

QA76.76  
I5808

UNIVERSIDAD DE LOS ANDES

Facultad de Ingeniería  
Maestría en Matemática Aplicada a la Ingeniería



www.bdigital.ula.ve

“METODOLOGÍAS MATEMÁTICAS PARA EL MODELADO  
Y PREDICCIÓN DE LA DINÁMICA DEL VRS EN LA  
POBLACIÓN”

TRABAJO ESPECIAL DE GRADO PRESENTADO POR

LCDO. JOSÉ QUERALES

COMO REQUISITO FINAL

PARA OBTENER EL TÍTULO DE MAESTRÍA

EN MATEMÁTICA APLICADA A LA INGENIERÍA

ÁREA DE CONOCIMIENTO: MATEMÁTICA APLICADA

TUTOR: DR. GILBERTO GONZÁLEZ PARRA

COTUTOR: DR. DIEGO ARANDA

DONACION

Mérida, Venezuela.

Septiembre de 2012

SERBIULA  
Tullo Fobres Cordero

[www.bdigital.ula.ve](http://www.bdigital.ula.ve)

*Primeramente a Dios, mis padres y mi esposa.*

Reconocimiento-No comercial-Compartir igual

# AGRADECIMIENTOS

Gracias a Dios por darme estos talentos para realizar este trabajo, a toda mi familia especialmente a mis padres, a mi esposa bella por su gran motivación y apoyo. A todos los profesores de la Maestría, por transmitirme sus conocimientos especialmente a mi tutor el Prof. Gilberto González. Gracias a todos por su gran apoyo y colaboración. Por último al CDCTHA ULA a través del proyecto I-1331-12-05-EM por su colaboración.

[www.bdigital.ula.ve](http://www.bdigital.ula.ve)

# “METODOLOGÍAS MATEMÁTICAS PARA EL MODELADO Y PREDICCIÓN DE LA DINÁMICA DEL VRS EN LA POBLACIÓN”

Lcdo. José Querales

## RESUMEN

En este trabajo se aplicaron los clasificadores bayesianos de Navie Bayes (*NB*) para predecir la primera semana del brote de la enfermedad estacionaria del Virus Respiratorio Sincitial (*VRS*) utilizando los datos correspondiente a los niños menores de cinco años de la ciudad de Bogotá (Colombia) del año 2005 hasta el 2010. Una predicción fiable de la semana del inicio de estos brotes, utilizando bases de datos climatológicas, podría ayudar a los hospitales y a la población a prepararse mejor para los grandes brotes causados por el *VRS*. En este trabajo, utilizamos el clasificador de Naive Bayes (*NB*) donde las variables usadas son temperatura promedio ( $C^{\circ}$ ), temperatura máxima( $C^{\circ}$ ), temperatura mínima( $C^{\circ}$ ), presión atmosférica (hPa), precipitación (mm), viento (Km/h), humedad máxima (%) y humedad mínima (%). Utilizando los clasificadores de Naive Bayes se construyeron 1020 modelos para los periodos de seis años (2005-2010), cuatro años (2005-2008) y dos años (2009-2010) respectivamente. Se logró determinar que las variables: humedad mínima, velocidad del viento y temperatura mínima son las que aparecen con mayor frecuencia en los modelos con mejor rendimiento para la predicción del brote del *VRS* en la ciudad Bogotá.

Por otra parte, en este trabajo también se utilizó el modelo matemático *SIRS*, basado en ecuaciones diferenciales para describir y entender los brotes epidémicos del *VRS* en la población de la ciudad de Bogotá. Utilizando el modelo *SIRS* se logró describir los datos reales correspondientes al *VRS* durante el periodo 2005-2010. Adicionalmente, se utilizó un modelo Multi-Agentes para simular y describir, desde un punto de vista estocástico, los brotes del *VRS* en la ciudad de Bogotá. Es importante mencionar que se lograron resultados similares a lo obtenidos mediante al modelo determinístico *SIRS*. Sin embargo, se comprobó la mayor versatilidad del modelo Multi-Agentes donde se puede obtener una mayor cantidad de escenarios.

# ÍNDICE

---

<b>Agradecimientos</b>	<b>i</b>
<b>Introducción</b>	<b>1</b>
<b>1 Virus Respiratorio Sincitial (VRS)</b>	<b>5</b>
1.1 ¿Qué es el VRS? . . . . .	5
1.2 ¿Como se contrae este virus? . . . . .	7
1.3 Relación del clima con el VRS . . . . .	7
1.4 Datos del clima e infectados por el VRS . . . . .	9
<b>2 Clasificadores de Naive Bayes (NB)</b>	<b>13</b>
2.1 Redes Bayesianas . . . . .	13
2.2 Construcción de las redes bayesianas . . . . .	14
2.3 Evidencias y Propagación de Probabilidades . . . . .	14
2.4 Separación $D$ . . . . .	15
2.4.1 Conexiones en Serie . . . . .	15
2.4.2 Conexiones Divergentes . . . . .	16
2.4.3 Conexiones convergentes . . . . .	16
2.5 Clasificadores de Naive Bayes (NB) . . . . .	17
2.6 La curva ROC . . . . .	18
2.7 Capacidad de Discriminación de un Clasificador . . . . .	19
2.8 Probabilidades de Clasificación de los Marcadores . . . . .	20
2.9 Valores Predictivos . . . . .	23
2.10 Espacio ROC . . . . .	24

---

<b>3</b>	<b>Clasificadores (NB) para predecir el inicio de brote del VRS en Bogotá</b>	<b>27</b>
3.1	Selección de las variables de las variables climatológicas . . . . .	27
3.1.1	Correlación de variables climatológicas . . . . .	29
3.2	Clasificador de Naive Bayes utilizando las variables climatológicas	32
3.3	Determinación de la semana de brote . . . . .	36
3.4	Resultados . . . . .	38
3.4.1	Evaluación del rendimiento de los modelos . . . . .	38
3.4.2	Clasificación de los modelos . . . . .	39
3.4.3	Modelos que utilizan los datos del periodo de 6 años . . . . .	45
3.4.4	Modelos que utilizan los datos del periodo de 4 años . . . . .	45
3.4.5	Modelos que utilizan los datos del periodo de 2 años . . . . .	46
3.4.6	Rendimiento individual de las variables climatológicas . . . . .	47
3.5	Análisis general de los modelos . . . . .	49
3.5.1	Mejor periodo . . . . .	50
3.5.2	Semanas de antelación . . . . .	51
3.5.3	Modelos por variables climatológicas . . . . .	51
3.6	Conclusiones . . . . .	52
<b>4</b>	<b>Modelo SIRS estacional para el VRS</b>	<b>55</b>
4.1	Modelado matemático de epidemias . . . . .	55
4.2	Modelo SIR . . . . .	55
4.3	Modelo SIRS . . . . .	57
4.4	Análisis Cuantitativo . . . . .	59
4.4.1	Estados estacionarios del sistema de ecuaciones diferenciales	59
4.5	Modelos SIRS con estacionalidad forzada . . . . .	60
4.6	Estimación de parámetros . . . . .	61
4.6.1	Mínimos cuadrados . . . . .	62
4.6.2	Algoritmo de Nelder-Mead . . . . .	63
4.7	Resultados . . . . .	63
4.8	Conclusión . . . . .	65
<b>5</b>	<b>Sistema de Multi-Agentes (SMA)</b>	<b>67</b>
5.1	Motivación . . . . .	67

---

5.2	Un <i>SMA</i> con arquitectura equivalente a un sistema EDO . . . . .	69
5.2.1	Agente . . . . .	70
5.2.2	Modelos Basados en agentes . . . . .	72
5.3	Modelo SIRS con Multi-Agentes . . . . .	72
5.3.1	Algoritmo de evolución del <i>SMA</i> . . . . .	73
5.4	Resultados del modelo <i>SIRS</i> con Multi-Agentes . . . . .	74
5.5	Conclusión . . . . .	77
<b>6</b>	<b>Conclusiones</b>	<b>79</b>
	<b>Referencias</b>	<b>83</b>
<b>A</b>	<b>Apendice</b>	<b>91</b>

[www.bdigital.ula.ve](http://www.bdigital.ula.ve)

## ÍNDICE DE TABLAS

---

1.1	Casos de <i>VRS</i> en niños menores de 5 años de edad por semana (2005)	10
1.2	Casos de <i>VRS</i> en niños menores de 5 años de edad por semana (2006)	10
1.3	Casos de <i>VRS</i> en niños menores de 5 años de edad por semana (2007)	11
1.4	Casos de <i>VRS</i> en niños menores de 5 años de edad por semana (2008)	11
1.5	Casos de <i>VRS</i> en niños menores de 5 años de edad por semana (2009)	12
1.6	Casos de <i>VRS</i> en niños menores de 5 años de edad por semana (2010)	12
2.1	Matriz de confusión.	19
3.1	Matriz de correlación.	31
3.2	Puntos obtenidos en cada gráfica.	39
3.3	Modelos aceptados y rechazados de la semana 0 (6 años)	39
3.4	Modelos aceptados y rechazados de las semana 1 (6 años)	40
3.5	Modelos aceptados y rechazados de las semana 2 (6 años)	40
3.6	Modelos aceptados y rechazados de las semana 3 (6 años)	41
3.7	Modelos aceptados y rechazados de las semana 0 (4 años)	41
3.8	Modelos aceptados y rechazados de las semana 1 (4 años)	42
3.9	Modelos aceptados y rechazados de las semana 2 (4 años)	42
3.10	Modelos aceptados y rechazados de las semana 3 (4 años)	43
3.11	Modelos aceptados y rechazados de las semana 0 (2 años)	43
3.12	Modelos aceptados y rechazados de las semana 1 (2 años)	43
3.13	Modelos aceptados y rechazados de las semana 2 (2 años)	43
3.14	Modelos aceptados y rechazados de las semana 3 (2 años)	43
3.15	Modelos con buen desempeño con variables climatológicas de 6 años	44

3.16 Modelos con buen desempeño con variables climatológicas de 4 años	44
3.17 Modelos con buen desempeño con variables climatológicas de 2 años	44
3.18 Modelos de mejor desempeño (zona noreste) de 6 años	45
3.19 Modelos de mejor desempeño (zona noreste) de 4 años	46
3.20 Modelos de mejor desempeño (zona noreste) de 2 años	46
3.21 Porcentaje de los modelos de la temporada de 6 años	49
3.22 Porcentaje de los modelos de la temporada de 4 años	49
3.23 Porcentaje de los modelos de la temporada de 2 años	49
3.24 Modelos de mejor rendimiento para cada periodo (2, 4 y 6 años).	50
3.25 Mejores modelos de cada semana 0,1,2 y 3 antes del inicio del brote.	51
3.26 Promedio de las variables climatológicas	52
4.1 Parámetros estimados para distintas regiones del mundo	64
A.1 Resultados semana 0 del periodo de 6 años	92
A.2 Resultados semana 1 del periodo de 6 años	94
A.3 Resultados semana 2 del periodo de 6 años	96
A.4 Resultados semana 3 del periodo de 6 años	97
A.5 Resultados semana 0 del periodo de 4 años	98
A.6 Resultados semana 1 del periodo de 4 años	100
A.7 Resultados semana 2 del periodo de 4 años	102
A.8 Resultados semana 3 del periodo de 4 años	103
A.9 Resultados semana 0 del periodo de 2 años	104
A.10 Resultados semana 1 del periodo de 2 años	104
A.11 Resultados semana 2 del periodo de 2 años	105
A.12 Resultados semana 3 del periodo de 2 años	106

## ÍNDICE DE FIGURAS

---

1.1	Comparación de las vías respiratorias . . . . .	6
1.2	Datos del Instituto Nacional de Salud de la ciudad de Bogotá . . . . .	11
2.1	Ejemplo de tres nodos conectados en serie. . . . .	15
2.2	Ejemplo de una conexión divergente. . . . .	16
2.3	Ejemplo de una conexión convergente. . . . .	16
2.4	Markov Blanket de un Nodo A. . . . .	17
2.5	Gráficos ROC básicos. . . . .	24
3.1	Diagrama de dispersión. . . . .	30
3.2	Estructura del clasificador <i>NB</i> en relación a la semana 0 . . . . .	34
3.3	Estructura del clasificador <i>NB</i> en relación a la semana 1 . . . . .	34
3.4	Estructura del clasificador <i>NB</i> en relación a la semana 2 . . . . .	34
3.5	Estructura del clasificador <i>NB</i> en relación a la semana 3 . . . . .	34
3.6	Análisis al comienzo de la semana de brote <i>VRS</i> . . . . .	37
3.7	FVP contra FFP en cada semana del periodo de 6 años . . . . .	40
3.8	FVP contra FFP en cada semana del periodo de 4 años . . . . .	41
3.9	FVP contra FFP en cada semana del periodo de 2 años . . . . .	42
4.1	Diagrama que describe la dinámica del <i>VRS</i> . . . . .	58
4.2	Comparación de datos reales y simulados (EDO) . . . . .	66
5.1	Proceso que sigue cada agente en cada iteración . . . . .	71
5.2	Comparación del SMA con la ecuación diferencial en dos años. . . . .	75
5.3	Comparación del SMA con la ecuación diferencial en tres años. . . . .	76

- 5.4 Promedio de varias simulaciones con *SMA* y el método *SIRS* clásico. 76
- 5.5 Comparación del promedio en *SMA*, EDO y datos reales en 4 años 77

[www.bdigital.ula.ve](http://www.bdigital.ula.ve)

# INTRODUCCIÓN

A través de la historia la humanidad ha soportado diferentes epidemias. Para 1580 fue documentada la primera epidemia de gripe de la humanidad llamada El Gran Catarro [1] se originó en Asia, de donde pasó a Europa y después a América. En apenas seis semanas, toda Europa fue afectada, y se dice que sólo se libró de la enfermedad el 20 % de la población. Entre los fallecidos se encontraba la reina Ana de Austria, esposa de Felipe II. En 1889 se dio inicio a una pandemia gripal en Siberia, propagándose rápidamente por toda Rusia y un año después se diseminó por toda Europa, las Américas y el resto del mundo infectando al 70 % de la población [2]. La primera pandemia del siglo XX, conocida como la Gripe Española ocasionada por el virus de Influenza A de sero-tipo H1N1 [3], se presentó entre 1918-1920; ocasionó 40 millones de muertes en el mundo y en Bogotá-Colombia 1900 muertes, afectando especialmente a la población económicamente activa entre los 20 y 40 años de edad [4]. En este mismo siglo, se repite la pandemia de gripe, en los años 1957-1958, siendo responsable un nuevo sero-tipo del virus de Influenza A H2N2 [3]. Y finalmente a inicios del siglo XXI, en marzo de 2009 una nueva pandemia de influenza azotó a la humanidad, apodada como Gripe Porcina de sero-tipo H1N1, que afectó a 1935 personas en la ciudad de Bogotá D. C.

El virus sincitial respiratorio fue aislado por primera vez en 1956 por Morris y colaboradores en un grupo de chimpancés que padecían coriza, denominándola Chimpanzee Coryza Agent. Al año siguiente Chanock y su equipo aislaron el mismo agente en dos niños diagnosticados uno de neumonía y el otro de laringitis estridulosa, denominándose a partir de ese momento virus sincitial respiratorio, debido al efecto citopático característico del virus.

Hoy en día, una de las principales causas de mortalidad de niños y adultos ma-

yores en el mundo, especialmente en países en desarrollo, son las infecciones respiratorias agudas (IRAB) del tracto respiratorio inferior, provocando aproximadamente un tercio de todas las muertes estimadas en niños menores de 5 años [5, 6, 7].

La epidemiología ambiental es parte de un nuevo paradigma que aborda desde un enfoque sistémico e interdisciplinario, el estudio de parámetros ambientales en regiones donde una enfermedad puede transmitirse. Constituye una innovadora línea de trabajo que permite, a través de datos relevados por satélites u otros equipos, la caracterización a gran escala de macrofactores que condicionan la reproducción, desarrollo y mantenimiento de un virus. Mediante la conjugación de datos alfa-numéricos de diversas fuentes con datos netamente espaciales como las imágenes de satélite, se puede considerar clasificar las variables climáticas con relación al comienzo de una epidemia. El principal objetivo en este trabajo es combinar todas las variables climatológicas que están correlacionadas con el Virus Respiratorio Sincitial (*VRS*).

Los clasificadores bayesianos han sido utilizados en distintos campos como la inteligencia artificial y el diagnóstico médico [8, 9]. Debido a la cantidad de datos que se almacenan hoy en día sobre los datos meteorológicos y la correlación que existe con las epidemias respiratorias se sugiere en este trabajo construir modelos usando los clasificadores bayesianos de Naive Bayes (*NB*) para predecir el inicio del brote del *VRS* en la ciudad de Bogotá. En estos modelos se proponen como variables predictoras distintos datos climatológicos de los años 2005 hasta el 2010. Las mismas proveen una forma compacta de representar el conocimiento y métodos exigibles de razonamiento, basados en las teorías probabilísticas, capaces de predecir el valor de variables no observadas y explicar las observadas.

Por otra parte, en este trabajo se realiza el ajuste del modelo matemático *SIRS*, a través de procesos estadísticos (mínimos cuadrados) y de optimización (algoritmo de Nelder-Mead), en relación a la dinámica poblacional referentes a la descripción epidemiológica de las enfermedades transmisibles como el *VRS*, para realizar este proceso se emplean datos reales sobre los niños menores de 5 años de edad infectados por el *VRS* en la ciudad de Bogotá desde el año 2005 hasta el 2010, donde se comparan los resultados obtenidos a través del modelo *SIRS* con los datos

reales, mostrando la pertinencia de estos modelos en los análisis epidemiológicos del *VRS*.

En la ausencia de sistemas para la detección fiable de las epidemias, los modelos y sistemas computacionales se han convertido en una importante herramienta de información para los responsables de la salud pública y el público en general. Estos modelos pueden ayudar en la creación de una visión sobre el comportamiento general de los brotes de enfermedades infecciosas mediante el análisis de la propagación en una población dada. Uno de los enfoques utilizados en gran escala en los últimos años, para la creación de modelos son los Sistemas Multi-agente (*SMA*), que son una poderosa técnica para las aplicaciones de modelado y simulación a los problemas del mundo real [10].

En este trabajo se propone un modelo basado en Multi-Agentes para la simulación y la comprensión de la dinámica del *VRS* en la población de niños menores de 5 años en la ciudad de Bogotá. El modelo propuesto utiliza para simular el proceso de transmisión del *VRS* varias realizaciones del proceso estocástico subyacente. De forma general, los modelos multi-agentes, son más versátiles, pero tienen un mayor costo computacional en comparación con los modelos basados en ecuaciones diferenciales. Finalmente, se desea analizar las diferencias entre ambos modelos propuestos en este trabajo con la finalidad de observar la conveniencia de cada modelo para futuras investigaciones.

# Virus Respiratorio Sincitial (VRS)

---

Las infecciones víricas del tracto respiratorio son una causa importante de morbilidad y mortalidad, en especial durante la infancia. El virus respiratorio sincitial (VRS) es la principal causa de patología de las vías aéreas inferiores en niños menores de 2 años, siendo la causa del 70% de los episodios de bronquiolitis, especialmente en sus formas graves. Su diseminación es tan eficaz, que casi la totalidad de la población ha sido infectada por el VRS en los primeros años de la vida. Sin embargo, la inmunidad tras la primera infección por el VRS no es completa, pudiendo observarse con frecuencia reinfecciones a lo largo de la vida que pueden afectar también a niños mayores y adultos.

## 1.1 ¿Qué es el VRS?

Es un virus de cadena simple de ARN (ácido ribonucleico) en sentido negativo de la familia de los Paramixovirus, la cual incluye virus respiratorios comunes como los que causan sarampión y parotiditis. El Virus sincitial respiratorio es miembro de la subfamilia de Pneumovirus. Está constituido por 10 proteínas fundamentales de las cuales dos son asociadas a la envoltura y son las más importantes en el desarrollo de inmunidad protectora frente a este virus. La proteína G es responsable de la adhesión del virus a las células del epitelio respiratorio y la proteína F facilita la fusión. La infección natural con el VRS no induce inmunidad y las personas pueden volver a infectarse. A veces un niño puede infectarse sintomáticamente más de una vez en la misma temporada del VSR.

La mayoría de las infecciones ocasionadas por el VRS, en la ciudad de Bogotá,

ocurren durante los meses de en la ciudad de Bogotá, durante los meses de Marzo, Abril, Mayo y Junio, esto debido en parte por la temporada de lluvias que ocurren año tras año en esta parte del trópico americano. Normalmente, en niños mayores de 5 años y adultos menores de 65 años, ocasionan cuadros virales simples como es el resfrío común, aun que en este rango de edad existen personas que terminan hospitalizadas por complicaciones en las vías del tracto respiratorio inferior (ver Figura 1.1) ocasionándoles hasta la muerte, estos caso atípicos son conocidos como infección respiratoria aguda grave inusitado (IRAG Inusitado); sin embargo, en niños menores de 5 años puede causar IRAG desencadenando hasta la muerte. El VRS infecta a la mayoría de los niños durante sus primeros 5 años de vida. Una vez una persona se infecte por el VRS y se recupere de éste, adquiere una inmunidad de corta duración y es normal que se infecte en el transcurso de su vida nuevamente con el virus [11].

Una enfermedad causada por el virus respiratorio sincitial puede ser especialmente grave en bebés de poco tiempo de nacidos o prematuros, aquellos con enfermedades del pulmón o del corazón y en los que tienen el sistema inmunológico debilitado. En bebés y niños pequeños, es una causa común de neumonía (pulmonía) y bronquiolitis (inflamación de las vías respiratorias en los pulmones). Los adultos y niños mayores pueden padecer bronquitis [12].

El virus respiratorio sincitial infecta a la mayoría de los niños durante los 3 primeros años de vida. La inmunidad contra él es de corta duración y es común padecerlo de nuevo.

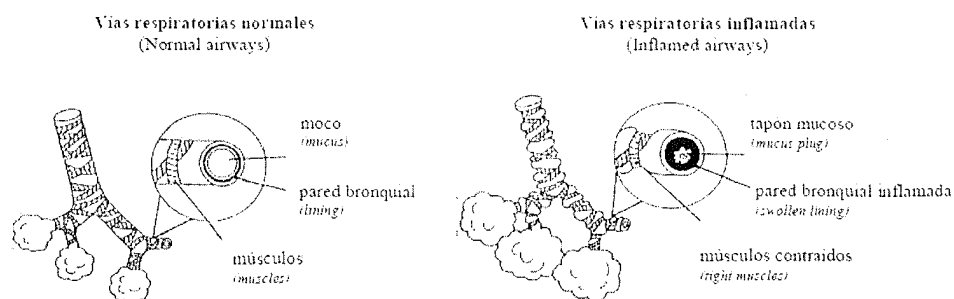


FIGURA 1.1: Comparación de las vías respiratorias

## 1.2 ¿Como se contrae este virus?

El contagio del *VRS* se presenta de persona a persona, especialmente durante los primeros días de los síntomas ocasionados por la infección. La enfermedad tiene una ventana epidemiológica o de latencia aproximadamente de 2 a 8 días, después de haber estado en contacto con el virus, una vez cumplido este tiempo, la persona presenta la enfermedad. Un individuo puede contraer el virus a través de un contacto directo con una persona infectada por el *VRS*. Esto se debe que cuando una persona tose, estornuda o habla, emite partículas de saliva y mucosas que contienen patógenos. Estas partículas aerosolizadas son de diversos tamaños, algunas de ellas del orden de los micrones que pueden alcanzar la región alveolar, al ser inhaladas por un individuo y de esta forma infectarlo. El virus puede vivir en la superficie de los objetos por horas y en las manos por más de media hora, es por ello que se recomienda el lavado de manos de las personas infectadas por virus respiratorios en general [12].

## 1.3 Relación del clima con el *VRS*

Las infecciones respiratorias siguen siendo más frecuentes en niños menores de 5 años y en menos medida en niños entre los 6 y 12 años (primera infancia y segunda infancia correspondientemente) [13, 14], las cuales representa el 30% en consulta externa y entre un 10 y 30% de las hospitalizaciones que ocurren en un año [13]. No hay diferencias en su incidencia entre países desarrollados y en vías de desarrollo; aunque sí existen diferencias significativas en la mortalidad, porque en los países desarrollados ésta, es más baja. Mientras que una persona que se encuentran entre el rango de edad (6-64 años) tienen de 1 y 3 infecciones de vías altas por año, los menores de 5 años tienen entre 8 y 10 infecciones de vías altas por año, de las cuales una tercera parte se extenderán a las vías inferiores ocasionando un IRAG [13].

Las infecciones respiratorias de vías bajas (neumonía) son la principal causa de hospitalización pediátrica, llegando a representar hasta un 51,2% de las hospitalizaciones, lo que corresponde a 27 niños por cada mil en el año [15]. En Estados

Unidos anualmente se reportan epidemias en las épocas de primavera y especialmente durante el invierno, casos de bronquitis, bronquiolitis y neumonías sobrecargan los servicios de urgencia pediátricos que generan picos máximos de ocupación hospitalaria, conllevando a la saturación de los servicios de salud. Las principales infecciones respiratorias están relacionadas a microorganismos especialmente los virus y dentro de éstos, el que toma mayor importancia es el *VRS* por su patogenicidad, siendo el principal responsable de los brotes epidémicos, que suelen presentarse con periodicidad anual y que persisten por varios meses [16, 17, 18, 15, 14]. Adicionalmente, en niños menores de 1 año el *VRS* es el principal causante de bronquiolitis (75% aproximadamente) y es un agente muy importante productor de neumonías (25%) [16, 17, 18, 15, 14].

