



Universidad del Desarrollo

Facultad de Ingeniería

IMPACTO VAC-COVID19-CHILE: ANÁLISIS DE LA VACUNACIÓN Y SU EFECTO EN LA DINÁMICA PANDÉMICA

Correlaciones y Tendencias en la Era de la Vacunación: Un Estudio Detallado de Casos,
Hospitalizaciones y Mortalidad en Chile

POR: RICARDO MIRANDA ARAYA Y SEBASTIAN DANKER GALDAMES

Proyecto de grado presentado a la Facultad de Ingeniería de la
Universidad del Desarrollo para optar al grado académico de Magíster en
Data Science

PROFESOR GUÍA:

PhD. MAURICIO HERRERA MARIN

Diciembre 2023

SANTIAGO

Aprender a transformar datos en información es
la clave para el futuro.

AGRADECIMIENTO

Es con un profundo sentido de gratitud que nosotros, Ricardo y Sebastian, nos dirigimos a ustedes al concluir nuestro viaje en el Máster en Ciencia de Datos. Estos 18 meses han sido un período de intenso aprendizaje, desafíos y crecimiento personal y profesional, y sabemos que no hubiéramos podido alcanzar este hito sin su inestimable apoyo.

A nuestras familias: Gracias por ser nuestro soporte incondicional en los momentos más exigentes. Vuestro aliento, comprensión y paciencia han sido fundamentales para mantener nuestro enfoque y motivación. Vuestra creencia en nosotros ha sido una fuente constante de fuerza.

A nuestros profesores: Su dedicación, conocimiento y pasión por la enseñanza han sido inspiradores. Nos han equipado no solo con habilidades técnicas en ciencia de datos, sino también con una mentalidad crítica y un enfoque ético hacia nuestro campo. Las lecciones aprendidas van más allá del aula y nos acompañarán a lo largo de nuestra carrera.

Este viaje no solo ha sido académico; ha sido una experiencia enriquecedora que nos ha enseñado sobre la perseverancia, el trabajo en equipo y la importancia de la comunidad. Estamos agradecidos por haber tenido la oportunidad de aprender de cada uno de ustedes y por la comunidad que hemos formado juntos.

A medida que avanzamos hacia la próxima etapa de nuestras vidas profesionales, llevamos con nosotros no solo el conocimiento adquirido, sino también los recuerdos y las relaciones forjadas durante este tiempo. Esperamos poder aplicar lo aprendido y contribuir significativamente a nuestra área.

Gracias una vez más por su invaluable apoyo y por ser parte de esta importante etapa de nuestras vidas.

TABLA DE CONTENIDO

RESUMEN.....	1
1. INTRODUCCIÓN.....	2
2. TRABAJO RELACIONADO	3
3. HIPÓTESIS Y OBJETIVOS.....	5
4. DATOS Y METODOLOGÍA.....	7
4.1. DATOS	7
4.2. METODOLOGÍA	8
5. RESULTADOS	9
6. CONCLUSIONES.....	11
BIBLIOGRAFÍA	13

Resumen

El estudio propone que la campaña de vacunación masiva en Chile ha llevado a una notable disminución en la incidencia de casos, hospitalizaciones y defunciones relacionadas con COVID-19. Utilizando datos del Ministerio de Ciencia y Salud del Gobierno de Chile y aplicando técnicas avanzadas de ciencia de datos, el análisis busca demostrar la efectividad de la vacunación y fortalecer la confianza pública en las vacunas. Los objetivos específicos incluyen analizar la correlación entre la tasa de vacunación y la incidencia de casos, hospitalizaciones y muertes, y evaluar la efectividad de diferentes tipos de vacunas utilizadas en Chile. El estudio utiliza modelado estadístico y técnicas de machine learning para asegurar resultados precisos y fiables, contribuyendo al conocimiento sobre la eficacia de las campañas de vacunación y sirviendo como herramienta para la planificación de estrategias de salud pública en el futuro.

1. Introducción

La pandemia de COVID-19, una crisis sanitaria global provocada por el coronavirus SARS-CoV-2, ha representado un desafío sin precedentes para la salud pública, la economía y la sociedad. En este contexto, la campaña de vacunación se ha convertido en el eje central de las estrategias para combatir la propagación del virus. Este estudio se enfoca en Chile, un país que ha destacado por su rápida y extensiva implementación de un programa de vacunación contra COVID-19. A través de un análisis exhaustivo, esta investigación busca evaluar el impacto de la vacunación en la incidencia de casos, hospitalizaciones y defunciones relacionadas con el virus. Utilizando datos gubernamentales y aplicando técnicas avanzadas de ciencia de datos, el proyecto tiene como objetivo no solo medir la efectividad de la vacunación, sino también proporcionar insights cruciales para la toma de decisiones en políticas de salud pública. Este análisis pretende ser un aporte significativo en la comprensión de cómo las intervenciones de vacunación pueden moldear el curso de una pandemia, ofreciendo lecciones valiosas para futuras emergencias sanitarias.

