Proyecto – Entrega final

Grupo 13: Freddy Rodrigo Mendoza Ticona, William Alexander Romero Bolívar, Maria Paula Salamanca Delgado, Jorge Oswaldo Suárez Rodríguez.

Detección e identificación de clústeres por afectación de dengue en Colombia: una aproximación desde el Aprendizaje no Supervisado

Resumen

El dengue es una enfermedad viral transmitida por la picadura de mosquitos infectados que se caracteriza por producir fiebre, dolor corporal, pérdida del apetito y, en casos graves, sangrado de mucosas. La población más vulnerable a esta enfermedad son niños y adultos mayores, sin embargo, puede afectar a cualquier grupo demográfico. Según el Instituto Nacional de Salud (INS), a julio de 2022, en Colombia se han presentado 34.017 casos, dentro de los cuales el 52,4% presentaron signos de alarma o graves (1,2).

Es posible prevenir y controlar su propagación concientizando a la población para evitar la proliferación de mosquitos que transmiten esta enfermedad. Así, a través de un análisis de clústeres se estratificaron 1.120 municipios de Colombia según el riesgo de incidencia de dengue de acuerdo con sus características demográficas (edad, población), socioeconómicas (desempleo, estratos), y climáticas (temperatura, precipitaciones), encontrando que existe un mayor riesgo de tasa de incidencia de dengue en municipios de clima cálido.

Los resultados de este proyecto permitirían dirigir recursos hacia la capacitación y concientización de la población para la prevención y control del dengue eficientemente en aquellas zonas de mayor riesgo de incidencia del dengue.

Introducción

Según la Organización Mundial de la Salud (OMS), se estima que anualmente el número de personas infectadas por dengue llega a 390 millones en el mundo y en riesgo de infección, a 3,9 billones (3). Adicionalmente, según la Organización Panamericana de la Salud (OPS), la circulación del dengue en Suramérica es hiperendémica, siendo Brasil y Colombia los que reportan el mayor número de casos; tan solo para el 2019 Colombia reportó 127.000 casos (4).

Los mosquitos *aegypti* son el principal vector de transmisión del dengue y están presentes en casi todos los países de Suramérica. El ciclo de transmisión se da por el contagio de la sangre de un humano infectado a un humano susceptible a través de la picadura del mosquito *Aedes* hembra que se reproduce cerca de las casas, poniendo huevos en recipientes de agua estancada. Se ha revisado factores asociados al aumento de la incidencia de infección por dengue, como el aumento global de la temperatura que aumenta el tiempo que el mosquito permanece infeccioso; el aumento estacional de las precipitaciones que contribuye a una mayor densidad de mosquitos; y condiciones de hacinamiento que incrementan la fracción de la población susceptible (5,6).

En estudios anteriores sobre el dengue se han utilizado algoritmos de aprendizaje no supervisado como los métodos de agrupamiento que son adecuados para la visualización de enfermedades, en especial los basados en la densidad para separar las regiones que tienen alta y baja densidad. En el estudio de Shaukat, et al, el análisis de datos de Pakistán por DBSCAN encontró que el dengue atacó principalmente la ciudad de Jhelum y Tehsil Jhelum, lo que permite enfocar estrategias de prevención de la enfermedad (7). Otro estudio realizado en la India encontró conglomerados de casos de dengue en Delhi usando un algoritmo de agrupamiento DBSCAN; estos puntos críticos se caracterizaron por ser grupos socioeconómicos bajos, estar cerca al río Yamuna, lagos y lugares con agua estancada (8).

Una revisión de la literatura realizada en Colombia encontró que, a nivel mundial en general, se ha usado modelos de aprendizaje supervisado con el objetivo de predicción de dengue, como de regresión logística con datos demográficos, clínicos y de laboratorio, así como modelos de regresión lineal, Random Forest y Support Vector Machine, con datos socioeconómicos, demográficos, climáticos y ambientales (9). También, estudios de análisis espacio-temporal como el realizado en Cali, Colombia, en el cual se encontró que el nivel socioeconómico, la densidad poblacional, proximidad a talleres con neumáticos, viveros de plantas y sistemas de alcantarillado, están relacionados con la enfermedad (10).

