



**UNIVERSIDAD
DE ANTIOQUIA**

**MODELO PREDICTIVO PARA LA CLASIFICACIÓN DE DONANTES DE
SANGRE EN EL BANCO DE SANGRE DE LA ESCUELA DE MICROBIOLOGÍA
DE LA UNIVERSIDAD DE ANTIOQUIA, MEDELLÍN. 2019**

Autor

Laura María Rodríguez Arango

**UNIVERSIDAD DE ANTIOQUIA
FACULTAD NACIONAL DE SALUD PÚBLICA
MEDELLIN, COLOMBIA**

2019



Modelo predictivo para la clasificación de donantes de sangre en el Banco de Sangre de la Escuela de Microbiología de la Universidad de Antioquia, Medellín.

2019

Laura María Rodríguez Arango

Trabajo de Grado para optar el Título de

GERENTE DE SISTEMAS DE INFORMACIÓN EN SALUD

Asesor

Carlos Alberto Tangarife Villa

Universidad De Antioquia

Facultad Nacional De Salud Pública

“Héctor Abad Gomez”

Medellin, Colombia

2019

Agradezco a el profesor Carlos Alberto Tangarife Villa por su apoyo continuo; al personal administrativo que con su gestión permitió la revisión y evaluación oportuna de este trabajo; a mis padres, esposo e hijo por brindarme el ánimo y tranquilidad necesaria, y a todos aquellos altos mandos de Consultoría en Gestión de Riesgo-Suramericana S.A, por otorgarme el tiempo necesario para la culminación efectiva de este proyecto.

Tabla de Contenido

Lista de tablas	5
Memoria de figuras	5
Glosario	6
Resumen	7
1. Introducción	8
2. Planteamiento del problema y justificación	9
3. Objetivos	13
3.1 General.....	13
3.2 Específicos.....	13
4. Marco Teórico	14
4.1 Marco conceptual.....	14
4.1.1 Donación de sangre.....	14
4.1.2 Machine Learning- ML (Aprendizaje automático de máquinas).....	26
4.1.3 Técnicas de machine learning.....	34
4.2 Marco referencial.....	41
4.3 Marco metodológico.....	47
3.3.1 Colectar Datos.....	47
3.3.2 Preprocesar los datos.....	47
3.3.3 Elegir el modelo/ crear modelo.....	48
3.3.4 Entrenar nuestra máquina.....	48
3.3.5 Evaluación.....	49
3.3.6 Afinación de parámetros (parameter tuning).....	49
3.3.7 Predicción o Inferencia y uso del modelo.....	49
4.4 Marco legal.....	50
5 Consideraciones éticas	52
6 Metodología	55
7. Resultados	61
8. Conclusión	64
9. Recomendaciones	65
10. Bibliografía	67

Lista de tablas

Tabla 1 Los estudios sociales y psicológicos que investigan los conductores de las donaciones de sangre (19).....	46
Tabla 2 Recursos documentales para el manejo y control en los Bancos de Sangre	51
Tabla 3. Variables contenidas en base de datos entregada por el Banco de Sangre de la Escuela de Microbiología de la Universidad de Antioquia	56
Tabla 4 Variables construidas para el diseño del modelo predictivo de donantes de sangre.	60
Tabla 5 Exactitud de los Modelos implementados de acuerdo al set de evaluación	63

Memoria de figuras

Ilustración 1 Proceso estandarizado para donar sangre.....	20
Ilustración 2 Etapa generica para llevar a cabo un modelo de Machine Learning.....	47
Ilustración 3 Registros excluidos de la base de datos según el motivo de omisión	57

Memoria de Gráficos

Grafico 1 Distribución donantes por edad	
Grafico 2 Distribución donantes por sexo	61
Grafico 3 Distribución donantes por año	62
Grafico 4 Donantes por tipo de sangre y RH	
Grafico 5 Donantes por clasificación de frecuencia.....	62
Grafico 6 Algoritmos usados en la modelación	63

Glosario

BANCO DE SANGRE: Es todo establecimiento o dependencia con Licencia Sanitaria de Funcionamiento para adelantar actividades relacionadas con la obtención, procesamiento y almacenamiento de sangre humana destinada a la transfusión de la sangre total o en componentes separados, a procedimientos de aféresis y a otros procedimientos preventivos, terapéuticos y de investigación. Tiene como uno de sus propósitos asegurar la calidad de la sangre y sus derivados. (1)

DONACIÓN DE SANGRE: Consiste en aceptar una extracción de la propia sangre, por parte de personal médico, de forma voluntaria sin exigir ninguna compensación monetaria. (2)

DONANTE VOLUNTARIO: Persona que dona sangre por su propia voluntad, con el deseo de ayudar y no recibe pago por ello, en efectivo o de otra forma que sea considerada como sustituto del dinero. (3)

SANGRE: Tejido líquido que recorre el organismo, a través de los vasos sanguíneos transportando las células necesarias para llevar a cabo las funciones vitales (respirar, formar sustancias, defenderse de agresiones). La cantidad de sangre de una persona está en relación con su edad, peso, sexo y altura. Una persona adulta tiene entre 4,5 y 6 litros de sangre, es decir, un 7% de su peso corporal. (4)

TRANSFUSIÓN: Es la transferencia de sangre o componentes sanguíneos de un sujeto (donante) a otro (receptor). (5)

Resumen

Machine Learning es una disciplina de la Inteligencia Artificial que crea sistemas que aprenden automáticamente identificando patrones complejos en millones de datos, lo que es importante en esta década donde día a día abundan los datos. Este estudio tiene como objetivo elaborar un modelo de aprendizaje automático o de máquinas (machine learning -ML) para la clasificación de donantes de sangre voluntarios en donante frecuente o no frecuente, del Banco de Sangre de la Escuela de Microbiología de la Universidad de Antioquia.

Para ello se tuvo en cuenta una base de datos extraída del sistema de información de registro de donantes de Sangre de dicho banco y se aplicaron varios tipos de algoritmos (modelos) con el fin de experimentar y finalmente escoger uno o varios de acuerdo a su rendimiento o ajuste.

Entre los algoritmos implementados, los que mejor se ajustaron con los datos de Test fueron: Regresión logística, k-Nearest Neighbors, Gaussian Naive Bayes- GNB y Maquinas de soporte de vector -SVM; estos se ajustan todos en un 93%.

No obstante, también encontrará conceptos necesarios para el entendimiento y comprensión del experimento.

1. Introducción

La sangre es un recurso público y componente de uso terapéutico para muchos eventos, de ahí la importancia que exista sangre suficiente, lo más segura posible, asequible y oportuna, con el fin de satisfacer las necesidades de los pacientes que la requieran.

Actualmente en Colombia los bancos de sangre públicos, privados y hospitalarios atraviesan una situación crítica, puesto que los niveles de donación son bajos.

En Medellín, específicamente en el Banco de Sangre de la Escuela de Microbiología de la Universidad de Antioquia ubicado en la Clínica León XIII, se hace necesario la construcción de un modelo predictivo que clasifique de manera inteligente a individuos que cuenten con unas características específicas como factores influyentes para la donación efectiva de la sangre, que aporte a la eficiencia del sistema y a la planificación de campañas de promoción de la donación voluntaria y repetitiva. En necesario porque allí no existe un modelo que permita clasificar los donantes en frecuentes o no frecuentes.

El año en el que más se presentaron donaciones fue en el 2013, sin embargo, el año que más tuvo donantes fue el 2014; es decir; aunque el 2013 ocupó el segundo lugar en número de donantes de todos los periodos evaluados, fue el que obtuvo mayor número de donaciones, por consiguiente, se infiere que los donantes de ese año hicieron más número de donaciones. El tipo de Sangre y RH que más donantes obtuvo en el transcurso de los 7 años evaluados fue la o+ seguido de la A+. El 84.5% de los donantes son No frecuentes, pues realizaron solo 1 donación en el periodo de los 7 años. De los seis métodos (tipos de modelo) implementados para el estudio, cuatro (Regresión, Logística, LDA; GNB y SVM) obtuvieron un ajuste del 93%.

2. Planteamiento del problema y justificación

La donación de sangre es un procedimiento médico por el cual a un individuo se le realiza una extracción de sangre ya sea para transfundir a otra persona o para elaborar medicamentos (6); en el modelo de donación colombiano solo se usa para transfundir a otras personas.

Alguno de los factores que influyen y superan los aspectos técnicos de la obtención de sangre, son: la libertad de donar o no, la motivación para donar, el reconocimiento social, comprensión del procedimiento y sus implicaciones para el donante y para la salud del posible receptor de la sangre, por ello el abastecimiento y buena parte de la seguridad de la sangre dependen de donaciones voluntarias y habituales y de una adecuada selección de los donantes que dependerá también de una serie de preguntas que buscan identificar factores de riesgo que pueden vulnerar la salud ya sea del donante o del receptor. (7)

Para medir la cantidad de donantes que hay en un lugar se usa la tasa de donación total que nos indica la cantidad de donantes que hay por cada x cantidad de personas en un lugar específico; sea un territorio país o ciudad o incluso a nivel mundial. Según información actualizada de la Cruz Roja Colombiana; la tasa de donación de sangre por cada 1000 personas en el mundo es de es de 32,6 en los países de ingresos altos, 15,1 en los de ingresos medios altos, 8,1 en los de ingresos medios bajos y 4,4 en los de ingresos bajos, en Colombia durante el año 2019 se captaron 896.107 donantes, un promedio mensual de 74.600 donantes, y a mitad del año 2020 se habían captado en promedio 56.600 donantes. Sería bueno contar con una herramienta que complemente dicha tasa suministrando la probabilidad de que una persona donante voluntaria vuelva a hacerlo

una próxima vez convirtiéndose en un donante voluntario habitual, a través de un Modelo predictivo como técnica de predicción de escenarios futuros de acuerdo a las tendencias y comportamiento de datos de ocurrencia histórica reales. Este analiza cómo intervienen los factores en el tiempo, cómo están asociados e identifica cuales factores son decisivos en el proceso de la solución de un problema (6).

Modelo predictivo es un nombre dado a una colección de técnicas matemáticas con la meta en común de encontrar una relación matemática entre un objetivo, respuesta o variable dependiente y varios factores de predicción o variables independientes, con el objetivo de medir valores futuros de estos factores de predicción e insertándolos en la relación matemática para predecir valores futuros de la variable objetivo. Como estas relaciones nunca son perfectas en la práctica, es deseable dar una medición de dudeza en las predicciones. (8)

Para realizar este modelo se deben identificar aquellos factores que interceden en la predicción, se pueden agrupar en tres categorías: aquellos que tienen poca posibilidad de afectar el resultado, aquellos con cierta certeza para afectar los resultados y deben ser considerados en el modelo y los que se encuentran en el medio, que pueden influir o no en el resultado final, siendo necesario identificar a través de una serie de técnicas si deben ser incluidas en el modelo. (8)

El modelamiento predictivo se implementa a través de Machine Learning (Aprendizaje automático) que es una de las ramas de la Inteligencia artificial, en éste los algoritmos realizan predicciones al mismo tiempo que aprenden de los datos y se considera una extensión moderna del análisis predictivo. Muchos algoritmos existentes y nuevos de

Machine learning se aplican para derivar predicciones que guían las decisiones en tiempo real con menos dependencia de la intervención humana. (9)

Como parte del diseño del modelo predictivo, se realiza el diseño de base de datos necesario para almacenar la data para el modelo, este diseño incluye la creación de procesos de transformación y carga, y funciones para realizar transformaciones de la información, necesarias para poder realizar la predicción. Así también se debe proceder al diseño de la estructura de datos de la base de datos que soporte la entrada y almacenamiento de información luego de la predicción. En síntesis, en cuanto a la solución a nivel de repositorio de datos se realiza la definición estructural completa necesaria para definir como recibir la información que alimenta el modelo y como guardarla de tal manera que sea útil para el uso que se propone del modelo.(9)

Existen estudios; definidos en el marco referencial que evidencian que a nivel mundial han tratado de predecir de alguna manera la donación de sangre, estos están enfocados a un modelo de predicción basado en la intención y hasta el momento no se ha observado alguno que acoja variables que puedan predecir la probabilidad de donar o volver a hacerlo convirtiéndose en un donante voluntario habitual tomando como base la primera donación, también existen páginas web, plataformas, blogs y otros recursos que incitan a la utilización de Machine Learning para la construcción de modelos que aporten a la analítica predictiva como factor resolutivo ante situaciones inciertas como por ejemplo; no tener la certeza de que un donante de sangre puede convertirse en donante habitual.

En Medellín, específicamente en el Banco de Sangre de la Escuela de Microbiología de la Universidad de Antioquia ubicado en la Clínica León XIII, no existe en la actualidad

un modelo predictivo automatizado que permita clasificar los donantes en frecuentes o no frecuentes. Es por ello que se hace necesario el análisis de una base de datos de donantes voluntarios del Banco de sangre mencionado e identificar cuáles son las variables predictoras para lograr ajustar un modelo, es decir construir un modelo predictivo que clasifique de manera inteligente a individuos que cuenten con unas características específicas como factores influyentes para la donación efectiva de la sangre, y que aporte a la eficiencia del sistema.

Este modelo puede servir como complemento a la tasa de donación de Sangre y clasifica a una persona como donante voluntario habitual (frecuente) o no frecuente, luego de haber donado una vez, en consecuencia, ayudaría también al proceso de gestión, ya que a través del mismo se podría inferir el número de donantes o unidades de sangre que serían recogidas en un determinado periodo futuro, así mismo puede ayudar para el diseño de campañas de promoción de la donación voluntaria y repetitiva.

3. Objetivos

3.1 General

Elaborar un modelo de machine learning para la clasificación de donantes de sangre voluntarios del Banco de Sangre de la Escuela de Microbiología de la Universidad de Antioquia.