2. Trabajo Relacionado

Título del Proyecto: Impacto Vac-Covid19-Chile

Resumen del Proyecto:

Este documento acompaña al código fuente de nuestro proyecto, Impacto Vac-Covid19-Chile, el cual está alojado en GitHub. El objetivo de este proyecto es la vacunación masiva es un factor crítico en la reducción de la transmisión del virus. A través de este trabajo, buscamos explicar los beneficios o la importancia de las vacunas y su comportamiento en Chile para futuras pandemias.

Enlace al Código Fuente en GitHub: <https://github.com/SDanker/IMPACTO-VAC-COVID19-CHILE>

Detalles del Código:

El repositorio de GitHub contiene todos los archivos y recursos utilizados para el desarrollo de nuestro proyecto. Esto incluye Jupyter notebook de nuestro trabajo en 3 ocasiones siendo el tercero el último y más completo, las bases de datos utilizados y textos de nuestro proyecto de título. Hemos asegurado que el código esté bien comentado y documentado para facilitar su comprensión y reutilización.

Colaboradores:

Este proyecto ha sido un esfuerzo colaborativo de Ricardo Miranda A. y Sebastian Danker G. Agradecemos a todos los que han contribuido con su tiempo, experiencia y conocimientos.

Instrucciones de Uso:

Para aquellos interesados en utilizar o contribuir a nuestro proyecto, recomendamos revisar el archivo README.md en nuestro repositorio de GitHub. Este archivo contiene instrucciones detalladas sobre cómo configurar el entorno, ejecutar el código y comprender las funcionalidades del proyecto.

Comentarios y Contribuciones:

Estamos abiertos a recibir comentarios y sugerencias para mejorar nuestro proyecto. Si estás interesado en colaborar o tienes alguna pregunta, no dudes en abrir un issue en el repositorio de GitHub o contactarnos directamente.

3. Hipótesis y Objetivos

La presente investigación se centra en examinar el impacto de la campaña de vacunación masiva contra el COVID-19 en Chile, proponiendo como hipótesis central que esta ha conducido a una disminución notable en la incidencia de casos, hospitalizaciones y defunciones relacionadas con el virus. El estudio se propone no solo demostrar la efectividad de la vacunación como herramienta clave en la lucha contra la pandemia, sino también fortalecer la confianza pública en las vacunas, proporcionar datos valiosos para la toma de decisiones informadas por parte de las autoridades sanitarias y ofrecer una perspectiva clara sobre el avance hacia la recuperación y normalidad post-pandémica.

La hipótesis se basa en la premisa de que la vacunación masiva es un factor crítico en la reducción de la transmisión del virus, la severidad de los casos y, en última instancia, la mortalidad asociada al COVID-19. Para probar esta hipótesis, el estudio se enfocará en el análisis detallado de los datos epidemiológicos disponibles, examinando las tendencias y patrones en la incidencia de casos, hospitalizaciones y muertes antes y después de la implementación de la campaña de vacunación en Chile.

Entre los objetivos específicos del estudio se incluyen:

1. **Analizar la correlación entre la tasa de vacunación y la incidencia de casos, hospitalizaciones y muertes por COVID-19:** Esto implicará la recopilación y el análisis de datos para determinar cómo las tasas de vacunación se relacionan con las tendencias en los indicadores claves de la pandemia. Se buscará entender si las regiones o grupos demográficos con mayores tasas de vacunación muestran una disminución más marcada en estos indicadores.
2. **Evaluar la efectividad de diferentes tipos de vacunas administradas en Chile:** Dado que en Chile se han utilizado diversas vacunas, el estudio buscará comparar su efectividad. Esto incluirá el análisis de la respuesta inmunitaria, la duración de la protección ofrecida y la eficacia contra distintas variantes del virus.

Este análisis se realizará utilizando métodos avanzados de ciencia de datos, incluyendo el modelado estadístico y técnicas de machine learning, para garantizar resultados precisos y fiables. Se espera que los hallazgos del estudio contribuyan significativamente al conocimiento existente sobre la eficacia de las campañas de vacunación y sirvan como una herramienta valiosa para la planificación de estrategias de salud pública en el futuro, tanto en Chile como en otros contextos internacionales.

