portada

Resumen

Aquí va el rezumen

Índice

Contenido

[Resumen 3](#_Toc42426384)

[Índice 5](#_Toc42426385)

[1. Introducción 13](#_Toc42426386)

[**1.1.** Conceptos previos 15](#_Toc42426387)

[**1.2.** Antecedentes 16](#_Toc42426388)

[**1.3.** Situación actual 18](#_Toc42426389)

[**1.4.** Motivación 20](#_Toc42426390)

[**1.5.** Objetivos 20](#_Toc42426391)

[**1.6.** Estructura de la memoria 21](#_Toc42426392)

[2. Contexto informático 23](#_Toc42426393)

[**2.1.** Teoría de agentes 25](#_Toc42426394)

[**2.2.** ¿Qué es un agente? 25](#_Toc42426395)

[**2.3.** Características del entorno 27](#_Toc42426396)

[**2.4.** Sistemas multiagentes 27](#_Toc42426397)

[2.4.1. Aplicaciones de los sistemas multiagentes 28](#_Toc42426398)

[2.4.2. Herramientas para su construcción 29](#_Toc42426399)

[**2.5.** Modelos de propagación de enfermedades 30](#_Toc42426400)

[3. Metodología y herramientas de trabajo 32](#_Toc42426401)

[**3.1.** Metodología 33](#_Toc42426402)

[**3.2.** Herramientas de trabajo 34](#_Toc42426403)

[3.2.1. R y RStudio 34](#_Toc42426404)

[3.2.2. Microsoft Excel 34](#_Toc42426405)

[3.2.3. Github 35](#_Toc42426406)

[3.2.4. Netlogo 35](#_Toc42426407)

[4. Modelado 1 38](#_Toc42426408)

[**4.1.** Requisitos modelo 1 40](#_Toc42426409)

[4.1.1. Requisitos funcionales 40](#_Toc42426410)

[4.1.2. Requisitos no funcionales 42](#_Toc42426411)

[**4.2.** Diseño modelo 1 43](#_Toc42426412)

[**4.3.** Simulaciones de prueba 50](#_Toc42426413)

[5. Modelado 2 51](#_Toc42426414)

[6. Conclusiones y mejoras futuras 51](#_Toc42426415)

[Referencias 53](#_Toc42426416)

[7. Apéndices 57](#_Toc42426417)

[**7.1.** Manual de instalación 59](#_Toc42426418)

[7.1.1. Requisitos previos 59](#_Toc42426419)

[7.1.2. Apertura de los modelos 60](#_Toc42426420)

[**7.2.** Manual de uso Modelo 1 63](#_Toc42426421)

[**7.3.** Manual de uso Modelo 2 67](#_Toc42426422)

[8. Anexo de datos 70](#_Toc42426423)

Tabla de ilustraciones

[Ilustración 1. Estadísticas de la epidemia de sida en 2009 [5] 18](#_Toc42426431)

[Ilustración 2. Zonas más afectadas por la epidemia de ébola (2014-2016) [1] 19](#_Toc42426432)

[Ilustración 3. Captura realizada a la página de ElPaís el 5 de abril de 2020 [6] 20](#_Toc42426433)

[Ilustración 4. Mapa de casos confirmados por COVID-19 [1] 21](#_Toc42426434)

[Ilustración 5. Representación de un agente en su entorno [7] 27](#_Toc42426435)

[Ilustración 6. Ecuaciones del modelo SI [8] 31](#_Toc42426436)

[Ilustración 7. Ecuaciones del modelo SIS [8] 31](#_Toc42426437)

[Ilustración 8. Ecuaciones del modelo SIR [8] 32](#_Toc42426438)

[Ilustración 9. Fases de CRISP-DM [1] 34](#_Toc42426439)

[Ilustración 10. Ventana de inicio de Netlogo 44](#_Toc42426440)

[Ilustración 11. Entrada del número de personas 45](#_Toc42426441)

[Ilustración 12. Botón setup de la simulación 45](#_Toc42426442)

[Ilustración 13. Botón go para movimiento de los agentes 45](#_Toc42426443)

[Ilustración 14. Seleccionador de la forma 46](#_Toc42426444)

[Ilustración 15. Ventana de entrada del porcentaje de infectados inicial 47](#_Toc42426445)

[Ilustración 16. Ventanas de entrada de porcentaje de aparición de los síntomas 48](#_Toc42426446)

[Ilustración 17. Creación de síntomas de los agentes 48](#_Toc42426447)

[Ilustración 18. Carga y muestra de tasa de letalidad y recuperación 49](#_Toc42426448)

[Ilustración 19. Ventanas de entrada de porcentaje de recuperación y fallecimiento 49](#_Toc42426449)

[Ilustración 20. Ventana de entrada de porcentaje de contagio 49](#_Toc42426450)

[Ilustración 21. Deslizador tiempo de recuperación 49](#_Toc42426451)

[Ilustración 22. Porcentajes referidos a las mascarillas 50](#_Toc42426452)

[Ilustración 23. Imagen obtenida de la página de Netlogo [10] 60](#_Toc42426453)

[Ilustración 24. Imagen para descargar Netlogo [10] 60](#_Toc42426454)

[Ilustración 25. Opciones de descarga de Netlogo [10] 61](#_Toc42426455)

[Ilustración 26. Interfaz inicial de Netlogo 62](#_Toc42426456)

[Ilustración 27. Deslizador de la velocidad de simulación 65](#_Toc42426457)

[Ilustración 28. Interfaz inicial del Modelo 1 66](#_Toc42426458)

[Ilustración 29. Variables del conjunto de datos de pacientes de COVID19 73](#_Toc42426459)

[Ilustración 30. 20 primeros síntomas de los pacientes con COVID19 75](#_Toc42426460)

[Ilustración 31. 6 primeras filas del conjunto de datos creados (1) 76](#_Toc42426461)

[Ilustración 32. 6 primeras filas del conjunto de datos creados (2) 76](#_Toc42426462)

[Ilustración 33. Histograma de la variable edad 76](#_Toc42426463)

[Ilustración 34. Distribución de la población con el síntoma fiebre 77](#_Toc42426464)

[Ilustración 35. Síntomas y sus porcentajes de aparición en la población 77](#_Toc42426465)

[Ilustración 36. Otros porcentajes de utilidad 78](#_Toc42426466)

1. Introducción
   1. Conceptos previos

Antes de comenzar con el tratamiento de situaciones pasadas y en este caso, de la situación actual que se está viviendo en el mundo, sería recomendable aclarar diferentes aspectos relacionados entre sí:

* Brote epidémico: se trata de la aparición repentina de una enfermedad debida a una infección producida en un lugar específico y en un momento determinado. Un ejemplo claro de esta situación puede ser una intoxicación alimentaria.
* Epidemia: se cataloga como epidemia cuando una enfermedad se propaga activamente debido al descontrol del brote y que éste se mantiene en el tiempo, provocando un aumento en el número de casos en un área geográfica concreta.
* Pandemia: para la declaración de estado de pandemia se deben cumplir dos requisitos principalmente, que el brote afecte epidémico afecte a más de un continente y que los casos provocados en cada país no sean importados, sino provocados por transmisión comunitaria.
* Número reproductivo básico (R0): se trata del número medio de personas contagiadas por un individuo infectado que se introduce en una población susceptible de contraer la enfermedad.
* Tamaño de la epidemia: se trata del número total de individuos afectados.

Así, por ejemplo, el caso más reciente de todos y que se está viviendo actualmente, se encuentra la enfermedad COVID-19 o [coronavirus 2 del síndrome respiratorio agudo grave (SARS-CoV-2)](https://es.wikipedia.org/wiki/SARS-CoV-2) la cual se transmitió por animales comercializados en mercados mayoristas de la ciudad de Wuhan, China y que se puede tratar en sus inicios de la enfermedad como un brote. Debido a su expansión por toda China pasó a ser una epidemia. Por último, conforme se fue transmitiendo por todos los continentes, la OMS acabó reconociendo la pandemia global el 11 de marzo de 2020. Todo esto se recoge en las palabras de [Ángel Gil](https://gacetamedica.com/?s=angel+gil), catedrático de Medicina Preventiva y Salud Pública de la Universidad Rey Juan Carlos: “Mientras los casos eran importados y el foco epidémico estaba localizado en China la situación era calificada de epidemia pero en el momento en que salta a otros países y empieza a haber contagios comunitarios en más de un continente se convierte en pandemia”.

* 1. Antecedentes

A continuación, se van a exponer diferentes enfermedades que fueron en su momento respectivas epidemias o pandemias:

* Peste de Justiniano

Cuando el Imperio bizantino se encontraba en uno de sus momentos de mayor esplendor, una epidemia de peste sacudió el mandato de Justiniano. Se trata de la primeria epidemia de peste de la que se tiene constancia. La enfermedad se expandió por toda Constantinopla, una ciudad de unos 800000 habitantes, con gran rapidez, afectando a un 40% de la población, incluido el emperador Justiniano.

* Peste negra

El brote más importante sufrido por la humanidad de peste negra tuvo lugar a mediados del siglo XIV (1346-1353). Se conocía debida a sus antecedentes, pero no se tenía constancia de tratamiento para curarla. Esto junto a la velocidad a la que se expandió, dejó números estremecedores. La península ibérica perdió entre el 60% y 65% de la población, mientras que la región italiana de la Toscana, entre el 50% y 60%. En resumen, la población europea se disminuyó en 50 millones de personas.

* Viruela

Su nombre hace referencia a las pústulas que surgen en la piel de la persona que la padece. Su aparición diezmó la población mundial, llegando a presentar un 30% de mortalidad, incluso alcanzando el nuevo mundo de la mano de los exploradores en el siglo XVIII, consiguiendo infectar y desfigurar a millones de personas.

* Gripe española

En marzo de 1918, durante la Primera Guerra Mundial, se dio el primer caso en Estados Unidos. Fue una virulenta cepa del virus de la gripe que se extendió por todos los frentes mientras que las tropas se repartían en ellos. Se estima una tasa de mortalidad de un 10% a un 20%, muriendo de 20 millones de personas a 50 millones.

* Gripe asiática

En la península de Yunán, China, surgió en 1957 el virus de la gripe A (H2N2) de procedencia aviar. Para ese año, la Organización Mundial de la Salud (OMS) creada en 1948, se encargaba de crear vacunas anuales para paliar los efectos de la gripe. Aun teniendo en cuenta los avances médicos, esta pandemia registró un millón de muertos en todo el planeta.