3.2 Específicos

- Elaborar el preprocesamiento de los datos de los donantes voluntarios de sangre del Banco de Sangre de la Escuela de Microbiología de la U de A la cual está ubicada en la Clínica León XIII de Medellín.
- Diseñar un modelo que clasifique los donantes de sangre voluntarios de la Escuela de Microbiología de la U de A en frecuentes o no frecuentes de acuerdo a la frecuencia con que donan.
- Desarrollar el modelo de clasificación para los donantes de sangre voluntarios de la Escuela de Microbiología de la U de A.
- Validar el modelo de clasificación de donantes de sangre voluntarios de la Escuela de Microbiología de la U de A.

4. Marco Teórico

Los siguientes marcos definen conceptos relevantes para el proyecto, normatividad vigente y referencias de estudios como elementos importantes a tener en cuenta como guía para el avance y documentación del trabajo.

4.1 Marco conceptual

A continuación, se presentan los conceptos relacionados que permiten sustentar el desarrollo del proyecto y describir y organizar las ideas más relevantes.

4.1.1 Donación de sangre.

La vida podría ser entendida como toda forma de existencia que suponga un ciclo de nacimiento, desarrollo, reproducción y muerte. La importancia de la vida reside en el don que uno recibe al nacer y que se transforma en aquello que permite, en definitiva, existir. Así, entra en cuestión la idea de cómo cuidar y proteger esa vida. Queda en claro que si no se tiene vida no se existe, por lo tanto, todo lo que se hace y realice para cuidarla será importante. (10)

Siendo el Ministerio de Salud la entidad encargada de reglamentar los aspectos relacionados con la salud pública en Colombia y específicamente con la donación de sangre, ha expedido una serie de decretos que reglamentan este servicio. Con el decreto 1571 de 1993 se realiza un gran avance dentro de la reglamentación de la donación de sangre en Colombia. Lo dispuesto en este decreto se aplica a todos los establecimientos

dedicados a la extracción, procesamiento, conservación y, transporte de sangre y sus derivados. Además, trata de reglamentar las actividades relacionadas con la obtención, donación, conservación, procesamiento, almacenamiento, transfusión y suministro de sangre humana y sus componentes.(11)

Dentro de esta norma se definen los bancos de sangre, los centro de procesamiento de plasma y sangre, un hemoderivado, entre otros; se define un donante de sangre como: “Persona que, previo el cumplimiento de los requisitos señalados en este Decreto, da sin retribución económica y a título gratuito y para fines preventivos, terapéuticos, de diagnóstico o de investigación, una porción de su sangre en forma voluntaria, libre y consciente.”(11)

También se define que “la donación de sangre es una actividad sin ánimo de lucro, será realizada con fines terapéuticos, de diagnóstico en seres humanos y para investigación científica”. La donación de sangre sólo puede hacerse en instituciones médico asistenciales, servicios de medicina transfusional y bancos de sangre que hayan obtenido licencia sanitaria de funcionamiento para tal fin, expedida por la autoridad sanitaria competente. (11)

En este decreto no se define ningún incentivo formal para los donantes de sangre, debido a que esta actividad es definida con un carácter meramente solidario, el cual no debe ser remunerado. Solo se exige que los bancos de sangre de categoría A, de origen público, deban promover y desarrollar programas y convenios tendientes a estimular la donación voluntaria y altruista de sangre. Además, se exige que las Direcciones de Salud deban desarrollar programas que incentiven la donación de sangre.(11)

Según estudios clínicos la donación de sangre es una buena alternativa para preservar la salud de la persona que recibe la transfusión, así como también la del emisor; Esta actividad permite limpiar la sangre, reduce las posibilidades de sufrir ataques cardiacos y accidentes cardiovasculares. Las personas que padecen problemas con el colesterol, pueden depurar triglicéridos afectados, permitiendo la formación de lípidos buenos. (11)

Dentro de las ventajas de la donación se encuentra la eliminación de hierro que sobra en el cuerpo, facilitando las labores del corazón y las venas, beneficiando el sistema circulatorio. Los donantes pueden prevenir enfermedades coronarias o circulatorias, si adoptan el hábito de donar sangre cada 6 meses durante 6 años. (11)

A nivel mundial se recolectan 108 millones unidades de sangre. De las cuales el 50% provienen de países de altos ingresos o desarrollados, esta cifra es alarmante, en la medida en que estos Estados solo representan el 15% de la población mundial. (11)

Países como Chile, Bolivia, Perú, y Ecuador han realizado una labor importante en la región en materias relacionadas con la donación de sangre. Tienen un papel importante en la promoción de las donaciones. En Ecuador, la promoción de los bancos de sangre arroja resultados positivos, en la medida en que el país cumple en un 100% en el suministro de sangre según una encuesta realizada por la Organización Panamericana de la Salud. Dentro de este mismo estudio Chile se destaca con un 75% de suministro de sangre, mientras que Colombia se mantiene en un 50%. Sin embargo, Perú no ha logrado destacarse en esta materia obteniendo un 20% en este estudio. Con base en estos

porcentajes, se puede concluir que la promoción de los bancos de sangre es de vital importancia para el desarrollo de las donaciones voluntarias en la región. (11)

Adicionalmente los países de la región han implementado un trabajo intersectorial con el fin de optimizar los resultados obtenidos en los bancos de sangre. Este trabajo ha ido de la mano con la sensibilización, educación y movilización de la población, hasta el punto de crear y fomentar cultura ciudadana frente a esta actividad. Además, la región latinoamericana ha buscado trabajar de manera conjunta con actores como la sociedad civil. No obstante, Perú es el país donde estos trabajos han pasado a un segundo plano. Por esta razón es de suma importancia que en Colombia se adopte medidas que incentiven a la población de la importancia de la donación de sangre. (11)

Dentro del sistema internacional, Colombia se ubica en un nivel medio, a raíz del número de donantes que poseen sus bancos de sangre. Por esta razón, es de vital importancia que se incentiven las donaciones permanentes a través de medidas ajenas al dinero. (10)

Las donaciones de sangre tienen la virtud de salvar vidas y preservar la salud de las personas. Estas transfusiones pueden ser necesitadas en cualquier momento, por lo que el país debe contar con un banco suficiente para responder a emergencias de este tipo. Zonas urbanas como rurales deben estar dotadas en esta materia, con la finalidad de evitar muertes, quebrantos y sufrimientos en los pacientes. (11)

En Colombia para el año 2017 se captaron 830.291 unidades de sangre total a partir de 998.773 donantes potenciales, lo que representa un incremento en la captación del 1,6% respecto a 2016 a pesar de que el número potencial de donantes se redujo en 0,1%. Según las proyecciones de crecimiento demográfico publicadas por el DANE se esperaba un

aumento vegetativo de la población entre 15 - 64 años del 1,2% (32.141.917 habitantes en 2016 vs 32.540.011 en 2017), por lo que se infiere que se elevó el número de donantes repetitivos en 2017. (8)

En el año 2002 se crea el Plan Nacional de Sangre (PNS) dependiente del Ministerio de Salud de la Nación con el propósito de organizar el Sistema Nacional de Sangre (SNS) y los objetivos de mejorar la Seguridad Transfusional y alcanzar la Autosuficiencia en componentes y derivados de la sangre.(12)

El Plan Nacional de Sangre ha impactado en un área sensible de la Salud Pública instalando al recurso sangre humana como una política de Estado con una presencia permanente en el Consejo Federal de Salud (COFESA) donde intervienen, además del Ministro de Salud de la Nación, todos los ministros de las veinticuatro jurisdicciones.(12)

La salud del donante está protegida porque la entrevista previa a la donación permite seleccionar claramente quiénes no pueden hacerlo porque pueden verse afectados. Es el caso, por ejemplo, de quienes tengan afecciones cardíacas, respiratorias, neurológicas, renales o hepáticas. La salud del receptor está protegida porque es un médico o profesional de la salud el que decide si el donante está en condiciones de donar o no. (12)

A todos los donantes potenciales se les realiza una entrevista confidencial y, sobre la base de las respuestas de una guía muy detallada, el profesional de acuerdo al criterio médico y las normas vigentes decidirá si el donante es apto para ofrecer su sangre o si debe ser diferido temporaria o definitivamente. Se consulta al donante acerca de sus hábitos y vida personal, viajes, vida sexual, utilización de drogas o medicamentos. Esto se debe a que

ciertas situaciones pudieron haberlo expuesto al riesgo de contraer infecciones graves transmisibles por sangre. Estos requisitos se establecen con un trabajo exhaustivo de las asociaciones científicas nacionales y mundiales y están en continua actualización. (12)

Cuando una persona se acerca a donar sangre es muy importante que reciba una atención integral que contemple, no sólo cada uno de los aspectos relativos al proceso de la donación de sangre en sí mismo, sino todo lo que significa entrar en contacto con un servicio de salud. Esto quiere decir que la donación de sangre lo va a vincular a un control general de su estado de salud y, a su vez, que si en ese control se detecta alguna situación que requiere atención o cuidado, será derivado a un centro adecuado. La donación de sangre es un proceso que implica diversos pasos. Estos tienen una explicación y una secuencia establecida, especialmente pensada para reforzar la seguridad del donante y del receptor. Esos pasos se pueden sintetizar en la **Ilustración 1 Proceso estandarizado para donar sangre** (12)

La “salud social” debe lograrse en un esfuerzo mancomunado de todos los actores. Al Estado le corresponde una firme presencia para cumplir con su responsabilidad indelegable de asegurar la equidad y la eficacia social, orientando y conduciendo con inteligencia, capacidad y transparencia.(12)

Numerosas investigaciones y publicaciones demuestran, a nivel nacional y mundial, que la donación más segura es aquella aportada por personas no coaccionadas, que eligen libre y voluntariamente donar su sangre sin esperar que los llamen y sin esperar nada a cambio. Estos documentos enfatizan la necesidad de identificar y elaborar estrategias que

promuevan este tipo de donaciones ya que los donantes voluntarios y habituales garantizan la mayor seguridad transfusional. (12)



Ilustración 1 Proceso estandarizado para donar sangre

En el contexto regional, la Organización Panamericana de la Salud (OPS), en trabajo conjunto con los Ministerios de Salud de la Región, identificó las brechas que es necesario eliminar en los países para mejorar la disponibilidad, la seguridad, la calidad y el uso de la sangre para transfusiones. El 41° Consejo Directivo de la OPS, celebrado en 1999, adoptó la Resolución CD 41 R15, que insta a los Estados Miembro a que “promuevan el desarrollo de Programas Nacionales de Sangre con base en la donación voluntaria, altruista y repetida de sangre como uno de los indicadores de desarrollo humano de la población y de la garantía de la calidad”. (12)

4.1.1.1 Tipos de Donantes de Sangre.

La transfusión es la única opción para las personas enfermas que requieren sangre; por eso, la donación es tan necesaria en el sistema sanitario. Existen diferentes perfiles de donantes: donantes voluntarios y habituales, donantes de reposición y dadores remunerados.(12)

Donante voluntario.

La donación voluntaria de sangre es importante porque sus componentes y derivados son elementos insustituibles dado que no pueden ser reproducidos de manera artificial. La sangre no se puede “fabricar” en un laboratorio: los seres humanos dependemos de nosotros mismos para ayudarnos. Si bien es probable que hayamos escuchado acerca de los avances de la ciencia en relación con la producción de sangre artificial, esto está muy lejos de la realidad. Los componentes de la sangre (glóbulos rojos, glóbulos blancos, plaquetas y plasma) cumplen funciones tan específicas que, hasta el momento, es imposible fabricarlos. Los donantes voluntarios y habituales son aquellos que donan su sangre de manera anónima, no remunerada y periódica, sin que nadie se lo solicite, sin responder a la necesidad de un paciente específico. Los donantes de reposición son aquellos que donan solo en respuesta al pedido de un familiar, amigo o conocido que lo necesita. (12)

Donante remunerado.

Los dadores remunerados son aquellas personas que venden su sangre. No se consideran donantes, dado que su acción no es motivada por un fin altruista. El pago a dadores de sangre está prohibido por la Ley de Sangre. (12)

Donante Voluntario habitual, repetitiva / de repetición.

Es la que se obtiene de una persona que dona sangre o componentes por lo menos 2 veces en los últimos 12 meses, por su propia voluntad y no recibe ningún pago, ya sea en forma de dinero en efectivo o en especie que podría ser considerado un sustituto del dinero. (8)

Donante voluntario convocado.

Donante voluntario, que por sus características fenotípicas e inmunohematológicas es llamado por un banco de sangre a donar para un paciente que dadas sus características fenotípicas requiere esta donación por ser la más compatible para él. (8)

Donante Dirigido.

Persona que dona sangre para un receptor determinado; ejemplo: madre que dona plaquetas para su hijo, donante de riñón que además dona su sangre para el receptor del órgano. La sangre debe llegar a la persona para la cual se donó y si este no es el caso se deben considerar como de reposición. (8)

Donante dador por reposición.

Dona sangre o componentes, previo proceso de sensibilización y educación, cuando es requerida por un miembro de su familia o comunidad sin que haya exigencia de por medio para la prestación del servicio. (8)

4.1.1.2 Tipos de procedimiento.

La donación de sangre es un acto realizado bajo la responsabilidad de un médico o profesional de la salud y esto es para proteger la salud del donante y la salud del receptor. El profesional que actúa debe tener certeza de que el donante está en pleno estado de salud pues, de este modo, la donación no le ocasionará ningún inconveniente y el receptor recibirá el componente de sangre “más seguro”, que lo ayudará a recuperar su salud.

