



Construcción de un modelo para predecir la mortalidad en pacientes en diálisis renal

Autor:

Lic. Ezequiel Scordamaglia

Director:

Nombre del Director (pertenencia)

*Esta planificación fue realizada en el curso de Gestión de proyectos
entre el 17 de octubre de 2023 y el ———COMPLETAR———.*

Índice

1. Descripción técnica-conceptual del proyecto a realizar	5
2. Identificación y análisis de los interesados	6
3. Propósito del proyecto	7
4. Alcance del proyecto	7
5. Supuestos del proyecto.	7
6. Requerimientos	8
7. Historias de usuarios (<i>Product backlog</i>).	8
8. Entregables principales del proyecto	9
9. Desglose del trabajo en tareas	10
10. Diagrama de Activity On Node.	10
11. Diagrama de Gantt	11
12. Presupuesto detallado del proyecto	14
13. Gestión de riesgos	14
14. Gestión de la calidad	15
15. Procesos de cierre	15

Registros de cambios

Revisión	Detalles de los cambios realizados	Fecha
0	Creación del documento	17 de octubre de 2023
1	Se completa hasta el punto 4 inclusive	26 de octubre de 2023
2	Se completa hasta el punto 9 inclusive	3 de noviembre de 2023
3	Se completa hasta el punto 12 inclusive	13 de noviembre de 2023

Acta de constitución del proyecto

Buenos Aires, 17 de octubre de 2023

Por medio de la presente se acuerda con el Lic. Ezequiel Scordamaglia que su Trabajo Final de la Carrera de Especialización en Inteligencia Artificial se titulará “Construcción de un modelo para predecir la mortalidad en pacientes en diálisis renal”, y consistirá en el desarrollo de un modelo predictivo, como prueba piloto, que prediga el nivel de riesgo de mortalidad de un paciente que se encuentre en tratamiento de diálisis renal. Este desarrollo se enmarca en la actualización tecnológica de la organización Grupo DUAM, donde se busca aplicar inteligencia artificial y análisis de datos a nuevos desarrollos que ofrezcan valor agregado a los clientes. El mismo tendrá un presupuesto preliminar estimado de 600 h de trabajo y \$3.864.000, con fecha de inicio 17 de octubre de 2023 y fecha de presentación pública —COMPLETAR—.

Se adjunta a esta acta la planificación inicial.

Dr. Ing. Ariel Lutenberg
Director posgrado FIUBA

Eugenio Bellia
Grupo DUAM

Nombre del Director
Director del Trabajo Final

1. Descripción técnica-conceptual del proyecto a realizar

Este proyecto, planteado como trabajo final, se realizará para la organización Grupo DUAM, la cual desarrolla software a medida para distintos clientes. El objetivo principal de este proyecto es incursionar en desarrollos que utilicen las nuevas tecnologías, como son la inteligencia artificial y el análisis de datos, con el fin de proporcionar soluciones innovadoras a problemas preexistentes. En esta línea, se tomó la decisión de colaborar con uno de sus clientes, una empresa médica que opera centros de atención en todo el país, para crear un primer prototipo que haga uso de estas tecnologías y que, al mismo tiempo, les aporte un valor agregado. Ellos poseen un sistema de gestión de pacientes donde registran datos médicos asociados a cada paciente, como son estudios médicos, hospitalizaciones, medicamentos prescritos, y muchos otros. Es fundamental en la medicina conocer el estado de salud de los pacientes y evaluar los riesgos asociados a ellos. Esto le permite a los profesionales de la salud ajustar el tratamiento y los medicamentos que prescriben. Pero predecir los riesgos que puede tener un paciente no es tarea sencilla. La complejidad de las condiciones de salud y la gran cantidad de datos clínicos disponibles hacen extremadamente difícil realizar una predicción de mortalidad de manera manual. En este contexto, surge la idea de ofrecerles desarrollar un modelo predictivo de mortalidad que pueda aprovechar el poder de la inteligencia artificial y del análisis de datos, para realizar un pronóstico rápido y preciso del nivel de riesgo que tiene un paciente.

