivo primario de estas bases de datos es, como su nombre indica, almacenar grandes cantidades de datos organizados siguiendo un determinado esquema o modelo de datos que facilite su almacenamiento, recuperación y modificación, pero no así su posterior uso o análisis. En muchos casos los registros almacenados son demasiados grandes y complejos como para analizar [Kantardzic, 2003].

Una posible herramienta a utilizar para tratar los grandes volúmenes de información almacenados en dichas bases de datos es la técnica de Minería de Datos. El termino *Minería de Datos (Data Mining)* puede ser definido como la manera no trivial de extracción de información no implícita, previamente desconocida y potencialmente útil, de una base de datos [Frawley, *et al*, 1992]. Minería de datos representa la posibilidad de buscar exhaustivamente dentro de un gran volumen de datos, información y conocimiento que pueden resultar de mucho valor. Es considerada uno de los puntos más importantes de los sistemas expertos de base de datos, y uno de los desarrollos más prometedores en la industria del manejo de la información [Cartagenova, 2005].

En el marco internacional el uso de la minería de datos aplicado a la inteligencia criminal ha tenido un gran crecimiento en los últimos años en EEUU [Chen *et al*, 2004]. A modo de ejemplo, ha sido mencionado como el método mediante el cual supuestamente las fuerzas armadas estadounidenses habrían identificado al líder de los ataques del 11 de Septiembre de 2001, Mohamed Atta, junto a otros 3 terroristas como posibles miembros de una célula de *Al Qaeda* operando en ese país más de un año previo a los ataques.

En este contexto, el objetivo principal de este trabajo es estudiar la factibilidad y valor agregado de aplicar técnicas de minería de datos para caracterizar a la población ubicada en las cárceles argentinas, obteniendo patrones de comportamientos y conductas, a fines de poder extraer conclusiones para la prevención de delitos.

## **1.2 MOTIVACIONES Y DESTINATARIOS**

Luego de leer proyectos de colegas relacionados con la minería de datos e información criminal [Perversi, 2007; Valenga, 2007], me pareció muy interesante poder seguir dicha línea de investigación pero haciendo foco no en los delitos propiamente dichos sino precisamente en las personas que cometen dichos hechos delictuosos. Se espera que mediante el presente trabajo se pueda contribuir para llevar a cabo tareas y actividades en materia de política criminal, que logren disminuir hechos delictivos a través la presentación, descripción y aplicación de la minería de datos como herramienta que ayuda al análisis y tratamiento de la información criminal.

Es importante destacar, que no se encontraron registros sobre trabajos o proyectos en donde se aplicaran herramientas como la minería de datos en poblaciones carcelarias, lo cual representa una motivación extra.

El presente informe se encuentra dirigido a aquellas personas involucradas en la creación de política criminal como a los interesados en conocer la utilidad de esta técnica para la modernización de las prácticas de información criminal en Argentina.

## **1.3 COMPOSICIÓN DEL TRABAJO**

El presente trabajo se encuentra estructurado de la siguiente manera.

### **Capítulo 1: Introducción.**

Se da una breve descripción a las razones que llevaron a la realización del proyecto final. A su vez se nombran las motivaciones, destinatarios y como se organiza el trabajo.

### **Capítulo 2: Estado de la cuestión.**

Comienza con una introducción a la minería de datos (MD) y la metodología elegida para llevar a cabo el proyecto de MD. Continúa con aplicaciones de la MD en información criminal, tanto a nivel nacional como internacional. Luego sigue con la descripción de la información criminal existente en Argentina. Finalmente, se da un diagnóstico de la situación carcelaria en Argentina, mostrando todos los problemas existentes en dicho régimen penitenciario.

### **Capítulo 3: Descripción del problema.**

Presenta al problema que se quiere resolver y los principales motivos de su elección.

#### **Capítulo 4: Solución**

En el presente capítulo se propone una solución al problema planteado en el capítulo 3. Se decide aplicar la metodología CRISP-DM al trabajo, pasando por todas sus fases.

#### **Capítulo 5: Conclusiones y Reflexiones.**

Se presentan las principales conclusiones del proyecto. Se escriben reflexiones sobre el tema investigado y el conocimiento adquirido.

#### **Capítulo 6: Futuras líneas de investigación.**

El presente capítulo contempla el abordaje de las futuras líneas de investigación relacionadas al proyecto desarrollado.

#### **Capítulo 7: Aportes realizados.**

Se mencionan los aportes realizados por el proyecto realizado.

#### **Capítulo 8: Referencias.**

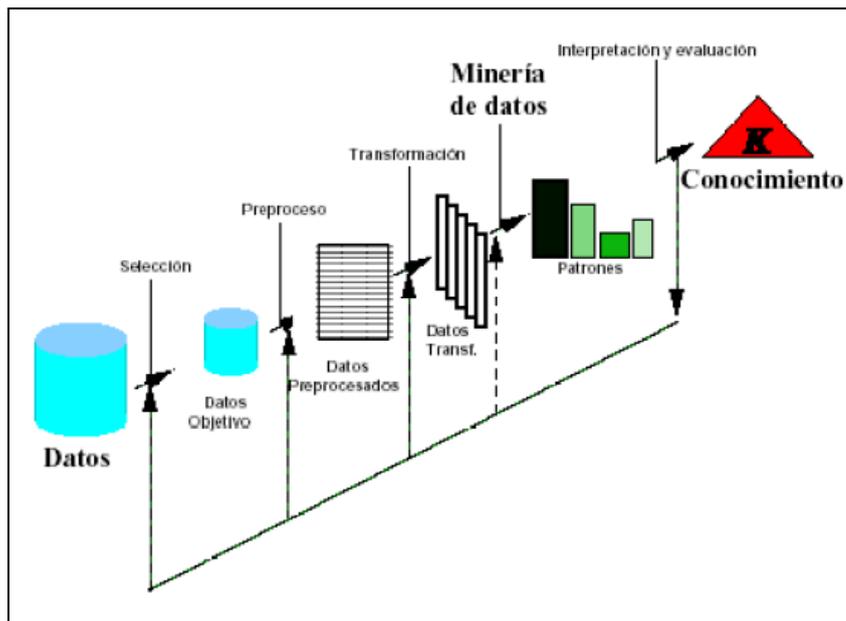
Bibliografía utilizada para la elaboración de la tesis.

#### **Capítulo 9: Anexos.**

## 2. ESTADO DE LA CUESTIÓN

### 2.1 INTRODUCCIÓN A LA MINERÍA DE DATOS

El termino minería de datos es una etapa dentro de un proceso mayor llamado *Extracción de Conocimiento en Bases de Datos (Knowledge Discovery in Databases o KDD)* [Figura 2.1], aunque ambos términos se usan de manera indistinta. Lo que en verdad hace la minería de datos es reunir las ventajas de varias áreas como la estadística, la inteligencia artificial, la computación gráfica, las bases de datos y el procesamiento masivo, principalmente usando como materia prima las bases de datos. Una definición tradicional [Fayyad *et al*, 1996] podría ser la siguiente: “*Un proceso no trivial de identificación válida, novedosa, potencialmente útil y entendible, de patrones comprensibles que se encuentran ocultos en los datos*”



**Figura 2.1.** Proceso de *Knowledge Discovery in Databases*.

El creciente volumen de datos en todas las áreas de aplicación humana demanda nuevas y poderosas técnicas de transformación de los datos en conocimiento útil. Se puede decir que la minería de datos es un avance en dicha cuestión. Busca generar información similar a la que podría producir un experto humano, que además satisfaga el Principio de Comprensibilidad (utilizar lenguaje adecuado a la temática del trabajo) [Britos *et al*, 2005].

El objetivo de la minería de datos es descubrir comportamientos interesantes como lo son patrones, asociaciones, cambios, anomalías y estructuras significativas a partir de grandes cantidades de datos almacenados en bases de datos, *data warehouses* o

cualquier otro medio de almacenamiento de información [Britos *et Al*, 2005].

La minería de datos produce, mediante diversas técnicas y algoritmos, cinco tipos de información:

- Clustering o agrupamiento
- Inducciones o clasificaciones
- Asociaciones
- Secuencias
- Pronósticos

### **2.1.1 Algoritmos de Minería de Datos**

Se define *algoritmo de minería de datos* [Thearling, 2007] al mecanismo utilizado para la creación de modelos. Para crear un modelo, un algoritmo analiza primero un conjunto de datos, buscando patrones y tendencias específicos. Después, el algoritmo utiliza los resultados de este análisis para definir los parámetros del modelo de minería de datos.

Los algoritmos de minería de datos [Weiss & Indurkha, 1998] se clasifican en dos grandes categorías:

- Supervisados (o predictivos): Predicen el valor de un atributo (etiqueta) de un conjunto de datos, conocidos otros atributos (atributos descriptivos). A partir de datos cuya etiqueta se conoce se induce una relación entre dicha etiqueta y otra serie de atributos. Esas relaciones sirven para realizar la predicción en datos cuya etiqueta es desconocida. Se desarrolla en dos fases: Entrenamiento y prueba.
- No Supervisados (o de descubrimiento de conocimiento): Cuando una aplicación no es lo suficientemente madura no tiene el potencial necesario para una solución predictiva. En ese caso hay que utilizar métodos no supervisados o de descubrimiento de conocimiento, los cuales descubren comportamientos, patrones y tendencias en los datos actuales (no utilizan información histórica). El descubrimiento sirve para llevar a cabo acciones y obtener un beneficio de ellas.

En la tabla 2.1 se pueden observar algunas de las técnicas de minería de datos de ambas categorías, supervisados y no supervisados.

Supervisados	No Supervisados
Arboles de Decisión	Detección de desvíos
Inducción Neuronal	Segmentación
Regresión	Agrupamiento ("clustering")
Series Temporales	Reglas de Asociación
	Patrones Secuenciales

**Tabla 2.1.** Clasificación de las técnicas de minería de datos.

### **2.1.1.1 Algoritmos de Clustering o Agrupamiento**

El clustering consiste en agrupar un conjunto de datos sin tener clases predefinidas, basándose en la similitud de los valores de los atributos de los distintos datos. Este tipo de algoritmo, tal como lo muestra la tabla 2.1, se realiza en forma no supervisada, ya que no se saben de antemano las clases del conjunto de datos de entrenamiento. El clustering identifica regiones densamente pobladas, de acuerdo a alguna medida de distancia, en un gran conjunto de datos multidimensional [Chen & Han, 1996]. Se basa en maximizar la similitud de las instancias en cada cluster y minimizar la similitud entre clusters [Han & Kamber, 2001].

El análisis de clusters [Cartagenova, 2005] es utilizado en numerosas aplicaciones tales como reconocimiento de patrones, análisis de datos, procesamiento de imágenes e investigaciones de mercado. Como función de la MD, el análisis de clusters puede ser utilizado como una herramienta independiente para obtener una visión de la distribución de los datos, para observar las características de cada cluster y enfocar un análisis más exhaustivo hacia un grupo o cluster determinado. Alternativamente, puede servir como un paso del preprocesamiento de los datos para otros algoritmos, como por ejemplo, el de clasificaciones en el cual se trabajaría luego sobre los clusters originados.

El clustering de datos esta en continuo desarrollo, y debido a los grandes volúmenes de datos almacenados en las bases de datos, el análisis de clusters se ha vuelto muy importante en los estudios de MD.

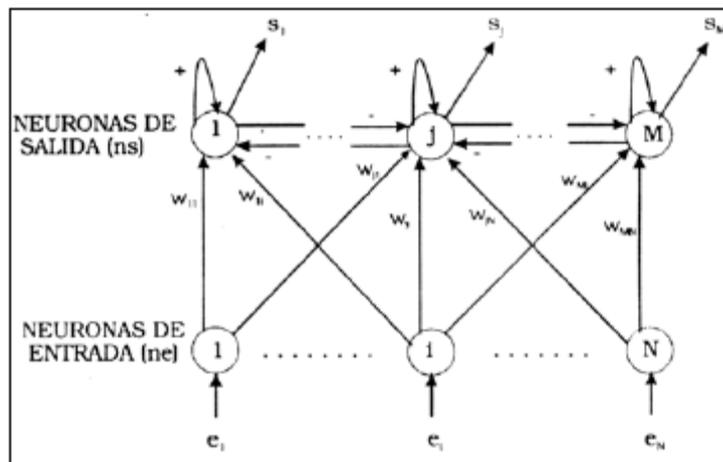
#### **2.1.1.1.1 SOM**

Una de las técnicas más utilizadas para realizar clustering es la de SOM (*Self Organizing Maps*). Este modelo no supervisado de red neuronal fue creado por Teuvo Kohonen en 1982, basándose en el funcionamiento de las neuronas biológicas. Vale recordar, que las redes neuronales son modelos que intentan reproducir el comportamiento del cerebro. Del mismo modo que aquél, realiza una simplificación, averiguando cuales son los elementos relevantes del sistema. El objetivo de Kohonen era demostrar que un estímulo externo por sí solo, suponiendo una estructura propia y

una descripción funcional del comportamiento de la red, era suficiente para forzar la formación de los mapas.

El modelo tiene dos variantes, LVQ (*Learning Vector Quantization*) y TPM (*Topology Preserving Map*) o SOM (*Self Organizing Map*). Ambas se basan en el principio de formación de mapas topológicos para establecer características comunes entre las informaciones (vectores) de entrada a la red, aunque difieren en las dimensiones de éstos, siendo de una sola dimensión en el caso de LVQ y bidimensional e incluso tridimensional en la red SOM o TPM.

El modelo presenta dos capas con N neuronas de entrada y M de salida [Figura 2.2]. Cada una de las N neuronas de entrada se conecta a las M de salida a través de conexiones hacia delante (*feedforward*). Entre las neuronas de la capa de salida, existen conexiones laterales de inhibición (peso negativo) implícitas, a pesar de no estar conectadas, cada una de estas neuronas va a tener cierta influencia sobre sus vecinas. El valor que se asigne a los pesos de las conexiones *feedforward* entre las capas de entrada y salida ( $W_{ij}$ ) durante el proceso de aprendizaje de la red va a depender precisamente de esta interacción lateral. La influencia que cada neurona ejerce sobre las demás es función de la distancia entre ellas, siendo muy pequeñas cuando están muy alejadas.



**Figura 2.2.** Estructura de una red de Kohonen.

Los puntos que están cerca unos de los otros en el espacio original de entrada son “mapeados” a neuronas cercanas en SOM. Las neuronas usualmente forman un mapa bidimensional, por lo que el mapeo transforma un problema de muchas dimensiones en el espacio, a un plano.

### 2.1.1.1.2 K-Means

Otro de los algoritmos más utilizados cuando se habla de clustering es el K-Medias o *K-Means*. Se trata de un método de agrupamiento por vecindad en el que se parte de un número determinado de prototipos y de un conjunto de ejemplos a agrupar sin etiquetar. Es el método más popular de los métodos de agrupamiento “por partición”.

La idea de *K-Means* es situar a los prototipos o centros en el espacio, de forma que los datos pertenecientes al mismo prototipo tengan características similares [Moody & Draken, 1989]. Todo ejemplo nuevo, una vez que los prototipos han sido correctamente situados, es comparado con estos y asociado a aquel que sea el más próximo, en los términos de una distancia previamente elegida. Normalmente, se utiliza la distancia euclidiana.

La etapa de entrenamiento del método puede ser lenta dependiendo del número de puntos a clasificar y de la dimensión del problema. Una vez superada esta fase, la clasificación de nuevos datos se torna rápida, debido a que la comparación de distancias es solo con los prototipos.

El objetivo que se busca mediante el algoritmo *K-Means* es minimizar la varianza total intra-grupo o la función de error cuadrático [Figura 2.3]:

$$V = \sum_{i=0}^K \sum_{j \in S_i} |x_j - \mu_i|^2$$

**Figura 2.3.** Error cuadrático K-Means

Donde existen K grupos  $S_i$   $i=1,2,\dots, k$  y  $\mu_i$  es el punto medio o centroide de todos los puntos  $X_j \in S_i$ .

*K-Means* comienza particionando los datos en k subconjuntos no vacíos, aleatoriamente o usando alguna heurística. Luego calcula el centroide de cada partición como el punto medio del cluster y asigna cada dato al cluster cuyo centroide sea el más próximo. Luego los centroides son recalculados para los grupos nuevos y el algoritmo se repite hasta la convergencia, la cual es obtenida cuando no haya más datos que cambien de grupo de una iteración a otra.

Para calcular el centroide más cercano a cada punto se debe utilizar una función de distancia. Para datos reales se suele utilizar la distancia euclidiana. Para datos categóricos se debe establecer una función específica de distancia para ese conjunto de

datos. Algunas de las opciones son utilizar una matriz de distancias predefinidas o una función heurística.

Dado  $k$ , el algoritmo *K-Means* se implementa en los siguientes 4 pasos [Ale, 2005]:

1. Particionar los objetos en  $k$  subconjuntos no vacíos.
2. Computar los centroides de los clusters de la partición corriente. El centroide es el centro (punto medio) del cluster.
3. Asignar cada objeto al cluster cuyo centroide sea más cercano.
4. Volver al paso 2, y parar cuando no existan más reasignaciones.

### **2.1.1.2 Algoritmos de Clasificación: Inducción**

Los algoritmos de clasificación se utilizan para clasificar un conjunto de datos basado en los valores de sus atributos [Servente & García-Martínez, 2002].

Este proceso encuentra las propiedades comunes entre un conjunto de objetos y los clasifica en diferentes clases de acuerdo al modelo utilizado. Para construir el mismo se utiliza un conjunto de entrenamiento, en que cada instancia consiste en un conjunto de atributos y el valor de la clase a la cual pertenece. El objetivo de la clasificación es analizar los datos de entrenamiento y, mediante un método supervisado, desarrollar una descripción o un modelo para cada clase utilizando las características disponibles en los datos. Esta descripción o modelo es utilizado para clasificar otras instancias, cuya clase es desconocida. El método se conoce como supervisado debido a que, para el conjunto de entrenamiento, se conoce la clase a la cual pertenece cada instancia y se lo indica al modelo si la clasificación que realiza es correcta [Chen *et al*, 1996].

Los algoritmos más utilizados para la clasificación son los algoritmos de inducción. Aún cuando existen varios enfoques para los algoritmos de inducción, se trabajará con aquellos que generan árboles de decisión conocida como la familia TDIT (*Top Down Induction Trees*). Cada hoja del árbol posee un nombre de clase, cada nodo interior especifica una evaluación en un atributo y cada rama representa un resultado de esa evaluación. La calidad de un árbol depende de la precisión de la clasificación y del tamaño del árbol. En particular se hará hincapié en los algoritmos **ID3**, **C4.5** y **CHAID**.

#### **2.1.1.2.1 ID3**

El ID3 (*Iterative Dichotomiser 3*) desarrollado en los años ochenta por Ross Quinlan, es un sistema de aprendizaje supervisado que construye árboles de decisión a partir de un

conjunto de ejemplos [Servente & García Martínez, 2002]. Estos ejemplos están constituidos por un conjunto de atributos y un clasificador o clase. Los dominios de los atributos y de las clases deben ser discretos. Además, las clases deben ser disjuntas. Las primeras versiones del ID3 generaban descripciones únicamente para dos clases, como ser positiva y negativa. En las versiones posteriores, se eliminó esta restricción, pero se mantuvo la restricción de las clases disjuntas.

El ID3 genera descripciones que clasifican a cada uno de los ejemplos del conjunto de entrenamiento. El nivel de precisión en la clasificación generalmente es alto. Sin embargo, el sistema tiene algunas desventajas. Los atributos y clases deben ser discretos y no pueden ser continuos. Además, aún cuando se cuenta con conocimientos de dominio o conocimientos previos, el sistema no hace uso de ellos. A veces, los árboles son demasiado frondosos, lo cual conlleva una difícil interpretación. En esos casos pueden ser transformados en reglas de decisión para hacerlos más comprensibles.

Hay varias razones para podar los árboles generados por los métodos de TDIDT: la sobregeneralización, evaluación de atributos poco importantes o significativos, y el gran tamaño del árbol. Si el árbol es demasiado grande, se dificulta la interpretación, con lo cual hubiera sido lo mismo utilizar un método de caja negra.

Existen dos enfoques para podar árboles [Ochoa, 2004]:

- Pre-poda (*pre-pruning*): detiene el crecimiento del árbol cuando la ganancia de información producida al dividir un conjunto no supera un umbral determinado
- Post-poda (*post-pruning*): se aplica sobre algunas ramas una vez que se ha terminado.

#### **2.1.1.2.2 C4.5**

El C4.5 [Quinlan, 1993] es una extensión del ID3 que acaba con muchas de sus limitaciones. Por ejemplo, permite trabajar con valores continuos para los atributos, separando los posibles resultados en dos ramas: una para aquellos  $A \leq N$  y otra para  $A > N$ . Además, los árboles son menos frondosos porque cada hoja no cubre una clase en particular sino una distribución de clases, lo cual los hace menos profundos y frondosos.

El C4.5 genera un árbol de decisión a partir de los datos mediante particiones realizadas recursivamente, según la estrategia de profundidad-primero (*depth-first*) [Servente & García Martínez, 2002]. Antes de cada partición de datos, el algoritmo considera todas las pruebas posibles que pueden dividir el conjunto de datos y selecciona la prueba que resulta en la mayor ganancia de información o en la mayor proporción de ganancia de información. Para cada atributo discreto, se considera una prueba con  $n$  resultados, siendo  $n$  el número de valores posibles que puede tomar el atributo. Para cada atributo

continuo, se realiza una prueba binaria sobre cada uno de los valores que toma el atributo en los datos.

Son numerosas las mejoras que presenta C4.5 en relación a ID3 [Lopez Takeyas, 2005]:

- Evitar sobreajuste de los datos.
- Determinar que tan profundo debe crecer el árbol de decisión.
- Reducir errores en la poda.
- Condicionar la Post-Poda.
- Manejar atributos continuos.
- Escoger un rango de medida apropiado.
- Manejar datos de entrenamiento con valores faltantes.
- Manejar de atributos con diferentes valores.
- Mejorar la eficiencia computacional.

### **2.1.1.2.3 CHAID**

CHAID fue diseñada a partir de una técnica conocida como *Detección Automática de Interacciones (AID: Automatic Interaction Detection)*. De aquí obtuvo su nombre, Detección Automática de Interacciones con Prueba  $\chi^2$ , CHAID (*Chi Squared Automatic Interaction Detection*) [Hartigan, 1975]. Es tanto una técnica de explotación de datos como un modelo estadístico. Forma parte, al igual que el ID3 y C4.5, de los modelos llamados árboles de decisión y se utiliza para realizar clasificaciones y pronósticos. La finalidad de la predicción es pronosticar el objetivo o resultado según una futura serie de criterios o variables independientes.

El modelo utiliza el algoritmo CHAID para dividir en grupos los registros que presenten la misma probabilidad de resultado, basándose en los valores de las variables independientes. El algoritmo parte de un nodo raíz y se va bifurcando en nodos descendientes hasta llegar a los nodos hoja, donde finaliza la ramificación. El mismo se sirve de la Prueba de la  $\chi^2$  (*chi squared test*) para determinar si se debe continuar con la ramificación y, en caso afirmativo, qué variables independientes usar. Esta prueba se lleva a cabo mediante la tabulación cruzada entre el resultado y cada una de las variables independientes. El resultado es un “valor-p”. Este valor representa la probabilidad de que la hipótesis nula sea correcta. Luego, los “valores-p” para cada tabulación cruzada de todas las variables independientes se clasifican, y si el mejor (el valor más pequeño) se encuentra bajo un umbral determinado, se realiza una ramificación del nodo raíz en esa ubicación.

Las valoraciones y ramificaciones se prolongan hasta construir un árbol. Cuanto más se alarguen las ramas, menos variables independientes habrá disponibles ya que el

alargamiento se lleva a cabo precisamente con las variables. La bifurcación llega a su fin cuando el mejor “valor-p” ya no se encuentra bajo dicho umbral determinado. Los nodos hoja del árbol son aquellos que no han sufrido ramificaciones.

Entre los puntos fuertes del CHAID figura la construcción sencilla del modelo, que permite manejar variables independientes tanto categóricas como continuas. CHAID requiere grandes masas de datos para asegurarse de que la cantidad de observaciones de los nodos hoja es significativa.

### **2.1.1.3 Selección de atributos**

Existen varias razones para realizar la selección de atributos [Hernández Orallo *et al*, 2004].

- Reducir el tamaño de los datos eliminando los irrelevantes o redundantes.
- Eliminar atributos que contienen numerosos datos erróneos o faltantes.
- Mejorar la calidad del modelo centrándose en atributos relevante.
- Expresar el modelo resultante en función de menos variables mejorando la comprensión.
- Reducir la dimensionalidad a fin de representar los datos visualmente.

El problema de hacer uso de todos los atributos disponibles consiste en que la mayoría de las técnicas de minería de datos pueden actuar incorrectamente entre tantos campos en un espacio que al tener alta dimensionalidad resulta estar más desierto (cuando hay irrelevantes, redundantes o con valores erróneos). Se busca obtener modelos que se ajusten a particularidades de los datos de entrenamiento y no de los datos en general.

Existen dos reglas generales para eliminar atributos, en especial los nominales:

- Eliminación de claves candidatas: la regla es eliminar cualquier atributo que pueda ser clave primaria de la tabla (sea candidata o parte de la clave candidata, parcial o totalmente). Estos campos agregan un detalle al dominio de datos que puede ser especialmente problemático para tareas de clasificación o regresión.
- Eliminación de atributos dependientes: cuando existen dependencias funcionales entre atributos se intenta normalizar en varias tablas. Dado que los datos a minar pueden provenir de una “vista minable” (tabla lógica con todos los atributos relevantes) que los ha desnormalizado, se debe evaluar si realmente se necesitan todos los atributos, ya que muchos pueden ser redundantes. No eliminar los atributos dependientes puede perjudicar la generación de reglas de asociación y las de agrupamiento.

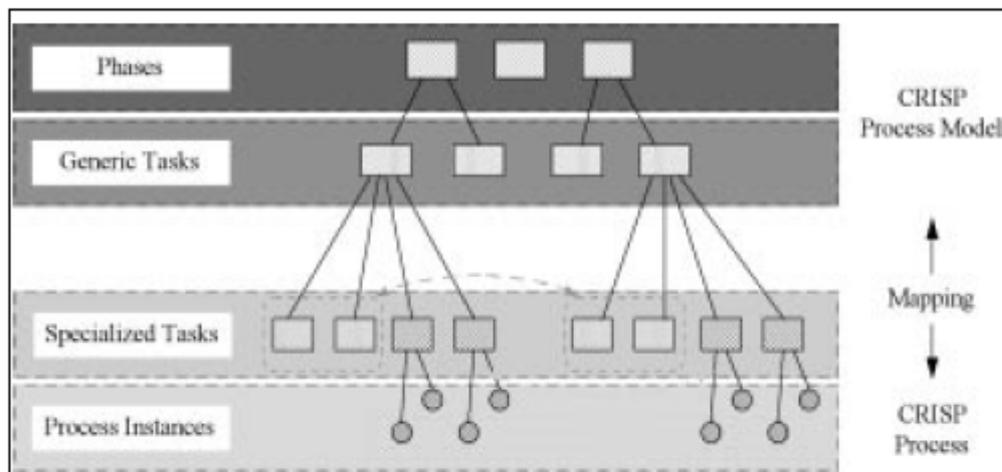
Por otra parte, se utilizan dos métodos para seleccionar atributos:

- Métodos de filtro: se filtran los atributos irrelevantes antes de cualquier proceso de minería de datos. Las técnicas son fundamentalmente estadísticas (medidas de información, distancia, dependencias o inconsistencias). El criterio para establecer el subconjunto óptimo de atributos se basa en medidas de calidad previa que se calculan a partir de los datos.
- Métodos envolventes (*Wrappers*): la selección óptima de atributos se evalúa respecto a la calidad de un modelo de minería de datos o estadístico extraído a partir de los datos (utilizando algún método de validación). Este tipo de técnicas requieren más tiempo que las otras ya que para evaluar hay que entrenar un modelo.

## **2.2 CRIPS-DM: METODOLOGÍA DE PROYECTOS DE MINERÍA DE DATOS**

CRISP-DM (*Cross Industry Standard Process for Data Mining*) es una metodología creada en el año 1996, diseñada para desarrollar proyectos de minería de datos.

La misma consiste en una serie de actividades o tareas descritas en cuatro niveles de abstracción (desde lo general a lo específico): fases, tareas genéricas, tareas específicas e instancias del proceso [Figura 2.4].



**Figura 2.4.** Los 4 niveles del CRISP-DM. [CRIPS-DM, 2007]

En el nivel superior el proceso de minería de datos está organizado en fases; cada fase consiste en un número de tareas genéricas de segundo nivel. A este segundo nivel de lo llama genérico porque se pretende que sea lo suficientemente general para cubrir todas las posibles situaciones del proceso de minería de datos. A su vez, se espera que sea

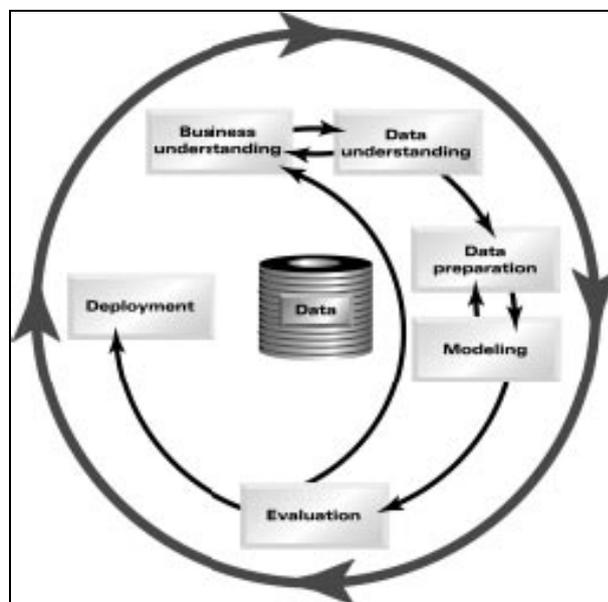
estable para que el modelo sea válido para nuevas técnicas de modelado aún no desarrolladas.

El tercer nivel (tareas específicas) es el lugar para describir como las acciones genéricas deben ser llevadas a cabo en ciertas situaciones específicas. Por ejemplo, si en el segundo nivel se tiene la tarea general “limpieza de datos”, en el tercer nivel se detallan las tareas que tienen que desarrollarse para un caso específico, como puede serlo la “limpieza de datos numéricos”, o “limpieza de datos categóricos”.

La descripción de fases y tareas, como pasos discretos llevados a cabo en un orden específico, representan una secuencia idealizada de eventos. En la práctica, muchas de las tareas pueden llevarse adelante en distinto orden y será necesario volver atrás varias veces para repetir ciertas acciones.

El cuarto nivel, instancias de proceso, es un registro de las acciones, decisiones y resultados de un proyecto de minería de datos específico. Está organizado de acuerdo a las tareas definidas en los niveles superiores, pero representan lo que sucede en instancias particulares durante el proceso de minería de datos.

El ciclo de vida del proyecto de minería de datos, según CRISP-DM, consta de 6 fases [Figura 2.5]. Es importante aclarar que la secuencia de las fases no es rígida, por lo que cuando sea requerido los proyectistas deberán moverse entre ellas.



**Figura 2.5:** CRISP-DM. Ciclo de vida del proyecto. [CRISP-DM, 2007]

Dependiendo del resultado de cada fase, se llevará a cabo la siguiente fase correspondiente. En la figura 2.5, las flechas representan las más importantes y

frecuentes dependencias entre fases. La circunferencia exterior simboliza la naturaleza cíclica de los proyectos propiamente dichos.

### **2.2.1 Objetivos de cada fase en la metodología CRISP-DM.**

Cada una de las fases de la presente metodología apunta a una serie de objetivos principales.

#### *Fase I - Comprensión del negocio*

Esta primera fase se focaliza en entender los objetivos del negocio y los requerimientos desde la perspectiva del negocio, para incluir estos conocimientos dentro de la definición del problema de Exploración de información y el diseño del plan preliminar para lograr los objetivos.

#### *Fase II - Comprensión de los datos*

La fase de comprensión de los datos comienza con la recolección inicial de datos y prosigue con actividades que apuntan a la familiarización de los datos, identificar problemas de calidad y detectar relaciones interesantes entre los mismos que permitan generar hipótesis sobre información oculta.

#### *Fase III - Preparación de los datos*

Esta fase contempla un conjunto de actividades destinadas a la construcción del *dataset* a partir de los datos iniciales. Esta fase implica múltiples tareas que pueden desarrollarse al mismo tiempo sin un orden estricto. Estas tareas incluyen la selección, limpieza y transformación de tablas, registros y atributos para poder ingresarlas en la herramienta de modelado.

#### *Fase IV - Modelado*

En esta fase se seleccionarán y aplicarán varias técnicas de modelado, como así también, opcionalmente, se podrán determinar los valores de los parámetros y variables de calibración. Para esta tarea generalmente se puede contar con más de una técnica que realice la misma función. Algunas técnicas pueden tener requerimientos específicos en cuanto a la conformación de los datos, lo cual puede hacer que se deba volver a la fase de preparación de los datos para realizar algún ajuste.

### *Fase V - Evaluación*

En esta fase se verifica que los resultados obtenidos en la fase de modelado sean de alta calidad desde la perspectiva del análisis de datos. Antes de realizar la implementación del modelo, es importante evaluar que los resultados del modelado y revisar los pasos realizados para la construcción del modelo y verificar que estos sean apropiados en función de los objetivos del negocio.

### *Fase VI - Implementación*

El final del proyecto, generalmente, no se encuentra en la creación del modelo. En la mayoría de los casos el analista de datos debe generar un informe final para ser presentado al cliente o en caso de que el propio cliente desee realizar el informe final se deberá transferir a éste el conocimiento para que pueda hacer una correcta interpretación de los datos.

## **2.3 MINERÍA DE DATOS APLICADA A LA INFORMACIÓN CRIMINAL**

A partir de los ataques terroristas del 11 de Septiembre de 2001 a las torres gemelas en la ciudad de Nueva York, EE.UU., la cantidad de información recolectada, reunida y compilada por las agencias locales, estatales y federales (CIA, FBI, etc.) de dicho país fue incrementándose significativamente, buscando de esta forma lograr prevenir futuros ataques. El manejo y capacidad de análisis de dicha información, mediante la creación de diversas herramientas, técnicas y sistemas, tampoco se ha quedado atrás [McCue, 2005].

Las mismas herramientas y técnicas que son usadas para determinar riesgo crediticio, descubrir fraudes e identificar hábitos de consumos en las personas, entre otros usos, pueden ser utilizadas para la lucha contra la delincuencia, terrorismo y la protección de la seguridad interior.

La minería de datos es muy aplicable en la información criminal, ya que se trata de una herramienta muy potente que le permite a los investigadores del crimen que poseen poco entrenamiento como analistas de datos explorar grandes bases de datos de manera rápida y eficiente [Fayyad & Uthurusamy, 2002]. Al ser las modalidades criminales cada vez más complejas y dinámicas, se cree que es sumamente útil la utilización de herramientas específicas para poder tener un mejor tratamiento de la información criminal. El uso de la minería de datos aplicado a la inteligencia criminal ha tenido un gran crecimiento en los últimos años en EE.UU. [Chen *et al.*, 2004].

Finalmente, con respecto a la aplicación de minería de datos en poblaciones penitenciarias, no se ha encontrado estudio alguno tanto en Argentina como a nivel internacional. Por lo tanto, se cree que esta es una muy buena oportunidad para aplicar minería de datos en información relativa a las cárceles de la Argentina.

A continuación se describen diversos proyectos de aplicación de la minería de datos en información criminal.

### **2.3.1 Proyecto Able Danger**

En el año 1999, los Comandos de Operaciones Especiales de EE.UU., buscaban la forma de acrecentar la lucha contra el terrorismo luego de los ataques que sufrieron sus embajadas en Kenya y Tanzania por parte de la red *Al Qaeda*.

*Able Danger* [Harris, 2005] fue el nombre en código del proyecto desarrollado por una de las agencias de inteligencia de los EE.UU., la IDC (*Information Dominance Center*). Mediante el mismo, supuestamente se habrían identificado al líder de los ataques del 11 de Septiembre de 2001, Mohamed Atta, junto a otros 3 terroristas como posibles miembros de una célula de *Al Qaeda* operando en ese país más de un año previo a los ataques.

El proyecto, basado en la minería de datos, contenía un gran volumen de información (2.5 terabytes), lo cual representaba aproximadamente al 12% de todas las páginas impresas guardadas en las bibliotecas del Congreso Norteamericano.

El resultado del *Able Danger* trajo a la luz diagramas y mapas que mostraban nombres de sospechosos, los relacionaban e interconectaban. Entre estos se encontraba Mohamed Atta, que lo vinculaba con un terrorista aún detenido en una prisión federal, quien a su vez había tenido participación en los ataques al *World Trade Center* en el año 1993. Los resultados fueron obtenidos sobre la base de búsquedas intensas de decenas de miles de páginas Web que contenían las palabras clave como por ejemplo "*Al Qaeda*" o "*Bin Laden*". Luego, una herramienta de visualización de datos, creaba un mapa tridimensional mostrando cuales palabras aparecían más frecuentemente y como se relacionaban. Nombres, ubicaciones, capacidades y hasta las fuentes de financiamiento de la red fue alguna de la información detectada.

En la figura 2.6 se puede observar un ejemplo de los resultados obtenidos mediante la Minería de Datos llevada a cabo durante el proyecto *Able Danger*.

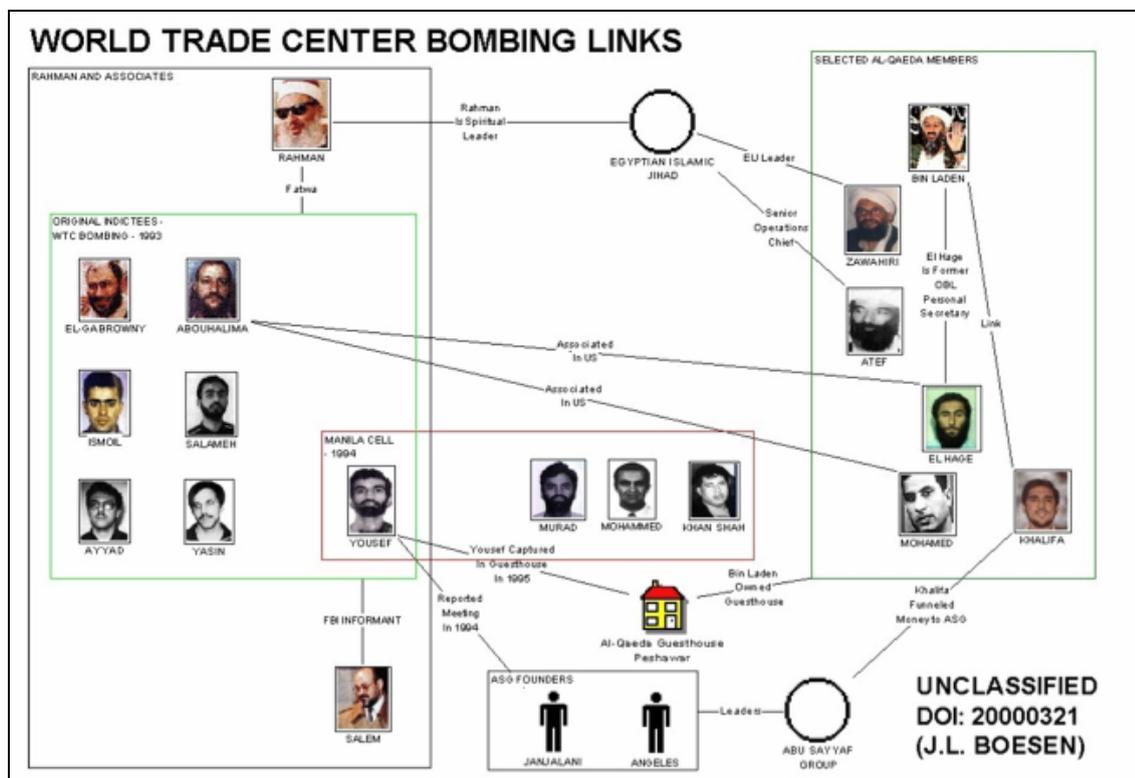


Figura 2.6. Mapas de relaciones del *Able Danger* [Lance, 2007].

Sin embargo, cuando se creía tener la herramienta que permitiría encontrar a los terroristas antes de que estos ejecuten sus actos, al mismo tiempo se estaba frente a una importante línea legal. La metodología básica de búsqueda sugería que todos los ciudadanos de los EE.UU. podrían aparecer en ella. Esta fue una de las principales causas por las cuales la IDC fue obligada a destruir cualquier información recolectada sobre ciudadanos estadounidenses dentro de los 90 días. Así fue, como en la primavera del año 2000, el proyecto quedó relegado y ningún rastro del mismo pudo ser rescatado. Nunca la CIA ni el FBI pudo tener en sus oficinas toda la información recolectada y procesada.

Para más información acerca de este proyecto remitirse al “*Able Danger Investigation Report*” [Departamento de Defensa EE.UU., 2006], en donde se encuentran publicados todos los cuadros de relaciones realizados.

### **2.3.2 Proyecto COPLINK – Universidad de Arizona**

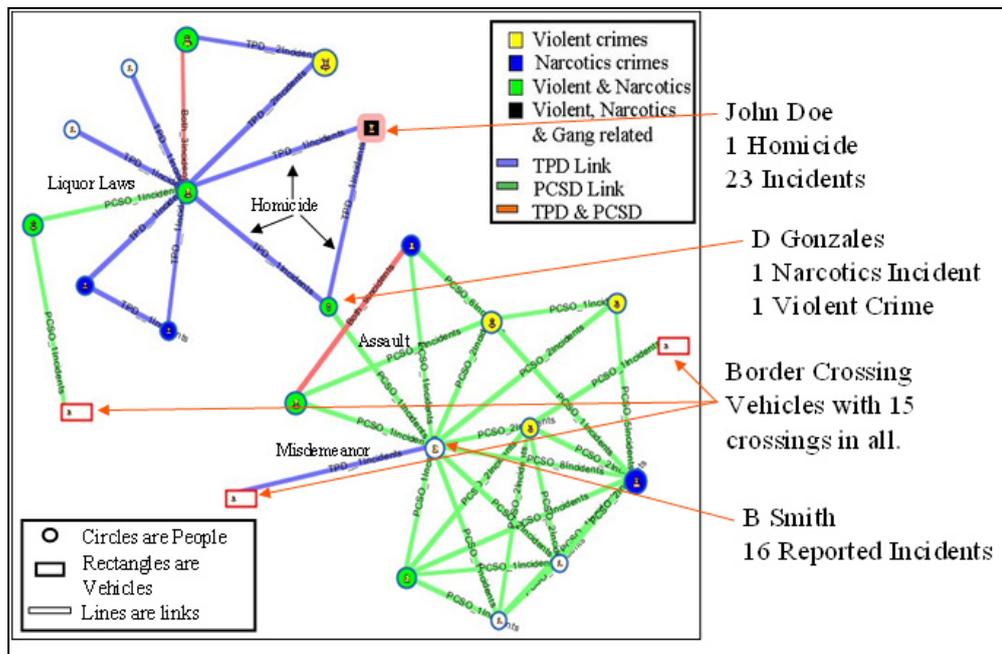
Este producto, creado en 1997 por el Laboratorio de Inteligencia Artificial de la Universidad de Arizona (Tucson, EE.UU.), se lanza con el objetivo de proveer soporte al análisis y toma de decisiones para la identificación de comportamientos criminales, relaciones y patrones, ayudando a resolver crímenes de distinta índole y detección de potenciales ataques terroristas [Artificial Intelligence Lab, 2007].

Este interesante desarrollo posee diversas aplicaciones, permitiendo trabajar con datos provenientes de diversas fuentes y analizando la información a través de una red segura y “amigable”.

Entre las aplicaciones más interesantes del *COPLINK* se pueden nombrar a las siguientes:

- **Seguridad Fronteriza (*COPLINK BorderSafe*)**

Este proyecto busca lograr que las distintas agencias generadores de información participen en una plataforma piloto para la investigación y análisis de información cruzada proveniente de distintas jurisdicciones. Registros del Departamento de Policía de Tucson, Departamento del condado de Pima y Patrullas fronterizas han sido integrados a esta plataforma para extraer conclusiones acerca de la actividad de redes de traficantes de drogas como también de “cruzadores” de fronteras (*border crossing vehicles*).



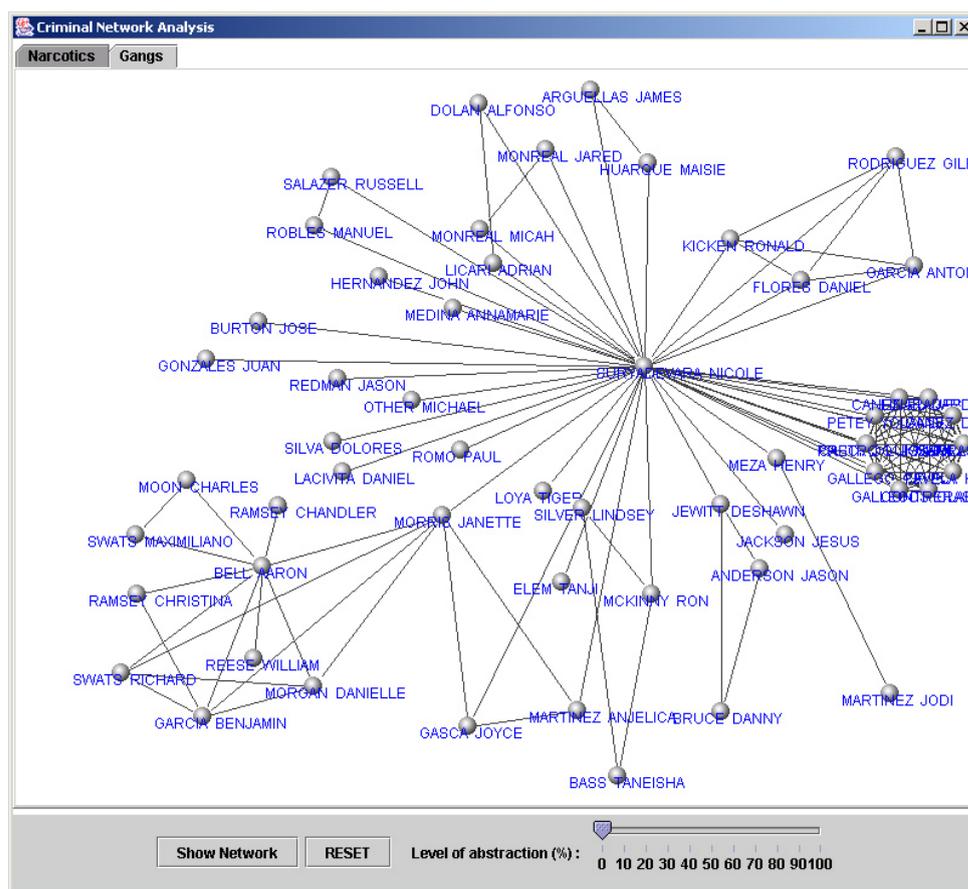
**Figura 2.7.** Red criminal obtenida con el *COPLINK BoarderSafe*.

