



Universidad
Internacional
de Andalucía

TÍTULO

**APROXIMACIÓN AL USO DE HERRAMIENTAS DE APRENDIZAJE
AUTOMÁTICO PARA LA PREDICCIÓN DE EFECTOS ADVERSOS EN
PACIENTES CON INMUNOTERAPIA**

AUTOR

Adolfo Paradela Carreiro

	Esta edición electrónica ha sido realizada en 2025
Director	Dr. Eduardo Guzmán De los Riscos
Institución	Universidad Internacional de Andalucía
Curso	<i>Diploma de Especialización en Digitalización e Innovación en Salud (2022-23)</i>
©	Adolfo Paradela Carreiro
©	De esta edición: Universidad Internacional de Andalucía
Fecha documento	2023



Universidad
Internacional
de Andalucía



**Atribución-NoComercial-SinDerivadas
4.0 Internacional (CC BY-NC-ND 4.0)**

Para más información:

<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>

<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>

Diploma Universitario en Digitalización e Innovación en salud

Aproximación al uso de herramientas de
aprendizaje automático para la predicción de
efectos adversos en pacientes con inmunoterapia.

Autor: Adolfo Paradela Carreiro

Director: Prof. Eduardo Guzmán De los Riscos

Curso académico: 2022/2023

INDICE

AUTORIZACIÓN DE DIFUSIÓN	3
RESUMEN.....	4
ABSTRACT	6
1. INTRODUCCIÓN	7
1.1 Los sistemas de notificación voluntaria de efectos adversos	7
1.2 Sistema de notificación de eventos adversos (FAERS) de la FDA.....	8
1.3 El aprendizaje automático (“Machine Learning”)	8
1.4 La Inteligencia Artificial explicable	10
1.5 Métodos de minería de datos y aprendizaje automático en la detección de efectos adversos.	10
1.6 Weka y el algoritmo J48	11
1.7 Los Inhibidores de los puntos de control inmunitarios (ICIs).....	12
1.8 Efectos adversos de los ICIs.....	13
1.9 Justificación de la memoria	14
1.10 Objetivos	14
2. METODOS.....	16
3. RESULTADOS	17
3.1 Descarga de los datos desde FAERS	17
3.2 Creación de la BD local	19
3.3 Selección y depuración de los datos correspondientes a los fármacos ICIs	20
3.4 Análisis convencional de la información utilizando herramientas de hoja de cálculo	23
3.5 Análisis del “data set” mediante herramientas de aprendizaje automatizado	27
3.5.1 Análisis Subset1: pt(efecto), sex, prod_ai (clase). Algoritmo J48.	28
3.5.2 Análisis Subset1: pt(efecto), sex, prod_ai (clase). Comparativo con modelo convencional.....	31
3.5.3 Análisis Subset2: pt(efecto), age_gr, prod_ai (clase). Algoritmo J48.....	33
	1

4. DISCUSIÓN	35
4.1 Consideraciones derivadas de los datos de partida.....	35
4.2 Consideraciones derivadas del análisis con herramientas de hoja de cálculo.....	36
4.3 Consideraciones derivadas del uso del algoritmo J48	36
4.4 Limitaciones del estudio.....	37
4.5 Conclusiones.....	37
BIBLIOGRAFÍA	39
ANEXO 1: Resultados J48 sobre subset1	41

AUTORIZACIÓN DE DIFUSIÓN

El abajo firmante, matriculado en el Diploma de Digitalización e Innovación en Salud autoriza a las entidades coordinadoras DUDIS a difundir y utilizar con fines académicos, no comerciales y mencionando expresamente a sus autores el presente Trabajo de Fin de Diploma, realizado en el curso académico 2022/2023 bajo la dirección del Prof. Eduardo Guzmán De los Riscos, con el objeto de incrementar la difusión, uso e impacto del trabajo y garantizar su preservación y acceso a largo plazo.



Adolfo Paradela Carreiro

RESUMEN

La FDA ha puesto a disposición del público una herramienta que permite descargar los datos del sistema de notificación de eventos adversos (FAERS), que recoge información proporcionada por la industria farmacéutica, los profesionales y los pacientes, sobre efectos adversos de medicamentos (EAM).

Se descargó localmente la totalidad de los registros correspondientes a 2022, lo que permitió una aproximación a las técnicas de Machine Learning (ML) partiendo de una amplísima colección de datos, obtenidos en vida real, de los que se seleccionaron los correspondientes a los Inhibidores de los puntos de control inmunitario (ICIs): avelumab, atezolizumab, cemiplimab, durvalumab, nivolumab, pembrolizumab e ipilimumab. Durante el proceso de extracción y depuración de los datos se han constatado las limitaciones intrínsecas a un sistema de notificación voluntaria de efectos adversos.

El conjunto de datos depurado se analizó utilizando herramientas de hoja de cálculo y de aprendizaje automatizado (algoritmo J48 implementado sobre Weka).

El análisis con herramientas propias de una hoja de cálculo (Excel), fundamentalmente tablas dinámicas, es viable para “data set” de hasta un millón de registros, aunque puede ser complejo y lento. En este estudio, fue posible conocer la distribución de las notificaciones de efectos adversos por medicamento y sexo, grupos de edad, etc. También fue posible conocer y ordenar los EAM más notificados y los medicamentos implicados.

Se aplicó también un modelo de aprendizaje automático basado en el algoritmo J48 implementado en el aplicativo Weka. Los resultados, cuya comprensión se facilitó al disponer como referencia del análisis con hoja de cálculo, fueron útiles cuando se buscó la asociación entre efecto adverso, sexo y medicamento, ya que permiten visualizar (y en cierta medida cuantificar), para cada efecto adverso el medicamento más frecuentemente asociado. El modelo tiene un funcionamiento peor cuando se analiza la relación efecto adverso-grupo de edad y medicamento, debido a la menor calidad de los datos de partida.

Esta memoria pretende poner en práctica los conocimientos adquiridos por el alumno a lo largo del curso. Por ello, la extrapolación de los resultados a una potencial utilidad en práctica clínica, está sujeta a múltiples limitaciones, tanto derivadas de los datos de partida como de la idoneidad del algoritmo empleado. No obstante, puede servir de base para estudios posteriores, con algoritmos más sofisticados, que permitan crear un sistema predictivo inteligente del riesgo de desarrollar efectos adversos graves en pacientes a tratamiento con este grupo farmacológico.

Palabras clave: Efectos adversos, Aprendizaje automático, Inmunoterapia, Inteligencia Artificial.

ABSTRACT

The FDA has made available to the public a tool that allows for the downloading of data from the Adverse Event Reporting System (FAERS), which collects information provided by the pharmaceutical industry, professionals, and patients, about adverse drug reactions (ADR).

The entirety of the records corresponding to 2022 was downloaded locally, which allowed for an approach to Machine Learning (ML) techniques starting from an extensive collection of real-world data, from which the ones corresponding to Immune Checkpoint Inhibitors (ICIs) were selected: avelumab, atezolizumab, cemiplimab, durvalumab, nivolumab, pembrolizumab, and ipilimumab. During the process of extraction and data cleaning, the intrinsic limitations of a voluntary adverse effect notification system were noted.

The cleaned dataset was analyzed using spreadsheet and machine learning tools (J48 algorithm implemented on Weka).

The analysis with common spreadsheet tools (Excel), mainly pivot tables, is feasible for datasets of up to a million records, although it can be complex and slow. In this study, it was possible to know the distribution of adverse effect notifications by medication and sex, age groups, etc. It was also possible to know and order the most reported ADR and the drugs involved.

A machine learning model was also applied, based on the J48 algorithm implemented in the Weka application. The results, whose understanding was facilitated by having the spreadsheet analysis as a reference, were useful when looking for the association between adverse effect, sex, and drug, as they allow visualizing (and to some extent quantifying), for each adverse effect, the most frequently associated drug. The model performs worse when analyzing the adverse effect-age group and medication relationship, due to the lower quality of the initial data.

This report aims to put into practice the knowledge acquired by the student throughout the course. Therefore, the extrapolation of the results to a potential utility in clinical practice, is subject to multiple limitations, both derived from the initial data and from the suitability of the algorithm used. Nevertheless, it can serve as a basis for subsequent studies, with more sophisticated algorithms, that allow for the creation of an intelligent predictive system of the risk of developing serious adverse effects in patients being treated with this group of pharmaceuticals.

Keywords: Adverse effects, Machine Learning, Immunotherapy, Artificial Intelligence.

1. INTRODUCCIÓN

1.1 Los sistemas de notificación voluntaria de efectos adversos

Un efecto adverso a medicamentos (EAM) es una respuesta negativa no intencionada provocada por la administración de un fármaco. En los Estados Unidos, las RAM representaron casi el 6 % de todos los pacientes hospitalizados en 2011, con un costo de miles de millones de dólares y generando una morbilidad y mortalidad significativas (1).

La farmacovigilancia implica una serie de métodos y técnicas para la detección, evaluación, comprensión y prevención de los efectos adversos de los medicamentos o cualquier otro problema relacionado con los medicamentos. Incluye distintos métodos: Notificación espontánea o voluntaria, estudios de cohortes, estudios de casos y controles, los propios ensayos clínicos, estudios de utilización de medicamentos, etc. También se incluyen los Métodos de minería de datos y aprendizaje automático, que pueden utilizarse para buscar patrones en grandes conjuntos de datos(2).

Destacan, por su sencillez y amplitud, los sistemas de notificación voluntaria de efectos adversos, también conocidos como sistemas de notificación espontánea o pasiva, son mecanismos implementados en muchos países para recoger y analizar información sobre los posibles efectos adversos asociados con el uso de medicamentos y otros productos sanitarios.

Estos sistemas suelen ser administrados por agencias reguladoras de medicamentos, como la “Administración de Alimentos y Medicamentos” de los Estados Unidos (FDA)(3), la Agencia Europea de Medicamentos (EMA)(4) o la Organización Mundial de la Salud (OMS)(5).

La notificación voluntaria permite que los profesionales de la salud, los pacientes y otros informantes (como los cuidadores y los miembros de la familia) reporten sospechas de reacciones adversas que pueden estar asociadas con un medicamento o producto sanitario. Estos informes pueden ser indicios de posibles relaciones causales entre un medicamento y un evento adverso.

Estos sistemas dependen de la notificación voluntaria de los efectos adversos y, por lo tanto, pueden estar sujetos a subnotificación, es decir, no todos los eventos adversos se reportan. Además, la información recogida a través de estos sistemas no siempre puede establecer una relación causal definitiva entre el producto sanitario y el efecto adverso observado.

A pesar de sus limitaciones, estos sistemas son una herramienta importante para el seguimiento de la seguridad de los medicamentos después de su aprobación y comercialización, y pueden ayudar a identificar problemas de seguridad que no se observaron en los ensayos clínicos previos a la comercialización.

1.2 Sistema de notificación de eventos adversos (FAERS) de la FDA

FAERS, o Sistema de Notificación de Eventos Adversos de la FDA (FDA Adverse Event Reporting System), es una base de datos que contiene información sobre eventos adversos e informes de problemas de medicamentos que son enviados a la FDA(3).

FAERS se utiliza para la vigilancia post-comercialización de medicamentos aprobados por la FDA. Estos informes son voluntarios y pueden ser enviados por profesionales de la salud, consumidores y fabricantes. Los fabricantes están obligados a enviar un informe a FAERS si tienen conocimiento de un evento adverso serio asociado con uno de sus productos.

Los informes a FAERS pueden ayudar a la FDA a identificar nuevos riesgos relacionados con el uso de los medicamentos y a tomar medidas para mejorar la seguridad de los mismos. Sin embargo, un informe en FAERS no indica que el medicamento haya causado necesariamente el evento adverso.

FAERS está adherido a la guía internacional de informes de seguridad emitida por la Conferencia Internacional sobre Armonización (ICH E2B). Además, los eventos adversos y los errores de medicación se codifican utilizando los términos del Diccionario médico para actividades reguladoras (MedDRA)

Además, la FDA proporciona acceso al público a FAERS a través de una herramienta llamada "FAERS Public Dashboard". Esto permite a los usuarios buscar y visualizar datos de FAERS, así como descargar la totalidad, o parte, de la información de la base de datos.

1.3 El aprendizaje automático ("Machine Learning")

Es una rama de la inteligencia artificial (IA) que se basa en la idea de que los sistemas pueden aprender de los datos, identificar patrones y tomar decisiones con mínima intervención humana. Se centra en el desarrollo de programas informáticos que pueden acceder a los datos y utilizarlos para aprender por sí mismos.

En el aprendizaje automático, los algoritmos son entrenados en un conjunto de datos o "datos de entrenamiento". A través de este entrenamiento, los algoritmos aprenden a predecir o clasificar nuevos datos. El aprendizaje puede ser supervisado, no supervisado, o por refuerzo(6).

Aprendizaje supervisado: El algoritmo aprende a predecir resultados a partir de un conjunto de datos etiquetado. Cada ejemplo en el conjunto de datos tiene un valor de salida correspondiente (la "etiqueta") que se desea predecir. Algunos ejemplos de este tipo de algoritmos son las máquinas de vectores de soporte (*Support Vector Machine*, SVM), redes Neuronales, Árboles de Decisión, Bosques Aleatorios, entre otros.

Aprendizaje no supervisado: El algoritmo encuentra patrones en un conjunto de datos sin etiquetas. Algunos ejemplos de este tipo de algoritmos son: *Cluster Analysis* (o Análisis de Conglomerados) y Análisis de Componentes Principales (PCA, en inglés).

Aprendizaje por refuerzo: En este tipo de aprendizaje, un agente de software aprende a realizar acciones en un entorno para maximizar algún tipo de recompensa acumulativa.

Los elementos clave del aprendizaje automático incluyen:

Modelo: Un modelo es un sistema que toma datos como entrada (input) y produce una salida (output). Se construyen utilizando algoritmos.

Algoritmo: Herramienta que se utiliza para descubrir los patrones en los datos que luego se pueden utilizar para crear un modelo. Hay muchos tipos diferentes de algoritmos.

Entrenamiento: Proceso de enseñar a un modelo a hacer predicciones proporcionándole datos. Durante el entrenamiento, el modelo ajusta sus parámetros internos para minimizar el error entre sus predicciones y los verdaderos valores.

Validación y prueba: Una vez que un modelo ha sido entrenado, es importante verificar su rendimiento con datos diferentes a los utilizados durante el entrenamiento.

El aprendizaje automático tiene numerosas aplicaciones en medicina (7),(8) incluyendo ayuda al diagnóstico, predicción de enfermedades, medicina personalizada, etc.:

- Ayuda a la detección de enfermedades y diagnósticos: Los algoritmos de aprendizaje automático pueden ser entrenados para reconocer patrones en imágenes médicas, o para detectar signos de enfermedades.
- Predicción de riesgo de desarrollar enfermedades: Los modelos de aprendizaje automático pueden usar datos de pacientes, como su historial médico, genética y estilo de vida, para predecir el riesgo de desarrollar ciertas enfermedades en el futuro.
- Medicina personalizada e interpretación de datos genómicos: El aprendizaje automático puede ayudar a personalizar los tratamientos para los pacientes en función de sus características individuales.

1.4 La Inteligencia Artificial explicable

La inteligencia artificial explicable (Explainable AI, o XAI, por sus siglas en inglés) se refiere a la aplicación de sistemas de inteligencia artificial que ofrecen explicaciones comprensibles y rastreables de sus decisiones y acciones. Este enfoque está diseñado para aumentar la confianza y la adopción de la IA en el ámbito como el de la atención médica, permitiendo a los profesionales de la salud entender el razonamiento detrás de una predicción o recomendación realizada por un sistema de IA(9).

La necesidad de XAI en medicina surge de la propia complejidad inherente de la atención médica y de la importancia de tomar decisiones clínicas basadas en evidencia y juicio clínico. Las decisiones médicas impactan significativamente en la vida y la salud de los pacientes, por lo que es vital que los médicos y otros profesionales de la salud puedan entender cómo y por qué un sistema de IA ha llegado a una determinada conclusión o recomendación(10),(11).

Además, la XAI también puede facilitar la adopción de la IA en la atención médica al aumentar la confianza de los profesionales de la salud en estos sistemas.

Por otra parte, a nivel regulatorio, el “Reglamento Europeo de Dispositivos Médicos de Diagnóstico In Vitro”(12) incluye explícitamente el software en sus requisitos, lo que tiene especial trascendencia dado que muchos de ellos involucran algoritmos de aprendizaje automático para el análisis de datos y el apoyo a la toma de decisiones. Esto puede aumentar la dificultad de aplicar procedimientos de “machine learning” o “Deep learning”, al precisar de complementos explicativos que permitan a los usuarios comprender la esencia del funcionamiento de los algoritmos utilizados. Una IA confiable debe empoderar a los profesionales biomédicos para asumir la responsabilidad de su toma de decisiones, lo que implica la necesidad de métodos de IA explicables (13).