En este trabajo, se propone partir por la creación de un modelo predictivo de mortalidad que sirva de ejemplo y permita ir desarrollando toda la infraestructura necesaria para que el modelo quede disponible para los usuarios. Se utilizará una plataforma de gestión de modelos y se desarrollará una interfaz por servicio web y un proceso automatizado que solicite las predicciones. Para el desarrollo de este modelo predictivo puntual, se contará con información de miles de pacientes que fueron atendidos por esta empresa médica y sirven de referencia para el proceso de entrenamiento. Se deberán procesar los datos disponibles, entrenar varios modelos de *machine learning* y *deep learning*, evaluar sus métricas y seleccionar los candidatos que logren los mejores resultados. El desarrollo de este tipo de modelos predictivos, administrados por una plataforma de gestión integral y accesibles por servicio web, permitirá adentrarse en un mercado en constante evolución que demanda soluciones innovadoras. En la figura 1 se muestra un diagrama de la solución propuesta para resolver el caso piloto.

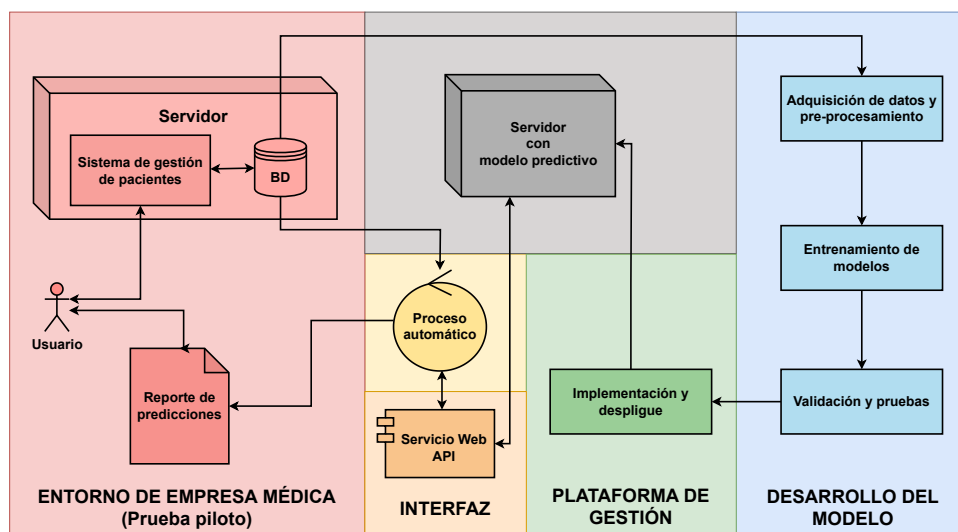


Figura 1. Diagrama en bloques del sistema

Si bien parte del proyecto se limitará a dar solución a un problema puntual en el ámbito médico, el objetivo final es poder brindar soluciones a distintos clientes, incorporando los beneficios de las nuevas tecnologías. El uso de la inteligencia artificial y el análisis de datos está en auge actualmente, y muchas industrias de distintos campos ya están logrando resultados exitosos. En la medicina ya se han utilizado para la predicción del desarrollo de enfermedades como la diabetes o el cáncer, y también para detectar formas extrañas en imágenes. Esto se debe a que los modelos utilizados tienen la capacidad de reconocer patrones en grandes conjuntos de datos. Para poder generar predicciones es necesario utilizar datos históricos en el entrenamiento, donde los modelos descubren relaciones entre ellos. Luego, por métodos estadísticos, logran predecir eventos futuros basados en dichas relaciones. Dado que ya se cuenta con los datos históricos de miles de pacientes, este tipo de tecnología podría servir perfectamente para resolver la problemática actual de este cliente. El desarrollo de este proyecto tendría dos resultados importantes: dotar a la organización de los conocimientos y herramientas necesarias para resolver problemas que requieran el uso de las nuevas tecnologías, y obtener un producto que resuelva una problemática real, ayudando al personal médico a guiar su trabajo sobre los pacientes más críticos y, en última instancia, salvar vidas.