* Gripe de Hong Kong

Tan solo 10 años más tarde de superar la última pandemia de gripe, surgió en Hong Kong una variación de la gripe A (H3N2). Esta nueva cepa provocó un millón de muertos.

* Virus de Inmunodeficiencia Adquirida (VIH)

Más comúnmente conocido como SIDA. Sus primeros casos tuvieron lugar en 1981 y comenzó a extenderse, centrando los esfuerzos de las organizaciones de la salud. Se piensa que tuvo un origen animal y se podría describir como un agotamiento del sistema inmunológico. Como pasa en todos estos casos, el desconocimiento inicial del virus provocó una rápida expansión.

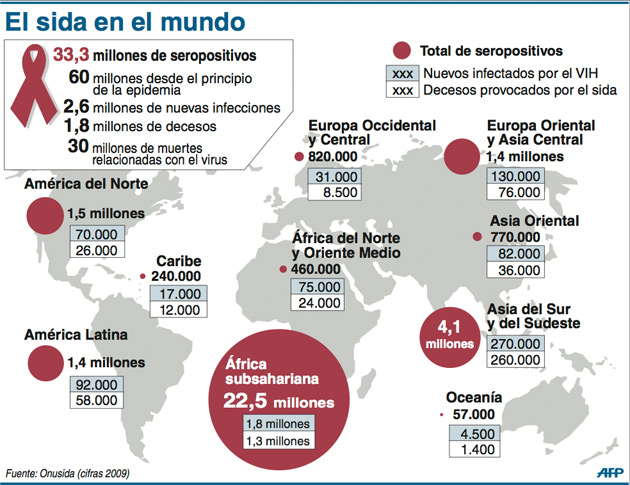


Ilustración 1. Estadísticas de la epidemia de sida en 2009 [5]

Como se observa en la “Ilustración 1. Estadísticas de la epidemia de sida en 2009 [5]”, el continente más afectado es África. Esto se puede deber al bajo desarrollo presente en él respecto a otros continentes, alcanzando cerca de dos millones de personas afectadas.

* Ébola

Se trata de uno de los casos más allegados en el tiempo. El virus del ébola se descubrió por primera vez en 1976 en África. Se trata de una enfermedad grave que a menudo es mortal para el ser humano. El brote que tuvo lugar entre los años 2014 y 2016, ha sido el más importante, alcanzando las mismas cifras que los demás juntos. Las víctimas de este brote se localizaron especialmente en África.

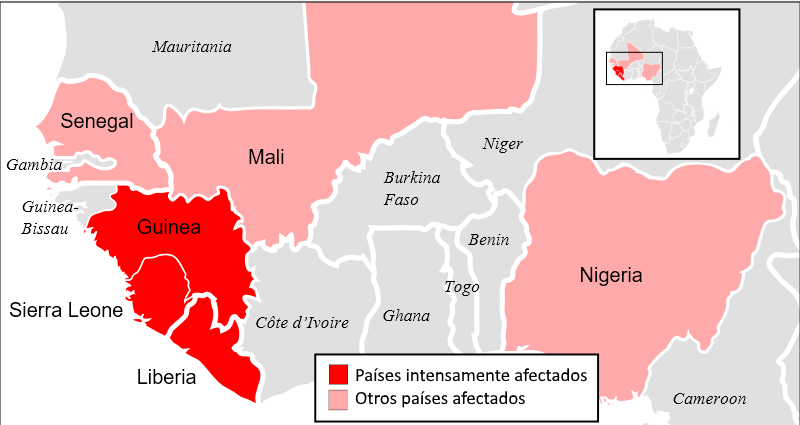


Ilustración 2. Zonas más afectadas por la epidemia de ébola (2014-2016) [1]

Además, caben destacar otras epidemias como pueden ser las de la meningitis, dengue o zika.

* 1. Situación actual

En estos días, el mundo se encuentra inmerso en una pandemia que prácticamente ha afectado a todos los países del mundo ya.

El COVID-19 o [coronavirus 2 del síndrome respiratorio agudo grave (SARS-CoV-2)](https://es.wikipedia.org/wiki/SARS-CoV-2) fue identificado por primera vez el 1 de diciembre de 2019 en la ciudad china de Wuhán al reportarse un número de personas con neumonía sin saberse la causa. La transmisión del virus se produce por las gotas emitidas al hablar, toser o estornudar.

Con el objetivo de frenar la expansión del virus, los gobiernos de todas las zonas del mundo están decretando restricciones de viajes, cuarentenas, confinamientos en casa entre otras medidas. Esta pandemia está teniendo un impacto a nivel mundial en diferentes aspectos como pueden ser el socioeconómico, cultural, medioambiental…

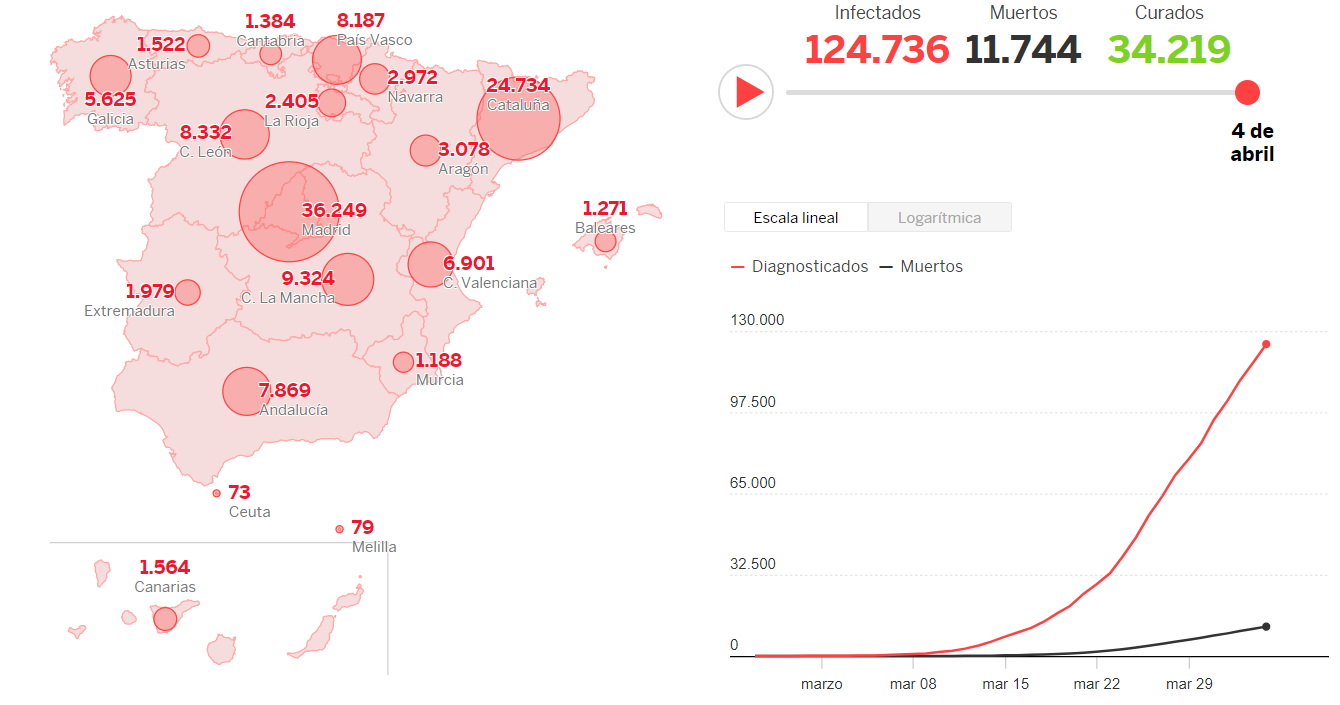


Ilustración 3. Captura realizada a la página de ElPaís el 5 de abril de 2020 [6]

En la “Ilustración 3. Captura realizada a la página de ElPaís el 5 de abril de 2020 [6]”, se pueden observar los casos de COVID-19 existentes en España hasta el 4 de abril de 2020. España es uno de los países más afectados en la actualidad por este virus. A continuación, se muestra una imagen de la expansión mundial del COVID-19.

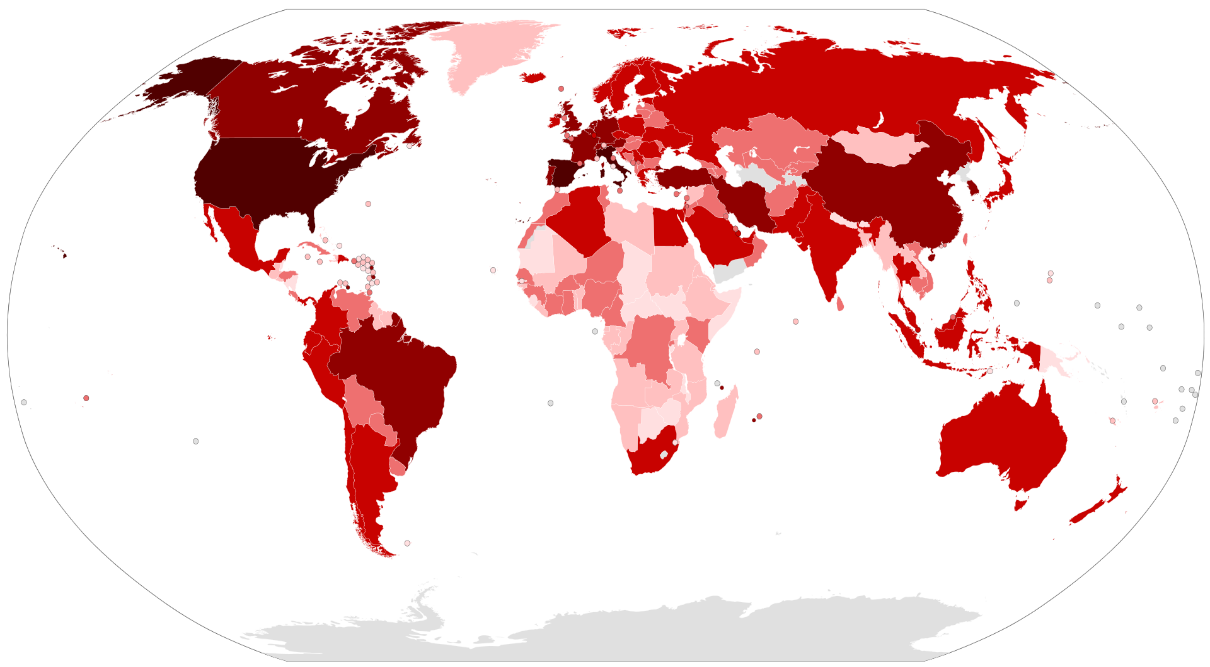


Ilustración 4. Mapa de casos confirmados por COVID-19 [1]

En la “Ilustración 4. Mapa de casos confirmados por COVID-19 [1]”, se observan cómo de afectados se encuentran los países según el color. A más oscuro es el color, más afectado se encuentra el país.