1.5 Métodos de minería de datos y aprendizaje automático en la detección de efectos adversos.

El aprendizaje automático, ha demostrado ser valioso en la detección de efectos adversos por medicamentos (EAM), ya que sus algoritmos pueden analizar grandes conjuntos de datos, identificar patrones y predecir EAM basándose en múltiples factores, como la edad del paciente, su historial médico y la combinación de medicamentos que está tomando.

A continuación, se presentan algunos ejemplos:

- Minería de informes de EAM: las notificaciones espontáneas de eventos adversos son importantes para la detección de EAM y generan una gran cantidad de información, que puede ser analizada por los algoritmos de aprendizaje automático.
- Minería de textos de registros médicos electrónicos: el aprendizaje automático puede ser utilizado para extraer y analizar esta información para detectar posibles EAM.
- Predicción de EAM: Los algoritmos de aprendizaje automático también pueden ser utilizados para predecir la probabilidad de que un paciente experimente un EAM basándose en sus características individuales y el perfil de medicamentos que está tomando.

En este sentido, Hae Reong Kim y col(2) realizaron una revisión sistemática de la literatura basada en artículos publicados entre 2015 y 2020. Las palabras clave utilizadas fueron métodos estadísticos, de aprendizaje automático y de aprendizaje profundo para detectar señales EAM. Los autores revisaron manualmente 72 artículos, de los cuales 51 y 21 se clasificaron como que abordaban métodos estadísticos y de aprendizaje automático, respectivamente. En los estudios de aprendizaje automático, DrugBank fue la base de datos más utilizada, seguida de las bases de datos EMR, SIDER y FAERS.

1.6 Weka y el algoritmo J48

Weka (*Waikato Environment for Knowledge Analysis*) es un software popular de código abierto desarrollado por la Universidad de Waikato en Nueva Zelanda, que proporciona una colección de algoritmos de aprendizaje automático y herramientas de minería de datos. Se utiliza principalmente para la enseñanza, la investigación y el desarrollo de aplicaciones.

Weka proporciona algoritmos para todas las tareas de minería de datos, incluyendo preparación de datos, clasificación, regresión, *clustering*, asociación de reglas, y selección de atributos. Los algoritmos se pueden aplicar directamente a un conjunto de datos o se pueden llamar desde Java.

Además, proporciona una interfaz gráfica fácil de usar que permite a los usuarios cargar conjuntos de datos, ejecutar algoritmos y visualizar los resultados y las evaluaciones de los modelos. Weka es compatible con casi todos los formatos de datos y también permite transformar y limpiar los datos antes de su análisis(14),(15)

Por otra parte, el algoritmo J48 es una implementación en Java del algoritmo C4.5, ambos utilizados en el campo del aprendizaje automático para generar árboles de decisión. Fue creado por Ross Quinlan(16), quien también desarrolló el algoritmo ID3. Los árboles de

decisión son estructuras que permiten tomar decisiones basadas en ciertas reglas o condiciones.

El algoritmo C4.5, y por extensión el J48, se utiliza para clasificar instancias, es decir, asignar una categoría a una nueva observación basándose en las características de la observación y en el conocimiento adquirido a partir de un conjunto de datos de entrenamiento. Pueden manejar tanto atributos categóricos como continuos. También pueden manejar datos con valores faltantes y son capaces de realizar poda del árbol para evitar el sobreajuste. Se ha utilizado en múltiples disciplinas y situaciones, incluyendo el estudio de los efectos adversos por medicamentos(17),(18)

1.7 Los Inhibidores de los puntos de control inmunitarios (ICIs)

Los inhibidores de puntos de control inmunitario son un grupo de medicamentos que se utiliza en la inmunoterapia del cáncer(19). Interactúan con ciertas proteínas que actúan como "interruptores" de las células T inmunitarias. Estos interruptores, o "puntos de control", normalmente ayudan a las células T a distinguir entre las células propias del cuerpo (a las que no deben atacar) y las células ajenas o dañinas (a las que deben atacar).

Los procesos neoplásicos, sin embargo, pueden "engañar" a estas células T utilizando estos puntos de control. Por ejemplo, algunas células cancerosas pueden producir una proteína (ligando) llamada PD-L1, que se une a otra proteína (receptor) llamada PD-1 en las células T. Cuando PD-1 y PD-L1 se unen, la célula T recibe una señal para "apagarse" y no atacar a la célula que produce PD-L1.

Los inhibidores de los puntos de control inmunitario son anticuerpos monoclonales que se unen a la proteína PD-1 en las células T, bloqueando así la interacción entre PD-1 y PD-L1. Al hacerlo, estos medicamentos permiten que las células T "ignoren" la señal de las células cancerosas y continúen atacándolas.

Otro punto de control inmunitario, llamada CTLA-4, actúa también como interruptor de apagado de células T. Los medicamentos que se dirigen frente a CTLA-4 también pueden ayudar a las células T a ignorar las señales de las células cancerosas y a seguir atacándolas.

Los medicamentos más destacados son los siguientes(20):

- Inhibidores de PD-1: Nivolumab, Pembrolizumab, Cemiplimab
- Inhibidores de PD-L1: Atezolizumab, Durvalumab, Avelumab,
- Inhibidores de CTLA-4: Ipilimumab

Estos medicamentos han demostrado ser muy efectivos para ciertos tipos de cáncer, pero no funcionan para todos los pacientes. Además, estos medicamentos pueden tener efectos secundarios graves, incluyendo reacciones inmunitarias que pueden dañar órganos del cuerpo.

1.8 Efectos adversos de los ICIs

Los efectos adversos a nivel inmunitario debidos a ICIs pueden afectar a diversos niveles como el endocrino, sistema nervioso central, ocular, cardíaco, respiratorio, gastrointestinal, hepatobiliar, piel y tejidos subcutáneo, musculoesquelético, renal o general. En algunos casos esta toxicidad puede llegar a ser grave y conllevar la suspensión del tratamiento (los llamados efectos adversos grado III y IV) y en algunos casos puede causar la muerte del paciente(21).

Wang DY y col(21) analizaron los efectos adversos fatales de pacientes con estos medicamentos en base los efectos adversos reportados en la base de datos de "World and Health organization" y la revisión de las publicaciones en relación a efectos adversos fatales. Los investigadores reportaron un total de 613 efectos adversos fatales. De un total de 193 muertes en pacientes con ICIs tipo anti-CTLA-4, la mayoría fueron debidas a colitis (135 [70%]). Sin embargo, en los 333 pacientes con ICIs anti PD-1/PD-L1-las muertes más frecuentes fueron debidas a pneumonitis (333 [35%]), hepatitis (115 [22%]) y efectos neurotóxicos (50 [15%]).

Un aspecto que no ha sido contemplado en los diferentes ensayos clínicos que llevaron a la aprobación de los diferentes ICIs es la posible diferencia en eficacia y seguridad entre el sexo masculino y femenino. En el meta-análisis de Conforti F(22) con 11.351 pacientes con tumores avanzados se detectaron diferencias de eficacia según sexo, observándose una mayor eficacia de los ICIs en mujeres que en hombres. Sin embargo, en el meta-análisis de Jing Y(23), realizado en base a los estudios publicados y a la base de datos de la "Food and drug administration Adverse Event Reporting System (FAERS)" no encontraron diferencias en cuanto a efectos adversos inmunitarios entre sexos.

Por otra parte, la utilización de la Inteligencia Artificial en salud se considera imprescindible para seguir avanzando en una medicina más personalizada y más segura para los pacientes(24). En este sentido, existen publicaciones que han logrado construir modelos predictivos de efectos adversos a través de bases de datos públicas dónde los pacientes y/o los profesionales pueden notificar efectos adversos, como puede ser la citada base FAERS(25).

1.9 Justificación de la memoria

La FDA ha puesto a disposición del público una herramienta que permite descargar los datos del sistema de notificación de eventos adversos (FAERS), que recoge información proporcionada por la industria farmacéutica, los profesionales y los pacientes, sobre efectos adversos de medicamentos. Ello permite disponer de un conjunto de datos que se aproxima mucho al concepto de “big data”, el cual podría servir para aplicar modelos de inteligencia artificial (IA). En este contexto, el aprendizaje automático (ML=Machine Learning), una rama de la IA que emplea algoritmos para construir modelos capaces de hacer predicciones o tomar decisiones sin haber sido programados específicamente para ello, podría ser una herramienta especialmente útil.

La presente memoria supone una aproximación a las técnicas de ML partiendo de una amplísima colección de datos, obtenidos en vida real, de especial interés para complementar el perfil de seguridad descrito en los ensayos clínicos y para estimar los riesgos específicos en pacientes concretos que podrían recibirlos en el futuro, contribuyendo a mejorar la seguridad de los tratamientos.

Aunque se descargará la totalidad de la información reportada a FAERS durante el 2022, en esta memoria se estudiará un grupo específico de fármacos: los ICIs, por las razones expuestas anteriormente. Además, al ser un conjunto limitado, se podrá hacer un análisis tradicional, basado en el uso de hojas de cálculo, que nos proporcione un escenario frente al que comparar los resultados que nos proporcionen los algoritmos y, de este modo, facilitar su comprensión por parte de un usuario con una experiencia muy limitada.

1.10 Objetivos

- 1.- Obtener de la base de datos FAERS la totalidad de las notificaciones reportadas en 2022 e incorporarlas a una base de datos propia (Microsoft Access) que permita su manejo a nivel local. De este conjunto, extraer los datos relativos a los pacientes a tratamiento con Inhibidores de los puntos de control inmunitario (ICIs): avelumab, atezolizumab, cemiplimab, durvalumab, nivolumab, pembrolizumab e ipilimumab), para su análisis y depuración (“data curation”).
- 2.- Analizar los datos depurados y estructurados correspondientes a los efectos adversos notificados en pacientes a tratamiento con ICIs, mediante herramientas convencionales (hojas de cálculo).

3.- Elaborar un modelo predictivo, usando técnicas de aprendizaje automatizado, a partir de los efectos adversos recopilados en el sistema FAERS de pacientes tratados con ICIs, y comparar sus resultados con los obtenidos a través del objetivo 2.

2. METODOS

Se descargaron las tablas disponibles en el sistema FAERS del año 2022 (utilizando para ello el “FDA Adverse Event Reporting System (FAERS) Quarterly Data Extract Files”(26), en formato ASCII.

Los datos de cada cuatrimestre se importaron a una base de datos relacional (Microsoft Access) para su análisis y explotación. Dado que la aplicación Microsoft Access tiene un límite máximo de 2GB, fue necesario utilizar dos bases de datos independientes, fusionadas posteriormente en una única BD con una tamaño compatible con Access (eliminando el contenido de ciertas tablas especialmente grandes, una vez que se procesó su información).

A partir de las dos bases de datos resultantes, se seleccionarían los datos correspondientes al grupo de fármacos ICIs en los que han sido reportados como “primary suspect drug”, exportándolos a un archivo excel. Sobre este último aplicativo, se hizo una revisión manual, subsanando, en la medida de lo posible, deficiencias en los mismos (data curation).

Sobre la base del archivo Excel depurado se realizó un análisis tradicional de sus datos, empleando fundamentalmente herramientas de las hojas de cálculo, sobre todo tablas dinámicas y filtros.

La hoja de cálculo con los datos depurados fue exportada a un archivo de tipo “csv” para su posterior incorporación al software Weka 3: Machine Learning Software in Java(27). La versión utilizada fue la 3.8.6. Utilizando dicho software, el “data set” fue analizado mediante el algoritmo de clasificación (J48). Los resultados fueron analizados e interpretados comparando con los resultados obtenidos a partir del análisis convencional mediante tablas dinámicas.

3. RESULTADOS

3.1 Descarga de los datos desde FAERS

A partir de FDA Adverse Event Reporting System (FAERS) Quarterly Data Extract Files(26), se realizaron 4 descargas de datos, correspondientes a cada trimestre de 2022.

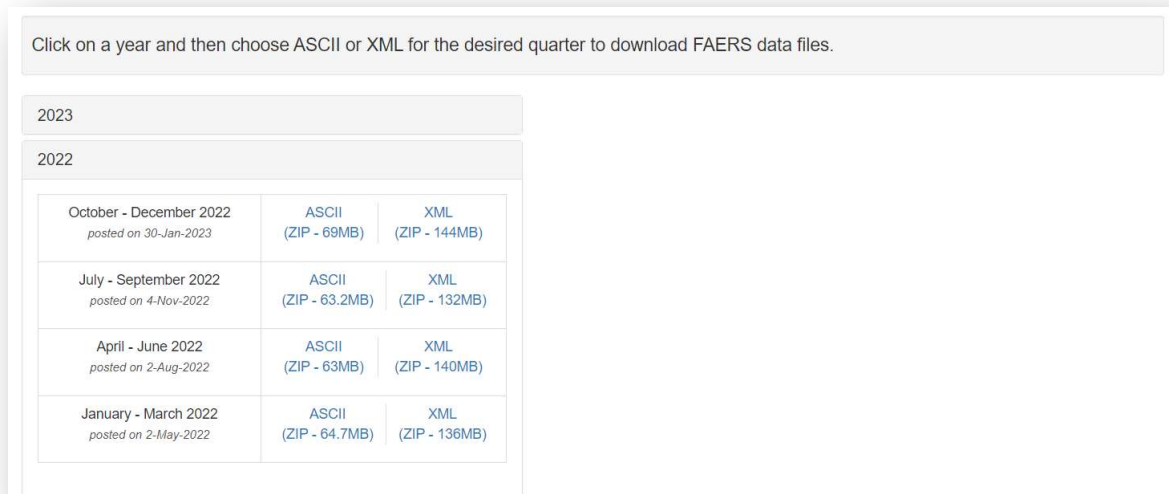


Ilustración 1: Zona de descargas FAERS

Se obtuvieron 4 archivos comprimidos (faers_ascii_2022Q4, faers_ascii_2022Q3, faers_ascii_2022Q2 y faers_ascii_2022Q1) en formato zip. Cada archivo comprimido contiene lo siguiente:

1. ASCII, que contiene los datos ASCII y los archivos informativos.
2. XML, que contiene los datos XML y los archivos informativos.
3. README.DOC, un archivo informativo.
4. Preguntas Frecuentes

También se incluye la estructura de la base de datos FAERS, que es necesario tener en cuenta para conocer el significado de las tablas y sus relaciones.

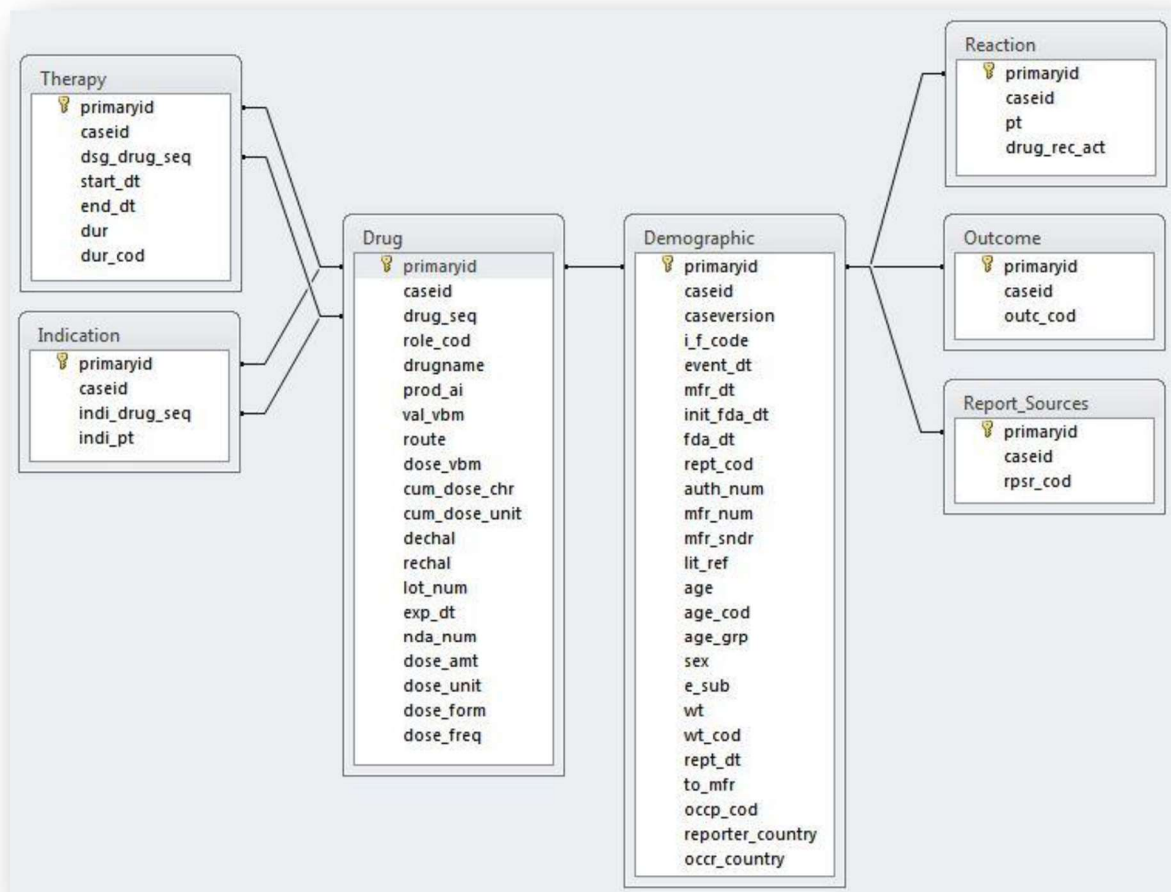


Ilustración 2: Estructura BD FAERS

Los archivos csv (ASCII) se corresponden a las tablas que se indican en la ilustración 2:

1. DEMOyyQq.TXT contiene información demográfica y administrativa del paciente, un registro único para cada reporte de evento.
2. DRUGyyQq.TXT contiene información de medicamentos/biológicos para tantos medicamentos como se informaron para el evento (1 o más por evento).
3. REACyyQq.TXT contiene todos los términos del "Diccionario médico para actividades regulatorias" (MedDRA) codificados para el evento adverso (1 o más). Para obtener más información sobre MedDRA, comuníquese con la mesa de ayuda de MSSO en mssohelp@meddra.org. El sitio web es www.meddra.org.
4. OUTCyyQq.TXT contiene los resultados del paciente para el evento (0 o más).
5. RPSRyyQq.TXT contiene fuentes de informes para el evento (0 o más).
6. THERyyQq.TXT contiene las fechas de inicio y finalización de la terapia con medicamentos para los medicamentos informados (0 o más por medicamento por evento).