2. Identificación y análisis de los interesados

Rol	Nombre y Apellido	Organización	Puesto
Cliente	Eugenio Bellia Fabio Rosellini	Grupo DUAM	Directores
Responsable	Lic. Ezequiel Scordamaglia	FIUBA	Alumno
Colaboradores	Colaborador 1 Colaborador 2	Empresa médica	Gerente de sistemas Asesor de presidencia
Orientador	Nombre del Director	pertenencia	Director Trabajo final
Usuario final	Personal médico y administrativo	Empresa médica	Usuario del sistema

- Cliente: está a favor del desarrollo de este proyecto para impulsar la organización a nuevos mercados y/o ofrecerle desarrollos innovadores a los clientes actuales.
- Colaboradores: el asesor de presidencia no tiene mucho tiempo para dedicarle al proyecto, pero puede aportar conocimientos médicos para seleccionar las características más importantes que tengan relación con la mortalidad de los pacientes. Está a favor del desarrollo del proyecto y aportará lo que sea necesario. El gerente de sistemas tampoco tiene mucho tiempo para dedicarle al proyecto, pero es quien dará el permiso para la utilización de los datos. Como no solicitó el desarrollo del proyecto puede que demore en responder si se le consulta algo referido a la infraestructura.
- Orientador: Puede ayudar mucho en el tratamiento de los datos antes de entrenar los modelos.
- Usuario final: Necesitará que las predicciones que realiza el modelo estén disponibles en todo momento.

3. Propósito del proyecto

El propósito principal de este proyecto es dotar a la organización de nuevas herramientas y capacidades en el ámbito de la inteligencia artificial y el análisis de datos, para poder ofrecer soluciones innovadoras a clientes nuevos o existentes, y poder insertarse en un mercado que se encuentra en constante crecimiento.

4. Alcance del proyecto

El presente proyecto incluye principalmente la instalación y configuración de una plataforma de gestión de modelos, que permita administrar versiones y ejecutar acciones como el despliegue y el re-entrenamiento de modelos. Como prueba piloto, se desarrollará un modelo de predicción de mortalidad, que se incorporará a la plataforma de gestión, y también una interfaz que actúe como nexo entre el modelo predictivo y el usuario final. Para el desarrollo del modelo, se realizarán las tareas de extracción de los datos, su procesamiento, el entrenamiento de los modelos y la evaluación de sus métricas. Una vez que se cuente con un modelo entrenado, se realizarán las tareas referidas al despliegue del modelo. Con el fin de automatizar la llamada al modelo, se desarrollará también un proceso que envíe automáticamente pedidos de predicciones de todos los pacientes activos cada cierto tiempo al modelo, para que el usuario final cuente con un reporte actualizado en todo momento. El resultado del modelo podrá ser binario (hay riesgo o no hay riesgo), o categórico, devolviendo un grado de riesgo para ese paciente. Esto se definirá luego de realizar un análisis de los datos disponibles y de los modelos seleccionados para el entrenamiento.

El presente proyecto no incluye el desarrollo de una plataforma de gestión, sino que se elegirá una existente que cumpla con los requerimientos del proyecto. No se desarrollará una interfaz web orientada al usuario final, sino que se desarrollará un servicio web que sea utilizado por procesos automáticos. No se incluirá un proceso de re-entrenamiento automático del modelo. No se incluye la instalación del modelo predictivo en el entorno productivo de la empresa médica, sino que para el proyecto presentado para este posgrado, se trabajará en un entorno propio de Grupo DUAM que incluirá la plataforma de gestión de modelos, el modelo predictivo y la interfaz por servicio web. Lo único que se instalará en el entorno de la empresa médica será el proceso que recupere datos de los pacientes y llame al servicio web cada cierto período de tiempo para recuperar las predicciones. Esto debe aclararse ya que la instalación en producción de toda la infraestructura no es el objetivo principal de este proyecto, y podría dejarse a cargo del equipo de sistemas de la empresa médica.

5. Supuestos del proyecto

Para el desarrollo del presente proyecto se supone que:

- Se dispone de tiempo en horario laboral para avanzar en el proyecto.
- Se dispone de los equipos necesarios para realizar el procesamiento de los datos y el entrenamiento de los modelos.
- Se dispone de un ambiente donde poder instalar y configurar la plataforma de gestión, disponibilizar la interfaz por web service y desplegar el modelo para realizar predicciones.

