



UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

DESARROLLO DE UNA METODOLOGÍA PARA CARACTERIZAR Y PREDECIR EL  
RIESGO CRIMINAL MEDIANTE LA GENERACIÓN DE MODELOS  
ESPACIO-TEMPORALES EMPÍRICOS BASADOS EN MANEJO DE DATOS.

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIAS DE LA  
INGENIERÍA, MENCIÓN ELÉCTRICA

JUAN PABLO RODRÍGUEZ SALAVARRÍA

PROFESOR GUÍA:  
MARCOS ORCHARD CONCHA

MIEMBROS DE LA COMISIÓN:  
JORGE SILVA SÁNCHEZ  
CARLOS MUÑOZ POBLETE

SANTIAGO DE CHILE  
2017



RESUMEN DE LA MEMORIA PARA OPTAR  
AL TÍTULO DE MAGÍSTER EN CIENCIAS DE LA INGENIERÍA, MENCIÓN ELÉCTRICA  
POR: JUAN PABLO RODRÍGUEZ SALAVARRÍA  
FECHA: 2017  
PROF. GUÍA: SR. MARCOS ORCHARD CONCHA

## DESARROLLO DE UNA METODOLOGÍA PARA CARACTERIZAR Y PREDECIR EL RIESGO CRIMINAL MEDIANTE LA GENERACIÓN DE MODELOS ESPACIO-TEMPORALES EMPÍRICOS BASADOS EN MANEJO DE DATOS.

Con el incremento de la delincuencia se ha hecho necesario que las fuerzas de seguridad busquen más herramientas que les permitan analizar los procesos delictivos y evaluar el posible riesgo criminal, con el fin de poder determinar las acciones tácticas, asignación de recursos policiales y planificaciones a corto y largo plazo para enfrentar este problema.

Una de las formas de realizar un análisis de los delitos que ocurren en una ciudad es mediante el análisis de patrones criminales, a través del modelamiento y simulación de datos que entreguen la información necesaria para el estudio de cada una de las variables que intervienen en cada hecho. De esta forma se podrá caracterizar al criminal, su forma de actuar, sus preferencias; factores que servirán para poder predecir hechos que atenten contra de la seguridad ciudadana.

En la actualidad existen en el mercado códigos y programas que permiten realizar predicciones de qué zona podrá tener un alto índice de criminalidad, pero los costos de estos programas son muy elevados.

La propuesta de este proyecto es generar y verificar un método para caracterizar y predecir el riesgo criminal de un área urbana determinada, usando una base de datos generada en forma artificial. Para la generación de la base de datos se usa un algoritmo que permite al usuario seleccionar un número de zonas urbanas a considerar, las mismas que están constituidas por servicios públicos o privados de todo tipo, y que una vez que se han elegido las zonas el algoritmo generará en forma aleatoria delitos asociados a cada servicio.

Los datos generados son utilizados para verificar una metodología de caracterización del riesgo criminal espacial. Pero no basta con esta distribución inicial, ya que el proceso delictivo está en constante cambio. Por esta razón es necesario simular con la inclusión de la variable temporal, mediante la inserción de nueva información de eventos delictivos y la aplicación de un modelo dinámico, donde se obtendrá una distribución espacial del riesgo actualizada. Para obtener un resultado óptimo del modelo espacial, se utiliza un algoritmo de optimización para encontrar la actualización adecuada de la distribución, siendo el resultado final el que se usará en la etapa de predicción.

Para la etapa de predicción se realiza un agrupamiento o “clustering” de la nueva información (nuevos eventos delictivos), incluidos en la etapa temporal, con la finalidad de encontrar los posibles centros de las zonas de riesgo (atractores criminales). Una vez que se tienen los centros, por medio de un modelo empírico de un atractor criminal y con una determinada ventana de tiempo, se genera una distribución de predicción del riesgo criminal.



*A Paty, Luis, Marcia y Gaby.*



# Agradecimientos

Primero quiero agradecer a Dios porque sin Él nada hubiera sido posible. A mis papás y hermana, que sin ellos no hubiéramos podido terminar con mi esposa esté gran logro de nuestras vidas, gracias por apoyarnos en los momentos más difíciles, gracias por estar conmigo en esta etapa de mi vida.

A mi esposa, mi compañera de vida, aprendo cada día que estar contigo es la fortaleza de mi vida, por todo lo que me das, por todo lo que me otorgas, por tu amor incondicional, por tu apoyo y comprensión en esta etapa de nuestras vidas, gracias mi amor.

Quiero agradecer a mis grandes amigos y maestros Fredy y Tedy, que estuvieron conmigo en los momentos más difíciles, siempre con palabras de apoyo, escuchándome, siempre les estaré muy agradecido mis cuates.

A mis compañeros del grupo de laboratorio, en especial a Francisco que me ayudó desde el primer día que llegué a la universidad que, aunque estábamos en áreas totalmente diferentes siempre hubo mucha apertura de su parte en ayudarme en todo y de igual forma cuando estuvimos trabajando en la misma área y con este proyecto tesis, mil gracias.

Siempre estaré muy agradecido con el Profe Marcos, me recibió en su área de investigación en un momento muy difícil de mis estudios, siempre tuvo la apertura de ayudarme desde el primer día en que empecé a trabajar con él y por todo su apoyo y guía para la culminación de este trabajo muchas gracias.

Finalmente quiero agradecer a la comisión por sus aportes en mi trabajo de tesis y a todos mis amigos de Ecuador y de todos los países que tuve la gran suerte de conocer, todos hicieron que mi vida esté más llena de grandes experiencias.





# Tabla de Contenido

|   |           |
|---|-----------|
| <b>1. Introducción</b>  | <b>1</b>  |
| <b>2. Marco Teórico y Estado del Arte</b>   | <b>5</b>  |
| 2.1. Teoría sobre el Análisis de la Delincuencia. . . . .   | 5         |
| 2.2. Generadores del Crimen y Atractores del Crimen. . . . .  | 7         |
| 2.3. Hot-Spot . . . . .   | 8         |
| 2.4. Modelos de Mixtura de Gaussianas (GMM) . . . . .   | 8         |
| 2.5. Inferencia Bayesiana e Integración de Monte Carlo . . . . .  | 9         |
| 2.6. Muestreo de Importancia (Importance Sampling) . . . . .  | 11        |
| 2.7. Re-muestreo (Resampling) . . . . .   | 12        |
| 2.8. Particle Swarm Optimization ( <i>PSO</i> ) . . . . .   | 13        |
| 2.9. K-means Clustering . . . . .   | 15        |
| 2.10. Método para evaluar los modelos: HRP vs. TIP . . . . .  | 18        |
| <b>3. Descripción de la metodología empleada para la caracterización y predicción del riesgo criminal</b> | <b>19</b> |
| 3.1. Área de Interés y Datos Geo-referenciados a los Servicios . . . . .                                  | 20        |
| 3.2. Modelos Espaciales de Riesgo . . . . .   | 21        |
| 3.3. Evolución Temporal. . . . .  | 22        |
| 3.4. Predicción. . . . .  | 26        |
| <b>4. Análisis y Discusión de los Resultados Obtenidos</b>  | <b>28</b> |
| 4.1. Área de interés seleccionada para el proyecto . . . . .  | 28        |
| 4.2. Generación de modelos espaciales de riesgo . . . . .   | 29        |
| 4.3. Evolución Temporal. . . . .  | 30        |
| 4.4. Predicción. . . . .  | 35        |
| 4.5. Evaluación de los Resultados. . . . .  | 39        |
| <b>5. Conclusión</b>  | <b>42</b> |
| <b>Bibliografía</b>   | <b>45</b> |

# Índice de Tablas

|  |    |
|--|----|
| 4.1. Resultados Finales con PSO para 50 iteraciones y 30 partículas. . . . .   | 33 |
| 4.2. Resultados de la Evaluación de la Predicción (ventana 131) . . . . .      | 40 |
| 4.3. Resultados de la Evaluación de la Predicción (ventana 417) . . . . .      | 40 |
| 4.4. Resultados de la Evaluación de la Predicción (ventana 399 datos). . . . . | 41 |

# Índice de Ilustraciones

|   |    |
|---|----|
| 2.1. Ejemplo de gráfico HRP vs. TIP [35]. . . . .   | 18 |
| 3.1. Ejemplo de representación de los servicios y sus delitos asociados mediante GMM. . . . .                                   | 22 |
| 3.2. Resultado final de la aproximación de la distribución a priori después de aplicar Important Sampling y Resampling. . . . . | 23 |
| 3.3. Movimiento de las partículas hacia las nuevas observaciones. . . . .   | 25 |
| 4.1. Mapa de la Comuna de Santiago Centro. . . . .  | 29 |
| 4.2. Área de Interés con los servicios seleccionados y sus delitos asociados. . . . .   | 29 |
| 4.3. Modelo GMM de las zonas de riesgo . . . . .  | 30 |
| 4.4. Grillado del mapa para la colocación de las muestras. . . . .  | 30 |
| 4.5. Distribución uniforme de las muestras. . . . .   | 31 |
| 4.6. Resampling. . . . .  | 31 |
| 4.7. Área de Interés con las partículas (rojo) y las nuevas observaciones (verde). . . . .                                      | 32 |
| 4.8. Gráfica HRP vs. TIP con el criterio de HRP 15 % y TIP=0.9375 . . . . .   | 34 |
| 4.9. Gráfica HRP vs. TIP con el criterio de HRP 15 % y TIP=0.8900 . . . . .   | 34 |
| 4.10. Área de Interés con datos a priori (rojo), datos posteriori (azul) y nuevas observaciones (verde). . . . .                | 35 |
| 4.11. Resultado del Índice de Dunn. . . . .   | 35 |
| 4.12. Resultado del Índice de Davies-Bouldin. . . . .   | 36 |
| 4.13. Resultado del Índice Silhouette. . . . .  | 36 |
| 4.14. Resultado del Índice Sum of The Square Error (SSE). . . . .   | 37 |
| 4.15. Resultado de la Función de Costo. . . . .   | 37 |
| 4.16. Clustering de las nuevas observaciones (verde) con sus respectivos centroides (cruz fucsia) para k=2. . . . .             | 38 |
| 4.17. Partículas atraídas hacia los centroides con un horizonte de predicción igual a 400. . . . .                              | 39 |
| 4.18. Distribución de Predicción. . . . .   | 39 |



# Capítulo 1

## Introducción

El miedo al crimen es creado por situaciones y escenarios que hacen que la gente se sienta vulnerable a la victimización [4]. Los escenarios urbanos y el miedo que genera el crimen son construcciones humanas, siendo un subproducto de esos ambientes construidos para satisfacer los requerimientos de la vida cotidiana (hogares, barrios residenciales, locales comerciales, oficinas, construcciones gubernamentales, etc.); dependiendo como se encuentran distribuidos, pueden tener un enorme impacto en nuestros niveles de miedo, ya sea por las cantidades, tipos y horarios en los que ocurren los crímenes [4].

Los crímenes son creados por la interacción de un potencial delincuente con un potencial objetivo en un escenario, que lo hacen más fácil, seguro y rentable [4]. Según estudios, se ha visto que el “medio ambiente”, a menudo complementado por los hechos delictivos ocurridos en el pasado de un determinado sector, es un factor que atrae a la delincuencia, es decir existe una conexión entre la tasa de crímenes y los atributos de la comunidad [3][30].

Las condiciones del ambiente son un atractor para la delincuencia [4], esto es, que la falta de seguridad, venta de drogas, prostitución, son factores que vuelven a un determinado sector o región atractiva para el delincuente [30]. También hay que considerar que los criminales seleccionan la locación donde cometerán un delito de acuerdo a sus preferencias o dependiendo de la utilidad percibida [30], siendo este un factor más psicológico del crimen. Una forma de explicar las acciones criminales, considerando lo anterior, es la “teoría ecológica”. Esta teoría busca describir las motivaciones y actos del crimen basado en las características generales de un ambiente en particular [30].

En la actualidad existe una gran variedad de proyectos sobre criminalidad, que tienen como objetivo entender cómo actúa el criminal, qué patrones se hacen presentes en cada una de sus acciones, cuáles son sus preferencias, predicción del delito. Existe también una cooperación entre varias ciencias para encontrar soluciones a esta problemática social. Lo que buscan es predecir los actos criminales, de esta forma las fuerzas de seguridad podrán mantener el control y la tranquilidad en las calles; además, el propósito del uso de estos predictores es que se reduzcan los costos de operaciones y la gestión óptima de sus recursos. Todo esto beneficiará a la comunidad mejorando el nivel de vida de las ciudades.

Dentro de los proyectos que existen en la actualidad está la memoria de la Srta. Paulina Flores [15], donde se presenta un método empírico para la caracterización del riesgo, además una caracterización de la evolución temporal de los focos de criminalidad y finalmente con la distribución del riesgo obtenido de la evolución temporal se implementa la predicción de focos de criminalidad. El método empleado utiliza información geo-referenciada de eventos criminales y servicios para generar una distribución espacial de riesgo a través de la suma de kernels Gaussianos. La caracterización de la evolución temporal se realiza mediante un algoritmo basado en métodos secuenciales de Monte Carlo, introduciendo en forma secuencial nuevos eventos y aproximando una distribución mediante la posición de partículas en el plano, donde las partículas son el resultado de la estimación de la distribución de la información geo-referenciada. Una vez que se obtiene la distribución del riesgo en la evolución temporal se define un modelo de predicción, actualizando con nuevos datos en el corto y largo plazo, modelando los cambios a medida que el entorno cambia, de esta forma se tiene que los eventos futuros ocurren dentro de áreas de riesgo.

Los métodos empleados en la memoria tienen ciertas limitaciones, principalmente en el ajuste del modelo dinámico y en la predicción. Respecto al ajuste del modelo dinámico la metodología presentada en la memoria se basa principalmente en el criterio del usuario, de esta forma es complicado obtener el ajuste óptimo. En cuanto a la predicción, se tiene la limitante que con ciertas ventanas de tiempo aparecen partículas fuera del perfil geográfico definido. Además, es necesario probar para una posible extensión del proyecto, los modelos planteados en la memoria, de esta forma ver si puede existir una generalización en los métodos empleados en la caracterización y en la predicción.

Con estos antecedentes se propone en este proyecto optimizar el procedimiento de ajuste del modelo temporal y una nueva metodología de predicción de las áreas de riesgo criminal. El uso de la “*teoría ecológica*” como base para el análisis y predicción del riesgo criminal de una zona determinada en este proyecto permitirá abordar el estudio de zonas donde existe altos índices de criminalidad. La razón del uso de esta teoría se debe al hecho de que es muy compleja la predicción de que una determinada persona pueda ser víctima del crimen, por este motivo es más factible predecir el lugar donde puede ocurrir algún evento delictivo [33].

En la actualidad, con las nuevas herramientas de localización es posible obtener información geo-referenciada de los probables sectores, los mismos que se usan para el proceso de encontrar patrones espacio-temporales de los incidentes, así se sabrá cuándo y dónde un nuevo crimen puede ocurrir [16]. Para la memoria [15] se usaron datos reales de la comuna de la Florida, pero para este proyecto fue necesario generar la base de datos en forma artificial, debido a la confidencialidad de la información sobre eventos delictivos [15]; la base de datos consta con datos geo-referenciados de eventos delictivos y servicios.

Con esta información se logra aproximar una distribución de riesgo criminal, obteniendo una representación espacial del crimen mediante el uso de GMM (Gaussian Mixture Model) [16]. En la etapa temporal en forma secuencial se ingresa nueva información de eventos delictivos, donde se puede apreciar como la distribución de riesgo inicial cambia dependiendo de la ubicación de la nueva información [15]. Este procedimiento se basa en la metodología empleada en la memoria de la Srta. Paulina Flores.

En el procedimiento anterior, es necesario optimizar cómo cambia la distribución de riesgo

inicial, que en la memoria se lo realiza mediante un ajuste empírico del “campo de atracción” que existe entre las nuevas observaciones y la información a priori que se presenta según el modelo. Como propuesta de este proyecto se realiza este procedimiento mediante el algoritmo de optimización PSO (optimización por enjambre de partículas).