Recordemos que la sangre tiene tres componentes que son útiles para ser transfundidos a los pacientes: los glóbulos rojos, las plaquetas y el plasma. La posibilidad de separar estos elementos, llamados “hemocomponentes”, posibilita que una misma unidad de sangre pueda tener tres receptores diferentes. Por eso se dice que con cada donación de sangre se salvan tres vidas: a cada persona se le transfunde sólo el elemento que necesita.(12)

Existen varios procedimientos que se realizan en los Bancos de Sangre, algunos de ellos son:

Aféresis.

Es el procedimiento mediante el cual se extrae sangre de un donante con el objeto de obtener uno de sus componentes, reinfundiéndole el resto de los componentes no separados. (1)

Leucoféresis.

Es el procedimiento mediante el cual se extrae del donante; sangre total, con el objeto de obtener concentrado de leucocitos, con o sin plaquetas, y reinfundirle los glóbulos rojos y el plasma no utilizado, con o sin plaquetas. (1)

Flebotomía terapéutica.

Es el procedimiento mediante el cual se extrae sangre a un paciente con el objeto de reducir el exceso de eritrocitos. (1)

Transfusión sanguínea.

Es el procedimiento por medio del cual, previa formulación médica y practicadas las pruebas de compatibilidad a que haya lugar, se le aplica sangre total o alguno de sus componentes a un paciente con fines terapéuticos o preventivos. (1)

Autotransfusión o transfusión autóloga.

Es un procedimiento mediante el cual se transfunde a una persona la sangre total o los componentes que previamente haya donado para tal fin. (1)

4.1.1.3 Requisitos para Donar Sangre.

El artículo 28, capítulo 3 del Decreto 1571 de 1993, expone los requisitos a tener en cuenta para donar Sangre. Por ser la salud un bien de interés público, donar sangre es un deber de solidaridad social que tienen las personas y, por ningún motivo, podrá ser remunerado. (1)

Para la donación deberán acreditarse los siguientes requisitos y condiciones:

- a) Ser mayor de 18 años y menor de 65 años.
- b) Que el acto de donación sea consciente, expreso y voluntario por parte del donante.
- c) Que tenga un peso mínimo de 50 kilogramos.

- d) Practicar valoración física con el fin de verificar: que la temperatura, la presión arterial y el pulso se encuentren dentro de rasgos normales.
 - e) Ausencia de signos, síntomas o antecedentes de enfermedades infecciosas que se transmitan por vía transfusional.
 - f) Ausencia de embarazo.
 - g) Que no haya donado sangre total durante un lapso no menos de cuatro (4) meses.
 - h) Que previa determinación posea valores de hemoglobina y de hematocrito dentro de los rasgos normales.
 - i) Que no haya recibido dentro del último año, transfusiones de sangre o de sus componentes.
 - j) No haber sido vacunado dentro de los 15 días anteriores a la donación.
 - k) No estar utilizando medicamentos contraindicados para la donación señalados en el manual de normas técnicas que expida el Ministerio de Salud.
 - l) Ausencia de signos, síntomas o antecedentes de alcoholismo, drogadicción, de enfermedades infecciosas transfusionales, así como también de enfermedades crónicas o degenerativas que comprometan la salud del donante y/o del receptor, establecidos por interrogatorio y/o por examen físico.
 - m) Y otras que determine el Manual de Normas Técnicas que expida el Ministerio de Salud.
- (1)

4.1.1.4 Frecuencia para donar Sangre de acuerdo a la edad y el sexo

Es importante tener en cuenta las siguientes observaciones en cuanto a la frecuencia para donar sangre de acuerdo al sexo y edad

La edad mínima para donar sangre es de 18 años y la máxima de 65 años.

- Aquellas personas que se encuentran entre 18 y 30 años

Mujeres: Pueden donar cada 6 meses, es decir, 2 veces al año.

Hombres: Pueden donar cada 4 meses, es decir, 3 veces al año

- Los mayores de 30 años y hasta 65 años

Mujeres: Pueden donar cada 4 meses, es decir 3 veces al año, y

Hombres: Pueden donar cada 3 meses, es decir, 4 veces al año.

(8)

4.1.2 Machine Learning- ML (Aprendizaje automático de máquinas).

Se considera, principalmente, un subcampo de la Inteligencia Artificial e incluso de la Informática, desde algunas perspectivas. ML tiene conceptos que han sido derivados y tomados de múltiples campos desde hace ya un largo periodo de tiempo, lo que lo convierte en un verdadero campo multidisciplinar o interdisciplinar. Los principales dominios o campos asociados incluyen: Inteligencia artificial, Procesamiento Natural del Lenguaje, Minería de Datos, Matemáticas, Estadística y las Ciencias de la Computación (13).

Tom Mitchell en su libro “Machine Learning” cubre el campo del aprendizaje automático como el estudio de algoritmos que permiten que los programas de computadora mejoren automáticamente a través de la experiencia, es decir, algoritmos que aprenden y mejoran

“solos” gracias a la experiencia. El libro está destinado a apoyar los cursos de pregrado de nivel superior y de posgrado de nivel introductorio en aprendizaje automático.

Machine Learning es la disciplina científica, rama de la inteligencia artificial que es capaz de crear sistemas que se adaptan a través de la experiencia, es decir, de reconocer los errores y rectificarlos sin estar explícitamente programado.(13)

Se dice que un programa de computadora aprende de la experiencia E con respecto a alguna clase de tareas T y medida de rendimiento P si su desempeño en las tareas en T medido por P mejora con la experiencia E . (14)

La Inteligencia Artificial es un campo de estudio que busca estimular el comportamiento inteligente en términos de procesos computacionales. Es entonces un sistema computarizado que exhibe un comportamiento que comúnmente se considera que requiere inteligencia, es la ciencia de hacer que las máquinas hagan cosas que requerirían inteligencia si las hiciera el hombre. El padre fundador de la IA, Alan Turing, define esta disciplina como: la ciencia y la ingeniería de la fabricación de máquinas inteligentes, especialmente programas informáticos inteligentes. En estas definiciones, el concepto de inteligencia se refiere a algún tipo de capacidad para planificar, razonar y aprender, sentir y construir algún tipo de percepción del conocimiento y comunicarse en el lenguaje natural.

Christopher Bishop en su libro “Reconocimiento de patrones y Aprendizaje automático” realiza un trabajo bastante bueno al proporcionar implícitamente una definición centrada en el algoritmo del aprendizaje automático (clasificación de dígitos).

El resultado la ejecución del algoritmo de aprendizaje automático se puede expresar como una función $y(x)$ que toma una nueva imagen de dígito x como entrada y que genera un vector de salida y , codificados de la misma manera que los vectores objetivo. La forma precisa de la función $y(x)$ se determina durante la fase de entrenamiento también conocida como aprendizaje sobre la base de los datos de entrenamiento. Una vez que el modelo está entrenado, puede determinar la identidad de las nuevas imágenes de dígitos, que se dice que comprenden un conjunto de prueba. La capacidad de categorizar correctamente los ejemplos nuevos que difieren de los utilizados para la capacitación se conoce como generalización. En aplicaciones prácticas, la variabilidad de los vectores de entrada será tal que los datos de entrenamiento pueden abarcar solo una pequeña fracción de todos los posibles vectores de entrada, por lo que la generalización es un objetivo central en el reconocimiento de patrones. (15)

Las aplicaciones de minería de datos pueden identificar tendencias y comportamientos, no sólo para extraer información, sino también para descubrir las relaciones en bases de datos que pueden identificar comportamientos que no son muy evidentes tales como asociaciones, secuencias, clasificaciones, agrupaciones y pronósticos. Los algoritmos de minería de datos son utilizados por las técnicas de aprendizaje automático o Machine Learning y se convierten en un apalancador del desarrollo y la transformación organizacional, debido a que mediante estas técnicas es posible detectar automáticamente patrones de información sobre procesos productivos, desempeño operacional, mercados y clientes de las empresas que a diario genera datos en grandes cantidades. El éxito de los negocios depende por lo general de la habilidad para ver nuevas tendencias o cambios en las tendencias. (13)

El aprendizaje automático proporciona conocimientos de los datos que no se podrían obtener con soluciones de Procesamiento Analítico en Línea-OLAP, debido a que permite explotar grandes bases de datos, para encontrar de manera automatizada patrones y relaciones e inferir reglas a partir de ellos para predecir el comportamiento a futuro; estos patrones y reglas se utilizan para orientar la toma de decisiones y predecir su efecto. (13)

4.1.2.1 Métodos /Algoritmos de Machine Learning.

Por lo general, los métodos de Machine Learning se pueden clasificar de múltiples formas bajo diferentes características. Seguidamente se detallan algunas de las clasificaciones:

- 1) Métodos basados en la cantidad de supervisión humana en el proceso de aprendizaje: (Aprendizaje supervisado, Aprendizaje no supervisado, Aprendizaje semi-supervisado y Aprendizaje reforzado)
2. Métodos basados en la capacidad de aprender a partir de muestras de datos incrementales: (Aprendizaje por lotes y Aprendizaje en línea) y
- 3) Métodos basados en su enfoque a la generalización a partir de muestras de datos: (Aprendizaje basado en instancias y Aprendizaje basado en modelos), seguidamente se describen los cuatro métodos basados en la cantidad de supervisión humana en el proceso de aprendizaje. (13)

Métodos de aprendizaje supervisado.

Los métodos de Aprendizaje Supervisados incluyen algoritmos de aprendizaje que toman muestras de datos (Datos de Entrenamiento) y salidas asociadas (Categorías o

Respuestas) con cada muestra de datos durante el proceso de entrenamiento de dicho modelo. (13)

Este conocimiento aprendido se puede usar en el futuro para predecir una salida, con cualquier nueva muestra de datos de entrada X que previamente no se conocía o que no se había visto durante el proceso de entrenamiento del modelo. Estos métodos se denominan supervisados porque el modelo aprende en muestras de datos donde las respuestas/etiquetas de salida ya se conocen de antemano en la fase de entrenamiento. Los métodos de aprendizaje supervisado son principalmente de dos clases, según el tipo de tareas de ML que pretendan resolver pueden ser de clasificación o de regresión. (13)

La clasificación es una subcategoría de aprendizaje supervisado donde el objetivo es predecir las etiquetas de clases categóricas de las nuevas observaciones, basada en observaciones pasadas. Estas etiquetas de clase son valores discretos que pueden entenderse como pertenencia a grupos definidos. (13)

Un segundo tipo de aprendizaje supervisado es la predicción de resultados continuos, que también se denomina Análisis de Regresión, en dicho Análisis dada una serie de variables predictoras (explicativas) y una variable de respuesta continua, se requiere encontrar una relación entre esas variables que nos permita predecir un resultado. Es decir, utilizan atributos o características de datos de entrada y sus correspondientes valores de salida numérica continua (también llamados respuesta), para aprender relaciones y/o asociaciones específicas entre las entradas y sus salidas correspondientes. Con este modelo, se puede predecir respuestas para observaciones nuevas no conocidas, proceso que es similar a la clasificación, pero con salidas numéricas continuas. (16)

Métodos de aprendizaje no supervisado.

Están más relacionados con tratar de extraer información valiosa de los datos, en lugar de tratar de predecir algún resultado a partir de la información de los mismos a través de clasificaciones supervisadas previamente disponibles.

Hay más incertidumbre en los resultados del aprendizaje no supervisado que en los supervisados, sin embargo, en los No supervisados se puede obtener mucha información que anteriormente no estaba disponible.

Los métodos de Aprendizaje No Supervisados se pueden categorizar en las siguientes áreas: Agrupación, Reducción de dimensionalidad, Detección de anomalías y Asociación.

(16)

Los métodos de agrupación. Son métodos de aprendizaje automático que intentan encontrar patrones de similitud y relaciones entre muestras de un conjunto de datos, y luego agrupan estas muestras en varios grupos, de modo que cada clúster o grupo de muestras de observaciones tiene cierta similitud, según los atributos o características inherentes.(17)

Estos métodos no están supervisados por completo, ya que intentan agrupar los datos mediante la observación de las características de los datos sin ninguna clasificación previa, supervisión o conocimiento sobre los atributos de dichas observaciones. (17)

Existen varios tipos de métodos de agrupamiento que se pueden clasificar bajo los siguientes enfoques principales.

- Métodos basados en centroides: K-means y K-medoids
- Métodos de agrupamiento jerárquico: Aglomeración y División (Ward, propagación de afinidad).
- Métodos de agrupación basados en distribuciones: Modelos de mezclas gaussianas.
- Métodos basados en densidades: dbscan y óptica. Además de esto, tenemos varios métodos que recientemente entraron en el panorama de la agrupación, como el abedul y clarans. (17)

Métodos de aprendizaje semi-supervisado.

Este método utiliza datos de entrenamiento tanto etiquetados como no etiquetados: normalmente una pequeña cantidad de datos etiquetados junto a una gran cantidad de datos no etiquetados. El aprendizaje semi-supervisado se encuentra entre el aprendizaje no supervisado (sin datos de entrenamiento etiquetados) y el aprendizaje supervisado (con todos los datos de entrenamiento etiquetados). Los investigadores del campo del aprendizaje automático han descubierto que los datos no etiquetados, cuando se utilizan junto a una pequeña cantidad de datos etiquetados, pueden mejorar de forma considerable la exactitud del aprendizaje. La adquisición de datos etiquetados para resolver un problema suele requerir un agente humano capacitado para clasificar de forma manual los ejemplos de entrenamiento. El coste asociado al proceso de etiquetado puede hacer que un conjunto de entrenamiento totalmente etiquetado sea inviable, mientras que la adquisición de datos sin etiquetar es relativamente poco costoso. En estos casos, el aprendizaje semi-supervisado puede ser muy útil. (17)

Métodos de aprendizaje reforzado.