- No hay urgencia en el desarrollo del modelo predictivo.
- Se tiene acceso a los datos médicos de los pacientes.
- Se cuenta con un conjunto inicial de datos lo suficientemente grande y representativo para entrenar los modelos de inteligencia artificial de manera efectiva.
- Se cuenta con el apoyo de un responsable médico que dé soporte en la selección de variables médicas a utilizar en el entrenamiento de los modelos y también en la interpretación clínica de las predicciones realizadas.
- Se mantendrá la protección de datos sensibles de los pacientes y de nuestro cliente en todo momento.

6. Requerimientos

1. Requerimientos funcionales

- 1.1. La plataforma de gestión de modelos deberá permitir desplegar modelos en diversos ambientes.
- 1.2. La interfaz por servicio web deberá recibir datos médicos de uno o varios pacientes y devolver las predicciones asociadas a ellos.
- 1.3. El modelo predictivo deberá tener una precisión de al menos un 75 %.
- 1.4. El proceso que solicita predicciones y genera el reporte al usuario deberá poder ejecutarse automáticamente cada cierto período de tiempo.
- 1.5. El reporte de predicciones que le llegue al usuario final deberá tener un formato claro y comprensible.

2. Requerimientos de datos a utilizar

- 2.1. Los datos a utilizar para el entrenamiento del modelo deben estar disponibles para los procesos automáticos que re-entrenen al modelo.
- 2.2. Durante el entrenamiento del modelo se deberá resguardar la confidencialidad de los datos de los pacientes.

3. Requerimientos de documentación

- 3.1. Redactar una memoria técnica con la información del proyecto.
- 3.2. La documentación de la interfaz por servicio web deberá incluir la lista de métodos disponibles con su detalle.
- 3.3. La documentación del modelo predictivo incluirá información sobre el origen de los datos utilizados para el entrenamiento, las características que se usaron, el detalle del modelo seleccionado y la información que haya sobre la explicabilidad del modelo.

7. Historias de usuarios (*Product backlog*)

Para poder estimar la cantidad de puntos de historia que representa cada historia de usuario se usarán los siguientes criterios y grados:

- Dificultad
 - Baja: 1 punto
 - Media: 3 puntos
 - Alta: 5 puntos
- Complejidad
 - Baja: 2 puntos
 - Media: 5 puntos
 - Alta: 7 puntos
- Incertidumbre
 - Baja: 1 punto
 - Media: 5 puntos
 - Alta: 10 puntos

El puntaje final de cada historia de usuario será calculado como el número de la secuencia de Fibonacci más próximo mayor o igual a la suma de las calificaciones en los 3 criterios.

- Como director de la empresa Grupo DUAM, quiero contar con un ambiente en el cual se puedan incluir nuevos modelos predictivos que resuelvan las problemáticas de mis clientes, que sea sencilla y fácil de configurar. ($3+5+10 = 18 \rightarrow 21$ puntos)
- Como desarrollador de la empresa Grupo DUAM, quiero poder reutilizar los componentes que se desarrollen en proyectos de otros clientes. ($3+2+5 = 10 \rightarrow 13$ puntos)
- Como médico de centro de la empresa médica, quiero contar con un reporte actualizado de riesgos de mortalidad de los pacientes de mi centro, para poder realizar ajustes en los tratamientos y medicaciones prescritas. ($1+5+1 = 7 \rightarrow 8$ puntos)
- Como gerente de sistemas de la empresa médica, quiero poder realizar el despliegue del modelo predictivo en producción de manera simple y sencilla. ($3+5+1 = 9 \rightarrow 13$ puntos)

8. Entregables principales del proyecto

Los entregables del proyecto son:

- Documentación del servicio web.
- Documentación del modelo predictivo.
- Modelo predictivo de mortalidad.
- Plataforma de gestión de modelos correctamente configurada.
- Interfaz por servicio web.
- Proceso que solicite predicciones y arme un reporte para el usuario de forma automática.
- Informe final.