La figura 2.7 es un típico ejemplo de actividad de una red criminal, integrando los registros de distintas agencias. En este caso, la red es usada para identificar vehículos asociados al sospechoso de un homicidio.

- **Análisis de Redes Criminales (COPLINK Criminal Network Analysis)**

Es una herramienta que pretende facilitar la investigación de crímenes como terrorismo, tráfico de drogas y crímenes cometidos por bandas autoorganizadas. Se busca determinar el comportamientos de bandas, lideres y ver como se relacionan y operan entre sí para llevar a cabo varios tipos de actividades ilegales [Chen et al., 2004].

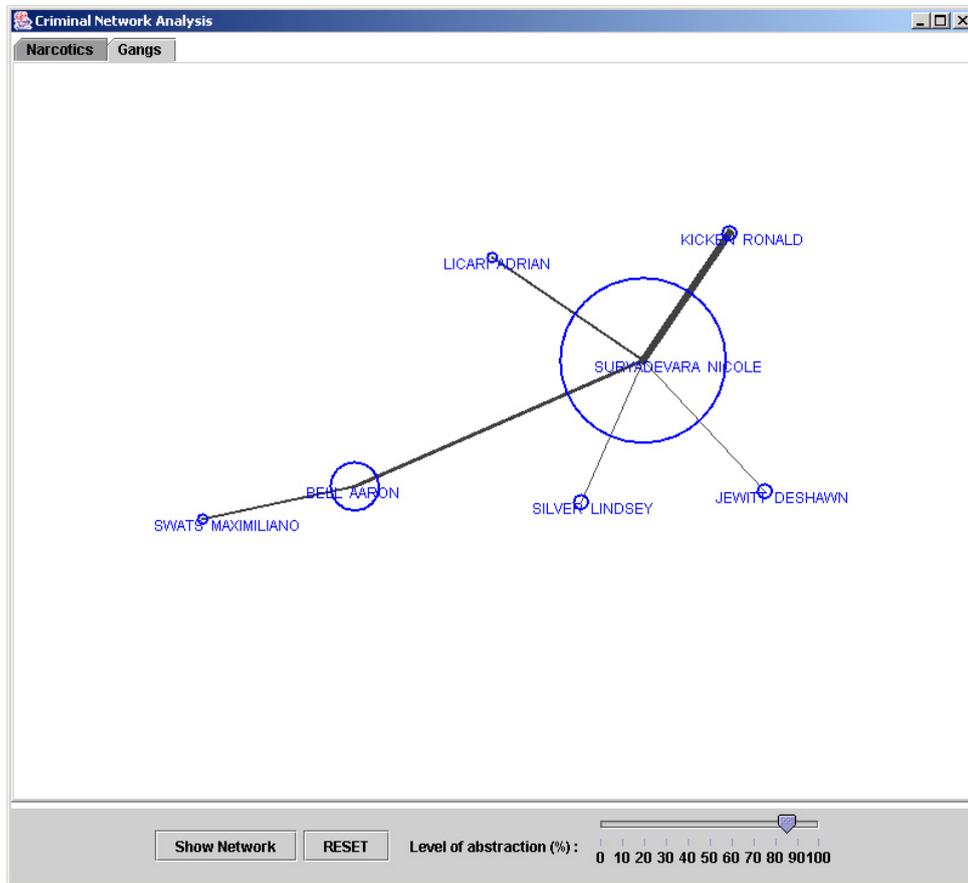
Se utilizan técnicas de *Concept Space* para crear una red de potenciales sospechosos. El vínculo entre dos sospechosos está dado por la cantidad de veces en que ambos fueron identificados en el mismo incidente. Se dividen en subgrupos mediante el *clustering* jerárquico y *block modeling* para detectar patrones de interacción entre los subgrupos. También se calculan distancias centrales (grados, intersecciones y cercanías) para detectar a los miembros clave de cada subgrupo, como lo son los líderes.



**Figura 2.8.** Red criminal obtenida con el COPLINK Criminal Network Analysis.

La figura 2.8 muestra a los nodos representados por los criminales con sus respectivos nombres. Las líneas conectores representan relaciones entre los criminales.

Se puede hacer foco en las redes individuales aumentando el porcentaje de abstracción (*level of abstraction %*) obteniéndose la figura 2.9. En ella se pueden observar círculos que representan grupos. El tamaño de los círculos es proporcional al número de miembros de los distintos grupos. Cada grupo está etiquetado con el nombre del líder del grupo.



**Figura 2.9.** El líder, su grupo y red. COPLINK Criminal Network Analysis.

Existen otras aplicaciones dentro del *COPLINK* como la *Extracción de Entidades* automáticamente desde los reportes policiales y la *Detección de Identidades Múltiples* en grandes bases de datos.

## **2.4 INFORMACIÓN CRIMINAL EN ARGENTINA**

A continuación se presenta, a modo de referencia, una descripción del estado actual de la información criminal en Argentina.

### **2.4.1 Dirección Nacional de Política Criminal**

A partir de la sanción de la Ley 25.266 se le otorgó a la Dirección Nacional de Política Criminal (DNPC), creada en el año 1991 y dependiente del Ministerio de Justicia y Derechos Humanos, la potestad no sólo de diseñar y producir las estadísticas de criminalidad, sino también las estadísticas sobre el sistema penal en la Argentina. En este sentido, esta nueva legislación procura un principio de organización para la producción de las fuentes estadísticas de conocimiento sobre la cuestión criminal y abre una oportunidad para generar fuertes avances en la materia.

El análisis de las estadísticas sobre criminalidad puede ser dividido en dos grandes áreas: las estadísticas oficiales, que toman como fuente el registro de agencias estatales (Policía, Poder Judicial, Penitenciarias) y los estudios de victimización, que se basan en encuestas poblacionales.

La DNPC cuenta con cinco sistemas de información que se utilizan como fuente para la generación de estadísticas oficiales:

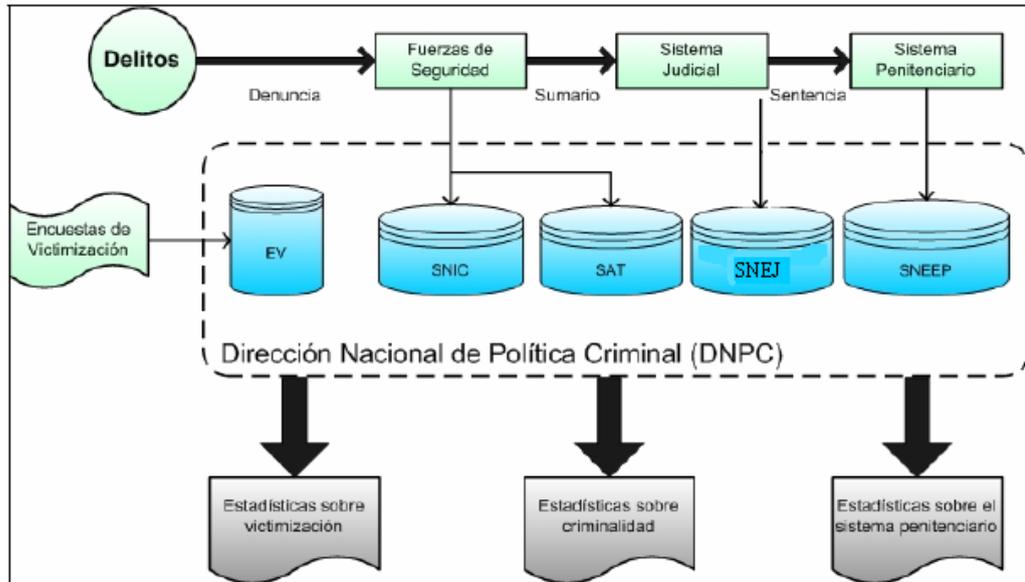
- Sistema Nacional de Información Criminal (SNIC).
- Sistema de Alerta Temprano (SAT).
- Sistema Nacional de Estadísticas Judiciales (SNEJ).
- Encuestas de Victimización.
- Sistema Nacional de Estadísticas sobre Ejecución de la Pena (SNEEP).

#### **2.4.1.1 Fuentes Oficiales de Información y Estudios de Victimización.**

Las fuentes oficiales de información criminal penal son, como se dijo anteriormente, los sistemas Policial, Judicial y al final de la “cadena”, el Penitenciario.

Los delitos ingresan, en forma de denuncias, a través del Sistema Policial. Luego pasan por el Sistema Judicial, en donde se observa cuantitativamente la imagen de un embudo, cuya boca es la institución policial, y se va angostando a medida en que se van atravesando las diversas etapas del proceso penal. En esta instancia no se presenta ningún tipo de información sobre las víctimas de los delitos, a diferencia de las

estadísticas policiales. Sobre los potenciales ofensores, sólo se presenta la información referida a si se encuentran detenidos o no durante en el proceso penal. Finalmente se tiene al Sistema de Información Penitenciaria, donde la calidad de dicha información es la que proviene del sistema judicial [Figura 2.10].



**Figura 2.10.** Flujo de información criminal. [DNPC, 2007]

Tal como el título del proyecto lo indica, se trabajará con la información proveniente del sistema penitenciario únicamente, dejando de lado la información proveniente del SNIC (Sistema Nacional de Información Criminal) como también del SNEJ (Sistema Nacional de Estadísticas Judiciales). Se cree conveniente, igualmente, describir brevemente a estos dos últimos sistemas en carácter informativo como así también los métodos no oficiales de información denominados estudios de victimización.

#### **2.4.1.1.1 Sistema Nacional de Información Criminal (SNIC)**

En el año 1999 la Dirección Nacional de Política Criminal implementó el *Sistema Nacional de Información Criminal* como parte de un programa de recolección, producción y análisis de informaciones estadísticas sobre el delito en la Argentina pretendiendo asumir la compleja tarea de medir el delito. Este sistema recopila la información registrada por las fuerzas de seguridad (policías provinciales, policía federal, gendarmería y prefectura naval) sobre la cantidad de sumarios iniciados mensualmente por hechos delictuosos.

Las policías y fuerzas de seguridad en la Argentina pueden tomar conocimiento de la realización de una conducta que presuntamente coincide con la tipificada como delito por parte de la ley penal por dos mecanismos:

- Como consecuencia de la actividad de prevención del delito (la vigilancia en sus diversas formas).
- Como consecuencia de la denuncia realizada por un ciudadano que da lugar a la actividad de investigación del delito.

El segundo mecanismo es cuantitativamente más importante que el primero.

En suma, los reportes que se realizan utilizando como fuente los hechos presuntamente delictuosos registrados por las policías y demás fuerzas de seguridad, no incluyen los delitos no denunciados ni los que ingresan directamente por algún tribunal judicial (Juzgado, Fiscalía, Cámara). Esto quiere decir, que existe una amplia gama de delitos que, al no ser denunciados por sus víctimas ni captados por la policía, no ingresan al circuito judicial y, por lo tanto, **no son registrados en la estadística oficial**. Estos hechos abarcan principalmente a gran parte de robos, hurtos, lesiones y delitos sexuales, que por diferentes razones no son denunciados (porque la víctima no quiere perder el tiempo, no cree que se pueda aclarar el hecho, etc.). Asimismo también hay una elevada cantidad de delitos que son muy difíciles de captar por el sistema de justicia por deficiencias propias de los organismos de control (como el caso de los denominados “delitos complejos”). Por tal motivo las variaciones que se producen en las cifras que reflejan estos reportes se limitan a las variaciones detectadas por los propios organismos que la registran sin que esto, necesariamente, implique un cambio real en la ocurrencia de delitos (**Cifras Negras del Delito**).

Los registros mensuales se realizan en base a una planilla diseñada a tal fin, que las policías y fuerzas de seguridad envían a la DNPC [Anexo 1]. La tipificación de los hechos presuntamente delictuosos en esta planilla, continuando con una tradición de relevamiento llevado a cabo por el Registro Nacional de Reincidencia y Estadística (antiguo consolidador de estadísticas a nivel Nacional), se encuentra fuertemente asociada a las figuras y tipos penales descritos en el Código Penal. Los datos disponibles han sido agrupados siguiendo en parte la tipificación del Código Penal (la excepción ha sido los Delitos contra el Estado y la Comunidad) de modo tal que puedan a simple vista identificarse grupos según los bienes jurídicos afectados:

#### **1<sup>er</sup> Grupo - Delitos contra las personas:**

Todos aquellos tipos de delitos que con un grado mayor o menor de violencia afectan la vida de las personas.

#### **2<sup>do</sup> Grupo - Delitos contra la honestidad y el honor:**

Las denuncias por violación han sido individualizadas para poder dar cuenta de este dato de manera desagregada.

**3<sup>er</sup> Grupo - *Delitos contra la propiedad*:**

Reúne todos aquellos hechos que afectan este bien jurídico, se trate de robos, hurtos (incluyendo las tentativas en ambos casos) así como todos los otros delitos que han sido agrupados en el Código Penal bajo este ítem.

**4<sup>to</sup> Grupo - *Delitos contra el Estado y la Comunidad*:**

Incluye todos aquellos delitos que afectan al Estado en mayor o menor medida.

**Otros Grupos:**

Los otros grupos son Delitos contra la libertad, Delitos contra el estado civil, Delitos previstos en leyes especiales, con la excepción de la ley 23.737 de *Estupefacientes* que, por tratarse de aquella que dentro de las leyes especiales concentra la mayor frecuencia de hechos, ha sido considerada aparte (en el interior del país sólo están informados los casos en los que participaron las policías provinciales, restando los datos de las delegaciones en el interior de la Policía Federal Argentina).

Adicional al SNIC, existe el Sistema de Alerta Temprana (SAT). Este se nutre de cuatro planillas complementarias a la del SNIC que relevan información detallada sobre cuatro aspectos en particular:

- homicidios dolosos.
- homicidios culposos en accidentes de tránsito.
- Suicidios.
- delitos contra la propiedad.

En los primeros tres casos se releva información puntual de cada hecho, mientras que el último consta de sumalizaciones parciales según distintas variables [García Martínez *et al*, 2007].

**2.4.1.1.2 Sistema Nacional de Estadísticas Judiciales (SNEJ)**

Las estadísticas judiciales sobre la criminalidad son el conjunto de las informaciones cuantificadas sobre los comportamientos efectivamente producidos en la vida social, presuntamente delictuosos, que nacen de la actividad de definición y registro por parte de las instituciones judiciales penales, en el marco del desarrollo de los procesos penales [Sozzo, 2000]. En la Argentina, existe una administración federal de justicia penal con jurisdicción sobre todo el territorio nacional y competencia sobre determinados tipos de delitos, una administración “nacional” de justicia penal cuya jurisdicción es la Ciudad

de Buenos Aires y veinticuatro provinciales administraciones de justicia penal. Esto hace que cada una de dichas instituciones estatales produzca su propia información sobre los comportamientos presuntamente delictuosos procesados.

Existe, entonces, una pulverización de esta fuente estadística de conocimiento sobre la cuestión criminal en torno a, por un lado, el poder judicial federal y por el otro, los diversos poderes judiciales provinciales. Esto se traduce en una diversidad de criterios para la recolección y producción de información estadística en manos de oficinas administrativas de los diversos poderes judiciales, dependientes de las distintas cortes supremas o tribunales supremos [Sozzo, 2000].

Por otro lado, a diferencia de lo que sucede con las estadísticas policiales, aún no se ha avanzado demasiado en la creación de un sistema centralizado de estadísticas judiciales para todo el país. Actualmente en la DNPC se está elaborando un diseño para la creación y funcionamiento de un Sistema Nacional de Estadísticas Judiciales automatizado [Sozzo, 2000].

#### **2.4.1.1.3 Encuestas de Victimización**

A fin de evaluar en forma más certera el impacto que tienen los datos que registra la policía (SNIC), es importante el aporte que han brindado los estudios de victimización, especialmente a los fines de percibir el nivel de denuncias de los delitos. Si bien estos estudios tienen objetivos y metodologías distintas a las estadísticas policiales, pueden ser de suma utilidad para complementar los resultados obtenidos por estas últimas.

Las encuestas de victimización nos permiten observar que es muy alta la cantidad de delitos que no capta el sistema de justicia y, por ende, el valor de la estadística oficial se encuentra condicionado. Estas limitaciones no implican que las estadísticas policiales no sean de gran valor sino más bien que su interpretación debe realizarse teniendo en cuenta los condicionamientos aludidos. En otras palabras, la información que surge de las estadísticas oficiales presentan fuertes limitaciones que deben ser tenidas en cuenta para no tergiversar su interpretación, pero esto no le resta importancia ni utilidad.

Estos estudios tienen muchos beneficios, entre los que se destaca que, en muchos casos, son más representativas que las estadísticas oficiales ya que incluyen aquellos delitos que no ingresaron al sistema de justicia. En efecto, una de las cuestiones que abordan las encuestas de victimización son la cantidad de delitos que la gente dijo haber sufrido en el período de estudio y su nivel de denuncia. Es decir que, sobre la base de estos estudios, se puede conocer cual es el porcentaje de delitos que realmente son denunciados (el denominado fenómeno de la “**cifras negras del delito**”). Por otro lado, otra de las características destacadas de las encuestas es que permiten acceder al

conocimiento del “problema criminal” desde el punto de vista de las propias **víctimas** y no de las **agencias estatales**.

De todas maneras, los estudios de victimización también tienen sus limitaciones, en especial en cuanto a su alcance, tanto territorial como a las cuestiones que abarca. En ese sentido, las encuestas se realizan básicamente en grandes centros urbanos y los delitos que más óptimamente capta son los delitos contra la propiedad (en especial robos y hurtos). En relación a otros delitos (homicidios, delitos sexuales, violencia familiar, delitos complejos), no resulta un instrumento muy efectivo ya que son conductas difíciles de abordar en entrevistas de estas características [Olaeta, 2006].

#### **2.4.1.1.4 Sistema Nacional de Estadísticas sobre Ejecución de la Pena (SNEEP)**

El Sistema Nacional de Estadísticas sobre Ejecución de la Pena (SNEEP) es un aporte vital para el desarrollo de un diagnóstico válido sobre el estado actual de la cuestión carcelaria a fines de aplicar una política criminal eficaz.

El Sistema Nacional de Estadísticas sobre Ejecución de la Pena (SNEEP) fue implementado en el año 2002 por la Dirección Nacional de Política Criminal con el fin de contar con información periódica y uniforme acerca de la población penal privada de libertad en la República Argentina.

En nuestro país existen instituciones penitenciarias a nivel federal, el Servicio Penitenciario Federal, con unidades ubicadas en todo el país, y a nivel provincial, veintitrés provincias poseen instituciones penitenciarias propias organizadas generalmente bajo la forma de un servicio penitenciario provincial o esquema análogo. Es decir, que se trata de una pluralidad de instituciones ubicadas bajo competencias jurídicas y políticas diferentes [DNPC, 2003]. Además, en nuestro país se ha constatado en los últimos años una creciente cantidad de casos de personas privadas de libertad alojadas en lugares no aptos a tal fin tal, como comisarías u otros destacamentos policiales. Esta compleja organización dificultó el diseño de un sistema válido de recopilación periódica de datos comparables para todo el país, que tuviera las mismas unidades de análisis y cuente con el rigor científico necesario.

En el campo de las estadísticas sobre el sistema penal, como se observó en el punto 2.2.1.1, el SNEEP abarca uno de los tres ejes básicos de información, el correspondiente a la agencia penitenciaria; mientras que las otras dos agencias, policial y judicial, son abordadas por el Sistema Nacional de Información Criminal (SNIC) y el Sistema Nacional de Estadísticas Judiciales (SNEJ), respectivamente [Figura 2.1]. En tal sentido, los resultados que nacen de la aplicación de estos sistemas de información permiten

trabajar en un diagnóstico acerca de la criminalidad desde la particular mirada de las agencias productoras de datos. Se cree que el desarrollo del SNEEP permite afrontar de modo sustentable el diseño de políticas en materia penitenciaria tanto a nivel federal como provincial.

#### **2.4.1.1.4.1 Metodología del Sistema Nacional de Estadísticas sobre Ejecución de la Pena**

El SNEEP tiene como destinatarios a todas las unidades de detención penal de país, dependientes del sistema federal y provincial. Es decir no se incluye en esta etapa a las personas detenidas en comisarías y a los niños y jóvenes privados de libertad que no se encuentran en una unidad de ejecución penal. Cabe aclarar que si bien las dependencias policiales no deben considerarse plazas carcelarias, en tanto no reúnen las condiciones requeridas para los lugares de cumplimiento de pena y/o prisión preventiva [CELS, 2005] constituyen una realidad insoslayable a la hora de analizar la situación del sistema penitenciario en nuestro país.

El relevamiento de la información se realiza en forma anual a través de dos cuestionarios que se completan en cada unidad de detención. El diseño de los instrumentos de recolección de datos se llevó a cabo teniendo como referencia la ley 24.660.

El primero de los instrumentos consiste en una serie de cuadros básicos con información agrupada, tomando como unidad de análisis al **establecimiento**, en base a las personas alojadas en el establecimiento en cuestión durante el último año. Los datos requeridos son los siguientes [Anexo 2]:

- Cantidad total de detenidos
- Situación legal
- Egresos
- Cantidad de visitas
- Cantidad de alteraciones al orden
- Cantidad de suicidios
- Cantidad de fallecidos
- Cantidad de fugas
- Cantidad de evasiones

El segundo instrumento consiste en un censo sobre el total de la población detenida al día 31 de diciembre de cada año en el establecimiento. Aquí la unidad de análisis son

las **personas** detenidas a esa fecha. En el anexo 3 se presenta la información que se recava sobre cada interno mediante el censo de la población.

Cabe destacar que la planilla se llena mediante una codificación de las variables, por lo que es importante aclarar que el desarrollo de las estadísticas criminales en las unidades de detención tiene sus limitaciones.

La falta de recursos y de personal capacitado, conjuntamente con una poca tradición en el manejo y producción de información por parte de las instituciones penitenciarias, complica la tarea de recolección de datos. Esta deficiencia trae como consecuencia la falta de datos o la existencia de ciertas inconsistencias en la información remitida en algunos casos particulares. En este sentido, la Dirección Nacional de Política Criminal tendrá que reforzar los esfuerzos en lo referente a las tareas de capacitación del personal y cada organismo provincial y federal competente en esta materia tendrá que apoyar el esfuerzo que realizan los encargados de confeccionar las estadísticas en cada instituto.

La información remitida presenta limitaciones propias de la fuente de origen. Se trata de un relevamiento en el que la visión de los servicios penitenciarios o policías locales acerca de las personas detenidas resulta fundamental. En tal sentido, la principal fuente de información es el legajo de cada interno por lo que existen datos que no pueden ser profundizados y otros que se encuentran sesgados por la propia fuente. De todas maneras, se cree que tener presente las limitaciones referidas no le resta importancia al censo en tanto se interprete con las debidas precauciones metodológicas y teóricas [Dirección Nacional de Política Criminal, 2005].

Finalmente, la oficina central de cada servicio penitenciario, o policía local, es la encargada de recopilar las planillas de todos sus establecimientos y remitirlas a la Dirección Nacional de Política Criminal en donde se vuelcan los datos en una base de datos diseñada para tal fin. Con los principales resultados obtenidos del análisis de la información se realizan informes que son remitidos a las autoridades nacionales y provinciales competentes y se publican en la página Web del Ministerio de Justicia y Derechos Humanos.

Como se comenta en el punto 2.4.1.1 el trabajo se basará en la caracterización de la población ubicada en los establecimientos penitenciarios de la Argentina. Los datos a utilizar se analizarán en la sección **4.2.2. Compresión de los Datos**.

#### **2.4.1.1.5 Sistema Unificado de Registros Criminales (SURC)**

En el año 2003 se crea el SURC a cargo del Ministerio de Justicia, Seguridad y Derechos Humanos, con el fin de integrar y unificar en una sola base de datos el manejo de información criminal y lograr servir al Mapa del Delito. Era sumamente necesario para ese entonces lograr interconectar las bases de datos criminales de las fuerzas policiales, de seguridad y fiscalías del país, mediante una red que permitiría realizar una consulta *on line* a un banco de datos alimentado en tiempo real. El Mapa del Delito iba a estar disponible en Internet para que pudiera ser consultado por todos los ciudadanos.

La iniciativa de la Secretaría de Seguridad Interior pretendía, en una primera etapa, interconectar la Capital Federal; en la segunda, las fiscalías de la zona norte del conurbano; luego, extender la experiencia al primer y segundo cordón del conurbano y al resto del país. Y la última etapa del proyecto consistía en interconectar el banco de datos con diversas bases del MERCOSUR para la persecución de delitos transnacionales como el terrorismo, el narcotráfico y el robo y contrabando de vehículos importados.

Los objetivos de mejora del SURC consistían en [Valenga, 2007]:

- Interconectar y sincronizar los sistemas judiciales, del Ministerio Público, las policías y Fuerzas de Seguridad para llevar a cabo investigaciones a nivel barrial y urbano.
- Fortalecer el esfuerzo nacional de Policía, el Sistema de Seguridad Interior, la Inteligencia Criminal con su vinculación al Sistema Penal, y de todos ellos, con los países del MERCOSUR.
- Hacer uso de mapas del delito basados en Sistemas de Información Geográfica (SIG) como herramienta para automatizar tareas de recopilación de datos, el acceso y análisis de los mismos en tiempo real.
- Agilizar los procedimientos, facilitando el acceso a la información por parte de todos los actores involucrados en el sistema.
- Estandarizar la información sobre delitos y su uso para las estrategias de investigación.
- Sistematizar los procesos administrativos de los expedientes y de recopilación de información sobre hechos criminales en una fiscalía.
- Sistematizar la relación entre los hechos, los imputados (individuales, grupos o bandas delictivas), con los denunciantes o víctimas.

El sistema permitiría, a partir de los resultados proporcionados por las búsquedas de hechos con características similares, individualizar y responsabilizar a personas imputadas por otros hechos en las que pudieren haber participado.

La base de datos del SURC almacenaba todo tipo de información diversa relacionada a hechos delictivos [Perversi ,2007; Valenga, 2007]:

- **Registro de Hechos:** características principales del hecho (lugar, delito denunciado, comisaría interviniente).
- **Registro de denunciantes:** toda aquella información que permite identificar a la víctima o denunciante.
- **Registro de autores identificados:** identidad, características, historial criminal e imágenes de los autores. Aquí se planificaba guardar lo necesario para la identificación del autor, desarrollando el primer banco de imágenes de Autores Individualizados.
- **Registro de autos No identificados:** descripción fisonómica de los NN tal como la contextura física, edad aproximada, estatura aproximada, tonada, señas particulares, tipo de rostro, color del pelo, frases o comentarios.
- **Registro de Elementos Sustraídos:** dispone de la información de aquellos elementos que fueron sustraídos en el hecho delictivo.
- **Registro de autos robados:** registra toda la información relativa a vehículos robados, como el número de patente, motor, chasis, color, señas particulares del automotor, marca, modelo, etc.
- **Registro de armas secuestradas:** registros necesarios para identificar armas vinculando esta base de datos con otros sistemas.
- **Registro de evidencias:** descripción de todo aquello considerado huella o pista en la escena del crimen.
- **Mapas del delito:** aportando la información desde la perspectiva geográfica permitiendo generar Mapas del Delito. Estos mapas permitirán utilizar y administrar más eficientemente los recursos de las fuerzas policiales y las fiscalías en la lucha contra el delito

Este proyecto, tal como se dijo anteriormente, comprendía varias etapas progresivas. Sin embargo, con la salida en el año 2004 del entonces Ministro de Justicia, Seguridad y Derechos Humanos, Gustavo Béliz, el proyecto se congela, sin conocimiento alguno del paradero actual del mismo.

#### **2.4.1.1.6 Mapa del delito de la Ciudad Autónoma de Buenos Aires**

El Mapa del delito de la Ciudad Autónoma de Buenos Aires (CABA) [Behar & Lucilli, 2003] surge a partir de un convenio entre el Centro de Información Metropolitana de la Facultad de Arquitectura, Diseño y Urbanismo de la UBA y el Ministerio Público Fiscal de la Nación. Este proyecto nace con el objetivo de generar un procedimiento metodológico para la detección espacial de los delitos buscando reducir el incremento de los mismos que se venía dando en la CABA en el año 2003.

La fuente sobre la cual se fundamenta el Mapa del delito es una base única de hechos delictivos generada por el Ministerio Público Fiscal de la Nación (MPF). Dicha base consta de más de 200,000 registros anuales, alimentada en forma permanente y sistemática. Cada registro contiene la información acerca del lugar geográfico en que se produjo el hecho, fecha, tipo de delito tipificado por el Código Penal, comisarías y fiscalías intervinientes, número de actuación, cantidad de víctimas, etc.

Por otro lado se cuenta con el Sistema de Información Territorial del Área Metropolitana de Buenos Aires (SIT/AMBA), generada en el Centro de Información Metropolitana (CIM) a partir de la tecnología de Sistemas de Información Geográfica (SIG). Esta tecnología sirve de apoyo al Sistema de Información Territorial del AMBA. La unión de estos dos componentes, la base alfanumérica de delitos (MPF) y la base cartográfica digital (CIM), permiten generar diferentes salidas gráficas para visualizar la problemática delictiva, descubrir las áreas de concentración, sus tendencias, su estacionalidad y poder vincularlas con las variables territoriales (urbanísticas, socioeconómicas, calidad de vida).

El concepto en el que se basan los Mapas del delito es el de “áreas calientes”. Estas últimas son concentraciones de incidentes dentro de un área geográfica limitada que aparece en un lapso de tiempo.

En la figura 2.11 se muestran las distintas pantallas de la base cartográfica digital CIM.

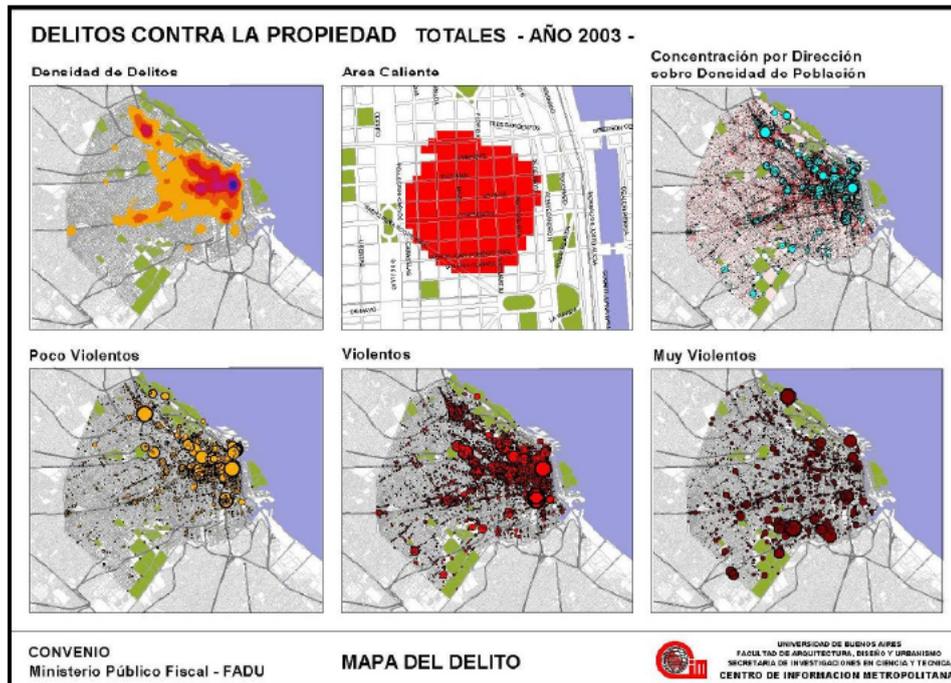


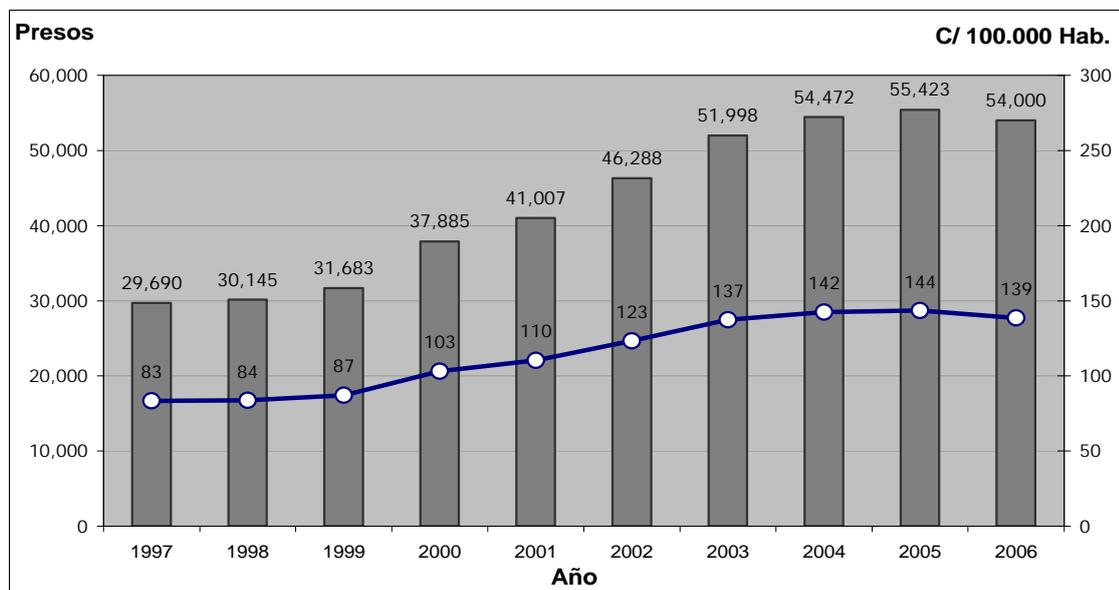
Figura 2.11. Mapa de los Delitos contra la Propiedad [Behar & Lucilli, 2003].

## 2.5 DIAGNÓSTICO DEL SERVICIO PENITENCIARIO EN ARGENTINA

Esta parte del estudio tiene por objetivo analizar brevemente el funcionamiento actual del servicio penitenciario en su conjunto, detallando indicadores y aspectos claves que se deben tener en cuenta a la hora de caracterizar a la población carcelaria en Argentina.

### 2.5.1 Evolución de la Población Penal en Argentina

La situación penitenciaria representa desde hace ya varios años un tema de especial preocupación en la Argentina. El crecimiento en las tasas de delitos registrado a lo largo de la década del noventa fue acompañado por un fuerte aumento en la población privada de libertad. Entre los años 1997 y 2006 dicha población aumentó alrededor del 83%, afectado profundamente por la crisis político social vivida entre los años 2001 y 2002. Como se puede observar en la figura 2.12, la cantidad de presos en el año 2005 llegaba a un pico 55.423 personas, con un índice de reclusos cada 100.000 habitantes. Estos incrementos han instalado el tema como uno de los puntos inevitables de estudio dentro de la política criminal [Dirección Nacional de Política Criminal, 2005].

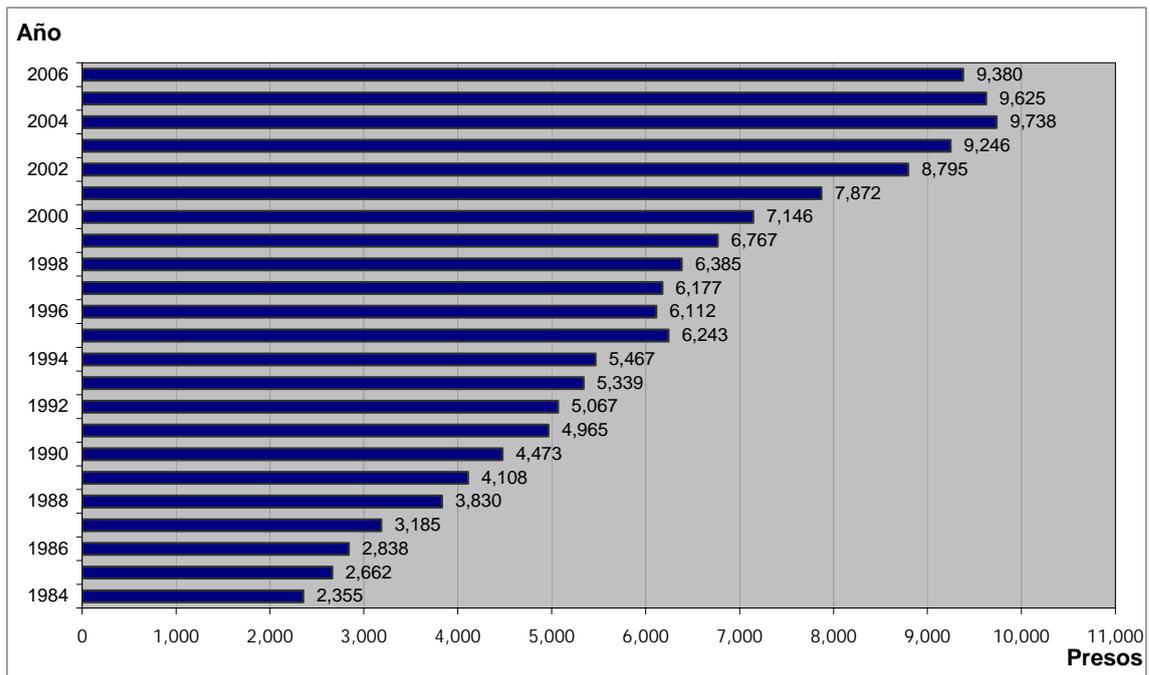


**Figura 2.12.** Evolución de la Población Carcelaria en Argentina. [DNPC, 2006]

Vale destacar, que la estadística oficial sobre población penitenciaria (SNEEP) comenzó en el año 2002. De los años anteriores no se cuenta con datos de algunas provincias.

Por otro lado, en nuestro país existen dos tipos de instituciones penitenciarias según como estén organizadas y bajo que competencias jurídicas se encuentran. Las primeras a nivel federal, Servicio Penitenciario Federal, que poseen unidades ubicadas en varias provincias del país. El segundo tipo es a nivel provincial, y son veintitrés provincias que poseen instituciones penitenciarias propias, organizadas generalmente bajo la forma de un servicio penitenciario provincial. A estos dos tipos de instituciones habría que sumarle las comisarías u otros destacamentos policiales, ya que en los últimos años se ha constatado que hay una gran cantidad de casos de personas privadas de libertad alojadas en este tipo de establecimiento no aptos para tal fin. Para fines del 2005 existía alrededor de 8000 personas detenidas en comisarías alrededor de todo el país [Anexo 4]. Se tiene, por lo tanto, una gran cantidad de instituciones ubicadas bajo competencias jurídicas y políticas diferentes.

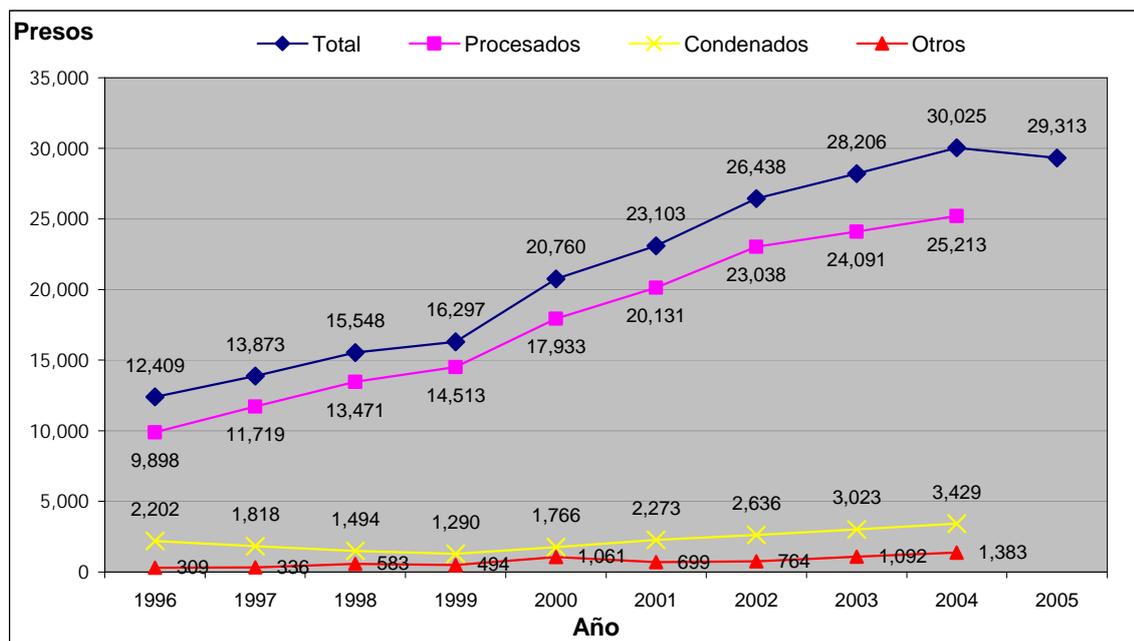
A nivel federal, la población del Servicio Penitenciario Federal (SPF) fue creciendo sostenidamente a través de los años (promedio del 8% anual) [Figura 2.13], hasta llegar a fines de 2006 con un crecimiento acumulado del casi 300% desde el año 1984. Nuevamente, es destacable el crecimiento que se dio entre los años 2001 (10%) y 2002 (12%).



**Figura 2.13.** Evolución de la Población Carcelaria del SPF. [DNPC, 2006]

Un estudio realizado por el Ministerio de Justicia en base a técnicas estadísticas, proyecta que el número de detenidos en establecimientos federales llegará a 15.000 en el 2015, y se incrementaría a unos 22.000 para el año 2025 [Prieto, 2005].

A nivel provincial se cuenta con información que indica que en la **Provincia de Buenos Aires**, desde el año 1996 hasta el 2005, la población carcelaria creció un 136% [Figura 2.14], mientras que la población general lo hizo en un 10%. En el Anexo 5 se puede observar la evolución de esta población penal en cada una de las provincias. Este incremento se produjo tanto en los establecimientos penitenciarios como en las comisarías bonaerenses, siendo estos últimos, lugares no aptos para el cumplimiento de penas.



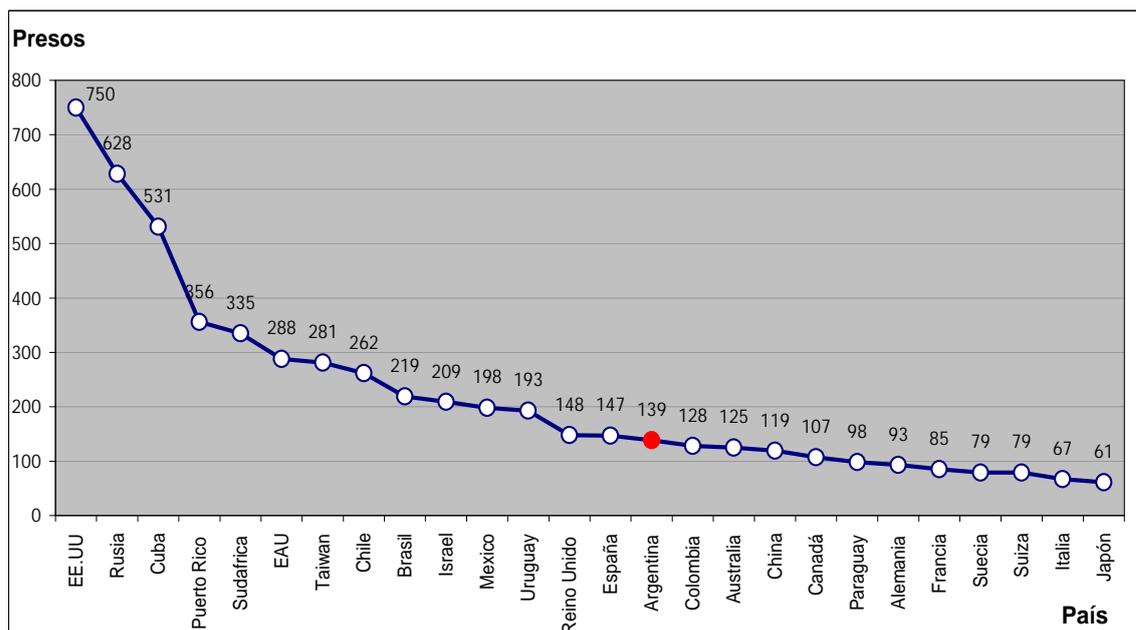
**Figura 2.14.** Evolución de la Población Carcelaria del Servicio Penitenciario Bonaerense

Como se puede observar en la figura 2.14, a principio de esta década las cárceles y comisarías bonaerenses alojaban 20.760 presos (16.567 en unidades y 5748 en comisarías), y la tasa de encarcelamiento era de 148 cada cien mil habitantes bonaerenses. Mientras que cinco años después, los detenidos llegaron a 29.313, con una tasa de encarcelados de aproximadamente 200 cada cien mil habitantes. Se observa, por lo tanto, que las cárceles bonaerenses alojan claramente a más del 50% de las personas detenidas del país.

En lo que respecta a la proporción procesados/condenados, se analizará el tema cuando se hable acerca de la situación penal de los detenidos en el apartado 2.5.4.

### **2.5.2 Situación a nivel internacional**

Durante el año 2002 existían en Argentina 123 internos en cárceles por cada 100,000 habitantes, incrementándose esta cifra en el año 2006 a 139 [Figura 2.15]. Si se tomase en cuenta a las personas detenidas en comisarías o destacamentos policiales, la tasa se incrementaría a 159. Esta tasa es muy similar a países europeos como España (147) y el Reino Unido (148), aunque menor a la de países sudamericanos como Uruguay (193), Brasil (219) y Chile (262). En cuanto a los países que presentan mayor cantidad de presos por cada 100.000 habitantes, se tiene a Estados Unidos, con una población de aproximadamente 2.300.000 internos (tasa = 750 presos c/ 100.000 hab.) y a Rusia con 628 reclusos por 100.000 habitantes. Se puede encontrar más detalles respecto a la población penal mundial en el anexo 6.



