PROPUESTA PARA UNA TÉCNICA DE AGRUPAMIENTO DIFUSO DE EVENTOS CRIMINALES BASADO EN LA ESTRUCTURA URBANA

Luigi Santiago Fajardo Toloza 20111005090



Universidad Distrital Francisco José de Caldas
Facultad de Ingeniería
Bogota D.C. - Colombia
23 de mayo de 2017

Trabajo de Grado
Requisito para optar por el título de Ingeniero Electrónico.

Luigi Santiago Fajardo Toloza 20111005090

Director:
Miguel Alberto Melgarejo Rey
Profesor Asociado
Laboratorio de Automática e Inteligencia Computacional
Facultad de Ingeniería

Universidad Distrital Francisco José de Caldas
Facultad de Ingeniería
Bogota D.C. - Colombia

"Para explicar toda la naturaleza no basta ni un hombre ni una edad completa. En su lugar, lo mejor es que el hombre busque un poco de verdad y certeza, dejando el resto para los demás, para los que vendrán, con conjeturas y sin dar nada por hecho."

Sir. Isaac Newton

Nota de Aceptación:

Firma del Revisor del Trabajo de Grado

Firma del Director del Trabajo de Grado

Bogotá D.C. 2017

A mis padres Alejandro e Irene por levantarse cada día
con el propósito de sacar esta familia adelante.
A mi loco y mi flaca por ser el motor que me inspira
a cada mañana superarme en la vida.
A Don Fabio y Doña Eunice por enseñarme día a día
que la vida difícil también es bella.

Resumen

Este libro reúne el desarrollo de la investigación correspondiente a un trabajo de grado realizado desde inicios del año 2015 hasta inicios del año 2017 por parte de un estudiante aspirante al título de Ingeniero Electrónico en colaboración con el grupo de investigación LAMIC (Laboratorio de Automática e Inteligencia Computacional).

Una de las áreas de estudio del grupo de investigación son la lógica y los conjuntos difusos y las aplicaciones que se pueden desarrollar en campos tanto científicos como sociales. Uno de estos casos es el reciente interés por abordar las teorías de análisis del crimen desde el análisis de los conjuntos difusos.

Para este trabajo se busca obtener un método que permita determinar la cantidad de agrupaciones en las que un conjunto de datos referentes a una conducta criminal ocurrida en un periodo de tiempo deban ser separados para así analizar cada conjunto como un independiente. Para desarrollar este método de determinación del numero de agrupaciones, es indispensable tener como base la estructura urbana donde han ocurrido estos sucesos.

Primariamente se seleccionan los datos de la conducta criminal que va a ser separada en agrupamientos individuales, de la misma manera se selecciona la estructura urbana que será separada en la misma cantidad de agrupamientos que los sucesos criminales. Para la etapa de agrupamiento se utilizan los algoritmos de agrupamiento difuso Fuzzy C-Means y Gustafson Kessel. Cada algoritmo se ejecuta desde un mínimo de 2 agrupamientos hasta un numero determinado por la primera percepción de divisibilidad que el diseñador de la técnica tenga sobre el conjunto correspondiente a la estructura urbana.

Posteriormente las particiones difusas obtenidas por los algoritmos de agrupamiento son evaluadas mediante técnicas estadísticas para la comparación de agrupamientos difusos, estas técnicas se basan en las diferencias presentadas para las distancias entre los diferentes conjuntos de datos obtenidos en los agrupamientos. El procesamiento de estos estadísticos sobre los agrupamientos entrega la cantidad de agrupamientos donde el conjunto de sucesos criminales y el conjunto de datos de estructura urbana tienen la mayor similitud.

Al desarrollar la técnica, esta se debe probar con dos diferentes áreas de estudio y dos tipos de crímenes diferentes, de esta manera se ejecutan

pruebas sobre el rendimiento de la técnica en el aspecto computacional, la validación de la técnica para mas de un escenario, y por ultimo, estos dos escenarios serán separados en diferentes ventanas de tiempo, de manera que se pueda validar tanto el funcionamiento de la técnica para diferentes periodos de tiempo, además de comprobar que el comportamiento de estos sucesos se repite tanto en diferentes periodos de tiempo, como en diferentes escalas de tiempo.

Los resultados obtenidos muestran que la técnica permite obtener el numero de agrupamientos en el cual debe ser separado el conjunto de datos de sucesos criminales para que su estudio se acople a la estructura urbana donde han ocurrido, de manera que conociendo la cantidad de áreas de estudio donde han ocurrido los crímenes, cada una puede ser analizada de manera individual con el objetivo de obtener patrones mas acertados sobre la aparición y comportamientos de los crímenes en dicha zona. Por ultimo, a comprobar que para los casos de estudio analizados, el comportamiento a menor escala de tiempo es similar al comportamiento a mayor escala, se puede deducir que la técnica permite obtener en el caso de menor escala. que representa menor cantidad de datos, resultados similares a su escala mayor, de manera que se puede hablar de una técnica que permite la reducción de costos computacionales altos en su operación.

Agradecimientos

En primera instancia agradezco a la causalidad que me ha llevado hasta este punto de mi historia, espero que los retos y acontecimientos que sigan presentándose en mi vida se rijan como lo han hecho hasta este momento, por la hermosa ley de causa y efecto.
Agradezco a la Universidad Distrital Francisco José de Caldas por permitirme ser parte de su gloriosa historia, por ser mi hogar durante estos maravillosos años en los que aprendí que el medio mas grande que tenemos para luchar contra aquellos que no quieren cambiar el mundo es el conocimiento.
A la Ingeniería Electrónica por reafirmar mi amor hacia los números e inculcarme el amor hacia todo lo que estos me permiten crear.
A Alejandro, Irene, Daniel y Juanita por las risas, los consejos, los momentos alegres que nos unen día a día y los momentos difíciles que nos han hecho mas fuertes y mas sabios. Solamente espero que me alcance el tiempo para devolverles todo lo que me han dado y que sepan que ustedes son el impulso que activa día a día este sistema.
A Fabio, Eunice, Hugo, Luz y Sebastian por estar siempre que lo he necesitado, por las enriquecedoras charlas que he tenido y que sigo teniendo con cada uno de ustedes. Por enseñarme que cada persona es un mundo que jamás terminaras de conocer.
A Brian Diaz, Fernando Arias, Ivonne Roa, Miguel Tuberquia, Nicolas Silva y Santiago Galindo. Dicen que los amigos son la familia que uno escoge, recordando las risas entre clases, pasillos y laboratorios, las lágrimas, los consejos, la fortaleza y el apoyo que nos hemos brindado unos a otros durante estos años me doy cuenta que la elección de mi familia no ha podido ser más acertada.
Al profesor Miguel Melgarejo por permitirme trabajar junto a él en este proyecto que ha redescubierto en mi la verdadera naturaleza de lo que buscar resolver la ingeniería en la sociedad. Por sus charlas, por sus consejos, por compartir conmigo su visión del mundo y de hacia donde enfocar los esfuerzos en pro de la humanidad.
No tengo mas que agradecimientos hacia este capitulo de mi vida que termina, puesto que a pesar de que no todos los momentos fueron perfectos, logre entender que en la vida se debe aceptar las cosas que no se pueden cambiar, tener el valor de cambiar aquellas que se pueden y por ultimo, tener la sabiduría para distinguir la diferencia entre ambas.

Índice general

Índice de Figuras ix

Índice de Tablas xvii

1. Introducción **1**
- **1.1. Planteamiento del problema** . . . 2
- **1.2. Solución propuesta** . . . 4
- **1.3. Contenido del libro** . . . 5

2. Marco de referencia **6**
- **2.1. Teorías de Análisis del Crimen** . . . 6
- **2.2. Conjuntos Difusos** . . . 8
  - **2.2.1. Lógica Difusa** . . . 8
  - **2.2.2. Grado de Pertenencia Difusa** . . . 9
  - **2.2.3. Operaciones en Conjuntos Difusos** . . . 9
    - **2.2.3.1. Unión** . . . 10
    - **2.2.3.2. Intersección** . . . 10
    - **2.2.3.3. Complemento** . . . 10
- **2.3. Métodos de Agrupamiento de Datos** . . . 10
  - **2.3.1. Cluster** . . . 10
  - **2.3.2. Algoritmo Fuzzy C-Means Clustering(FCM)** . . . 11
    - **2.3.2.1. Función Objetivo C-Means** . . . 12
    - **2.3.2.2. Algoritmo** . . . 12
  - **2.3.3. Algoritmo de Gustafson-Kessel(GK)** . . . 14
- **2.4. Validación de Clúster** . . . 16
  - **2.4.1. Evaluación interna** . . . 16
    - **2.4.1.1. Índice de Davies–Bouldin** . . . 17
    - **2.4.1.2. Índice de Dunn** . . . 17
  - **2.4.2. Evaluación externa** . . . 17
    - **2.4.2.1. Índice de Rand** . . . 18
    - **2.4.2.2. Índice Jaccard** . . . 18
    - **2.4.2.3. Índice de Fowlkes–Mallows** . . . 19
- **2.5. Índices de Validación de Agrupamiento Difuso** . . . 19
  - **2.5.1. Coeficiente de Partición(PC)** . . . 19
  - **2.5.2. Entropía de Clasificación(CE)** . . . 20
- **2.6. Resumen** . . . 20

3. Diseño de una Técnica de Agrupamiento Difuso de Eventos Criminales Basado en la Estructura Urbana **21**
3.1. Solución Propuesta 21
3.2. Selección de Los Conjuntos de Datos 22
3.2.1. Proyección de Mercator 23
3.3. Algoritmos de Agrupamiento Difuso 25
3.4. Métodos de Validación de Clúster 26
3.4.1. Índice Rand para Agrupamientos Difusos 26
3.4.2. Índice Rand para diferentes conjuntos de datos . 28
3.5. Algoritmo Propuesto 28
3.6. Resumen 29

4. Resultados **30**
4.1. Conjuntos de Datos 30
4.2. Elección del numero de grupos 31
4.3. Resultados 33
4.3.1. Rendimiento de los algoritmos de agrupamiento . 33
4.3.2. Rendimiento de los algoritmos Índice de Validación de Clúster 34
4.3.3. Índices de Validación de Clúster 35
4.3.3.1. Bogotá D.C.,CO 35
4.3.3.2. San Francisco,CA,USA 40
4.4. Análisis Generales 44

5. Conclusiones y trabajos futuros **46**
5.1. Resumen 46
5.2. Aportes Originales 47
5.3. Trabajos futuros 48

A. Anexos **50**

B. Anexos **62**

C. Anexos **69**

Bibliografía **74**

Índice de figuras

2.1. Agrupamiento obtenido para un conjunto aleatorio de datos. 11
2.2. Algoritmo FCM aplicado para la distribución de 8 agrupaciones sobre una data generada aleatoriamente. 13
2.3. Algoritmo GK aplicado para la distribución de 8 agrupaciones sobre una data generada aleatoriamente. 15

3.1. Diagrama de la Solución Propuesta. 22
3.2. Conjunto de Datos de la Malla Vial en la ciudad de Bogotá D.C., Colombia. 24
3.3. Conjunto de Datos de los Hurtos a Celulares en la ciudad de Bogotá D.C., Colombia. 24
3.4. Conjunto de Datos de la Malla Vial en la ciudad de San Francisco,CA,USA. 24
3.5. Conjunto de Datos de los Hurtos a Residencias en la ciudad de San Francisco,CA,USA. 24
3.6. Conjunto de Datos de la Malla Vial y los Hurtos en la ciudad de Bogotá D.C.,Colombia 25
3.7. Conjunto de Datos de la Malla Vial y los Hurtos en la ciudad de San Francisco,CA,USA. 25

4.1. Funciones Objetivo para el Agrupamiento mediante el algoritmo FCM en la ciudad de Bogotá D.C.,CO. 36
4.2. Funciones Objetivo para el Agrupamiento mediante el algoritmo GK en la ciudad de Bogotá D.C.,CO. 38
4.3. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 18 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 39
4.4. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 18 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . 39
4.5. Agrupamiento obtenido mediante el algoritmo Gustafson Kessel para 3 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 39
4.6. Agrupamiento obtenido mediante el algoritmo Gustafson Kessel para 3 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . . 39

4.7. Funciones Objetivo para el Agrupamiento mediante el algoritmo FCM en la ciudad de Bogotá D.C.,CO. 41
4.8. Funciones Objetivo para el Agrupamiento mediante el algoritmo GK en la ciudad de San Francisco,CA,USA. . 43
4.9. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 3 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA,USA. . . . 44
4.10. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 3 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA,USA. 44
4.11. Agrupamiento obtenido mediante el algoritmo Gustafson Kessel para 2 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA,USA. . . . 44
4.12. Agrupamiento obtenido mediante el algoritmo Gustafson Kessel para 2 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA,USA. 44

A.1. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 2 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 50
A.2. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 2 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . . 50
A.3. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 3 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 51
A.4. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 3 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . . 51
A.5. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 4 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 51
A.6. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 4 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . . 51
A.7. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 5 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 51
A.8. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 5 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . . 51
A.9. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 6 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 52
A.10. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 6 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . . 52

A.11.Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 7 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 52
A.12.Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 7 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . . 52
A.13.Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 8 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 52
A.14.Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 8 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . . 52
A.15.Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 9 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 53
A.16.Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 9 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . . 53
A.17.Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 10 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 53
A.18.Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 10 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . 53
A.19.Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 11 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 53
A.20.Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 11 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . 53
A.21.Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 12 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 54
A.22.Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 12 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . 54
A.23.Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 13 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 54
A.24.Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 13 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . 54
A.25.Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 14 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 54
A.26.Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 14 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . 54

A.27.Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 15 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 55
A.28.Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 15 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . . 55
A.29.Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 16 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 55
A.30.Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 16 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . . 55
A.31.Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 17 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 55
A.32.Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 17 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . . 55
A.33.Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 18 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 56
A.34.Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 18 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . . 56
A.35.Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 2 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 56
A.36.Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 2 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . . 56
A.37.Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 3 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 56
A.38.Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 3 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . . 56
A.39.Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 4 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 57
A.40.Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 4 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . . 57
A.41.Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 5 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 57
A.42.Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 5 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . . 57

A.43.Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 6 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 57
A.44.Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 6 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . . 57
A.45.Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 7 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 58
A.46.Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 7 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . . 58
A.47.Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 8 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 58
A.48.Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 8 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . . 58
A.49.Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 9 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 58
A.50.Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 9 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . . 58
A.51.Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 10 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 59
A.52.Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 10 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . 59
A.53.Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 11 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 59
A.54.Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 11 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . 59
A.55.Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 12 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 59
A.56.Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 12 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . 59
A.57.Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 13 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 60
A.58.Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 13 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . 60

A.59. Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 14 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 60
A.60. Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 14 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . 60
A.61. Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 15 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 60
A.62. Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 15 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . 60
A.63. Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 16 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 61
A.64. Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 16 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . 61
A.65. Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 17 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 61
A.66. Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 17 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . 61
A.67. Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 18 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO. 61
A.68. Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 18 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO. . 61

B.1. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 2 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA. . . . 62
B.2. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 2 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA. 62
B.3. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 3 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA. . . . 63
B.4. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 3 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA. 63
B.5. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 4 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA. . . . 63

B.6. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 4 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA. 63
B.7. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 5 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA. . . . 63
B.8. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 5 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA. 63
B.9. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 6 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA. . . . 64
B.10. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 6 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA. 64
B.11. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 7 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA. . . . 64
B.12. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 7 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA. 64
B.13. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 8 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA. . . . 64
B.14. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 8 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA. 64
B.15. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 9 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA. . . . 65
B.16. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 9 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA. 65
B.17. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 10 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA. . 65
B.18. Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 10 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA. 65
B.19. Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 2 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA. . . . 65
B.20. Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 2 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA. 65

B.21. Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 3 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA. . . . 66
B.22. Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 3 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA. 66
B.23. Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 4 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA. . . . 66
B.24. Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 4 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA. 66
B.25. Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 5 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA. . . . 66
B.26. Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 5 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA. 66
B.27. Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 6 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA. . . . 67
B.28. Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 6 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA. 67
B.29. Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 7 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA. . . . 67
B.30. Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 7 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA. 67
B.31. Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 8 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA. . . . 67
B.32. Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 8 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA. 67
B.33. Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 9 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA. . . . 68
B.34. Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 9 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA. 68
B.35. Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 10 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA. . 68
B.36. Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 10 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA. 68

Índice de cuadros

4.1. Conjuntos de datos utilizados en la Técnica propuesta . 31
4.2. Criterios de elección para el Limite Superior en la ejecución de la técnica 32
4.3. Algoritmos ejecutados 32
4.4. Cantidad de ejecuciones para los algoritmos utilizados en la técnica desarrollada 33
4.5. Tiempo promedio de ejecución para cada algoritmo de agrupamiento difuso. 33
4.6. Número de ejecución elegida para el para el procesamiento posterior por sus resultados en el agrupamiento difuso 34
4.7. Tiempo promedio de ejecución para cada algoritmo índice de validación de Clúster. 34
4.8. Resultados de las Funciones Objetivo ejecutadas para para el algoritmo Fuzzy C-Means en la ciudad de Bogotá D.C.,CO 35
4.9. Resultados de las Funciones Objetivo ejecutadas para para el algoritmo Gustafson Kessel en la ciudad de Bogotá D.C.,CO 37
4.10. Funciones Objetivo para el Agrupamiento mediante el algoritmo FCM en la ciudad de San Francisco,CA.USA . 40
4.11. Funciones Objetivo para el Agrupamiento mediante el algoritmo GK en la ciudad de San Francisco,CA.USA . . 42

C.1. Tiempos empleados ejecución Algoritmo Fuzzy C-Means sobre Bogotá D.C.,CO 70
C.2. Tiempos empleados ejecución Algoritmo Gustafson Kessel sobre Bogotá D.C.,CO 71
C.3. Tiempos empleados ejecución Algoritmo Fuzzy C-Means sobre San Francisco,CA,USA 72
C.4. Tiempos empleados ejecución Algoritmo Gustafson Kessel sobre San Francisco,CA,USA 73

Capítulo 1

Introducción

"La mayoría de las personas gastan más tiempo y energías en hablar de los problemas que en afrontarlos."

Henry Ford

La humanidad se encuentra ante una de las épocas más proliferas de su historia reciente. En el último siglo se ha logrado equiparar la cantidad de avances tecnológicos de la última mitad del milenio anterior, y de esta manera, la humanidad ha pasado de centrar sus esfuerzos científicos en la búsqueda del origen de las cosas, hacia la resolución de los problemas que la aquejan. Dicho de otra manera, se ha dejado de querer responder la preguntar de ¿de dónde venimos? Para responder la de ¿hasta dónde podremos llegar?

Y esta misma concepción de la resolución de los problemas como eje del progreso de la humanidad los ha llevado a querer atacar dichos problemas desde todas las áreas del conocimiento, ya sean del conocimiento científico como el social, se busca solucionar los problemas que genera tanto la naturaleza, como los creados por la misma humanidad. Y es que una de las razones que han motivado a este cambio en el eje del progreso de la humanidad es el reconocimiento que el paso de la especie humana por la tierra ha dejado grandes avances pero a su vez una gran cantidad de problemas por resolver, siendo uno de ellos el daño que los humanos causan entre ellos mismos en su afán por obtener algo que desean, estos sucesos llamados crímenes, son acciones humanas que han acompañado de manera negativa el surgir y crecimiento de todas las

civilizaciones en la historia antigua y reciente.
La humanidad históricamente ha intentado comprender este accionar desde el conocimiento social, pero ante la falta de respuestas acertadas sobre su origen se ha intentado recientemente asignar esta tarea al conocimiento científico, donde analizar los factores que permiten las ejecución de un crimen desde los sistemas complejos ha tomado una acogida reciente desde una de las ramas del conocimiento científico como lo es la ingeniería.
De modo que a partir del análisis de crímenes como conjuntos de datos que se desplazan sobre un espacio R2 en un periodo de tiempo especifico, considerando la evolución de estos datos como señales que se mueven a través del tiempo, permite ejecutar herramientas propias de la ingeniería que pueden llegar a describir el comportamiento de estos datos.
En este proyecto se diseña una técnica de agrupamiento de eventos criminales basados en la estructura urbana en la cual ocurren estos sucesos. Se parte desde el agrupamiento mediante técnicas difusas del conjunto de crímenes y del conjunto de calles que componen la estructura urbana de estudio de manera individual, para posteriormente mediante técnicas de análisis de similitud de conjuntos difusos obtener la cantidad de agrupaciones que asemejan los dos conjuntos de datos. Finalmente, se ejecuta la técnica obtenida para dos áreas urbanas de estudio, cada una con diferentes modalidades de crímenes ocurridas en un periodo de tiempo, pero analizadas durante diferentes ventanas de tiempo, analizando los resultados de dichos experimentos con el fin de comprobar la funcionalidad de la técnica desarrollada.

1.1. Planteamiento del problema

El crimen no es algo que se descubre en la naturaleza, pertenece a la invención humana. Pertenece al dominio de los sistemas artificiales. No puede ser explicado con afirmaciones universales, leyes o hipótesis, es cambiante y obedece a relaciones, decisiones y propósitos humanos caracterizados por actuar de maneras que no han sido observadas, que innovan constantemente. Obedece además a las condiciones en las que se desarrolla, a la historia que lo construyó en el tiempo y a los lugares específicos en los que ocurre [1]. El hurto es un crimen que deja

cada día 74 víctimas tan solo en la ciudad de Bogotá[2]. Según cifras oficiales, en esta ciudad en el 2014 se registraron 27190 hurtos a personas, representando un 65 % de los delitos que se cometen en la capital del país. Estas cifras que representan la victimización de la ciudadanía, muestran también que un 44 % de las personas consideran este tipo de hurto como el que más concurre y afecta a la ciudad[2].

Estas cifras muestran una alta percepción de inseguridad en los habitantes de la ciudad. Es una obligación de los gobiernos distritales la implementación de políticas públicas que reduzcan estos crímenes a tasas que permitan a los ciudadanos sentirse tranquilos[3].

Uno de los interrogantes es desde donde atacar este problema. Esta pregunta ha llevado al aumento de no solamente el pie de fuerza, que efectúa la tarea de evitar los crímenes en la calle, también ha permitido la investigación y el análisis de patrones, conductas y situaciones que permitan determinar patrones que puedan existir en el crimen[3].

Uno de los retos fundamentales en el análisis del crimen es el reconocimiento de patrones[4]. Existen métodos para detectar zonas geográficas de alta actividad, las cuales son agrupadas por regiones de interés llamados grupos[5]. El análisis de grupos es una técnica usada ampliamente en investigación científica. Es usada en geografía [6], salud pública [7], criminología [5], entre otras áreas de interés [8]. La detección de grupos en una zona específica requiere de un conocimiento de las componentes(i.e. calles, carreras, puentes, parques, obras en construcción) y características(i.e. iluminación, concurrencia, presencia horaria) de dicha zona[5]. Revisando la literatura relacionada con la búsqueda de patrones criminales, se observa que un problema con la distribución en grupos en la búsqueda de patrones es que esta técnica no tiene en cuenta la información que entrega la forma urbana de la misma.

