

# Aplicación de un modelo de aprendizaje no supervisado para la identificación de perfiles en riesgo de hospitalización por efectos adversos causados por la vacuna contra el Covid-19.

Nicolle Pereira

Universidad Nacional Mayor de San Marcos  
Lima, Peru

Martina Guillermo

IETSI, ESSALUD  
Lima, Peru

Paulo Vela

Universidad Peruana Cayetano Heredia  
Lima, Peru

David Guevara

Universidad Nacional de San Antonio de  
Abad del Cuzco  
Lima, Peru

Angel Ramirez

Universidad Nacional Mayor de San Marcos  
Lima, Peru

Kathy Bolivar

DAS-MINSA  
Lima, Peru

Leidy Pola

Universidad Nacional Mayor de San Marcos  
Lima, Peru

Alvaro Rodríguez

Universidad Nacional de Trujillo  
Lima, Peru

Eduardo Apari

Universidad Nacional Mayor de San Marcos  
UTEC  
Lima, Peru

**Abstract—** La pandemia de COVID-19 se encuentra a punto de iniciar la tercera ola en el Perú, poniendo en riesgo el avance de las campañas de inmunización para contrarrestar el aumento de contagios por nuevas variantes. Frente a ello, se ha incrementado la tasa de vacunación sin monitorear la aparición de efectos adversos, los cuáles pueden llevar a la hospitalización de los pacientes en riesgo, perjudicando su calidad de vida y generando gastos adicionales para el estado. **Objetivo:** Determinar los perfiles de riesgo de hospitalización en pacientes con efectos adversos provocados por inmunización contra COVID-19 aplicando un modelo de aprendizaje no supervisado. **Metodología:** A partir de un dataset de efectos adversos de la vacuna Pfizer obtenido de VAERS, se aplicaron modelos de aprendizaje no supervisado K-Means y Hierarchical Clustering para determinar posibles agrupamientos relacionados a términos AESI. **Resultados:** El dataset final para procesamiento contó con 1373 observaciones, tras analizarlos con algoritmos de Natural Language Processing. El número óptimo de clústeres del dataset final fue 3, encontrado usando el “Método del Codo”. PCA y Hierarchical clustering mostraron resultados similares. **Conclusiones:** Los modelos aplicados son capaces de identificar 3 clusters definidos respecto a la edad y tiempo de inicio de eventos adversos post-vacunación, en relación a los días de hospitalización experimentados, siendo el cluster más relevante aquel donde los pacientes registraban edades mayor a los 50 años e inicio de eventos adversos menor a los 50 días.

**Keywords—** Machine Learning, Unsupervised Learning, Adverse reactions, AESIS, Vaccination, Covid-19

## I. INTRODUCCIÓN

En diciembre del 2019, un nuevo coronavirus denominado como SARS-CoV-2 surgió en la ciudad de Wuhan, China, el cual ha ido propagándose rápidamente por todo el mundo, teniendo presencia en más de 181 países y territorios. En la actualidad (agosto del 2021), la Organización Mundial de la Salud (OMS) ha reportado más de 214 millones de casos confirmados y más de 4 millones de muertes a nivel mundial; convirtiéndose así en el principal problema de salud pública al momento (OMS, 2021).

Entre las diferentes estrategias para combatir esta nueva pandemia, se encuentra el desarrollo de vacunas que son elementos indispensables para la salud pública a nivel global, dado que brindan protección a la población al generar inmunidad adquirida frente a múltiples enfermedades infecciosas, previniendo su propagación (Gao et al, 2020). En el marco de la pandemia de COVID-19, se han desarrollado

en distintos países más de 270 vacunas candidatas para contrarrestar el avance del virus, y brindar inmunización frente a las consecuencias propias de esta enfermedad. Sin embargo, la aplicación de vacunas frecuentemente es acompañada por la aparición de efectos adversos, los cuáles pueden ocasionar síntomas de gravedad que implique el ingreso a hospitalización, generando gastos adicionales al sistema de salud y perjuicios a la calidad de vida del paciente.

La vigilancia de seguridad de la vacuna del COVID-19 facilita la detección temprana, investigación y análisis de eventos adversos posteriores a la inmunización (AEFIS) y eventos adversos de especial interés (AESIS) para asegurar una respuesta rápida y adecuada, y su objetivo es disminuir el impacto negativo de los eventos en la salud de las personas tras su participación en los programas de inmunización, sosteniendo la confianza de los trabajadores de salud y la población en general. En EEUU, el Centro de Control de Enfermedades (CDC), a través del Sistema de Notificación de Eventos Adversos a las Vacunas (VAERS), monitorea la seguridad de las vacunas detectando patrones de notificación inusuales o inesperados de los eventos adversos reportados. Asimismo, en el Perú, la Dirección General de Medicamentos Insumos y Drogas (DIGEMID) a través del Centro Nacional de Farmacovigilancia y Tecnovigilancia (CENAFyT) vigila la seguridad de las vacunas para la COVID-19, sin embargo, la información colectada no es pública.