**Figura 2.15.** Población Carcelaria en el mundo. [ICPS, 2007]

La alta tasa de presos que registra Estados Unidos [Figura 2.15] se puede explicar de la siguiente manera. Las políticas de los 70 y 80 estuvieron impulsadas por el miedo al delito y por el hecho de que éste figuraba como tema de importancia central en todas las encuestas nacionales [Morris & Rosales, 2007]. Hoy en día, menos del cinco por ciento de la población norteamericana considera al crimen como el mayor problema que enfrenta Estados Unidos. La explicación es muy simple. Estados Unidos duplicó la población carcelaria. De casi un millón de presos que había a comienzos de los 80, hoy pasaron a ser dos millones. Con el doble de delincuentes en prisión, el crimen se redujo a la mitad. En el año 1971, el número de presos era aún mucho menor, 200.000 personas.

Claro que para lograr la reducción de hechos delictuosos en EE.UU., se impusieron sentencias obligatorias a los criminales reincidentes. Aquellos que cometieran tres delitos fueron automáticamente sentenciados a prisión perpetua sin ninguna posibilidad de excarcelación. A su vez, el entonces presidente Bill Clinton aprobó partidas masivas de dinero para la construcción de prisiones, lo cual permitió remediar la superpoblación y evito la liberación de criminales debido a la saturación de las cárceles. Otras de las medidas tomadas fue la imposición de tácticas policiales novedosas. Se enfatizó el arresto por delitos aparentemente menos graves que en ocasiones revelaron evidencias de crímenes más serios. Finalmente, se prosiguió a realizar un mayor seguimiento y control de los delincuentes en libertad condicional, invirtiendo buena parte del tiempo en buscar y aprehender a aquellos que no acudían a las presentaciones judiciales obligatorias.

Latinoamérica protagoniza, en este tema, una nueva paradoja: a pesar de que las tasas de criminalidad son más altas, las cárceles tienen menos cantidad de reclusos que en los Estados Unidos.

### **2.5.3 Sobrepopulación y hacinamiento en Argentina**

Las condiciones de detención tanto en el ámbito federal como en la provincia de Buenos Aires, continúan siendo preocupantes. El encierro en condiciones de hacinamiento y sobrepoblación pone en serio riesgo la vida y la salud de los detenidos, y constituye, sin lugar a dudas, el sometimiento a un trato cruel, inhumano y degradante. La situación es particularmente grave en la provincia de Buenos Aires, donde el trato inhumano de los detenidos así como su sometimiento a torturas es el resultado de la violencia sistemática y estructural que organiza la vida intramuros, incompatible con un ordenamiento democrático de la sociedad.

El problema de la sobrepoblación en las unidades penitenciarias se encuentra estrechamente vinculado al fenómeno de la violencia al interior de las cárceles. Si bien estos fenómenos, el de la sobrepoblación y el de la violencia, no son únicamente característicos de las cárceles en América Latina, sí es cierto que en nuestro continente desarrollan niveles particularmente altos. Situación que en el caso de la Argentina se presenta tanto a nivel federal como provincial [DNPC, 2005]. Es innegable que el hacinamiento genera un constante incremento de la tensión entre los internos, que deriva en problemas de convivencia en un espacio insuficiente e inadecuado para el desarrollo de la persona. Fomenta, además, la propagación de enfermedades de todo tipo, producto de la falta de higiene de los establecimientos y de adecuada atención médica [CELS, 2005].

La promiscuidad que se experimenta en el interior de los establecimientos hace suponer que los internos se exponen a una alta probabilidad de contraer enfermedades infectocontagiosas, particularmente SIDA. No se puede saber con exactitud el número de infectados puesto que la realización del correspondiente análisis debe contar con la autorización del detenido, negada en la mayoría de los casos. Para el año 2001, se ha detectado la afección del 7% de la población carcelaria, superando casi en doce veces a la tasa estadounidense (0,6%) y superior al porcentaje de personas adultas infectadas en toda América Latina (entre 0.5% y 1%) [Petroni, 2005]

En cuanto a los regímenes de ejecución de penas, el exceso de población carcelaria imposibilita la asignación de tareas laborales a todos los internos y participación en actividades educativas y recreativas, obstaculizando su reinserción laboral. Los presos no pueden trabajar y ganarse su dinero porque “las plazas de trabajo no alcanzan” [Alconada Mon, 2004].

### **2.5.3.1 Sobrepopulación. Dimensión del problema**

El **Servicio Penitenciario Federal** administra 29 cárceles (contando los complejos federales), y 10 alcaldías en 11 provincias de la República Argentina [Anexo 7]. El índice de sobrepoblación ha ido disminuyendo a través de los años, pasando de un 3,3% en 2004 a tener capacidad ociosa en 2006 debido a la reducción en la población penal perteneciente al SPF. Sin embargo hay que tener en cuenta que existían en el año 2004 varias de las unidades que alojaban una cantidad superior a la permitida, tales como la unidad 3 (50% de sobrepoblación), unidad 20 (82%), unidad 2 (14%), unidad 13 (16%), entre otras. Para el año 2006, había unos pocos alojamientos con mayores plazas ocupadas de las disponibles, entre ellas la unidad 13 que aumento su tasa (25% de sobrepoblación), unidad 17 (8.5%) y unidad 2 que pudo reducir el ratio en 10 puntos aproximadamente (4.2%) En el anexo 8 se pueden observar los distintos porcentajes de sobrepoblación de los últimos 4 años censados para cada provincia e instituto del Servicio Penitenciario Federal.

Claro que la situación del SPF es casi ejemplar si se la compara con lo que pasa en el resto de las provincias. Por lo tanto resulta elemental analizar por separado la situación en la provincia de Buenos Aires y el resto de las provincias.

A fines del 2006, las **cárceles de Buenos Aires** alojaban a 23.878 internos, casi el 50% del total del país, que era de 54.000 [Anexo 5]. Teniendo en cuenta que la capacidad disponible a esa fecha era de 22.164 plazas, la sobrepoblación llegaba al 7,7% [Anexo 9]. Estos números aumentan si se tienen en cuenta las personas ubicadas en comisarías de la provincia. Para fines de 2005, había unas 4592 personas alojadas alrededor de las 340 dependencias de la policía de la provincia de Buenos Aires. Si bien significaba una mejoría con respecto a fines de 2002 cuando las comisarías de la provincia tuvieron su pico de población (7.507 personas), son cifras que marcan claramente el problema de la sobrepoblación y el hacinamiento en el sistema penitenciario bonaerense [Ministerio Seguridad Prov. Bs.As., 2004]. Frente a esta situación es preciso destacar nuevamente que las dependencias policiales no deben considerarse plazas carcelarias, en tanto no reúnen las condiciones requeridas para los lugares de cumplimiento de pena y/o prisión preventiva. La presencia de personas detenidas en estos espacios es claramente ilegal. Ni las comisarías se encuentran en condiciones edilicias de mantener en ellas por un tiempo prolongado a personas detenidas, ni el personal policial está capacitado para hacer frente a esta situación [CELS, 2005].

En lo que refiere al resto de las provincias, la situación se vuelve aún peor que en Buenos Aires. Analizando los números informados por las provincias acerca de capacidades y detenidos, se observa que a fines del 2006 la sobrepoblación llegaba al 11.36%. Se puede ver que aproximadamente 50 de los 121 establecimientos penitenciarios distribuidos por todas las provincias se encuentran sobresaturados.

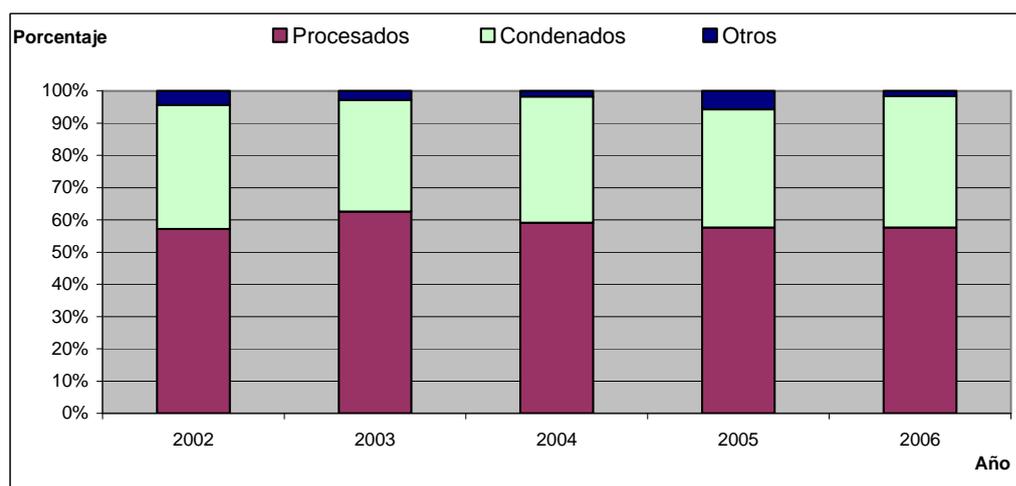
### **2.5.4 Situación penal de los detenidos**

En su gran mayoría, quienes deben soportar esta situación de encierro en condiciones inhumanas e ilegítimas resultan ser personas detenidas solo preventivamente. Es decir, que se encuentran a la espera de una resolución judicial que determine si resultan o no culpables de los delitos que se les imputan. A su vez, la situación se torna más grave si se repara en el tiempo en que se prolongan los procesos penales y, por ende, también la prisión preventiva. Este tema es de suma importancia, ya que existe todo un régimen de progresividad (etapas de observación, tratamiento y prueba) en la ejecución de la pena que es aplicable solo a los condenados.

Según el censo sobre población carcelaria en penitenciarías del 2004, el 60% de los detenidos se encontraba procesado, mientras que solo el 40% había recibido condena. La situación es peor si se tiene en cuenta que los números en la provincia de Buenos Aires hablan de solo un 20% de condenados y un 80% de procesados. Si se consideran a los detenidos en comisarías bonaerenses, la proporción aumenta a casi el 84% de procesados, contra solo un 10% de condenados y un 6% de inimputables y menores, tal como se observó anteriormente en la figura 2.14. Finalmente, en lo que respecta al SPF, se tiene la misma proporción entre condenados y procesados.

El porcentaje de personas procesadas en prisión preventiva en la Provincia de Buenos Aires sólo es superado en América del Sur por el de Paraguay (92,7%), encontrándose muy por encima de los porcentajes de Uruguay (77,2%), Ecuador (68,3%) y Perú (63,2%) [Petroni, 2005].

La situación parece no haber cambiado tanto para el año 2006, como se puede observar en la siguiente figura, ya que las estadísticas oficiales hablan de un 57% de procesados, 41% de condenados y un 2% inimputables o menores [DNPC, 2006].



**Figura 2.16.** Proporción Procesados/Condenados a nivel Nacional. [DNPC, 2006]

La separación de procesados y condenados es un principio elemental del tratamiento carcelario. Los procesados se encuentran en la particular situación de incertidumbre que deriva del hecho de que el proceso que los tiene detenidos no ha sido definido aún, desconocen cuanto tiempo deberá seguir en esa condición, cual será la resolución final y, por cierto, hasta tanto se lo condene, mantienen su condición de inocentes. Por tales circunstancias, la inclusión de procesados en planes de trabajo y educación intensos resulta por demás conflictiva [Petrone, 2005].

### **2.5.5 Costo promedio de cada preso.**

Cada preso alojado en alguno de los establecimientos del SPF le cuesta al Estado alrededor de \$ **4,285**. Este número resulta por si elocuente, pero lo es aún más si se considera que, lejos de redundar en un servicio ejemplar, el hacinamiento, la superpoblación y la promiscuidad conviven a diario dentro del servicio, tal como se describió en los puntos anteriores. El cálculo está hecho sobre la base del presupuesto oficial del año 2006, dividiendo lo destinado al SPF y la cantidad de detenidos en establecimientos federales a fin de dicho año.

Lo curioso es que si se tiene en cuenta el presupuesto del año 2004, alrededor del 80% del presupuesto del SPF estaba afectado para gastos de personal y pago a retirados y pensionados. Como se observa en el anexo 10, unos 288 millones de pesos eran destinados a gastos de personal, sobre un total de 372 millos destinados entre otras cosas para la seguridad, manutención y rehabilitación de los internos para lograr una adecuada reinserción posterior en la sociedad. Para dicho año el SPF tenía más de 8100 agentes, lo que establece una relación penitenciario-detenido cercana al 1 a 1. La relación baja a 1 a 0,4 si se tienen en cuenta los que realizan servicio en las unidades penitenciarias. Esto último evidencia que gran parte de la dotación se dedica a tareas administrativas fuera de las cárceles [Petrone, 2005].

En la tabla 2.2 se puede observar la evolución del presupuesto destinado al SPF desde el año 2001 al 2006, y la cantidad promedio por detenido en el mismo período.

<b>PRESUPUESTO SPF</b>	<b>2001</b>	<b>2002</b>	<b>2003</b>	<b>2004</b>	<b>2005</b>	<b>2006</b>
Actividades Comunes al Servicio Penitenciario Federal (Programas 16,17 y 15)	\$46,587,807	\$26,150,299	\$24,890,053	\$61,127,136	Sin Datos	Sin Datos
Seguridad y Rehabilitación del interno	\$96,937,642	\$122,693,913	\$116,632,790	\$151,527,695		
Pagos a Retirados y Pensionados	\$89,424,056	\$95,518,206	\$100,135,408	\$125,856,449		
Cooperación Técnica y Financiera para la Laborterapia de Internos	\$7,660,004	\$5,894,010	\$6,255,851	\$7,670,190		
Formación y Capacitación	\$6,277,455	\$8,195,581	\$8,669,590	\$11,743,253		
Obras Penitenciarias	\$39,292,439	\$39,283,890	\$30,076,587	\$14,460,295		
<b>Presupuesto Total +Obras Penitenciarias</b>	<b>\$286,179,403</b>	<b>\$297,735,899</b>	<b>\$286,660,279</b>	<b>\$372,385,018</b>	<b>\$432,281,000</b>	<b>\$482,356,000</b>
<b>Cantidad Detenidos</b>	7,872	8,795	9,246	9,738	9,625	9,380
<b>Presupuesto Mensual por Detenido</b>	<b>\$3,030</b>	<b>\$2,821</b>	<b>\$2,584</b>	<b>\$3,187</b>	<b>\$3,743</b>	<b>\$4,285</b>

**Tabla 2.2.** Evolución del presupuesto SPF. [Petrone, 2005]

### **2.5.6 Trabajo en las cárceles**

Las estadísticas revelan que solo un 31% de los detenidos posee algún trabajo dentro de la cárcel. De este 31% de mano de obra ocupada, solo un 9% tiene un trabajo relacionado con actividades productivas. Haciendo foco en la situación legal, es de esperar que el porcentaje de no condenados que trabaja sea menor que el de los condenados. Tal como se dijo en el punto 2.5.4, la situación de los no condenados asociados a un régimen laboral resulta por demás conflictiva.

Situación Legal	Trabajo Remunerado		Trabajo Productivo	
	Si	No	Si	No
Condenados	35%	65%	8%	92%
No condenados	28%	72%	10%	90%

**Figura 2.17.** Distribución de trabajadores según Situación Legal.

Más allá del 9% de la población carcelaria que trabaja en actividades productivas, prácticamente el trabajo en las cárceles no existe. En rigor de verdad, no está sucediendo el mito de finales del siglo XIX que decía que todo el mundo iba a estar trabajando en las cárceles y después iba a volver al mercado, recuperado y más dócil [Daroqui, 2008].

### **2.5.7 Reflexiones acerca del diagnóstico del Servicio Penitenciario**

Es importante aclarar, que mediante el diagnóstico desarrollado anteriormente se busca que el lector pueda, en alguna medida, introducirse en esta sociedad tan particular como lo es la carcelaria. Se cree que para entender la problemática lo más a fondo posible, es sumamente importante dar a conocer estos indicadores.



### **3. DESCRIPCIÓN DEL PROBLEMA**

#### **3.1. TRATAMIENTO DE LA INFORMACIÓN**

En la actualidad, la Dirección Nacional de Política Criminal (DNPC) del Ministerio de Justicia y Derechos Humanos de la Nación (MJDHN) solo publica las estadísticas provenientes del Sistema Nacional de Estadísticas sobre Ejecución de la Pena sin tratamiento alguno de las mismas. A través de un análisis estadístico básico, en el cuál lo que se hace es mostrar las frecuencias de ocurrencia de cada aspecto analizado, se esta desaprovechando toda la información que proporcionan los datos.

En tiempos en los que los volúmenes de información son muy grandes, se hace necesario un tratamiento de la información más complejo, esperando encontrar relaciones subyacentes y comportamientos en los datos que no pueden identificarse mediante un tratamiento estadístico clásico. Las técnicas estadísticas clásicas se centran generalmente en técnicas confirmatorias, mientras que las técnicas de minería de datos son generalmente exploratorias, pudiendo validar comportamientos ya conocidos como también plantear nuevas hipótesis.

#### **3.2. EL PROBLEMA ESPECÍFICO**

Mientras que la cantidad de presos en establecimientos en la República Argentina continúa en ascenso continuo, lo mismo ocurre con las tasas de delitos registradas año tras año. Estamos frente a una situación en la cual los delincuentes entran a las penitenciarias para la readaptación y resocialización, y sin embargo cuando la mayoría de ellos cumplen sus penas y dejan las cárceles, vuelven a buscar al delito como medio de subsistencia. Existe dentro de estos establecimientos penitenciarios diversos factores como el ocio, la violencia, el hacinamiento, la promiscuidad, la soledad, que hacen que los detenidos sean despersonalizados moral, psíquica y físicamente.

Por lo tanto se pueden destacar dos frentes a atacar. En primer lugar, si se quiere readaptar al detenido a una sociedad de la cual ya no forma parte, es fundamental que se llegue a cumplir con los fines que se buscan mediante la aplicación de la sanción: prevención, retribución, readaptación, resocialización y futura ubicación social de aquél. De esta manera, esperaríamos que el preso pueda volver a integrarse de manera definitiva a la sociedad de la cual fue separado, sin necesidad marcada de volver a delinquir. En segundo lugar, se tiene que trabajar intensivamente en la prevención, es decir, evitar que nuevas personas cometan algún tipo de delito y se conviertan en los futuros presos. Si no se lograra esto último se estaría frente a un “*loop*” interminable.

Por lo tanto, uno de los **principales objetivos** que se pretende alcanzar mediante el presente proyecto es el siguiente:

*Analizar la factibilidad y valor agregado de aplicar minería de datos en poblaciones carcelarias logrando entender mejor los motivos que pudieron haber motivado a esta fracción de la sociedad a delinquir y extraer conclusiones que sirvan de apoyo para la creación de políticas criminales reparadoras intramuros y preventivas fuera de las cárceles.*

## **4. SOLUCIÓN**

### **4.1 PROPUESTA DE SOLUCIÓN**

Una vez planteados tanto el estado de la cuestión, como el problema que se requiere resolver, se continúa con la propuesta de solución al problema.

A los fines de caracterizar a la población carcelaria y encontrar conductas y patrones de comportamiento en los reclusos, se decide encarar la problemática de la siguiente manera:

1. Proceso de *clustering* o agrupamiento utilizando los atributos más significativos de los presos.
2. Análisis de los clusters obtenidos. Validación con especialistas en temática criminal.
3. Aplicación de algoritmos de inducción a cada cluster para la identificación de reglas de decisión que ayuden a explicar la composición de cada grupo.

De esta forma, mediante la utilización de algoritmos de agrupamiento y de clasificación, se esperan obtener comportamientos comunes entre los presos explorando los datos almacenados en el sistema SNEEP. Se busca generar grupos y reglas relevantes que ayuden a la toma de decisiones de la DNPC.

Se cree conveniente seguir una metodología de trabajo en minería de datos. Se utilizará como marco de trabajo a la **CRISP-DM**, explicada en el capítulo 2 “Estado de la Cuestión” y como herramienta informática al *Rapid-i*, un sólido programa de minería de datos libre uso (“*freeware*”) que posee gran cantidad de algoritmos.

### **4.2 APLICACIÓN DE CRISP-DM**

Para poder desarrollar el problema planteado de una forma adecuada, se aplica la metodología CRISP-DM. La misma permite realizar el trabajo en forma ordenada y consistente pasando por cada una de sus fases:

- **Fase I:** Compresión del Negocio.
- **Fase II:** Comprensión de los Datos.
- **Fase III:** Preparación de los Datos.
- **Fase IV:** Modelado.
- **Fase V:** Evaluación.
- **Fase VI:** Implementación.

## **4.2.1 Fase I: Comprensión del negocio**

### **4.2.1.1 Determinar los objetivos del negocio.**

La Dirección Nacional de Política Criminal (DNPC) fue creada con las siguientes finalidades:

- Realizar estudios e investigaciones estadísticas en el ámbito del sistema penal.
- Desarrollar y ejercer los lineamientos político-criminales llevados adelante por el Ministerio de Justicia de la Nación.
- Diseñar los programas específicos destinados a actuar sobre la problemática delictiva, tanto en la faz preventiva como punitiva, determinando las áreas prioritarias, así como los medios idóneos para la producción de las transformaciones que se considere necesarias con relación al tema.
- Formular y realizar los proyectos de investigación científica aplicada, directamente vinculados con el delito, sus consecuencias y la reacción punitiva estatal.
- Intervenir en todos los proyectos legislativos vinculados con temas de su competencia en el ámbito del Ministerio de Justicia de la Nación.

Mediante el cumplimiento de los objetivos del proyecto se espera concluir las siguientes necesidades:

- Identificar grupos de presos, de acuerdo a características comunes.
- Obtener las reglas que ayudan a comprender los distintos grupos formados.
- Implementar una herramienta que permita caracterizar a la población carcelaria cada cierto tiempo preestablecido y actualizar las medidas desarrolladas en lo que concierne a política criminal preventiva.

Los **criterios de éxito** del proyecto están relacionados en cumplir con estas necesidades, desde el punto de vista de encontrar patrones de conducta y comportamientos en los presos a través de la minería de datos. Se pretende entender los motivos que los llevaron a delinquir para proponer alternativas que permitan mejorar la forma en que estos individuos son reinsertados en la sociedad.

### **4.2.1.2 Evaluación de la Situación**

En los capítulos 2.4 “Información Criminal en Argentina” y 2.5 “Diagnóstico del Servicio Penitenciario en Argentina”, se realiza una buena evaluación de la situación actual. Los requisitos, supuestos y restricciones se presentan a continuación.

### ***Requisitos***

Los diversos requisitos identificables para avanzar con el proyecto son:

- Definir el alcance del proyecto.
- Aprender y entender la herramienta de minería de datos a utilizar.
- Contar con la base de población carcelaria en las primeras etapas para poder realizar corridas de prueba.
- Validar los resultados obtenidos en cada una de las pruebas con especialistas en el tema (jueces, fiscales, abogados y sociólogos).
- Reuniones periódicas con los especialistas de la Escuela de Postgrado de Ingeniería de Software del ITBA con el objetivo de analizar los datos y realizar la técnica de minería de datos más conveniente.

### ***Supuestos***

Entre los supuestos se puede nombrar a:

- Al utilizar una base de datos del año 2004, se supone que el comportamiento de esta población no ha cambiado significativamente desde entonces.
- Al tratarse de un tema tan complejo, como lo es llegar a entender el motivo por el cual una persona comete un delito, al momento de analizarse los grupos se toman diversos supuestos a validar por especialistas. Entre ellos se pueden nombrar al supuesto de que muchos de los hombres tienen hijos aún cuando la mayoría es soltera. Por otro lado, al no tener información sobre la historia de cada preso, se tuvo que suponer que en general esta última no influye de manera fuerte en las acciones de estos individuos.

### ***Restricciones***

En cuanto a las **restricciones**, se consideran las siguientes:

- No se posee información del detenido acerca de su historial familiar y cultural que ayude a caracterizarlo de una manera más completa en cuanto a patrones de conducta y generación de reglas.
- Existen en la base de datos campos inconsistentes, con valores erróneos o vacíos que afectan al análisis de los resultados. Esto hace que se tenga que preparar la información previamente llevando forzosamente a la reducción de la muestra poblacional (ver 4.2.3 Preparación de los datos).

### ***Expectativas***

- Lograr presentar dicho proyecto fuera del ámbito académico en el marco del proyecto final de Ingeniería Industrial del ITBA, con fines de prevención de delitos.
- Adquirir el conocimiento y entendimiento tanto de la técnica de minería de datos como de las herramientas informáticas utilizadas para desarrollar el proyecto. Se espera poder aplicar dicho conocimiento en otros campos y rubros a lo largo de mi carrera profesional.
- Ampliar mi conocimiento en cuestiones claves dentro de la política tanto criminal, como judicial y de derechos humanos.

#### **4.2.1.3 Plan de Proyecto**

Se describe el plan para llevar adelante el proyecto, según las fases que plantea la metodología CRISP-DM. El tiempo total estimado de realización es de 5 meses y corresponde a una persona trabajando “full time”.

##### ***Fase I – Comprensión del negocio:***

1. Búsqueda y análisis de la información criminal disponible en Argentina
2. Búsqueda, análisis y diagnóstico del servicio penitenciario en Argentina.
3. Estudiar aplicaciones de minería de datos en poblaciones carcelarias o con fines de prevención de delitos.
4. Analizar técnicas y algoritmos de minería de datos.

Tiempo estimado: 1 mes

##### ***Fase II – Comprensión de los datos:***

5. Adquirir los datos de la base SNEEP.
6. Recolectar los datos y estudiar el vínculo entre las tablas de la base de datos.
7. Decodificación y exploración de los datos mediante gráficos de barras y de torta.
8. Verificación de la calidad de los datos.

Tiempo estimado: 3 semanas

##### ***Fase III – Preparación de los datos:***

9. Preparación del *dataset*.
10. Selección de atributos o campos, limpieza y construcción de los datos

Tiempo estimado: 2 semanas

##### ***Fase IV – Modelado:***

11. Aplicar las técnicas y herramientas seleccionadas en la fase I.

Tiempo estimado: 1 mes

**Fase V – Evaluación:**

12. Evaluar el modelo de minería de datos obtenido.

13. Validar el modelo con especialistas del tema.

Tiempo estimado: 1 semana

**Fase VI – Implementación:**

14. Generar el informe final

Tiempo estimado: 1 mes

**4.2.2 Fase II: Comprensión de los datos**

Esta fase involucra las tareas que hacen a la comprensión de los datos, para lo cual, los datos se deben recolectar, organizar, describir, verificar y limpiar antes de realizar el análisis de los mismos. Esto puede consumir mucho tiempo y es crítico para el proyecto de exploración de información.

**4.2.2.1 Recolección de los datos iniciales**

Para la elaboración del proyecto se utiliza la base de datos de la totalidad de población carcelaria en Argentina a fines del año 2004. Dicha base de datos, conformada por más de 53,000 registros, fue provista para el Ministerio de Justicia y Derechos Humanos de la Nación Argentina bajo la realización y supervisión de la Dirección Nacional de Política Criminal (DNPC). Los datos a utilizarse son los recolectados por el Sistema Nacional de Estadísticos sobre Ejecución de la Pena (SNEEP).

La base de datos original se encuentra en formato *MS-Access* y esta compuesta por las siguientes tablas:

**Tabla Principal:**

- maeCenso [Tabla 4.1]: contiene los datos codificados en forma numérica asociados al censo penitenciario del año 2004, desde los datos concretos del preso antes del ingreso al establecimiento penitenciario hasta su comportamiento dentro del mismo.

**Tablas Secundarias:**

- maeTabCodigos [Tabla 4.2]: diccionario que contiene los atributos asociados a los códigos de la tabla principal.
- tabEstablec [Tabla 4.3]: es la tabla que contiene información relacionada a los establecimientos penitenciarios.
- TabProv [Tabla 4.4]: relacionada a la codificación de las provincias.

#### 4.2.2.2 Descripción de los datos iniciales

A continuación se muestra la estructura (Atributos/Descripción) de cada una de las tablas mencionadas en el punto anterior.

Campo	Descripción	Tipo de Dato
idReco	Clave Principal de acceso a registros	Autonumérico
AniCenso	Año de relevamiento del censo	Numérico
idProvincia	código de id. de provincia	Numérico
idEstabtec	código de id. establecimiento	Numérico
idInterno	código de identificación del interno	Texto
Edad	Edad	Numérico
Sexo	Sexo	Numérico
Nacionalidad	Nacionalidad	Numérico
estadoCivil	Estado Civil	Numérico
nivellInstruc	Nivel de Instrucción	Numérico
ulSitLaboral	Ultima situación laboral al momento del ingreso	Numérico
capLaboallIngresar	Capacitación laboral al momento del ingreso	Numérico
ullugarResid	Ultimo lugar de residencia	Numérico
ulProvResid	Ultima provincia de residencia	Numérico
Jurisdicción	Jurisdicción	Numérico
sitLegal	Situación legal	Numérico
fecDete	Fecha de Detención	Fecha/Hora
fecConde	Fecha de Condena	Fecha/Hora
estabDeProced	Establecimiento u origen de procedencia	Numérico
idDeli1	Tipo de delito por el cual ingresa al establecimiento 1ra. mención	Numérico
idDeli2	2da. mención	Numérico
idDeli3	3ra. mención	Numérico
idDeli4	4ta. mención	Numérico
idDeli5	5ta. mención	Numérico
hsTrabRemu	Cantidad de horas por trabajo remunerado	Numérico
partiProgLabo	Participación en algún programa de formación o capacitación laboral	Numérico
partiProgEdu	Participación en algún programa educativo	Numérico
partiActivDepo	Participación en algún programa de actividades recreativas o deportivas	Numérico
recAsistMedUIAnio	Indica si recibió asistencia médica en el ultimo año	Numérico
recVisitasUIAnio	Indica si recibió visitas en el ultimo año	Numérico
partiAlterOrden	Indica si participó de alteraciones del orden el último año	Numérico
tipInfracDisciplina	Tipo de infracción si tuvo infracciones disciplinarias	Numérico
sancionAplicada	Tipo de sanción disciplinaria aplicada	Numérico
califConducta	Calificación de la conducta del interno	Numérico
tenFugasEvas	Intentos de fugas o evasiones el último año	Numérico
tenSuicidio	Intentos de suicidios	Numérico
fueLesionado	Indica si fue lesionado el ultimo año	Numérico
duracCondena_A	Duración en años de la condena aplicada	Numérico
duracCondena_M	Duración en meses de la condena aplicada	Numérico
tipCondena	Tipo de condena: 1=Prisión, 2=Reclusión	Numérico
tieneMedSegu	Tiene medidas de seguridad	Numérico
esReincidente	Muestra el grado de reincidencia del interno	Numérico
periProgresiv	Tiene período de progresividad	Numérico
tuvoSaliTransi	Indica si tuvo salidas transitorias el ultimo año	Numérico
incRegSemilib	Está incorporado al régimen de semilibertad	Numérico
partiProgPreLib	Indica si participa en algún programa de pre-libertad	Numérico
prisDiscontinua	Indica si participa en algún programa de prisión discontinua	Numérico
semiDetencion	Indica si participa en algún programa de semi-detención	Numérico
tuvoReducPena	Muestra si se le aplicaron reducciones de condena el ultimo año	Numérico
mjTieneHijos	Posee hijos con ella en el establecimiento (solo mujeres)	Numérico

Tabla 4.1. macCenso.

Campo	Descripción	Tipo de Dato
nomVari	Cada uno de los atributos que aparece en la tabla maeCenso	Texto
Item	Posibilidades (código numerico) que puede tomar cada una de las <i>nomVari</i>	Numérico
Descri	Descripcion del Item	Texto

**Tabla 4.2.** maeTabCodigos.

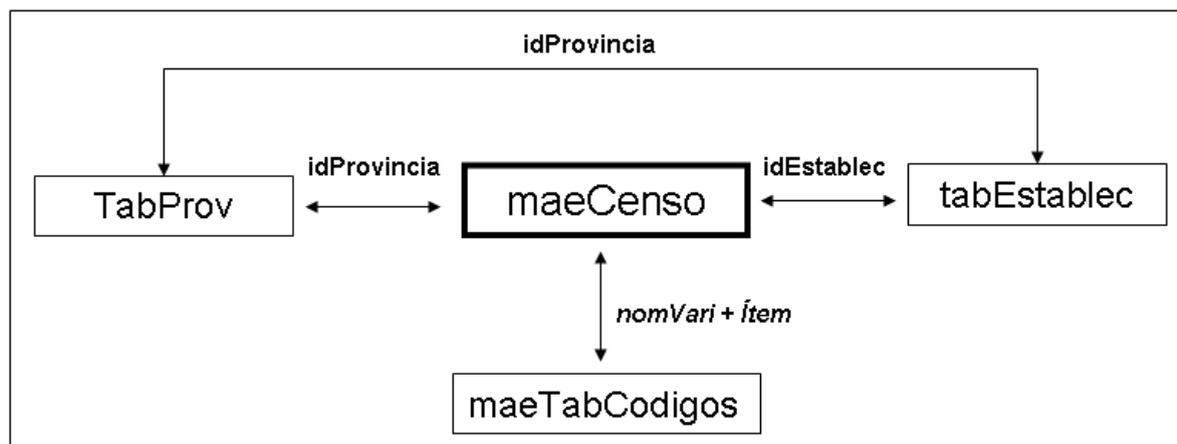
Campo	Descripción	Tipo de Dato
idProv	código de id. de provincia	Numérico
nroID	código de id. establecimiento	Autonumérico
Nombre	Nombre del Establecimiento	Texto
Reparti	N/A	Numérico
tipEstab	Tipo de Establecimiento según modalidad	Texto
direccion	Direccion Fisica del Establecimiento	Texto
cPostal	Codigo Postal	Texto
Fonos	Números de Telefono	Texto
Fax	Números de FAX	Texto
email	Direccion de E-mail	Texto
responsable	Responsable del Establecimiento	Texto
capacidad	Capacidad de alojamiento	Numérico
plazCellndiv	Poblacion ubicada en alojamientos individuales	Numérico
plazLocColec	Poblacion ubicada en alojamientos colectivos	Numérico

**Tabla 4.3.** tabEstablec.

Campo	Descripción	Tipo de Dato
Codigo	código de id. de provincia	Numérico
Descri	Nombre de la Provincia	Texto

**Tabla 4.4.** TabProv.

Las tablas principales y secundarias se vinculan entre sí de la siguiente manera:



**Figura 4.1.** Vinculación entre las tablas.

#### **4.2.2.2.1 Dataset seleccionado**

A priori, los campos o atributos que se eligen para generar el conjunto de datos (*dataset*) y efectuar la minería de datos son los siguientes:

Campo	Descripción	Tipo de Dato
Sexo	Sexo	Numérico
edad	Edad del detenido	Numérico
Nacionalidad	Nacionalidad	Numérico
estadoCivil	Estado Civil del detenido	Numérico
nivellInstruc	Nivel de Instrucción	Numérico
ulSitLaboral	Ultima situación laboral al momento del ingreso	Numérico
capLaboallngresar	Capacitación laboral al momento del ingreso	Numérico
ulLugarResid	Ultimo lugar de residencia (Urbano/Rural)	Numérico
ulProvResid	Ultima provincia de residencia	Numérico
sitLegal	Situación legal	Numérico
idDeli1	Tipo de delito por el cual ingresa al establecimiento 1ra. mención	Numérico
idDeli2	2da. mención	Numérico
hsTrabRemu	Cantidad de horas por trabajo remunerado	Numérico
partiProgLabo	Participación en algún programa de formación o capacitación laboral	Numérico
partiProgEdu	Participación en algún programa educativo	Numérico
partiActivDepo	Participación en algún programa de actividades recreativas o deportivas	Numérico
esReincidente	Muestra el grado de reincidencia del interno	Numérico

**Tabla 4.5.** *Dataset* preliminar seleccionado.

El presente *dataset* es preliminar. Consta de características de los presos al momento de ser detenidos, durante el ingreso del preso, y algunos campos que describen el comportamiento de los mismos ya dentro del establecimiento penitenciario. Los campos que presentan mayores inconsistencias y problemas de calidad son los que proceden de la información obtenida del interior del penal, como la conducta, infracciones, alteraciones del orden, etc. Debido a esto último, en principio, no serían utilizados para realizar el estudio. Más adelante, cuando se verifique la calidad de los datos, se profundizará en el tema.

#### **4.2.2.2.2 Decodificación de los campos**

Para la decodificación de los valores de la tabla principal en formato numérico se utilizan consultas (*Queries*) programadas en lenguaje SQL mediante el software *MS-Access*, según las relaciones de las distintas tablas [Figura 4.1]. Cabe mencionar, que se trabajará con atributos alfanuméricos.

Los valores que pueden tomar cada uno de los atributos seleccionados a priori pueden ser los siguientes:

nomVari	Item	Descri
Sexo	1	Masculino
Sexo	2	Femenino

**Tabla 4.6.** Sexo.

nomVari	Item	Descri
Nacionalidad	1	Argentina
Nacionalidad	2	Boliviana
Nacionalidad	3	Brasileña
Nacionalidad	4	Chilena
Nacionalidad	5	Paraguaya
Nacionalidad	6	Uruguaya
Nacionalidad	7	Peruana
Nacionalidad	8	Colombiana
Nacionalidad	9	Ecuatoriana
Nacionalidad	10	Española
Nacionalidad	11	Italiana
Nacionalidad	12	Inglesa
Nacionalidad	13	China
Nacionalidad	14	Sudafricana
Nacionalidad	15	Nigeriana
Nacionalidad	16	Otras

**Tabla 4.7.** Nacionalidad.

nomVari	Item	Descri
estadoCivil	1	Soltero
estadoCivil	2	Casado
estadoCivil	3	Viudo
estadoCivil	4	Separado o divorciado
estadoCivil	5	Separado de hecho
estadoCivil	6	Concubino

**Tabla 4.8.** Estado Civil.

nomVari	Item	Descri
nivellInstruc	1	Ninguno
nivellInstruc	2	Primario incompleto
nivellInstruc	3	Primario completo
nivellInstruc	4	Secundario incompleto
nivellInstruc	5	Secundario completo
nivellInstruc	6	Terciario incompleto
nivellInstruc	7	Terciario completo
nivellInstruc	8	Universitario incompleto
nivellInstruc	9	Universitario completo
nivellInstruc	10	EGB 1 Incompleto
nivellInstruc	11	EGB 1 Completo
nivellInstruc	12	EGB 2 Incompleto
nivellInstruc	13	EGB 2 Completo
nivellInstruc	14	EGB 3 Incompleto
nivellInstruc	15	EGB 3 Completo
nivellInstruc	16	Polimodal Incompleto
nivellInstruc	17	Polimodal Completo

**Tabla 4.9.** Nivel de Instrucción.

nomVari	Item	Descri
ulSitLaboral	1	Trabajador de tiempo completo
ulSitLaboral	2	Trabajador de tiempo parcial
ulSitLaboral	3	Desocupado

**Tabla 4.10.** Última Situación Laboral.

nomVari	Item	Descri
capLaboallngresar	1	Tiene algún oficio
capLaboallngresar	2	Tiene alguna profesión
capLaboallngresar	3	No tiene ni oficio ni profesión

**Tabla 4.11.** Capacitación Laboral al momento de ingreso.

idProvincia	Descri
1	Buenos Aires
2	Catamarca
3	Córdoba
4	Corrientes
5	Chaco
6	Chubut
7	Entre Rios
8	Formosa
9	Jujuy
10	La Pampa
11	La Rioja
12	Mendoza
13	Misiones
14	Neuquen
15	Rio Negro
16	Salta
17	San Juan
18	San Luis
19	Santa Cruz
20	Santa Fe
21	Santiago Del Estero
22	Tierra Del Fuego
23	Tucumán
24	Ciudad De Buenos Aires

**Tabla 4.12.** Última Provincia de Residencia.

nomVari	Item	Descri
sitLegal	1	Condenado
sitLegal	2	Procesado
sitLegal	3	Contraventor
sitLegal	4	Inimputable
sitLegal	5	Otra situación

**Tabla 4.13.** Situación Legal.

nomVari	Item	Descri
partiProgEdu	1	Si - educación formal - EGB
partiProgEdu	2	Si - educación formal - POLIMODAL
partiProgEdu	3	Si - educación formal - TERCIARIA
partiProgEdu	4	Si - educación formal - UNIVERSITARIA
partiProgEdu	5	Si - educación no formal
partiProgEdu	6	No participa de ningún programa educativo

**Tabla 4.14.** Participa en Programas educativos

nomVari	Item	Descri
idDeli1	1	Homicidios dolosos
idDeli1	2	Homicidios dolosos (tent.)
idDeli1	3	Homicidios Culposos
idDeli1	4	Lesiones Dolosas
idDeli1	5	Lesiones Culposas
idDeli1	6	Otros delitos contra las personas
idDeli1	7	Delitos contra el honor
idDeli1	8	Violaciones
idDeli1	9	Otros delitos contra la honestidad
idDeli1	10	Amenazas
idDeli1	11	Privación ilegítima de la libertad
idDeli1	12	Del. contra la lib. comet. por func. público
idDeli1	13	Otros delitos contra la libertad
idDeli1	14	Hurto y/o tentativa de hurto
idDeli1	15	Robo y/o tentativa de robo
idDeli1	16	Otros delitos contra la propiedad
idDeli1	17	Delitos contra la seguridad pública
idDeli1	18	Delitos c/el orden público
idDeli1	19	Delitos contra la seguridad de la nación
idDeli1	20	Delitos contra los poderes publicos
idDeli1	21	Delitos c/ la administracion pública
idDeli1	22	Delitos c/ la fe pública
idDeli1	23	Delitos c/ el estado civil
idDeli1	24	Infracción ley n° 23.737 (estupefacientes)
idDeli1	25	Infraccion ley n° 24.769 penal tributaria
idDeli1	26	Infraccion ley n° 13.944 incumplimiento de deberes
idDeli1	27	Delitos previstos en leyes especiales
idDeli1	28	Contravenciones

**Tabla 4.15.** Tipo de delito cometido.

nomVari	Item	Descri
hsTrabRemu	1	Hasta 10 hs. Semanales
hsTrabRemu	2	Hasta 20 hs. Semanales
hsTrabRemu	3	Hasta 30 hs. Semanales
hsTrabRemu	4	Hasta 40 hs. Semanales
hsTrabRemu	5	No tiene trabajo remunerado

**Tabla 4.16.** Cantidad horas de trabajo remunerado

nomVari	Item	Descri
partiProgLabo	1	Si
partiProgLabo	2	No

**Tabla 4.17.** Participa programa Laboral

nomVari	Item	Descri
ulLugarResid	1	Rural
ulLugarResid	2	Urbano

**Tabla 4.18.** Último lugar de residencia (Urbano/Rural)

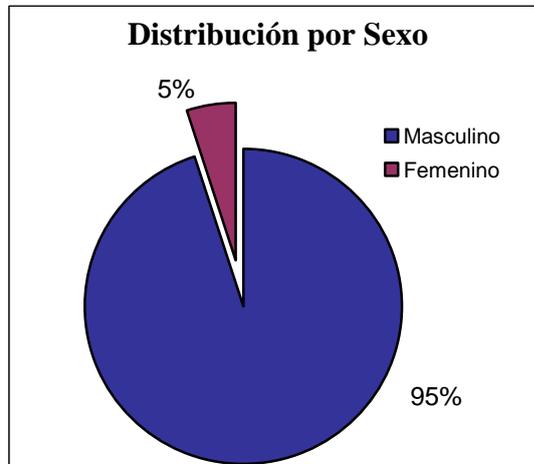
### 4.2.2.3 Exploración de los datos

Mediante el *dataset* preliminar seleccionado en el punto anterior, se lleva a cabo la exploración de los datos para familiarizarse con los mismos, y al mismo tiempo poder identificar problemas de calidad. Se comienzan a detectar faltantes de información en algunos campos que se presentan vacíos (“Sin datos”). En esta etapa no se realiza ningún tratamiento de los datos.

**[Sexo]:** Se puede observar la distribución del sexo entre los internos. Se nota claramente que los detenidos son su mayoría del sexo masculino.

Sexo	Presos
Masculino	50,408
Femenino	2,621
<b>Total</b>	<b>53,029</b>

**Tabla 4.19.** Sexo



**Figura 4.2.** Distribución por sexo.

**[Nacionalidad] / [Edad]:** Se puede observar que la nacionalidad predominante es la Argentina. Luego le siguen presos de distintos países limítrofes de la Nación.

Nacionalidad	Presos
Argentina	50,329
Boliviana	531
Paraguayana	424
Uruguaya	396
Chilena	364
Peruana	313
Otras	132
Brasileña	59
Colombiana	43
Italiana	25
Sudafricana	24
Española	23
China	20
Ecuatoriana	6
Inglesa	6
Nigeriana	3
Sin Datos	331
<b>Total</b>	<b>53,029</b>

**Tabla 4.20.** Distribución de nacionalidades.

Rango Edad	Presos
Menor 18	520
18-24	15,975
25-34	21,044
35-44	9,377
45-54	4,146
55-64	1,405
65 y mas	398
Sin Datos	164
<b>Total</b>	<b>53,029</b>

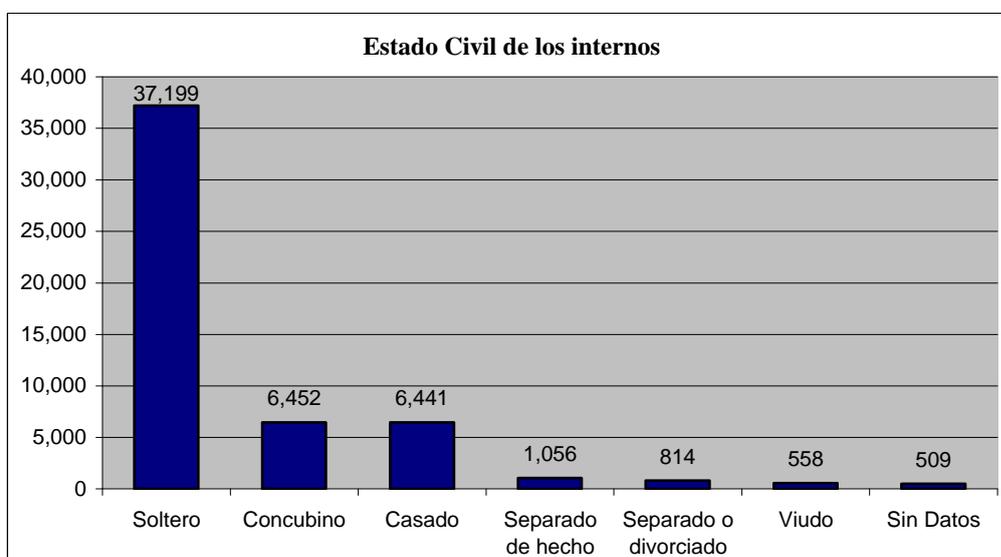
**Tabla 4.21.** Rango de edades

**[Estado Civil]:**

La mayor cantidad de los detenidos indicó que se encontraba soltero al momento de la detención. Luego, en proporciones similares, le siguen los casados y los concubinos.