En los Estados Unidos más de 91.000 niños son ingresados anualmente con infección respiratoria causada por el *VRS*, de estos 4.500 mueren aproximadamente y generan un coste anual estimado por hospitalización superior a 300 millones de dólares [19]. Ante todos estos datos, nos planteamos cuál es la incidencia real de la enfermedad causada por el *VRS* y si son las epidemias prolongadas o recortadas por las características climáticas peculiares de la ciudad de Bogotá D.C, Colombia.

Las IRAG son la principal causa de consultas médicas y hospitalizaciones de niños en el mundo [20]. Se le atribuye al *VRS* la mayor causa de infecciones respiratorias agudas bajas (IRAB) y de hospitalizaciones en niños menores de 5 años [21]. En los climas templados las infecciones por *VRS* se presentan como epidemias anuales, comenzando en otoño y termina en primavera. Sin embargo, el inicio de cada epidemia puede variar no sólo de año en año, sino también en las diferentes regiones [22, 23]. No se ha definido claramente que factores determinar la variación en la aparición de infecciones por *VRS* en una región, pero varios estudios han encontrado que los factores ambientales asociados con el *VRS* pueden ser la causa de las infecciones [24, 25, 26].

La temperatura ha sido inversamente asociada con las actividades de los virus respiratorios, aunque el mayor número de casos no necesariamente coincide con la menor temperatura [24, 25, 26]. Los patrones de incidencia estacionales son una

característica distintiva de los virus respiratorios. Históricamente, las bajas temperaturas han estado ligadas a eventos de la salud, y los aumentos de infecciones respiratorias durante el invierno han demostrado tener incidencia en el aumento de la mortalidad global [27]. Las variaciones geográficas relacionadas con la temperatura, estación lluviosa, y humedad relativa han sido identificadas en estudios epidemiológicos realizados en todo el mundo. En países con climas templados, como la Argentina, la prevalencia de infecciones respiratorias aumenta durante los meses más fríos y húmedos. Por ejemplo, el virus sincitial respiratorio e influenza están normalmente limitados al invierno, mientras que la incidencia de rinovirus y adenovirus se mantiene a lo largo de todo el año con incrementos periódicos en su frecuencia [28]. Por otro lado, en los países tropicales el aumento se relaciona con la estación lluviosa del año [29]. En Hong Kong, el pico de casos por *VRS* también ocurre durante la estación lluviosa, aunque para esa época su temperatura sea cálida [30].

Estudios anteriores sugieren que el inicio de la temporada del *VRS* depende de los datos del clima, por lo cual, es posible pensar en que el brote de la epidemia se puede predecir. Esta predicción podría ser útil en lugares del mundo donde no hay una continua vigilancia viral disponible. En éste trabajo se construyeron varios modelos con el fin de predecir el inicio del brote del *VRS* utilizando datos climatológicos de la ciudad de Bogotá de los años 2005 hasta el 2010.

## 1.4 Datos del clima e infectados por el *VRS*

Con la finalidad de aplicar los clasificadores de Naive Bayes (*NB*) para predecir el brote del *VRS* se utilizaran datos climatológicos. Los datos meteorológicos del año 2005 al 2010 de la ciudad de Bogotá D.C se obtuvieron de la Administración Nacional Oceánica y Atmosférica (NOAA) National Climatic Data Center (<http://cdo.ncdc.noaa.gov/CDO/CDO>), del aeropuerto "El Dorado".

El Laboratorio de la Salud Pública se encarga de diagnosticar los principales virus respiratorios (Influenza A, Influenza B, Virus Sincitial Respiratorio (*VSR*), Adenovirus y Parainfluenza 1,2 y 3) por la técnica de inmunofluorescencia indirecta.

Esta base de datos se obtuvo de la Secretaria Distrital de la Salud y el Instituto Nacional de Salud. En las tablas 1.1, 1.2, 1.3, 1.4, 1.5 y 1.6 observamos los datos de los casos del VRS en los niños menores de cinco años por semana epidemiológicas para el período del 2005-2010.

Semanas	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Casos	2	0	0	3	0	1	1	2	1	0	4	3	14	7
Semanas	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Casos	11	20	25	20	18	10	10	6	5	10	6	5	3	0
Semanas	29	30	31	32	33	34	35	36	37	38	39	40	41	42
Casos	0	3	1	1	1	0	1	0	0	0	0	1	0	0
Semanas	43	44	45	46	47	48	49	50	51	52				
Casos	0	0	1	0	0	2	1	1	1	0				

TABLA 1.1: Datos provenientes de la Secretaria de Salud Distrital de Bogotá para las semanas correspondientes al año 2005 de los casos de VRS en niños menores de 5 años de edad.

Semanas	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Casos	0	1	0	0	2	1	2	5	3	4	6	8	9	14
Semanas	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Casos	14	13	23	6	14	20	19	11	7	13	4	7	6	4
Semanas	29	30	31	32	33	34	35	36	37	38	39	40	41	42
Casos	2	2	0	0	1	3	1	0	1	0	0	0	0	0
Semanas	43	44	45	46	47	48	49	50	51	52				
Casos	0	0	0	0	1	0	0	0	2	0				

TABLA 1.2: Datos provenientes de la Secretaria de Salud Distrital de Bogotá para las semanas correspondientes al año 2006 de los casos de VRS en niños menores de 5 años de edad.

Semanas	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Casos	0	0	0	0	1	0	0	0	1	2	2	0	2	5
Semanas	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Casos	4	8	6	4	9	2	4	2	0	1	1	1	0	1
Semanas	29	30	31	32	33	34	35	36	37	38	39	40	41	42
Casos	0	0	0	0	0	0	0	3	5	1	1	0	1	0
Semanas	43	44	45	46	47	48	49	50	51	52				
Casos	0	0	1	1	1	2	2	1	5	3				

TABLA 1.3: Datos provenientes de la Secretaria de Salud Distrital de Bogotá para las semanas correspondientes al año 2007 de los casos de VRS en niños menores de 5 años de edad.

Semanas	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Casos	2	2	3	7	7	3	9	6	10	16	25	5	34	25
Semanas	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Casos	23	29	25	17	13	22	10	14	11	6	3	11	6	4
Semanas	29	30	31	32	33	34	35	36	37	38	39	40	41	42
Casos	4	7	6	5	2	3	2	1	1	3	1	0	2	1
Semanas	43	44	45	46	47	48	49	50	51	52				
Casos	7	1	3	2	2	2	7	1	2	3				

TABLA 1.4: Datos provenientes de la Secretaria de Salud Distrital de Bogotá para las semanas correspondientes al año 2008 de los casos de VRS en niños menores de 5 años de edad.

CASOS DE VRS EN SEMANAS EPIDEMIOLÓGICAS 2005-2010 EN NIÑOS MENORES DE 5 AÑOS

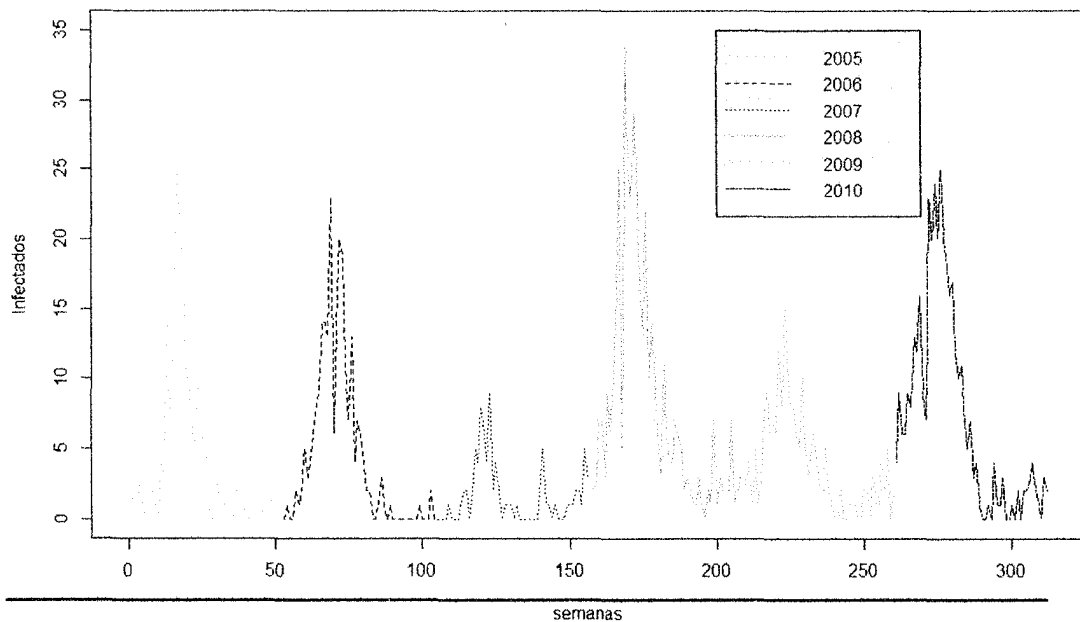


FIGURA 1.2: Datos del Instituto Nacional de Salud de la ciudad de Bogotá

Semanas	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Casos	2	3	4	1	5	1	3	7	9	6	7	6	12	8
Semanas	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Casos	15	11	8	8	6	5	10	5	3	5	6	4	2	2
Semanas	29	30	31	32	33	34	35	36	37	38	39	40	41	42
Casos	5	2	1	0	0	2	0	0	1	1	1	0	0	2
Semanas	43	44	45	46	47	48	49	50	51	52				
Casos	2	0	3	1	1	4	2	5	0	2				

TABLA 1.5: Datos provenientes de la Secretaria de Salud Distrital de Bogotá para las semanas correspondientes al año 2009 de los casos de VRS en niños menores de 5 años de edad.

Semanas	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Casos	4	9	6	6	9	8	13	12	16	9	7	23	20	24
Semanas	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Casos	20	25	20	19	16	17	12	10	11	8	5	7	3	4
Semanas	29	30	31	32	33	34	35	36	37	38	39	40	41	42
Casos	1	0	0	1	0	4	1	1	3	0	0	1	0	2
Semanas	43	44	45	46	47	48	49	50	51	52				
Casos	0	2	2	3	4	2	1	0	3	2				

TABLA 1.6: Datos provenientes de la Secretaria de Salud Distrital de Bogotá para las semanas correspondientes al año 2010 de los casos de VRS en niños menores de 5 años de edad.

Como se puede observar en la Figura 1.2 durante el periodo 2005-2010 existen picos o brotes epidemiológicos para el VRS, uno por cada año. De esta forma lo que se desea es poder crear un modelo matemático utilizando los clasificadores de Naive Bayes y los datos climatológicos para predecir el brote epidémico de VRS en la ciudad de Bogotá D. C. y así mejorar las medidas de promoción, prevención y control en la ciudad. Adicionalmente, se espera que esta metodología pueda ser aplicable a otras regiones con sus datos propios.

## Clasificadores de Naive Bayes (NB)

---

Los clasificadores de Naive Bayes(NB) son un tipo de grafos o redes Bayesianas que han sido de gran ayuda para resolver problemas en el área de la medicina y otros campos de la ciencia [31].

### 2.1 Redes Bayesianas

Una red bayesiana es un grafo acíclico dirigido en el que cada nodo representa una variable y cada arco una dependencia probabilística. La variable a la que apunta el arco es dependiente (en términos de causa-efecto) de la que está en el origen de éste. En cada variable se especifica la probabilidad condicional dados sus padres en caso de tenerlos. Una red bayesiana tiene al menos una raíz (nodo sin padre alguno) y un nodo terminal (nodo sin hijos).

En una red bayesiana cualquiera, el cálculo de la probabilidad conjunta de las  $n$  variables se puede representar como:

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | X_{j(i)}) \quad (2.1)$$

donde  $X_{j(i)}$  con  $j = 1, \dots, m$  es el conjunto de padres de la  $X_i$ .

Las redes bayesianas han sido utilizadas en diversas áreas de aplicación como por ejemplo el diagnóstico médico [31]. Las mismas proveen una forma compacta de representar el conocimiento y métodos flexibles de razonamiento, basados en las teorías probabilísticas, capaces de predecir el valor de variables no observadas y explicar las observadas [32].

Entre las características que poseen las redes bayesianas, se puede destacar que permiten aprender sobre relaciones de dependencia y causalidad, permiten combinar conocimiento con datos, evitan el sobre-ajuste de los datos y pueden manejar bases de datos incompletas [32].

El obtener una red bayesiana a partir de datos, es un proceso de aprendizaje que se divide en dos etapas: el aprendizaje estructural y el aprendizaje paramétrico. La primera de ellas, consiste en obtener la estructura de la red bayesiana, es decir, las relaciones de dependencia e independencia entre las variables involucradas. La segunda etapa, tiene como finalidad obtener las probabilidades a priori y condicionales requeridas a partir de una estructura dada [32].

## 2.2 Construcción de las redes bayesianas

La red bayesiana puede ser construida valiéndose exclusivamente del conocimiento experto, sin embargo se han publicado diversos métodos, para el aprendizaje desde un conjunto de datos [32]. Publicaciones recientes han propuesto métodos que combinan ambos enfoques [33]. En este trabajo la red bayesiana es construida utilizando los valores de las variables climatológicas y datos sobre la aparición del brote epidémico. Como se menciono anteriormente, la creación de una red bayesiana, se divide en dos etapas importantes:

### **Aprendizaje estructural**

Consiste en establecer la estructura de la red, es decir las relaciones de dependencia e independencia de las variables que la conforman.

### **Aprendizaje paramétrico**

Consiste en completar las tablas de probabilidades asociadas a cada nodo una vez establecida la estructura de la red.

## 2.3 Evidencias y Propagación de Probabilidades

Si tenemos la estructura de una red y las distribuciones de probabilidad para cada nodo, queremos determinar el cambio en estas probabilidades cuando los valores de algunas variables llegan a ser conocidos. El proceso de clasificar estas

variables de entrada y propagar sus efectos a través de la red es lo que se llama propagación de probabilidades. La propagación de probabilidades está relacionada al razonamiento probabilístico y consiste en propagar los efectos de la evidencias a través de la red para conocer la probabilidad a posteriori de las variables [34]. Existen otras características de la topología de la red que afectan la propagación de probabilidades y por lo tanto la predicción del modelo. La Separación  $D$  define un bloqueo en la propagación de las probabilidades en las circunstancias que son explicadas a continuación [34].

## 2.4 Separación $D$

Dos conjuntos de nodos,  $X$  e  $Y$ , son condicionalmente independientes, en función de un conjunto  $E$ , cuando todas las rutas no dirigidas que van de un nodo de  $X$  a un nodo de  $Y$  están bloqueadas. Dependiendo de la forma en que están conectados los nodos, una ruta puede estar bloqueada en las siguientes situaciones:

### 2.4.1 Conexiones en Serie

Considerando la situación de la Figura 2.1, la evidencia en  $A$  tiene influencia en  $B$  y esta a su vez tiene influencia en  $C$ . De la misma forma, la evidencia en  $C$  tiene influencia en la probabilidad de  $A$  a través de  $B$ . Por otra parte, si el estado de  $B$  es conocido, la ruta está bloqueada, por lo tanto  $A$  y  $C$  pasan a ser independientes. Entonces se puede decir que  $A$  y  $C$  están  $D$ -Separadas. La evidencia puede ser transmitida en una conexión en serie a menos que el estado de una de las variables dentro de la conexión sea conocido.

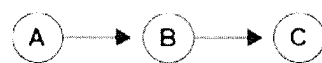


FIGURA 2.1: Ejemplo de tres nodos conectados en serie.

## 2.4.2 Conexiones Divergentes

En una conexión divergente como la que se muestra en la Figura 2.2 la influencia puede pasar a través de los hijos de  $A$ , a menos que  $A$  sea conocida (tenga evidencia), en cuyo caso se puede decir que  $B, C, \dots, E$  están  $D$ -Separadas dado  $A$ . La evidencia puede ser transmitida a través de una conexión divergente a menos que la variable que las conecta esté instanciada [34].

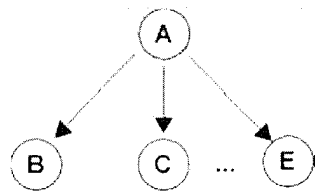


FIGURA 2.2: Ejemplo de una conexión divergente.

## 2.4.3 Conexiones convergentes

El análisis de la Figura 2.3 requiere una mayor atención. Si no se sabe nada acerca de  $A$ , excepto que puede ser inferida por el conocimiento de sus padres  $B, C, \dots, E$ , entonces los padres son independientes, es decir, una evidencia en uno no tiene ningún efecto sobre los otros. Sin embargo, si se conoce sobre la variable dependiente ( $A$ ), la evidencia en uno de sus padres hará variar la probabilidad de los otros [34]. La evidencia puede ser solo ser transmitida en una conexión convergente si la variable en la conexión o uno de sus descendientes ha recibido evidencia.

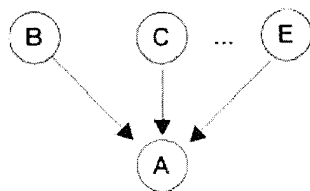


FIGURA 2.3: Ejemplo de una conexión convergente.

En general, una ruta está bloqueada (en función de nodos  $E$ ) si en ella se encuentra un nodo  $Z$  y se cumple:

- Z está en E y una de las flechas entra y la otra sale de Z.
- Z está en E y las dos flechas salen de Z.
- Ni Z ni sus descendientes están en E y las dos flechas de la ruta entran a Z.

Por otra parte, se define el Markov Blanket (ver Figura 2.4) para un nodo  $n$ , como el conjunto de la unión de los padres de  $n$ , los hijos de  $n$  y los padres de los hijos de  $n$  [34]. Con esta definición se puede concluir que si todas las variables en el Markov Blanket de  $A$ , están clasificadas, entonces  $A$  está  $D$ -Separada del resto de la red.

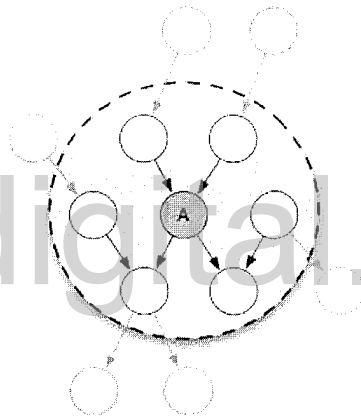


FIGURA 2.4: Markov Blanket de un Nodo A.

## 2.5 Clasificadores de Naive Bayes (NB)

El clasificador Naive Bayes [35] es un modelo de clasificación construido bajo la premisa de que todas las variables predictoras son condicionalmente independientes dada la variable clase  $C$ ; es decir, la conexión es del tipo divergente. Sin embargo, en el dominio de problemas reales esta suposición no se cumple en numerosas ocasiones. A pesar de ello el paradigma Naive Bayes en algunas ocasiones ha mostrado un buen funcionamiento en problemas de este tipo [36, 37]. El clasificador Naive Bayes aplica el teorema de Bayes para predecir para cada instancia  $x$ , la clase  $c \in C$  con la máxima probabilidad a posteriori esto es:

$$P(x|c) \propto P(c) \prod_{i=1}^n P(x_i|c), \quad (2.2)$$

donde  $P(x_i|c)$  representa la probabilidad condicional de  $X_i = x_i$  dado que  $C = c$ . Como resultado el clasificador Naive Bayes predice en base a:

$$c^* = \operatorname{argmax}_c P(c) \prod_{i=1}^n P(x_i|c), \quad (2.3)$$

donde  $\operatorname{argmax}_c$  es el máximo valor de todas las probabilidades calculadas con los datos de entrada  $x_i$ .

Asumimos para  $P(x_i|c)$  la distribución gaussiana (Normal) para cada predictor con media y varianza estimada de los datos y para el caso de la distribución  $P(c)$  una distribución uniforme. La librería **e1071** en R Development Core Team contiene la función **naiveBayes** que calcula el clasificador de Naive Bayes.

En este caso tenemos,

$$P(c) = \frac{1}{n}, \quad (2.4)$$

donde  $n$  es el número de datos relacionados con el clase  $c$ ,

$$P(X_i = x_i|c) = \frac{1}{S_i \sqrt{2\pi}} \exp\left[-\frac{(x_i - \bar{x}_i)^2}{2S_i^2}\right], \quad (2.5)$$

donde  $\bar{x}_i$  y  $S_i$  son la media y la varianza respectivamente de los valores de  $X_i$  en la clase  $c$ .

## 2.6 La curva ROC

La curva característica de funcionamiento del receptor (siglas en inglés ROC) es una técnica para visualizar, organizar y seleccionar clasificadores o marcadores basándose en la capacidad de discriminación de los mismos en todo el rango de puntos de corte. Para evaluar esta habilidad discriminatoria de un marcador, es común resumir la información de la curva ROC en un único valor global o índice [38].

## 2.7 Capacidad de Discriminación de un Clasificador

Un clasificador puede ser considerado como una variable explicativa que clasifica en cada caso o sujeto, la presencia o ausencia de una condición, que se considera la variable respuesta. La información que suministra un clasificador, depende de su nivel de medición: nominal si sus valores son categorías, binaria si es nominal con sólo dos categorías, ordinal si las categorías pueden ser ordenadas de alguna manera, discreta si puede tomar un número finito de posibles valores distintos y continua si puede tomar cualquier valor en un rango finito o infinito de valores. En este trabajo el clasificador considerado suministra valores discretos.

En situaciones de clasificación en las que se usan sólo dos clases, cada objeto, sujeto o caso es etiquetado con uno de los elementos del conjunto positivo (S), negativo (N), augurándose la clase a la que pertenece dicho caso. Algunos modelos de clasificación producen una salida continua, como la estimación de la probabilidad de un caso de pertenecer a una clase, situación en la que diferentes umbrales de decisión o puntos de corte pueden ser aplicados para predecir la clase a la que pertenece dicho caso, y otros producen apenas la etiqueta discreta de una clase, indicándose con esto la clase predicha de ese caso [38]. La aplicación de un marcador, en situaciones como la descrita, produce cuatro posibles resultados:

- Si el caso es positivo y es clasificado como positivo se cuenta como un verdadero positivo.
- Si el caso es positivo y es clasificado como negativo se cuenta como un falso negativo.

Clases hipotetizadas (Y)	Clases verdaderas (D)	
	P	N
Si	Verdadero Positivo	Falso Positivo
No	Falsos Negativos	Verdaderos Negativos

TABLA 2.1: Matriz de confusión.

- Si el caso es negativo y es clasificado como negativo se cuenta como un ver-

dadero negativo.

- Si el caso es negativo y es clasificado como positivo se cuenta como un falso positivo.

Dado un marcador y un conjunto de sujetos, una matriz de confusión o tabla de contingencia 2 x 2 puede ser construida para representar la disposición de dicho conjunto (ver Tabla 2.1).

## 2.8 Probabilidades de Clasificación de los Marcadores

Los números que se encuentran a lo largo de la diagonal principal de la matriz de confusión representada en la Tabla 2.1 representan las clasificaciones correctas y los que están a lo largo de la diagonal secundaria representan los errores (la confusión) entre las clases. Esta matriz es la base o soporte para varios indicadores comúnmente utilizados en sistemas de diagnóstico [38]:

- Fracción de Verdaderos Positivos (TPF) denominada también Tasa de Aciertos o Recall.

$$TPF = \frac{\text{Positivos clasificados correctamente (Verdaderos positivos)}}{\text{Total de positivos}} \quad (2.6)$$

- Fracción de Falsos Positivos (FPF) denominada también Tasa de Falsa Alarma.

$$FPF = \frac{\text{Negativos clasificados incorrectamente (Falsos positivos)}}{\text{Total de Negativos}} \quad (2.7)$$

- Precisión.

$$\text{Precision} = \frac{\text{Verdaderos Positivos}}{\text{Verdaderos Positivos} + \text{Falsos Positivos}} \quad (2.8)$$

- Exactitud.

$$\text{Exactitud} = \frac{\text{Verdaderos Positivos} + \text{Verdaderos Negativos}}{\text{Positivos} + \text{Negativos}} \quad (2.9)$$

El autor Fawcett [39] comenta que algunos términos adicionales asociados con las curvas ROC son:

- Sensibilidad = Fracción de Verdaderos Positivos.
- Especificidad (TNF) definida como la fracción siguiente:

$$\frac{\text{Verdaderos Negativos}}{\text{Falsos Positivos} + \text{Verdaderos Negativos}} = 1 - FPF \quad (2.10)$$

- Valor Predictivo Positivo = Precisión.

La autora Pepe [40] adapta la matriz de confusión presentada en la Tabla 2.1 usando:

- Una variable binaria denominada  $D$  para denotar el estado de la enfermedad, en vez de la clase verdadera:

$$D = \begin{cases} 1 & \text{Para Enfermedad Presente} \\ 0 & \text{Para Enfermedad No Presente} \end{cases}$$

- Una variable denominada  $Y$  que representa el resultado de la prueba, en vez de la clase hipotetizada:

$$Y = \begin{cases} 1 & \text{positivo Para Enfermedad} \\ 0 & \text{negativo Para Enfermedad} \end{cases}$$

y explica que como  $1 - TPF$  es igual a la Fracción de Falsos Negativos, el par  $(FPF; TPF)$  define las probabilidades con las que los errores son cometidos al utilizar el clasificador. Esta autora comenta que, aunque a menudo se utilicen los términos de tasa de falsos positivos y tasa de verdaderos positivos, no son en realidad tasas. Son probabilidades o fracciones y como tal deben ser denominadas [40]. Sobre esta matriz adaptada al campo de la medicina, la autora Pepe [40] define:

- Prevalencia de la población para la enfermedad como:

$$p = P[D = 1], \quad (2.11)$$

- Probabilidad general del error de clasificación en función del par (FPF; TPF) y de prevalencia de la enfermedad:

$$P[Y \neq D] = p(1 - TPF) + (1 - p)FPF, \quad (2.12)$$

En la investigación biomédica los términos sensibilidad y especificidad son con frecuencia utilizados para detallar el comportamiento del marcador. La autora Pepe [40] opina que, acompañando a esta información, deben ser reportados también los errores de clasificación ( $1 - TPF$ ) y  $FPF$  para describir la exactitud diagnóstica del marcador, en vez de la probabilidad general. Toda esta información ((FPF; TPF),(1-TPF) y FPF) la considera necesaria por dos razones fundamentales:

- La primera razón es que los costes y consecuencias de los dos tipos de errores son, usualmente, verdaderamente diferentes. Los errores Falsos Negativos, es decir sujetos enfermos no diagnosticados como tal, pueden resultar en individuos que no reciban el tratamiento que necesitan para su enfermedad. La consecuencia puede ser tan grave como la muerte. Los errores Falsos Positivos tienden a ser menos graves: personas sometidas a procedimientos adicionales de diagnóstico o incluso a tratamiento. Los impactos negativos incluyen los inconvenientes personales y el estrés, pero las consecuencias a largo plazo son usualmente menor [40].
- La segunda razón para preferir el par (FPF; TPF) en vez de la probabilidad general del error de clasificación, es que esta última depende de la prevalencia de la enfermedad. Para ilustrar el problema con esto pone el ejemplo de una prueba totalmente no-informativa, que clasifica todos los sujetos como negativos (libres de la enfermedad). Si la prevalencia de la enfermedad es baja, esta prueba tendrá una probabilidad general de error baja, pero claramente es una prueba inútil [40].

