



Clasificación de reclamos de usuarios

Autor:

Ing. Lucas Rivela

Director:

Mg. Lic. Rodrigo Cárdenas (FIUBA)

*Esta planificación fue realizada en el curso de Gestión de proyectos
entre el 28 de febrero de 2023 y el 24 de abril de 2023.*

Índice

1. Descripción técnica-conceptual del proyecto a realizar	5
2. Identificación y análisis de los interesados	6
3. Propósito del proyecto	6
4. Alcance del proyecto	6
5. Supuestos del proyecto.	7
6. Requerimientos	7
7. Historias de usuarios (<i>Product backlog</i>).	8
8. Entregables principales del proyecto	9
9. Desglose del trabajo en tareas	9
10. Diagrama de Activity On Node.	11
11. Diagrama de Gantt	13
12. Presupuesto detallado del proyecto	14
13. Gestión de riesgos	14
14. Gestión de la calidad	16
15. Procesos de cierre	17

Registros de cambios

Revisión	Detalles de los cambios realizados	Fecha
0	Creación del documento	28 de febrero de 2023
1	Se completa hasta el punto 5 inclusive	13 de marzo de 2023
2	Se completa hasta el punto 9 inclusive	20 de marzo de 2023
3	Se completa hasta el punto 12 inclusive	27 de marzo de 2023
4	Se completa el plan	3 de abril de 2023

Acta de constitución del proyecto

Buenos Aires, 28 de febrero de 2023

Por medio de la presente se acuerda con el Ing. Lucas Rivela que su Trabajo Final de la Carrera de Especialización en Inteligencia Artificial se titulará “Clasificación de reclamos de usuarios”, consistirá esencialmente en la implementación de un prototipo de procesamiento de lenguaje natural (PNL) para la clasificación de reclamos de usuarios, y tendrá un presupuesto preliminar estimado de 600 h de trabajo y U\$D7415, con fecha de inicio 28 de febrero de 2023 y fecha de presentación pública el 11 de diciembre de 2023.

Se adjunta a esta acta la planificación inicial.

Dr. Ing. Ariel Lutenberg
Director posgrado FIUBA

Lic. Nicolás Dagosta
Ualá

Mg. Lic. Rodrigo Cárdenas
Director del Trabajo Final

1. Descripción técnica-conceptual del proyecto a realizar

El objetivo de este proyecto será contar con modelos de inteligencia artificial (IA) que contribuyan a agilizar y mejorar la atención al cliente, mediante la categorización automática de los reclamos de los usuarios. Esto permitirá reducir la cantidad de personas necesarias para la clasificación manual, para que se puedan enfocar en resolver los casos más prioritarios.

Esta propuesta se alinea con la estrategia de la compañía de no tener sucursales abiertas al público, en la cual resulta muy importante que los canales de atención online sean rápidos y efectivos, para así poder asegurarle al cliente una buena calidad de servicio.

En la Figura 1 se presenta un diagrama de alto nivel de la solución. Se observa que en primera instancia, los reclamos son recibidos por correo y chat, y luego son registrados en Salesforce. Diariamente, estos datos son replicados en BigQuery (el Data Warehouse) a través del orquestador Apache Airflow. Estos dos últimos componentes están desplegados en Google Cloud Platform.