Estado Civil	Presos
Soltero	37,199
Concubino	6,452
Casado	6,441
Separado de hecho	1,056
Separado o divorciado	814
Viudo	558
Sin Datos	509
<b>Total</b>	<b>53,029</b>

**Tabla 4.22.** Estado Civil.



**Figura 4.3.** Estado Civil de los detenidos.

**[Nivel de Instrucción]:**

Dicho campo refleja el nivel de educación de los internos. Primario incompleto y completo, secundario incompleto y ningún tipo de educación representan el 92% de los casos.

Nivel de Instrucción	Presos
Primario completo	27,451
Primario incompleto	10,740
Secundario incompleto	7,029
Ninguno	3,328
Secundario completo	2,220
EGB 3 Incompleto	286
Universitario incompleto	268
Terciario incompleto	224
EGB 2 Incompleto	165
Terciario completo	146
EGB 3 Completo	136
Universitario completo	110
EGB 1 Incompleto	100
Polimodal Incompleto	89
EGB 1 Completo	49
EGB 2 Completo	46
Polimodal Completo	15
Sin Datos	627
<b>Total</b>	<b>53,029</b>

**Tabla 4.23.** Nivel de Instrucción de los detenidos.

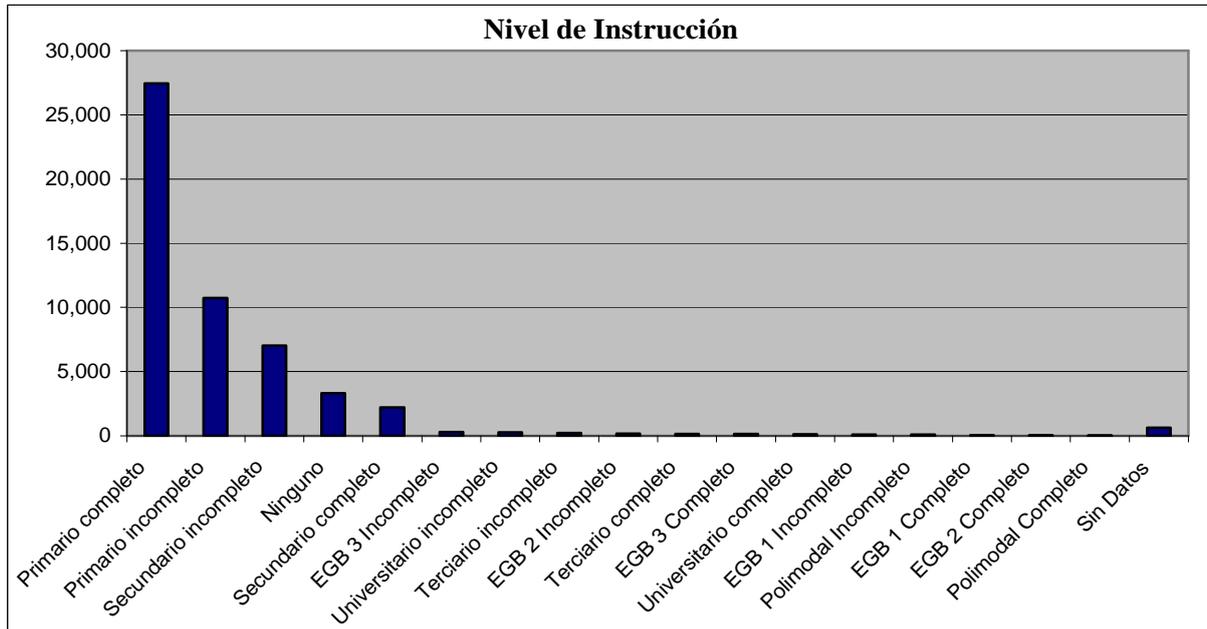


Figura 4.4. Nivel de Instrucción de los detenidos.

**[Última Situación Laboral]:**

Muestra el status laboral de los reclusos al momento de ingresar al penal. Se observa como sólo el 19% de los mismos poseía un trabajo de tiempo completo al ser detenido. Este campo está muy relacionado con el anterior, nivel de instrucción.

Situacion Laboral	Presos
Desocupado	21,342
Trabajador de tiempo parcial	20,831
Trabajador de tiempo completo	9,824
Sin Datos	1,032
<b>Total</b>	<b>53,029</b>

Tabla 4.24. Situación Laboral

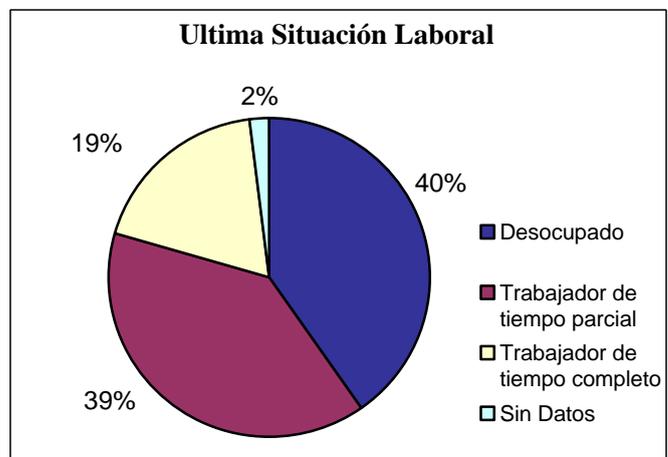


Figura 4.5. Situación Laboral de los detenidos.

**[Capacitación Laboral]:**

Refleja si los detenidos poseían alguna profesión u oficio al momento de ingreso al penal. Alrededor de la mitad de los detenidos no tenía ningún oficio ni profesión.

Situación Laboral	Presos
No tiene ni oficio ni profesión	25,202
Tiene algún oficio	20,290
Tiene alguna profesión	4,312
Sin Datos	3,225
<b>Total</b>	<b>53,029</b>

Tabla 4.25. Capacitación Laboral.

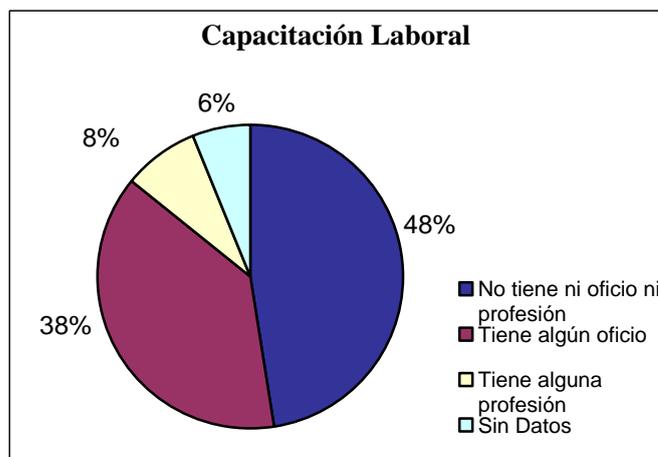


Figura 4.6. Capacitación Laboral de los detenidos.

[Tipo del lugar de residencia]:

Lugar de Residencia	Presos
Urbano	44,663
Rural	5,623
Sin Datos	2,743
<b>Total</b>	<b>53,029</b>

Tabla 4.26. Lugar de Residencia.

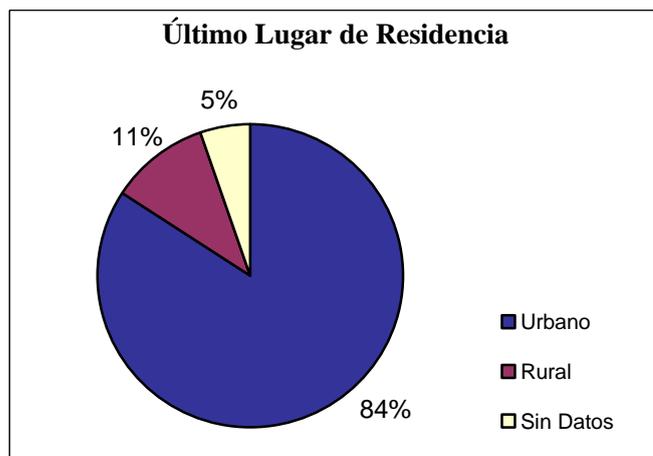


Figura 4.7. Lugar de Residencia de los detenidos.

[Situación Legal]: El presente campo es de real importancia al momento de estudiar a la población carcelaria en Argentina. La separación de procesados y condenados es un principio elemental del tratamiento carcelario. Hoy en día, solo el 40% de los detenidos posee una condena.

Situación Legal	Presos
Procesado	31,043
Condenado	21,054
Inimputable	674
Otra situación	258
<b>Total</b>	<b>53,029</b>

Tabla 4.27. Situación Legal.

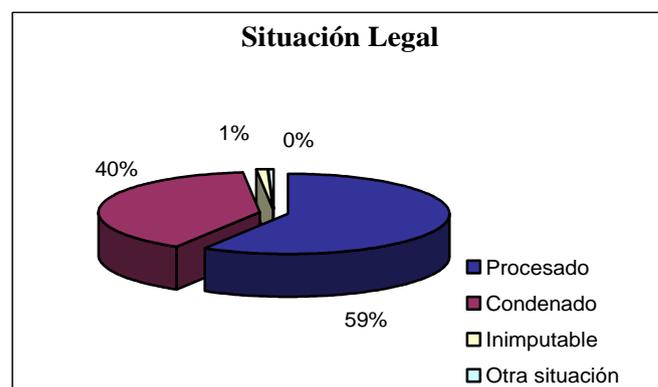


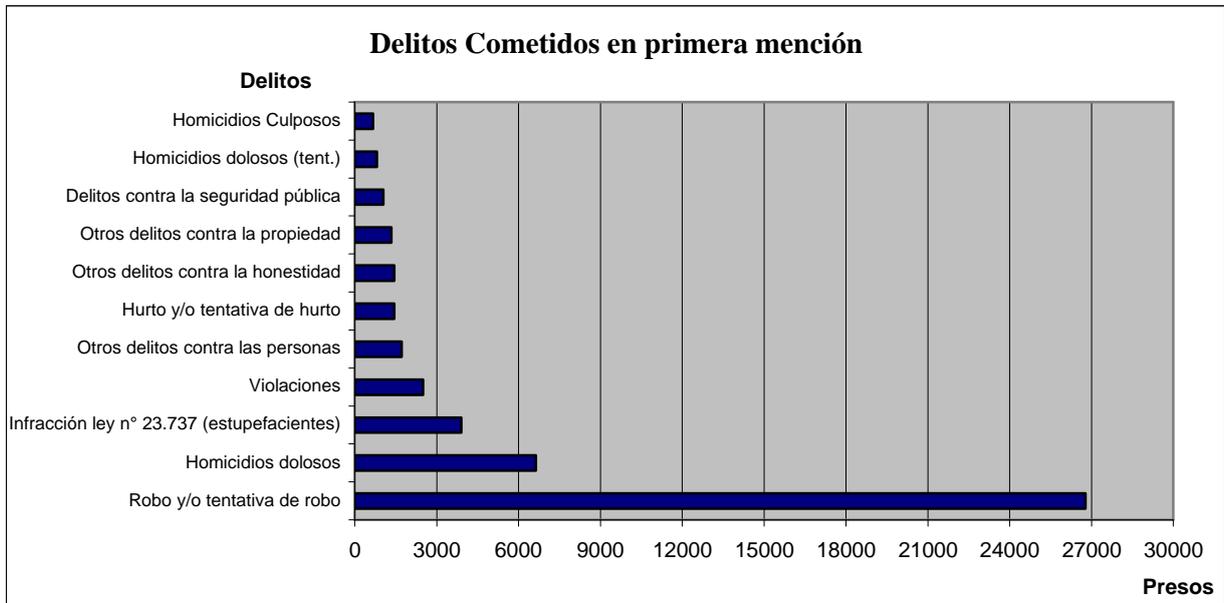
Figura 4.8. Situación legal de los detenidos.

**[Delito cometido]:**

Delito en primera mención por el cual la persona fue detenida según el código penal. Los delitos de robo, homicidios, estupefacientes y violaciones conforman el 87% de los casos.

Delito Cometido	Presos		
Robo y/o tentativa de robo	26,777	Delitos contra el honor	215
Homicidios dolosos	6,635	Otros delitos contra la libertad	194
Infracción ley n° 23.737 (estupefacientes)	3,906	Delitos c/ la fe pública	193
Violaciones	2,504	Delitos c/el orden público	155
Otros delitos contra las personas	1,722	Lesiones Culposas	123
Hurto y/o tentativa de hurto	1,453	Delitos previstos en leyes especiales	58
Otros delitos contra la honestidad	1,452	Infraccion ley n° 24.769 penal tributaria	24
Otros delitos contra la propiedad	1,345	Contravenciones	21
Delitos contra la seguridad pública	1,052	Delitos contra la seguridad de la nación	20
Homicidios dolosos (tent.)	819	Del. contra la lib. comet. por func. público	18
Homicidios Culposos	675	Delitos contra los poderes publicos	15
Privación ilegítima de la libertad	578	Infraccion ley n° 13.944 incumplimiento de c	13
Lesiones Dolosas	494	Delitos c/ el estado civil	9
Delitos c/ la administracion pública	438	Sin Datos	1,861
Amenazas	260	<b>Total</b>	<b>53,029</b>

**Tabla 4.28.** Delitos Cometidos en primera mención



**Figura 4.9.** Delitos Cometidos en primera mención

**[Trabajo remunerado]:**

Representa la proporción de internos que tienen un trabajo remunerado dentro del penal. Vale destacar que para tener acceso a un trabajo remunerado es necesario que el detenido tenga una condena firme.

Trabajo Remunerado	Presos
No tiene trabajo remunerado	30,539
Hasta 20 hs. Semanales	5,435
Hasta 40 hs. Semanales	5,099
Hasta 30 hs. Semanales	4,304
Hasta 10 hs. Semanales	1,672
Sin Datos	5,980
<b>Total</b>	<b>53,029</b>

Tabla 4.29. Trabajo Remunerado.

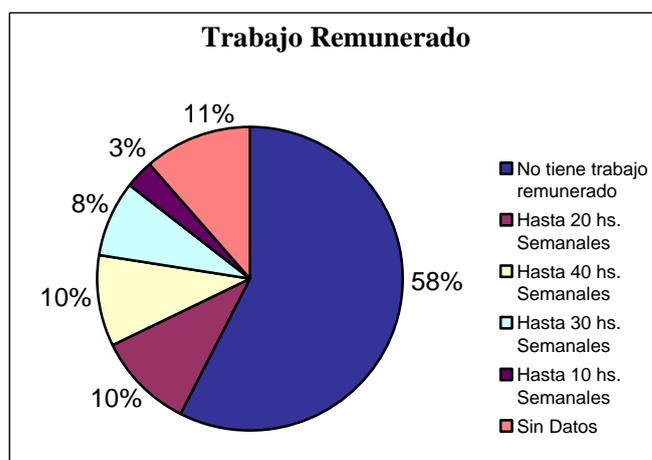


Figura 4.10. Trabajo remunerado en los internos.

**[Participación Programa Capacitación Laboral]:**

El 70% de los reclusos no participa de ningún programa de capacitación laboral. Esto último, sumado al nivel de instrucción y la capacitación laboral al ingreso, entrega un panorama no muy alentador para las personas que pertenecen a la sociedad carcelaria.

Programa capacitación laboral	Presos
No participa	37,344
Si participa	8,442
Sin Datos	7,243
<b>Total</b>	<b>53,029</b>

Tabla 4.30. Capacitación Laboral.

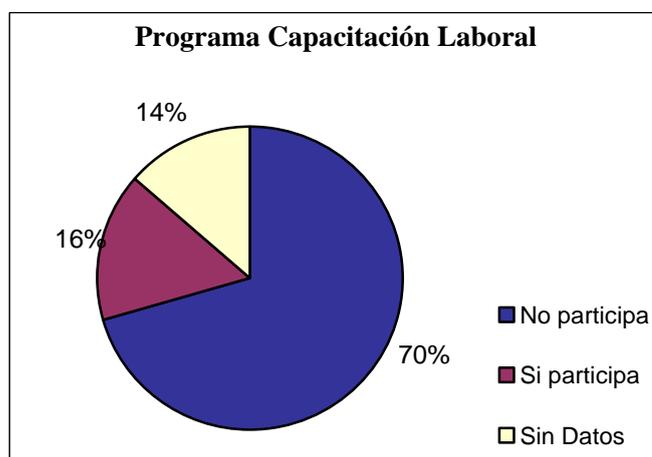


Figura 4.11. Participación en Capacitación Laboral.

**[Participación Programa Educativo]:**

Nuevamente, tal como ocurre en los programas laborales, se observa una muy baja participación en programas educativos.

Participacion en programa educativo	Presos
No participa de ningún programa educativo	29,514
Si - educación formal - EGB	9,615
Si - educación no formal	4,702
Si - educación formal - POLIMODAL	3,511
Si - educación formal - UNIVERSITARIA	312
Si - educación formal - TERCIARIA	136
Sin Datos	5,239
<b>Total</b>	<b>53,029</b>

Tabla 4.31. Programa Educativo.

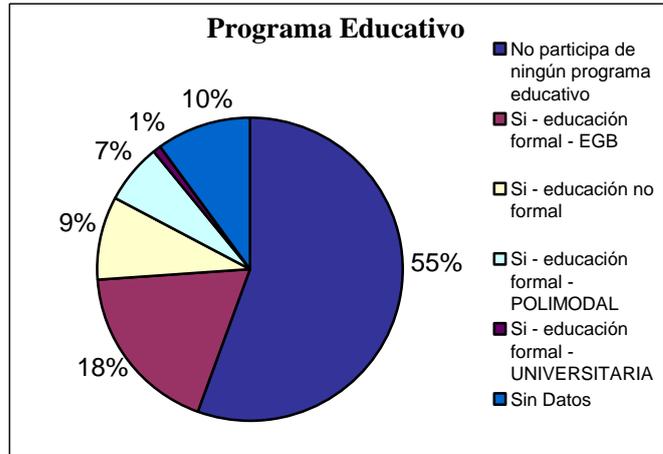


Figura 4.12. Participación en Programa Educativo.

**[Reincidencia]:**

Reincidencia	Presos
Primario	34,264
Reincidente (art. 50 CP)	6,669
Reiterante	2,747
Reincidencia múltiple (art. 52 CP)	389
Sin Datos	8,960
<b>Total</b>	<b>53,029</b>

Tabla 4.32. Reincidencia.

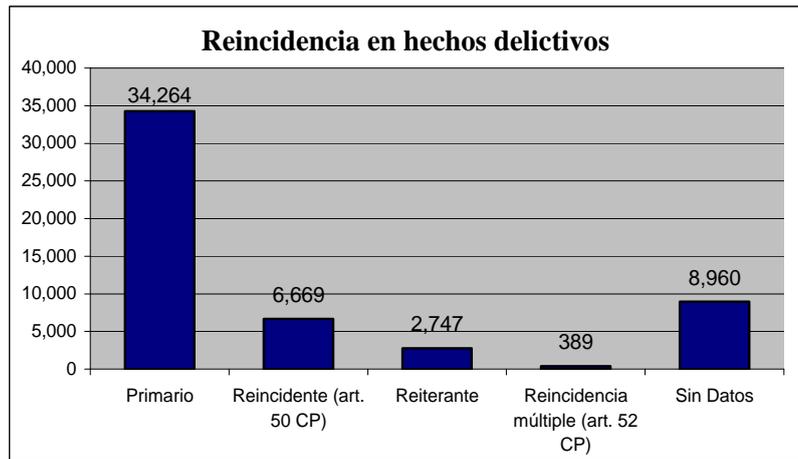


Figura 4.13. Reincidencia.

**4.2.2.4 Verificación de Calidad de los Datos**

Durante la exploración de los datos pudieron observar problemas en la calidad de los mismos, y a su vez, se encontraron varios campos que presentaban datos nulos o ausentes.

En primer lugar, tal como puede observarse en los gráficos del punto anterior, la totalidad de los mismos presentan registros que tienen valor nulo en alguno de sus campos. La cantidad de veces que aparecen estos campos vacíos varía entre los atributos de acuerdo al momento en que el dato haya sido censado: al ingreso del penal o una vez ya ingresado al mismo. Siendo estos últimos, los que presentan mayores problemas de calidad y ausencia de información.

Comenzando por el campo *edad*, existen 456 registros que contienen al número cero como la posible edad del detenido, algo que es realmente imposible. Lo mismo ocurre con otros 10 casos de personas que presentan edades entre los 2 y 13 años.

En lo que respecta al *estado civil* de las personas, de los 509 registros que presentan un nulo como atributo del campo, hay 166 que tienen un valor de *estado civil* = 0, pero que el mismo no es coherente con la codificación del censo, tal como lo muestra la tabla 4.8.

Inconsistencias similares en los datos ocurren con los relacionados a la *nacionalidad*, *última situación laboral*, *capacitación laboral*, *lugar de residencia*, *última provincia de residencia* y en menor medida el *nivel de instrucción*, *delito cometido*, *horas de trabajo remunerado*, *participación de programas laborales y educativos* y *reincidencia*.

Tanto la calidad de los datos, como el faltante de los mismos, son cuestiones muy importantes a tener en cuenta previo a realizar la minería en los datos. Esto se analizará en la **Fase III Preparación de los Datos**.

#### **4.2.3 Fase III: Preparación de los datos**

En esta fase se trabajará sobre la preparación del *dataset* definitivo, la selección y limpieza de los datos. Una vez hechas estas actividades se procede a construir los datos a los cuales les fueran necesarios algún tipo de modificación. Es decir, es en esta etapa en donde se prepara todo lo relativo a la base de datos, para poder luego ingresarlas en la herramienta de modelado.

##### **4.2.3.1 Preparación del Dataset**

La versión final del *dataset*, el cual consta de 40,928 registros sobre presos **masculinos**, a utilizar para realizar la minería de datos se presenta en la tabla 4.33.

Campo	Descripción	Tipo de Dato
Edad	Edad del detenido	Numérico
Estado Civil	Estado Civil del detenido	Texto
Nivel_Instr_Summ	Nivel de Instrucción agrupado	Texto
Ultima Situacion Laboral	Ultima situación laboral al momento del ingreso	Texto
Capacitacion Laboral	Capacitación laboral al momento del ingreso	Texto
Ultima Residencia	Ultimo lugar de residencia (Urbano/Rural)	Texto
Delito Cometido	Tipo de delito Agrupado - 1ra. mención	Texto

**Tabla 4.33.** *Dataset* definitivo.

A los fines de obtener mayor claridad en los resultados obtenidos una vez corridos los modelos, se procede a agrupar las variables de los atributos *nivel de instrucción* y *delito cometido*. Vale destacar que estos dos atributos poseen una gran cantidad de variables,

lo cual hace más difícil el análisis si no se lleva a cabo una agrupación de las mismas. Dicho agrupamiento se puede observar en el apartado 4.2.3.4 de **Construcción de los Datos**.

A su vez, para obtener los 40,928 registros, se tuvieron que agregar distintos filtros a los fines de obtener un *dataset* que fuera lo más limpio posible. Se profundiza el tema en el apartado 4.2.3.3 **Limpieza de los Datos**.

#### **4.2.3.2 Selección de los Datos**

Los campos que se considerarán para el presente estudio se eligen teniendo en cuenta los siguientes factores:

- Variables que puede tomar cada uno de los campos o atributos.
- Calidad de los datos.
- Importancia del atributo para el estudio en cuestión en cuanto a los objetivos del proyecto, información que entrega y significancia del mismo.

Los campos que se consideran para el presente estudio son los siguientes:

##### ***Sexo***

Campo alfanumérico que indica el sexo del detenido. Para el análisis se tienen en cuenta solo a las personas de sexo masculino por representar al 95% de la población penal.

##### ***Edad***

Campo numérico que indica la edad del detenido. Este atributo toma valores desde los 10 años hasta los 89 años.

##### ***Estado Civil***

Describe el estado civil en el que se encontraba el preso al momento de la detención. Los posibles estados son: Soltero, Casado, Concubino, Separado de hecho o Divorciado, Viudo.

##### ***Nivel de Instrucción***

Atributo que describe el nivel de educación del preso y situación de avance (completo o incompleto). Se procede agrupar dicho campo para llevarlo al viejo esquema del sistema educativo: Primario/ Secundario/Terciario/Universitario (ver 4.2.3.4. **Construcción de los Datos**).

### ***Capacitación Laboral***

Tal como indica la palabra, indica el nivel de capacitación que la persona posee laboralmente hablando. Este atributo puede tomar las siguientes variables: tiene oficio, tiene profesión, no tiene oficio ni profesión.

### ***Última Residencia***

Representa el tipo de residencia en el cual vivía el detenido. Toma dos valores posibles: Urbana o Rural.

### ***Delito Cometido***

Describe el tipo de delito cometido en primera mención por el detenido. Se agrupa a este campo según el código penal en los siguientes tipos de delitos: Delitos contra la Propiedad, Delitos contra las Personas, Delitos contra la integridad sexual y el Honor, Infracción ley N° 23.737 (estupefacientes), Contravenciones, Delitos c/ el estado civil, Delitos c/el Estado y la Comunidad, Delitos contra la libertad, Delitos previstos en leyes especiales, Infracción ley N° 13.944 incumplimiento de deberes, Infracción ley N° 24.769 penal tributaria, Otros delitos contra la honestidad. Debido a que los primeros 4 delitos tipificados por el código penal agrupan aproximadamente al 90% de las instancias, se procede a acotar el estudio a estos 4 delito principales.

A continuación se detallan los atributos no tenidos en cuenta para el análisis:

### ***Última provincia de residencia***

Se omite la utilización de este atributo ya que se busca encontrar un comportamiento general, sin hacer foco en cuestiones particulares por provincia. Más del 50% de los casos corresponde a la Provincia de Buenos Aires.

### ***Situación Legal***

Indica la situación legal en la cual se encuentra el detenido. Hay 4 situaciones posibles: Condenado, Procesado, Inimputable u Otra Situación. En un principio se pensó en no omitir este campo, sin embargo se cree que es un dato irrelevante para caracterizar, en una primera aproximación con minería de datos, la conducta que habría llevado a que las personas pertenecientes a esta población cometieran un delito.

### ***Trabajo Remunerado***

Informa acerca de la cantidad de horas que cada detenido pasa trabajando de manera remunerada. Los valores que toma son: no tiene trabajo remunerado, hasta 40 hs. semanales, hasta 30 hs. semanales, hasta 20 hs. semanales, hasta 10 hs. semanales y sin datos. Dado que se trata de un atributo asociado al comportamiento ya dentro de la cárcel y que el mismo presenta un 11% de nulos, se decide omitir dicho campo.

***Participación en programa de formación o capacitación laboral / Participación en algún programa educativo / Participación en algún programa de actividades recreativas o deportivas***

Hacen hincapié en la participación de los internos en los distintos programas del sistema penitenciario. Al igual que el atributo anterior, al ser información asociada a conductas dentro del establecimiento, se decide no utilizar estos campos.

**4.2.3.3 Limpieza de los Datos**

Existen registros que son excluidos del análisis por encontrarse alguna inconsistencia, tal como se especificó en el punto 4.2.2.4 de **Verificación de Calidad de los Datos**.

En lo que respecta al campo *Edad*, se procedió a eliminar los registros que tenían como valor 0, 2 o 13 años. En segundo lugar, mediante una consulta en *MS Access*, se generó la condición de que la base de datos no presentara ningún valor nulo para los atributos seleccionados en el dataset. A través de estas modificaciones, se obtuvo una base limpia de inconsistencias y errores de 40,928 registros.

**4.2.3.4 Construcción e integración de los Datos**

En la presente etapa se procede a la construcción de los datos que conforman el *dataset* definitivo para su posterior ingreso en la herramienta de modelado.

Tal como se comenta anteriormente, se realiza un agrupamiento de los distintos *Delitos Cometidos* teniendo en cuenta la tipificación del código penal. Como puede observarse en la tabla 4.34, los 28 delitos cometidos que figuran en la base de datos terminan agrupándose solo en 12. Se cree necesaria esta tarea a los fines de facilitar la comprensión de los resultados una vez realizadas las corridas experimentales.

Otro de los campos en donde es necesaria su reconstrucción es el campo del *Nivel de Instrucción*. A partir del año 1996, la Escuela Primaria y la Escuela Secundaria cambiaron a un nuevo sistema educativo:

- La Educación General Básica (EGB) de 9 años
- La Educación Polimodal de 3 años.

Delito Cometido	Delito Cometido Unificado
Contravenciones	Contravenciones
Delitos c/ el estado civil	Delitos c/ el estado civil
Del. contra la lib. comet. por func. público Delitos c/ la administracion pública Delitos c/ la fe pública Delitos c/el orden público Delitos contra la seguridad de la nación Delitos contra la seguridad pública Delitos contra los poderes publicos	Delitos c/el Estado y la Comunidad
Delitos contra el honor Violaciones	Delitos contra la integridad sexual y el Honor
Amenazas Otros delitos contra la libertad Privación ilegítima de la libertad	Delitos contra la libertad
Hurto y/o tentativa de hurto Otros delitos contra la propiedad Robo y/o tentativa de robo	Delitos contra la propiedad
Homicidios Culposos Homicidios dolosos Homicidios dolosos (tent.) Lesiones Culposas Lesiones Dolosas Otros delitos contra las personas	Delitos contra las personas
Delitos previstos en leyes especiales	Delitos previstos en leyes especiales
Infraccion ley n° 13.944 incumplimiento de deberes	Infraccion ley n° 13.944 incumplimiento de deberes
Infracción ley n° 23.737 (estupefacientes)	Infracción ley n° 23.737 (estupefacientes)
Infraccion ley n° 24.769 penal tributaria	Infraccion ley n° 24.769 penal tributaria
Otros delitos contra la honestidad	Otros delitos contra la honestidad
Sin Datos	Sin Datos

**Tabla 4.34.** Unificación de delitos según el Código Penal.

A los fines del estudio en cuestión se unificaron los campos según el viejo sistema:

NivelInstruc	Nivel Instrucción Unificado
Ninguno	Ninguno
Primario Completo	Primario Completo
Primario Incompleto	Primario Incompleto
Secundario Completo	Secundario Completo
Secundario Incompleto	Secundario Incompleto
EGB 1 Completo	
EGB 1 Incompleto	
EGB 2 Completo	
EGB 2 Incompleto	
EGB 3 Completo	
EGB 3 Incompleto	
Polimodal Incompleto	
Polimodal Completo	Terciario completo
Terciario completo	
Terciario Incompleto	Terciario Incompleto
Universitario Incompleto	Universitario Incompleto
Universitario Completo	Universitario Completo
Sin Datos	Sin Datos

**Tabla 4.35.** Unificación del campo Nivel de Instrucción.

Una vez terminada la Fase III, en donde se preparó el *dataset* definitivo, se seleccionaron, limpiaron y se modificaron los datos, se continúa con la Fase IV de Modelado.

#### **4.2.4 Fase IV: Modelado**

Es en esta fase en donde se seleccionan y aplican las técnicas de modelado. Al mismo tiempo se determinan los valores de los parámetros y variables de calibración.

Previo a seleccionar las técnicas y algoritmos para la explotación de la información, es necesario seleccionar el software a utilizar. La herramienta informática que se utiliza para llevar a cabo la presente fase es el *RapidMiner* o más conocido como *Rapid-i*. El programa seleccionado cuenta con las siguientes ventajas:

- Distribución 100% libre.
- Funciona sobre cualquier plataforma y sistema operativo.
- Los procesos de explotación de información son modelados en forma de árboles operadores, lo cual resulta fácil e intuitivo.
- Los árboles operadores o sub-árboles pueden ser guardados en formas de bloques para su posterior utilización.
- Posee más de 400 modelos de aprendizaje, evaluación, pre-procesamiento y post-procesamiento y operadores para visualización.
- Integra la librería del programa WEKA.
- La característica más importante es la capacidad de jerarquizar cadenas del operador y de construir complejos árboles de operadores.

Sin embargo, como todo software, presenta puntos flojos:

- La interfaz grafica no es del todo “amigable”.
- Salidas gráficas limitadas. Solo realiza la gráfica de una serie de datos elegida al azar, la cual está limitada a las 1000 muestras.
- No posee un “tutorial” fácil de navegar que permita aprender el programa de manera completa y rápida.
- Salidas poco claras para quienes no acostumbran a aplicar minería de datos.

En el anexo 14 se explican los conceptos básicos del programa y se incluyen ejemplos para entender las características operativas de la herramienta.

#### **4.2.4.1 Selección de la técnica de modelado**

En esta etapa, se seleccionará algunas de las técnicas y algoritmos desarrollados a lo largo del capítulo 2.1.1, perteneciente al Estado de la Cuestión.

Al estar trabajando con una base de más de 40,000 registros, se hace necesario realizar un agrupamiento o *clustering* de los datos para que el descubrimiento de la información oculta en ellos se torne lo menos difícil posible. El algoritmo que se utilizará es el *K-Means*, uno de los métodos más utilizados y el más popular de los métodos de agrupamiento “por partición”.

El próximo paso es el de validación de los grupos o *clusters* con los especialistas en temática criminal. Esta validación representa una real importancia para poder seguir adelante con la técnica de clasificación.

Previo a la clasificación, es necesario preseleccionar los atributos que serán utilizados posteriormente. Para la selección de atributos, tal como se menciona en el “Estado de la Cuestión”, existen métodos basados en filtros y aquellos llamados envolventes (*wrappers*). En el presente estudio, se utilizan ambos métodos de selección de atributos para validar la significancia de los mismos

Finalmente, se aplican algoritmos de inducción (clasificación) a cada cluster para la identificación de reglas de decisión que ayuden a explicar la composición de cada grupo formado. Para la clasificación se utilizan los algoritmos C4.5 y CHAID.

#### **4.2.4.2 Construcción del modelo y resultados experimentales**

Se procede a explicar cada uno de los modelos utilizados y sus resultados. A su vez, en donde corresponda, se explican los parámetros utilizados y configuraciones realizadas en el *Rapid-i* para el desarrollo de la construcción del modelo.

##### **4.2.4.2.1 Clustering**

A los fines de concretar el agrupamiento de los datos, se utiliza el algoritmo *K-Means* para obtener 4 grupos de presos a través de sus características más significativas.

La configuración de dicho algoritmo de agrupamiento exige que se elija un número, denominado semilla (*seed*), para realizar una distribución aleatoria inicial a partir de la cual el algoritmo comienza las sucesivas iteraciones. El número elegido para llevar a cabo el proceso es el que viene por defecto en el programa (*seed* = 10). Al ser el *K-Means* un algoritmo que se basa en minimizar la suma del error cuadrático, podría tomarse a ese error como punto de partida para la elección del número óptimo de la

semilla. Cabe destacar, que el *Rapid-i* no cuenta con alguna salida en la cual informe la suma del error cuadrático. Sin embargo, a modo de prueba, se utilizó otro programa de minería de datos para realizar corridas con diferentes números de semilla, dando como resultado al valor de semilla=10 como la que minimizaba el error cuadrático.

Una vez hecha la configuración y establecidos los parámetros, se introduce el *dataset* definido en el punto 4.2.3.1 al modelo de minería de datos para realizar el *clustering* correspondiente. A continuación se muestran los resultados obtenidos.

#### **4.2.4.2.2 Resultados del Clustering**

El resultado obtenido en *Rapid-i* una vez ejecutado *K-Means* con 4 clusters, se resume a continuación.

Atributo	Cluster 0	Cluster 1	Cluster 2	Cluster 3
Delito Cometido	Contra la Propiedad	Contra las Personas	Contra la Propiedad	Violación / Estupefactientes
Nivel de Instrucción	Primario Completo	Primario Completo	Primario Completo	Primario Completo
Ultima Situacion Laboral	Tiempo Parcial	Desocupado	Desocupado	Tiempo Parcial
Capacitación Laboral	Oficio	Ni Oficio ni Profesión	Ni Oficio ni Profesión	Oficio / Profesión
Estado Civil	Soltero	Soltero	Soltero	Casado
Lugar de Residencia	Urbana	Urbana	Urbana	Urbana
Edad Promedio	31	34	27	43
<b>Total</b>	<b>16849 (41%)</b>	<b>6513 (16 %)</b>	<b>14662 (36%)</b>	<b>2904 (7%)</b>

**Tabla 4.35.** Resultados del proceso de *Clustering*.

En primer lugar, y antes de analizar cada cluster por separado, ocurre en general que todos los clusters que se forman muestran un *nivel de instrucción* muy bajo: *Primario completo* es el más frecuente. Hay que tener en cuenta que los principales niveles de instrucción en la base son los siguientes (muy bajos por cierto):

Nivel de Instrucción	%
Primario completo	52%
Primario incompleto	20%
Secundario incompleto	14%
Ninguno	6%
Secundario completo	4%
<b>Total</b>	<b>97%</b>

**Tabla 4.36.** Distribución del Nivel de Instrucción.

Por lo tanto, es de esperar que en todas las corridas de *clustering*, cada uno de los grupos formados tenga un nivel de instrucción muy bajo. Muy pocas veces aparece como representativo de un cluster el nivel secundario incompleto. Aún así entre primario completo y secundario incompleto hay un mínimo paso.

#### 4.2.4.2.1 Gráficos de Barras para el análisis de clusters

A través de los gráficos de barras [Figura 4.14] se ayuda a comprender como se distribuyen las variables de los atributos entre los distintos clusters y medir el nivel de significancia de los mismos. Se espera que si todos los atributos fueran irrelevantes, se cumpliera la distribución en cada cluster de los atributos según la proporción 41%, 16%, 36% y 7%.

Si bien en los atributos *Nivel de Instrucción* y *Última Residencia* la distribución en los clusters es irrelevante, ya que se cumple la proporción 41% rojo (cluster 0) 16% gris (cluster 1) 36% turquesa (cluster 2) 7% azul (cluster 3), en los otros atributos se pueden encontrar interacciones significativas

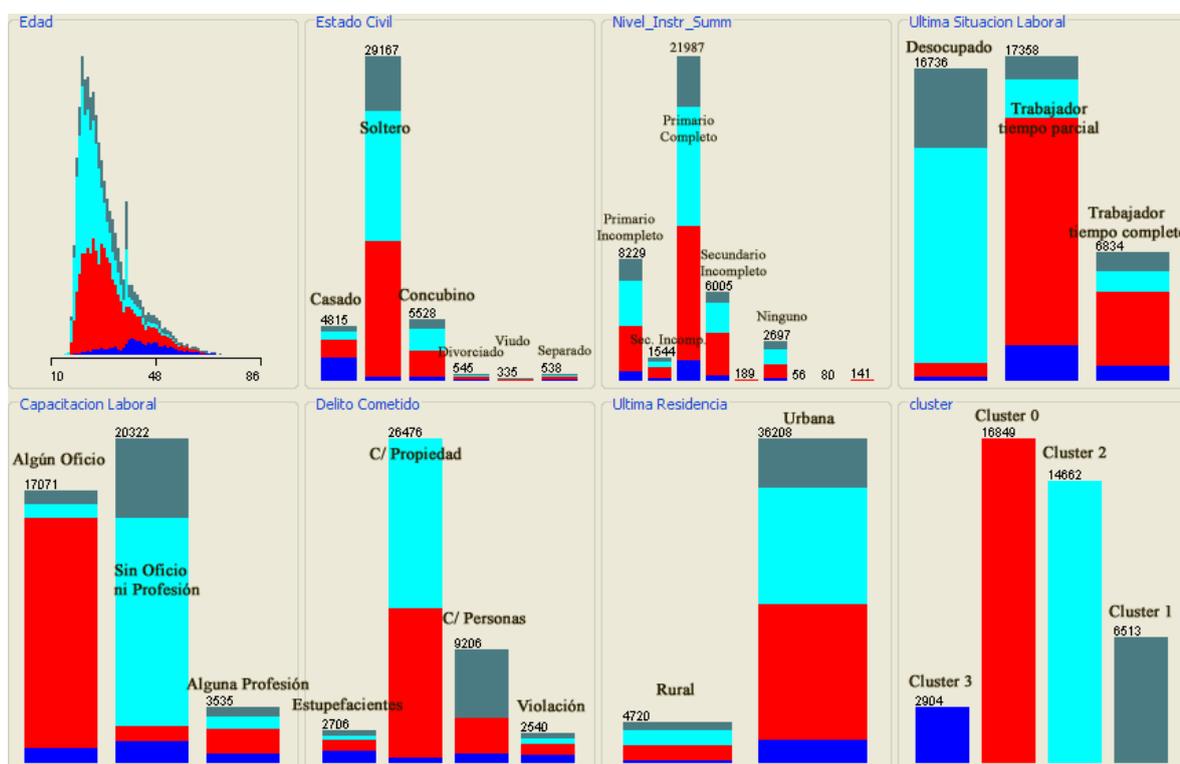
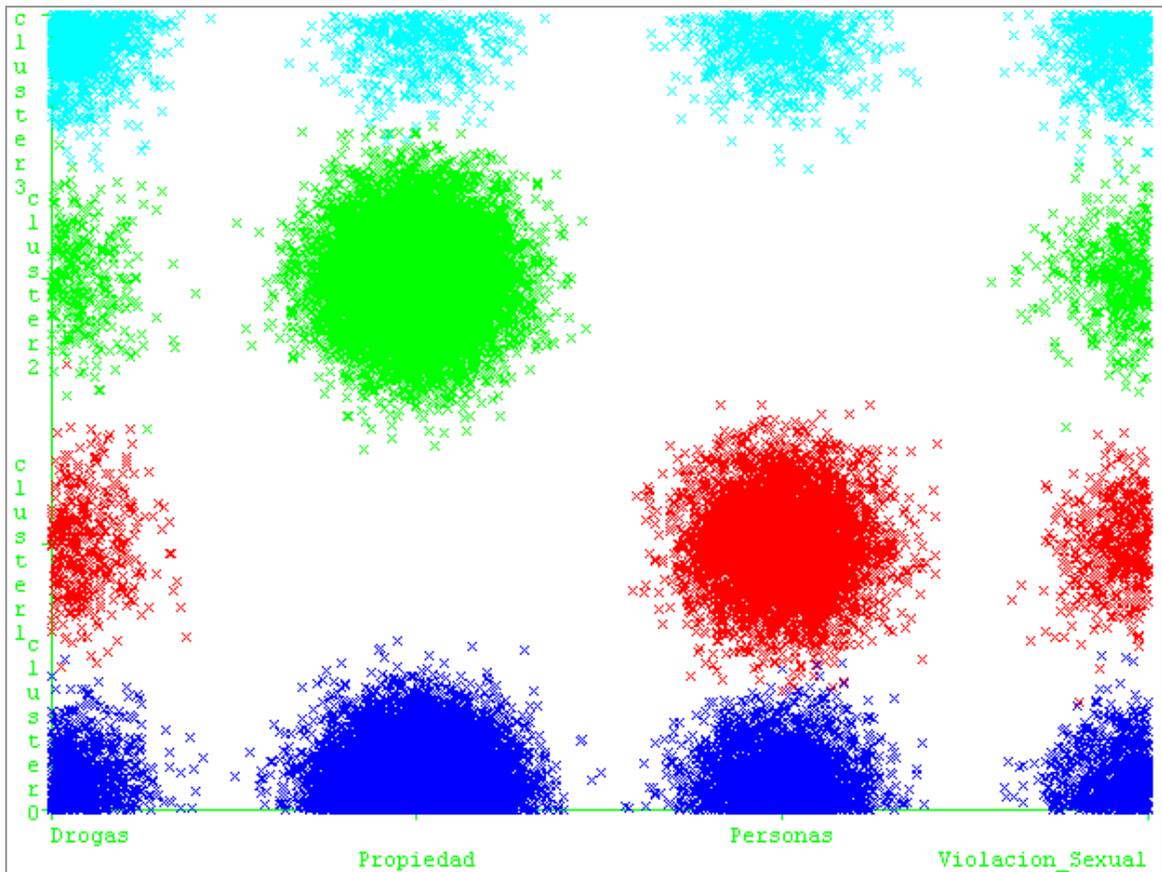


Figura 4.14. Gráfico de Barras. Distribución de las variables de los atributos en los clusters.

Los atributos en donde se aprecia buena interacción entre los clusters y las variables son los de *delito cometido*, *capacitación laboral* y *última situación laboral*. Se profundizará en estos atributos a continuación mediante la utilización de gráficos de dispersión.

