

Proyecto 1: Evolución del COVID-19 en Pacientes del Estado de Puebla, México Usando Cadenas de Markov

Miguel Angel Soto Hernandez

*Centro de Investigación en Computación, Instituto Politécnico Nacional
Ciudad de México, México
msotoh2021@cic.ipn.mx*

Resumen. En este proyecto se realiza un modelo de cadenas de Markov para el que se utilizaron los datos públicos proporcionados por parte del gobierno de México para pacientes que han experimentado síntomas o tuvieron la enfermedad SARS-COV-19. Los datos específicos utilizados para este modelo corresponden al estado de Puebla, sin embargo, el conjunto de datos comprende los 32 estados de la República Mexicana. Las probabilidades se calculan para comprender el impacto que esta nueva enfermedad llega a causar en la sociedad y por tal motivo el modelo proporciona información que sirve para evaluar el riesgo de estancia en un hospital.

Palabras clave: Cadenas de Markov, Probabilidad, COVID-19, Inferencia Bayesiana

I. Introducción

El virus SARS-COV-2 descubierto a finales del año 2019 que genera la enfermedad COVID-19 fue declarado como pandemia mundial por la Organización Mundial de la Salud (OMS) en Marzo de 2020. Esto causó una fuerte crisis en los distintos ámbitos de la sociedad. Sin embargo, la comunidad científica de todo el mundo comenzó a trabajar en diferentes áreas con la finalidad de entender como es que la enfermedad se propagaba, cuales eran los medios más fuertes de propagación, así como modelos que predecían con que rapidez se propagaría, y como se combatiría con la finalidad de tener control sobre esta.

Hubo muchas aproximaciones para entender como la enfermedad COVID-19 y para esto se estudiaron y se propusieron muchos modelos científicos y matemáticos. Este trabajo pretende crear un modelo que sea capaz de comprender las diferentes etapas que una persona pueda experimentar desde que acude a un centro de salud hasta que se recupera o muere, así como identificar los síntomas durante el transcurso de la enfermedad.

II. Conjunto de datos

Para llevar a cabo la creación del modelo de cadenas de Markov se utilizó el conjunto de datos de acceso libre que ofrece la Dirección General de Epidemiología de México, en específico se utilizó el conjunto de datos liberado el día 15 de Junio de 2021, el cual contiene 7,252,170 casos de la enfermedad a nivel nacional, es decir repartidos entre los 32 estados de la república y 201,122 casos específicos para el estado de Puebla con el cual se trabajará este proyecto. Este conjunto de

datos contiene 40 registros que contienen información como sexo, estado, síntomas, resultados de pruebas contra la enfermedad, el estado de cada paciente, entre otras.

II.A. Registros para el estudio

Los registros seleccionados para el estudio discutido en este trabajo son relacionados con la salud de los pacientes y la infraestructura médica necesaria para sus tratamientos:

- Tipo de paciente
- Clasificación final
- Comorbilidad
- Intubado
- Unidad de Cuidados Intensivos (ICU)
- Muerte

Tipo de paciente: Este registro cuenta con dos posibles valores, los cuales son hospitalizado correspondiente al número 2 o ambulatorio correspondiente al número 1.

Clasificación final: En este registro se refiere a la conclusión a la que llegaron los médicos para determinar si el paciente es positivo o no al virus del SARS-COV-2, en caso de ser clasificado como positivo se le asignó el número 1 y en caso de ser negativo se le asignó el número 0.

Comorbilidad: Para este registro se tomaron en cuenta distintos registros conglomerados a uno. Dichos registros corresponden a si el paciente tuvo alguna enfermedad comorbida, es decir, si padeció o padece diabétes, EPOC, asma, inmunosupresión, otra complicación, alguna enfermedad cardiovascular, obesidad, enfermedad renal crónica o tabaquismo.

Intubado: En este registro se toma en cuenta si el paciente fue intubado con el número 1 o no con el número 0.

