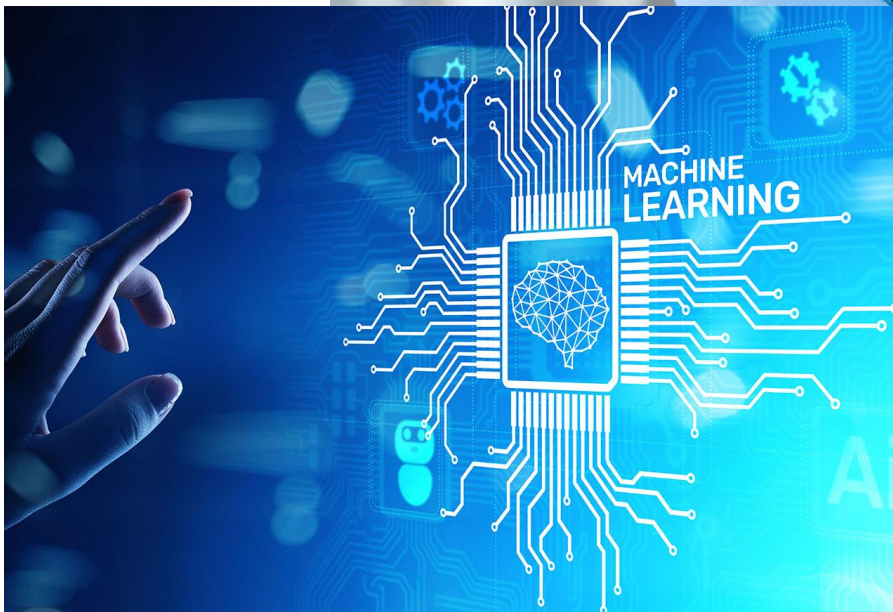


# Estudio de Inteligencia Artificial sobre las Vacunas H1N1 y Seasonal Flu



# Índice

1. Introducción
2. Breve sumario de productos o resultados obtenidos.
3. Contexto y justificación del Trabajo
  - 3.1. Objetivos del Trabajo.
  - 3.2. Enfoque y metodología.
  - 3.3. Planificación del Trabajo.
  - 3.4. Recursos de sistemas utilizados.
4. Creación del set de datos
  - 4.1. Estudio de datos necesarios.
  - 4.2. Fuente de datos.
  - 4.3. Recolección de los datos.
5. Optimización, normalización y calidad del set de datos.
6. Desarrollo del modelo predictivo
  - 6.1. Justificación del modelo seleccionado.
  - 6.2. Descripción del modelo.
7. Entrenamiento y evaluación del modelo predictivo.
8. Puesta en producción (API y aplicación cliente).
9. Conclusiones y mejoras a realizar, modelos alternativos.
10. Bibliografía / Webgrafía.
11. Presentación final

## 1. Introducción

La vacunación es una medida clave de salud pública para combatir las enfermedades infecciosas. Las vacunas proporcionan inmunización para los individuos y reducen la propagación de las enfermedades en la comunidad a través de la “inmunidad colectiva”. En este problema, se trata de predecir si las personas se vacunaron contra la gripe H1N1 y la gripe estacional usando la información que compartieron sobre sus antecedentes, opiniones y comportamientos de salud.

El problema se basa en los datos del National 2009 H1N1 Flu Survey, una encuesta telefónica realizada en Estados Unidos entre finales de 2009 y principios de 2010. La encuesta preguntó a los encuestados si se habían vacunado contra la gripe H1N1 y la gripe estacional, junto con preguntas sobre ellos mismos. Estas preguntas adicionales cubrían su contexto social, económico y demográfico, opiniones sobre los riesgos de enfermedad y la efectividad de la vacuna, y comportamientos para mitigar la transmisión.

El objetivo es comprender mejor cómo estas características se asocian con los patrones de vacunación personales. Este es un problema de aprendizaje supervisado de clasificación binaria, donde se debe predecir dos variables objetivo: si el encuestado recibió la vacuna H1N1 y la vacuna contra la gripe estacional.

Para resolver este problema, se puede utilizar cualquier método o algoritmo de aprendizaje automático. Se debe entrenar el modelo con el conjunto de datos de entrenamiento proporcionado por el sitio web del problema antes de eso hay que hacer una limpieza de los datos y evaluar su rendimiento con el conjunto de datos de prueba. El sitio web también proporciona un formato de envío y una métrica de evaluación (área bajo la curva ROC) para comparar los resultados con otros participantes.