9. Desglose del trabajo en tareas

1. Desarrollo del modelo predictivo de mortalidad (280 h).
 - 1.1. Investigación sobre variables médicas que tienen influencia en la mortalidad (20 h).
 - 1.2. Extracción de datos médicos de la base de datos (40 h).
 - 1.3. Procesamiento de datos médicos (40 h).
 - 1.4. Entrenamiento de modelos de *machine learning*(40 h).
 - 1.5. Entrenamiento de modelos de *deep learning*(40 h).
 - 1.6. Evaluación de métricas de los modelos (40 h).
 - 1.7. Selección de los modelos candidatos (20 h).
 - 1.8. Ajuste fino de los modelos seleccionados (40 h).
2. Desarrollo de la interfaz por servicio web (80 h).
 - 2.1. Diseño de la interfaz (20 h).
 - 2.2. Implementación de la interfaz (40 h).
 - 2.3. Pruebas sobre la interfaz para obtener predicciones (20 h).
3. Instalación y configuración de la plataforma de gestión de modelos (60 h).
 - 3.1. Investigación de plataformas de gestión de modelos disponibles (10 h).
 - 3.2. Instalación de plataforma de gestión seleccionada (30 h).
 - 3.3. Configuración de ambientes para despliegue de modelos (20 h).
4. Desarrollo del proceso automático para solicitar predicciones y generar reportes (80 h).
 - 4.1. Diseño del proceso que solicita predicciones automáticamente (20 h).
 - 4.2. Desarrollo del proceso que solicita predicciones automáticamente (40 h).
 - 4.3. Pruebas de automatización de llamadas al modelo (20 h).
5. Documentación (100 h).
 - 5.1. Documentación del servicio web (20 h).
 - 5.2. Documentación del modelo predictivo (30 h).
 - 5.3. Documentación de informe de avance (20 h).
 - 5.4. Informe final del proyecto (30 h).

Cantidad total de horas: 600 h

10. Diagrama de Activity On Node

A continuación, se realiza una breve descripción del Diagrama de Activity On Node que se muestra en la figura 2:

- Las tareas de desarrollo del modelo predictivo son las primeras que debe realizarse, ya que de ellas dependen otras actividades como el desarrollo de la interfaz o el desarrollo del proceso automático.

- Las tareas de instalación y configuración de la plataforma de gestión de modelos puede iniciarse junto con el desarrollo del modelo, pero no podrá finalizar hasta que se cuente con el modelo final.
- A medida que las tareas de desarrollo vayan finalizado, se documentarán los productos terminados.

