

# Guía 1. Definición Proyecto APT

## Asignatura Capstone

### A. PARTE I

#### 1. Antecedentes Personales

Nombre estudiante	Daniel Eduardo Santibáñez Mondaca
Rut	15.949.615-5
Carrera	Ingeniería en Informática
Sede	Viña del Mar

Nombre estudiante	Bastian Ignacio Fierro Solas
Rut	20.172.780-4
Carrera	Ingeniería en Informática
Sede	Viña del Mar

Nombre estudiante	Matías Andrés Zamorano Carvajal
Rut	20.839.591-2
Carrera	Ingeniería en Informática
Sede	Viña del Mar

## 2. Descripción Proyecto APT

<b>Nombre del proyecto</b>	<b>Opinión Salud Valparaíso: Monitoreo y Análisis en Tiempo Real</b>
<b>Área (s) de desempeño(s)</b>	<p><b>Ciencia de Datos</b> (Machine Learning, Deep Learning)</p> <p><b>Inteligencia de Negocios</b> (Big Data, Minería de Datos, Inteligencia de Negocios)</p> <p><b>Gestión de Proyectos Informáticos</b> (Gestión de Proyectos Informáticos, Gestión de Riesgos)</p> <p><b>Análisis y Planificación de Requerimientos Informáticos</b> (Ingeniería de Software, Diseño y Gestión de Requisitos)</p>
<b>Competencias</b>	<p><b>Gestionar proyectos informáticos, ofreciendo alternativas para la toma de decisiones de acuerdo a los requerimientos de la organización:</b> Este proyecto requerirá una gestión efectiva, desde la planificación hasta la ejecución y evaluación.</p> <p><b>Construir modelos de datos para soportar los requerimientos de la organización de acuerdo a un diseño definido y escalable en el tiempo:</b> El proyecto implica la creación de modelos de NLP y la manipulación de bases de datos grandes y complejas.</p> <p><b>Desarrollar una solución de software utilizando técnicas que permitan sistematizar el proceso de desarrollo y mantenimiento, asegurando el logro de los objetivos:</b> Este proyecto implica el desarrollo de un sistema completo, que integra NLP, Web Scrapping, y un portal de usuario.</p> <p><b>Analizar y Procesar Grandes Volúmenes de Datos (Big Data):</b> Este proyecto implica el manejo y análisis de grandes volúmenes de datos recopilados de diferentes fuentes, incluidos formularios y redes sociales.</p> <p><b>Aplicación de Técnicas de Machine Learning y Deep Learning:</b> Dado que el proyecto involucra el uso de un modelo pre entrenado de NLP (BERT) para analizar sentimientos, se aborda directamente la aplicación de técnicas avanzadas de machine learning y deep learning.</p> <p><b>Desarrollo de Soluciones de Inteligencia de Negocios:</b> La herramienta que se desarrollará permitirá a la Corporación Municipal de Valparaíso tomar decisiones informadas basadas en datos, que es un aspecto clave de la inteligencia de negocios.</p> <p><b>Gestión de Riesgos en Proyectos Informáticos:</b> El proyecto, siendo de naturaleza compleja y tecnológica, requerirá una evaluación y gestión proactiva de los riesgos que podrían afectar su ejecución exitosa.</p> <p><b>Comunicación Efectiva y Presentación de Resultados:</b> La naturaleza del proyecto requiere que los resultados del análisis se comuniquen de manera clara y efectiva a los stakeholders, lo cual implica habilidades de comunicación y presentación de datos.</p>