## 2.9 Valores Predictivos

En el trabajo presentado en [40], como alternativa para considerar la frecuencia de los errores de clasificación para cada estatus de la enfermedad, sugieren que la exactitud puede ser cuantificada como buenas si el resultado de la prueba diagnóstica puede predecir el verdadero estatus de la enfermedad, definiendo los valores predictivos como:

$$\text{Valor Predictivo Positivo} = PPV = P[D = 1 = Y = 1], \quad (2.13)$$

$$\text{Valor Predictivo Negativo} = NPV = P[D = 0 = Y = 0]. \quad (2.14)$$

Una prueba perfecta predecirá la enfermedad perfectamente con valores  $PPV = 1$  y  $NPV = 1$ , pero por otro lado estos valores indicarían una prueba inútil que no tiene información sobre el verdadero estatus de la enfermedad, ya que  $P[D = 1 = Y = 1] = P[D = 1]$  y  $P[D = 0 = Y = 0] = P[D = 0]$ , es decir  $PPV = p$  (ver ecuación 2.10) y  $NPV = 1$ .

Los valores predictivos dependen no sólo de la capacidad discriminatoria de la prueba en la clasificación de los sujetos, sino también de la prevalencia de la enfermedad. Un bajo  $PPV$  podría ser, simplemente, el resultado de una baja prevalencia de la enfermedad o podría ser debido a que a la prueba no refleja muy bien el verdadero estatus de la enfermedad, razón por la que los valores predictivos no son usados para cuantificar la exactitud inherente de la prueba, pero pueden ser utilizados para cuantificar el valor clínico de la prueba [40].

Hay una relación directa entre los valores predictivos y las probabilidades de clasificación, aunque el conocimiento de la prevalencia de la enfermedad también sea requerido para calcular un valor a partir del otro, por lo que en realidad la distribución conjunta de  $(D; Y)$  requiere tres parámetros. Parametrizaciones naturales serían:

1.  $(TPF; FPF; p)$ ,
2.  $(PPV; NPV; \tau)$  donde  $\tau = P[Y = 1]$  es la probabilidad de una prueba positiva.

## 2.10 Espacio ROC

Los gráficos ROC son gráficos bidimensionales en los que la Fracción de Verdaderos Positivos (TPF) es representada en el eje Y y la Fracción de Falsos Positivos (FPF) es representada en el eje X. Estas gráficas muestran las compensaciones relativas entre los beneficios (verdaderos positivos) y los costes (falsos positivos).

Un clasificador discreto es el que produce una salida que representa sólo la etiqueta de la clase. Cada clasificador discreto produce un par (FPF,TPF) correspondiente a un único punto en el espacio ROC [39]. En el gráfico A de la Figura 2.5 se muestran cinco marcadores discretos identificados con las letras A hasta E. Algunos puntos importantes del espacio ROC deben ser destacados, como:

- El punto (0,0) que representa la estrategia de nunca emitir una clasificación positiva. Un clasificador así no comete errores de falsos positivos, pero también no reconoce verdaderos positivos.
- El punto (1,1) que representa la estrategia opuesta de emitir incondicionalmente clasificaciones positivas.
- El punto (0,1) que representa la clasificación perfecta. En el gráfico A de la figura 2.5, el punto D es un marcador con una actuación perfecta.

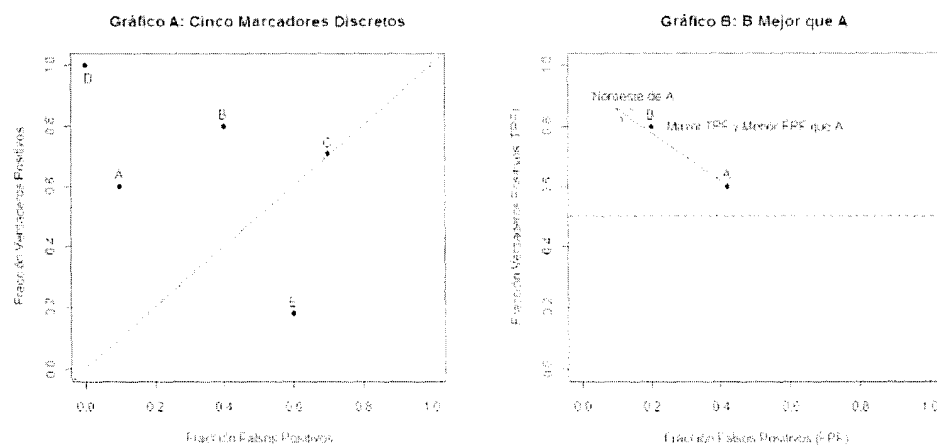


FIGURA 2.5: Gráficos ROC básicos.

En el trabajo [39] se sugiere que un punto en el espacio ROC es mejor que otro si está al noroeste (alta fracción de verdaderos positivos, baja fracción de falsos positivos o ambas) del primero (en el gráfico B de la Figura 2.5 el punto B es mejor que el A). Los clasificadores que figuren en el lado izquierdo de un gráfico ROC, cerca del eje X, podrían ser considerados conservadores (clasifican positivamente sólo con una fuerte evidencia, por lo que tienen una baja fracción de falsos positivos y también, a menudo, una baja fracción de verdaderos positivos). Por otra parte los modelos (puntos) que se encuentran en la parte superior del lado derecho pueden ser considerados liberales, es decir, clasifican positivamente con una débil evidencia, por lo que clasifican casi todos los positivos correctamente, pero a menudo, tienen altas fracciones de falsos positivos. En el gráfico A de la Figura 2.5, el modelo (punto) A es más conservador que el modelo (punto) B.

La diagonal  $y = x$ , representa la estrategia de adivinar al azar una clase. Si un marcador clasifica aleatoriamente positivo la mitad de las veces, se puede esperar que obtenga mitad positivos y mitad negativos, produciendo el punto (0.5, 0.5) en el espacio ROC; similarmente si clasifica al azar positivo 90% de las veces produciría el punto (0.9, 0.9), es decir un clasificador aleatorio produce un punto ROC que se desliza sobre la diagonal (hacia atrás y hacia adelante) basado en la frecuencia con la que adivina la clase positiva. Así, la actuación del marcador C (gráfico A de la Figura 2.5) es virtualmente aleatoria.

Cualquier clasificador que aparezca en el triángulo inferior derecho tiene una actuación o comportamiento peor que la clasificación al azar, por tanto, usualmente este triángulo está vacío en los gráficos ROC. Si se invierten los resultados de este clasificador en todos los casos, sus verdaderos positivos se convierten en falsos negativos y sus falsos positivos se convierten en verdaderos negativos. Esto significa que cualquier clasificador que produzca un punto en el triángulo inferior derecho puede invertirse para producir un punto en el triángulo superior izquierdo. En el gráfico A de la Figura 2.5, E actúa peor que un clasificador aleatorio y de hecho es la negación de B. Resumiendo, se puede decir que:

- Cualquier clasificador sobre la diagonal no tiene información sobre la clase.
- Los clasificadores que están ubicados en el cuadro superior izquierdo (el

cuadro con los vértices  $(0.5,0.5);(0.5,1);(0,0.5)$  y  $(0,1)$  son los mejores modelos si se quiere estudiar varios modelos para una mismo estudio. En particular en nuestro trabajo se desean utilizar muchos modelos (255) para predecir el inicio del brote del *VRS* de la ciudad de Bogotá.

- Un clasificador debajo de la diagonal tiene información útil, pero aplicada de forma incorrecta.

El próximo capítulo será dedicado a utilizar, todo el contenido dado en este capítulo para predecir el estallido del *VRS*, en los niños menores de 5 años de la ciudad de Bogotá para el periodo del 2005 hasta el 2010, utilizando los datos de algunas variables climatológicos como variables predictoras. Para el análisis de los resultados utilizaremos el espacio ROC.

[www.bdigital.ula.ve](http://www.bdigital.ula.ve)

## Clasificadores (*NB*) para predecir el inicio de brote del *VRS* en Bogotá

---

En este capítulo se muestran las variables climatológicas a ser consideradas en los modelos clasificadores de *NB*. Es importante resaltar que la base de datos disponible en relación a los casos de *VRS* en niños menores de cinco años en la ciudad de Bogotá corresponde al periodo 2005-2010. De esta forma los modelos construidos permiten correlacionar los datos climatológicos de los años 2005 al 2010, con la base de datos del *VRS* para ese periodo.

Por otra parte, se presenta la construcción de los modelos de *NB* con la finalidad de predecir la primera semana de brote del *VRS* para niños menores de cinco años en la ciudad de Bogotá. Adicionalmente, se muestran los detalles para generar los 3060 modelos de *NB* para los distintos periodos considerados. Por último, se presentan los resultados de los distintos modelos de *NB* y se analizan cuales de ellos presentan los mejores rendimientos en cuanto a la predicción del brote del *VRS* en la ciudad de Bogotá.

### 3.1 Selección de las variables de las variables climatológicas

En este trabajo las variables que se tomaron para predecir la semana de inicio del brote del *VRS* mediante el clasificador de Naive Bayes fueron las climatológicas. Para seleccionar las variables climatológicas se revisaron varios artícu-

los, donde se tomaron en cuenta diversas variables climatológicas [41, 42, 43]. Se identificaron investigaciones que reportaron un efecto positivo o negativo para la relación de la actividad del VRS y las condiciones meteorológicas. En varias publicaciones observamos que las variables meteorológicas que tuvieron mejor correlación con el VRS, fueron las siguientes:

1. Humedad.
2. Punto de rocío.
3. La temperatura.
4. Sensación térmica.
5. La velocidad del viento.
6. La presión atmosférica.
7. La precipitación.
8. Exposición a la luz UV.

En este trabajo, se descartó la luz UV como una variable potencial porque no hay suficientes datos disponibles para su inclusión en nuestro modelo de predicción. Se descartó el punto rocío y sensación térmica debido a que estas variables pueden ser derivadas de otras variables ya incluidas en el modelo (temperatura con humedad y velocidad del viento con temperatura respectivamente). Las variables que se consideraron en nuestros modelos fueron; la humedad, de la cual se tomaran dos valores de esta (humedad relativa mínima y máxima). Adicionalmente, se tomaron tres mediciones para representar la temperatura (temperatura media diaria, temperatura mínima diaria y temperatura máxima diaria), presión atmosférica, velocidad del viento y precipitación. Finalmente, es importante mencionar que cada una de estas variables (8) serán representadas por una medición única diaria.

### 3.1.1 Correlación de variables climatológicas

Para analizar si es posible reducir el número de variables de los posibles modelos, usamos la matriz de correlación y un diagrama de dispersión. Para una mayor claridad en la explicación del trabajo, primero daremos sus definiciones:

#### Diagrama de dispersión

Es un tipo de diagrama matemático que utiliza las coordenadas cartesianas para mostrar los valores de dos variables para un conjunto de datos. Los datos se muestran como un conjunto de puntos, cada uno con el valor de una variable que determina la posición en el eje horizontal y el valor de la otra variable determinado por la posición en el eje vertical.

La Figura 3.1 muestra un diagrama de dispersión de todas las variables climatológicas generadas con los datos desde el año 2005 hasta el 2010. Observamos que los datos están dispersos, por tal razón podemos sugerir que las variables son independientes.

#### Matriz de correlación:

Sea un experimento de  $n$  variables  $X_1, X_2, \dots, X_n$ . Podemos ordenar en una matriz los diferentes coeficientes de correlación de cada variable con el resto de variables y consigo misma, obteniendo así una matriz con cada elemento  $r_{ij}$  dado por la siguiente ecuación [44]:

$$r_{ij} = \frac{S_{x_i x_j}}{S_{x_i} S_{x_j}}, \quad (3.1)$$

donde,

$$S_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad \text{y} \quad S_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

El resultado es una matriz simétrica, con los elementos  $r_{ii}$  de la diagonal principal igual a 1.

#### Propiedades de los elementos de la matriz

- El valor  $r$  está siempre comprendido entre -1 y 1.

- Si  $r = 1$  ó  $r = -1$  entonces los puntos de la muestra están situados en línea recta (correlación lineal perfecta).
- Si  $r$  está próximo a 1 ó a -1, habrá una asociación lineal fuerte entre ambas variables.
- Si  $r$  es cercano a 0, habrá una asociación lineal muy débil.
- El valor  $r$  no varía cuando en las variables se realiza un cambio de escala o de origen. Esto demuestra que  $r$  no tiene dimensión.

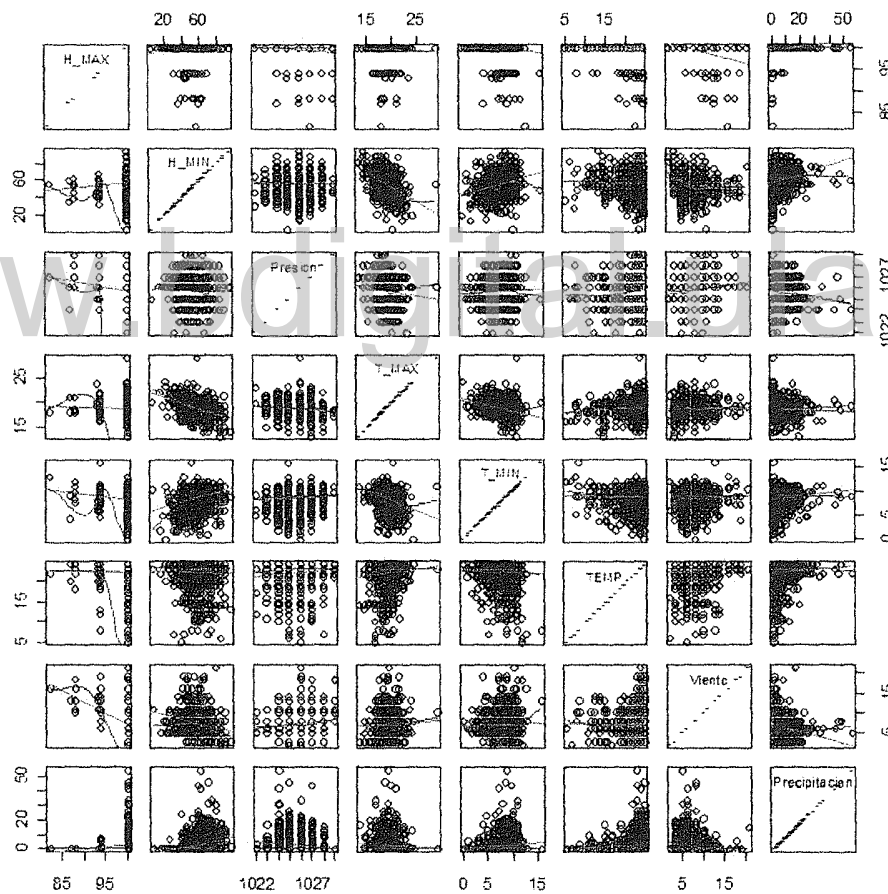


FIGURA 3.1: Diagrama de dispersión.

	Humedad MAX	Humedad Min,	Presión	Temp MAX	Temp Min,	Temp AVG	Viento,
Humedad MAX	1						
Humedad Min,	0.073	1					
Presión	-0.114	-0.021	1				
Temp MAX	-0.033	-0.439	-0.084	1			
Temp Min,	-0.086	0.289	0.036	-0.309	1		
Temp AVG	-0.003	-0.026	0.112	0.077	-0.036	1	
Viento,	-0.364	-0.243	0.133	0.026	0.039	-0.005	1
Precipitación	0.097	0.256	-0.104	-0.052	0.081	0.052	-0.208

TABLA 3.1: Matriz de correlación.

Observamos en la Tabla 3.1 que los ajustes de  $R^2$  son valores cercanos a cero; es decir podemos sugerir que las variables son independientes entre sí. Los intentos para utilizar métodos de selección de variables para reducir aún más su número en nuestros modelos no tuvieron éxito. Esta misma situación sucedió en el trabajo [8], donde se estudiaron varios modelos para predecir el inicio del brote del VRS en Salt Lake City, Utah, EEUU.

Utilizando la información que aporta el diagrama de dispersión y la matriz de correlación, se decidió incluir las siguientes variables como posibles predictoras del brote epidémico del VRS en la ciudad Bogotá:

1. Temperatura promedio (C°).
2. Temperatura máxima (C°).
3. Temperatura mínima (C°).
4. Humedad máxima (%).
5. Humedad mínima (%).
6. Presión atmosférica (hPa).
7. Precipitación (mm).
8. Velocidad del viento (Km/h).

### 3.2 Clasificador de Naive Bayes utilizando las variables climatológicas

Para analizar la efectividad de los clasificadores de *NB* se disponen de los datos climatológicos y del *VRS* correspondientes a 6 años (2005 hasta el 2010) de la ciudad de Bogotá. En este trabajo se realizarán tres estudios por separado con los datos de 2 años (2009 hasta el 2010), 4 años (2005 hasta el 2008) y 6 años (2005 hasta el 2010), para observar en cuáles semanas se obtienen los mejores modelos para predecir el brote del *VRS* usando la teoría de Espacio ROC [39]. La cantidad de modelos posibles utilizando los clasificadores de *NB* se calcularon tomando todas las combinaciones posibles de las ocho variables seleccionadas sin importar el orden de las mismas, por lo que al calcular se generan un total de 255 clasificadores *NB*. Este cálculo se obtuvo usando la combinatoria de 8 con los números desde 1 al 8 (ver Ecuación 3.2), de esta forma tenemos:

$$\binom{8}{8} + \binom{8}{7} + \binom{8}{6} + \binom{8}{5} + \binom{8}{4} + \binom{8}{5} + \binom{8}{4} + \binom{8}{3} + \binom{8}{2} + \binom{8}{1} = 255 \quad (3.2)$$

La efectividad de los 255 clasificadores de *NB* fueron calculados usando el software estadístico R Development Core Team en su versión 2.15.0 usando el paquete e1701 y otros [45].

#### Construcción de los clasificadores

Los clasificadores de *NB* se formaron con los distintos valores de las variables meteorológicas tomadas por días. El conjunto de datos meteorológicos se tomó en forma regresiva a partir del domingo siguiente después del inicio del brote del *VRS*. Como se muestra en la Figura 3.2, para el modelo de la semana 0 o la de inicio del brote, se tomaron 6 semanas (los 42 días anteriores al domingo siguiente a la semana del inicio del brote) en total como los datos de entrada para construir varios clasificadores de *NB* con las distintas variables meteorológicas. Estas semanas fueron utilizadas tomando en cuenta los siguientes pasos:

- Se usaron 4 semanas (28 días anteriores al domingo del inicio del brote) para la clase *S*, la cual indica la semana del brote.
- Se usaron 3 semanas (21 días de los cuales los primeros 14 días son los últimos utilizados en la clase *S* y los otros 7 días son de la semana anterior) para la clase *N*.
- Por último, se utilizaron 3 semanas (21 días de los cuales los primeros 14 días son los mismo que se usaron en la clase *N* nombrada anteriormente y los últimos 7 días son de la semana anterior) para la clase *N*.
- Tenemos en total 28 valores para la clase *S* y 42 valores para la clase *N*.

Los clasificadores (modelos) fueron construidos independientemente para predecir el estallido en la misma semana (semana 0), una semana antes del brote (semana 1), dos semanas en avance hacia atrás de la epidemia (semana 2) y tres semanas antes del brote (semana 3).

Para la construcción de los otros modelos para las semanas 1, 2 y 3 se procede de forma análoga a como se realizó con el modelo de la semana 0, pero tomando en cuenta los datos de entrada con la semanas en la cual se produjo el brote; es decir una semana antes del inicio del brote para la semana 1 (ver Figura 3.3), dos semanas de antelación a la semana del brote semana 2 (ver Figura 3.4) y por último tres semanas antes del inicio del brote para la semana 3 (ver Figura 3.5).

Para las distintas temporadas que se utilizaron en este trabajo se tiene que la cantidad de valores para cada una de las variables climatológicas son:

- Para el modelo que utiliza los datos de 6 años tenemos que son 168 ( $6 \text{ años} \times 7 \text{ días} \times 4 \text{ semanas}$ ) y 252 ( $6 \text{ años} \times 7 \text{ días} \times 6 \text{ semanas}$ ) valores para la clase *S* y *N* respectivamente para cada semana de antelación.
- Para el modelo que utiliza los datos de 4 años tenemos que son 112 ( $4 \text{ años} \times 7 \text{ días} \times 4 \text{ semanas}$ ) y 168 ( $4 \text{ años} \times 7 \text{ días} \times 6 \text{ semanas}$ ) valores para la clase *S* y *N* respectivamente para cada semana de antelación.

- Para el modelo que utiliza los datos de 2 años tenemos que son  $56 (2 \text{ años} \times 7 \text{ días} \times 4 \text{ semanas})$  y  $84 (2 \text{ años} \times 7 \text{ días} \times 4 \text{ semanas})$  valores para la clase S y N respectivamente para cada semana de antelación.

DATOS DE ENTRADA (días de la semanas)									BROTE SEMANA(0)	DATOS DE SALIDA
F-8	F-7	F-6	F-5	F-4	F-3	F-2	F-1			
SEMANAS DE BROTE										S
SEMANA ANTES DEL BROTE										N
2 SEMANAS ANTES DEL BROTE										N

FIGURA 3.2: Representación visual de cómo los datos de entrada están colocados en el clasificador NB en relación a la semana 0.

DATOS DE ENTRADA (días de la semanas)									BROTE SEMANA(1)	SEMANA(0)	DATOS DE SALIDA
F-8	F-7	F-6	F-5	F-4	F-3	F-2					
SEMANAS DE BROTE											S
SEMANA ANTES DEL BROTE											N
2 SEMANAS ANTES DEL BROTE											N

FIGURA 3.3: Representación visual de cómo los datos de entrada están colocados en el clasificador NB en relación a la semana 1.

DATOS DE ENTRADA (días de la semanas)									BROTE SEMANA(2)	F-1	SEMANA(0)	DATOS DE SALIDA
F-8	F-7	F-6	F-5	F-4	F-3							
SEMANAS DE BROTE												S
SEMANA ANTES DEL BROTE												N
2 SEMANAS ANTES DEL BROTE												N

FIGURA 3.4: Representación visual de cómo los datos de entrada están colocados en el clasificador NB en relación a la semana 2.

DATOS DE ENTRADA (días de la semanas)									BROTE SEMANA(3)	F-2	F-1	SEMANA(0)	DATOS DE SALIDA
F-8	F-7	F-6	F-5	F-4									
SEMANAS DE BROTE													S
SEMANA ANTES DEL BROTE													N
2 SEMANAS ANTES DEL BROTE													N

FIGURA 3.5: Representación visual de cómo los datos de entrada están colocados en el clasificador NB en relación a la semana 3.

### Ingreso de los datos en los clasificadores de *NB*

Anteriormente analizamos como se construyeron los clasificadores de *NB*, para poder predecir la semana del brote. Para cada temporada que se está estudiando se deben tomar unos datos de entrada, para observar si coinciden (sensibilidad) o no coinciden (especificidad) con los datos reales. Para el modelo de la semana 0 en cada temporada los datos de entrada fueron tomados de la siguiente manera:

- Para el modelo que utiliza los datos de 6 años se tomaron los domingos siguientes a la semana de inicio del brote de cada año; es decir se tienen un total de 6 valores climatológicos para observar si coinciden con la clase *S*.
- Para el modelo que utiliza los datos de 6 años también se tomaron los 6 domingos anteriores a la semana de inicio del brote de cada año; es decir se tienen un total de 36 valores climatológicos para observar si coinciden con la clase *N*.
- Para el modelo que utiliza los datos de 4 años se tomaron los domingos siguientes a la semana de inicio del brote de cada año; es decir se tienen un total de 4 valores climatológicos para observar si coinciden con la clase *S*.
- Para el modelo que utiliza los datos de 4 años también se tomaron los 6 domingos anteriores a la semana de inicio del brote de cada año; es decir se tienen un total de 24 valores climatológicos para observar si coinciden con la clase *N*.
- Para el modelo que utiliza los datos de 2 años se tomaron los domingos siguientes a la semana de inicio del brote de cada año; es decir se tienen un total de 2 valores climatológicos para observar si coinciden con la clase *S*.
- Para el modelo que utiliza los datos de 2 años también se tomaron los 6 domingos anteriores a la semana de inicio del brote de cada año; es decir se tienen un total de 12 valores climatológicos para observar si coinciden con la clase *N*.