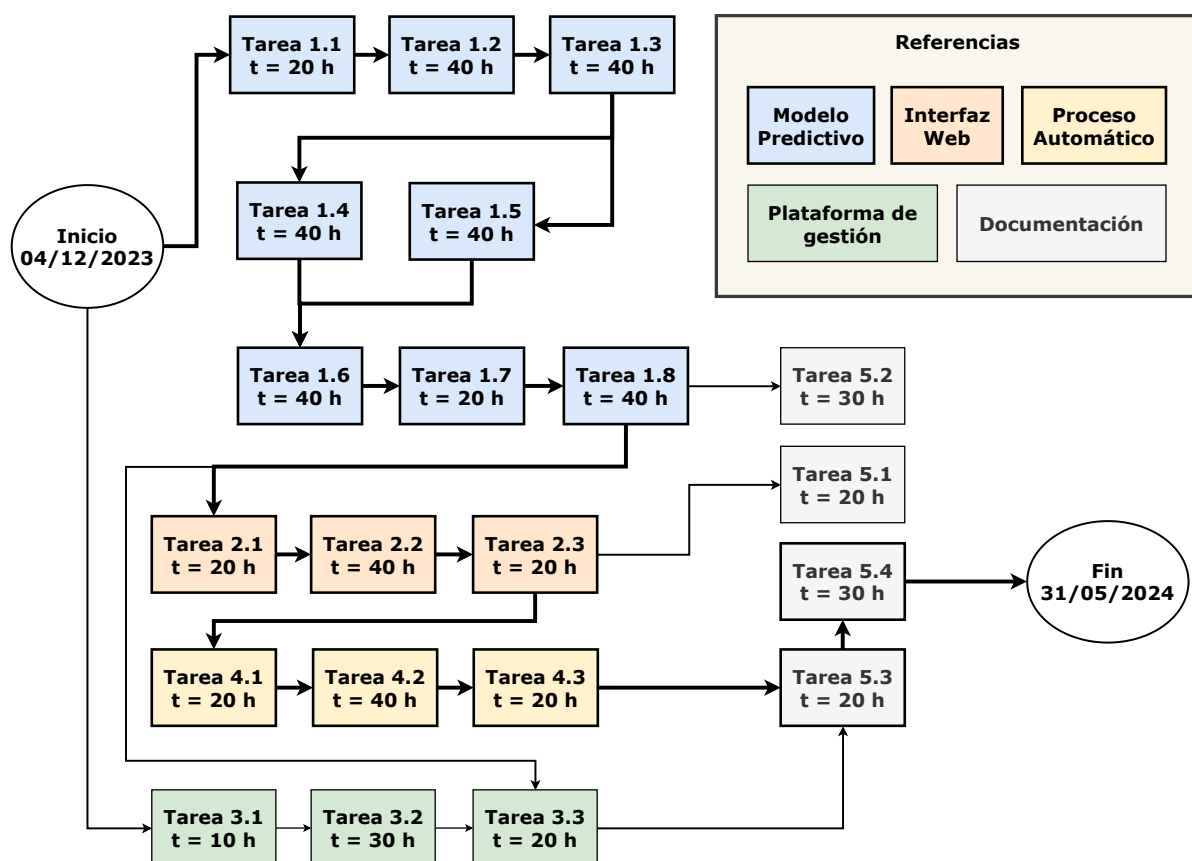


Figura 2. Diagrama de *Activity on Node* con su camino crítico marcado en negrita

El camino crítico nos muestra que como mínimo se requieren 450 horas para finalizar el proyecto.

11. Diagrama de Gantt

En la figura 3, se muestra la tabla del Diagrama de Gantt y en la figura 4 se muestran las tareas desplegadas sobre la línea de tiempo.

	Id	Nombre de tarea	Duración	Comienzo	Fin	Predecesoras
1	1	Investigación sobre variables médicas que tienen influencia en la mortalidad	20 horas	lun 4/12/23	jue 7/12/23	
2	2	Extracción de datos médicos de la Base de Datos	40 horas	vie 8/12/23	mar 19/12/23	1
3	3	Procesamiento de datos médicos	40 horas	mié 20/12/23	vie 29/12/23	2
4	4	Entrenamiento de modelos de Machine Learning	40 horas	lun 1/1/24	mié 10/1/24	3
5	5	Entrenamiento de modelos de Deep Learning	40 horas	jue 11/1/24	lun 22/1/24	3
6	6	Evaluación de métricas de los modelos	40 horas	mar 23/1/24	vie 2/2/24	4,5
7	7	Selección de los modelos candidatos	20 horas	sáb 3/2/24	vie 9/2/24	6
8	8	Ajuste fino de modelos seleccionados	40 horas	sáb 10/2/24	mié 21/2/24	7
9	9	Diseño de la interfaz	20 horas	jue 22/2/24	mar 27/2/24	8
10	10	Implementación de la interfaz	40 horas	mié 28/2/24	vie 8/3/24	9
11	11	Pruebas sobre la interfaz para obtener predicciones	20 días	lun 11/3/24	mar 26/3/24	10
12	12	Investigación de plataformas de gestión de modelos disponibles	10 horas	mié 27/3/24	jue 28/3/24	
13	13	Instalación de plataforma de gestión seleccionada	30 horas	vie 29/3/24	vie 5/4/24	12
14	14	Configuración de ambientes para despliegue de modelos	20 horas	lun 8/4/24	jue 11/4/24	13;8
15	15	Diseño del proceso que solicita predicciones automáticamente	20 horas	vie 12/4/24	mié 17/4/24	11
16	16	Desarrollo del proceso que solicita predicciones automáticamente	40 horas	jue 18/4/24	lun 29/4/24	15
17	17	Pruebas de automatización de llamadas al modelo	20 horas	mar 30/4/24	vie 3/5/24	16
18	18	Documentación del servicio web	20 horas	lun 6/5/24	jue 9/5/24	11
19	19	Documentación del modelo predictivo	30 horas	vie 10/5/24	vie 17/5/24	8
20	20	Documentación de informe de avance	20 horas	lun 20/5/24	jue 23/5/24	17;14
21	21	Informe final del proyecto	30 horas	vie 24/5/24	vie 31/5/24	20