Los métodos de Aprendizaje de Refuerzo son diferentes de los métodos convencionales supervisados o no supervisados.

En este contexto, tenemos un agente que queremos capacitar durante un periodo de tiempo para interactuar con un entorno específico y mejorar su rendimiento a lo largo del tiempo con respecto al tipo de acciones que realiza en el entorno. Normalmente, el agente comienza con un conjunto de estrategias o políticas para interactuar con el entorno. (17)

Al observar el entorno, toma una acción particular basada en una regla o política y observando el estado actual del entorno. En función de la acción, el agente obtiene una recompensa, que podría ser beneficiosa o perjudicial (en forma de penalización). Actualiza sus políticas y estrategias actuales si es necesario y este proceso iterativo continúa hasta que aprende lo suficiente sobre su entorno para obtener las recompensas deseadas. (17)

Los principales pasos de un método de aprendizaje de refuerzo se mencionan a continuación: (17)

- Preparar un agente con un conjunto de políticas iniciales y estrategia.
- Observar el entorno y el estado actual
- Seleccionar las políticas óptimas y realizar una acción
- Obtener una recompensa correspondiente (o una penalización)
- Actualizar las políticas si es necesario

- Repetir los pasos 2 a 5 de forma iterativa hasta que el agente aprenda las políticas más óptimas.

4.1.3 Técnicas de machine learning.

4.1.3.1 Regresión Logística.

La regresión logística es un tipo especial de regresión que se utiliza para explicar y predecir una variable categórica binaria, en función de varias variables independientes que a su vez puede ser cuantitativas o cualitativas. Por sus características, los modelos de regresión logística permiten dos finalidades:

- Cuantificar la importancia de la relación existente entre cada una de las covariables o variables independientes y la variable dependiente, lo que lleva implícito también clarificar la existencia de interacción y confusión entre covariables respecto a la variable dependiente (es decir, conocer la “odds ratio” para cada covariable). (17)
- Clasificar individuos dentro de las categorías (presente/ausente) de la variable dependiente, según la probabilidad que tenga de pertenecer a una de ellas dada la presencia de determinadas covariables. No cabe duda que la regresión logística es una de las herramientas estadísticas con mejor capacidad para el análisis de datos en investigación clínica y epidemiología, de ahí su amplia utilización.(17)

El objetivo primordial que resuelve esta técnica es el de modelar cómo influye en la Probabilidad de aparición de un suceso, habitualmente dicotómico, la presencia o no de Diversos factores y el valor o nivel de estos. También puede ser usada para estimar La probabilidad de aparición de cada una de las posibilidades de un suceso con más de dos categorías (politómico).(17)

El modelo de regresión lineal (Legendre, Gauss, Galton y Pearson) considera que, dado un conjunto de observaciones $\{Y_i, X_{i1}, \dots, X_{ip}\}_{i=1}^n$, la media μ de la variable respuesta Y se relaciona de forma lineal con la o las variables regresoras $X_1 \dots X_p$ acorde a la ecuación: $\mu_y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$

El resultado de esta ecuación se conoce como la línea de regresión poblacional, y recoge la relación entre los predictores y la media de la variable respuesta. Otra definición que se encuentra con frecuencia en los libros de estadística es: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \epsilon_i$

En este caso, se está haciendo referencia al valor de Y para una observación i concreta. El valor de una observación puntual nunca va a ser exactamente igual al promedio, de ahí que se añada el término de error ϵ_i .

En ambos casos, la interpretación de los elementos del modelo es la misma:

- β_0 : es la ordenada en el origen, se corresponde con el valor promedio de la variable respuesta Y cuando todos los predictores son cero.
- β_p : es el efecto promedio que tiene sobre la variable respuesta el incremento en una unidad de la variable predictora X_p , manteniéndose constantes el resto de las variables. Se conocen como coeficientes parciales de regresión.
- ϵ : es el residuo o error, la diferencia entre el valor observado y el estimado por el modelo. Recoge el efecto de todas aquellas variables que influyen en Y pero que no se incluyen en el modelo como predictores.

En la gran mayoría de casos, los valores β_0 y β_p poblacionales se desconocen, por lo que, a partir de una muestra, se obtienen sus estimaciones. $\hat{\beta}_0$ y $\hat{\beta}_p$. Ajustar el modelo consiste en estimar, a partir de los datos disponibles, los valores de los coeficientes de regresión que maximizan la verosimilitud (likelihood), es decir, los que dan lugar al modelo que con mayor probabilidad puede haber generado los datos observados.

El método empleado con más frecuencia es el ajuste por mínimos cuadrados ordinarios (OLS), que identifica como mejor modelo la recta (o plano si es regresión múltiple) que minimiza la suma de las desviaciones verticales entre cada dato de entrenamiento y la recta, elevadas al cuadrado.

El término "lineal" en los modelos de regresión hace referencia al hecho de que los parámetros se incorporan en la ecuación de forma lineal, no a que necesariamente la relación entre cada predictor y la variable respuesta tenga que seguir un patrón de recta.

Cuando la variable respuesta es binaria, es decir que solo son posibles dos valores; se codifica como 0 y 1, matemáticamente es posible ajustar un modelo de regresión lineal por mínimos cuadrados. Sin embargo, esta aproximación tiene dos problemas:

- Al generar una recta (hiperplano si hay múltiples variables), se pueden obtener valores predichos distintos de 0 y 1, lo que entra en contradicción con la definición de la variable respuesta binaria.
- Si se quiere interpretar las predicciones del modelo como probabilidades de pertenencia a cada clase, no se cumpliría la condición de que toda probabilidad ha de estar dentro del intervalo $[0,1]$, ya que podrían obtenerse valores fuera de este.

Para evitar estos problemas, la regresión logística (David Cox 1958) transforma el valor devuelto por la regresión lineal con una función cuyo resultado siempre está comprendido entre 0 y 1.

Existen varias funciones que cumplen esta descripción, una de las más utilizadas es la función logística (también conocida como función sigmoide): $\text{sigmoide} = \sigma(Y) = 1 / (1 + e^{-Y})$

Para valores de Y positivos muy grandes, e^{-Y} es aproximadamente 0, por lo que el valor de la función sigmoide es 1. Para valores de Y negativos muy grandes, e^{-Y} tiende a infinito, por lo que el valor de la función sigmoide es 0.

Sustituyendo la Y de la ecuación anterior por la función de un modelo lineal $y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$, Se obtiene que: $P(y=1|X=x) = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}$ donde $P(y=1|X=x)$ puede interpretarse como la probabilidad de que la variable respuesta Y adquiera el valor 1 (clase de referencia), dado los predictores X_1, X_p

El modelo resultante tiene los coeficientes de regresión en los exponentes, por lo que no es un modelo lineal y no puede ajustarse con la estrategia descrita al inicio del documento. Para evitar este inconveniente; la expresión obtenida tiene la característica de ser siempre positiva, ya que la función exponencial solo toma valores positivos y, el cociente de valores positivos es siempre positivo.

Esto hace posible aplicarle el logaritmo: $\ln\left(\frac{p(y=1|X=x)}{p(y=0|X=x)}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$

Al realizar la transformación, en el lado derecho se obtiene la ecuación de un modelo lineal. El término de la izquierda resulta ser el logaritmo de un cociente de probabilidades, lo que se conoce como razón de probabilidad (log of odds).

Como resultado de este proceso se consigue convertir un problema de clasificación no lineal, en un problema de regresión lineal que sí puede ajustarse mediante los métodos convencionales.

Una vez obtenidos los coeficientes del modelo ($\beta_0, \beta_1, \dots, \beta_p$) se puede obtener la probabilidad de que una nueva observación pertenezca a la clase $y=1$ con la ecuación: $p(y=1|X=x) = \frac{e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}}$

(18)

4.1.3.2 Árboles de Decisión.

Un árbol de decisión es una estructura en la cual cada nodo interno denota una prueba sobre uno o varios atributos, cada rama representa una salida de la prueba y los nodos hoja representan clases. (17)

La característica principal de los árboles de decisión es que son modelos de caja blanca en los cuales se puede ver directamente la frecuencia de aparición de cada atributo. Además, le permite al experto conocer el atributo con mayor poder de clasificación, es decir, aquel que se localice en el nodo raíz (Mazo, Bedoya,2010). (17)

Hay varios tipos de árboles de decisión, los más conocidos son:

- ID3 (Iterative Dichotomiser 3)
- C4.5 (successor of ID3)
- CART (Classification And Regression Tree)
- CHAID (CHi-squared Automatic Interaction Detector).

De estos tipos de árboles el más utilizado es CART ya que es un tipo de árbol de decisión binario, esto quiere decir que en cada nodo sólo se pueden dividir los datos en dos grupos. CART es capaz de manejar tanto datos numéricos como categóricos, e incluso posee sofisticados métodos para tratar con valores perdidos, como por ejemplo el uso de variables sustitutas o surrogate variables como se denomina en inglés. Otra de las características de CART es que se puede utilizar tanto en problemas de clasificación como de regresión. (17)

4.1.3.3 Máquinas de Soporte Vectorial.

Las Máquinas de Soporte Vectorial (del inglés Support Vector Machine) fueron Desarrolladas en 1995 por Vladimir Vapnik cuando propone un modelo matemático para la resolución de problemas de clasificación y regresión el cual llamaron Modelo MSV. Están basadas en la teoría de aprendizaje estadístico que permiten resolver problemas de clasificación y regresión de manera eficiente (Cuba, 2010), (Dexheimera et al. (2007), (Rengifo, Juménez, 2010). (17)

Clasificación Linealmente Separable.

Dentro de los conjuntos de datos que abarcamos con las máquinas de soporte vectorial, encontramos aquellos que son linealmente separables, y en ellos, minimizar la función de costos resulta muy sencillo. La hipótesis de partida es que las clases son linealmente separables y por ello existen infinitos hiperplanos que separan las muestras de una clase, de la otra. Los puntos del espacio que caen dentro de cada uno de estos hiperplanos son los que satisfacen la siguiente expresión (Reginfo, 2010).(17)

Clasificación Linealmente no Separable.

Cuando los datos no son linealmente separables, existe la posibilidad de transformar los datos a un espacio de mayor dimensión utilizando una función donde se encontrará un hiperplano que los pueda separar. La frontera de decisión resultante en el espacio de entrada ya no será lineal y vendrá dada por otro tipo de función que pueda ser polinómica de grado superior a 1, gaussiana, sigmoide, entre otras, dichas funciones se conocen como funciones núcleo o “Kernel”. Las muestras una vez proyectadas, pueden usarse como un nuevo conjunto de entrenamiento, de esta forma se buscará una frontera lineal en el

espacio, dicha frontera otra en el espacio de partida cuya forma dependerá de la función de proyección empleada y de las muestras del conjunto de entrenamiento. Como se nota ya los datos en la formulación de la máquina. (17)

4.1.3.4 k-Nearest Neighbors- K-NN.

El algoritmo **k-Nearest Neighbors** (vecinos más cercanos) o **k-NN** es posiblemente el algoritmo de aprendizaje automático más simple. k-NN es un ejemplo típico de un aprendiz perezoso. Se llama perezoso no debido a su aparente simplicidad, sino porque no aprende una función discriminativa de los datos de entrenamiento, sino que memoriza el conjunto de datos de entrenamiento.

Según Raschka & Mirjalili (2017), el algoritmo KNN en sí mismo es bastante sencillo y se puede resumir mediante los siguientes pasos:

- Elegir el número de k y una métrica de distancia.
- Encontrar a los k vecinos más cercanos de la muestra que queremos
- clasificar.
- Asignar la etiqueta de la clase por mayoría de votos. (17)

4.1.3.5 Gaussian Naive Bayes (GNB).

Bayes ingenuos **gaussianos** son Bayes ingenuos simples con una suposición típica de que las características continuas asociadas con cada clase se distribuyen de acuerdo con una distribución normal (o gaussiana). El clasificador Naive Bayes, un modelo generativo, utiliza la regla de Bayes para aprender $P(Y | X)$ donde $X = \langle X_1, \dots, X_n \rangle$, y lo hace haciendo una suposición de independencia condicional que reduce drásticamente el número de parámetros a estimar al modelar la probabilidad $P(X | Y)$. (17)

4.1.3.6 Análisis discriminante lineal-ADL.

Sir Ronald Fisher publicó desarrollado análisis discriminante lineal (LDA) en 1936, lo que hace que sea una de las primeras técnicas de clasificación. Adopta un enfoque diferente a las probabilidades de estimación mediante el modelado de la distribución de las características por separado en cada una de las clases de respuesta (de donante frecuente vs donante no frecuente). El modelo es muy similar en forma a la regresión logística cuando la distribución se asume normal en las características.

Sagun James, Witten et al. (2013), cuando las clases de respuesta tienen distribución heterogénea, el logit tenderá a ser inestable. LDA tiende a no tener este problema si el número de observaciones es pequeño y la distribución de características se distribuye normalmente en cada una de las clases. (19)

4.2 Marco referencial

Son muchas las aplicaciones que se encuentran de técnicas de Machine learning en la salud, este tipo de técnicas han cobrado relevancia cuando se tienen grandes volúmenes de datos y facilitan procesos tales como: diagnóstico, tratamiento, seguimiento de enfermedades, iniciativas de medicina personalizada y en procesos de promoción y prevención de la salud entre otros. Los algoritmos de Machine Learning se aprovechan de la llegada del Big Data creando sistemas que aprenden automáticamente. Dicho en otras palabras, sistemas que aprenden mediante ejemplos. (20)

Se han desarrollado proyectos para la detección de diferentes tipos de cáncer y otras enfermedades, para el descubrimiento de nuevas drogas, para clasificación del dolor y del triage, para evaluar tiempos de espera, para la asignación de recursos médicos, monitoreo de estilos de vida y el establecimiento de alertas proactivas que permitan actuar anticipadamente para evitar lesiones, accidentes o inclusive la muerte o también para advertir sobre los factores de riesgo de enfermedad en poblaciones grandes. (20)

(Prieto, 2012) present un nuevo procedimiento informático que ayuda al diagnóstico precoz de la enfermedad de Alzheimer mediante la clasificación automática de imágenes de resonancia magnética nuclear de pacientes sanos, con deterioro cognitivo leve y pacientes que sufren la enfermedad de Alzheimer, los resultados obtenidos en la clasificación de imágenes ofrecen una sensibilidad y especificidad del 98,7 por ciento para la clasificación entre personas sanas y con enfermedad de Alzheimer, y del 80 y 96 por ciento, respectivamente, cuando se incluye en la clasificación a sujetos con deterioro cognitivo leve.