4. Datos y Metodología

4.1. Datos

El estudio se fundamenta en un conjunto de datos exhaustivo y minuciosamente curado, proporcionado por el Ministerio de Ciencia y Salud del Gobierno de Chile. Este conjunto de datos presenta información detallada y actualizada sobre la incidencia del COVID-19 en Chile, abarcando tanto a individuos vacunados como a aquellos que no han recibido la vacuna. Este recurso de datos es crucial para el análisis, ya que incluye una amplia variedad de indicadores clave: número de casos confirmados de COVID-19, hospitalizaciones, admisiones a Unidades de Cuidados Intensivos (UCI) y fallecimientos. Además, los datos están meticulosamente desglosados por estado de vacunación y grupos de edad, lo que permite realizar un análisis diferenciado y más preciso.

Al contar con información oficial y directa de las autoridades sanitarias, el análisis se beneficia de una mayor precisión y fiabilidad. Estos datos no solo reflejan la situación epidemiológica actual del país, sino que también proporcionan una perspectiva integral de cómo la vacunación ha influido en la evolución de la pandemia en diferentes segmentos de la población chilena.

Además, el uso de esta fuente de datos gubernamentales asegura que el estudio se alinea con los estándares nacionales e internacionales en términos de recopilación y tratamiento de la información sanitaria. La minuciosa clasificación de los datos por grupos de edad y estado de vacunación permite realizar un análisis más detallado y específico, crucial para entender las dinámicas de la transmisión del virus y la eficacia de las vacunas en distintos segmentos poblacionales.

Este abordaje metódico y riguroso en la selección y análisis de los datos proporcionados por el Ministerio es fundamental para alcanzar conclusiones robustas y significativas acerca del impacto de la vacunación en la trayectoria de la pandemia en Chile.

4.2. Metodología

1. Limpieza y Preprocesamiento de Datos:

La metodología de este estudio se inició con una fase crítica de limpieza y preprocesamiento de datos. Este proceso comenzó con la importación cuidadosa de los datos proporcionados por el Ministerio de Ciencia y Salud del Gobierno de Chile. Posteriormente, se revisaron los datos para identificar y manejar valores faltantes, una práctica esencial para garantizar la integridad y coherencia del conjunto de datos. Además, se realizó una corrección y estandarización de los tipos de datos para asegurar la interpretación y utilización adecuada de todas las variables en análisis futuros. Durante esta fase, también se crearon nuevas variables que permitieron una exploración más detallada y específica de ciertos aspectos del conjunto de datos.

2. Análisis Exploratorio de Datos (EDA):

Tras la limpieza y el preprocesamiento, se procedió con un Análisis Exploratorio de Datos. Esta etapa implicó el uso de técnicas de visualización de datos avanzadas para identificar posibles valores atípicos, así como para verificar la coherencia y adecuación de los tipos de datos. Las visualizaciones, tales como gráficos y diagramas, fueron herramientas clave para revelar patrones, tendencias y anomalías en los datos. Este análisis exploratorio permitió obtener una visión preliminar del conjunto de datos, facilitando la formulación de hipótesis y la definición de las direcciones para los análisis estadísticos más profundos.

3. Modelado y Análisis Estadístico:

La siguiente fase del estudio involucró el modelado estadístico y el análisis predictivo. Utilizando métodos avanzados de ciencia de datos, se analizaron las relaciones entre la vacunación y los diferentes indicadores de la pandemia (casos, hospitalizaciones, ingresos a UCI y fallecimientos). Este análisis buscó no solo establecer correlaciones, sino también explorar causas y efectos potenciales, proporcionando una comprensión más profunda del impacto de la campaña de vacunación en la trayectoria de la pandemia en Chile.

4. Interpretación y Presentación de Resultados:

Finalmente, los resultados obtenidos fueron interpretados en el contexto del panorama más amplio de la pandemia de COVID-19 en Chile. Se puso especial énfasis en la presentación clara y precisa de los hallazgos, asegurando que los insights fueran accesibles tanto para las autoridades sanitarias como para el público en general. Esta fase también incluyó la discusión de las implicaciones de los resultados para las futuras políticas de salud pública y estrategias de vacunación en Chile.

Esta metodología integral garantiza un enfoque riguroso y basado en datos para entender el impacto de la vacunación contra el COVID-19, proporcionando insights valiosos y prácticos para la gestión de la pandemia en Chile.