Unidad de Cuidados Intensivos (ICU): Para este registro se tomó en cuenta si el paciente pasó por el área de cuidados intensivos con el número 1 o no con el número 0.

Sexo: Con este registro se toma en consideración el sexo del paciente, el cuál puede ser 0 para sexo no definido, 1 para mujeres y 2 para hombres.

Muerte: En este registro se tomó en cuenta la fecha de deceso del paciente, en caso de no tener se asigna un 0 como valor y en caso de tener se asigna un 1.

II.B. Partición de datos

De los 7,252,170 casos alrededor de la república mexicana se particionaron de las siguiente manera con la finalidad de obtener las cadenas de markov correspondientes a:

- Hombres:
 - Resultado negativo a la prueba de COVID-19
 - Resultado positivo a la prueba de COVID-19 con cormobilidad

- Resultado positivo a la prueba de COVID-19 sin cormobilidad
- Mujeres:
 - Resultado negativo a la prueba de COVID-19
 - Resultado positivo a la prueba de COVID-19 con cormobilidad
 - Resultado positivo a la prueba de COVID-19 sin cormobilidad

Esta partición se muestra de manera específica en la tabla 1:

| Partición | Elementos |
|--------------------|-----------|
| República Mexicana | 7,252,170 |
| Puebla | 201,122 |
| Hombres | 98,781 |
| Mujeres | 102,341 |

Tabla 1. Partición del conjunto de datos

El conjunto de datos fue descargado en formato CSV y fue particionado usando scripts del lenguaje de programación Python.

III. Modelo de Procesos de Markov

El marco de referencia utilizado en este trabajo fue el Proceso de Markov [1]. El proceso de Markov es un modelo estocástico, que es una herramienta matemática usada para modelar fenómenos random dependientes del tiempo. La propiedad de Markov establece que la distribución del siguiente estado x_{n+1} depende sólo del estado actual x_n y no depende de los anteriores $x_{n-1}, x_{n-2}, \dots, x_1$.

III.A. Estados del proceso de Markov

Dado el proceso de Markov fueron definidas siete estados, los cuales son: Enfermo, Ambulatorio, Hospitalizado, Unidad de Cuidados Intensivos (UCI), Intubado, Recuperado o Vivo y Muerto. Los estados de Recuperado o Vivo y Muerto son nuestros estados absorbentes, ya que una vez que se alcance uno de estos dos estados, se permanecerá indefinidamente en estos.

Enfermo. Todos los casos del conjunto de datos son originados de acuerdo a los datos que maneja el centro de salud. Las personas enfermas con síntomas que los relacionen con la enfermedad de COVID-19. Sin embargo, esto no significa que el paciente tenga la enfermedad, pero esto lo hace un caso probable. En este modelo el tener la enfermedad es el estado de inicio para cada caso.

Ambulatorio. Algunas personas enfermas no pueden permanecer en los hospitales debido a que sus síntomas no son lo suficientemente graves y por lo tanto se les permite recuperarse en casa.

Hospitalizado. Por otro lado, algunas personas enfermas se les pide que permanezcan en el hospital para recibir su tratamiento. Esto suele pasar cuando sus síntomas son muchos y muy graves.

UCI. Este estado representa a los pacientes enfermos de gravedad que reciben tratamiento en una unidad de cuidados intensivos donde los vigilan con mayor cautela pero no se encuentran intubados para poder respirar.

Intubados. Este estado es representado por pacientes que pierden la habilidad de respirar por si mismos así que es necesario que reciban intubación o ayuda mecánica para seguir respirando.

Recuperados o Vivos. Estado representado por pacientes que se recuperaron de la enfermedad. Este es uno de nuestros dos estados absorbentes.

Muerto. Estado representado por pacientes que no se recuperaron de la enfermedad y murieron. Este es el segundo y último estado absorbente del modelo.