El objetivo del presente estudio es generar modelos utilizando aprendizaje automatizado para determinar los perfiles de riesgo de hospitalización en paciente con efectos adversos provocados por inmunización contra COVID-19, esto con el fin de disminuir tanto los casos graves como decesos.

## II. METODOLOGÍA

Se utilizó la base de datos del VAERS del año 2021, de la cual se tomaron dos datasets (VAERSDATA, VAERSVAX) que fueron unidos por el código de cada paciente y se realizó el análisis exploratorio para cada variable. La data luego fue preprocesada aplicando algoritmos de procesamiento de lenguaje natural a fin de obtener los reportes calificados como AESI. Después se realizó un último corte de la data que reportaba eventos AESI. Finalmente se aplicaron dos algoritmos no supervisados de machine learning (Kmeans y PCA) para agrupar las características principales de los pacientes hospitalizados con eventos AESI. El diagrama del estudio se detalla en la **figura 1**.

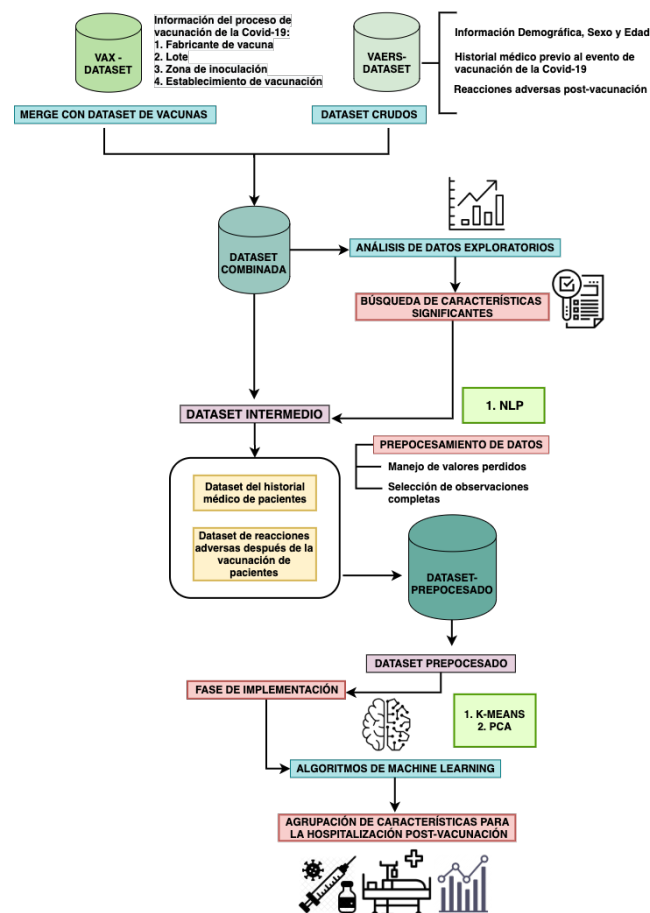


Fig.1 Diseño de Investigación

## 1. Obtención de bases de datos.

Los conjuntos de datos fueron obtenidos de la página web de VAERS (<https://vaers.hhs.gov/data/datasets.html>) en formato CSV (Comma separated values), tomando en consideración la última actualización hasta el 31 de julio del 2021. El conjunto de datos públicos VAERS descargables consta de 3 archivos de datos separados:

- **VAERSDATA.CSV:** proporciona una descripción detallada de los datos proporcionados en cada campo.
- **VAERSVAX.CSV:** proporcione la información restante de la vacuna (nombre de la vacuna, fabricante, número de lote, ruta, lugar y número de dosis administradas anteriormente), para cada una de las vacunas enumeradas. VAX\_DOSE se suspendió en el formulario VAERS 2.

Además en todos los datasets se observan datos de la siguiente naturaleza:

- NUM (float64) - datos numéricos
- CHAR (object) - texto o datos de "carácter"
- DATE (not formatted) = campos de fecha en formato "mm/dd/AAAA"

De estos, aproximadamente el 85-90% de los informes de eventos adversos de la vacuna se refieren a eventos relativamente menores, como fiebre o enrojecimiento e hinchazón en el lugar de la inyección. Los informes restantes (menos del 15%) describen eventos graves, como hospitalizaciones, enfermedades potencialmente mortales o muertes.

La **Tabla 1** (adjunta en ANEXOS) muestra la arquitectura del conjunto de datos, con una descripción de cada característica y el formulario de datos asociado y los valores de restricción.

## 2. Colección de Datos.

El dataset utilizado fue obtenido de la plataforma del VAERS, teniendo en cuenta la actualización hasta el 31 de julio del 2021. La recolección de datos fue basada en la versión del formulario en línea del VAERS.

## 3. Preparación de bases de datos y Análisis Exploratorio.

Se realizaron cortes y reducción de observaciones para cumplir con nuestro objetivo, el cual se describe en la **figura 2**.