Estas componentes y características que entrega la forma urbana de una zona son usadas desde técnicas que buscan obtener patrones basados en los atributos estadísticos y en la densidad espacial(i.e. Kernel Density Estimation) que posee la misma[9][10][11]. A pesar de que los resultados obtenidos por técnicas que basadas en la densidad espacial presentan resultados favorables, estas técnicas probabilísticas tienen deficiencias

tales como la dependencia del tamaño que se desea analizar, puesto que los resultados difieren en gran medida dependiendo del área que se analice[12]. Otro inconveniente es la falta de relación entre los conjuntos cercanos que se analizan bajo las técnicas probabilísticas, puesto que estas consideran cada grupo de eventos como independiente de sus pares cercanos, y la información que aportan las agrupaciones cercanas puede ser necesaria para el análisis de patrones criminales[10].

La información que entregan calles, edificios, parques y terrenos en abandono, pueden reflejar posibles comportamientos de la zona que se pretende clasificar. Conociendo el desarrollo y los alcances que produce incluir la forma urbana en la búsqueda de patrones de conducta criminal, surge la pregunta que se busca resolver con este trabajo: ¿Cómo se puede agrupar eventos criminales teniendo en cuenta la forma urbana por medio de técnicas basadas en la Inteligencia Computacional?

1.2. Solución propuesta

A partir del planteamiento del problema presentando en la sección anterior, se propone el diseño y la implementación de una técnica que permite determinar el numero de agrupaciones en que debe ser clasificado un conjunto de datos de crímenes basado en la estructura del área donde han ocurrido previamente estos sucesos.
Primariamente se seleccionan las técnicas de agrupamiento de datos que puedan ser aplicada al agrupamiento de eventos criminales.
De la misma manera, se seleccionan los métodos estadísticos en el análisis de datos espaciales que permitan conocer las similitudes entre las características que entregan la forma urbana y los eventos criminales de una zona específica.
Se deben incluir los estadísticos seleccionados en las técnicas seleccionadas para el agrupamiento de eventos criminales, de manera que el resultado que entreguen estos estadísticos sobre los datos agrupados definan el numero de agrupamientos correctos para el conjunto de datos.
La técnica obtenida se prueba sobre los eventos criminales referentes al Hurtos de Celulares a personas en la ciudad de Bogotá D.C., Colombia y los Hurtos a Residencias en la ciudad de San Francisco, California,

USA. Esto con el fin de determinar la respuesta de la técnica para diferentes escenarios de tiempo, lugar, crimen y de la misma manera comparar los rendimientos computacionales de las técnicas estadísticas y de agrupamiento que actúan sobre la técnica desarrollada.

1.3. Contenido del libro

Este libro condensa el planteamiento y desarrollo del presente trabajo de grado. En el siguiente capítulo se encuentra el marco de referencia, donde se encuentran de manera introductoria los temas que sirven como base para el desarrollo de este proyecto, de manera que funcionan como material introductorio para aquellos lectores que deseen obtener un primer acercamiento a los sistemas complejos, las teorías de análisis del crimen, las técnicas de agrupamiento difuso de datos, las técnicas de análisis de similitud de conjuntos de datos, entre otros temas abarcados en este trabajo.
El tercer capitulo se centra en el diseño de la técnica basada en la estructura urbana. En este capitulo se presenta tanto la descripción de los pasos presentados en la solución propuesta, como la justificación del desarrollo de cada etapa del trabajo.
El cuarto capitulo presenta los resultados obtenidos para los experimentos en los cuales se prueba la funcionalidad de la técnica desarrollada para los diferentes escenarios planteados en la solución propuesta.
Por ultimo, en el quinto capitulo se presentan las conclusiones obtenidas en el desarrollo del trabajo, los aportes obtenidos y el enfoque propuesto para las investigaciones que pueden nacer a partir de los resultados obtenidos en este trabajo. Al final del libro se encuentra la bibliografía que se ha referenciado durante todo el libro.

Capítulo 2

Marco de referencia

"Los grandes conocimientos engendran las grandes dudas."

Aristóteles

El objetivo de esta sección consiste en brindar un primer acercamiento a los conceptos relacionados y el problema planteado en secciones anteriores, por otro lado, se pretende entregar una posible primera respuesta a la pregunta planteada para el proyecto. Esta sección abarca temas tales como teorías de análisis del crimen, métodos estadísticos para el análisis de conjuntos de datos, métodos de agrupamiento de datos y métodos que permitan la validación de la cantidad de agrupaciones obtenidas.

2.1. Teorías de Análisis del Crimen

El crimen ha existido desde la creación de las civilizaciones, como dice [1], el crimen no se descubre en la naturaleza, es una invención humana, y esta característica de invención humana le ha permitido cambiar y expandirse a través de los tiempos junto a las poblaciones que coexisten con él.

Esta constante expansión del crimen en las sociedades es la que ha generado que el hombre busque estrategias que permitan en primera medida disminuir el crimen presente en las sociedades, en búsqueda de

en algún momento, erradicar el crimen. Pero esta expansión mencionada previamente hace que sea una tarea difícil analizar cada crimen como un evento independiente de su entorno, puesto que esta tarea no se lograría con la rapidez con la que el crimen sigue aumentando.

Esta razón hace que se busque analizar el crimen que ocurre en un lugar como una sucesión de hechos que ha permitido que ocurra dicho evento, es decir, un crimen ocurre porque existen tres factores principales, una víctima que desconoce su estado en el instante del evento, un atacante que conoce su estado, el estado de la víctima, y las condiciones del entorno, y un entorno donde no existe un organismo que actué como impedimento y resistencia para el atacante en el evento[13].

Además de lo mencionado previamente, la ocurrencia de un crimen se relaciona con los eventos previos ocurridos en su entorno, de manera que un crimen puede ocurrir con mayor probabilidad en lugares donde previamente han ocurrido eventos de esta índole. Esta búsqueda de un análisis del crimen visto desde un conjunto de eventos hace que los gobiernos se pongan en la tarea de recolectar los eventos ocurridos en sistemas informáticos(Bases de datos), esta recolección de eventos permite un análisis en conjunto del crimen para la búsqueda de rutinas y patrones presentes[14].

Como se ha mencionado previamente existen dos variables elementales en la ocurrencia de un crimen, el tiempo y el espacio donde ocurre el crimen, esto permite observar que el análisis del crimen se debe ser interpretado como un análisis espacio-temporal de una serie de eventos en conjunto.

El mapeo del crimen es un subconjunto del análisis del crimen que busca comprender la naturaleza geográfica en la que ocurre el crimen. Según [14], el mapeo del crimen busca cumplir con objetivos en del análisis del crimen tales como facilitar el análisis visual y estadístico de la naturaleza espacial del crimen y otro tipo de eventos, permitir a los analistas relacionar información dispersa basadas en variables comunes, proveer de mapas que ayuden a comunicar los resultados de los análisis.

El análisis espacial se define en [15] como el conjunto de técnicas donde sus resultados dependen de la localización de los objetos u eventos analizados, siendo necesario conocer la localización y los atributos de los objetos. Según [16], uno de los métodos enfocados en el mapeo del crimen es el análisis de Hotspot(áreas de alta concentración de eventos criminales), método usado en los Sistemas Geográficos de Información(GIS), donde la cantidad de eventos ocurridos se representan como puntos dentro de elipses de formas variantes, mediante esta técnica se pueden observar visualmente patrones del crimen para una zona de estudio. El aumento del tamaño de un Hotspot depende directamente del aumento de la zona de estudio, puesto que dependiendo del tamaño del área a analizar, el Hotspot puede ser desde un evento particular, hasta un país entero, siendo estos límites poco considerados en el análisis espacio-temporal.

Para el trabajo presente se prestará especial atención en el análisis de Hotspot, esto debido a que uno de los principales interrogantes que se pretende resolver es cuales son los factores que influyen en la forma que tiene una zona de interés(Hotspot) en el área de estudio.

2.2. Conjuntos Difusos

Lotfi A. Zadeh define los conjuntos difusos en [17] como un conjunto de elementos contenidos de manera parcial, es decir, que la propiedad de que un elemento x pertenezca al conjunto $A(x \in A)$ puede ser cierta con un grado parcial de verdad. Este grado de pertenencia es una proposición en el contexto de la lógica difusa, y no de la lógica usual binaria, que sólo admite dos valores: cierto o falso.

2.2.1. Lógica Difusa

Formulada por el ingeniero y matemático Lotfi A. Zadeh en [17], se basa en lo relativo de lo observado como posición diferencial. Este tipo de lógica toma dos valores aleatorios, pero contextualizados y referidos entre sí. De esta manera, por ejemplo, una persona que mida dos metros

es claramente una persona alta, si previamente se ha tomado el valor de persona baja y se ha establecido en un metro. Ambos valores están contextualizados a personas y referidos a una medida métrica lineal.

Esta lógica permite adaptarse de mejor manera a las situaciones y expresiones usadas cotidianamente en el mundo real. De manera que permite caracterizar y definir un valor cuantificable para expresiones desde aquellas que definen personas tales como "es una persona muy baja","ese sujeto muy flaco", hasta aquellas que definen situaciones especificas como "la velocidad de ese coche es muy rápida".

2.2.2. Grado de Pertenencia Difusa

Cada elemento que pertenece a un conjunto difuso lo hace de manera parcial, de modo que para cada elemento x que pertenezca al conjunto difuso A, debe existir un indicador que mide dicho grado de pertenencia al conjunto.[17] Para cada conjunto difuso, existe asociada una función de pertenencia para sus elementos, que indica en qué medida el elemento forma parte de ese conjunto difuso.
El grado de pertenencia de x a A, o probabilidad de pertenecer al conjunto, se mide con un número real $\mu_A(x)$ comprendido entre 0 y 1, aunque puede tener el valor de uno de estos dos.
De forma rigurosa, el valor correspondiente a cada elemento define una función indicatriz $\mu_A(x) : X \to [0, 1]$, donde X representa el conjunto universal del que el conjunto A toma sus elementos.
Si el valor de esta función es 0, x no pertenece a A. Si es 1, entonces $x \in A$ totalmente, y si $0 < \mu_A(x) < 1$ entonces x pertenece a A de una manera parcial [18].

2.2.3. Operaciones en Conjuntos Difusos

Tal como sucede en la teoría de conjuntos basados en la lógica clásica, en la teoría de conjuntos difusos se definen también las operaciones de unión, intersección, diferencia y complemento [18]. Las ecuaciones (2.1),(2.2) y (2.3) definen las operaciones mencionadas.

2.2.3.1. Unión

$$\mu_{A\cup B} = \max_{\mu_A(x),\mu_B(x)} \tag{2.1}$$

2.2.3.2. Intersección

$$\mu_{A\cap B} = \min_{\mu_A(x),\mu_B(x)} \tag{2.2}$$

2.2.3.3. Complemento

$$\mu_{\overline{A}} = 1 - \mu_A(x) \tag{2.3}$$

2.3. Métodos de Agrupamiento de Datos

El principal objetivo del análisis de agrupaciones es la clasificación de los objetos dependiendo de las similitudes que pueden existir entre ellos, y la organización de los datos en grupos o Clusters. Los métodos de agrupamiento son métodos no supervisados, puesto que no utilizan identificadores de clase anteriores. La principal ventaja de la agrupación es detectar la estructura subyacente de los datos, esto con el objetivo de clasificar datos, reconocer patrones, reducir modelos y optimizarlos, entre otros[19].
Los tipos de clasificaciones obtenidas se relacionan con el enfoque algorítmico de las técnicas de agrupamiento. Entre estas se encuentran el particionamiento, la jerarquía, los métodos de la teoría de grafos y los métodos basados en la función objetivo, entre otros[19].

2.3.1. Cluster

Dependiendo el objetivo de la agrupación de los datos se pueden entregar diferentes definiciones para Cluster, la que se acerca a los objetivos de esta investigación es la que plantea un Cluster como un grupo de

objetos que presentan una mayor similitud entre ellos que entre los miembros de otros grupos. Cuando hablamos de similitud, nos referimos estrictamente a la similitud matemática que pueda existir entre los objetos de los grupos[19].
En espacios métricos, la similitud se puede determinar por medio del concepto de distancia. Dicha distancia puede ser medida entre los datos del conjunto a agrupar, o entre los datos y un objetivo específico del grupo. Como se mencionó previamente, estos métodos son no supervisados, por esto, este objetivo específico generalmente no se conoce previamente, y es buscado por los métodos de agrupamiento de manera simultánea con la agrupación de los datos. En la Figura 2.1 se observa un agrupamiento aplicado para un conjunto de datos obtenidos aleatoriamente.

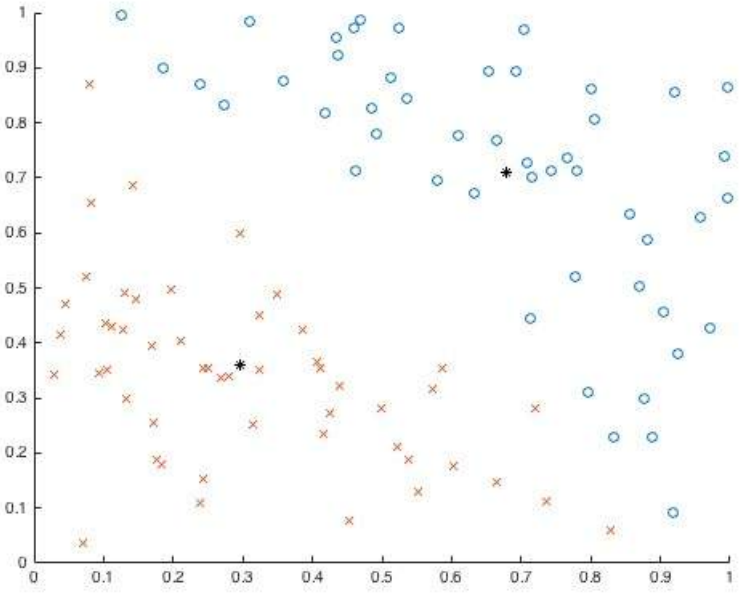


FIGURA 2.1: Agrupamiento obtenido para un conjunto aleatorio de datos.

2.3.2. Algoritmo Fuzzy C-Means Clustering(FCM)

El algoritmo Fuzzy C-Means(FCM) es la base de la mayoría de los métodos de agrupamiento difuso basados en la optimización de la función objetivo c-means, este algoritmo fue desarrollado Bezdek en [20] y [21].

2.3.2.1. Función Objetivo C-Means

Desarrollada por Bezdek en [22], la función objetivo, base en la mayoría de algoritmos de agrupamiento difuso es la siguiente:

$$J(Z;U,C) = \sum_{i=1}^{c}\sum_{k=1}^{N}(\mu_{ik})^m||Z_k - c_i||_B^2 \tag{2.4}$$

donde

$$Z = Z_1, Z_2, ...Z_N \tag{2.5}$$

son los datos que se pretende agrupar,

$$U = [\mu_{ik}] \in M_{fc}0 < M_{ik} < 1 \tag{2.6}$$

es la matriz de partición difusa de Z, encargada de asignar a cada valor de Z un valor de pertenencia para cada uno de los grupos,

$$C = [c_1, c_2, ..., c_c], c_i \in R^2 \tag{2.7}$$

es el vector de centros que se van a determinar con el algoritmo,

$$D_{ikB}^2 = ||Z_k - ci||_B^2 = (Z_k - ci)^T B(Z_k - c_i) \tag{2.8}$$

es una norma determinada por la elección de la matriz B, para el caso del algoritmo FCM, la norma es la distancia euclidiana, y

$$m \in (1, \infty) \tag{2.9}$$

determina la difusidad de las agrupaciones resultantes, de modo que mientras es más grande el valor de m, mas difusos son las agrupaciones resultantes. El valor de la función objetivo C-Means es el error cuadrático medio obtenido al representar las clases C de los centros c_i.

2.3.2.2. Algoritmo

El procedimiento consiste que para el conjunto de datos Z, se minimice iterativamente la función objetivo 2.4 hasta obtener una partición difusa óptima. Conociendo la condición para la creación de las agrupaciones, el procedimiento es el siguiente[23]:

1. **Determinar c y m.**

2. **Determinar el criterio de parada, entre los más utilizados se encuentra:**

   - **Un número máximo de iteraciones.**
   - **Una variación mínima para la matriz U.**

3. **Calcular los centros de cada agrupación mediante la siguiente ecuación:**

$$c_i = \frac{\sum_{k=1}^{N}(\mu_{ik})^m Z_k}{\sum_{k=1}^{N}(\mu_{ik})^m}; 1 \leq i \leq c \tag{2.10}$$

4. **Actualizar la matriz de partición difusa $U = [\mu_{ik}]$ mediante la siguiente ecuación:**

$$\mu_{ik} = (\sum_{j=1}^{c}(\frac{D_{ik}}{D_{jk}})^{\frac{2}{m-1}})^{-}1; 1 \leq k \leq n; 1 \leq i \leq c \tag{2.11}$$

5. **Si el criterio de parada ha sido cumplido, se finaliza el algoritmo, de lo contrario, regresar a 3.**

La figura 2.2 muestra los agrupamientos obtenidos para un conjunto de datos mediante el algoritmo de agrupamiento difuso Fuzzy C-Means para 8 grupos.

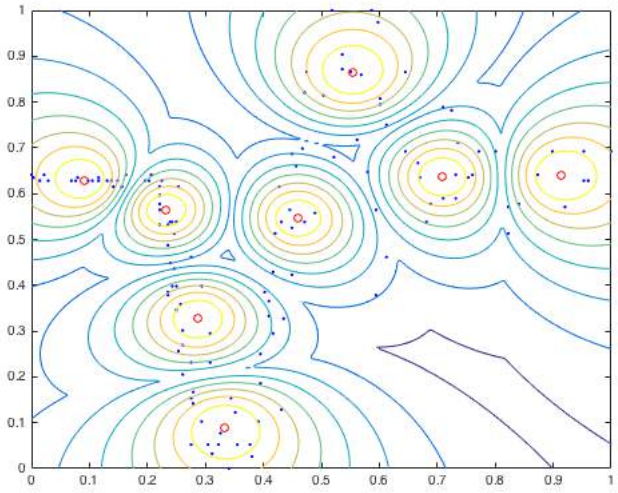


FIGURA 2.2: Algoritmo FCM aplicado para la distribución de 8 agrupaciones sobre una data generada aleatoriamente.

2.3.3. Algoritmo de Gustafson-Kessel(GK)

El algoritmo Gustafson-Kessel (GK), es una extensión del algoritmo Fuzzy C-Means, difiriendo de este en la elección de una norma diferente B_i para cada clase en lugar de la B constante en 2.8. Este algoritmo, aunque no tanto como el FCM, se suele usar en la obtención de modelos difusos, dado que las clases hiperelipsoidales que busca detectan de forma bastante correcta los comportamientos cuasi-lineales de los diversos regímenes de funcionamiento que pueden existir en un conjunto de datos[23].

El algoritmo GK extiende el algoritmo básico FCM eligiendo una norma diferente Bi para cada grupo, de esta manera 2.8 se convierte en [24]:

$$D^2_{ikB_i} = ||Z_k - ci||^2_{B_i} = (Z_k - c_i)^T B_i (Z_k - c_i) \tag{2.12}$$

Las matrices que definen las normas se consideran ahora variables que podrían optimizar la función 2.4. De esta manera, la norma para cada grupo será adaptada dependiendo de las características de cada matriz. Sea $B = B_1, B_2, \ldots, B_c$ el vector que contiene las c normas. La nueva función a minimizar será:

$$J(Z; U, C, B) = \sum_{i=1}^{c} \sum_{k=1}^{N} (\mu_{ik})^m D^2_{ikB_i} \tag{2.13}$$

cumpliendo con 2.6, 2.7 y 2.9. Para obtener una solución viable, se requiere que B_i sea limitada de alguna manera. La forma más habitual es fijar el determinante de B_i, lo que es equivalente a optimizar la forma de la clase manteniendo su volumen constante:

$$|B_i| = \rho_i, \rho > 0 \tag{2.14}$$

con ρ_i constante para cada clase. La expresión que se obtiene mediante multiplicadores de Lagrange es:

$$B_i = [\rho_i det F_i]^{\frac{1}{n}} F_i^{-} 1 \tag{2.15}$$

siendo F_i la matriz de covarianzas de la clase i definida por:

$$F_i = \frac{\sum_{k=1}^{N}(\mu_{ik})^m(Z_k - c_i)(Z_k - c_i)^T}{\sum_{k=1}^{N}(\mu_{ik})^m} \tag{2.16}$$

Una vez determinadas las clases hiper-elipsoidales, si lo que se desea es conseguir un modelo difuso de Takagi-Sugeno, es necesario ajustar dichas clases a estructuras lineales[25], siguiendo la información proporcionada por el eje mayor de dicha clase y mediante el uso de, por ejemplo, el algoritmo de mínimos cuadrados.

El algoritmo GK puede ser aplicado bajo los mismos parámetros del algoritmo Fuzzy C-Means mencionados en 2.3.2.2. La figura 2.3 muestra los agrupamientos obtenidos para un conjunto de datos mediante el algoritmo de agrupamiento difuso Gustafson-Kessel para 8 grupos.

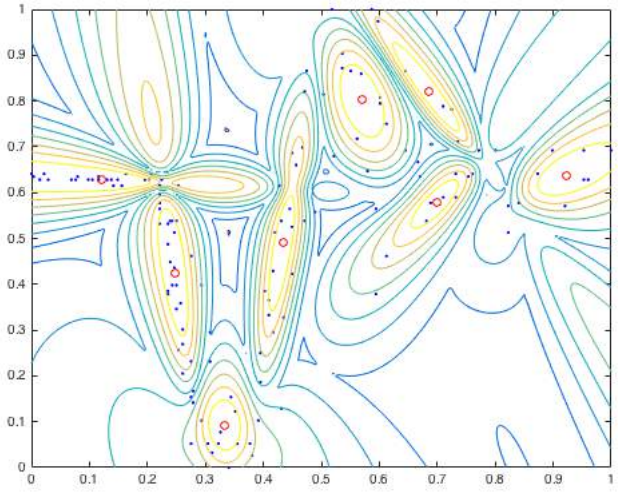


FIGURA 2.3: Algoritmo GK aplicado para la distribución de 8 agrupaciones sobre una data generada aleatoriamente.

Además de los algoritmos de agrupamiento difuso mencionados previamente, se destacan los algoritmos fuzzy c-varieties (FCV)[22], fuzzy c-regression models (FCRM)[26], fuzzy c-elliptotypes(FCE)[22] y el algoritmo adaptive fuzzy c-regression models (AFCR)[27].

2.4. Validación de Clúster

El proceso de determinar el número de agrupaciones óptimo se llama Validación de Clúster ó Análisis de Agrupamientos. Dicho de otra manera, el número de agrupaciones que mejor representan la clasificación de un conjunto de datos puede ser determinado mediante ciertos parámetros aplicados sobre los resultados obtenidos en la agrupación de estos mismos previamente.
La literatura entrega varias medidas propuestas para validar la semejanza entre dos grupos obtenidos por un algoritmo de agrupamiento. Dichas medidas se usan frecuentemente para comparar los resultados que entregan diferentes algoritmos agrupando un conjunto de datos.
Cabe aclarar que no existe una medida universal para el análisis de los agrupamientos obtenidos por los algoritmos, en cambio se podría considerar que cada una de estas medidas puede dar un aporte al análisis dependiendo del objetivo buscado y tanto de las condiciones, como del algoritmo utilizado.