III.B. Procesos observados en la cadena de Markov

Es cierto que una cadena de Markov de tiempo discreto puede ajustarse a partir de los datos observados mediante su análisis estadístico. Con esto en mente, la información de los registros de nuestro conjunto de datos se utiliza para determinar el proceso observado para cada uno de los casos cerrados, es decir, para cada caso que haya llegado a un estado absorbente. A continuación se muestran los procesos observados en la cadena de Markov:

Enfermo \rightarrow Ambulatorio \rightarrow Recuperado o Vivo
Enfermo \rightarrow Ambulatorio \rightarrow Muerto
Enfermo \rightarrow Hospitalizado \rightarrow Recuperado o Vivo
Enfermo \rightarrow Hospitalizado \rightarrow Muerto
Enfermo \rightarrow Hospitalizado \rightarrow Intubado \rightarrow Recuperado o Vivo
Enfermo \rightarrow Hospitalizado \rightarrow Intubado \rightarrow Muerto
Enfermo \rightarrow Hospitalizado \rightarrow UCI \rightarrow Recuperado o Vivo
Enfermo \rightarrow Hospitalizado \rightarrow UCI \rightarrow Muerto
Enfermo \rightarrow Hospitalizado \rightarrow Intubado \rightarrow UCI \rightarrow Recuperado o Vivo
Enfermo \rightarrow Hospitalizado \rightarrow Intubado \rightarrow UCI \rightarrow Muerto

III.C. Ajuste del proceso de la cadena de Markov

Los 201,122 procesos observados se utilizan para ajustar el modelo de la cadena de Markov de tiempo discreto para cada una de las particiones. Este proceso de ajuste fue realizado bajo el lenguaje de programación Python. La librería utilizada para realizar las secuencias fue 'markovchain' y 'PyDTMC' con las funciones 'markovchain' y 'MarkovChain'.

Se generan 2 tablas de probabilidades para buscar la muerte o la recuperación de cada estado del modelo de cadena de Markov para las particiones consideradas. En este caso las tablas corresponden a mujeres como se puede observar en la figura 1 y hombres como se puede observar en la figura 2. Para cada tabla se generan 3 gráficos donde se puede ver de una mejor manera la cadena de Markov. Los niveles de confianza indicados para los tres modelos de la figura 1 es de 0,94 y para los correspondientes de la figura 2 es 0,93.

IV. Discusión

En esta sección se da una interpretación de los modelos de cadenas de Markov en tiempo discreto. El modelo obtenido de las particiones negativas al SARS-COV-2 se utiliza como base. El gráfico del proceso para este grupo de individuos se muestra en las figuras 3 y 4.

Podemos observar que para las figuras 3 y 4 las personas enfermas que no están hospitalizadas (estado ambulatorio) tienen una recuperación completa. De igual manera podemos observar que para las personas que si son hospitalizadas las probabilidades de que se recuperen comienzan a

| | MUJERES | | | | | |
|---------------|---------------------|--------|-----------------|--------|---------------------|--------|
| | POSITIVO A COVID-19 | | | | NEGATIVO A COVID-19 | |
| | COMORBILIDAD | | NO COMORBILIDAD | | | |
| | RECUPERADO | MUERTO | RECUPERADO | MUERTO | RECUPERADO | MUERTO |
| ENFERMO | 0.7962 | 0.2038 | 0.9571 | 0.0429 | 0.9827 | 0.0173 |
| AMBULATORIO | 0.9661 | 0.0339 | 0.9951 | 0.0049 | 0.9982 | 0.0018 |
| HOSPITALIZADO | 0.5236 | 0.4764 | 0.7997 | 0.2003 | 0.8562 | 0.1438 |
| UCI | 0.3721 | 0.6279 | 0.7055 | 0.2945 | 0.8013 | 0.1987 |
| INTUBADO | 0.115 | 0.885 | 0.4815 | 0.5185 | 0.5852 | 0.4148 |