(Wub, Zheng, Chen, 2011) reportan la utilización de técnicas de inducción de reglas y técnicas de análisis de grupos en el servicio de Urgencias del hospital de Taiwán, donde analizaron la asignación del personal médico, la clasificación que le han realizado a los pacientes, así como los costos generados en la sala de urgencia, a fin de identificar si estaban aplicando bien los recursos asignados. (21)

Por ejemplo, en el MIT se han estudiado los patrones de comportamiento a través de teléfonos inteligentes para predecir si una persona está enferma y el tipo de problema de salud que tiene, también para detectar si una zona geográfica puede ser vulnerable a una

epidemia o para el tratamiento del estrés, la depresión y la ansiedad. En otro estudio en el que se empleó el diagnóstico asistido por computadora (CAD) para revisar las mamografías iniciales de mujeres que desarrollaron posteriormente cáncer de mama, la computadora detectó el 52% de los cánceres hasta un año antes de que las mujeres fueran diagnosticadas oficialmente, lo que deja evidencia de la eficiencia de este tipo de técnicas. (19). En Costa de Marfil y América Latina se han implementado algoritmos de aprendizaje automático para construir mapas de pobreza que sirvan de soporte a la política pública con base en la información de telefonía celular con niveles de precisión de más de 70%. (22)

Es evidente que el auge de ML permite que se aumente diariamente su aplicación en salud, y es por esta razón que existen áreas donde se ha comprobado su eficiencia y por lo tanto uno de los retos principales es complementar los modelos existentes para aumentar la precisión de las predicciones en las áreas en las que se ha logrado identificar un mejor desempeño es en las tareas de pronóstico, interpretación de imágenes y en la precisión diagnóstica donde mediante su uso se puede reducir el error humano.(23)

Eamonn Ferguson en su artículo “Predictores del comportamiento futuro: una revisión de la literatura psicológica sobre la donación de sangre”, publicado en 1996 explora la eficacia relativa de diferentes modelos teóricos para predecir el comportamiento futuro de la donación de sangre. El comportamiento estudiado fue la cantidad de donaciones de sangre futuras. La intencionalidad, a partir de la teoría de la acción planificada / razonada, surgió como el mejor predictor del comportamiento futuro de los donantes, pero pareció ofrecer muy poco en la forma de sugerir intervenciones. El poder predictivo de la

intencionalidad se redujo a medida que aumentaba el intervalo de tiempo entre su medición y el registro del comportamiento real del donante. (24)

Un estudio publicado en Pubmed en el año 2014, realizado en la Universidad de Vasco, en España tenía como objetivo predecir las intenciones de donación de sangre y el comportamiento entre los donantes de sangre Españoles, probando una teoría extendida del modelo de comportamiento planificado. Este estudio planteó un problema importante para las agencias de recolección de sangre de todo el mundo. Investigaciones previas que utilizaron un enfoque de teoría aumentada del comportamiento planificado (TPB) demostraron que la actitud, la norma subjetiva, la autoeficacia, la norma moral, el remordimiento anticipado, la ansiedad por donaciones de donaciones de sangre previas y la auto identidad como donante de sangre predicen donantes experimentados. Las intenciones, la autoeficacia, la norma moral y el arrepentimiento anticipado pueden influir en el comportamiento real de donación de sangre de las personas. (25)

Zuckerman, M publicó en la Revista de Personalidad y Psicología Social el artículo “Comparación de tres modelos para predecir el comportamiento altruista” y demostró que el comportamiento de donación se predijo mejor como una función de intenciones y actitudes, mientras que las intenciones se describieron mejor como una función de las actitudes, las normas sociales y las normas morales. La comparación de los resultados con los de varios estudios previos sugirió que la combinación específica de variables que mejor predecía el comportamiento altruista dependía de las condiciones particulares bajo las cuales se formaban las predicciones (intervalo de tiempo, creencias de la persona y la observación de su comportamiento). (26)

En el año 2013 es publicado un artículo de Doggenweiler I, titulado “Determinantes psicosociales de la intención de donación de órganos en una muestra chilena” se relaciona un estudio construye un modelo predictivo de las intenciones de realizar conductas de donación, basado en la adaptación y extensión de la teoría de la acción razonada (TRA). Este modelo teórico se ha utilizado en múltiples trabajos en los que se busca a los determinantes de las conductas de donación de órganos. La TRA es un modelo utilizado para la comprensión y predicción de la conducta individual y voluntaria, a la cual se le atribuye como su principal y más directo determinante la intención. (27)

El Grupo AIA de Barcelona- España dispone de una experiencia de más de 15 años en el desarrollo de modelos predictivos en distintos sectores y actualmente está aplicando esta experiencia en el sector sanitario desarrollando modelos predictivos que permitan una mejor gestión de los pacientes crónicos. (28)

Existen páginas web que estimulan a la construcción de modelos predictivos a través de competencias de Machine Learning como lo es el caso de UCI- Machine Learning repository (Centro de aprendizaje automático y sistemas inteligentes) en donde se presenta un caso de estudio de aplicación de Data Analytics para la predicción del número de donantes de sangre en un determinado periodo de tiempo, en un escenario concreto. Para el análisis se toman los datos de donaciones de una competición de DrivenData, en los que se recoge, de manera sencilla, algunas características de los donantes, indicando además si fueron a donar sangre durante un mes en concreto (marzo de 2007). Dichos datos provienen de Taiwán, concretamente de encuestas que se realizan dentro de un vehículo para donar sangre que recorre distintas universidades.(28)

En consecuencia los autores Deepti Bahel, Prerana Ghosh, Arundhyoti Sarkar, Matthew A. Lanham en un estudio llamado “Predicting Blood Donations Using Machine Learning Techniques” ; centraron la construcción de un sistema basado en datos para el seguimiento y la predicción de potenciales donantes de sangre, investigando el uso de diversas técnicas de clasificación binaria para estimar la probabilidad de que una persona donara sangre o no en marzo de 2007 teniendo en cuenta la base de su comportamiento en donaciones pasadas. (19)

Estos autores realizaron una revisión de estudios similares que se relacionan a continuación:

Autores	Métodos	Datos	Conclusión
Godin, conner et al. 2007	Regresión Logística	Encuesta (2070 donantes frecuentes, 161 donantes nuevos.	Se observa que en los donantes frecuentes; la intención, el control percibido, el arrepentimiento anticipado, moral, edad influyen en la frecuencia de donación. En los donantes nuevos; solo la intención y la edad.
Sojka y Sojka 2008	Estadísticas descriptivas	Encuesta (531 participantes)	Se encontró que los motivadores generales para donar sangre eran: Influencia de amigos (47.2%) Solicitudes de los medio (23.5%), y aquellos para donaciones continuas o frecuentes; altruismo (40.3%), responsabilidad social (17.7%) e influencia de amigos (17.9%)
Masser, Bednall et al. 2012	Modelos de ecuaciones estructurales	Encuesta (263 participantes de Encuesta de seguimiento (182 donantes)	Se encontró que las normas morales, ansiedad por la donación y la identidad del donante predijeron indirectamente la intención a través de la actitud.
Masser, Bednall et al. 2012	Análisis de ruta	Encuesta (256 participantes)	Su modelo extendido de la teoría del comportamiento planificado- TPB mostró que la intención fue predicha por las actitudes, el control percibido y la auto identificación

Tabla 1 Los estudios sociales y psicológicos que investigan los conductores de las donaciones de sangre (19)

3.3 Marco metodológico

Se tienen en cuenta los siguientes pasos para la construcción del modelo

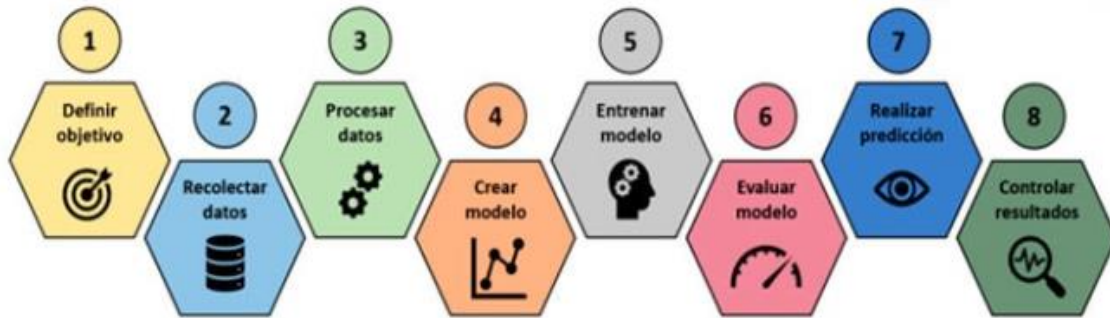


Ilustración 2 Etapas para llevar a cabo un modelo de Machine Learning

3.3.1 Colectar Datos.

Los datos se obtienen de fuentes de datos secundarias, se debe establecer si es necesaria la utilización de APIs para la extracción y carga de datos. (29).

3.3.2 Preprocesar los datos.

La preparación de datos implica verificar la calidad de los datos disponibles, explorarlos con el fin encontrar valores atípicos y/o duplicados, reducirlos, separarlos en grupos de entrenamiento y evaluación o encontrar las características relevantes para la predicción y establecer los procedimientos de transformación e imputación. (29)

La primera tarea dentro del preprocesamiento de los datos consiste en estandarizar los nombres de los campos (eliminando espacios, acentos y signos de puntuación entre otras

cosas) y recodificar todos los valores numéricos de la base de datos en su correspondiente valor categórico o nominal. (29)

Una vez ya eliminados aquellos campos no relevantes para el desarrollo del modelo, se procede al manejo de los valores perdidos, lo cual se realiza en dos frentes: El primero, eliminar aquellos campos de mala calidad, es decir, los que posean un gran porcentaje de registros sin información, y el segundo, eliminar/imputar aquellos registros con un rango manejable de valores perdidos. (29)

Posteriormente los datos se deben separar en dos conjuntos, el primero asociado al conjunto de entrenamiento del modelo, y un segundo conjunto asociado a la validación.

3.3.3 Elegir el modelo/ crear modelo.

Debido a que existen diferentes algoritmos en este paso se debe elegir de acuerdo con el objetivo cual se utilizará. (29)

3.3.4 Entrenar nuestra máquina.

En esta etapa de acuerdo con los métodos de aprendizaje y los seleccionar los algoritmos se debe utilizar el set de datos de entrenamiento para alimentar la máquina y mejorar incrementalmente la predicción mediante la extracción de la información relevante para la predicción. Es importante precisar que en esta se debe llevar a cabo un proceso iterativo donde después de los resultados obtenidos se debe ajustar el modelo. (29)

3.3.5 Evaluación.

Después de comprobar el conocimiento adquirido por el algoritmo se procede a evaluar su precisión en las predicciones y lo debemos comprobar de acuerdo con los datos del set de evaluación. En el caso de que la exactitud sea menor o igual al 50% debemos descartar el modelo, pero si se alcanza un 80% debemos considerar el modelo. (29)

3.3.6 Afinación de parámetros (parameter tuning).

Si durante la evaluación no se obtuvo el desempeño esperado es decir una buena tasa de aciertos en las predicciones y no se alcanza el mínimo deseado se debe proceder con el proceso de overfitting (ó underfitting) y se debe proceder nuevamente con el paso 4, mediante la reconfiguración de los parámetros o hiperparámetros del modelo. Algunos de los parámetros que se ajustarán serán: el número de iteraciones, la tasa de aprendizaje, el máximo error permitido. Es importante resaltar que cada algoritmo permite ajustar parámetros diferentes. Como por ejemplo el número de capas ocultas o el número de neuronas. (29)

3.3.7 Predicción o Inferencia y uso del modelo.

Después del proceso de afinamiento se procederá con el uso del modelo en producción con la nueva información obtenida en el sistema de información para poder predecir los resultados en el mundo real. (29)

Durante el proceso se utiliza los servicios Python Anaconda, ya que contiene una serie de herramientas para el procesamiento y análisis de datos. Se usarán los algoritmos apropiados para la construcción de modelos predictivos a través de varios algoritmos.