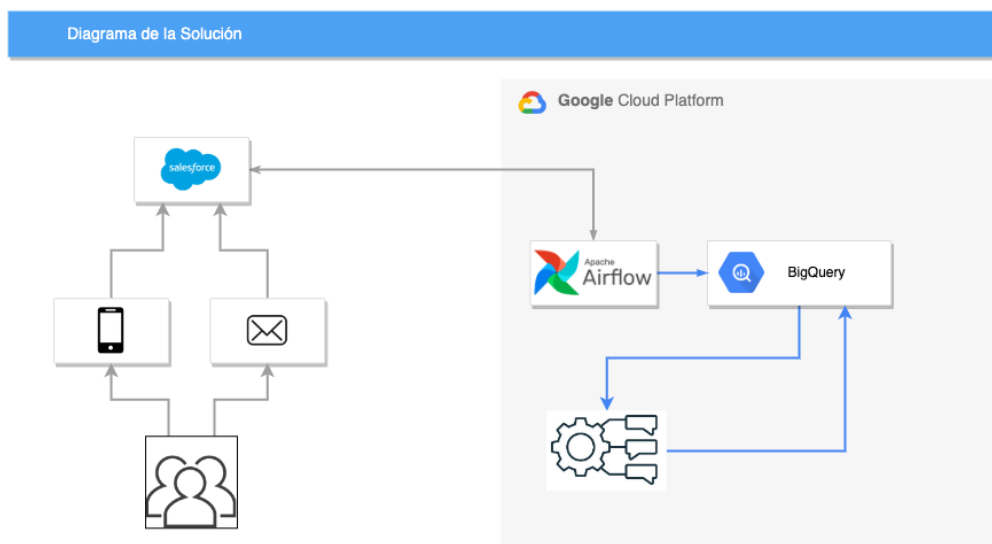


Figura 1. Diagrama de alto nivel.

Con este proyecto se busca desarrollar y entrenar un modelo clasificador, con los casos de 2021 y 2022 que se encuentran en BigQuery y hayan sido etiquetados manualmente. Una vez desarrollado, se desplegará en la misma nube para realizar las inferencias sobre los nuevos reclamos.

Acerca de la empresa:

Ualá es una *fintech* argentina que brinda su servicio de billetera digital a través de una aplicación móvil. Además, provee una tarjeta prepaga Mastercard para poder operar la cuenta. Los servicios principales que provee son:

- Enviar y recibir dinero desde cualquier cuenta bancaria.
- Realizar compras nacionales o internacionales con la tarjeta.
- Extraer efectivo.

- Pagar servicios.
- Pedir préstamos o cuotificar consumos.
- Realizar ventas a través de *mPOS*, *QR* o link de pago.
- Realizar inversiones.

2. Identificación y análisis de los interesados

Rol	Nombre y Apellido	Organización	Puesto
Cliente	Lic. Nicolás Dagosta	Ualá	Gerente de Machine Learning
Impulsor	Lic. Nicolás Dagosta	Ualá	Gerente de Machine Learning
Responsable	Ing. Lucas Rivela	FIUBA	Alumno
Orientador	Mg. Lic. Rodrigo Cárdenas	FIUBA	Director Trabajo final
Usuario final	Área de atención al cliente	Ualá	-

- Impulsor: está interesado en los *insights* que pueda darle este proyecto a nivel tecnológico.
- Orientador: cuenta con una gran experiencia en PNL y tiene buena predisposición.

3. Propósito del proyecto

Optimizar el tiempo que lleva el proceso de clasificación de reclamos y consultas mediante su automatización; y por otro lado, reducir la cantidad de personas necesarias para supervisar esta tarea, de manera tal que puedan enfocarse en el proceso de resolución.

4. Alcance del proyecto

El alcance de este proyecto está orientado a desarrollar un prototipo de solución de software que incluirá las siguientes actividades:

- Obtención de los datos: se corresponde con el análisis de las fuentes de datos disponibles y su integración al proceso que se plantea desarrollar.
- Análisis exploratorio de los datos: se corresponde con las actividades necesarias para generar nuevos *insights*, que sirvan para guiar el desarrollo de los modelos.
- Modelado: se corresponde con la generación de variables a partir de los datos disponibles, que luego serán utilizadas para entrenar los modelos.
- Entrenamiento: se corresponde con la generación de distintos modelos de IA, y su comparación a través de métricas, para encontrar el que mejor se adapte a la problemática de negocio. Estas comparaciones se presentarán en un informe de resultados de los modelos.
- Despliegue: se corresponde con el diseño de la infraestructura necesaria para ejecutar los modelos de IA y su despliegue en un entorno de desarrollo (no productivo).

- Documentación: se corresponde con los documentos de soporte que explican el proceso de modelado, entrenamiento y despliegue de los modelos.