5. Resultados

1. Modelo de Regresión Lineal (Primer Modelo):

- Incidencia de casos: $\text{Casos} = 429.54 \times \text{Tasa de vacunación} - 10153.17$ 2.
- Incidencia de UCI: $\text{UCI} = 12.64 \times \text{Tasa de vacunación} - 358.10$ 3.
- Incidencia de fallecimientos: $\text{Fallecimientos} = 6.43 \times \text{Tasa de vacunación} - 157.66$

Estas ecuaciones representan cómo varían las incidencias de casos, UCI y fallecimientos según las tasas de vacunación

2. Modelos de Machine Learning (Segundo Modelo):

- **Modelo de Máquinas de Soporte Vectorial (SVM) y Modelo de Bosque Aleatorio (Random Forest):** Para la predicción de la incidencia de casos sin vacunación en función de la tasa de vacunación, se obtuvieron los siguientes resultados:
 - Error Cuadrático Medio (MSE) para SVM: 205,681,274.86
 - Coeficiente de Determinación (R^2) para SVM: -0.3559 (indicando un modelo no adecuado)

- MSE para Random Forest: 5,703,877.01
- (R^2) para Random Forest: 0.9624 (indicando un buen ajuste)

3. Validación Cruzada y Análisis de Importancia de Características (Tercer Modelo):

- **Modelo de Bosque Aleatorio con Validación Cruzada:**
 - Puntuaciones (R^2) en validación cruzada: [-4.856, 0.725, -0.666, -0.270, -19.790]
 - Media de las puntuaciones (R^2): -4.9714
 - Desviación estándar de las puntuaciones (R^2): 7.6511
 - Importancia de la característica (solo una en este caso): [1.0]

Estos resultados proporcionan una visión cuantitativa de los modelos empleados en el análisis. Mientras que el modelo de Random Forest muestra un buen ajuste para la predicción de la incidencia de casos no vacunados, la validación cruzada indica una variabilidad significativa en el rendimiento del modelo. El modelo SVM, por otro lado, parece no ser adecuado para esta tarea particular.

6. Conclusiones

La pandemia de COVID-19 ha presentado desafíos sin precedentes a nivel global, y la implementación de programas de vacunación ha sido un pilar fundamental en la lucha contra este virus. Examinando cómo la tasa de vacunación influye en la dinámica de la enfermedad, particularmente en términos de incidencia de casos, hospitalizaciones y admisiones en UCI. A través de técnicas avanzadas de análisis de datos y modelización predictiva, el estudio busca entender la eficacia de la vacunación y su papel en la contención de la pandemia. A continuación, se presentan las conclusiones clave derivadas de este análisis.

1. Relación entre Tasa de Vacunación e Incidencia de COVID-19:

- El estudio demuestra una correlación significativa entre la tasa de vacunación y la reducción en la incidencia de casos confirmados, hospitalizaciones, y admisiones en UCI, lo que sugiere que una mayor tasa de vacunación contribuye efectivamente a la disminución de estos indicadores.

2. Eficacia de los Modelos Predictivos:

- Los modelos de Machine Learning, especialmente el Modelo de Bosque Aleatorio, mostraron ser herramientas potentes para predecir la incidencia de casos de COVID-19 en base a la tasa de vacunación, con un (R^2) alto de 0.9624, indicando una buena capacidad predictiva.
- Sin embargo, la variabilidad en las puntuaciones (R^2) observada en la validación cruzada sugiere que el modelo puede comportarse de manera inconsistente bajo diferentes conjuntos de datos, lo que destaca la necesidad de cautela en su interpretación y aplicación.

3. Importancia de la Tasa de Vacunación:

- El modelo de Bosque Aleatorio, centrado únicamente en la tasa de vacunación como variable predictiva, reafirma la importancia crítica de esta variable en la modelización de los impactos de COVID-19, aunque también señala la posible necesidad de considerar otras variables para capturar la complejidad total de la dinámica de la enfermedad.

En conclusión, el análisis subraya la importancia de la vacunación en la gestión de la pandemia de COVID-19, proporcionando evidencia cuantitativa de su efecto en la reducción de la incidencia de la enfermedad. Al mismo tiempo, resalta la utilidad y también los límites de los modelos predictivos en la interpretación de tendencias epidemiológicas, lo que puede ser crucial para la formulación de políticas de salud pública y estrategias de intervención. Estos resultados refuerzan la necesidad de continuar con las campañas de vacunación y de combinarlas con otras medidas de salud pública para controlar eficazmente la pandemia.

Bibliografía

Las bases iniciales que iniciaremos nuestro modelo son desde el Github de Ministerio de Ciencia, Tecnología, Conocimiento, e Innovación (<https://github.com/MinCiencia>) del repositorio Datos-COVID19 (<https://github.com/MinCiencia/Datos-COVID19>). Las bases son las siguientes

Data Product 88 - Vacunación por fabricante Total de vacunas por fabricante.

Data Product 89 - Incidencia de casos según estado de vacunación, grupo de edad, y semana epidemiológica El producto contiene información sobre los casos confirmados, ingresos a UCI y defunciones, según estado de vacunación, agrupado por tramo etario.

Data Product 90 - Distribución de casos según antecedentes de vacunación y semana epidemiológica. El producto contiene información sobre los casos confirmados, ingresos a UCI y defunciones, según estado de vacunación, agrupado por semana epidemiológica.