## 2. Breve resumen de productos o resultados obtenidos.

## 3. Contexto y justificación del Trabajo

### 3.1. Objetivos del Trabajo.

Estos son mis objetivos a conseguir en el presente trabajo:

- Predecir si las personas recibieron las vacunas contra la gripe H1N1 y la gripe estacional usando la información que compartieron sobre sus antecedentes, opiniones y comportamientos de salud.
- Comprender mejor cómo estas características se asocian con los patrones de vacunación personales para orientar los futuros esfuerzos de salud pública.
- Aprender y explorar el mundo de las competiciones de ciencia de datos.

### 3.2. Enfoque y metodología.

### 3.3. Planificación del Trabajo.

- Entrega 1: 18/05/2023 (JUE) - Entrega del dataset.
- Entrega 2: 24/05/2023 (MIE) - Entrenamiento y evaluación del modelo predictivo.
- Entrega 3: 30/05/2023 (MAR) - Entrega final.
- Entrega 4: 01/06/2023 (JUE) - Presentaciones.

### 3.4. Recursos de sistemas utilizados.

Para realizar la siguiente tarea solo hemos necesitado los datasets de la pagina de la competición, un ordenador portátil (**RAM** de 16 Gb, **Procesador** - AMD Ryzen 7 5700U with Radeon Graphics - 1.80 GHz, **Tarjeta Gráfica** - AMD Radeon, **Disco duro** - 500 gb, **Sistema Operativo** - Windows 11) y el Visual Studio Code.

## 4. Creación del set de datos

### 4.1. Estudio de datos necesarios.

Los datos necesarios son un conjunto de datos de entrenamiento con 36 columnas, que incluyen el identificador del encuestado, 35 características relacionadas con sus antecedentes, opiniones y comportamientos de salud (*dftrain*), y dos variables objetivo que indican si ha recibido la vacuna contra la gripe H1N1 y la vacuna contra la gripe estacional(*dftrain\_y*). Un conjunto de datos de prueba con las mismas 36 columnas que el conjunto de entrenamiento, excepto las dos variables objetivo que se deben predecir(*dftest*). Un formato de envío que especifica cómo se deben presentar las

predicciones para cada variable objetivo y cada identificador del encuestado(*dfsubmission*).

## 4.2. Fuente de datos.

La fuente de datos en el problema de Flu Shot Learning en el DataDriven es, <https://www.drivendata.org/competitions/66/flu-shot-learning/page/210/>

## 4.3. Recolección de los datos.

La recolección de los datos en el problema de Flu Shot Learning en el DataDriven es, <https://www.drivendata.org/competitions/66/flu-shot-learning/page/210/>

## 5. Optimización, normalización y calidad del set de datos.

La optimización, normalización y calidad del set de datos en el problema de Flu Shot Learning dependen de cómo cada uno procese y prepare los datos para el modelado. Mi estrategia es:

- Optimización: reducir el tamaño de los datos de diversas maneras como puede ser la eliminación de las columnas o filas innecesarias, transformando los tipos de datos.
- Normalización: escalar o estandarizar los valores numéricos para que tengan un rango o una media y una desviación estándar comunes, codificar las variables categóricas.
- Calidad: verificar la consistencia, integridad, precisión y validez de los datos, identificar y corregir los errores, anomalías o valores atípicos, etc.

En mi caso utilicé pandas para limpiar y transformar los datos, eliminé las columnas con más del 40% de valores faltantes, sustituí los valores faltantes usando la media de la variable, y normalizó los datos usando en principio el StandardScaler.

## 6. Desarrollo del modelo predictivo

### 6.1. Justificación del modelo seleccionado.

6.2. Descripción del modelo.

7. Entrenamiento y evaluación del modelo predictivo.

8. Puesta en producción (API y aplicación cliente).

## puedes hacer que te vaya haciendo preguntas de si o no y cuando te haga todas las preguntas te enseñe toda la predicción.

9. Conclusiones y mejoras a realizar, modelos alternativos.

10. Bibliografía / Webgrafía.

<https://www.drivendata.org/competitions/66/flu-shot-learning/data/>

<https://lazypredict.readthedocs.io/en/latest/usage.html>

[https://scikit-learn.org/stable/supervised\\_learning.html#supervised-learning](https://scikit-learn.org/stable/supervised_learning.html#supervised-learning)

11. Presentación final

**GIT :** [https://github.com/Ruben11040/Proyectos\\_Colab/tree/main/Proyecto\\_Final](https://github.com/Ruben11040/Proyectos_Colab/tree/main/Proyecto_Final)