

Evaluación de un algoritmo de agrupamiento de flujo de datos para el análisis de trayectorias GPS

Evaluation of a data flow clustering algorithm for GPS trajectory analysis

Reyes Zambrano, Gary; Zambrano Meza, Edison Steeven; León Granizo, Oscar; Crespo León, Christopher



Gary Reyes Zambrano

gary.reyesz@ug.edu.ec

Universidad de Guayaquil, Ecuador

Edison Steeven Zambrano Meza

edisson.zambranom@ug.edu.ec

Universidad de Guayaquil, Ecuador

Oscar León Granizo

oscar.leong@ug.edu.ec

Universidad de Guayaquil, Ecuador

Christopher Crespo León

christopher.crespol@ug.edu.ec

Universidad de Guayaquil, Ecuador

Ecuadorian Science Journal

GDEON, Ecuador

ISSN-e: 2602-8077

Periodicidad: Semestral

vol. 6, núm. 2, 2022

esj@gdeon.org

Recepción: 22 Marzo 2022

Aprobación: 02 Septiembre 2022

URL: <http://portal.amelica.org/ameli/journal/606/6063133004/>

DOI: <https://doi.org/10.46480/esj.6.2.99>

Los autores mantienen los derechos sobre los artículos y por tanto son libres de compartir, copiar, distribuir, ejecutar y comunicar públicamente la obra sus sitios web personales o en depósitos institucionales, después de su publicación en esta revista, siempre y cuando proporcionen información bibliográfica que acredite su publicación en esta revista.



Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-SinDerivar 4.0 Internacional.

Cómo citar: Reyes Zambrano, G., Zambrano Meza, E. S., León-Granizo, O., & Crespo León, C. (2023). Evaluación de un algoritmo de agrupamiento de flujo de datos para el análisis de trayectorias GPS. *Ecuadorian Science Journal*, 6(2), 33-45. <https://doi.org/10.46480/esj.6.2.99>

Resumen: Contexto. El gran volumen de datos que hoy en día generan los diversos dispositivos GPS, originan que el análisis de esos datos sea un campo de investigación de interés actual. Los algoritmos de agrupación identifican patrones en un conjunto masivo de datos. En el ámbito de la transportación en el caso de los vehículos que circulan por las calles de una ciudad, estos algoritmos ayudan a identificar congestionamientos, flujos de tráfico comunes, eventos frecuentes como paradas que realizan los vehículos, entre otros. Una trayectoria GPS está definida por un conjunto de ubicaciones geográficas, cada una de las cuales se encuentra representada por su latitud y longitud, en un instante de tiempo. Existen trayectorias GPS vehiculares que hoy en día son generadas por los sistemas inteligentes de transportación en las ciudades. Este trabajo evalúa un algoritmo de agrupación de flujo de datos, llamado dylee en el procesamiento y análisis de dos conjuntos de datos de las ciudades de Guayaquil-Ecuador y Roma-Italia. Los resultados que se obtuvieron de la evaluación de los dos conjuntos de datos utilizados permiten identificar patrones vehiculares que se presentan en distintos instantes de tiempo, pudiéndose evidenciar rangos de velocidades comunes. Además, se realizan mediciones para determinar la calidad de los grupos resultantes, los resultados obtenidos son satisfactorios.

Palabras clave: Clustering, Trayectorias GPS, Algoritmos, Flujo de datos.

Abstract: Context: The large volume of data generated today by various GPS devices makes the analysis of this data a field of current research interest. Clustering algorithms identify patterns in a massive set of data. In the field of transportation, in the case of vehicles circulating through the streets of a city, these algorithms help to identify congestion, common traffic flows, frequent events such as vehicle stops, among others. A GPS trajectory is defined by a set of geographic locations, each of which is represented by its latitude and longitude at an instant in time. There are vehicular GPS trajectories that are nowadays generated by intelligent transportation systems in cities. This work evaluates a data flow clustering algorithm called Dylee in the processing and analysis of two data sets from the cities of Guayaquil-Ecuador and Rome-Italy. The results obtained from the evaluation of the two data sets using Dylee allow the identification of vehicular patterns that occur at different time

instants, being able to show common speed ranges. In addition, measurements are made to determine the quality of the resulting groups, the results obtained are satisfactory.

Keywords: Clustering, GPS Trajectories, Algorithms, Data flow.

INTRODUCCIÓN

El avance de la tecnología ha proporcionado muchas comodidades para la sociedad en distintos ámbitos y campos técnicos (Suasnabas et al., 2016). Actualmente se cuenta con tecnologías para facilitar la recolección de información, como los dispositivos del sistema de posicionamiento global (GPS), (Juca Maldonado & Burgo Bencomo, 2016) que se puede encontrar múltiples artefactos de manera cotidiana o en el ámbito profesional, comúnmente vistos en vehículos, Smartphone, relojes inteligentes y entre otros, donde permite seguir sus trayectorias (Tusa et al., 2016). Estos instrumentos suministran grandes cantidades volúmenes de registros a las bases de datos, donde se adquiere información geoespacial (Reyes Zambrano & Cedeño Pozo, 2018).