Sin embargo, no se encontró literatura donde se utilice el aprendizaje no supervisado para la estratificación del riesgo de dengue en Colombia, por lo cual, un análisis de clústeres por municipios de Colombia aportaría al caso de estudio. Así, este trabajo tiene como propósito estratificar el riesgo de infección por localización mediante una unidad de medida, como la tasa de incidencia de dengue del análisis de clústeres. Lo anterior sería de utilidad para las entidades gubernamentales y funcionarios de salud pública, al poder diferenciar zonas con mayor riesgo que requieran toma de decisiones sobre una intervención de control de propagación de esta enfermedad.

Metodología

Para el desarrollo de este proyecto se cuenta con una base de datos con información sobre 1.121 municipios y 1.017 características, como datos demográficos y socioeconómicos, casos acumulados de dengue del 2007 al 2019 y por semana epidemiológica e información climática como mediciones de temperatura y precipitaciones mensuales. Esta información se obtuvo de las bases de datos del Sistema de Vigilancia en Salud Pública (SIVIGILA) del Instituto Nacional de Salud (INS) y del Departamento Administrativo Nacional de Estadística (DANE).

Variable	**Descripción**	**Tipo dato**
Municipality code	Código del municipio	numérico
Municipality	Descripción de municipio	texto
Population (2007-2019)	Población del municipio por año	numérico
Cases (2007-2019)	Casos dengue reportados en el municipio x año	numérico
Age0-4(%)	Porcentaje de población menor de 4 años	numérico
Age5-14(%)	Porcentaje de población entre 5 y 14 años	numérico
Age15-29(%)	Porcentaje de población entre 15 y 29 años	numérico
Age>30(%)	Porcentaje de población mayor de 30 años	numérico
AfrocolombianPopulation(%)	Porcentaje de población afrocolombiana	numérico
IndianPopulation(%)	Porcentaje de población indígena	numérico
PeoplewithDisabilities(%)	Porcentaje de población con discapacidades (física, psicológica o mental)	numérico
Peoplewhocannotreadorwrite(%)	Porcentaje de población con que no puede leer/escribir	numérico
Secondary/HigherEducation(%)	Porcentaje de población que tiene educación secundaria	numérico
Employedpopulation(%)	Porcentaje de población empleada	numérico
Unemployedpopulation(%)	Porcentaje de población desempleada	numérico
Peopledoinghousework(%)	Porcentaje de población que realizan trabajo doméstico	numérico
Retiredpeople(%)	Porcentaje de población jubilada	numérico
Men(%)	Porcentaje de población masculina	numérico
Women(%)	Porcentaje de población femenina	numérico
Householdswithoutwateraccess (%)	Porcentaje de viviendas sin acceso a agua	numérico
Householdswithoutinternetaccess(%)	Porcentaje de viviendas sin acceso a internet	numérico
Buildingstratification1(%)	Porcentaje de viviendas estrato 1	numérico
Buildingstratification2(%)	Porcentaje de viviendas estrato 2	numérico
Buildingstratification3(%)	Porcentaje de viviendas estrato 3	numérico
Buildingstratification4(%)	Porcentaje de viviendas estrato 4	numérico
Buildingstratification5(%)	Porcentaje de viviendas estrato 5	numérico

Buildingstratification6(%)	Porcentaje de viviendas estrato 6	numérico
NumberofhospitalsperKm2	Número de hospitales por km2 en el municipio	numérico
NumberofhousesperKm2	Número de casas por km2 en el municipio	numérico
TEMPERATURE_mm_AA	Temperatura promedio en el municipio por mes mm del año AA (desde 2007 hasta 201812)	numérico
PRECIPITATION_mm_AA	Precipitaciones en el municipio por mes por mes mm del año AA (desde 2007 hasta 201812)	numérico
AAAA/wi	Casos de dengue reportados x semana w(i) (1-52) del año AA (desde el año 2007 hasta 201912)	numérico

A continuación, se encuentran algunas estadísticas de estas variables:

	Age0-4(%)	Age5-14(%)	Age15-29(%)	Age>30(%)	Afrocolombian Population(%)	Indian Population(%)	Peoplewith Disabilities(%)	Peoplewhocannot readorwrite(%)	Secondary/Higher Education(%)	Employedpopulation(%)	Unemployedpopulation(%)	Peopledoing housework(%)	Retiredpeople (%)
count	1120	1120	1120	1120	1120	1120	1120	1120	1120	1120	1120	1120	1120
mean	7.313339	17.851643	24.617893	50.21725	7.052393	7.360429	9.907688	8.918884	44.003848	33.311098	4.459554	22.121187	1.19817
std	1.980428	4.165485	2.944725	7.752337	19.319731	19.075145	5.611397	3.850673	9.310005	7.687468	2.885017	5.645678	1.415711
min	3.28	9.27	13.98	19.13	0	0	0	1.47	15.02	0.83	0.5	6.47	0
25%	5.88	14.99	22.79	45.1825	0.18	0.03	5.73	6.3275	37.73	29.2575	2.53	18.6875	0.26
50%	6.965	17.015	24.93	51.045	0.41	0.14	8.695	8.46	42.555	34.26	3.925	21.8	0.64
75%	8.47	19.8725	26.52	56.035	1.6175	2.715	13.13	10.855	49.29	38.4975	5.6175	24.6775	1.5425
max	23	48.15	47.18	69.5	97.3	98.58	48.48	34.28	77.33	51.58	37.2	51.7	10.17

	Cases2007	Cases2008	Cases2009	Cases2010	Cases2011	Cases2012	Cases2013	Cases2014	Cases2015	Cases2016	Cases2017	Cases2018	Cases2019
count	1120	1120	1120	1120	1120	1120	1120	1120	1120	1120	1120	1120	1120
mean	36.582143	32.2125	46.378571	138.61875	26.675893	47.751786	111.308929	93.544643	85.56875	89.2625	22.364286	38.841964	113.23125
std	229.21036	163.028782	286.540394	784.956306	111.958658	211.543485	661.101501	399.807836	510.722281	797.613186	137.836907	162.624276	443.491883
min	0	0	0	0	0	0	0	0	0	0	0	0	0
25%	0	0	0	1	0	0	1	1	1	1	0	0	1
50%	1	2	2	10	2	3	9	8	8	8	2	2	10
75%	9	10	12.25	54	12	23	51.25	51.25	44	37.25	10	14	75
max	5223	2816	6335	15570	1724	3128	17539	5855	14523	19414	3710	2687	8258

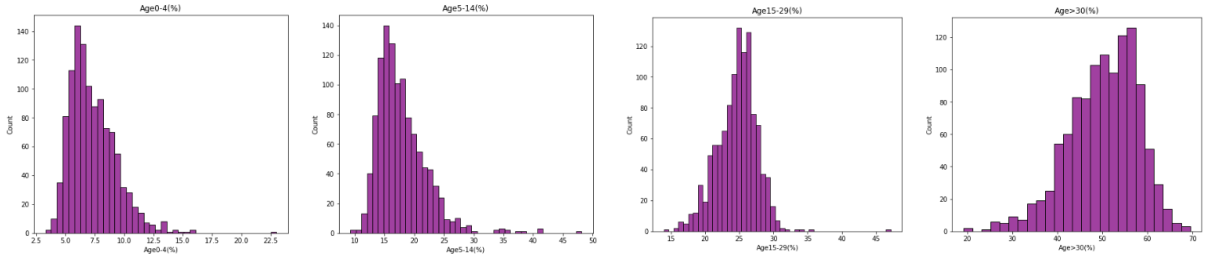
	Population2007	Population2008	Population2009	Population2010	Population2011	Population2012	Population2013	Population2014	Population2015	Population2016	Population2017	Population2018	Population2019
count	1.12E+03	1.12E+03	1.12E+03	1.12E+03	1.12E+03	1.12E+03	1.12E+03	1.12E+03	1.12E+03	1.12E+03	1.12E+03	1.12E+03	1.12E+03
mean	3.83E+04	3.87E+04	3.92E+04	3.96E+04	4.00E+04	4.03E+04	4.07E+04	4.11E+04	4.14E+04	4.18E+04	4.23E+04	4.31E+04	4.41E+04
std	2.35E+05	2.37E+05	2.40E+05	2.42E+05	2.44E+05	2.45E+05	2.47E+05	2.48E+05	2.49E+05	2.50E+05	2.51E+05	2.54E+05	2.60E+05
min	0.00E+00	1.27E+02	1.27E+02	1.29E+02	1.44E+02	1.59E+02	1.72E+02	1.92E+02	2.10E+02	2.24E+02	2.43E+02	2.79E+02	2.88E+02
25%	6.37E+03	6.51E+03	6.58E+03	6.63E+03	6.67E+03	6.58E+03	6.49E+03	6.52E+03	6.43E+03	6.43E+03	6.42E+03	6.58E+03	6.64E+03
50%	1.20E+04	1.22E+04	1.23E+04	1.25E+04	1.26E+04	1.25E+04	1.27E+04	1.27E+04	1.27E+04	1.27E+04	1.27E+04	1.28E+04	1.28E+04
75%	2.39E+04	2.46E+04	2.48E+04	2.51E+04	2.54E+04	2.57E+04	2.60E+04	2.63E+04	2.63E+04	2.62E+04	2.64E+04	2.69E+04	2.75E+04
max	6.87E+06	6.94E+06	7.00E+06	7.07E+06	7.12E+06	7.16E+06	7.20E+06	7.23E+06	7.25E+06	7.29E+06	7.33E+06	7.41E+06	7.59E+06