En una zona urbana pueden existir ciertos lugares que atraen más que otros a la delincuencia, a estos se les puede considerar como un “atractor criminal” [4]. Ciertas características que permitan al delincuente lograr su cometido con mayor facilidad, incluyendo las posibles rutas de escape después de delinquir ya lo hace un lugar propenso a la delincuencia. Con esta idea en el proyecto se trata de implementar un modelo empírico de un “atractor criminal”, de esta forma se podrá analizar en el tiempo cómo ciertos puntos pueden ser futuros focos de riesgo criminal. El uso de este modelo más el uso del algoritmo de clustering K-means, se determinarán las futuras zonas de riesgo, donde K-means definirá los centros de las zonas riesgo y el modelo empírico simulará en el tiempo como se comportan las partículas frente al atractor, donde su movimiento dependerá si se encuentra en la zona de influencia del atractor.

Finalmente, en el proyecto se implementará y analizará cada uno de los métodos descritos y los resultados determinarán que tan factibles son los métodos propuestos.

## Hipótesis

- Es factible caracterizar espacialmente la distribución de riesgo espacial a través de un Modelo de Mixtura de Gaussianas (GMM), utilizando datos generados sintéticamente.
- Es factible comprobar que el método por optimización del modelo dinámico espacial de riesgo obtiene mejores resultados que el método empírico.
- Es factible realizar la predicción del riesgo criminal mediante el algoritmo K-means y la definición de un modelo empírico de “atractor criminal”.

## Objetivos

### Objetivo general

Generar una metodología basada en la relación entre eventos criminales y locación de servicios que proporcione un modelo espacio-temporal que permita caracterizar y predecir zonas de riesgo criminal dentro de un área determinada.

### Objetivos específicos

- Desarrollar una metodología para la implementación de un modelo espacial de riesgo criminal utilizando una base de datos compuesta por servicios y delitos.
- Desarrollar una metodología que permita actualizar de forma óptima el modelo espacial del riesgo criminal cada vez que nuevos eventos delictivos sean incluidos en el tiempo.

- Desarrollar una metodología, que en ausencia de nuevos datos de crímenes, realice predicción del riesgo criminal dentro de la zona urbana determinada.

## Alcance

El presente proyecto tiene como alcance generar la caracterización del riesgo criminal por medio del uso de datos de delitos y servicios asociados a un área urbana, mediante el modelamiento espacio-temporal del riesgo delictivo. Con esta representación se puede realizar un análisis de la evolución del crimen y de las posibles formas de enfrentar a la delincuencia en áreas urbanas.

Es necesario tener en cuenta que, con una caracterización inicial de los datos de los servicios y sus delitos asociados no basta, porque en la vida real cada día se sigue obteniendo más información de hechos delictivos que han ocurrido y por lo tanto es necesario ir actualizando la información con el propósito de tener datos actuales de la actividad criminal. Con el uso de un modelo espacial del riesgo criminal se realizará la actualización de la información y con la ayuda del algoritmo PSO se optimiza el resultado del modelo, de esta forma se puede representar la influencia de los nuevos delitos en la caracterización inicial. Una vez que se actualizado la información y no se cuenta con más información, la predicción de posibles zonas de riesgo criminal ayudará a tener información que sirva para enfrentar a la delincuencia. Para esto se pretende usar el algoritmo de clustering K-means con el cual cuál se obtiene los lugares (centroides) que se pueden definir como posibles “atractores del crimen”. Además, con el modelo de “atractor criminal” implementado se definirá la forma (perfil) de las zonas de riesgo criminal.

Por la naturaleza del proyecto, hay que considerar que este trabajo no es una herramienta para ser usada directamente por organismos de seguridad, debido a que para su uso comercial es necesario de una implementación más amigable para el usuario, que pueda ser usada para cualquier tipo de área urbana y realizar otros tipos de estudios de la realidad delictiva de una ciudad; la idea que se presenta es una propuesta más que sirve para poder llegar a la generalización de los métodos y modelos que se presentan.

Finalmente, hay que tomar en cuenta que para esta aplicación se consideró una base de datos sintética de servicios con sus delitos asociados, por lo tanto los modelos presentados no son una generalización para el análisis del riesgo criminal, el resultado de la simulación es un acercamiento para determinar las zonas riesgo criminal que pueden aparecer en el tiempo, teniendo como precedente un historial delictivo de una zona, siendo este un resultado preliminar que sirva como aporte para el análisis delictivo.



# Capítulo 2

## Marco Teórico y Estado del Arte

Para el análisis del comportamiento de la delincuencia en un sector, se han presentado diversas ideas con el propósito de comprender como operan los delincuentes, que incide en el cometimiento de un delito, estudios de comportamiento de los delincuentes, estudios sobre manejo de recursos policiales. Todo esto ayuda a que las fuerzas de seguridad puedan hacer frente a la delincuencia.

En esta sección se incluyen los principales enfoques de los modelos criminales presentados hasta el momento y las principales definiciones, que servirán para respaldar las ideas propuestas en este proyecto.

### 2.1. Teoría sobre el Análisis de la Delincuencia.

Día a día se escucha como la delincuencia aumenta en el país. Las técnicas y modalidades que emplea el delincuente de igual forma cambian y se adaptan para que puedan lograr su objetivo. Ante esto las fuerzas de seguridad deben mejorar todas sus herramientas para combatir a los delincuentes, valiéndose hoy en día de las matemáticas, la psicología y muchas otras ciencias, que ayudan a explicar el comportamiento de los delincuentes, y así poder enfrentarlos de mejor manera.

A lo largo del tiempo se han generado muchas propuestas y los investigadores en esta área han visto que tratar de interpretar el comportamiento de los delincuentes es la mejor forma de saber cómo y cuándo realizarán su próximo atraco. La dificultad de este proceso radica en que el comportamiento del ser humano es tan complejo como para predecir exactamente cuando va a cometer su próximo ilícito; entonces lo que se trata es de analizar varias variables que intervienen en cada uno de los hechos delictivos tales como patrones de comportamiento, el ambiente en el que se desarrollan estos hechos y en general el análisis del *modus operandi* [33] de los delincuentes.

En la actualidad una de las propuestas que se presentan es el análisis del riesgo criminal. El riesgo criminal es la probabilidad de que se produzca un evento delictivo, considerando la

relación que existe entre ese evento y los anteriores [8]. El proyecto Modelado del Riesgo de Terreno (*Risk Terrain Modeling - RTM*) [8], es un acercamiento a la estimación del riesgo en el cual a un mapa se le separa en capas, representando la influencia y la intensidad del factor de riesgo criminal en cada lugar. Todas las capas son combinadas para producir un mapa compuesto del riesgo del terreno ("*risk terrain*") que se representa en un mapa generado por un sistema de información geográfica (Geographical Information System - *GIS*). Este método se construye sobre los principios del mapeo por Hot-Spot (método para georreferenciar geográficamente las zonas donde se cometen mayor número de delitos [19]), criminología ambiental y problemas en el ámbito policial; todo esto orientado a producir mapas que muestren donde se presentan las condiciones ideales o que conducen a que ocurran los crímenes en un futuro [8].

Otra de las teorías que se presentan es la del "*Perfil Geográfico*" [22] es una técnica que analiza los patrones espaciales relacionados con la locación, usado principalmente con crímenes en serie, con el propósito de predecir la locación donde ocurrirán los eventos, se basa en la combinación de perfiles geográficos ponderados por efecto de las funciones, las cuales son adaptivamente ajustadas mediante la teorías de aprendizaje Bayesiano [22]. Una de las teorías más utilizadas en la actualidad es la de Hot-Spot, que mediante el uso de datos de incidentes criminales se genera un plano formando clusters y asume que son constantes para cualquier horizonte de predicción, es decir que en el método no existe una relación entre los crímenes y el ambiente en el tiempo, si existiera un cambio en el ambiente el método no puede indicar un cambio en los patrones del crimen [35][18].

Existen algunos métodos que se enfocan en el análisis del ambiente en donde ocurren los delitos (así como RTM), la teoría que se encarga del análisis del entorno de un crimen es la "*Teoría Ecológica*", la cual busca describir las motivaciones y actos criminales basándose en las características generales de un ambiente en particular [30]. Otro de los modelos que en la actualidad son analizados son los modelos aditivos generalizados (*Generalized Additive Model - GAM*) los cuales combinan información espacial, temporal y diversas características para predicción [30][35].

Dentro de la teoría criminológica, existe la intervención de varias ciencias como la sociología, psicología, medicina, antropología, matemática, física y química, apoyándose de manera indirecta con el derecho penal y otras ciencias de carácter penal o forense[24] . Con las ideas que presentan es posible proponer modelos que ayuden a interpretar patrones de posibles comportamientos de los delincuentes, que según un análisis de las actividades cotidianas, los infractores convergerán hacia blancos adecuados en el proceso de sus rutinas diarias. Se analizan factores como sus rutas de viaje, lugares de trabajo o residencia, es decir todas las áreas donde un infractor encontrará oportunidad. Estas ideas proporcionan el fundamento para generar un perfil geográfico básico [24]. Si se quiere profundizar en teorías como estas se puede revisar en [28][21].

Para este proyecto se considera la idea de "*Teoría Ecológica*", porque el planteamiento que presenta esta teoría es una gran herramienta para realizar un análisis de la relación que hay entre el riesgo criminal y el ambiente, pero antes hay que considerar que el escenario que se presenta es dinámico y está influenciado por fuerzas como los "atractores del crimen" y los "generadores del crimen" [4], donde ambos contribuyen para el análisis de áreas con cierta

intensidad delictual. A continuación se analiza cada uno de estos factores que influyen en la dinámica delictual de un área.

## **2.2. Generadores del Crimen y Atractores del Crimen.**

Los crímenes son creados por la interacción de un potencial delincuente con un potencial objetivo en un escenario, que hace más fácil, seguro y rentable el crimen [4]. Según estudios, se ha visto que el “medio ambiente”, a menudo complementado por el pasado delincencial, es un factor que atrae a la delincuencia, es decir, existe una conexión entre la tasa de crímenes y los atributos de la comunidad [3][30].

Existen 4 grandes grupos urbanos para el estudio del crimen, que se describen a continuación [4]:

### **2.2.1. Generadores del crimen.**

Los generadores del crimen son áreas en particular por las que se sienten atraídos los delincuentes, debido a la aglomeración de personas, ya sea por razones ajenas o por cualquier nivel particular de motivación criminal que pudieran tener o por algún delito en particular que podrían cometer [4].

### **2.2.2. Atractores del crimen.**

Los atractores del crimen son en particular lugares, áreas, vecindarios, distritos, los cuales crean una oportunidad bien conocida por el criminal, donde existe una fuerte motivación por parte del delincuente, entendiendo que los delincuentes son atraídos porque conocen de las oportunidades, en particular de algunos tipos de crimen [4].

Los crímenes en esas locaciones a menudo son cometidos por delincuentes que no pertenecen a esa área. Los delincuentes están fuertemente motivados a viajar relativamente largas distancias en búsqueda de un objetivo [4]. La atracción es creada por una característica “ecológica” [6][5], a menudo suplementado por la intención del delincuente y su pasado histórico, estableciendo la locación como un lugar conocido para ir y cometer una clase de crimen [4].

### **2.2.3. Sitios neutrales del crimen.**

Las áreas neutrales del crimen no atraen a la intención de los delincuentes, porque ellos esperan realizar un robo en un área en particular según sus intereses, ni tampoco ellos producen crímenes con el propósito de crear oportunidades criminales en esas áreas [4].

#### 2.2.4. Generadores de miedo.

El miedo al crimen es en general un miedo a ser atacado, de sufrir daño físico, de sufrir una intrusión que destruya la privacidad y dignidad de las personas. El miedo es mayor para una persona que es potencialmente vulnerable, como cuando se encuentra sola en espacios públicos de los cuales no tiene conocimiento de sus alrededores, en el momento que cruza por las calles y ve la presencia de potenciales atracadores o cuando hay signo de “problemas” (ventanas rotas, grafitis) en el área [4].

### 2.3. Hot-Spot

Una vez analizado cada uno de los modelos y teorías sobre la criminalidad, es necesario analizar una de las formas en que se representan cada uno de los resultados ya sea de un modelamiento, caracterización o predicción, y una de las formas es el trazado del mapa criminal (*mapping crime*).

Una de las principales funciones del trazado del mapa criminal es detectar las áreas con gran densidad de delitos y mostrar la relación que existe entre el delito y una determinada zona geográfica, siendo esta una de las herramientas más usadas por la policía. La policía principalmente lo usa cuando tiene que identificar zonas de alto riesgo criminal. De esta forma pueden analizar e intervenir según como se presente dicha información, y así poder actuar de mejor manera frente a estos delitos [19].

Una característica del crimen es que no se distribuye de manera uniforme en los mapas, es decir que tiende a acumularse en algunas zonas mientras en otras no se registra [19]. Esta información es utilizada por la gente, sin saberlo, en sus actividades diarias; incluso la decisión de donde vivir, comprar o estacionar el auto se basa en este tipo de conocimiento [15].

La reunión de toda esta información da como resultado áreas en donde más se concentran los crímenes, estas áreas son los “*Hot-Spot*” [19]. Aunque no existe una definición clara de lo que son, se entiende generalmente como un área que tiene un número de eventos delictuales mayor que el promedio o también como un área donde las personas tienen un riesgo de victimización mayor que el promedio [19].

Pero existen otras formas de representar y agrupar esta información además de los Hot-Spot, a continuación se presenta otra herramienta de igual forma muy utilizada y que en este proyecto fue de mucha utilidad al momento de la caracterización del riesgo criminal.

### 2.4. Modelos de Mixtura de Gaussianas (GMM)

GMM es una función de densidad de probabilidad paramétrica, representada como una suma ponderada de las densidades de los componentes Gaussianos, utilizadas comúnmente

como un modelo paramétrico de la distribución de probabilidad de mediciones o características diversas, matemáticamente se define como [36][16][15]

$$G(\vec{x}) = \sum_{i=1}^M \alpha_i f_i(\vec{x}), \quad (2.1)$$

donde  $\sum_{i=1}^M \alpha_i = 1$  y  $M > 1$ .

Por lo tanto, una GMM es una suma ponderada de M kernels, donde cada kernel  $f_i(\vec{x})$  es una distribución Gaussiana [16]

$$f_i(\vec{x}) = \frac{1}{(2\pi)^{D/2} |\sum_i|^{1/2}} \exp \left\{ -\frac{1}{2} (\vec{x} - \vec{\mu}_i)' \sum_i^{-1} (\vec{x} - \vec{\mu}_i) \right\}, \quad (2.2)$$

donde  $\mu_i$  y  $\sum_i$  son el vector de medias y la matriz de covarianza de la  $i$ -th mixtura de Gaussianas [16].

Existen 3 formas de GMMs, dependiendo cómo es seleccionada su matriz de covarianza [16]:

- *Primer caso:* hay una matriz de covarianza por componente Gaussiana (varianza nodal).
- *Segundo caso:* la matriz de covarianza es la misma para todas las componentes Gaussianas (gran covarianza).
- *Tercer caso:* es una simple matriz de covarianza compartida por todas las Gaussianas del modelo (covarianza global).

Mediante esta técnica es posible representar una base de datos como una sola distribución de probabilidad, la misma que está formada por la mezcla de varias Gaussianas, siendo esta la representación del área de riesgo criminal, la misma que se compone de curvas de nivel donde se representa el grado de criminalidad según se vaya acercando al foco (centro) del área de riesgo.

## 2.5. Inferencia Bayesiana e Integración de Monte Carlo

El enfoque Bayesiano proporciona un marco general para estimar las variables desconocidas de un sistema dinámico, a través de una actualización secuencial [16]. Este proceso consta de dos etapas, donde cada una se ejecuta cada vez que existe una nueva observación, con el respectivo cálculo de la PDF (Probability Density Function) en cada etapa. La primera etapa consiste en calcular una PDF a priori, la cual requiere una representación estocástica del espacio de estados actual del sistema. En la segunda etapa se incorporan las nuevas observaciones de corrección de la PDF a priori, a través de su verosimilitud, así se tendrá una PDF a posteriori [16].