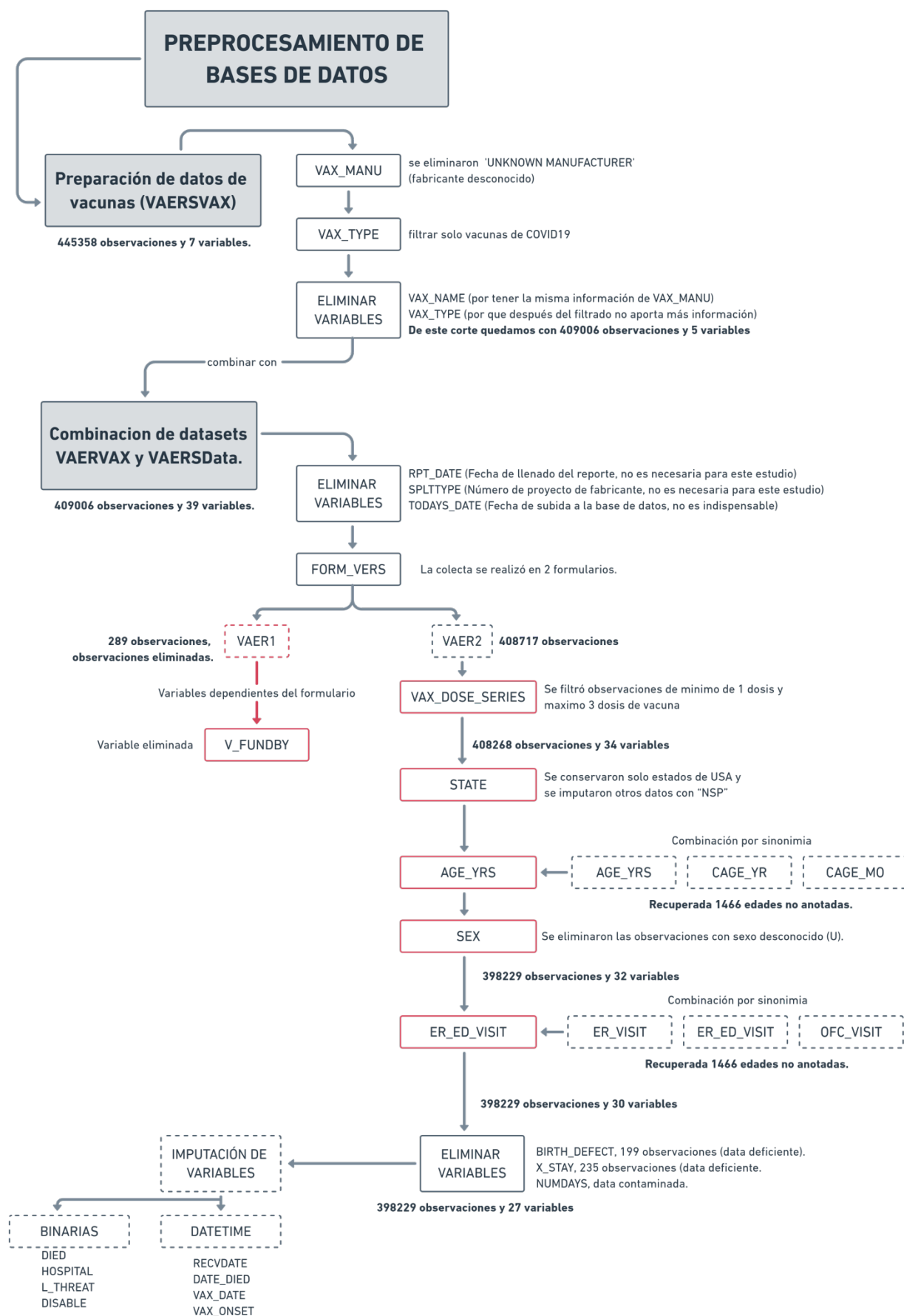


Fig.2. Pre-procesamiento de datos

### 3.1. Preparación de datos de vacunas (VAERSVAX).

La base de datos contenía 445358 observaciones y 7 variables.

- A. En VAX\_MANU se eliminaron las observaciones con 'UNKNOWN MANUFACTURER' (fabricante desconocido).
- B. VAX\_TYPE para filtrar exclusivamente las vacunas de COVID19.
- C. Una vez filtrado se eliminaron las siguientes variables: VAX\_NAME (por tener la misma información de VAX\_MANU) y VAX\_TYPE (por que después del filtrado no aporta más información)

Luego del filtrado se obtuvieron 409006 observaciones y 5 variables.

### 3.2. Combinación de los datasets VAERVAX (vacuna) y VAERSData (data).

La combinación fue basada en el ID único (VAERSID). La base de datos contenía 409006 observaciones y 39 variables.

Se eliminaron variables que no se consideraban relevantes para el objetivo de la investigación:

- RPT\_DATE (Fecha de llenado del reporte, no es necesaria para este estudio ya que preferimos la fecha de carga del reporte), SPLTTYPE (Número de proyecto de fabricante, no es necesaria para este estudio) y TODAYS\_DATE (Fecha de subida a la base de datos, no es indispensable).

Por la colección de datos en la variable FORM\_VERS, se observó la presencia de 2 formularios (descritos previamente) de los cuales pudimos obtener la siguiente información.