	TEMPERATURE_jan_07	TEMPERATURE_feb_07	TEMPERATURE_mar_07	TEMPERATURE_apr_07	TEMPERATURE_may_07	TEMPERATURE_jun_07	TEMPERATURE_jul_07	TEMPERATURE_aug_07	TEMPERATURE_sep_07	TEMPERATURE_oct_07	...	TEMPERATURE_mar_18	TEMPERATURE_apr_18	TEMPERATURE_may_18	TEMPERATURE_jun_18	TEMPERATURE_jul_18	TEMPERATURE_aug_18	TEMPERATURE_sep_18	TEMPERATURE_oct_18	TEMPERATURE_nov_18	TEMPERATURE_dec_18
count	1120	1120	1120	1120	1120	1120	1120	1120	1120	1120	...	1120	1120	1120	1120	1120	1120	1120	1120	1120	1120
mean	21.341494	21.369348	21.624123	21.427352	21.394515	20.985997	21.221093	20.691514	20.881906	20.394404	...	21.909892	21.321128	21.253347	21.282962	21.535259	21.546901	21.557522	21.255658	21.368144	21.347864
std	5.690491	5.7895	5.813286	5.737872	5.670817	5.830323	5.945628	5.843421	5.731512	5.622461	...	5.741662	5.714293	5.596556	5.72358	5.868096	5.778783	5.70219	5.538864	5.535728	5.610472
min	3.927416	3.718099	4.207906	4.065408	4.02966	3.317243	2.965087	2.681081	3.291453	3.387867	...	4.317282	3.887283	3.471066	3.404483	2.927327	3.381603	3.892235	3.970158	4.306131	4.383894
25%	17.045011	17.02751	17.274697	17.10996	17.248967	16.751721	17.0725	16.426404	16.64545	16.118646	...	17.55801	16.97379	17.071008	17.081442	17.333293	17.261581	17.322357	17.038166	17.125349	17.057504
50%	21.795407	21.870771	22.119152	21.881226	21.893436	21.486656	22.081644	21.316279	21.413619	20.775802	...	22.416773	21.765229	21.771181	21.884009	22.476355	22.202146	22.127749	21.634492	21.672488	21.834375
75%	26.950056	27.07109	27.14519	26.780102	26.670794	26.111782	26.21806	25.890016	26.220641	25.947726	...	27.380697	26.639163	26.40791	26.348177	26.617457	26.589537	26.820132	26.733839	26.870083	26.829089
max	29.710881	30.431314	30.537816	29.604315	30.158815	30.600778	30.807735	30.164066	29.77935	28.815366	...	30.129071	29.420981	29.578173	30.500952	30.539854	30.80686	30.257563	29.229948	29.038523	28.674008