2.4.1. Evaluación interna

Cuándo el resultado de un agrupamiento es evaluado basado en los datos agrupados por él, esta recibe el nombre de evaluación interna. Estos métodos asignan la mayor puntuación al algoritmo que produce grupos con alta similitud entre sus componentes y baja similitud entre los grupos.[28]
Por tanto, las medidas de evaluación internas son más convenidas para conseguir alguna idea en situaciones donde un algoritmo actúa mejor que otro, pero esto no implicará que un algoritmo produce resultados más válidos que otro.[29] La validez medida por tal índice depende de la clase de estructura existente en el conjunto de datos. Un algoritmo diseñado para alguna clase de modelos no tiene ninguna posibilidad si el conjunto de dato contiene un conjunto radicalmente diferente de modelos, o si la evaluación mide un criterio radicalmente diferente.[29]
A continuación se muestran algunos de los métodos basados en criterio interno para la evaluación de agrupamientos de datos.

2.4.1.1. Índice de Davies–Bouldin

Desarrollado en [30], puede ser calculado por la siguiente ecuación:

$$DB = \frac{1}{n}\sum_{i=1}^{n} \max_{j \neq i} \left(\frac{\sigma_i + \sigma_j}{d(c_i, c_j)} \right) \tag{2.17}$$

donde n es el número de grupos, c_x es el centroide de grupo x, σ_x es la distancia media de todos los elementos en grupo al centroide x, y $d(c_i, c_j)$ es la distancia entre centroides c_i y c_j.c_x Desde algoritmos que producen grupos con bajas distancias intra-grupo(alta semejanza intra-grupo) y alta distancia inter-grupo (baja semejanza inter-grupo) tendrá un bajo índice de Davies–Bouldin, el algoritmo de agrupamiento que produce la colección de grupos con menores índices de Davies–Bouldin está considerado el algoritmo mejor basado en este criterio.

2.4.1.2. Índice de Dunn

Desarrollado en [31], el índice de Dunn tiene como objetivo identificar grupos con un alta separación entre ellos. Está definido como la proporción entre la mínima distancia inter-grupo y la máxima distancia intra-grupo. Para cada partición de grupo, el índice de Dunn se obtiene apartir de la siguiente ecuación:

$$D = \frac{\min_{1 \leq i < j \leq n} d(i,j)}{\max_{1 \leq k \leq n} d'(k)} \tag{2.18}$$

donde $d(i,j)$ representa la distancia entre grupos i y j, y $d'(k)$ mide la distancia intra-grupo del grupo k. La distancia inter-grupo $d(i,j)$ entre dos grupos pueden ser cualquier número de medidas de distancia, como la distancia entre los centroides de los grupos. De modo parecido, la distancia intra-grupo $d'(k)$ puede ser medida de diferentes formas, como la distancia máxima entre cualquier par de elementos en el grupo k.

2.4.2. Evaluación externa

Para la evaluación externa, los resultados de un agrupamiento son evaluados basados en datos que no fueron usados para el agrupamiento, dichos datos pueden ser etiquetas de clases conocidas o marcadores

externos. Estas etiquetas constan de un conjunto de pre-elementos clasificados, y estos conjuntos son a menudo creados por expertos. Estos tipos de métodos de evaluación miden qué cercano está el agrupamiento del conjunto a las clases predeterminadas. [32] Además, de un punto de vista de descubrimiento del conocimiento, la reproducción de conocimiento sabido puede no necesariamente ser el resultado esperado.33 En el escenario especial de Agrupamiento con restricciones, donde meta información (como etiquetas de clase) está utilizada ya en el proceso de agrupamiento, el control de la información para propósitos de evaluación es no trivial.[33]
Algunas de las medidas de análisis de algoritmos de agrupamientos basadas en el criterio externo se muestran a continuación:

2.4.2.1. Índice de Rand

Desarrollado en [34], el índice de Rand analiza la similitud presente entre los grupos resultantes de la aplicación de los algoritmos de agrupamiento y las clasificaciones de los marcadores previamente establecidos. Visto desde otro punto, la medida Rand se expresa como el porcentaje de decisiones correctas hechas por el algoritmo. Puede ser computado utilizando la fórmula siguiente:

$$RI = \frac{a+d}{a+b+c+d} \tag{2.19}$$

Dónde a es el número de ciertos positivos, d es el número de ciertos negativos, b es el número de falsos positivos, y c es el número de falsos negativos. Una característica del indice de Rand consiste en que los falsos positivos y los falsos negativos tienen el mismo peso.

2.4.2.2. Índice Jaccard

Desarrollado en [35], el índice de Jaccard suele cuantificar la semejanza entre dos conjuntos de datos. Tomando un valor entre 0 y 1, donde un índice de 1 significa que los dos conjuntos son idénticos, y un índice de 0 indica que los conjuntos no tienen elementos comunes, el índice de

Jaccard está definido por la siguiente ecuación:

$$J(A,B) = \frac{a}{a+b+c} \tag{2.20}$$

Esto es sencillamente el número de los elementos únicos comunes a ambos conjuntos divididos por el número total de elementos únicos en ambos conjuntos.[36]

2.4.2.3. Índice de Fowlkes–Mallows

Desarrollado en [37], el índice de Fowlkes-Mallows computa la similitud entre los agrupamientos obtenidos por el algoritmo correspondiente y las clasificaciones de los marcadores. Mientras el valor del índice sea mayor, los grupos serán mas similares a las clasificaciones de los marcadores. El índice de Fowlkes–Mallows se obtiene mediante la siguiente ecuación:

$$FM = \sqrt{\frac{a}{a+b} \cdot \frac{a}{a+c}} \tag{2.21}$$

Dónde a es el número de ciertos positivos, b es el número de falsos positivos, y c es el número de falsos negativos.

2.5. Índices de Validación de Agrupamiento Difuso

Adicionalmente a los índices presentados en 2.4, existen índices propios para la validación de agrupamientos obtenidos mediante las técnicas de agrupamiento difusas.

2.5.1. Coeficiente de Partición(PC)

El Coeficiente de Partición mide la cantidad de "solapamiento" entre cada agrupación obtenida en el clustering. Fue definido por Bezdek en [22] de la siguiente manera:

$$PC_{(c)} = \frac{1}{N} \sum_{i=1}^{c} \sum_{j=1}^{N} (\mu_{ij})^2 \tag{2.22}$$

donde μ_{ij} es el número de miembros de datos del punto j en el grupo i. La desventaja del Coeficiente de Partición se encuentra en la falta

de conexión directa a alguna propiedad que puedan tener los propios datos. Según la teoría propuesta para el PC, el número óptimo de agrupamientos está en el valor máximo de PC, que para un agrupamiento apropiado estará cerca de 1.

2.5.2. Entropía de Clasificación(CE)

La Entropía de Clasificación mide la difusidad de la partición de cada agrupación individualmente, este método es similar al Coeficiente de Partición. Definido por Bezdek en [20] de la siguiente manera:

$$CE_{(c)} = -\frac{1}{N}\sum_{i=1}^{c}\sum_{j=1}^{N}\mu_{ij}log(\mu_{ij})^2 \tag{2.23}$$

donde μ_{ij} es el número de miembros de datos del punto j en el grupo i. Para la teoría propuesta para el CE, el número óptimo de agrupamientos está en el valor mínimo de CE, que para un agrupamiento apropiado estará cerca de 0. Además de los índices de validación de agrupamiento difuso mencionados previamente, se destacan el Índice de Partición(SC)[38], el Índice de Separación(S)[38] y el Índice de Xie y Benis(XB)[38].

2.6. Resumen

En este capítulo se presentaron los conceptos referentes a teorías de análisis del crimen, conjuntos difusos, métodos de agrupamiento de datos, haciendo énfasis en los métodos de agrupamiento difuso como el algoritmo Fuzzy C-Mean y el algoritmo Gustafson-Kessel, por último, los Índices de Validación de Clústers junto con algunos Métodos propios para la validación de Clústers obtenidos por métodos de agrupamiento difuso.

Capítulo 3

Diseño de una Técnica de Agrupamiento Difuso de Eventos Criminales Basado en la Estructura Urbana

"La naturaleza puede ser corregida, enmendada, pues de no ser así quedaríamos sepultados bajo los prejuicios. Sin eso no habría ni un solo gran hombre."

Crimen y Castigo - Fiódor Dostoyevski

El capítulo presente contiene el diseño conceptual de cada una de las fases que componen la solución propuesta en 1.2.

3.1. Solución Propuesta

La figura 3.1 muestra un diagrama con la solución planteada al problema presentado en 1.1.
Como se puede observar, los datos de Hurtos y Características de la cuidad, llamados Malla Vial a partir de este momento, son ingresados a los algoritmos de agrupamiento difuso 2.3.

Los resultados obtenidos por la ejecución de estos algoritmos son procesados por las técnicas de validación de Clúster, finalmente, estas técnicas indicaran la cantidad de agrupamientos donde los dos conjuntos de datos procesados han sido distribuidos de la manera mas adecuada.

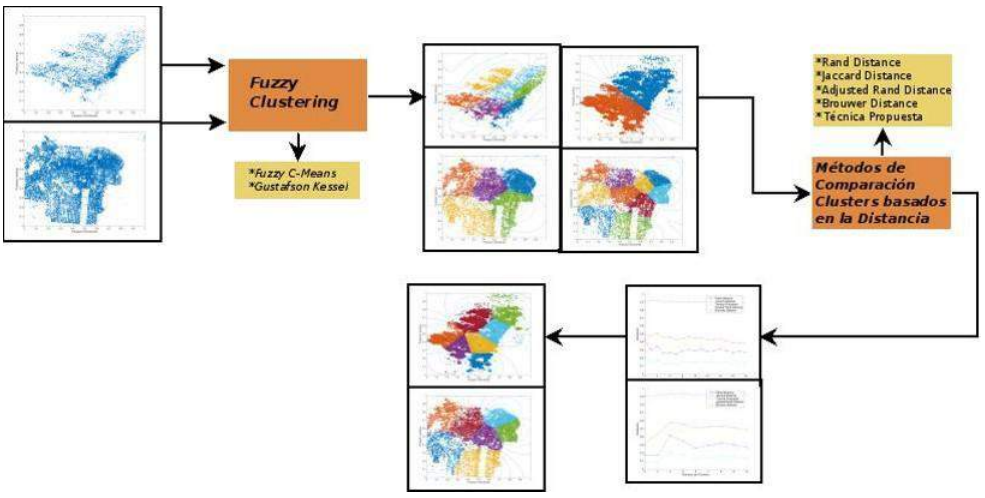


FIGURA 3.1: Diagrama de la Solución Propuesta.

3.2. Selección de Los Conjuntos de Datos

El primer paso consiste en la selección de los conjuntos de datos que serán procesados por la Técnica de agrupamiento propuesta. Para esto es necesario disponer de dos conjuntos de datos diferentes, el primero de estos es la intersección de puntos de las calles y carreras que componen el área de estudio, por otra parte, el segundo conjunto de datos consiste en la ubicación espacial de los eventos criminales ocurridos en el área de estudio para la técnica desarrollada.

Los conjuntos de datos de la intersección de calles y carreras son en su mayoría suministrados por los mecanismos gubernamentales encargados de la organización geográfica de las áreas de estudio, dos ejemplos

de estos suministros se encuentran en [39] y [40], estos mecanismos gubernamentales disponen los conjuntos de datos a la comunidad para su análisis y conocimiento.
Por otro lado, los conjuntos de datos correspondientes a los eventos criminales pueden tanto estar disponibles al publico, como estar protegidos por los organismos de control encargados de su recolección. Para el caso de la ciudad de San Francisco,CA, los datos correspondientes a eventos criminales utilizados en este trabajo se obtienen de [39]. Por otra parte, los datos utilizados correspondientes a eventos criminales ocurridos en la ciudad de Bogotá D.C., fueron suministrados por la Fundación Ideas para la Paz(F.I.P.). Una de las áreas programáticas de esta fundación es la de seguridad y política criminal, esta área promueve el estudio pormenorizado de los fenómenos delictivos y apoya la armonización de herramientas conceptuales, normativas y organizacionales que articulen planes y estrategias diferenciales en materia de seguridad y justicia.[41] [1]

3.2.1. Proyección de Mercator

Al obtener nuestros conjuntos de datos de Malla Vial y Eventos Criminales es necesario conocer el sistema geográfico en el cual estos se encuentran.
La mayoría de datos disponibles se encuentran en el sistema de coordenadas geográficas, este sistema se caracteriza por utilizar la latitud y longitud como indicativos del la posición de un lugar especifico. Pero debido a que este sistema tiene en cuenta la superficie elipsoide de la tierra, es necesario convertir los datos hacia un sistema geográfico que represente la proyección cartográfica de los datos en un espacio $R2$, de esta manera se usa la proyección de Mercator, esta proyección desarrollada por Gerardus Mercator en 1569 es utilizada por aplicaciones web de cartografía como Google Maps, Bing Maps, OpenStreetMap, entre otros. La definición propuesta por Mercator para un punto geográfico a partiendo de la longitud λ y la latitud ϕ es la siguiente[42]: Siendo λ_0

[1] La Fundación Ideas para la Paz, creada en 1999 por un grupo de empresarios Colombianos, busca aportar a la comprensión de las conflictividades de nuestro país, ejerciendo el liderazgo necesario para movilizar a ciudadanos, servidores públicos y empresarios alrededor del objetivo común de construir una sociedad en paz `http://www.ideaspaz.org`

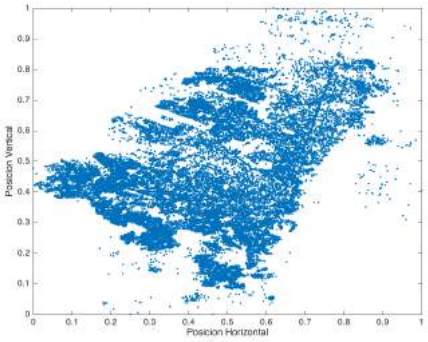


FIGURA 3.2: Conjunto de Datos de la Malla Vial en la ciudad de Bogotá D.C., Colombia.

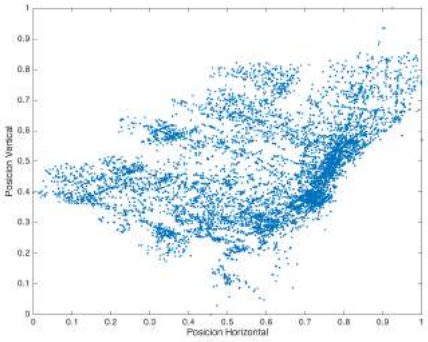


FIGURA 3.3: Conjunto de Datos de los Hurtos a Celulares en la ciudad de Bogotá D.C., Colombia.

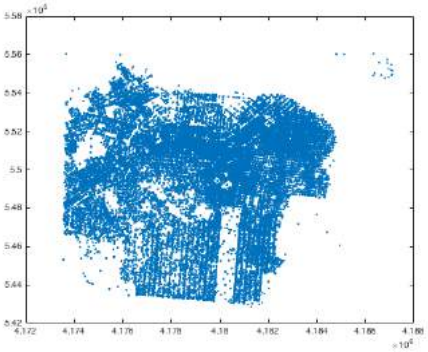


FIGURA 3.4: Conjunto de Datos de la Malla Vial en la ciudad de San Francisco,CA,USA.

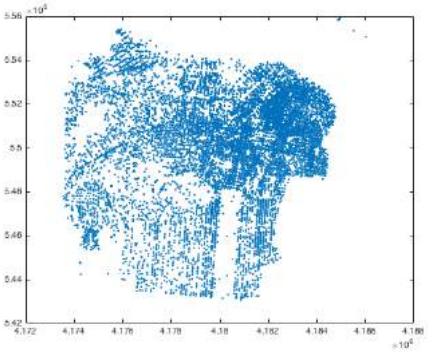


FIGURA 3.5: Conjunto de Datos de los Hurtos a Residencias en la ciudad de San Francisco,CA,USA.

la longitud central del mapa:

$$x = \lambda - \lambda_0 \tag{3.1}$$

$$y = \ln(\tan\phi + \sec\phi) \tag{3.2}$$

Luego del procesamiento, los datos se encuentran en el mismo sistema geográfico y se puede proceder a procesar los mismos en los algoritmos de agrupamiento difuso. Las figuras 3.2,3.3,3.4 y 3.5 representan los conjuntos de datos de Malla Vial y Eventos criminales en las ciudades de Bogotá D.C.,CO. y San Francisco,CA,USA. Mientras que las figuras 3.6 Y 3.7 representan los datos de los eventos criminales sobre la Malla vial.

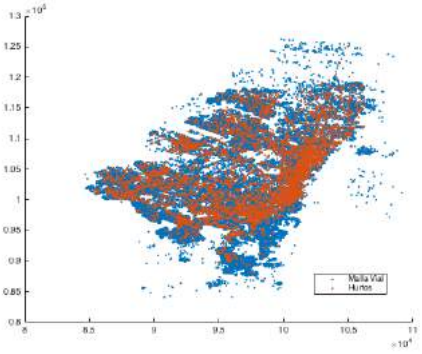


FIGURA 3.6: Conjunto de Datos de la Malla Vial y los Hurtos en la ciudad de Bogotá D.C.,Colombia

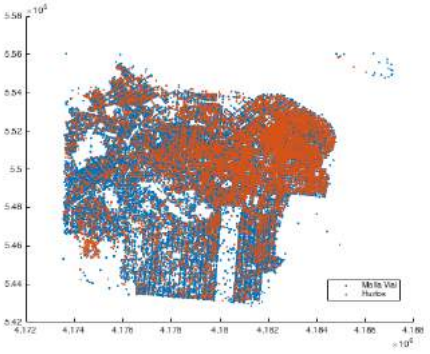


FIGURA 3.7: Conjunto de Datos de la Malla Vial y los Hurtos en la ciudad de San Francisco,CA,USA.

3.3. Algoritmos de Agrupamiento Difuso

Los algoritmos de agrupamiento difuso que utiliza la técnica son los mencionados en 2.3.2 y 2.3.3, como se menciona en dichas secciones, estos algoritmos requieren de unos parámetros de iniciales en su funcionamiento.

Entre estos parámetros destaca el numero de agrupaciones en las que se desea separar el conjunto de datos. Este parámetro varia desde 2 agrupaciones hasta las seleccionadas por quien ejecuta la técnica para una especifica. De modo que dependiendo del primer análisis que haga el usuario que va a ejecutar la técnica sobre los conjuntos de datos, definirá la cantidad de ocasiones que se ejecute el algoritmo.

Para la técnica desarrollada se eligen un total de 100 ejecuciones del algoritmo desde 2 agrupamientos hasta la cantidad N seleccionada por el usuario de manera que el algoritmo es ejecutado en un total de $(N-1)*100$ veces. De estas 100 ejecuciones generales se obtiene la mejor ejecución, para esta decisión se toma el conjunto de ejecuciones en los cuales las el promedio de ejecuciones desde 2 hasta N llego a la función objetivo del algoritmo en el menor tiempo.

Posterior a esta ejecución para cada conjunto de datos, se obtienen las agrupaciones y las funciones de pertenencia para cada uno de los datos en cada una de las ejecuciones. Con estos datos obtenidos se procede a validar la mejor cantidad de agrupaciones basada en la forma urbana.

3.4. Métodos de Validación de Clúster

La ejecución mencionada en la sección anterior genera como resultado las funciones de pertenencia para cada uno de los datos de los conjuntos de Malla Vial y de Eventos Criminales en cada una de las ejecuciones que ha decidido el usuario de la técnica.
El paso siguiente consiste en el procesamiento de estos resultados en un método de validación que permita comparar en el mismo instante los resultados del agrupamiento en un numero de grupos especifico, de manera que se comparan los resultados del agrupamiento en 2 grupos para los dos conjuntos, posteriormente para 3 grupos y de esta manera sucesivamente hasta llegar a comparar todos los grupos que ha elegido el usuario de la técnica.
Se elige utilizar el índice Rand (sección 2.4.2.1), debido a que permite analizar la similitud entre los resultados de el algoritmo en diferentes ejecuciones, pero para este caso es necesario tener dos factores en cuenta, el primer de estos consiste en que el índice Rand no se encuentra inicialmente planteado para agrupamientos difusos, mientras que por otra parte, tampoco contempla que los conjuntos de datos sean diferentes. Estas dos inquietudes se resuelven mediante la expansiones presentadas a continuación.

3.4.1. Índice Rand para Agrupamientos Difusos

A pesar de que la literatura ofrece diferentes maneras de expandir el índice Rand hacia los agrupamientos difusos, se decide utilizar el propuesto en [43], esto debido a su sencilla implementación y bajo costo computacional.
El autor propone la expansión de la siguiente manera:
Dada una partición difusa $P = (P_1, P_2, P_3, ..., P_n)$ para un conjunto de datos X, cada dato de X puede ser caracterizado por su vector de pertenencias $P(x) = (P_{1(x)}, P_{2(x)}, P_{3(x)}, ..., P_{n(x)})$, donde $P_{i(x)}$ representa la pertenencia de x para la agrupación P_i.
De este modo, para definir una relación de equivalencia difusa para X en términos de su pertenencia se utiliza el siguiente concepto:

$$E_P(X, X') = 1 - ||P(X) - P(X')|| \tag{3.3}$$

Pero dado que son dos particiones las que se pretenden comparar, es necesario definir la concordancia entre las dos particiones de la siguiente manera:

$$Concordancia(X, X') = 1 - |E_P(X, X') - E_Q(X, X')| \tag{3.4}$$

De manera análoga, se define la discordancia entre las dos particiones como:

$$Discordancia(X, X') = |E_P(X, X') - E_Q(X, X')| \tag{3.5}$$

Por ultimo, al definir la concordancia y la discordancia que de las particiones obtenidas, se define la medida de distancia difusa como la suma normalizada de las discordancias:

$$d(P, Q) = \frac{\sum_{X,X'} |E_P(X, X') - E_Q(X, X')|}{n(n-1)/2} \tag{3.6}$$

Basados en los conceptos de concordancia y relación de equivalencia difusa se definen nuevamente los índices a,b,c,d utilizados en Índice Rand:

$$a = (1 - |E_P(X, X') - E_Q(X, X')| * (|E_P(X, X') * E_Q(X, X')|) \tag{3.7}$$

$$b = max(E_P(X, X') - E_Q(X, X'), 0) \tag{3.8}$$

$$c = max(E_Q(X, X') - E_P(X, X'), 0) \tag{3.9}$$

$$d = (1 - |E_P(X, X') - E_Q(X, X')| * (1 - (E_P(X, X') * E_Q(X, X'))) \tag{3.10}$$

Esta nueva definición de los índices a,b,c,d permite aplicar índices propuestos en la literatura como otras funciones objetivo que se ejecutan en la técnica. Entre las técnicas desarrolladas en la literatura, se adaptan las siguientes:

- **Índice Jaccard [35]**

$$JI = \frac{a}{a + b + c} \tag{3.11}$$

- **Índice Browuer [44]**

$$BI = \frac{2(ad - bc)}{c^2 + b^2 + 2ad + (a + d)(b + c)} \tag{3.12}$$

- **Índice Rand Ajustado [34]**

$$ARI = \frac{a}{\frac{1}{2} * (b + c)} \tag{3.13}$$

Por ultimo se propone tomar los valores de los índices a,b,c y d obtenidos para las agrupaciones difusas y aplicar nuevamente la definición general del Índice Rand, esto con el objetivo de obtener un índice adicional que pueda llegar a ser una función objetivo de menor valor:

$$ÍndicePropuestoRI = \frac{a + d}{a + b + c + d} \tag{3.14}$$

3.4.2. Índice Rand para diferentes conjuntos de datos

El segundo problema que presenta el Índice Rand para este caso de estudio consiste en que los conjuntos de datos que se comparan contienen una cantidad diferente de datos, esto debido a que el conjunto de datos de la Malla Vial siempre sera mayor al conjunto de datos de Eventos Criminales.
Para este caso se utiliza la propuesta desarrollada por el profesor Miguel Melgarejo [45], la cual consiste en realizar una expansión difusa a los dos conjuntos de datos sobre un universo que los contenga a los dos, de manera que al pertenecer los dos conjuntos de datos al mismo universo, sera posible aplicar los índices mencionados previamente.
El ultimo paso consiste en buscar el numero de grupos para los cuales la función objetivo ha entregado el valor mínimo, de esta manera se conoce cual es el numero de grupos donde los agrupamientos obtenidos tienen mayor similitud entre los dos conjuntos de datos.