**4.2.4.2.2 Gráficos de Dispersión para el análisis de clusters**

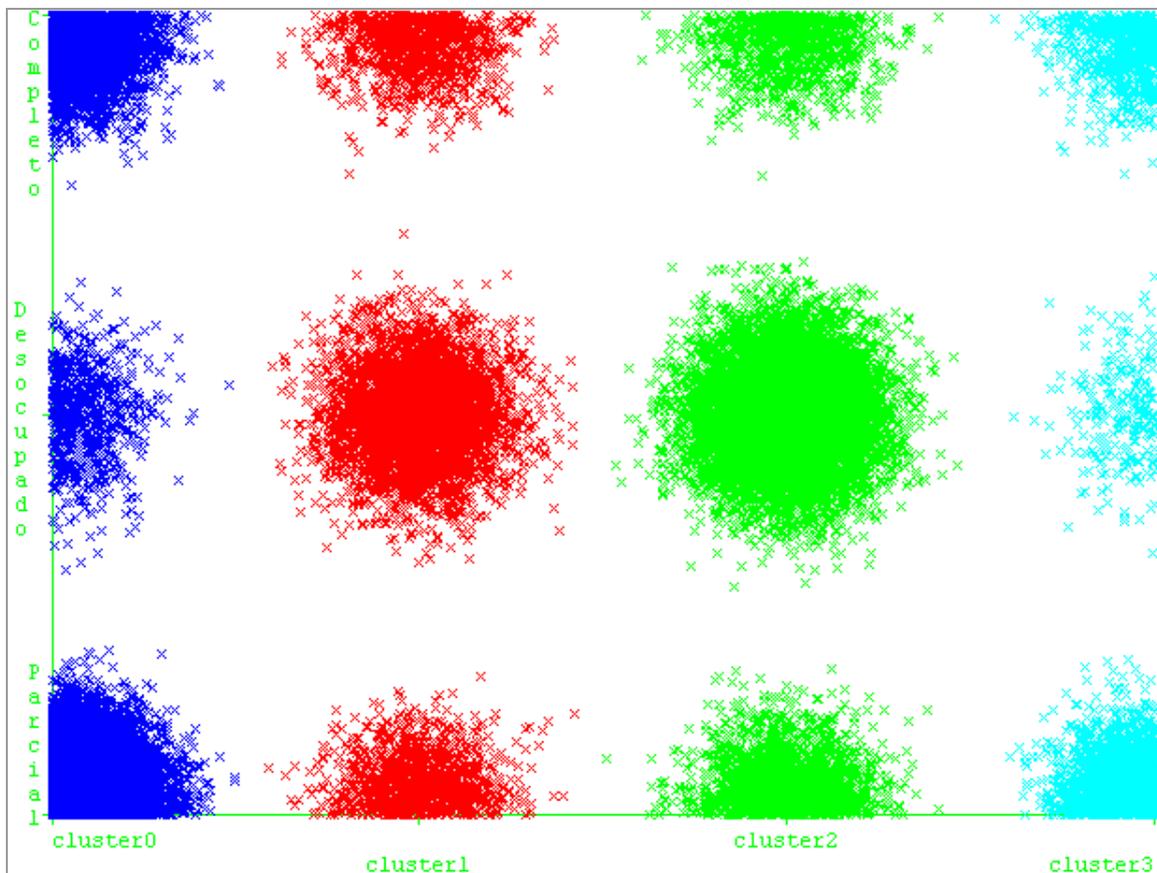
En la figura 4.15 se observa que mientras que los delitos contra la propiedad están fuertemente concentrados en los clusters 0 (72%) y 2 (95%), el cluster 1 presenta una gran concentración de individuos que cometieron delito contra las personas (95%). Entre paréntesis figura la frecuencia de ocurrencia porcentual de cada variable de los atributos dentro de cada cluster. Con respecto al cluster 3, los delitos se encuentran distribuidos más homogéneamente, siendo los delitos contra la integridad sexual (23%), estupefacientes (35%) y personas (26%) los que predominan en el mismo. Por lo tanto, ya se puede tener una primera aproximación a los delitos que caracterizarían a cada grupo.



**Figura 4.15.** Distribución de los clusters según *Delito Cometido*

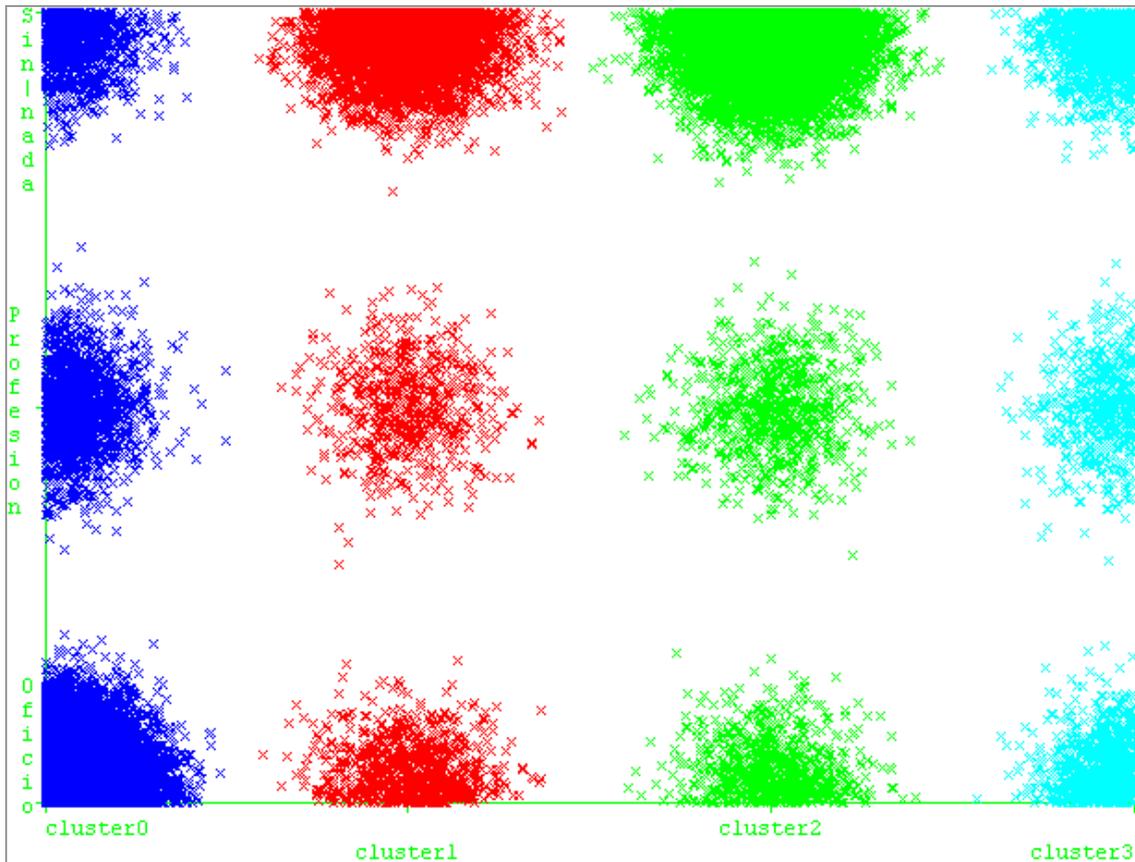
En lo que respecta al atributo *última situación laboral* [Figura 4.16], se observa que los cluster 1 y cluster 2 están representados en su mayoría por la situación de desocupado, con un 65% y 79% respectivamente. Si bien estos dos grupos la situación de desocupado presenta una gran concentración, existen instancias que presentan en una misma proporción la situación de trabajadores de tiempo parcial y completo.

En cuanto a los clusters 0 y 3, están mayormente caracterizados por trabajadores de tiempo parcial (72% y 65% respectivamente) y en menor proporción, de tiempo completo (23% y 27%).



**Figura 4.16.** Distribución de los clusters según *Última Situación Laboral*

Con respecto a la capacitación laboral, el cluster 1 y el cluster 2 presentan una distribución similar [Figura 4.17], con una alta concentración en *sin oficio ni profesión* (77% y 89%), seguido por *oficio* y luego *profesión*. En contraposición, el cluster 0 está mayormente polarizado hacia *oficio* (85%), con una baja población de individuos *sin oficio ni profesión* (9%) y *profesión* (6%). Finalmente, como se observa en la figura inferior, el cluster 3 se encuentra homogéneamente distribuido para este atributo.



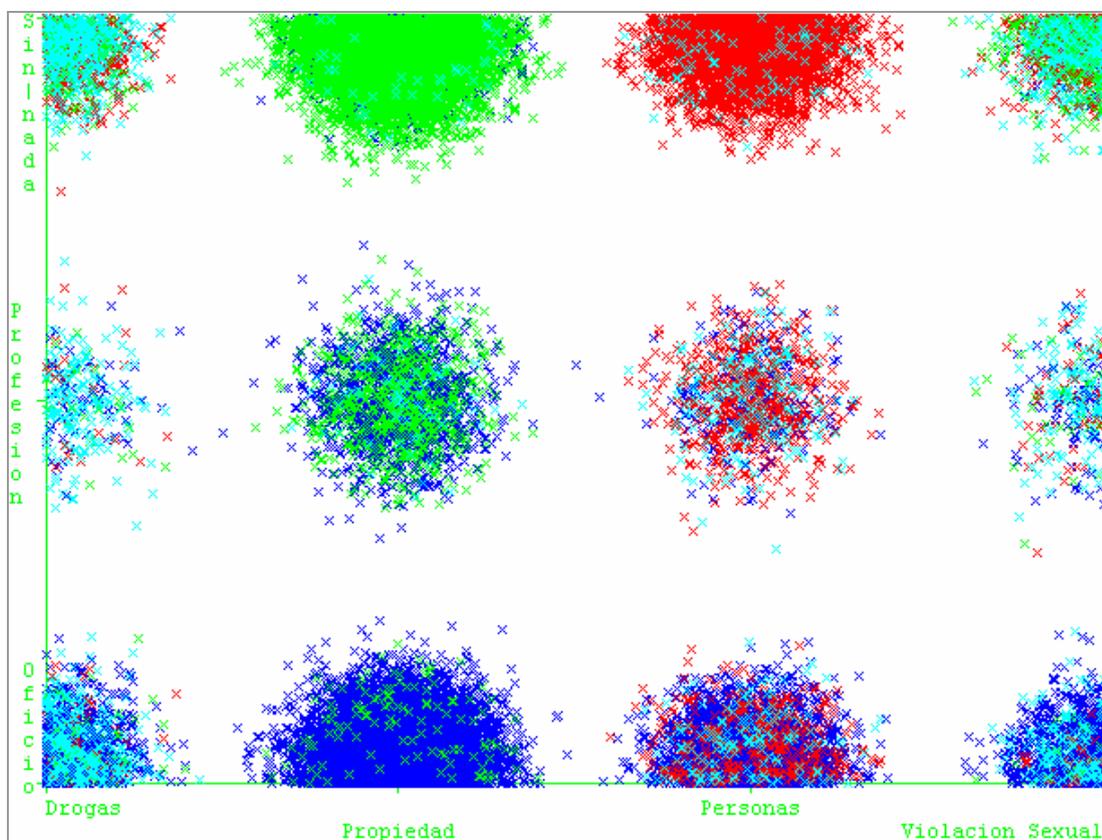
**Figura 4.17.** Distribución de los clusters según *Capacitación Laboral*

A continuación se analizan, también con gráficos de dispersión, las interacciones entre 2 tipos de atributos y los clusters. Con este análisis se busca ir teniendo una mejor interpretación de los grupos formados para poder luego dar una caracterización más completa de los mismos.

Siguiendo con la serie de colores de los gráficos anteriores (cluster 0= azul, cluster 1= rojo, cluster 2= verde, cluster 3= turquesa), se observa una fuerte interacción entre *delito contra la propiedad, sin oficio ni profesión* y cluster 2 (verde) [Figura 4.18]. Lo mismo ocurre con el cluster 0 (azul), *delito contra la propiedad*, aunque en este caso aparece *oficio* para el atributo capacitación laboral. Por lo tanto, ya podríamos diferenciar a estos dos clusters según el nivel de capacitación laboral. Mientras que en ambos, los presos delinquieron contra la propiedad, en el cluster 2 no poseían un oficio ni una profesión al momento de la detención, mientras que en cluster 0 tenían algún oficio.

Otra interrelación importante es la que se da entre *delito contra las personas, no tiene oficio ni profesión* y cluster 1 (rojo). Confirma esto último, que el cluster 1 estaría caracterizado por delincuentes que delinquieron contra las personas. Existe, de todas formas, una pequeña interacción entre cluster 0 (azul), *delito contra las personas y oficio*.

Con respecto al cluster 3 (turquesa), se observa que la interacción de *delito de estupefacientes y violación sexual* distribuido homogéneamente en *oficio y sin oficio ni profesión*, y en una pequeña proporción en los que tienen alguna *profesión*.

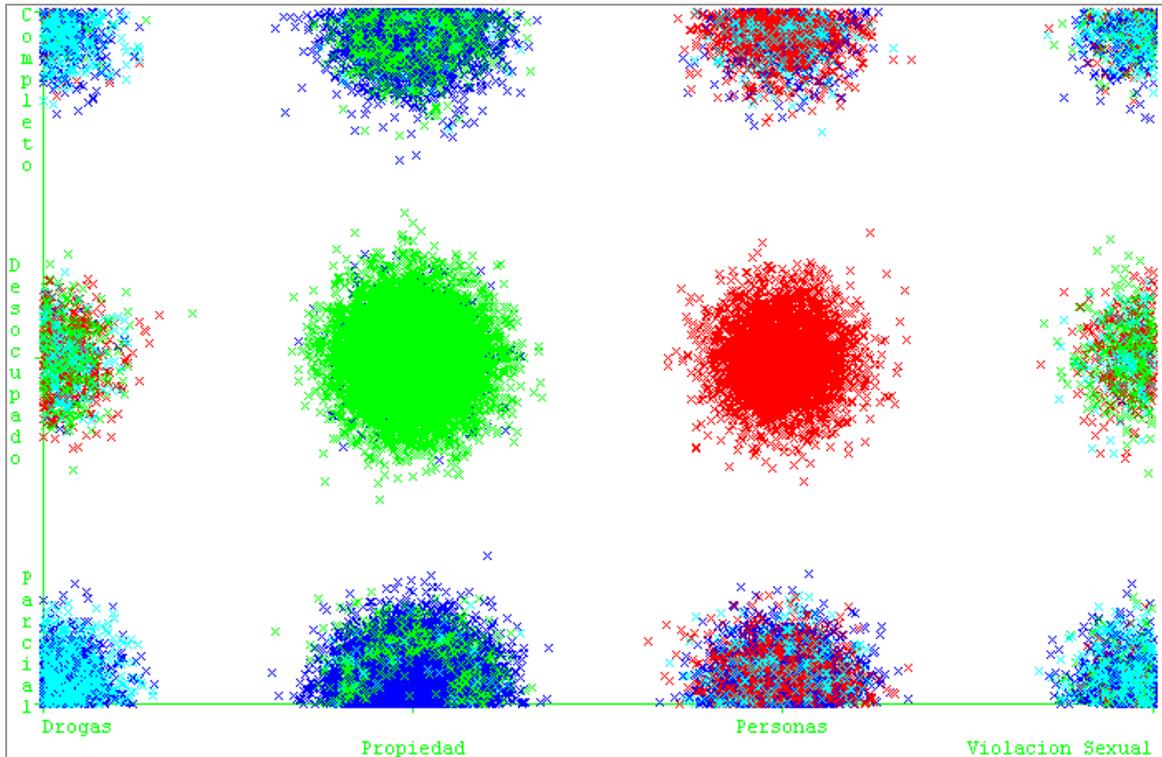


**Figura 4.18.** Interrelación *Delito Cometido- Capacitación Laboral-Clusters*

En la figura 4.19 se destaca la importante relación que hay entre *delito contra la propiedad, desocupado*, y cluster 2 (verde). El cluster 0 (azul) está interrelacionado en mayor medida con *trabajo de tiempo completo y de tiempo parcial*.

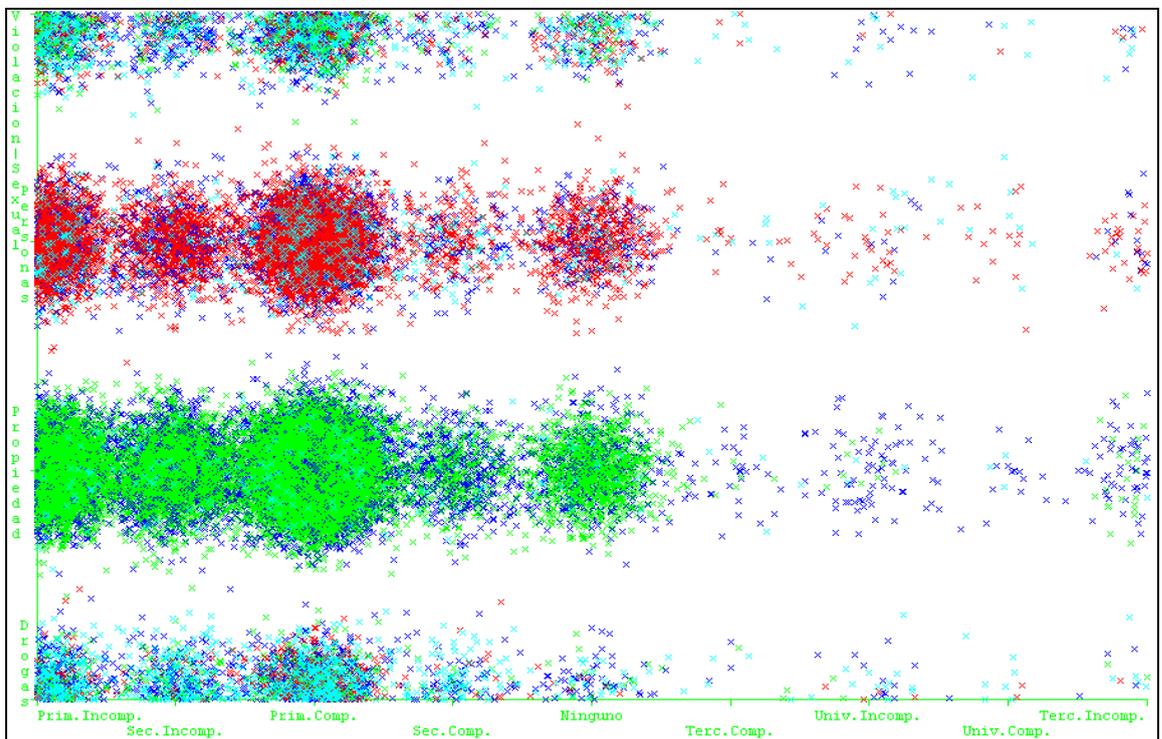
Se puede observar interacción entre *delitos contra las personas, desocupado* y cluster 2 (rojo). Aún así, también se observan algunas interrelaciones entre este último cluster, *personas, y trabajo de tiempo parcial y de tiempo completo*.

Con respecto al cluster 3 (turquesa), se observa interacción de *delito de estupefacientes y violación sexual* distribuido mayormente en *trabajo de tiempo parcial* y en menor proporción *trabajo de tiempo completo*.



**Figura 4.19.** Interrelación *Delito Cometido- Última Situación Laboral-Clusters*

Es importante mostrar mediante la gráfica de la figura 4.20, que la mayoría de los delitos son cometidos por personas con un nivel de instrucción bajo. Existen muy pocos casos en los que los que hayan cometido delitos sean personas con niveles de educación superior al secundario completo.



**Figura 4.20.** Interrelación *Delito Cometido- Nivel de Instrucción-Clusters*

#### 4.2.4.2.2.3 Interpretación de los clusters

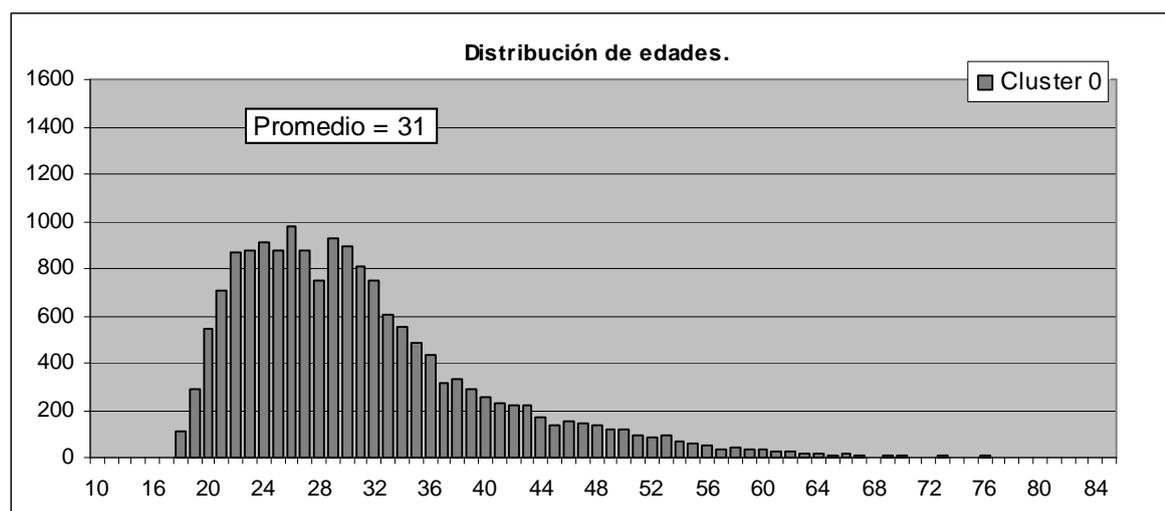
En base a la información que surge tanto del análisis de los gráficos anteriores como de la investigación en el tema y consulta con especialistas, se da una primera interpretación de cada cluster formado.

Es muy conveniente destacar, según lo hablado con especialistas, que claramente para poder realizar un análisis más profundo y exacto, habría que tener acceso a la historia individual de cada uno, como lo es conocer el origen familiar y cultural. Sin embargo, se cree que mediante el presente estudio se puede dar una muy buena aproximación.

A continuación se presenta el análisis de cada uno de los clusters formados.

#### Interpretación Cluster 0

- **Tamaño del cluster:** 16849 (41% del total de la muestra).
- **Edad Promedio** = 31 años.
- **Delito Cometido** = Delitos contra la propiedad (90% de Robo y/o Tentativa de robo).
- **Estado Civil** = Soltero (73% del cluster).
- **Nivel de Instrucción** = Primario Completo (54%, 40% restante entre ninguno y Sec. Incompleto).
- **Última Situación Laboral** = Trabajo tiempo Parcial (72% del cluster, 23% de tiempo completo).
- **Capacitación Laboral** = Poseían algún oficio (85% del cluster).
- **Último lugar de residencia** = Urbana (90% del cluster).



**Figura 4.21.** Distribución de edades Cluster 0.

Se observa un cluster de solteros en cual el promedio de edad es de 31 años [Figura 4.21]. El nivel de instrucción que predomina es de primario completo, aunque la distribución del nivel de instrucción del resto de la población del cluster es muy pobre también. Son personas, en su mayoría, trabajadoras de tiempo parcial y que poseían algún oficio previo al ingreso al establecimiento penitenciario. Aún así, delinquieron **contra la propiedad**, generalmente delitos de robo y/o tentativa de robo.

Para que se cometa el delito de robo concurren diferentes factores y múltiples circunstancias individuales. En este caso, lo más probable es que se tratándose de personas que teniendo algún trabajo parcial o también llamadas “changas” puedan juntar algo de plata para cubrir algunas de sus necesidades básicas. El resto, lo obtienen robando. Aún siendo solteros en su mayoría, muchas veces tienen varios hijos, lo que acentúa la necesidad de salir a robar. Pero no se puede dejar de advertir, que no se debe únicamente a una “necesidad marcada”, sino que el robo probablemente se deba a otros factores como la exclusión del trabajo formal, la exclusión social, de salud, de educación, de vivienda, de participación social, sumado a haber crecido en medios hostiles en que los delitos de robo y violencia no les eran ajeno. Estos factores, muchas veces se traducen en una violación a la propiedad ajena.

A la edad promedio de 31 años, muchos de estos detenidos son reincidentes. Heredan la cultura del robo, y no pueden desprenderse de ella, lo que los lleva a delinquir aún cuando tienen algún tipo de oficio y trabajo. A su vez, ocurre que trabajan generalmente en negro, por lo cual estos trabajos de tiempo parcial pueden llegar a ser muy inestables sufriendo despidos prematuros e imprevistos.

#### **Posible solución al problema:**

Al ser gente que tiene una educación pobre, sin duda que habría que hacerse hincapié en este sentido. No es una característica importante en este grupo el desempleo, ya que solo un 4% se encuentra en el estado de desocupación total. Una forma de fomentar la educación entre familias de bajos recursos es mediante la implementación de un eficaz sistema de becas (ocurre en Brasil) que premie el esfuerzo y calificaciones y que permita que las personas no dejen sus estudios secundarios para salir a trabajar.

Al mismo tiempo se los puede instruir en algún otro oficio para que tengan distintas salidas y estén el mayor tiempo del día ocupados. Es fundamental que se les de a este grupo de personas seguridad en sus trabajos, para que no tengan que recurrir al robo como única salida al desempleo.

### Interpretación Cluster1

- **Tamaño del cluster:** 6513 (16% del total de la muestra).
- **Edad Promedio** = 34 años.
- **Delito Cometido** = Delitos contra las Personas (86% del cluster).

Delitos contra las Personas	%
Homicidios dolosos	63%
Otros delitos contra las personas	17%
Homicidios dolosos (tentativa.)	7%
Homicidios Culposos	6%
Lesiones Dolosas	4%
Lesiones Culposas	1%
<b>Total</b>	<b>100%</b>

Tabla 4.37. Apertura de Delitos contra las Personas

- **Nivel de Instrucción** = Primario Completo (54%, 42% restante entre ninguno y Sec. Incompleto).
- **Estado Civil** = Soltero (76%).
- **Última Situación Laboral** = Desocupado (65%).
- **Capacitación Laboral** = No tiene ni oficio ni profesión (77%).
- **Último lugar de residencia** = Urbana (85%).

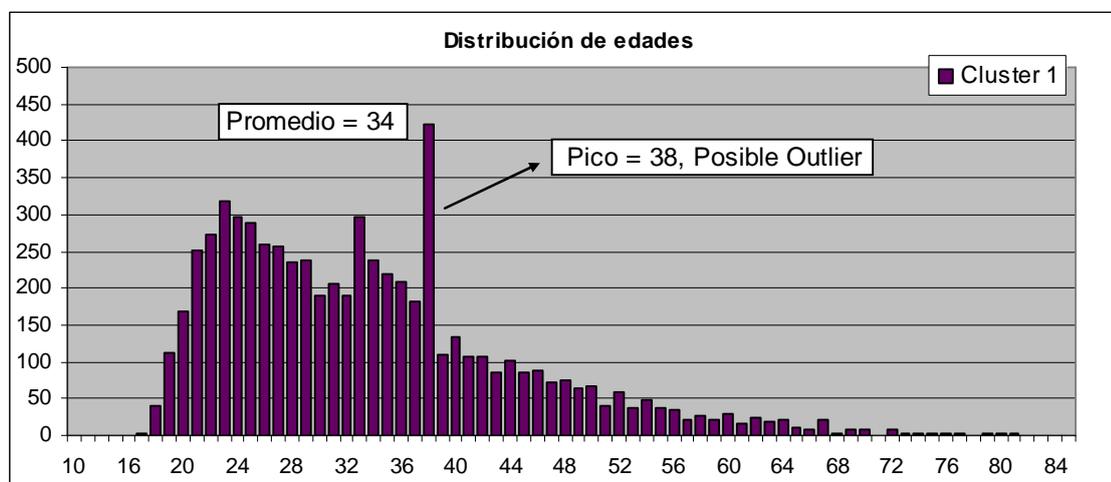


Figura 4.22. Distribución de edades Cluster 1.

Personas de mayor edad que la del cluster 0 (promedio de 34 y pico de 38) [Figura 4.22], con pobre formación educativa y solteros. Totalmente **desocupados y sin oficio ni profesión alguna**.

Se cree que la desesperación del no “hacer nada” (mayoría desocupados) sumado a “nada que perder” (mayoría solteros), conjuntamente con los problemas que estos individuos pueden padecer, los llevo a cometer algún **delito contra las personas**. En un

ámbito de violencia social como el que se vive hoy en día, este tipo de personas pierde el valor de la vida, y día tras día tienen menos esperanzas de conseguir algún trabajo debido a la edad y a la nula formación profesional. Las drogas, el alcohol y el fácil acceso a las armas, son actores principales en este tipo de grupo. Podría tratarse, también, de **homicidios en ocasión de robo**.

Hay una muy pequeña proporción de personas mayores a 33 años que se encuentra con un trabajo de tiempo completo/parcial y algún oficio o profesión. Aún así, delinquen contra las personas. Sin lugar a duda, que en estos delitos penales como el homicidio y las lesiones, siempre juega el factor psicológico del agresor, y los altos niveles de violencia que se perciben en las escuelas, vía pública, eventos deportivos y desde luego de la violencia familiar. O sea, la falta de trabajo no es siempre condición necesaria para que una persona cometa un homicidio.

En principio, al presente cluster se lo identificaría con las personas que cometen delitos contra las personas de cualquier tipo (mayormente homicidios dolosos), asociados con la exclusión total a un régimen laboral.

#### **Posibles soluciones al problema:**

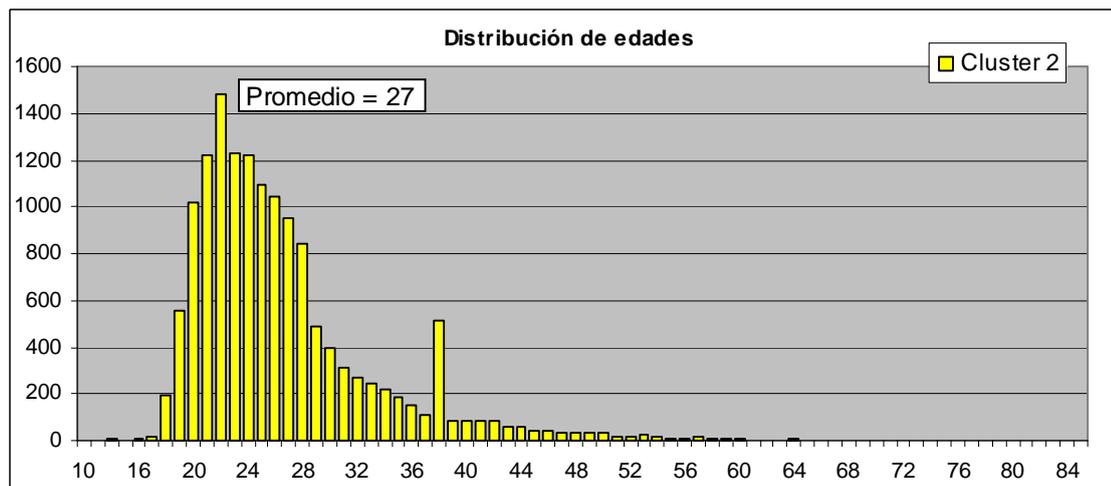
Programas de inclusión, reinserción y capacitación laboral para gente mayor. Es necesario que a este grupo de personas se les puedan brindar una oportunidad de obtener un trabajo digno.

A su vez, habría que prestarle más atención a los casos de violencia doméstica, laboral, de vecindad, etc. De esta forma, se quiere tratar de advertir la violencia presente y ofrecer contención y ayuda a los agresores antes que la violencia crezca hasta convertirse en homicidio.

#### **Interpretación Cluster 2**

- **Tamaño del cluster:** 14662 (36% del total de la muestra).
- **Edad Promedio** = 27 años.
- **Delito Cometido** = Delitos contra la propiedad (92% de Robo y/o Tentativa de robo).
- **Estado Civil** = Soltero (80% del cluster).
- **Nivel de Instrucción** = Primario Completo (55%, 42% restante entre ninguno y Sec. Incompleto).
- **Última Situación Laboral** = Desocupado (79% del cluster, 14% trabajo parcial).
- **Capacitación Laboral** = No tenía oficio ni profesión (89% del cluster).

- **Último lugar de residencia** = Urbana (88% del cluster).



**Figura 4.23.** Distribución de edades Cluster 2.

Este cluster tiene la letal combinación de agrupar a personas jóvenes de entre 20 y 30 años, en su mayoría solteras (80%), con un nivel de instrucción bajísimo (sólo el 55% concluyó los estudios primarios) junto a una muy alta tasa de desempleo (79% desocupado y 14 % trabajo de tiempo parcial). Es decir, se está hablando de que un 93% de estas personas jóvenes no posee un trabajo formal de jornada completa. Esto último, sumado a la carencia de un oficio u profesión, coloca a esta población como un gran riesgo: sin fuente de ingresos y dificultades para reinserirse laboralmente no resulta extraño que el delito que cometen es contra la propiedad.

A su vez, la muy pequeña proporción de estos jóvenes individuos que tienen algún oficio o profesión, lamentablemente no poseía trabajo al momento de la detención, lo que podría ser la causante que los haya llevado a delinquir contra la propiedad.

Este cluster estaría identificado con el estado de sedentarismo de los jóvenes que se encuentran apartados del común de la gente, y no tienen acceso en primer lugar a una educación formal que les ayude a abrirles las puertas del ámbito laboral.

Importante tener en cuenta que es uno de los clusters más grandes junto con el cluster 0 y más preocupantes aún por tratarse de gente joven.

### **Posible solución:**

Este grupo de gente todavía está a tiempo de cambiar y tener una vida digna fuera de las drogas y cárceles. Eficientes políticas de estado, como programas intensivos de educación formal y capacitación laboral es una posible solución para esta futura generación del país.

**Interpretación Cluster 3**

- **Tamaño del cluster:** 2904 (7% del total de la muestra).
- **Edad Promedio** = 43 años
- **Delito Cometido** = Estupefacientes + Integridad Sexual (60%)

Delitos Cometidos Cluster 3	%
Infracción ley N° 23.737 (estupefacientes)	35%
Delitos contra las personas	26%
Delitos contra la integridad sexual y el Honor	23%
Delitos contra la propiedad	15%
Total	100%

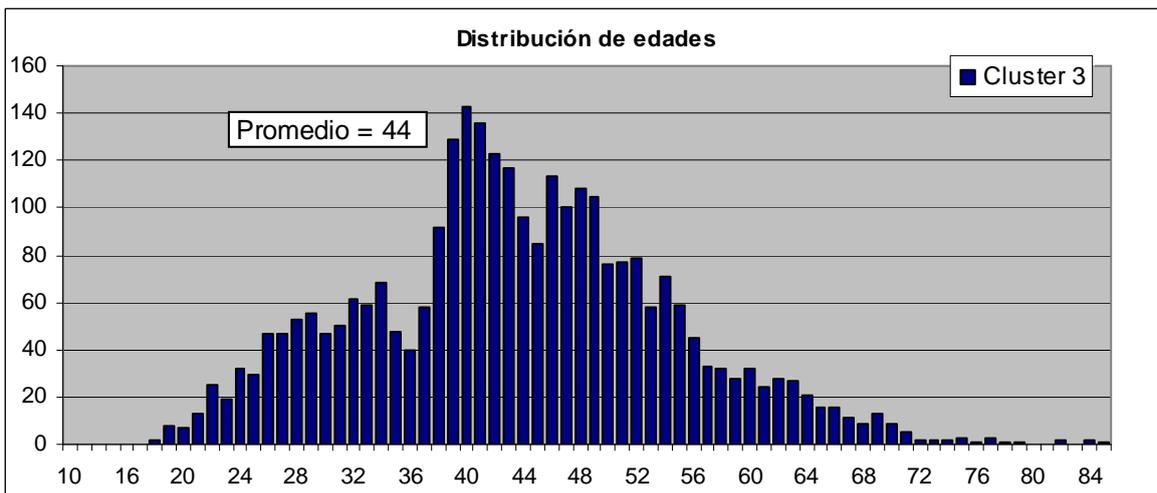
**Tabla 4.38.** Delitos Cluster 3.

- **Estado Civil** = Casados (72% del cluster).
- **Nivel de Instrucción** = Primario Completo (47%, 50% restante entre ninguno y Sec. Incompleto, 3% con nivel alto de educación).
- **Última Situación Laboral** = Trabajo Tiempo parcial (65% del cluster, 27% trabajo tiempo completo).
- **Capacitación Laboral** = No tiene oficio ni Profesión (47% del cluster).

Capacitación Laboral	%
No tiene ni oficio ni profesión	47%
Tiene algún oficio	34%
Tiene alguna profesión	20%
Total	100%

**Tabla 4.39.** Capacitación Laboral Cluster 3.

- **Último lugar de residencia** = Urbana (89% del cluster).



**Figura 4.24.** Distribución de edades Cluster 3.

Se trata del cluster con menor cantidad de instancias y al mismo tiempo el más difícil de analizar debido a sus distintas variantes y reglas.

Sin duda, no son los delitos contra la propiedad los que marcan a este cluster. Los delitos que predominan son el de violación, tráfico de drogas y contra las personas (totalizan el 85%).

Está representado por gente de mayor edad que los anteriores, con una media de 43 años, en donde la mayoría (73%) esta casada y nuevamente el nivel de instrucción es muy bajo. Estos individuos poseían algún trabajo de tiempo parcial (56%) y de tiempo completo (27%) y su capacitación laboral se encuentra dividida entre las 3 categorías. En este cluster se pueden observar distintas características entre los detenidos, siendo el nivel de capacitación laboral el punto de partida en el árbol de decisión.

- Tienen oficio/profesión y trabajan parcialmente o completamente, mayores a 33/38 años. Sin embargo delinquen contra las personas, violan y trafican drogas. Generalmente están casados o viven en concubinato.
- Los que no tienen oficio ni profesión, y delinquen contra la integridad sexual y el honor y estupefacientes. En este caso, también generalmente están casados.

Como se comenta más arriba, es difícil con esta información establecer algún patrón en este cluster. Se debería tener acceso, por ejemplo, a más información relacionada con los orígenes de estas personas. Sin embargo, al realizar las corridas de inducción, se encuentran reglas interesantes.

Se observa que los que cometen delitos contra la integridad sexual, en su amplia mayoría violación, generalmente son personas casadas, mayores a los 32 años y que poseían trabajo y oficio en muchos casos. Se trata de personas enfermas, muchas veces con desórdenes psicológicos y físicos, con personalidades psicópatas y perversas. Estos agresores tienen patrones de comportamiento y de personalidad diferentes a cualquier otro autor de otros delitos. Es difícil, por lo tanto, lograr prevenir este tipo de delito, ya que suele ser muy complicado detectar a un potencial violador.

*No se puede dejar de aclarar que la mayoría de los casos de abuso sexual, especialmente los que tienen como víctima a niños y niñas, ocurren dentro del hogar, y los victimarios suelen ser los padres, padrastros, abuelos o tíos. Hace falta derribar el mito de que el peligro está afuera, ya que si bien hay casos, por ejemplo, de menores violados por desconocidos que los sorprenden en la vía pública, la mayoría de los casos son intrafamiliares [Lalurette, 2008].*

Para estos casos es necesario, que el tema de los abusos sexuales deje de ser un tema tabú, que los habitantes tomemos conciencia de este flagelo que existe en la sociedad y que en la mayoría de los casos se da dentro de las familias. Es necesario acentuar la prevención, y en el tratamiento de los agresores y víctimas.

En cuanto a los que participan en delitos de estupefacientes, que conforman el 35 % de este cluster, nos encontramos con gente que siempre tiene algún tipo de trabajo y en muchas ocasiones oficio o profesión. El delito de los estupefacientes es diferente a los demás delitos penales. Desde ya no es lo mismo un narcotraficante que un poseedor de drogas para consumo personal, por lo cual merecerán una diversa atención. En el último caso, el de consumo personal, muchas veces se trata de gente que tiene una adicción, que atenta contra la posibilidad que tiene un sujeto normal de comprender que está quebrantando la ley. Es por ello, que en algunos países, lentamente se está avanzando en la despenalización de este delito.

#### **4.2.4.2.3 Selección de Atributos**

*Rapid-i* cuenta con una serie de métodos para realizar una preselección de los atributos que serán utilizados posteriormente en los algoritmos de inducción o clasificación. Tal como se comentó al comienzo del presente apartado, se utilizan métodos de filtros y envolventes (*wrappers*).

##### **4.2.4.2.3.1 Filtros**

Complementando lo dicho en el estado de la cuestión, se basan en la idea de dar una calidad de los atributos. Se suelen aplicar haciendo un ranking de atributos.

A continuación se muestran los métodos de filtros aplicados con sus respectivos resultados en forma de ranking de atributos.

<b>GainRatioAttributeEval</b>	
<b>Attribute</b>	<b>Weight</b>
Capacitacion Laboral	0.405
Delito Cometido	0.366
Ultima Situacion Laboral	0.291
Estado Civil	0.105
Edad	0.054
Nivel_Instr_Summ	0.007
Ultima Residencia	0.004

**Tabla 4.40.** Ranking de Atributos

<b>InfoGainAttributeEval</b>	
<b>Attribute</b>	<b>Weight</b>
Capacitacion Laboral	0.540
Delito Cometido	0.512
Ultima Situacion Laboral	0.432
Edad	0.168
Estado Civil	0.139
Nivel_Instr_Summ	0.012
Ultima Residencia	0.002

**Tabla 4.41.** Ranking de Atributos

ReliefFAttributeEval	
Attribute	Weight
Delito Cometido	0.365
Capacitacion Laboral	0.352
Ultima Situacion Laboral	0.309
Estado Civil	0.070
Edad	0.042
Ultima Residencia	-0.001
Nivel_Instr_Summ	-0.005

**Tabla 4.42.** Ranking de Atributos

ChiSquaredAttributeEval	
Attribute	Weight
Delito Cometido	28429.192
Capacitacion Laboral	26250.296
Ultima Situacion Laboral	21184.971
Estado Civil	11536.309
Edad	10607.634
Nivel_Instr_Summ	748.152
Ultima Residencia	135.467

**Tabla 4.43.** Ranking de Atributos

Se puede ver que mediante los 4 métodos de selección de atributos se confirma lo dicho anteriormente sobre la poca relevancia en los clusters que tienen los atributos *Última Residencia* y *Nivel de Instrucción*. Sin embargo, cada uno de los métodos sugirió que no se descartara a ninguno de ellos.

#### **4.2.4.2.3.2 Wrappers**

Este método seleccionan los atributos en función de la calidad del modelo de minería de datos asociado a los atributos utilizados.

Se utiliza el *WrapperSubsetEval* y como clasificador al algoritmo *J48*. El resultado del mismo sugirió la exclusión del atributo *Nivel de Instrucción*. Sin embargo, se cree conveniente continuar con la clasificación utilizando a dicho atributo debido a que ninguno de los métodos por filtros lo descartó.

#### **4.2.4.2.4 Clasificación**

Para la clasificación se utilizan los algoritmos **C4.5** y **CHAID**. Se esperaba que cada algoritmo en particular arrojara reglas distintas. Sin embargo se obtienen reglas muy similares en los árboles de decisión, por no decir iguales, al realizar inducción sobre el atributo *cluster* obtenido en el proceso de *clustering*. Esto demuestra la consistencia de los datos en ambos modelos.

##### **4.2.4.2.4.1 Clasificación a través del C4.5**

Se estableció un nivel de confianza para la poda del árbol de decisión del 50% y se fijó en 2 el mínimo de instancias por hoja.

El resultado obtenido sobre un total de 40,928 registros agrupados en los 4 clusters es el siguiente:

- Instancias clasificadas correctamente: 40,913 (99.96%).
- Instancias clasificadas incorrectamente 15 (0.04%).

El árbol resultante tiene un tamaño de 392 elementos y 276 hojas y la matriz de confusión obtenida a través del *Rapid-i* queda constituida de la siguiente manera:

	true cluster3	true cluster0	true cluster2	true cluster1	class precision
pred. cluster3	2,902	2	1	7	99.66%
pred. cluster0	-	16,847	1	2	99.98%
pred. cluster2	1	-	14,660	-	99.99%
pred. cluster1	1	-	-	6,504	99.98%
class recall	99.93%	99.99%	99.99%	99.86%	

**Tabla 4.43.** Matriz de confusión obtenida a través del algoritmo C4.5.

Si la clasificación hubiera sido perfecta se esperaría encontrar únicamente elementos en la diagonal. En otras palabras, la matriz de confusión permite entender cual es el error que está cometiendo el algoritmo al intentar clasificar a todos los registros.

Por otro lado se testea al modelo mediante un set de criterios de selección, a través del llamado *Performance Vector*. Los resultados del mismo para el algoritmo C4.5 se muestran a continuación:

Criterio	Valor
accuracy:	99.96%
classification_error:	0.04%
absolute_error:	0.001 +/- 0.017
correlation:	0.998
squared_correlation:	0.997

**Tabla 4.44.** *Performance Vector* del algoritmo C4.5.

#### **4.2.4.2.4.2 Clasificación a través del CHAID**

Se corrió el algoritmo CHAID con un nivel de confianza máximo del 50% para realizar la poda del árbol de decisión y se estableció en 2 el mínimo tamaño de todas las hojas.

El resultado obtenido sobre un total de 40,928 registros agrupados en los 4 clusters es el siguiente:

- Instancias clasificadas correctamente: 40,683 (99.4%).
- Instancias clasificadas incorrectamente 245 (0.6%).

La matriz de confusión obtenida queda constituida de la siguiente manera:

	true cluster3	true cluster0	true cluster2	true cluster1	class precision
pred. cluster3	2,750	3	1	49	98.11%
pred. cluster0	27	16,832	24	-	99.70%
pred. cluster2	39	14	14,637	-	99.64%
pred. cluster1	88	-	-	6,464	98.66%
class recall	94.70%	99.90%	99.83%	99.25%	

**Tabla 4.45.** Matriz de confusión obtenida a través del algoritmo CHAID.

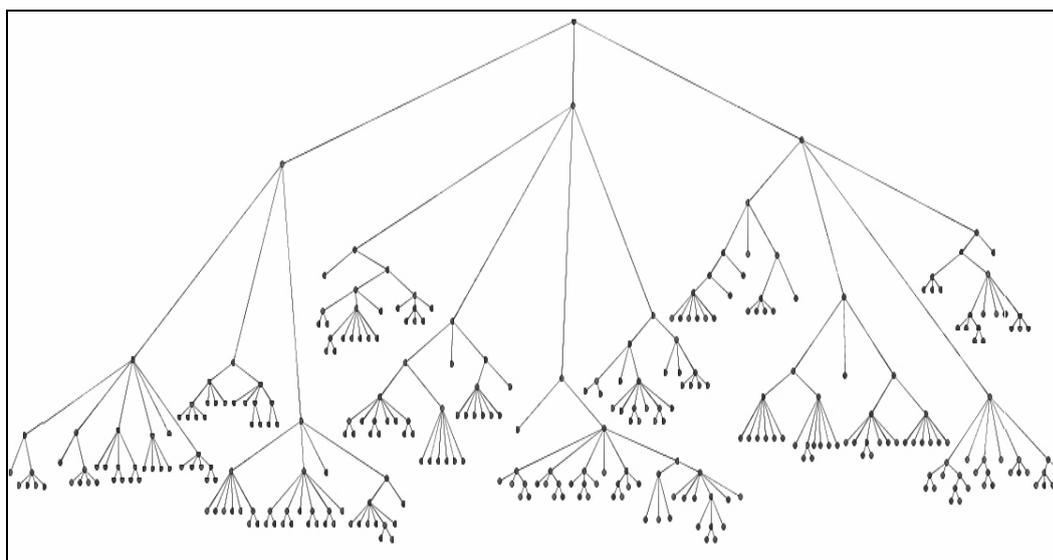
Los resultados del *Performance Vector* para el algoritmo CHAID se muestran a continuación:

Criterio	Valor
accuracy:	99.40%
classification_error:	0.60%
absolute_error:	0.008 +/- 0.066
correlation:	0.974
squared_correlation:	0.949

**Tabla 4.46.** *Performance Vector* del algoritmo CHAID.

#### 4.2.4.2.4.3 Árbol de decisión

Vale recordar que no son las salidas gráficas uno de los puntos fuertes del *Rapid-i*. Más allá de esta desventaja, se puede observar que el árbol de decisión obtenido mediante el algoritmo CHAID es por demás frondoso, aún cuando una de las características del dicho algoritmo es la de realizar la “poda” del mismo.