4.4 Marco legal

Aunque en la Constitución Política Colombiana no está contemplada de manera explícita la donación de sangre, la política nacional de sangre realizada por el Ministerio de Protección Social, que contempla la donación de sangre como un acto altruista y voluntario hace referencia más que todo a reglamentos técnicos para bancos de sangre y guías clínicas par la aplicación de sangre. Esta política nacional de sangre evita la descoordinación y descontrol en los bancos de sangre, maneja el flujo de pacientes y vigila los procedimientos de transfusión. Además indica la forma correcta en que debe realizarse una transfusión de sangre y controla que se le dé un uso adecuado a la sangre. (30)

La Ley 919 del año 2004 prohíbe la comercialización de componentes de la anatomía humana para trasplantes, y lo tipifica además como delito. Esta ley contempla en su artículo número 1 que toda persona que decida ser donante deberá hacerlo siempre por razones humanitarias y que está prohibida cualquier forma de pago por dicha sangre. Aclara que las instituciones como bancos de tejidos y de medula ósea son las que pueden cobrar exámenes necesarios del receptor y el donante, hospitalización, transporte, cuidado médico entre otras, pero por ningún motivo cobro por el fluido sanguíneo. (30)

Los Banco de Sangre cuentan con normas y reglamentaciones que propenden por vigilar el tema de la donación de sangre en bancos de sangre, EPS, IPS y ESE. Algunas de estas se encuentran relacionadas en la **Tabla 2 Recursos documentales para el manejo y control en los Bancos de Sangre.**

Documento legal	Objeto
Circular 050	La Organización Mundial de la Salud y la Organización Panamericana de la Salud definen la obtención de sangre humana como un acto de interés social y sanitario, y resaltan que la donación voluntaria no debe depender de ninguna remuneración. (30)
Decreto 1571 de 1993	Por el cual se reglamenta parcialmente el Título IX de la Ley 09 de 1979, en cuanto a funcionamiento de establecimientos dedicados a la extracción, procesamiento, conservación y transporte de sangre total o de sus hemoderivados, se crean la Red Nacional de Bancos de Sangre y el Consejo Nacional de Bancos de Sangre y se dictan otras disposiciones sobre la materia. (1)
Política Nacional de Sangre	Esta política fue formulada para responder a los problemas de Colombia en materia de sangre y componentes sanguíneos. Es por esto que el Ministerio de la Protección Social hizo público el documento de la Política Nacional de Sangre, permitiendo así la articulación de los distintos actores involucrados con el tema de donación voluntaria y altruista de sangre y transfusión de componentes sanguíneos en beneficio de la comunidad. (31)
Guía rápida para tomar decisiones en Medicina Transfuncional – Documento Técnico – INS	Es una guía para la reserva de sangre en procedimientos quirúrgicos. Hace hincapié en que las necesidades allí identificadas deben ser individualizadas dependiendo del estado clínico del paciente y de acuerdo con el juicio médico particular en cada situación. (32)
Manual de Hemovigilancia – INS – 2010.	Marco de referencia para el programa de Colombia que busca mejorar el perfil de seguridad de toda la cadena transfusional y mejorar la seguridad y esperan lograr un impacto positivo en los procesos de notificación de los eventos adversos asociados a la donación y transfusión de sangre en Colombia, al sistematizar la información derivada de ello, presentar un análisis estadístico completo e implementar medidas oportunas y eficaces que eviten la ocurrencia o recurrencia de algunos de estos eventos.(33)
Res 0901 de 1996 Manual de normas técnicas, administrativas y de procedimientos en Bancos de Sangre	Suministra al sistema de salud, la información necesaria para la toma de decisiones. Unifica criterios, procedimientos y técnicas que permiten la adecuada estructura, organización y funcionamiento de los Bancos de Sangre y Servicios de transfusión. (34)
Decreto 559 de 1991	Por el cual se reglamentan parcialmente las leyes 09 de 1979 y 10 de 1990, en cuanto a la prevención control y vigilancia de las enfermedades transmisibles, especialmente lo relacionado con la infección por el Virus de Inmunodeficiencia Humana (HIV) y el Síndrome de Inmunodeficiencia Adquirida (SIDA), y se dictan otras disposiciones sobre la materia. (34)

Tabla 2 Recursos documentales para el manejo y control en los Bancos de Sangre

5 Consideraciones éticas

De acuerdo con los principios establecidos en la Declaración de Helsinki y en la Resolución 8430 de octubre 4 de 1993: y debido a que esta investigación es considerada como investigación sin riesgo, y en cumplimiento con los aspectos mencionados con el Artículo 6 de la presente resolución, este estudio se desarrollará conforme a los siguientes criterios:

La Facultad Nacional de Salud de Pública Solicitara el permiso con los respectivos criterios al Banco de Sangre de la Escuela de Microbiología de la U de A, de esta manera se obtendrá la base de datos para poder realizar el análisis pertinente de esta investigación es de suma importancia aclarar que este estudio será conservado de forma confidencial y sólo será usada con fines académicos, no se hablará de una persona en particular.

Cuando se habla de los sistemas autónomos se debe respetar los derechos de la libertad y la dignidad humana para esto se opera de cierta manera, se sabe que la tecnología está transformando considerablemente el ámbito de las profesiones y es el principal detonante de dicha transformación. Cada vez hay más empresas que dedican sus investigaciones a la creación de una gran variedad de sistemas, máquinas, herramientas, etc., que tienen como finalidad la recopilación y almacenamiento de gran cantidad de conocimientos y habilidades que pueden mejorar o dar soluciones a problemas que afectan a los seres humanos.

Este proyecto se plantea en el sector de salud, cumpliendo con la responsabilidad social y ética al recopilar los datos de los pacientes sin atentar contra su privacidad y derecho a la protección de información: esto quiere decir que en la información recopilada de

entrenamiento y validación del modelo, no se almacenaran datos personales como nombre, apellidos, dirección, teléfonos y demás datos de dicha índole. Es así que el proyecto fue desarrollado con la confidencialidad e integridad de la información brindada por la Escuela de Microbiología de la U de A cuya base de datos es utilizada como sustento del entrenamiento y validación del modelo predictivo. Adicionalmente, el entrenamiento del modelo se realizó con un conjunto de datos aleatorio, garantizando la imparcialidad frente a los resultados obtenidos en la ejecución de la predicción.

Esta investigación cuenta con un modelo de aprendizaje automático preciso y efectivo que brinda total transparencia para todos los involucrados, cumpliendo con esta transparencia se tienen en cuenta los 7 principios para la transparencia y responsabilidad de los algoritmos los cuales serán mencionados a continuación:

Conciencia: los propietarios, diseñadores o programadores tienen que ser conscientes de las posibles consecuencias de uso y los posibles sesgos a los que se puede incurrir.

Acceso y corrección: la diseñadora del prototipo tiene que tener los mecanismos para corregir los perjuicios individuales debido a los sesgos detectados en los algoritmos.

Responsabilidad: la institución puede ser responsable de las decisiones tomadas para sus algoritmos, aunque no sean capaces de explicar el cómo se ha llegado a esas decisiones.

Explicación: especialmente cuando afecta las políticas públicas, en las investigaciones cuando se utilizan algoritmos para la toma de decisiones deben explicar los procedimientos seguidos por los algoritmos y las decisiones que toma.

Procedencia de los datos: se tiene que explicar las características de los datos de entrenamiento (de donde los sacaron como los obtuvieron), así como una exploración de los posibles sesgos del algoritmo debido al uso de estos datos. Para evitar problemas de privacidad, se puede restringir el acceso a estos datos a personas no autorizadas explícitamente.

Auditoria: hay que guardar los modelos, los algoritmos, los datos y las decisiones para poder escoger el modelo que mejor explica la investigación.

Validación y testeo: se debe hacer una validación rigurosa del modelo y documentarlo. También deben correr el test para detectar los posibles sesgos del algoritmo.

Por último es importante mencionar que hay que generar educación y conciencia sobre los riesgos potenciales del mal uso y sobre los principios del diseño que busca el bienestar humano de la sociedad.

6 Metodología

Para la elaboración del modelo de Machine Learning se siguieron los pasos descritos en el marco metodológico.

Se descargaron los registros del sistema de registro de donantes usado en la Escuela de Microbiología de la U de A, que toma como entrada la encuesta de selección de donantes.

Estos se descargaron en diferentes segmentos de datos de la siguiente manera:

- Libro de Excel con datos de donantes entre los periodos de 1 de septiembre de 2012 y 30 de septiembre de 2013
- Libro de Excel con datos de donantes entre los periodos de 1 de octubre de 2013 y 31 de diciembre de 2014
- Libro de Excel con datos de donantes entre los periodos de 1 de enero de 2015 y 31 de diciembre de 2015
- Libro de Excel con datos de donantes entre los periodos de 1 de enero de 2016 y 31 de diciembre de 2016
- Libro de Excel con datos de donantes entre los periodos de 1 de enero de 2017 y 31 de diciembre de 2017
- Libro de Excel con datos de donantes entre los periodos de 1 de enero de 2018 y 31 de diciembre de 2018
- Libro de Excel con datos de donantes entre los periodos de 1 de enero de 2019 y 30 de septiembre de 2019

Un total de 7 libros que se consolidan conformando una base de datos con 100.473 registros de donaciones.

Las Variables contenidas se relacionan en la **Tabla 3. Variables contenidas en base de datos entregada por el Banco de Sangre de la Escuela de Microbiología de la Universidad de Antioquia**

Nombre de la Variable	Descripción
TIPDOC	Tipo de Identificación del donante de sangre
NUMDOC	Número de identificación del donante de sangre
PESO	Peso en Kg del donante de sangre
ABO	Indica si el donante de sangre es tipo de sangre A, B, O, AB
RH	Indica si el RH del donante de sangre es + o -
FECHAEXTRACCION	Indica la fecha en que se dio la donación de sangre
CIUDAD	Ciudad donde vive el donante de sangre
DOMICILIO	Dirección donde reside el donante de sangre
SEXO	sexo "FEMENINO" o "MASCULINO"
EDAD	Indica la edad que tiene el donante
DEFRECUENCIA	Donante "REPETITIVO", "NO REPETITIVO" o " PRIMERA VEZ"
OCUPACION	Indica la profesión u oficio que tiene el donante de sangre
FECNAC	Indica la fecha en que nació el donante de sangre.

Tabla 3. Variables contenidas en base de datos entregada por el Banco de Sangre de la Escuela de Microbiología de la Universidad de Antioquia

Se obtiene el conjunto de datos con 100.473 registros equivalente a 75.606 donantes.

Una vez estudiadas sus variables y registros se toma la decisión de omitir/excluir algunos de sus registros dadas inconsistencias en el registro o porque no se pertenecen al periodo de tiempo estudiado. Posteriormente se analizan los datos en un gestor de base de datos con el fin de realizar exploraciones y preparar la base para el modelo.

En la **Ilustración 3 Registros excluidos de la base de datos según el motivo de omisión** se puede observar el número de registros no tenidos en cuenta en la exploración.

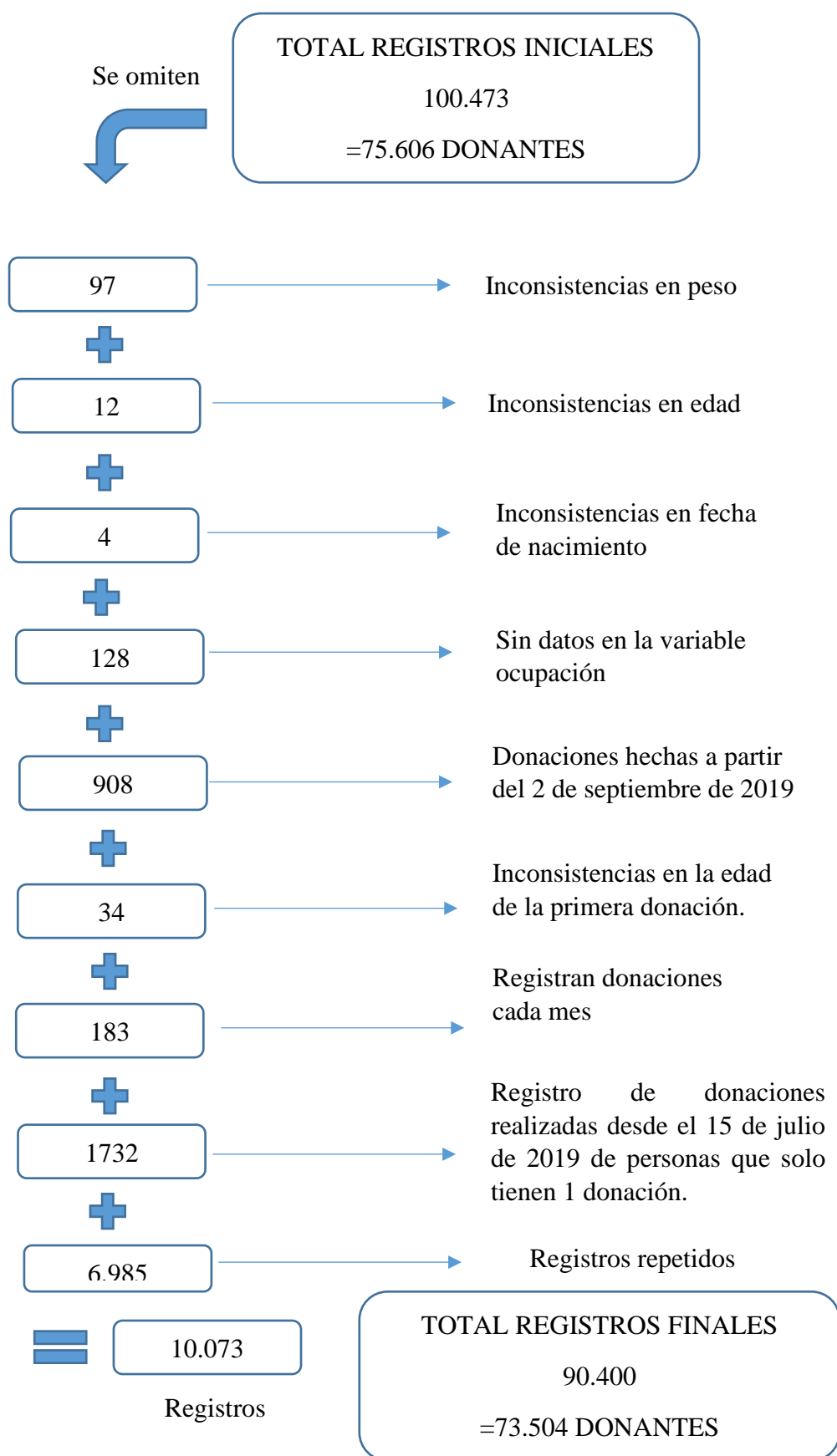


Ilustración 3 Registros excluidos de la base de datos según el motivo de omisión

Los 1732 registros de donaciones, equivalente a 1732 donantes se tienen en cuenta posteriormente para la validación del modelo que clasificará a los donantes en: Donantes Frecuentes o No frecuentes. Estos donantes (1732) no fueron tenidos en cuenta en el modelo, dado que no cumplen con el periodo transcurrido necesario para volver a donar.