### 3. Fundamentación Proyecto APT

<p><b>Relevancia del proyecto APT</b></p>	<p><b>¿Por qué escogiste este tema?</b> El equipo escogió este tema porque identificó una oportunidad para usar la tecnología de manera significativa para mejorar la atención en los Centros de Salud Familiar en Valparaíso.</p> <p><b>¿Por qué es relevante para el campo laboral de tu carrera?</b> El proyecto es relevante porque fusiona áreas críticas de la informática aplicada, tales como la ciencia de datos, la gestión de proyectos y el desarrollo de software, en una solución con un genuino impacto social. Además, el equipo cree que puede generar un impacto positivo en los pacientes, ya que sentirán que sus comentarios son tomados en cuenta y reflejados en las mejoras de los centros de salud.</p> <p><b>¿Dónde se ubica la situación que vas a abordar?</b> La situación se aborda en los Centros de Salud Familiar de Valparaíso, Chile.</p> <p><b>¿Cuáles son las características principales de ese lugar?</b> Valparaíso es una comuna portuaria, con una población diversa y con centros de salud que atienden a una amplia variedad de pacientes.</p> <p><b>¿A quiénes afecta o impacta la situación que vas a abordar?</b> Afecta a los pacientes (o familiares) que usan los Centros de Salud Familiar, el personal médico y administrativo de los centros, y la Corporación Municipal de Valparaíso.</p> <p><b>¿Cuál sería el aporte de valor (real o simulado) de tu Proyecto APT para el contexto laboral y/o social en que se sitúa?</b> El proyecto podría mejorar significativamente la calidad de la atención en los Centros de Salud Familiar, identificando áreas donde se requiere atención y mejoras, lo que podría llevar a una mayor satisfacción del paciente (o sus familiares) y una operación más eficiente de los centros.</p>
<p><b>Descripción del Proyecto APT</b></p>	<p><b>Objetivo:</b> Desarrollar una herramienta de análisis de sentimientos basada en técnicas de NLP que permita monitorear y evaluar las percepciones de los pacientes respecto a los Centros de Salud Familiar en Valparaíso.</p> <p><b>Descripción Breve:</b> El proyecto consistirá en diseñar y desarrollar un sistema que recopile datos de diversas fuentes, incluyendo un formulario de opinión y redes sociales. Estos datos serán procesados y analizados usando un modelo de NLP pre entrenado, como BERT, para evaluar los sentimientos y las opiniones de los pacientes. Los resultados se presentarán en un formato accesible y útil para que los responsables de los Centros de Salud puedan tomar decisiones informadas y proactivas para mejorar la atención. El proyecto será gestionado siguiendo buenas prácticas de gestión de proyectos informáticos, con un enfoque en la entrega de valor real y tangible para la comunidad de Valparaíso.</p>
<p><b>Pertinencia del proyecto con el perfil de egreso</b></p>	<p>El Proyecto APT se alinea directamente con los perfiles de egreso de la carrera del equipo, ya que involucra competencias críticas en las áreas de Ciencia de Datos, Inteligencia de Negocios, Gestión de Proyectos Informáticos y Análisis y Planificación de Requerimientos Informáticos. El diseño, desarrollo y gestión de una herramienta de monitoreo de sentimientos para los Centros de Salud Familiar de Valparaíso requiere habilidades analíticas avanzadas (manejo de Big Data, Machine Learning y Deep Learning), gestión de proyectos efectiva y estratégica (evaluación y mitigación de riesgos, planificación de proyectos), y una fuerte capacidad para analizar y planificar requerimientos de software. Las competencias seleccionadas son esenciales para</p>