Los datos de entrada para los modelos de semana 1, 2 y 3 son análogos, pero con los valores climatológicos respectivos a la semana que le corresponde. Un único clasificador NB es inducido para cada posible combinación (255 modelos) de las variables en cada temporada de 2, 4 y 6 años, y para los modelos de 0, 1, 2 y 3 semanas antes de la epidemia, generando un total de 3060 clasificadores de NB. Los resultados de los rendimientos de cada uno de los 3060 modelos incluyen los valores de la sensibilidad y especificidad para las temporadas de 2, 4 y 6 años.

### 3.3 Determinación de la semana de brote

En primer lugar es necesario determinar un estándar de referencia para la fecha de inicio del brote para analizar los modelos de cada año, para esto se necesita la ayuda de las gráficas de casos de infectados contra semanas epidemiológicas en cada uno de los años.

La base de datos del VRS de la ciudad de Bogotá es enriquecida por la Vigilancia Centinela que se realiza a través de las instituciones centinelas que han sido seleccionadas de común acuerdo entre el ente Territorial (Secretaría Distrital de Salud) y el Instituto Nacional de Salud (Colombia), para monitorear “Enfermedad Similar a Influenza” y las relacionadas con “Infección Respiratoria Aguda Grave”. En el Laboratorio de Salud Pública del Distrito que corresponde a la ciudad de Bogotá se diagnostican los principales virus respiratorios (Influenza A, Influenza B, Virus Sincitial Respiratorio (VSR), Adenovirus y Parainfluenza 1,2 y 3) por la técnica de Inmunofluorescencia indirecta. Entre los principales resultados tenemos:

- Determinación del número de casos y porcentaje de positividad del VRS por semana epidemiológica en niños menores de 5 años.
- Determinación del porcentaje del VRS entre los virus respiratorios detectados por semana epidemiológica en la Vigilancia Centinela.

En la Figura 3.6 muestra los casos de VRS en niños menores de 5 años de edad en la ciudad de Bogotá (Colombia). Utilizando estos datos hemos estimado las siguientes semanas de inicio del brote para cada año:

1. En la semana 13 en el año 2005.
2. En la semana 12 en el año 2006.
3. En la semana 14 en el año 2007.
4. En la semana 13 en el año 2008.
5. En la semana 13 en el año 2009.
6. En la semana 12 en el año 2010.

Es importante mencionar que los datos históricos que se observan en la Figura 3.6 muestran una cierta consistencia en la aparición del brote. Sin embargo, debido a los cambios climáticos estos datos a futuro podrían cambiar, como ha sucedido con la aparición de enfermedades en diferentes regiones del mundo.

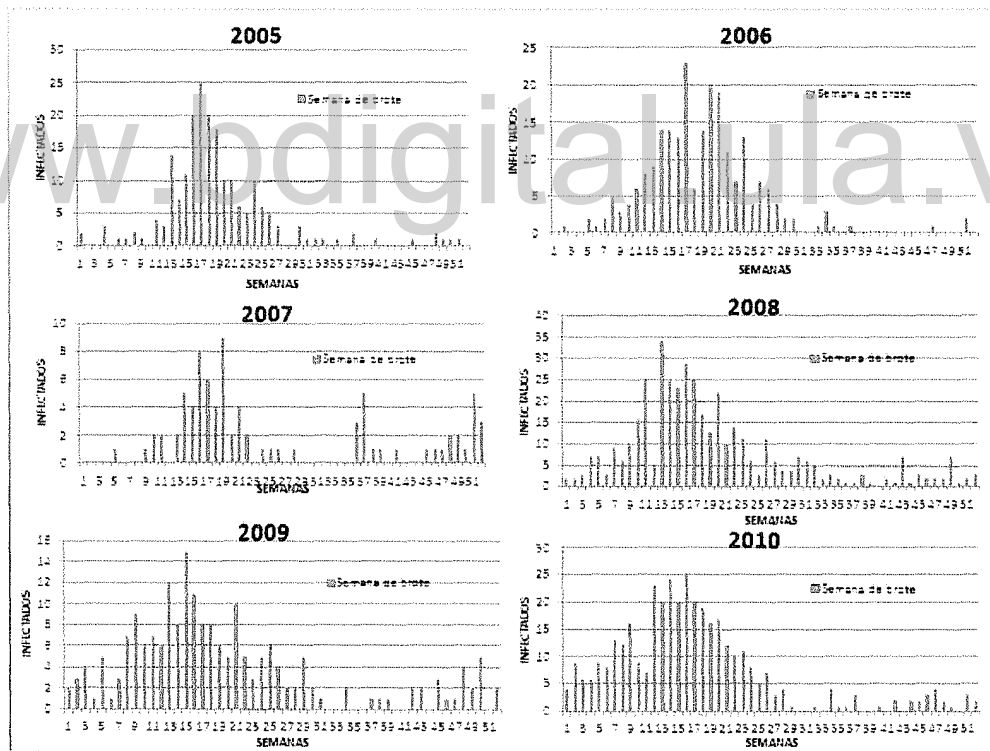


FIGURA 3.6: Gráficas utilizadas para analizar el comienzo de la semana de brote VRS.

## 3.4 Resultados

### 3.4.1 Evaluación del rendimiento de los modelos

Los 3060 modelos fueron evaluados utilizando los datos meteorológicos diarios y los datos sobre la aparición del brote del VRS por semana para cada año. Los múltiples resultados que se obtuvieron se pueden observar en las tablas del Anexo A. Los resultados correspondientes a los valores de sensibilidad y especificidad son utilizados como referencia para evaluar el rendimiento de cada modelo estudiado. Con la finalidad de tener una mayor claridad de los resultados en las tablas del Anexo A usaremos las siguientes abreviaturas para las variables nombradas entre paréntesis:

- Sens (Sensibilidad).
- Esp (Especificidad).
- TempAVG (Temperatura promedio).
- TempMAX (Temperatura máxima).
- TempMIN (Temperatura mínima).
- HumMax (Humedad máxima).
- HumMin (Humedad mínima).
- Vie (Velocidad del Viento).
- Pres (Presión atmosférica).
- Precip (Precipitación).
- REP (veces que se repite el resultado).

### 3.4.2 Clasificación de los modelos

Para decidir cuales de los modelos (Anexo A) presentan mejores rendimiento para predecir el brote usaremos la gráfica de fracción falso positivo (complemento de la especificidad) contra fracción verdadero positivo (sensibilidad). En las Figuras 3.7, 3.8 y 3.9, podemos observar los tipos de resultados o puntos, los cuales representan todos los valores obtenidos en relación a la fracción de verdadero positivo y falso positivo. Un resumen de los puntos obtenidos se muestra en la Tabla 3.2.

Modelos	DATOS	PUNTOS OBTENIDOS
Semana 0	6 años	28
Semana 1	6 años	83
Semana 2	6 años	63
Semana 3	6 años	29
Semana 0	4 años	33
Semana 1	4 años	57
Semana 2	4 años	43
Semana 3	4 años	31
Semana 0	2 años	13
Semana 1	2 años	10
Semana 2	2 años	19
Semana 3	2 años	14

TABLA 3.2: Puntos obtenidos en cada gráfica.

Utilizando el espacio ROC [39] tenemos que la zona de aceptación es la región sobre la diagonal; es decir cuando  $\{(x, y) \in [0, 1] \times [0, 1] / y > x\}$  y la zona de rechazo es la que esta por debajo de la diagonal; es decir, cuando  $\{(x, y) \in [0, 1] \times [0, 1] / y \leq x\}$ . En las Tablas desde la 3.3 hasta la 3.14, identificaremos los modelos aceptados y rechazos. En estas tablas también podemos observar las veces que aparece cada variable climática en los modelos.

Modelos	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	Total
ACEPTADOS	115	116	108	112	128	120	123	118	226
RECHAZADOS	13	12	20	16	0	8	5	10	29

TABLA 3.3: Modelos aceptados y rechazados de los clasificadores de la semana 0 utilizando el periodo de 6 años (2005 hasta 2010).

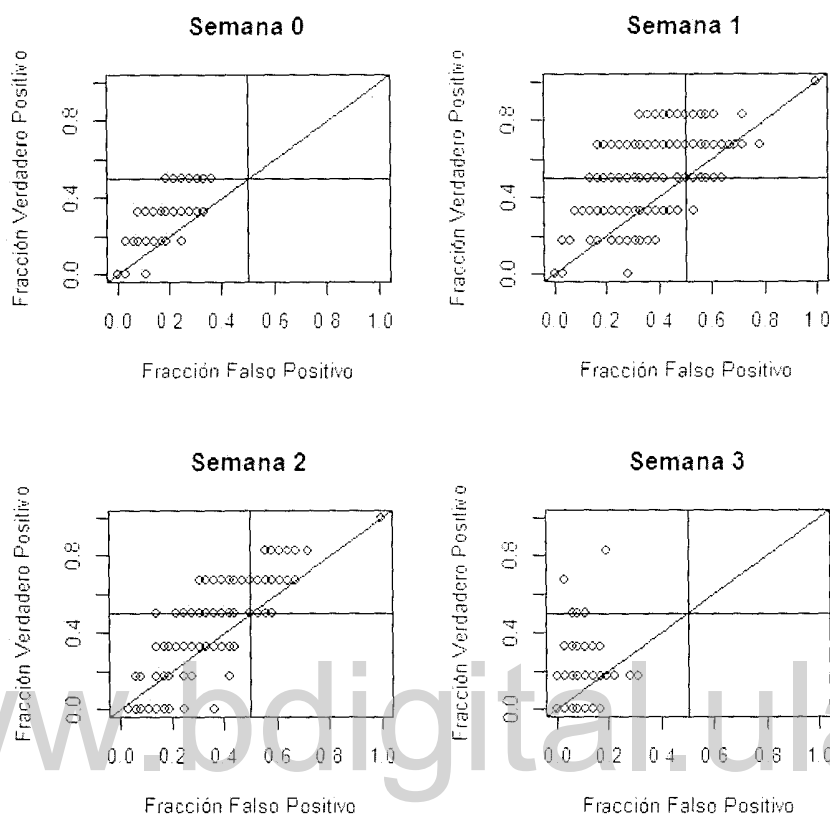


FIGURA 3.7: La fracción de verdaderos positivos (eje Y) se gráfica contra la fracción de falsos positivos (eje X). Las veces que se repite cada punto en el periodo de 6 años puede observarse en la ultima columna de las Tablas A.1, A.2, A.3 y A.4.

Modelos	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	Total
ACEPTADOS	92	108	94	94	124	84	87	100	190
RECHAZADOS	36	20	34	34	4	44	41	28	65

TABLA 3.4: Modelos aceptados y rechazados de los clasificadores de la semana 1, utilizando el periodo de 6 años (2005 hasta 2010).

Modelos	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	Total
ACEPTADOS	59	80	80	86	87	88	73	43	139
RECHAZADOS	69	48	48	42	41	40	55	85	116

TABLA 3.5: Modelos aceptados y rechazados de los clasificadores de la semana 2 utilizando el periodo de 6 años (2005 hasta 2010).

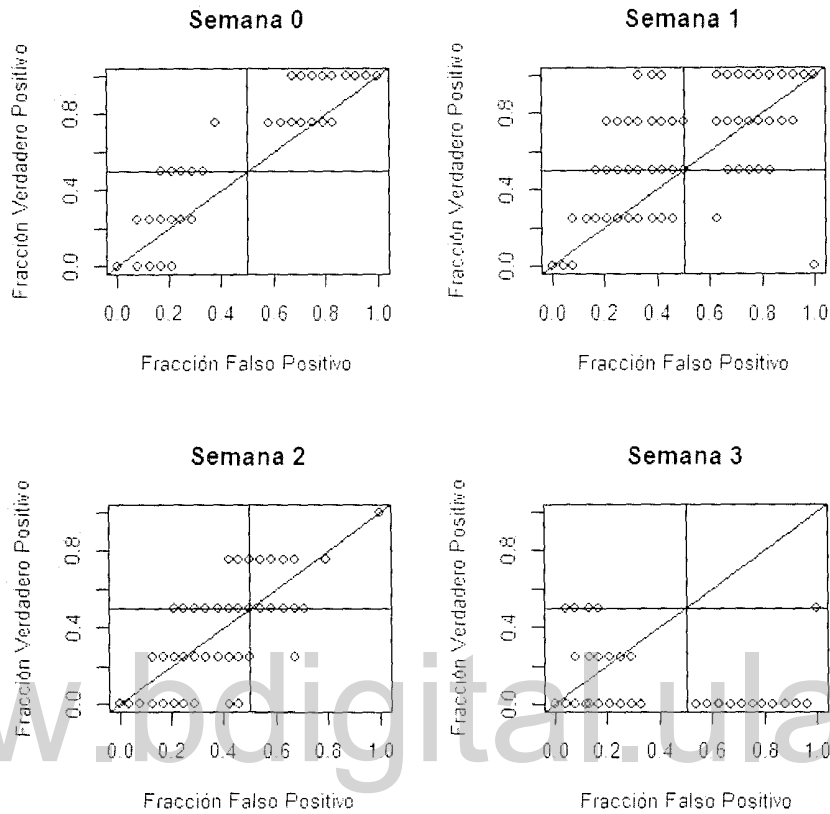


FIGURA 3.8: La fracción de verdaderos positivos (eje Y) se gráfica contra la fracción de falsos positivos (eje X). Las veces que se repite cada punto en el periodo de 4 años puede observarse en la ultima columna de las Tablas A.5,A.6,A.7 y A.8.

Modelos	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	Total
ACEPTADOS	101	61	72	49	72	68	46	52	123
RECHAZADOS	21	67	56	79	56	60	82	76	132

TABLA 3.6: Modelos aceptados y rechazados de los clasificadores de la semana 3 utilizando el periodo de 6 años (2005 hasta 2010).

Modelos	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	Total
ACEPTADOS	116	106	116	117	115	117	127	113	219
RECHAZADOS	12	22	12	11	13	11	1	15	36

TABLA 3.7: Modelos aceptados y rechazados de las semana 0 utilizando el periodo de 4 años (2005 hasta 2008).

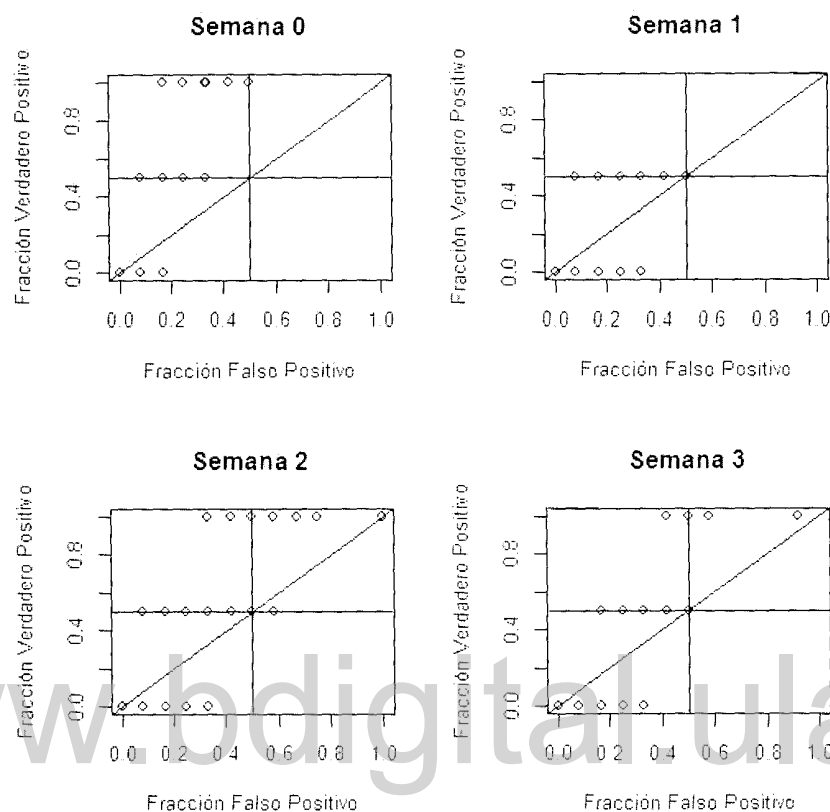


FIGURA 3.9: La fracción de verdaderos positivos (eje Y) se gráfica contra la fracción de falsos positivos (eje X). Las veces que se repite cada punto en el periodo de 2 años puede observarse en la ultima columna de las Tablas A.9, A.10, A.11 y A.12.

Modelos	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	Total
ACEPTADOS	91	99	99	86	117	88	104	97	188
RECHAZADOS	37	29	29	42	11	40	24	31	67

TABLA 3.8: Modelos aceptados y rechazados de las semana 1 utilizando el periodo de 4 años (2005 hasta 2008).

Modelos	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	Total
ACEPTADOS	50	72	70	63	91	55	55	56	121
RECHAZADOS	78	56	58	65	37	73	73	72	134

TABLA 3.9: Modelos aceptados y rechazados de las semana 2 utilizando el periodo de 4 años (2005 hasta 2008).

Modelos	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	Total
ACEPTADOS	40	20	28	2	19	26	16	20	41
RECHAZADOS	88	108	100	126	109	102	112	108	214

TABLA 3.10: Modelos aceptados y rechazados de las semana 3 utilizando el periodo de 4 años (2005 hasta 2008).

Modelos	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	Total
ACEPTADOS	127	128	126	128	128	123	126	128	250
RECHAZADOS	1	0	2	0	0	3	2	0	5

TABLA 3.11: Modelos aceptados y rechazados de las semana 0 utilizando el periodo de 2 años (2009 hasta 2010).

Modelos	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	Total
ACEPTADOS	32	43	52	16	44	31	29	30	75
RECHAZADOS	96	85	76	112	84	97	99	98	180

TABLA 3.12: Modelos aceptados y rechazados de las semana 1 utilizando el periodo de 2 años (2009 hasta 2010).

Modelos	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	Total
ACEPTADOS	61	68	73	74	62	108	82	53	138
RECHAZADOS	96	85	76	112	84	97	99	98	117

TABLA 3.13: Modelos aceptados y rechazados de las semana 2 utilizando el periodo de 2 años (2009 hasta 2010).

Modelos	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	Total
ACEPTADOS	39	37	49	58	42	49	69	10	87
RECHAZADOS	83	91	79	70	86	79	59	118	168

TABLA 3.14: Modelos aceptados y rechazados de las semana 3 utilizando el periodo de 2 años (2009 hasta 2010).

Como podemos observar en la ultima columna de las Tablas 3.3, 3.7 y 3.11, los modelos de la semana cero de cada estudio (6, 4 y 2 años) son los más aceptados. Es importante mencionar que entre los modelos que se encuentran en la zona de aceptación, solo utilizaremos para el análisis, de datos los modelos que están

ubicados al noreste de la gráfica fracción falso positivo contra fracción verdadero positivo, es decir, el cuadro superior izquierdo de cada gráfica [39]. De esta forma entre los modelos aceptados podemos realizar una subclasificación más exigente en cuanto al rendimiento de los modelos para predecir el brote del VRS. Asumiremos que los modelos de mejor rendimiento están en el área del cuadrado con los vértices (0.5,0.5); (0.5,1); (0,0.5) y (0,1) (zona noreste). En las tablas 3.15, 3.16 y 3.17, se pueden observar las veces que aparece cada variable meteorológica en los modelos de buen rendimiento.

Posteriormente, analizaremos cuales variables se encuentra más correlacionadas con la aparición del brote utilizando los modelos de mejor desempeño.

Semana	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	Total
Semana 0	81	70	76	75	120	67	78	70	138
Semana 1	57	70	56	53	98	52	56	69	135
Semana 2	30	38	37	37	53	46	36	25	70
Semana 3	0	3	6	0	2	5	1	0	6

TABLA 3.15: Modelos con buen desempeño (zona noreste) con variables climatológicas, correspondiente a los datos de 6 años (2005 hasta el 2010).

Semana	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	Total
Semana 0	45	36	40	5	50	34	52	40	74
Semana 1	39	48	43	5	61	39	44	38	85
Semana 2	37	54	47	40	72	42	44	41	84
Semana 3	15	8	6	2	7	14	1	7	15

TABLA 3.16: Modelos con buen desempeño (zona noreste) con variables climatológicas, correspondiente a los datos de 4 años (2005 hasta el 2008).

Semana	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	Total
Semana 0	127	128	126	128	128	123	126	128	250
Semana 1	32	43	52	16	44	31	29	30	75
Semana 2	59	63	62	60	54	98	70	49	124
Semana 3	39	37	49	57	42	49	68	10	85

TABLA 3.17: Modelos con buen desempeño (zona noreste) con variables climatológicas, correspondiente a los datos de 2 años (2009 hasta el 2010).

### 3.4.3 Modelos que utilizan los datos del periodo de 6 años (2005-2010).

Es importante observar que con el conjunto de datos de seis años, tenemos que la sensibilidad y la especificidad solo pueden tener los siguientes resultados:

$$\text{Sensibilidad} = \frac{i}{6}, \quad \text{con } i = 0, 1, 2, \dots, 6. \quad (3.3)$$

$$\text{Especificidad} = \frac{j}{36}, \quad \text{con } j = 0, 1, 2, \dots, 36. \quad (3.4)$$

Para estudiar los modelos de mejor desempeño, se analizó la Tabla 3.18. En esta tabla se observa que mientras más lejos se está de la semana 0 los modelos de mejor desempeño disminuyen.

Modelos	ZONA NORESTE
Semana 0	138
Semana 1	135
Semana 2	70
Semana 3	6

TABLA 3.18: Modelos de mejor desempeño (zona noreste) utilizando los datos de 6 años.

### 3.4.4 Modelos que utilizan los datos del periodo de 4 años (2005-2008)

Para los modelos que utilizan los datos de 4 años, tenemos que la sensibilidad y la especificidad puede solo tener los resultados:

$$\text{Sensibilidad} = \frac{i}{4}, \quad \text{con } i = 0, 1, 2, \dots, 4. \quad (3.5)$$

$$\text{Especificidad} = \frac{j}{24}, \quad \text{con } j = 0, 1, 2, \dots, 24. \quad (3.6)$$

Para el caso de los datos de 4 años, observamos en la Tabla 3.19 que los mejores modelos no tienen un comportamiento regular como en el caso anterior (datos de

6 años); ya que la semana 1 es la que tiene mayor cantidad de modelos con buen rendimiento. También se puede observar que la diferencia con los resultados de la semana 0 no son tan relevantes.

Modelos	ZONA NORESTE
Semana 0	74
Semana 1	85
Semana 2	84
Semana 3	15

TABLA 3.19: Modelos de mejor desempeño (zona noreste) utilizando los datos de 4 años.

### 3.4.5 Modelos que utilizan los datos del periodo de 2 años (2009-2010)

En los últimos modelos estudiados con los datos del 2009 y 2010 es importante observar que por el reducido conjunto de datos, la sensibilidad y la especificidad sólo puede tener los valores siguientes:

$$\text{Sensibilidad} = \frac{i}{2}, \quad \text{con } i = 0, 1, 2. \quad (3.7)$$

$$\text{Especificidad} = \frac{j}{12}, \quad \text{con } j = 0, 1, 2, \dots, 12. \quad (3.8)$$

Para este caso en los modelos de mejor desempeño (ver Tabla 3.20), ocurre igual que en el caso de los datos de 6 años, que la semana 0 es la que tiene mayor cantidad de modelos de mejor rendimiento.

Modelos	ZONA NORESTE
Semana 0	250
Semana 1	75
Semana 2	124
Semana 3	85

TABLA 3.20: Modelos de mejor desempeño (zona noreste) utilizando los datos de 2 años.

Observamos en las Ecuaciones (3.7) y (3.8), que en este caso la probabilidad que los resultados estén en la zona noreste es muy alta, ya que basta tener un acierto para la sensibilidad y al menos siete aciertos para la especificidad. En la Tabla 3.20 se puede observar que los modelos de buen rendimiento para la semana 0 son 250 de los 255, y en las otras semanas los modelos tienen un mayor acierto que los modelos de otros periodos.

### 3.4.6 Rendimiento individual de las variables climatológicas

La importancia relativa de una variable meteorológica contribuyendo a los mejores modelos de desempeño, se definió como el porcentaje que la variable aparece en los modelos de mejor desempeño (zona noreste). En las Tablas 3.21, 3.22 y 3.23, se presentan los porcentajes para cada variable climatológica. En este trabajo, se definió que un modelo tiene un buen desempeño si los resultados están por encima del 50 % (0.5) para la sensibilidad y la especificidad. Para los modelos que consideran el periodo de los años 2005 hasta el 2010, los porcentajes mínimos para la sensibilidad es 50 % (0.5) y especificidad es 53 % (0.53). De esta forma, tenemos que la humedad mínima es la variable que aparece con mayor frecuencia en los modelos de buen rendimiento que predicen en la misma semana del brote (semana 0), con una semana de antelación (semana 1) y para las dos semanas de antes del inicio del brote (semana 2). Sin embargo, es menos representativa en los modelos de mejor rendimiento para predecir el brote con tres semanas de antelación. Por otra parte, la temperatura mínima y el viento son las variables donde se obtuvo mejores resultados en los modelos de mejor rendimiento (zona noreste) que predicen el brote con tres semanas de antelación. Por último, podemos observar que las variables que tienen un promedio mayor al 50 % en todos los mejores modelos de las semanas son la temperatura máxima, temperatura mínima, viento y humedad mínima. Por otra parte la variable que tiene menor promedio en todos los modelos de buen rendimiento es la precipitación.