7. INDllyQq.TXT contiene todos los términos del "Diccionario médico para actividades regulatorias" (MedDRA) codificados para las indicaciones de uso (diagnósticos) de los medicamentos informados (0 o más por medicamento por evento).

Los campos de las tablas se describen en el documento "ASC_NTS.DOC" FILE FOR THE QUARTERLY DATA EXTRACT (QDE) FROM THE FDA ADVERSE EVENT REPORTING SYSTEM (FAERS). Se puede localizar en la carpeta denominada "ASCII" de cada descarga cuatrimestral, con el nombre: ASC_NTS.pdf

3.2 Creación de la BD local

Los archivos correspondientes a cada tabla se importaron desde Microsoft Access V.2019. Posteriormente se agruparon las tablas correspondientes a los 4 cuatrimestres en archivos únicos con el contenido anual. Para ello se utilizó el "asistente de importación de datos de texto, como en el siguiente ejemplo, aplicado a la tabla DRUGS de un cuatrimestre:

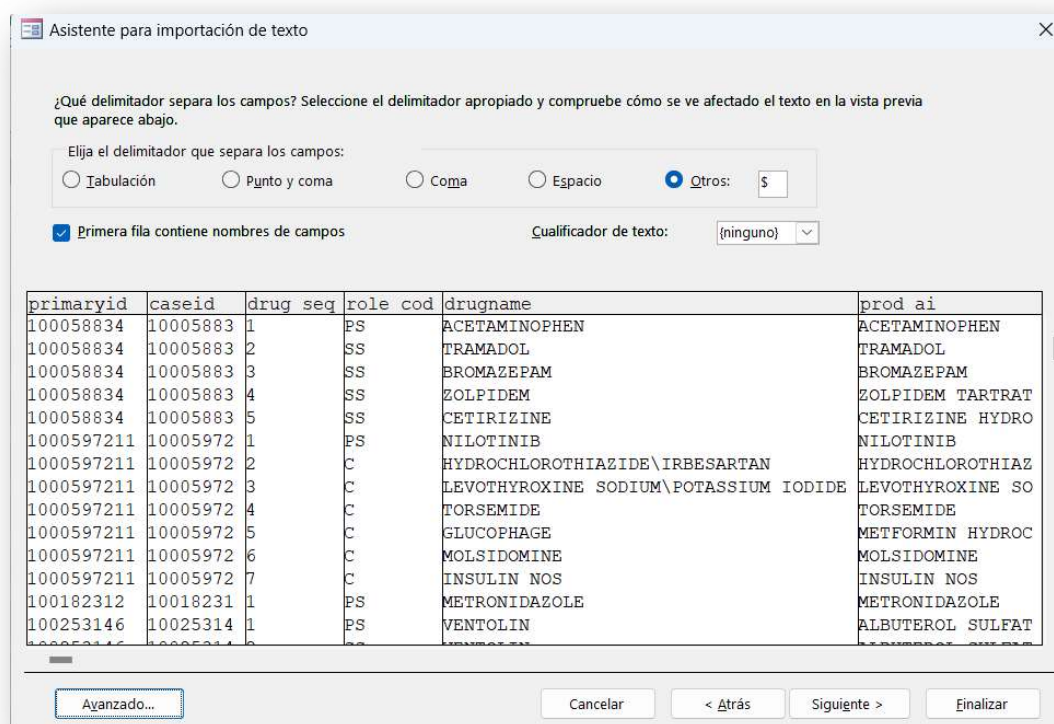


Ilustración 3: Ejemplo del asistente de importación de texto

El total de registros de la tabla DRUG_2022 fue de 7.664.702, que se corresponden con 1.702.222 casos (un mismo caso reportado puede implicar a más de un medicamento). Las relaciones entre las 7 tablas principales se establecieron teniendo en cuenta la información proporcionada por la propia FAERS.

Debido a limitaciones de Microsoft Access, que no permite manejar bases de datos de más de 2GB, fue necesario utilizar dos BD, una denominada “DRUGS_TOTAL” con el contenido de la tabla DRUGS y otra, denominada como “FAERS_LOCAL” con las otras 6 tablas + la tabla DRUGS limitada a los fármacos ICIs, como se describe a continuación.

3.3 Selección y depuración de los datos correspondientes a los fármacos ICIs

A partir de las dos bases de datos resultantes, se seleccionarían los datos correspondientes al grupo de fármacos ICIs en los que han sido reportados como “primary suspect drug”, exportándolos a un archivo Excel, obteniendo un total de 29.260 notificaciones

Para ello, en la BD que contiene la tabla DRUGS se aplicó la siguiente consulta de selección:

```
SELECT  DRUG22_final.primaryid,  DRUG22_final.caseid,  DRUG22_final.drug_seq,  DRUG22_final.role_cod,
DRUG22_final.drugname, DRUG22_final.prod_ai, DRUG22_final.val_vbm, DRUG22_final.route, DRUG22_final.dose_vbm,
DRUG22_final.cum_dose_chr,  DRUG22_final.cum_dose_unit,  DRUG22_final.dechal,  DRUG22_final.rechal,
DRUG22_final.lot_num,      DRUG22_final.exp_dt,      DRUG22_final.nda_num,      DRUG22_final.dose_amt,
DRUG22_final.dose_unit, DRUG22_final.dose_form, DRUG22_final.dose_freq INTO DRUG22_IPC

FROM DRUG22_final

WHERE      (((DRUG22_final.role_cod)="PS")      AND      ((DRUG22_final.prod_ai)="NIVOLUMAB"      Or
(DRUG22_final.prod_ai)="AVELUMAB"      Or      (DRUG22_final.prod_ai)="ATEZOLIZUMAB"      Or
(DRUG22_final.prod_ai)="CEMIPLIMAB"      Or      (DRUG22_final.prod_ai)="DURVALUMAB"      Or
(DRUG22_final.prod_ai)="PEMBROLIZUMAB" Or (DRUG22_final.prod_ai)="IPILIMUMAB");
```

Los datos seleccionados fueron copiados en la base de datos FAERS_LOCAL, en la tabla denominada DRUGS22_IPC, donde se continuó el procesamiento de la información: La tabla DRUGS22_IPC se relacionó con las tablas Demo22_final e Indi22_final, para extraer la información complementaria, incluida la demográfica y de indicaciones, de cada caso. Se usó la consulta de selección:

```
SELECT  DRUG22_IPC.primaryid,  DRUG22_IPC.caseid,  DRUG22_IPC.drug_seq,  DRUG22_IPC.role_cod,
DRUG22_IPC.drugname,      DRUG22_IPC.prod_ai,      DRUG22_IPC.route,      DRUG22_IPC.dose_vbm,
DRUG22_IPC.cum_dose_chr,  DRUG22_IPC.cum_dose_unit,  DRUG22_IPC.dechal,  DRUG22_IPC.rechal,
DRUG22_IPC.dose_amt, DRUG22_IPC.dose_unit, DRUG22_IPC.dose_form, DRUG22_IPC.dose_freq, DEMO22_final.age,
DEMO22_final.age_cod,  DEMO22_final.age_grp,  DEMO22_final.sex,  DEMO22_final.wt,  DEMO22_final.wt_cod,
DEMO22_final.occp_cod, INDI22_final.indi_pt

FROM (DRUG22_IPC LEFT JOIN DEMO22_final ON DRUG22_IPC.primaryid = DEMO22_final.primaryid) LEFT JOIN
INDI22_final ON (DRUG22_IPC.drug_seq = INDI22_final.indi_drug_seq) AND (DRUG22_IPC.primaryid =
INDI22_final.primaryid);
```

A continuación se hizo una nueva consulta, relacionando los resultados anteriores (RESULTADO1) con la tabla de efectos adversos descritos (Reac22_final), obteniendo la tabla RESULTADO2 con 86053 registros (este número es superior al de notificaciones, ya

que en una misma notificación pueden describirse varios efectos adversos. La consulta utilizada fue:

```
SELECT RESULTADO1.primaryid, RESULTADO1.caseid, RESULTADO1.drug_seq, RESULTADO1.role_cod,
RESULTADO1.drugname, RESULTADO1.prod_ai, RESULTADO1.route, RESULTADO1.dose_vbm,
RESULTADO1.cum_dose_chr, RESULTADO1.cum_dose_unit, RESULTADO1.dechal, RESULTADO1.rechal,
RESULTADO1.dose_amt, RESULTADO1.dose_unit, RESULTADO1.dose_form, RESULTADO1.dose_freq,
RESULTADO1.age, RESULTADO1.age_cod, RESULTADO1.age_grp, RESULTADO1.sex, RESULTADO1.wt,
RESULTADO1.wt_cod, RESULTADO1.occp_cod, RESULTADO1.indi_pt, REAC22_final.pt INTO RESULTADO2

FROM RESULTADO1 LEFT JOIN REAC22_final ON RESULTADO1.primaryid = REAC22_final.primaryid;
```

Finalmente, se restringió el estudio solo a aquellas reacciones reportadas como graves, es decir, las correspondientes a los siguientes valores del campo outc_cod de la tabla OUTC_22: DE Death, LT Life-Threatening, HO Hospitalization-Initial or Prolonged y DS Disability. Con ello, se obtuvo un conjunto de datos con 71.672 registros, correspondientes a 24.969 notificaciones.

A continuación se exportaron a un archivo Excel los campos (atributos): “pt” (reacción adversa), “age_grp” (grupo de edad), “sex” (sexo) y la clase “prod_ai” (principio activo).

Utilizando el programa Microsoft Excel, se realizó una revisión de los datos exportados, para detectar deficiencias o datos incompletos, sobre todo en el atributo “sex” y en “age_grp”.

En cuanto al atributo “sex”, existían valores como “?” o “UKN” en el 4% de los registros. En lo que respecta a “age group”, figuraban valores como “?” en el 13,5% de los registros. Aquellos casos donde fue posible completar o deducir la información incompleta (a partir de otros datos), fueron corregidos. Aquellas situaciones donde no se pudo obtener la información, fueron excluidas. Por otra parte, en el atributo “pt” no solo figuran efectos adversos, por lo que se excluyeron también los registros con los siguientes valores en “pt”: Accidental death, Apparent death, Brain death, Cell death, Clinical death, Death, Disease progression, Malignant neoplasm progression, Neoplasm progression, Tumour hyperprogression, Tumour pseudoprogession. Una vez realizados los procesos de depuración anteriores, quedaron finalmente un total de 65.703 registros, correspondientes a 21.904 casos.

Un primer análisis del “data set” anterior, indica que el número de efectos adversos diferentes descritos con los medicamentos del grupo anterior es de 4091, una cifra demasiado alta para que sea operativo su análisis. Ha de tenerse en cuenta que muchos de estos efectos solo han sido reportados unas pocas veces (o incluso una sola vez), por lo que es conveniente una optimización del “data set” que facilite su análisis.

Para ello, se obtuvo una relación de los efectos adversos y el número de veces que son descritos. Ello permite tabularlos y ordenarlos de mayor a menor, calculando los % y los %acumulados (Ilustración 4).

pt(Efecto)	num.notificaciones	%	%Acum
Diarrhoea	1171	1,78	1,78
Pyrexia	1012	1,54	3,32
Fatigue	970	1,48	4,80
Decreased appetite	852	1,30	6,10
Rash	822	1,25	7,35
Nausea	682	1,04	8,38
Hypertension	626	0,95	9,34
Hypothyroidism	617	0,94	10,28
Pneumonia	614	0,93	11,21
Anaemia	593	0,90	12,11
Malaise	592	0,90	13,01
Interstitial lung disease	564	0,86	13,87
Intentional product use issue	563	0,86	14,73
Dyspnoea	539	0,82	15,55
Platelet count decreased	537	0,82	16,37
Asthenia	515	0,78	17,15
Neutrophil count decreased	496	0,75	17,91
Hepatic function abnormal	443	0,67	18,58
Febrile neutropenia	436	0,66	19,24
Vomiting	433	0,66	19,90
Acute kidney injury	430	0,65	20,56
Adrenal insufficiency	419	0,64	21,20
Arthralgia	403	0,61	21,81
Constipation	402	0,61	22,42
Colitis	401	0,61	23,03

Ilustración 4: Efectos adversos ordenados por su frecuencia

A partir de la tabla recogida en la ilustración 4, se pudo observar que con 279 efectos adversos era posible cubrir el 80% de todos los efectos adversos notificados, por lo que el “data set” se limitó a los registros que implicaban alguno de esos efectos adversos, sin que ello supusiese una pérdida excesiva de información relevante. El “data set” final contenía 40.643 registros, que se corresponden a 15.961 casos notificados. Los atributos (campos) incluyen: “pt” (reacción adversa), “age_grp” (grupo de edad), “sex” (sexo) y la clase “prod_ai” (principio activo).

3.4 Análisis convencional de la información utilizando herramientas de hoja de cálculo

Sobre la base del archivo Excel conteniendo el “data set” depurado (40.643 registros) de los efectos adversos de los ICIs, se hizo un análisis tradicional de sus datos, empleando herramientas propias de una hoja de cálculo, fundamentalmente tablas dinámicas y filtros.

En primer lugar, se analiza la distribución del número de notificaciones por principio activo, teniendo en cuenta el sexo (ilustración 5) y el grupo de edad (ilustración 6):

Medicamento	Núm. Notificaciones
ATEZOLIZUMAB	7806
F	3381
M	4425
AVELUMAB	580
F	162
M	418
CEMPIIMAB	146
F	45
M	101
DURVALUMAB	1442
F	416
M	1026
IPILIMUMAB	4664
F	1323
M	3341
NIVOLUMAB	13669
F	4591
M	9078
PEMBROLIZUMAB	12336
F	7117
M	5219
Total general	40643

Ilustración 5: Distribución del número de notificaciones, según medicamento y sexo

Medicamento	Núm. Notificaciones
ATEZOLIZUMAB	7806
?	337
A	2829
E	4639
T	1
AVELUMAB	580
?	39
A	124
E	417
CEMPIIMAB	146
?	41
A	39
E	61
T	5
DURVALUMAB	1442
?	130
A	528
E	779
I	5
IPILIMUMAB	4664
?	123
A	1740
C	1
E	2798
I	2
NIVOLUMAB	13669
?	596
A	5394
C	23
E	7615
I	1
T	40
PEMBROLIZUMAB	12336
?	2295
A	4046
C	2
E	5976
T	17
Total general	40643

Ilustración 6: Distribución del número de notificaciones, según medicamento y grupo de edad

El significado de los códigos de los grupos de edad es el siguiente: N Neonate; I Infant; C Child; T Adolescent; A Adult; E Elderly; ? Desconocido.

A diferencia del sexo, que está definido en la totalidad de los pacientes, la información sobre los grupos de edad es más errática, con un número considerable de situaciones en las que es desconocido y una alta concentración de notificaciones en personas adultas o de edad avanzada (Ilustración 7), algo lógico desde el punto de vista clínico, pero que implica que sea una variable poco útil para la clasificación.