**Figura 4.25.** Estructura del árbol completo obtenido mediante el algoritmo CHAID.

El nodo principal es el atributo *Capacitación Laboral*. Las ramas correspondientes a *Sin oficio ni profesión* y *Profesión* se dirigen al nodo referido al atributo *Delito Cometido* [Figura 4.26]. La rama correspondiente a *Oficio* se dirige al atributo situación laboral. De esta manera se puede seguir “leyendo” el árbol, partiendo de cada uno de los nodos, pasando por las ramas hasta llegar a las hojas. El árbol que se presenta en la figura 4.26 es un modelo resumido del mismo. Debido al tamaño que exhibe, cada una de las ramas se representa por separado en los anexos 11, 12 y 13.

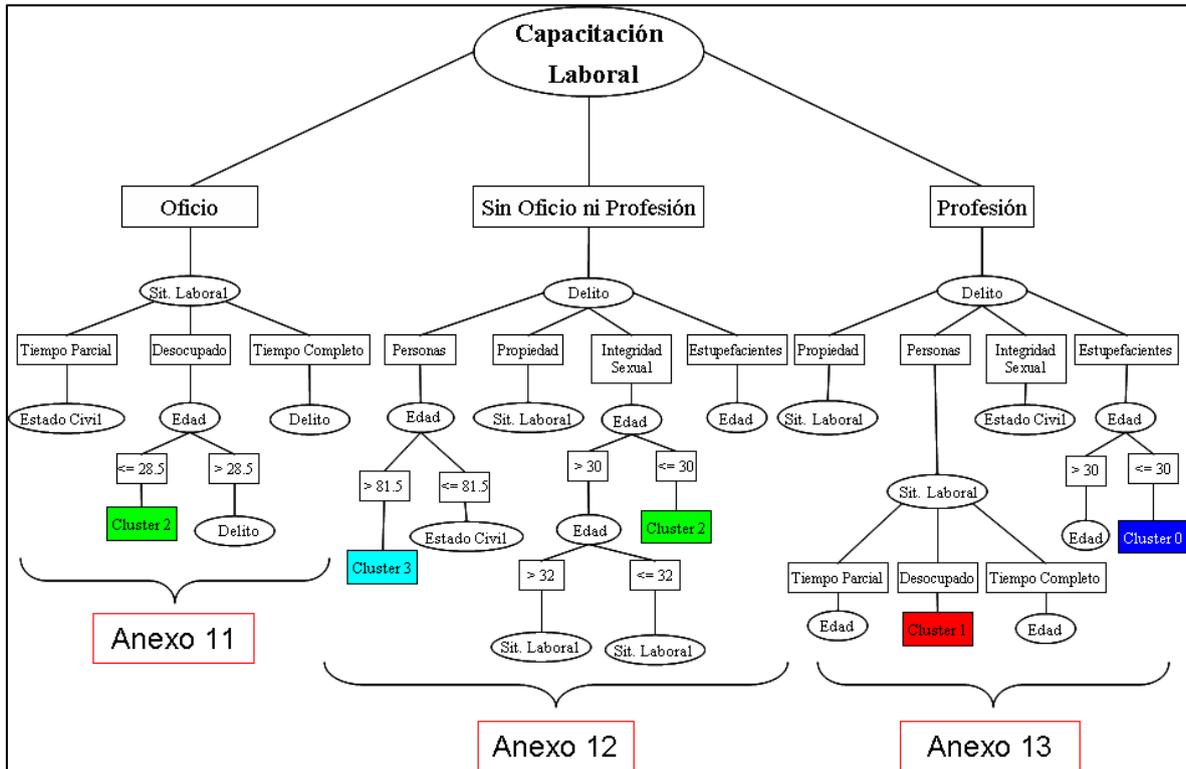


Figura 4.26. Árbol resumido obtenido mediante el algoritmo CHAID.

#### 4.2.4.2.4.4 Reglas de Clasificación

La finalidad del árbol de decisión del presente estudio es la creación de reglas de clasificación que ayuden a explicar y validar la composición de cada cluster formado. Se destaca una vez más, la similitud entre las reglas al utilizarse los algoritmos C4.5 y CHAID.

Las reglas extraídas del árbol de decisión se dividen según cada cluster para tener un panorama más sencillo. Así, por lo tanto, se obtiene que el 20% de la totalidad de las reglas generadas por el árbol clasifican a más del 80% de las instancias [Tabla 4.47].

Clusters	Total de Reglas por cluster	Reglas Seleccionadas	%	Instancias por Cluster	Instancias por Regla	%
Cluster 0	53	6	11%	16,849	13,657	81%
Cluster 1	51	6	12%	6,513	4,826	74%
Cluster 2	17	3	18%	14,662	12,606	86%
Cluster 3	57	18	32%	2,904	2,063	71%
<b>Total</b>	<b>178</b>	<b>33</b>	<b>19%</b>	<b>40,928</b>	<b>33,152</b>	<b>81%</b>

**Tabla 4.47.** Total de reglas vs. Reglas seleccionadas.

A continuación se detallan las principales reglas de clasificación, en orden decreciente según cantidad de instancias que clasifica correctamente, para cada uno de los clusters obtenidos.

### **Reglas Principales Cluster 0**

#### **Regla 1**

Si Capacitación Laboral = Tiene algún oficio  
 Y Última Situación Laboral = Trabajador de tiempo parcial  
 Y Estado Civil = Soltero  
 Entonces Cluster 0 (7516)

#### **Regla 2**

Si Capacitación Laboral = Tiene algún oficio  
 Y Última Situación Laboral = Trabajador de tiempo completo  
 Y Delito Cometido = Delitos contra la propiedad  
 Entonces Cluster 0 (2615)

#### **Regla 3**

Si Capacitación Laboral = Tiene algún oficio  
 Y Última Situación Laboral = Trabajador de tiempo parcial  
 Y Estado Civil = Concubino  
 Entonces Cluster 0 (1314)

#### **Regla 4**

Si Capacitación Laboral = Tiene algún oficio  
 Y Última Situación Laboral = Trabajador de tiempo parcial  
 Y Estado Civil = Casado  
 Y Delito Cometido = Delitos contra la propiedad  
 Entonces Cluster 0 (696)

#### **Regla 5**

Si Capacitación Laboral = Tiene algún oficio  
 Y Última Situación Laboral = Desocupado  
 Y Edad >29  
 Y Delito Cometido = Delitos contra la propiedad  
 Entonces Cluster 0 (695)

**Regla 6**

Si Capacitación Laboral = Tiene alguna profesión  
Y Delito Cometido = Delitos contra la propiedad  
Y Última Situación Laboral = Trabajador de tiempo parcial  
Y Edad <38  
Entonces Cluster 0 (821)

**Reglas Principales Cluster 1**

**Regla 1**

Si Capacitación Laboral = No Tiene oficio ni profesión  
Y Delito Cometido = Delitos contra las Personas  
Y Edad < 82  
Y Estado Civil = Soltero  
Entonces Cluster 1 (3310)

**Regla 2**

Si Capacitación Laboral = No Tiene oficio ni profesión  
Y Delito Cometido = Delitos contra las Personas  
Y Estado Civil = Concubino  
Y Edad < 39  
Entonces Cluster 1 (514)

**Regla 3**

Si Capacitación Laboral = No Tiene oficio ni profesión  
Y Delito Cometido = Infracción ley N° 23.737  
Y Edad > 30  
Y Última Situación Laboral = Desocupado  
Y Estado Civil = Soltero  
Entonces Cluster 1 (290)

**Regla 4**

Si Capacitación Laboral = Tiene alguna profesión  
Y Delito Cometido = Delito contra las personas  
Y Última Situación Laboral = Trabajador de tiempo completo  
Y Edad < 39  
Entonces Cluster 1 (268)

**Regla 5**

Si Capacitación Laboral = Tiene algún oficio  
Y Última Situación Laboral = Desocupado  
Y Edad >29  
Y Delito Cometido = Delitos contra las Personas  
Entonces Cluster 1 (226)

**Regla 6**

Si Capacitación Laboral = Tiene algún oficio  
Y Última Situación Laboral = Trabajador de tiempo completo  
Y Delito Cometido = Delitos contra las Personas

Y Edad >33  
Y Estado Civil = Soltero  
Entonces Cluster 1 (218)

### **Reglas Principales Cluster 2**

#### **Regla 1**

Si Capacitación Laboral = No tiene ni oficio ni profesión  
Y Delito Cometido = Delitos contra la propiedad  
Y Última Situación Laboral = Desocupado  
Entonces Cluster 2 (9822)

#### **Regla 2**

Si Capacitación Laboral = No tiene ni oficio ni profesión  
Y Delito Cometido = Delitos contra la propiedad  
Y Última Situación Laboral = Trabajador de tiempo parcial  
Y Edad <29  
Entonces Cluster 2 (1961)

#### **Regla 3**

Si Capacitación Laboral = Tiene algún oficio  
Y Última Situación Laboral = Desocupado  
Y Edad <29  
Entonces Cluster 2 (823)

### **Reglas Principales Cluster 3**

#### **Regla 1**

Si Capacitación Laboral = No tiene ni oficio ni profesión  
Y Delito Cometido = Infracción ley N° 23.737 (estupefacientes)  
Y Última Situación Laboral = Trabajador de tiempo parcial  
Entonces Cluster 3 (290)

#### **Regla 2**

Si Capacitación Laboral = No tiene ni oficio ni profesión  
Y Delito Cometido = Delitos contra la integridad sexual y el Honor  
Y Edad > 32  
Y Última Situación Laboral = Trabajador de tiempo parcial  
Entonces Cluster 3 (207)

#### **Regla 3**

Si Capacitación Laboral = Tiene algún oficio  
Y Última Situación Laboral = Trabajador de tiempo parcial  
Y Estado Civil = Casado  
Y Delito Cometido = Infracción ley N° 23.737 (estupefacientes)  
Entonces Cluster 3 (201)

#### **Regla 4**

Si Capacitación Laboral = Tiene algún oficio  
Y Última Situación Laboral = Trabajador de tiempo parcial

Y Estado Civil = Casado  
Y Delito Cometido = Delitos contra las personas  
Y Edad > 38  
Entonces Cluster 3 (200)

**Regla 5**

Si Capacitación Laboral = No tiene ni oficio ni profesión  
Y Delito Cometido = Delito contra la propiedad  
Y Última Situación Laboral = Trabajador de tiempo parcial  
Y Edad > 29  
Y Estado Civil = Casado  
Entonces Cluster 3 (165)

**Regla 6**

Si Capacitación Laboral = Tiene algún oficio  
Y Última Situación Laboral = Trabajador de tiempo completo  
Y Delito Cometido = Delito contra las personas  
Y Edad > 33  
Y Estado Civil = Casado  
Entonces Cluster 3 (137)

**Regla 7**

Si Capacitación Laboral = No tiene ni oficio ni profesión  
Y Delito Cometido = Delito contra las personas  
Y Edad < 82  
Y Estado Civil = Casado  
Y Última Situación Laboral = Trabajador de tiempo parcial  
Entonces Cluster 3 (131)

**Regla 8**

Si Capacitación Laboral = Tiene algún oficio  
Y Última Situación Laboral = Trabajador de tiempo parcial  
Y Estado Civil = Casado  
Y Delito Cometido = Delitos contra la integridad sexual y el Honor  
Y Edad > 38  
Entonces Cluster 3 (119)

**Regla 9**

Si Capacitación Laboral = Tiene alguna profesión  
Y Delito Cometido = Delitos contra la integridad sexual y el Honor  
Y Estado Civil = Concubino  
Y Edad > 37  
Entonces Cluster 3 (100)

**Regla 10**

Si Capacitación Laboral = No tiene ni oficio ni profesión  
Y Delito Cometido = Infracción ley N° 23.737 (estupefacientes)  
Y Edad < 30  
Y Última Situación Laboral = Trabajador de tiempo parcial  
Entonces Cluster 3 (95)

**Regla 11**

Si Capacitación Laboral = Tiene algún oficio  
Y Última Situación Laboral = Trabajador de tiempo completo  
Y Delito Cometido = Infracción ley N° 23.737 (estupefacientes)  
Y Estado Civil = Casado  
Entonces Cluster 3 (90)

**Regla 12**

Si Capacitación Laboral = Tiene algún oficio  
Y Última Situación Laboral = Trabajador de tiempo completo  
Y Delito Cometido = Delitos contra la integridad sexual y el Honor  
Y Estado Civil = Casado  
Y Edad > 38  
Entonces Cluster 3 (73)

**Regla 13**

Si Capacitación Laboral = No tiene ni oficio ni profesión  
Y Delito Cometido = Delitos contra la integridad sexual y el Honor  
Y Edad > 32  
Y Última Situación Laboral = Desocupado  
Y Estado Civil = Casado  
Entonces Cluster 3 (72)

**Regla 14**

Si Capacitación Laboral = No tiene ni oficio ni profesión  
Y Delito Cometido = Delito contra la propiedad  
Y Última Situación Laboral = Trabajador de tiempo parcial  
Y Estado Civil = Concubino  
Y Edad > 38  
Entonces Cluster 3 (61)

**Regla 15**

Si Capacitación Laboral = Tiene alguna profesión  
Y Delito Cometido = Delito contra la propiedad  
Y Última Situación Laboral = Trabajador de tiempo completo  
Y Edad > 38  
Y Estado Civil = Casado  
Entonces Cluster 3 (60)

**Regla 16**

Si Capacitación Laboral = Tiene alguna profesión  
Y Delito Cometido = Delito contra las personas  
Y Última Situación Laboral = Trabajador de tiempo completo  
Y Edad > 39  
Y Estado Civil = Casado  
Entonces Cluster 3 (59)

**Regla 17**

Si Capacitación Laboral = Tiene alguna profesión  
Y Delito Cometido = Delito contra las personas

Y Última Situación Laboral = Trabajador de tiempo parcial  
 Y Edad < 33  
 Y Estado Civil = Casado  
 Entonces Cluster 3 (53)

**Regla 18**

Si Capacitación Laboral = No tiene ni oficio ni profesión  
 Y Delito Cometido = Delito contra la propiedad  
 Y Última Situación Laboral = Trabajador de tiempo completo  
 Y Edad > 36  
 Y Estado Civil = Casado  
 Entonces Cluster 3 (45)

**4.2.4.3 Evaluar el modelo.**

A fines de evaluar el modelo, se aplica una inducción directa sobre uno de los atributos para la obtención de las reglas de clasificación. O sea, se omite el paso de *clustering* de los datos.

En este caso, se toma como clase al atributo *Delito Cometido* ya que junto con *Capacitación Laboral* son los campos más relevantes del *dataset*. Esto último se puede verificar en la sección de selección de atributos (capítulo 4.2.4.2.2).

Como algoritmo clasificador se utiliza al C4.5 obteniendo los siguientes resultados:

- Instancias clasificadas correctamente: 28,216 (68.94%).
- Instancias clasificadas incorrectamente: 12,712 (31.06%).

La matriz de confusión queda conformada de la siguiente forma:

	true Infracción ley N° 23.737 (estupefacientes)	true Delitos contra la propiedad	true Delitos contra las personas	true Delitos contra la integridad sexual y el Honor	class precision
pred. Infracción ley n° 23.737 (estupefacientes)	358	110	91	65	57.37%
pred. Delitos contra la propiedad	2,108	25,805	7,332	1,760	69.73%
pred. Delitos contra las personas	204	476	1,682	344	62.16%
pred. Delitos contra la integridad sexual y el Honor	36	85	101	371	62.56%
class recall	13.23%	97.47%	18.27%	14.61%	

**Tabla 4.48.** Matriz de confusión obtenida con el algoritmo C4.5 y clase *Delito Cometido*.

Tal como se observa en la tabla 4.49, los valores correspondientes al vector de *performance* son significativamente peores que las corridas anteriores:

Criterio	Valor
accuracy:	68.94%
classification_error:	31.06%
absolute_error:	0.456 +/- 0.262
correlation:	0.296
squared_correlation:	0.088

**Tabla 4.49.** *Performance Vector* del algoritmo C4.5

En comparación a los dos primeros modelos en donde ambos casos mostraban un poder clasificador superior al 99%, en este último solo se pudo clasificar en forma correcta solo al 68.94% de las instancias.

Esta disminución en la eficacia clasificatoria no significa que pueda haber generado reglas inconsistentes o incorrectas. Puede ocurrir, que dentro de aquellas instancias en las que no pudo clasificar correctamente haya reglas subyacentes que resulten relevantes.

#### **4.2.5 Fase V: Evaluación**

La evaluación desde todo punto de vista resulta ser muy positiva. Se comenzó el proyecto desconociendo totalmente al tema en cuestión y mediante las técnicas de minería de datos se pudo llegar a conclusiones y reflexiones interesantes, de acuerdo a la opinión vertida por los expertos del dominio.

Otro punto a favor es que la interpretación de los resultados obtenidos a través de los algoritmos utilizados, tanto para el agrupamiento de los datos como para la clasificación, se encuentra alineada con la opinión de los especialistas en materia criminal, penal, del derecho y social. Todos los especialistas coinciden en un punto clave:

*“Para lograr un mejor entendimiento de las conductas y comportamientos del detenido, es sumamente importante conocer la historia familiar y cultural del mismo”.*

Lamentablemente, hoy en día no se cuenta con dicha información en el sistema SNEEP.

Conclusiones de las principales reglas obtenidas a partir del árbol de decisión y conocimientos adquiridos:

- La mayoría de las personas presas se encuentran privadas de su libertad por haber cometido el delito de robo o hurto.
- Existe una distribución del nivel de instrucción muy pobre entre los detenidos en establecimientos penitenciarios.

- Aún cuando una persona posee un oficio o profesión y un trabajo de tiempo parcial/completo, comete delitos contra la propiedad. Podría estar relacionado con la cultura del robo que trae innata.
- La mayoría de las personas de más de 30 años que delinquen contra la propiedad son reincidentes.
- Los individuos que delinquen contra las personas son en su mayoría gente que se encuentra desocupada e incapacitada laboralmente.
- El nivel de capacitación laboral esta fuertemente relacionada con el delito de robo y hurto en jóvenes menores a 29 años.
- Un joven de edad entre 20 y 29 años, sin capacitación laboral ni trabajo de tiempo completo, corre mucho riesgo de que sea puesto preso por delinquir contra la propiedad.
- En general, quienes cometen delitos contra la integridad sexual y el honor (violación), son personas casadas, mayor a 32 años, con trabajo de tiempo parcial o completo y que poseen un oficio o profesión.
- Quienes son procesados por infracción a ley N° 23.737 (estupefacientes), son en su mayoría personas con algún tipo de trabajo y capacitación laboral.

## **5. CONCLUSIONES Y REFLEXIONES**

### **5.1 CONCLUSIONES DEL PROYECTO**

El presente proyecto ha demostrado la factibilidad y valor agregado al aplicar sistemas inteligentes de procesamiento de información, como la minería de datos, a la información criminal en Argentina. En particular, se destacan los resultados obtenidos al utilizar métodos de explotación de información en poblaciones carcelarias con el fin de extraer conocimiento de los datos.

Queda demostrado que la minería de datos es una herramienta muy potente que permite explorar grandes bases de datos de manera rápida y eficiente, sin necesidad de ser un experto en el tema a investigar. En mi caso particular, sin entrenamiento ni experiencia alguna en las técnicas utilizadas, sumado al desconocimiento del tema encarado, me permitió comprender cuales son los principales motivos que llevan a una persona a delinquir y así lograr pensar alternativas para la prevención de dichos hechos delictivos.

Lógicamente, previo a la aplicación de la minería de datos, se tuvo que realizar una fuerte investigación sobre estadísticas vinculadas al crimen, al servicio penitenciario y a los procesos penales en Argentina. Esto fue el punto de partida del proyecto, el cual basa su factibilidad en los siguientes puntos:

- Existe muchísima información que no esta siendo aprovechada para la extracción de conocimiento útil (patrones, comportamientos, anomalías, etc.) en toda su dimensión.
- Existen programas de minería de datos de libre uso, relativamente fáciles de aprender y utilizar.
- No es necesario ser un experto ni en temas de sistemas inteligentes ni en el tema a explorar.
- Sin embargo, es condición necesaria contar con la experiencia de especialistas en la temática a analizar que validen las conclusiones extraídas de los modelos de minería de datos.

Se destaca la participación de especialistas (fiscal, juez y abogados) en el tema que permitieron ampliar el conocimiento sobre los delincuentes y sus comportamientos. Sin la ayuda de estos, junto con sus opiniones y soporte, hubiera sido muy difícil, o hasta imposible, lograr captar detalles para los cuales un Ingeniero Industrial no está preparado.

Los resultados experimentales lograron obtener 4 grupos de presos con distintas características y conductas permitiendo validar conocimientos preexistentes, pero esta

vez, con una justificación apoyado por los datos. A su vez, mediante las reglas de clasificación, se logró caracterizar a los presos en base a sus atributos más relevantes. Es importante destacar, que a diferencia de las interpretaciones subjetivas en que pueden incurrir las personas, las reglas de clasificación son evidenciadas con los registros almacenados en la base de datos

## **5.2 REFLEXIONES**

Desde el punto de vista de la Ingeniería Industrial podría verse al sistema penitenciario como una especie de proceso productivo. En este caso, el flujo entrante al sistema serían los presos que ingresan a los establecimientos penitenciarios para ser “reprocesados”, luego de que se les detectara alguna falla (haber cometido el delito). Claramente, como a todo “producto” que necesita ser reprocesado, es necesario que se cuenten con las herramientas adecuadas para llevar a cabo dicha tarea, y que cada uno de los pasos se cumpla a la perfección para que no se vuelva a incurrir en el mismo error o falla.

Lamentablemente, aquí no estamos hablando de bienes materiales. Estamos frente a seres humanos, que por diversos motivos delinquieron y necesitan de ayuda del Estado para poder readaptarse y reubicarse en una sociedad de la cual fueron apartados. El siguiente párrafo, sintetiza brevemente lo que está ocurriendo hoy en día con las personas que se encuentran privadas de su libertad y ayuda a comprender parte de la problemática existente:

*“Se habla de readaptar socialmente a los condenados, se debate sobre la mejor forma que deben asumir los llamados patronatos deliberados, cuando previamente se han degradado en la cárcel de procesados a tantos y tantos seres que son los mismos que ahora se quieren readaptar. ¿Por que entonces no comenzar por donde corresponde: no despersonalizando moral, psíquica y físicamente al encausado y, sobre todo, logrando que el hombre que ha caído no se contamine ante sus propios ojos, porque eso significa llenarlo de resentimiento, por una parte, y, por la otra, dar razón a su rebelión?”* [Neuman & Irurzun, 1994].

Como se pudo observar a lo largo del proyecto, la falta de instrucción en los detenidos, tanto educativa como laboral, es alarmante. La educación debe ser integral y llenar los 3 ciclos, en el caso que correspondiere, para armar al hombre preso con un bagaje de conocimientos que le permitan obtener mejores posibilidades cuando recobre su libertad. Hoy en día esto no se está cumpliendo, es muy bajo, por no decir nulo, el volumen de presos que participan en actividades educativas y laborales. Pero no solo hay que prestarle atención a la educación de los detenidos, sino que también hay que hacer hincapié en aquellas personas que no poseen instrucción y que en muchos casos,

junto con la falta de posibilidades laborales, terminan incurriendo en algún delito contra la propiedad.

Es hora que la dirigencia política modifique su postura frente a estos aspectos. La educación nunca fue un tema determinante para ganar las elecciones, ni en el centro, ni en el sur de América. Las escuelas rara vez son importantes en las elecciones nacionales, provinciales o incluso en las municipales.

Otro importante punto a tener en cuenta es el costo que un preso tiene para la sociedad. No solo el Estado gasta mucho dinero por cada preso, sino que se está perdiendo la oportunidad de utilizar a esas personas como mano de obra productiva. Toda esa gran cantidad de dinero podría destinarse a planes de capacitación, becas para estudio, programas sociales-educativos, y así, en un sin fin de aplicaciones. Si no se logra la reinstrucción y capacitación de estos delincuentes, las probabilidades de que vuelvan a delinquir una vez que recobren la libertad es muy alta. Así lo muestra el índice de reincidencia, el cual habla de que un 40% de los presos vuelve a reingresar a los establecimientos penitenciarios al cabo de 6 años.

Si lugar a duda, la inseguridad pasó a ser un tema que preocupa en la actualidad a la mayoría de las personas. Hay que tratar de afrontar el problema de una manera dual. No basta con la publicación de estadísticas sin tratamiento alguno, tampoco alcanzaría con resolver el “caos” penitenciario, sino que habría que desarrollar nuevas y mejores políticas criminales preventivas. Una ayuda en tal sentido es mediante la creación de sistemas de información de mayor calidad como herramienta fundamental para la toma de esas decisiones. Cuanto más acertada sea la información recolectada, mejores serán las decisiones tomadas y habrá más probabilidad de reducir el delito. Aún así, es necesario que esto sea acompañado por políticas de estado efectivas y duraderas en el tiempo.

Es importante recalcar, que no se trata de defender, mediante este proyecto y reflexiones, a las personas que cometieron un delito. Sino, más bien, se trata darle una mirada desde otro punto de vista. El punto de vista que puede tener un Ingeniero Industrial.



## **6. FUTURAS LÍNEAS DE INVESTIGACIÓN**

Se cree muy importante y de gran valor agregado continuar con la aplicación de minería de datos como técnica de soporte para la toma de decisiones en lo que respecta a políticas criminales y del tipo social.

Si bien durante el presente proyecto se investigó el comportamiento de los imputados en diferentes delitos, se hizo especial foco en las conductas y características asociadas al preso previo al ingreso al establecimiento penitenciario. Sería valioso poder continuar investigando a los detenidos, pero desde la visión intramuros, analizando sus desenvolvimientos dentro del penal o establecimiento de detención. Para ello es necesario hacer uso de una gran cantidad de atributos que fueron descartados por presentar valores o datos faltantes. Mejorando la metodología de recolección de los datos se lograría tener acceso a los valores faltantes, y de esta manera, se podrían evaluar la totalidad de los registros pudiendo así obtener nuevos grupos y reglas de clasificación mediante las técnicas de minería de datos.

Siguiendo con la información necesaria para llegar a mejores conclusiones, sería muy valioso poder llegar a tener acceso a aquella información familiar relevante de los detenidos. Existe una gran relación entre las acciones de la gran mayoría de los presos y sus pasados familiares y culturales.

Finalmente, se propone expandir el uso de esta técnica, aplicándola en las distintas bases de datos de los entes generadores de información. Sería de gran utilidad a los fines de encontrar relaciones entre distintos hechos, lograr vincular la base del SNEEP con la información proveniente del SNIC y su Sistema de Alerta Temprana (SAT). De esta manera, se lograría tener toda la información referente a un hecho delictivo consolidada en un solo sistema de información.



## **7. APORTES REALIZADOS**

El presente proyecto realizó su aporte en el *WICC 2008-X Workshop de Investigadores en Ciencias de la Computación*, bajo la supervisión de los profesores M. Ing. Paola Britos y Dr. Ramón García Martínez. Dicho certamen es organizado por la Facultad de Ingeniería de la Universidad Nacional de La Pampa, que se llevará a cabo los días 5 y 6 de Mayo de 2008 en la ciudad de General Pico, La Pampa.



## **8. REFERENCIAS**

- Ale, J., 2005. *Análisis de Clusters*.
- Aleonada Mon, H., 2004. *Una Sorda Batalla tras las rejas*.  
[http://www.lanacion.com.ar/archivo/Nota.asp?nota\\_id=593630](http://www.lanacion.com.ar/archivo/Nota.asp?nota_id=593630) .  
Diario La Nación. Buenos Aires. Acceso Noviembre 2007.
- Behar, A.M. y Lucilli, P., 2003. *Mapa del delito de la Ciudad Autónoma de Buenos Aires*. Terceras Jornadas de Jóvenes Investigadores. Instituto Gino Germani.  
[http://www.iigg.fsoc.uba.ar/Jovenes\\_investigadores/3JornadasJovenes/Template/s/Eje%20Poder%20y%20Dominacion/Behar%20-%20Poder.pdf](http://www.iigg.fsoc.uba.ar/Jovenes_investigadores/3JornadasJovenes/Template/s/Eje%20Poder%20y%20Dominacion/Behar%20-%20Poder.pdf)  
Acceso Diciembre 2007.
- Britos, P., Hossian, A., García-Martínez, R. y Sierra, E., 2005. *Minería de Datos Basada en Sistemas Inteligentes*. Editorial Nueva Librería. Buenos Aires. ISBN 987-1104-30-8.
- Cartagenova, S. G., 2005. *Detección Automática de Reglas de Asociación*. Trabajo Final de Especialidad en Ingeniería en sistemas Expertos, Instituto Tecnológico de Buenos Aires (ITBA).
- Centro de Estudios Legales y Sociales (CELS), 2005. *Colapso del Sistema Carcelario*. Editorial Siglo Veintiuno Editores. Buenos Aires. ISBN: 987-1220-35-9.
- Chen, H., Chung, W., Xu, J., Wang, G., Qin, Y., Chau, M., 2004. *Crime Data Mining: A General Framework and Some Examples*. IEEE Computer Society, vol. 37, no.4, pp. 50-56.
- Chen, H. y Han J., 1996. *Data Mining: An overview from database perspective*. IEEE Transactions on Knowledge and Data Engineering.
- COPLINK, 2007. Artificial Intelligence Lab. Management Information Systems. Universidad Arizona. <http://ai.bpa.arizona.edu> . Acceso Noviembre 2007.
- Corcoran, J. J., Wilson I. D., Ware J. A. 2003. Predicting de geo-temporal variations of crime and disorder. *International Journal of Forecasting*.
- CRISP-DM, 2007. *CRISP 1.0 Process and User Guide*. Cross Industry Standard Process for Data Mining. <http://www.crisp-dm.org/CRISPWP-0800.pdf>  
Acceso Enero 2008.
- Daroqui, A., 2008. "Trabajo y educación son una especie de ficción dentro de la cárcel". <http://www.clarin.com/suplementos/zona/2008/02/24/z-03615.htm>  
Diario Clarín. Buenos Aires. Acceso Enero 2008.
- Dirección Nacional de Política Criminal. <http://www.polcrim.jus.gov.ar/>. Acceso Octubre 2007.

- Fayyad, U. M. y Uthurusamy R., 2002. *Evolving Data Mining into Solutions for Insights*. ACM. Páginas 28-31.
- Fayyad, U.M., Piatetsky-Shapiro G., Smyth, P., Uthurusamy, R., 1996. *Advances in Knowledge and data mining*. Cambridge (Massachussets): AAAI/MIT Press.
- Frawley W., Piatetsky-Shapiro G., Matheus C. *Knowledge Discovery in Databases: An Overview*. AI Magazine: pp. 213-228. ISSN 0738-4602.
- Han, J. y Kamber, M., 2001. *Data Mining: Concepts and techniques*. Morgan Kauffmann Publishers. Edición 2001.
- Harris, S., 2005. *Army Project illustrates promise, shortcomings of data mining*. National Journal.  
[http://www.govexec.com/story\\_page\\_pf.cfm?articleid=32944&printerfriendlyve rs=1](http://www.govexec.com/story_page_pf.cfm?articleid=32944&printerfriendlyve rs=1) .Acceso Octubre 2007.
- Hartigan, J.A, 1975. *Clustering Algorithms*. New York 1975.
- Hernández Orallo, J., Ramírez Quintana, M.J., Ferri Ramírez, C., 2004. *Introducción a la Minería de Datos*. Editorial Pearson. ISBN 84-205-40919.
- Igarzábal de Nistal, M.A., 2003. *El Mapa del Delito para la Ciudad Autónoma de Buenos Aires*. Centro de Información Metropolitana. Facultad de Arquitectura, Diseño y Urbanismo. UBA. República Argentina.  
<http://www.cexeci.org/IX%20CONFIBSIG/Comunicaciones/turismo,%20estudios%20sociales%20y%20geomarketing/Igarzabal%20de%20Nistal,%20A.pdf>  
Acceso Diciembre 2007.
- ICPS, International Centre for Prison Studies (ICPS), 2007.  
<http://www.umds.ac.uk/depsta/rel/icps/home.html> . Acceso Noviembre 2007.
- Kantardzic, M., 2003. *Data Mining: Concepts, Models, Methods, and Algorithms*. John Wiley & Sons. ISBN 0471228524.
- Lance, P., 2007. *Defense Intelligence Analysis Center (DIAC)*.  
<http://www.peterlance.com/Peter%20Lance/3.21.00%20Link%20Chart.html>  
Acceso Diciembre 2007.
- Lalurette, S., 2008. *Creció 50% el abuso sexual de menores*.  
[http://www.lanacion.com.ar/archivo/nota.asp?nota\\_id=990034](http://www.lanacion.com.ar/archivo/nota.asp?nota_id=990034). Acceso Marzo 2008.
- Lopez Takeyas, B, 2005. *Algoritmo C4.5*. Instituto Tecnológico de Nuevo Laredo.  
[http://www.itnuevolaredo.edu.mx/takeyas/Apuntes/Inteligencia%20Artificial/Apuntes/tareas\\_alumnos/C4.5/C4.5\(2005-II-B\).pdf](http://www.itnuevolaredo.edu.mx/takeyas/Apuntes/Inteligencia%20Artificial/Apuntes/tareas_alumnos/C4.5/C4.5(2005-II-B).pdf) . Acceso Marzo 2008.
- McCue, C., 2005. *Data Mining and Predictive Analytics: Battlespace Awareness for the War of Terrorism*. Defense Intelligence Journal. Páginas 47-63.

- Molina Felix, L. C., 2002. *Data Mining: Torturando a los datos hasta que confiesen*. Universidad Abierta de Cataluña.  
<http://www.uoc.edu/web/esp/art/uoc/molina1102/molina1102.html>  
Acceso Diciembre 2007
- Molteni, S., 2007. *Detección de Patrones de Producción Educativa Basada en Minería de Datos*. Tesis de Grado en Ingeniería Industrial. Instituto Tecnológico de Buenos Aires. <http://www.itba.edu.ar/capis/webcapis/tesisgrado/molteni-tesisdegradoingenieria.pdf>  
Acceso Noviembre 2007.
- Moody, J. y Darken, C., 1989. *Fast Learning in networks of locally tuned processing units*. Neural Computation. 1989.
- Morris, D. y Rosales, L., 2007. *Poder político, maestros y policías*. Revista Noticias. 13 de Octubre del 2007. Páginas 74-80.
- Neuman E. y Irurzun V., 1994. *La Sociedad Carcelaria*. Editorial Despalma. Buenos Aires 1994.
- Ochoa, M. A. 2004. *Herramientas Inteligentes para la Explotación de Información*. Trabajo Final de Especialidad en Ingeniería en sistemas Expertos, Instituto Tecnológico de Buenos Aires (ITBA).
- Oficina del Inspector General del Departamento de Justicia de los EE.UU., 2006. *Able Danger Investigation Report*. <http://www.milnet.com/archives/DOD-IG-Able-Danger.pdf> . Acceso Diciembre 2007.
- Olaeta, H., 2006. *Estadísticas Criminales y Sistemas de Información*. Instituto Latinoamericano de Seguridad y Democracia (Ilsed).  
[http://www.ilsed.org/index.php?option=com\\_docman&task=doc\\_download&gid=121](http://www.ilsed.org/index.php?option=com_docman&task=doc_download&gid=121). Acceso Diciembre 2007.
- Perversi, I., 2007. *Aplicación de Minería de Datos para la Exploración y Detección de Patrones Delictivos en Argentina*. Tesis de Grado en Ingeniería Industrial. Instituto Tecnológico de Buenos Aires.  
<http://www.itba.edu.ar/capis/webcapis/tesisgrado/PERVERSI-tesisdegradoingenieria.pdf> Acceso Noviembre 2007.
- Petrone, D., 2005. *Cárceles Sanas y Limpias: hacia un nuevo régimen penitenciario*. Unidos por la Justicia. Buenos Aires. ISBN: 987-21857-0-0.
- Prieto, C., 2005. *La situación Penitenciaria en Argentina*. Foro de Profesionales Latinoamericanos de Seguridad.  
[http://www.forodeseguridad.com/artic/discipl/disc\\_4062.htm](http://www.forodeseguridad.com/artic/discipl/disc_4062.htm) . Acceso Octubre 2007.
- Quinlan, J., 1993. *Programs for Machine Learning*. Morgan Kaufmann Publishers. . Edición 1993.

Servente, M.; García-Martínez, R., 2002. *Algoritmos TDIDT Aplicados a la Minería Inteligente*. <http://www.fi.uba.ar/laboratorios/lsi/R-ITBA-26-datamining.pdf>  
Acceso Enero 2008.

Sozzo, M., 2000. *Pintando a traves de numeros. Fuentes Estadísticas de Conocimiento y Gobierno Democrático de la Cuestión Criminal en la Argentina*. Instituto Latinoamericano de Seguridad y Democracia (Ilsed).  
[http://www.ilsed.org/index.php?option=com\\_docman&task=doc\\_view&gid=159&Itemid=44](http://www.ilsed.org/index.php?option=com_docman&task=doc_view&gid=159&Itemid=44) Acceso Diciembre 2007.

Thearling, K., 2007. *An Overview of Data Mining Techniques*.  
<http://www.thearling.com/text/dmtechniques/dmtechniques.htm>. Acceso Enero 2008.

Valenga, F., Perversi, I., Fernández, E., Merlino, H., Rodríguez, D., Britos, P., García-Martínez, R., 2007. *La Estadística Criminal y el Aporte de la Minería de Datos*. En Kaminsky, G., Kosovsky, D., Kessler, G. *El Delito en la Argentina Post-crisis*. Pág. 11-24. Editado por la Friedrich Ebert Stiftung

Weiss, S. M. y Indurkha, N., 1998. *Predictive Data Mining. A practical Guide*. Morgan Kauffmann Publishers. San Francisco 1998.

## 9. ANEXOS

### ANEXO 1

### INSTRUMENTO DE RECOLECCIÓN DE DATOS DEL SNIC

 <b>Ministerio de Justicia</b>		S N I C		Dirección Nacional de Política Criminal			
PLANILLA DE HECHOS DELICTIVOS							
Provincia:			Mes:				
Departamento:			Año:				
Municipalidad:							
Seccional:							
Tipo de Delito	Cantidad de hechos			Cantidad de Víctimas			
	Por denuncia particular	Por interv. policial	Total de hechos	Masculino	Femenino	No consta	Total de víctimas
1. Homicidios dolosos							
2. Homicidios dolosos en grado de tentativa							
3. Homicidios culposos en accidentes de tránsito							
4. Homicidios culposos por otros hechos							
5. Lesiones dolosas							
6. Lesiones culposas en accidentes de tránsito							
7. Lesiones culposas por otros hechos							
8. Otros delitos contra las personas							
9. Delitos contra el honor							
10. Violaciones							
11. Otros delitos contra la integridad sexual							
12. Delitos contra el estado civil							
13. Amenazas							
14. Otros delitos contra la libertad							
15. Robos (excluye los agravados por el resultado de lesiones y/o muertes)							
16. Tentativas de robo (excluye las agravadas por el res.de lesiones y/o muertes)							
17. Robos agravados por el resultado de lesiones y/o muertes							
18. Tentativas de robo agravado por el resultado de lesiones y/o muertes							
19. Hurtos							
20. Tentativas de hurto							
21. Otros delitos contra la propiedad							
22. Delitos contra la seguridad pública							
23. Delitos contra el orden público							
24. Delitos contra la seguridad de la nación							
25. Delitos contra los poderes públicos y el orden constitucional							
26. Delitos contra la administración pública							
27. Delitos contra la fe pública							
28. Ley 23.737 (estupefacentes)							
29. Otros delitos previstos en leyes especiales							
30. Figuras contravencionales							
31. Suicidios (consumados)							
Lugar y fecha de remisión			Firma y sello del responsable				
<p><b>IMPORTANTE:</b> ESTE FORMULARIO DEBE SER DEVUELTO DENTRO DE LOS DIEZ PRIMEROS DIAS POSTERIORES AL MES A QUE SE REFIERE LA INFORMACIÓN. ORIGINAL PARA LA DIRECCION NACIONAL DE POLITICA CRIMINAL, SARMIENTO 329 6° PISO (1041) CIUDAD DE BUENOS AIRES.</p> <p>POR CONSULTAS COMUNICARSE AL 11 4328 4674 (FAX) O AL 11 4328 3015 INT. 2659</p>							

Figura 9.1. Instrumento de recolección de datos del SNIC [DNPC, 2007].

**ANEXO 2**

**INSTRUMENTO DE RECOLECCIÓN DE DATOS PARA ESTABLECIMIENTOS DE EJECUCIÓN DE PENAS PRIVATIVAS DE LIBERTAD - SNEEP**

1. PROVINCIA:.....

2. REPARTICIÓN .....

3. NOMBRE DEL ESTABLECIMIENTO DE EJECUCIÓN DE PENAS PRIVATIVAS DE LA LIBERTAD:.....

4. TIPO DE ESTABLECIMIENTO DE EJECUCIÓN DE PENAS PRIVATIVAS DE LA LIBERTAD (Art. 176 L. N° 24.660):.....

5. DIRECCIÓN POSTAL: .....

6. TELÉFONO / FAX:.....

7. CORREO ELECTRÓNICO:.....

8. RESPONSABLE DESIGNADO PARA LA CONFECCIÓN DE LA ESTADÍSTICA:.....

9. CAPACIDAD FÍSICA DEL ALOJAMIENTO (Art. 59 L. N° 24.660):.....

9.A. Cantidad de plazas/camas en celdas individuales:.....

9.B. Cantidad de plazas/camas en locales colectivos:.....

**CUADRO 1**  
POBLACIÓN PRIVADA DE LA LIBERTAD POR JURISDICCIÓN, SITUACIÓN LEGAL Y SEXO

Situación Legal	Provincial		Nacional		Federal		TOTAL
	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	
Procesados							
Condenados							
Inimputables							
Contraventores							
Otros alojados							
<b>TOTAL</b>							

**CUADRO 2**  
CANTIDAD DE INGRESOS DURANTE EL ÚLTIMO AÑO

PERSONAS INGRESADAS EN EL ÚLTIMO AÑO	MASCULINO	FEMENINO	TOTAL

**Figura 9.2.** Instrumento de recolección de datos del SNEEP [DNPC, 2007].

**ANEXO 2 - CONTINUACIÓN**

CUADRO 3								
EGRESOS DE PROCESADOS DURANTE EL ULTIMO AÑO								
MOTIVOS DE EGRESOS	PROVINCIAL		NACIONAL		FEDERAL		TOTAL	
	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.		
ABSOLUCIÓN								
CAMBIO SITUACION LEGAL								
ENTREGA A PADRES								
EVASIÓN								
EXCARCELACIÓN								
FALLECIMIENTO								
FALTA DE MÉRITO								
FUGA								
INDULTO								
SOBRESEIMIENTO								
TRASLADOS								
VIGILADA								
NO ESPECIFICADOS								
OTROS MOTIVOS								
TOTAL								

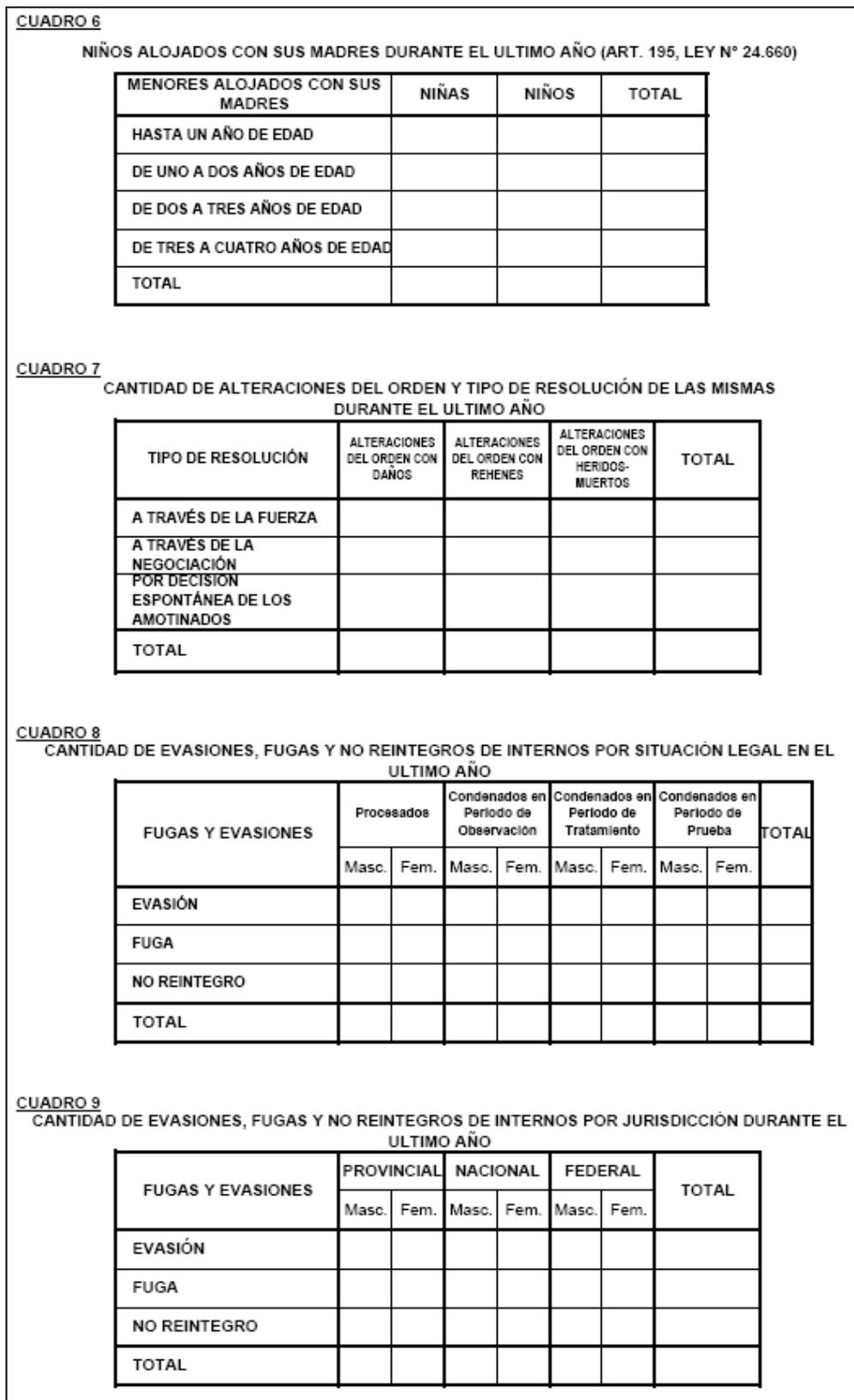
CUADRO 4								
EGRESOS DE CONDENADOS DURANTE EL ULTIMO AÑO								
MOTIVOS DE EGRESOS	PROVINCIAL		NACIONAL		FEDERAL		TOTAL	
	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.		
AGOTAMIENTO DE LA PENA								
EVASIÓN								
FALLECIMIENTO								
FUGA								
INDULTO								
LIBERTAD CONDICIONAL ART. 13 C.P.								
LIBERTAD CONDICIONAL ART. 53 C.P.								
LIBERTAD ASISTIDA ART. 54 L.E.P.P.								
PRISIÓN DOMICILIARIA ART. 33 L.E.P.P.								
TRASLADOS								
OTROS MOTIVOS								
TOTAL								

CUADRO 5									
CANTIDAD DE VISITANTES DURANTE EL ULTIMO AÑO									
VISITANTES	Procesados		Condenados en Periodo de Observación		Condenados en Periodo de Tratamiento		Condenados en Periodo de Prueba		TOTAL
	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	

Figura 9.3. Instrumento de recolección de datos del SNEEP [DNPC, 2007].