Figura 3. Tabla del Diagrama de Gantt

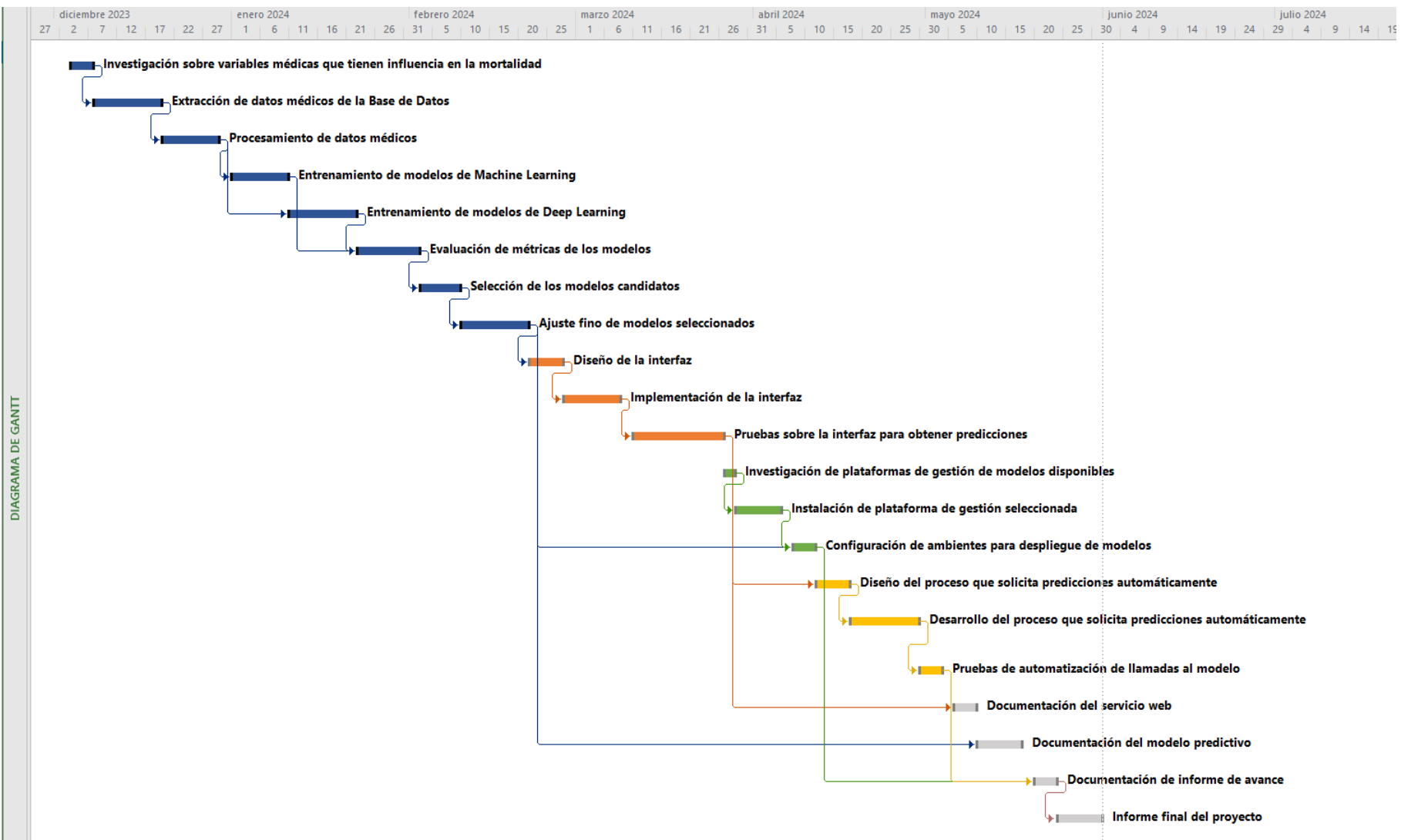


Figura 4. Diagrama de Gantt

12. Presupuesto detallado del proyecto

COSTOS DIRECTOS			
Descripción	Cantidad	Valor unitario	Valor total
Mano de Obra	600 h	\$4.600/h	\$2.760.000
SUBTOTAL			\$2.760.000
COSTOS INDIRECTOS			
Descripción	Cantidad	Valor unitario	Valor total
40 % Costos Indirectos	600 h	\$1.840/h	\$1.104.000
SUBTOTAL			\$1.104.000
TOTAL			\$3.864.000

13. Gestión de riesgos

a) Identificación de los riesgos (al menos cinco) y estimación de sus consecuencias:

Riesgo 1: detallar el riesgo (riesgo es algo que si ocurre altera los planes previstos de forma negativa)

- Severidad (S): mientras más severo, más alto es el número (usar números del 1 al 10). Justificar el motivo por el cual se asigna determinado número de severidad (S).
- Probabilidad de ocurrencia (O): mientras más probable, más alto es el número (usar del 1 al 10). Justificar el motivo por el cual se asigna determinado número de (O).

Riesgo 2:

- Severidad (S):
- Ocurrencia (O):

Riesgo 3:

- Severidad (S):
- Ocurrencia (O):

b) Tabla de gestión de riesgos: (El RPN se calcula como $RPN = S \times O$)

Riesgo	S	O	RPN	S*	O*	RPN*

Criterio adoptado: Se tomarán medidas de mitigación en los riesgos cuyos números de RPN sean mayores a...

Nota: los valores marcados con (*) en la tabla corresponden luego de haber aplicado la mitigación.

c) Plan de mitigación de los riesgos que originalmente excedían el RPN máximo establecido:

Riesgo 1: plan de mitigación (si por el RPN fuera necesario elaborar un plan de mitigación). Nueva asignación de S y O, con su respectiva justificación: - Severidad (S): mientras más severo, más alto es el número (usar números del 1 al 10). Justificar el motivo por el cual se asigna determinado número de severidad (S). - Probabilidad de ocurrencia (O): mientras más probable, más alto es el número (usar del 1 al 10). Justificar el motivo por el cual se asigna determinado número de (O).