El presente proyecto no incluye:

- La adaptación de los modelos a nuevas categorías de reclamos que puedan surgir durante el desarrollo o una vez finalizado el desarrollo. Se utilizarán casos resueltos de 2021 y 2022 y las categorías a considerar serán las que existan hasta ese momento.
- La integración de los modelos y su despliegue con el ambiente productivo.
- El soporte de la infraestructura desplegada en el entorno de desarrollo.

5. Supuestos del proyecto

Para el desarrollo del presente proyecto se supone que:

- La empresa mantendrá sus operaciones al menos hasta finalizar el desarrollo.
- Se Mantendrá la relación laboral con la empresa al menos hasta finalizar el desarrollo.
- El desarrollador asignado a este proyecto no tendrá enfermedades graves que puedan alterar los plazos estipulados para las actividades planteadas.
- El área de atención al cliente no realizará cambios en su *CRM* que puedan afectar el proceso de réplica de los casos hacia BigQuery, ni que puedan afectar las categorías sobre las cuales se desarrollan los modelos.
- El área de Data de la empresa no cambiará su *Cloud Provider* ni modificará su proceso de réplica hacia BigQuery.
- Se tendrá acceso a infraestructura cloud para el entrenamiento y despliegue de los modelos.
- Las consultas realizadas al área de atención al cliente se resolverán en plazos razonables.

6. Requerimientos

1. Requerimientos funcionales

- 1.1. El sistema debe poder detectar la categoría de un reclamo escrito en lenguaje natural.
- 1.2. El sistema debe poder detectar la categoría de una consulta escrita en lenguaje natural.
- 1.3. El usuario debe poder utilizar los resultados de la clasificación desde una base de datos.
- 1.4. El proceso debe ser capaz de interpretar errores de ortografía.
- 1.5. El proceso debe ser capaz de adaptarse a distinta cantidad de palabras en el mensaje.
- 1.6. La solución debe ejecutarse en forma *batch*, corriendo diariamente y tomando los casos del día anterior.

2. Requerimientos no funcionales

- 2.1. El sistema debe estar desarrollado en lenguaje Python.
- 2.2. El código debe ser versionado con Git.
- 2.3. La solución debe estar desplegada sobre infraestructura de Google Cloud Platform.
- 2.4. La salida de los modelos debe ser almacenada en BigQuery.
- 2.5. El proceso debe ser ejecutado a través del orquestador Apache Airflow.

3. Requerimientos de testing

- 3.1. Se deben generar métricas de performance de los modelos con el dataset de entrenamiento y de prueba.

4. Requerimientos de documentación

- 4.1. Se debe confeccionar un documento con el diseño de la arquitectura de alto nivel.
- 4.2. Se debe confeccionar un documento con el diseño de los modelos de IA.
- 4.3. Se debe confeccionar un documento que especifique los datos que consumen los modelos y su origen.

7. Historias de usuarios (*Product backlog*)

Roles:

- Área de atención al cliente
- Área de *Machine Learning*

Historias de usuario:

1. Como área de atención al cliente quiero clasificar automáticamente los reclamos y consultas para poder brindar una mejor atención a los usuarios.
2. Como área de atención al cliente quiero que el proceso corra diariamente para poder clasificar los reclamos a día cerrado.
3. Como área de atención al cliente quiero poder consumir los resultados de la clasificación desde una base de datos para poder integrarlo con el sistema utilizado en el área.
4. Como área de *Machine Learning* quiero tener documentación que detalle la implementación de la solución para que el proyecto se ajuste a nuestras prácticas de desarrollo.
5. Como área de *Machine Learning* quiero tener métricas de performance de los modelos para tener una referencia y poder monitorear desvíos en el proceso en el futuro.

Criterio para calcular los *story points* de cada historia:

- Criterio A: Cantidad de trabajo a realizar

- Criterio B: Complejidad del trabajo a realizar
- Criterio C: Incertidumbre del trabajo a realizar

Pesos para los criterios:

Nivel	Criterio A	Criterio B	Criterio C
Bajo	1	1	1
Medio	2	3	5
Alto	5	5	8

Story points:

Historia	Criterio A	Criterio B	Criterio C	Peso resultante	Fibonacci
1	5	5	8	18	21
2	2	1	1	4	5
3	2	3	1	6	8
4	2	1	1	4	5
5	2	3	1	6	8

8. Entregables principales del proyecto

Los entregables del proyecto son:

- Plan de proyecto.
- Código fuente (queda reservado para Ualá).
- Modelos de inteligencia artificial (queda reservado para Ualá).
- Documento con el diseño de arquitectura de alto nivel.
- Documento con el diseño de los modelos de inteligencia artificial.
- Documento que especifica los datos que consumen los modelos y su origen.
- Documento con las métricas de evaluación de los modelos.
- Informe de avance.
- Memoria del trabajo.