Figura 1: Probabilidad para buscar la muerte o la recuperación de cada estado del modelo de cadena de Markov para las particiones consideradas en mujeres

| | HOMBRES | | | | | |
|---------------|---------------------|--------|-----------------|--------|---------------------|--------|
| | POSITIVO A COVID-19 | | | | NEGATIVO A COVID-19 | |
| | COMORBILIDAD | | NO COMORBILIDAD | | | |
| | RECUPERADO | MUERTO | RECUPERADO | MUERTO | RECUPERADO | MUERTO |
| ENFERMO | 0.7363 | 0.2637 | 0.9409 | 0.0591 | 0.9711 | 0.0289 |
| AMBULATORIO | 0.9484 | 0.0516 | 0.9937 | 0.0063 | 0.9969 | 0.0031 |
| HOSPITALIZADO | 0.4611 | 0.5389 | 0.7576 | 0.2424 | 0.8115 | 0.1885 |
| UCI | 0.3562 | 0.6438 | 0.6275 | 0.3725 | 0.6919 | 0.3081 |
| INTUBADO | 0.1473 | 0.8527 | 0.4949 | 0.5051 | 0.5417 | 0.4583 |

Figura 2: Probabilidad para buscar la muerte o la recuperación de cada estado del modelo de cadena de Markov para las particiones consideradas en hombres

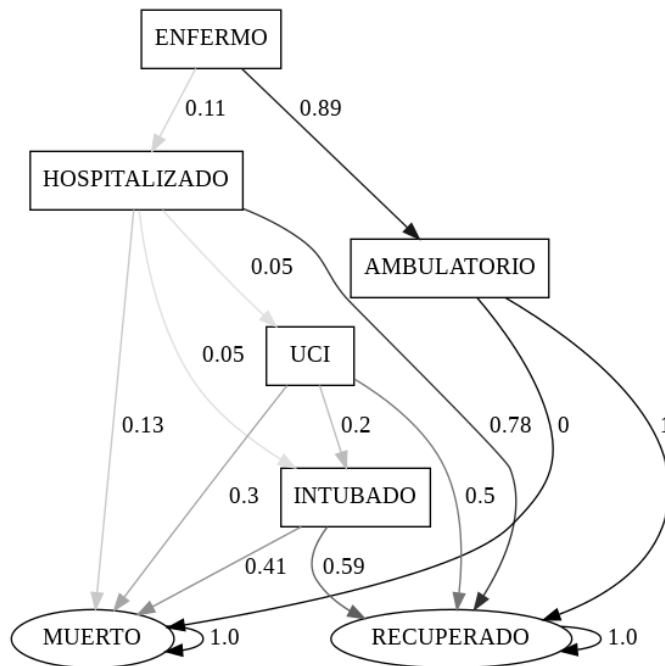


Figura 3: El grafico para el proceso de Markov para los casos negativos en SARS-COVID-2 en mujeres

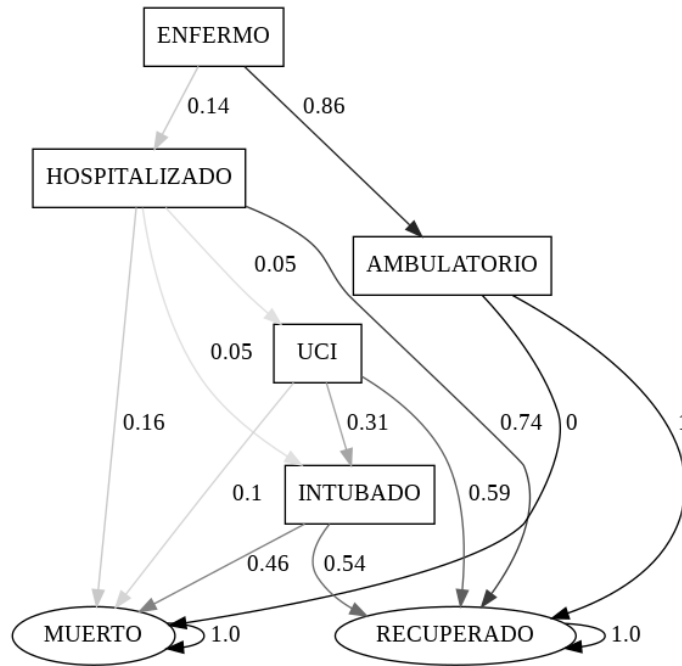


Figura 4: El grafico para el proceso de Markov para los casos negativos en SARS-COVID-2 en hombres