Riesgo 2: plan de mitigación (si por el RPN fuera necesario elaborar un plan de mitigación).

Riesgo 3: plan de mitigación (si por el RPN fuera necesario elaborar un plan de mitigación).

14. Gestión de la calidad

Elija al menos diez requerimientos que a su criterio sean los más importantes/críticos/que aportan más valor y para cada uno de ellos indique las acciones de verificación y validación que permitan asegurar su cumplimiento.

- Req #1: copiar acá el requerimiento.
 - Verificación para confirmar si se cumplió con lo requerido antes de mostrar el sistema al cliente. Detallar
 - Validación con el cliente para confirmar que está de acuerdo en que se cumplió con lo requerido. Detallar

Tener en cuenta que en este contexto se pueden mencionar simulaciones, cálculos, revisión de hojas de datos, consulta con expertos, mediciones, etc. Las acciones de verificación suelen considerar al entregable como “caja blanca”, es decir se conoce en profundidad su funcionamiento interno. En cambio, las acciones de validación suelen considerar al entregable como “caja negra”, es decir, que no se conocen los detalles de su funcionamiento interno.

15. Procesos de cierre

Establecer las pautas de trabajo para realizar una reunión final de evaluación del proyecto, tal que contemple las siguientes actividades:

- Pautas de trabajo que se seguirán para analizar si se respetó el Plan de Proyecto original:
 - Indicar quién se ocupará de hacer esto y cuál será el procedimiento a aplicar.
- Identificación de las técnicas y procedimientos útiles e inútiles que se emplearon, y los problemas que surgieron y cómo se solucionaron: - Indicar quién se ocupará de hacer esto y cuál será el procedimiento para dejar registro.

- Indicar quién organizará el acto de agradecimiento a todos los interesados, y en especial al equipo de trabajo y colaboradores: - Indicar esto y quién financiará los gastos correspondientes.