* 1. Motivación

Como consecuencia de todas las pandemias surgidas en la historia de la humanidad y que por desgracia seguiremos sufriendo, los gobiernos de los países precisaron que era necesario la creación de un modelo informático el cual les permitiese saber cómo actuar en estos casos. De esta forma, se podrá estudiar el impacto de una enfermedad en una población con unas características determinadas pudiendo actuar de la mejor manera posible para así disminuir posibles daños.

Desde el punto de vista personal, a la hora de la elección de esta línea de trabajo (noviembre de 2019), me decanté por el tema de la simulación puesto que siempre me interesó dicho tema. Con el avance de los días y la aparición de este virus, mi interés fue creciendo a pasos agigantados.

* 1. Objetivos

El objetivo del proyecto es realizar una aplicación de escritorio usando Netlogo que, mediante el estudio previo de unos datos, y la inserción de parámetros varios, permita simular y visualizar el comportamiento de una enfermedad en una población determinada para ver cómo se desarrolla con el paso del tiempo.

En el primer modelado se permitirá la simulación y visualización de la enfermedad en una población en la que los agentes presenten actividades sencillas. En el segundo modelado se pretende mejorar la actividad del agente, además de incluir otros factores externos como pueden ser mascotas, vacunas o añadir hospitales. También se generarán gráficos y se mostrarán estadísticas sobre los resultados para su posterior análisis.

De esta forma, se podrá realizar un estudio sobre dicha epidemia para lograr disminuir sus efectos en la población, así como estudiar diferentes escenarios posibles.

* 1. Estructura de la memoria

La memoria estará constituida por los siguientes capítulos:

**Capítulo 2. Contexto informático**: en este capítulo se abordarán temas como la teoría de agentes, los sistemas multiagentes y los modelos de propagación de enfermedades.

**Capítulo 3. Metodología y herramientas de trabajo:** se describirá cómo se realizará la aplicación, así como las herramientas usadas.

**Capítulo 4. Modelado 1:** se realizará un primer modelado sencillo, para simular la expansión de una enfermedad (en este caso COVID19), así como los síntomas destacados y la inclusión de mascarillas

Capítulo 5. Modelado 2:

Capítulo 6. Conclusiones y futuras mejoras:

1. Contexto informático

En la introducción se ha comentado la parte sanitaria de este trabajo. A continuación, se va a explicar la parte informática, para ello se debe tratar todo lo relacionado con la teoría de agentes y el modelado en sistemas multiagentes.

* 1. Teoría de agentes

Los ordenadores, en general, no son bueno tomando decisiones. Cada acción realizada por una computadora es anticipada, planeada y programada por un informático. Por el contrario, si la acción no ha sido programada el sistema fallará en el mejor de los casos y en el peor, se romperá. De esta manera, las computadoras son aceptadas como máquinas que reciben órdenes y permiten realizar diversas aplicaciones, sin embargo, para un número en incremento de actividades se es necesario el uso de otro tipo de sistemas capaces de tomar decisiones por sí mismos. Estos son conocidos como **agentes**. De ellos deriva el paradigma de la **programación orientada a agentes (AOP)** en el que a los agentes se le inserta el conocimiento que tiene el mundo de ellos e interactúan los unos con los otros originando **sistemas multiagentes**.

Algunas ventajas del desarrollo con agentes son:

* Son amigables y adaptables al usuario.
* Presenta un coste bajo debido a la fácil reutilización.
* Reducción del mantenimiento: la funcionalidad puede cambiarse modificando los conocimientos, estrategias y objetivos o incluyendo nuevos agentes, nuevo conocimiento y nuevos objetivos en los agentes.
* Pueden ser compatibles con otras tecnologías como webs, base de datos…
  1. ¿Qué es un agente?

El término **agente** es un término que, según la bibliografía, se le da una definición diferente. Una manera de definir a un agente podría ser como un sistema informático, objeto de software o una entidad autónoma e interactiva situada en un ambiente, reaccionando a los estímulos que les transmiten.

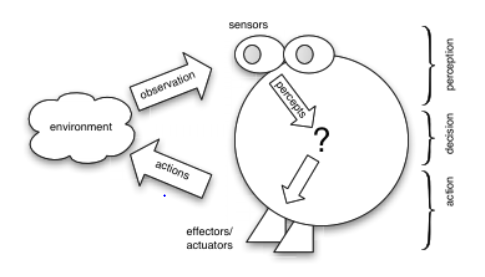


Ilustración 5. Representación de un agente en su entorno [7]

En la “Ilustración 5. Representación de un agente en su entorno [7]”, se representa a un agente inmerso en un entorno del que mediante unos sensores accede a información que le transmiten. Esta información va a ser procesada y finalmente se tomará una decisión que derivará en la realización de una acción por el agente.

Los agentes presentan unas reglas bases que les permiten interactuar con el entorno. También tienen reglas para cambiar reglar que les aportan adaptación. La característica fundamental es la capacidad de tomar decisiones de forma independiente, lo que es conocido como **autonomía**. Para ello, los agentes deben ser activos. Los agentes deben poseer las siguientes características:

* Son identificables, es decir, son individuos que presentan características y reglas para definir sus formas de actuar.
* Como se ha comentado anteriormente, un agente está inmerso en un entorno en el que puede interactuar con más agentes y otros objetos simulados del entorno. Los agentes pueden colaborar para realizar diferentes tareas (sistemas multiagentes).
* Deben presentar uno o varios objetivos a cumplir, acorde a sus comportamientos. Por ello, para cumplir sus objetivos realizarán determinadas acciones (**proactividad**).
* Son flexibles, puesto que presentan la habilidad de aprender y modificar su comportamiento. Presentan reglas que les permiten adaptarse a los cambios mediante la modificación de sus reglas de comportamiento.
  1. Características del entorno
* Accesible/ inaccesible

Un entorno será accesible cuando el agente sea capaz de adquirir información actualizada y precisa sobre él. Cuanto más accesible sea, más sencillo será la construcción de un agente.

* Determinista/ no determinista

En este tipo de entorno, cada acción tiene un único efecto garantizado.

* Episódico/ no episódico

El agente desempeñará una acción dependiendo del episodio en el que se encuentre.

* Estático/ dinámico

El entorno estático es aquel que no recibe ningún cambio independientemente del acto del agente. Por el contrario, el dinámico es aquel que presenta cambios.

* Discreto/ continuo

Un entorno discreto está compuesto por un número de acciones, mientras que el continuo es una única acción.

* 1. Sistemas multiagentes

Un sistema multiagente es aquel en el que interactúan un número de agentes. Cada agente presenta sus propios objetivos y motivaciones. Para interactuar de forma satisfactoria, los agentes necesitan las habilidades de cooperación, coordinación y negociación.

Según los agentes que compongan el sistema, éste presentará un tipo concreto de arquitectura:

* **Arquitectura reactiva pura**. Estos agentes presentan una conexión entre sensores y actuadores que hacen emerger una conducta inteligente. Su conducta se basa en estímulo-respuesta. Sus acciones no dependen del pasado, dependen únicamente del presente. Los agentes captan una observación del entorno y actúan acorde al estado presente. Por ejemplo: un termostato.
* **Arquitecturas reactivas con estado interno.** Existe un modelo del entorno que les permite decidir la reacción al estímulo. Pueden interpretar el entorno mediante una función que poseen y así poder realizar la mejor respuesta. El estado se actualiza mediante la combinación de la percepción y el estado interno.
* **Arquitecturas deliberativas**. Este tipo de arquitectura está orientada por objetivos. El funcionamiento se basa en el razonamiento, mediante decisiones lógicas, además de tener un modelo interno del entorno. La finalidad es revolver las tareas de los humanos, que les dicen qué hacer, pero no cómo. El agente elige sus propios objetivos y cómo alcanzarlos.

Dentro de este tipo destacan los agentes BDI (Beliefs, Desires, Intentions). Presentan un razonamiento práctico. Se basan en la **deliberación** (decidir los objetivos a conseguir) y en el **razonamiento** (cómo conseguirlos). En el proceso de decisión intervienen 3 elementos:

1. Creencias (Beliefs): saber la visión del mundo
2. Deseos (Desires): opciones según las creencias
3. Intenciones (Intentions): objetivos a perseguir. Este elemento permite influir en decisiones futuras y realizar restricciones en el razonamiento futuro.

* **Arquitecturas híbridas.** Este tipo presentan diferentes capas de deción dónde se combinan las aproximaciones reactivas y deliberativas. El objetivo será maximizar las ventajas de los dos modelos intentando evitar sus inconvenientes.
  + 1. Aplicaciones de los sistemas multiagentes

A continuación, se van a nombrar diferentes ámbitos en los que se usan los sistemas multiagentes:

* Son utilizados en servicios de información en Internet
* Ayudan a la recuperación y extracción de información
* Permiten simular negociaciones
* Son implementados en equipos móviles y PCs domésticos
* Redes públicas de telecomunicaciones
* Se utilizan en la simulación de sistemas dinámicos
* Muchos juegos actuales se apoyan en dicha tecnología
* Robótica
  + 1. Herramientas para su construcción

Las herramientas o lenguajes para la programación de agentes deben contar con unas características determinadas. En ellos, se define los principales componentes del sistema (agentes) así como sus responsabilidades y capacidades y cómo interaccionan entre ellos. Destacan:

* **JADE** (Java Agent Development Framework): se trata de una plataforma dedicada al desarrollo de agentes e implementada en JAVA. es un [middleware](https://es.wikipedia.org/wiki/Middleware) que facilita el desarrollo de sistemas multi-agente bajo el estándar [FIPA](https://es.wikipedia.org/wiki/FIPA) (organismo para el desarrollo y establecimiento de estándares de software para [agentes](https://es.wikipedia.org/wiki/Agente_(software)) heterogéneos que interactúan y sistemas basados en agentes) para lo cual crea múltiples contenedores destinados a los agentes, cada uno de los cuales puede ejecutarse en uno o en varios sistemas. Además, proporciona un entorno de ejecución con herramientas gráficas para la monitorización y administración y bibliotecas para la creación de agentes.
* **Repast**: se trata de un conjunto de herramientas de modelado y simulación basado en agentes, gratuito, de código abierto y multiplataforma. Repast fue desarrollado en la Universidad de Chicago y presenta múltiples implementaciones en diferentes idiomas. También presenta características adaptativas integradas tales como algoritmos genéticos y regresión.
* **MASON**: es un entorno de simulación de sistemas multiagentes. Fue desarrollado por la Universidad de George Mason y su primera publicación fue en 2003. Hoy en día, sigue actualizándose periódicamente.
* **Swarm**: se trata de un paquete de código abierto que permite la simulación de modelado basado en sistemas multiagentes, siendo útil para simular la interacción entre agentes y su comportamiento colectivo emergente.
* **Netlogo**: es un lenguaje de programación orientado a agentes y un entorno de modelado programable de múltiples agentes. Se encuentra orientado a la enseñanza, por lo que no es tan potente como otras herramientas dedicadas a la investigación. Se puede o descargar la aplicación de escritorio que incluye una biblioteca con diferentes modelados o utilizar la aplicación web. En su página, incluye diversos tutoriales para iniciarse, así como documentación varia. Esta será la herramienta utilizada para crear nuestro modelo.
  1. Modelos de propagación de enfermedades

Los modelos matemáticos son herramientas muy utilizadas en la actualidad. Presentan objetivos como describir, explicar y predecir fenómenos y procesos en áreas tan variadas como medicina, biología, fisiología, epidemiologías entre otras muchas.

