

# Guía1. Definición Proyecto APT

## Asignatura Capstone

### A. PARTE I

#### 1. Antecedentes Personales

A continuación, se presenta una tabla en la que debes completar la información solicitada.

Nombre estudiante	<b>Patricia Aldunce Silva</b> <b>Francisco González Santana</b> <b>Cindy Cañete Aguirre</b>
Rut	<b>Patricia Aldunce Silva 18.047.660-1</b> <b>Francisco González Santana 17.623.180-7</b> <b>Cindy Cañete Aguirre 16.716.058-1</b>
Carrera	<b>Ingeniería en Informática</b>
Sede	<b>Padre Alonso Ovalle</b>

## 2. Descripción Proyecto APT

Optimizar la experiencia del cliente mediante un nuevo "Dashboard" con una interfaz más intuitiva y accesible. Este Dashboard integrará prácticas de machine learning y análisis predictivo de datos, proporcionando información más precisa y accionable para mejorar la toma de decisiones estratégicas

Nombre del proyecto	Proyecto DataLab Duoc
Área (s) de desempeño(s)	<ul style="list-style-type: none"><li>• Área de software: Desarrollo de software</li><li>• Área de datos: Ciencia de datos – Ingeniería de datos</li></ul>
Competencias	<ul style="list-style-type: none"><li>• Ofrecer propuestas de solución informática analizando de forma integral los procesos de acuerdo con los requerimientos de la organización.</li><li>• Desarrollar la transformación de grandes volúmenes de datos para la obtención de información y conocimiento de la organización a fin de apoyar la toma de decisiones y la mejora de los procesos de negocio, de acuerdo con las necesidades de la organización.</li><li>• Gestionar proyectos informáticos, ofreciendo alternativas para la toma de decisiones de acuerdo con los requerimientos de la organización.</li></ul>

### 3. Fundamentación Proyecto APT

A continuación, se presentan distintos campos que debes completar con la información solicitada. Esta sección busca que describas en detalle tu proyecto y justifiques su relevancia y pertinencia.

Relevancia del proyecto APT	<p>El tema fue escogido porque DuocLaboral.cl es una plataforma (bolsa de empleo de Duoc UC) la cual contiene mucha data, asociada a temáticas de empleabilidad, lo cual es parte de uno de los objetivos del modelo Educativo Institucional y parte del plan de desarrollo.</p> <p>Como público, va en beneficio directo a las 9 Escuelas que conforman/son parte de Duoc UC (Administración y Negocios, Comunicación, Construcción, Diseño, Gastronomía, Informática y Telecomunicaciones, Ingeniería y Recursos Naturales, Salud y Turismo y Hospitalidad) dado que la reportería les aporta en la toma de decisiones informadas con relación a temáticas de empleabilidad.</p> <p>Este proyecto permitirá potenciar la actual reportería de la plataforma DuocLaboral.cl, integrando técnicas de minerías de datos y machine learning</p>
Descripción del Proyecto APT	<p>Mejorar la precisión y calidad del análisis de datos y desarrollar modelos predictivos que faciliten la toma de decisiones estratégicas para las escuelas. Este proyecto busca ofrecer una visión integral del mercado laboral, alineando sus capacidades con las oportunidades disponibles.</p> <p>Este objetivo enfatiza tanto la calidad del análisis como el impacto estratégico en la motivación y el registro de los alumnos.</p>
Pertinencia del proyecto con el perfil de egreso	<p><b>Área de Software: Desarrollo de Software:</b> El proyecto APT está enfocado en el desarrollo de un <b>**Dashboard**</b> que mejorará la experiencia del cliente a través de una interfaz más amigable y con la integración de prácticas como el <b>*machine learning*</b>. Esto se alinea directamente con el perfil de egreso que busca formar profesionales capaces de desarrollar soluciones de software que se adapten a las necesidades organizacionales, promoviendo la innovación y optimización de procesos.</p> <p><b>Área de Datos: Ciencia de Datos e Ingeniería de Datos:</b> El proyecto implica la implementación de análisis de datos predictivos, lo cual requiere de competencias en ciencia de datos. Este enfoque es esencial para la toma de decisiones basadas en</p>

datos, que es un componente clave del perfil de egreso de la carrera.

**Análisis de Procesos y Requerimientos:** El proyecto demanda la capacidad de analizar los procesos actuales de la organización para proponer una solución informática adecuada. Esta competencia es crítica, ya que para implementar un \*Dashboard\* efectivo, es necesario entender profundamente cómo se usan los datos en la toma de decisiones actualmente y qué mejoras pueden introducirse.