disminuir. Sin embargo, hay una gran probabilidad de que se si son hospitalizados se recuperen con un 0,78 y un 0,74 para mujeres y hombres respectivamente. Siguido esta línea de hospitalización de pacientes, podemos observar que mientras más tiempo pasen en el centro médico, ya sea en una UCI o intubado reducen sus pobabilidades de recuperarse hasta casi equipararlas a las probablidades de muerte. Donde si pasan por una UCI las probabilidades de recuperarse bajan a 0,5 y 0,59 para mujeres y hombres respectivamente. Por último, si pasan por una UCI y además son intubados las porbabilidades de recuperarse son de 0,59 y de 0,54 para mujeres y hombres respectivamente.

Mientras que para las figuras 5 y 6 las personas enfermas que no están hospitalizadas (estado ambulatorio) tienen una recuperación completa en las mujeres y un 0,99 de probabilidad de recuperación en los hombres. De igual manera podemos observar que para las personas que si son hospitalizadas las probabilidades de que se recuperen comienzan a disminuir. Sin embargo, hay una gran probabilidad de que se si son hospitalizados se recuperen con un 0,64 y un 0,51 para mujeres y hombres respectivamente. Siguido esta línea de hospitalización de pacientes, podemos observar que mientras más tiempo pasen en el centro médico, ya sea en una UCI o intubado reducen sus pobabilidades de recuperarse hasta casi equipararlas a las probablidades de muerte. Donde si pasan por una UCI las probabilidades de recuperarse bajan a 0,2 tanto para mujeres como para hombres. Por último, si pasan por una UCI y además son intubados las porbabilidades de recuperarse son de 0,48 y de 0,49 para mujeres y hombres respectivamente.

Por último, podemos observar que para las figuras 7 y 8 las personas enfermas que no están hospitalizadas (estado ambulatorio) tienen una recuperación de 0,97 y 0,95 para mujeres y hombres respectivamente. De igual manera podemos observar que para las personas que si son hospitalizadas las probabilidades de que se recuperen comienzan a disminuir. Sin embargo, hay una probabilidad de que se si son hospitalizados se recuperen con un 0,47 y un 0,4 para mujeres y hombres respectivamente. Siguido esta línea de hospitalización de pacientes, podemos observar que mientras más tiempo pasen en el centro médico, ya sea en una UCI o intubado reducen sus pobabilidades de recuperarse hasta casi equipararlas a las probablidades de muerte. Donde si pasan por una UCI