Estos modelos se basan en ecuaciones diferenciales y básicamente existen:

* **SI**: los individuos pueden estar sanos (susceptibles) o infectados.

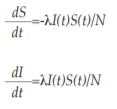


Ilustración 6. Ecuaciones del modelo SI [8]

* **SIS**: estos individuos pasan por los estados sanos, infectados y sanos de nuevo.

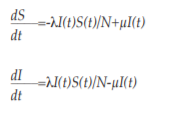


Ilustración 7. Ecuaciones del modelo SIS [8]

* **SIR**: en este caso los individuos sanos pueden infectarse y luego pasan al estado recuperados cuando pasan la enfermedad.

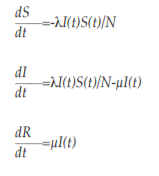


Ilustración 8. Ecuaciones del modelo SIR [8]

El parámetro λ es la tasa de contagio de la enfermedad, mientras que el parámetro μ es la tasa de recuperación.

1. Metodología y herramientas de trabajo
   1. Metodología

La metodología empleada en la realización del proyecto es CRISP-DM, abreviación de “Cross-industry standard process for data mining”. Esta metodología surgió en la década de los 90. En ella se realiza una descripción del ciclo de vida de un proyecto de análisis de datos, de la misma forma a como se realiza un desarrollo de software en ingeniería. Presenta unas fases no rígidas, es decir, pueden realizarse unas antes que otras, aunque normalmente sigan la siguiente disposición:

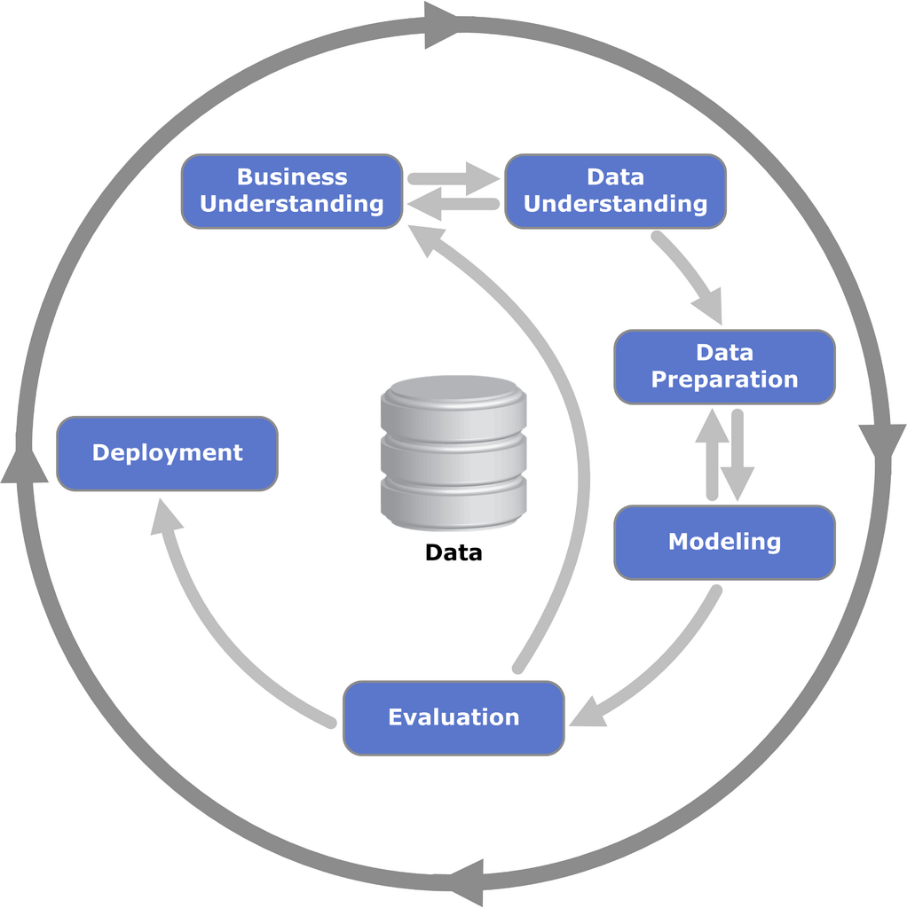


Ilustración 9. Fases de CRISP-DM [1]

A continuación, se van a explicar las fases:

* **Fase 1: Business Understanding**. En esta fase se estudian las necesidades del cliente o el negocio a trabajar. Además, también se estudian los objetivos a alcanzar.
* **Fase 2: Data Understanding.** Se realiza un estudio de los datos para ver cómo se encuentran organizados. Se realizan diferentes actividades para descubrir problemas de calidad en los datos o un determinado conocimiento previo de los datos.
* **Fase 3: Data Preparation.** Se procede a realizar un análisis de datos seleccionando las características necesarias para el conjunto de datos final. Se realizan tareas como selección de tablas, registros y atributos…
* **Fase 4: Modelling.** Se aplican técnicas de modelado para abordar el problema en cuestión. Para la realización de estas técnicas es necesario, muchas veces, volver a los datos debido a ciertos requerimientos específicos.
* **Fase 5: Evaluation.** Se realiza una evaluación para obtener resultados. Se evalúan todos los pasos para ver si existe alguna cuestión de negocio que no ha sido suficientemente considerada.
* **Fase 6: Deployment.** Se trata de la puesta en producción del producto o de la presentación al cliente para que así pueda usarlo.
  1. Herramientas de trabajo
     1. R y RStudio

**R** se trata de un lenguaje de programación orientado al análisis estadístico, siendo un lenguaje muy utilizado en investigación científica (minería de datos, machine learning, bioinformática…). Una de sus ventajas es la posibilidad de cargar librerías y paquetes de diversas funcionalidades. La versión utilizada ha sido la 3.6.2.

El entorno de desarrollo usado para programar en R ha sido **RStudio**, donde se dispone de una consola y de un editor de sintaxis con la función de ejecutar el código. La versión usada ha sido la 1.2.5033.

R y RStudio han sido utilizados en el proyecto para leer los datos, realizar el análisis de estos y conformar el conjunto de datos que finalmente sería utilizado para el modelado.

* + 1. Microsoft Excel

Microsoft Excel se trata de la hoja de cálculo de Microsoft para Windows, MacOS… Presenta diversas características como funciones de cálculo o representaciones gráficas. La versión utilizada es 2004. Junto a R y RStudio ha sido utilizada para el análisis de los datos y para realizar la normalización de los síntomas de los pacientes del conjunto de datos como se verá en el análisis de datos.

* + 1. Github

Para el control de las versiones se ha usado **Github**, una conocida plataforma de desarrollo colaborativo para alojar diversos proyectos. Se permiten crear diversos repositorios y realizar diferentes ramas, además de recuperar los archivos subidos a cada repositorio.

Para acceder a dicha plataforma se pueden usar dos vías, su página web [9] o mediante la aplicación de escritorio **Github Desktop**, la cual ha sido empleada en dicho trabajo con la versión 2.4.3.

* + 1. Netlogo

Por último, pero no por eso menos importante, aparece la plataforma **Netlogo** la cual ha sido utilizada para realizar el modelado del sistema multiagente con el que se realiza la simulación de la expansión de una epidemia.

Esta plataforma ha sido explicada en el Capítulo 2.4.2. Al igual que ocurre con Github, a ella se puede acceder mediante su aplicación web [10] y con su aplicación de escritorio, que puede ser descargada desde su página. En este trabajo se ha utilizado la aplicación de escritorio para no depender del funcionamiento de su web. La versión empleada ha sido Netlogo 6.1.1.

1. Modelado 1

Como se ha comentado anteriormente en “3.1 Metodología”, la metodología empleada va a ser la CRISP-DM. Para comenzar el desarrollo del primer modelado, se va a necesitar un conjunto de datos que aporte las suficientes características para poder simular la epidemia en cuestión.

En este caso, con el tema del COVID19 muy reciente, los datos escogidos van a ser sobre esta enfermedad. Así, para el primer modelado se va a necesitar un conjunto de datos de pacientes en el que se muestre la información sobre los síntomas mostrados al padecer la enfermedad. Este conjunto de datos aporta otra gran cantidad de información que no va a ser empleada ni en este modelado ni en el siguiente.

Otra información necesaria para este modelo va a ser la posibilidad de un ciudadano de conseguir una mascarilla y la probabilidad de recuperación, fallecimiento o contagio. Para estos tres últimos datos, se aportará un archivo con los datos de España y otro con los del mundo.

El modelo matemático empleado será el SIR con la inclusión de fallecidos.