Algoritmos de agrupamiento

El agrupamiento de datos es una técnica muy utilizada a la hora de identificar características comunes entre instancias de un mismo problema (Madhulatha, 2012). Actualmente las técnicas de agrupamiento brindan soluciones identificando grupos de datos en áreas muy diversas, tales como salud, finanzas, telecomunicaciones, agricultura y transporte, entre otras (A. Jain, 2009).

A lo largo del tiempo investigadores en todo el mundo han propuesto mejoras a las limitantes identificadas en algunas técnicas como en (Bindiya M Varghese, 2013) en otros casos se han mejorado las técnicas para que funcionen en un contexto específico como, por ejemplo, para la minería de datos espacial (Tork, 2012) o para el análisis de trayectorias GPS (Mazimpaka & Timpf, 2016).

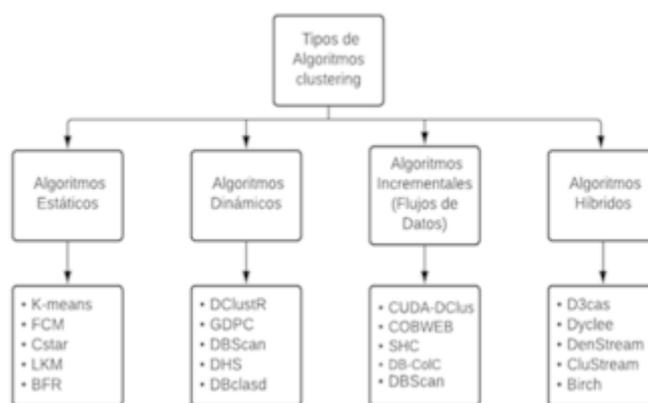


FIGURA 1.
Taxonomía de los tipos de algoritmos clustering.

Fuente: Autores.

En la figura 1 se muestran una taxonomía de los tipos de algoritmos clustering, existen muchas clasificaciones en la literatura revisada, pudiendo utilizarse cualquier clasificación dependiendo del tipo de investigación a realizarse.

Trayectorias

Una trayectoria GPS está definida por un conjunto de ubicaciones geográficas, cada una de las cuales se encuentra representada por su latitud y longitud, en un instante de tiempo. Es un conjunto de puntos que representan el espacio-tiempo, quedando definido como se observa en la ecuación 1:

$$\text{puntos } (p_0, p_1, \dots, p_n), \text{ donde } p_i = (x_i, y_i, t_i) \text{ y } x_i, y_i, t_i \in R$$

$$\text{desde } i = 0, \dots, N \text{ y } t_0 < t_1 < \dots < t_n$$

Cada punto posee tres valores, que son: latitud (x), longitud (y) y tiempo (t). Al tener un conjunto de puntos, el trazo sobre estos en orden de tiempo (t) es el correspondiente a la trayectoria del objeto dado. Cada punto puede tener un conjunto de valores adicionales según sea necesario, tales como velocidad, altura o estado que presenta el objeto en el tiempo (t) que se obtiene o analiza de la información de este, en la figura 2 Se representa una trayectoria en un espacio 2D.

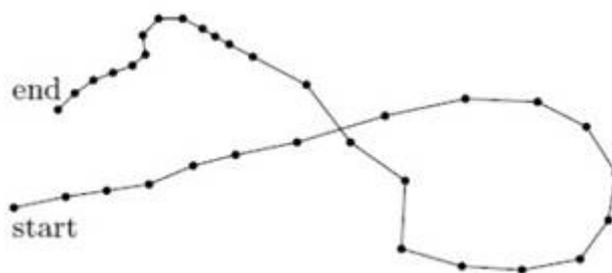


FIGURA 2.

Representación de una trayectoria en un espacio 2D como puntos a intervalos de tiempo iguales.

Fuente: Autores.

Como podemos observar se representa una trayectoria en un espacio 2D como puntos a intervalos de tiempos iguales, notándose la existencia de un punto inicial (start) y de un punto final (end) los cuales limitan la trayectoria.

Algoritmos de agrupamiento de trayectorias GPS

Los algoritmos de agrupamiento son fundamentales en el momento de definir el enfoque de una investigación, en este trabajo el enfoque es sobre agrupamiento de trayectorias vehiculares. De acuerdo a lo interpretado por (Vera, 2018) “clustering es una técnica no supervisada que tiene como tarea generar agrupaciones de acuerdo a la conexión de las semejanzas o diferencia que existen entre los objetos que se analizan.

En la literatura se han identificado una serie de soluciones que realizan agrupamiento de trayectorias. Estas soluciones mejoran algoritmos tradicionales de agrupamiento adaptándolos a un contexto en particular o aplicabilidad y utilizando medidas de similitud entre trayectorias no convencionales (C. Riquelme, J., Ruiz, R., & Gilbert, K.,2006).

(Yuan & Raubal, 2014) Adicionalmente combina dos paradas consecutivas con la misma ubicación geográfica y un intervalo temporal corto, haciendo uso del método cuantitativo de Yuan (Laohakiat, S., & Sa-ing, V.,2021).

En el algoritmo Tra-DBScan (L. X. Liu, J. T. Song, B. Guan, Z. X. Wu, and K. J., 2012) se utiliza el algoritmo DBScan, haciendo uso de una fase de segmentación de trayectorias, realizando una partición de

las trayectorias en tramos y utilizando la distancia de Hausdorff como medida de similitud. Esta métrica considera la distancia perpendicular, angular y paralela de los distintos puntos GPS que son parte de los segmentos de las trayectorias.