**ANEXO 2- CONTINUACIÓN**



**Figura 9.4.** Instrumento de recolección de datos del SNEEP [DNPC, 2007].

**ANEXO 2- CONTINUACIÓN**

CUADRO 10							
CANTIDAD DE SUICIDIOS DE INTERNOS POR JURISDICCION DURANTE EL ULTIMO AÑO							
SUICIDIOS	PROVINCIAL		NACIONAL		FEDERAL		TOTAL
	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	

CUADRO 11											
CANTIDAD DE SUICIDIOS DE INTERNOS POR SITUACION LEGAL DURANTE EL ULTIMO AÑO											
SUICIDIOS	Procesados		Condenados		Inimputables		Contravent.		Otros alojados		TOTAL
	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	

CUADRO 12											
CANTIDAD DE INTERNOS FALLECIDOS SEGUN SITUACION LEGAL DURANTE EL ULTIMO AÑO											
INTERNOS FALLECIDOS	Procesados		Condenados		Inimputables		Contravent.		Otros alojados		TOTAL
	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	
POR HECHOS DE VIOLENCIA PRODUCIDOS EN EL ESTABLECIMIENTO POR OTROS INTERNOS											
POR HECHOS DE VIOLENCIA PRODUCIDOS EN EL ESTABLECIMIENTO POR AGENTES DEL MISMO											
POR OTRAS CAUSAS (EXCLUIDOS SUICIDIOS)											
TOTAL											

CUADRO 13							
CANTIDAD DE INTERNOS FALLECIDOS POR JURISDICCION DURANTE EL ULTIMO AÑO							
INTERNOS FALLECIDOS	Provincial		Nacional		Federal		TOTAL
	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.	
POR HECHOS DE VIOLENCIA PRODUCIDOS EN EL ESTABLECIMIENTO POR OTROS INTERNOS							
POR HECHOS DE VIOLENCIA PRODUCIDOS EN EL ESTABLECIMIENTO POR AGENTES DEL MISMO							
POR OTRAS CAUSAS (EXCLUIDOS SUICIDIOS)							
TOTAL							

CUADRO 14			
CANTIDAD DE LESIONES Y/O FALLECIMIENTOS COMO RESULTADO DE ALTERACIONES DEL ORDEN DURANTE EL ULTIMO AÑO			
PERSONAS LESIONADAS / FALLECIDAS EN ALTERACIONES DEL ORDEN			TOTAL
	MASCULINO	FEMENINO	
AGENTES LESIONADOS			
AGENTES FALLECIDOS			
INTERNOS LESIONADOS			
INTERNOS FALLECIDOS			
TOTAL			

Figura 9.5. Instrumento de recolección de datos del SNEEP [DNPC, 2007].



**ANEXO 3- CONTINUACIÓN**  
**CENSO PENITENCIARIO - CODIFICACIÓN DE VARIABLES**

**(1) Nacionalidad**

- 1- Argentina
- 2- Boliviana
- 3- Brasileña
- 4- Chilena
- 5- Paraguaya
- 6- Uruguaya
- 7- Peruana
- 8- Colombiana
- 9- Ecuatoriana
- 10- Española
- 11- Italiana
- 12- Inglesa
- 13- China
- 14- Sudafricana
- 15- Nigeriana
- 16- Otras

**(2) Estado Civil**

- 1- Soltero
- 2- Casado
- 3- Viudo
- 4- Separado o Divorciado
- 5- Separado de Hecho
- 6- Concubino

**(3) Nivel de Instrucción**

- 1- Ninguno
- 2- Primario Incompleto
- 3- Primario Completo
- 4- Secundario Incompleto
- 5- Secundario Completo
- 6- Terciario Incompleto
- 7- Terciario Completo
- 8- Universitario Incompleto
- 9- Universitario Completo
- 10- EGB 1 Incompleto
- 11- EGB 1 Completo
- 12- EGB 2 Incompleto
- 13- EGB 2 Completo
- 14- EGB 3 Incompleto
- 15- EGB 3 Completo
- 16- Polimodal Incompleto
- 17- Polimodal Completo

**(4) Situación laboral al momento del ingreso**

- 1- Trabajador de tiempo completo

- 2- Trabajador de tiempo parcial
- 3- Desocupado

**(5) Capacitación Laboral al momento del ingreso**

- 1- Tenía algún oficio
- 2- Tenía alguna profesión
- 3- No tenía ni oficio, ni profesión

**(6) Lugar de residencia al momento del ingreso**

- 1- Rural
- 2- Urbano

**(7) Provincia de residencia al momento del ingreso**

- 1- Buenos Aires
- 2- Catamarca
- 3- Córdoba
- 4- Corrientes
- 5- Chaco
- 6- Chubut
- 7- Entre Ríos
- 8- Formosa
- 9- Jujuy
- 10- La Pampa
- 11- La Rioja
- 12- Mendoza
- 13- Misiones
- 14- Neuquén
- 15- Río Negro
- 16- Salta
- 17- San Juan
- 18- San Luis
- 19- Santa Cruz
- 20- Santa Fe
- 21- Santiago del Estero
- 22- Tierra del Fuego
- 23- Tucumán
- 24- Ciudad de Buenos Aires

**(8) Jurisdicción Judicial**

- 1- Provincial - Buenos Aires
- 2- Provincial - Catamarca
- 3- Provincial - Córdoba
- 4- Provincial - Corrientes
- 5- Provincial - Chaco

- 6- Provincial - Chubut
- 7- Provincial - Entre Ríos
- 8- Provincial - Formosa
- 9- Provincial - Jujuy
- 10- Provincial - La Pampa
- 11- Provincial - La Rioja
- 12- Provincial - Mendoza
- 13- Provincial - Misiones
- 14- Provincial - Neuquén
- 15- Provincial - Río Negro
- 16- Provincial - Salta
- 17- Provincial - San Juan
- 18- Provincial - San Luis
- 19- Provincial - Santa Cruz
- 20- Provincial - Santa Fe
- 21- Provincial - Santiago del Estero
- 22- Provincial - Tierra del Fuego
- 23- Provincial - Tucumán
- 24- Nacional - Ciudad de Buenos Aires
- 25- Federal

**(9) Situación Legal**

- 1- Condenado
- 2- Procesado
- 3- Contraventor
- 4- Inimputable
- 5- Otra

**(10) Establecimiento de Procedencia**

- 1- Ingreso Directo
- 2- Derivación de otro Establecimiento de Ejecución Penal
- 3- Derivación de una Institución Policial (alojados más de 4 días)
- 4- Derivación de una Institución dependiente de una fuerza de seguridad (Prefectura o Gendarmería)

**(11) Tipo de delito por el cual ingresó al Establecimiento**

**Delitos contra las Personas**

- 1- Homicidios Dolosos
- 2- Homicidios Dolosos (tentativa)
- 3- Homicidios Culposos
- 4- Lesiones Dolosas
- 5- Lesiones Culposas
- 6- Otros Delitos contra las personas
- 7-Delitos contra el honor

**Delitos contra la integridad sexual**

- 8- Violaciones
- 9- Otros delitos contra la honestidad

**Delitos contra la libertad**

- 10- Amenazas
- 11- Privación ilegítima de la libertad
- 12- Delito contra la libertad cometido por funcionario público
- 13- Otros delitos contra la libertad

**Delitos contra la propiedad**

- 14- Hurto y/o tentativa de hurto
- 15- Robo y/o tentativa de robo
- 16- Otros delitos contra la propiedad
- 17- Delitos contra la seguridad pública
- 18- Delitos C/ el orden público
- 16- Otros delitos contra la propiedad
- 17- Delitos contra la seguridad pública
- 18- Delitos C/ el orden público
- 19- Delitos C/ la seguridad de la nación
- 20- Delitos C/ los poderes públicos
- 21- Delitos C/ la administración pública
- 22- Delitos C/ la Fe.
- 23- Delitos C/ el Estado Civil.
- 24- Infracción Ley N° 23.737 (Estupefacientes)

- 25- Infracción Ley N° 24.769 Penal Tributaria
- 26- Infracción Ley N° 13.944 Incumplimiento de deberes
- 27- Delitos Previstos en leyes especiales.
- 28- Contravenciones 19-Delitos C/ la seguridad de la nación

**(12) Trabajo Remunerado según cantidad de horas**

- 1- Hasta 10 hs. Semanales
- 2- Hasta 20 hs. Semanales
- 3- Hasta 30 hs. Semanales
- 4- Hasta 40 hs. Semanales
- 5- No tiene trabajo remunerado

**(13) ¿El interno participó regularmente de algún programa de Formación o Capacitación Laboral durante el último año?**

- 1- SI
- 2- NO

**(14) ¿El interno participó regularmente de algún Programa Educativo durante el último año?**

- 1-SI- Educación Formal- EGB
- 2-SI- Educación Formal- Polimodal
- 3-SI- Educación Formal- Terciaria
- 4-SI- Educación Formal - Universitaria
- 5-SI- Educación NO Formal
- 6- No participa de ningún programa educativo

**(15) ¿El interno participó regularmente de Actividades Recreativas y/o Deportivas durante el último año?**

- 1- SI

- 2- NO

**(16) ¿El interno recibió Asistencia Médica durante el último año?**

- 1- SI
- 2- NO

**(17) ¿El interno recibió Visitas durante el último año?**

- 1- SI
- 2- NO

**(18) ¿El interno participó de alguna Alteración del Orden durante el último año? Indicar tipo de Alteración del Orden**

- 1- SI – Alteración del orden con Heridos o Muertos
- 2- SI – Alteración del orden con Rehenes
- 3- SI – Alteración del orden con Daños
- 4- SI – Alteración del orden sin heridos, ni rehenes, ni daños.
- 5- NO participó de ninguna alteración

**(19) Tipo de Infracción disciplinaria cometida por el interno durante el últ. Año**

- 1- Faltas graves
- 2- Faltas medias
- 3- Faltas leves
- 4- NO cometió Infracción disciplinaria

**(20) Tipo de Sanciones Disciplinarias impuesta al interno**

- 1- Traslado a otro establecimiento de régimen más severo
- 2- Traslado a otra sección de régimen más severo
- 3- Permanencia en su alojamiento individual o en celdas cuyas condiciones no agraven

ilegítimamente la detención hasta 7 fines de semana sucesivos o alternados  
 4- Permanencia en su alojamiento individual o en celdas cuyas condiciones no agraven ilegítimamente la detención hasta 15 días ininterrumpidos  
 5- Suspensión o restricción total o parcial de los derechos reglamentarios hasta 15 días  
 6- Exclusión de la actividad común hasta 15 días  
 7- Exclusión de la actividad recreativa o deportiva hasta 10 días  
 8- Amonestación  
 9- Otra sanción

**(22) ¿El interno tuvo intentos de Fugas o Evasiones durante el último año?**

1- SI, tentativa de Evasión  
 2- SI, tentativa de Fuga  
 3- No

**(23) ¿El interno tuvo intentos de Suicidio durante el último año?**

1- SI  
 2- NO

**(24) ¿El interno fue lesionado en el establecimiento en el último año? Motivos**

1- SI, por hechos de violencia con otros internos  
 2- SI, por hechos de violencia con agentes del establecimiento  
 3- SI, por otros motivos  
 4- No fue lesionado

**(25) ¿El interno está condenado a Prisión o reclusión Perpetua?**

1- SI  
 2- NO

**(26) ¿El interno tiene Medida de Seguridad (art. 52 CP)?**

1- SI  
 2- NO

**(27) Reincidencia**

1- Primario  
 2- Reiterante  
 3- Reincidente (art. 50 CP)  
 4- Reincidencia múltiple (art. 52 CP)

**(28) Períodos de la Progresividad del Régimen Penitenciario en el último año.**

**Artículo 12, Ley N° 24.660**

1- Período de observación  
 2- Período de tratamiento  
 3- Período de prueba  
 4- Período de libertad condicional  
 5- Se aplica otra caracterización  
 6- No se aplica ninguna

**(29) ¿Tiene o tuvo salidas transitorias en el último año? Art. 16 y 19, Ley N° 24.660**

1- NO le fueron otorgadas  
 2- SI goza de salidas transitorias  
 3- Suspendidas durante el último año  
 4- Revocadas durante el último año

**(30) ¿El Interno está incorporado al Régimen de Semilibertad?**

**Artículos 15, 19 y 23 Ley N° 24.660**

1- NO está incorporado al régimen de Semilibertad  
 2- SI está incorporado

3- Suspendido durante el último año  
 4- Revocado durante el último año

**(31) ¿El Interno está incorporado a un Programa de Prelibertad? Artículo 30, Ley N° 24.660**

1- SI está incorporado  
 2- No está incorporado

**(32) ¿El Interno cumple la pena bajo la modalidad de Prisión Discontinua?**

**Artículos 36 y 49, Ley N° 24.660**

1- No está incorporado  
 2- SI está incorporado  
 3- Revocada durante el último año  
 4- Renuncia del interno durante el último año

**(33) ¿Al Interno se le otorgó la Semidetención? Artículos 39 a 42, Ley N° 24.660**

1- No tiene Semidetención  
 2- SI, Prisión Diurna  
 3- SI, Prisión Nocturna  
 4- Revocada durante el último año  
 5- Renuncia del interno durante el último año

**(34) ¿El interno tuvo reducción de pena en el último año?**

1- SI – Hasta 6 meses  
 2- SI – Hasta 1 año  
 3- SI – Más de 1 año  
 4- NO tuvo reducción de pena

**(35) Para Mujeres: ¿Tiene hijos alojados con ella? Artículo 195, Ley N° 24.660**

1- SI  
 2- NO

**ANEXO 4**

**POBLACIÓN DETENIDA EN COMISARIAS – AÑO 2005**

PROVINCIA	DETENIDOS CÁRCELES	DETENIDOS COMISARIAS	TOTAL
SPF	9625		9,625
BUENOS AIRES	24721	4,592	29,313
CATAMARCA	356	107	463
CÓRDOBA	5484	141	5,625
CORRIENTES	700	267	967
CHACO	961	96	1,057
CHUBUT	933	-	933
ENTRE RIOS	304	140	444
FORMOSA	442	221	663
JUJUY	186	67	253
LA PAMPA	2464	-	2,464
LA RIOJA	1074	20	1,094
MENDOZA	595	No informó	595
MISIONES	630	98	728
NEUQUEN	1707	15	1,722
RIO NEGRO	723	35	758
SALTA	405	147	552
SAN JUAN	116	54	170
SAN LUIS	2217	No informó	2,217
SANTA CRUZ	467	105	572
SANTA FE	956	1,279	2,235
SANTIAGO DEL ESTERO	115	45	160
TIERRA DEL FUEGO	147	-	147
TUCUMÁN	95	79	174
GENDARMERIA (Junio 2005)		367	367
PREFECTURA (Junio 2005)		59	59
	55,423	7,934	63,357

**Tabla 9.1.** Población detenida en Comisarías [DNPC, 2005].

Relevamiento basado en datos proporcionados por el SPF y los servicios penitenciarios o policías provinciales.

**ANEXO 5**

**EVOLUCIÓN POBLACIÓN PENITENCIARIA POR PROVINCIA – AÑO 2006**

PROVINCIA / AÑO	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
SPF	6,177	6,385	6,767	7,146	7,872	8,795	9,246	9,738	9,625	9,380
BUENOS AIRES	11,527	12,460	13,190	15,012	16,990	18,931	22,983	23,449	24,721	23,878
CATAMARCA	107	139	166	166	171	243	252	277	356	358
CÓRDOBA	3,283	3,475	3,854	4,196	4,582	4,926	5	5,661	5,484	5,162
CORRIENTES	237	220	202	217	252	286	312	405	700	795
CHACO	s/d	s/d	s/d	s/d	s/d	116	476	991	961	819
CHUBUT	202	254	320	406	448	310	147	158	147	109
ENTRE RÍOS	536	643	649	705	773	881	838	986	933	873
FORMOSA	245	217	260	269	277	283	260	299	304	279
JUJUY	288	318	368	440	s/d	429	387	456	442	443
LA PAMPA	s/d	s/d	s/d	s/d	s/d	255	66	94	115	135
LA RIOJA	110	107	106	121	112	119	179	218	186	197
MENDOZA	1,214	1,415	1,759	2	2,183	2,479	2,543	2,559	2,464	2,482
MISIONES	521	560	637	651	755	701	775	1,023	1,074	1,076
NEUQUEN	466	491	466	490	483	874	945	575	595	593
RIO NEGRO	546	409	487	806	564	611	516	571	630	652
SALTA	1,280	1,385	1,410	1,490	1,548	1,575	1,733	1,787	1,707	1,839
SAN JUAN	406	360	397	520	556	585	672	605	723	691
SAN LUIS	s/d	s/d	s/d	s/d	s/d	312	389	429	405	422
SANTA CRUZ	128	130	97	93	609	140	262	215	116	88
SANTA FE	1,532	1,758	1,842	1,784	2,176	2,289	2,567	2,564	2,217	2,181
SANTIAGO DEL ESTE	386	332	451	559	s/d	334	228	465	467	478
TIERRA DEL FUEGO	54	66	99	228	166	150	100	94	95	92
TUCUMÁN	445	497	513	632	490	664	822	853	956	978
<b>TOTAL</b>	<b>29,690</b>	<b>31,621</b>	<b>34,040</b>	<b>35,933</b>	<b>41,007</b>	<b>46,288</b>	<b>46,703</b>	<b>54,472</b>	<b>55,423</b>	<b>54,000</b>
Comisaría s Bs. As.	2,346	3,088	3,107	5,748	6,113	7,507	5,223	6,576	4,592	

Tabla 9.2. Evolución de la Población Penitenciaria en Argentina [DNPC, 2006].

**ANEXO 6**

**POBLACIÓN CARCELARIA EN EL MUNDO**

<b>Pais y Año medición</b>	<b>Población Carcelaria</b>	<b>Tasa c/100.000</b>
Estados Unidos (Jun-2006)	2,245,189	750
Rusia (Jul-2007)	889,598	628
Cuba (Nov-2006)	60,000	531
Bielorrusia (Ene-2006)	41,538	426
Puerto Rico (Jul-2006)	14,239	356
Kazakhstan (Ene-2007)	51,538	348
Ucrania (Feb-2007)	160,046	345
Sudafrica (May-2007)	159,961	335
Emiratos Arabes Unidos (Ene-2004)	8,927	288
Taiwan (Abr-2007)	64,279	281
Chile (May-2007)	43,723	262
Tailandia (Dic-2006)	161,844	249
Brasil (Jun-2007)	419,551	219
Israel (Dic-2004)	13,909	209
Mexico (May-2007)	216,290	198
Uruguay (Ago-2006)	6,947	193
Inglaterra y Gales (Jul-2007)	80,229	148
España (Jul-2007)	66,129	147
<b>Argentina (Dic-2006)</b>	<b>54,000</b>	<b>139</b>
Colombia (May-2007)	60,158	128
Australia (Jun-2006)	25,790	125
China (Dec-2005)	1,565,771	119
Canadá (Mar-2005)	34,244	107
Paraguay (Jun-2006)	6,264	98
Alemania (Nov-2006)	76,629	93
Francia (Sep-2006)	52,009	85
Suecia (Oct-2006)	7,175	79
Suiza (Sep-2006)	5,888	79
Italia (Dic-2006)	39,348	67
Japón (Dic-2006)	77,932	61

**Tabla 9.3.** Población Penitenciaria en Mundial [ICPS, 2007].

## **ANEXO 7**

### **SERVICIO PENITENCIARIO FEDERAL: CÁRCELES Y ALCALDÍAS**

- Unidad 2 (cárcel de Devoto), Ciudad Autónoma de Buenos Aires
  - Unidad 3, Instituto Correccional de Mujeres, Localidad de Ezeiza, Prov.de Bs.As.
  - Unidad 4, Colonia Penal de Santa Rosa, Provincia de La Pampa
  - Unidad 5, Colonia Penal Subprefecto Rocha, Provincia de Río Negro
  - Unidad 6, Instituto de Seguridad, Rawson, Chubut
  - Unidad 7, Prisión Regional del Norte, Resistencia, Chaco
  - Unidad 9, Prisión Regional del Sur, Provincia de Neuquén
  - Unidad 10, Cárcel de Formosa
  - Unidad 11, Colonia Penal de Roque Sáenz Peña, Chaco
  - Unidad 12, Colonia Penal de Viedma, Río Negro
  - Unidad 13, Instituto Correccional de Mujeres Nuestra Sra. Del Carmen, Santa Rosa, La Pampa
  - Unidad 14, cárcel de Esquel, Chubut
  - Unidad 15, cárcel de Río Gallegos, Santa Cruz
  - Unidad 17, Colonia Penal de Candelaria, Misiones
  - Unidad 19, Instituto Correccional Abierto de Ezeiza, Provincia de Buenos Aires.
  - Unidad 20, Instituto Psiquiátrico de Varones, Ciudad de Buenos Aires
  - Unidad 21, Centro de Enfermedades Infectocontagiosas, Ciudad de Buenos Aires.
  - Unidad 24 y 26, Complejo Federal para jóvenes adultos, Marcos Paz, Prov.de Bs.As.
  - Unidad 25, Correccional Abierto de General Pico, La Pampa
  - Unidad 27, Servicio Psiquiátrico Central de Mujeres, Ciudad de Buenos Aires
  - Unidad 28, Centro de detención Judicial, Ciudad de Buenos Aires
  - Unidad 29, Alcaldía Federal, Ciudad de Buenos Aires
  - Unida 30, Instituto de Jóvenes Adultos, Dr. Julio Alfonsín
  - Unida 31, Centro Federal de detención de mujeres, Ezeiza, Provincia de Bs. As.
  - Unidad 32, Centro de Detención Judicial de Mendoza
  - Complejos Penitenciarios I y II, sitios en las localidades de Ezeiza y Marcos Paz, respectivamente. Las Unidades 1 y 16 eran las cárceles de Encausados de la Capital Federal y la de Jóvenes Adultos, respectivamente; la unidad 22 ya no funciona como establecimiento penitenciario y su espacio es el que hoy corresponde a la cárcel de Contraventores de Buenos Aires; la unidad 23 era una cárcel de jóvenes que funcionaba en el ámbito del Establecimiento de Devoto.
- 
- Alcaldía Correccional Juncal, Ciudad de Buenos Aires
  - Alcaldía Penal Coronel Paiva, Ciudad de Buenos Aires
  - Alcaldía Penal Inspector General R. Pettinato, Ciudad de Buenos Aires
  - Alcaldía Correccional Lavalle, Ciudad de Buenos Aires
  - Unidad 18, Casa de Preegreso, Ciudad de Buenos Aires
  - Alcaldía Federal de Lomas de Zamora, Provincia de Buenos Aires;
  - Alcaldía Federal de Salta;
  - Alcaldía Federal de Jujuy;

**ANEXO 8**
**SOBREPOBLACIÓN SPF 2003**

PROVINCIA	E STABLECIMIENTO O	CAPACIDAD	POBLACION ALOJADA	SOBREPOBLACION
SPF BUENOS AIRES	SPF - COMPLEJO PENITENCIARIO FEDERAL I - EZEIZA-	1680	1665	-1.5%
	SPF - COMPLEJO PENITENCIARIO FEDERAL II -MARCOS PAZ-	1300	1410	-6.0%
	SPF - CENTRO FED. DE DET. DE MUJERES (U 31)	236	234	-8.6%
	SPF - INSTITUTO CORREC. ABIERTO DE EZEIZA (U.19)	248	210	-15.3%
	SPF - CENTRO FED. DE TRAT. ESPECIAL PARA JOV. ADULTOS	44	32	-27.3%
	SPF - INSTITUTO FED. PARA JOV. ADULTOS (U.24)	152	66	-56.6%
	SPF - INST. JOV. ADULTO "JUAN C. LANDO"(U.26)	39	57	46.2%
	SPF - INSTITUTO CORRECCIONAL DE MUJERES (U.03)	374	397	6.1%
	SPF - PRENSION REGIONAL DEL NORTE (U.07)	500	475	-5.0%
	SPF - COLONIA PENAL DE PRESID. ROQUE SAENZ PEÑA (U.11)	140	135	-3.6%
SPF CHUBUT	SPF - INSTITUTO DE SEGURIDAD Y RESOCIALIZACION (U.06)	500	475	-5.0%
	SPF - CARCEL DE ESQUEL "SUBALC. ABEL R. MUÑOZ"; (U.14)	92	58	-35.9%
SPF FORMOSA	SPF - CARCEL DE FORMOSA (U.10)	120	121	0.8%
	SPF - ALCALDIA FEDERAL DE JUJUY	80	72	-10.0%
SPF LA PAMPA	SPF - COLONIA PENAL DE SANTA ROSA - LA PAMPA (U.04)	342	304	-11.1%
	SPF - INST. CORREC. DE MUJ. NTRA. SRA. DEL CARMEN" (U.13)	70	74	5.7%
	SPF - INST. CORREC. ABIERTO DE GRAL PICO (U.25)	23	24	4.3%
	SPF - INSTITUTO DE JOVENES ADULTOS (U.30)	21	15	-28.6%
	SPF - CENTRO DE DETENCION JUDICIAL DE MENDOZA (U.32)	24	148	516.7%
SPF MISIONES	SPF - COLONIA PENAL DE CANDELARIA (U.17)	156	181	16.0%
	SPF - PRENSION REGIONAL DEL SUR (U.09)	237	245	3.4%
SPF RIO NEGRO	SPF - COLONIA PENAL "SUBP. MIGUEL ROCHA" (U.05)	330	293	-11.2%
	SPF - COLONIA PENAL DE VIEDMA (U.12)	155	125	-19.4%
SPF SALTA	SPF - ALCALDIA FEDERAL DE SALTA	30	28	-3.3%
SPF SANTA CRUZ	SPF - CARCEL DE RIO GALLEGOS (U.15)	108	93	-13.9%
	SPF - INSTITUTO DE DETENCION DE LA CAP. FED (U.02)	1800	1846	2.6%
SPF CIUDAD BUENOS AIRES	SPF - CASA DE PREGRESO "DR. JOSE INGENIEROS" (U.18)	6	5	-16.7%
	SPF - SERVICIO PSIQUIATRICO CENTRAL DE VARONES (U.20)	85	126	48.2%
	SPF - CENTRO PENIT. DE ENFERMEDADES INFECCIOCONTAGIOSAS (U.	61	20	-67.2%
	SPF - SERVICIO PSIQUIATRICO CENTRAL DE MUJERES (U.27)	25	16	-36.0%
		9,208	8,952	-2.8%

**Tabla 9.4.** Sobrepoblación Servicio Penitenciario Federal [DNPC, 2003].

**ANEXO 8 - CONTINUACIÓN**  
**SOBREPOBLACIÓN SPF 2004**

PROVINCIA	ESTABLECIMIENTO	CAPACIDAD	POBLACION ALOJADA	SOBREPOBLACION
SPF BUENOS AIRES	SPF - COMPLEJO PENITENCIARIO FEDERAL I - EZEIZA	1690	1645	-2.7%
	SPF - COMPLEJO PENITENCIARIO FEDERAL II - MARCOS PAZ	1500	1505	0.3%
	SPF - U. 3 INSTITUTO CORRECCIONAL DE MUJERES	416	624	50.0%
	SPF - U.19 INSTITUTO CORRECCIONAL ABIERTO DE EZEIZA	380	372	-2.1%
	SPF - CENTRO FED. DE TRAT. ESPECIAL P/ JÓVENES ADULTOS (AREA CRD)	44	23	-47.7%
	SPF - U.24 INSTITUTO FEDERAL PARA JÓVENES ADULTOS	152	143	-5.9%
	SPF - U.26 INSTITUTO PARA JÓVENES ADULTOS "JUAN C. LANDO"	39	29	-25.6%
	SPF - U.31 CENTRO FEDERAL DE DETENCIÓN DE MUJERES	256	250	-2.3%
	SPF - U. 7 PRISION REGIONAL DEL NORTE	500	490	-2.0%
	SPF - U.11 COLONIA PENAL DE PRESID. ROQUE SAENZ PEÑA	140	141	0.7%
SPF CHUBUT	SPF - U. 6 INSTITUTO DE SEGURIDAD Y RESOCIALIZACION	515	509	-1.2%
	SPF - U.14 CARCEL DE ESQUEL "SUBALCAIDE ABEL R. MUÑOZ"	92	84	-8.7%
SPF FORMOSA	SPF - U.10 CARCEL DE FORMOSA	120	114	-5.0%
SPF JUJUY	SPF - ALCALDIA FEDERAL DE JUJUY	80	70	-12.5%
	SPF - U. 4 COLONIA PENAL DE SANTA ROSA - LA PAMPA	322	302	-6.2%
SPF LA PAMPA	SPF - U.13 INSTITUTO CORRECCIONAL DE MUJERES "NTRA. SRA. DEL CARMEN	70	81	15.7%
	SPF - U.25 INSTITUTO CORRECCIONAL ABIERTO DE GENERAL PICO	47	23	-51.1%
	SPF - U.30 INSTITUTO DE JOVENES ADULTOS	21	18	-14.3%
	SPF - U.17 COLONIA PENAL DE CANDELARIA	176	176	0.0%
SPF NEUQUEN	SPF - U. 9 PRISION REGIONAL DEL SUR	250	249	-0.4%
SPF RIO NEGRO	SPF - U. 5 COLONIA PENAL "SUBPREFECTO MIGUEL ROCHA"	330	333	0.9%
	SPF - U.12 COLONIA PENAL DE VIEDMA	198	179	-9.6%
SPF SALTA	SPF - ALCALDIA FEDERAL DE SALTA	30	30	0.0%
SPF SANTA CRUZ	SPF - U.15 CARCEL DE RIO GALLEGOS	104	99	-4.8%
	SPF - U. 2 INSTITUTO DE DETENCIÓN DE LA CAPITAL FEDERAL	1800	2053	14.1%
SPF CIUDAD BUENOS AIRES	SPF - U.18 CASA DE PREGRESO "DR. JOSE INGENIEROS"	8	9	12.5%
	SPF - U.20 SERVICIO PSIQUIATRICO CENTRAL DE VARONES	87	158	81.6%
	SPF - U.21 CENTRO PENIT. DE ENFERMEDADES INFECCIOSAS	37	17	-54.1%
	SPF - U.27 SERVICIO PSIQUIATRICO CENTRAL DE MUJERES	25	12	-52.0%
		9,429	9,738	3.3%

Tabla 9.5. Sobrepopulación Servicio Penitenciario Federal [DNPC, 2004].

**ANEXO 8- CONTINUACIÓN**  
**SOBREPOBLACIÓN SPF 2005**

PROVINCIA	ESTABLECIMIENTO	CAPACIDAD	POBLACION ALOJADA	SOBREPOBLACION
SPF BUENOS AIRES	SPF - COMPLEJO PENITENCIARIO FEDERAL I - EZEIZA	1676	1576	-6.0%
	SPF - COMPLEJO PENITENCIARIO FEDERAL II - MARCOS PAZ	1500	1493	-0.5%
	SPF - U. 3 INS TITUTO CORRECCIONAL DE MUJERES	824	725	-12.0%
	SPF - U.19 INS TITUTO CORRECCIONAL ABIERTO DE EZEIZA	380	340	-10.5%
	SPF - CENTRO FED DE TRAT ESPECIAL P/ JÓVENES ADULTOS (AREA CRD)	44	30	-31.8%
	SPF - U.24 INS TITUTO FEDERAL PARA JÓVENES ADULTOS	152	137	-9.9%
	SPF - U.26 INS TITUTO PARA JÓVENES ADULTOS "JUAN C. LANDO"	39	27	-30.8%
	SPF - U.31 CENTRO FEDERAL DE DETENCIÓN DE MUJERES	256	229	-10.5%
	SPF - U. 7 FREION REGIONAL DEL NORTE	500	478	-4.4%
	SPF - U.11 COLONIA PENAL DE PRESID. ROQUE SABENZ PEÑA	140	147	5.0%
SPF CHUBUT	SPF - U. 6 INS TITUTO DE SEGURIDAD Y RESOCIALIZACION	450	450	0.0%
	SPF - U.14 CARCEL DE ESQUEL "SUBALCAIDE ABEL R. MLÑOZ"	92	79	-14.1%
SPF FORMOSA	SPF - U.10 CARCEL DE FORMOSA	120	106	-11.7%
	SPF - U.22 CARCEL FEDERAL DE JUJUY	89	111	24.7%
SPF LA PAMPA	SPF - U. 4 COLONIA PENAL DE SANTA ROSA - LA PAMPA	322	299	-7.1%
	SPF - U.13 INS TITUTO CORRECCIONAL DE MUJERES "NTRA. SRA. DEL CARMEN	76	74	-2.6%
	SPF - U.25 INS TITUTO CORRECCIONAL ABIERTO DE GENERAL PICO	47	16	-66.0%
	SPF - U.30 INS TITUTO DE JOVENES ADULTOS	21	20	-4.8%
	SPF - U.17 COLONIA PENAL DE CANDELARIA	176	185	5.1%
	SPF - U. 9 FREION REGIONAL DEL SUR	250	245	-2.0%
SPF RIO NEGRO	SPF - U. 5 COLONIA PENAL "SUBPREFECTO MIGUEL ROCHA"	330	313	-5.2%
	SPF - U.12 COLONIA PENAL DE VIEDMA	198	169	-14.6%
SPF SALTA	SPF - U.23 CARCEL FEDERAL DE SALTA	30	34	13.3%
	SPF - U.15 CARCEL DE RIO GALLEGOS	104	95	-8.7%
SPF CIUDAD BUENOS AIRES	SPF - U. 2 INS TITUTO DE DETENCIÓN DE LA CAPITAL FEDERAL	1800	2095	16.4%
	SPF - U.18 CASA DE PREGRESO "DR. JOSE INGENIEROS"	8	8	0.0%
	SPF - U.20 SERVICIO PSIQUIATRICO CENTRAL DE VARONES	100	110	10.0%
	SPF - U.21 CENTRO PENIT. DE ENFERMEDADES INFECCIOCONTAGIOSAS	37	21	-43.2%
	SPF - U.27 SERVICIO PSIQUIATRICO CENTRAL DE MUJERES	25	13	-48.0%
		9,786	9,625	-1.6%

**Tabla 9.6.** Sobrepopulación Servicio Penitenciario Federal [DNPC, 2005].

**ANEXO 8 - CONTINUACIÓN**  
**SOBREPOBLACIÓN SPF 2006**

PROVINCIA	ESTABLECIMIENTO	CAPACIDAD	POBLACION ALOJADA	SOBREPOBLACION
SPF BUENOS AIRES	SPF - COMPLEJO PENITENCIARIO FEDERAL I - EZEIZA	1817	1619	-10.9%
	SPF - COMPLEJO PENITENCIARIO FEDERAL II - MARCOS PAZ	1500	1530	2.0%
	SPF - U 3 INSTITUTO CORRECCIONAL DE MUJERES	811	696	-14.2%
	SPF - U.19 INSTITUTO CORRECCIONAL ABIERTO DE EZEIZA	380	339	-10.8%
	SPF - CENTRO FED DE TRAT ESPECIAL P/ JÓVENES ADULTOS (AREA CRD)	44	27	-38.6%
	SPF - U.24 INSTITUTO FEDERAL PARA JÓVENES ADULTOS	152	129	-15.1%
	SPF - U.26 INSTITUTO PARA JÓVENES ADULTOS "JUAN C. LANDO"	39	28	-28.2%
	SPF - U.31 CENTRO FEDERAL DE DETENCION DE MUJERES	256	239	-6.6%
	SPF - U. 7 PRISION REGIONAL DEL NORTE	500	431	-13.8%
	SPF - U.11 COLONIA PENAL DE PRESID. ROQUE SAENZ PEÑA	140	113	-19.3%
	SPF CHUBUT	510	430	-15.7%
	SPF - U. 14 CARCEL DE ESQUEL "SUBALCAIDE ABEL R. MUÑOZ"	91	62	-31.9%
	SPF FORMOSA	120	101	-15.8%
SPF JUJUY	151	126	-16.6%	
SPF LA PAMPA	SPF - U. 4 COLONIA PENAL DE SANTA ROSA - LA PAMPA	336	291	-13.4%
	SPF - U.13 INSTITUTO CORRECCIONAL DE MUJERES "NTRA. SRA. DEL CARMEN	76	95	25.0%
	SPF - U.25 INSTITUTO CORRECCIONAL ABIERTO DE GENERAL PICO	47	15	-68.1%
	SPF - U.30 INSTITUTO DE JÓVENES ADULTOS	21	17	-19.0%
	SPF - U.17 COLONIA PENAL DE CANDELARIA	176	191	8.5%
SPF NEUQUEN	250	221	-11.6%	
SPF RIO NEGRO	SPF - U. 5 COLONIA PENAL "SUBPREFECTO MIGUEL ROCHA"	350	296	-15.4%
	SPF - U.12 COLONIA PENAL DE VIEDMA	326	247	-24.2%
SPF SALTA	SPF - U.23 CARCEL FEDERAL DE SALTA	30	30	0.0%
SPF SANTA CRUZ	SPF - U.15 CARCEL DE RIO GALLEGOS	99	86	-13.1%
SPF CIUDAD BUENOS AIRES	SPF - U. 2 INSTITUTO DE DETENCION DE LA CAPITAL FEDERAL	1800	1882	4.6%
	SPF - U.18 CASA DE PREENGRESO "DR. JOSE INGENIEROS"	8	8	0.0%
	SPF - U.20 SERVICIO PSIQUIATRICO CENTRAL DE VARONES	100	104	4.0%
	SPF - U.21 CENTRO PENIT. DE ENFERMEDADES INFECCIOSAS	30	18	-40.0%
	SPF - U.27 SERVICIO PSIQUIATRICO CENTRAL DE MUJERES	25	9	-64.0%
		10,185	9,380	-7.9%

**Tabla 9.5.** Sobrepopulación Servicio Penitenciario Federal [DNPC, 2006].

**ANEXO 9**

**SOBREPOBLACIÓN SERVICIO PENITENCIARIO BONAERENSE – AÑO 2006**

PROVINCIA	ESTABLECIMIENTO	CAPACIDAD	DETENIDOS	SOBREPOBLACION	
BUENOS AIRES	UNIDAD 1 - LISANDRO OLMOS	1,740	1,727	-0.7%	
	UNIDAD 2 - SIERRA CHICA	1,550	1,303	-15.9%	
	UNIDAD 3 - SAN NICOLAS	400	463	15.8%	
	UNIDAD 4 - BAHIA BLANCA	592	468	-20.9%	
	UNIDAD 5 - MERCEDES	717	711	-0.8%	
	UNIDAD 6 - DOLORES	327	327	0.0%	
	UNIDAD 7 - AZUL	266	302	13.5%	
	UNIDAD 8 - LOS HORNOS	157	168	7.0%	
	UNIDAD 9 - LA PLATA	1,250	1,217	-2.6%	
	UNIDAD 10 - MELCHOR ROMERO	280	284	1.4%	
	UNIDAD 11 - BARADERO	66	56	-15.2%	
	UNIDAD 12 - JOAQUIN GORINA	98	20	-79.6%	
	UNIDAD 13 - JUNIN	748	557	-25.5%	
	UNIDAD 14 - GENERAL ALVEAR	48	23	-52.1%	
	UNIDAD 15 - BATAN	1,386	1,227	-11.5%	
	UNIDAD 16 - JUNIN	140	138	-1.4%	
	UNIDAD 17 - URDAMPILLETA	450	424	-5.8%	
	UNIDAD 18 - JOAQUÍN GORINA	582	267	-54.1%	
	UNIDAD 19 - SAAVEDRA	650	584	-10.2%	
	UNIDAD 20 - TRENQUEN LAUQUEN	80	57	-28.8%	
	UNIDAD 21 - CAMPANA	750	724	-3.5%	
	UNIDAD 22 - HO.G.A.M.	52	29	-44.2%	
	UNIDAD 23 - FLORENCIO VARELA	772	778	0.8%	
	UNIDAD 24 - FLORENCIO VARELA	750	751	0.1%	
	UNIDAD 25 - LISANDRO OLMOS	95	238	150.5%	
	UNIDAD 26 - LISANDRO OLMOS	94	109	16.0%	
	UNIDAD 27 - SIERRA CHICA	140	128	-8.6%	
	UNIDAD 28 - MAGDALENA	646	572	-11.5%	
	UNIDAD 30 - GENERAL ALVEAR	1,550	1,528	-1.4%	
	UNIDAD 31 - FLORENCIO VARELA	499	487	-2.4%	
	UNIDAD 32 - FLORENCIO VARELA	468	438	-6.4%	
	UNIDAD 33 - LOS HORNOS (MUJERES)	338	245	-27.5%	
	UNIDAD 34 - MELCHOR ROMERO	362	354	-2.2%	
	UNIDAD 35 - MAGDALENA	850	760	-10.6%	
	UNIDAD 36 - MAGDALENA	650	548	-15.7%	
	UNIDAD 37 - BARKER	650	562	-13.5%	
	UNIDAD 38 - SIERRA CHICA	650	644	-0.9%	
	UNIDAD 39 - ITUZAINGÓ	400	375	-6.3%	
	UNIDAD 41 - CAMPANA	300	100	-66.7%	
	UNIDAD 42 - FLORENCIO VARELA	102	171	67.6%	
	UNIDAD 44 - MAR DEL PLATA	45	96	113.3%	
	UNIDAD 45 - MELCHOR ROMERO	250	178	-28.8%	
	UNIDAD 50 - MAR DEL PLATA	96	52	-45.8%	
	UNIDAD 51 - MAGDALENA	32	55	71.9%	
	UNIDAD 52 - AZUL	96	61	-36.5%	
	Diferencias Registro Gral. de Internos del S.P. de la Prov. de Bs As.			3,572	0.0%
			<b>22,164</b>	<b>23,878</b>	<b>7.7%</b>

**Tabla 9.6.** Sobrepopulación Servicio Penitenciario Bonaerense [DNPC, 2006].

**ANEXO 10**

**APERTURA DEL PRESUPUESTO 2004 DEL SPF SEGÚN CONCEPTO**

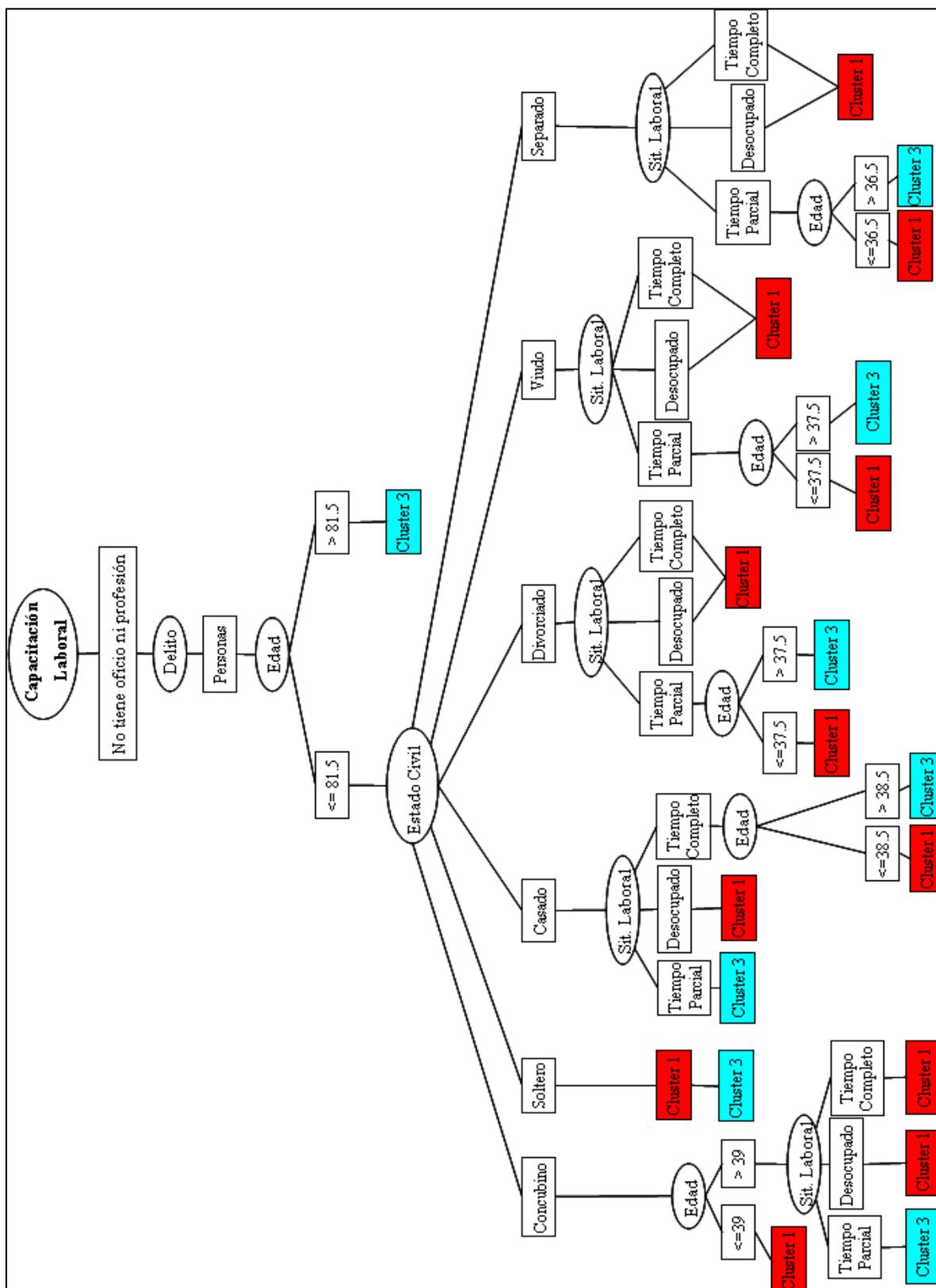
<b>Actividades Comunes al Servicio Penitenciario Federal (Programas 16,17 y 15)</b>	Gastos de Personal	\$22,935,690
	Servicios no personales	\$12,290,829
	Bienes de consumo	\$11,009,522
	Bienes de uso	\$645,095
	Transferencias	\$14,246,000
		<b>\$61,127,136</b>
<b>Seguridad y Rehabilitación del interno</b>	Gastos de Personal	\$128,772,059
	Servicios no personales	\$4,238,059
	Bienes de consumo	\$18,095,960
	Bienes de uso	\$420,905
		<b>\$151,526,983</b>
<b>Pagos a Retirados y Pensionados</b>	Gastos de Personal	\$1,107,277
	Servicios no personales	\$1,625
	Transferencias	\$124,737,300
		<b>\$125,846,202</b>
<b>Cooperación Técnica y Financiera para la Laborterapia de Internos</b>	Gastos de Personal	\$ 686,190
	Servicios no personales	\$ 4,772,200
	Bienes de consumo	\$ 1,414,000
	Bienes de uso	\$ 798,000
		<b>\$ 7,670,390</b>
<b>Formación y Capacitación</b>	Gastos de Personal	\$10,389,974
	Servicios no personales	\$1,308
	Bienes de consumo	\$998,971
	Transferencias	\$353,000
		<b>\$ 11,743,253</b>
	<b>Sub-Total</b>	<b>\$357,913,964</b>
	(+) Obras Penitenciarias	\$14,460,295
	<b>Total</b>	<b>\$372,374,259</b>

Figura 9.6. Presupuesto destinado al SPF [Petrone, 2005].



