MINISTERIO DE ECONOMÍA, INDUSTRIA Y COMPETITIVIDAD

CDTI Centro para el Desarrollo Tecnológico Industrial

DANTIA

TECNOLOGÍA

MEMORIA GENERAL DEL PROYECTO

CollaborativeHealth

PREVENCIÓN DE ENFERMEDADES INFECCIOSAS BASADA EN EL ANÁLISIS INTELIGENTE EN RRSS Y PARTICIPACIÓN CIUDADANA

2. OBJETIVOS

El objetivo principal del presente proyecto se centra en la creación una plataforma de monitorización de enfermedades infecciosas a través del análisis de datos procedentes de i) comentarios y contribuciones en redes sociales como Twitter y Facebook, ii) datos provenientes de fuentes oficiales y iii) a partir de la participación activa de la ciudadanía para ayudar con la planificación operacional a las administraciones públicas, entes sanitarios y ciudadanía a monitorizar posibles casos que puedan afectar a la sanidad pública.

La información recolectada será presentada a profesionales sanitarios de instituciones públicas mediante indicadores de rendimiento (*KPIs*) para que puedan tomar medidas estratégicas acerca de cómo actuar ante posibles nuevos brotes o epidemias.

2.1 Objetivos técnicos específicos del proyecto.

Los objetivos técnico-específicos del proyecto son:

- **(OB1) Desarrollo de una ontología en el dominio de las enfermedades infecciosas.**

  Disponer de una ontología de nivel general y/o varias ontologías aplicadas a enfermedades concretas ofrece una base sólida para conocer términos asociados a éstas. Este objetivo es la búsqueda, evaluación y construcción (en caso de no encontrar fuentes apropiadas) de ontologías relacionadas con enfermedades infecciosas y sus síntomas.

  Para ello, será necesario reflejar los diferentes umbrales, ponderaciones y normas que regirán la creación de contenidos por parte de los ciudadanos, así como las pruebas de calidad establecidas para los mismos: muestreos, validaciones, chequeos, contra-chequeos, etc.

  Se dedicará una particular dedicación a la elaboración de métricas precisas de respetabilidad de los usuarios dada la vulnerabilidad de la plataforma a la incorporación de “evidencias” no contrastadas.

- **(OB2) Desarrollo de una plataforma de extracción de datos relacionados con el dominio de enfermedades infecciosas en redes sociales.**

  Esta plataforma incluye la recogida, clasificación y almacenamiento de estos datos en repositorios para su posterior análisis Este sistema podrá ser configurado mediante distintas fuentes de datos, siendo *Twitter* la más importante debido a sus características previamente mencionadas.

  En este proceso se fijarán los esquemas necesarios para poder añadir al sistema otras fuentes de datos además de las inicialmente implantadas. Para ello se aplicarán técnicas de *Procesamiento del Lenguaje Natural*. Se aplicarán técnicas de reconocimiento de entidades nombradas, extracción de información y minería

de opiniones. Además, se desarrollarán nuevas técnicas sobre el análisis de los datos y el descubrimiento de nuevos patrones relativos a factores estacionales y geográficos.

- **(OB3) Desarrollo de un sistema de filtrado de las fuentes obtenidas a través de parámetros de confianza y reputación basado en *BlockChain***

  Con el objetivo de poder priorizar las fuentes más relevantes y poder descartar falsos positivos, además de hacer transparente y pública la información obtenida. Será necesario el diseño de algoritmos de evaluación de la confianza de las distintas fuentes de datos obtenidas, así como de la calidad de las publicaciones realizadas en las redes sociales, teniendo en cuenta la fiabilidad de pruebas aportadas y la experiencia y reputación de los usuarios.

- **(OB4) Desarrollo de un sistema software para el análisis de datos obtenidos** por las distintas técnicas de extracción de datos estructurados y no estructurados a través de un enfoque de *Semantic deep learning* usando para ello la ontología del dominio de enfermedades infecciosas con el objetivo de identificar posibles pautas que indiquen casos que deban de ser alertados a los profesionales de la salud pública.

  Además de la creación de un repositorio de datos dónde aplicar técnicas de aprendizaje a posteriori, y mejorar el modelo de predicción conforme a nuevos datos. Aquí también se desarrollarán distintas tecnologías de procesamiento del lenguaje natural como extracción de información y minería de opiniones.

- **(OB5) Desarrollo del módulo de participación activa de la ciudadanía relacionada con la obtención de datos de enfermedades infecciosas dentro de una determinada área.**

  Los usuarios de la aplicación dispondrán de una herramienta que les permitirá y guiará a la hora de reportar posibles circunstancias que tengan que ver con distintas enfermedades infecciosas.

  Este módulo dispondrá de un sistema de reputación de los usuarios basado en estrategias de gamificación. Como objetivo adicional, este sistema proveería un catálogo de *FAQs* para cada tipo de enfermedad infecciosa monitorizada y así poder servir de referencia para los ciudadanos en función del estadio en el que se encuentren.