* 1. Requisitos modelo 1
     1. Requisitos funcionales

|  |  |
| --- | --- |
| Requisitos | Descripción |
| Creación de agentes | El usuario podrá crear un determinado número de agentes y con una forma una vez se hayan completado todas las variables de entrada. Deberá pulsar el botón **setup.** |
| Elección de los síntomas | El usuario podrá escoger la probabilidad de aparición de cada síntoma. |
| Elección porcentaje infección inicial | El usuario deberá introducir por pantalla el porcentaje de infectados al inicio de la simulación. |
| Elección de tasa de recuperación, contagio y fallecimiento | El usuario podrá introducir por pantalla las diferentes tasas posibles en forma de porcentaje. |
| Elección de tiempo de recuperación | El usuario escogerá el tiempo de recuperación de la enfermedad. |
| Inclusión de mascarillas | El usuario tendrá la posibilidad de incluir mascarillas, así como introducir el porcentaje de protección de éstas. |
| Carga de datos de contagiado, recuperados y fallecidos | El usuario podrá cargar un archivo con el número de infectados, recuperados y fallecidos. |
| Mostrar tasa de fallecimientos y de recuperación | El usuario podrá mostrar la tasa de fallecimientos y de recuperación cuando haya cargado los datos correspondientes. |
| Ver tasa r0 | El usuario podrá ver en un monitor cómo va variando la tasa r0 a lo largo de la simulación. |
| Ver porcentajes de recuperados, contagiados, fallecidos y susceptibles de la población | El usuario podrá ver diferentes porcentajes de interés dentro de la población. |
| Hacer una simulación | El usuario podrá realizar la simulación de la enfermedad una vez se hayan creado los agentes de simulación Deberá pulsar el botón **go.** |
| Ver una simulación | El usuario podrá visualizar la expansión de la enfermedad en una población. |
| Exportar interfaz | El usuario podrá crear un archivo de imagen de la simulación. |
| Exportar datos de los agentes | El usuario podrá crear un archivo .csv con los datos de los agentes. Esta opción es aportada por el mismo programa. |

Tabla 1. Requisitos funcionales del modelo 1

* + 1. Requisitos no funcionales

|  |  |
| --- | --- |
| Requisitos | Descripción |
| Usabilidad | La interfaz será fácil de usar e intuitiva. El tiempo de aprendizaje del usuario no será demasiado. |
| Dependibilidad | La aplicación no depende de un servidor externo o de conexión a internet. |
| Mantenimiento | El código se encontrará propiamente comentado, recalcando las partes de más difícil comprensión. |
| Documentación | La aplicación tiene un manual de usuario para poder realizar la simulación de forma sencilla |

Tabla 2. Requisitos no funcionales modelo 1

* 1. Diseño modelo 1

En este apartado se va a comentar cómo se ha realizado el modelo 1. Para empezar, este tipo de modelado que se va a realizar solo necesita de un tipo de agente (los humanos). En el 5. Modelado 2 se verá que para este modelo harán falta varios tipos de agentes más.

Cuando se crea un proyecto nuevo en Netlogo, se abre una pestaña como la siguiente:

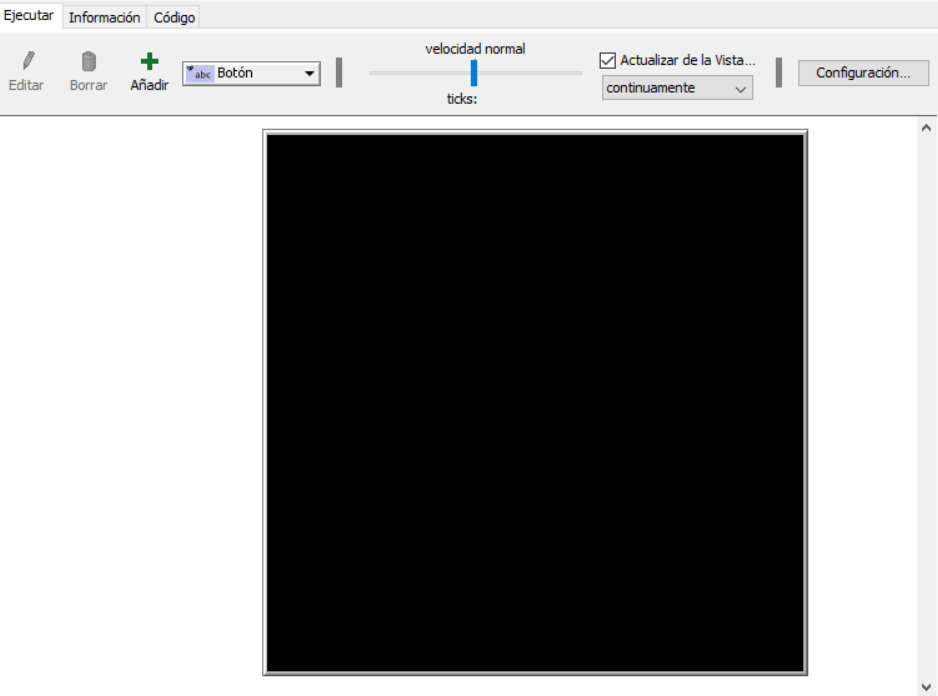


Ilustración 10. Ventana de inicio de Netlogo

Lo ideal antes de continuar, sería ajustar el tamaño de la ventana del modelado, con el fin de garantizar que se pueda ver la simulación, así como todas las ventanas de entrada y los monitores. En el cuadrado negro que se observa en la "Ilustración 10. Ventana de inicio de Netlogo” es donde se va a realizar la simulación. El proceso utilizado ha sido muy sencillo, se empezó añadiendo los agentes sin ningún tipo de funcionalidad y a partir de ahí, mediante una forma progresiva, se fueron complementado las acciones que podían realizar los agentes y las características de estos.

Para crear un número exacto de agentes, se va a añadir una entrada de número, con el fin de que el usuario pueda decidir cuántas personas van a participar en la simulación.



Ilustración 11. Entrada del número de personas

El siguiente paso realizado fue comprobar si se creaban correctamente los agentes, para lo que se añadirá un botón setup, cuya función es “llamar” al procedimiento de creación de agentes.



Ilustración 12. Botón setup de la simulación

También, se quería ver si se producía el movimiento de los agentes. De esta forma se incluirá el botón go, que es el encargado de simular el movimiento de los agentes en el mapa. Este botón permanecerá desactivado hasta la creación de los agentes. Además, no hará falta pulsar cada vez que se quiera avanzar en la simulación.



Ilustración 13. Botón go para movimiento de los agentes

Hasta ahora, se ha realizado lo esencial para la simulación. El siguiente paso será ir añadiendo los diferentes atributos de los agentes. En este entorno de desarrollo, se distinguen atributos de las “tortugas” (forma de denominar de Netlogo a los agentes de simulación) y atributos globales de la simulación.

Los atributos añadidos a los agentes serán:

* Variable booleana que indicará si se encuentra infectado de la enfermedad.
* Variable booleana que indicará si ha curado de la enfermedad.
* Variable booleana que indicará si es susceptible de contraer la enfermedad.
* Variable numérica que dará los días que tarda un agente en recuperarse.
* Variable numérica que contará el número de días que lleva contagiado un agente
* A continuación, se añadirán tantas variables booleanas como síntomas se hayan encontrado en la exploración de los datos (----anexo datos----).
* Por último, otra variable booleana que indicará si un agente ha podido conseguir una mascarilla o no.
* Otras variables de los agentes serán el tamaño (predeterminado en 1.5), la forma y el color. La forma podrá ser escogida en un seleccionador entre un punto, una silueta humana o una tortuga.



Ilustración 14. Seleccionador de la forma

El color variará según si el agente se encuentra infectado, es susceptible o se ha curado. Para ello, se ha configurado un procedimiento de asignación del color.

Como se puede observar, no se ha añadido una variable que indique si el agente ha fallecido o no. Esto es debido a que el mismo entorno facilita una instrucción (die) para eliminar a los agentes que fallezcan.

Como se ha nombrado anteriormente, también existen variables globales de la simulación:

* Variable numérica que guarda el número de contagios en la simulación.
* Variable numérica que guarda el número de recuperaciones en la simulación.
* Variable numérica que guarda el número de fallecidos en la simulación.
* Variable numérica que guarda el porcentaje de contagios en la simulación.
* Variable numérica que guarda el porcentaje de curados en la simulación.
* Variable numérica que guarda el porcentaje de fallecidos en la simulación.
* Variable numérica que guarda el porcentaje de susceptibles en la simulación.
* Variable numérica que guarda el porcentaje de recuperación según los datos introducidos.
* Variable numérica que guarda el porcentaje de fallecidos según los datos introducidos.
* Variable numérica de apoyo para guardar el número de casos según los datos introducidos.
* Variable numérica de apoyo para guardar el número de recuperaciones según los datos introducidos.
* Variable numérica de apoyo para guardar el número de fallecidos según los datos introducidos.
* Variable numérica que almacena el número de infectados en el instante previo.
* Variable numérica para almacenar la tasa r0
* Variable numérica donde se guarda la tasa de recuperación
* Variable numérica donde se guarda la tasa de nuevas infecciones

Ahora, el paso correspondiente sería añadir una entrada para calcular el número de infectados sobre la población inicial. La infección inicial se hará en el mismo procedimiento de creación de agentes, de tal forma que se generará un número flotante aleatorio de 0 a 100 y si este número es menor que el porcentaje introducido, este agente en cuestión se infectará. Esta forma de calcular la infección es realmente la usada en la mayoría de procesos “aleatorios” de este modelo de simulación. Además, cuando se produzca la infección, se le asignará al agente los síntomas correspondientes según los porcentajes.



Ilustración 15. Ventana de entrada del porcentaje de infectados inicial

Lo próximo a implementar serían los síntomas. Como se comenta en ----anexo datos----, en el conjunto de datos hay una variable dedicada exclusivamente a los síntomas, aunque algunos síntomas son nombrados de diversas formas por lo que se realizó una normalización de los síntomas.



Ilustración 16. Ventanas de entrada de porcentaje de aparición de los síntomas

Las ventanas se encuentran inicializadas con los porcentajes obtenidos del conjunto de datos. Como se ha nombrado ya, la creación de síntomas se basará en números aleatorios.

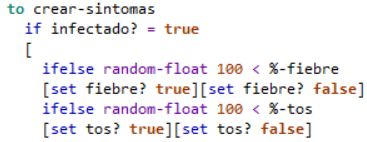


Ilustración 17. Creación de síntomas de los agentes

En la “Ilustración 17. Creación de síntomas de los agentes” se puede ver como para poder crear los síntomas el agente debe haber sido infectado previamente. A continuación, para cada uno de los síntomas, si el número generado es menor que el porcentaje en cuestión, dicho agente presentará ese síntoma. Si por el contrario el número es mayor, no lo presentará.

Una vez incluidos los síntomas, se añadirá un botón para cargar los datos de infectados, recuperados y fallecidos en un archivo .txt (Ej: 564 465 50) para poder después mostrar las tasas de recuperación y letalidad.

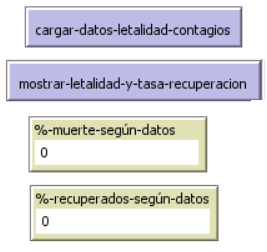


Ilustración 18. Carga y muestra de tasa de letalidad y recuperación

Estos datos calculados podrán ser usados en la simulación en las siguientes ventanas de entrada.