- **VAERS1** - Impreso (289 observaciones)
- **VAERS2** - Online (408717 observaciones)

La versión impresa tiene una cantidad no significativa en comparación de los otros datos, lo que nos permite no considerarla como parte de este estudio.

En consecuencia de la remoción de VAERS1 (Formularios 1), existen otras variables que solo fueron recolectadas con este formulario con este formato, de las cuales se analizó lo siguiente:

- V\_FUNDBY (Tipos de financiamiento usado para la vacunación). Se eliminó por poseer data insuficiente.
- VAX\_DOSE\_SERIES -- Número de dosis administradas. Esta variable no es confiable debido a su discontinuación en el formulario. A pesar de eso se realizó una limpieza a través de la remoción de observaciones (fila entera) donde se observó más de 3 dosis de vacunas.

Importante notar que en estos casos excedentes, se observaron por fabricante de vacuna:

1. MODERNA (644 observaciones con excedentes de 3 dosis)

2. PFIZER-BIONTECH (585 observaciones con excedentes de 3 dosis)
3. JANSSEN (85 observaciones con excedentes de 3 dosis)

Además, se reportó 11 fallecidos (tanto para Moderna y Pfizer) y 17 hospitalizados por todas las vacunas. Una vez usado el filtro, se conservó la variables para separar por grupos de dosis de vacunas.

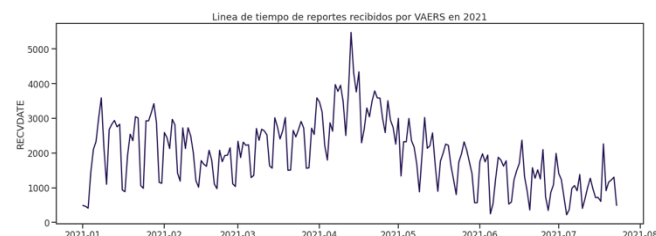
- HISTORY (Condiciones crónicas y enfermedades duraderas reportadas) Al momento de analizar la información se vio una cantidad de datos considerable teniendo un valor significativo en el dataset, por lo que no fue eliminada y fue procesada por un algoritmo de Lenguaje Natural (NLP, descrito más adelante).
- ER\_VISIT (Visita a sala de emergencia o doctor). Debido a la similitud en contenido con otras variables fue combinada (ver más adelante).

Luego de este filtrado se **obtuvieron 408268 observaciones y 34 variables.**

### 3.3. Filtrado por variables.

La base de datos contenía **408268 observaciones y 34 variables.**

1. RECVDATE (fecha de reporte recibido).



2. STATE. Se conservaron solo estados de USA, otros territorios y datos perdidos fueron cambiados por NSP.
3. AGE\_YRS, CAGE\_YR, CAGE\_MO. (Edad, Edad y meses aparentes)

Se combinó CAGE\_YR (edad en años estimada) y CAGE\_MO (sumatorio de edad estimada) para recuperar la edad en caso no haya sido reportada en AGE\_YRS. Se convirtieron las edades en números integrales y todo se estableció en la variable AGE\_YRS. Por este filtro se pudieron recuperar 1466 edades no anotadas.

4. SEX. Se eliminaron los record con sexo desconocido (U). Luego de aplicar este filtro se obtuvieron 398229 observaciones y 32 variables con 285023 (Femenino) y 113206 (Masculino).
5. DIED, DATE\_DIED. Se confirmó la relación entre el reporte de muerte y la fecha de muerte y se mejoró datos erróneos.
6. ER\_VISIT, ER\_ED\_VISIT, OFC\_VISIT. (Todas relacionadas a visitar al médico o a una

sala de emergencias). En el caso de las dos primeras variables en mención tienen la misma información pero fueron colectadas en diferentes formularios. Después de este filtrado se obtuvieron 398229 observaciones y 30 variables.

7. HOSPITAL, HOSPDAYS. Se imputó con 0 los valores faltantes y se cambió Y por 1 en HOSPITAL, y considerando que “Todo paciente hospitalizado (Y) debe al menos haber pasado 1 día de hospitalización.” Se realizaron las correcciones a HOSPDAYS. El límite de los días de hospitalización fue de hasta 120 días.
8. L\_THREAT. Se corrigieron los datos faltantes y se le dio un trato de variable binaria.
9. DISABLE. Idem a L\_THREAT.
10. BIRTH\_DEFECT, X\_STAY. Ambas fueron eliminadas, debido a que no representan un valor significativo a toda la base de datos (solo 199 pacientes en BIRTH DEFECT y 235 en X\_STAY), además no especifica el defecto de nacimiento por lo que no es indispensable para este estudio, ni la cantidad de días de hospitalización excedente. Después de este filtrado se obtuvieron 398229 observaciones y 28 variables.
11. RECOVD. Se imputaron los datos faltantes con valores U (desconocidos) de periodos de recuperación de los pacientes.
12. VAX\_DATE, VAX\_ONSET. (Fecha de vacunación, Fecha de evento adverso) Se corrigieron múltiples errores, corregidos y mapeados desde la información de fecha de reporte del formulario.