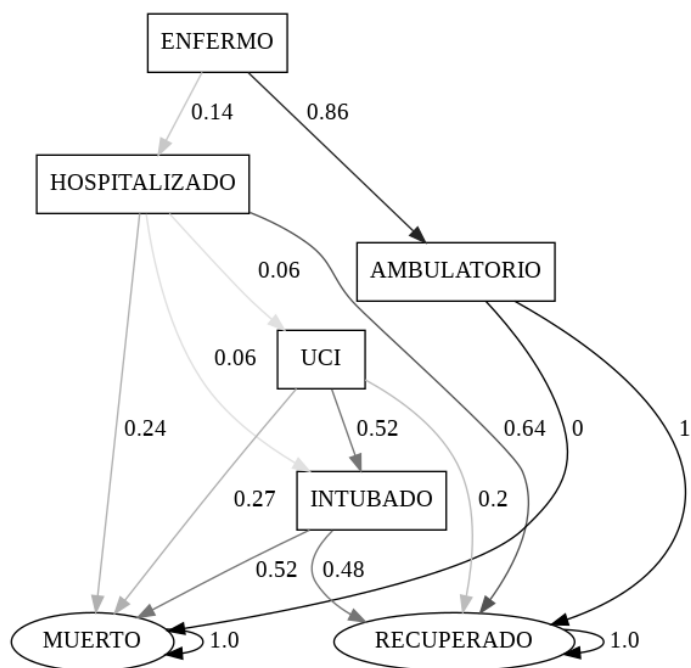


Figura 5: El grafico para el proceso de Markov para los casos positivos en SARS-COVID-2 y no comorbilidad en mujeres

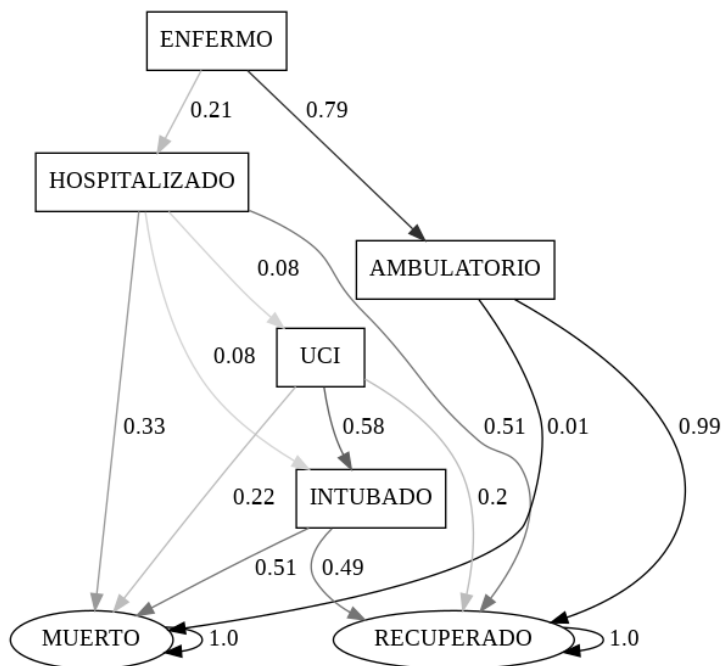


Figura 6: El grafico para el proceso de Markov para los casos positivos en SARS-COVID-2 y no comorbilidad en hombres

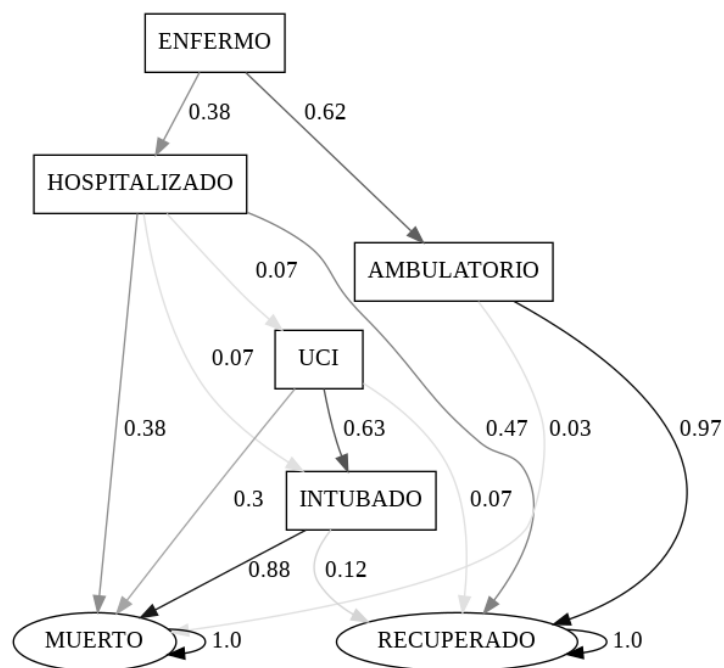


Figura 7: El grafico para el proceso de Markov para los casos positivos en SARS-COVID-2 y comorbilidad en mujeres