3.5. Algoritmo Propuesto

Ahora se presenta un resumen de los pasos aplicados en el método desarrollado.

1. **El primer paso consiste en comparar los conjuntos de datos que se pretenden procesar en los algoritmos de agrupamiento difuso, si**

estos se encuentran en el sistema geográfico tradicional, es necesario realizar una proyección de Mercator sobre los mismos para que estos queden proyectados sobre el espacio $R2$.

2. **Al obtener los datos se debe elegir la cantidad de agrupamientos sobre los que se pretende comparar, el valor mínimo sera siempre 2 mientras que el valor máximo lo decide el usuario que ejecuta la técnica.**

3. **El siguiente paso consiste en procesar los algoritmos de agrupamiento difuso la cantidad de veces que se ha seleccionado en el paso anterior, posteriormente se debe elegir el grupo que ha entregado de manera mas rápida y acertada los agrupamientos.**

4. **Al obtener las funciones de pertenencia de los algoritmos de agrupamiento difuso se procede a procesar estos resultados en los índices de validación de Clúster, esto teniendo en cuenta la expansión del índice Rand para agrupamientos difusos y la expansión difusa para conjuntos de diferente tamaño.**

5. **Por ultimo, se comparan los valores obtenidos por las funciones objetivo y el numero de grupos donde se obtenga el valor mínimo de la función objetivo, sera el numero de agrupaciones donde se presenta la mayor similitud entre los agrupamientos de Malla Vial y Eventos Criminales.**

3.6. Resumen

En este capitulo se presento el diagrama con la solución propuesta para el problema planteado en la sección 1.1, junto con el diseño conceptual de la técnica propuesta. Se mostró la descripción y las razones de la elección de cada unos de pasos que tiene la técnica desarrollada y por ultimo se mostró un resumen del funcionamiento de la técnica.

Capítulo 4

Resultados

"Si la libertad significa algo, será, sobre todo, el derecho a decirle a la gente aquello que no quiere oír."

Rebelión en la granja - George Orwell

En el siguiente capitulo se presenta la descripción detallada de las pruebas realizadas sobre la técnica presentada en la sección anterior y de los resultados obtenidos por las mismas. Se presentan dos casos de estudio independientes los cuales son el Hurtos a Celulares en la Ciudad de Bogotá D.C., Colombia y los Hurtos Residencias en la ciudad de San Francisco, California, Estados Unidos de América, cada uno de estos casos ha sido ejecutado y evaluado de manera individual por la tecnica propuesta, utilizando los algoritmos de agrupamiento difuso Fuzzy C-Means(FCM) y Gustafson Kessel(GK).

4.1. Conjuntos de Datos

Cada uno de los casos de estudio tiene dos conjuntos de datos diferentes que corresponden a la Malla Vial y a los Eventos Criminales ocurridos en una ventana de tiempo especifica.
Para el caso de Bogotá D.C., los datos de Malla Vial se encuentran disponibles en [40], mientras que el conjunto de datos de Hurtos a Celulares es suministrado por la Fundación Ideas para la Paz(F.I.P.), fundación encargada de analizar las dinámicas del conflicto armado y urbano en

Colombia.
La cantidad de datos suministrados para la Malla Vial corresponde a 28489, mientras que para los Hurtos a Celulares se tienen 9390 datos, estos datos corresponden a un total de 1440 Días, correspondidos en el periodo de tiempo desde Enero de 2011, hasta Mayo de 2015.
Para el caso de San Francisco,CA, tanto los datos de Malla Vial como los de Hurtos a Residencias son obtenidos de [39].
La cantidad de datos suministrados para la Malla Vial corresponde a 66323, mientras que para los Hurtos a Residencias se tienen 15386 datos, estos datos, al igual que los del caso de Bogotá D.C., corresponden a un total de 1440 Días, correspondidos en el periodo de tiempo desde Enero de 2011, hasta Mayo de 2015.
La tabla 4.1 muestra un resumen de los conjuntos de datos utilizados en la ejecución de la técnica.

Datos Utilizados	**Bogotá,D.C.,CO**	**San Francisco,CA,USA**
Malla Vial	28489	66323
Hurtos	9390	15386
Tipificación	Hurtos a Personas con objetivo de Celulares	Hurtos a Residencias
Periodo de Tiempo	1440 Días(Enero 2011-Mayo 2015)	1440 Días (Enero 2011-Mayo 2015)

CUADRO 4.1: Conjuntos de datos utilizados en la Técnica propuesta

4.2. Elección del numero de grupos

La elección del limite superior de grupos en los cuales se agruparan los conjuntos de datos depende del primer análisis que ejecuta el usuario sobre los datos, de manera que dependerá completamente de él la cantidad de grupos y de esta manera la cantidad de ejecuciones del algoritmo.
Para el caso de Bogotá D.C., se tomó inicialmente las 20 localidades en las cuales esta distribuida el área de estudio, posteriormente de estas 20 se elimina la localidad de Sumapaz debido a que esta área no aporta ningún evento criminal en el periodo de tiempo analizado. Por otro lado, la localidad de la Candelaria al estar incluida dentro de la localidad

Santa Fe es descartada de igual manera.
Para el caso de la ciudad de San Francisco,CA., se eligen 10 grupos como limite superior debido a que el área de estudio se encuentra dividida en 10 distritos policiales. La tabla 4.2 muestra un resumen de la cantidad de grupos elegidos para cada caso y una pequeña justificación para dicha elección.

Datos Utilizados	**Bogotá D.C.,CO**	**San Francisco,CA,USA**
Limite de Agrupamientos Superior e Inferior	2-18	2-10
Justificación de la decisión	Se toman las 20 localidades que tiene la ciudad y se quitan Sumapaz por no aportar datos en el periodo de tiempo y La Candelaria por estar incluida dentro de la localidad Santa Fe.	Se toman los 10 distritos policiales en los que está dividida la ciudad.

Cuadro 4.2: Criterios de elección para el Limite Superior en la ejecución de la técnica

La tabla 4.3 muestra los algoritmos empleados por la tecnica junto con su desarrollador y el Entorno de Desarrollo Integrado(IDE) en el cual ha sido ejecutado.

Algoritmos Utilizados	**Desarrollador**	**Entorno de Desarrollo Integrado(IDE)**
Fuzzy C-Means Clustering	Balazs Balasko, Janos Abonyi, Balazs Feil	MatLab
Gustafson Kessel Clustering	Balazs Balasko, Janos Abonyi, Balazs Feil	MatLab
Algoritmo de Ejecución de los agrupamientos	Luigi S. Fajardo T.	MatLab
Algoritmo Fuzzy Rand Index Expandido para múltiples conjuntos de datos.[45]	Ing. Miguel A. Melgarejo R.	MatLab
Algoritmo Fuzzy Adjusted Rand Index, Fuzzy Brouwer, Técnica Propuesta.	Luigi S. Fajardo T.	MatLab

Cuadro 4.3: Algoritmos ejecutados

La tabla 4.4 presenta la cantidad de ejecuciones para cada uno de los algoritmos utilizados en la técnica desarrollada y en caso de ser necesario, el criterio de elección para la ejecución de algunos de estos algoritmos.

Algoritmos Utilizados	**Cantidad de Procesamientos**	**Criterio de Elección**
Algoritmo de Ejecución de los agrupamientos	100 Para cada conjunto de datos, 200 por ciudad	Se elige el grupo que entrega la mínima función objetivo en el menor tiempo.
Algoritmo Fuzzy Rand Index Expandido para múltiples conjuntos de datos.	1 para cada técnica de agrupamiento, 2 por ciudad	-
Algoritmo Fuzzy Adjusted Rand Index, Fuzzy Brouwer, Técnica Propuesta.	1 para cada técnica de agrupamiento, 2 por ciudad	-

CUADRO 4.4: Cantidad de ejecuciones para los algoritmos utilizados en la técnica desarrollada

4.3. Resultados

Luego de la ejecución de la técnica para las áreas de estudio seleccionadas se tienen los siguientes resultados:

4.3.1. Rendimiento de los algoritmos de agrupamiento

En la tabla 4.5 se observa el promedio de las ejecuciones realizadas para cada uno de los algoritmos de agrupamiento difuso en cada una de las áreas de estudio.

Algoritmo De Agrupamiento Difuso	**Bogotá, CO**	**San Francisco, CA,USA**
Fuzzy C-Means	85 s/Ejecución	55 s/Ejecución
Gustafson Kessel	90 s/Ejecución	65 s/Ejecución

CUADRO 4.5: Tiempo promedio de ejecución para cada algoritmo de agrupamiento difuso.

Como se puede observar, la elección de 18 grupos como limite superior para las ejecuciones en el área de Bogotá D.C.,CO., generan un

mayor tiempo en promedio que las ejecuciones en el área de San Francisco,CA,USA., donde se utilizaron 10 grupos como limite superior. Teniendo en cuenta que la cantidad de datos en el conjunto de datos del área de San Francisco,CA,USA. representa el doble de los datos en el conjunto de datos del área de Bogotá D.C.,CO., se demuestra que tiene un mayor peso el limite de grupos elegidos que la cantidad de datos procesados. Para observar el resultado de cada una de las ejecución, remitirse a la sección de Anexos.
Por otro lado, la tabla 4.6 muestra la ejecución que entrego el mejor resultado, y por lo tanto, la que entrego el conjunto de funciones de pertenencia que fueron procesadas por los índices de validación de Clúster.

Algoritmo De Agrupamiento Difuso	Bogotá, CO	San Francisco, CA
Fuzzy C-Means	34	16
Gustafson Kessel	73	82

CUADRO 4.6: Número de ejecución elegida para el para el procesamiento posterior por sus resultados en el agrupamiento difuso

4.3.2. Rendimiento de los algoritmos Índice de Validación de Clúster

La tabla 4.7 muestra los resultados del rendimiento para cada uno de los índices de validación de clúster ejecutados en la técnica.

Índice de Validación de Clúster	Bogotá D.C.,CO	San Francisco,CA,USA
Rand Distance	22.77 s/Ejecución	15.82 s/Ejecución
Jaccard Distance	24.65 s/Ejecución	16.75 s/Ejecución
Ajusted Rand Distance	20.21 s/Ejecución	15.55 s/Ejecución
Brouwer Distance	25.38 s/Ejecución	16.84 s/Ejecución
Técnica Propuesta	19.89 s/Ejecución	14.38 s/Ejecución

CUADRO 4.7: Tiempo promedio de ejecución para cada algoritmo índice de validación de Clúster.

Nuevamente se observa que la ejecución de los índices de validación de clúster tarda un tiempo mayor en su aplicación sobre el área de Bogotá que sobre San Francisco, de modo que el limite superior sigue siendo un factor determinante en el costo computacional que pueda tener la técnica desarrollada.

4.3.3. Índices de Validación de Clúster

4.3.3.1. Bogotá D.C.,CO

La tabla 4.8 contiene los resultados obtenidos por los índices de validación para el algoritmo Fuzzy C-Means en la ciudad de Bogotá D.C.,CO.

Número de Grupos	Índice Rand	Índice Jaccard	Índice Propuesto	Índice Rand Ajustado	Índice Browuer
2	0.42936941	0.4942745	0.21405932	0.39786096	0.914935612
3	0.30260235	0.48090568	0.15067578	0.38651033	0.916315532
4	0.34942974	0.50800216	0.17408948	0.4114428	0.916482776
5	0.25880085	0.45535832	0.12877503	0.36074144	0.908746077
6	0.26966478	0.46199366	0.134207	0.36651461	0.911190095
7	0.2463477	0.43168233	0.12254846	0.33764535	0.904183639
8	0.28788835	0.44775225	0.14331878	0.35240382	0.903423855
9	0.3068675	0.46770933	0.15280836	0.37148921	0.907145711
10	0.29382886	0.44906861	0.14628904	0.35457218	0.903282154
11	0.31163767	0.45127069	0.15519345	0.35574708	0.90121789
12	0.28278402	0.42959179	0.14076662	0.33464958	0.89741714
13	0.29216876	0.42946613	0.14545899	0.33504881	0.895915804
14	0.31816264	0.43950528	0.15845593	0.3446421	0.895246778
15	0.29181305	0.41322911	0.14528113	0.31889514	0.889312555
16	0.26444835	0.38830381	0.13159879	0.29612497	0.88611459
17	0.28448109	0.39451291	0.14161516	0.30193387	0.885263759
18	0.26861958	0.37968695	0.1336844	0.28814905	0.88241958

CUADRO 4.8: Resultados de las Funciones Objetivo ejecutadas para para el algoritmo Fuzzy C-Means en la ciudad de Bogotá D.C.,CO

Como se puede observar, el índice propuesto es el que presenta los valores mínimos de todas las funciones objetivo evaluadas. Por otra parte, para el caso del procesamiento del algoritmo de agrupamiento difuso Fuzzy C-Means, se tiene que el valor mínimo de los índices de validación, por tanto el numero de grupos en el cual se tiene la mayor similitud entre los conjuntos de datos de Malla Vial y Hurtos a Celulares, se obtiene para 18 agrupamientos, que para esta área de estudio es su limite

superior elegido.

La figura 4.1 muestra la evolución de los valores de las funciones objetivo evaluadas a medida que aumenta el numero de grupos elegidos para los algoritmos de agrupamiento.
Esta figura corrobora que el índice propuesto es el que genera los menores valores en su función objetivo.

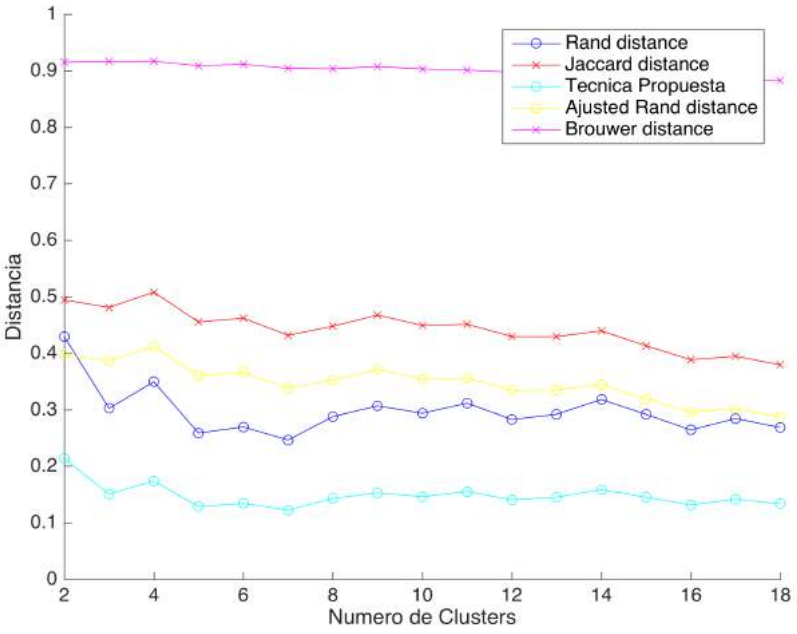


FIGURA 4.1: Funciones Objetivo para el Agrupamiento mediante el algoritmo FCM en la ciudad de Bogotá D.C.,CO.

La tabla 4.9 contiene los resultados obtenidos por los índices de validación para el algoritmo Gustafson Kessel en la ciudad de Bogotá D.C.,CO.

Número de Grupos	**Índice Rand**	**Índice Jaccard**	**Índice Propuesto**	**Índice Rand Ajustado**	**Índice Browuer**
2	0.73690011	0.63948153	0.36782467	0.55823327	0.936197514
3	0.2394546	0.4447829	0.11910191	0.35377066	0.906596498
4	0.38253782	0.53312127	0.19064352	0.43668153	0.91756024
5	0.24770692	0.43155598	0.12322807	0.33830386	0.902966895
6	0.30360776	0.45919894	0.15117849	0.36240859	0.905142475
7	0.33283397	0.49150804	0.16579159	0.39510939	0.911652296
8	0.29879383	0.45611054	0.14877152	0.36297244	0.903933091
9	0.28745062	0.43964653	0.14309992	0.34511733	0.900308261
10	0.28773849	0.43366539	0.14324385	0.3407307	0.898440751
11	0.33279149	0.46427767	0.16577035	0.36970717	0.901669341
12	0.3298787	0.46159594	0.16431396	0.36722891	0.901570973
13	0.30394622	0.43772516	0.15134772	0.3410031	0.898266868
14	0.3044764	0.43138716	0.15161281	0.33621368	0.89575059
15	0.30059088	0.42220769	0.14967005	0.32740829	0.892367212
16	0.29158794	0.40546541	0.14516858	0.31328437	0.888336018
17	0.29373811	0.40225147	0.14624366	0.31055264	0.886018613
18	0.30417125	0.41717782	0.15146023	0.32190352	0.889549054

CUADRO 4.9: Resultados de las Funciones Objetivo ejecutadas para para el algoritmo Gustafson Kessel en la ciudad de Bogotá D.C.,CO

Para este caso donde se se procesa el algoritmo de agrupamiento difuso Gustafson Kessel, se tiene que el valor mínimo de los índices de validación se obtiene para 3 agrupamientos.
Por otro lado, tal como ocurre en el caso anterior para la ciudad de Bogotá, el índice propuesto es el que genera los valores mínimos entre las funciones objetivo.
La figura 4.2 muestra la evolución de los valores de las funciones objetivo evaluadas a medida que aumenta el numero de grupos elegidos para los algoritmos de agrupamiento.

Esta figura corrobora que el índice propuesto es el que genera los menores valores en su función objetivo.

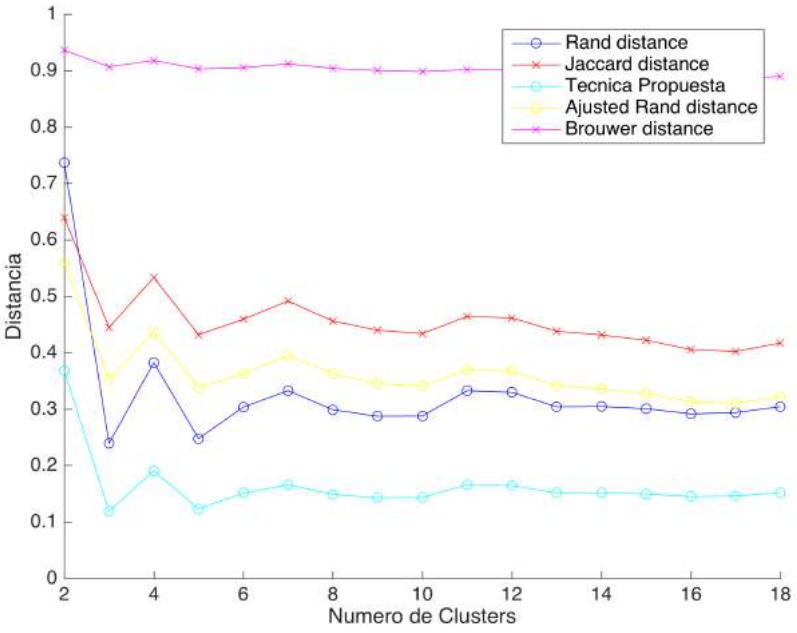


FIGURA 4.2: Funciones Objetivo para el Agrupamiento mediante el algoritmo GK en la ciudad de Bogotá D.C.,CO.

Ahora se procede a comparar los valores donde se ha encontrado el mínimo de la función objetivo para cada uno de los algoritmos de agrupamiento. Para el primer caso de estudio que corresponde a la ciudad de Bogotá, se tiene que para el agrupamiento difuso con el algoritmo Fuzzy C-Means el valor mínimo de la función objetivo se obtiene en 18 grupos con un valor de 0.1336844. Por otra parte, el agrupamiento difuso con el algoritmo Gustafson Kessel presenta un valor mínimo de la función objetivo cuando se obtienen 3 grupos con un valor de 0.11910191.
De esta manera se concluye que el grupo que presenta el valor mínimo y por tanto la mayor similitud entre los conjuntos de Malla Vial y Hurtos para la ciudad de Bogotá, es el agrupamiento de 3 grupos mediante el algoritmo Gustafson Kessel.
Las figuras 4.3,4.4,4.5 y 4.6 muestran los agrupamientos obtenidos para los conjuntos de Malla Vial y Hurtos a Celulares en la ciudad de Bogotá para los cuales el índice de validación ha llegado a su valor mínimo en cada uno de los algoritmos de agrupamiento. Para consultar los agrupamientos obtenidos en cada una de las ejecuciones para el área de Bogotá ver sección Anexos.

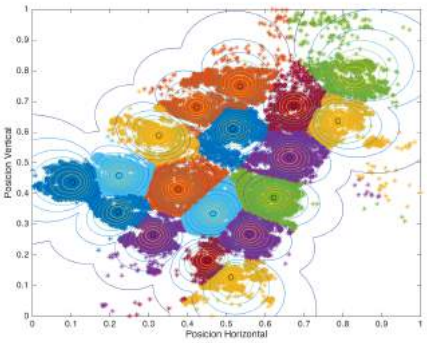


FIGURA 4.3: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 18 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

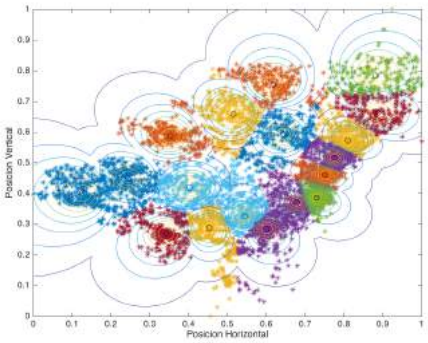


FIGURA 4.4: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 18 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

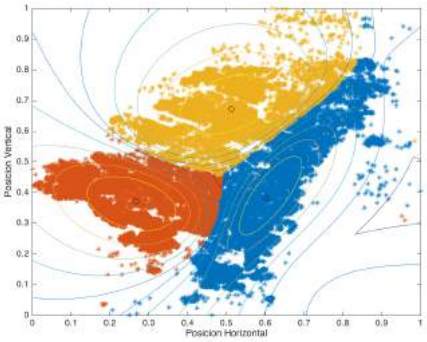


FIGURA 4.5: Agrupamiento obtenido mediante el algoritmo Gustafson Kessel para 3 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

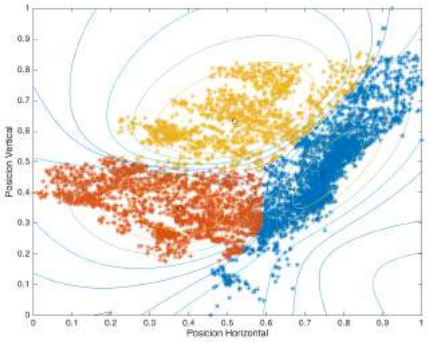


FIGURA 4.6: Agrupamiento obtenido mediante el algoritmo Gustafson Kessel para 3 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

4.3.3.2. San Francisco,CA,USA

La tabla 4.10 contiene los resultados obtenidos por los índices de validación para el algoritmo Fuzzy C-Means en la ciudad de San Francisco,CA,USA.