14. ADMINBY. (Establecimiento donde se inoculó la vacuna). Completa y conservada.
15. VAX\_LOT, VAX\_ROUTE, VAX\_SITE. (Lote de vacuna, Ruta de aplicación de la vacuna y sitio anatómico de inoculación de la vacuna) Se analizaron los datos perdidos y se completó la información.

Después del análisis y filtrado se obtuvieron **398229 observaciones y 27 variables**.

#### 4. Selección de variables y pre-procesamiento de datos.

##### 4.1. Corte de datos en base a los objetivos de estudio.

Se realizaron los siguientes cortes para análisis final de los datos.

1. De la variable VAX\_MANU se filtraron solo variables de la vacuna “PFIZER”.
2. De la variable HOSPITAL se filtraron solo los hospitalizados (1).

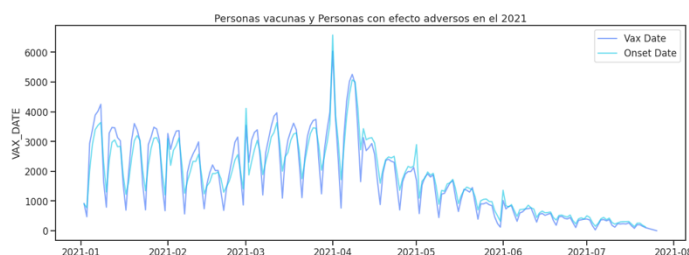
Después de este corte, se eliminaron las variables en mención y además se eliminó DATEDIED por insuficiencia de datos.

Después del análisis y filtrado se obtuvieron **10434 observaciones y 27 variables**.

##### 4.2. Uso de Lenguaje Natural para registro de eventos AESI.

El conjunto de datos final se obtuvo aplicando un algoritmo de procesamiento del lenguaje natural para encontrar una coincidencia aproximada de cadenas de texto (string matching), haciendo uso de la biblioteca fuzzywuzzy, usando el método `fuz_ratio > 85`.

Se combinó todas las variables que reciben información textual ('HISTORY',



**El gráfico muestra las fechas de vacunación y la presencia de efectos adversos.**



**El gráfico muestra las 4 fechas reportadas por las bases de datos en estudio, se puede observar una cantidad constante de muertes en todo el periodo de estudio.**

13. NUMDAYS. (Número de días desde el día del efecto adverso) Eliminada, debido a la cantidad excesiva de datos erróneos (desde 1 hasta 36896).

'SYMPTOM\_TEXT', 'LAB\_DATA', 'OTHER\_MEDS', 'CUR\_ILL', 'ALLERGIES', 'PRIOR\_VAX') en la variable 'INFO' y se utilizó la biblioteca de procesamiento de lenguaje natural fuzzywuzzy (versión 0.18.0) para detectar patrones de coincidencia de texto (string matching) usando un diccionario diseñado para detectar evento AESI, este diccionario basado en sinónimos de términos MedDRA, se observa en el dataset bajo el nombre de las variables: 'infarct\_myocardic', 'anaphylactic', 'coagulopathy', 'covid19', 'guillain', 'miller', 'polyneuropathy', 'kawasaki', 'multisystem', 'myocarditis\_pericarditis', 'narcolepsia', 'pregnancy', 'epilepsy\_consulsions', 'stroke', 'transverse myelitis', 'adrs', 'autoimmuned', 'adem', 'multiplesclerosis', 'opticneuritis', 'cidp', 'encephalitis', 'myelitis', 'encephalomyelitis', 'meningoencefalitis', 'meningitis', 'encephalopathy', 'ataxia', 'noanaphylactic'.

Luego de la verificación (mediante la sumatoria en una variable 'DECISION') se eliminaron toda observación que no

presentaba ningún evento AESI. Una vez filtrada la información se eliminó también la variable 'INFO'.

Finalmente se obtuvieron **10434 observaciones y 27 variables**.

## 5. Modelamiento con algoritmos de Aprendizaje Automatizado no supervisado

Los algoritmos de agrupación en clústeres se utilizan comúnmente para la estructura y categorización de datos, así como para la compresión de datos y la creación de modelos. Por otro lado el algoritmo de PCA se utilizó para reducir el número de variables de forma que pasemos a tener el mínimo número de nuevas variables y que representen a todas las antiguas variables de la forma más representativa posible.

### 5.1 Algoritmo K-Means.

El algoritmo k-means es un proceso de partición claro. Separa (N) objetos de datos en (K) conjuntos de grupos para obtener una baja similitud cruzada y una alta especificidad entre grupos. El diámetro medio de los puntos de un grupo, también conocido como centroide del grupo, se utiliza para evaluar la similitud del grupo.

Primero se eligió k puntos al azar como la media del grupo (centro). Después de eso, a todas las variables se asignaron "k" grupos con la medida de distancia más corta entre sus centroides y objetos. La media se revisó antes de asignarla a todos las variables. Esta actualización continúa hasta que la asignación sea segura.

A partir de la librería "Scikit Learn", se importó el modelo K-Means para la identificación de clústeres de riesgo frente a factores AESI. Se utilizaron los parámetros por default para realizar la identificación de grupos clústeres. Para evaluar los clústeres encontrados por el modelo K-Means, se aplicó adicionalmente Hierarchical clustering y se hizo un Principal Component Analysis (PCA) para corroborar las agrupaciones de los perfiles identificados.