	PRECIPITATION_jan_07	PRECIPITATION_feb_07	PRECIPITATION_mar_07	PRECIPITATION_apr_07	PRECIPITATION_may_07	PRECIPITATION_jun_07	PRECIPITATION_jul_07	PRECIPITATION_aug_07	PRECIPITATION_sep_07	PRECIPITATION_oct_07	...	PRECIPITATION_mar_18	PRECIPITATION_apr_18	PRECIPITATION_may_18	PRECIPITATION_jun_18	PRECIPITATION_jul_18	PRECIPITATION_aug_18	PRECIPITATION_sep_18	PRECIPITATION_oct_18	PRECIPITATION_nov_18	PRECIPITATION_dec_18
count	1120	1120	1120	1120	1120	1120	1120	1120	1120	1120	...	1120	1120	1120	1120	1120	1120	1120	1120	1120	1120
mean	58.939316	43.138888	140.990806	280.6461	271.626818	156.403747	213.566671	251.010376	170.347437	321.097562	...	102.798618	237.488985	293.902968	185.121259	156.038725	149.250247	186.633616	271.943875	196.227458	49.733369
std	79.846343	37.177589	89.884983	102.101643	113.774602	86.729283	181.552088	148.273573	104.824351	105.299413	...	65.409966	81.514146	131.746465	100.536602	110.824689	98.508155	102.198787	95.393597	113.600392	57.995459
min	0.113365	0.198215	0.701972	27.129646	59.491062	18.991796	21.740707	30.423084	33.799904	95.840744	...	0.599829	19.097061	37.572857	21.781021	16.594635	23.58654	36.369263	104.506844	45.653301	0.902079
25%	9.822512	22.464324	89.069555	211.460246	178.799721	93.212273	95.087875	133.428032	86.434338	252.648006	...	66.392992	177.73811	195.790398	113.920891	77.161709	81.617582	116.061552	201.03244	109.864227	17.721469
50%	37.6359	35.83193	142.010719	278.711731	264.532791	134.02597	155.569702	218.900551	150.956848	313.188736	...	101.733345	228.947937	275.260513	157.643913	117.835323	120.328899	157.802322	260.294006	185.442703	37.102827
75%	73.97994	51.815425	185.343494	326.028084	338.819244	189.300114	289.582605	331.352974	229.326088	370.52459	...	136.059479	280.5961	364.824768	226.10923	192.111206	197.785351	242.241825	316.978779	258.00985	60.799553
max	549.786499	238.397308	615.918823	688.654114	773.122009	548.34613	1357.88269	913.841675	661.933777	821.908569	...	418.169952	576.553894	923.395142	584.70575	627.695618	709.576599	773.638	631.59967	829.757263	609.078796

	2007/w01	2007/w02	2007/w03	2007/w04	2007/w05	2007/w06	2007/w07	2007/w08	2007/w09	2007/w10	...	2019/w43	2019/w44	2019/w45	2019/w46	2019/w47	2019/w48	2019/w49	2019/w50	2019/w51	2019/w52
count	1120	1120	1120	1120	1120	1120	1120	1120	1120	1120	...	1120	1120	1120	1120	1120	1120	1120	1120	1120	1120
mean	0.590179	0.392857	0.4375	0.521429	0.525893	0.649107	0.586607	0.554464	0.586607	0.545536	...	2.344643	2.436607	2.674107	2.569643	2.583929	2.496429	2.66875	2.702679	2.352679	1.010714
std	3.720687	2.62032	3.014061	3.332549	3.723851	4.42625	3.522125	3.647404	3.344343	2.570833	...	9.859014	10.502457	11.845824	11.093983	11.342507	10.709337	11.901746	11.982486	10.623279	4.85186
min	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
25%	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
50%	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
75%	0	0	0	0	0	0	0	0	0	0	...	1	1	1	1	1	1	1	1.25	1	0
max	82	63	63	70	100	107	66	73	59	43	...	146	174	174	182	187	186	212	223	242	106

Con lo anterior, se encontró que, de acuerdo con la edad de los habitantes de los municipios, el 30% de la población es vulnerable al dengue.



Se encontró que el municipio de Puebloviejo no cuenta con información sobre temperatura y precipitaciones, por lo cual se decidió eliminar de la base este registro teniendo en cuenta que se consideran variables importantes para el análisis de incidencia del dengue. De esta manera, se cuenta con una base de datos limpia con información sobre 1.120 municipios y 1.017 características.

A esta base, se le aplicó el algoritmo K-medias para la serie temporal de casos y población (del 2007 al 2019) y se creó una única serie temporal que es la tasa de casos de dengue por población para realizar un K-medias utilizando Dynamic Time Warping (DTW) como medida de distancia, el resultado es la agrupación de municipios con similar tasa de casos de dengue en el tiempo (2007-2019). Además, se estimaron clústeres de la base con el último dato disponible de las series de tiempo (2019) a través de Clustering Jerárquico y DBSCAN.