9. Desglose del trabajo en tareas

Se detallan a continuación todas las actividades que se harán en el proyecto para dar cumplimiento a los requerimientos y su duración estimada:

1. Planificación general del proyecto (50 h)

- 1.1. Redacción de la descripción técnica-conceptual y propósito. (10 h)
- 1.2. Definición del alcance, requerimientos y entregables. (10 h)
- 1.3. Definición de tiempos y presupuesto. (15 h)
- 1.4. Definición de la gestión de riesgos, calidad y procesos de cierre. (15 h)
2. Relevamiento (30 h)
 - 2.1. Reuniones con el área de atención al cliente. (10 h)
 - 2.2. Análisis de las fuentes de datos. (10 h)
 - 2.3. Análisis del proceso de clasificación de reclamos y consultas. (10 h)
3. Análisis exploratorio de los datos (35 h)
 - 3.1. Análisis de la distribución de categorías. (10 h)
 - 3.2. Análisis de la longitud de los mensajes. (10 h)
 - 3.3. Análisis del procesamiento del texto. (15 h)
4. Desarrollo (290 h)
 - 4.1. Programación de *scripts* para obtener datos de entrada de los modelos. (10 h)
 - 4.2. Desarrollo del pipeline de preprocesamiento del texto. (90 h)
 - Módulo para remover partes comunes de los mensajes. (15 h)
 - Módulo para remover palabras redundantes. (15 h)
 - Módulo para remover símbolos. (10 h)
 - Módulo para corregir faltas de ortografía. (40 h)
 - Módulo para armar secuencias. (10 h)
 - 4.3. Desarrollo y entrenamiento de *embeddings*. (60 h)
 - Investigación herramientas de *embeddings*. (20 h)
 - Selección de herramienta de *embeddings*. (20 h)
 - Entrenamiento de *embeddings*. (20 h)
 - 4.4. Desarrollo y entrenamiento de modelos. (130 h)
 - Investigación de arquitecturas de IA que apliquen al problema. (40 h)
 - Entrenamiento de modelo baseline. (10 h)
 - Desarrollo de *scripts* para modelos avanzados. (25 h)
 - Entrenamiento de modelos avanzados - Parte A. (25 h)
 - Entrenamiento de modelos avanzados - Parte B. (25 h)
 - Desarrollo de módulo de escritura de resultados en BigQuery. (5 h)
5. Evaluación y pruebas (35 h)
 - 5.1. Obtención de métricas de los modelos. (30 h)
 - 5.2. Selección del mejor modelo. (5 h)
6. Despliegue (50 h)
 - 6.1. Diseño de la arquitectura del *pipeline*. (10 h)
 - 6.2. Desarrollo de GitHub Actions. (5 h)
 - 6.3. Desarrollo de la imagen de Docker. (20 h)
 - 6.4. Desarrollo de DAG (Directed Acyclic Graph) de clasificación. (15 h)

7. Documentación (30 h)

- 7.1. Elaboración de documento de arquitectura alto nivel. (10 h)
- 7.2. Elaboración de documento con el diseño de los modelos de IA. (10 h)
- 7.3. Elaboración de documento con los datos consumidos. (10 h)

8. Cierre del proyecto (80 h)

- 8.1. Redacción del informe de avance. (20 h)
- 8.2. Redacción de memoria final. (40 h)
- 8.3. Elaboración de presentación para exposición final. (20 h)

Cantidad total de horas: (600 h)

10. Diagrama de Activity On Node

En la Figura 2 se muestra el diagrama *Activity On Node*.