Número de Grupos	Índice Rand	Índice Jaccard	Índice Propuesto	Índice Rand Ajustado	Índice Browuer
2	0.17385009	0.3132795	0.08629966	0.23707853	0.899271448
3	0.18702602	0.44159782	0.08288762	0.35524622	0.915379862
4	0.41588374	0.58306106	0.20731648	0.49353626	0.92965691
5	0.34584452	0.55733797	0.17229687	0.46829015	0.930191192
6	0.26183377	0.5107713	0.1302915	0.42179717	0.92474569
7	0.29383629	0.52489361	0.14629275	0.43591873	0.926489731
8	0.32532323	0.53399036	0.16203622	0.44481875	0.926472346
9	0.30567293	0.51663206	0.15221107	0.42695799	0.922726767
10	0.26942734	0.48363409	0.13408828	0.39408271	0.919322316

CUADRO 4.10: Funciones Objetivo para el Agrupamiento mediante el algoritmo FCM en la ciudad de San Francisco,CA.USA

De manera análoga a la ejecución en el área de Bogotá, los resultados en para el área de San Francisco muestran que el índice propuesto es el que presenta los valores mínimos de todas las funciones objetivo evaluadas. Por otra parte, para el caso del procesamiento del algoritmo de agrupamiento difuso Fuzzy C-Means, se tiene que el valor mínimo de los índices de validación se obtienen para 3 agrupamientos.

La figura 4.7 muestra la evolución de los valores de las funciones objetivo evaluadas a medida que aumenta el numero de grupos elegidos para los algoritmos de agrupamiento.

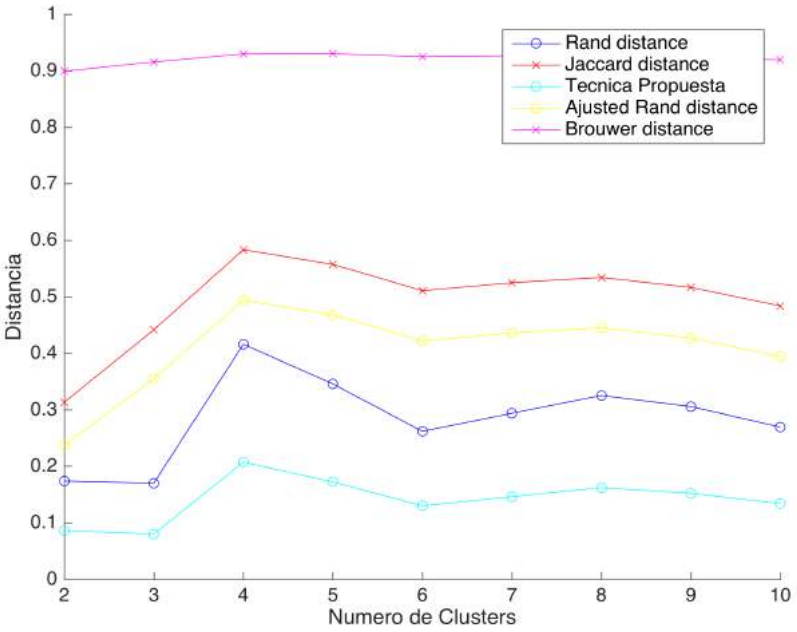


FIGURA 4.7: Funciones Objetivo para el Agrupamiento mediante el algoritmo FCM en la ciudad de Bogotá D.C.,CO.

La tabla 4.11 contiene los resultados obtenidos por los índices de validación para el algoritmo Gustafson Kessel en la ciudad de San Francisco,CA,USA.

Número de Grupos	**Índice Rand**	**Índice Jaccard**	**Índice Propuesto**	**Índice Rand Ajustado**	**Índice Browuer**
2	0.17385009	0.3132795	0.08629966	0.23707853	0.899271448
3	0.1702	0.44159782	0.08868472	0.35524622	0.915379862
4	0.41588374	0.58306106	0.20731648	0.49353626	0.92965691
5	0.34584452	0.55733797	0.17229687	0.46829015	0.930191192
6	0.26183377	0.5107713	0.1302915	0.42179717	0.92474569
7	0.29383629	0.52489361	0.14629275	0.43591873	0.926489731
8	0.32532323	0.53399036	0.16203622	0.44481875	0.926472346
9	0.30567293	0.51663206	0.15221107	0.42695799	0.922726767
10	0.26942734	0.48363409	0.13408828	0.39408271	0.919322316

CUADRO 4.11: Funciones Objetivo para el Agrupamiento mediante el algoritmo GK en la ciudad de San Francisco,CA.USA

Para el caso del procesamiento del algoritmo de agrupamiento difuso Gustafson Kessel, se tiene que el valor mínimo de los índices de validación se obtienen para 2 agrupamientos.

La figura 4.8 muestra la evolución de los valores de las funciones objetivo evaluadas a medida que aumenta el numero de grupos elegidos para los algoritmos de agrupamiento.
Ahora se procede a comparar los valores donde se ha encontrado el mínimo de la función objetivo para cada uno de los algoritmos de agrupamiento. Para el primer caso de estudio que corresponde a la ciudad de San Francisco, se tiene que para el agrupamiento difuso con el algoritmo Fuzzy C-Means el valor mínimo de la función objetivo se obtiene en 3 grupos con un valor de 0.08288762. Por otra parte, el agrupamiento difuso con el algoritmo Gustafson Kessel presenta un valor mínimo de la función objetivo cuando se obtienen 2 grupos con un valor de 0.08629966.
De esta manera se concluye que el grupo que presenta el valor mínimo y por tanto la mayor similitud entre los conjuntos de Malla Vial y Hurtos para la ciudad de San Francisco, es el agrupamiento de 3 grupos

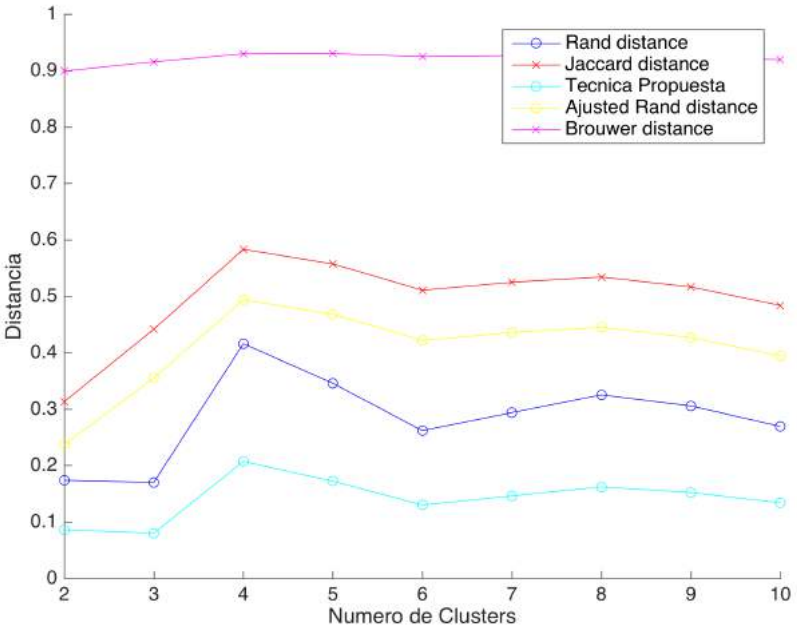


FIGURA 4.8: Funciones Objetivo para el Agrupamiento mediante el algoritmo GK en la ciudad de San Francisco,CA,USA.

mediante el algoritmo Fuzzy C-Means.

Las figuras 4.9,4.10,4.11 y 4.12 muestran los agrupamientos obtenidos para los conjuntos de Malla Vial y Hurtos a Celulares en la ciudad de San Francisco para los cuales el índice de validación ha llegado a su valor mínimo en cada uno de los algoritmos de agrupamiento. Para consultar los agrupamientos obtenidos en cada una de las ejecuciones para el área de San Francisco ver sección Anexos.

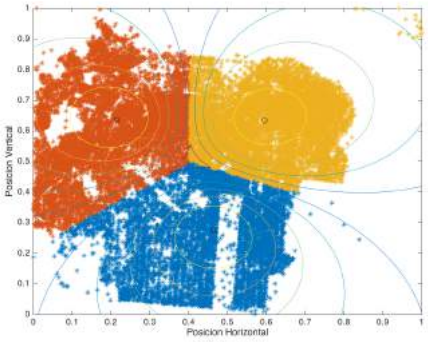


FIGURA 4.9: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 3 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA,USA.

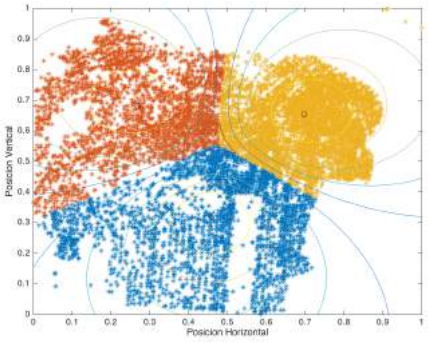


FIGURA 4.10: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 3 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA,USA.

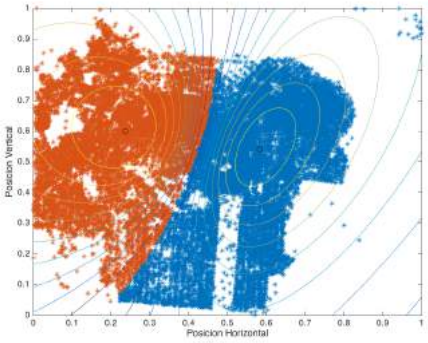


FIGURA 4.11: Agrupamiento obtenido mediante el algoritmo Gustafson Kessel para 2 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA,USA.

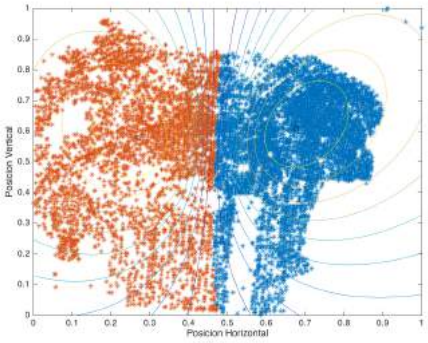


FIGURA 4.12: Agrupamiento obtenido mediante el algoritmo Gustafson Kessel para 2 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA,USA.

4.4. Análisis Generales

A partir de los resultados obtenidos por las ejecuciones de la técnica propuesta, se presentan los siguientes análisis de los resultados obtenidos:

- **Para las dos ciudades procesadas por la técnica se obtiene que la mejor distribución para el periodo de tiempo seleccionado se da en 3 agrupamientos, pero cada uno obtiene los mejores resultados a partir de diferentes algoritmos de agrupamiento difuso; puesto que mientras que para Bogotá se obtiene con el algoritmo Gustafson**

Kessel, para San Francisco se obtiene con el algoritmo Fuzzy C-Means.

- **Analizando las agrupaciones obtenidas en la ciudad de Bogotá, podemos observar que el mejor resultado divide la ciudad por 4 vías de suma importancia para la ciudad como lo son la autopista Norte, la Avenida el Dorado, La Avenida NQS y el tramo Sur de la Avenida Boyacá.**
- **Por otra parte, en la ciudad de San Francisco el mejor resultado divide la ciudad por la Calle Market y la Calle Guerrero, calles que atraviesan la ciudad de tramo a tramo.**

Capítulo 5

Conclusiones y trabajos futuros

"La conclusión es que sabemos muy poco y sin embargo es asombroso lo mucho que conocemos. Y más asombroso todavía que un conocimiento tan pequeño pueda dar tanto poder."

Bertrand Russell

5.1. Resumen

En este trabajo se presento una propuesta para una técnica de agrupamiento difuso para eventos criminales basada en la forma urbana del área en la que ocurren estos eventos.
En primer lugar se presentaron los conceptos referentes a teorías de análisis del crimen. Posteriormente se presentaron los conceptos referentes a los conjuntos difusos y la lógica difusa, dando paso así a los métodos de agrupamiento de datos donde se destaca el concepto del Cluster, de los métodos de agrupamiento se hace énfasis en los métodos de agrupamiento difuso, dentro de los cuales se destacan los algoritmos Fuzzy C-Means y Gustafson Kessel. Posteriormente se presentan los conceptos referentes a los métodos de validación de Cluster, entre los que destaca el Índice Rand.
Se mostró la propuesta de una técnica que al recibir dos conjuntos de datos correspondientes a la Malla Vial de un área especifica y a unos eventos criminales ocurridos en dicha área durante un periodo de tiempo determinado, se procesan mediante algoritmos de agrupamiento

difuso para posteriormente ser comparados bajo índices de validación de Cluster que determinaran en que cantidad de agrupamientos los dos conjuntos de datos presentan la mayor similitud. Para esta sección se propone un índice nuevo de validación de Cluster basado en el índice Rand para agrupamientos difusos.
Posteriormente se mostró un conjunto de pruebas ejecutadas para probar el funcionamiento de la técnica desarrollada para las ciudades de Bogotá D.C., Colombia y San Francisco, Estados Unidos de América, donde se procesaron en la técnica los eventos criminales correspondientes a Hurtos a Celulares y Hurtos a Residencias.
Las pruebas ejecutadas consistieron en un procesamiento de 100 iteraciones para cada uno de los conjuntos de datos, estos variando en la cantidad máxima de grupos decididos por el usuario que ejecuta las pruebas. Posteriormente se eligió el mejor grupo de conjuntos de datos agrupados dependiendo de su rendimiento, luego de elegir estos conjuntos de datos agrupados se utilizaron los métodos de validación de Cluster para crear las funciones objetivo que deciden mediante su valor mínimo, el numero de agrupaciones en las que los conjuntos de datos procesados poseen mayor similitud.
Finalmente, conociendo el número de grupos que presentan la mayor similitud para cada algoritmo, se elige entre estos dos el que represente el verdadero valor mínimo de todas las ejecuciones efectuadas sobre el conjunto de datos.

5.2. Aportes Originales

- **En este trabajo se creo un algoritmo que permite escoger entre un grupo de ejecuciones de algoritmos de agrupamiento la mejor ejecucion basada en el tiempo que este tarda en llegar a su función objetivo.**
- **Se planteo una técnica que permite conocer el numero correcto de grupos en el cual debe ser distribuido un conjunto de datos de eventos criminales teniendo en cuenta las caracteristicas espaciales del lugar donde han ocurrido dichos eventos.**

- **Se propone un índice para la validación de agrupamientos difusos basado en el índice Rand y en su expansión a los agrupamientos difusos.**

5.3. Trabajos futuros

Partiendo del desarrollo de este proyecto y de los resultados obtenidos en el mismo, se plantean los siguientes trabajos futuros relacionados con el análisis de agrupamientos difusos y la aplicación de estos en el análisis del crimen:

- **Aplicación de la técnica desarrollada en otras ciudades con el fin de comprobar el funcionamiento y buscar la generalización de un método. Ciudades como Filadelfia, Oklahoma City, Los Ángeles, Boston y New York, USA. Londres, UK. Berlin, GER., entre otras, tienen a la disponibilidad del publico los datos de Malla Vial y Eventos criminales para ser aplicados a la técnica.**
- **Aplicación de la técnica mediante el uso de otros algoritmos de agrupamiento difuso, esto con el fin de comprobar que las formas que entregan los algoritmos GK y FCM son las mas adecuadas para distribuir el area de estudio de la técnica. Dentro de los algoritmos de agruoamiento difuso se proponen el Fuzzy C-Varieties (FCV), Fuzzy C-Regression Models (FCRM), Fuzzy C-Elliptotypes (FCE), y el Adaptive Fuzzy C-Regression Models (AFCR).**
- **Aplicación de la técnica para otras tipificaciones de crimen tales como Asesinato, Violación, Hurtos a tiendas bajo la modalidad de atraco, Hurtos bajo la modalidad de cosquilleo, Hurtos de Vehículos, Atracos a vehículos de repartición de dinero, Atracos a Bancos, entre otros.**
- **Proponer otra técnica para la comparación de las agrupaciones obtenidas, entre estas se propone tratar las agrupaciones obtenidas como imágenes para compararlas bajo técnicas de procesamiento de imágenes y tratar las agrupaciones como figuras geométricas obtenidas al aplicar técnicas de Convex Hull.**

- **Análisis de la técnica para diferentes periodos de tiempo en las áreas que han sido analizadas en este trabajo, esto con el fin de comprobar si existe un fenómeno fractal en el tiempo. Los periodos de tiempo propuestos son Año, Mes, Semana, Día y Horas especificas.**

- **Análisis de cada agrupamiento como una sección de la ciudad estudiada como un conjunto independiente con características propias que no se presentan en sus similares. Teniendo la certeza de que se ha elegido el correcto numero de agrupaciones para el conjunto de datos procesados por la técnica, se puede iniciar el análisis de patrones criminales para cada agrupamiento como un conjunto de datos independiente de los eventos ocurridos en los agrupamientos adyacentes a el.**

- **Posteriormente, realizando predicciones a partir de la data que se tiene en cada agrupamiento y comparando estos resultados con la data disponible en los próximos meses se podría dar un veredicto sobre si la manera en que una ciudad se ha expandido a través del tiempo ha influido e influye en los crímenes que ocurren sobre dicho lugar.**

Apéndice A

Anexos

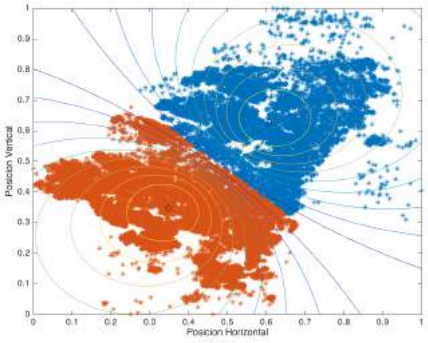


FIGURA A.1: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 2 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

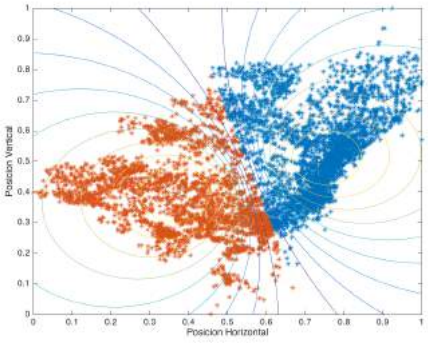


FIGURA A.2: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 2 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

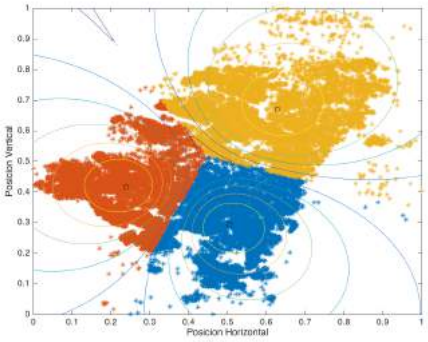


FIGURA A.3: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 3 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

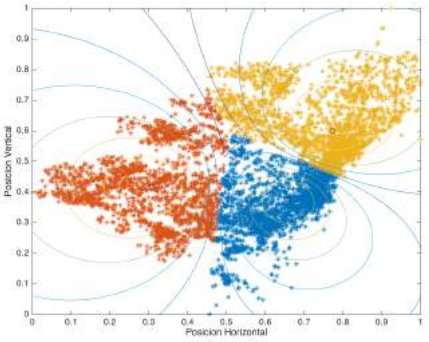


FIGURA A.4: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 3 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

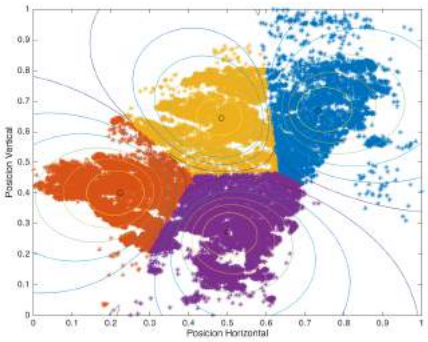


FIGURA A.5: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 4 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

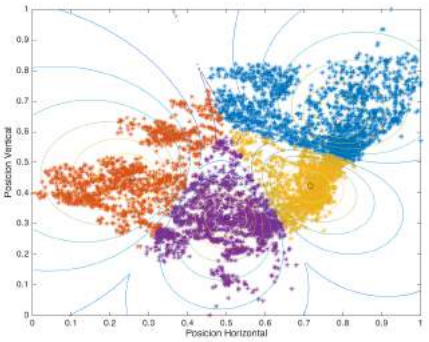


FIGURA A.6: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 4 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

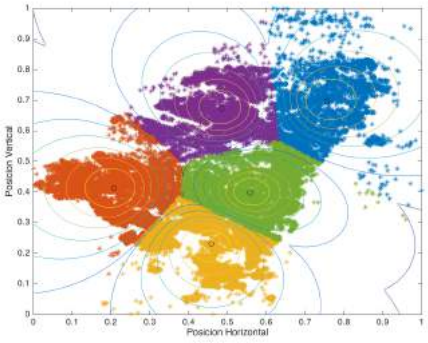


FIGURA A.7: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 5 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

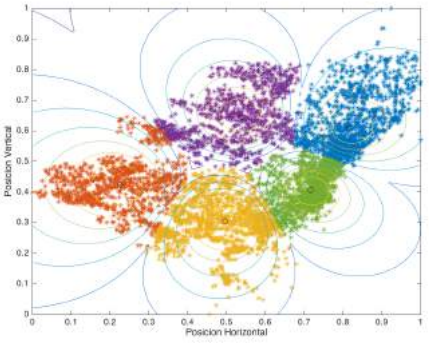


FIGURA A.8: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 5 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

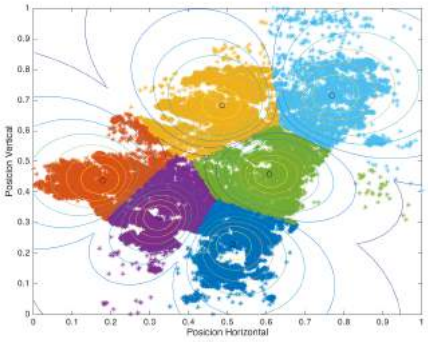


FIGURA A.9: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 6 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