Grupo de edad	Núm. Notificaciones
?	3561
A	14700
C	26
E	22285
I	8
T	63
Total general	40643

Ilustración 7: Número de notificaciones según grupo de edad

También se analizó, mediante una tabla dinámica, el recuento de cada efecto adverso descrito y su distribución según el principio activo (Ilustración 8).

Número de notificaciones	Etiquetas de co							
Efecto adverso	ATEZOLIZUMAB	AVELUMAB	CEMIPLIMAB	DURVALUMAB	IPILIMUMAB	NIVOLUMAB	PEMBROLIZUMAB	Total general
Abdominal discomfort	4	1	0	0	4	27	23	59
Abdominal distension	15	0	0	2	3	31	17	68
Abdominal pain	54	4	1	5	6	80	56	206
Abdominal pain upper	23	1	0	3	3	50	43	123
Acute kidney injury	101	7	1	9	47	142	108	415
Adrenal disorder	8	3	0	1	18	20	22	72
Adrenal insufficiency	103	3	2	5	108	104	69	394
Adrenocorticotrophic hormone deficiency	12	3	0	2	46	41	32	136
Alanine aminotransferase increased	79	4	2	5	36	62	37	225
Alopecia	29	1	0	4	4	32	31	101
Altered state of consciousness	14	4	0	1	23	20	18	80
Anaemia	226	6	0	23	34	134	120	543
Anaphylactic reaction	12	0	0	3	7	27	8	57
Anxiety	4	0	0	3	0	48	17	72
Arthralgia	39	3	2	9	33	123	174	383
Arthritis	15	0	0	2	13	34	26	90
Ascites	120	1	0	2	4	53	39	219
Aspartate aminotransferase increased	75	3	2	3	31	66	37	217
Asthenia	68	8	1	16	16	219	167	495
Atrial fibrillation	39	3	1	0	7	58	33	141
Autoimmune disorder	2	0	0	1	2	14	11	30
Autoimmune hepatitis	15	2	1	1	6	24	10	59
Back pain	33	4	1	7	12	98	48	203
Blister	4	0	0	2	3	49	16	74
Blood alkaline phosphatase increased	19	0	0	2	1	9	9	40
Blood bilirubin increased	36	0	1	0	8	15	12	72
Blood creatine phosphokinase increased	15	2	1	4	8	23	25	78
Blood creatinine increased	54	3	0	0	20	26	45	148
Blood glucose increased	9	2	1	3	3	15	29	62
Blood pressure decreased	4	5	1	2	10	27	19	68
Blood pressure increased	10	4	0	3	1	92	102	212
Blood thyroid stimulating hormone increased	4	1	1	0	2	14	42	64
Brain oedema	9	0	0	2	11	18	11	51
Weight decreased	33	2	1	6	12	207	133	394
Weight increased	7	1	0	4	1	41	28	82
White blood cell count decreased	127	1	0	37	29	59	61	314
Total general	7806	580	146	1442	4664	13669	12336	40643

Ilustración 8: Recuento de efectos adversos según principio activo

Si la tabla anterior se ordena de mayor a menor número de notificaciones se tendrá la información de los efectos adversos más relevantes desde el punto de vista de su frecuencia de notificación (En la ilustración 9 se muestran las 30 primeras).

Efecto adverso	ATEZOLIZUMAB	AVELUMAB	CEMIPLIMAB	DURVALUMAB	IPILIMUMAB	NIVOLUMAB	PEMBROLIZUMAB	Total general
Diarrhoea	139	15	4	24	121	454	361	1118
Pyrexia	259	11	2	26	137	278	243	956
Fatigue	97	15	6	17	28	410	361	934
Decreased appetite	132	7	1	19	92	318	242	811
Rash	72	6	2	23	171	267	236	777
Nausea	100	4	2	12	53	283	203	657
Hypertension	125	11	0	5	11	89	356	597
Hypothyroidism	89	14	1	15	78	144	250	591
Pneumonia	156	5	5	22	67	212	107	574
Malaise	84	8	1	8	97	144	223	565
Anaemia	226	6	0	23	34	134	120	543
Dyspnoea	107	5	2	27	24	209	143	517
Interstitial lung disease	143	20	0	29	105	106	113	516
Platelet count decreased	147	5	0	46	38	67	197	500
Asthenia	68	8	1	16	16	219	167	495
Neutrophil count decreased	179	1	0	73	47	55	67	422
Acute kidney injury	101	7	1	9	47	142	108	415
Vomiting	80	6	0	19	32	167	111	415
Hepatic function abnormal	85	6	0	15	95	100	107	408
Febrile neutropenia	170	2	0	41	46	80	56	395
Adrenal insufficiency	103	3	2	5	108	104	69	394
Weight decreased	33	2	1	6	12	207	133	394
Arthralgia	39	3	2	9	33	123	174	383
Constipation	94	1	0	7	25	166	87	380
Pruritus	31	8	2	5	48	173	109	376
Pneumonitis	75	1	4	59	37	104	86	366
Colitis	65	5	1	7	60	121	105	364
Immune-mediated enterocolitis	24	1	0	4	86	161	85	361
Neutropenia	88	0	0	11	30	55	138	322
General physical health deterioration	65	11	1	6	20	142	72	317
Total general	7806	580	146	1442	4664	13669	12336	40643

Ilustración 9: Recuento de efectos adversos según principio activo (ordenado por frecuencia)

Posteriormente se añadió la variable sexo a las tablas dinámicas mencionadas, obteniéndose la ilustración 10, que una vez ordenada de mayor a menor frecuencia del efecto adverso da lugar a la ilustración 11 (se muestran las primeras 15)

Efecto adverso	Medicamento								
Etiquetas de fila	ATEZOLIZUMAB	AVELUMAB	CEMIPLIMAB	DURVALUMAB	IPILIMUMAB	NIVOLUMAB	PEMBROLIZUMAB	Total general	
Abdominal discomfort	4	1			4	27	23	59	
F	2				2	11	13	28	
M	2	1				2	10	31	
Abdominal distension	15			2	3	31	17	68	
F	4				1	5	12	22	
M	11			1	3	26	5	46	
Abdominal pain	54	4	1	5	6	80	56	206	
F	29	2	1	3	2	31	34	102	
M	25	2		2	4	49	22	104	
Abdominal pain upper	23	1		3	3	50	43	123	
F	10	1		2	2	20	29	64	
M	13			1	1	30	14	59	
Acute kidney injury	101	7	1	9	47	142	108	415	
F	39	2		3	11	39	49	143	
M	62	5	1	6	36	103	59	272	
Adrenal disorder	8	3		1	18	20	22	72	
F	6				8	5	5	24	
M	2	3		1	10	15	17	48	
Adrenal insufficiency	103	3	2	5	108	104	69	394	
F	35	2	1	1	28	23	33	123	
M	68	1	1	4	80	81	36	271	
Adrenocorticotrophic hormone deficiency	12	3		2	46	41	32	136	
F	5			2	11	9	12	39	
M	7	3			35	32	20	97	
White blood cell count decreased	127	1		37	29	59	61	314	
F	63			6	7	20	27	123	
M	64	1		31	22	39	34	191	
Total general	7806	580	146	1442	4664	13669	12336	40643	

Ilustración 10: Recuento de efectos adversos según principio activo y sexo

Efecto adverso	Medicamento							
Etiquetas de fila	ATEZOLIZUMAB	AVELUMAB	CEMPIPLIMAB	DURVALUMAB	IPILIMUMAB	NIVOLUMAB	PEMBROLIZUMAB	Total general
Diarrhoea	139	15	4	24	121	454	361	1118
F	67	2		9	36	136	256	506
M	72	13	4	15	85	318	105	612
Pyrexia	259	11	2	26	137	278	243	956
F	127	2		9	35	93	157	423
M	132	9	2	17	102	185	86	533
Fatigue	97	15	6	17	28	410	361	934
F	47	3	2	8	11	140	237	448
M	50	12	4	9	17	270	124	486
Decreased appetite	132	7	1	19	92	318	242	811
F	47	1		2	27	96	135	308
M	85	6	1	17	65	222	107	503
Rash	72	6	2	23	171	267	236	777
F	29	1		7	50	94	140	321
M	43	5	2	16	121	173	96	456
Nausea	100	4	2	12	53	283	203	657
F	47	4		3	22	104	138	318
M	53		2	9	31	179	65	339
Hypertension	125	11		5	11	89	356	597
F	54	6		4	3	25	284	376
M	71	5		1	8	64	72	221
Hypothyroidism	89	14	1	15	78	144	250	591
F	32	5		5	24	45	180	291
M	57	9	1	10	54	99	70	300
Pneumonia	156	5	5	22	67	212	107	574
F	57	1		2	30	69	44	203
M	99	4	5	20	37	143	63	371
Malaise	84	8	1	8	97	144	223	565
F	28	5		3	28	32	153	249
M	56	3	1	5	69	112	70	316
Anaemia	226	6		23	34	134	120	543
F	126	1		7	9	47	61	251
M	100	5		16	25	87	59	292
Dyspnoea	107	5	2	27	24	209	143	517
F	54		1	7	7	70	69	208
M	53	5	1	20	17	139	74	309
Interstitial lung disease	143	20		29	105	106	113	516
F	37	3		4	22	16	27	109
M	106	17		25	83	90	86	407
Platelet count decreased	147	5		46	38	67	197	500
F	64			11	13	27	150	265
M	83	5		35	25	40	47	235
Asthenia	68	8	1	16	16	219	167	495
F	29	2	1	4	6	73	97	212
M	39	6		12	10	146	70	283
Total general	7806	580	146	1442	4664	13669	12336	40643

Ilustración 11: Recuento de efectos adversos según principio activo y sexo (ordenado por frecuencia)

3.5 Análisis del “data set” mediante herramientas de aprendizaje automatizado

La hoja de cálculo con los datos depurados fue exportada a un archivo de tipo “csv” para su posterior incorporación al software Weka 3: Machine Learning Software in Java (27).

Ha de tenerse en cuenta que la exportación desde Excel ha de hacerse, estando en la hoja que contiene los datos, a través de la opción “guardar como” y seleccionando el tipo CSV (MS-DOS). Además, dado que Excel usa como delimitador de campos el “;” en lugar de “,” es necesario abrir el archivo csv mediante un editor de texto plano (como el propio block de notas de Windows) y remplazar todos los “;” por “,”. Si no se realiza de esta manera, Weka no será capaz de importarlo correctamente.

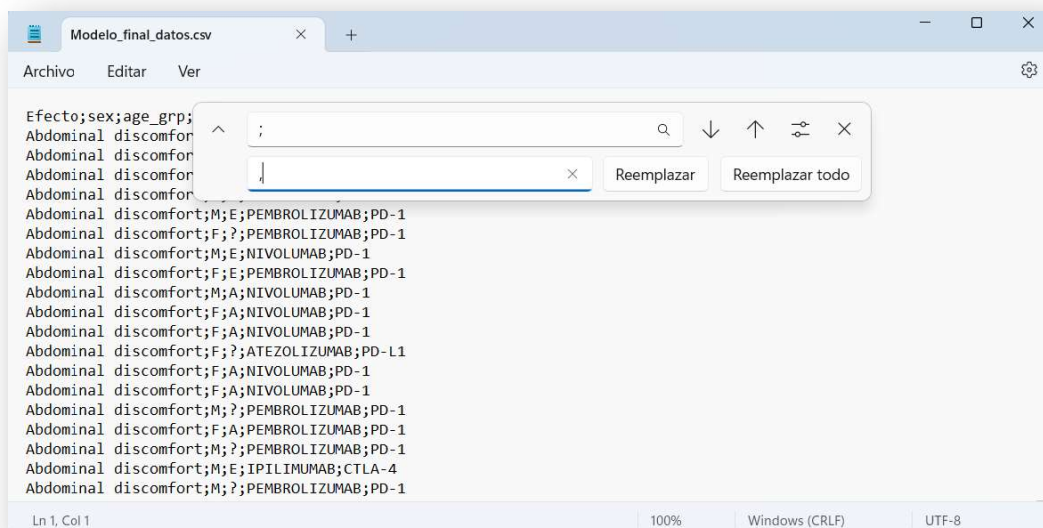


Ilustración 12: Ajustes de la exportación desde Excel a fichero csv

Los campos (atributos) seleccionados han sido: pt (renombrado “Efecto”), sex, age_grp y prod_ai (renombrado “clase”). Se añadió un campo adicional, denominado “mecanismo”, que no forma parte de los datos recogidos en FAERS. Se refiere al mecanismo por el que actúan los ICIs: Sobre el PD-1 (Nivolumab, Pembrolizumab, Cemiplimab), sobre el PD-L1 (Atezolizumab, Avelumab, Durvalumab) o sobre el CTLA-4 (Ipilimumab).

Para evitar tener que usar filtros en el aplicativo weka se prepararon, a partir del “data set” anterior, distintos subset, solo con los atributos que se analizaban en cada momento:

3.5.1 Análisis Subset1: pt(efecto), sex, prod_ai (clase). Algoritmo J48.

Se trata de analizar, con el algoritmo J48, la posibilidad de predecir, dado un efecto adverso observado en un paciente (varón o mujer), cual es el medicamento ICI responsable. De esa manera alertaría, en un paciente que va a ser tratado con ICIs y que tiene factores de riesgo para desarrollar ese efecto adverso, cual sería el medicamento del grupo a evitar (siempre que se disponga de otras alternativas factibles para la indicación)

Los datos del subset fueron cargados en Weka (versión 3.8.6), a través de su aplicación “explorer” (ilustración 13).

Este data set se procesó con el algoritmo j48, manteniendo los parámetros por defecto de la aplicación Weka. En el Anexo 1 se muestra el resultado completo.