Análogamente, aquí definimos un modelo de buen desempeño utilizando los datos de 4 años, aquellos que tienen los valores de sensibilidad superiores al 54 % (0.54) y una especificidad de mayor que 53 % (0.53). Para los modelos de mejor

desempeño, la presión atmosférica y la humedad mínima son las variables que pueden verse con mayor frecuencia para predecir la semana del brote. La humedad mínima es la variable que aparece con mayor frecuencia en los modelos de mejor rendimiento, que predicen con una y dos semanas de antelación al inicio del brote. Para los modelos que predicen con tres semanas de antelación al estallido de la epidemia, las variables que aparecen con más frecuencia son la temperatura promedio y el viento. Por último, podemos observar que las variables que tienen un promedio mayor al 50% en todos los modelos de semana, son temperatura promedio, temperatura máxima, temperatura mínima, viento y humedad mínima. Por otra parte, la variable que tiene menor promedio en todos los modelos es la precipitación.

Finalmente, para definir los mejores modelos que utilizan el periodo de dos años, asumiremos que deben superar valores del 50% y 58% para la sensibilidad y especificidad respectivamente. Para la semana de inicio del brote (semana 0), se obtuvo que de los 255 modelos, un total de 250 podemos considerarlos con buen rendimiento de predicción, por lo tanto no se puede identificar la variable climatológica que tiene mayor relación al brote del VRS. La temperatura mínima es la variable que está mejor representada (69%) en los mejores modelos que predicen una semana antes del brote. En el modelo que predice con dos semanas de antelación al estallido del brote la variable que aparece con más frecuencia es la velocidad del viento con un 79%. La presión atmosférica con 80% es la que aparece mayor frecuencia en los modelos de acierto en la tercera semana antes del comienzo de la epidemia. Por último podemos observar que las variables que tienen un promedio mayor al 50% en todas las semanas son temperatura máxima, temperatura mínima, velocidad del viento, humedad mínima y presión atmosférica. Por otra parte, la variable que tiene menos promedio en todas las semanas es la precipitación.

Semana	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip
Semana 0	59 %	51 %	55 %	54 %	87 %	49 %	57 %	51 %
Semana 1	46 %	56 %	45 %	42 %	78 %	42 %	45 %	55 %
Semana 2	43 %	54 %	53 %	53 %	76 %	66 %	51 %	36 %
Semana 3	0 %	50 %	100 %	0 %	33 %	83 %	17 %	0 %

TABLA 3.21: Porcentaje de los modelos de la temporada de 6 años que incluyen cada variable en los modelos de mejor de rendimiento (Sensibilidad mínima del 50 % y una especificidad mínima del 53 %).

Semana	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip
Semana 0	61 %	49 %	54 %	7 %	68 %	46 %	70 %	54 %
Semana 1	46 %	56 %	51 %	6 %	72 %	46 %	52 %	45 %
Semana 2	44 %	64 %	56 %	48 %	86 %	50 %	52 %	49 %
Semana 3	100 %	53 %	40 %	13 %	47 %	93 %	7 %	47 %

TABLA 3.22: Porcentaje de los modelos de la temporada de 4 años que incluyen cada variable en los modelos de mejor de rendimiento (Sensibilidad mínima del 50 % y una especificidad mínima del 54 %).

Semana	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip
Semana 0	51 %	51 %	50 %	51 %	51 %	49 %	50 %	51 %
Semana 1	43 %	57 %	69 %	21 %	59 %	41 %	39 %	40 %
Semana 2	48 %	51 %	50 %	48 %	44 %	79 %	56 %	40 %
Semana 3	46 %	44 %	58 %	67 %	43 %	58 %	80 %	12 %

TABLA 3.23: Porcentaje de los modelos de la temporada de 2 años que incluyen cada variable en los modelos de mejor rendimiento (Sensibilidad mínima del 50 % y una especificidad mínima del 58 %).

### 3.5 Análisis general de los modelos

Como hemos logrado observar que, podemos clasificar los resultados por tres categorías:

- Datos utilizados por los periodos:

1. 2 años (2009-2010).

2. 4 años (2005-2008).
  3. 6 años (2005-2010).
- Semana de antelación (0, 1, 2 y 3) al inicio del brote.
  - Variables climatológicas.

De las tres categorías que nombramos anteriormente los mejores resultados son los siguientes:

### 3.5.1 Mejor periodo

Para este trabajo como utilizamos los datos de los años 2005 hasta el 2010 para un total de 6 años de datos de las variables climatológicas y los contagiados por el VRS, estos datos los dividimos en tres bases de datos, las cuales son 2 años (2009 hasta el 2010), 4 años (2005 hasta el 2008) y 6 años tomando todos los años. De estos tres conjuntos de datos se generan 1020 modelos por cada periodo utilizado. En la Tabla 3.24 se presenta la cantidad de modelos con buen rendimiento (zona noreste) para cada una de los periodos utilizados.

Observamos en la Tabla 3.24 que los mejores resultados se produjeron con los datos que utilizan el periodo 2009 hasta el 2010. Este resultado parece lógico, debido al número de combinaciones de valores (ver ecuaciones 3.8, 3.7) de la sensibilidad y especificidad, con relación a los otros dos periodos (ver ecuaciones 3.3, 3.4, 3.5 y 3.6). También podemos observar que los modelos que utilizan los datos del periodo de 4 años producen menos modelos de buen rendimiento, a pesar de tener menos combinaciones de resultados con relación a la sensibilidad y la especificidad (ver ecuaciones 3.3, 3.4), que el periodo de 6 años. Podemos decir que no existe una regla que permita declarar que entre menor número de datos o combinaciones, mejor serán los rendimientos de los modelos.

Años	Mejores Modelos
2	534
4	258
6	349

TABLA 3.24: Modelos de mejor rendimiento para cada periodo (2, 4 y 6 años).

### 3.5.2 Semanas de antelación

Para los resultados de los modelos por semana de antelación al inicio del brote podemos observar en la Tabla 3.25, que el modelo de mejor rendimiento fue el que predice la misma semana de brote (semana 0). Este resultado nos sugiere que la predicción con muchas semanas de antelación, tiene un mayor déficit, y que los resultados obtenidos coinciden con los datos reales con un porcentaje de 60 % (462 aciertos de 765 modelos).

Años	Mejores Modelos
Semana 0	462
Semana 1	295
Semana 2	278
Semana 3	106

TABLA 3.25: Mejores modelos de cada semana 0,1,2 y 3 antes del inicio del brote.

En la Tabla 3.25 observamos que los modelos que utilizan la semana de inicio del brote son los que mejor resultados obtenidos. También que mientras más lejos están los datos de la semana 0, los modelos de mejor rendimiento van disminuyendo.

### 3.5.3 Modelos por variables climatológicas

Por último, para el análisis de los distintos modelos generados estudiamos la relación que tienen las variables climáticas con los modelos de mejor desempeño. Para este análisis usamos el promedio de los modelos que utilizan todos los periodos y semanas de antelación al brote. Un resumen de estos promedios pueden ser observados en la Tabla 3.26. Es importante mencionar que de los 255 modelos que utilizan distintas combinaciones de las variables climáticas solo 128 modelos contienen cada variable. De esta forma en la Tabla 3.26 se puede observar que las variables mas relevantes son:

- Humedad mínima.
- Velocidad del viento.

- Temperatura mínima.

La variable humedad mínima aparece en 62 % de los modelos de mejor rendimiento, demostrando una correlación con el comienzo del brote. Para la variable velocidad del viento aparece en el 59 % de los modelos de mejor rendimiento. Por último, la variable bajas temperatura tiene un 57 % de aparición en los modelos de mejor desempeño.

TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip
49 %	53 %	57 %	34 %	62 %	59 %	48 %	40 %

TABLA 3.26: Promedio de las variables climatológicas en todas la temporadas y semanas.

### 3.6 Conclusiones

En este capítulo se construyeron varios clasificadores de NB para predecir la semana de brote del VRS de la ciudad de Bogotá, utilizando la base de datos climatológicos y el número de casos de infectados de niños menores de 5 años de los años 2005 hasta el 2010, dividiéndola en tres periodos 2 años (2009-2010), 4 años (2005-2008) y 6 años (2005-2010). Los resultados obtenidos, basándose en la teoría ROC para los periodos de 6 y 2 años fueron los mejores (ver tablas 3.18 y 3.20), donde estos coinciden con el inicio real del brote (casos de infectados). Sin embargo, para los modelos que utilizan el periodo de 2 años es difícil determinar las variables que tienen mayor relación con el inicio del brote del VRS, por esta razón consideramos que el periodo de 6 años es el que tuvo mejores resultados. Como podemos ver en la Tabla 3.19, el periodo de los 4 años no genera una cantidad de resultados relevante, por esta razón no se tomo en cuenta estos resultados para el estudio de mejor periodo. Por otra parte, tenemos que la humedad mínima con 62 % de promedio es la variable más significativa en los modelos de mejor rendimiento para la predicción el brote. También se observó, los distintos modelos proporcionados con los datos diarios con las variables velocidad del viento (59 %) y bajas temperaturas (57 %) se observó mayor acierto en las distintas clasificaciones; es decir, los modelos que tienen un mejor rendimiento; este hecho se repite

en distintos estudios realizados en diferentes regiones del mundo [43, 30, 42]. Por otra parte, la variable precipitación no tiene un mayor relevancia para el inicio del brote; aunque los distintos reportes hospitalarios la precipitación es la que tiene mayor correlación con el inicio del brote del *VRS* [11, 12]. Por último, podemos observar (ver Tabla 3.25) la semana 0 con 462 aciertos de 765 posibles es la que obtuvo mejores resultados en las clasificaciones, donde debemos acotar que todas las semanas tenían la misma cantidad de datos para su construcción y análisis.

Para futuros estudios se sugiere incluir bases de datos de más años donde se podría conseguir mejoras de los modelos de predicción, y también incluir otras variables en relación a la contaminación del aire (concentración de  $\text{CO}$  y  $\text{SO}_2$ ). De esta forma se recomienda a la autoridades sanitarias recolectar los datos relacionados al *VRS*.

[www.bdigital.ula.ve](http://www.bdigital.ula.ve)

## Modelo SIRS estacional para el VRS

---

### 4.1 Modelado matemático de epidemias

El modelado matemático de epidemias consiste en el uso del lenguaje y herramientas matemáticas para explicar y predecir el comportamiento de agentes infecciosos y potencialmente dañinos a poblaciones humanas o animales.

#### Modelos deterministas

Son modelos matemáticos donde las mismas entradas producirán invariablemente las mismas salidas, no contemplándose la existencia del azar ni el principio de incertidumbre. Debido a esto los modelos deterministas son usados para tratar enfermedades que afectan a poblaciones grandes y a menudo surgen representados a través de ecuaciones diferenciales [46].

### 4.2 Modelo SIR

En 1927, Kermack y McKendrick [46] formularon un modelo matemático basado en ecuaciones diferenciales bastante general y complejo para describir la epidemia de malaria que sufriera la India en 1906. Estos autores establecieron en este caso, los siguientes postulados básicos:

- La enfermedad que iban a estudiar debía ser viral o bacteriana y ser transmitida directamente por el contacto persona a persona.

- Al inicio de la epidemia solamente una fracción de la población era contagiosa.
- La población sería una población cerrada, y a excepción de las pocas personas inicialmente enfermas, todas las demás eran susceptibles de enfermarse.
- El individuo sufre el curso completo de la enfermedad para al final recuperarse (sanar) adquiriendo inmunidad o morir.
- La población total de las personas sería constante o sin dinámica demográfica.

Existen varios modelos matemáticos epidemiológicos clásicos entre ellos están el *SI*, *SIR*, *SIRS*, *SEIRS*, etc. En el modelo matemático *SIR* se denota por  $N(t)$  la población total en el tiempo  $t$  en que el brote epidémico puede ocurrir y por  $S(t)$ ,  $I(t)$  y  $R(t)$  a los individuos en estado Susceptible, Infeccioso, Recuperado o muerto al tiempo  $t$  respectivamente. De esta forma, en el modelo propuesto por Kermack y McKendrick [46],  $N(t) = S(t) + I(t) + R(t)$  y viene analíticamente dado por:

$$\begin{aligned}\dot{S}(t) &= -\beta S(t)I(t), \\ \dot{I}(t) &= \beta S(t)I(t) - \gamma I(t), \\ \dot{R}(t) &= \gamma I(t),\end{aligned}\tag{4.1}$$

donde:

- $S(t)$ , es el número de personas susceptibles.
- $I(t)$ , es el número de personas infectadas.
- $R(t)$ , es el número de personas recuperadas.
- $\beta$ , es la tasa de infectados.
- $\gamma$ , la tasa de recuperados.

El valor crítico que gobierna la evolución temporal de estas ecuaciones se llama  $R_0$  ("número reproductivo básico") [47].

$$R_0 = \frac{\beta S(t)}{\gamma} \quad (4.2)$$

El número reproductivo básico  $R_0$  se define como el número de infectados secundarios causados por una única infección primaria, en otras palabras determina el número de infecciones causados por una sola persona infectada antes de su muerte o recuperación. Por lo tanto cuando mas alto es su valor, mayor es el número de las nuevas infecciones.

En el trabajo presentado en [48], si  $R_0 < 1$  cada persona que contrae la enfermedad pueden infectar al menos una persona antes de morir o de estar recuperado, por lo que el brote desaparece ( $\dot{I}(t) < 0$ ). Cuando  $R_0 > 1$ , cada persona que contrae la enfermedad infectará a más de una persona, por lo que la epidemia se extiende ( $\dot{I}(t) > 0$ ). La Ecuación (4.2), se aplica sólo para el modelo básico de Kermack y McKendrick presentado, para otros modelos SIR que tienen diferente forma, la fórmula para  $R_0$  también será diferente.

### 4.3 Modelo SIRS

El modelo matemático epidemiológico SIRS es un extensión del modelo SIR, en donde los individuos recuperados pueden perder la inmunidad a la enfermedad y volver a formar parte del grupo de susceptibles. En este trabajo utilizaremos un modelo SIRS para modelar la dinámica del VRS en la población de niños menores de 5 años de edad en la ciudad de Bogotá. En particular el flujo de un grupo a otro y las suposiciones básicas están dadas a continuación:

- El contagio es proporcional al producto de la población de susceptibles por la proporción de infecciosos sobre la población total. Esta suposición es llamada generalmente ley de la acción de masas.
- Todos los individuos son igualmente susceptibles.
- La población es lo suficientemente grande como para garantizar una descripción determinista.

- La población esta homogéneamente mezclada, es decir, cada individuo tienen la misma probabilidad de entrar en contacto con cualquier otro.
- La tasa de mortalidad  $\mu$  de todas las clases se asume igual, despreciando la tasa de mortalidad debido al VRS.
- La tasa de natalidad se asume igual a  $\mu$ .
- El valor  $\rho$  es el tiempo medio que permanece un individuo en la clase  $I(t)$ .
- Los individuos que se encuentran en la clase  $R(t)$ , vuelven a ser susceptibles a una velocidad  $\gamma$ , debido a que son posibles las reinfecciones.
- Es importante mencionar que la inmunidad celular es la responsable de la eliminación del virus en la infección primaria, pero esta inmunidad desaparece en un periodo de pocos meses [49, 50].

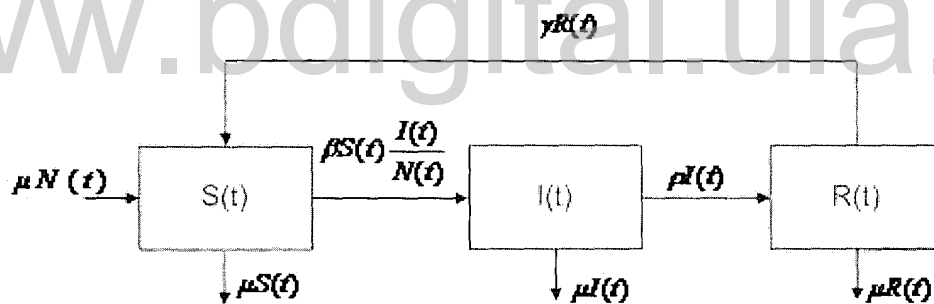


FIGURA 4.1: Diagrama del modelos que describe la dinámica del Virus respiratorio Sincitial en una población de niños menores de 5 años, en la ciudad de Bogotá D. C.

Como se aprecia en la Figura 4.1, la población total es igual a la suma de las subpoblaciones; luego  $N(t) = S(t) + I(t) + R(t)$ . Luego el modelo SIRS basado un sistema de ecuaciones diferenciales esta dado analíticamente por:

$$\begin{aligned}
 \dot{S}(t) &= \gamma R(t) + \mu N(t) - \mu S(t) - \beta S(t) \frac{I(t)}{N(t)}, S(0) = S_0 \\
 \dot{I}(t) &= \beta S(t) \frac{I(t)}{N(t)} - \mu I(t) - \rho I(t), I(0) = I_0 \\
 \dot{R}(t) &= \rho I(t) - \mu R(t) - \gamma R(t), R(0) = R_0,
 \end{aligned}
 \tag{4.3}$$

donde  $S_0 > 0$ ,  $I_0 > 0$  y  $R_0 > 0$ . La dinámica de este modelo *SIRS*, viene dada en su mayoría por la tasa de transmisión:

$$\beta S(t) \frac{I(t)}{N(t)}, \quad (4.4)$$

donde  $\beta$  es el promedio de contactos efectivos que tiene un infectado para transmitir el virus, el factor  $\frac{I(t)}{N(t)}$  es la probabilidad que un individuo de la población tenga contacto con un infectado y  $S(t)$  es la subpoblación de individuos susceptibles al virus. El producto de estos tres términos (ver Ecuación 4.4) es el valor esperado de nuevas infecciones por unidad de tiempo en la población.

## 4.4 Análisis Cuantitativo

### 4.4.1 Estados estacionarios del sistema de ecuaciones diferenciales

Los estados estacionarios del modelo *SIRS* se calculan igualando las ecuaciones del sistema (4.3) a cero. Es decir,

$$\begin{aligned} \gamma R(t) + \mu N(t) - \mu S(t) - \beta S(t) \frac{I(t)}{N(t)} &= 0, \\ \beta S(t) \frac{I(t)}{N(t)} - \mu I(t) - \rho I(t) &= 0, \\ \rho I(t) - \mu R(t) - \gamma R(t) &= 0, \end{aligned} \quad (4.5)$$

De la tercera ecuación del sistema (4.5) si hacemos a  $I(t) = 0$ , entonces  $R(t) = 0$ , ahora como  $N(t) = S(t) + R(t) + I(t)$ , tenemos el estado estacionario:

$$(S_0^*, R_0^*, I_0^*) = (N^*, 0, 0) \quad (4.6)$$

Este estado estacionario es conocido en la literatura como punto de equilibrio libre de enfermedad, debido a que toda la población es susceptible. El otro estado estacionario es el punto de equilibrio endémico el cual esta dado por:

$$\begin{aligned}
 S_2^* &= \frac{N^*(\mu + \rho)}{\beta}, \\
 I_2^* &= \frac{N^*(\gamma + \mu)\xi}{\beta\psi}, \\
 R_2^* &= \frac{N^*\rho\xi}{\beta\psi},
 \end{aligned} \tag{4.7}$$

donde,  $N^*$  es la población en el estado estacionario,  $\xi = (\beta - \mu - \rho)$  y  $\psi = (\gamma + \mu + \rho)$ , esta solución se obtiene resolviendo el sistema (4.5). El número básico de reproducción o el término  $R_0$  para este modelo SIRS se define como:

$$R_0 = \frac{\beta}{\rho + \mu} \tag{4.8}$$

#### 4.5 Modelos SIRS con estacionalidad forzada

Es bien conocido que la probabilidad de contagiarse con una enfermedad no es constante en el tiempo. Por experiencia se conoce que ciertas enfermedades son más frecuentes en invierno y otras en el verano dependiendo del clima. De igual forma, para enfermedades de la niñez existe una gran correlación con el calendario escolar. Por ejemplo, en las épocas de vacaciones la probabilidad de contagiarse con estas enfermedades es menor. Debido a esto es plausible asumir para varias clases de enfermedades incluyendo el VRS una fuerza de infección con una tasa de contagio periódica [51].

Como se ha mencionado anteriormente la probabilidad de contagiarse con una enfermedad no es constante en el tiempo. Debido a esto es plausible asumir en el modelo un  $\beta$  no constante en el tiempo para la enfermedad del VRS. De esta forma, el modelo (4.3) pierde el punto de equilibrio endémico, debido a que la población de infectados oscila en el tiempo. Con esta modificación, el modelo SIRS puede ajustarse a datos reales del VRS correspondientes a la ciudad de Bogotá, Colombia. Por lo tanto al usar una tasa de contagio periódica, el modelo SIRS (4.3) con estacionalidad forzada esta dado por:

$$\begin{aligned}
\dot{S}(t) &= \gamma R(t) + \mu N(t) - \mu S(t) - \beta(t)S(t)\frac{I(t)}{N(t)}, S(0) = S_0 > 0 \\
\dot{I}(t) &= \beta(t)S(t)\frac{I(t)}{N(t)} - \mu I(t) - \rho I(t), I(0) = I_0 > 0 \\
\dot{R}(t) &= \rho I(t) - \mu R(t) - \gamma R(t), R(0) = R_0 > 0,
\end{aligned} \tag{4.9}$$

Las enfermedades respiratorias y en especial el VRS tienen picos de casos positivos en la población al menos una vez al año debido a la posible estacionalidad del virus [51]. Por lo tanto, con la finalidad de generar las fluctuaciones en la población de infectados se considera al parámetro  $\beta(t)$  como una función periódica:

$$\beta(t) = b_0 \left( 1 + b_1 \cos \left( \frac{2(\pi + \Phi)}{T} \right) \right), \tag{4.10}$$

con  $b_0 > 0$ ,  $0 < b_1 \leq 1$  y  $0 \leq \Phi \leq 2\pi$ . Donde  $T$  representa el periodo de la enfermedad,  $b_1$  representan la estacionalidad de la transmisión,  $b_0$  representa la tasa de contacto y  $\Phi$  representa la densidad de individuos infectados. Es así que el modelo sin importar la tasa de transmisión, se ve forzado a tener soluciones periódicas [52].

## 4.6 Estimación de parámetros

Todos los datos fueron registrados en el sistema de información del Laboratorio de Salud Pública SILASP, posteriormente la información fue extraída, consolidada y tabulada. Adicionalmente, se obtuvieron frecuencias simples de las variables de interés, porcentajes y gráficas. El proceso de estimación de los parámetros  $b_0$ ,  $b_1$ ,  $\Phi$  y el factor de escala  $s$  es realizado utilizando el método de mínimos cuadrados en conjunto con el algoritmo de Nelder-Mead para conseguir el valor óptimo de los parámetros a estimar [53]. Todo el proceso de estimación de los parámetros se realiza utilizando el programa matemático MATHEMATICA versión 7.0.

### 4.6.1 Mínimos cuadrados

Es una técnica de análisis numérico encuadrada dentro de la optimización matemática, en la que, dados un conjunto de pares ordenados: (variable independiente, variable dependiente) y una familia de funciones, se intenta encontrar la función, dentro de dicha familia, que mejor se aproxime a los datos (un "mejor ajuste"), de acuerdo con el criterio de mínimo error cuadrático. Desde un punto de vista estadístico, un requisito implícito para que funcione el método de mínimos cuadrados es que los errores de cada medida estén distribuidos de forma aleatoria. La técnica de mínimos cuadrados se usa comúnmente en el ajuste de curvas. Muchos otros problemas de optimización pueden expresarse también en forma de mínimos cuadrados, minimizando la energía o maximizando la entropía.

#### En forma matemática

Sea  $(x_k, y_k)_{k=1}^n$  un conjunto de  $n$  puntos en el plano real, y sea  $f_j(x)_{j=1}^m$  una base de  $m$  funciones linealmente independientes en un espacio de funciones. Queremos encontrar una función  $f(x)$  que sea combinación lineal de las funciones base, de modo que  $f(x_k) \approx y_k$ , esto es:

$$f(x) = \sum_{j=1}^m c_j f_j(x),$$

Por lo tanto, se trata de hallar los  $m$  coeficientes  $c_j$  que hagan que la función aproximante  $f(x)$  dé la mejor aproximación para los puntos dados  $(x_k, y_k)_{k=1}^n$ . El criterio de "mejor aproximación" puede variar, pero en general se basa en aquél que minimice una acumulación del error individual (en cada punto) sobre el conjunto total. En primer lugar, el error (con signo positivo o negativo) de la función  $f(x)$  en un solo punto,  $(x_k, y_k)$ , se define como:

$$e_k = y_k - f(x_k),$$

pero tratamos de medir y minimizar el error en todo el conjunto de la aproximación,  $(x_k, y_k)_{k=1}^n$ . Por lo tanto, el método de los mínimos cuadrados se basa en la minimizar el error cuadrático medio; es decir la ecuación:

$$E_{cm}(f) = \sqrt{\frac{\sum_{k=1}^n (e_k)^2}{n}}.$$

#### 4.6.2 Algoritmo de Nelder-Mead

El método Nelder-Mead busca encontrar un mínimo local de una función de muchas variables y no requiere derivadas ni integrales. Nelder-Mead evalúa los valores de la función en los vértices de un triángulo, el peor vértice, donde  $F(x,y)$  es más grande se descarta y se reemplaza con otro vértice con lo que se forma un nuevo triángulo y la búsqueda continúa. Este proceso va generando una serie de triángulos o simplex, para los cuales los valores de la función se van haciendo más pequeños cada vez. Con los simplex reducidos se encuentran las coordenadas del mínimo local. Este método es efectivo y compacto [54].

### 4.7 Resultados

En este apartado se presentan los resultados numéricos del ajuste del modelo estacional *SIRS* (4.9) para la dinámica real del *VRS* en la población de Bogotá, Colombia.

#### Ajuste del modelo SIRS

Los datos utilizados del *VRS* corresponden a los niños menores de 5 años en la ciudad de Bogotá, Colombia. En particular, se utilizaron los datos reportados en el periodo 2005-2010 por el LSP de Vigilancia Epidemiológica SIVIGILA.