A razón de que en la base de datos de donantes de sangre del Banco de Sangre de la Escuela de Microbiología existen varios registros de donaciones de la misma persona en la misma fecha en el caso de aféresis, se crea una llave conformada por las variables: Número de documento y fecha de extracción que permite el conteo por registro único de donante y única fecha.

Así obtenemos de 73.504 donantes, 90.891 donaciones.

Se construyen las variables relacionadas en **Tabla 4 Variables construidas para el diseño del modelo predictivo de donantes de sangre.**

Variable	Descripción	Se construye
peso_prom	Peso promedio del donante de sangre	Identificando a través de la función Promedio, el peso promedio registrado en n registros de donación por cada donante de sangre. Dado que no se distribuye normal se estandariza a través de la formula $Z = (x - U) / DE$
edad actual	Indica la edad actual del donante de sangre	Utilizando la función hoy así; (función hoy – fecha de nacimiento)/365. Dado que no se distribuye normal se estandariza a través de la formula $Z = (x - U) / DE$
edad_primer_donacion	Indica la edad que tenía el donante de sangre en su primera donación	Con la función SIFECHA, se calcula la edad del donante de sangre en su primera donación tomando como inicio la fecha de su primera donación y como fin el 01/09/2019.
Grupo_edad_primer_donacion	Indica si el donante tiene entre 18 y 30 años o entre 31 y 65 años	Con la función SI, se nombra como “Entre 18 y 30” aquellos donantes que su edad de primer donación es ≥ 18 y ≤ 30 , y “Entre 31 y 65” aquellos

Variable	Descripción	Se construye
		donantes que en su primera donación tenían edad >30 y <=65.
Edad_ult_donacion	Indica la edad que tenía el donante de sangre en su última donación	Con la función SIFECHA, se calcula la edad del donante de sangre en su última donación tomando como inicio la fecha de última donación y como fin el 01/09/2019. Expresado en años.
Donacion_total	Indica el número de donaciones totales para cada donante.	Se obtiene con la sumatoria de todas las donaciones realizadas
Cada_cuanto_donar_enmeses	Indica el número de meses que deben transcurrir para donar sangre de acuerdo al sexo y a la edad.	Se establece con la función SI anidada, si los criterios se cumplen de acuerdo al sexo y edad; así: mujeres >=30 años cada 4 meses, <30 cada 6 meses, hombres >=30 cada 3 meses y <30 cada 4 meses, lanza el número de meses que deben transcurrir para que la persona vuelva a donar sangre.
Fr_posible_donacion_anual	Indica el número de veces que una persona puede donar en el año, de acuerdo al sexo y a la edad.	Se calcula dividiendo 12 en Cada_cuanto_donar_enmeses
n_donesperada_aSep2019	Indica el número de donaciones esperadas desde la primera donación hasta el 01/09/2019	Se toma en cuenta un maestro construido con el número de meses evaluado (85), el sexo, grupo de edad y cada_cuanto_donar_enmeses. Se estima el cociente entre N° de mes transcurrido y cada_cuanto_donar_enmeses + 1 (que es la donación en el momento 0). Así obtendremos para cada caso el número de donaciones esperadas. Posterior a esto se construye una llave conformada por el N° de mes, sexo y grupo de edad para llevar el valor estimado a la tabla de datos que tendrá la llave de meses_primer_donacion, sexo y grupo_edad_primer_donacion.
Meses_primer_donacion	Indica el número de meses transcurridos desde la primera donación	Se estima con la función SIFECHA, tomando como inicio la fecha de la primera donación y como fin el 01/09/2019, expresado en meses
meses_ult_donacion	Indica el número de meses transcurridos desde la última donación	Se estima con la función SIFECHA, tomando como inicio la fecha de la última donación y como fin el 01/09/2019, expresado en meses
prop_donacion_aSep2019	Indica la proporción de donaciones con respecto a las esperadas a septiembre	Se calcula dividiendo las donaciones totales, sobre las donaciones esperadas al periodo evaluado

Variable	Descripción	Se construye
Promedio_dias_donacion	Indica cuantos días tarda un donante que ha tenido más de 1 donación, en volver a donar.	En un gestor de base de datos se corre la frecuencia en días que hay entre una donación y otra para cada uno de los donantes. Una vez obtenemos la suma, la dividimos entre el número de donaciones totales para obtener así la frecuencia promedio en días.
Tipo donante	Indica si un donante es NO FRECUENTE: 0 FRECUENTE: 1	Tomando como base la variable días_donación_total y promedio días de donación se marca NO FRECUENTE = 1 donación o >a 2 donaciones con promedio_dias_donación > 421 días FRECUENTE: > 2 donaciones con promedio_dias_donación <= 421 días
Dia_donacion	Indica el día en que la persona donó sangre	Se extrae de la fecha de la primera donación.
Grupo_ocupacion	Indica si la persona esta empleada o desempleada	Se obtiene recodificando la variable ocupación. Se toman las ama de casa, estudiantes, pensionados y desempleados como 0 (desempleados) y los demás como 1(empleados)

Tabla 4 Variables construidas para el diseño del modelo predictivo de donantes de sangre.

Una vez lista la base de datos y como parte de la implementación del modelo, se procede a realizar la modelación usando algoritmos para cada uno de los métodos, se realizan pruebas de algoritmos a diferentes métodos para evaluar los resultados, y determinar si lo predicho es correcto o no, se hace un análisis de los resultados de acuerdo a la precisión de los algoritmos a través de una prueba para conjuntos diferentes de datos con el propósito de garantizar la precisión del algoritmo elegido a través de dos fases: la primera fase siendo la de entrenamiento del modelo con los datos del set de entrenamiento (70% de los datos) y la segunda siendo la validación real con los datos del Set de Test (30% de los datos), donde finalmente se elige la modelación cuyo método obtenga el mayor ajuste en el Set de Test, adicional se podrían probar los modelos con aquellos datos de donantes nuevos a partir del 15 de Julio de 2019 (1732) y esperar si estos se convierten en la manera como son clasificados.

7. Resultados

Antes de relacionar los resultados de la modelación, se hace importante contextualizar al lector sobre el resumen de la información, ya que esta dará un panorama de la población estudiada.

En el periodo del 1 septiembre de 2012 a 30 de septiembre de 2019 hubo un total de 75.606 donaciones, sin embargo y excluyendo aquellos registros de donantes a partir del 1 de septiembre y aquellos de donantes que solo tienen una donación a partir del 15 de julio de 2019, obtenemos un total de 73.504 donaciones, de las cuales el 51.4% (37.769) son de personas entre 18 y 30 años de edad y el 53.4% mujeres.

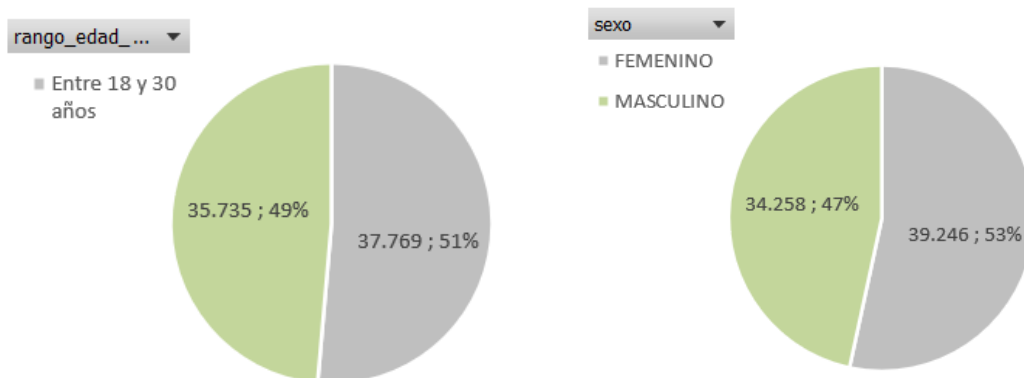


Grafico 1 Distribución donantes por edad Grafico 2 Distribución donantes por sexo

El año en el que más se presentaron donaciones fue en el 2013 (17.210 – 12355 donantes), sin embargo, el año que más tuvo donantes fue el 2014 (12569 -16.191 donaciones); es decir; aunque el 2013 ocupó el segundo lugar en número de donantes de todos los periodos evaluados, fue el que obtuvo mayor número de donaciones, por consiguiente, se infiere que los donantes de ese año hicieron más número de donaciones. Así mismo y

observando años completos, el que tuvo menor número de donaciones fue el año 2016 (10.677 donaciones- 9.003 donantes).

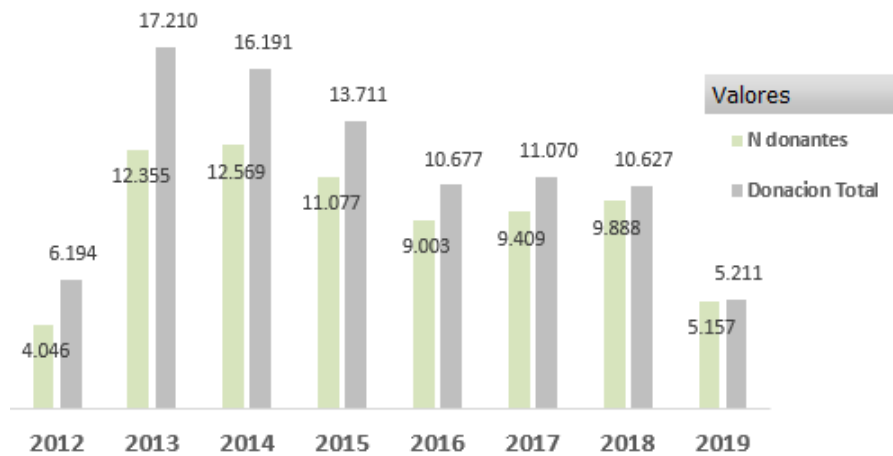


Grafico 3 Distribución donantes por año

El tipo de Sangre y RH que más donantes obtuvo en el transcurso de los 7 años evaluados fue la o+ con un 54% (39.708), seguido de la A+ con un 27% (19.781). El 84.5% (62.148) de los donantes son No frecuentes, han realizado solo 1 donación en el periodo de los 7 años, y subiría a 92.2 % si sumamos aquellos que en su promedio de frecuencia en días para donar superan los 421 días- valor medio estimado para la muestra.

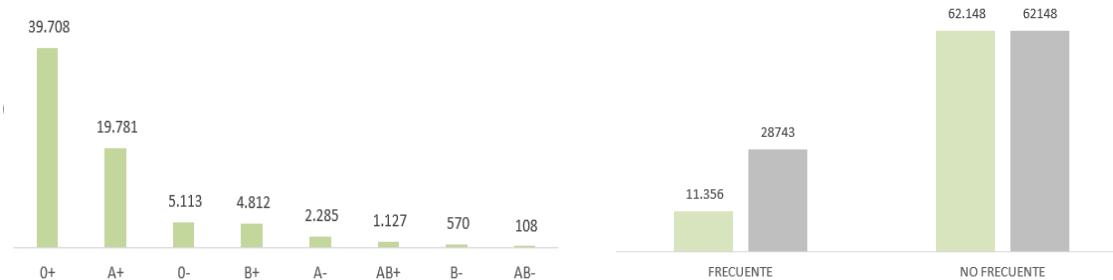


Grafico 4 Donantes por tipo de sangre y RH Grafico 5 Donantes por clasificación de frecuencia

Para finalizar y brindar resultados del modelo, se recuerda que se hizo uso de Python y su librería Scikit-learn.

Python es uno de los lenguajes de programación que domina dentro del ámbito de la estadística, data mining y machine learning, innumerables usuarios pueden implementar

sus algoritmos, cada paquete tiene una sintaxis, estructura e implementación propia. Scikit-learn, es una librería de código abierto que unifica bajo un único marco los principales algoritmos y funciones, facilitando en gran medida todas las etapas de preprocesado, entrenamiento, optimización y validación de modelos predictivos. (35)

se puede observar los seis métodos (tipos de modelo) implementados para el estudio, en donde cuatro (Regresión, Logística, LDA; GNB y SVM) obtuvieron un ajuste del 93%.

Tipo de Modelo	Exactitud del modelo según set		
	Entrenamiento	Test	EMC test
Logistic Regression classifier	0.92	0.93	0.07
Decision Tree classifier	0.99	0.89	0.11
K-NN classifier	0.92	0.92	0.08
LDA classifier	0.92	0.93	0.07
GNB classifier	0.92	0.93	0.07
SVM classifier	0.92	0.93	0.07

Tabla 5 Exactitud de los Modelos implementados de acuerdo al set de evaluación

Esto quiere decir que dichos modelos clasificarían con una exactitud del 93% los donantes en Frecuentes o No frecuentes.