- **(OB6) Desarrollo de una plataforma software para la monitorización, en tiempo real, de enfermedades infecciosas dentro de una determinada área** para ayudar a las organizaciones a la actuación temprana de este tipo de brotes.

  Este sistema permitiría crear a los operarios vistas con distintos indicadores de rendimiento (*KPIs*) que les permitan monitorizar de forma eficiente según las necesidades de cada caso. Los usuarios con acceso al sistema podrán configurar un cuadro de mando (*dashboard*) con diferentes métricas y vistas (gráficas de tiempo, mapas de calor, etc.) y la posibilidad de generar informes en distintos formatos para su posterior análisis.

El sistema también proveerá un sistema de alertas configurable que permitirá a las autoridades competentes recibir alertas cuando se infieran posibles causas de un determinado brote.

La arquitectura de la plataforma está dividida en seis módulos principales, que son los siguientes:

1. **Módulo de monitorización de redes sociales.** Desarrollo de un sistema de extracción de datos relacionados con las enfermedades infecciosas a través de datos no estructurados, procedentes de redes sociales. Para ello, se emplearán técnicas de *Minería de Opiniones* y Procesamiento del Lenguaje Natural.
2. **Módulo *crawler* de extracción de datos oficiales.** Desarrollo de un sistema de extracción de datos relacionados con las enfermedades infecciosas a través de datos no estructurados, procedentes de redes sociales. Para ello, se emplearán técnicas de *Minería de Opiniones* y Procesamiento del Lenguaje Natural.
3. **App multi-dispositivo de participación ciudadana.** El módulo a desarrollar consiste en una aplicación multi-dispositivo que permita a los ciudadanos la identificación y notificación de sucesos relacionados con las enfermedades infecciosas.
4. **Módulo de sistema de filtrado y confianza basado en BlockChain**. Esta Esta Esta tarea consiste en el desarrollo y validación de un conjunto algoritmos de clasificación que permitirán priorizar y descartar entre los obtenidos a través de la herramienta de participación ciudadana y los datos provenientes de las redes sociales. Este sistema se basará en la tecnología de cadena de bloques (*Blockchain*) creando un listado público y transparente de fuentes registradas.
5. **Módulo *Semantic Deep Learning* para el análisis inteligente de datos, dirigido por ontologías.** Este módulo consiste en el diseño de una solución que permita analizar los datos, desde un enfoque semántico, con el fin de extraer patrones significativos. A nivel técnico, se creará una API rest para la correcta explotación de los mismos.
6. **Módulo para la monitorización inteligente de enfermedades en tiempo real.** Este módulo consiste en el diseño de una aplicación multidispositivo que permita a los profesionales de la salud monitorizar, de forma activa, los datos inferidos relacionados con las enfermedades infecciosas. Bajo un enfoque de Semantic Deep-Learning, se procesarán los datos obtenidos a través de los módulos anteriores con el objetivo de poder realizar modelos estadísticos que serán mostrados a los profesionales de la salud mediante un *Dashboard* configurable.

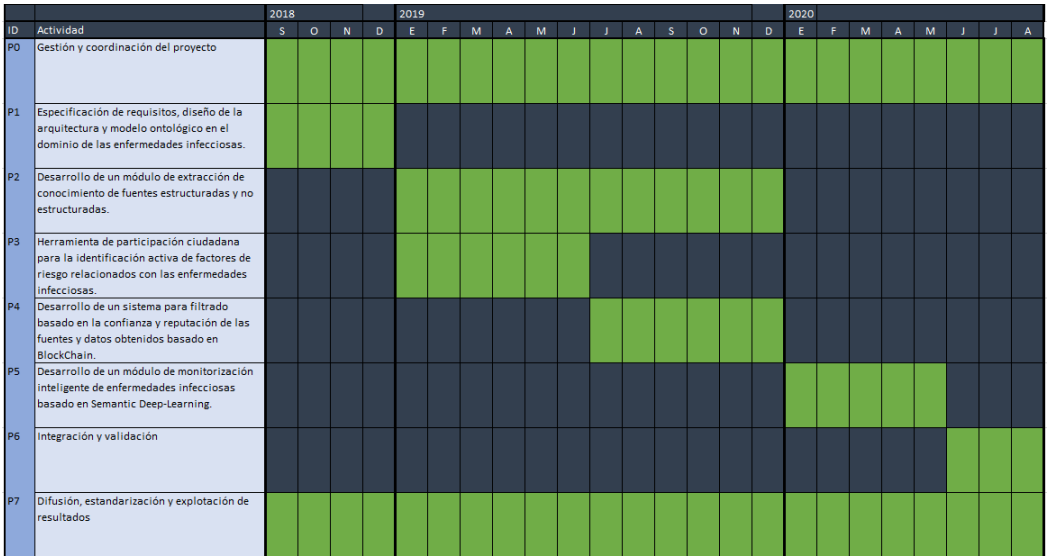

Figura 2. Cronograma general del proyecto