	resolver la problemática de analizar y responder de manera más efectiva a las percepciones y sentimientos del público respecto a los servicios de salud, que es una cuestión crítica para mejorar la calidad y la eficiencia de los servicios de salud.
<b>Relación con los intereses profesionales</b>	Los intereses profesionales del equipo se centran en la aplicación de técnicas avanzadas de análisis de datos para resolver problemas reales y tangibles en diversos sectores, incluido el de la salud. El equipo está particularmente interesado en cómo la inteligencia de negocios y la ciencia de datos pueden informar y mejorar las decisiones organizacionales y políticas. El Proyecto APT refleja estos intereses, permitiéndonos abordar un problema real que tiene un impacto directo en la calidad de los servicios de salud y en la vida de las personas. Además, este proyecto brindará al equipo una experiencia invaluable en la gestión de proyectos, fundamental para el desarrollo de los miembros del equipo como líderes en futuros roles profesionales.
<b>Factibilidad de desarrollo del Proyecto APT</b>	<p><b>Duración del semestre (18 semanas):</b> Este tiempo debería ser suficiente para desarrollar, entrenar y validar un modelo de NLP, además de diseñar e implementar el sistema de recopilación de datos y el portal del usuario, asumiendo que el proyecto se gestione de manera efectiva.</p> <p><b>Horas asignadas a la asignatura (4 por día, excepto domingos):</b> Esto suma un total de 24 horas por semana, que proporciona un tiempo de trabajo razonable para el proyecto.</p> <p><b>Materiales requeridos:</b> Todos los materiales necesarios (datos, software, hardware, acceso a Internet) están disponibles y son adecuados para el proyecto.</p> <p><b>Factores externos que facilitan su desarrollo:</b> La disponibilidad de datos y la cooperación del cliente son factores clave que facilitarán el desarrollo del proyecto.</p> <p><b>Factores externos que dificultan su desarrollo y soluciones propuestas:</b></p> <p>Cantidad de datos: La solución propuesta es el relleno artificial de los datos basado en los datos originales para el entrenamiento del modelo, lo cual es una estrategia común en proyectos de machine learning.</p> <p>Cliente abandonando: En caso de que el cliente abandone el proyecto, la solución sería terminar el proyecto de manera independiente y dejarlo en un estado funcional, sin atarlo a un cliente específico. Esto permitiría que el sistema desarrollado se ofrezca a múltiples clientes potenciales en el futuro.</p>

## B. PARTE II

### 4. Objetivos

<b>Objetivo general</b>	<p>Desarrollar e implementar un sistema de monitoreo de sentimientos basado en técnicas de procesamiento del lenguaje natural (NLP) y deep learning que permita analizar y comprender las percepciones y opiniones de la comunidad respecto a los Centros de Salud Familiar en Valparaíso, con el propósito de proporcionar a la Corporación Municipal de Valparaíso información valiosa y accionable para mejorar la calidad y eficiencia de los servicios de salud.</p>
<b>Objetivos específicos</b>	<p><b>Recopilar y Preparar Datos:</b> Recolectar datos de opiniones y percepciones de los ciudadanos a través de un formulario en línea y técnicas de Web Scrapping de redes sociales.</p> <p>Preprocesar y limpiar los datos recolectados, incluyendo la generación de datos artificiales para enriquecer el conjunto de entrenamiento.</p> <p><b>Desarrollar y Entrenar un Modelo de NLP:</b> Adaptar y entrenar un modelo pre entrenado (BERT) con PyTorch, usando una GPU local, para analizar y clasificar los sentimientos y temas predominantes en las opiniones recopiladas.</p> <p><b>Implementar una Interfaz de Usuario:</b> Diseñar y desarrollar una interfaz de usuario accesible y fácil de usar que permita a los ciudadanos enviar sus opiniones y percepciones a través de un formulario en línea. Esto, además, complementará la recolección de datos para el conjunto de entrenamiento.</p> <p><b>Analizar y Visualizar Resultados:</b> Implementar herramientas de visualización y análisis de datos que permitan a la Corporación Municipal de Valparaíso interpretar los resultados y detectar áreas de mejora en los Centros de Salud Familiar.</p> <p><b>Gestionar el Proyecto de Forma Eficiente:</b> Planificar, monitorear y controlar las actividades del proyecto, gestionando riesgos y asegurando que se cumplan los plazos y objetivos establecidos.</p> <p><b>Evaluar y Refinar el Modelo:</b> Evaluar constantemente la calidad y eficiencia del modelo de NLP desarrollado, realizando ajustes y mejoras según los feedback y los resultados obtenidos en las pruebas.</p> <p><b>Entregar un Informe Final y Documentación:</b> Compilar un informe detallado que presente los resultados del proyecto, incluyendo análisis de datos, métricas del modelo y recomendaciones para la Corporación Municipal de Valparaíso. Además, producir documentación completa del código y del sistema desarrollado.</p>

## 5. Metodología

### Descripción

#### Metodología y Abordaje del Problema

Para abordar el problema identificado, se emplea una metodología ágil, específicamente Scrum, la cual es ideal para proyectos de desarrollo donde las condiciones pueden cambiar rápidamente y requiere adaptación. Scrum permitirá iteraciones rápidas, entregas incrementales de valor y la adaptabilidad a los cambios.

