



Reto Hackathon **en INTELIGENCIA ARTIFICIAL**



Articulación de la ECCL
Estrategias Inmersivas FP
Bootcamp y Hackathon

Contexto del Reto:

En el sistema de salud de Colombia, las Instituciones Prestadoras de Servicios de Salud (IPS) desempeñan un papel fundamental en la administración de vacunas a los pacientes dentro del régimen de salud. Su función principal incluye la prestación directa de servicios de salud y la ejecución de campañas de vacunación. Las IPS son responsables de aplicar las vacunas a los pacientes, tanto en atención preventiva como en situaciones específicas como pandemias, asegurando que se siga el Programa Ampliado de Inmunización (PAI), el plan nacional de vacunación en Colombia.

Para cumplir con su misión, las IPS mantienen un registro detallado de los pacientes, incluyendo el historial de vacunación. Esto les permite seguir los esquemas de vacunación adecuados, programar citas para dosis subsecuentes o refuerzos, y hacer seguimiento a pacientes con condiciones especiales como niños, adultos mayores o personas con enfermedades crónicas. Además, deben reportar a las Entidades Administradoras de Planes de Beneficios (EPS) los servicios prestados, incluyendo la administración de vacunas. Estas entidades garantizan la cobertura y el reembolso de los servicios dentro del sistema de salud, ya sea en el régimen contributivo o subsidiado.

En el contexto de este reto, la Institución Prestadora de Servicios de Salud, identificada como Vacunación Total para mantener su anonimato, enfrenta serias dificultades en la gestión de la información de sus pacientes. La misión de Vacunación Total es garantizar que al 2026 se complete el esquema total de vacunación para el 90% de la población. Actualmente, la institución no puede identificar con precisión cuáles pacientes han completado su esquema de vacunación ni cuándo deben ser citados para nuevas dosis. Además, carece de un sistema eficiente para distinguir a los pacientes con condiciones especiales que requieren un seguimiento más detallado y personalizado.

Vacunación Total también necesita mejorar la claridad y precisión de sus reportes a las Entidades Administradoras de Planes de Beneficios (EPS), especialmente en lo relativo a los servicios prestados y la administración de vacunas. El reto busca aplicar herramientas de analítica de datos e inteligencia artificial para resolver estos problemas, optimizando la gestión de la información y la calidad de los servicios de salud. Los participantes deben comprometerse a manejar toda la información relacionada con los pacientes con estricta confidencialidad, conforme a la Ley 1581 de 2012 sobre protección de datos personales y la Ley 23 de 1981 que regula el derecho a la confidencialidad de la información clínica. Cualquier divulgación no autorizada de esta información será sancionada conforme a la normativa vigente.

El esquema de vacunación se puede observar a continuación:



Reto Hackathon

en INTELIGENCIA ARTIFICIAL



Articulación de la ECCL
Estrategias Inmersivas FP
Bootcamp y Hackathon

Edad	Vacuna	Dosis	Enfermedad Que Previene
Recién nacido	Antituberculosa (BCG)	Única	Meningitis tuberculosa
	Hepatitis B (Hep B)	Recién Nacido	Hepatitis B
A partir de los 2 meses	Pentavalente	Difteria - tos ferina - tétanos (DPT)	Difteria - tos ferina - tétanos
		Haemophilus influenzae tipo B	Meningitis y otras enfermedades causadas por Haemophilus influenzae tipo b
		Hepatitis B (Hep B)	Hepatitis B
	Rotavirus		Diarrea por rotavirus
	Polio		Poliomielitis
	Neumococo		Neumonías y Meningitis
A partir de los 4 meses	Pentavalente	Difteria - tos ferina - tétanos (DPT)	Difteria - tos ferina - tétanos
		Haemophilus influenzae tipo B	Meningitis y otras enfermedades causadas por Haemophilus influenzae tipo b
		Hepatitis B (Hep B)	Hepatitis B
	Rotavirus		Diarrea por rotavirus
	Polio		Poliomielitis
	Neumococo		Neumonías y Meningitis
A partir de los 6 meses	Pentavalente	Difteria - tos ferina - tétanos (DPT)	Difteria - tos ferina - tétanos
		Haemophilus influenzae tipo B	Meningitis y otras enfermedades causadas por Haemophilus influenzae tipo b
		Hepatitis B (Hep B)	Hepatitis B
	Polio		Poliomielitis
A los 7 meses	Influenza Estacional		Enfermedad respiratoria causada por el virus de la Influenza
	Influenza Estacional		Enfermedad respiratoria causada por el virus de la Influenza
A los 12 meses	Sarampión Rubéola Paperas (SRP)		Sarampión, rubéola y paperas
	Varicela		Varicela
	Neumococo		Neumonías y Meningitis
	Hepatitis A		Hepatitis A
A los 18 meses	Pentavalente	Difteria - tos ferina - tétanos (DPT)	Difteria - tos ferina - tétanos
		Haemophilus influenzae tipo B	Meningitis y otras enfermedades causadas por Haemophilus influenzae tipo b
		Hepatitis B (Hep B)	Hepatitis B
	Polio		Poliomielitis
	Fiebre amarilla (FA)		Fiebre amarilla
	Sarampión Rubéola Paperas (SRP)		Sarampión, rubéola y paperas
A los 5 años	Difteria - tos ferina - tétanos (DPT)		Difteria - tos ferina - tétanos
	Polio		Poliomielitis
	Varicela		Varicela
	Varicela		Varicela
Niños de 9 años	Vacuna contra el Virus del Papiloma humano (VPH)	Única	Virus del Papiloma humano (VPH)
Niñas y mujeres de 9 a 17 años	Vacuna contra el Virus del Papiloma humano (VPH)	Única	Virus del Papiloma humano (VPH)
Mujeres en edad fértil (MEF) entre los 10 y 49 años	Toxide tetánico difterico del adulto (Td)	5 dosis Td1: dosis inicial Td2: a los 6 meses de Td1 Td3: a los 6 meses de Td2 Td4: al año de la Td3 Td5: al año de la Td4	Difteria, tétanos y tétanos neonatal
Población adulta mayor a 60 años	Influenza estacional	Una cada año para mayores de 60 años	Enfermedad respiratoria causada por el virus de la Influenza
Mujeres gestantes	Influenza estacional	Única a partir de la semana 14	Enfermedad respiratoria causada por el virus de la Influenza
	Tdap (Tétanos, Difteria, Tostferina Acelular)	Única a partir de la semana 26	Difteria - tos ferina - tétanos y Bordetella pertussis acelular

El Reto:

El reto para los participantes en este hackathon consiste en trabajar con un archivo de datos de 562.269 KB proporcionado por Vacunación Total. Este archivo presenta inconsistencias, duplicados y errores que deben ser corregidos para que la IPS pueda gestionar la información de manera efectiva. Los participantes deberán diseñar e implementar una solución basada en inteligencia artificial que incluya los siguientes pasos:

Carga y Limpieza de Datos: Aplicar técnicas de tokenización y vectorización mediante TF-IDF para la preparación de los datos.



Reto Hackathon

en INTELIGENCIA ARTIFICIAL



Articulación de la ECCL
Estrategias Inmersivas FP
Bootcamp y Hackathon

Normalización: Agrupar los registros utilizando el algoritmo KMeans para corregir y unificar los registros dentro de cada clúster.

Validación: Asegurar que los datos corregidos sean coherentes y limpios, listos para su uso en producción.

Visualización: Crear tableros de visualización en Power BI que permitan a la IPS gestionar la información corregida y tomar decisiones informadas sobre la identificación de pacientes con esquemas de vacunación incompletos y aquellos con condiciones especiales.

Desarrollo de Chatbot: Implementar un chatbot que utilice la base de datos refinada para responder consultas sobre los pacientes y demostrar la efectividad del proceso de corrección y unificación de datos. La validación del chatbot se basará en su capacidad para interactuar de manera precisa y eficiente con la información.

Este reto está diseñado para mejorar la gestión y la calidad de los servicios de salud ofrecidos por Vacunación Total, apoyando la misión de alcanzar un 90% de cumplimiento del esquema total de vacunación para 2026.

Objetivos del Reto:

- **Agrupación de Registros Similares mediante Clusterización:** Utilizar algoritmos de clusterización para identificar y agrupar registros similares, facilitando la detección de duplicados y variaciones en los nombres y apellidos.
- **Vectorización de Nombres y Apellidos:** Emplear técnicas de procesamiento de lenguaje natural (NLP) para convertir los nombres y apellidos en vectores numéricos, permitiendo la comparación y agrupación eficaz de los registros.
- **Unificación y Corrección de Registros:** Desarrollar una metodología que seleccione y unifique las versiones más correctas y coherentes de los nombres y apellidos dentro de cada clúster.