Por otra parte, en (Ferreira et al., 2013) se presenta una nueva técnica de agrupamiento de trayectorias que utiliza campos vectoriales para representar los centros de los grupos y proponer una definición de similitud entre trayectorias. El campo vectorial k-means, aunque recuerda al algoritmo k-means se trata de una formulación nueva para la agrupación de trayectorias, que puede encontrar patrones globales que no son aparentes a nivel local en trayectorias ruidosas o incluso trayectorias parciales, además de ser altamente paralelizable lo que le permite fácilmente escalar a grandes conjuntos de datos.

En (Q. Yu, Y. Luo, C. Chen, and S. Chen, 2019) se propone un modelo de trayectoria mejorado y se presenta un nuevo algoritmo de agrupación de trayectorias, con una medida de similitud de trayectorias que calcula la distancia entre dos trayectorias basado en múltiples características de la trayectoria, logrando maximizar la similitud de las trayectorias en el mismo grupo.

Hoy en día continúan los esfuerzos de investigación como se evidencia en (Y. Yang, J. Cai, H. Yang, J. Zhang, and X. Zhao, 2020) que propone un nuevo algoritmo de agrupamiento de trayectoria (TAD), entre sus principales ventajas están las siguientes: Introduce un factor de tolerancia al ruido, que puede evaluar dinámicamente y reducir el impacto del ruido en el proceso de ejecución del algoritmo; Integra las características de tiempo y espacio en una misma función de medición para mejorar la precisión del agrupamiento; Construye una nueva función de densidad, que distingue con mayor precisión diferentes tipos de puntos en trayectorias y encuentra estancias de trayectorias basada en análisis de datos de densidad espacio-temporal. Dos nuevas métricas: NMAST (Capacidad de movimiento y tiempo de permanencia en el vecindario) y un factor NT (tolerancia de ruido) se definen en este algoritmo.

En (Reyes et al., 2020) se define un método de segmentación de trayectorias GPS que utiliza el algoritmo Kmeans para el agrupamiento. El método tiene un componente de segmentación de trayectorias y un componente de agrupamiento de sub-trayectorias. La primera divide las trayectorias GPS en sub-trayectorias caracterizadas por no presentar cambios de dirección significativos, lo que favorece su análisis.

Algoritmos de agrupamiento de trayectorias GPS que procesan flujos de datos

Una de las aplicabilidades importantes al momento de analizar flujo de tráfico en una ciudad, es identificar patrones dentro agrupaciones de trayectoria de manera incremental, ya que estos algoritmos obtienen grandes cantidades de información. El aprendizaje incremental se refiere a las estrategias de aprendizaje en línea que funcionan con recursos de memoria limitados. (Gepperth & Hammer, 2016) Estos volúmenes de datos se denominan "flujos de datos" (data streams), porque se generan de forma secuencial a gran velocidad. (Molina & Hasperué, 2018).

Algoritmo de Clustering de flujos de datos a evaluar

Esta investigación evalúa un algoritmo de clustering llamado Dyclee, este algoritmo tiene la capacidad de procesar flujos de datos en entornos dinámicos. Este algoritmo está conformado por dos etapas, una primera etapa donde se realiza un agrupamiento por distancia y una segunda etapa donde se realiza un agrupamiento por densidad. La lectura de datos se puede realizar de manera incremental (flujos de datos) permitiendo obtener agrupamientos dinámicos en cada evolución.

El algoritmo Dyclee fue implementado utilizando el lenguaje de programación R adaptando la lectura y procesamiento de trayectorias y el agrupamiento por la dimensión de la velocidad. Se realizaron mediciones en 8 instantes de tiempo secuenciales que para propósitos de nuestro trabajo denominamos evoluciones,

logrando obtener rangos comunes de velocidades. De la misma manera se realizaron mediciones de dos métricas de calidad para agrupamientos: Silhouette y Davies Bouldin y los resultados fueron satisfactorios.

MATERIALES Y MÉTODOS

En éste trabajo se utilizó el algoritmo Dyclee para procesar diferentes conjuntos de datos que corresponden a trayectorias GPS para lograr identificar patrones de velocidad. Para ello la metodología utilizada permitió definir dos pasos:

1. La definición de celdas para el área de procesamiento.
La adaptación para que el algoritmo Dyclee procese trayectorias y realice el agrupamiento por la dimensión de velocidad.

El algoritmo de DyClee procesa datos de manera dinámica, en el caso de este algoritmo se optó por realizar flujos de datos los cuales se los denominarán evoluciones que ingresarán al procesamiento a partir de determinados periodos de tiempo, se utiliza por defecto el valor de 3, este valor está en unidades de minutos e indica la duración que tendrá cada evolución y los datos que ingresarán corresponderán a datos que se encuentren en ese rango de minutos de acuerdo a cada evolución que se realizarán de manera secuencial.