La inferencia Bayesiana es el proceso de ajuste de un conjunto de datos a un modelo de probabilidad, condensando el resultado de los parámetros del modelo por medio de una distribución de probabilidad y las cantidades no observadas tal como predicciones para nuevas observaciones [17]. La característica esencial de los métodos Bayesianos es su uso explícito de probabilidades para cuantificar la incertidumbre en inferencias basadas en análisis de datos estadísticos [17].

El proceso de análisis de datos Bayesiano puede ser representado en los tres pasos siguientes [17]:

- El ajuste de un modelo de probabilidad completo: o una distribución de probabilidad conjunta para todas las cantidades observables y no observables en un problema. El modelo debe ser consistente con el conocimiento acerca del problema científico subyacente y el proceso de recopilación de datos.
- Acondicionamiento en los datos observados: cálculo e interpretación de la apropiada distribución a posteriori. Mediante la incorporación de nuevas observaciones, modificando la distribución a priori a través de su verosimilitud, así lo que se obtiene es una distribución a posteriori [16].
- Evaluar el ajuste del modelo y las implicaciones de la distribución posterior resultante: ¿cuán bueno fue el ajuste del modelo, son sustancialmente razonables las conclusiones y cuan sensibles son los resultados al modelamiento asumido en el primer paso? En respuesta, uno puede alterar o expandir el modelo y repetir los tres pasos.

Cuando las estimaciones son en línea, la actualización de la distribución a posteriori se lo hace introduciendo en forma secuencial nuevas observaciones. Si la función de los procesos de observación y transición de estado se pueden representar mediante un modelo de espacio-estado lineal y Gaussiano, es posible obtener una expresión óptima para la evolución de las distribuciones a posteriori mediante el Filtro de Kalman; sin embargo esta distribución no siempre es Gaussiana, siendo el Filtro de Kalman no muy adecuado para este caso [15][16].

Los métodos convencionales de Monte Carlo han sido extensamente usados para aproximar integrales y densidades de probabilidad [15]. El problema de su uso en inferencia Bayesiana es que no es directo, ya que involucra una secuencia de densidades de probabilidad variantes en el tiempo, mientras que los métodos convencionales asumen que la densidad objetivo es invariante [15].

Para resolver este problema se emplea el método secuencial de Monte Carlo, el cuál permite incorporar las observaciones para mejorar las estimaciones realizadas recursivamente [15]. Una de las técnicas usadas para este propósito es el Filtro de Partículas (*FP*), que es una técnica numérica para aproximar integrales multidimensionales con respecto a una secuencia de medidas de probabilidad [15][1].

La idea principal es representar la densidad a posteriori por un conjunto de muestras aleatorias con pesos asociados y calcular las estimaciones en base a esas muestras; por lo tanto, definiendo el concepto de partícula es posible realizar la inferencia estadística, que corresponde a la dupla  $\{x_{0:k}^i, \omega_k^i\}$ , donde  $x_{0:k}^i$  es una realización de la función de distribución de probabilidad de estado del proceso, a la que se le asocia un escalar  $\omega_k^i$ , que es una versión muestreada de la función de distribución de probabilidad [15].

## 2.6. Muestreo de Importancia (Importance Sampling)

Desafortunadamente el muestreo de una función de probabilidad a posteriori no es muy efectivo cuando es multivariable, no estándar y conocida sólo hasta una constante de proporcionalidad [14]. Para este caso, es muy difícil el uso de un modelo con las características de la GMM (Sección 1.4), por esta razón el muestreo de importancia es una solución de muestreo.

En estadística, el muestreo de importancia (Importance Sampling - *IS*) es una técnica general para la estimación de las propiedades de una distribución en particular utilizando muestras generadas a partir de una distribución de interés [15]. La idea fundamental sugiere que para ahorrar recursos, se puede enfocar la atención del proceso de muestreo hacia regiones de “importancia” [14].

El planteamiento propone utilizar un conjunto de muestras aleatorias y sus pesos respectivos, para aproximar la distribución deseada en base a ellos. Al considerar un número de muestras suficientemente grande, la aproximación calculada se asemejará a la distribución real [15]. La idea es resolver el problema mediante el muestreo de una PDF alternativa, la misma que se conoce en la literatura como *distribución de importancia* y se denota por  $q(x_{0:k}|y_{1:k})$  [16].

El método que se deriva de lo anteriormente explicado se inicia considerando  $\{x_{0:k}^i, w_k^i\}_{i=1}^{N_s}$ , un conjunto de muestras aleatorias (partículas), que caracterizan una distribución a posteriori  $p(x_{0:k}|y_{1:k})$  y donde los pesos normalizados son  $\sum_i \omega_k^i = 1$ , por lo tanto la densidad a posteriori se tiene que es aproximadamente [1]

$$p(x_{0:k}|y_{1:k}) \approx \sum_{i=1}^{N_s} \omega_k^i \delta(x_{0:k} - x_{0:k}^i). \quad (2.3)$$

Si se asume que  $p(x_{0:k}|y_{1:k}) \propto \pi(x_{0:k}|y_{1:k})$ , siendo la idea calcular los pesos en forma adecuada, es una pdf muy difícil de muestrear, pero  $\pi(x_{0:k}|y_{1:k})$  puede ser evaluada. También se tiene si  $x^i \sim q(x_{0:k}|y_{1:k})$ ,  $i=1, \dots, N_s$ , las muestras pueden ser generadas desde la *importance density*  $q(\cdot)$  [1], entonces

$$\omega_k^i = \frac{p(x_{0:k}^i|y_{1:k})}{q(x_{0:k}^i|y_{1:k})} \propto \frac{\pi(x_{0:k}^i|y_{1:k})}{q(x_{0:k}^i|y_{1:k})}. \quad (2.4)$$

El “arte” del *IS* es seleccionar la distribución de importancia  $q(\cdot)$  la cual aproxime lo más posible a  $p(\cdot)$  [7]. Este es el factor principal que afecta el rendimiento de este enfoque, desde las variables que deben ser descritas a partir de  $q(\cdot)$  que esbozan la distribución objetivo. Usando los conceptos de *IS*, se puede aproximar la distribución a posteriori con una función base finita discreta [7].

Pero para una función de importancia, se ha probado que la varianza de los pesos sólo puede crecer en el tiempo [1], es decir que después de un cierto número de pasos temporales

todas las partículas excepto una tienen pesos despreciables [1]. Todo esto tiene un efecto negativo sobre la precisión del método, para evitar este fenómeno se aplica *Resampling*.

## 2.7. Re-muestreo (Resampling)

Uno de los mayores problemas con en el *IS* es el llamado fenómeno de degeneración (*degeneracy phenomenon*) [14][1][7]. En términos prácticos, esto implica que después de un cierto número de pasos temporales, todas las partículas excepto una tienen pesos despreciables [1]. Esto ocurre porque la varianza de los pesos de importancia aumenta en el tiempo, lo que se hace imposible evitar la degradación de los pesos [7].

Para evitar este fenómeno es necesario introducir una etapa llamada de re-muestreo (*resampling*). El concepto fundamental de la teoría del re-muestreo es preservar las partículas con mayor peso (es decir mayor probabilidad) mientras se descarta aquellas con un menor peso [7]. Como resultado de esta etapa, a partir de la población  $\{x_k^i, w_k^i\}$ , se obtiene el nuevo conjunto  $\{x_k^{i*}, \omega_k^i = \frac{1}{N_s}\}$  con pesos uniformes [14].

La muestra resultante es una muestra i.i.d. (independiente e idénticamente distribuidas) de densidad discreta, por lo que los pesos se redefinen como  $\omega_k^i = \frac{1}{N_s}$  [15]. Una posible implementación se muestra en el Algoritmo 1.

**Input:** A set of particles with degeneracy phenomenon  $[(x_k^{j*}, w_k^j, i^j)_{j=1}^{N_s}]$   
**Output:** A new set of particles with  $w_k$  weight  $[(x_k^i, w_k^i)_{i=1}^{N_s}]$   
Initialize the CDF:  $c_1 = 0$ ;  
**for**  $i=2:N_s$  **do**  
    | Constructed CDF:  $c_i = c_{i-1} + w_k^i$ ;  
**end**  
Start at the bottom of CDF:  $i = 1$  ;  
Draw a starting point:  $u_i \sim \mathbb{U}[0, N_s^{-1}]$  ;  
**for**  $j=1:N_s$  **do**  
    | Move along the CDF:  $u_j = u_1 + N_s^{-1}(j - 1)$ ;  
    | **while**  $u_j > c_i$  **do**  
        |  $j^* = i + 1$ ;  
    | **end**  
    | Assign sample:  $x_k^{j*} = x_k^i$ ;  
    | Assign weight:  $w_k^j = N_s^{-1}$ ;  
    | Assign parent:  $i^j = i$ ;  
**end**

**Algorithm 1:** Algoritmo de Re-muestreo



## 2.8. Particle Swarm Optimization (*PSO*)

En la Sección 1.4 se presenta la forma de una distribución Gaussiana (Ecuación 1.2), donde es necesario ajustar la matriz de covarianza para propósitos de que la evolución temporal sea la más óptima [15]. Una forma de ajustar la matriz covarianza es encontrando los valores óptimos mediante *PSO*.

*Particle Swarm Optimization (PSO)* fue desarrollado por Kennedy y Eberhart [12], como un algoritmo de optimización estocástico basado en modelos de simulación social. El algoritmo emplea una población de puntos de búsqueda que se mueven estocásticamente en un espacio de búsqueda [26]. El desarrollo de este algoritmo se basa en conceptos y reglas que gobiernan socialmente poblaciones organizadas en la naturaleza, tal como las bandadas de pájaros, bancos de peces y rebaños de animales [26][29].

En *PSO* cada una de las simples entidades (partículas) se encuentran en un espacio de búsqueda de algún problema o función, y cada una evalúa la función objetivo según la posición en la que se encuentre en el espacio de búsqueda [29]. Entonces, cada partícula determina su movimiento a través del espacio de búsqueda por la combinación de algunos aspectos del historial de la posición actual y mejor localización (best-fitness) de cada uno de los miembros del enjambre sumado con algunas perturbaciones aleatorias [29].

La próxima iteración toma lugar después de que todas las partículas han sido movidas, como un enjambre o como una bandada de pájaros en busca de comida; siendo el movimiento que describen para encontrar un óptimo de la función fitness [29]. Hay que tomar en cuenta que teniendo sólo una partícula no es posible resolver el problema, ya que el proceso sólo se da cuando hay interacción entre varias partículas [29].

Si suponemos que el espacio de búsqueda es  $D$ -dimensional, entonces la posición de la  $i$ -ésima partícula del enjambre se puede representar por un vector  $D$ -dimensional,  $X_i = (x_{i1}, x_{i2}, \dots, x_{iD})^T$ . La velocidad (cambio de posición) de esta partícula puede ser representado por otro vector  $D$ -dimensional  $V_i = (v_{i1}, v_{i2}, \dots, v_{iD})^T$ . La mejor posición previamente visitada por la  $i$ -ésima partícula se puede expresar como  $P_i = (p_{i1}, p_{i2}, \dots, p_{iD})^T$  [27]. Definiendo  $g$  como el índice de la mejor partícula en el enjambre y los superíndices denotan el número de iteraciones, entonces el enjambre es manipulado de acuerdo a las siguientes ecuaciones [26][27]:

$$v_{id}^{n+1} = v_{id}^n + cr_1^n (p_{id}^n - x_{id}^n) + cr_2^n (p_{gd}^n - x_{id}^n), \quad (2.5)$$

$$x_{id}^{n+1} = x_{id}^n + v_{id}^{n+1}, \quad (2.6)$$

donde  $d=1,2,\dots,D$ ;  $i=1,2,\dots,N$ , y  $N$  es el tamaño del enjambre;  $c$  es una constante positiva, llamada *constante de aceleración*;  $r1$ ,  $r2$  son números aleatorios distribuido uniformemente en  $[0,1]$ ; y  $n$  determina el número de iteración [27].

Las Ecuaciones 1.5 y 1.6, definen la versión inicial del *PSO*. El mecanismo de control

de velocidad fue impuesto por un valor  $V_{max}$ . Este valor es un umbral de la velocidad, si la partícula excede este humbral la velocidad se fija  $V_{max}$ . Fue necesario definir este límite porque a valores muy altos de velocidad puede resultar que las partículas se muevan a buenas soluciones, pero para valores pequeños hay la posibilidad de tener una exploración insuficiente del espacio de búsqueda [27].

Por lo tanto, para evitar este problema, fue necesario incorporar un parámetro  $\omega$  previo a la velocidad de la partícula [27]. Por esta razón en las últimas versiones de *PSO* fueron cambiadas las Ecuaciones 1.5 y 1.6 por las siguientes ecuaciones [27]:

$$v_{id}^{n+1} = \chi (\omega v_{id}^n + c_1 r_1^n (p_{id}^n - x_{id}^n) + c_2 r_2^n (p_{gd}^n - x_{id}^n)), \quad (2.7)$$

$$x_{id}^{n+1} = x_{id}^n + v_{id}^{n+1}, \quad (2.8)$$

donde  $\omega$  es el factor de inercia;  $c_1$ ,  $c_2$  son dos constantes positivas llamadas parámetros cognitivo y social respectivamente y  $\chi$  es un factor de constricción (*constriction factor*), el cual es usado alternativamente a  $\omega$  para limitar la velocidad [27][26].

### 2.8.1. Parámetros del *PSO*

**Factor de Inercia ( $\omega$ )** : este factor es empleado para el control del impacto de las previas velocidades de las partículas en la actual [27][9].  $\omega$  también lo que hace es regular el “*trade-off*” entre la habilidad de exploración global (*wide-ranging*) y local (*nearby*) del enjambre [27].

Un valor alto de  $\omega$  facilita la exploración global (búsqueda de nuevas áreas), mientras que un valor bajo facilita la exploración local, es decir un buen sintonizado de este factor determina el área de búsqueda actual [26][27]. En forma experimental el valor de  $\omega$  se ha determinado como mejor valor 1.2, pero también para el ajuste se realiza pruebas disminuyendo desde 1.2 hacia 0 [27].

**Valores Cognitivo y Social ( $c1$  y  $c2$ )** : aunque no es un valor crítico para la convergencia, como lo es  $\omega$ , el apropiado sintonizado puede resultar en un convergencia rápida y una evasión del mínimo local [27][9]. Por defecto estos valores se usan como  $c1 = c2 = 2$ , pero en experimentos más recientes se ha visto que el mejor criterio de selección es  $c1 + c2 \ll 4$  [27].

**Parámetros  $r1$  y  $r2$**  : son usados para mantener la diversidad de la población, se encuentran uniformemente distribuidos en el rango de  $[0,1]$  [27].

**Número de Partículas** : el número de partículas que se usa está en el rango de  $20 \leq \text{número de partículas} \leq 40$  [9].

**Número de Iteraciones** : el número iteraciones utilizadas es como mínimo 500, pero si se va a utilizar el algoritmo ON-LINE se usa entre 10 y 50 iteraciones [9].

## 2.9. K-means Clustering

Una forma de agrupar y definir grupos de datos es usando *Clustering*. Con esta característica, mediante la agrupación de cierto tipo de datos en este caso delitos, pueden definirse posibles áreas con ciertas características que se pueden considerar zonas de riesgo.

*Clustering* es un procedimiento de agrupación de serie de vectores de acuerdo con un criterio. Esos criterios son por lo general la distancia o similitud. La cercanía se define en términos de una determinada función de distancia, como la euclídea, aunque existen otras más robustas o que permiten extenderlas a variables discretas [31][25].

Generalmente, los vectores de un mismo grupo (o *clústers*) comparten propiedades comunes. El conocimiento de los grupos puede permitir una descripción sintética de un conjunto de datos multidimensional complejo. De ahí su uso en minería de datos. Esta descripción sintética se consigue sustituyendo la descripción de todos los elementos de un grupo por la de un representante característico del mismo [25].