**Transformación de Datos en Conocimiento:** Desarrollar un Dashboard con análisis predictivo implica trabajar con grandes volúmenes de datos y transformarlos en información útil. Esto es clave para apoyar la toma de decisiones en la organización, alineándose con la competencia de generar conocimiento a partir de datos.

**Gestión de Proyectos Informáticos:** Para llevar a cabo este proyecto, es necesario gestionar eficientemente los recursos y tiempos, asegurando que la solución propuesta cumpla con los requerimientos de la organización. Esta competencia es fundamental para garantizar que el proyecto se complete exitosamente y dentro de los parámetros establecidos.

El proyecto APT requiere de las competencias seleccionadas para abordar la problemática de mejorar la toma de decisiones mediante un Dashboard avanzado. Sin una adecuada gestión de datos, análisis de requerimientos, y transformación de datos en conocimiento, el proyecto no podría cumplir sus objetivos de manera efectiva. Las competencias seleccionadas no solo son necesarias, sino que son fundamentales para el éxito del proyecto y su alineación con el perfil de egreso de la carrera.

Relación con los intereses profesionales	<p>Patricia Aldunce: El proyecto se alinea perfectamente con mis intereses profesionales, ya que se enfoca en el análisis de grandes volúmenes de datos para descubrir insights valiosos y generar un impacto significativo.</p> <p>Cindy Cañete: El proyecto se alinea con mis intereses profesionales, dado que me permitirá desplegar los conocimientos asociados a la transformación de grandes volúmenes de datos, para la obtención y análisis de información y apoyar en la toma de decisiones estratégicas de las organizaciones.</p> <p>Francisco González: El proyecto se relaciona directamente con mi interés de generar soluciones a problemáticas o necesidades, es importante poder llegar a una solución innovadora que pueda aportar al negocio.</p>
Factibilidad de desarrollo del Proyecto APT	<p>El proyecto es viable dentro del marco del curso debido a la disponibilidad de datos adecuados para realizar técnicas avanzadas de minería de datos y machine learning. Además, su alcance está bien definido para completarse en un período de 10 semanas.</p> <p>La disponibilidad de datos históricos y variables externas relevantes, así como herramientas de software adecuadas (como Python y Power BI) facilita la implementación del proyecto.</p>

#### 4. Alcance General

Segmentación, Modelado Predictivo y Visualización de Prácticas Profesionales en Duoc Laboral

Que se espera	
<b>Segmentación y análisis de patrones</b>	Se aplicarán técnicas de clustering para segmentar las prácticas profesionales disponibles en la plataforma Duoc Laboral. La segmentación se llevará a cabo en base a variables clave, como tipo de práctica, salario ofrecido, duración de la práctica, ubicación geográfica, área de especialización, género de los postulantes y su condición de discapacidad. Esta segmentación permitirá identificar tendencias y patrones en las ofertas de prácticas, proporcionando información valiosa para adaptar la oferta educativa de Duoc UC y desarrollar programas que respondan a las necesidades actuales del mercado laboral, fomentando la inclusión y equidad.
<b>Creación de modelos predictivos</b>	Se implementarán modelos de machine learning para predecir el comportamiento del mercado laboral en relación con las vacantes y ofertas de prácticas profesionales. Los modelos se basarán en datos históricos de la bolsa de empleo Duoc Laboral, así como en variables externas relevantes, como la situación económica del país y las tendencias actuales. Este análisis predictivo permitirá anticipar la demanda futura de prácticas en diferentes sectores, fortaleciendo la vinculación con el sector empresarial y asegurando que la formación ofrecida esté alineada con las necesidades emergentes del mercado, con un enfoque en la equidad e inclusión.
<b>Clasificación de oferta y perfiles de usuarios</b>	Se emplearán técnicas de clasificación para construir perfiles detallados de los usuarios tanto estudiantes como titulados que utilizan Duoc Laboral. Estos perfiles incluirán datos demográficos, intereses profesionales y comportamiento en la plataforma, permitiendo realizar un análisis de tendencias para identificar patrones de comportamiento. Este conocimiento ayudará a los usuarios principales (Escuela Duoc UC, Direcciones Centrales y Sedes) a ajustar sus estrategias y servicios según las necesidades observadas, así como a sugerir ofertas que se ajusten a las habilidades y aspiraciones de cada usuario.
<b>Visualización y reportes dinámicos</b>	Se desarrollarán Dashboard interactivos en Power BI (u otra plataforma) para presentar datos históricos y simulaciones basadas en los modelos predictivos. Estos Dashboard serán actualizados automáticamente mediante scripts programados que integrarán nuevos datos conforme estén disponibles, permitiendo una visualización dinámica y en tiempo real de las métricas más relevantes para la toma de decisión.