Definición de celdas para el área de procesamiento

Para poder enfocar los análisis de trayectorias en determinadas ubicaciones en una ciudad se realizó el trazo de celdas dentro del área a procesar. Para ello, tras haber obtenido la cantidad de ejecuciones necesarias para abarcar todo el conjunto de datos, se puede empezar a definir las celdas y el área que ocupará cada celda, tanto para el largo como el ancho, usando la proporción de $n/10000$ en donde n será la longitud de los lados de cada celda como se observa en la ecuación 2.

Valor de la longitud a usar= $n/(100\ 000)=200/100000=0.002$ (2)

Por ejemplo, el siguiente cálculo define el valor a utilizar para un largo de 200 metros y un ancho de 200 metros: El resultado es 0.002 que representa un largo de 200 metros y un ancho de 200 metros por lo que el área definida quedaría de tamaño 200x200 m², si se traslada esta área a un mapa y se visualiza, se puede visualizar el área que ocupará cada celda, como se observa en la figura 3.

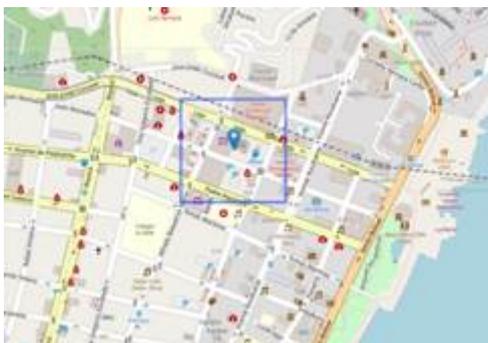


FIGURA. 3.
Área de procesamiento
Fuente: Autores.

Una vez que se definió el tamaño de la celda dentro del área de procesamiento, se implementó el algoritmo a evaluar, llamado Dyclee, realizando su adaptación para el procesamiento y análisis de trayectorias GPS.

Algoritmo Dyclee adaptado

Para poder procesar las trayectorias GPS se realizó una adaptación del algoritmo DyClee. En la tabla 1 se identifican algunas definiciones de Dyclee y que fueron utilizadas en la adaptación.

TABLA 1.
Algunas definiciones usadas por DyClee

Elementos	Definición
Micro agrupamiento	Un μ -clúster (μC) es un hipercuadro d-dimensional que representa el conjunto de muestras de datos que se asignan a él desde su momento de su creación.
Hipercaja	La estructura de DyClee se basa en hipercajas que sirven como representaciones geométricas del conjunto de muestras de datos, es utilizada para determinar el área que existe entre un microclúster y/o algún punto espacial
Densidad de un grupo	La densidad del μ -cluster se denota ρ y corresponde al número de muestras de datos que se le asignan sobre su hiper volumen V
Centroide	Es el punto central de cualquier sistema geométrico, en DyClee se lo utiliza para definir el punto central en donde coinciden cada uno de los elementos que ingresan al algoritmo a procesarse y ser formados por algún microclúster.

Fuente: Autores.

Para este trabajo se tuvo que modificar el parámetro que se menciona a continuación:

RelativeSize: Especifica el tamaño relativo que alcanzarán los μC con respecto a la Hipercaja. Permite crear microclusters de diferentes densidades y a su vez permite que se reconozcan regiones de bajas densidades que se encuentran conectadas. Debe tomar valores entre 0 y 1; en donde los valores cercanos a 1 creará microclústers de tamaños cercanos a la hipercaja y si es cercano a 0, tomará tamaños más pequeños. El tamaño de los microclusters se establece como una fracción del rango de la Hipercaja. Este parámetro es obligatorio, es definido por el usuario para DyClee y se ajusta de forma experimental.

En este artículo, en la primera etapa del algoritmo, en vez de utilizar directamente las ubicaciones GPS y su densidad, se consideraron las velocidades de los tramos de trayectorias que se encuentran comprendidos en cada celda. De esta forma, para un periodo de tiempo dado, cada celda será representada por la velocidad promedio de los tramos de trayectorias que contiene. Esto implica recortar las trayectorias adecuadamente considerando que las variaciones de velocidad pueden llevar a que un vehículo circulando a muy alta velocidad no sea registrado (o posea muy pocas ubicaciones GPS) al pasar por una celda. Además, es importante considerar que las velocidades deben ser promediadas considerando los vehículos y no la cantidad de ubicaciones registradas. En lo referido al tamaño de los microclusters, el valor del parámetro "Relative Size" especifica el tamaño relativo del parámetro "HiperBox" con respecto al área a procesar. Es decir que a medida que su valor se reduce la cantidad de microclusters aumenta y viceversa.

En su segunda etapa, el algoritmo analiza las densidades de los microclústers conformados y los clasifica en dos categorías: densos y semidensos. A partir de los microclusters densos comienza a unir aquellos que

se encuentren directamente conectados. Dos microclustres se considerarán directamente conectados si la distancia máxima a la que pueden encontrarse los centroides de dos microclusters, con densidades similares, no supera el valor del parámetro "HiperBox". Es decir que, a medida que se incrementa el valor del parámetro "HiperBox", la cantidad de clusters será menor y, por lo tanto, los rangos de velocidades en el área de interés tendrán mayor amplitud.