### 5.2. Algoritmo para PCA.

Es un método matemático que se utiliza para reducir el número de variables de forma que pasemos a tener el mínimo número de nuevas variables y que representen a todas las antiguas variables de la forma más representativa posible. Es decir, reducimos el número de variables a dos o tres nuevas, representando los datos originales en el plano o en un gráfico de 3 dimensiones y, así, se visualiza de forma gráfica un resumen de nuestros datos.

Para aplicar este análisis se consideraron las variables "días de hospitalización", "edad", y "días desde la vacunación hasta el inicio de síntomas".

## III. RESULTADOS

A partir de la librería "Scikit Learn", se importó el modelo K-Means para la identificación de clústeres de riesgo frente a factores AESI. Al aplicar el método del codo para visualizar la cantidad de clústeres que podían hallarse en el dataset, para un K-Means con los siguientes parámetros: **n\_clusters=k**, **max\_iter=300** y **random\_state=42**. La curva que se obtuvo se visualiza en **Figura 3**. Esta nos indica que el número óptimo de clústeres a seleccionar para nuestro dataset es de 3.

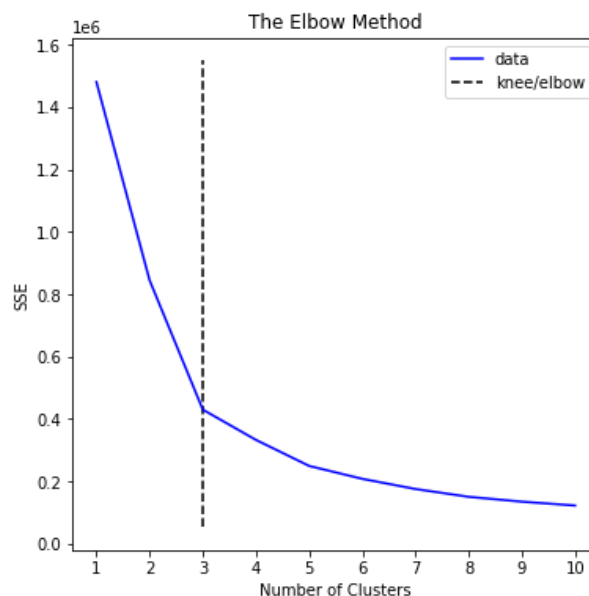


Fig. 3 Método del codo para optimización de número de clústeres

Seguidamente, se realizó la identificación de clústeres utilizando como variables constantes la edad y los días de hospitalización, variando para cada caso el término AESI correspondiente, e.g. en la **Figura 4**.

age\_yrs vs vax\_dose\_series vs hospsdays

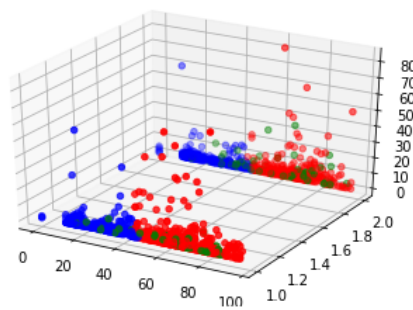
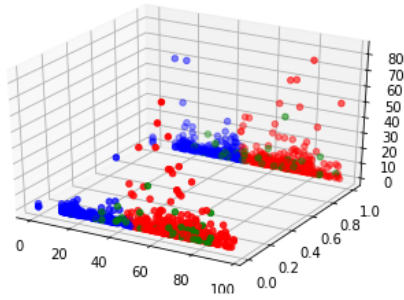


Fig. 4. Clústers identificados al utilizar las variables edad, N° de dosis, y días de hospitalización

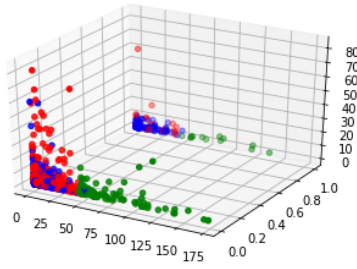


age\_yrs vs sex vs hospdays



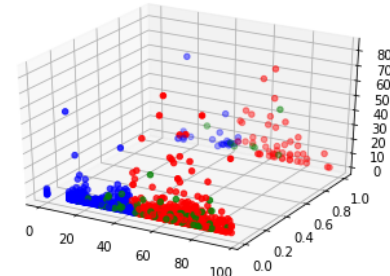
**Fig. 5. Agrupaciones formadas al considerar edad, sexo, y días de hospitalización**

days\_to\_onset\_symptoms vs myocarditis\_pericarditis vs hospdays



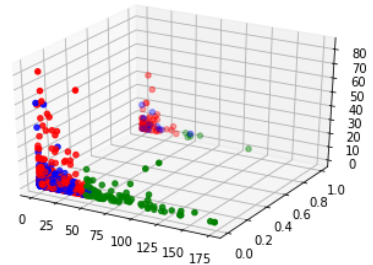
**Fig. 6. Agrupaciones formadas al considerar días desde inicio de síntomas, miocarditis/pericarditis y días de hospitalización.**