Los algoritmos usados fueron

```
import pandas as pd
import numpy as np
from sklearn import linear_model
from sklearn import model_selection
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
import matplotlib.pyplot as plt
import seaborn as sb

Model=pd.read_pickle ('dt_classifier.pkl')
clf2 = joblib.load("dt_classifier.pkl")
result= clf2.predict(X_new_donac)
result
```

Grafico 6 Algoritmos usados en la modelación

8. Conclusión

La sangre es un recurso público y sigue siendo un componente de uso terapéutico para muchos eventos que requieren manejo de oxígeno y compensación de volumen, de ahí la importancia que exista sangre suficiente, lo más segura posible, asequible y oportuna. Todo con el fin de satisfacer las necesidades de los pacientes que la requieran. Por ello la importancia de contar con un modelo que pueda clasificar los donantes en frecuentes o no frecuentes y así apoyar la planificación de los servicios y campañas en los Bancos de Sangre, especialmente en el Banco de Sangre de la Escuela de Microbiología de la Universidad de Antioquia.

Se ha comparado la exactitud o ajuste de varios algoritmos o métodos de clasificación con el fin de saber si se puede predecir si una persona donante voluntaria puede convertirse en frecuente o no. Entre los algoritmos implementados, los que mejor se ajustaron con los datos de Test fueron: Regresión logística, k-Nearest Neighbors, Gaussian Naive Bayes- GNB, análisis discriminante lineal –ADL y Maquinas de soporte de vector –SVM; estos se ajustan todos en un 93%, el algoritmo de Árbol de decisión fue el que menos se ajustó cuando se validó con los datos de Test (89%) aun cuando fue el que mejor se ajustó con los datos de entrenamiento (99%).

Los modelos implementados tomaron como variable respuesta el tipo de donante (Frecuente o no Frecuente) y variables predictoras el sexo, edad estandarizada, peso estandarizado, ocupación, clasificación sanguínea y día de donación (sexo, edad, peso, ocupación_rec22, ABO_RH y dia_donacion). Estos modelos cuentan con un desempeño en la validación considerado como bueno.

9. Recomendaciones

Este estudio se realizó con una base de datos extraída del sistema de información de registro de donantes de la Escuela de Microbiología de la Universidad de Antioquia, el cual toma como entrada la encuesta estandarizada y unificada para Colombia para la selección de donantes de sangre. Dicha encuesta debe ser diligenciada directamente por el donante potencial y está conformada por un total de 29 preguntas relacionadas con antecedentes médicos y conductas de riesgo para ITTs. La encuesta (física, digital o electrónica) es uno de los soportes fundamentales para conocer el destino final dado a cada uno de los hemocomponentes producidos a partir de una unidad de sangre extraída (trazabilidad). Por lo tanto, puede ser considerada como una historia clínica y debe ser archivada (en forma: física, digital o electrónica) según el tiempo y mecanismo establecido por la normatividad vigente, con acceso limitado al personal de salud autorizado, en condiciones que garanticen su integridad, sin alteración de la información y debe velarse por su custodia. (8)

Por lo anterior, solo es posible trabajar con las variables otorgadas por el Banco de Sangre y realizar las transformaciones correspondientes sin afectar la originalidad y calidad del dato.

Se recomienda para futuros estudios poder analizar otro tipo de variables que si bien están relacionadas en la encuesta, no están registradas para todos los casos en el sistema de información, como:

- Si ha donado sangre anteriormente y a quien, ya que con esto se puede discriminar el tipo de donante.

- Si va a donar sangre de manera altruista, para reponer unidades transfundidas, o para sí mismo (autóloga), con el fin de mitigar el ruido que generan donaciones que a juicio subjetivo pueden ser no altruistas.
- El tipo de procedimiento: si es Aféresis, leucoferesis, flebotomía, plaquetas, sangre total, entre otras, esto con el fin de mitigar el ruido que se genera en las frecuencias de donación, ya que para algunos procedimientos puede realizarse cada mes y esto dificulta la identificación de la donación de sangre total.
- Si presentó alguna reacción adversa, ya que esta variable puede ser altamente influyente para que un donante se convierta en habitual o repetitivo.
- El barrio donde vive el donante, ya que no todos los registros cuentan con este y podría ser un factor influyente en la donación.

Entre otras variables que representarían novedades en los estudios de predicción.

10. Bibliografía

1. Colombia. Santafé de Bogota DC. Ministerio de salud. Por el cual se reglamenta parcialmente el Título IX de la Ley 09 de 1979, en cuanto a funcionamiento de establecimientos dedicados a la extracción, procesamiento, conservación y transporte de sangre total o de sus hemoderivados, se crean la Red Nacional de Bancos de Sangre y el Consejo Nacional de Bancos de Sangre y se dictan otras disposiciones sobre la materia. Diario oficial No 40.989. Decreto 1571 de 1993. 1993.
2. Gustavo Gollan D, Yedlin G, Kohan P, Maschio M, Fernández G. Criterios para la selección de donantes de sangre [Internet]. Vol. 1, Ministerio de Salud. Buenos Aires Argentina; 2015. Available from: <http://www.msal.gov.ar/disahe/images/stories/pdf/guia-criterios-seleccion-donantes.pdf>
3. Ministerio de Salud. Manual Donantes de Sangre [Internet]. San Salvador; 2010. Available from: http://asp.salud.gob.sv/regulacion/pdf/manual/manual_donantes_sangre.pdf
4. Banco de sangre Fundación. Fundación Banco de Sangre y Tejidos de las Islas Baleares [Internet]. 2016 [cited 2019 Sep 9]. p. 7. Available from: http://www.donasang.org/que-es-la-sang/es_index.html
5. Organización Mundial de la Salud. Transfusión de Sangre [Internet]. 2014. p. 1. Available from: https://www.who.int/topics/blood_transfusion/es/
6. Espino Timón C. Análisis predictivo técnicas y modelos utilizados y aplicaciones del mismo [Internet]. 2017. Available from: <http://openaccess.uoc.edu/webapps/o2/bitstream/10609/59565/6/caresptimTFG0117memoria.pdf>

7. Instituto Nacional de Salud. Lineamiento técnico para la selección de donantes de sangre en Colombia. Bogotá; 2018. p. 8.
8. Ruiz F. Modelamiento matemático aplicado al modelamiento espacial. Br J Psychiatry [Internet]. 1966;112(483):211–2. Available from: <http://www.sociedadelainformacion.com/14/modelamiento.pdf>
9. Digital H. Diferencia entre el aprendizaje automático y el análisis predictivo [Internet]. 03. 2018 [cited 2019 Sep 9]. p. 1. Available from: <https://www.digitalhouse.com/noticias/diferencias-aprendizaje-automatico-analisis-predictivo-datos/>
10. Yutang L. La importancia de vivir [Internet]. 6 Edición. Jiménez R, editor. Buenos Aires Argentina: Sudamericana; 1943. 225 p. Available from: <https://www.pasadofuturo.com/archivos/librolaimportanciadevivirlinyutang01.pdf>
11. Gnecco Zuleta AJ. Proyecto de Ley para la Donación de Sangre. In: Proyecto de ley para incentivar la donación de sangre en el país [Internet]. Colombia; 2018. p. 8. Available from: [http://leyes.senado.gov.co/proyectos/images/documentos/Textos/Radicados/proyectos de ley/2018 - 2019/PL 093-18 Donacion de Sangre.pdf](http://leyes.senado.gov.co/proyectos/images/documentos/Textos/Radicados/proyectos%20de%20ley/2018%20-%202019/PL%20093-18%20Donacion%20de%20Sangre.pdf)
12. Fernández Cid, Gabriela. Fontana D. Donación de sangre voluntaria y habitual [Internet]. 1st ed. Seoane S, Serrano A, editors. Buenos Aires Argentina: Blanco Gonzalo; 2011. 1–77 p. Available from: <http://www.msal.gob.ar/images/stories/ryc/graficos/0000000888cnt-cuadernillo-educacion.pdf>
13. Laudon KC, Laudon JP. Sistemas de procesamiento de transacciones [Internet]. Decimosegu. Cruz castillo LM, Hernández Gutiérrez B, editors. Sistemas de

- Información Gerencial, 12va Edición. México; 2012. 643 p. Available from:
<https://juanantonioleonlopez.files.wordpress.com/2017/08/sistemas-de-informacion-gerencial-12va-edicion-kenneth-c-laudon.pdf>
14. Tom Mitchell. Aprendizaje automático , McGraw Hill, 1997.
 15. Christopher M. Bishop. Reconocimiento de patrones y aprendizaje automático , Springer, 2006.
 16. Stephen M. Machine Learning An Algorithmic Perspective [Internet]. 2nd ed. Herbrich R, Graepel T, editors. Berlin: Stephen Marsland; 2014. 457 p. Available from: [https://doc.lagout.org/science/Artificial Intelligence/Machine learning/Machine Learning_ An Algorithmic Perspective %282nd ed.%29 %5BMarsland 2014-10-08%5D.pdf](https://doc.lagout.org/science/Artificial%20Intelligence/Machine%20learning/Machine%20Learning%20An%20Algorithmic%20Perspective%202nd%20ed.%20%5BMarsland%202014-10-08%5D.pdf)
 17. Russell R. Machine learning step-by-step guide to implement machine learning algorithms with Python [Internet]. 1st ed. 2018. 106 p. Available from: <http://booksdescr.org/item/index.php?md5=D161EE832B8007A058CD006DD67E388E>
 18. Joaquín Amat Rodrigo. Regresión logística con Python Noviembre, 2020
 19. Deepti Bahel, Prerana Ghosh, Arundhyoti Sarkar, Matthew A. Lanham. Predicting Blood Donations Using Machine Learning Techniques. 2018. Purdue University, Department of Management, 403 W. State Street, West Lafayette, IN 47907
 20. Swamynathan M. Mastering Machine Learning with Python in Six Steps [Internet]. 1st ed. Green T, Suresh J, editors. New York; 2017. 374 p. Available from: <https://tanthiamhuat.files.wordpress.com/2018/04/mastering-machine-learning-with-python-in-six-steps.pdf>
 21. Madan A, Cebrian M, Lazer D, Pentland A. Social sensing for epidemiological

- behavior change. UbiComp'10 - Proc 2010 ACM Conf Ubiquitous Comput [Internet]. 2010;291–300. Available from: <http://web.media.mit.edu/~cebrian/sensing.pdf>
22. Napoleón C. Huellas digitales para prevenir el estado de Salud [Internet]. Banco Interamericano de Desarrollo. 2017 [cited 2019 Sep 12]. p. 2. Available from: <https://clustersalud.americaeconomia.com/opinion/netflix-amazon-y-las-politicas-publicas-en-salud>
 23. Obermeyer Z. Prediciendo el futuro: Big Data, aprendizaje automático y medicina clínica. 2016;10. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5070532/>
 24. License CC, Version D. DESIGN4HEALTH. In: Seemann K, editor. Conferencia del diseño 4 de Salud [Internet]. Australia; 2018. p. 357. Available from: https://vbn.aau.dk/ws/portalfiles/portal/288036476/D4H2017_Proceedings_Master_File_print.pdf
 25. Aldamiz-Echevarria C, Aguirre-Garcia MS. Modelo de Comportamiento de los donantes de sangre y estrategias de Marketing para retenerles y atraerles. Rev Lat Am Enfermagem [Internet]. 2014;22(3):467–75. Available from: http://www.scielo.br/pdf/rlae/v22n3/es_0104-1169-rlae-22-03-00467.pdf
 26. Zuckerman M. Comparación de tres modelos para predecir el comportamiento altruista. 1978;36:20. Available from: <https://psycnet.apa.org/doiLanding?doi=10.1037%2F0022-3514.36.5.498>
 27. Doggenweiler I, Guic E. Determinantes psicosociales de la intención de donación de órganos en una muestra Chilena. Rev Med Chil [Internet]. 2014;142(1):27–33. Available from: <https://scielo.conicyt.cl/pdf/rmc/v142n1/art05.pdf>
 28. Bamberg F. Modelo de Futuro de Gestion de la Salud. xxv Aniversario la ley Gen

- Sanid [Internet]. 2011;1:241. Available from: http://www.consorci.org/media/upload/pdf/docs_norm/el_modelo_de_futuro_de_gestion_de_la_salud.pdf
29. Appelbaum B. Las fases que define un Modelo Predictivo [Internet]. Febrero. 2018 [cited 2019 Sep 16]. p. 5. Available from: <https://blog.datary.io/las-fases-que-definen-un-modelo-predictivo/>
 30. Ministerio de Salud. Procedimientos En Bancos De Sangre Normas Tecnicas Y Administrativas [Internet]. Colombia; 2017 p. 105. Available from: [https://www.minsalud.gov.co/Documentos_y_Publicaciones/MANUAL DE NORMAS TECNICAS ADMINISTRATIVAS Y DE PROCEDIMIENTOS PARA BANCOS DE SANGRE.pdf](https://www.minsalud.gov.co/Documentos_y_Publicaciones/MANUAL_DE_NORMAS_TECNICAS_ADMINISTRATIVAS_Y_DE_PROCEDIMIENTOS_PARA_BANCOS_DE_SANGRE.pdf)
 31. Ministerio de la Protección Social (2005). Política Nacional de Sangre. Bogotá, p.3.
 32. Instituto Nacional de Salud. Guía rápida para la toma de decisiones en Medicina Transfuncional. Bogotá: Imprenta Nacional de Colombia; 2010 p. 22.
 33. Instituto Nacional de Salud. Manual de Hemovigilancia. Bogotá: Imprenta Nacional de Colombia; 2010. p. 4.
 34. Colombia. Santafé de Bogotá. Diario oficial. Normas técnicas servicio de Banco de Sangre. Resolución 0901 de 1996.
 35. Joaquín Amat Rodrigo. Machine learning con Python y Scikit-learn. 2020 actualizado Octubre2020) Available from: https://www.cienciadedatos.net/documentos/py06_machine_learning_python_scikitlearn.html