- La ruta crítica es resaltada mediante flechas rojas.
- Las tareas están expresadas en horas.
- A pesar de contar con una sola persona asignada al proyecto, se ejecutan tareas relacionadas en paralelo.
- Los colores de las cajas representan las tareas principales del WBS de la sección 9.

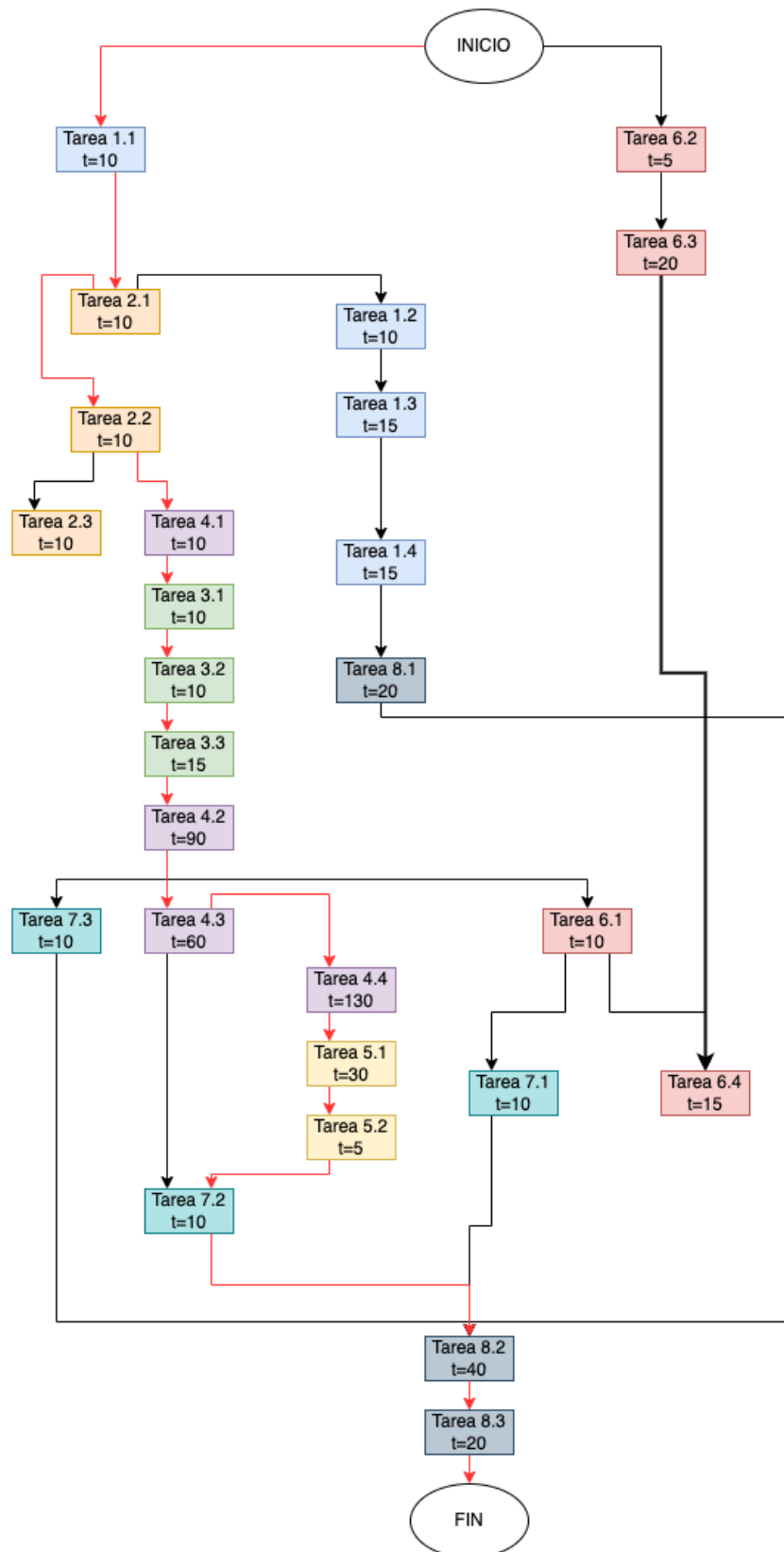


Figura 2. Diagrama de *Activity on Node*.