age\_yrs vs guillain vs hospdays



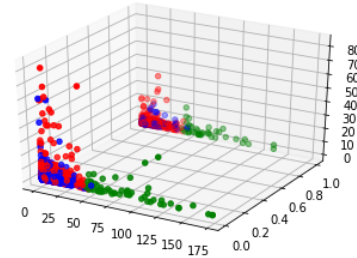
**Fig. 7. Agrupaciones formadas al considerar edad, guillian barré, y días de hospitalización**

days\_to\_onset\_symptoms vs stroke vs hospdays



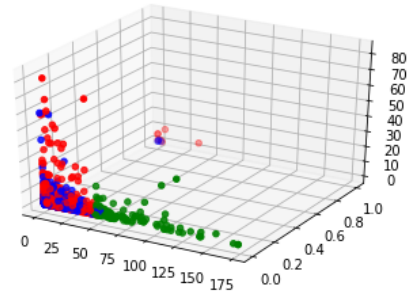
**Fig.8. Agrupaciones formadas al considerar días desde inicio de síntomas, accidente cerebro vascular y días de hospitalización.**

days\_to\_onset\_symptoms vs coagulopathy vs hospdays

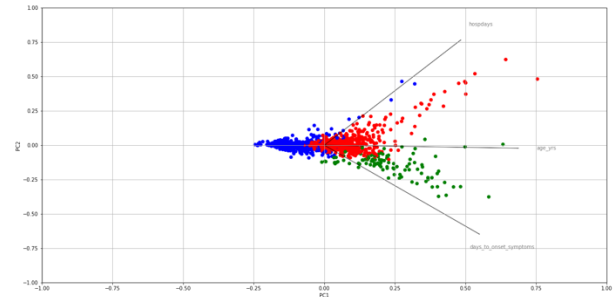


**Fig.9. Agrupaciones formadas al considerar días desde inicio de síntomas, coagulopatías y días de hospitalización.**

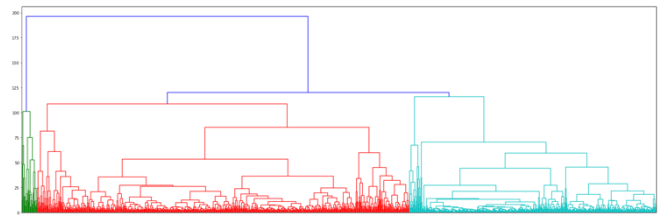
days\_to\_onset\_symptoms vs anaphylactic vs hospdays



**Fig.10. Agrupaciones formadas al considerar días desde inicio de síntomas, reacciones alérgicas y días de hospitalización.**



**Fig. 11. Componentes principales “Días de hospitalización”, “edad”, y “Días desde inicio de síntomas”.**



**Fig. 12. Dendrograma por Agrupamiento Jerárquico**

## IV. CONCLUSIONES

Tras evaluar los resultados obtenidos a partir de la aplicación de un modelo de aprendizaje no supervisado, llegamos a las siguientes conclusiones:

1. Se obtuvo un dataset de efectos adversos reportados tras la aplicación de la vacuna Pfizer de pacientes hospitalizados, el cual fue extraído de la plataforma VAERS del CDC de los Estados Unidos. Dicho conjunto de datos contaba inicialmente con aproximadamente 400 000 datos.
2. Se ensambló un dataset consolidado para procesamiento, el cuál incluía 22 variables de términos AESI, además de información relacionada al

paciente. Se aplicó un algoritmo de procesamiento de lenguaje natural “fuzzywuzzy” a la columna de descripción de eventos adversos de los pacientes hospitalizados post-vacunación. Este dataset final tuvo 1373 observaciones.

3. Los modelos aplicados son capaces de identificar 3 clústeres definidos respecto a la edad y tiempo de inicio de eventos adversos post-vacunación, en relación a los días de hospitalización experimentados. Al aplicar el modelo de aprendizaje no supervisado K-Means, el cluster más relevante fue aquel en que los pacientes registraban edades mayor a los 50 años e inicio de eventos adversos menor a los 50 días, ya que registran los mayores tiempos de hospitalización. La formación de los clústeres, fue comprobada también mediante agrupamiento jerárquico y PCA (Análisis de componentes principales).

## V. RECOMENDACIONES

- Se recomienda realizar una aplicación similar a datasets locales para conocer si la identificación de perfiles varía según la localidad del proceso de inmunización.
- Se recomienda incluir otros signos además de términos AESI para encontrar su asociación con algún target como días de hospitalización o muerte.

- Incluir en la muestra casos que no mostraron efectos secundarios permitiría realizar predicciones hacia objetivos como la personalización de la vacuna para rangos etarios con comorbilidades pre-establecidas.
- Se recomienda la aplicación de modelos de aprendizaje automático para predecir el sexo y/o rango de edad basado en los síntomas presentados post aplicación de las vacunas.

## AGRADECIMIENTOS

Agradecemos principalmente a FONDECYT y la Universidad de Ingeniería y Tecnología por el otorgamiento de la beca que nos permitió el desarrollo de este proyecto. También un agradecimiento a nuestro asesor el Dr. Furukawa, por el asesoramiento y apoyo.