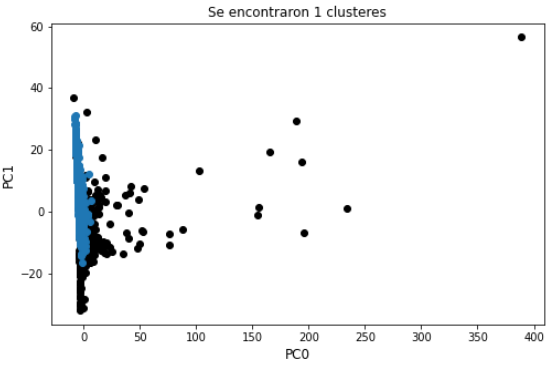
Resultados

Al aplicar K-medias sobre la serie de tiempo de contagios y población, se obtuvieron tres clústeres, entre los cuales se identificó uno que contiene los municipios con mayor número de contagios por población. Estos municipios se caracterizan por ser de clima cálido y húmedo, además de no ser pequeños en comparación con ciudades intermedias.

	Municipality	Municipality code	indice_total	cluster_ts
1050	Fortul	81300	0.020372	0
518	Nilo	25488	0.017301	0
690	Castilla La Nueva	50150	0.011041	0
1055	Aguazul	85010	0.010755	0
984	Melgar	73449	0.010042	0
687	Acacías	50006	0.008943	0
960	Alvarado	73026	0.008869	0
981	Lérida	73408	0.008670	0
607	Aipe	41016	0.008577	0
688	Barranca de Upía	50110	0.008402	0
973	Espinal	73268	0.008287	0
706	Puerto López	50573	0.007699	0
710	San Carlos de Guaroa	50680	0.007435	0
617	Hobo	41349	0.007164	0
1107	San José del Guaviare	95001	0.007164	0
698	Guamal	50318	0.007114	0
1070	Tauramena	85410	0.007080	0
1054	Yopal	85001	0.007030	0
1109	El Retorno	95025	0.007016	0

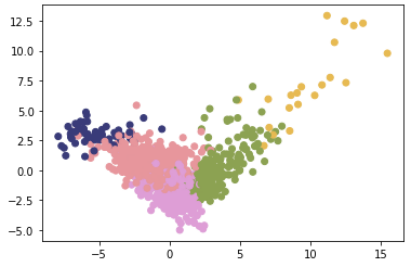
Por otro lado, se estimó un modelo por DBSCAN a una base reducida de 73 componentes principales que explican el 98% de la varianza total de la base de datos de 2019 y se encontró que los 458 municipios que se clasificaron como ruido o outliers se caracterizan por tener tanto altas como bajas tasas de incidencia de dengue.

	Municipality	Municipality code	Casos_total	Temperatura_promedio	Precipitacion_promedio
9	Anorí	5040	75	22.745777	266.232825
599	San José del Palmar	27660	21	22.383884	260.546915
82	Remedios	5604	252	25.867615	260.301663
5	Amalfi	5031	32	22.061463	257.702365
117	Vegachí	5858	208	24.055471	254.445760
...	...	...	...	...	...
297	Susacón	15774	1	10.410217	80.896387
865	Concepción	68207	35	8.622600	77.888286
195	Tunja	15001	23	11.981617	77.742075
288	Siachoque	15740	0	10.423938	77.338384
227	El Cocuy	15244	1	8.628353	74.024666



Por último, se estimó un modelo por Clustering Jerárquico a una base reducida de 21 componentes principales que explican el 95% de la varianza total, obteniendo cinco clústeres, entre los cuales se identificó uno que contiene los municipios con mayor tasa de contagios por población. Estos municipios se caracterizan por ser de clima cálido y húmedo, resultados muy similares a los obtenidos por K-medias sobre series de tiempo.

	Municipality code	Municipality	AfrocolombianPopulation(%)	Cases2019	Population2019	ratioPoblacionCasos	cluster
514	25438	Medina	0.18	260	8064	0.032242	3
1088	91001	Leticia	0.88	1676	48918	0.034261	3
713	50689	San Martín	0.75	953	26571	0.035866	3
856	68147	Capitanejo	0.39	250	5479	0.045629	3
996	73671	Saldaña	0.09	688	14450	0.047612	3



De acuerdo con lo anterior, se encuentra que a través de los algoritmos K-medias en series de tiempo y Clustering Jerárquico se logran identificar clústeres de municipios que tienen mayor riesgo de presentar brotes de dengue, a diferencia de DBSCAN, con el cual se clasificaron estos municipios como atípicos. Teniendo en cuenta que el interés del proyecto es la identificación de las zonas en las que se deberían concentrar las campañas de concientización de prevención y control contra el dengue, los algoritmos K-medias sobre series de tiempo y Clustering Jerárquico proporcionan los resultados esperados.