11. Diagrama de Gantt

En la figura 3 se muestra el diagrama de Gantt del proyecto.

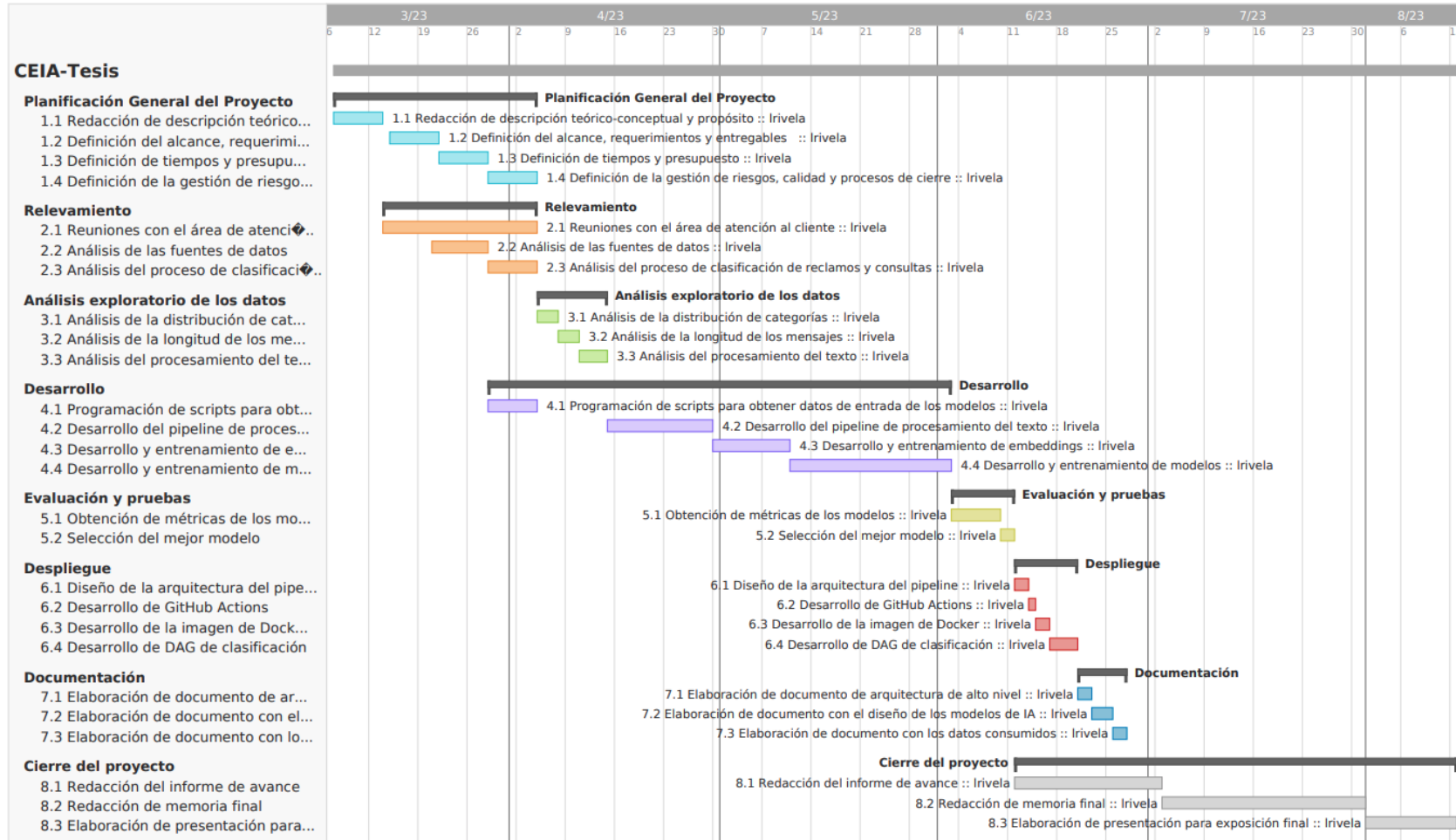


Figura 3. Diagrama de Gantt.