Metodología del Reto:

La metodología empleada durante el hackathon se basa en la Formación Basada en Retos complementada con Design Thinking, siguiendo un enfoque práctico y colaborativo:

1. Preprocesamiento de los Datos:

Carga y Limpieza Inicial de Datos: Utilizar herramientas como pandas para cargar la base de datos, eliminar duplicados exactos, y normalizar el texto.

Tokenización: tomar nombres y apellidos y tokenizarlos para facilitar la comparación.



Reto Hackathon

en INTELIGENCIA ARTIFICIAL



Articulación de la ECCL
Estrategias Inmersivas FP
Bootcamp y Hackathon

2. Vectorización y Representación de Datos:

Vectorización mediante TF-IDF: Convertir los nombres y apellidos en vectores numéricos usando TF-IDF, optimizando la comparación de registros.

Uso de Joblib: Implementar la paralelización para manejar grandes volúmenes de datos eficientemente.

3. Clusterización No Supervisada:

Aplicación de KMeans: Utilizar KMeans para agrupar registros en clústeres basados en similitudes de vectores, ajustando parámetros para optimizar la agrupación.

4. Unificación y Corrección de Registros dentro de Clústeres:

Análisis de Clústeres: Revisar cada clúster para identificar las versiones más comunes de nombres y apellidos.

Unificación de Registros: Desarrollar un algoritmo para unificar los registros dentro de cada clúster, seleccionando las versiones más coherentes.

5. Validación y Exportación de Resultados:

Validación: Realizar una validación de los registros unificados para garantizar la precisión.

Exportación: Guardar los registros unificados en un archivo CSV o en la base de datos de producción, asegurando la consistencia de la base de datos final.

6. Visualización: Crear tableros de visualización en Power BI que permitan a la IPS gestionar la información corregida y tomar decisiones informadas sobre la identificación de pacientes con esquemas de vacunación incompletos y aquellos con condiciones especiales.

7. Desarrollo de Chatbot: Implementar un chatbot que utilice la base de datos refinada para responder consultas sobre los pacientes y demostrar la efectividad del proceso de corrección y unificación de datos. La validación del chatbot se basará en su capacidad para interactuar de manera precisa y eficiente con la información.



Reto Hackathon

en INTELIGENCIA ARTIFICIAL



Articulación de la ECCL
Estrategias Inmersivas FP
Bootcamp y Hackathon

Entregables del hackathon

- **Informe Ejecutivo:**

Descripción del Proceso: Documento detallado que describa cada paso realizado durante el reto, desde el preprocesamiento de los datos hasta la Visualización y desarrollo del chatbot. Incluirá el enfoque metodológico, las técnicas de inteligencia artificial aplicadas, y los resultados obtenidos en cada etapa.

Análisis de Resultados: Explicación de cómo las decisiones tomadas impactaron la calidad de los datos y cómo estas mejoras pueden influir en la toma de decisiones empresariales.

- **Tableros de Visualización en Power BI:**

Presentación de Resultados: Desarrollo de tableros interactivos en Power BI que muestren claramente los resultados obtenidos después de la unificación y corrección de los registros. Los tableros deben visualizar la mejora en la calidad de los datos, destacando la reducción de duplicados y errores.

Sustentación ante Jurados: Los aprendices deberán presentar y defender sus tableros de visualización frente a un equipo de jurados, explicando cómo sus soluciones contribuyen a una mejor gestión de la base de datos y cómo esto impacta en la toma de decisiones.

- **Chatbot Entrenado:**

Desarrollo y Entrenamiento del Chatbot: Creación de un chatbot que interactúe utilizando la base de datos corregida y clusterizada. Este chatbot deberá ser capaz de responder consultas sobre los datos, mostrando la efectividad del proceso de corrección y unificación.

Validación del Chatbot: Demostración de la precisión del chatbot para interactuar con los datos refinados, evidenciando cómo maneja la información corregida y cómo puede ser utilizado en escenarios empresariales reales.

¡Prepárate para un desafío emocionante que fortalecerá tus habilidades en inteligencia artificial y te acercará al dominio de la gestión avanzada de datos!