#### Fases del Proyecto:

##### 1. Definición y Planificación:

- Objetivo: Establecer el alcance del proyecto, identificar las principales tareas y planificar las iteraciones o Sprint.
- Tareas: Crear el product backlog, definir criterios de aceptación para cada tarea, estimar tiempos y esfuerzos.

##### 2. Diseño y Desarrollo:

- Objetivo: Diseñar y construir el sistema propuesto.
- Tareas: Diseño de la interfaz de usuario, recopilación y generación de datos, desarrollo del modelo de NLP, implementación del formulario y herramientas de visualización.

##### 3. Pruebas y Ajustes:

- Objetivo: Evaluar el funcionamiento del sistema y realizar los ajustes necesarios.
- Tareas: Realizar pruebas unitarias, de integración y de usuario. Refinar el modelo de NLP y solucionar problemas identificados.

##### 4. Entrega y Evaluación:

- Objetivo: Presentar el sistema a la Corporación Municipal de Valparaíso, recoger feedback y realizar las mejoras finales.
- Tareas: Entrega de una versión preliminar, recopilación de feedback, ajustes finales y entrega del sistema definitivo.

#### Roles y Responsabilidades:

##### - Daniel Santibañez (Scrum Master / Data Engineer):

- Garantizar que el equipo siga los principios y prácticas de Scrum.
- Eliminar obstáculos que impidan al equipo alcanzar sus objetivos.
- Facilitar las reuniones de Scrum (daily standups, sprint planning, sprint review y sprint retrospective).
- Proteger al equipo de interferencias externas.
- Además, colaborará en todas las tareas del proyecto al igual que los otros integrantes.

**- Bastian Fierro (Data Scientist):**

- Participar activamente en la recopilación y preprocesamiento de datos.
- Colaborar en el diseño y desarrollo de la interfaz de usuario y en la implementación del modelo de NLP.
- Participar en las pruebas, identificar errores y colaborar en las soluciones.
- Aportar en la entrega y presentación del sistema.

**- Matías Zamorano (Ingeniero de Software/Analista de Datos):**

- Contribuir en la recopilación de datos y en el diseño del modelo de NLP.
- Participar en el diseño de la interfaz de usuario y en las pruebas.
- Colaborar en la definición de criterios de aceptación y en la evaluación del proyecto.
- Aportar en la entrega y presentación del sistema.

**- Renzo Vergara (Product Owner / Administrador)**

- Definir y priorizar las características del producto basado en las necesidades de la Corporación Municipal de Valparaíso y de otros stakeholders.
- Mantener y priorizar el Product Backlog, garantizando que el equipo esté trabajando en las historias y tareas que ofrecen el máximo valor al proyecto.
- Ser el punto principal de comunicación entre los stakeholders y el equipo de desarrollo, garantizando que todas las partes estén alineadas con la visión del producto.
- Aceptar o rechazar los resultados al final de cada sprint, asegurando que cumplan con los criterios de aceptación definidos.
- Aportar información y clarificaciones al equipo durante el desarrollo, garantizando que las dudas y obstáculos relacionados con los requisitos sean abordados rápidamente.
- Participar activamente en las reuniones de Scrum, incluyendo sprint planning, sprint review y sprint retrospective.
- Continuamente recibir feedback de los usuarios y stakeholders para ajustar y mejorar el producto.
- Colaborar estrechamente con el Scrum Master y el equipo para garantizar la entrega de un producto de alta calidad en el tiempo establecido.

**- Mónica Riveros Sobenes (Stakeholders):**

- Brinda detalles esenciales y retroalimentación sobre el proyecto.
- Participa activamente en decisiones clave y se mantiene en comunicación constante con el equipo.
- Verifica que los entregables estén alineados con las expectativas y ayuda al equipo a superar obstáculos relacionados con sus intereses.

**- Usuario (Pacientes, familiares y visitantes del centro de salud):**

- Interactuar con la plataforma final y utilizar sus funciones para enviar sugerencias, reclamos y calificar la recomendación del centro.