Conjunto de datos utilizados

Los conjuntos de datos a utilizar son muestras representativas de los conjuntos de datos de trayectorias GPS de Guayaquil-Ecuador y de Roma-Italia. Para el conjunto de datos de Guayaquil dado que se trata de una parte del conjunto de trayectorias, para efectos de análisis se seleccionaron los registros en un horario comprendido entre las 16:30 hasta las 18:30 horas Como resultado de este proceso de filtrado se obtuvieron 29503 registros que representan 206 trayectorias de todo el conjunto de datos. Para el conjunto de datos de Roma, se efectuó el análisis en el horario comprendido entre las 18:00 hasta las 20:00 horas. Como resultado de este proceso de filtrado se obtuvieron 33793 registros que representan 137 trayectorias de todo el conjunto de datos.

TABLA 2.
Cantidad de registros en los conjuntos de datos

Conjunto de datos	Cantidad de trayectorias	Cantidad de puntos o registros
Guayaquil	206	29503
Roma	137	33793

Fuente: Autores.

En la Tabla 2, se puede visualizar los conjuntos de datos con su correspondiente cantidad de trayectoria y la cantidad de puntos o registros generados y utilizados, en los experimentos diseñados.

Los conjuntos de datos tienen similares características, las columnas utilizadas se observan en la tabla 3.

TABLA 3.
Columnas que contiene los conjuntos de datos

Columnas	Descripción
Latitud	Representa la latitud para formar nuestra coordenada geográfica de la trayectoria.
Longitud	Representa la longitud para formar nuestra coordenada geográfica de la trayectoria.
Tiempo	Fecha y hora de la recolección de cada registro.
ID	Identificador de los conjuntos de clúster.

Fuente: Autores.

Definición de experimentos

Para validar este trabajo se definieron dos experimentos con cada uno de los conjuntos de datos a utilizar. En cada experimento se realizan las mismas parametrizaciones iniciales con respecto al relative size, hipercaja, período de tiempo, evoluciones. Para cada período de tiempo consecutivo, evoluciones, se calcula las velocidades mínimas, máximas y promedio para realizar una interpretación de los rangos de velocidades. También por cada evolución se obtiene la cantidad de grupos densos, semi-densos y poco densos, así como también los indicadores de silhouette y Davies bouldin que indican la calidad de los grupos.

RESULTADOS Y DISCUSIÓN

Experimento 1. Algoritmo DyClee - Roma

La configuración de DyClee se realiza mediante parámetros necesarios. El valor para el parámetro "Relative Size" fue de 0.2, a partir de este valor y en base al área de procesamiento se define el valor de las dimensiones de la "hipercaja". Se dividieron los 30557 registros del conjunto de datos de Roma-Italia en 8 bloques de 15 minutos cada uno y se los analizó en forma consecutiva.

Como podemos observar en la tabla 4 la información representada para cada celda consiste principalmente en la velocidad promedio de los vehículos registrados en su interior durante el periodo de tiempo analizado. Para cada grupo se indican sus velocidades mínimas, máxima y promedio, así como el desvío de las velocidades pertenecientes a dicho grupo. Existen 8 evoluciones para el agrupamiento realizado con la adaptación del algoritmo DyClee de las cuales podemos resaltar que la velocidad promedio más alta se encuentra en la evolución 7 con un valor de 96,35 esta velocidad se da en un grupo atípico el cual está representado con valores negativo, por otra parte, la velocidad promedio más baja se encuentra en la evolución 6 con un valor de 1,71 (Ver Tabla 4).

En la Tabla 5 podremos observar los resultados de algunos indicadores de agrupamiento evaluados durante las 8 evoluciones, al final de la tabla se observa las respectivas sumatorias y promedios. A lo largo de las evoluciones el algoritmo obtiene 5 grupos densos-semidensos de forma simultánea. También se destaca que en la evolución 1, 4 y la 5 se obtienen 4 grupos pocos densos o atípicos mientras que en la evolución 2, 3, 6, 7 y 8 se obtienen 5 grupos pocos densos o atípicos, dando como sumatoria 37 grupos poco densos. La evolución con más puntos en grupos densos es la 7 con un valor de 4766 con respecto a otras que obtuvieron un puntaje relativamente menor. El valor de Silhouette más alto en la primera etapa se ubica en la evolución 4 con 0,76060, sin embargo, el valor más alto de la segunda etapa es de 0,64525 y se ubica en la primera evolución. El valor más alto de Davies Bouldin etapa 1 está ubicado en la evolución 2 con un valor de 0,46921 y de la etapa 2 se ubica en la misma evolución 2 con un valor de 0,58556 (Ver Tabla 5).

Experimento 2. Algoritmo DyClee – Guayaquil

La configuración de DyClee utilizada para el experimento con el conjunto de datos de Guayaquil es similar a la utilizada para el conjunto de datos de Roma.

Como podemos observar en la tabla 6 la información representada para cada celda consiste principalmente en la velocidad promedio de los vehículos registrados en su interior durante el periodo de tiempo analizado. Para cada grupo se indican sus velocidades mínimas, máxima y promedio, así como el desvío de las velocidades pertenecientes a dicho grupo. Existen 8 evoluciones para el agrupamiento realizado con la adaptación del algoritmo DyClee de las cuales podemos resaltar que la velocidad promedio más alta se encuentra en la evolución 5 con un valor de 106,57 esta velocidad se da en un grupo atípico el cual está representado con

valores negativo, por otra parte, la velocidad promedio más bajo se encuentra en la misma evolución 5 con un valor de 4,87 (Ver Tabla 6).