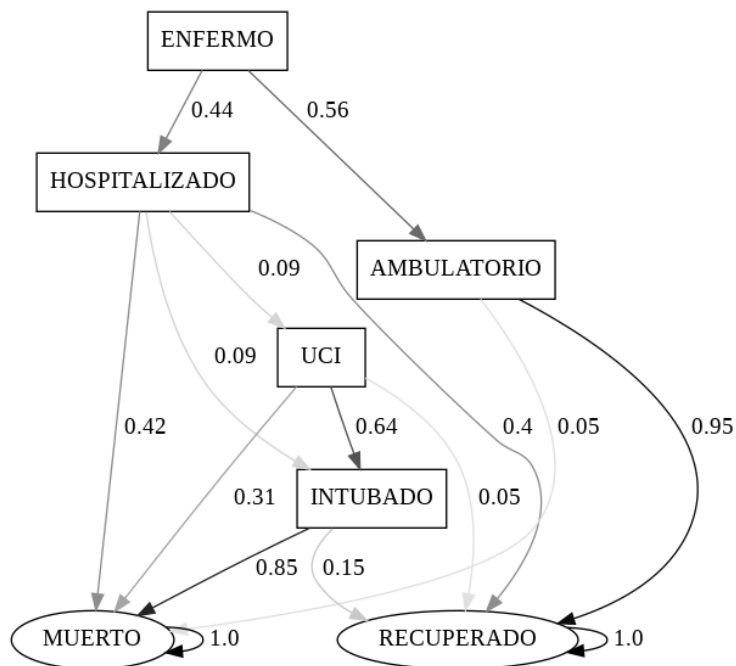


Figura 8: El grafico para el proceso de Markov para los casos positivos en SARS-COVID-2 y comorbilidad en hombres

las probabilidades de recuperarse bajan a 0,07 y 0,05 para mujeres y hombres respectivamente. Por último, si pasan por una UCI y además son intubados las probabilidades de recuperarse son de 0,12 y de 0,15 para mujeres y hombres

V. Conclusiones

El proceso de las cadenas de Markov puede ser usado para observar el comportamiento de los pacientes para una enfermedad con tantos datos como lo es el COVID-19, con esto podemos modelar probabilidades que nos pueden sugerir cuales son los grupos de personas más vulnerables ante la enfermedad, así como también nos indica la probabilidad de que el paciente se recupere si se llega a seguir una ruta determinada a partir del tratamiento que se le esté asignando o no.

En este modelo en particular se puede observar que las probabilidades de recuperación de las personas con comorbilidad ya sean mujeres u hombres tienen una gran disminución en comparación con los que no lo tienen. Esto hace sentido ya que las personas que pertenecen a este grupo son un grupo vulnerable ya que cuentan con antecedentes médicos que pueden agravar la situación de la enfermedad.

A pesar de que este trabajo se haya realizado a partir de los datos de un solo estado de la república mexicana, esto se puede realizar a nivel nacional o para realizar comparativas contra como otros estados lo han estado sobrellevando y así obtener las diferencias y ver que se puede hacer mejor y así mejorar las probabilidades de que las personas se recuperen.

Referencias

- [1] Ricardo C. Villarreal-Calva, Ponciano J. Escamilla-Ambrosio, Abraham Rodríguez-Mota y Juan M. Ramírez-Cortés. "Evolution of COVID-19 Patients in Mexico City Using Markov Chains". En: *Telematics and Computing*. Ed. por Miguel Félix Mata-Rivera, Roberto Zagal-Flores y Cristian Barria-Huidobro. Cham: Springer International Publishing, 2020, págs. 309-318. ISBN: 978-3-030-62554-2.