Ilustración 19. Ventanas de entrada de porcentaje de recuperación y fallecimiento

Agregados a estas ventanas, se encontrará la que permitirá incluir el porcentaje de contagio en la población.

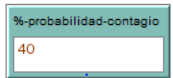


Ilustración 20. Ventana de entrada de porcentaje de contagio

La última información que queda por recibir de un agente es el tiempo de recuperación que va a necesitar.



Ilustración 21. Deslizador tiempo de recuperación

Para obtener el tiempo de recuperación de un agente, se hallará un número entre 0 y el doble del tiempo introducido.

La siguiente implementación será la manera de infectar de los contagiados, la posibilidad de fallecer y de recuperarse y la posibilidad de obtener mascarillas.

Para obtener una mascarilla y que éstas tengan un efecto, se introducirán por pantalla dos porcentajes más:

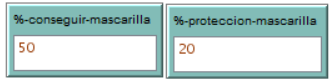


Ilustración 22. Porcentajes referidos a las mascarillas

Para la obtención de una mascarilla el agente tiene que ser susceptible y la conseguirá de forma aleatoria (se generará un número y si éste es menor al porcentaje, la obtendrá).

En el proceso de recuperación, por cada día pasado el contado de tiempo de infección aumentará. Si esta variable es mayor que el tiempo de recuperación de dicho agente, este podrá recuperarse. Con el fallecimiento pasa lo mismo, si la infección lleva en el agente más días que lo estipulado en su tiempo de recuperación, este podrá fallecer.

Cuando un agente infectado se mueva en el entorno, será capaz de infectar a otros agentes cercanos, siempre y cuando, éstos no estén infectados ni hayan pasado la enfermedad ya. Luego, hay que tener en cuenta que ciertos agentes pueden presentar mascarillas, reduciendo así su probabilidad de ser contagiado.

Respecto al movimiento y las acciones de los agentes, todos podrán moverse (en este modelo) con libertad, independientemente de si están o no infectados. Además, los contagiados podrán infectar y recuperarse o fallecer. Los susceptibles tendrán la posibilidad de conseguir mascarillas.

Para acabar, se deberán calcular los porcentajes de infectados, recuperados, susceptibles y fallecidos en la población, así como la tasa r0. Esta tasa r0 será calculada siguiendo las ecuaciones del modelo matemático de propagación de enfermedades SIR, para lo que se tendrá que tener en cuenta la tasa de nuevos infectados y recuperados por cada día (unidad de tiempo de la simulación) como se puede ver en la “Ilustración 8. Ecuaciones del modelo SIR [8]”.

Otro aspecto introducido es la exportación de interfaz, que permite guardar la interfaz del programa en un instante determinado.

* 1. Simulaciones de prueba

En este apartado se realizarán dos simulaciones de prueba, aunque realmente el modelado que va a permitir tener varios escenarios es el siguiente, debido a la inclusión de varios parámetros más.

En esta **primera simulación**, no se van a incluir mascarillas y se van a usar los porcentajes de los síntomas obtenidos. También, se usarán los porcentajes de recuperación y de fallecimientos de los datos de España del 3 de junio de 2020 (archivo “covid19\_españa\_03\_06.txt”). Se establecerá un tiempo de recuperación de 30 días y una probabilidad de contagio de 40%. El número de personas iniciales será de 100 con una infección inicial del 5%.

Los datos obtenidos al final de esta simulación son:

* Tasa r0: 1.98
* Porcentaje de recuperados: 79%
* Porcentaje de fallecidos: 17%
* Porcentaje de infectados: 0%
* Porcentaje de susceptibles: 4%

En esta **segunda simulación**, la probabilidad de conseguir mascarillas será del 50% y disminuirán un 30% la posibilidad de contagiarse. Los demás parámetros son los mismos que la simulación anterior para ver la comparación.

Los datos obtenidos al final de esta simulación son:

* Tasa r0: 1.98
* Porcentaje de recuperados: 79%
* Porcentaje de fallecidos: 13%
* Porcentaje de infectados: 0%
* Porcentaje de susceptibles: 8%

En lo que a los datos obtenidos se refiere, no existen unas diferencias abismales, pero en las gráficas de la población, sí se notan grandes cambios.

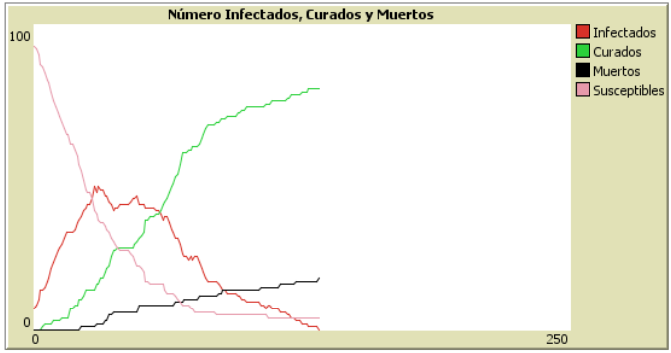


Ilustración 23. Gráfica de población de la simulación 1 del modelo 1

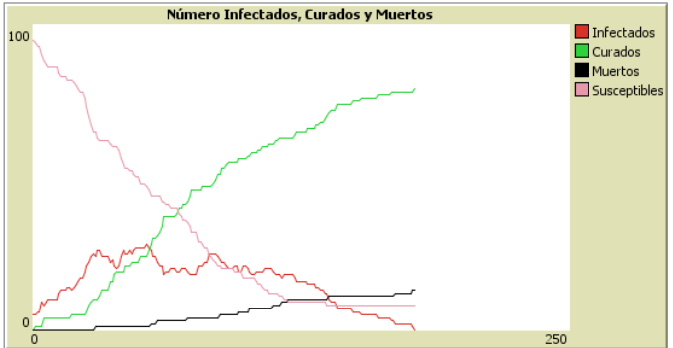


Ilustración 24. Gráfica de población de la simulación 2 del modelo 1

Se puede observar que la curva de infectados en la simulación 1 alcanza su pico en muy poco tiempo, mientras que en la simulación 2, el aumento de los contagios se hace de una forma progresiva.

Estas diferencias habrían provocado que en la simulación 1 los hospitales se hubiesen saturado puesto que habría más contagiados que camas libres. Por el contrario, en la simulación 2, al ir aumentando lentamente el número de contagios, los hospitales no se verían sobresaturados.

Todos los archivos generados en estas simulaciones se encontrarán en la carpeta “modelo 1”, en la que habrá dos carpetas con los archivos de cada simulación y con el modelado en cuestión

1. Modelado 2
2. Conclusiones y mejoras futuras

Referencias

E, B. (2002). *Agent-based modeling: methods and techniques for simulating human systems.*

González Arroyo, M. (2017). *Modelización y simulación en epidemias.* Universidad Complutense de Madrid: Trabajo Fin de Grado.

Juliana. (s.f.). *Imágenes y Noticias*. Recuperado el 15 de Abril de 2020, de https://imagenesnoticias.com/estadisticas-del-sida-en-el-mundo-del-2009/

Mariano Zafra, P. R. (5 de Abril de 2020). *ElPaís*. Recuperado el 5 de Abril de 2020, de https://elpais.com/sociedad/2020/04/09/actualidad/1586437657\_937910.html

Montesinos-López OA, H.-S. C. (2007). Modelos matemáticos para enfermedades infecciosas. *Salud Publica Mex*(49), 218-226.

Netlogo. (s.f.). Obtenido de https://ccl.northwestern.edu/netlogo/

North, C. M. (2006). *Tutorial on Agent-Based Modeling and Simulation PART 2: How to Model with Agents.* California.

OMS. (10 de Febrero de 2020). Recuperado el 13 de Abril de 2020, de https://www.who.int/es/news-room/fact-sheets/detail/ebola-virus-disease

Pané, G. H. (25 de Marzo de 2020). *National Geographic*. Recuperado el 9 de Abril de 2020, de https://historia.nationalgeographic.com.es/a/grandes-pandemias-historia\_15178/5

Pulido, S. (12 de Marzo de 2020). *GacetaMédica*. Recuperado el 9 de Abril de 2020, de https://gacetamedica.com/investigacion/cual-es-la-diferencia-entre-brote-epidemia-y-pandemia/

Schiffmann, A. (s.f.). Obtenido de https://ncov2019.live/

Stackoverflow. (s.f.). Obtenido de www.stackoverflow.com

Ternero, M. d. (s.f.). Obtenido de http://www.dte.us.es/personal/mcromero/masredes/docs/SMARD.0910.mas.pdf

Weiss, G. (2013). *Multiagent Systems* (2nd ed.). Cambridge: Mass: MIT Press.

*Wikipedia*. (s.f.). Obtenido de http://wikipedia.org/

1. Apéndices
   1. Manual de instalación
      1. Requisitos previos

Para usar la aplicación de simulación de epidemias, ya sea el modelo 1 o el modelo 2, es necesario tener instalado el programa Netlogo o usar su versión web. Para los dos casos es necesario acceder a su página [10].

Una vez se haya entrado, se dispondrá de las dos opciones, como se observa en la ilustr…



Ilustración 25. Imagen obtenida de la página de Netlogo [10]

En este caso, se usará la aplicación de escritorio puesto que al cargar los modelos diseñados en la aplicación web se han experimentado diversos errores. Se pulsa en **Download Netlogo** y la página te redireccionará a otra en la que se podrá añadir diversa información, aunque no es necesaria, siendo lo más relevante la elección de la versión a descargar.



Ilustración 26. Imagen para descargar Netlogo [10]

En este caso ha sido usada la versión 6.1.1. Al pulsar en **Download** se cargará la siguiente página, en la que se podrá elegir para qué sistema operativo descargar Netlogo:

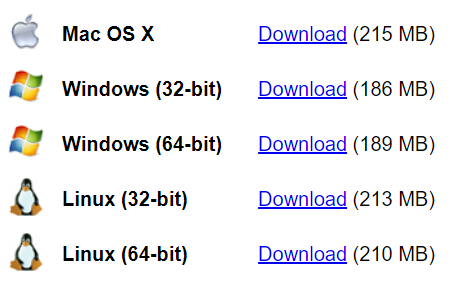


Ilustración 27. Opciones de descarga de Netlogo [10]

Una vez se seleccione la opción necesitada se iniciará la descarga del programa y lo único que habrá que hacer será esperar hasta que termine. Como se puede observar, es un programa que no ocupa mucho, así que no es necesario tener grandes cantidades de memoria en el disco duro.