- Participar, si se les solicita, en pruebas de usuario o en sesiones de feedback para ayudar a mejorar el sistema antes de su lanzamiento oficial.

Dado que todos los integrantes realizan labores similares (excepto el product owner), es esencial que el equipo mantenga una comunicación constante para sincronizar esfuerzos, compartir conocimientos y garantizar que el proyecto avance de manera uniforme. Las reuniones diarias (daily stand ups) facilitadas por Daniel Eduardo Santibáñez Mondaca como Scrum Master serán cruciales para esto.

## 6. Evidencias

Tipo de evidencia (avance o final)	Nombre de la evidencia	Descripción	Justificación
Avance	Mapa de Actores	Gráfico que identifica y categoriza a todos los individuos, grupos o entidades involucrados o afectados por el proyecto.	Fundamental para entender quiénes son las partes interesadas, sus interacciones y cómo podrían impactar o ser impactados por el proyecto.
Avance	Mapa Mental	Representación gráfica que organiza información, ideas y conceptos relacionados con el proyecto en forma de diagrama.	Ayuda a visualizar la estructura del proyecto, las interrelaciones de sus componentes y facilita la comprensión global del mismo.
Avance	Roadmap	Documento visual que muestra los objetivos principales y las etapas del proyecto en una línea de tiempo.	Ofrece una visión clara de la dirección del proyecto, los hitos clave y las fechas previstas, permitiendo una planificación y seguimiento eficaz.
Avance	Visión y Roles de Scrum	Descripción del propósito global del proyecto y la definición de los roles dentro del marco de trabajo de Scrum.	Es esencial para garantizar que todos los miembros del equipo comprendan su papel y la dirección general que debe tomar el proyecto.
Avance	Épicas e Historias de Usuarios	Documento que detalla las características grandes (épicas) y las tareas más pequeñas y específicas (historias de usuario) que se deben realizar.	Asegura que se cubran todas las necesidades y requisitos del proyecto, proporcionando una base para la planificación y desarrollo.
Avance	Gestión de Riesgos	Análisis de los posibles riesgos asociados con el proyecto y planes para mitigar o responder a esos riesgos.	Reduce la probabilidad de sorpresas negativas, permitiendo que el proyecto avance de manera más fluida y previsible.
Avance	Definición del Proyecto APT	Documento que establece el alcance, objetivos, entregables y limitaciones del Proyecto APT.	Proporciona un punto de referencia claro para todas las partes interesadas, asegurando que todos estén alineados en cuanto a lo que se espera del proyecto.



## Guía Estudiante - Definición Proyecto APT

### Fase 1



Avance	Formulario prototipo	Primer diseño y funcionalidad del formulario de opinión que se usará para recolectar feedback.	Mostrar el progreso en la interfaz de usuario y las funcionalidades previstas para la recolección de datos.
Avance	Datos recopilados (preliminar)	Conjunto inicial de datos recolectados a través del formulario y Web Scrapping.	Dar cuenta de la capacidad de recolección de datos y la eficiencia de las herramientas utilizadas.
Avance	Modelo de NLP (primer entrenamiento)	Resultados iniciales de entrenamiento y ajuste del modelo BERT con los datos recolectados.	Verificar el progreso y la precisión inicial del modelo en identificar y categorizar sentimientos.
Final	Formulario final	Versión final del formulario de opinión con todas sus funcionalidades implementadas.	Confirmar que se ha desarrollado una herramienta útil y funcional para recolectar opiniones del público.
Final	Datos recopilados (final)	Conjunto total de datos recolectados a lo largo del proyecto, incluyendo datos artificiales y reales.	Mostrar la riqueza y variedad de datos con los que se ha trabajado y que han alimentado el modelo de NLP.
Final	Modelo de NLP (entrenamiento final)	Resultados finales del modelo después de todos los ciclos de entrenamiento, ajuste y optimización.	Presentar la precisión, eficacia y resultados del modelo en la identificación y categorización de sentimientos.
Final	Reporte de análisis de sentimientos	Documento que resume las principales tendencias, palabras clave, problemas recurrentes y opiniones recolectadas.	Entregar una visión comprensiva y resumida de la percepción pública sobre los Centros de Salud Familiar de Valparaíso.
Final	Feedback del cliente	Respuestas y opiniones de la Corporación Municipal de Valparaíso sobre la utilidad y eficacia del proyecto entregado.	Validar la pertinencia, utilidad y calidad del proyecto a través de las opiniones del cliente final.