En el modelo matemático epidemiológico estacional *SIRS* (4.9) los únicos parámetros que necesitan estimarse son  $b_0$ ,  $b_1$  y  $\Phi$ . Los demás parámetros son obtenidos de la literatura médica relativa al *VRS*. Se utilizan los valores  $\rho = 1.8 \text{ años}^{-1}$  y  $\gamma = 36 \text{ años}^{-1}$  [55].

Es importante mencionar que las series de datos correspondientes al *VRS* para niños menores de cinco años en Bogotá no corresponden exactamente a la subpoblación de infectados  $I(t)$ , ya que solo una fracción  $s$  a ser determinada asiste

al médico quien reporta el caso. Adicionalmente, la población  $I(t)$  contiene todos los casos de infectados con el VRS y una porción de ellos pertenecen al grupo de niños menores de cinco años. Por otra parte, las condiciones iniciales de las subpoblaciones son desconocidas  $S(0)$ ,  $I(0)$  y  $R(0)$ . Una forma de considerar condiciones cercanas a la realidad es asumir que el VRS ha tenido una dinámica similar en la población en los años previos al período 2005-2010 donde se reportan los datos. De esta forma, los brotes epidémicos se consideran a partir de cuando el modelo estacional propuesto presenta oscilaciones estables.

El proceso de estimación de los parámetros  $b_0$ ,  $b_1$ ,  $\Phi$  y el factor de escala  $s$  es realizado utilizando el método de mínimos cuadrados y el algoritmo de Nelder-Mead para conseguir el valor óptimo de los parámetros a estimar [53]. Los resultados para el ajuste del modelo matemático a los datos de VRS en Bogotá son los siguientes:  $b_0 = 36.5$ ,  $b_1 = 0.45$ ,  $\Phi = 5.7$  y un valor de escala  $s = 31800$ . Es importante resaltar que cada región tiene sus propias características. Por ejemplo, en la Tabla 4.1 se presenta un resumen con los valores de los parámetros para distintas regiones del mundo. Posteriormente, el modelo matemático epidemiológico SIRS con estacionalidad forzada es comparado con los datos correspondientes al VRS para niños menores de cinco años en Bogotá. Esta comparación se puede observar en la Figura 4.8 correspondiente al periodo 2005-2010. Como se puede apreciar el modelo matemático explica de una forma razonable el comportamiento cíclico de los casos de niños infectados con el VRS. Se puede notar que durante el año 2007 hubo un comportamiento un poco atípico en relación a los años precedentes y posteriores, ya que se reportaron una menor cantidad de casos. Sin embargo, en el año posterior 2008 se ve un repunte en los casos [56].

Región	$b_0$	$b_1$	$\phi$
Valencia (España)	36.4	0.38	1.07
Gambia	60	0.16	0.15
Florida	62	0.10	0.14
Finlandia	44	0.36	0.60
Singapur	77	0.14	0.28

TABLA 4.1: Parámetros estimados para distintas regiones del mundo del modelo estacional SIRS para el VRS[52, 55]

## 4.8 Conclusión

En este capítulo se propone el uso del modelo matemático epidemiológico SIRS con estacionalidad forzada por un armónico simple, para dar explicación a la dinámica del Virus Respiratorio Sincitial en la población de niños menores de cinco años en Bogotá. Colombia. El modelo es ajustado utilizando específicamente los datos reportados por el LSP de Vigilancia Epidemiológica SIVIGILA correspondientes al periodo 2005-2010. El modelo matemático SIRS con estacionalidad forzada propuesto aquí, explica de una forma razonable el comportamiento cíclico de los casos de niños infectados con el Virus Respiratorio Sincitial en la ciudad de Bogotá. Colombia.

Específicamente, se puede mencionar en cuanto a los resultados numéricos que los parámetros se encuentran en concordancia con los obtenidos en la región de Valencia, España, para el periodo 2001-2005 [52]. Es importante resaltar que el factor de escala  $s$  ha resultado con un mayor valor en Bogotá. Esto debido a que Bogotá es una ciudad con más habitantes y/o que los habitantes de Bogotá asisten menos al médico en caso de presentar síntomas del VRS. En relación a este último punto pensamos que puede ser debido a recomendaciones por parte de los organismos de salud en Colombia donde se sugiere permanecer en casa cuando se presentan cuadros gripales no severos.

Es importante añadir que la función periódica utilizada para el forzamiento estacional necesita ser estudiada con mayor profundidad. Por ejemplo, su relación con variables del clima podría traer información relevante para entender la propagación del VRS en la población infantil y adulta mayor [8]. Adicionalmente, podemos mencionar que el modelo estacional SIRS es conveniente para los datos reportados de niños. Sin embargo, si los reportes incluyen adultos mayores podría ser más conveniente usar un modelo SIRS estructurado por edades [57]. En este modelo alternativo se debe resaltar que al aumentar los parámetros del modelo, la certeza de las predicciones mejora a cambio de un costo en complejidad en el modelado y estimación de los parámetros.

Finalmente, se debe resaltar que el modelo determinístico presentado en este capítulo es válido solo para poblaciones suficientemente grandes. Inclusive en caso de

poblaciones grandes también es necesario ser precavido. Por ejemplo, la tasa de contacto periódica puede producir valores infinitesimales y así algunos resultados pueden ser inválidos [58]. Futuras líneas de investigación incluyen la incorporación al modelo determinístico de factores estocásticos con la finalidad de explicar las variaciones interanuales de los niños infectados con el VRS.

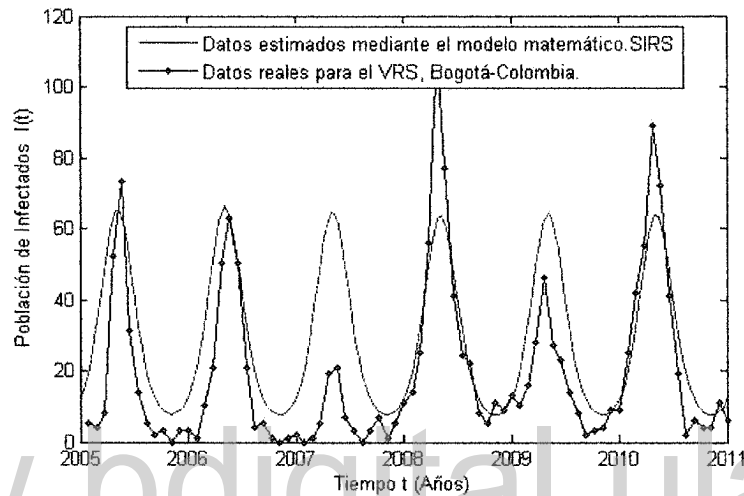


FIGURA 4.2: Datos reales para el VRS y datos estimados mediante el modelo matemático SIRS estacional (4.9). propuesto. Datos reportados de VRS en niños menores de 5 años por el LSP de Vigilancia Epidemiológica SIVIGILA correspondientes al periodo 2005-2010, Bogotá D. C.

## Sistema de Multi-Agentes (SMA)

---

En este capítulo se propone un modelo basado en Multi-Agentes para la simulación y la comprensión de la dinámica del *VRS* en la población de niños menores de 5 años en la ciudad de Bogotá. El modelo propuesto utiliza para simular el proceso de transmisión del *VRS* varias realizaciones del proceso estocástico subyacente.

Las simulaciones utilizando modelos basados en sistemas Multi-Agentes se han incrementado recientemente en diversas áreas de la ciencia. El modelo Multi-Agentes propuesto en este trabajo, está fundamentado en el modelo matemático *SIRS* estacional presentado en el capítulo anterior. Por otra parte, el modelo Multi-Agentes construido en este capítulo puede ser visto como un proceso estocástico a diferencia del modelo determinista del capítulo anterior. Como se ha mencionado en diversos trabajos, los sistemas Multi-Agentes tienen una mayor flexibilidad para modelar situaciones reales, por lo cual en muchos casos los resultados son más confiables. Sin embargo, esta mayor flexibilidad de los modelos Multi-Agentes, viene por lo general a un costo añadido de complejidad en la construcción del modelo. Adicionalmente, es importante mencionar que la teoría para estudiar los modelos basados en ecuaciones diferenciales está mucho más desarrollada, facilitando en muchos casos su análisis [59].

### 5.1 Motivación

Los modelos matemáticos epidemiológicos determinísticos como el modelo *SIRS*, han sido ampliamente utilizados para el análisis y la comprensión de epi-

demias. Adicionalmente, estos modelos han servido para examinar los efectos de diferentes estrategias para combatir las enfermedades que atacan una población [60]. Sin embargo, en este trabajo surge la idea de modelar la dinámica del *VRS* mediante un sistema Multi-Agentes, para poder explicar y reflejar de una forma más exacta el origen de los datos reales correspondiente al *VRS* en la población. Como se pudo observar en el capítulo anterior, los datos reales relacionados al *VRS* no presentan un comportamiento regular, como el que generalmente presentan los resultados de los modelos basados en ecuaciones diferenciales. Por esta razón, es necesario buscar modelos alternativos como los modelos Multi-Agentes, para poder explicar mediante un modelo la irregularidad en los datos reales relacionados al *VRS*.

En este capítulo, se desea determinar la eficacia del uso de un modelo Multi-Agentes particular para la simulación y la comprensión de la dinámica del *VRS* en la población de niños menores de 5 años en la ciudad de Bogotá. Adicionalmente, se realiza una comparación con el modelo matemático determinístico *SIRS* para determinar el modelo más idóneo para modelar la dinámica del *VRS*.

Es importante mencionar que utilizando el enfoque de los modelos Multi-Agentes, el conocimiento puede ser representado de forma más simple, utilizando las características de las enfermedades, tales como la tasa infección, duración de la enfermedad, tasa de mortalidad, entre otros. Otra, ventaja de utilizar esta técnica es generar un modelo general como base para un futuro desarrollo de un modelo más robusto. Adicionalmente, con los modelos Multi-Agentes se tienen la capacidad de construir situaciones diferentes simplemente cambiando el valor de algunas variables, mientras que los modelos matemáticos es necesario cambiar sus ecuaciones generalmente, haciendo de esta tarea más costosa [61]. Por lo tanto, se piensa que el enfoque del modelo Multi-Agentes puede ser una buena alternativa para la simulación de epidemias.

## 5.2 Un sistema Multi-Agente con arquitectura equivalente a un sistema EDO

Existen diversas formas o arquitecturas para construir un sistema Multi-Agente. Un ejemplo interesante es el trabajo presentado en [62], donde se desarrolla un SMA con una arquitectura equivalente a un sistema de ecuaciones diferenciales ordinarias (EDO) para modelar la leucemia mieloide crónica a nivel de poblaciones celulares. En [62] consideran un sistema de la forma:

$$\begin{aligned}x_1 &= f_1(x_1, x_2, \dots, x_n, t), \\x_2 &= f_2(x_1, x_2, \dots, x_n, t), \\&\vdots \\x_n &= f_n(x_1, x_2, \dots, x_n, t),\end{aligned}\tag{5.1}$$

donde cada  $x_i$  es una subpoblación con  $i = 1, 2, \dots, n$ , los autores proponen el uso de dos máquinas; "Vivir-Morir" y "Reproducir-Abortar", donde en función de las percepciones de los agentes se toman las decisiones. Estas decisiones son tomadas en función de probabilidades, las cuales dependen de las  $f_i$  del sistema de EDO original y del paso del tiempo utilizado en las simulaciones de Monte Carlo. El entorno considerado para el agente es una malla de celdas donde en cada celda puede haber un solo agente [63]. Es importante mencionar que en este modelo de la leucemia mieloide crónica considerado en [62], ninguna subpoblación de células inicialmente es cero, lo cual permite que cualquiera de las subpoblaciones sea modificada. Finalmente, es interesante mencionar que la metodología propuesta en [62] se convierte paradójicamente en el método de Euler para resolver ecuaciones diferenciales. Este, hecho no es nada obvio, ya que la metodología desarrollada en [62] hace uso de la distribución binomial y de dos máquinas de estados para el modelo Multi-Agente.

Dado que nuestro modelo SIRS está basado en un sistema de EDO, en este trabajo se intentó utilizar la metodología propuesta en [62]. En la metodología propuesta las máquinas de estado actúan sobre cada agente. Sin embargo, si no existen agentes de alguna subpoblación, esta no podría ser modificada. Esta situación trae como consecuencia que si alguna subpoblación o clase no tiene agentes, esta

no se vera alterada durante la evolución del sistema. Por otra parte, en el modelo *SIRS* para el *VRS* la subpoblación de recuperados  $R(t)$  en general puede ser igual a cero y luego incrementarse debido a la llegada de agentes infectados que se han recuperado. Una alternativa para simular correctamente este escenario, es utilizar cada término de la funciones  $f_i$  para generar las probabilidades de transición de estado. Esta alternativa implica un costo computacional mayor debido a la cantidad de términos que componen cada función  $f_i$ . Sin embargo, esta forma de cambios de estados simulan mejor lo que sucede en el mundo real, donde cada evento nacer, morir, infectarse o recuperarse son eventos independientes. Por último es importante mencionar, que diversos diseños de la arquitectura del *SMA* son validos y posibles de implementar.

La arquitectura equivalente a un sistema de *EDO* dado que se propone, se basa en un modelo Multi-Agente o *SMA* que se construye a partir de dicho sistema de *EDO*. En primer lugar, se asume como base el siguiente modelo *SIRS*, utilizado en el capítulo anterior:

$$\begin{aligned}\dot{S}(t) &= \gamma R(t) + \mu N(t) - \mu S(t) - \beta S(t) \frac{I(t)}{N(t)}, S(0) = S_0 > 0 \\ \dot{I}(t) &= \beta S(t) \frac{I(t)}{N(t)} - \mu I(t) - \rho I(t), I(0) = I_0 > 0 \\ \dot{R}(t) &= \rho I(t) - \mu R(t) - \gamma R(t), R(0) = R_0 > 0,\end{aligned}\tag{5.2}$$

A partir de este sistema de ecuaciones diferenciales se construye un *SMA*. A continuación definimos los elementos que se necesitan para construirlo:

### 5.2.1 Agente

Un agente es una entidad en un entorno virtual capaz de actuar, reproducirse, y en general tener un comportamiento autónomo. Un agente, en función de la percepciones y de sus recursos, interacciona con el entorno y con otros agentes [61]. En la Figura 5.1 puede verse el esquema que sigue un agente en cada iteración. Primero percibe, luego decide y por ultimo actúa.

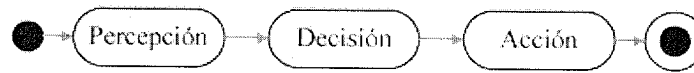


FIGURA 5.1: Proceso que sigue cada agente en cada iteración

En la arquitectura basada en agentes que se propone, las variables de estado de la *EDO* de la ecuación no corresponden directamente con agentes, sino con las subpoblaciones de susceptibles  $S(t)$ , infectados  $I(t)$  y recuperados  $R(t)$ . De esta forma tenemos:

- El número de especies de diferentes Agentes del *SMA* corresponde al número de variables de correspondiente a la población.
- La suma del número de individuos de cada especie en el *SMA* corresponde a las variables de estado del sistema *EDO*  $S$ ,  $I$  y  $R$ .

### Percepción

Cada agente recibe en cada iteración una serie de datos que son recogidos del entorno, aunque de forma indirecta, pueden ser datos que provienen de otros agentes. Para transformar el sistema de *EDO* en un *SMA*, cada agente percibe del entorno el número de agentes que existen de cada clase [62].

### Decisión

En función de las percepciones cada agente toma unas decisiones. Estas decisiones se toman en función de las probabilidades de sus máquinas de estados internas, que pueden tener o no memoria [64]. La toma de decisiones forma la parte central de la evolución del *SMA*. Cada agente, tiene distintas máquinas de estados con probabilidades de transición entre los estados. Las máquinas de estados son dependientes del número de agentes de cada especie y del tiempo. Estas máquinas se basan en los distintos términos de las  $f_i$  del sistema de *EDO* original.

## 5.2.2 Modelos Basados en agentes

La simulación basada en agentes se centra en los componentes básicos del sistema, en nuestro caso son los agentes que pertenecen a las clases o subpoblaciones  $S(t)$ ,  $I(t)$  y  $R(t)$ . El proceso de abstracción se hace sobre cada componente básico individualmente, no sobre el sistema en su conjunto. Para construir un modelo basado en agentes debemos identificar los componentes básicos del sistema real, y abstraer sus propiedades individuales (los parámetros de la ecuación diferencial) y las interacciones que se producen entre ellos (las probabilidades).

## 5.3 Modelo SIRS con Multi-Agentes

En este trabajo se plantea un enfoque para el modelado determinístico *SIRS* basado en un Sistema Multi-Agentes (*SMA*). Como se ha mencionado anteriormente, el modelo epidemiológico determinista *SIRS*, solo permite reproducir resultados con perfiles regulares o suaves, los cuales difieren de los datos reales de población infectada con el *VRS*. Adicionalmente, el modelo determinista describe las epidemias con valores medios sin tener en cuenta la variación de la población o que la población tiene valores discretos [65]. Por otra parte, el modelo *SIRS* puede ser desarrollado utilizando un *SMA*, el cual permite generar resultados con perfiles irregulares debido a la estructura subyacente discreta y estocástica del *SMA*. De esta forma, podemos encontrar que los resultados de la simulación con modelos de red pueden ser muy diferente que el modelo determinista subyacente [65]. Es importante señalar, que en el mundo real el seguimiento detallado de las actividades individuales (agentes) es complejo y requiere una gran cantidad de recursos humanos. Por ejemplo, el seguimiento de las actividades individuales y la construcción de una red de contacto se ha intentado para la ciudad de Portland [66].

En este trabajo asumimos que todos los agentes están relacionados y pueden contagiarse del *VRS*. Adicionalmente, las simulaciones del *SMA* se realizan utilizando el método de Monte Carlo. Como condición inicial se asume un estado inicial con una pequeña fracción de individuos infectados y ningún recuperado, de for-

ma tal de simular el inicio de un brote del *VRS*. Por ultimo, debemos mencionar que las simulaciones se realizan utilizando un total de 7.609.424 individuos (población de Bogotá), lo cual hace que el SMA sea computacionalmente costoso para simular. En la siguiente sección presentaremos los resultados más importantes de las simulaciones con el SMA de la dinámica del *VRS* en la población de Bogota.

### 5.3.1 Algoritmo de evolución del SMA

- Inicializar los valores numéricos a los distintos parámetros del SMA, entre estos parámetros tenemos la cantidad de agentes  $N$  y el estado en que se encuentra inicialmente cada agente, fijándolos en cada subpoblación;  $S(0)$ ,  $I(0)$  y  $R(0)$ . Es importante mencionar, que la asignación de los estados iniciales no se hace aleatoriamente como ocurre en otros trabajos.
- Posterior, a la preparación inicial se comienza la simulación del SMA utilizando un tamaño de paso en el tiempo  $\Delta t$ .
- La dinámica de interacciones entre los agentes es necesaria describirla, para entender el proceso de evolución del SMA. En cada paso se selecciona aleatoriamente un agente del sistema, como el muestra a continuación:
  - En el caso de estar en el estado susceptible, este agente podrá morir o infectarse con probabilidades  $\Delta t\mu$  y  $\Delta tb_0(I(t)/N)(1 + b_1\cos((2\pi t + \phi)))$  respectivamente. Es importante mencionar que los eventos morir o infectarse son independientes.
  - Si el agente seleccionado esta en un estado infectado entonces podrá morir o recuperarse con probabilidades  $\Delta t\mu$  y  $\Delta t\gamma$  respectivamente.
  - Por último si el agente seleccionado esta en el estado recuperado puede morir o perder su inmunidad con probabilidades  $\Delta t\mu$  y  $\Delta t\rho$  respectivamente.
- Este proceso de selección sin avanzar en el tiempo se realiza en  $N$  oportunidades. Luego se avanza en el tiempo con un paso  $\Delta t$  y se repite el proceso de selección de agentes.

Al cumplir el tiempo de simulación, el SMA ha presenta una tripleta de valores de las subpoblaciones para cada paso del tiempo  $\Delta t$ . Como se puede observar el SMA se ha intentado mantener lo más simple posible. De esta forma tenemos solo cuatro funciones morir, infectarse, recuperarse y perdida de inmunidad. Por otra parte, estas funciones dependen entre otros parámetros del tamaño de paso  $\Delta t$ , el cual al incrementarse también incrementa la probabilidad de aplicar estas funciones.

## 5.4 Resultados del modelo SIRS con Multi-Agentes

Esta sección está dedicada a presentar los resultados de las simulaciones del SMA del modelado de la dinámica del VRS a nivel poblacional. En esta sección se desea encontrar explicaciones de una forma adecuada a la no uniformidad de la serie temporal de datos reales en relación con el VRS. Las simulaciones se realizan utilizando distintas probabilidades extraídas de los términos de las funciones  $f_i$  del sistema de EDO (5.2).

En el SMA se utilizaron diferentes valores para los parámetros del modelo, tales como la tasa de contagio, tasa de natalidad(mortalidad), tasa de recuperación, tasa de perdida de inmunidad y condiciones iniciales. La variación de todos estos parámetros se realiza con el fin de encontrar los mejores ajustes del SMA a los datos reales.

Debido al gran costo computacional de las simulaciones del SMA utilizando el método de Monte Carlo, estas se realizaron con valores para los parámetros cercanos a los valores que minimizan el error cuadrático, para el modelo SIRS determinista del capítulo anterior. Las simulaciones fueron realizadas en equipos con procesadores Intel(R) Core(TM)i5-2400 CPU con 4 núcleos, velocidad del reloj de 3.10GHz, sistema operativo Linux (Ubuntu) y con el programa MATLAB R2010b [67]. Es importante mencionar que con tan solo 10 realizaciones de Monte Carlo, un tamaño de paso  $\Delta t = 0.002$  y una simulación del SMA de 4 años, el tiempo de computo sobrepaso las 200 horas.

En las Figuras 5.2 y 5.3, se puede observar la dinámica de la población de infectados  $I(t)$ , utilizando los modelos de Multi-Agentes y el de ecuaciones diferencia-

les. Se puede observar, que el modelo basado en agentes produce resultados con perfiles irregulares, similares a los datos reales y donde los valores que toma la subpoblación de infectados  $I(t)$  son discretos. Por otra parte, el modelo basado en ecuaciones diferenciales producen resultados numéricos con perfiles suaves donde  $I(t)$  toma valores continuos positivos. Es importante hacer notar que los resultados de ambos modelos no son exactamente iguales, pero dada la estructura del modelo SMA los resultados son similares. En la Figura 5.4, se compara la media de los resultados del modelo Multi-Agentes con el modelo basado en un sistema de EDO, utilizando 10 realizaciones de Monte Carlo y un tamaño de paso en el tiempo  $\Delta t = 0.002$ . Estos resultados son muy similares debido a la utilización de un paso pequeño y al uso de la media de varias realizaciones. Es de esperarse que al disminuir el tamaño de paso en el tiempo y aumentar la cantidad de realizaciones, la media del SMA se acercara al modelo del sistema de EDO.

Por último, tenemos en la Figura 5.5 los datos reales del VRS y los resultados de las simulaciones del SMA y el sistema de EDO para la dinámica del VRS. Se puede observar que ambos modelos generan comportamientos análogos a los datos reales. El ajuste de ambos modelos puede ser mejorado, pero no es el objetivo de este trabajo, donde lo que se desea es mostrar la factibilidad del uso de estos modelos y mostrar la ventaja que tienen los modelos Multi-Agentes.

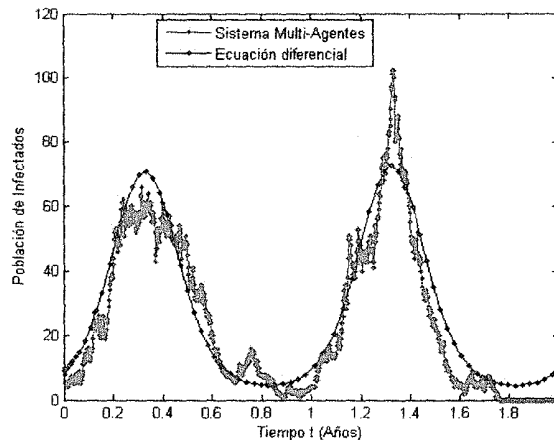


FIGURA 5.2: Comparación del SMA con la ecuación diferencial en dos años.

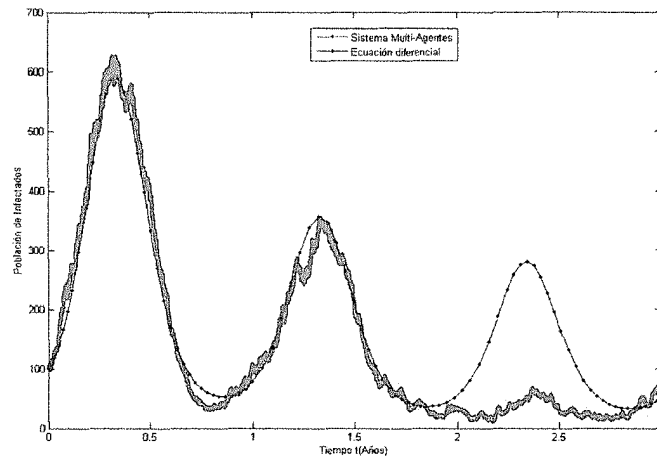


FIGURA 5.3: Comparación del SMA con la ecuación diferencial en tres años.

www.bdigital.ula.ve

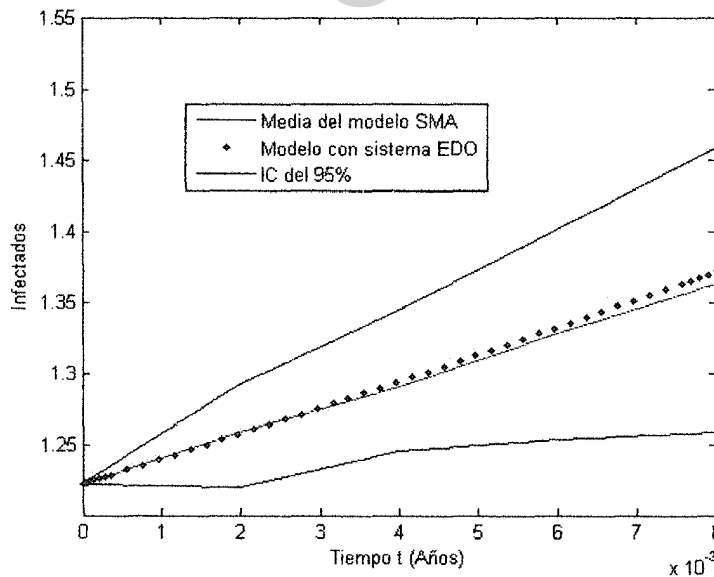


FIGURA 5.4: Promedio de varias simulaciones con SMA y el método SIRS clásico.