En la tabla 7 podremos observar los resultados de algunos indicadores de agrupamiento evaluados durante las 8 evoluciones. Las evoluciones 1, 2 y 7 tiene 4 grupos densos, las evoluciones 3 y 6 tienen 5 grupos densos, la evolución 4 tiene 6 grupos, la evolución 5 tiene 7 grupos y la evolución 8 tiene 3 grupos totalizando 38 grupos densos-semidensos. También se destaca que en la evolución 2, 3, 6 y 7 se obtienen 4 grupos pocos densos o atípicos, mientras que en la evolución 1 tiene 2 grupos, la evolución 4 tiene 5, la evolución 5 tiene 6 y la evolución 8 tiene 3 totalizando 32 grupos densos o atípicos. La evolución más alta de puntos en grupos densos es la 4 con un valor de 7983 con respecto a otras que obtuvieron un puntaje relativamente menor. El valor de Silhouette más alto en la primera etapa se ubica en la evolución 1 con un valor de 0,56771, de igual manera, el valor más alto de la segunda etapa se ubica en la misma evolución es de 0,71126. Por otro lado, el valor más alto de Davies Bouldin etapa 1 está ubicado en la evolución 4 con un valor de 0,46138 y de la etapa 2 se ubica en la evolución 6 con un valor de 0,60718 (Ver Tabla 7).

TABLA. 4.
Resultados obtenidos de los experimentos en el conjunto de datos de Roma en 8 periodos consecutivos de tiempo.

Evolución 1							Evolución 2						
Gru	Cant.		Vel	Vel.	Vel.		Grupo	Cant.	Cant.	Vel	Vel.	Vel.	
po	Puntos	CantVeh	Min	Mas	prom	Desv	Puntos	Veh	Min	Mas	prom	Desv	
2	1486	43	0	86	2,85	6,79	1	2079	45	0	92	2,38	6,42
1	849	41	0	111	17,42	14,83	2	1033	49	0	106	16,24	13,49
3	247	31	0	121	29,11	17,33	3	428	43	0	104	27,12	16,27
4	63	16	0	111	41	22,48	-4	121	27	0	129	58,11	30,82
-5	35	11	9	113	69,9	27,11							
Evolución 3							Evolución 4						
Gru	Cant.		Vel	Vel.	Vel.		Gru	Cant.	Cant.	Vel	Vel.	Vel.	
po	Puntos	CantVeh	Min	Mas	prom	Desv	po	Puntos	Veh	Min	Mas	prom	Desv
1	2809	49	0	58	2,77	6,66	2	2253	48	0	56	2,88	6,95
2	999	44	0	98	20,3	13,92	1	885	49	0	120	20,42	14,56
4	174	24	0	128	33,05	18,82	3	235	32	0	128	35,08	18,36
3	88	17	0	122	45,17	25,04	4	13	3	33	90	59,46	17,56
-4	20	7	4	126	69,75	29,75	-5	12	6	34	118	87,96	40,19
Evolución 5							Evolución 6						
Gru	Cant.	Cant.	Vel	Vel.	Vel.		Gru	Cant.	Cant.	Vel	Vel.	Vel.	
po	Puntos	Veh	Min	Mas	prom	Desv	po	Puntos	Veh	Min	Mas	prom	Desv
2	2680	55	0	76	2,1	5,66	4	2057	53	0	54	1,71	4,75
3	949	53	0	88	15	13,28	2	1308	61	0	94	13,46	12,32
1	503	44	0	95	26,37	17	3	554	55	0	129	25,12	7,35
4	133	23	0	128	39,44	23,45	1	264	40	0	123	36,22	23,29
-5	67	15	0	119	72,61	26,37	-5	54	16	0	126	84,79	14,8
Evolución 7							Evolución 8						
Gru	Cant.	Cant.	Vel	Vel.	Vel.		Gru	Cant.	Cant.	Vel	Vel.	Vel.	
po	Puntos	Veh	Min	Mas	prom	Desv	po	Puntos	Veh	Min	Mas	prom	Desv
2	2571	55	0	57	1,76	5,09	1	5091	68	0	94	1,89	5,33
1	1218	58	0	111	15,64	13,77	2	1341	58	0	129	17,83	14,09
3	542	49	0	117	28,36	17,82	3	426	49	0	101	32,84	18,21
5	45	11	0	106	40,49	24,88		83	24	0	129	66,68	24,85
4	68	17	0	124	50	27,85							
-6	9	7	29	128	96,35	16,8							

Fuente: Autores.