12. Presupuesto detallado del proyecto

A continuación, se detalla la composición del presupuesto del proyecto expresado en dólares estadounidenses y pesos argentinos:

COSTOS DIRECTOS				
Descripción	Cantidad	Valor unitario	Valor total (USD)	Valor total (AR\$)
Horas de ingeniería	600 horas	\$10 USD	\$6000 USD	\$2.340.000 ARS
Procesamiento GCP	200 horas	\$1 USD	\$200 USD	\$78000 ARS
Consultas BigQuery	1 unidad	\$10 USD	\$10 USD	\$3900 ARS
SUBTOTAL			\$6210 USD	\$2.421.900 ARS
COSTOS INDIRECTOS				
Descripción	Cantidad	Valor unitario	Valor total (USD)	Valor total (AR\$)
Notebook	1 unidad	\$1200 USD	\$1200 USD	\$468000 ARS
Cloud Storage 20GB	1 unidad	\$5 USD	\$5 USD	\$1950 ARS
SUBTOTAL			\$1205 USD	\$469950 ARS
TOTAL			\$7415 USD	\$2.891.850 ARS

Nota: los valores fueron convertidos a pesos argentinos tomando la cotización del dólar MEP al día 30 de marzo de 2023, cuya tasa de conversión es 390 AR\$/USD.

13. Gestión de riesgos

Los riesgos asociados al proyecto y sus planes respectivos planes de mitigación son los siguientes:

Riesgo 1: no contar con datos distribuidos uniformemente entre las categorías.

- Severidad (S): este riesgo tiene una severidad media-alta porque puede influir en la performance final de los modelos de IA. (6)
- Probabilidad de ocurrencia (O): la probabilidad de este riesgo es alta porque hay categorías de reclamos que suceden mucho más a menudo que otras. (7)

Riesgo 2: no contar con la infraestructura cloud de la empresa para entrenar los modelos de IA.

- Severidad (S): este riesgo tiene una severidad media-alta porque puede alargar las tareas de modelado y entrenamiento. (8)
- Ocurrencia (O): la probabilidad de este riesgo es baja porque la utilización de los recursos fue pactada previamente. (4)

Riesgo 3: decisión del área de Data de la empresa de cambiar de *Cloud Provider*.

- Severidad (S): este riesgo tiene una severidad baja porque no requeriría realizar grandes cambios al utilizar herramientas *open source* para el despliegue. (4)

- Ocurrencia (O): la probabilidad de este riesgo es baja porque la adopción de la plataforma es estratégica. (2)

Riesgo 4: no poder realizar la clasificación automática con un nivel de precisión aceptable.

- Severidad (S): este riesgo tiene una severidad alta porque tener una mala precisión podría hacer que los modelos de IA no se usen. (8)
- Ocurrencia (O): la probabilidad de este riesgo es baja porque antes de este proyecto se realizó una prueba piloto para algunas categorías que fueron combinadas y tuvo buenos resultados. (2)

Riesgo 5: no contar con herramientas de preprocesamiento de texto en español.

- Severidad (S): este riesgo tiene una severidad media-alta porque puede influir en la performance final de los modelos de IA y también porque puede alargar las tareas de preprocesamiento al no poder utilizar herramientas ya desarrolladas. (8)
- Ocurrencia (O): la probabilidad de este riesgo es baja ya que se realizó una investigación previa y hay algunas bibliotecas de Python que pueden ser utilizadas para texto en español. (3)

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

Criterio adoptado: se tomarán medidas de mitigación en los riesgos cuyos números de RPN sean mayores a 30.

Riesgo	S	O	RPN	S*	O*	RPN*
1	6	7	42	3	7	21
2	8	4	32	6	4	24
3	4	2	8	-	-	-
4	8	2	16	-	-	-
5	8	3	24	-	-	-

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: se utilizarán técnicas de balanceo de clases.

- Severidad (S*): la severidad baja considerablemente, ya que al aplicar estas técnicas se intenta disminuir los efectos del desbalanceo en el entrenamiento del modelo. (3)
- Probabilidad de ocurrencia (O): la probabilidad de ocurrencia se mantiene igual. (7)

Riesgo 2: se utilizarán alternativas a Google Cloud, como Paperspace, Google Colab o incluso usando hardware físico.