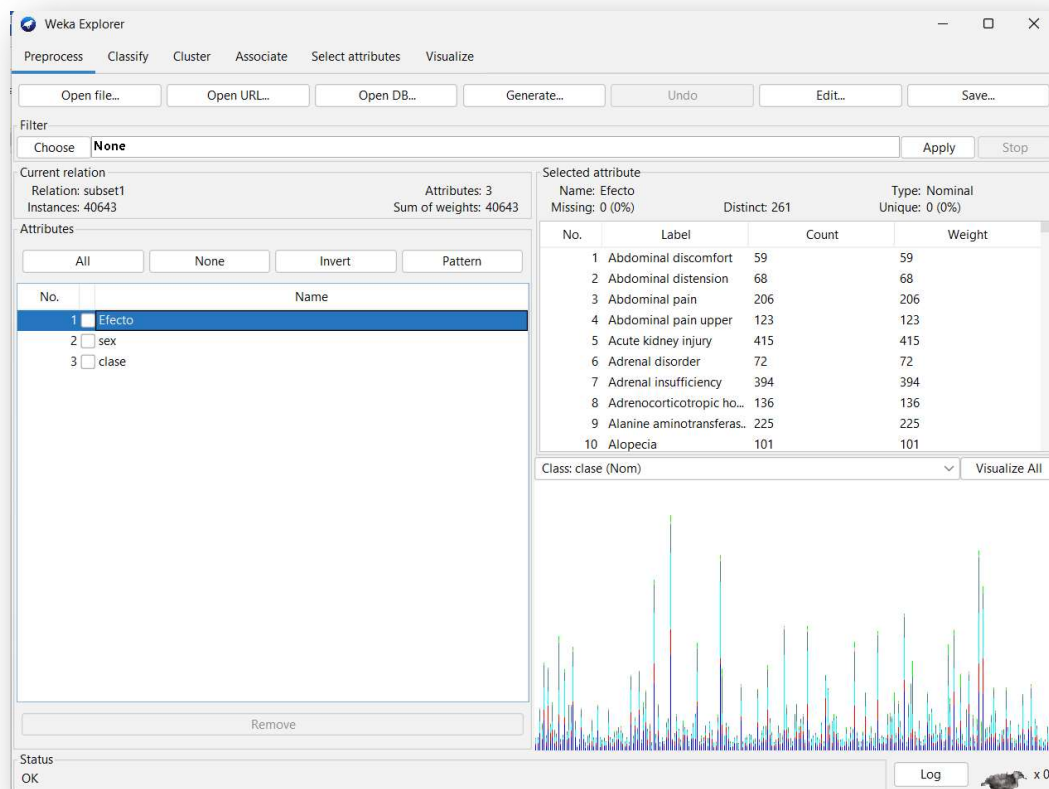


Ilustración 13: Weka explorer con subset1

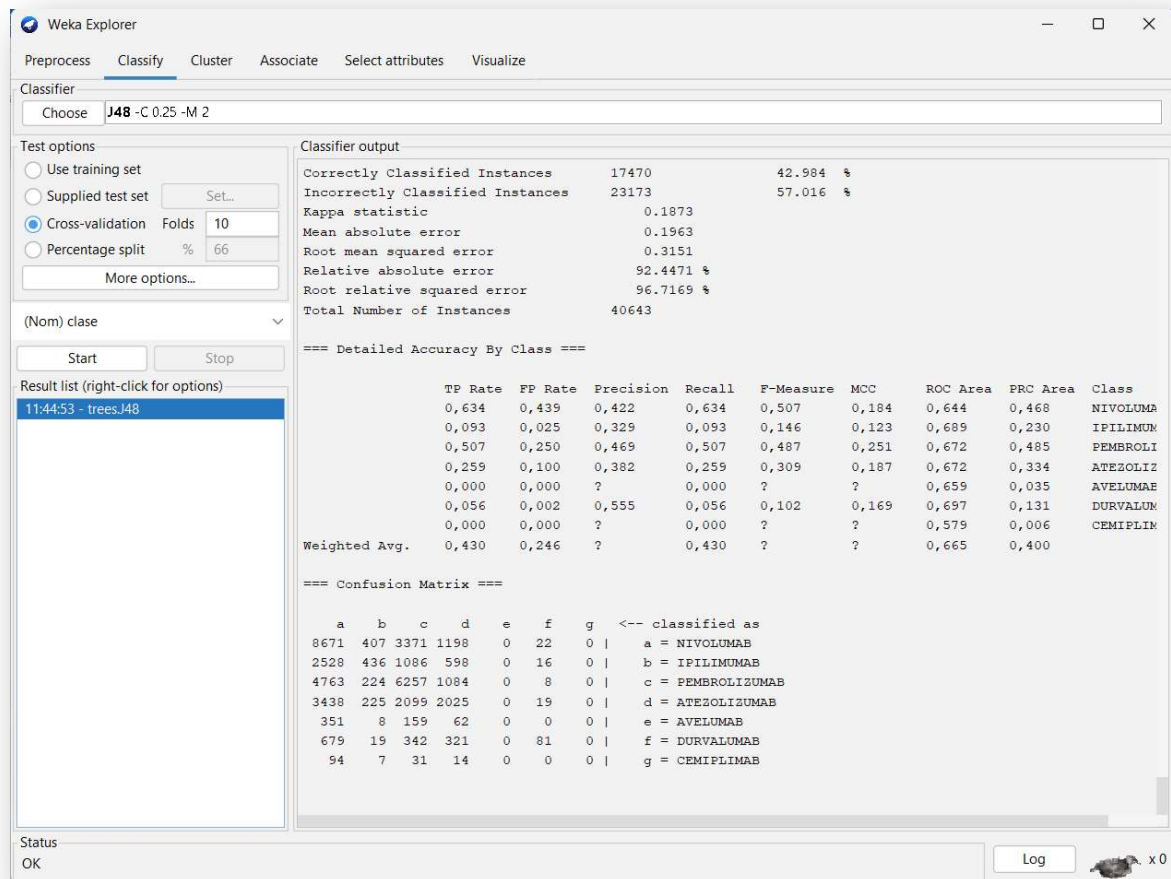


Ilustración 14: Métricas J48 sobre subset1

El algoritmo elabora también un árbol muy complejo, con 383 hojas (algo esperable dado el considerable número de efectos adversos que se analizan). La representación gráfica del mismo, aunque factible, por su complejidad no es posible hacerla de manera legible en el tamaño de papel habitual.

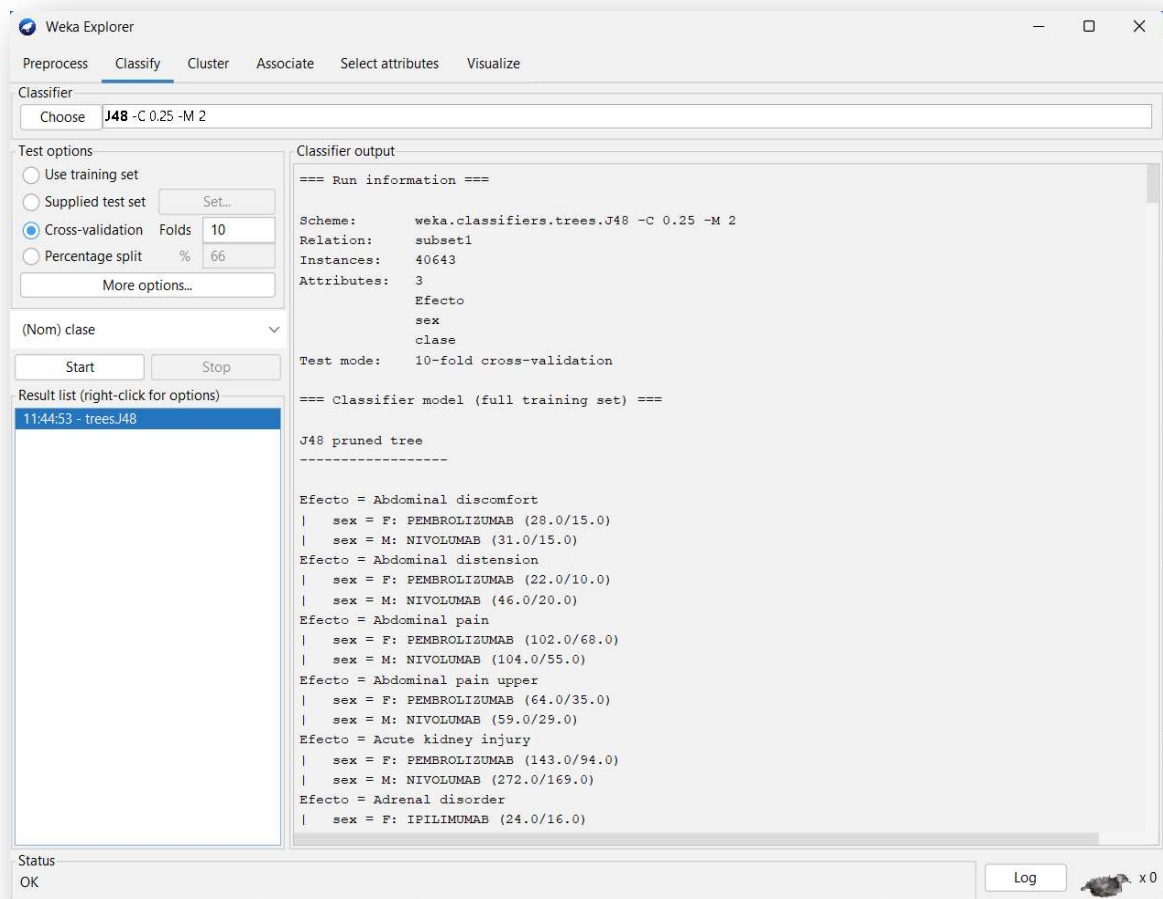


Ilustración 15: Arbol J48 sobre subset1

3.5.2 Análisis Subset1: pt(efecto), sex, prod_ai (clase). Comparativo con modelo convencional.

Los resultados del algoritmo J48 fueron analizados e interpretados comparando con los resultados obtenidos a partir del análisis convencional mediante herramientas de hoja de cálculo realizado en el punto 3.4.

El total de instancias clasificadas, según los resultados del algoritmo, ha sido de 40643, que se corresponde con el 100% de las notificaciones. De ellas, la predicción es correcta en el 43% de los casos. Aunque es un valor bajo, se ha de contextualizar analizando su significado. Que la predicción sea correcta en el 43% de los casos (17.470 notificaciones) significa que, en promedio, el algoritmo asocia correctamente el efecto adverso a un único medicamento. Que la predicción sea incorrecta en el 57% de los casos (23.173 notificaciones) implica que en estas notificaciones el medicamento asociado es alguno de los otros 6 del grupo. Es comprensible que no puedan alcanzarse valores de % de predicción muy altos, ya que se analiza un grupo de fármacos muy similares entre sí.

Este hecho se aprecia mejor analizando algunos efectos adversos concretos (los más frecuentes):

a) En el caso de “Diarrea”, el algoritmo indica lo siguiente:

Efecto = Diarrhoea

| sex = F: PEMBROLIZUMAB (506.0/250.0)

| sex = M: NIVOLUMAB (612.0/294.0)

En este caso, el efecto se asocia con mayor probabilidad a Pembrolizumab, en el caso de mujeres: de las 506 notificaciones en mujeres, Pembrolizumab figuraba en 256 notificaciones (50,5%), y alguno de los otros (nivolumab, atezolizumab, cemiplimab, ipilimumab, avelumab o durvalumab) en las 250 restantes (40,5%).

En el caso de hombres, el medicamento que se asocia con mayor probabilidad a este efecto es el nivolumab, ya que figura en 318 notificaciones (52,0%), y alguno de los otros en los 294 casos restantes (48,0%)

En base a lo anterior se deduce que, en el caso de hombres, la utilización de pembrolizumab entraña un mayor riesgo de desarrollar este efecto. En el caso de mujeres, es el nivolumab el que se asocia a un mayor riesgo.

b) En el caso de “pirexia”, J48 indica:

Efecto = Pyrexia

| sex = F: PEMBROLIZUMAB (423.0/266.0)

| sex = M: NIVOLUMAB (533.0/348.0)

En este caso, en mujeres, Pembrolizumab se asocia con el efecto adverso en 157 casos (37,1%) y el resto de los medicamentos en 266 casos (62,9%). En el caso de hombres, se asocia con Nivolumab en 185 casos (34,7%) y con el resto en 348 casos (65,3%). Aunque se pueden extraer conclusiones análogas a las del caso anterior, la asociación efecto adverso-medicamento es más débil.

c) Pneumonía

Efecto = Pneumonia: NIVOLUMAB (574.0/362.0)

En este caso, el algoritmo no encuentra diferencias en función del sexo. Por tanto, para el conjunto de hombres y mujeres, 212 notificaciones se asocian a nivolumab (36,9%) y 362 al resto (63,1%). Además, la asociación Pneumonía-Pembrolizumab, aunque mayoritaria, es débil.

Como puede observarse en los resultados del apartado 3.4, las conclusiones de los ejemplos también se obtienen del análisis mediante herramientas de hoja de cálculo, aunque de una manera más compleja y que precisa más tiempo. El algoritmo J48, en las condiciones descritas, hace el análisis en unos pocos segundos y, además, se podría aplicar a “data set” mucho más voluminosos, que incluso no puedan analizarse en una hoja de cálculo, dadas sus limitaciones en el número de filas.

3.5.3 Análisis Subset2: pt(efecto), age_gr, prod_ai (clase). Algoritmo J48.

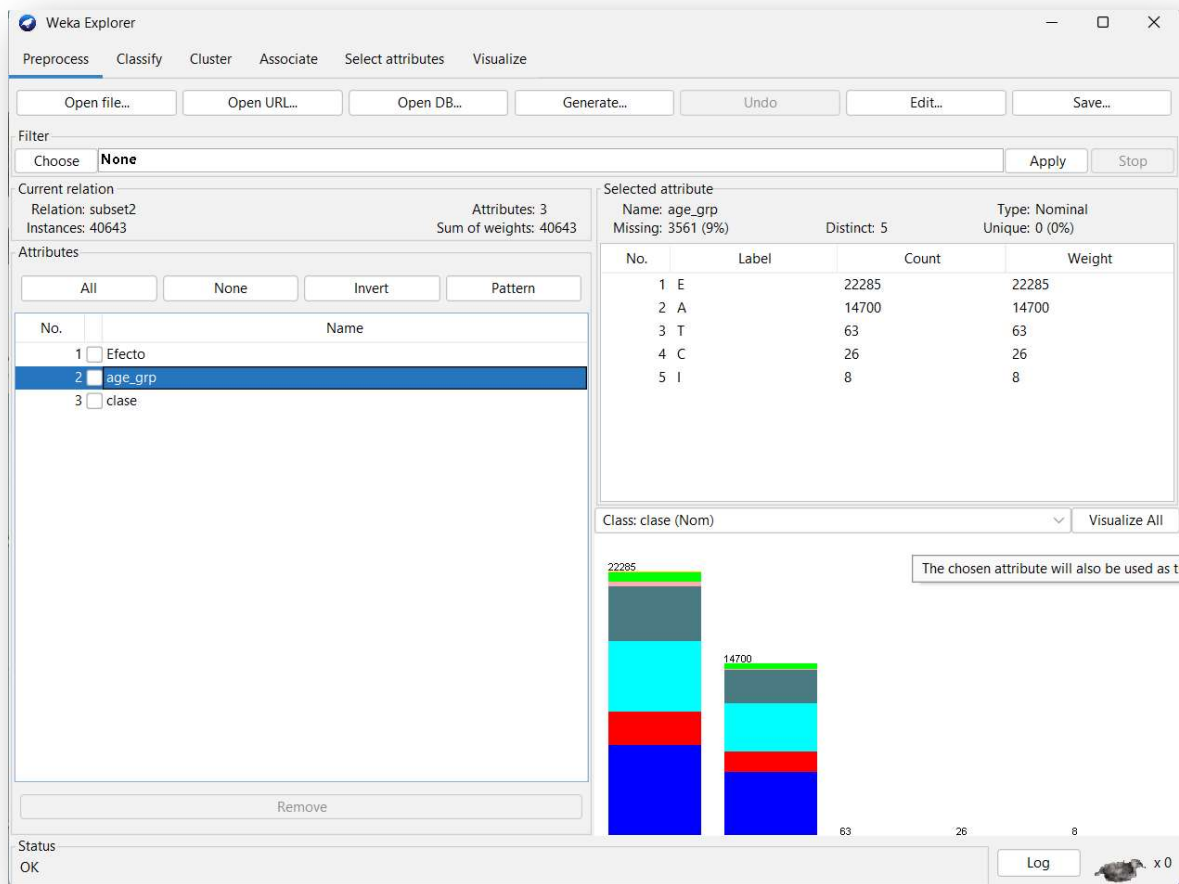


Ilustración 16: Arbol J48 sobre subset2

El resultado de aplicar el algoritmo J48 aparece en la Ilustración 17.

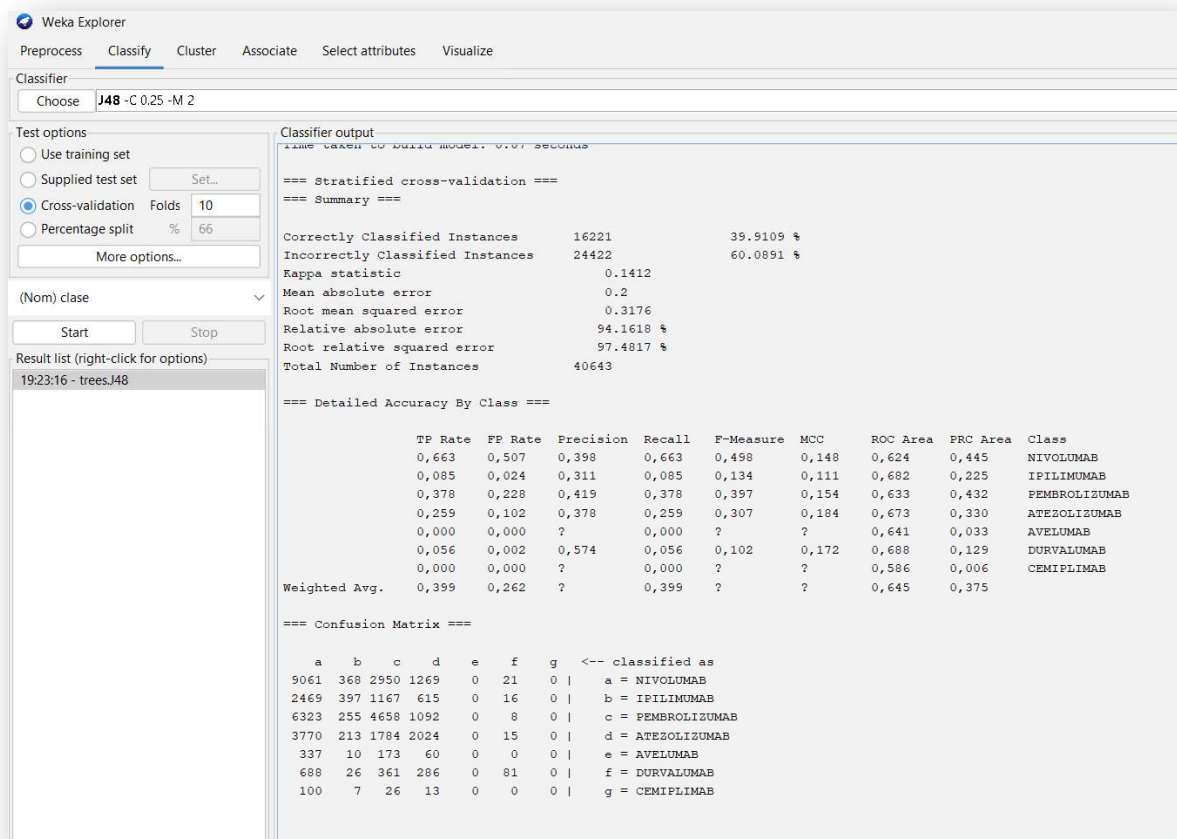


Ilustración 17: Métricas J48 sobre subset2

Se aprecia que la capacidad predictiva del modelo es mucho menor que en el punto anterior. Aunque construye un árbol con 461 hojas, es de poca utilidad. Se comprende al observar los datos de partida: Solo hay una cantidad importante de notificaciones en el caso de pacientes adultos y ancianos, lo que limita su capacidad predictiva para el conjunto de todos los grupos de edad.