TABLA 5.
Algoritmo Dyclee - Roma

	Evo. 1	Evo. 2	Evo. 3	Evo. 4	Evo. 5	Evo. 6	Evo. 7	Evo. 8	Sumatoria	Promedios
Trayectorias	137									
Puntos en consulta	33793									
Puntos a procesar	4055	3861	3665	3712	4220	4616	4851	4813	33793	
Grupos Densos/Semidensos	5	5	5	5	5	5	5	5	40	
Puntos en grupos densos	4030	3808	3643	3691	4185	4531	4766	4756	33410	
Grupos Pocos Densos/Atípico	4	5	5	4	4	5	5	5	37	
Puntos en grupos poco densos	25	53	22	21	35	85	85	57	383	
Silhouette Etapa 1 por Grupos	0,51967	0,46852	0,49508	0,76060	0,48989	0,48845	0,47786	0,48437		0,52306
Silhouette Etapa 2 por Grupos	0,64525	0,58456	0,62014	0,61366	0,61494	0,61110	0,59404	0,60846		0,61152
Davies Bouldin Etapa 1	0,43612	0,46921	0,44529	0,44384	0,43849	0,44604	0,46239	0,44860		0,44875
Davies Bouldin Etapa 2	0,54381	0,58556	0,55331	0,54653	0,54565	0,55647	0,58162	0,55950		0,55906

Fuente: Autores.

TABLA. 6.
Resultados obtenidos de los experimentos en el conjunto de datos de Guayaquil en 8 periodos consecutivos de tiempo.

Evolución 1							Evolución 2						
Gru po	Cant. Puntos	Cant Veh	Vel Min	Vel. Mas	Vel. prom	Desv	Grupo	Cant. Puntos	Cant Veh	Vel Min	Vel. Mas	Vel. prom	Desv
2	41	3	0	43,5	8,61	12,16	1	421	21	0	41,47	6,25	9,17
1	32	3	24,75	55,57	39,62	7,93	-1	127	12	0	52,68	18,34	13,84
-1	29	1	40,31	65,84	57,21	7,45	2	231	22	5,37	59,21	35,3	11,79
							3	256	19	22,73	68,96	50,93	8,28
							4	263	22	27,2	84,03	63,11	10,72
							-3	36	4	69	98,88	84,89	7,3
Evolución 3							Evolución 4						
Gru po	Cant. Puntos	Cant Veh	Vel Min	Vel. Mas	Vel. prom	Desv	Gru po	Cant. Puntos	Cant Veh	Vel Min	Vel. Mas	Vel. prom	Desv
2	2781	67	0	46,62	5,78	8,5	2	4139	87	0	67,41	7,01	9,03
-1	350	28	0	50,37	17,94	11,9	4	911	52	0	61,95	23,29	12,12
3	1201	58	0	63,15	33,3	10,92	3	967	56	0	80,83	36,39	13,25
1	1933	65	2,04	91,99	49,53	9,54	1	1118	47	6,6	108,48	51,15	11,3
4	688	44	22,96	117,8	63,73	10,18	5	164	14	33,66	106,41	67,05	11,47
-3	52	8	46,43	96,83	83,14	9,09	-4	75	8	53,18	159,01	100,8	9,58
Evolución 5							Evolución 6						
Gru po	Cant. Puntos	Cant Veh	Vel Min	Vel. Mas	Vel. prom	Desv	Gru po	Cant. Puntos	Cant Veh	Vel Min	Vel. Mas	Vel. prom	Desv
3	3912	67	0	53,06	4,87	7,12	2	2078	45	0	62,65	5,62	6,99
1	985	45	0	65,43	14,93	12,01	1	338	27	0	72,94	20,43	12,43
2	1094	58	0	91,18	28,38	13,49	3	364	26	0	66,97	35,12	11,7
5	510	24	10,55	128,68	45,78	11,91	-3	212	20	6,87	127,3	69,36	11,41
6	264	15	5,7	148,3	56,31	13,41							
4	155	15	7,89	190,56	68,13	16,3							
-5	65	12	24,14	192,4	106,57	25,61							
Evolución 7							Evolución 8						
Gru po	Cant. Puntos	Cant Veh	Vel Min	Vel. Mas	Vel. prom	Desv	Gru po	Cant. Puntos	Cant Veh	Vel Min	Vel. Mas	Vel. prom	Desv
2	1836	30	0	86,27	6,17	9,45	2	535	12	0	49,44	5,78	7,28
-2	22	2	0	53,66	17,11	15,07	-1	158	9	0	50,38	23,89	12,68
3	528	24	0	73,69	27,25	12,62	1	211	12	5,27	86,52	42,76	13,94
1	303	18	17,97	91,96	47,01	11,1							
4	94	8	21,24	80,23	57,31	13,91							
-3	13	2	57,6	111,52	92,43	4,5							

Fuente: Autores.

TABLA. 7.
Algoritmo Dyclee - Guayaquil

	Evo. 1	Evo. 2	Evo. 3	Evo. 4	Evo. 5	Evo. 6	Evo. 7	Evo. 8	Sumatoria	Promedios
Trayectorias	206									
Puntos en consulta	29503									
Puntos a procesar	102	1658	6905	8053	6409	2879	2729	764	29499	
Grupos Densos/Semidensos	4	4	5	6	7	5	4	3	38	
Puntos en grupos densos	95	1418	6501	7893	6388	2823	2602	665	28385	
Grupos Pocos Densos/Atipico	2	4	4	5	6	4	4	3	32	
Puntos en grupos poco densos	7	240	404	160	21	56	127	99	1114	
Silhouette Etapa 1 por Grupos	0,56771	0,13528	0,48221	0,45862	0,45816	0,45016	0,48053	0,52065		0,44417
Silhouette Etapa 2 por Grupos	0,71126	0,27388	0,60316	0,57612	0,57430	0,56524	0,60103	0,64669		0,56896
Davies Bouldin Etapa 1	0,36062	0,09601	0,45170	0,46138	0,45164	0,48929	0,45373	0,41165		0,39700
Davies Bouldin Etapa 2	0,42975	0,18912	0,56267	0,57156	0,56129	0,60718	0,56596	0,51607		0,50045

Fuente: Autores.