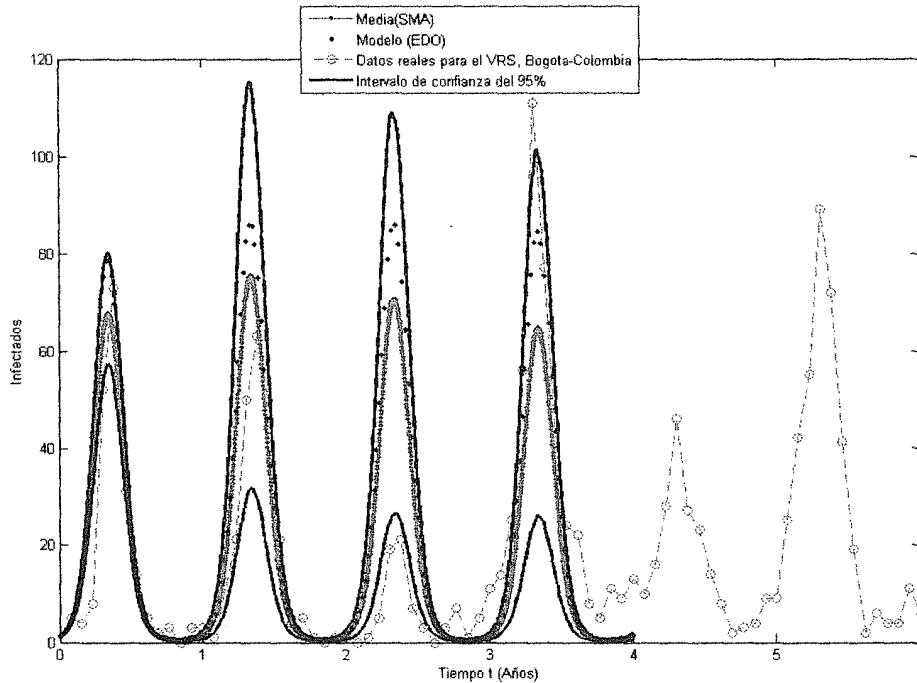


FIGURA 5.5: Comparación del Promedio de varias simulaciones con SMA con el método SIRS clásico y los datos reales durante 4 años (2005-2008)

## 5.5 Conclusión

En este capítulo se transformó el modelo epidemiológico *SIRS* basado en un sistema de EDO a un modelo basado en un SMA equivalente. Estos modelos se utilizaron para modelar la dinámica del *VRS* y comparar las características de cada enfoque. En el modelo Multi-Agentes, obtuvimos que para cada estado inicial y unos parámetros idénticos con relación al modelo *SIRS* se generan distintos resultados. Este comportamiento aleatorio, permite comprender mejor cómo es posible que en situaciones reales equivalentes, las dinámicas de una enfermedad en una población o en un paciente difieren fuertemente una de otra. En el modelo *SIRS*, la solución que se obtiene ante un mismo estado inicial y un mismo conjunto de valores de parámetros es única. Esto hace que dado un escenario inicial

concreto, solamente pueda desarrollarse una sola dinámica.

En el modelo basado en un *SMA* la población de infectados es modelada de un modo más realista, ya que cada variable del sistema *SIRS* toma solamente valores enteros positivos. De esta forma, el *SMA* modela las variables de un modo discreto. Adicionalmente, en el *SMA* es posible estimar la probabilidad de cada posible solución y dar la probabilidad de un evento concreto, ya que se dispone de las probabilidades de transición en cada instante. Esta propiedad es especialmente importante cuando se quiere conocer la probabilidad de que un agente infectado se recupere, infecte o muera. Por otra parte, como hemos podido constatar el modelado con *SMA* requiere de una mayor cantidad de detalles, incrementado de esta forma su complejidad. En cuanto a la implementación de las simulaciones los costos computacionales del modelo Multi-Agentes es mayor al tradicional sistema de EDO.

Finalmente, debemos mencionar que se logro modelar la dinámica del *VRS* mediante un sistema Multi-Agentes, para poder explicar y reflejar la irregularidad en los datos reales relacionados al *VRS*. El ajuste del modelo puede ser mejorado de diversas formas, entre ellas podemos citar; modificar los valores de los parámetros del modelo, utilización de otros tipos de agentes y arquitecturas. pero no es el objetivo de este trabajo, donde lo que se desea es mostrar la factibilidad del uso de estos modelos y mostrar una gran ventaja que tienen los modelos Multi-Agentes. Futuros trabajos podrían incluir información mas detallada sobre la estructura de contactos entre individuos en la ciudad de Bogotá en conjunto con datos meteorológicos, lo cual incrementaría la exactitud del modelo y su complejidad.

# Conclusiones

---

En este trabajo se propuso la construcción de varios clasificadores de *NB* para predecir la semana de brote del *VRS* de la ciudad de Bogotá, utilizando la base de datos climatológicos y el número de casos de infectados de niños menores de 5 años de los años 2005 hasta el 2010, dividiéndola en tres periodos 2 años (2009-2010), 4 años (2005-2008) y 6 años (2005-2010). Los resultados obtenidos, se basaron de la teoría ROC para los periodos de 6 y 2 años fueron buenos, ya que la mayoría de los resultados coincide con la semana de inicio del brotes con los datos reales (casos de infectados). Sin embargo, para los modelos que utilizan el periodo de 2 años es difícil determinar las variables que tienen mayor relación con el inicio del brote del *VRS*, por esta razón se considero que el periodo de 6 años es el que tuvo mejores resultados. Como podemos ver en la Tabla 3.19, el periodo de los 4 años no genera una cantidad de resultados relevante, por está razón no se tomo en cuenta estos resultados para el estudio de mejor periodo. Por otra parte, tenemos que la humedad mínima con 62% de promedio es la variable más significativa en los modelos de mejor rendimiento para la predicción el brote. También se observó, los distintos modelos proporcionados con los datos diarios con las variables velocidad del viento (59%) y bajas temperaturas (57%) se observó mayor acierto en las distintas clasificaciones; es decir, los modelos que tienen un mejor rendimiento; este hecho se repite en distintos estudios realizados en diferentes regiones del mundo [43, 30, 42]. Por último podemos observar los modelos con la semana 0 con 462 aciertos son los que produjeron mejores resultados en las clasificación, donde todos los modelos de las semanas tenían la misma cantidad de datos para su construcción y análisis.

Para futuros estudios se sugiere incluir bases de datos de más años donde se podría conseguir mejoras de los modelos de predicción, y también incluir otras variables en relación a la contaminación del aire (concentración de CO y SO<sub>2</sub>). De esta forma se recomienda a las autoridades sanitarias recolectar los datos relacionados al VRS.

Se demuestra que los modelos de clasificadores Naive Bayes (NB) basados en los datos meteorológicos disponibles en tiempo real tienen el potencial para ser utilizados como eficaces modelos predictivos. Estos modelos pueden ser capaces de predecir la semana que ocurre el brote del VRS. Su utilidad clínica se pondrá a prueba durante los próximos cinco años en la ciudad de Bogotá.

También en este trabajo se propuso el uso del modelo matemático epidemiológico SIRS con estacionalidad forzada por un armónico simple para dar explicación a la dinámica del Virus Respiratorio Sincitial en la población de niños menores de cinco años en Bogotá D. C.-Colombia. El modelo fue ajustado utilizando específicamente los datos reportados por el LSP de Vigilancia Epidemiológica SIVIGILA correspondientes al periodo 2005-2010. El modelo matemático SIRS con estacionalidad forzada propuesto aquí explicó de una forma razonable el comportamiento cíclico de los casos de niños infectados con el Virus Respiratorio Sincitial en la ciudad de Bogotá D. C. Colombia.

Es importante mencionar que la actual función periódica utilizada para el forzamiento estacional necesita ser estudiada con mayor profundidad. Por ejemplo su relación con variables del clima podría traer información relevante para entender la propagación del Virus Respiratorio Sincitial en la población infantil y adulta mayor [8]. Adicionalmente, podemos mencionar que el modelo estacional SIRS resultó conveniente para los datos reportados de niños. Sin embargo, si los reportes incluyen adultos mayores podría ser más conveniente usar un modelo SIRS estructurado por edades [57]. En este modelo alternativo se debe resaltar que al aumentar los parámetros del modelo la certeza de las predicciones mejora a un costo de complejidad en el modelado y estimado de valores de los parámetros.

Por otra parte, en este trabajo también se transformó el modelo epidemiológico SIRS basado en un sistema de EDO a un modelo basado en un SMA equivalen-

te. Este modelo fue utilizado para modelar la dinámica del *VRS* y comparar las características de cada enfoque. En el modelo Multi-Agentes, obtuvimos que para cada estado inicial y unos parámetros idénticos con relación al modelo *SIRS* se generaron distintos resultados. Este comportamiento aleatorio, permite comprender mejor cómo es posible que en situaciones reales equivalentes, las dinámicas de una enfermedad en una población o en un paciente difieren fuertemente una de otra. En el modelo *SIRS* (EDO), la solución que se obtuvo ante un mismo estado inicial y un mismo conjunto de valores de parámetros es única. Esto hace que dado un escenario inicial concreto, solamente pueda desarrollarse una sola dinámica.

En el modelo basado en un *SMA* la población de infectados fue modelada de un modo más realista, ya que cada variable del sistema *SIRS* toma solamente valores enteros positivos. De esta forma, el *SMA* modela las variables de un modo discreto. Por otra parte, como hemos podido constatar el modelado con *SMA* requiere de una mayor cantidad de detalles, incrementado de esta forma su complejidad. En cuanto a la implementación de las simulaciones los costos computacionales del modelo Multi-Agentes es mayor al sistema de EDO.

Por último, debemos mencionar que se logró modelar la dinámica del *VRS* mediante un sistema Multi-Agentes, para poder explicar y reflejar la irregularidad en los datos reales relacionados al *VRS*. El ajuste del modelo puede ser mejorado de diversas formas, entre ellas podemos citar; modificar los valores de los parámetros del modelo, utilización de otros tipos de agentes y arquitecturas. Como comentario final podemos decir que se logró la factibilidad del uso de los modelos Multi-Agentes y mostrar la ventaja que tienen este tipo de modelos. Futuros trabajos podrían incluir información más detallada sobre la estructura de contactos entre individuos en la ciudad de Bogotá en conjunto con datos meteorológicos, lo cual incrementaría la exactitud del modelo y su complejidad.

## REFERENCIAS

---

- [1] Gil P. *Medicina Preventiva y Salud Pública*. 10a edition, 2003.
- [2] Fauci A. Morens D., Taubenberger J. The persistent legacy of the 1918 influenza virus. *N Engl J Med*, 361(3):225–229, 2009.
- [3] Fred G. Meléndez A. Bernardo F. Martínez A., Manrique A. The 1918 influenza epidemic in Bogota. *Dynamis*, 27:287–307, 2007.
- [4] Jaramillo A. Moyano L. Osorio E. Nuñez L., Aranda D. Chronology of a pandemic: the new influenza a (H1N1) in Bogota, 2009-2010. *Rev Salud Publica (Bogota)*, 13(3):480–91, 2011.
- [5] Schmunis G. Yunes J. Benguigui Y., López F. Infecciones respiratorias en niños. *Organización Panamericana de la Salud, Washington, D.C*, 1997.
- [6] Savulescu C. Larrauri A. López T., Jiménez S. Influenza-related mortality in Spain, 1999-2005. *Gac Sanit*, 26(4):325–9.
- [7] Kimpen J. Van Woensel J., Van Aalderen W. Viral lower respiratory tract infection in infants and young children. *BMJ*, 327:36–40, 2003.
- [8] Gesteland P. Maloney C. Staes C. Facelli J. Walton N., Poynton M. Predicting the start week of respiratory syncytial virus outbreaks using real time weather variables. *Medical Informatics and Decision Making, Utha*, 2010.
- [9] Morales V. Dinora A. Clasificadores bayesianos en la selección embrionaria en tratamientos de reproducción asistida.

- [10] Gursky E. Nyce J. Morin M. Stomgren M. Holm E. Ekberg J. Timpka T., Eriksson H. Population-based simulations of influenza pandemics: validity and significance for public health policy. *Bulletin of the World Health Organization*, 2009.
- [11] Hospital Rafael Uribe Uribe E.S.E. Guia para la atención bronquiolitis. [online] <http://www.eserafaeluribe.gov.co/Educacion/downloads/>, 2011.
- [12] Children's Hospitals and Clinics of Minnesota. Patient family education. [online] <http://www.childrensmn.org/manuals/PFS/condill/018671.pdf>.
- [13] Roqueta M. Cabot A. Vall O. García O., Casado M. Protocolos prácticos de pediatría. *Ed Wander*, 1989.
- [14] Graham N. The epidemiology of acute respiratory infections in children and adults: a global perspective. *Epidemiol Rev* 1990, 1990.
- [15] Liptak G. McConnochie K., Roghmann K. Hospitalization for lower respiratory tract illness in infants: variation in rates among counties in new york state and areas within monroe county. *J Pediatrics*, pages 126:220–229, 1996.
- [16] Mufson M Belshe R, Van Voris L. Impact of viral respiratory diseases on infants and young children in a rural and urban area of southern West Virginia. *Am J Epidemiol*, 117(4):467–474, 1983.
- [17] Robertson C. Phelan P., Olinsky A. The epidemiology of acute respiratory infections. *Oxford: Blackwell Scientific Publication*, pages 27–51, 1994.
- [18] Clyde W. Denny F. Acute lower respiratory tract infections in non hospitalized children. *Pediatrics*, pages 108:635–646, 1986.
- [19] Meissner H. Economic impact of viral respiratory disease in children. *J Pediatrics*, pages S17–S21, 1994.
- [20] Mullholand K. Global burden of acute respiratory infections in children : implications for interventions. *Pediatric Pulmonology*, page 36: 469–474, 2003.

- [21] Szilagyi P. Walker F. Griffin M. Weinberg G. Coulen C. Poehling K. Shone L. Balter S. Hall C. Erdman D. Wooten K. Schwartz B. Iwane M., Edwards K. Population-based surveillance for hospitalizations associated with respiratory syncytial virus, influenza virus, and parainfluenza viruses among young children. *Pediatric*, 113(6):1758–1764, 2004.
- [22] Bresee J. Anderson L. Mullins J., Lamonte A. Substantial variability in community respiratory syncytial virus season timing. *Pediatric Infectious Disease journal*, page 22: 857–862., 2003.
- [23] Terletskaia E. Defining the timing of respiratory syncytial virus (rsv) outbreaks: an epidemiological study. *BMC Infectious Diseases*, page 5:20., 2005.
- [24] McLaren L. Florman A. The effect of altitude and winter on the occurrence of outbreaks of respiratory syncytial virus infections. *Journal of Infectious Diseases*, page 1401–1402., 1988.
- [25] Trabelsi A. Freymuth F. Fodha I., Vabret A. Epidemiological and antigenic analysis of respiratory syncytial virus in hospitalised tunisian children, from 2000 to 2002. *J Med Virol*, 72(4):683–7, 2004.
- [26] Tan T. Chua K. Hooi P. Chan P., Chew F. Seasonal variation in respiratory syncytial virus chest infection in the tropics. *Pediatric Pulmonol*, 34(1):47–51, 2002.
- [27] Cross K. Fleming D., Pannell R. Mortality in children from influenza and respiratory syncytial virus. *J Epidemiol Community Health*, 59:586–90, 2005.
- [28] Edmunds W. Joseph C. Gerard E. George R. Cooper D., Smith G. The contribution of respiratory pathogens to the seasonality of nhs direct calls. *J Infect*, 55:240–8, 2007.
- [29] Lee B. Shek L. Epidemiology and seasonality of respiratory tract virus infections in the tropics. *Paediatric Respiratory Reviews*, 4:105–11, 2003.

- 
- [30] Sung R. Seasonal patterns of respiratory syncytial virus infection in Hong Kong: a preliminary report. *Journal of Infectious Diseases*, page 156: 527–528., 1987.
- [31] Larrañaga P. Learning bayesian networks from data. some applications in biomedicine. *IDAMAP*, 2002.
- [32] Steck H. Bayesian belief networks for data mining. *Proceedings of the 2. Workshop on Data Mining und Data Warehousing*, 1996.
- [33] Russell G. Cheng J. Learning bayesian belief network classifiers: Algorithms and system. *Proceedings of the 14th Biennial Conference of the Canadian Society on Computational Studies of Intelligence: Advances in Artificial Intelligence*, 1999.
- [34] Finn J. *Bayesian Networks and Decision Graphs, Statics for Engineering and Information Science*. 2002.
- [35] Minsky M. Steps toward artificial intelligence. In *Computers and Thought*, pages 406–450. McGraw-Hill, 1961.
- [36] Pazzani M. Domingos P. On the optimality of the simple bayesian classifier under zero-one loss, 1997.
- [37] Yu-Shan S. Tjen-sien L., Wei-yin L. An empirical comparison of decision trees and other classification methods. Technical report, 1998.
- [38] Noguera T. Metodología ROC en la evaluación de medidas antropométricas como marcadores de la hipertensión arterial. aplicación a población gallega adulta. Master's thesis, Facultad de Matemáticas Máster en Técnicas Estadísticas, Universidad Santiago de Compostela, 2010.
- [39] Mandeville P. Noyola E. An introduction to roc analysis. *27(8)*, pp. 861–874, 2006.
- [40] Pepe Sullivan M. The statistical evaluation of medical tests for classification and prediction. *New York*, 2003.
-

- [41] Mandeville P. Noyola D. Effect of climatological factors on respiratory syncytial virus epidemics. *Cambridge University Press*, 2008.
- [42] Sarwo H. Linehan M. Djelantik I. Mercer D. Moniaga V. Moulton L. Widjaya A. Muljati P. Gessner B. Steinhoff M. Omer S., Sutanto A. Climatic, temporal, and geographic characteristics of respiratory syncytial virus disease in a tropical island population. *Cambridge University Press*, 136:1319–1327, 2008.
- [43] Auais A. Demmler G. Krishan S. Van Caesele P. Singleton R. Broor S. Parveen S. Avendano L. Parra J. Chavez S. Murguía T. Simoes E. Shaha S. Welliver R. Yusuf S., Piedimonte G. The relationship of meteorological conditions to the epidemic activity of respiratory syncytial virus. *journal of the Royal Statistical Society. Series A (General)*, 135(7):1077–1090, 2007.
- [44] Peña D. *Análisis de datos Multivariante*. 2 edition, 2001.
- [45] Meyer D. Naive Bayes classifier. *R Development Core Team*, 2011.
- [46] McKendrick A. Kermack W. Contributions to the mathematical theory of epidemics-i. *Bulletin of Mathematical Biology*, 53(2-1):33 – 55, 1991.
- [47] Brito H. Larios L. Acao L. Betancourt J., Rivero J. Simulación de un brote de enfermedad respiratoria aguda en el municipio de Camagüey. *Revista Archivo Medicos de Camagüey*, 13:0 – 0, 06 2009.
- [48] McKendrick Weisstein E., Kermack. in mathworld—a wolfram web resource. <http://mathworld.wolfram.com/Kermack-McKendrickModel.html>, 2009.
- [49] Chua H. Chan P., Chew T. Patogenia de las infecciones respiratorias por virus. *Revista del Instituto Nacional de Enfermedades Respiratorias*, 15(4):239–254, 2002.
- [50] De la Hoz F. Rico A. Alvis L. Velandia M. Porras A., Alvis N. Costo efectividad de la vacunación contra influenza en menores de 2 años y mayores de 65 años en Colombia. *Rev Salud Pública, Colombia*, 2009.

- [51] Mardones R. Velozo L. Parra G. Arce J., Moncada R. Secuelas post-infección por adenovirus en niños: Evaluación con tomografía computada. *Revista Chilena de Radiología*, (4):154–163, 2002.
- [52] Moraño J. Arenas A., González G. Stochastic modeling of the transmission of respiratory syncytial virus (rsv) in the region of Valencia, Spain. *Biosystems .Valencia, España*, (3):206–212, 2009.
- [53] Lupe S. Aranda D. González G., Arenas A. Modeling the epidemic waves of AH1N1/09 influenza around the world. *Spatial and Spatio-temporal Epidemiology*, (4):219–226, 2011.
- [54] Mead R. Nelder J. A simplex method for function minimization. *Computer journal*, 1965.
- [55] Arenas T. Abraham J. Mathematical modelling of virus RSV: qualitative properties, numerical solutions and validation for the case of the region of Valencia. *Universidad politécnica de Valencia*, 2010.
- [56] Milligan P. Weber A., Weber M. Modeling epidemics caused by respiratory syncytial virus (RSV). *Mathematical Biosciences*, 172(2):95 – 113, 2001.
- [57] Arenas A. González G., Villanueva R. An age structured model for obesity prevalence dynamics in populations. *Rev.MVZ Córdoba*, 15(2):2051–2059, 2010.
- [58] Bartlett M. Measles periodicity and community size. *Journal of the Royal Statistical Society Series A-General*, 120(1):48–70, 1957.
- [59] Modelling the influence of human behaviour on the spread of infectious diseases: a review. *The Royal of Society*, (50).
- [60] Hethcote H. The mathematics of infectious diseases. *SIAM Review*, 42:599–653, 2000.
- [61] Ferber J. Multi-agent systems. *Addison-Wesley*, 1999.

- 
- [62] Gordillo J. Calvo I. Guerrero Z., González A. Simulación basada en SMA de sistemas originalmente representados con EDO. *Revista iberoamericana de automática e informática industrial*, 8(4):323–33, 2011.
- [63] Molina J. Modelado y simulación mediante sistemas multiagente de la evacuación de personas de edificaciones de varios pisos en presencia de obstáculos, 2010.
- [64] Parunak H. Brueckner S. Applying metareasoning to swarming systems: Approaches and architectures. *Self-Adaptive and Self-Organizing Systems Workshops, IEEE International Conference on*, 1:317–322, 2010.
- [65] Øksendal Bernt K. Stochastic differential equations: An introduction with applications. *Springer, Berlin*, 2003.
- [66] Duch W. Visualization of hidden node activity in neural networks: I. visualization methods. In *In Proceedings of the international conference on artificial intelligence and soft computing*, pages 38–43. Physica Verlag, Springer, 2004.
- [67] MATLAB. version 7.10.0 (r2010a). 2010.