## 7. Plan de Trabajo

Plan de Trabajo Proyecto APT						
Competencia o unidades de competencias	Nombre de Actividades /Tareas	Descripción Actividades/Tareas	Recursos	Duración de la actividad	Responsable	Observaciones
Desarrollo Web	Diseño del Formulario	Crear el diseño y la estructura del formulario de opinión.	Python, Flask, HTML, CSS	2 Sprint	Matías Zamorano	Se debe validar el diseño con los usuarios
NLP y Machine Learning	Configuración de BERT	Ajustar el modelo BERT para su entrenamiento con datos específicos.	PyTorch, CUDA, GPU RTX 4080	1 Sprint	Bastian Fierro	Requiere revisión constante
Web Scrapping	Recolección de datos	Extraer datos de opiniones de redes sociales y otros sitios web.	Python, BeautifulSoup	2 Sprint	Daniel Santibáñez	Respetar términos de uso de sitios web
Gestión de Datos	Generación de datos	Crear datos artificiales basados en información real para enriquecer el entrenamiento.	Python	1 Sprint	Daniel Santibáñez	Datos deben ser revisados y validados
NLP y Machine Learning	Entrenamiento de modelo	Entrenar y ajustar el modelo BERT con los datos recopilados y generados.	PyTorch, CUDA, GPU RTX 4080	2 Sprint	Bastian Fierro	Monitorizar progreso del modelo
Comunicación	Reportes de avance	Elaborar informes periódicos sobre el progreso del proyecto para la corporación.	Microsoft Word/Google Docs	Continuo	Daniel Santibáñez	Mantener a stakeholders informados
Evaluación y QA	Testeo del modelo	Realizar pruebas para evaluar la precisión y eficiencia del modelo en datos no vistos.	Set de datos de prueba	1 Sprint	Matías Zamorano	Iterar en base a resultados
Desarrollo Web	Implementación final	Puesta en marcha del formulario final y herramientas asociadas para su uso en producción.	Servidor, Flask	1 Sprint	Daniel Santibáñez	Validar con usuarios finales

## 8. Roadmap

### OPINIÓN SALUD VALPARAÍSO: MONITOREO Y ANÁLISIS EN TIEMPO REAL

#### Épicas

- 1. Diseño e Implementación del Formulario de Opinión
- 2. Configuración y Entrenamiento del Modelo BERT
- 3. Recopilación de Datos a través de Web Scrapping
- 4. Generación de Datos Artificiales
- 5. Evaluación y Testeo del Modelo NLP
- 6. Actividades y/o documentación del proyecto

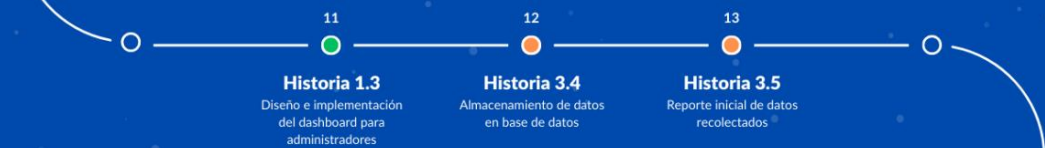
Sprint 0: 07 de Agosto -> 02 de Septiembre



Sprint 1: 04 de Septiembre -> 16 de Septiembre



Sprint 1: 18 de Septiembre -> 30 de Septiembre



Sprint 2: 02 de Octubre -> 14 de Octubre



Sprint 2: 16 de Octubre -> 28 de Octubre



Sprint 3: 30 de Octubre -> 18 de Noviembre



Sprint 3: 20 de Noviembre -> 02 de Diciembre