*K-MEANS* es un método de agrupamiento, que tiene como objetivo la participación de un conjunto de datos en  $k$  grupos, en el que cada observación pertenece al grupo mas cercano al centro de los clusters [20][31][25]. Esto da lugar a una compartimentación del espacio de datos en celdas de Voronoi. K-Means clasifica un determinado conjunto de datos a través de un cierto número de racimos o grupos. La idea principal es definir  $k$  centroides (denotado como el “objeto medio” o el punto medio en el espacio del cluster), uno para cada grupo [20]. Estos centroides deben colocarse de la mejor manera, debido a que un buen resultado dependerá de la localización de los mismos. El siguiente paso es tomar cada punto que pertenece a un determinado conjunto de datos y asociarlos al centroide más cercano [25].

El proceso de agrupamiento en k-means se inicia determinando los centroides o centros de esos grupos [25]. Para determinar los centroides hay dos alternativas prácticas: la primera es tomar de forma aleatoria  $k$  “objetos” o puntos como centroides iniciales y la segunda es tomar los primeros  $k$  “objetos” en secuencia. Luego el algoritmo de clustering ejecuta los siguientes tres pasos hasta alcanzar el criterio de convergencia, es decir que los objetos no se muevan del grupo: [31][25]

- Primero, se determinan a los centroides iniciales de acuerdo al número de cluster esperado.
- Después, se determina la distancia de cada objetivo con relación a los centroides.
- Finalmente, se agrupan los objetivos con base en la distancia mínima.

### 2.9.1. Valor óptimo de Clústers

Encontrar cual puede ser el centro en cada cluster (que para este caso serán los centros de las zonas de riesgo) es determinante, debido a que estos serán los puntos donde se localizarán los atractores criminales.

Para encontrar el número de centroides es necesario saber cuál es el valor óptimo de clusters, para este propósito se usan herramientas como los índices de validación de clusters, los más utilizados para encontrar el número óptimo de clusters son [32]: el índice de Dunn, Davies-Bouldin, Silhouette, función de costo de K-Means y SSE (Sum Squared Error).

#### Índice de Dunn.

Es un índice que identifica los clusters cuán separados y compactos están, el objetivo de este índice es maximizar la distancia inter-cluster mientras se minimiza la distancia intra-cluster [32]. El índice de Dunn se define de la siguiente forma [32]:

$$DU_k = \min_{i=1,\dots,k} \left\{ \min_{j=1+1,\dots,k} \left( \frac{\text{diss}(c_i, c_j)}{\max_{m=1,\dots,k} \text{diam}(c_m)} \right) \right\}, \quad (2.9)$$

donde  $\text{diss}(c_i, c_j) = \min_{x \in c_i, y \in c_j} \|x - y\|$  es la diferencia entre clústers  $c_i$  y  $c_j$  y  $\text{diam}(C) = \max_{x,y \in C} \|x - y\|$  es la función intra-cluster (o diámetro) de los clusters. Por lo tanto, significa que compacto y bien separados existen los clústers. El índice nos dice el valor máximo observado para  $k$  igual al más probable número de clusters en el conjunto de datos [32].

#### Índice Davies-Boulding.

El índice Davies-Boulding [32] identifica los clusters que tan lejos están el uno del otro y que tan compactos están. Se define de la siguiente forma [32]

$$DB_k = \frac{1}{k} \sum_{i=1}^k \max_{j=1,\dots,k, i \neq j} \left\{ \frac{\text{diam}(c_i) + \text{diam}(c_j)}{\|c_i - c_j\|} \right\}, \quad (2.10)$$

donde para este caso, el diámetro de un cluster es definido como [32]:

$$\text{diam}(c_i) = \left( \frac{1}{n_i} \sum_{x \in c_i} \|x - z_i\|^2 \right)^{1/2}, \quad (2.11)$$

con  $n_i$  el número de puntos y  $z_i$  el centroide de los clusters  $c_i$ . El objetivo de este índice es obtener clusters con la mínima distancia intra-clúster; es decir, el índice minimizado cuando se encuentra el mejor número de clústers [32].

## Índice Silhouette.

El estadístico Silhouette (SI) [32], es otra forma de estimar el número de grupos en un conjunto de datos. El índice calcula para cada punto un ancho dependiendo si pertenece a alguno de los clusters. El ancho de este índice es el promedio de las observaciones [32].

Tenemos

$$SI_k = \frac{1}{n} \sum_{i=1}^n \frac{(b_i - a_i)}{\text{máx}(a_i, b_i)}, \quad (2.12)$$

donde  $n$  es el total de número de puntos,  $a_i$  es el promedio de la distancia entre el punto  $i$  y todos los otros puntos en su propio cluster y  $b_i$  es la mínima diferencia promedio entre  $i$  y los puntos en otros clusters. Finalmente, la partición con el más alto SI es tomado como el óptimo [32].

## Sum of The Square Error (SSE).

Es el cálculo del error de cada punto, es decir su distancia euclidiana a los centroides más cercanos, y se calcula la suma total de los errores al cuadrado [32]. El SSE formalmente se define como

$$SSE = \sum_{i=1}^K \sum_{x \in C_i} \text{dist}(c_i, x)^2, \quad (2.13)$$

donde  $x$  es un objeto,  $C_i$  es el  $i$ th cluster,  $c_i$  es el centroide del cluster  $C_i$ ,  $K$  es el número de clusters y la  $dist$  es la distancia Euclidiana entre los objetos en el espacio euclideo [32].

## Función de Costo.

La función de costo se define como

$$C(z, \mu) = \sum_i \|x_i - \mu_{z_i}\|^2. \quad (2.14)$$

Llamado también “*elbow method*”, es un método que es el cuadrado del error o la distancia entre los centroides y los datos pertenecientes a un determinado cluster y finalmente para el resultado se saca la media de los valores resultantes de  $C$ , que mientras se aumenta el número de clusters la función decrece, y donde existe un cambio en la función como un “*codo*” (*elbow*) es el valor óptimo de clusters.

## 2.10. Método para evaluar los modelos: HRP vs. TIP

La forma de evaluar el modelo del riesgo criminal, consiste en contrastar las áreas de alta probabilidad predichas por el modelo, con un número de crímenes que actualmente ocurrieron en esas áreas [16]. Un modelo se dice que es bueno mientras los incidentes en una ventana de tiempo son proporcional a sus áreas de riesgo predichas [16]. La caracterización del modelo de riesgo en un tiempo  $t_j$  está dado por  $HRP_\theta$  versus  $TIP_\theta$  método propuesto por [35], donde

$$HRP_\theta = \frac{||\{a_i | \mathbb{P}(\text{inci}_{a_i, t_j} = 1) > \theta\}||}{||\{a_i\}||} \quad (2.15)$$

$$TIP_\theta = \frac{||\{\text{inci}_{a_i, t_j} = 1 | a_i \in \{a_i | \mathbb{P}(\text{inci}_{a_i, t_j} = 1) > \theta\}\}||}{||\{\text{inci}_{a_i, t_j} = 1\}||}, \quad (2.16)$$

en que el *TIP* (*Test Incident Percentage*) representa los incidentes desde un conjunto de prueba que se llevaron acabo en áreas de alto riesgo ; *HRP* (*High Risk Percentage*) representa el porcentaje de las áreas de alto riesgo predichas por el modelo;  $|| \cdot ||$  representa el tamaño de un set,  $\theta$  es un umbral que se encuentra entre  $[0,1]$  y  $\mathbb{P}(\text{inci}_{a_i, t_j} = 1)$  es la probabilidad de que ocurran incidentes criminales en un área con subdivisión  $a_i$  y en una ventana de tiempo  $t_j$  [16]. Ambas medidas son calculadas para diferentes  $\theta$  y son graficados para obtener una representación similar a la curva de operación característica (ROC) (ver Figura 1.1) [13].

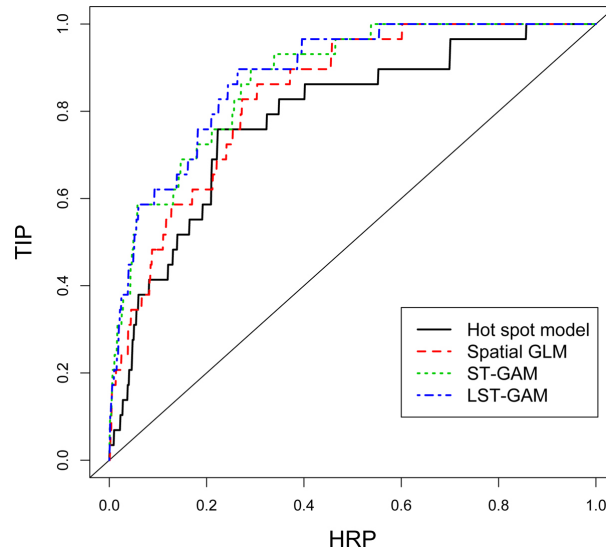


Figura 2.1: Ejemplo de gráfico HRP vs. TIP [35].

Si muchos incidentes criminales toman lugar en áreas con alta probabilidad de riesgo, la curva será muy cercana a la esquina superior izquierda y en caso opuesto será una curva similar a una relación lineal [16]. La medida de calidad del modelo es el área bajo la curva (*AUC*) descrita por la relación HRP vs. TIP.

## Capítulo 3

# Descripción de la metodología empleada para la caracterización y predicción del riesgo criminal

En esta sección se presenta el procedimiento para la caracterización y predicción de las zonas con un alto riesgo delictivo. El criterio que se muestra es la generación automática de distribuciones espaciales de riesgo asociado a la actividad delictual, usando una base de datos generada en forma artificial.

El tipo de datos que se usan son eventos delictivos geo-referenciados, asociados a un servicio determinado [15]. Los servicios pueden ser cualquier tipo de asentamiento urbano como hospitales, colegios, parques, supermercados, farmacias, entre otros; incluso todo tipo de lugar geográfico con la presencia de ríos y lagos [15].

La ocurrencia de un delito está determinado por varios factores, como: factores geoespaciales, la hora del día en que ocurren, preferencias de los delincuentes, características de los lugares donde ocurren los crímenes o utilidad percibida de la locación; todas estas características favorecen a que una determinada zona tenga más riesgo de ser vulnerada por un delincuente [30].

Para este proyecto se consideran los factores geoespaciales y la teoría “ecológica”, la misma que busca describir las motivaciones y actos criminales basado en las características generales de un ambiente en particular [30].

El procedimiento se inicia con la creación de una base de datos de eventos delictivos asociados a servicios, que sirve para la caracterización inicial del riesgo criminal. Este resultado es útil para la generación automática de distribuciones espaciales de riesgo asociado a la actividad delictual, este mecanismo de caracterización temporal se lo realiza mediante la inclusión iterativa y cronológica de nuevos eventos (nuevas observaciones) [15].

En la generación automática de las distribuciones aparece un elemento determinante que es el “campo de atracción” entre partículas. El campo de atracción es una función (distribución

Gaussiana bivariada) que produce los cambios al modelo dinámico generado en [15]. Parte de la definición de esta función (ver Ecuación 2.2) es el ajuste de un elemento, la matriz de covarianza ( $\Sigma$ ).  $\Sigma$  es la que define un umbral de acción, que indica el límite aproximado para el cual las nuevas observaciones influirán en el movimiento de las partículas [15]. En [15] se menciona que los valores de esta matriz son determinados en forma empírica, pero para este trabajo se propone encontrar los valores óptimos de esta matriz, con el propósito de que los cambios en el modelo sean los que se requiere para obtener un buen resultado.

Para la estrategia de predicción se presenta una propuesta que se basa primero en el agrupamiento (*clustering*) de la nueva información de delitos, teniendo así una posible región de riesgo y los centros de esas regiones. Después mediante el uso de la idea de un atractor criminal [4] se genera un modelo que simula el comportamiento de las partículas frente a un atractor y finalmente la generación de una distribución de predicción.

### 3.1. Área de Interés y Datos Geo-referenciados a los Servicios

Ciertas características del área geográfica son muy importantes al momento de analizar cómo la delincuencia podría actuar en ciertas zonas, como en sectores comerciales, bancarios, de entretenimiento, residenciales; considerando qué tan significativas son esas características se podrán analizar y determinar por qué es más alto o bajo el riesgo de ser víctima de la delincuencia.

Hay que tomar en cuenta que para propósitos de esta tesis se trabajó con un área más acotada, debido al alto costo computacional que implica al momento de procesar y registrar los datos de los delitos y los servicios; siendo suficiente el perfil geográfico seleccionado para alcanzar los objetivos propuestos.

Para este proyecto se generó una base de datos con servicios y sus delitos asociados, generados en forma aleatoria tanto la ubicación de los servicios como la ubicación de los delitos. Una vez que se generan por todo el mapa el usuario puede seleccionar las áreas con los servicios y automáticamente se generan los delitos asociados a cada servicio seleccionado. Se consideró la generación de la base de datos de esta forma porque así el usuario puede seleccionar según crea conveniente las zonas y la distribución de los servicios.

La base de datos debe contar con estas tres variables para que se pueda representar el riesgo criminal, para este caso es necesario [16]:

- El área  $A \subseteq \mathbb{R}^2$ , la cual debe ser definida ya que esta será el dominio de la función de la distribución de probabilidad que se va a definir. Para esto es necesario considerar ciertos factores, como: analizar porque razón se va a generar un modelo, la capacidad computacional con la que se cuenta, examinar la información con que se cuenta.
- Dentro de una zona habitable existen servicios de toda índole, servicios  $\{s_i\}_{i=1}^M$  presentes en  $A$ , necesarios para el desarrollo de la vida cotidiana de las personas, pero hay que considerar que para el análisis existe una relación entre el área y el número de servicios



que se consideran para el estudio. Si el área es muy extensa y existen pocos servicios la representación del riesgo criminal no será adecuada; además si el área es muy pequeña y tiene muchos servicios el modelo generado no mejorará y habrá una sobrecarga en el procesamiento de los datos.

- Los eventos criminales geo-referenciados o delitos  $\{d_j\}_{j=1}^D$ , ocurren en un intervalo de tiempo  $T \subseteq \mathfrak{R}^+$ , siendo una información muy importante al momento de considerar estos registros la cantidad de eventos con los que cuenta, ya que para el proyecto será muy significativo si se tiene un número considerable de eventos. Además dependiendo del análisis los modelos pueden variar, ya sea si se genera para uno o más tipos de delitos (si se necesita un modelo de riesgo general) o si se considera en conjunto todos los crímenes (crímenes de alto impacto como: homicidio, robo a personas o casas).

Los tipos de delitos que se consideran para generar el modelo dinámico usado en el proyecto son delitos de mayor connotación social (DMCS) [10][15]. Los DMCS se consideran delitos como: robos con violencia o intimidación, robos por sorpresa, robos de vehículos, robos de accesorios de vehículos, robos en lugar habitado, robos en lugar no habitado, hurtos, lesiones leves, lesiones graves, homicidios y violaciones [10]. Se consideran este tipo de delitos porque se pueden modelar en grupo, obteniendo un solo modelo útil para la caracterización del riesgo criminal [15].

## 3.2. Modelos Espaciales de Riesgo

La generación de un modelo de riesgo criminal tiene como finalidad la creación de un modelo probabilístico que caracterice el riesgo asociado a la actividad delictual, a partir de datos de servicios y eventos delictuales, y que pueda ser utilizado para la predicción de los focos de criminalidad en un área determinada [15]. Los eventos, más específicamente la varianza en el espacio de sus localizaciones, permiten aproximar una distribución normal bivariada en cada servicio, y mediante la generación de un modelo por mezcla de Gaussianas (GMM) se genera la caracterización probabilística del riesgo final [15].

La razón del uso de GMM en este proyecto es porque permite caracterizar la distribución del riesgo del área seleccionada, de tal forma que el usuario pueda interpretar la información que se presenta, ya sea como niveles de riesgo, la localización y el tamaño del área de riesgo.