**ANEXO 12 A**

**RAMA “SIN OFICIO NI PROFESIÓN – DELITO CONTRA LA PERSONAS”**



**Figura 9.8.** Rama “Sin Oficio ni Profesión – Delito contra las Personas”.

**ANEXO 12 B**

**RAMA “SIN OFICIO NI PROFESIÓN – DELITO CONTRA LA PROPIEDAD”**

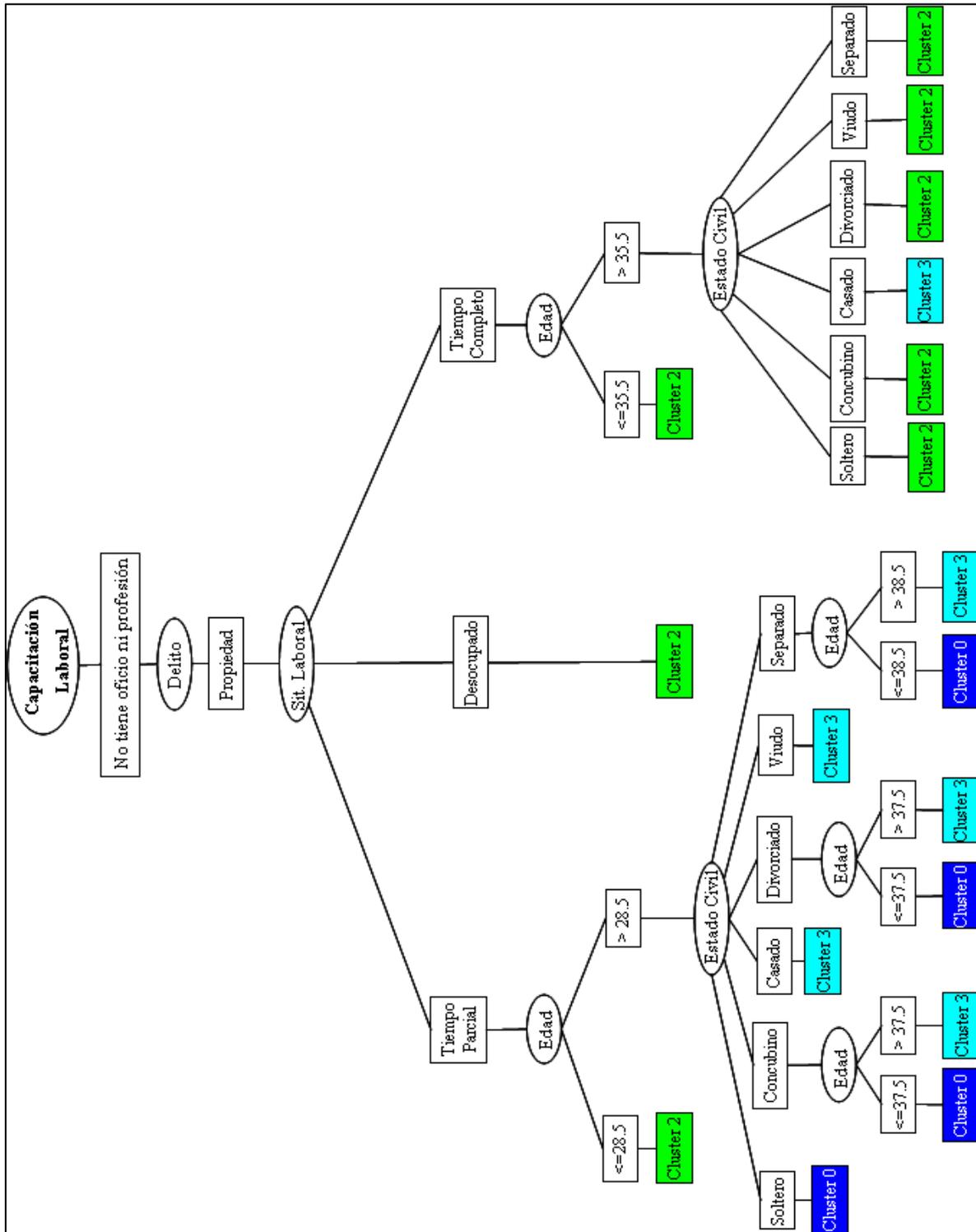


Figura 9.9. Rama “Sin Oficio ni Profesión – Delito contra la Propiedad”.

**ANEXO 12 C**

**RAMA “SIN OFICIO NI PROFESIÓN – DELITO CONTRA LA INTEGRIDAD SEXUAL Y EL HONOR”**

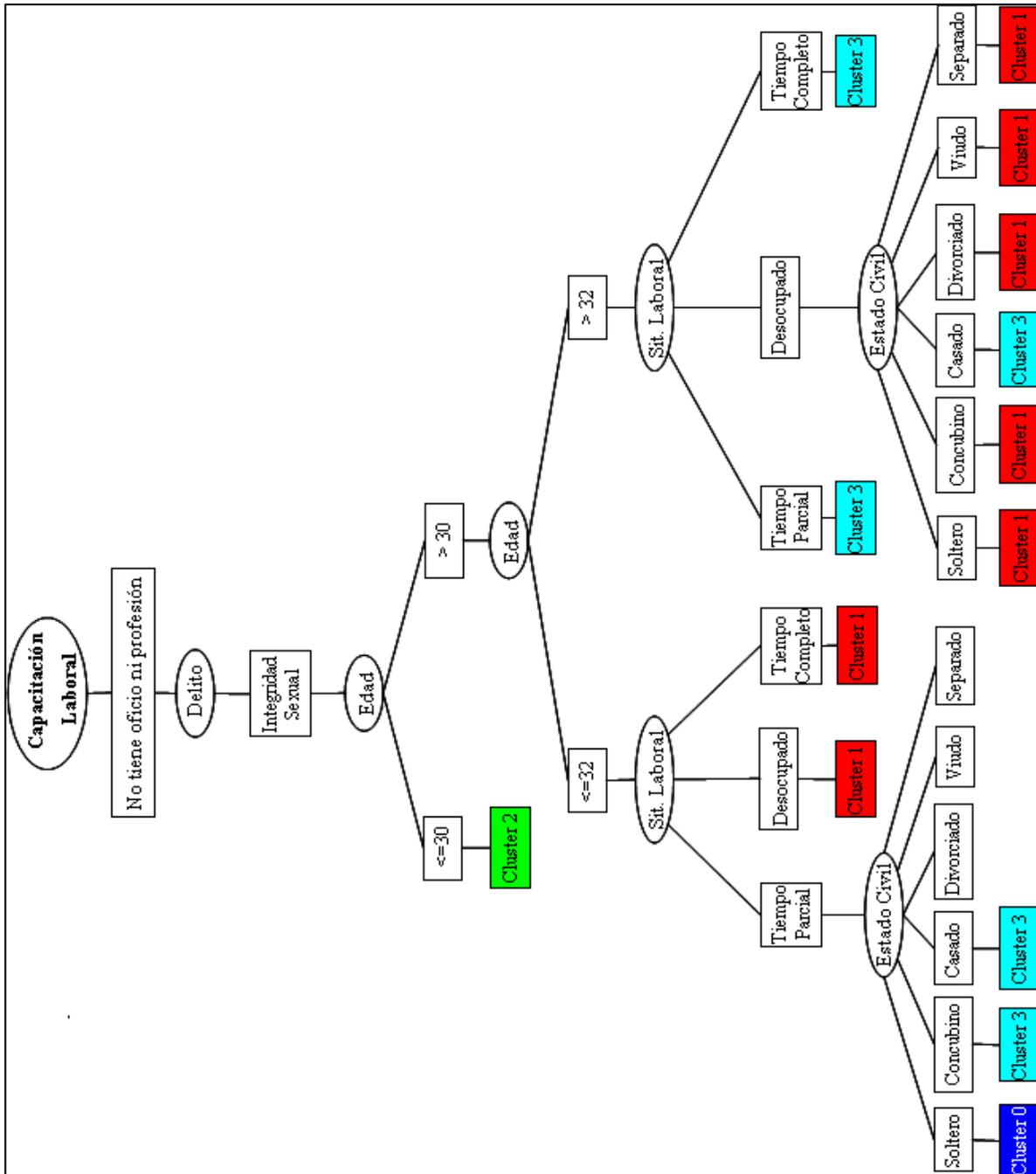
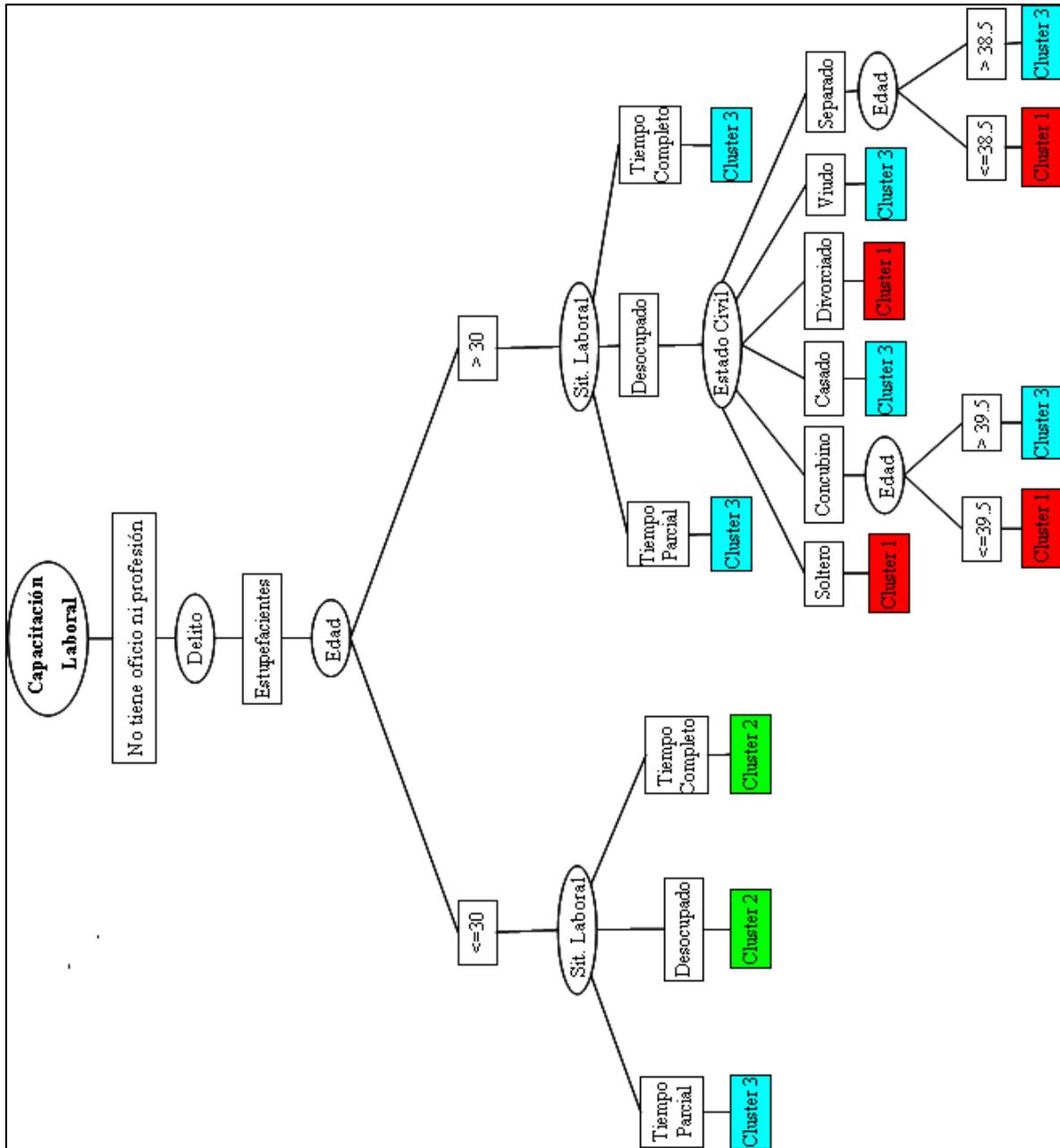


Figura 9.10. Rama “Sin Oficio ni Profesión – Delito contra la Integridad Sexual y el Honor”.

**ANEXO 12 D**

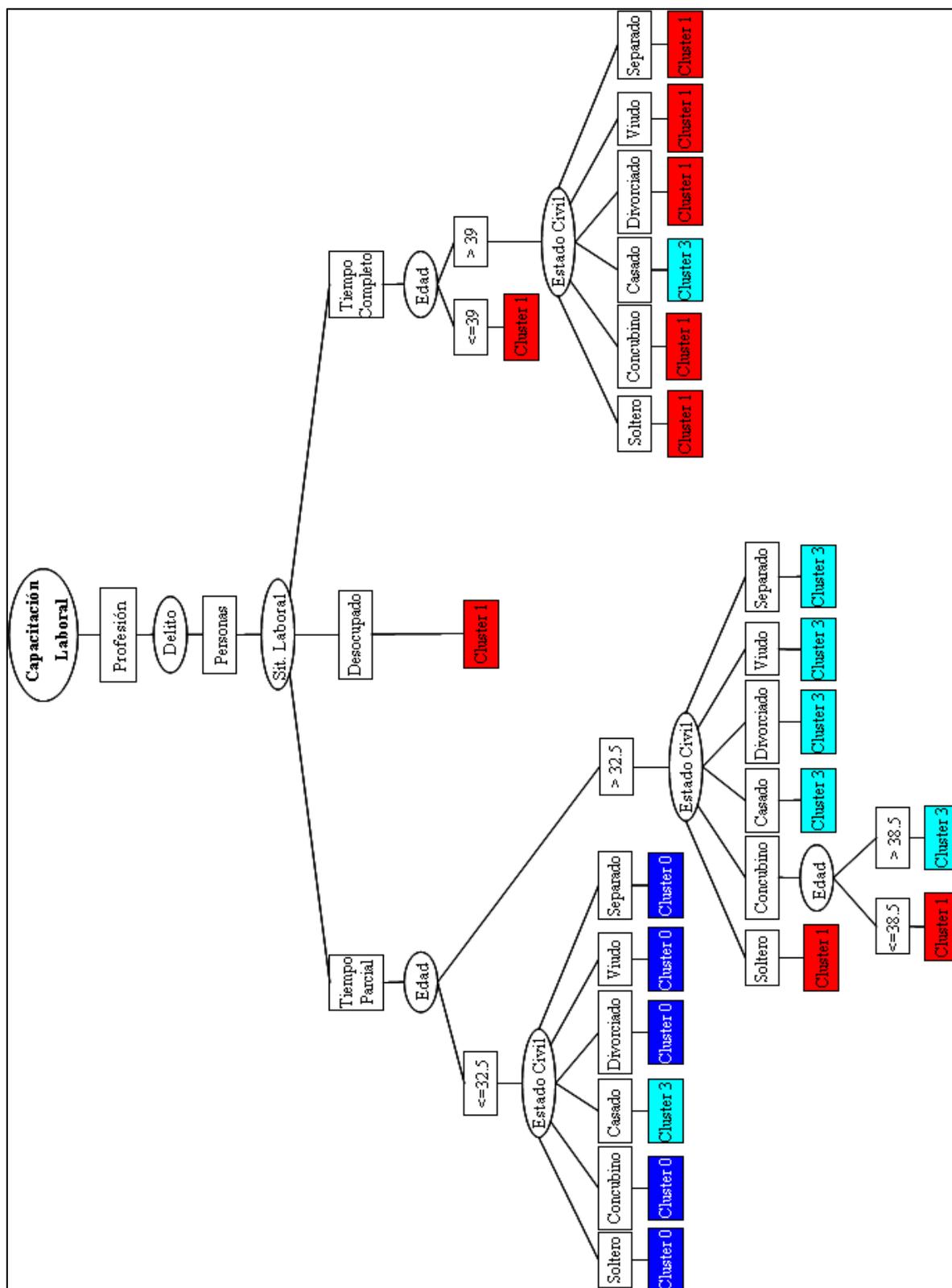
**RAMA “SIN OFICIO NI PROFESIÓN – DELITO CONTRA LEY N° 23.737 DE ESTUPEFACIENTES”**



**Figura 9.11.** Rama “Sin Oficio ni Profesión – Delito contra Estupeficientes”.

**ANEXO 13 A**

**RAMA “PROFESIÓN – DELITO CONTRA LAS PERSONAS”**



**Figura 9.12.** Rama “Profesión– Delito contra las Personas”.

**ANEXO 13 B**

**RAMA “PROFESIÓN – DELITO CONTRA LA PROPIEDAD”**

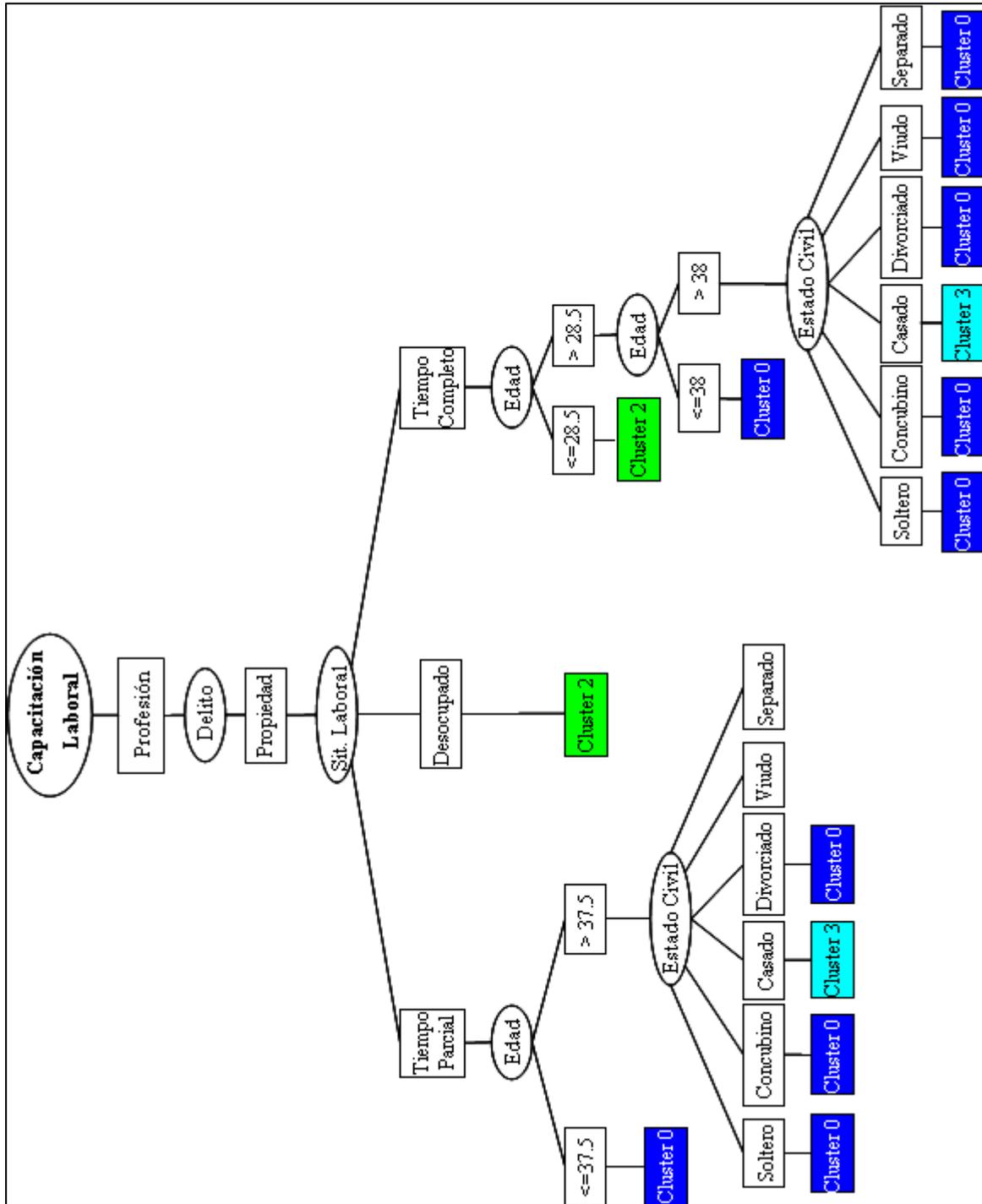
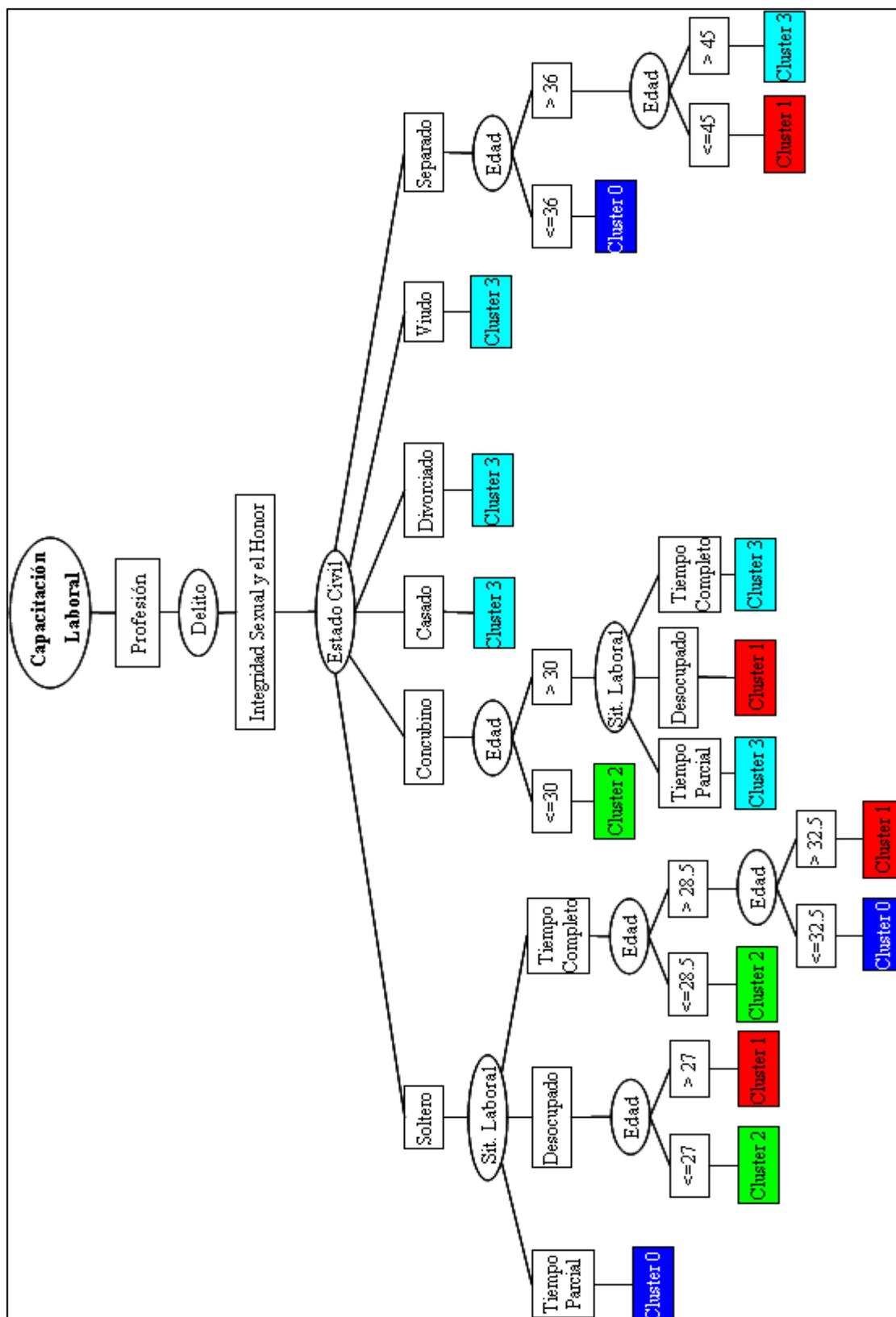


Figura 9.13. Rama “Profesión – Delito contra la Propiedad”.

**ANEXO 13 C**

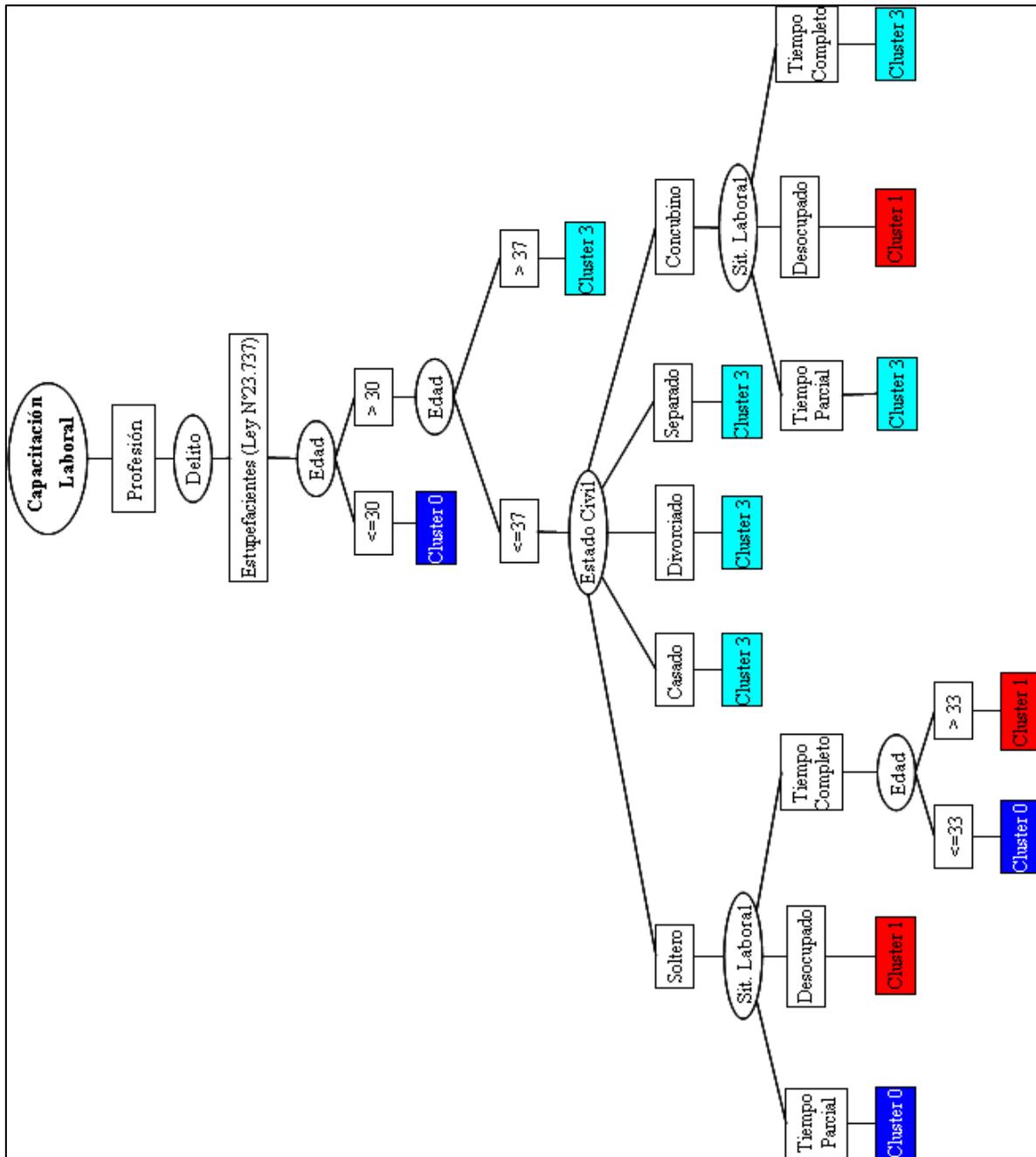
**RAMA “PROFESIÓN – DELITO CONTRA LA INTEGRIDAD SEXUAL”**



**Figura 9.14.** Rama “Profesión – Delito contra la Integridad Sexual”.

**ANEXO 13 D**

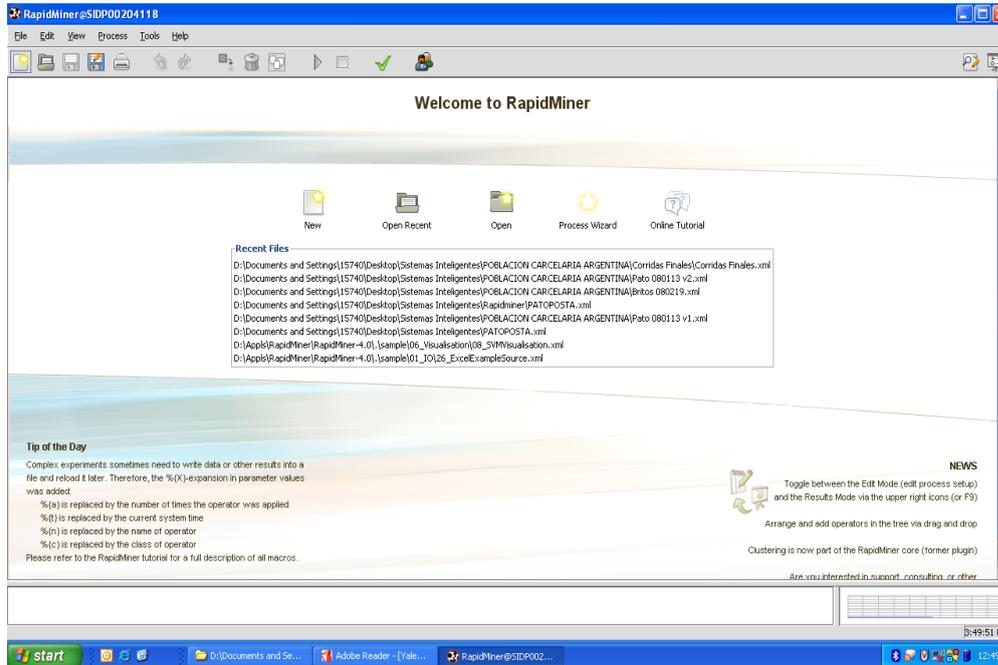
**RAMA “PROFESIÓN – DELITO DE ESTUPEFACIENTES”**



**Figura 9.15.** Rama “Profesión - Delito contra la ley N° 23.737 de Estupefacentes”.

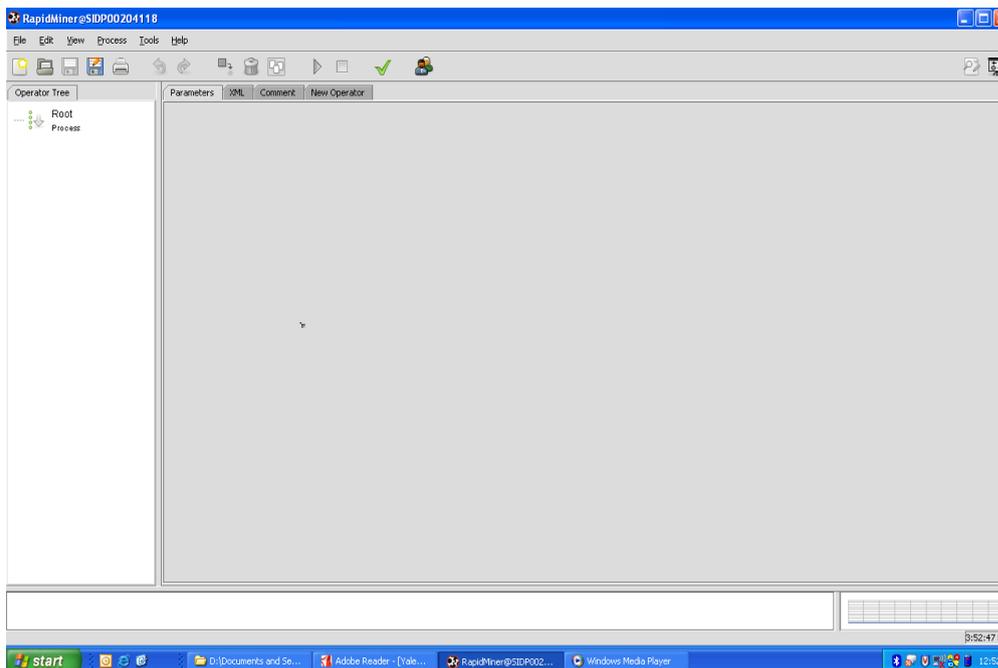
## **ANEXO 14**

### **INTRODUCCIÓN AL RAPIDMINER (YALE)**



**Figura 9.16.** Pantalla de Bienvenida del *RapidMiner*.

Una vez elegida alguna de las opciones de la pantalla de Bienvenida [Figura 9.16] como lo puede ser iniciar nuevos procesos, abrir existentes, utilizar asistente de creación, etc., se pasa a la siguiente pantalla [Figura 9.17]:



**Figura 9.17.** Pantalla Principal del *RapidMiner*.

Una vez dentro de la pantalla principal del programa, se pueden realizar las siguientes acciones:

- Crear Operadores:

- Usando el icono de nuevo operador. 
- Usando el botón derecho del *mouse* ubicándose en la lengüeta del Árbol de Operadores sobre cualquier operador que lo permita, se puede elegir un nuevo operador.

- Reemplazar Operadores:

- Usando el botón derecho del *mouse* ubicándose en la lengüeta del Árbol de Operadores sobre cualquier operador que lo permita, se puede reemplazar un nuevo operador.

- Insertar Bloques de Operadores guardados:

- Usando el icono de bloque guardado. 
- Usando el botón derecho del *mouse* ubicándose en la lengüeta del Árbol de Operadores sobre cualquier operador que lo permita, se agrega un bloque de operadores.

- Eliminar Operadores:

- Usando el icono de remoción de operadores. 
- Usando el botón derecho del *mouse* ubicándose en la lengüeta del Árbol de Operadores sobre cualquier operador que lo permita, se elige remover operador.

- Manejo de parámetros:

- A partir de la lengüeta de parámetros se puede pasar de modo normal a modo experto con el icono correspondiente: 

Se observa en la figura 9.18 como se incorpora el *dataset* al modelo. Importante destacar, que el presente software tiene la capacidad de permitir que el *input operator* acepte bases en distintos formatos. De la misma manera, en la figura 9.19, se muestra como se van incorporando al modelo los distintos algoritmos de minería de datos. En el caso de la presente figura, se está agregando un algoritmo basado en árboles de decisión, que tal como se explica anteriormente, es un método supervisado de aprendizaje.

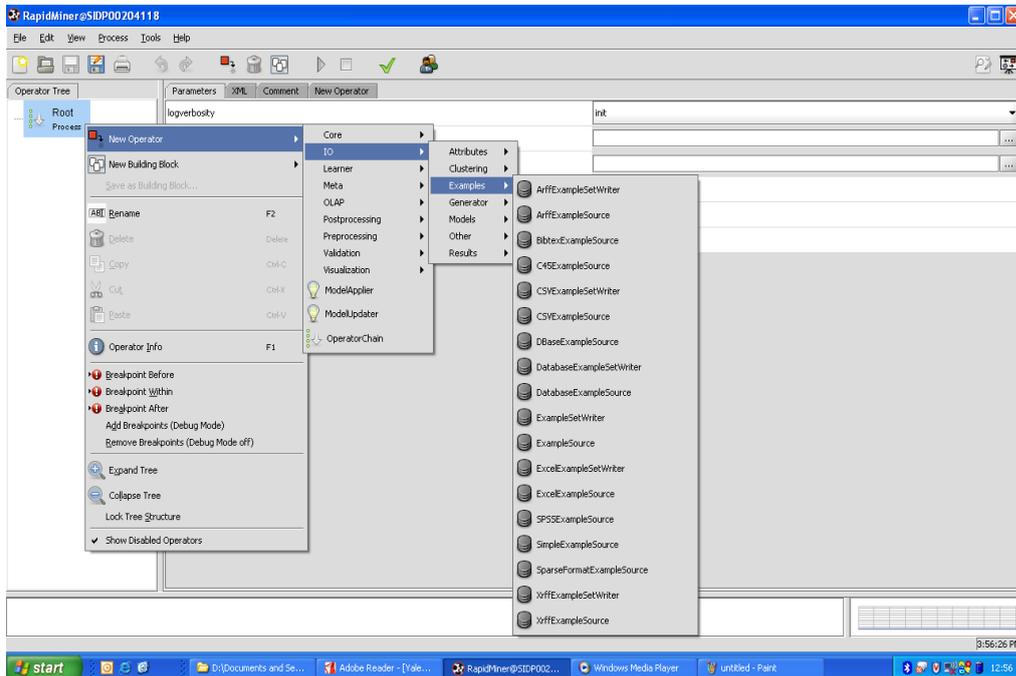


Figura 9.18. Agregado del *dataset* al modelo de minería de datos.

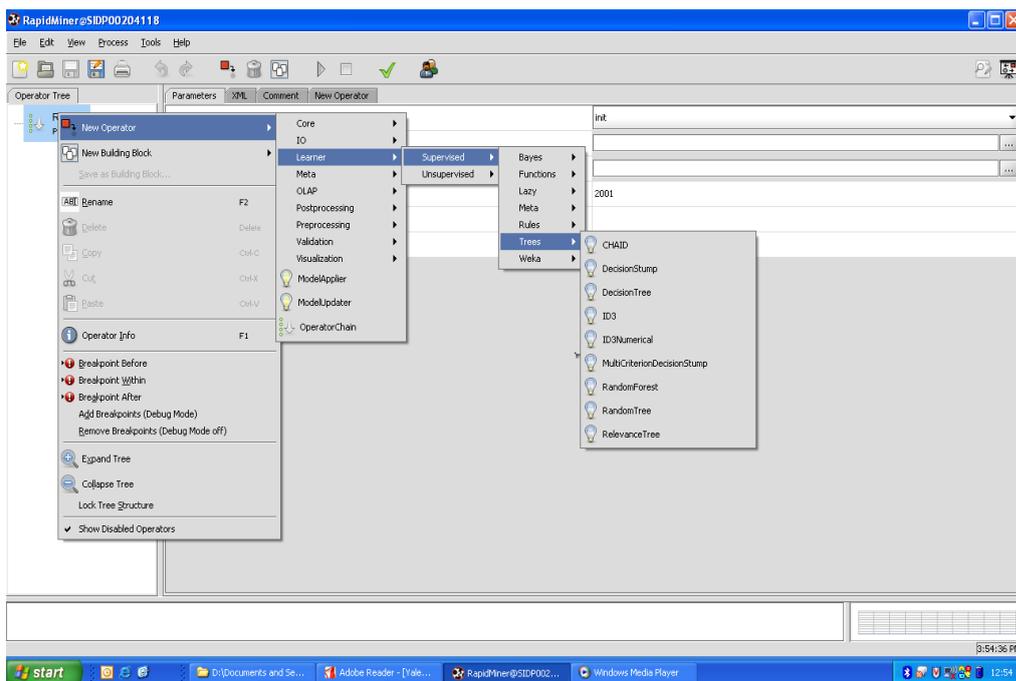


Figura 9.19. Agregado del algoritmo de minería de datos.

Una vez cargado el *dataset*, como agregados los operadores de aprendizaje necesarios para el proceso, se recomienda seguir los siguientes pasos:

- Validación de la definición de Proceso:
  - Una vez que se agregaron todos los operadores necesarios para un proceso, es posible chequear si estos fueron jerarquizados como

corresponde y si tienen sus entradas (*inputs*) correctas y sus propiedades bien *seteadas*.

- Para realizar la validación del proceso se utiliza el siguiente icono: 
- Los resultados se observan en el visor de mensajes (pie de la pantalla) y en caso de que existiera algún tipo de error con la estructura y jerarquía de los operadores y/o el proceso, aparece un signo de exclamación junto al operador.
- Ejecución del proceso:
  - Una vez validado el proceso, la ejecución consiste en hacer *click* en el icono de ejecución: 
  - Se puede detener el proceso mediante el icono de *stop*: 
- Agregados de Puntos de Corte (BreakPoint):
  - Si nos ubicamos en alguno de los operadores en la lengüeta del árbol de operadores, usando el botón derecho del *mouse*, se puede elegir colocar un *Breakpoint* para que cuando la ejecución se inicie se pueda frenar en dicho punto y luego reanudarla.
- Evaluación de Resultados:
  - Una vez terminada la ejecución, los resultados se pueden observar automáticamente en pantalla, en lo que se denomina Modo de Resultados (*Result Mode*). 
  - De la misma forma, se puede volver a la pantalla de edición (*Edit Mode*) con el siguiente icono: 
  - Desde la vista de resultados es posible, dependiendo del tipo de resultados que arroja el modelo, obtener gráficos de distinto tipo.

A modo de ejemplo, se muestra a continuación la figura 9.20, en donde se encuentra desarrollado un árbol de operadores a punto de ser ejecutado para realizar el proceso de clasificación con su correspondiente análisis de *performance*.

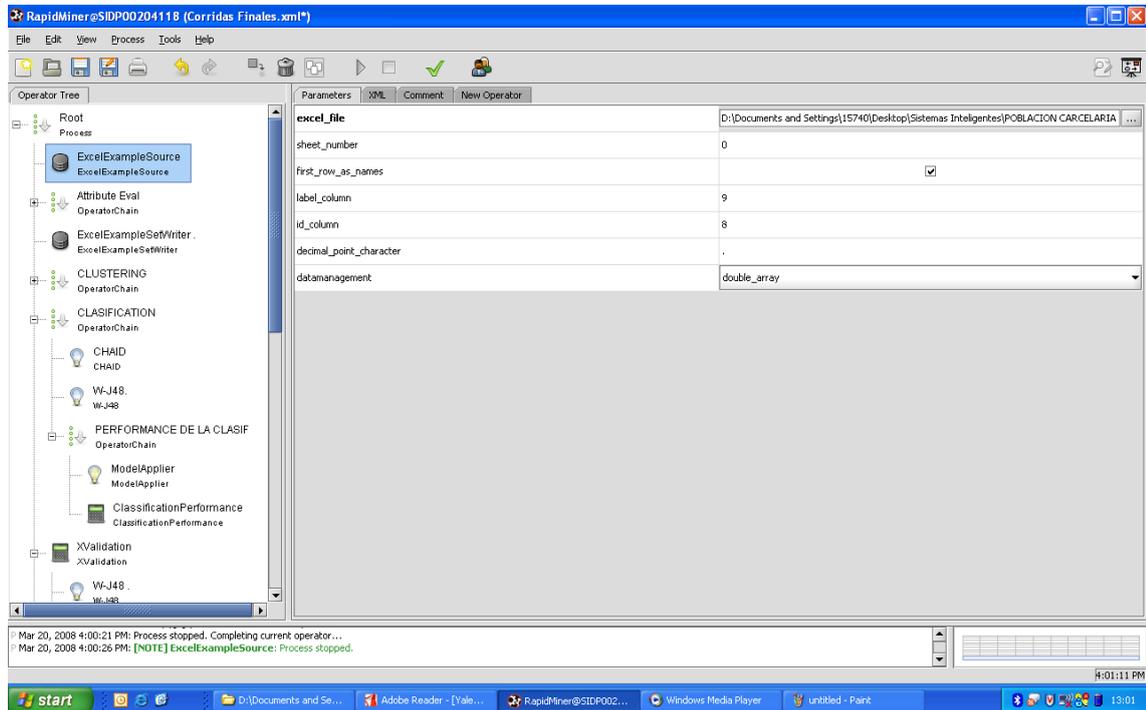


Figura 9.20. Ejemplo de árbol de operadores.