[www.bdigital.ula.ve](http://www.bdigital.ula.ve)

## Apéndice A

## Apendice

Sens	Esp	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	REP
0.5	0.78	22	13	14	7	24	6	14	13	27
0.5	0.75	18	21	17	20	36	17	19	20	38
0.5	0.67	2	2	2	2	0	2	2	1	2
0.5	0.72	17	24	25	34	44	30	29	25	49
0.5	0.81	10	4	10	2	8	2	2	5	10
0.5	0.69	10	6	6	8	8	8	10	5	10
0.5	0.64	2	0	2	2	0	2	2	1	2
0.3	0.81	4	7	5	4	5	10	7	6	13
0.3	0.75	6	5	3	4	0	5	4	2	7
0.3	0.78	5	8	10	9	3	14	7	8	15
0.3	0.72	2	2	2	0	0	2	0	1	2
0.3	0.69	2	2	0	2	0	2	2	1	2
0.3	0.83	3	2	0	2	0	1	2	2	3
0.3	0.86	0	0	2	2	0	2	2	1	2
0.3	0.89	0	2	3	0	0	5	4	3	5
0.3	0.92	0	0	0	1	0	2	2	2	2
0.17	0.92	0	5	1	2	0	3	0	3	6
0.17	0.89	5	5	2	0	0	0	3	4	7
0.17	0.97	3	0	0	3	0	0	3	7	7
0.17	0.94	0	0	0	2	0	4	2	2	4
0.17	0.86	4	8	4	6	0	3	7	6	13
0.3	0.67	4	2	4	4	0	4	0	2	4
0.17	0.83	2	6	8	2	0	2	2	4	8
0.17	0.81	4	4	6	5	0	1	0	4	8
0.17	0.75	0	0	1	0	0	0	0	0	1
0	1	3	0	0	3	0	0	3	0	6
0	0.97	0	0	0	1	0	1	0	0	1
0	0.89	0	0	1	1	0	0	0	0	1

TABLA A.1: Resultados utilizando la semana 0 para los modelos que utilizan los datos de 6 años (2005 hasta 2010)

Sens	Esp	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	REP
0.83	0.67	1	1	0	1	0	1	0	1	1
0.83	0.64	2	2	1	1	2	1	1	1	2
0.83	0.61	0	0	1	0	0	0	0	0	1
0.83	0.58	1	2	2	2	2	0	1	1	2
0.83	0.56	2	3	2	4	4	1	3	3	4
0.83	0.53	2	3	0	5	5	3	2	2	5
0.83	0.5	1	1	2	6	5	1	3	2	6
0.83	0.47	1	0	1	1	1	0	0	0	1
0.83	0.44	2	2	3	3	3	2	1	0	3
0.83	0.42	2	2	7	7	7	5	5	0	7
0.83	0.39	1	2	2	3	1	0	2	0	3
0.83	0.28	0	0	1	1	0	0	0	0	1
0.67	0.83	0	0	0	0	1	0	0	1	1
0.67	0.81	1	2	0	0	2	0	0	2	2
0.67	0.78	2	3	1	0	4	1	0	0	4
0.67	0.75	0	0	1	1	3	0	2	3	3
0.67	0.72	2	4	2	1	6	2	3	4	6
0.67	0.69	1	7	1	1	7	4	4	4	7
0.67	0.67	2	1	1	1	3	2	2	1	4
0.67	0.64	1	0	1	1	2	0	1	0	2
0.67	0.61	2	4	2	4	5	3	2	2	6
0.67	0.58	1	2	1	1	2	1	1	0	2
0.67	0.56	1	3	4	3	4	4	3	3	5
0.67	0.53	2	0	1	3	3	2	2	3	3
0.67	0.5	3	2	2	4	3	3	4	3	4
0.67	0.47	2	2	3	5	4	3	4	4	5
0.67	0.44	5	7	5	8	3	5	7	5	8
0.67	0.42	1	1	1	3	2	2	3	1	3
0.67	0.39	2	5	2	5	1	3	4	0	5
0.67	0.36	0	2	2	1	1	0	2	0	2
0.5	0.86	1	1	1	0	1	0	0	0	2
0.5	0.83	2	0	1	0	2	0	0	1	2
0.5	0.81	3	5	6	0	5	1	2	5	8
0.5	0.78	5	2	3	2	6	3	3	6	8
0.5	0.75	3	6	5	0	4	3	2	5	7
0.5	0.72	4	4	3	1	6	4	3	2	7
0.5	0.69	2	3	2	1	3	2	1	2	3
0.5	0.67	2	1	2	1	2	1	3	2	4
0.5	0.64	2	3	3	2	3	2	3	3	5
0.5	0.61	1	1	1	1	0	0	0	1	1

Sens	Esp	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	REP
0.5	0.58	2	2	2	2	2	3	2	2	4
0.5	0.53	3	2	2	4	1	4	3	4	4
0.3	0.92	1	2	2	0	0	0	0	1	2
0.3	0.89	1	1	1	0	0	0	0	1	1
0.3	0.86	2	2	2	0	1	1	1	3	3
0.3	0.83	3	1	0	1	1	2	1	3	3
0.3	0.81	1	0	0	0	0	2	0	2	2
0.3	0.78	3	2	1	1	0	2	0	2	3
0.3	0.75	0	1	1	0	0	2	0	1	2
0.3	0.72	1	1	1	0	0	1	0	0	1
0.3	0.69	1	1	1	0	1	1	1	1	1
0.17	0.97	1	1	0	0	0	0	0	0	1
0.17	0.95	4	3	2	2	0	0	0	7	7
0.17	0.86	1	0	0	0	0	1	0	0	1
1	0	0	0	0	1	0	0	0	0	1
0.67	0.3	1	0	1	1	0	0	0	0	1
0.67	0.31	1	0	0	1	0	0	1	0	1
0.67	0.28	0	0	0	1	0	0	1	0	1
0.67	0.22	1	0	0	1	0	0	0	0	1
0.5	0.5	1	0	1	1	1	1	1	1	1
0.5	0.47	2	1	2	3	0	3	3	0	3
0.5	0.44	2	1	1	3	1	2	3	1	3
0.5	0.42	1	0	2	2	0	0	2	2	2
0.5	0.39	1	2	2	2	0	1	1	0	2
0.5	0.36	1	0	3	3	0	1	2	0	3
0.3	0.67	2	0	4	0	2	2	3	1	4
0.3	0.64	1	0	2	0	0	1	1	0	2
0.3	0.61	1	0	0	1	0	1	0	0	1
0.3	0.58	1	0	1	0	0	1	1	0	1
0.3	0.56	4	4	2	8	0	8	4	5	9
0.3	0.53	2	1	3	4	0	4	1	4	4
0.3	0.47	2	1	2	2	0	2	0	0	2
0.17	0.83	1	2	0	0	0	2	0	2	2
0.17	0.78	0	1	0	0	0	1	1	1	1
0.17	0.75	0	0	0	0	0	1	1	1	1
0.17	0.72	2	1	0	0	0	2	2	2	2
0.17	0.69	0	1	1	0	0	2	2	1	2
0.17	0.67	3	3	2	0	0	4	4	2	4
0.17	0.64	1	1	3	0	0	3	3	1	3
0.17	0.61	1	1	1	0	0	1	1	0	1
0	1	2	0	1	0	0	0	0	0	2
0	0.97	2	0	0	0	0	0	2	4	4
0	0.72	0	0	0	0	0	1	1	0	1

Sens	Esp	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	REP
0.83	0.44	0	1	1	3	1	3	3	0	3
0.83	0.39	1	1	1	1	0	1	1	0	1
0.83	0.36	2	1	5	5	3	5	4	0	5
0.83	0.3	0	1	1	2	0	2	2	0	2
0.83	0.28	0	0	1	2	0	0	1	0	2
0.67	0.69	2	3	2	2	3	1	1	1	3
0.67	0.67	0	1	2	2	3	1	1	2	3
0.67	0.64	1	2	0	0	3	3	3	0	3
0.67	0.61	2	2	1	1	3	2	2	1	3
0.67	0.58	1	1	0	1	1	0	0	0	1
0.67	0.56	0	1	0	1	1	0	0	0	1
0.67	0.53	2	2	0	5	4	5	1	0	6
0.67	0.5	1	1	1	3	3	3	2	0	4
0.67	0.47	3	2	2	3	3	2	1	0	3
0.67	0.44	6	8	5	9	5	4	6	0	10
0.67	0.42	1	1	3	4	4	3	2	0	4
0.67	0.39	4	7	8	8	3	3	6	0	9
0.67	0.36	1	1	2	3	1	1	3	0	3
0.5	0.86	0	0	0	0	1	1	0	0	1
0.5	0.78	0	2	1	2	3	2	0	2	3
0.5	0.75	2	3	3	1	4	4	3	3	5
0.5	0.72	1	0	1	0	1	0	0	0	1
0.5	0.69	0	2	2	3	2	3	1	2	4
0.5	0.67	2	5	4	4	4	4	4	3	7
0.5	0.64	6	6	8	5	7	6	6	5	10
0.5	0.61	3	4	6	5	5	6	6	5	8
0.5	0.58	4	3	3	2	3	3	3	1	4
0.5	0.56	3	0	3	0	2	2	3	0	3
0.3	0.86	0	2	0	0	1	2	1	1	2
0.3	0.83	2	2	2	0	3	3	1	2	4
0.3	0.81	1	4	2	1	3	1	0	2	4
0.3	0.78	2	2	1	2	2	4	2	3	4
0.3	0.75	4	5	4	3	3	3	2	3	6
0.3	0.72	1	1	1	1	0	1	0	1	1
0.3	0.69	1	1	1	2	1	1	1	2	2
0.17	0.94	0	1	1	0	0	1	0	1	1
0.17	0.92	0	0	1	0	1	1	0	1	1
0.17	0.86	0	1	1	0	0	1	1	1	2
1	0	0	0	0	1	0	0	0	0	1
0.67	0.3	1	0	3	2	1	0	1	0	3
0.5	0.5	1	1	1	1	0	1	0	0	1

Sens	Esp	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	REP
0.5	0.47	0	1	1	0	1	1	1	0	1
0.5	0.44	2	1	2	2	2	1	1	0	2
0.5	0.42	2	0	1	2	1	0	2	0	2
0.3	0.67	3	1	2	1	0	3	2	2	3
0.3	0.64	1	3	2	2	0	2	3	2	4
0.3	0.61	0	1	1	1	0	1	1	1	1
0.3	0.58	2	2	2	1	0	1	2	1	2
0.3	0.56	1	1	2	0	0	0	2	0	2
0.17	0.83	1	0	1	0	1	1	1	1	1
0.17	0.81	2	2	1	1	3	2	4	3	4
0.17	0.75	2	0	2	2	1	1	2	2	3
0.17	0.72	2	1	0	2	0	1	1	1	2
0.17	0.58	1	0	1	0	0	0	1	0	1
0	0.97	0	9	4	5	8	8	8	23	23
0	0.94	0	1	2	4	2	3	0	5	5
0	0.92	5	4	2	0	3	1	6	8	8
0	0.89	19	9	6	2	5	3	5	15	19
0	0.86	14	6	5	6	7	8	5	12	16
0	0.83	7	3	5	5	4	2	3	7	8
0	0.81	1	2	1	0	1	0	2	1	2
0	0.75	1	0	0	1	1	0	1	1	1
0	0.64	1	0	1	1	0	0	1	1	1

TABLA A.3: Resultados utilizando la semana 2 para los modelos que utilizan los datos de 6 años (2005 hasta 2010)

Sens	Esp	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	REP
0.83	0.81	0	0	1	0	0	0	0	0	1
0.67	0.97	0	0	1	0	0	1	0	0	1
0.5	0.94	0	0	1	0	1	1	0	0	1
0.5	0.92	0	2	2	0	1	2	1	0	2
0.5	0.89	0	1	1	0	0	1	0	0	1
0.3	0.97	1	1	0	0	0	1	0	0	1
0.3	0.94	2	1	2	0	2	2	0	0	2
0.3	0.92	3	3	4	0	1	1	1	0	4
0.3	0.89	5	4	5	2	1	4	3	0	5
0.3	0.86	5	4	5	1	2	2	2	0	5
0.3	0.83	1	1	1	1	0	1	0	0	1
0.17	1	6	0	0	2	0	0	2	3	6
0.17	0.97	1	1	0	0	0	1	0	1	1
0.17	0.94	13	5	6	3	15	12	1	9	16
0.17	0.92	25	18	18	15	22	18	12	16	32
0.17	0.89	19	7	11	10	13	13	13	10	22
0.17	0.86	20	13	14	15	14	8	11	13	22
0.17	0.83	21	11	9	15	6	9	18	11	21
0.17	0.81	3	3	2	3	2	1	3	2	3
0.17	0.78	1	1	0	1	0	1	1	0	1
0.17	0.72	1	0	0	1	0	0	1	1	1
0.17	0.69	1	0	0	1	0	0	1	0	1
0	1	0	4	0	4	0	4	0	6	11
0	0.97	0	3	1	4	3	6	6	7	10
0	0.94	0	16	10	12	14	12	15	15	27
0	0.92	0	14	19	21	19	19	26	22	35
0	0.89	0	8	10	10	3	7	4	6	12
0	0.86	0	5	3	5	7	1	5	5	8
0	0.83	0	2	2	2	2	0	2	1	2

TABLA A.4: Resultados utilizando la semana 3 para los modelos que utilizan los datos de 6 años (2005 hasta 2010)

Sens	Esp	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	REP
1	0.3	2	2	1	2	0	2	2	2	2
1	0.29	13	7	5	14	2	7	14	7	14
1	0.25	9	7	14	20	15	11	18	15	20
1	0.21	9	15	14	24	9	11	14	13	25
1	0.17	11	7	13	17	8	7	6	8	18
1	0.13	2	5	5	6	3	2	0	3	6
1	0.08	0	2	4	5	2	2	0	3	5
1	0.04	0	2	2	7	2	2	0	7	7
0.75	0.63	1	0	1	0	0	1	1	1	1
0.75	0.42	1	1	0	1	1	1	1	1	1
0.75	0.38	2	2	1	2	2	2	1	0	2
0.75	0.33	5	5	1	7	7	6	5	2	7
0.75	0.29	3	1	1	5	5	3	4	0	5
0.5	0.83	0	0	2	0	0	1	2	1	2
0.5	0.79	5	5	5	2	4	2	4	4	8
0.5	0.75	15	9	13	0	10	7	8	7	17
0.5	0.71	18	15	14	3	24	13	22	18	30
0.5	0.67	6	7	5	0	12	10	15	9	16
0.25	0.92	3	1	2	1	1	7	2	3	8
0.25	0.875	2	0	3	0	0	4	2	1	4
0.25	0.83	3	6	2	1	4	5	2	2	8
0.25	0.79	6	7	8	0	4	11	4	6	13
1	0	0	1	1	5	1	1	0	0	5
0.75	0.25	1	1	0	3	3	2	0	1	3
0.75	0.21	1	2	1	2	2	0	0	0	2
0.75	0.17	1	0	0	1	1	0	0	0	1
0.25	0.75	1	4	1	0	5	4	0	3	5
0.25	0.71	0	2	2	0	1	2	0	2	2
0	1	2	0	0	0	0	0	0	2	3
0	0.92	3	4	3	0	0	0	1	2	6
0	0.88	2	3	0	0	0	1	0	2	3
0	0.83	1	3	3	0	0	0	0	2	3
0	0.79	0	2	1	0	0	1	0	1	3

TABLA A.5: Resultados utilizando la semana 0 para los modelos que utilizan los datos de 4 años (2005 hasta 2008)

Sens	Esp	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	REP
1	0.67	0	1	1	0	1	0	1	1	1
1	0.63	1	0	2	0	2	0	2	0	2
1	0.58	1	0	1	0	1	0	1	0	1
1	0.38	1	0	1	1	1	1	1	1	1
1	0.3	4	1	1	4	4	3	3	2	4
1	0.29	8	12	15	17	14	9	16	10	17
1	0.25	2	2	5	6	2	4	6	2	6
1	0.21	8	9	12	20	16	2	16	11	22
1	0.17	0	3	1	4	1	2	3	1	4
1	0.13	1	2	2	4	4	0	1	0	4
1	0.08	1	2	0	2	1	0	1	0	2
1	0.04	1	1	0	2	1	0	0	1	2
0.75	0.79	1	0	0	0	3	1	0	2	3
0.75	0.75	2	3	0	0	5	1	0	2	5
0.75	0.71	2	3	0	0	3	2	0	1	3
0.75	0.67	3	6	3	1	7	2	6	5	8
0.75	0.63	1	4	1	0	3	2	4	2	4
0.75	0.58	3	3	0	0	4	2	4	0	4
0.75	0.54	1	1	2	0	2	2	3	1	4
0.75	0.5	1	2	2	0	3	3	3	0	3
0.75	0.38	2	0	1	2	0	2	2	2	2
0.75	0.3	8	6	7	11	6	8	6	9	11
0.75	0.29	2	1	2	7	5	7	4	5	7
0.5	0.83	0	1	1	0	0	0	0	0	1
0.5	0.79	6	4	6	3	6	4	3	8	11
0.5	0.75	5	4	5	0	6	2	1	4	7
0.5	0.71	6	9	9	1	6	6	6	7	14
0.5	0.67	0	3	2	0	2	3	1	1	3
0.5	0.63	2	3	3	0	3	4	2	2	4
0.5	0.58	2	0	1	0	3	3	3	2	3
0.5	0.54	2	1	4	0	1	2	4	0	4
0.25	0.92	3	3	2	0	0	0	0	6	6
0.25	0.875	3	3	2	0	0	0	0	1	3
0.25	0.83	3	2	2	0	0	6	0	6	6
0.25	0.79	5	4	3	1	1	5	1	2	6

Sens	Esp	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	REP
1	0	1	2	0	5	1	0	0	1	5
0.75	0.25	2	5	2	7	4	5	2	4	7
0.75	0.21	6	5	1	10	3	7	4	3	10
0.75	0.17	2	3	3	4	0	1	0	2	4
0.75	0.13	2	1	2	3	0	0	0	1	3
0.75	0.08	2	1	0	2	0	0	0	1	2
0.5	0.5	1	0	0	0	0	2	2	0	2
0.5	0.3	0	1	2	2	0	2	1	2	3
0.5	0.29	1	2	2	2	0	2	0	1	2
0.5	0.25	2	1	0	2	0	2	0	1	2
0.5	0.21	2	0	2	2	0	1	0	1	2
0.5	0.17	1	0	1	1	0	0	0	0	1
0.25	0.75	2	1	1	0	1	1	1	2	2
0.25	0.71	2	3	2	0	0	5	3	2	5
0.25	0.67	3	3	3	1	1	4	4	4	5
0.25	0.63	2	0	3	0	1	3	3	3	3
0.25	0.58	2	1	2	0	0	3	2	0	3
0.25	0.54	0	0	1	0	0	1	1	0	1
0.25	0.38	1	0	1	1	0	1	0	1	1
0	1	2	0	1	0	0	0	0	0	2
0	0.96	1	0	0	0	0	0	0	1	1
0	0.92	0	0	0	0	0	0	1	1	1

TABLA A.6: Resultados utilizando la semana 1 para los modelos que utilizan los datos de 4 años (2005 hasta 2008)

Sens	Esp	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	REP
0.75	0.58	2	4	6	3	6	0	0	3	6
0.75	0.54	0	0	0	1	1	0	1	0	1
0.75	0.5	1	3	1	4	3	0	2	0	4
0.75	0.46	2	1	3	3	3	1	1	0	3
0.75	0.42	3	4	4	4	2	1	2	0	4
0.75	0.38	0	4	6	5	2	2	2	0	6
0.75	0.3	1	0	4	3	1	0	1	0	4
0.5	0.79	0	1	0	0	3	2	1	1	3
0.5	0.75	0	1	0	0	1	0	1	1	1
0.5	0.71	1	2	1	0	4	4	2	1	4
0.5	0.67	9	13	6	6	15	7	4	11	16
0.5	0.63	10	12	12	11	15	10	11	13	19
0.5	0.58	9	10	12	12	18	11	13	9	20
0.5	0.54	5	8	9	3	6	8	9	2	10
0.25	0.88	1	2	3	3	4	3	3	6	7
0.25	0.83	1	3	0	2	2	2	0	3	4
0.25	0.79	5	4	3	3	5	4	2	6	9
1	0	0	0	0	1	0	0	0	0	1
0.75	0.21	0	0	0	1	0	0	1	0	1
0.5	0.5	6	7	7	9	7	9	9	2	11
0.5	0.46	2	2	2	5	4	5	5	0	5
0.5	0.42	2	0	4	4	3	3	4	0	4
0.5	0.38	1	2	3	4	0	3	4	0	4
0.5	0.3	2	0	1	2	0	1	2	0	2
0.5	0.29	0	1	0	1	0	1	1	0	1
0.25	0.75	7	7	3	3	8	7	7	8	13
0.25	0.71	2	3	3	2	1	3	3	4	5
0.25	0.67	5	3	3	4	1	3	2	4	5
0.25	0.63	5	7	5	2	0	3	6	2	9
0.25	0.58	2	1	1	2	0	2	2	2	2
0.25	0.54	3	5	4	4	0	4	2	2	6
0.25	0.5	1	1	0	1	0	0	1	0	1
0.25	0.3	1	0	0	1	0	0	1	0	1

Sens	Esp	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	REP
0	1	0	0	0	0	0	1	0	0	1
0	0.96	3	4	3	4	4	8	5	13	16
0	0.92	9	6	3	2	5	5	4	11	13
0	0.88	10	3	5	5	2	7	6	11	14
0	0.83	6	3	2	0	2	2	2	5	6
0	0.79	5	1	2	2	0	2	2	2	5
0	0.75	3	0	2	2	0	2	1	2	3
0	0.71	1	0	2	2	0	2	0	2	2
0	0.58	1	0	2	2	0	0	2	2	2
0	0.54	1	0	1	0	0	0	1	0	1

TABLA A.7: Resultados utilizando la semana 2 para los modelos que utilizan los datos de 4 años (2005 hasta 2008)

[www.bdigital.ula.ve](http://www.bdigital.ula.ve)

Sens	Esp	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	REP
0.5	1	1	0	0	0	0	0	0	0	1
0.5	0.96	2	0	1	0	1	2	0	0	2
0.5	0.92	4	3	3	0	1	4	0	1	4
0.5	0.88	4	3	1	1	3	4	0	3	4
0.5	0.83	4	2	1	1	2	4	1	3	4
0.25	0.92	6	3	6	0	3	2	1	4	7
0.25	0.875	6	4	5	0	2	3	2	3	6
0.25	0.83	7	4	6	0	4	4	7	3	7
0.25	0.79	6	1	5	0	3	3	5	3	6
0.25	0.75	1	1	0	0	0	1	1	0	1
0.25	0.71	3	2	1	0	2	3	3	2	3
0	1	1	0	0	0	0	0	0	2	2
0	0.96	0	1	2	0	1	1	0	3	4
0	0.92	2	9	11	1	12	11	6	6	19
0	0.88	7	18	12	2	12	11	8	12	25
0	0.83	8	11	8	1	13	7	15	11	21
0	0.79	0	0	2	0	2	1	2	1	3
0	0.75	2	3	0	0	2	5	6	5	6
0	0.71	3	3	1	1	2	3	4	4	5
0	0.67	1	0	0	0	0	0	2	0	2
0	0.46	8	9	10	12	10	9	11	4	12
0	0.42	9	7	14	16	11	10	14	11	16
0	0.38	15	11	13	24	14	15	14	10	24
0	0.33	13	15	13	24	12	8	13	11	25
0	0.29	3	4	6	12	5	6	8	9	12
0	0.25	8	9	4	16	6	7	5	8	17
0	0.21	4	1	2	6	2	0	0	2	6
0	0.17	0	1	1	3	0	3	0	3	3
0	0.13	0	1	0	1	1	0	0	1	1
0	0.08	0	1	0	4	1	1	0	3	4
0	0.04	0	1	0	3	1	0	0	0	3

TABLA A.8: Resultados utilizando la semana 3 para los modelos que utilizan los datos de 4 años (2005 hasta 2008)

Sens	Esp	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	REP
1	0.83	15	16	9	26	1	14	9	21	27
1	0.75	36	38	39	67	54	32	39	42	70
1	0.67	7	5	12	9	10	6	13	1	13
1	0.66	1	1	1	0	0	1	1	0	1
1	0.58	2	0	2	0	0	1	2	0	2
1	0.5	1	0	0	0	0	0	0	0	1
0.5	0.92	15	12	7	17	2	17	12	14	32
0.5	0.83	38	41	33	9	50	39	31	41	78
0.5	0.75	10	15	21	0	11	12	17	9	22
0.5	0.67	2	0	2	0	0	1	2	0	4
0	1	1	0	0	0	0	2	0	0	2
0	0.92	0	0	1	0	0	2	1	0	2
0	0.83	0	0	1	0	0	1	1	0	1

TABLA A.9: Resultados utilizando la semana 0 para los modelos que utilizan los datos de 2 años (2009 hasta 2010)

www.bdigital.ula.ve

Sens	Esp	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	REP
0.5	0.92	15	24	25	14	25	3	16	15	35
0.5	0.83	7	15	11	2	11	8	7	8	17
0.5	0.75	6	4	8	0	6	12	4	5	15
0.5	0.67	3	0	5	0	1	5	2	2	5
0.5	0.58	1	0	3	0	1	3	0	0	3
0.5	0.5	0	0	0	0	0	1	0	0	1
0	1	58	43	39	74	34	23	46	60	102
0	0.92	24	32	21	34	33	48	28	27	51
0	0.83	11	9	15	3	15	20	20	9	20
0	0.75	3	1	1	1	2	5	5	2	6

TABLA A.10: Resultados utilizando la semana 1 para los modelos que utilizan los datos de 2 años (2009 hasta 2010)

Sens	Esp	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	REP
1	0.67	5	5	7	5	0	13	13	4	13
1	0.58	3	0	3	3	2	5	4	0	5
1	0.5	6	7	7	11	1	8	8	0	13
1	0.42	2	3	8	11	7	8	10	0	11
1	0.3	0	2	2	2	1	2	2	0	2
1	0.25	0	0	1	1	0	0	0	0	1
0.5	0.92	14	19	12	7	1	24	10	23	24
0.5	0.83	7	7	9	12	11	14	4	14	15
0.5	0.75	10	9	10	8	14	19	14	11	22
0.5	0.67	7	8	7	5	13	8	10	1	17
0.5	0.58	7	8	7	9	12	7	7	0	15
1	0	0	0	0	1	0	0	0	0	1
0.5	0.5	3	1	3	6	6	2	4	0	6
0.5	0.42	1	4	3	4	3	2	0	0	4
0	1	40	35	32	25	26	15	18	59	72
0	0.92	14	13	7	9	11	1	14	11	18
0	0.83	7	5	8	7	8	0	8	5	12
0	0.75	2	1	2	1	1	0	2	0	3
0	0.67	0	1	0	1	0	0	0	0	1

TABLA A.11: Resultados utilizando la semana 2 para los modelos que utilizan los datos de 2 años (2009 hasta 2010)

Sens	Esp	TempAVG	TempMAX	TempMIN	HumMAX	HumMIN	Vie	Pres	Precip	REP
1	0.58	5	9	7	10	7	0	11	0	11
1	0.5	9	8	8	18	6	12	18	0	18
1	0.42	0	0	0	0	0	0	1	0	1
1	0.08	0	0	0	1	0	0	0	0	1
0.5	0.83	3	3	9	10	6	8	9	9	10
0.5	0.75	3	2	4	2	2	7	4	1	7
0.5	0.67	5	5	6	11	4	11	1	0	13
0.5	0.58	14	10	15	6	17	11	25	0	26
0.5	0.5	0	4	3	1	5	6	6	0	7
0	1	8	9	6	2	7	1	1	1	14
0	0.92	72	68	62	56	62	64	48	113	129
0	0.83	4	3	4	5	8	5	3	3	10
0	0.75	3	5	3	3	1	3	1	1	5
0	0.67	2	2	1	3	3	0	0	0	3

TABLA A.12: Resultados utilizando la semana 3 para los modelos que utilizan los datos de 2 años (2009 hasta 2010)

[www.bdigital.ula.ve](http://www.bdigital.ula.ve)

[www.bdigital.ula.ve](http://www.bdigital.ula.ve)