El procedimiento se inicia con el área de interés  $A$ , donde se especifican los servicios  $\{s_i\}_{i=1}^M$  y los delitos  $\{d_j\}_{j=1}^D$ , de la forma en que se analizó en la Sección 3.1. Una vez que se define el área con toda la información geo-referenciada de los servicios y delitos, se posicionan en el mapa; asociando cada evento delictivo a cada servicio es posible calcular la covarianza en la localización de los delitos, medida que se usa para aproximar la distribución bivariada centrada en el servicio [15].

Por lo tanto, para cada servicio se realiza el cálculo del riesgo asociado, ya que mediante la asociación de los delitos a cada servicio, es posible calcular la varianza en la localización de los delitos la misma que sirve para aproximar una distribución bivariada centrada en el servicio [15]. Por lo tanto, se tendrá una distribución de riesgo aproximada por una distribución

normal bivariada, donde su media ( $\mu_i$ ) es la localización del servicio y su matriz de covarianza ( $\Sigma_i$ ) está dada por la varianza empírica de la localización de los eventos delictuales que ocurren dentro de un rango a su alrededor. El rango se define por el radio,  $r_i \in \mathbb{R}^+$  de un círculo centrado en la coordenada del servicio. Para que un evento sea asociado a un servicio debe encontrarse a una distancia euclidiana ( $dist_{ij} = \sqrt{(d_{j_x} - s_{i_x})^2 + (d_{j_y} - s_{i_y})^2}$ ) igual o menor del radio definido como rango ( $dist_{ij} \leq r_i$ ) [16].

Con todos los elementos se realiza el cálculo de la varianza de las localizaciones ( $\Sigma$ ). La matriz de covarianza es utilizada para aproximar una distribución normal bivariada (ver Ecuación 1.2). Tomando en cuenta que los datos cumplen con el rango establecido (radio del área definido), si existen pocos datos, es decir que cerca al servicio existió una baja tasa de victimización o si los datos se repiten, es decir que existe una victimización reiterada; para estos casos es necesario aumentar el rango y por lo tanto agregar más datos hasta que  $\Sigma$  sea invertible [15].

Finalmente, una vez que están definidos todos los elementos necesarios para la caracterización, se obtiene un modelo *GMM*,  $PDF_{prior}(\vec{x}) = \sum_{i=1}^M \alpha_i \cdot f_i(\vec{x})$  donde  $\alpha_i = \frac{1}{M} \Rightarrow \sum_{i=1}^M \alpha_i = 1$ , que es una distribución de probabilidad (distribución a priori) que caracteriza espacialmente en conjunto los servicios y sus delitos asociados, es decir, la caracterización del riesgo del área de interés (ver Figura 3.1) [15] [16].



Figura 3.1: Ejemplo de representación de los servicios y sus delitos asociados mediante GMM.

### 3.3. Evolución Temporal.

Una vez determinada la distribución espacial de riesgo, no es suficiente para que exista una caracterización completa, por lo tanto, es necesario incluir variables temporales, así se adaptará mejor a los cambios en las condiciones del medio que genera la delincuencia [15].

La inclusión de nuevos datos delictivos permite tener una distribución actualizada y por lo tanto nueva información sobre el riesgo criminal.

En esta etapa lo que se busca es la caracterización de la evolución temporal de la distribución espacial de riesgo asociada a un área de interés, por medio de una metodología inspirada en métodos secuenciales de Monte Carlo (Sección 2.5), y el uso de un modelo dinámico definido en [15]. Dentro de todo este procedimiento se insertarán en forma cronológica y sucesiva la información de nuevos delitos (nuevas observaciones), obteniéndose un esquema de cambio on-line[15] [16]. El uso de los métodos secuenciales de Monte Carlo permiten obtener una estimación de la distribución inicial a priori, es decir una estimación de la distribución que permite reducir el costo computacional al momento de procesar los datos.

Primero se inicia con la obtención de muestras de la distribución de densidad de probabilidad que sean representativas, para esto se realiza un muestreo de importancia (*important sampling*—ver Sección 2.6) [16]. Esta operación es necesaria debido a que es más sencillo manipular las muestras de la distribución a priori, por el bajo costo computacional que conlleva realizar esta operación. El método que se aplica para cumplir con esta meta es mediante una malla[14] [15], la misma que sirve para discretizar el dominio de la función (PDF a priori).

Como es necesario tener una aproximación de la distribución a priori, después de aplicar el algoritmo de *Important Sampling*, se obtendrán zonas con más alta probabilidad donde existen más partículas y menos partículas en las zonas de menos probabilidad, con la respectiva asignación de los pesos; al final para obtener la aproximación de la distribución se aplica el algoritmo *Resampling* (ver Figura 3.2).

Una vez que se obtiene la aproximación de la distribución a priori, se puede apreciar que está compuesta por partículas que es necesario definir como se comportaran dentro del área definida.

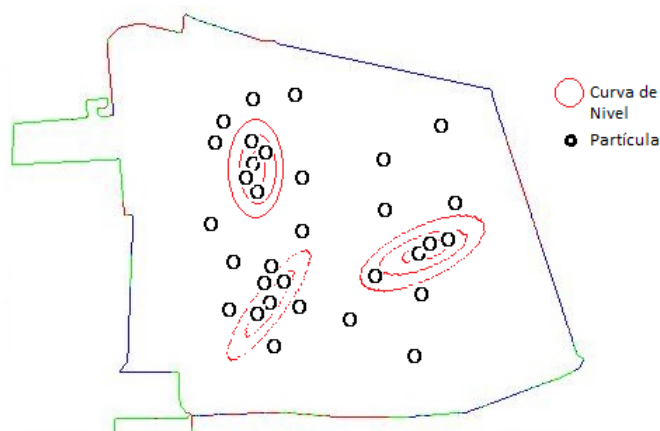


Figura 3.2: Resultado final de la aproximación de la distribución a priori después de aplicar Important Sampling y Resampling.

### 3.3.1. Modelo Dinámico.

Con la introducción de nuevos datos geo-referenciados de crímenes (nuevas observaciones- $Y$ ), las partículas van a modificar su posición; su movimiento está determinado por un modelo dinámico, necesario para que las partículas se desplacen en el plano. Con el modelo dinámico se debe cumplir que [15]:

- las partículas lejanas a la observación deben mantenerse alejadas y ser afectadas sólo por el ruido de proceso;
- las partículas cercanas a la observación deben mantenerse cerca y deben ser afectadas sólo por el ruido de proceso; y
- las partículas a una distancia prudente (parámetro de diseño) de la observación deben acercarse a ella y a la vez ser afectadas por el ruido de proceso.

El propósito de usar el modelo dinámico es analizar cómo es la influencia de los nuevos delitos en la caracterización inicial del riesgo, es decir cómo cambia la distribución en el tiempo con la introducción de nueva información, de esta forma se introduce la variable temporal en el proceso.

El modelo dinámico que define el movimiento de las partículas está dado por[15]:

$$x(k) = x(k - 1) + f(d)\{y(k) - x(k - 1)\} + w(k - 1), \quad (3.1)$$

siendo  $x(k)$  la posición de la partícula en el tiempo  $k$ ,  $x(k - 1)$  la posición de la partícula en el instante anterior,  $y(k)$  es la observación en el tiempo  $k$ ,  $w(k - 1)$  es el ruido de proceso y  $\{y(k) - x(k - 1)\}$  es la expresión que entrega dirección de movimiento y su magnitud [15].

La función  $f(d)$  depende de la diferencia  $d$ , entre la observación y la partícula que está siendo estudiada [15]. El ajuste de esta función es muy importante, porque de ella dependerá el funcionamiento dinámico. Para definir  $f(d)$  se lo hace por medio de una Gaussiana bivariada, ajustada de tal forma que el máximo de ésta sea la unidad [15], entonces tenemos

$$g(d) = \frac{1}{2\pi|\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2}(\vec{x} - \vec{y})^T \Sigma^{-1} (\vec{x} - \vec{y}) \right\}, \quad (3.2)$$

$$f(d) = \frac{g(d)}{\max(g(d))}. \quad (3.3)$$

Para que esté completamente definida la función  $f(d)$ , es necesario que se defina la matriz de covarianza que se encuentra dentro de  $g(d)$ , dado por la Ecuación 3.2. La matriz de covarianza ( $\Sigma$ ) es una matriz diagonal, cuyos parámetros son parte del diseño y sus valores dependerán de cuánto se desea que la partícula se acerque a la nueva observación (Figura 3.3) [16][15].

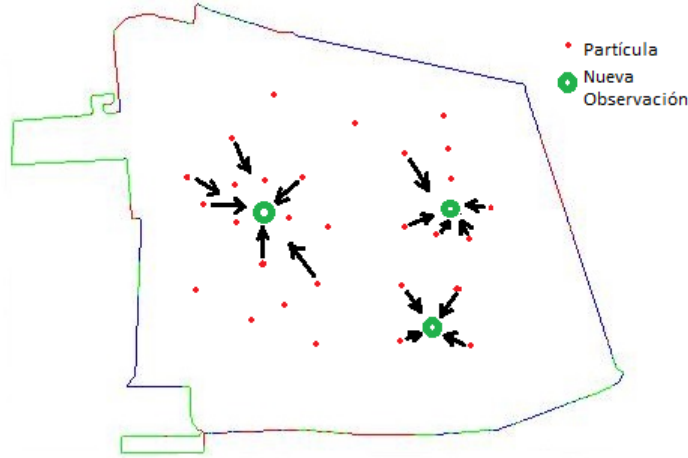


Figura 3.3: Movimiento de las partículas hacia las nuevas observaciones.

### 3.3.2. Optimización de los valores de la matriz de covarianza $\Sigma$ .

El ajuste de  $\Sigma$  es muy importante, debido a que define el rango dentro del cual las partículas se moverán hacia las nuevas observaciones. Para esta etapa se define este parámetro, dicha variable es parte del diseño y depende del área de interés; pero en general los valores de la diagonal de  $\Sigma$  depende mucho del criterio del usuario, propuesta presentada en [15]. Pero esta forma de encontrar los valores de la matriz hace que el resultado este muy supeditado a la experiencia del usuario, por esta razón, como propuesta de este proyecto es encontrar los valores de esta matriz diagonal  $\Sigma$ , para que el movimiento de las partículas (“campo de fuerza”) sea óptimo. Para encontrar los valores óptimos de la matriz se usa el algoritmo *PSO* (Particle Swarm Optimization–Sección 2.8), este es un algoritmo de optimización que mediante la inteligencia de enjambre encuentra los valores óptimos, para este caso, de la diagonal de la matriz  $\Sigma$ . Este algoritmo cuenta con características que lo hacen apropiado para el objetivo a lograr; características como: es un algoritmo de rápida convergencia en comparación con GA (Genetic Algorithm) y Temple Simulado [9], ocupa menos recursos computacionales y es muy eficiente al momento de optimizar en un amplio rango de funciones [12].

Con el uso de *PSO* lo que se trata es encontrar los valores óptimos de  $\Sigma$ , considerando como función objetivo al *TIP* (ver Sección 1.10)

$$TIP = \frac{\text{número de eventos que caen en los sectores elegidos}}{\text{número total de eventos}}. \quad (3.4)$$

Como propuesta se considera, que mediante el uso de *PSO*, maximizar a la función objetivo *TIP* (Ecuación 3.4) y considerar un *HRP* del 15%. En general el algoritmo lo que hace es generar varios resultados de distribuciones a posteriori con cada valor de la matriz de covarianza que genera el algoritmo *PSO*, con una ventana igual al número de nuevas observaciones, en un rango determinado para los valores que puede tomar la diagonal de  $\Sigma$  (para este caso entre  $1 \times 10^{-12}$  y  $1 \times 10^{-3}$ ). Cada distribución a posteriori es evaluada con el criterio de *HRP* vs. *TIP*, considerando el criterio de *HRP* del 15% y maximizando el valor del *TIP* en ese punto. El algoritmo realizará estas operaciones hasta encontrar el valor más alto del

TIP o el más cercano a 1. Con el resultado del TIP más alto encontrado por el algoritmo se puede apreciar cual distribución fue la que obtuvo este valor y por lo tanto cuales fueron los valores de la matriz de covarianza que se usaron para generar esta distribución, siendo estos valores los óptimos para  $\Sigma$ .

En las anteriores operaciones hay que considerar que existe un problema en los resultados debido al ruido del proceso, por esta razón se obtiene el promedio de todos los posibles valores de TIP de cada distribución generada dentro del *PSO*, de esta forma se trabaja con un valor promedio resultante del TIP, siendo esta medida útil para el procedimiento. Esta operación se realiza debido a la aleatoriedad presente en los resultados, así se intenta disminuir la variación presente en los resultados finales del TIP y los valores de la matriz de covarianza.

Finalmente con el mejor resultado que nos entregue *PSO* se usa en  $\Sigma$  del modelo de la etapa de evolución temporal, obteniendo así la distribución a posteriori, considerando una ventana de tiempo igual al número de nuevas observaciones que fueron insertadas en forma secuencial.

### 3.4. Predicción.

La etapa de predicción se inicia con el agrupamiento de las nuevas observaciones ( $Y$ ), que fueron introducidas en forma secuencial en la etapa de evolución temporal. Para el agrupamiento de los datos se usó el algoritmo de *clustering K-means*. Mediante este procedimiento se encuentra el número de centroides de las posibles zonas de riesgo criminal. La razón del uso de K-means es porque permite determinar los centros (centroides de los clusters) de las posibles zonas de riesgo y la ubicación del atractor criminal, además definirá a priori las áreas de riesgo criminal donde se localizarán los focos de criminalidad.

Para encontrar el número óptimo de *clusters* y centroides que permita tener un agrupamiento de todos los datos sin perder datos, se usó los índices de validación de *clusters* (Sección 2.9.1). Estas herramientas nos entregaran el número óptimo  $K$  o número de *clusters*, necesarios de definir al inicio del algoritmo *clustering*.

Una vez encontrados los centroides ( $\mu_2$ ) de los posibles hot-spot de predicción, es necesario definir como las partículas son “atraídas” hacia los centroides o atractores del crimen (sección 2.2.2), donde la posición final de las partículas, dentro de una ventana de tiempo, determinarán la distribución de predicción del área de riesgo criminal. Para el desarrollo de este planteamiento se propone el modelo de un atractor criminal, que se presenta a continuación

$$x_1(k+1) = x_1(k) + \alpha \sum_{i=1}^m \left( \frac{\mu_{2_i} - x_1(k)}{\|\mu_{2_i} - x_1(k)\|} \right) \cdot \omega_i \cdot N \left( x_1(k); \mu_{2_i}, \sum 2_i \right), \quad (3.5)$$

donde  $m$  es el número de centroides,  $x_1(k+1)$  es la posición de la partícula que se mueve hacia el atractor,  $x_1(k)$  es la posición en el instante anterior, la expresión  $\left( \frac{\mu_{2_i} - x_1(k)}{\|\mu_{2_i} - x_1(k)\|} \right)$  es

el cálculo de un vector unitario, el mismo que da la dirección para el movimiento de las partículas y  $N(x_1(k); \mu_{2_i}, \Sigma_{2_i})$  es una función Gaussiana que da la “fuerza” con la que se va a mover la partícula y depende de su posición, la media y la matriz de covarianza asociado al atractor.

El modelo del atractor presentado tiene como finalidad modelar un “atractor criminal”, donde los centros del atractor encontrados son puntos donde convergen los delitos (partículas) en el tiempo. Los delitos convergerán según la posición a la que se encuentren respecto a los centros, es decir su movimiento se dará mientras más cerca esté de alguno de los centros. El área de incidencia del atractor depende de la función Gaussiana definida en el modelo.