## B. PARTE II

### 5. Objetivos

En este apartado debes definir objetivos generales y específicos del Proyecto APT. Es importante aclarar que los objetivos se deben plantear en forma clara, concisa y sin dar mayores explicaciones, es decir, deben entenderse por sí solos. Se sugiere redactarlos utilizando un verbo en infinitivo, pues ello obliga a precisar acciones concretas.

Objetivo general	Mejorar la experiencia del cliente mediante el desarrollo de un Dashboard interactivo que permita la toma de decisiones informadas a través del análisis de datos predictivos y prácticas de machine learning
Objetivos específicos	<p><b>Diseñar y desarrollar una interfaz de usuario amigable para el Dashboard, que facilite el acceso y comprensión de los datos relevantes para la toma de decisiones.</b></p> <p>Implementar algoritmos de machine learning para el análisis predictivo de datos, proporcionando información útil y accionable a los usuarios.</p> <p><b>Integrar y transformar grandes volúmenes de datos de la organización para generar informes y visualizaciones que apoyen la toma de decisiones estratégicas.</b></p> <p>Validar y optimizar el rendimiento del Dashboard a través de pruebas con usuarios, asegurando que cumple con los requerimientos de la organización y mejora los procesos de negocio.</p>

## 6. Metodología

En el siguiente apartado deberás describir la metodología, propia de tu disciplina, que utilizarás para resolver el proyecto APT antes descrito, incluyendo las etapas y métodos de trabajo.

### Descripción de la Metodología

Determinamos como equipo que la mejor metodología es ágil, debido al tiempo que hay para desarrollar, se necesita enfocar historias de usuarios cortas, pero concisas para que entreguen una funcionalidad funcional. El equipo se compone por tres integrantes:

1. Francisco González: Data Scientist - Software Developer
  - Descripción de su Rol Data Scientist:
    - ✓ Recopila, analiza e interpreta grandes volúmenes de datos
    - ✓ Debe de tener una colaboración continua con el PO y equipo TI
    - ✓ Desarrollo iterativo de modelos
    - ✓ Modelado Predictivo
  - Descripción de su Rol Software Developer:
    - ✓ Revisión de código y mejora continua
    - ✓ Pruebas y validación
2. Cindy Cañete: Analista de datos - BI Developer – Product Owner
  - Descripción de su Rol Analista de datos – BI Developer:
    - ✓ Recopilación y limpieza de datos
    - ✓ Análisis exploratorio
    - ✓ Modelado de datos
    - ✓ Desarrollo de Dashboards e informes
    - ✓ Administración de plataformas BI
  - Descripción de su Rol Product Owner:
    - ✓ Visión del producto
    - ✓ Crear y priorizar el backlog
    - ✓ Especificación de Historias de Usuarios
    - ✓ Toma de decisiones
3. Patricia Aldunce: Scrum Master - Analista de datos
  - Descripción de su Rol Scrum Master:
    - ✓ Facilitación del proceso Scrum
    - ✓ Eliminación de impedimentos
    - ✓ Apoyo al Product Owner
    - ✓ Protección al equipo
    - ✓ Recopilación y limpieza de datos
    - ✓ Análisis exploratorio
    - ✓ Modelado de datos



## 7. Evidencias

A continuación, describe qué evidencias serán evaluadas en el informe de avance y en el informe final de tu proyecto APT. Estas evidencias deben ser acordadas con tu docente. Se entenderá por evidencia los productos que se desarrollen durante el proyecto y cuyo propósito sea visibilizar o documentar cómo se ha implementado el trabajo.

Tipo de evidencia (avance o final)	Nombre de la evidencia	Descripción	Justificación
Avance	Comprensión del negocio	Definir los objetivos del proyecto en términos de negocio, recopilando las ideas, sus experiencias con lo que tienen en la actualidad y lo que esperan con la innovación propuesta por parte del equipo TI	Esta recopilación permite identificar las carencias y fortalezas actuales de la organización, ayudando a priorizar los esfuerzos en aquellas áreas que generen un mayor impacto positivo
Avance	Comprensión de los datos	Recolectar y explorar los datos disponibles, evaluando la calidad de los datos con los siguientes ítems: <ol style="list-style-type: none"> <li>1. Identificación de datos faltantes</li> <li>2. Distribución de variables</li> <li>3. Relación entre variables</li> <li>4. Transformación</li> <li>5. Visualizaciones iniciales básicas de histogramas y demás gráficos necesarios</li> </ol>	La recolección y exploración de datos son etapas críticas para garantizar que el análisis posterior se base en datos precisos, relevantes y de alta calidad
Avance	Preparación de los datos	Los datos estarán listos para su modelado con las siguientes preparaciones: <ol style="list-style-type: none"> <li>1. Limpieza de datos</li> <li>2. Transformación de variables</li> <li>3. Manejo de Outliers</li> <li>4. Integración de datos</li> <li>5. Exportación de datos preparados</li> </ol>	La preparación de los datos es una etapa clave en el desarrollo de modelos de machine learning y análisis predictivo