CONCLUSIONES

Al finalizar el trabajo se ha logrado la implementación del Algoritmo Dyclee para que procese trayectorias vehiculares utilizando la dimensión de velocidad. A si mismo se ha logrado configurar un área de procesamiento para los 2 conjuntos de datos utilizados en donde se definieron celdas que proporcionan resúmenes de los datos de las trayectorias que se envían al algoritmo de clustering para la identificación de grupos.

Una vez realizado los experimentos y el análisis de los resultados con los dos conjuntos de datos esto es Roma y Guayaquil, se lograron identificar patrones con respecto a la velocidad promedio más alta existentes en las 8 evoluciones.

También con respecto a los indicadores de agrupamiento luego de realizados los experimentos se identificaron las evoluciones con las respectivas cantidades de grupos densos-semidensos, así como también las cantidades de grupos pocos densos-atípicos. Con respecto a las métricas de calidad Silhouette y Davies Bouldin, estas presentan resultados con respecto a la cantidad de los agrupamientos que son satisfactorios.

Como trabajo futuro se plantea poder incorporar a los experimentos y análisis otros conjuntos de datos y realizar el agrupamiento incorporando la dimensión de la ubicación de las trayectorias GPS para de esta forma enriquecer el análisis previo a una toma de decisiones.

REFERENCIAS BIBLIOGRÁFICA

- Bindiya M Varghese, U. A. P. J. (2013). Spatial Clustering Algorithms - an Overview. *Asian Journal of Computer Science and Information Technology*, 3(1).
- Ferreira, N., Klosowski, J. T., Scheidegger, C. E., & Silva, C. T. (2013). *Vector Field k-Means*. 32(3).
- Gepperth, A., & Hammer, B. (2016). Incremental learning algorithms and applications. *ESANN 2016 - 24th European Symposium on Artificial Neural Networks*, 357–368.
- Juca Maldonado, F., & Burgo Bencomo, O. B. (2016). lended Learning as a model for universities use Moodle. *Investigación, Tecnología e Innovación*, 8(EE SE-Article), 27–34. <https://doi.org/10.53591/iti.v8iEE.140>
- Madhulatha. (2012). An overview of clustering methods. *Intelligent Data Analysis*, 11(6), 583–605. <https://doi.org/10.3233/ida-2007-11602>

- Mazimpaka, J. D., & Timpf, S. (2016). Trajectory data mining: A review of methods and applications. *Journal of Spatial Information Science*, 13(2016), 61–99. <https://doi.org/10.5311/josis.2016.13.263>
- Molina, R., & Hasperué, W. (2018). *D3CAS#: un Algoritmo de Clustering para el Procesamiento de Flujos de Datos en Spark*. 452–461. http://sedici.unlp.edu.ar/bitstream/handle/10915/73223/Documento_completo.pdf-PDFA.pdf?sequence=1&isAllowed=y
- Reyes, G., Lanzarini, L., Hasperué, W., & Bariviera, A. F. (2020). GPS trajectory clustering method for decision making on intelligent transportation systems. *Journal of Intelligent and Fuzzy Systems*, 38(5), 5529–5535. <https://doi.org/10.3233/JIFS-179644>
- Reyes Zambrano, G., & Cedeño Pozo, A. (2018). Algorithm GR-B for the compression of vehicular trajectories. *Jurnaly*, 39(28).
- Suasnabas, L. S., Rubira, A. K., Casilla, I. N., & Schreiber, M. J. (2016). *Uso de la tecnología en la educación universitaria ecuatoriana* (p. 12).
- Tork, H. F. (2012). Spatio-Temporal Clustering Methods Classification. *Doctoral Symposium on Informatics Engineering (DSIE'2012), December*, 1–12. <https://doi.org/10.13140/RG.2.1.3812.7204>
- Tusa, F., Maza, J., & Briceño, X. (2016). La construcción de una educación del Buen Vivir a partir del estudio de las tecnologías de la comunicación y la información. In *Investigación, Tecnología e Innovación* (Vol. 8, Issue 8, pp. 97–106). <https://doi.org/10.53591/iti.v8i8.141>
- Vera, J. (2018). ANÁLISIS DE TRAYECTORIAS VEHICULARES GPS PARA EVALUAR SU CALIDAD DE AGRUPAMIENTO UTILIZANDO ALGORITMOS CLUSTERING DE MINERÍA DE DATOS. *Universidad de Guayaquil*, 83. http://repositorio.ug.edu.ec/bitstream/redug/41488/1/T-ZAMBRANO_ZAMBRANO JOSSELYN JAMILE.pdf
- Yuan, Y., & Raubal, M. (2014). Measuring similarity of mobile phone user trajectories- a Spatio-temporal Edit Distance method. *International Journal of Geographical Information Science*, 28(3), 496–520. <https://doi.org/10.1080/13658816.2013.854369>