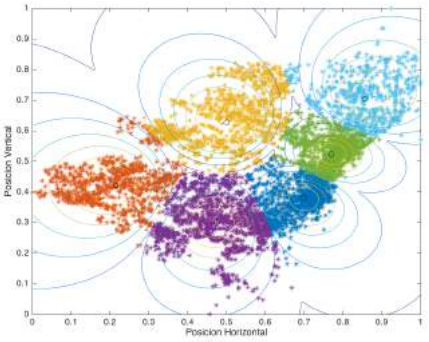


FIGURA A.10: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 6 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

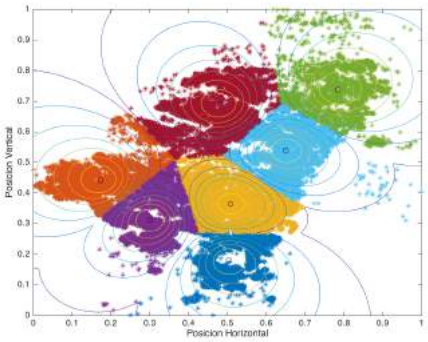


FIGURA A.11: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 7 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

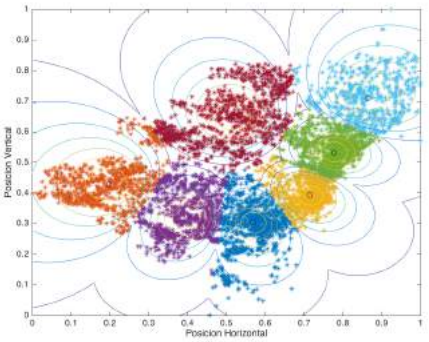


FIGURA A.12: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 7 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

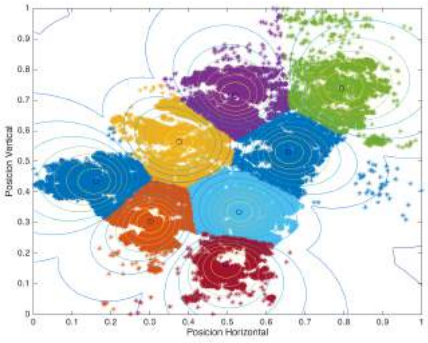


FIGURA A.13: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 8 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

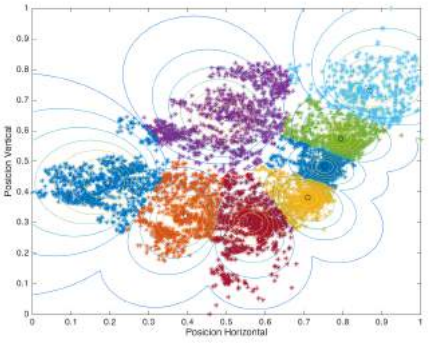


FIGURA A.14: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 8 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

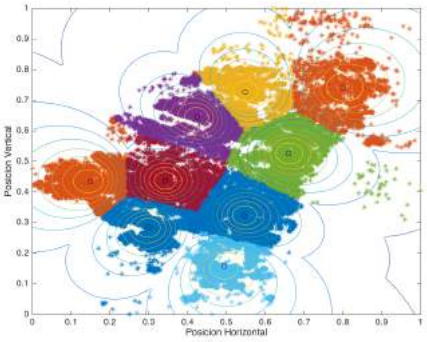


FIGURA A.15: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 9 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

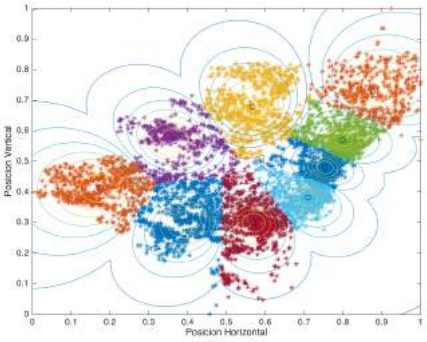


FIGURA A.16: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 9 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

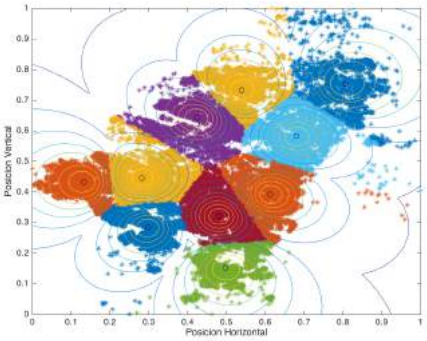


FIGURA A.17: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 10 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

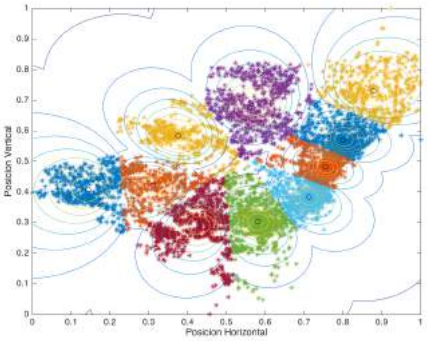


FIGURA A.18: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 10 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

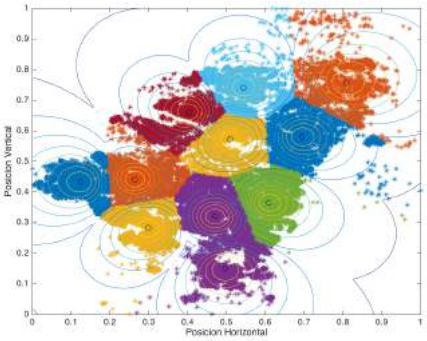


FIGURA A.19: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 11 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

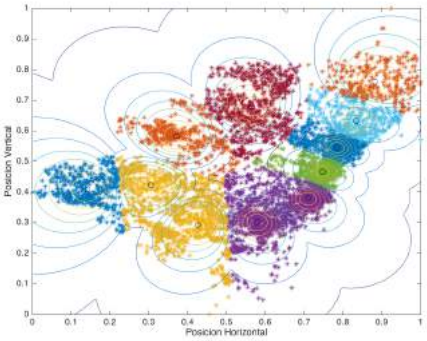


FIGURA A.20: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 11 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

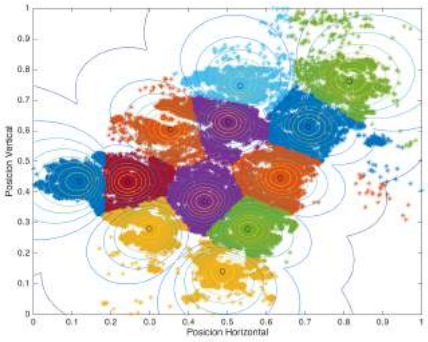


FIGURA A.21: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 12 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

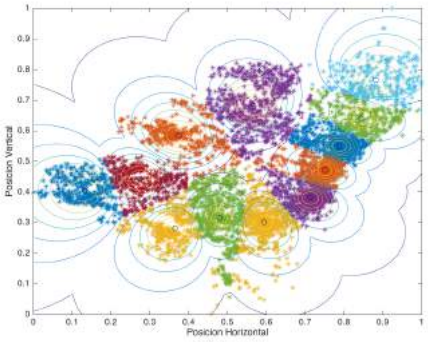


FIGURA A.22: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 12 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

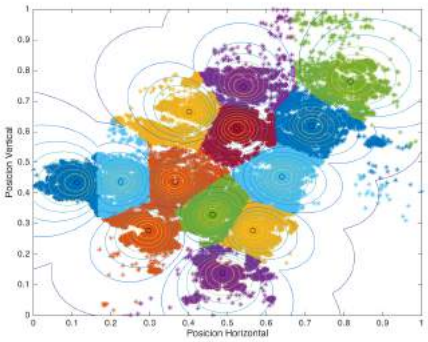


FIGURA A.23: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 13 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

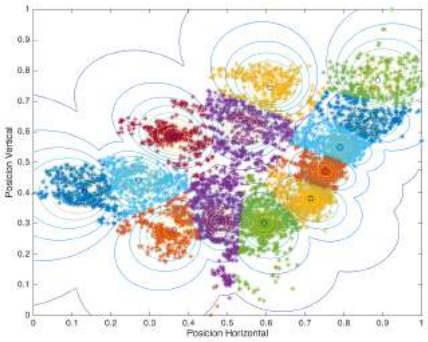


FIGURA A.24: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 13 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

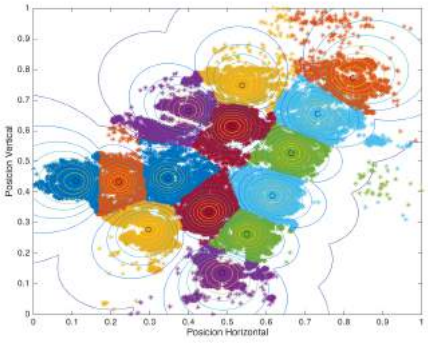


FIGURA A.25: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 14 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

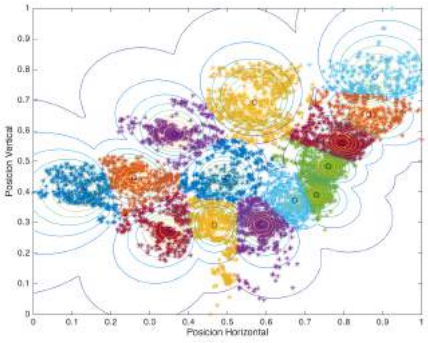


FIGURA A.26: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 14 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

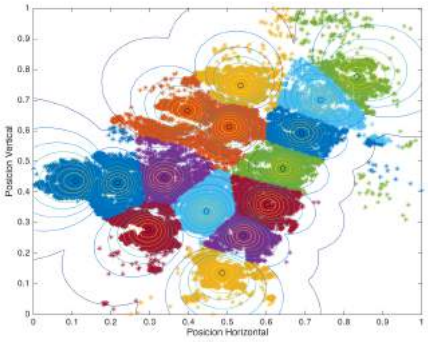


FIGURA A.27: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 15 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

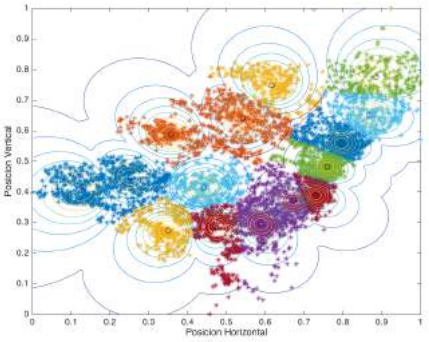


FIGURA A.28: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 15 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

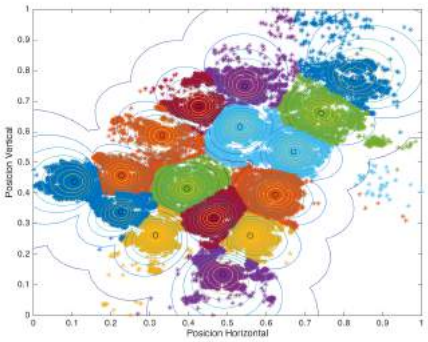


FIGURA A.29: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 16 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

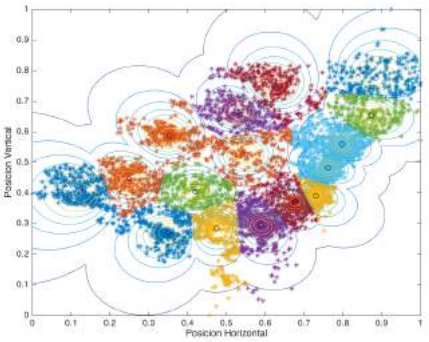


FIGURA A.30: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 16 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

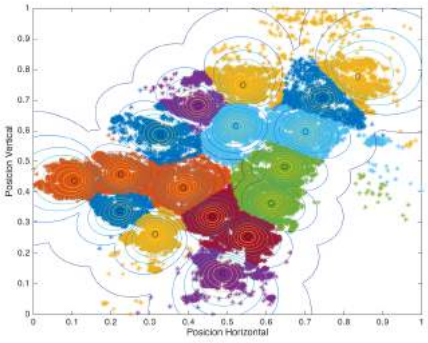


FIGURA A.31: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 17 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

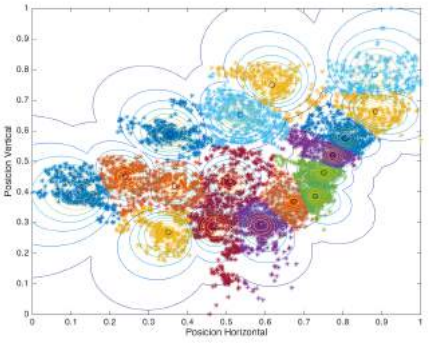


FIGURA A.32: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 17 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

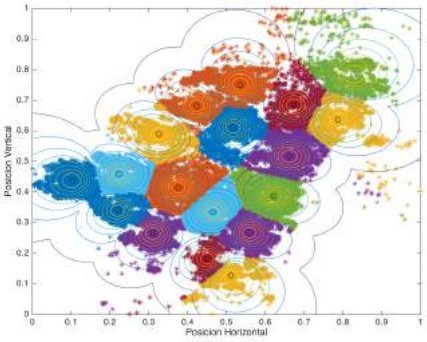


FIGURA A.33: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 18 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

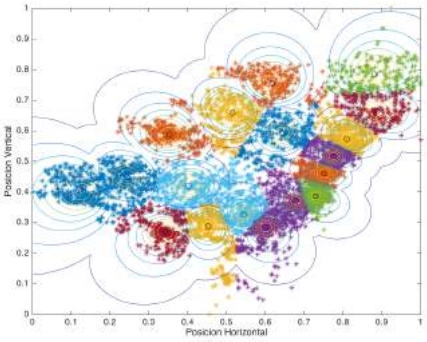


FIGURA A.34: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 18 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

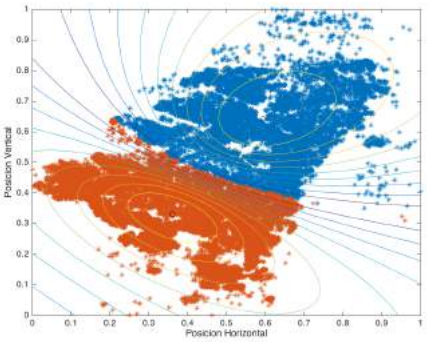


FIGURA A.35: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 2 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

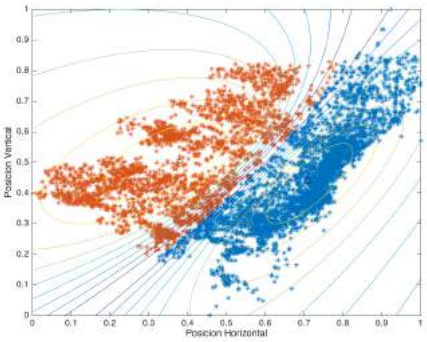


FIGURA A.36: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 2 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

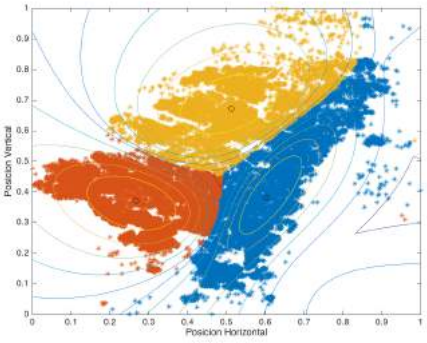


FIGURA A.37: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 3 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

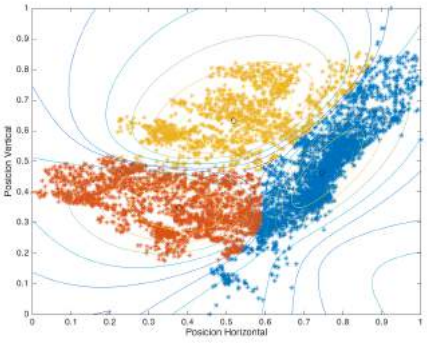


FIGURA A.38: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 3 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

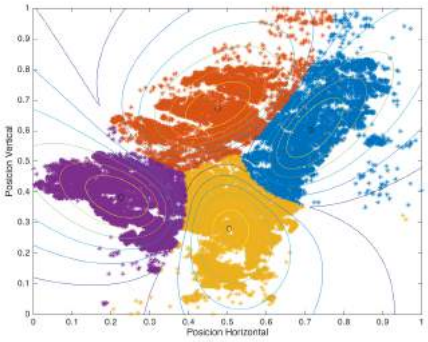


FIGURA A.39: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 4 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

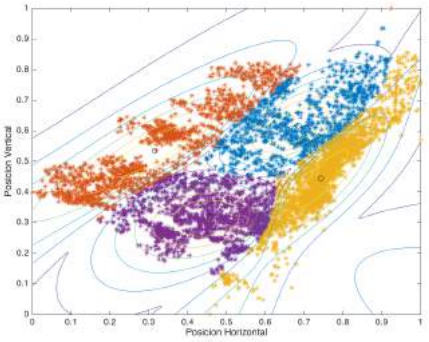


FIGURA A.40: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 4 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

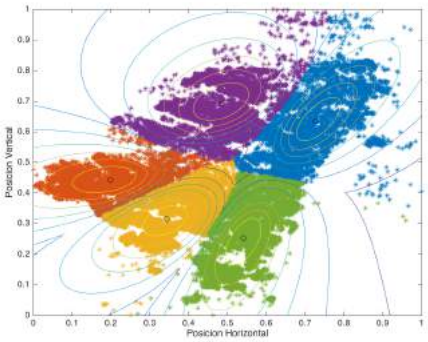


FIGURA A.41: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 5 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

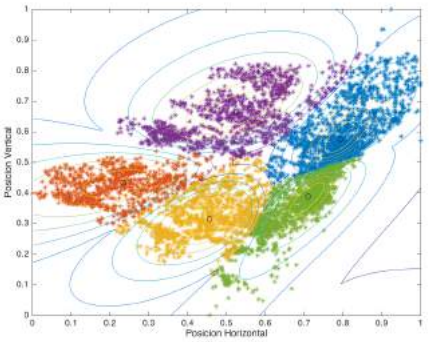


FIGURA A.42: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 5 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

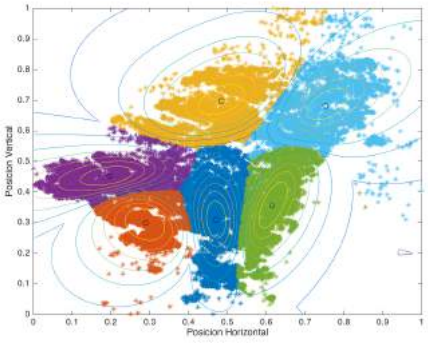


FIGURA A.43: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 6 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

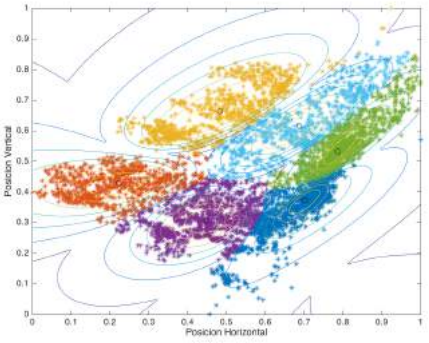


FIGURA A.44: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 6 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

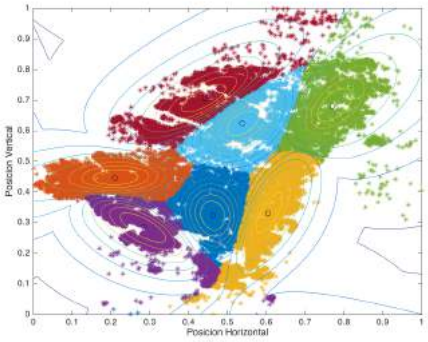


FIGURA A.45: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 7 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

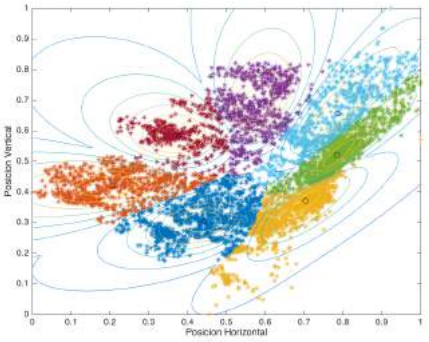


FIGURA A.46: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 7 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

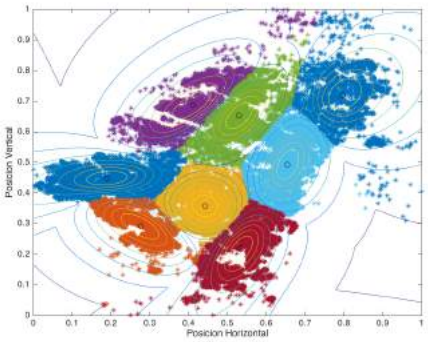


FIGURA A.47: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 8 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

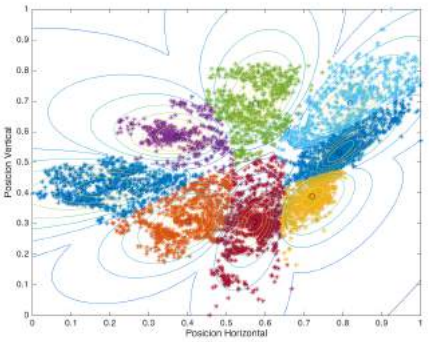


FIGURA A.48: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 8 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

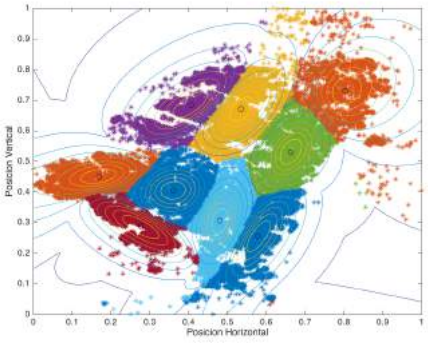


FIGURA A.49: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 9 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

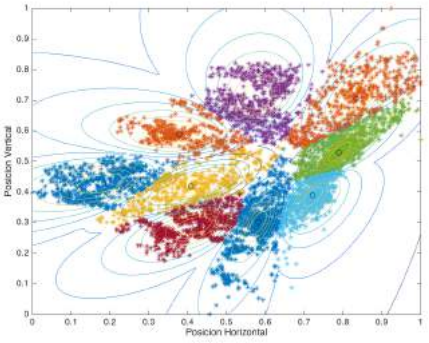


FIGURA A.50: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 9 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

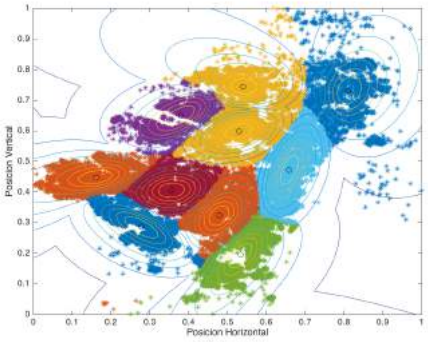


FIGURA A.51: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 10 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