Después, usando los nuevos eventos introducidos en la evolución temporal, se genera una  $GMM_{new-obs}$  donde su vector de medias es  $\mu_{2_i}$ , que son los centroides encontrados con el *clustering* y  $\Sigma_{2_i}$  es la matriz de covarianza generada con las nuevos eventos. El factor  $\alpha$  es un valor que se define para ajustar el movimiento de las partículas según la escala del mapa, así se puede apreciar que son atraídas hacia los centroides;  $\omega_i$  son los pesos del  $i$ -ésimo cluster, donde la propuesta es

$$\omega_i = \frac{\text{número de observaciones en cluster } i}{\text{número total de observaciones}}. \quad (3.6)$$

Una vez que las partículas son atraídas hacia los centroides, se genera la distribución de predicción. Para que las partículas sean atraídas se consideró un horizonte de predicción igual al número de observaciones, para este caso 400. Finalmente para apreciar de mejor manera el resultado de la predicción se la representa mediante una GMM de predicción.

# Capítulo 4

## Análisis y Discusión de los Resultados Obtenidos

### 4.1. Área de interés seleccionada para el proyecto

El área considerada para desarrollar el trabajo fue extraída del mapa de la comuna de Santiago Centro (Figura 4.1) obtenido de la página del portal de la Municipalidad de Santiago [11], generado a partir de este un mapa artificial de coordenadas  $-70.66$  de longitud y  $-33.45$  de latitud; estas coordenadas sirvieron para referenciar el área de interés, los servicios, los delitos asociados a cada uno de los servicios y referenciar las nuevas observaciones.

Se seleccionó esta comuna debido a que su área es más pequeña ( $23,20Km^2$ ) y su perfil es más sencillo de procesar para el algoritmo. Se realizaron pruebas con una ciudad de mayor área ( $290Km^2$ ) y un perfil mucho más complejo como es la ciudad de Quito (Ecuador), donde el algoritmo tuvo problemas en el procesamiento y un mayor costo computacional al momento de procesar los datos de los delitos.

Respecto a los datos utilizados para este proyecto, se generó una base de datos con servicios y sus delitos asociados, generados en forma aleatoria. El algoritmo propuesto permite seleccionar los servicios y el número de áreas que el usuario crea conveniente, para este caso se seleccionaron los servicios de 3 zonas urbanas y como siguiente paso el algoritmo genera en forma aleatoria los delitos asociados a cada servicio con sus posiciones en el mapa.

En la Figura 4.2 se muestra el mapa generado con 72 servicios (cruces negras) y 3710 eventos (puntos rojos) delictivos asociados a los servicios. De todos estos eventos generados se tomaron 3310 para la caracterización y 400 para la etapa de evolución temporal.





Figura 4.1: Mapa de la Comuna de Santiago Centro.

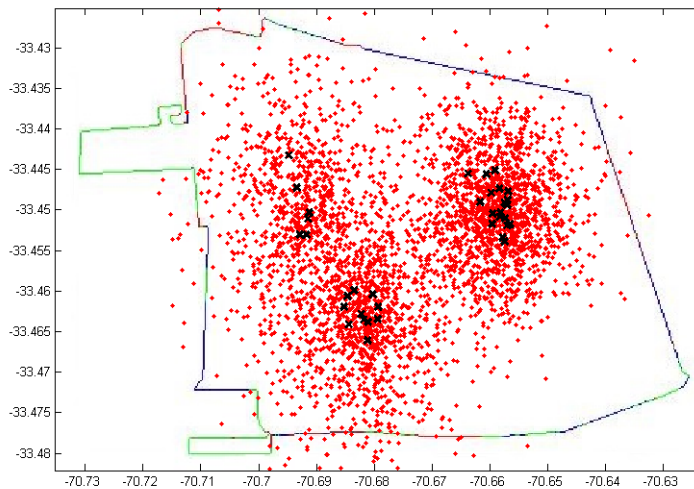


Figura 4.2: Área de Interés con los servicios seleccionados y sus delitos asociados.

## 4.2. Generación de modelos espaciales de riesgo

Una vez que se tienen todos los datos necesarios para la caracterización es necesario representarlos de una forma en que se puedan interpretar y analizar las zonas de riesgo, la mejor forma es mediante *Gaussian Mixture Models*(GMM).

Como se explicó en la sección anterior cada servicio será representado por una distribución normal bivariada, de esta forma se genera una GMM, así se tiene una caracterización del riesgo criminal con la base de datos generada y una representación de los focos de criminalidad en cada uno de los sectores seleccionados (Figura 4.3).

En la Figura 4.3 se puede apreciar la representación de los focos de criminalidad, se los

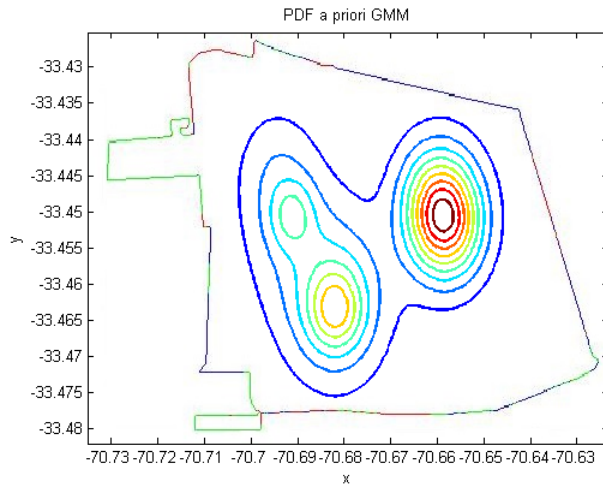


Figura 4.3: Modelo GMM de las zonas de riesgo

representa mediante curvas de nivel, donde los focos más significativos se muestran con colores rojo donde existe un nivel alto de riesgo y que progresivamente disminuye el riesgo los colores van cambiando hasta un color azul.

### 4.3. Evolución Temporal.

En la etapa de evolución temporal las operaciones se inician con la caracterización del riesgo criminal generada a partir de los datos generados. A partir de esta caracterización se va a obtener una aproximación mediante el uso de *muestreo de importancia* y *re-muestreo* (ver Sección 3.3). Para la etapa de *muestreo de importancia* se inicia con un grilla, generando cuadrículas con un valor de una cuadra igual a 150 metros, siendo en total 23x16 que subdivisiones del mapa (Figura 4.4).

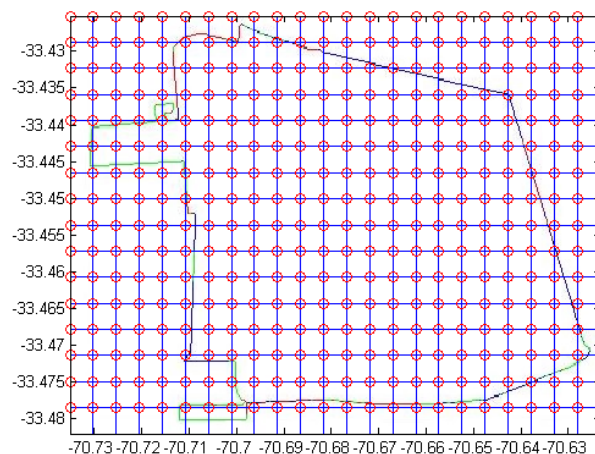


Figura 4.4: Grillado del mapa para la colocación de las muestras.

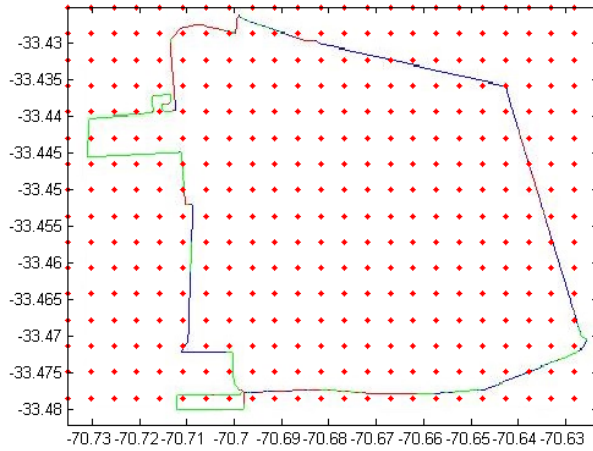


Figura 4.5: Distribución uniforme de las muestras.

De esta forma se discretiza el dominio de la función, para después evaluar el valor de la función en cada punto de la malla, de esta forma se selecciona muestras representativas de la pdf.

A continuación se obtiene una distribución uniforme (Figura 4.5) de las partículas, de esta forma se generan las muestras a partir de la densidad de importancia (Sección 2.6), las mismas que tienen un mayor o menor peso asignado según la distribución *a priori*. Entonces una vez que se tiene la distribución uniforme de las muestras se aplica *resampling*, obteniendo así una aproximación de la distribución *a priori* con 368 partículas (Figura 4.6).

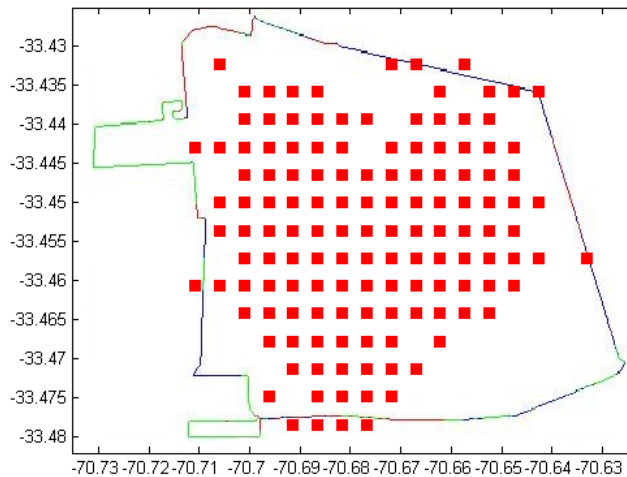


Figura 4.6: Resampling.

Utilizando el modelo dinámico definido por la Ecuación 3.1 (Sección 3.3) se definen los parámetros  $f(d)$  por la Ecuación 3.3 y el ruido  $w(k-1)$  de media nula y una varianza de una cuadra multiplicado por 0.1 (10% del ruido total). Este ajuste se realizó porque el ruido empleado en la Memoria [15] era demasiado alto y las partículas se salían del límite del mapa.

Entonces, con la aproximación de la distribución a priori se introducen nuevos eventos delictivos (400 nuevos eventos delictivos) en forma secuencial, obteniéndose un esquema más real e idóneo para la simulación (Figura 4.7).

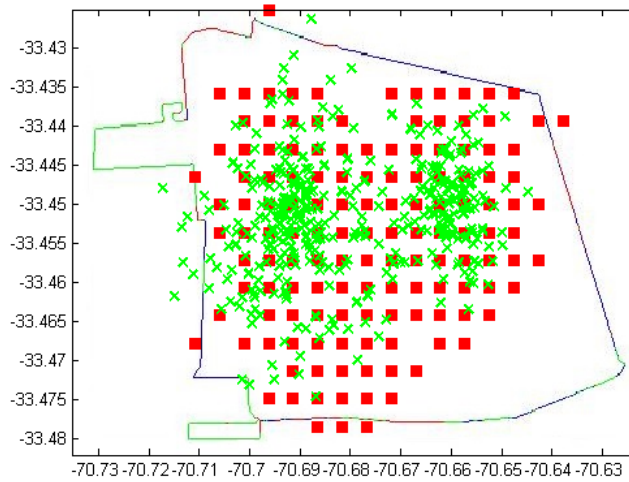


Figura 4.7: Área de Interés con las partículas (rojo) y las nuevas observaciones (verde).

En la etapa de evolución temporal se aplica el modelo presentado en la Ecuación 3.1, pero antes es necesario el ajuste de  $\Sigma$ , ya que los valores de sus componentes definen cuánto una nueva observación afecta al entorno de ésta, entonces como propuesta de este trabajo se optimiza los valores de la diagonal de la matriz  $\Sigma$  con PSO (ver Sección 3.3).

El criterio que se aplica es *HRP* del 15%, y maximizando su respectivo *TIP*, donde la curva tiende a acercarse a la esquina superior izquierda del eje, siendo esto indicativo de un buen modelo [34].

Los resultados obtenidos en cada una de las pruebas con el PSO, para 50 iteraciones y 30 partículas, son los siguientes:

| Número de pruebas | Valores diagonal matriz | TIP(max) |
|-------------------|-------------------------|----------|
| <b>Prueba 1</b>   | 1,33168E-05             | 0,9425   |
| <b>Prueba 2</b>   | 1,45079E-12             | 0,89     |
| <b>Prueba 3</b>   | 6,83232E-10             | 0,915    |
| <b>Prueba 4</b>   | 8,49872E-11             | 0,9375   |
| <b>Prueba 5</b>   | 1,61503E-05             | 0,945    |
| <b>Prueba 6</b>   | 3,54687E-12             | 0,9375   |
| <b>Prueba 7</b>   | 6,87905E-11             | 0,84     |
| <b>Prueba 8</b>   | 3,3459E-12              | 0,915    |
| <b>Prueba 9</b>   | 1,31518E-10             | 0,915    |
| <b>Prueba 10</b>  | 1,72434E-12             | 0,84     |
| <b>Prueba 11</b>  | 1,53693E-10             | 0,9375   |
| <b>Prueba 12</b>  | 4,46479E-09             | 0,9375   |
| <b>Prueba 13</b>  | 1,53777E-11             | 0,9375   |
| <b>Prueba 14</b>  | 1,03991E-11             | 0,89     |
| <b>Prueba 15</b>  | 1,80509E-12             | 0,9375   |
| <b>Prueba 16</b>  | 7,87733E-06             | 0,915    |
| <b>Prueba 17</b>  | 1,20916E-07             | 0,84     |
| <b>Prueba 18</b>  | 6,98925E-06             | 0,8675   |
| <b>Prueba 19</b>  | 1,39071E-05             | 0,94     |
| <b>Prueba 20</b>  | 1,40197E-11             | 0,9375   |

Tabla 4.1: Resultados Finales con PSO para 50 iteraciones y 30 partículas.

En la Tabla 4.1 se puede ver cada uno de los resultados en cada prueba. De las 20 pruebas, el mejor resultado se da en la *Prueba 5*, ya que de todas las pruebas ésta es en la que se obtiene el *TIP* más alto, siendo estos valores de la diagonal de la matriz

$$\Sigma = \begin{bmatrix} 1.615025e-5 & 0 \\ 0 & 1.615025e-5 \end{bmatrix},$$

que es la que se utiliza para las pruebas del modelo dinámico .

Realizando un análisis de los valores TIP resultantes, fue muy útil obtener el área bajo la curva (AUC) de la porción de la gráfica y los valores de AUC de cada una, que se pueden apreciar en las Figuras 4.8 y 4.9, ya que de esta forma el valor del AUC permite discriminar entre los TIPs óptimos que son iguales, pudiéndose utilizar el AUC también como una función objetivo, pero el uso del TIP permite que el cálculo sea más rápido para después obtener el AUC de la porción.