## REFERENCIAS

- [1] Alimadadi, A., Aryal, S., Manandhar, I., Munroe, P. B., Joe, B., & Cheng, X. (2020). Artificial intelligence and machine learning to fight COVID-19.
- [2] La OMS caracteriza a COVID-19 como una pandemia - OPS/OMS | Organización Panamericana de la Salud [Internet]. [citado 31 de mayo de 2021]. Disponible en: <https://www.paho.org/es/noticias/11-3-2020-oms-caracteriza-covid-19-como-pandemia>
- [3] Mottaqi, M. S., Mohammadipanah, F., & Sajedi, H. (2021). Contribution of machine learning approaches in response to Sars-CoV-2 infection. *Informatics in Medicine Unlocked*, 100526.



## ANEXOS

*Tabla. 1 Lista de AESI y términos de búsqueda*

AESI	Códigos MedDRA o Búsqueda de texto
Acute Myocardial Infarction	Myocardial Acute myocardial infarction Myocardial infarction Silent myocardial infarction Acute myocardial injury Acute myocardial ischemia
Anaphylaxis	Anaphylactic Anaphylactic reaction Anaphylactic shock Anaphylactoid reaction Anaphylactic shock
Coagulopathy	Coagulopathy Acquired amegakaryocytic thrombocytopenia Amegakaryocytic thrombocytopenia Axillary vein thrombosis Cavernous sinus thrombosis Cerebral venous thrombosis Deep vein thrombosis Disseminated intravascular coagulation Embolism venous Hepatic vein thrombosis Immune thrombocytopenia Intracranial venous sinus thrombosis Mesenteric vein thrombosis Portal vein thrombosis Pulmonary embolism Pulmonary thrombosis Pulmonary venous thrombosis Severe fever with thrombocytopenia syndrome Subclavian vein thrombosis Thrombocytopenia Thrombocytopenic purpura Thrombotic thrombocytopenic purpura Thrombosis Transverse sinus thrombosis Vena cava embolism Vena cava thrombosis Venous thrombosis
COVID-19 disease	COVID-19 Asymptomatic COVID-19 COVID-19 pneumonia
Death	Death
GBS	Guillian Guillain-Barre syndrome
Kawasaki's disease	kawasaki Kawasaki disease
Multisystem Inflammatory Syndrome in Children (MIS-C)	Multisystem Multisystem inflammatory syndrome in children

Multisystem Inflammatory Syndrome in Adults (MIS-A)	Multisystem Inflammatory Syndrome in Adults
Myocarditis / Pericarditis	Atypical mycobacterium pericarditis Autoimmune myocarditis Autoimmune pericarditis Bacterial pericarditis Coxsackie myocarditis Coxsackie pericarditis Cytomegalovirus myocarditis Cytomegalovirus pericarditis Enterovirus myocarditis Eosinophilic myocarditis Hypersensitivity myocarditis Immune mediated myocarditis Myocarditis Myocarditis bacterial Myocarditis helminthic Myocarditis infectious Myocarditis meningococcal Myocarditis mycotic Myocarditis post infection Myocarditis septic Viral myocarditis Pericarditis Pericarditis adhesive Pericarditis constrictive Pericarditis helminthic Pericarditis infective Pericarditis mycoplasma Pleuropericarditis Purulent pericarditis Viral pericarditis
Narcolepsy/ Cataplexy	Narcolepsy Cataplexy
Pregnancy and Prespecified Conditions	Abortion Aborted pregnancy Abortion complete Abortion early Abortion incomplete Abortion late Abortion missed Abortion spontaneous Abortion spontaneous complete Abortion spontaneous incomplete Abortion threatened Congenital anomaly Drug exposure during pregnancy Exposure during pregnancy Foetal death Maternal Exposure during pregnancy Stillbirth
Seizure/Convulsion	Atonic seizures Atypical benign partial epilepsy Autonomic seizure Clonic convulsion Complex partial seizures Convulsion in childhood Convulsion Convulsions local Epilepsy Epileptic encephalopathy Febrile convulsion

	Febrile infection-related epilepsy syndrome Generalised non-convulsive epilepsy Generalised onset non-motor seizure Generalised tonic-clonic seizure Grand mal convulsion Idiopathic generalised epilepsy Myoclonic epilepsy Neonatal seizure Partial seizures with secondary generalisation Partial seizures Petit mal epilepsy Seizure anoxic Seizure cluster Seizure like phenomena Seizure Simple partial seizures Status epilepticus Temporal lobe epilepsy Tonic clonic movements Tonic convulsion Tonic posturing
Stroke	Basal ganglia stroke Brain stem stroke Cerebellar stroke Cerebrovascular accident Embolic stroke Haemorrhagic stroke Haemorrhagic transformation stroke Ischaemic stroke Lacunar stroke Perinatal stroke Spinal stroke Thrombotic stroke Vertebrobasilar stroke
Transverse myelitis	Transverse myelitis

Fig. 13. Procesamiento de Lenguaje Natural