- Severidad (S*): la severidad baja a un nivel medio, ya que igualmente se podrá entrenar usando placa de video. (6)
- Probabilidad de ocurrencia (O): la probabilidad de ocurrencia se mantiene igual. (4)

14. Gestión de la calidad

- Req. #1.1, Req. #1.2, Req. #1.4 y Req. #1.5: el sistema debe poder detectar la categoría de un reclamo o consulta escrito en lenguaje natural, con o sin errores de ortografía y cantidad de palabras variable.
 - Verificación: se probarán mensajes de distintas longitudes, con y sin errores y se revisará que se asigne una categoría.
 - Validación: se revisarán con el cliente distintos mensajes de entrada y la categoría asignada.
- Req. #1.6: la solución debe ejecutarse en forma *batch*, corriendo diariamente y tomando los casos del día anterior.
 - Verificación: se implementará un *cronjob* que refleje una periodicidad diaria, corriendo por la mañana y tomando los casos del día cerrado inmediatamente anterior.
 - Validación: se dejará el *schedule* activado y se controlará al día siguiente la ejecución automática con el cliente.
- Req. #2.3: la solución debe estar desplegada sobre infraestructura de Google Cloud Platform.
 - Verificación: la solución se implementará usando componentes de infraestructura que hayan sido previamente desplegadas en el área de *Machine Learning*.
 - Validación: se controlará con el cliente que los componentes elegidos sean los adecuados.
- Req. #1.3 y #2.4: el usuario debe poder utilizar los resultados de la clasificación desde una base de datos y la salida de los modelos debe ser almacenada en BigQuery.
 - Verificación: se almacenarán los resultados en una tabla de BigQuery controlando que el esquema donde está la tabla sea para salidas de modelos de IA.
 - Validación: se controlará con el cliente que el dataset sea el correcto.
- Req. #2.5: el proceso debe ser ejecutado a través del orquestador Apache Airflow.
 - Verificación: se generará un DAG para Airflow cuya ejecución será probada en el ambiente de testing.
 - Validación: se ejecutará un GitHub Action que controlará las buenas prácticas de código referidas a los DAG y las validaciones se enviarán al cliente.
- Req. #3.1: se deben generar métricas de performance de los modelos con el dataset de entrenamiento y prueba.
 - Verificación: se calcularán distintas métricas de performance de cada modelo para los dos lotes.
 - Validación: se entregará un reporte con las distintas métricas al cliente para su revisión.

- Req. #4.1, Req. #4.2 y Req. #4.3: requerimientos de documentación.
 - Verificación: se procederá a realizar la documentación necesaria del proyecto.
 - Validación: se entregará la documentación al área de *Machine Learning* para su aprobación.

15. Procesos de cierre

Las pautas que darán cierre al proyecto son las siguientes:

- Pautas de trabajo que se seguirán para analizar si se respetó el Plan de Proyecto original:
Persona a cargo: Lucas Rivela

Procedimiento:

- Se analizará el cumplimiento de los requerimientos.
 - Se analizarán riesgos planeados y no planeados, y se evaluarán las técnicas de gestión de riesgos.
 - Se analizará en qué medida se cumplieron los plazos estipulados para cada etapa del proyecto.
- Identificación de las técnicas y procedimientos que se emplearon, los problemas que surgieron y cómo se solucionaron:
Persona a cargo: Lucas Rivela
Procedimiento: se incluirá en la memoria final del proyecto un apartado con las lecciones aprendidas, y un análisis de qué técnicas y procedimientos resultaron útiles y cuáles no.

- Acto de agradecimiento a todos los interesados:

Persona a cargo: Lucas Rivela

Procedimiento:

- Se incluirá un agradecimiento en la memoria del proyecto.
- Se invitará a la exposición del trabajo final a todos los interesados en el proyecto.
- Al final de la exposición se agradecerá a todas las personas involucradas, incluyendo *stakeholders* de la empresa, familiares, miembros del jurado, docentes y autoridades de la CEIA.