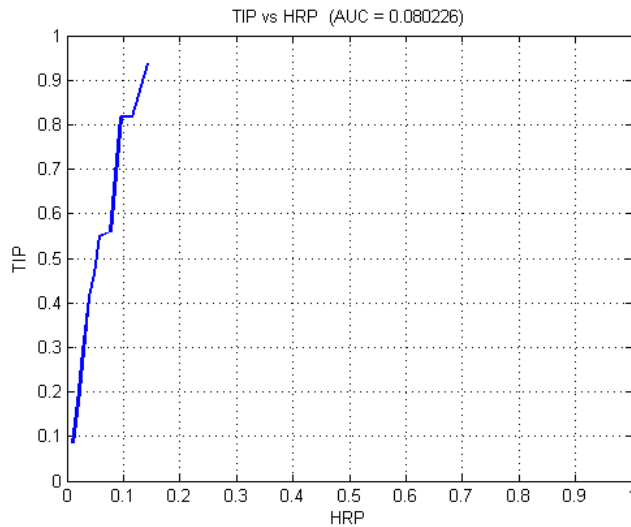


Figura 4.8: Gráfica HRP vs. TIP con el criterio de HRP 15% y TIP=0.9375

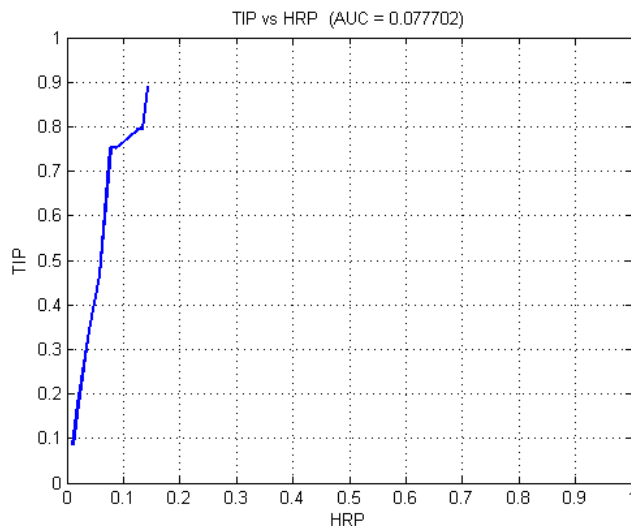


Figura 4.9: Gráfica HRP vs. TIP con el criterio de HRP 15% y TIP=0.8900

En la Figura 4.10 se puede apreciar el resultado aplicando el modelo dinámico, de la Ecuación 3.1, con la matriz  $\Sigma$  encontrada. Es evidente que ciertas partículas cambiaron de posición debido a la “atracción” de las nuevas observaciones que se encuentran dentro del rango de afectación; dicha atracción se puede evidenciar en el contraste que existe entre las partículas de color rojo (posición a priori) y las partículas de color azul (posteriori).

La distribución resultante (partículas en azul), se considera la distribución a posteriori, donde las partículas afectadas por el ruido de proceso es porque se encontraban muy cerca o muy lejos de las nuevas observaciones; mientras que las partículas que se encontraban a cierta distancia de las nuevas observaciones son afectadas por una campo de fuerza más el ruido del proceso, siendo estas partículas las idóneas para comenzar la etapa de predicción.

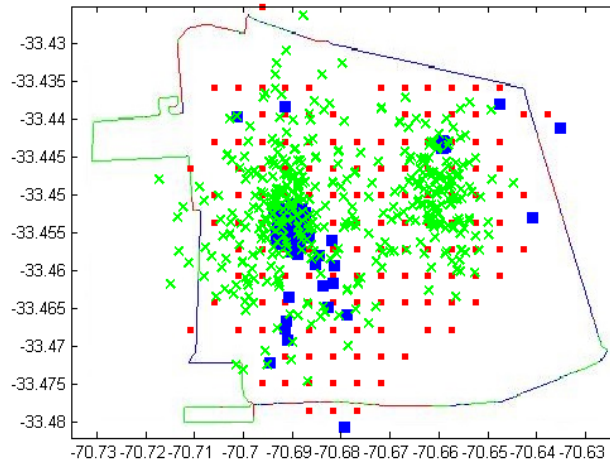


Figura 4.10: Área de Interés con datos a priori (rojo), datos posteriori (azul) y nuevas observaciones (verde).

#### 4.4. Predicción.

En esta etapa se inicia realizando un *clustering* con *K-means* de los nuevos eventos delictivos (nuevas observaciones introducidas en Sección 4.3), considerando 400 eventos para este propósito. Pero antes de realizar esta operación es necesario saber cual es el número de centroides que se va a definir.

Para encontrar el número de centroides es necesario saber cuál es el valor óptimo de *clusters*, para esto se realiza un análisis de los resultados los índices más utilizados (sección 2.9.1) para encontrar el número óptimo de clusters. Los resultados obtenidos con cada uno de los índices son:

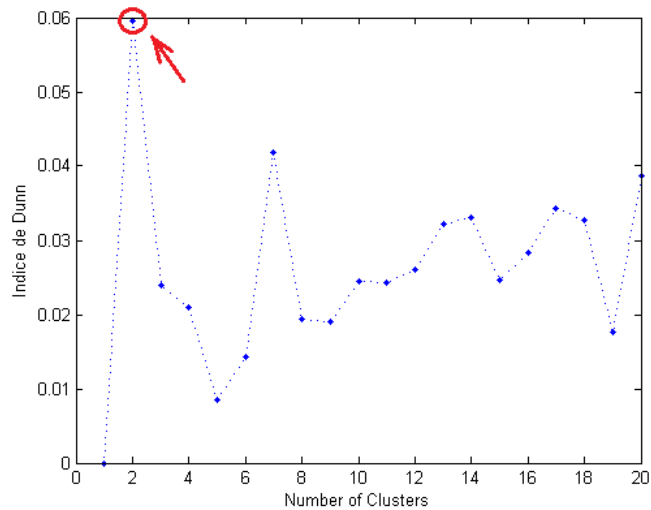


Figura 4.11: Resultado del Índice de Dunn.

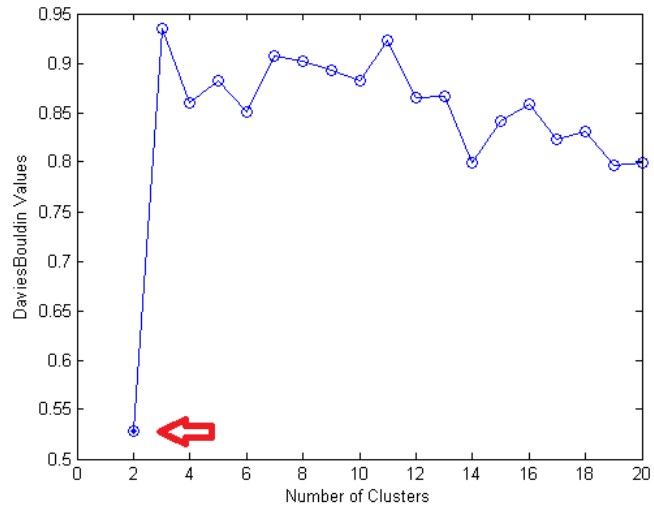


Figura 4.12: Resultado del Índice de Davies-Bouldin.

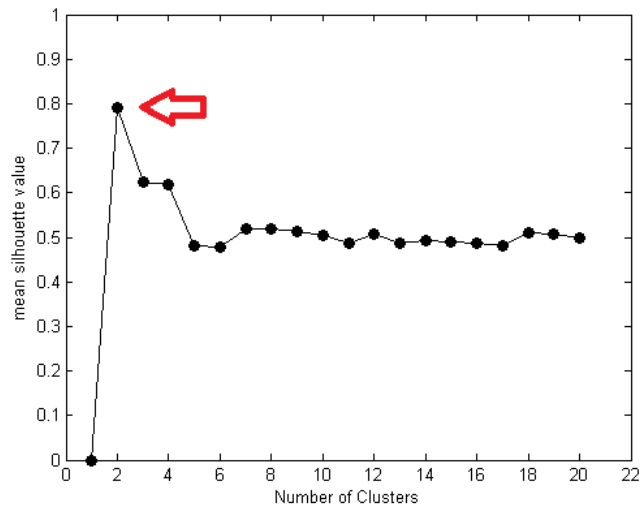


Figura 4.13: Resultado del Índice Silhouette.



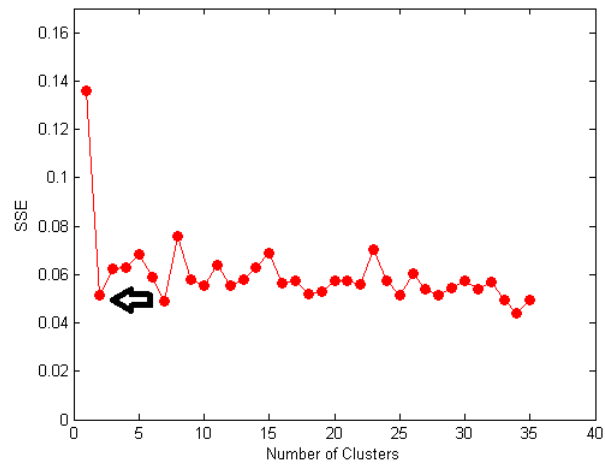


Figura 4.14: Resultado del Índice Sum of The Square Error (SSE).

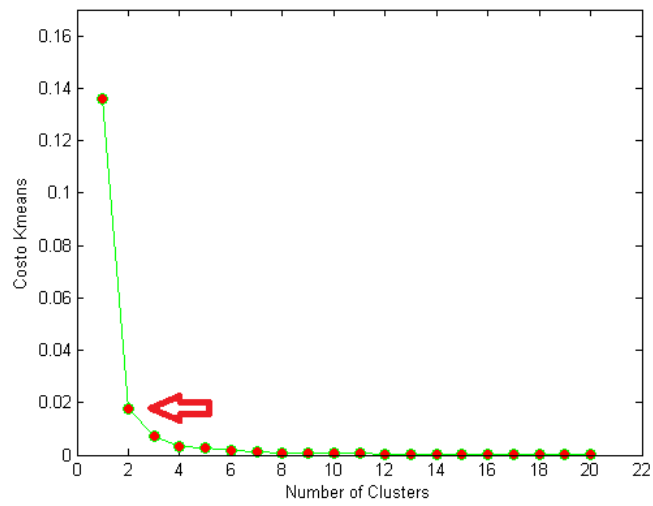


Figura 4.15: Resultado de la Función de Costo.

Considerando las definiciones de la Sección 2.9.1, el resultado en la Figura 4.11 con el índice Dunn nos da el máximo de la función para un valor de  $k = 2$ ; en la Figura 4.12 tenemos el índice de Davies-Bouldin, donde el mínimo valor de la función se da para  $k = 2$ ; para el resultado con el índice Silhouette, que se puede ver en la Figura 4.13, el mejor resultado se da para  $k = 2$ ; en la Figura 4.14, para el índice SSE el mejor resultado es para  $k = 2$ ; finalmente el resultado para la función de costo, Figura 4.15, se da en  $k = 2$  donde existe una inflexión en la función.

La selección de los índices, para seleccionar el resultado óptimo se considera la literatura recomendada [2][23][37], que menciona que los índices más utilizados son el índice de Dunn, Davies-Bouldin y Silhouette. Entonces, considerando los resultados presentados, se tiene los centroides de las nuevas observaciones en la Figura 4.16.

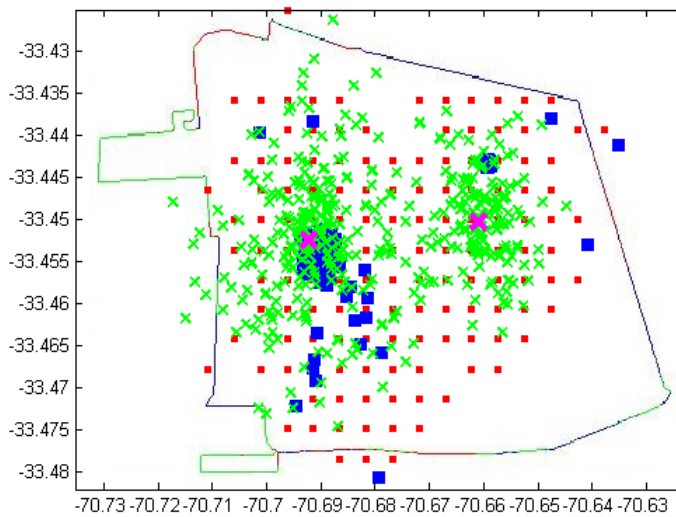


Figura 4.16: Clustering de las nuevas observaciones (verde) con sus respectivos centroides (cruz fucsia) para  $k=2$ .

Además el índice de Davies-Bouldin presenta los siguientes resultados en la ventana de comando de Matlab®:

```
DaviesBouldinEvaluation with properties (resultado de Matlab):
NumObservations: 400
InspectedK: [1x20 double]
CriterionValues: [1x20 double]
OptimalK: 2
```

Una vez que las partículas son atraídas (Figura 4.16), se genera la distribución de predicción (Figura 4.17). Para que las partículas sean atraídas se consideró un horizonte de predicción igual al número de observaciones, para este caso 400. Finalmente para apreciar de mejor manera el resultado de la predicción se la representa mediante una GMM de predicción (Figura 4.18).

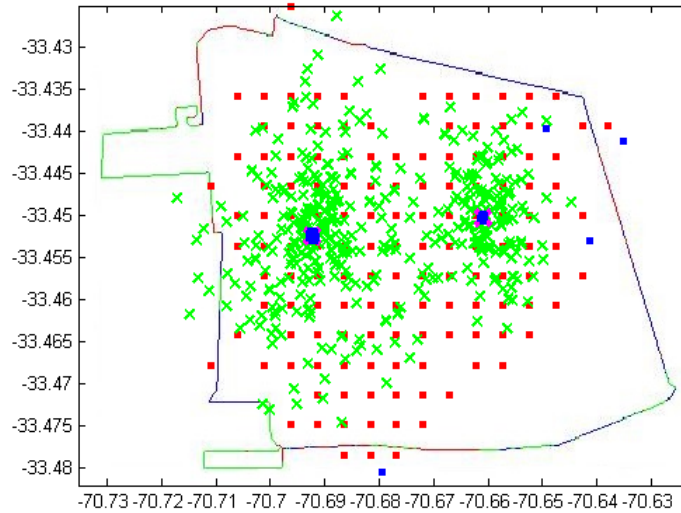


Figura 4.17: Partículas atraídas hacia los centroides con un horizonte de predicción igual a 400.

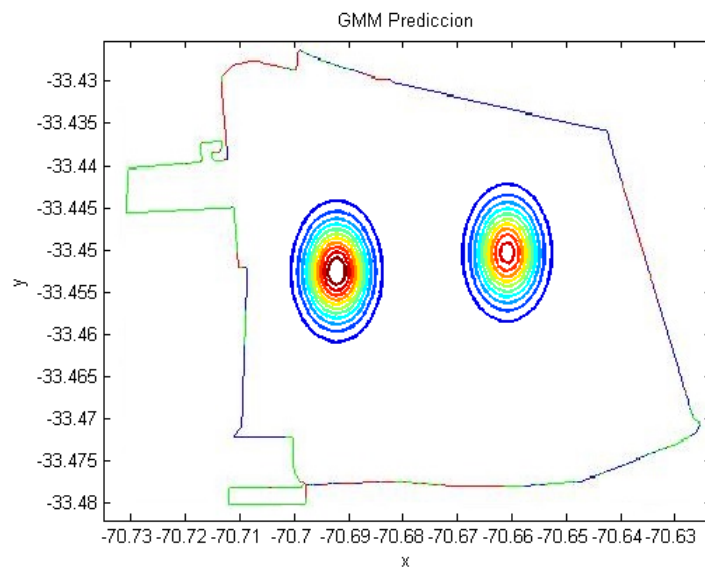


Figura 4.18: Distribución de Predicción.

## 4.5. Evaluación de los Resultados.

El procedimiento para evaluar los resultados de la predicción es:

- Primero se realiza una división del área  $A \subseteq \mathbb{R}^2$  en sectores, donde cada sector es de  $150m^2$ .

- Después se calcula la probabilidad de riesgo en cada sector y se ordenan los sectores de mayor a menor riesgo.
- Por último se realiza una selección de los sectores candidatos en función de HRP del 15 % (selección del 15 % de las áreas de mayor riesgo).