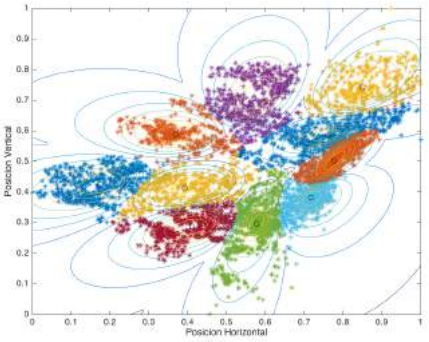


FIGURA A.52: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 10 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

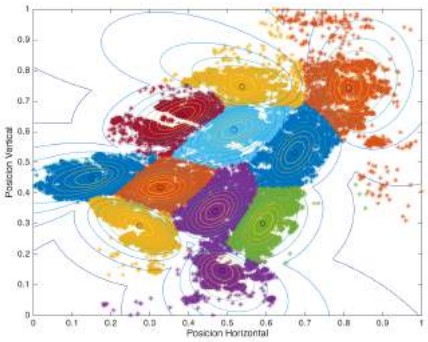


FIGURA A.53: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 11 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

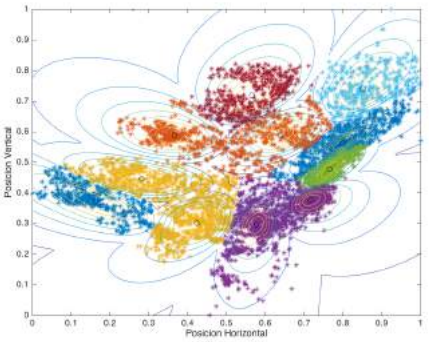


FIGURA A.54: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 11 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

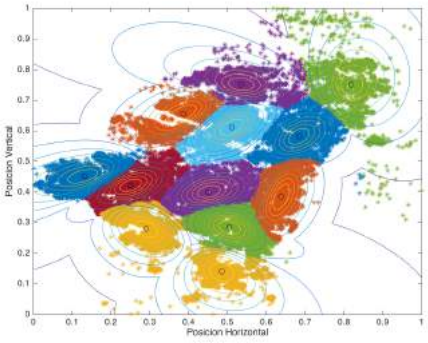


FIGURA A.55: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 12 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

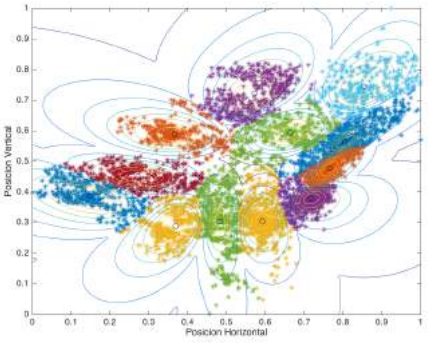


FIGURA A.56: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 12 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

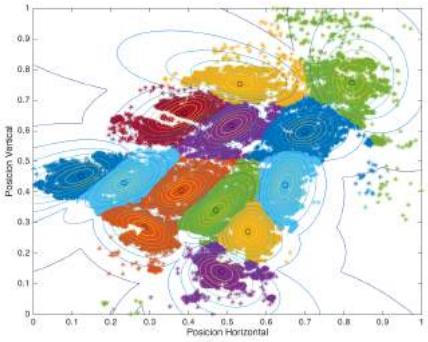


FIGURA A.57: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 13 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

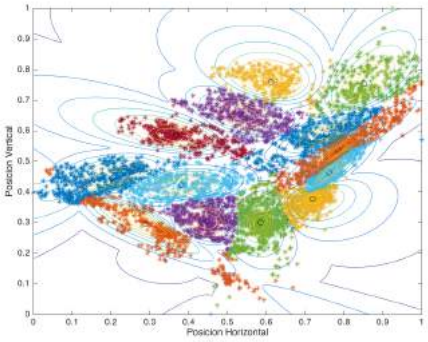


FIGURA A.58: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 13 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

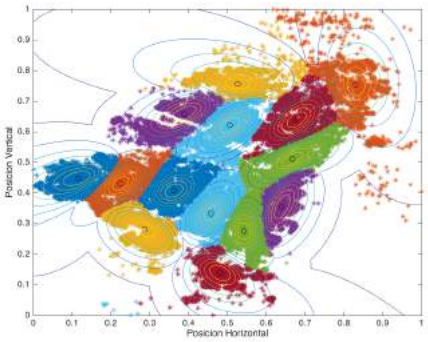


FIGURA A.59: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 14 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

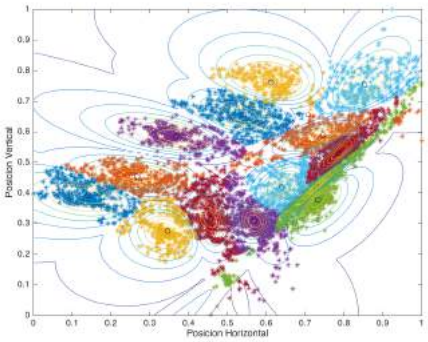


FIGURA A.60: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 14 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

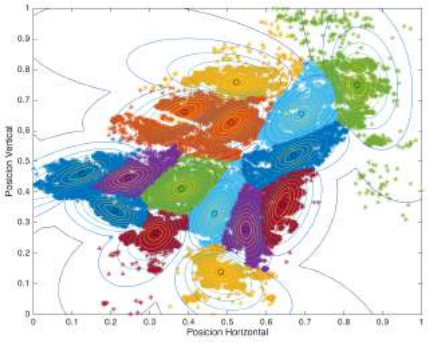


FIGURA A.61: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 15 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

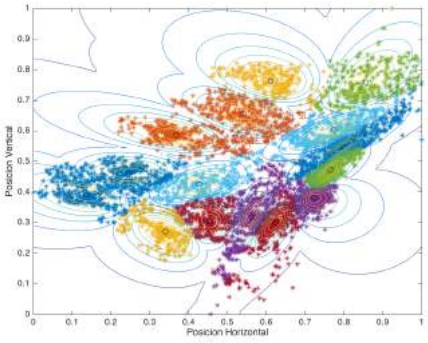


FIGURA A.62: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 15 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

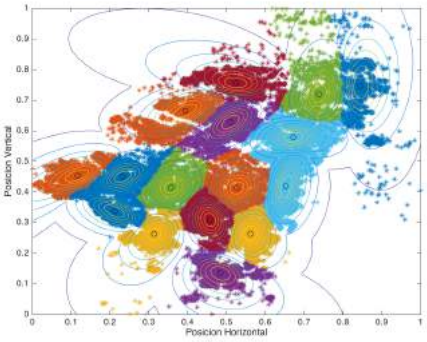


FIGURA A.63: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 16 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

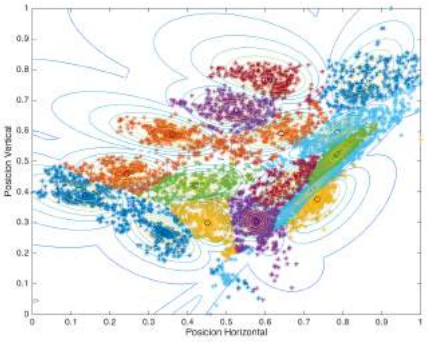


FIGURA A.64: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 16 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

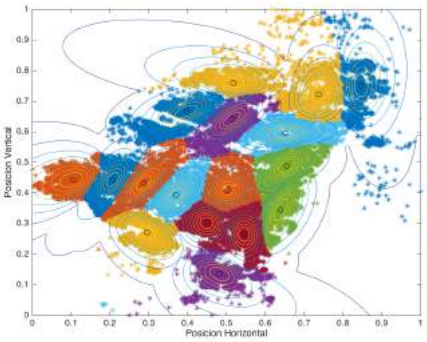


FIGURA A.65: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 17 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

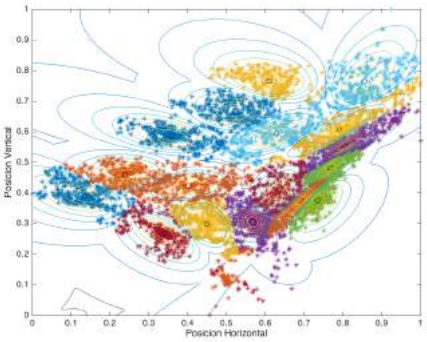


FIGURA A.66: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 17 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

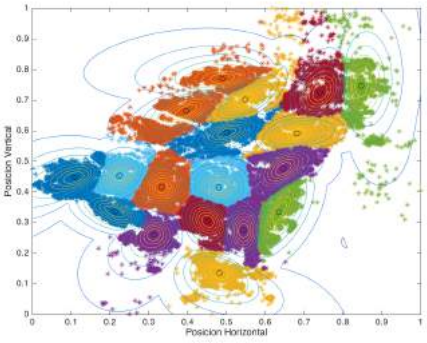


FIGURA A.67: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 18 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de Bogotá,D.C.,CO.

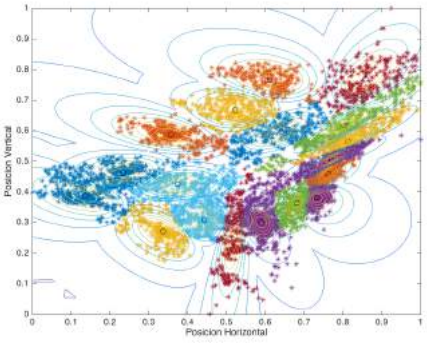


FIGURA A.68: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 18 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de Bogotá,D.C.,CO.

Apéndice B

Anexos

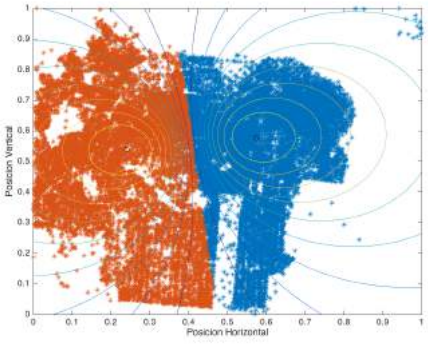


FIGURA B.1: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 2 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA.

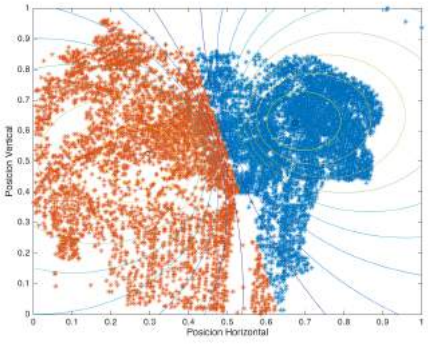


FIGURA B.2: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 2 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA.

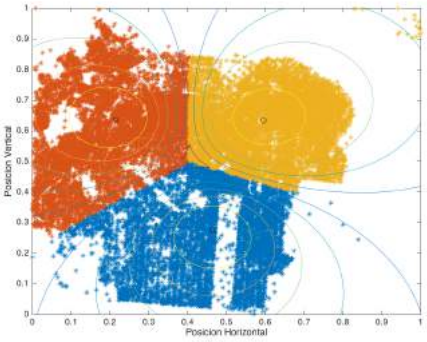


FIGURA B.3: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 3 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA.

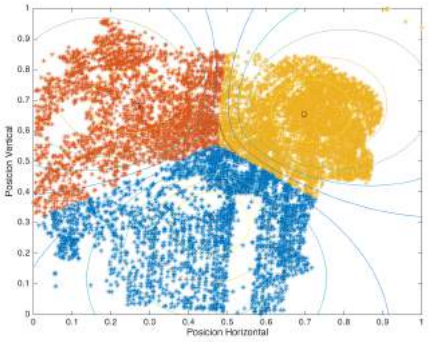


FIGURA B.4: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 3 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA.

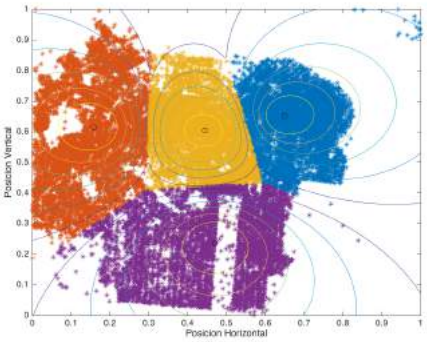


FIGURA B.5: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 4 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA.

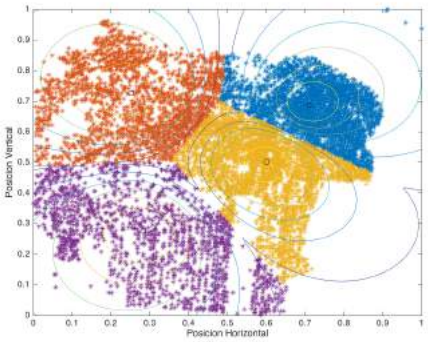


FIGURA B.6: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 4 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA.

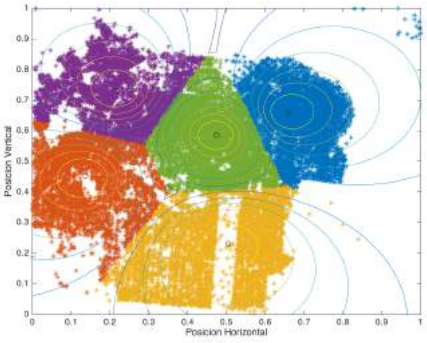


FIGURA B.7: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 5 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA.

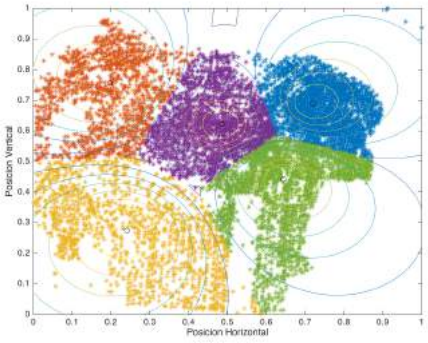


FIGURA B.8: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 5 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA.

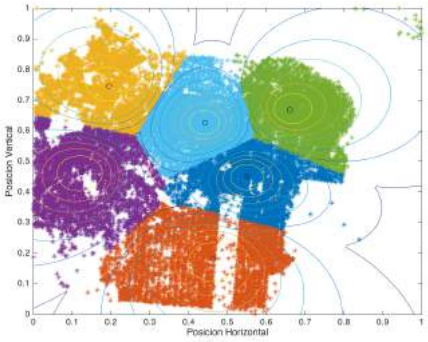


FIGURA B.9: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 6 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA.

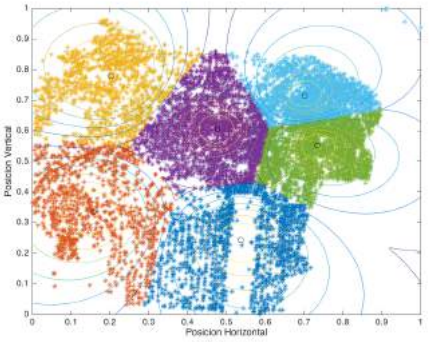


FIGURA B.10: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 6 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA.

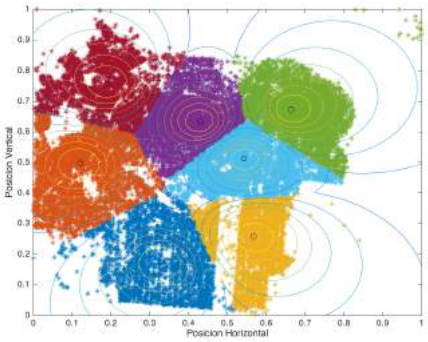


FIGURA B.11: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 7 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA.

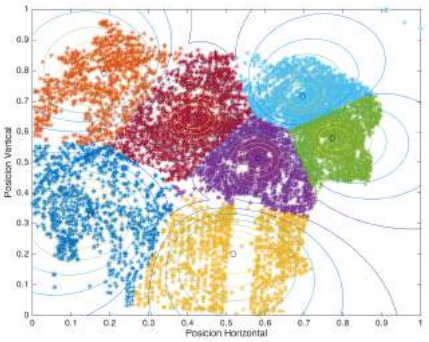


FIGURA B.12: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 7 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA.

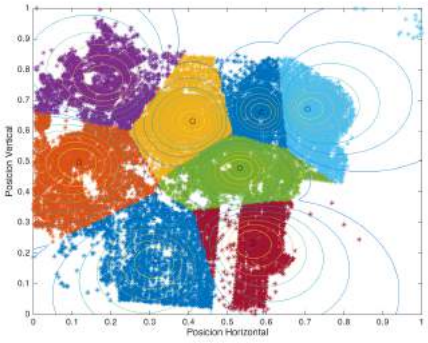


FIGURA B.13: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 8 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA.

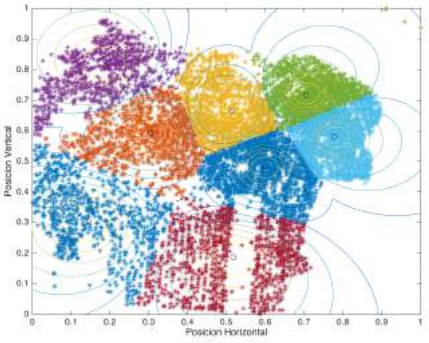


FIGURA B.14: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 8 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA.

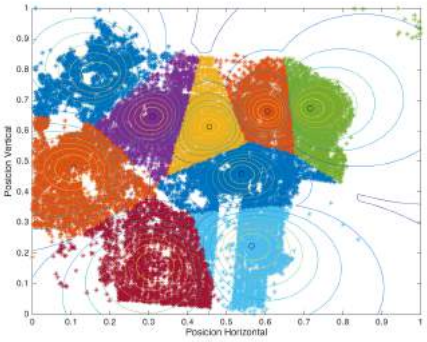


FIGURA B.15: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 9 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA.

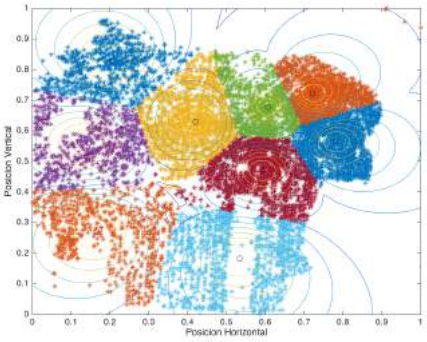


FIGURA B.16: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 9 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA.

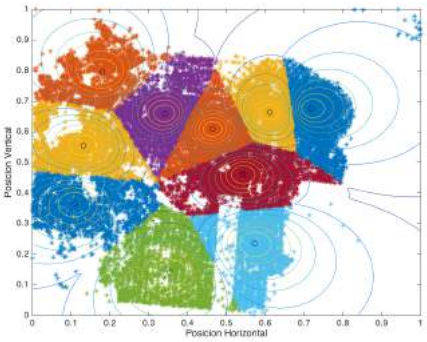


FIGURA B.17: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 10 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA.

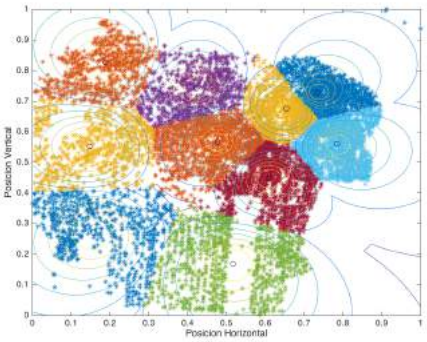


FIGURA B.18: Agrupamiento obtenido mediante el algoritmo Fuzzy C-Means para 10 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA.

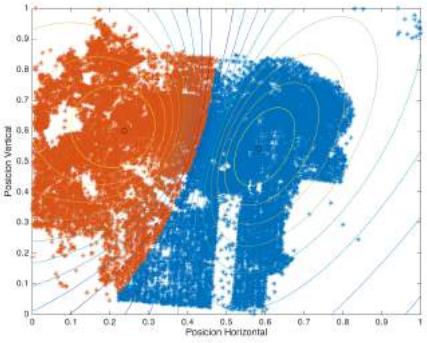


FIGURA B.19: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 2 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA.

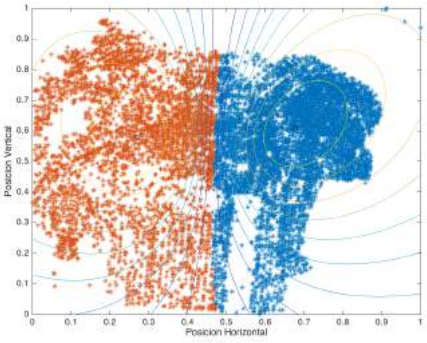


FIGURA B.20: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 2 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA.

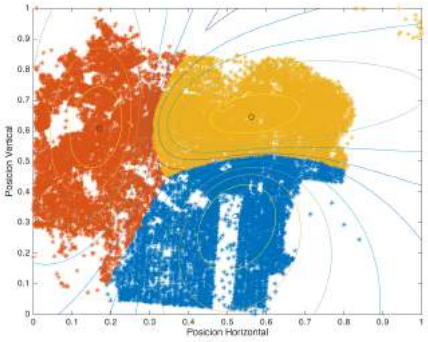


FIGURA B.21: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 3 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA.

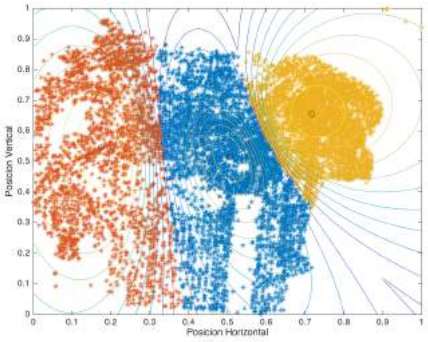


FIGURA B.22: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 3 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA.

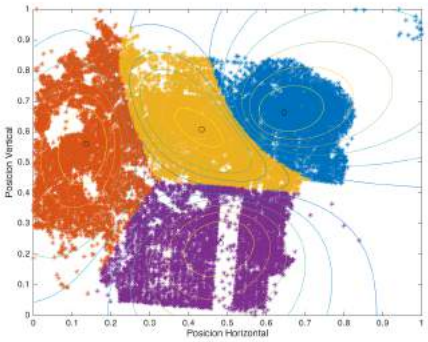


FIGURA B.23: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 4 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA.

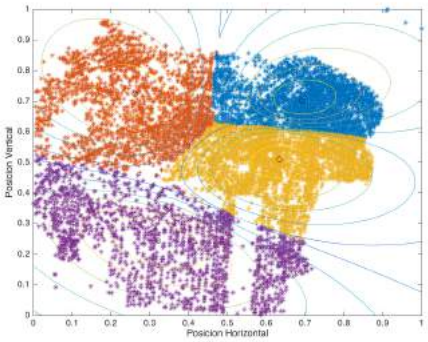


FIGURA B.24: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 4 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA.

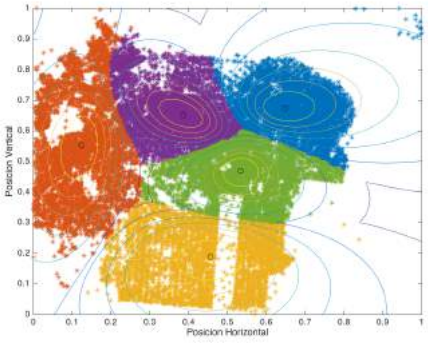


FIGURA B.25: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 5 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA.