Cuando se descargue, se inicia y se abrirá un ejecutador para realizar la instalación del programa.

* + 1. Apertura de los modelos

Una vez se inicie el programa Netlogo, la interfaz primera que se obtendrá es:

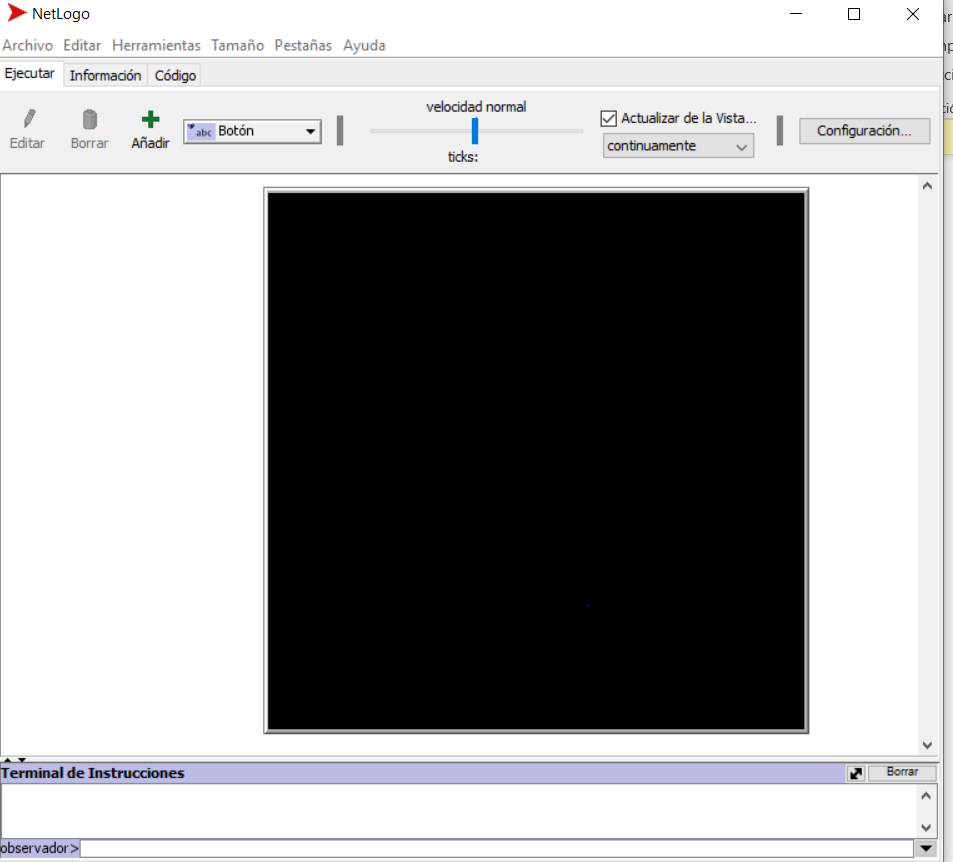


Ilustración 28. Interfaz inicial de Netlogo

Para abrir los modelos simplemente se pulsará en el desplegable **Archivo** para posteriormente pulsar en **Abrir** y seleccionar el modelo que quiera ser usado. Cuando se abra, ya se podrá pasar al Manual de uso Modelo 1o Manual de uso Modelo 2.

* 1. Manual de uso Modelo 1

1. Al haber abierto el Modelo 1 se iniciará una interfaz como la de la

En ella se pueden observar botones, desplegables (para seleccionar la forma del agente durante la simulación), monitores (para ver la evolución de determinados valores), entradas (para introducir valores para realizar la simulación), deslizadores (para elegir un valor determinado), gráficas (para ver la evolución respecto al tiempo de ciertos valores) y el campo de la simulación (donde se realiza la simulación de la epidemia).

1. Para comenzar se deberán cargar los datos de la letalidad y la tasa de recuperación pulsando el botón **cargar-datos-letalidad-contagios**. Estos datos irán en un archivo .txt en el que se pondrán de la siguiente manera: nºcasos nºrecuperados nºmuertos.

Ejemplo: 472 300 124

En el caso de la simulación se usarán los datos de España que se encuentran en el archivo “covid19.txt”.

1. Se pulsa el botón **mostrar-letalidad-y-tasa-recuperación**.
2. Se muestra en los monitores la tasa de letalidad y la tasa de recuperación.
3. Se pueden usar esos datos para rellenar la entrada **%-probabilidad-muerte** y **%-probabilidad-muerte**
4. Se rellenan los porcentajes de padecer un síntoma. Para ello se usarán los porcentajes de los datos de pacientes que se tiene.

***HACER REFERENCIA CRUZADA A IMAGEN DEL ANEXO DATOS***

1. Se rellenan los campos restantes tales como el **tiempo-recuperación** mediante un deslizador, el **número-personas** que da el número de agentes de la población inicial, se escoge en el seleccionador **forma** cómo van a ser vistos los agentes, se introduce el porcentaje de obtención de mascarillas (obtenido mediante los datos de repartición de mascarillas por comunidad autónoma en el archivo “ccaa\_covid19\_mascarillas.csv”) en la entrada **%-conseguir-mascarillas** y su disminución en la probabilidad de ser contagiado, así como esta misma.
2. Paso 7. Se pulsa el botón **setup** para inicializar los agentes.
3. Antes de comenzar la simulación, se recomienda bajar la velocidad de la ésta en un deslizador situado en la parte superior central de la interfaz:

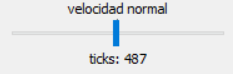


Ilustración 29. Deslizador de la velocidad de simulación

1. Se pulsa el botón **go,** iniciando la simulación.
2. Se pueden observar los resultados mediante una gráfica y viendo los porcentajes de infectados, muertos, recuperados y susceptibles.
3. Mediante el botón **exportar-interfaz** se puede conseguir una imagen de lo realizado.
4. Seleccionando Archivo > Exportar > Exportar mundo se puede conseguir un archivo .csv con en el que se ven las características de los diferentes agentes que intervienen en la simulación.

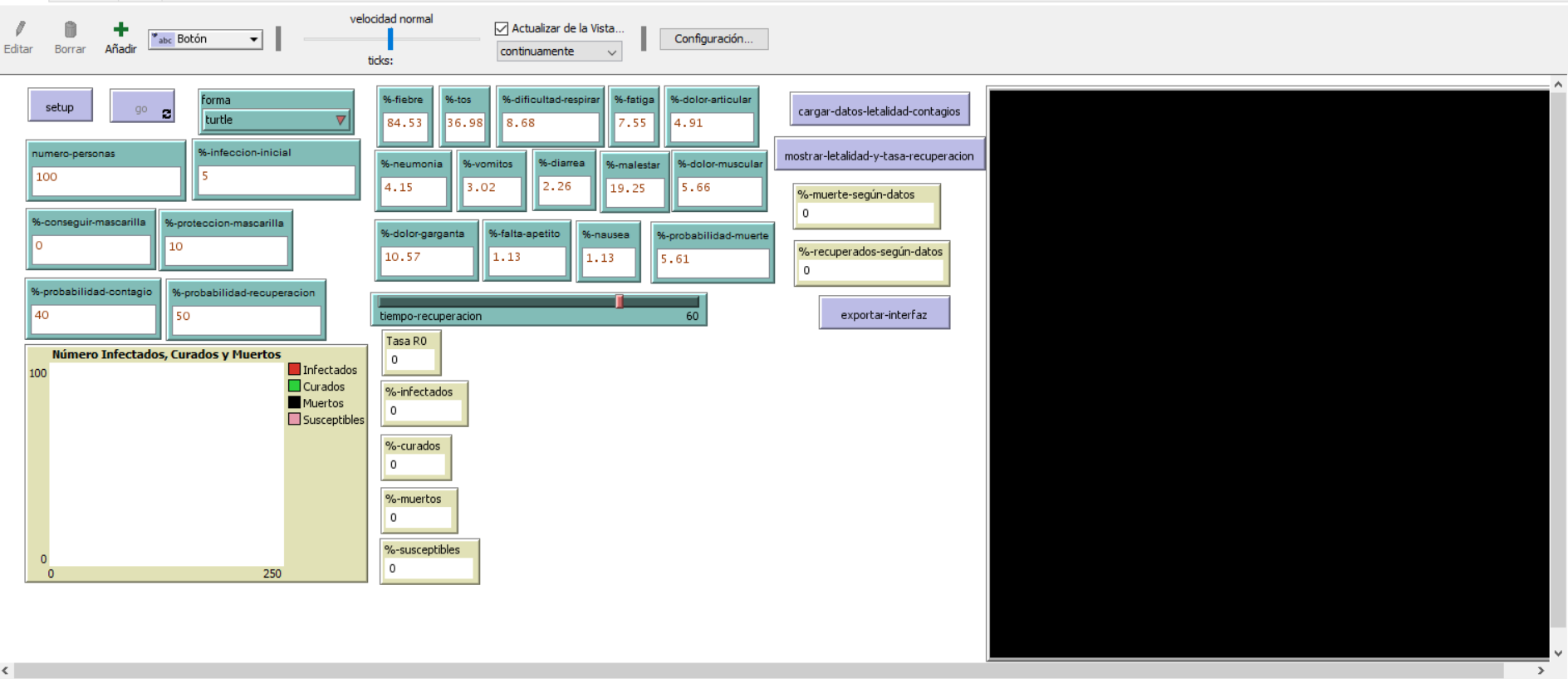


Ilustración 30. Interfaz inicial del Modelo 1

* 1. Manual de uso Modelo 2

1. Anexo de datos

Antes de comenzar, he de recalcar que el archivo íntegro del análisis de datos es “Análisis\_datos\_pacientes.docx”, en este anexo no se incluirá el archivo entero puesto que se va a realizar una especie de resumen incluyendo solo lo que es necesario y eliminando así la parte del código.

Una vez fueron cargados los datos de los pacientes, lo primero era ver cómo se encontraban organizados, es decir, qué variables presentaba cada uno.



Ilustración 31. Variables del conjunto de datos de pacientes de COVID19

Como se puede observar, hay más de 20 variables, de las cuáles no todas serán usadas, además de que alguna será cambiada para facilitar su uso.

A continuación, se van a explicar dichas variables:

-Id: número de identificación para cada paciente

-Case\_in\_country: se trata del número de caso de esa persona en su país.

-Reporting date: fecha en la que se notificó el caso.

-Summary: resumen general del paciente, si visita Wuhán, síntomas, cuándo aparecen los síntomas…

-Location: provincia/ estado de la que pertenece el individuo.