Las pruebas realizadas se hicieron generando 10 distribuciones de predicción, cada una es evaluada con tres grupos de datos simulados uno de 131, otro de 417 y otro 399 datos, que son grupos de datos muy cercanos o menores a la ventana de predicción con que se generaron las predicciones; con estos datos analizamos cuantos de estos datos caen en los sectores candidatos bajo el criterio de HRP del 15 %. Los resultados son:

| <b>Pruebas (pdfs de predicción)</b> | <b>TIP con HRP 15 %</b> |
|-------------------------------------|-------------------------|
| Predicción 1                        | 0.6336                  |
| Predicción 2                        | 0.8015                  |
| Predicción 3                        | 0.8015                  |
| Predicción 4                        | 0.9847                  |
| Predicción 5                        | 0.4351                  |
| Predicción 6                        | 0.8015                  |
| Predicción 7                        | 0.2137                  |
| Predicción 8                        | 0.1908                  |
| Predicción 9                        | 0.9847                  |
| Predicción 10                       | 0.4351                  |

Tabla 4.2: Resultados de la Evaluación de la Predicción (ventana 131)

| <b>Pruebas (pdfs de predicción)</b> | <b>TIP con HRP 15 %</b> |
|-------------------------------------|-------------------------|
| Predicción 1                        | 0.8657                  |
| Predicción 2                        | 0.4532                  |
| Predicción 3                        | 0.4317                  |
| Predicción 4                        | 0.6139                  |
| Predicción 5                        | 0.4556                  |
| Predicción 6                        | 0.4580                  |
| Predicción 7                        | 0.4460                  |
| Predicción 8                        | 0.4005                  |
| Predicción 9                        | 0.5659                  |
| Predicción 10                       | 0.4580                  |

Tabla 4.3: Resultados de la Evaluación de la Predicción (ventana 417)

| Pruebas (pdfs de predicción) | TIP con HRP 15 % |
|------------------------------|------------------|
| Predicción 1                 | 0.8521           |
| Predicción 2                 | 0.5138           |
| Predicción 3                 | 0.5038           |
| Predicción 4                 | 0.6190           |
| Predicción 5                 | 0.5088           |
| Predicción 6                 | 0.5163           |
| Predicción 7                 | 0.4286           |
| Predicción 8                 | 0.3659           |
| Predicción 9                 | 0.5589           |
| Predicción 10                | 0.5138           |

Tabla 4.4: Resultados de la Evaluación de la Predicción (ventana 399 datos).

En las tablas se presentan los resultados de evaluar cada una de las distribuciones de predicción generadas, donde el modelo de predicción se considera que es bueno si el TIP es alto. Existen buenos resultados como en la Tabla 4.2 (Predicción 9), donde el resultado máximo del TIP es muy cercano a 1. Cada una de las predicciones son evaluadas en diferentes conjunto de datos y se puede apreciar que en la mayoría de los casos muchos de los datos generados caen en los sectores elegidos considerando el criterio del HRP 15 %.

# Capítulo 5

## Conclusión

En este trabajo se presenta la caracterización del riesgo criminal de un área urbana, creada en base al perfil de la comuna de Santiago Centro Región Metropolitana. El perfil de esta comuna fue seleccionado ya que tiene una forma sencilla y práctica para el propósito de este proyecto. Para la obtención de la información necesaria para la simulación fue indispensable generar una base de datos artificial, debido a la confidencialidad de los datos reales, logrando un algoritmo que genera datos de servicios y sus delitos asociados.

La limitante de contar con una información de estas características es que los datos son muy dependientes del criterio del usuario al momento de seleccionar las zonas urbanas y los delitos asociados a cada servicio se generan en forma aleatoria a partir de una distribución Gaussiana centrada en el servicio; aunque la base de datos es generada de esta forma es útil para los propósitos de prueba y análisis de los algoritmos presentados en este trabajo.

Una parte de los algoritmos y modelos empleados en el proyecto son basados en la Memoria de título de la Srta. Paulina Flores. Principalmente en las ideas de la caracterización del riesgo criminal mediante una GMM, en la generación de la distribución a priori y a posteriori; resultados necesarios para todos los procesos de caracterización y predicción del riesgo criminal. Pero en las pruebas del modelo dinámico fue necesario reformular ciertas condiciones mencionadas en esta Memoria, como la varianza del ruido de proceso, que se recomienda de 3 cuadras, en cambio para este caso se usó de una cuadra siendo necesario también multiplicar por 0.1 a la varianza del ruido (10 % del ruido); esto se hizo porque en las pruebas se observó que las partículas se salían fuera del mapa generado, dificultándose la obtención de la distribución a posteriori. Después de este procedimiento se puede concluir que el modelo es sensible al cambio de ambiente y es necesario ajustar el modelo según el perfil en el que se aplique.

Como parte del modelo espacial, se implementó una forma de actualizar el modelo dinámico, mediante la búsqueda de una solución óptima con el uso del algoritmo PSO. El algoritmo ajusta los valores de la diagonal de la matriz de covarianza, de esta forma el movimiento o “atracción” entre las partículas y las nuevas observaciones se da dentro del rango definido por el modelo. Con esta propuesta ya no es necesario encontrar los valores de la matriz en forma empírica, es decir que la actualización del modelo espacial ya no estará supeditado al criterio

y experiencia del usuario, sino al resultado del algoritmo PSO. Hay que considerar que la operación de este algoritmo se puede ver afectada por el ruido de proceso y por lo tanto existe mucha variación en los resultados de la función objetivo. Por esta razón lo que se hizo fue obtener el promedio de los resultados de los valores TIP de cada una de las distribuciones, dentro de la función objetivo, que son obtenidos por medio del criterio de HRP del 15 % dentro del algoritmo de PSO; el resultado promedio de cada distribución será el resultado final de la función objetivo. De esta forma se trata de que los resultados no sean tan variables, pero dicha variabilidad no va a desaparecer por completo debido a que existe una gran varianza en los datos. El uso de PSO fue una herramienta muy útil, ya que permite reducir y automatizar la búsqueda de los valores que optimicen el movimiento de las partículas.

El valor resultante del AUC de las figuras 4.8 y 4.9 se puede usar para discriminar entre los resultados que coincidan en el valor del TIP (Tabla 4.1); aplicando el criterio de considerar el AUC se puede obtener un óptimo global para la matriz de covarianza considerando el que tenga el valor más alto de AUC entre ellos.

La metodología propuesta para la etapa de predicción fue usar el algoritmo K-means para agrupar los nuevos eventos criminales introducidos en la actualización del modelo dinámico, de esta forma se encuentran los centros de las posibles zonas de riesgo criminal. Con la idea de “atractor criminal” se generó un modelo empírico dinámico, con el cual los eventos criminales fueron atraídos según la cercanía al atractor.

Aplicando este procedimiento se pudo apreciar cómo fue posible generar y caracterizar la predicción de zonas de riesgo, concluyendo que la consideración del “atractor criminal” es válido para este propósito, ya que permite visualizar y analizar los hot-spot resultantes que se presentan en las gráficas resultantes (Figuras 4.17 y 4.18) y obteniendo resultados en la evaluación de la predicción (con el criterio HRP 15 %) de  $TIP = 0.857$  para una ventana de predicción de 417 y de  $TIP = 0.9847$  para una ventana de 131.

Analizando estos resultados, se puede apreciar cómo se evalúa cada uno de los modelos, según el número de datos generados. El TIP resultante es alto para algunos resultados de la evaluación, llegando en algunos casos a valores de TIP muy cercanos a uno, siendo un buen resultado en la evaluación del modelo de predicción generado. Los resultados se obtuvieron evaluando el modelo generando varias distribuciones de predicción con diferentes ventanas de tiempo.

Para encontrar el número de clusters se usaron los índices del valor óptimo de clusters. Estos métodos permitieron encontrar el número con el cual se estableció cuantos clusters o grupos definirá K-means, siendo el resultado  $k = 2$ ; fue analizado con varios tipos de índices para corroborar el resultado. Hay que considerar que con K-means se agrupan los nuevos eventos criminales, los mismos que son parte de la base de datos generada.

Considerando los proyectos que existen en la actualidad, los objetivos presentados en este proyecto tratan de dar solución a problemas que se presentan en otros proyectos, por ejemplo con la inclusión de la variable temporal en la etapa de evolución temporal, mediante el uso de K-means y un modelo de “atractor criminal” para predecir en una ventana de tiempo futura. Con el uso de estas herramientas se intenta mejorar los resultados obtenidos por el modelo temporal presentado en la memoria de la Srta. Paulina Flores y proponer un

nuevo método para la predicción de áreas de riesgo criminal. El ajuste del modelo temporal se propone en la memoria que sea realizada por el usuario, es decir que está supeditada a su experiencia y conocimiento; en este proyecto se presentó otra forma de realizar dicho procedimiento y es optimizando el ajuste para que el modelo sea óptimo al momento de simular la variable temporal. Con las pruebas realizadas en otro tipo de perfil geográfico se busca que la propuesta de la memoria sea una generalización y analizar cómo se comporta en otros ambientes. El problema que se presenta en la etapa de predicción en la memoria es, que si aumenta la ventana de tiempo empiezan aparecer datos fuera del perfil geográfico, con la propuesta presentada en este trabajo se evita ese problema, ya que con el modelo de “atractor” las partículas son atraídas a los centros de las áreas de riesgo que al final de la ventana de tiempo su posición definirá la distribución de predicción.

Como trabajo futuro se propone mejorar la operación del PSO frente al ruido de proceso introduciendo la función “Stretching”, la cual provee una forma de escape de un mínimo local cuando el PSO se estanca y no converge. Otra propuesta es realizar la caracterización y predicción del riesgo criminal mediante la teoría de modelado basado en agentes (Agent-Based Model); el cual consiste en representar un sistema como un conjunto de entidades de toma de decisiones autónomas (agentes), cada uno de los cuales evalúa su situación y toma de decisiones basado en un conjunto de reglas. Esta forma de modelar y simular permite incluir las acciones criminales en forma individual, el comportamiento criminal y el medioambiente; de esta forma se puede obtener una herramienta más completa para el análisis y predicción del crimen.



# Bibliografía

- [1] M Sanjeev Arulampalam, Simon Maskell, Neil Gordon, and Tim Clapp. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *Signal Processing, IEEE Transactions on*, 50(2):174–188, 2002.
- [2] James C Bezdek and Nikhil R Pal. Some new indexes of cluster validity. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 28(3):301–315, 1998.
- [3] Alfred Blumstein and Jacqueline Cohen. Characterizing criminal careers. *Science*, 237(4818):985–991, 1987.
- [4] Patricia Brantingham and Paul Brantingham. Criminality of place. *European journal on criminal policy and research*, 3(3):5–26, 1995.
- [5] Patricia L Brantingham and Paul J Brantingham. Nodes, paths and edges: Considerations on the complexity of crime and the physical environment. *Journal of Environmental Psychology*, 13(1):3–28, 1993.
- [6] Paul J Brantingham and Patricia L Brantingham. Environment, routine and situation: Toward a pattern theory of crime. *Advances in criminological theory*, 5:259–294, 1993.
- [7] James V Candy. *Bayesian signal processing: Classical, modern and particle filtering methods*, volume 54. John Wiley & Sons, 2011.
- [8] Joel M Caplan and Leslie W Kennedy. Risk terrain modeling compendium. *Rutgers Center on Public Security, Newark*, 2011.
- [9] D. Muñoz D. Sáez. Apuntes del ramo el 7012, algoritmos genéticos y optimización por enjambre de partículas. Departamento de Ingeniería Eléctrica, Universidad de Chile.
- [10] Gobierno de Chile. Glosario ministerio de desarrollo social. observatorio .ministeriode-sarrollosocial.gob.cl/glosario.php, 2015.
- [11] Municipalidad de Santiago. Portal de la municipalidad de santiago. <http://www.municipalidaddesantiago.cl/categorias/home/la-comuna/demografia/verPagina/plano-de-la-comuna-2>, 2016.
- [12] Russ C Eberhart, James Kennedy, et al. A new optimizer using particle swarm theory. In *Proceedings of the sixth international symposium on micro machine and human science*,

volume 1, pages 39–43. New York, NY, 1995.

- [13] Tom Fawcett. An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874, 2006.
- [14] Juan José Pantrigo Fernández. *Resolución de problemas de optimización dinámica mediante la hibridación entre filtros de partículas y metaheurísticas poblacionales*. PhD thesis, Universidad Rey Juan Carlos, 2005.
- [15] Paulina Flores. Modelación y predicción de focos de criminalidad basado en modelos probabilísticos. Master’s thesis, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, 2014.
- [16] Paulina Flores, Mario Vergara, Pablo Fuentes, Francisco Jaramillo, David Acuna, Aramis Perez, and Marcos Orchard. Modeling and prediction of criminal activity based on spatio-temporal probabilistic risk functions. Prognostics and Health Management Society, 2015.
- [17] Andrew Gelman, John B Carlin, Hal S Stern, and Donald B Rubin. *Bayesian data analysis*, volume 2. Chapman & Hall/CRC Boca Raton, FL, USA, 2014.
- [18] Matthew S Gerber. Predicting crime using twitter and kernel density estimation. *Decision Support Systems*, 61:115–125, 2014.
- [19] Alberto R Gonzales, Regina B Schofield, Sarah V Hart, John E Eck, Spencer Chainey, James G Cameron, Michael Leitner, and Ronald E Wilson. Mapping crime: Understanding hot spots. 2005.
- [20] Jiawei Han. *Micheline Kamber amd Simon Fraser University “Data Mining Concepts and Techniques” Morgan Kaufmann Publishers*. USA, 2001.
- [21] Dan Helms. Temporal analysis. *Exploring Crime Analysis: Readings on Essencial Skills. North Charleston, CS: International Association of Crime Analysis, Booksurge*, pages 220–262, 2004.
- [22] Renjie Liao, Xueyao Wang, Lun Li, and Zengchang Qin. A novel serial crime prediction model based on bayesian learning theory. In *2010 International Conference on Machine Learning and Cybernetics*, volume 4, pages 1757–1762. IEEE, 2010.
- [23] Yongguo Liu, Mao Ye, Jun Peng, and Hong Wu. Finding the optimal number of clusters using genetic algorithms. In *Cybernetics and Intelligent Systems, 2008 IEEE Conference on*, pages 1325–1330. IEEE, 2008.
- [24] Gregg O McCrary. Paulsen, derek j.; bair, sean; helms, dan (2010) tactical crime analysis; research and investigation crc press taylor & francis group, llc. *Investigative Sciences Journal*, 2(3), 2010.
- [25] Claudio Held Pablo Estévez. Inteligencia computacional: Validación de clustering. Departamento de Ingeniería Eléctrica, Universidad de Chile, 2013.

- [26] Konstantinos E Parsopoulos. *Particle Swarm Optimization and Intelligence: Advances and Applications: Advances and Applications*. IGI Global, 2010.
- [27] Konstantinos E Parsopoulos and Michael N. Vrahatis. Recent approaches to global optimization problems through particle swarm optimization. *Natural computing*, 1(2-3):235–306, 2002.
- [28] Derek Paulsen, Sean Bair, and Dan Helms. Pronosticar y predecir. *Índice*, page 128, 2012.
- [29] Riccardo Poli, James Kennedy, and Tim Blackwell. Particle swarm optimization. *Swarm intelligence*, 1(1):33–57, 2007.
- [30] Michael D Porter and Donald E Brown. Detecting local regions of change in high-dimensional criminal or terrorist point processes. *Computational statistics & data analysis*, 51(5):2753–2768, 2007.
- [31] David Stork Richard Duda, Peter Hart. *Pattern Classification*. Jhon WILEY and Sons, 2001.
- [32] Sandro Saitta, Benny Raphael, and Ian FC Smith. A bounded index for cluster validity. In *Machine learning and data mining in pattern recognition*, pages 174–187. Springer, 2007.
- [33] Shiju Sathyadevan, MS Devan, and S Surya Gangadharan. Crime analysis and prediction using data mining. In *Networks & Soft Computing (ICNSC), 2014 First International Conference on*, pages 406–412. IEEE, 2014.
- [34] Xiaofeng Wang and Donald E Brown. The spatio-temporal generalized additive model for criminal incidents. In *Intelligence and Security Informatics (ISI), 2011 IEEE International Conference on*, pages 42–47. IEEE, 2011.
- [35] Xiaofeng Wang and Donald E Brown. The spatio-temporal modeling for criminal incidents. *Security Informatics*, 1(1):1–17, 2012.
- [36] Guoshen Yu and Guillermo Sapiro. Statistical compressed sensing of gaussian mixture models. *IEEE Transactions on Signal Processing*, 59(12):5842–5858, 2011.
- [37] Xiao Zhang, Aichen Li, You Zhang, and Yongpeng Xiao. Validity of cluster technique for genome expression data. In *Control and Decision Conference (CCDC), 2012 24th Chinese*, pages 3737–3741. IEEE, 2012.