Una limitación de la implementación de K-medias sobre series de tiempo es que requiere mayores recursos computacionales si la serie es de gran longitud. Para futuras implementaciones, se podrían tener en cuenta las series de tiempo de temperatura y precipitación para la estimación de un modelo de K-medias de series de tiempo multidimensional. Adicionalmente, estos resultados podrían implementarse para un análisis de geolocalización que permitiera la identificación de puntos calientes en Colombia, además de servir como ruta de priorización por zonas y puntos de mayor concentración.

Conclusión

Con una base de 1.120 municipios de Colombia y sus características demográficas (edad, población), socioeconómicas (desempleo, estratos), y climáticas (temperatura, precipitaciones), se estimaron clústeres por K-medias para series de tiempo y Clustering Jerárquico. Con el primero, se encontró que los municipios como Fortul, Nilo, Castilla La Nueva, entre otros, son aquellos que tienen un mayor riesgo de brote del dengue y se caracterizan en especial por su clima cálido y baja población.

Por otro lado, a través de Clustering Jerárquico se encontró que municipios como Saldaña, Capitanejo, San Martin, Leticia, Medina, entre otros, son los municipios que tienen mayor riesgo de incidencia de dengue, en relación con la proporción de contagios sobre el total de su población. Asimismo, estos municipios se caracterizan por su clima húmedo y cálido. Por último, aunque a través de DBSCAN se esperaba encontrar zonas densas de municipios cuyas características permitieran la identificación de clústeres con mayor tasa de incidencia de dengue, no se obtuvieron resultados que estuvieran alineados con el objetivo de este proyecto, ya que se agruparon como atípicos tanto municipios con alta tasa de incidencia como de baja incidencia.

De esta manera, se encontró que los municipios de clima cálido y mayor humedad hacen parte de las zonas de mayor riesgo y, por tanto, son a estos municipios a los que se deberían dirigir recursos para la capacitación y concientización de la población para la prevención y control del dengue eficientemente.

Bibliografía

1 . Instituto Nacional de Salud. Protocolo de Vigilancia de Dengue [Internet]. INS 2022. [citado el 20 de agosto de 2022. Disponible en: https://www.ins.gov.co/buscadoreventos/Lineamientos/Pro_Dengue.pdf

2. Instituto Nacional de Salud. Informe de Evento Dengue [Internet]. INS 2022. Citado el 20 de agosto de 2022. Disponible en: https://www.ins.gov.co/buscadoreventos/Informesdeevento/DENGUE%20PE%20VII%202022.pdf

3. World Health Organization. Dengue and severe dengue [Internet]. WHO 2022. Citado el 04 de septimebre de 2022. Disponible en: https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue#:~:text=The%20number%20of%20dengue%20cases,affecting%20mostly%20younger%20age%20group.

4. Organización Panamericana de la Salud. Dengue [Internet]. OPS. 2020. Citado el 20 de agosto de 2022. Disponible en: https://www.paho.org/es/temas/dengue

5. Thomas SJ, Rothman AL. Dengue virus infection: Epidemiology. In: UpToDate, Shefner JM (Ed), UpToDate, Waltham, MA. (Accessed on September 04, 2022.)

6. Bhatt S, Gething PW, Brady OJ, et al. The global distribution and burden of dengue. *Nature*. 2013;496(7446):504-507. doi:10.1038/nature12060

7. Shaukat K, Masood N, Shafaat AB, Jabbar K, Shabbir H, Shabbir S. Dengue fever in perspective of clustering algorithms. arXiv preprint arXiv:1511.07353. 2015 Nov 23.

8. G. M. Nandana, S. Mala and A. Rawat, "Hotspot Detection of Dengue Fever Outbreaks Using DBSCAN Algorithm," *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2019, pp. 158-161, doi: 10.1109/CONFLUENCE.2019.8776916.

9. Hoyos W, Aguilar J, Toro M. Dengue models based on machine learning techniques: A systematic literature review. *Artif Intell Med*. 2021;119:102157. doi:10.1016/j.artmed.2021.102157

10. Delmelle E, Hagenlocher M, Kienberger S, Casas I. A spatial model of socioeconomic and environmental determinants of dengue fever in Cali, Colombia. *Acta Trop*. 2016;164:169-176. doi:10.1016/j.actatropica.2016.08.028