Para corroborar lo anterior, se describe el ejemplo correspondiente al efecto adverso “Diarrea”:

Efecto = Diarrhoea: NIVOLUMAB (1118.0/664.0)

Aunque el medicamento más frecuentemente asociado al efecto adverso es Nivolumab (45% de las notificaciones), no se puede establecer ninguna asociación con grupos de edad.

Por ello, el modelo que busca relacionar medicamento, efecto adverso y grupo de edad, es poco útil con el dataset proporcionado, debido a que las notificaciones están concentradas en solo 2 de los grupos, y el resto son marginales.

4. DISCUSIÓN

El “Sistema de Informes de Eventos Adversos de la FDA” (FDA Adverse Event Reporting System-FAERS), es una base de datos que contiene información sobre los eventos adversos e informes de medicación errónea que han sido reportados a la Administración de Alimentos y Medicamentos de Estados Unidos (FDA). La posibilidad de descargarla localmente abre múltiples posibilidades para estudiar relaciones entre medicamentos y eventos adversos, así como para comparar eventos adversos entre diferentes medicamentos dentro de un grupo farmacológico concreto. Permite seleccionar “data sets” para entrenar sistemas de aprendizaje automático, cuyos resultados pueden contrastarse con el procesamiento de esos “data set” por sistemas tradicionales, lo que ayuda a comprender mejor el funcionamiento y las limitaciones de los diferentes algoritmos, contribuyendo a lo que se ha denominado “inteligencia artificial explicable”, también conocida como XAI (eXplainable Artificial Intelligence). Esta es la filosofía que inspira la presente memoria.

Por otra parte, los inhibidores de los puntos de control inmunitarios (ICIs) son un tipo de inmunoterapia que ha revolucionado el tratamiento de varios tipos de cáncer en los últimos años. Sin embargo, dado que estos medicamentos modifican la respuesta inmunitaria del individuo, también pueden dar lugar a múltiples efectos adversos, que pueden variar en gravedad, desde síntomas leves hasta reacciones potencialmente mortales. Por ello, pueden constituir un buen grupo farmacológico para analizar las notificaciones relativas a ellos en el sistema FAERS.

En este contexto, y con una finalidad eminentemente didáctica (poner en práctica los conocimientos adquiridos durante el curso), se plantea esta memoria. La selección del grupo de medicamentos ICIs, aunque se ha descargado totalmente la BD de FAERS de 2022, obedece, además de su importancia clínica, a que da lugar a un conjunto de datos que es factible analizar tanto por métodos tradicionales como por algoritmos de aprendizaje automático.

4.1 Consideraciones derivadas de los datos de partida

Ha de tenerse en cuenta que, aunque FAERS puede ser una herramienta muy útil para el estudio de los efectos adversos de medicamentos, tiene limitaciones importantes, por una parte derivadas de que solo incluye notificaciones realizadas en el contexto de la FDA, y por otra, derivadas del hecho de depender de la notificación voluntaria de eventos adversos, por lo que puede estar sujeta a sesgos y puede no reflejar la verdadera

prevalencia de un evento adverso. Además, la presencia de un informe en FAERS no significa que el medicamento haya causado necesariamente el evento adverso reportado.

Además, sus registros, aunque son estructurados y se basan en el uso de “structured product labeling (SPL)” y códigos estandarizados, pueden estar incompletos o duplicados y no han sido verificados. Además, no pueden usarse para estimar la incidencia.

Este hecho se ha puesto de manifiesto en la presente memoria, encontrando, por ejemplo, campos como “edad” o “peso” donde, o bien no figuran, o bien tienen errores evidentes, por utilizar unidades diferentes, estar mal escritos, etc. El campo de “grupo de edad” no está cubierto en muchas ocasiones. Por otra parte, en el campo donde se recoge la denominación del efecto adverso, es frecuente que aparezcan otro tipo de información que no es un efecto adverso.

Por todo ello, es fundamental una revisión previa de los datos para depurarlos: una especie de “data curation” con los datos seleccionados, para mejorar su calidad. El grupo farmacológico elegido para esta memoria, medicamentos ICIs, al tener un número de registros compatible con una hoja de cálculo “Excel”, pudo depurarse y optimizarse con relativa facilidad.

4.2 Consideraciones derivadas del análisis con herramientas de hoja de cálculo.

Se ha hecho un primer análisis de los datos utilizando la hoja de cálculo “Excel”, dado que el número de registros lo permite y porque nos proporciona resultados fiables y trazables, que se podrán tomar como referencia cuando se analicen los obtenidos mediante herramientas de aprendizaje automatizado, lo cual se ajusta al carácter eminentemente docente de la memoria.

No obstante, las limitaciones en cuanto al número de registros que maneja Excel, así como lo tedioso y complejo que puede llegar a ser su análisis, implica que no es sistema adecuado para el manejo de grandes volúmenes de información (propios del “big data”).

4.3 Consideraciones derivadas del uso del algoritmo J48

Se ha elegido para esta memoria el algoritmo J48, por su versatilidad, relativa sencillez y porque permite una “visualización” gráfica de los resultados (árbol). No obstante, se es consciente de sus limitaciones, ya que es un algoritmo de clasificación, por lo que en temas como el planteado puede ser insuficiente y existir otros mejores (aunque probablemente más complejos y menos “gráficos”).

La principal dificultad que se ha planteado en esta memoria es que el número de efectos adversos (más de 250) y el número de clases (7 medicamentos) es muy amplio, lo que da

lugar a árboles muy complejos, imposibles de visualizar en el tamaño de papel habitual. Sin embargo, como el algoritmo también hace una descripción en texto del árbol, sigue siendo factible su utilización.

La capacidad para predecir si un efecto adverso es exclusivo de un medicamento en concreto, es limitada (en torno al 43%), pero ha resultado útil en nuestro contexto porque nos permite ver y cuantificar, en cada efecto adverso, cuál es el medicamento que se notifica con mayor frecuencia, incluso con la posibilidad de diferenciar según se trate de hombres o mujeres. El hecho de tener como referencia el análisis en Excel, facilitó comprender lo que representaban los resultados del algoritmo. Ello puede contribuir, a igualdad de otras variables, a elegir un determinado medicamento ICIs según el perfil de riesgos y el sexo del paciente.

4.4 Limitaciones del estudio

En esencia, esta memoria pretende poner en práctica los conocimientos adquiridos por el alumno a lo largo del curso. Por ello, la extrapolación de los resultados a una potencial utilidad en práctica clínica, está sujeta a múltiples limitaciones, muchas de ellas ya expuestas: Limitaciones derivadas del origen de los datos, limitaciones derivadas del algoritmo empleado, etc.

No obstante, la principal limitación, que no se solucionaría aunque se eligiesen algoritmos optimizados para el aprendizaje automatizado, deriva del hecho de que la información de una base de datos basada en notificaciones voluntarias no puede usarse para estimar la incidencia. Esto es especialmente relevante en este tipo de estudios, ya que el hecho de que un efecto adverso se haya notificado poco con un determinado medicamento no implica, necesariamente, que tenga una menor incidencia, ya que, por ejemplo, puede ser debido a que el medicamento se utilice menos en la práctica clínica, o lleve menos tiempo en el mercado. Aún así, puede servir de referencia para la realización de otros estudios.

4.5 Conclusiones

Se ha construido una base de datos local, a partir de todas las notificaciones reportadas en 2022 en el sistema FAERS. De este conjunto, se han extraído y depurado los datos relativos a los pacientes a tratamiento con Inhibidores de los puntos de control inmunitario (ICIs): avelumab, atezolizumab, cemiplimab, durvalumab, nivolumab, pembrolizumab e ipilimumab). Durante el proceso de extracción y depuración se han constatado las limitaciones intrínsecas a un sistema de notificación voluntaria de efectos adversos.

El análisis de los datos seleccionados (ICIs) con herramientas propias de una hoja de cálculo (Excel), fundamentalmente tablas dinámicas, es viable para “data set” de hasta un millón de registros, aunque puede ser complejo y lento. En este estudio, fue posible conocer la distribución de las notificaciones de efectos adversos por medicamento y sexo, grupos de edad, etc. También fue posible conocer y ordenar los efectos adversos más notificados y los medicamentos implicados.

Sobre el conjunto de datos seleccionado y depurado, se aplicó un modelo de aprendizaje automatizado basado en el algoritmo J48 implementado en el aplicativo Weka. Sus resultados, cuya comprensión se facilitó al disponer como referencia del análisis realizado con herramientas convencionales, fueron útiles cuando se buscó la asociación entre efecto adverso, sexo y medicamento, ya que permiten visualizar (y en cierta medida cuantificar), para cada efecto adverso el medicamento más frecuentemente asociado. El modelo tiene un funcionamiento peor cuando se analiza la relación efecto adverso-grupo de edad y medicamento, debido a la menor calidad de los datos de partida.

BIBLIOGRAFÍA

1. Lee CY, Chen YPP. Machine learning on adverse drug reactions for pharmacovigilance. *Drug Discov Today*. julio de 2019;24(7):1332-43.
2. Kim HR, Sung M, Park JA, Jeong K, Kim HH, Lee S, et al. Analyzing adverse drug reaction using statistical and machine learning methods: A systematic review. *Medicine (Baltimore)*. 24 de junio de 2022;101(25):e29387.
3. Research C for DE and. FDA Adverse Event Reporting System (FAERS) Public Dashboard. FDA [Internet]. 22 de octubre de 2021 [citado 8 de junio de 2023]; Disponible en: <https://www.fda.gov/drugs/questions-and-answers-fdas-adverse-event-reporting-system-faers/fda-adverse-event-reporting-system-faers-public-dashboard>
4. EMA. EudraVigilance [Internet]. European Medicines Agency. 2018 [citado 8 de junio de 2023]. Disponible en: <https://www.ema.europa.eu/en/human-regulatory/research-development/pharmacovigilance/eudravigilance>
5. World Health Organization. Quality Assurance and Safety of Medicines Team. Safety of medicines: a guide to detecting and reporting adverse drug reactions: why health professionals need to take action [Internet]. World Health Organization; 2002 [citado 8 de junio de 2023]. Report No.: WHO/EDM/QSM/2002.2. Disponible en: <https://apps.who.int/iris/handle/10665/67378>
6. Pedrero V, Reynaldós-Grandón K, Ureta-Achurra J, Cortez-Pinto E, Pedrero V, Reynaldós-Grandón K, et al. Generalidades del Machine Learning y su aplicación en la gestión sanitaria en Servicios de Urgencia. *Rev Médica Chile*. febrero de 2021;149(2):248-54.
7. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. *N Engl J Med*. 4 de abril de 2019;380(14):1347-58.
8. Vega MÁ, Mora LM, Badilla MV. Inteligencia artificial y aprendizaje automático en medicina. *Rev Medica Sinerg*. 1 de agosto de 2020;5(8):e557-e557.
9. Holzinger A, Biemann C, Pattichis CS, Kell DB. What do we need to build explainable AI systems for the medical domain? [Internet]. arXiv; 2017 [citado 16 de junio de 2023]. Disponible en: <http://arxiv.org/abs/1712.09923>
10. Chaddad A, Peng J, Xu J, Bouridane A. Survey of Explainable AI Techniques in Healthcare. *Sensors*. 5 de enero de 2023;23(2):634.
11. Theunissen M, Browning J. Putting explainable AI in context: institutional explanations for medical AI. *Ethics Inf Technol*. 2022;24(2):23.
12. REGLAMENTO (UE) 2017/ 746 DEL PARLAMENTO EUROPEO Y DEL CONSEJO - de 5 de abril de 2017 - sobre los productos sanitarios para diagnóstico in vitro y por el que se derogan la Directiva 98/ 79/ CE y la Decisión 2010/ 227/ UE de la Comisión.
13. Müller H, Holzinger A, Plass M, Brcic L, Stumptner C, Zatloukal K. Explainability and causability for artificial intelligence-supported medical image analysis in the context of the European In Vitro Diagnostic Regulation. *New Biotechnol*. 25 de septiembre de 2022;70:67-72.

14. Data Mining with Weka MOOC - Material [Internet]. [citado 17 de marzo de 2023]. Disponible en: <https://www.cs.waikato.ac.nz/ml/weka/mooc/dataminingwithweka/>
15. WekaManual-3-8-1.pdf [Internet]. [citado 9 de junio de 2023]. Disponible en: <https://sci2s.ugr.es/sites/default/files/files/Teaching/GraduatesCourses/InteligenciaDeNegocio/Curso19-20/WekaManual-3-8-1.pdf>
16. C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993 | SpringerLink [Internet]. [citado 9 de junio de 2023]. Disponible en: <https://link.springer.com/article/10.1007/BF00993309>
17. Sohn S, Kocher JPA, Chute CG, Savova GK. Drug side effect extraction from clinical narratives of psychiatry and psychology patients. *J Am Med Inform Assoc JAMIA*. diciembre de 2011;18 Suppl 1(Suppl 1):i144-149.
18. Yoo JH, Sharma V, Kim JW, McMakin DL, Hong SB, Zalesky A, et al. Prediction of sleep side effects following methylphenidate treatment in ADHD youth. *NeuroImage Clin*. 2020;26:102030.
19. Bagchi S, Yuan R, Engleman EG. Immune Checkpoint Inhibitors for the Treatment of Cancer: Clinical Impact and Mechanisms of Response and Resistance. *Annu Rev Pathol*. 24 de enero de 2021;16:223-49.
20. Shiravand Y, Khodadadi F, Kashani SMA, Hosseini-Fard SR, Hosseini S, Sadeghirad H, et al. Immune Checkpoint Inhibitors in Cancer Therapy. *Curr Oncol Tor Ont*. 24 de abril de 2022;29(5):3044-60.
21. Wang DY, Salem JE, Cohen JV, Chandra S, Menzer C, Ye F, et al. Fatal Toxic Effects Associated With Immune Checkpoint Inhibitors: A Systematic Review and Meta-analysis. *JAMA Oncol*. 1 de diciembre de 2018;4(12):1721-8.
22. Conforti F, Pala L, Bagnardi V, De Pas T, Martinetti M, Viale G, et al. Cancer immunotherapy efficacy and patients' sex: a systematic review and meta-analysis. *Lancet Oncol*. junio de 2018;19(6):737-46.
23. Jing Y, Zhang Y, Wang J, Li K, Chen X, Heng J, et al. Association Between Sex and Immune-Related Adverse Events During Immune Checkpoint Inhibitor Therapy. *J Natl Cancer Inst*. 1 de octubre de 2021;113(10):1396-404.
24. Bukhtiyarova O, Abderrazak A, Chiu Y, Sparano S, Simard M, Sirois C. Major areas of interest of artificial intelligence research applied to health care administrative data: a scoping review. *Front Pharmacol*. 2022;13:944516.
25. Arku D, Yousef C, Abraham I. Changing paradigms in detecting rare adverse drug reactions: from disproportionality analysis, old and new, to machine learning. *Expert Opin Drug Saf*. octubre de 2022;21(10):1235-8.
26. FAERS Quarterly Data Extract Files [Internet]. [citado 15 de marzo de 2023]. Disponible en: <https://fis.fda.gov/extensions/FPD-QDE-FAERS/FPD-QDE-FAERS.html>
27. Downloading and installing Weka - Weka Wiki [Internet]. [citado 13 de junio de 2023]. Disponible en: https://waikato.github.io/weka-wiki/downloading_weka/

ANEXO 1: Resultados J48 sobre subset1

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: subset1
Instances: 40643
Attributes: 3
 Efecto
 sex
 clase
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree

Efecto = Abdominal discomfort
| sex = F: PEMBROLIZUMAB (28.0/15.0)
| sex = M: NIVOLUMAB (31.0/15.0)
Efecto = Abdominal distension
| sex = F: PEMBROLIZUMAB (22.0/10.0)
| sex = M: NIVOLUMAB (46.0/20.0)
Efecto = Abdominal pain
| sex = F: PEMBROLIZUMAB (102.0/68.0)
| sex = M: NIVOLUMAB (104.0/55.0)
Efecto = Abdominal pain upper
| sex = F: PEMBROLIZUMAB (64.0/35.0)
| sex = M: NIVOLUMAB (59.0/29.0)
Efecto = Acute kidney injury
| sex = F: PEMBROLIZUMAB (143.0/94.0)
| sex = M: NIVOLUMAB (272.0/169.0)
Efecto = Adrenal disorder
| sex = F: IPILIMUMAB (24.0/16.0)
| sex = M: PEMBROLIZUMAB (48.0/31.0)
Efecto = Adrenal insufficiency
| sex = F: ATEZOLIZUMAB (123.0/88.0)
| sex = M: NIVOLUMAB (271.0/190.0)
Efecto = Adrenocorticotrophic hormone deficiency: IPILIMUMAB (136.0/90.0)
Efecto = Alanine aminotransferase increased: ATEZOLIZUMAB (225.0/146.0)
Efecto = Alopecia
| sex = F: PEMBROLIZUMAB (73.0/47.0)
| sex = M: NIVOLUMAB (28.0/17.0)
Efecto = Altered state of consciousness
| sex = F: PEMBROLIZUMAB (21.0/14.0)
| sex = M: IPILIMUMAB (59.0/38.0)
Efecto = Anaemia: ATEZOLIZUMAB (543.0/317.0)
Efecto = Anaphylactic reaction: NIVOLUMAB (57.0/30.0)
Efecto = Anxiety: NIVOLUMAB (72.0/24.0)
Efecto = Arthralgia
| sex = F: PEMBROLIZUMAB (200.0/86.0)
| sex = M: NIVOLUMAB (183.0/108.0)
Efecto = Arthritis
| sex = F: PEMBROLIZUMAB (39.0/25.0)
| sex = M: NIVOLUMAB (51.0/26.0)
Efecto = Ascites: ATEZOLIZUMAB (219.0/99.0)
Efecto = Aspartate aminotransferase increased: ATEZOLIZUMAB (217.0/142.0)
Efecto = Asthenia
| sex = F: PEMBROLIZUMAB (212.0/115.0)
| sex = M: NIVOLUMAB (283.0/137.0)
Efecto = Atrial fibrillation
| sex = F: PEMBROLIZUMAB (30.0/18.0)
| sex = M: NIVOLUMAB (111.0/63.0)
Efecto = Autoimmune disorder
| sex = F: PEMBROLIZUMAB (13.0/5.0)
| sex = M: NIVOLUMAB (17.0/8.0)
Efecto = Autoimmune hepatitis: NIVOLUMAB (59.0/35.0)
Efecto = Back pain: NIVOLUMAB (203.0/105.0)
Efecto = Blister: NIVOLUMAB (74.0/25.0)
Efecto = Blood alkaline phosphatase increased: ATEZOLIZUMAB (40.0/21.0)

Efecto = Blood bilirubin increased: ATEZOLIZUMAB (72.0/36.0)
 Efecto = Blood creatine phosphokinase increased
 | sex = F: PEMBROLIZUMAB (23.0/12.0)
 | sex = M: NIVOLUMAB (55.0/38.0)
 Efecto = Blood creatinine increased
 | sex = F: PEMBROLIZUMAB (50.0/28.0)
 | sex = M: ATEZOLIZUMAB (98.0/62.0)
 Efecto = Blood glucose increased: PEMBROLIZUMAB (62.0/33.0)
 Efecto = Blood pressure decreased
 | sex = F: PEMBROLIZUMAB (30.0/14.0)
 | sex = M: NIVOLUMAB (38.0/18.0)
 Efecto = Blood pressure increased
 | sex = F: PEMBROLIZUMAB (118.0/37.0)
 | sex = M: NIVOLUMAB (94.0/33.0)
 Efecto = Blood thyroid stimulating hormone increased: PEMBROLIZUMAB (64.0/22.0)
 Efecto = Brain oedema: NIVOLUMAB (51.0/33.0)
 Efecto = Cardiac arrest
 | sex = F: PEMBROLIZUMAB (24.0/14.0)
 | sex = M: NIVOLUMAB (41.0/26.0)
 Efecto = Cardiac disorder
 | sex = F: PEMBROLIZUMAB (20.0/10.0)
 | sex = M: NIVOLUMAB (27.0/11.0)
 Efecto = Cardiac failure
 | sex = F: PEMBROLIZUMAB (78.0/40.0)
 | sex = M: ATEZOLIZUMAB (123.0/87.0)
 Efecto = Cellulitis: NIVOLUMAB (69.0/44.0)
 Efecto = Cerebral haemorrhage
 | sex = F: PEMBROLIZUMAB (19.0/6.0)
 | sex = M: ATEZOLIZUMAB (37.0/21.0)
 Efecto = Cerebral infarction
 | sex = F: PEMBROLIZUMAB (28.0/14.0)
 | sex = M: NIVOLUMAB (65.0/44.0)
 Efecto = Cerebrovascular accident: NIVOLUMAB (91.0/46.0)
 Efecto = Chest pain
 | sex = F: PEMBROLIZUMAB (43.0/25.0)
 | sex = M: NIVOLUMAB (74.0/35.0)
 Efecto = Chills
 | sex = F: PEMBROLIZUMAB (42.0/20.0)
 | sex = M: NIVOLUMAB (73.0/45.0)
 Efecto = Cholangitis
 | sex = F: NIVOLUMAB (37.0/25.0)
 | sex = M: PEMBROLIZUMAB (46.0/27.0)
 Efecto = Cholangitis sclerosing
 | sex = F: PEMBROLIZUMAB (15.0/8.0)
 | sex = M: NIVOLUMAB (33.0/20.0)
 Efecto = Cholecystitis: PEMBROLIZUMAB (61.0/37.0)
 Efecto = Chronic obstructive pulmonary disease: NIVOLUMAB (48.0/26.0)
 Efecto = Colitis
 | sex = F: PEMBROLIZUMAB (154.0/101.0)
 | sex = M: NIVOLUMAB (210.0/132.0)
 Efecto = Colitis ulcerative
 | sex = F: PEMBROLIZUMAB (14.0/7.0)
 | sex = M: NIVOLUMAB (50.0/27.0)
 Efecto = Condition aggravated: NIVOLUMAB (163.0/66.0)
 Efecto = Confusional state
 | sex = F: PEMBROLIZUMAB (47.0/26.0)
 | sex = M: NIVOLUMAB (62.0/32.0)
 Efecto = Constipation: NIVOLUMAB (380.0/214.0)
 Efecto = Contusion: NIVOLUMAB (47.0/13.0)
 Efecto = Cough
 | sex = F: PEMBROLIZUMAB (99.0/53.0)
 | sex = M: NIVOLUMAB (124.0/63.0)
 Efecto = COVID-19: NIVOLUMAB (291.0/176.0)
 Efecto = C-reactive protein increased
 | sex = F: PEMBROLIZUMAB (44.0/22.0)
 | sex = M: IPILIMUMAB (58.0/39.0)
 Efecto = Cytokine release syndrome: IPILIMUMAB (130.0/76.0)
 Efecto = Cytopenia
 | sex = F: ATEZOLIZUMAB (19.0/10.0)
 | sex = M: PEMBROLIZUMAB (28.0/13.0)
 Efecto = Decreased appetite
 | sex = F: PEMBROLIZUMAB (308.0/173.0)
 | sex = M: NIVOLUMAB (503.0/281.0)
 Efecto = Deep vein thrombosis: NIVOLUMAB (79.0/49.0)

Efecto = Dehydration
 | sex = F: PEMBROLIZUMAB (93.0/55.0)
 | sex = M: NIVOLUMAB (133.0/81.0)
 Efecto = Delirium: NIVOLUMAB (47.0/30.0)
 Efecto = Depression: NIVOLUMAB (65.0/25.0)
 Efecto = Dermatitis: NIVOLUMAB (46.0/27.0)
 Efecto = Diabetes mellitus
 | sex = F: PEMBROLIZUMAB (51.0/29.0)
 | sex = M: NIVOLUMAB (83.0/44.0)
 Efecto = Diabetic ketoacidosis: NIVOLUMAB (154.0/93.0)
 Efecto = Diarrhoea
 | sex = F: PEMBROLIZUMAB (506.0/250.0)
 | sex = M: NIVOLUMAB (612.0/294.0)
 Efecto = Discomfort: NIVOLUMAB (52.0/9.0)
 Efecto = Disseminated intravascular coagulation
 | sex = F: NIVOLUMAB (31.0/20.0)
 | sex = M: IPILIMUMAB (53.0/37.0)
 Efecto = Dizziness: NIVOLUMAB (175.0/99.0)
 Efecto = Drug eruption
 | sex = F: PEMBROLIZUMAB (39.0/22.0)
 | sex = M: IPILIMUMAB (67.0/46.0)
 Efecto = Drug intolerance: PEMBROLIZUMAB (48.0/23.0)
 Efecto = Drug-induced liver injury: NIVOLUMAB (148.0/109.0)
 Efecto = Dry mouth: NIVOLUMAB (108.0/40.0)
 Efecto = Dry skin: NIVOLUMAB (95.0/31.0)
 Efecto = Dysgeusia
 | sex = F: PEMBROLIZUMAB (20.0/7.0)
 | sex = M: NIVOLUMAB (27.0/11.0)
 Efecto = Dyspepsia: NIVOLUMAB (80.0/29.0)
 Efecto = Dysphagia
 | sex = F: PEMBROLIZUMAB (47.0/27.0)
 | sex = M: NIVOLUMAB (82.0/41.0)
 Efecto = Dysphonia
 | sex = F: PEMBROLIZUMAB (93.0/18.0)
 | sex = M: NIVOLUMAB (83.0/49.0)
 Efecto = Dyspnoea: NIVOLUMAB (517.0/308.0)
 Efecto = Embolism
 | sex = F: ATEZOLIZUMAB (25.0/13.0)
 | sex = M: NIVOLUMAB (23.0/8.0)
 Efecto = Encephalitis: ATEZOLIZUMAB (72.0/50.0)
 Efecto = Encephalopathy: NIVOLUMAB (69.0/40.0)
 Efecto = Enterocolitis: NIVOLUMAB (124.0/79.0)
 Efecto = Epistaxis
 | sex = F: ATEZOLIZUMAB (23.0/10.0)
 | sex = M: NIVOLUMAB (40.0/25.0)
 Efecto = Erythema: NIVOLUMAB (137.0/79.0)
 Efecto = Erythema multiforme
 | sex = F: ATEZOLIZUMAB (35.0/22.0)
 | sex = M: IPILIMUMAB (65.0/41.0)
 Efecto = Eyelid ptosis: PEMBROLIZUMAB (45.0/20.0)
 Efecto = Facial paralysis
 | sex = F: PEMBROLIZUMAB (19.0/12.0)
 | sex = M: NIVOLUMAB (26.0/15.0)
 Efecto = Fall: NIVOLUMAB (169.0/81.0)
 Efecto = Fatigue
 | sex = F: PEMBROLIZUMAB (448.0/211.0)
 | sex = M: NIVOLUMAB (486.0/216.0)
 Efecto = Febrile neutropenia: ATEZOLIZUMAB (395.0/225.0)
 Efecto = Feeling abnormal: PEMBROLIZUMAB (71.0/26.0)
 Efecto = Fulminant type 1 diabetes mellitus: NIVOLUMAB (135.0/65.0)
 Efecto = Gait disturbance
 | sex = F: PEMBROLIZUMAB (41.0/16.0)
 | sex = M: NIVOLUMAB (75.0/40.0)
 Efecto = Gait inability
 | sex = F: PEMBROLIZUMAB (23.0/13.0)
 | sex = M: NIVOLUMAB (28.0/17.0)
 Efecto = Gastritis: NIVOLUMAB (44.0/29.0)
 Efecto = Gastrointestinal disorder: NIVOLUMAB (66.0/37.0)
 Efecto = Gastrointestinal haemorrhage: NIVOLUMAB (70.0/42.0)
 Efecto = Gastrointestinal perforation: PEMBROLIZUMAB (40.0/13.0)
 Efecto = General physical health deterioration: NIVOLUMAB (317.0/175.0)
 Efecto = Guillain-Barre syndrome: PEMBROLIZUMAB (50.0/34.0)
 Efecto = Haematochezia: NIVOLUMAB (63.0/36.0)
 Efecto = Haematuria

| sex = F: PEMBROLIZUMAB (28.0/13.0)
 | sex = M: NIVOLUMAB (42.0/30.0)
 Efecto = Haemoglobin decreased: PEMBROLIZUMAB (86.0/46.0)
 Efecto = Haemophagocytic lymphohistiocytosis: ATEZOLIZUMAB (76.0/53.0)
 Efecto = Haemoptysis: ATEZOLIZUMAB (51.0/34.0)
 Efecto = Haemorrhage: NIVOLUMAB (65.0/36.0)
 Efecto = Headache
 | sex = F: PEMBROLIZUMAB (167.0/81.0)
 | sex = M: NIVOLUMAB (121.0/59.0)
 Efecto = Heart rate increased: NIVOLUMAB (49.0/25.0)
 Efecto = Hepatic cytolysis
 | sex = F: PEMBROLIZUMAB (37.0/19.0)
 | sex = M: NIVOLUMAB (51.0/28.0)
 Efecto = Hepatic enzyme increased
 | sex = F: PEMBROLIZUMAB (66.0/34.0)
 | sex = M: NIVOLUMAB (54.0/32.0)
 Efecto = Hepatic failure
 | sex = F: PEMBROLIZUMAB (40.0/24.0)
 | sex = M: ATEZOLIZUMAB (83.0/52.0)
 Efecto = Hepatic function abnormal
 | sex = F: PEMBROLIZUMAB (133.0/83.0)
 | sex = M: NIVOLUMAB (275.0/203.0)
 Efecto = Hepatitis
 | sex = F: PEMBROLIZUMAB (85.0/53.0)
 | sex = M: NIVOLUMAB (72.0/46.0)
 Efecto = Hepatotoxicity: PEMBROLIZUMAB (92.0/46.0)
 Efecto = Herpes zoster
 | sex = F: PEMBROLIZUMAB (23.0/14.0)
 | sex = M: NIVOLUMAB (23.0/13.0)
 Efecto = Hypercalcaemia: NIVOLUMAB (54.0/32.0)
 Efecto = Hyperglycaemia
 | sex = F: ATEZOLIZUMAB (44.0/26.0)
 | sex = M: NIVOLUMAB (78.0/45.0)
 Efecto = Hyperkalaemia: ATEZOLIZUMAB (91.0/48.0)
 Efecto = Hypersensitivity: NIVOLUMAB (72.0/42.0)
 Efecto = Hypertension: PEMBROLIZUMAB (597.0/241.0)
 Efecto = Hyperthyroidism
 | sex = F: PEMBROLIZUMAB (112.0/60.0)
 | sex = M: NIVOLUMAB (99.0/61.0)
 Efecto = Hypoaesthesia
 | sex = F: PEMBROLIZUMAB (57.0/38.0)
 | sex = M: NIVOLUMAB (50.0/23.0)
 Efecto = Hypoalbuminaemia: ATEZOLIZUMAB (59.0/26.0)
 Efecto = Hypoglycaemia: NIVOLUMAB (44.0/23.0)
 Efecto = Hypokalaemia
 | sex = F: ATEZOLIZUMAB (107.0/54.0)
 | sex = M: NIVOLUMAB (65.0/39.0)
 Efecto = Hyponatraemia: ATEZOLIZUMAB (223.0/146.0)
 Efecto = Hypophysitis: NIVOLUMAB (114.0/61.0)
 Efecto = Hypopituitarism
 | sex = F: NIVOLUMAB (40.0/25.0)
 | sex = M: IPILIMUMAB (108.0/59.0)
 Efecto = Hypotension: NIVOLUMAB (171.0/98.0)
 Efecto = Hypothalamo-pituitary disorder: NIVOLUMAB (52.0/33.0)
 Efecto = Hypothyroidism
 | sex = F: PEMBROLIZUMAB (291.0/111.0)
 | sex = M: NIVOLUMAB (300.0/201.0)
 Efecto = Hypoxia
 | sex = F: ATEZOLIZUMAB (50.0/29.0)
 | sex = M: NIVOLUMAB (48.0/30.0)
 Efecto = Ileus: NIVOLUMAB (52.0/33.0)
 Efecto = Ill-defined disorder: PEMBROLIZUMAB (69.0/7.0)
 Efecto = Illness
 | sex = F: PEMBROLIZUMAB (39.0/15.0)
 | sex = M: NIVOLUMAB (46.0/26.0)
 Efecto = Immune thrombocytopenia: PEMBROLIZUMAB (64.0/41.0)
 Efecto = Immune-mediated adrenal insufficiency
 | sex = F: PEMBROLIZUMAB (19.0/10.0)
 | sex = M: IPILIMUMAB (39.0/25.0)
 Efecto = Immune-mediated dermatitis
 | sex = F: PEMBROLIZUMAB (46.0/21.0)
 | sex = M: IPILIMUMAB (85.0/48.0)
 Efecto = Immune-mediated encephalitis
 | sex = F: PEMBROLIZUMAB (18.0/9.0)