-Country: país del que pertenece el individuo.

-Gender: sexo del individuo.

-Age: edad del individuo.

-Symptom\_onset: fecha de aparición de síntomas.

-If\_onset\_approximated: fecha de aparición aproximada.

-hosp\_visit\_date: fecha de visita al hospital.

-International\_traveller: indica si el individuo ha viajado (1) o no (0) fuera de las fronteras de su país.

-Domestic\_traveller: indica si el individuo ha viajado (1) o no (0) dentro de las fronteras de su país.

-Exposure\_start: fecha de inicio de exposición a la enfermedad.

-Exposure\_end: fecha de fin de la exposición.

-Traveler: indica si el individuo ha viajado o no.

-Visiting\_Wuhan: indica si el individuo visita Wuhán

-From\_Wuhan: indica si el individuo proviene de Wuhán

-Death: indica si el individuo ha muerto.

-Recovered: indica si el individuo se ha recuperado de la enfermedad.

-Symtom: síntomas que han aparecido en el individuo.

-Source: fuente de donde se consiguen los datos.

-Link: url de los datos

En el análisis original, se comenzaba estudiando las distintas variables interesantes para los modelados. En este caso, se ha de comentar primero que no todos los pacientes tenían todos sus campos con datos, así que antes de realizar el estudio, se eliminaron aquellos pacientes que presentaban el género, la edad o los síntomas vacíos. En algunas otras características también había algunos pacientes con los campos vacíos, aunque esto era debido a que algunos países, por ejemplo, el no visitar un hospital lo denotaban con un NA o simplemente sin rellenar la celda del paciente. Por ello, antes de realizar este cribado había 3257 pacientes. Una vez hecho, restan 265.

El objetivo es montar un nuevo dataset con aquellos datos que van a ser necesarios para el modelado. Algunas de las variables que hay van a ser modificadas. Por ejemplo, el día de la visita al hospital va a ser cambiado por un booleano (verdadero, 1, o falso, 0) de si el paciente visitó el hospital. Se añaden las variables edad, género y síntomas.

Por último, antes de realizar la exploración de los datos, se advirtió que no todos los síntomas se encontraban dispuestos de la misma forma, por lo que se debían normalizar. A continuación, se ven algunos de los síntomas que aparecían en el conjunto de datos:

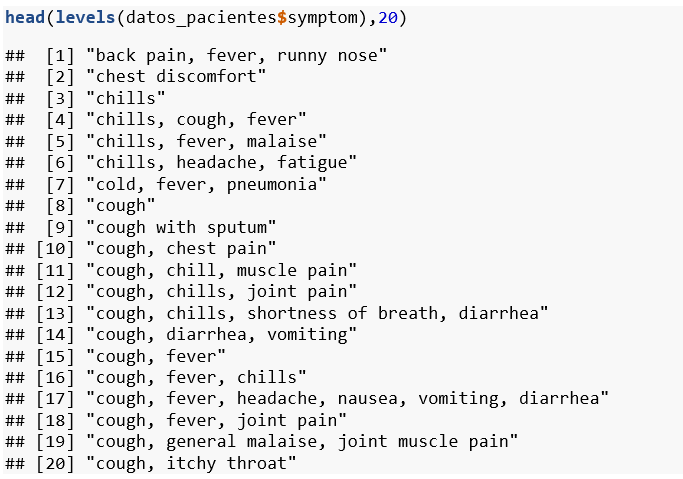


Ilustración 32. 20 primeros síntomas de los pacientes con COVID19

El siguiente paso era normalizar estos síntomas de tal forma que cada paciente tuviera una celda por síntoma y una variable booleana que indicara si presentó dicho síntoma o no. De esta manera, se generará un conjunto de datos donde para cada paciente aparezcan todos los síntomas y “true” si lo padece o “false” en el caso contrario. Se van a tener algunas consideraciones en la realización de esta normalización:

-La primera de ellas es que los datos provienen de diferentes países por lo que existen síntomas nombrados de diferentes formas. Así por ejemplo existían pacientes con el síntoma “tired” que en el dataset ha sido clasificado en “fatigue”, entre otros muchos ejemplos.

-La segunda es que en esta clasificación no se va a atender al grado del síntoma. Al normalizar los datos algunos eran del estilo “high fever”, el cual no tendría cabida en el dataset, por lo que se le añade un “true” a la variable “fever”.

-La tercera es que algunos síntomas podían ser englobados en otros y aparecían en pocos pacientes. Así que para no sobrecargar el dataset, no se ha añadido ese síntoma y al paciente que lo presentaba se le ha puesto el síntoma más general y que englobaba el mencionado anteriormente.

-Y, por último, también ocurría el caso contrario a la tercera consideración. Algunos pacientes tenían “flu symptoms”, por lo que se ha optado en dividirlo en los síntomas que aparecen en el dataset.

Ahora se verá como queda el conjunto de datos creado:

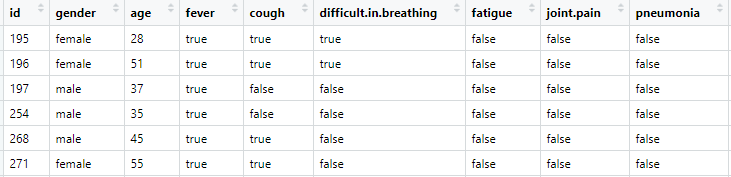


Ilustración 33. 6 primeras filas del conjunto de datos creados (1)

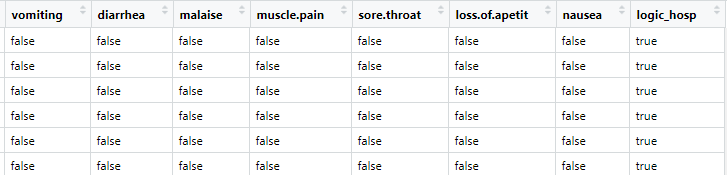


Ilustración 34. 6 primeras filas del conjunto de datos creados (2)

Para empezar la exploración de los datos, se iniciará con la edad. La media de esta variable es 54.39 y su mediana es 55. También se ha realizado el histograma:

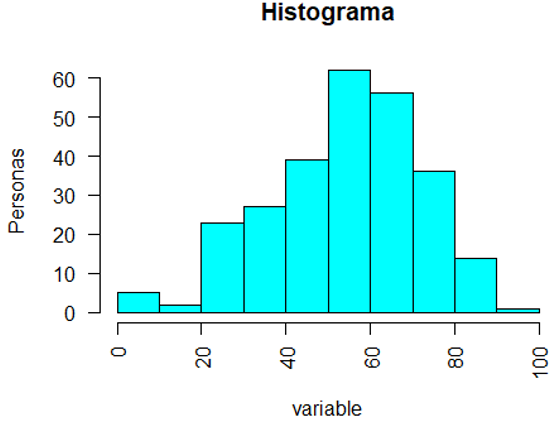


Ilustración 35. Histograma de la variable edad

Como es sabido, esta enfermedad afecta con mayor severidad a las personas de mayor edad, por lo que a la hora de realizar el modelado se añadirá un aumento en el porcentaje de fallecer para las personas mayores a 60 años,

Ahora se verá la distribución del síntoma fiebre.

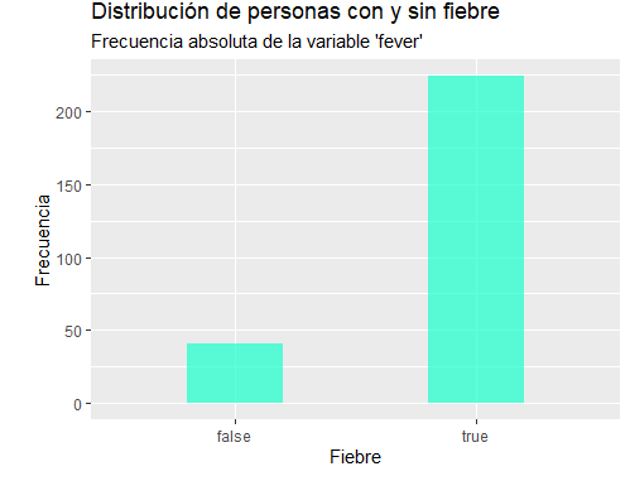


Ilustración 36. Distribución de la población con el síntoma fiebre

Se puede observar que este síntoma es muy frecuente en esta enfermedad.

Para ver el porcentaje exacto de cada síntoma, se ha realizado una tabla donde se ve cada síntoma y su porcentaje, que serán los introducidos tanto en el modelo 1 como en el 2.

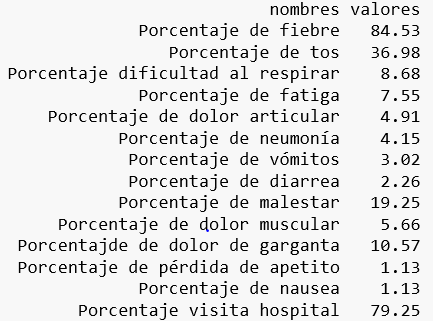


Ilustración 37. Síntomas y sus porcentajes de aparición en la población

El porcentaje de neumonía es tan bajo debido a que, como los pacientes eran de diferentes países, algunos incluían neumonía como un síntoma y otros lo añadían en el “summary” del paciente.

También, de estos datos se puede se puede obtener la probabilidad de fallecer, la probabilidad de recuperación y la de viajar:

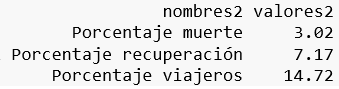


Ilustración 38. Otros porcentajes de utilidad

En \_\_\_\_\_\_\_\_\_ se observa que el porcentaje de fallecer y de recuperarse no son datos representativos, por lo que para la realización del modelado se usarán o los datos mundiales o los de España.

A día 3 de junio, en España el porcentaje de recuperación es de 62.67% y el de fallecimiento de 11.31%, mientras tanto a nivel mundial el porcentaje de recuperación es de 44.44% y el de fallecimiento de 5.91%. Esta variación se debe a que la enfermedad llegó antes a España que a otros lugares del mundo.

Por último, para ver la probabilidad de conseguir una mascarilla se usarán los datos en España. Para obtener el número de mascarillas se usa el siguiente enlace. En esta página se da información en forma de conjunto de datos sobre la epidemia del COVID19 en España. El número de mascarillas repartidas el día de la consulta al conjunto de datos (9 de mayo de 2020) era de 27605827 unidades, por lo que teniendo en cuenta que la población española es de 47100396 habitantes, se obtiene un porcentaje de 58.61%.