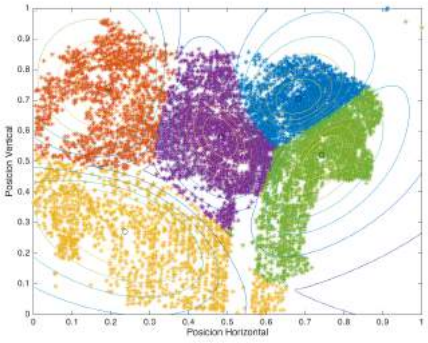


FIGURA B.26: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 5 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA.

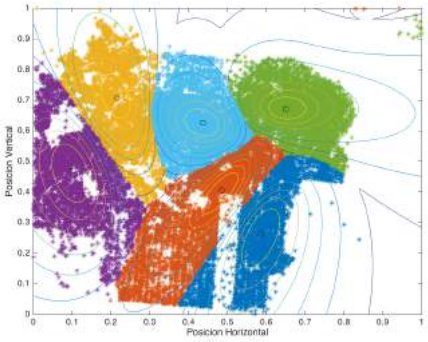


FIGURA B.27: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 6 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA.

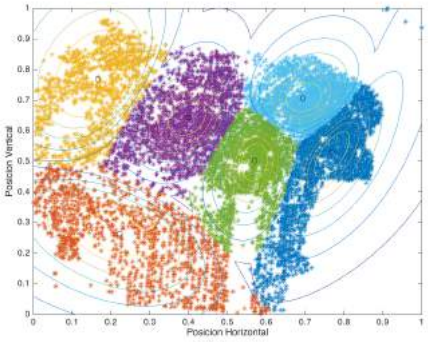


FIGURA B.28: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 6 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA.

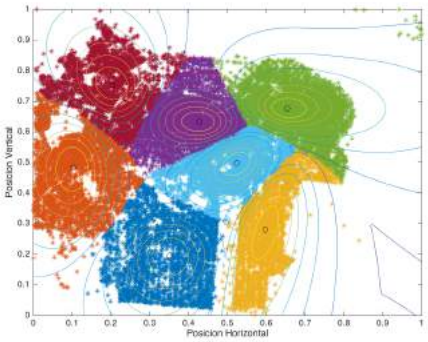


FIGURA B.29: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 7 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA.

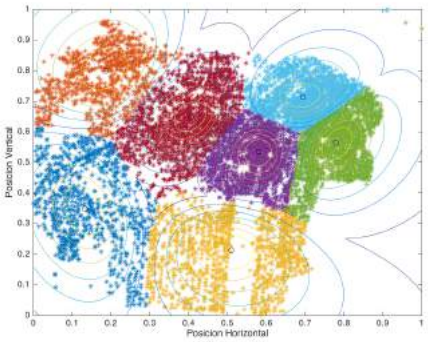


FIGURA B.30: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 7 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA.

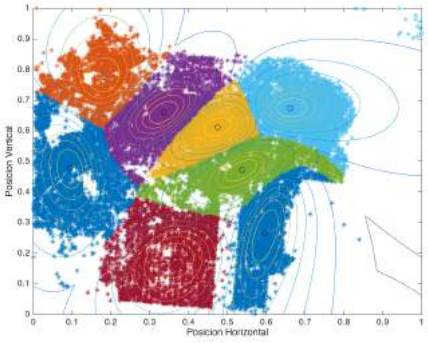


FIGURA B.31: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 8 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA.

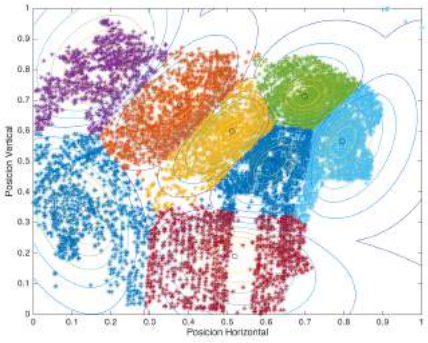


FIGURA B.32: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 8 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA.

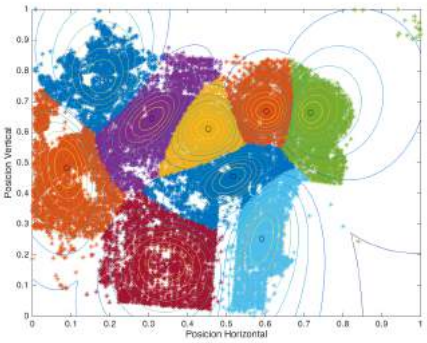


FIGURA B.33: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 9 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA.

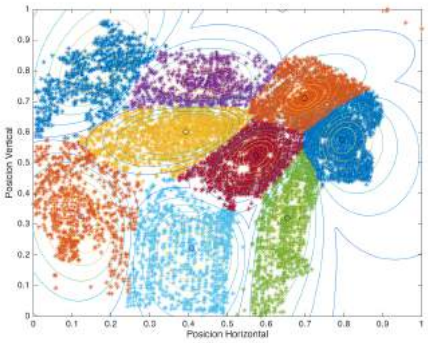


FIGURA B.34: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 9 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA.

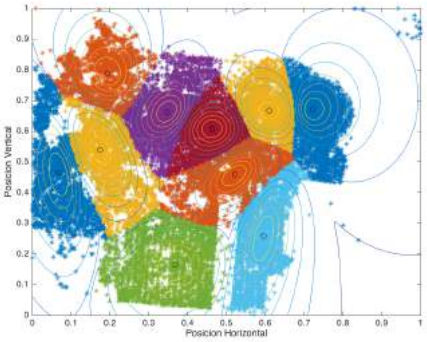


FIGURA B.35: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 10 agrupaciones para el conjuntos de datos de Malla Vial de la ciudad de San Francisco,CA.,USA.

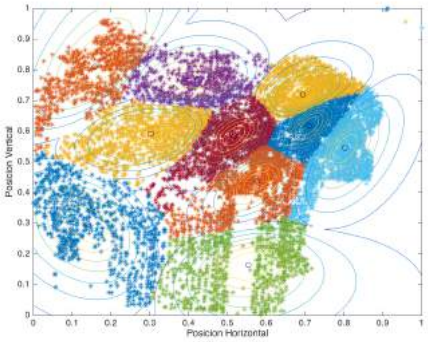


FIGURA B.36: Agrupamiento obtenido mediante el algoritmo Gustafson-Kessel para 10 agrupaciones para el conjuntos de datos de Hurtos a Celulares de la ciudad de San Francisco,CA.,USA.

Apéndice C

Anexos

Bogotá D.C., CO
Fuzzy C-Means

1	86.42067271	**51**	83.68773473
2	86.16901058	**52**	84.14938571
3	84.37422953	**53**	85.43757195
4	84.01173959	**54**	84.74132333
5	83.42254211	**55**	84.65260205
6	86.48122101	**56**	85.85777988
7	85.78709234	**57**	84.26817487
8	83.41616889	**58**	84.39628526
9	85.69224954	**59**	84.12750245
10	86.57998118	**60**	83.96206136
11	84.44151427	**61**	84.34962386
12	86.55074975	**62**	85.55142653
13	85.64439706	**63**	86.00175235
14	83.1925808	**64**	86.25137649
15	83.72554081	**65**	86.16072312
16	85.9904536	**66**	85.48549434
17	83.6668742	**67**	83.35490995
18	86.17068648	**68**	83.38801528
19	85.98629093	**69**	84.39089268
20	86.91599218	**70**	83.45146821
21	84.03209839	**71**	85.81692528
22	83.53918719	**72**	86.56603157
23	83.3932844	**73**	86.58127819
24	84.99386152	**74**	85.23240657
25	84.62348595	**75**	85.95320053
26	84.70885925	**76**	83.20029707
27	84.14115123	**77**	85.52183285
28	83.84600448	**78**	84.21142283
29	83.1092175	**79**	85.67252839
30	83.21322257	**80**	86.89703076
31	84.04260446	**81**	84.11089649
32	85.63344624	**82**	84.91299493
33	83.27096196	**83**	83.45313874
34	83.0000018	**84**	84.32009273
35	83.76455263	**85**	83.87600605
36	86.50668923	**86**	84.9379318
37	83.09411105	**87**	84.54704744
38	83.84006355	**88**	85.48928226
39	83.98038019	**89**	86.25543957
40	83.02870691	**90**	86.37910939
41	83.0583503	**91**	83.7158743
42	86.54228955	**92**	85.52418387
43	86.08921488	**93**	83.44540869
44	84.97429013	**94**	85.72666823
45	83.84745361	**95**	84.12441402
46	86.99464869	**96**	84.24819024
47	84.59602237	**97**	83.87495407
48	84.26213944	**98**	85.01332499
49	84.67028065	**99**	86.01859944
50	86.54405536	**100**	86.5584205

Cuadro C.1: Tiempos empleados ejecución Algoritmo Fuzzy C-Means sobre Bogotá D.C.,CO

Bogotá D.C., CO			
Gustafson Kessel			
1	89.94407347	**51**	90.75266177
2	88.75015329	**52**	90.36005763
3	91.32612442	**53**	89.94675125
4	88.95089772	**54**	89.90123075
5	88.72909532	**55**	91.43680196
6	91.12696927	**56**	90.11233908
7	88.4052828	**57**	88.06178481
8	88.70578735	**58**	88.4429127
9	88.99329788	**59**	89.0374178
10	91.2267299	**60**	88.22621438
11	88.24437285	**61**	88.67665287
12	89.06965599	**62**	91.802946
13	89.23360855	**63**	91.37021266
14	88.93656693	**64**	91.42506726
15	88.59747558	**65**	91.61894863
16	88.34214119	**66**	89.4989474
17	88.88103166	**67**	89.19582767
18	89.71190501	**68**	88.34937658
19	91.04079827	**69**	91.40460532
20	88.58673826	**70**	89.33309209
21	90.91057816	**71**	89.00869852
22	88.54717581	**72**	88.12050317
23	89.64598516	**73**	88.00003752
24	89.32728393	**74**	88.87378704
25	90.3787172	**75**	91.56577874
26	88.89419402	**76**	91.11025853
27	91.47660445	**77**	88.01575358
28	89.83918461	**78**	91.79648165
29	89.44324568	**79**	91.50064476
30	88.22460264	**80**	90.79031093
31	90.99999072	**81**	90.94828934
32	89.9432758	**82**	91.52603537
33	89.3356507	**83**	90.50386927
34	91.32570552	**84**	90.4945881
35	88.16882849	**85**	89.81452656
36	88.45315646	**86**	89.19131956
37	88.97735759	**87**	90.31240754
38	90.85215863	**88**	88.82777064
39	91.72452308	**89**	90.16518684
40	91.26962251	**90**	90.79444523
41	91.1786035	**91**	90.89723257
42	90.13918074	**92**	89.80502229
43	90.98211268	**93**	91.09523851
44	88.1766575	**94**	90.52837427
45	91.34184722	**95**	88.6111797
46	90.25503667	**96**	90.4319986
47	90.13376572	**97**	88.12990131
48	89.35110663	**98**	90.34258043
49	88.57058634	**99**	90.42411042
50	91.21320706	**100**	89.19196058

CUADRO C.2: Tiempos empleados ejecución Algoritmo Gustafson Kessel sobre Bogotá D.C.,CO

San Francisco,CA,USA			
Fuzzy C-Means			
1	54.967592	**51**	54.6786282
2	54.1595258	**52**	54.957429
3	54.9393501	**53**	55.309443
4	54.7352586	**54**	54.3564555
5	55.3649514	**55**	54.4467304
6	55.8473937	**56**	54.8511985
7	55.8842857	**57**	55.0257546
8	55.3751092	**58**	55.4892269
9	54.3630001	**59**	54.2865691
10	55.7193232	**60**	54.6434574
11	54.6860075	**61**	54.3320543
12	54.3755959	**62**	55.464879
13	55.2507196	**63**	55.9737538
14	54.6101684	**64**	55.9864756
15	55.2477726	**65**	54.1259801
16	54.0042113	**66**	55.5189816
17	55.53871	**67**	55.7576526
18	54.1261002	**68**	54.1486977
19	55.1703153	**69**	55.8788992
20	55.3760221	**70**	55.6836564
21	55.079532	**71**	54.4263558
22	54.093253	**72**	55.5099204
23	54.7588351	**73**	55.4901889
24	54.0637732	**74**	54.2015076
25	55.7934717	**75**	55.9475631
26	55.7593682	**76**	54.5298537
27	55.5857616	**77**	55.183032
28	54.4252703	**78**	55.6510703
29	54.1745445	**79**	54.916876
30	54.8002767	**80**	55.6003802
31	55.736658	**81**	54.5706343
32	55.4056702	**82**	54.9144948
33	55.5821749	**83**	54.1386826
34	54.2057638	**84**	54.9181774
35	55.2234952	**85**	55.0047908
36	54.4393902	**86**	55.6119829
37	54.7116754	**87**	55.8068032
38	55.2769039	**88**	54.7787056
39	55.1448935	**89**	54.0577527
40	55.3245008	**90**	54.3723984
41	54.2213368	**91**	54.722726
42	54.2051089	**92**	54.7995551
43	55.8239175	**93**	54.5646602
44	54.6673916	**94**	55.5143914
45	54.4593224	**95**	54.9467107
46	55.659513	**96**	54.0822313
47	55.508276	**97**	55.7457665
48	55.9078951	**98**	55.9195465
49	54.8746025	**99**	54.7168211
50	55.6759761	**100**	54.6288522

CUADRO C.3: Tiempos empleados ejecución Algoritmo Fuzzy C-Means sobre San Francisco,CA,USA

San Francisco,CA,USA Gustafson Kessel			
1	64.1850888	**51**	64.1326706
2	64.2553733	**52**	64.3786755
3	65.7456784	**53**	65.0900958
4	64.3067095	**54**	65.5537492
5	64.69854	**55**	65.8039816
6	65.1550701	**56**	65.2232095
7	64.0478085	**57**	65.4758671
8	65.9800251	**58**	65.264717
9	65.7348541	**59**	64.080437
10	64.0591428	**60**	65.939951
11	65.548919	**61**	64.5489685
12	64.3740577	**62**	65.5553935
13	65.7308586	**63**	64.4541952
14	64.2162613	**64**	64.2107489
15	64.5702214	**65**	65.7295384
16	65.2793418	**66**	65.7626057
17	64.4420445	**67**	65.1999429
18	64.3265439	**68**	64.5156042
19	65.7170318	**69**	65.5969152
20	64.1147791	**70**	64.8846772
21	65.7354612	**71**	65.4668452
22	64.757036	**72**	65.284996
23	65.4018108	**73**	65.5509435
24	65.4458539	**74**	64.0518302
25	65.9392362	**75**	65.9825749
26	64.2716239	**76**	65.2250753
27	65.1611002	**77**	65.1292992
28	65.5235268	**78**	64.0799484
29	64.9232283	**79**	64.926189
30	65.0412288	**80**	64.4489943
31	65.6542603	**81**	64.1311781
32	64.4209601	**82**	64.000278
33	65.3662184	**83**	65.8295325
34	64.687362	**84**	65.2048849
35	64.3141042	**85**	65.7875243
36	65.0427269	**86**	64.044931
37	64.8212412	**87**	64.6149899
38	64.7141278	**88**	64.1619249
39	64.5361741	**89**	64.621769
40	65.9551254	**90**	65.8041337
41	64.8736578	**91**	64.4272097
42	65.1817749	**92**	65.8443439
43	65.1565214	**93**	65.092674
44	64.6880202	**94**	64.9856736
45	64.5997695	**95**	65.8030456
46	65.0277899	**96**	64.4865156
47	65.3846479	**97**	65.3656134
48	65.7671215	**98**	64.1724244
49	64.0083274	**99**	64.2799606
50	64.5268262	**100**	65.3581054

CUADRO C.4: Tiempos empleados ejecución Algoritmo Gustafson Kessel sobre San Francisco,CA,USA

Bibliografía

[1] Olaya C. La irrelevancia del naturalismo criminológico: una invitación a la ingeniería. *Notas Criminológicas*, pages 6–9, 2014.

[2] Camara de Comercio de Bogota. Encuesta de percepción y victimización: Bogotá primer semestre 2015. 2015.

[3] González A. Campoy P. El modelo lotka-volterra y el çontrol-crime": policia y prevención del delito. *Notas Criminológicas*, pages 11–14, 2014.

[4] Ratcliffe J. Crime mapping: Spatial and temporal challenges. *Weisburd D, Piquero A (eds) Handbook of quantitative criminology, Springer, New York*, pages 5–24, 2010.

[5] T. H. Grubesic. On the application of fuzzy clustering for crime hot spot detection. *J. Quant. Criminol*, 22(1):77–105, 2006.

[6] Ord. J. K. Getis, A. The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24(3):189–206, 1992.

[7] Gotway C. A. Waller, L. A. Applied spatial statistics for public health data. *New York: Wiley*, 2004.

[8] Rousseeuw P. J. Kaufman, L. Finding groups in data: An introduction to cluster analysis. *New York: Wiley-Interscience*, 2005.

[9] Murray A. T. Grubesic T. H., Wei R. Spatial clustering overview and comparison: Accuracy, sensitivity, and computational expense. *Ann. Assoc. Am. Geogr.*, 104(6):1134–1156, 2014.

[10] Bertozzi L. Brantingham P. J. Woodworth J. T., Mohler G. O. Nonlocal crime density estimation incorporating housing information. *Philos. Trans. A. Math. Phys. Eng. Sci.*, 372:1–15, 2014.

[11] **L. Smith. Incorporating spatial information into density estimates and street gang models. page 127, 2012.**

[12] **Andresen M. Malleson N. Spatio-temporal crime hotspots and the ambient population. *Crime Sci*, 4(1):1–8, 2015.**

[13] **M. R. Nobles C. A. Franklin, T. W. Franklin and G. A. Kercher. Assessing the effect of routine activity theory and self-control on property, personal, and sexual assault victimization. *Crim. Justice Behav*, 39(10):1296–1315, Oct 2012.**

[14] **R. Boba. Crime analysis and crime mapping. california, e.u.a. *Sage Publications Inc.*, 2005.**

[15] **M. Goodchild. Towards an enumeration and classification of gis. proceedings. *CIS 87: the Research Agenda*, (11):67–77, 1987.**

[16] **La Vigne N. G. Groff, E. R. Forecasting the future of predictive crime mapping. *Crime Prevention Studies*, (13):29–58, 2002.**

[17] **Zadeh L. A. Fuzzy sets. *Information and Control*, (8), 1965.**

[18] **H. J. Zimmermann. Fuzzy set theory—and its applications. Springer Science Business Media, 2001.**

[19] **R. Kruse T. Runkler. F. Hoppner, F. Klawonn. Fuzzy cluster analysis: Methods for classification, data analysis and image recognition. *Wiley, Chichester*, 1999.**

[20] **J. C. Bezdek. Cluster validity with fuzzy sets. *Journal of Cybernetics*, (3):58–73, 1974a.**

[21] **J. C. Bezdek. Numerical taxonomy with fuzzy sets. *Journal of Mathematical Biology*, (1):57–71, 1974b.**

[22] **Bezdek J. C. Pattern recognition with fuzzy objective function algorithms. *Ed. Plenum Press*, 1987.**

[23] **Sala A. Díez J., Navarro J. Algoritmos de agrupamiento en la identifiacion de modelos borrosos. *Revista Iberoamericana de Automática e Informática Industrial*, 1(2):32–41, 2004.**

[24] **Kessel W. C. Gustafson E. E. Fuzzy clustering with a fuzzy covariance matrix. *IEEE CDC, San Diego, California*, pages 761–766, 1979.**

[25] Navarro J. L. Díez J. L. Fuzzy models of complex systems by means of clustering techniques. ***Proc. 2nd Intelligent Systems in Control and Measurement***, pages 147–153, 1999.

[26] Bezdek J. C. Hathaway R. J. Switching regression models and fuzzy clustering. ***Transactions on Fuzzy Systems***, 1(3):195–204, 1993.

[27] Nakamori Y. Ryoke M. Simultaneous analysis of classification and regression by adaptive fuzzy clustering. ***Japanese Journal of Fuzzy Theory and Systems***, 8(1):99–113, 1996.

[28] Raghavan P. Schütze H. Manning C. Introduction to information retrieval. Cambridge University Press.

[29] Estivill V. Why so many clustering algorithms — a position paper. ***ACM SIGKDD Explorations Newsletter***, 4:65–75, 2002.

[30] Bouldin D. Davies D. A cluster separation measure. ***IEEE Transactions on Pattern Analysis and Machine Intelligence***, page 224–227, 1979.

[31] J. Dunn. Well separated clusters and optimal fuzzy partitions. ***Journal of Cybernetics***, 4:95–104, 1974.

[32] Kriegel H. Kröger P. Müller E. Schubert E. Seidl T. Zimek A. Färber I., Günnemann S. On using class-labels in evaluation of clusterings. 2010.

[33] Moulavi D. Campello R. J. G. B. Zimek A. Sander J. Goebel R. Pourrajabi, M. Model selection for semi-supervised clustering. ***Proceedings of the 17th International Conference on Extending Database Technology***, page 331–342, 2014.

[34] Rand W. M. Objective criteria for the evaluation of clustering methods. ***Journal of the American Statistical Association (American Statistical Association)***, 66:846–850, 1971.

[35] Jaccard P. Distribution de la flore alpine dans le bassin des dranses et dans quelques régions voisines. ***Bulletin de la Société Vaudoise des Sciences Naturelles***, 37:241–272, 1901.

[36] The probabilistic basis of jaccard's index of similarity. ***Systematic biology***, 45(3):380–385, 1996.

[37] **Mallows C. L. Fowlkes E. B. A method for comparing two hierarchical clusterings. *Journal of the American Statistical Association*, page 553–569.**

[38] **Bezdek J. Clarke L. Silbiger M. Arrington J. Murtagh R. Bensaid A., Hall L. Validity-guided (re)clustering with applications to imige segmentation. *IEEE Transactions on Fuzzy Systems*, 4: 112–123, 1996.**

[39] **Data sf, bay area cities. urlhttps://data.sfgov.org/Government/Bay-Area-Cities-Zipped-Shapefile-Format-/nghj-u9xk/data.**

[40] **Unidad administrativa especial de catastro distrital. urlhttps://www.ideca.gov.co/es/servicios/mapa-de-referencia.**

[41] **F. F. Paz. Seguridad y política criminal. 5 2016. URL** `http://www.ideaspaz.org/foundation/work-areas/6`.

[42] **American Cartographic Association. Geographers and cartographers urge end to popular use of rectangular maps. *American Cartographer*, 16(3):222–223, 1989.**

[43] **Henzgen S. Senge R. Hüllermeier E., Rifqi M. Comparing fuzzy partitions: A generalization of the rand index and related measures. *IEEE TRANSACTIONS ON FUZZY SYSTEMS*, 20(3):546–556, 6 2012.**

[44] **Brouwer R. Extending the rand, adjusted rand, and jaccard indices to fuzzy partitions. *J. Intell. Inf. Syst.*, 32, 2009.**

[45] **Melgarejo M. Documento, código en matlab para la expansión del Índice rand para agrupamientos difusos. 2017.**