| sex = M: NIVOLUMAB (27.0/18.0)
 Efecto = Immune-mediated enterocolitis: NIVOLUMAB (361.0/200.0)
 Efecto = Immune-mediated hepatic disorder: PEMBROLIZUMAB (301.0/133.0)
 Efecto = Immune-mediated hepatitis: NIVOLUMAB (129.0/71.0)
 Efecto = Immune-mediated hypothyroidism: PEMBROLIZUMAB (136.0/38.0)
 Efecto = Immune-mediated lung disease: NIVOLUMAB (182.0/116.0)
 Efecto = Immune-mediated myocarditis
 | sex = F: PEMBROLIZUMAB (55.0/19.0)
 | sex = M: NIVOLUMAB (107.0/61.0)
 Efecto = Immune-mediated myositis: NIVOLUMAB (84.0/49.0)
 Efecto = Immune-mediated pancreatitis: PEMBROLIZUMAB (46.0/23.0)
 Efecto = Immune-mediated renal disorder: PEMBROLIZUMAB (57.0/19.0)
 Efecto = Immune-mediated thyroiditis
 | sex = F: PEMBROLIZUMAB (24.0/13.0)
 | sex = M: NIVOLUMAB (28.0/14.0)
 Efecto = Infection
 | sex = F: PEMBROLIZUMAB (54.0/33.0)
 | sex = M: NIVOLUMAB (62.0/42.0)
 Efecto = Inflammation: PEMBROLIZUMAB (58.0/30.0)
 Efecto = Infusion related reaction: NIVOLUMAB (115.0/79.0)
 Efecto = Insomnia
 | sex = F: PEMBROLIZUMAB (45.0/18.0)
 | sex = M: NIVOLUMAB (57.0/23.0)
 Efecto = Interstitial lung disease: ATEZOLIZUMAB (516.0/373.0)
 Efecto = Intestinal obstruction: NIVOLUMAB (88.0/36.0)
 Efecto = Jaundice
 | sex = F: PEMBROLIZUMAB (18.0/11.0)
 | sex = M: ATEZOLIZUMAB (30.0/17.0)
 Efecto = Large intestine perforation
 | sex = F: NIVOLUMAB (24.0/14.0)
 | sex = M: ATEZOLIZUMAB (21.0/13.0)
 Efecto = Leukopenia: NIVOLUMAB (92.0/41.0)
 Efecto = Liver disorder: IPILIMUMAB (280.0/189.0)
 Efecto = Liver function test increased: PEMBROLIZUMAB (45.0/29.0)
 Efecto = Loss of consciousness
 | sex = F: PEMBROLIZUMAB (26.0/12.0)
 | sex = M: NIVOLUMAB (44.0/21.0)
 Efecto = Lung disorder
 | sex = F: PEMBROLIZUMAB (25.0/17.0)
 | sex = M: NIVOLUMAB (85.0/49.0)
 Efecto = Lymphadenopathy: PEMBROLIZUMAB (60.0/26.0)
 Efecto = Lymphocyte count decreased: ATEZOLIZUMAB (92.0/21.0)
 Efecto = Malaise
 | sex = F: PEMBROLIZUMAB (249.0/96.0)
 | sex = M: NIVOLUMAB (316.0/204.0)
 Efecto = Memory impairment: NIVOLUMAB (52.0/22.0)
 Efecto = Mucosal inflammation
 | sex = F: PEMBROLIZUMAB (37.0/19.0)
 | sex = M: NIVOLUMAB (35.0/17.0)
 Efecto = Multiple organ dysfunction syndrome: NIVOLUMAB (73.0/47.0)
 Efecto = Muscle spasms
 | sex = F: PEMBROLIZUMAB (26.0/7.0)
 | sex = M: NIVOLUMAB (36.0/12.0)
 Efecto = Muscular weakness: NIVOLUMAB (172.0/118.0)
 Efecto = Myalgia: PEMBROLIZUMAB (171.0/94.0)
 Efecto = Myasthenia gravis: PEMBROLIZUMAB (165.0/105.0)
 Efecto = Myelosuppression: PEMBROLIZUMAB (275.0/136.0)
 Efecto = Myocardial infarction: NIVOLUMAB (82.0/47.0)
 Efecto = Myocarditis
 | sex = F: PEMBROLIZUMAB (86.0/56.0)
 | sex = M: NIVOLUMAB (149.0/96.0)
 Efecto = Myositis: PEMBROLIZUMAB (146.0/92.0)
 Efecto = Nasopharyngitis: NIVOLUMAB (67.0/33.0)
 Efecto = Nausea
 | sex = F: PEMBROLIZUMAB (318.0/180.0)
 | sex = M: NIVOLUMAB (339.0/160.0)
 Efecto = Nephritis: NIVOLUMAB (44.0/30.0)
 Efecto = Neuropathy peripheral: PEMBROLIZUMAB (227.0/134.0)
 Efecto = Neutropenia: PEMBROLIZUMAB (322.0/184.0)
 Efecto = Neutrophil count decreased: ATEZOLIZUMAB (422.0/243.0)
 Efecto = Non-small cell lung cancer: IPILIMUMAB (92.0/37.0)
 Efecto = Oedema: PEMBROLIZUMAB (65.0/39.0)
 Efecto = Oedema peripheral
 | sex = F: PEMBROLIZUMAB (56.0/30.0)

| sex = M: NIVOLUMAB (89.0/52.0)
 Efecto = Oesophagitis: NIVOLUMAB (46.0/23.0)
 Efecto = Oral pain
 | sex = F: PEMBROLIZUMAB (24.0/10.0)
 | sex = M: NIVOLUMAB (23.0/6.0)
 Efecto = Oxygen saturation decreased
 | sex = F: PEMBROLIZUMAB (25.0/15.0)
 | sex = M: NIVOLUMAB (47.0/26.0)
 Efecto = Pain
 | sex = F: PEMBROLIZUMAB (127.0/73.0)
 | sex = M: NIVOLUMAB (149.0/80.0)
 Efecto = Pain in extremity
 | sex = F: PEMBROLIZUMAB (87.0/47.0)
 | sex = M: NIVOLUMAB (84.0/48.0)
 Efecto = Palmar-plantar erythrodysesthesia syndrome
 | sex = F: PEMBROLIZUMAB (100.0/23.0)
 | sex = M: NIVOLUMAB (100.0/52.0)
 Efecto = Pancreatitis
 | sex = F: ATEZOLIZUMAB (39.0/25.0)
 | sex = M: NIVOLUMAB (50.0/34.0)
 Efecto = Pancytopenia
 | sex = F: ATEZOLIZUMAB (61.0/36.0)
 | sex = M: PEMBROLIZUMAB (64.0/42.0)
 Efecto = Paraesthesia: NIVOLUMAB (73.0/34.0)
 Efecto = Pemphigoid: NIVOLUMAB (130.0/55.0)
 Efecto = Pericardial effusion
 | sex = F: PEMBROLIZUMAB (38.0/21.0)
 | sex = M: NIVOLUMAB (56.0/28.0)
 Efecto = Peripheral sensory neuropathy: ATEZOLIZUMAB (45.0/18.0)
 Efecto = Peripheral swelling: PEMBROLIZUMAB (104.0/56.0)
 Efecto = Platelet count decreased
 | sex = F: PEMBROLIZUMAB (265.0/115.0)
 | sex = M: ATEZOLIZUMAB (235.0/152.0)
 Efecto = Pleural effusion
 | sex = F: PEMBROLIZUMAB (89.0/52.0)
 | sex = M: NIVOLUMAB (149.0/101.0)
 Efecto = Pneumocystis jirovecii pneumonia: PEMBROLIZUMAB (50.0/32.0)
 Efecto = Pneumonia: NIVOLUMAB (574.0/362.0)
 Efecto = Pneumonia aspiration
 | sex = F: PEMBROLIZUMAB (15.0/8.0)
 | sex = M: NIVOLUMAB (92.0/61.0)
 Efecto = Pneumonia bacterial: NIVOLUMAB (65.0/44.0)
 Efecto = Pneumonitis
 | sex = F: ATEZOLIZUMAB (116.0/85.0)
 | sex = M: NIVOLUMAB (250.0/171.0)
 Efecto = Pneumothorax
 | sex = F: PEMBROLIZUMAB (26.0/13.0)
 | sex = M: NIVOLUMAB (58.0/38.0)
 Efecto = Productive cough: NIVOLUMAB (48.0/25.0)
 Efecto = Proteinuria
 | sex = F: PEMBROLIZUMAB (103.0/30.0)
 | sex = M: ATEZOLIZUMAB (96.0/35.0)
 Efecto = Pruritus: NIVOLUMAB (376.0/203.0)
 Efecto = Psoriasis: PEMBROLIZUMAB (63.0/37.0)
 Efecto = Pulmonary embolism: NIVOLUMAB (168.0/103.0)
 Efecto = Pulmonary oedema: PEMBROLIZUMAB (53.0/37.0)
 Efecto = Pulmonary toxicity
 | sex = F: ATEZOLIZUMAB (23.0/14.0)
 | sex = M: NIVOLUMAB (126.0/93.0)
 Efecto = Pyrexia
 | sex = F: PEMBROLIZUMAB (423.0/266.0)
 | sex = M: NIVOLUMAB (533.0/348.0)
 Efecto = Radiation pneumonitis: DURVALUMAB (141.0/60.0)
 Efecto = Rash
 | sex = F: PEMBROLIZUMAB (321.0/181.0)
 | sex = M: NIVOLUMAB (456.0/283.0)
 Efecto = Rash maculo-papular
 | sex = F: ATEZOLIZUMAB (55.0/28.0)
 | sex = M: NIVOLUMAB (53.0/37.0)
 Efecto = Rash pruritic
 | sex = F: PEMBROLIZUMAB (50.0/21.0)
 | sex = M: NIVOLUMAB (54.0/29.0)
 Efecto = Renal disorder: NIVOLUMAB (110.0/73.0)
 Efecto = Renal failure

| sex = F: PEMBROLIZUMAB (55.0/31.0)
 | sex = M: NIVOLUMAB (115.0/66.0)
 Efecto = Renal impairment: PEMBROLIZUMAB (305.0/124.0)
 Efecto = Respiratory failure
 | sex = F: PEMBROLIZUMAB (68.0/46.0)
 | sex = M: NIVOLUMAB (117.0/81.0)
 Efecto = Rhabdomyolysis: NIVOLUMAB (56.0/39.0)
 Efecto = Rheumatoid arthritis
 | sex = F: PEMBROLIZUMAB (41.0/17.0)
 | sex = M: NIVOLUMAB (39.0/22.0)
 Efecto = Secondary adrenocortical insufficiency: NIVOLUMAB (93.0/56.0)
 Efecto = Seizure: NIVOLUMAB (105.0/64.0)
 Efecto = Sepsis: NIVOLUMAB (299.0/189.0)
 Efecto = Septic shock
 | sex = F: PEMBROLIZUMAB (34.0/21.0)
 | sex = M: NIVOLUMAB (72.0/46.0)
 Efecto = Skin disorder
 | sex = F: PEMBROLIZUMAB (52.0/23.0)
 | sex = M: IPILIMUMAB (90.0/60.0)
 Efecto = Skin exfoliation: NIVOLUMAB (58.0/23.0)
 Efecto = Skin toxicity: PEMBROLIZUMAB (50.0/25.0)
 Efecto = Skin ulcer: NIVOLUMAB (51.0/20.0)
 Efecto = Somnolence
 | sex = F: PEMBROLIZUMAB (48.0/20.0)
 | sex = M: NIVOLUMAB (75.0/22.0)
 Efecto = Stevens-Johnson syndrome
 | sex = F: NIVOLUMAB (28.0/18.0)
 | sex = M: PEMBROLIZUMAB (47.0/32.0)
 Efecto = Stomatitis
 | sex = F: PEMBROLIZUMAB (137.0/63.0)
 | sex = M: NIVOLUMAB (128.0/76.0)
 Efecto = Syncope
 | sex = F: ATEZOLIZUMAB (38.0/23.0)
 | sex = M: NIVOLUMAB (44.0/28.0)
 Efecto = Tachycardia: NIVOLUMAB (48.0/28.0)
 Efecto = Taste disorder: NIVOLUMAB (140.0/48.0)
 Efecto = Thrombocytopenia
 | sex = F: PEMBROLIZUMAB (154.0/71.0)
 | sex = M: NIVOLUMAB (159.0/100.0)
 Efecto = Thrombosis
 | sex = F: PEMBROLIZUMAB (29.0/15.0)
 | sex = M: NIVOLUMAB (41.0/20.0)
 Efecto = Thyroid disorder
 | sex = F: PEMBROLIZUMAB (75.0/31.0)
 | sex = M: NIVOLUMAB (75.0/47.0)
 Efecto = Thyroiditis
 | sex = F: PEMBROLIZUMAB (51.0/24.0)
 | sex = M: NIVOLUMAB (68.0/37.0)
 Efecto = Toxic epidermal necrolysis: PEMBROLIZUMAB (54.0/32.0)
 Efecto = Toxicity to various agents: PEMBROLIZUMAB (65.0/30.0)
 Efecto = Transaminases increased: PEMBROLIZUMAB (45.0/28.0)
 Efecto = Tremor
 | sex = F: PEMBROLIZUMAB (27.0/10.0)
 | sex = M: NIVOLUMAB (30.0/15.0)
 Efecto = Tubulointerstitial nephritis: PEMBROLIZUMAB (117.0/74.0)
 Efecto = Tumour haemorrhage
 | sex = F: PEMBROLIZUMAB (11.0/6.0)
 | sex = M: ATEZOLIZUMAB (37.0/20.0)
 Efecto = Type 1 diabetes mellitus
 | sex = F: PEMBROLIZUMAB (58.0/38.0)
 | sex = M: NIVOLUMAB (93.0/58.0)
 Efecto = Type 2 diabetes mellitus
 | sex = F: PEMBROLIZUMAB (14.0/5.0)
 | sex = M: NIVOLUMAB (34.0/24.0)
 Efecto = Urinary tract infection: ATEZOLIZUMAB (212.0/129.0)
 Efecto = Urticaria
 | sex = F: PEMBROLIZUMAB (39.0/18.0)
 | sex = M: NIVOLUMAB (35.0/21.0)
 Efecto = Uveitis: NIVOLUMAB (69.0/29.0)
 Efecto = Vision blurred
 | sex = F: PEMBROLIZUMAB (32.0/17.0)
 | sex = M: NIVOLUMAB (31.0/17.0)
 Efecto = Visual impairment: NIVOLUMAB (46.0/20.0)
 Efecto = Vitiligo: PEMBROLIZUMAB (49.0/24.0)

Efecto = Vomiting
 | sex = F: PEMBROLIZUMAB (205.0/130.0)
 | sex = M: NIVOLUMAB (210.0/105.0)
 Efecto = Weight decreased
 | sex = F: PEMBROLIZUMAB (152.0/67.0)
 | sex = M: NIVOLUMAB (242.0/86.0)
 Efecto = Weight increased: NIVOLUMAB (82.0/41.0)
 Efecto = White blood cell count decreased: ATEZOLIZUMAB (314.0/187.0)

Number of Leaves : 383

Size of the tree : 506

Time taken to build model: 0.12 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	17470	42.984 %
Incorrectly Classified Instances	23173	57.016 %
Kappa statistic	0.1873	
Mean absolute error	0.1963	
Root mean squared error	0.3151	
Relative absolute error	92.4471 %	
Root relative squared error	96.7169 %	
Total Number of Instances	40643	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,634	0,439	0,422	0,634	0,507	0,184	0,644	0,468	NIVOLUMAB
	0,093	0,025	0,329	0,093	0,146	0,123	0,689	0,230	IPILIMUMAB
	0,507	0,250	0,469	0,507	0,487	0,251	0,672	0,485	PEMBROLIZUMAB
	0,259	0,100	0,382	0,259	0,309	0,187	0,672	0,334	ATEZOLIZUMAB
	0,000	0,000	?	0,000	?	?	0,659	0,035	AVELUMAB
	0,056	0,002	0,555	0,056	0,102	0,169	0,697	0,131	DURVALUMAB
	0,000	0,000	?	0,000	?	?	0,579	0,006	CEMPIPLIMAB
Weighted Avg.	0,430	0,246	?	0,430	?	?	0,665	0,400	

=== Confusion Matrix ===

a	b	c	d	e	f	g	<-- classified as
8671	407	3371	1198	0	22	0	a = NIVOLUMAB
2528	436	1086	598	0	16	0	b = IPILIMUMAB
4763	224	6257	1084	0	8	0	c = PEMBROLIZUMAB
3438	225	2099	2025	0	19	0	d = ATEZOLIZUMAB
351	8	159	62	0	0	0	e = AVELUMAB
679	19	342	321	0	81	0	f = DURVALUMAB
94	7	31	14	0	0	0	g = CEMPIPLIMAB