<b>Avance</b>	<b>Modelado</b>	<p>Para el modelado se entregará lo siguiente:</p> <ol style="list-style-type: none"> <li>1. Justificación de selección del modelo</li> <li>2. Configuración de parámetros</li> <li>3. Proceso de entrenamiento</li> </ol>	<p>La entrega de los elementos relacionados con el modelado del proyecto es fundamental para documentar y explicar claramente las decisiones tomadas durante la construcción de los modelos predictivos. La entrega de los elementos relacionados con el modelado del proyecto es fundamental para documentar y explicar claramente las decisiones tomadas durante la construcción de los modelos predictivos.</p>
<b>Avance</b>	<b>Evaluación del modelo</b>	<p>Para la evaluación de entregará lo siguiente:</p> <ol style="list-style-type: none"> <li>1. Rendimiento del modelo en el conjunto de datos</li> <li>2. Identificación de errores comunes</li> <li>3. Impacto de errores</li> <li>4. Interpretabilidad del modelo</li> </ol>	<p>La evaluación del modelo es una fase crucial para garantizar que la solución desarrollada cumple con los objetivos del proyecto y proporciona resultados precisos y útiles para la toma de decisiones de negocio.</p>
<b>Final</b>	<b>Despliegue</b>	<p>Para el despliegue se entregará todo lo anterior documentado y adicional:</p> <ol style="list-style-type: none"> <li>1. Estrategia de implementación</li> <li>2. Guía de instalación</li> <li>3. Pruebas de aceptación</li> <li>4. Manual de usuario</li> </ol>	<p>El despliegue de un modelo de una solución analítica es una fase crítica para asegurar que los resultados obtenidos en el desarrollo sean aplicables y útiles en un entorno real de producción.</p>

## 8. Plan de Trabajo

En la siguiente tabla define la planificación de tu Proyecto APT de acuerdo con lo requerido.

Plan de Trabajo Proyecto APT						
Competencia o unidades de competencias	Nombre de Actividades/Tareas	Descripción Actividades/Tareas	Recursos	Duración de la actividad	Responsable	Observaciones
Ofrecer propuestas de solución informática, analizando de forma integral los procesos de acuerdo con los requerimientos de la organización.	Definir caso de uso	Describir el caso de uso asociado a la reportería de DuocLaboral, especificando cómo interactúan los usuarios con el dashboard	Documentación de requisitos (reportería y data duoclaboral)	2 semanas	Cindy Cañete Patricia Aldunce	Puede requerir revisión por parte de los stakeholders.
Ofrecer propuestas de solución informática, analizando de forma integral los procesos de acuerdo a los requerimientos de la organización.	Identificar de stakeholders y sus necesidades	Identificar a las partes interesadas de Duoc UC y la VRA; y documentar sus expectativas y necesidades.	Reuniones, entrevistas.	2 semanas	Cindy Cañete	Involucrar a todas las áreas relevantes.
Gestionar proyectos informáticos, ofreciendo alternativas para la toma de decisiones de acuerdo a los requerimientos de la organización.	Establecer objetivos	Definir objetivos claros y medibles para el proyecto, en base a las necesidades identificadas.	Documentos de planificación, herramientas de colaboración (Jira)	1 semana	Patricia Aldunce	Alinear con stakeholders.

Desarrollar la transformación de grandes volúmenes de datos para la obtención de información y conocimiento de la organización a fin de apoyar.	Seleccionar variables objetivo	Seleccionar las variables clave que se usarán en el análisis predictivo.	Reportería duoclaboral (excel y power bi)	3 semanas	Francisco González Patricia Aldunce	Validar la relevancia de las variables.
Desarrollar la transformación de grandes volúmenes de datos para la obtención de información y conocimiento de la organización a fin de apoyar.	Explorar de manera inicial los datos	Analizar los datos para identificar patrones y posibles errores o valores atípicos.	Python	2 semanas	Francisco González Patricia Aldunce Cindy Cañete	
Desarrollar la transformación de grandes volúmenes de datos para la obtención de información y conocimiento de la organización a fin de apoyar.	Limpiar y manejar datos	Limpiar y transformar los datos para asegurar su calidad y formato adecuado para el modelado.	Python	3 semanas	Francisco González Patricia Aldunce Cindy Cañete	
Desarrollar la transformación de grandes volúmenes de datos para la obtención de	Transformar variables	Crear nuevas variables o modificar los datos para que sean más adecuados	Python	3 semanas		Validar impacto del modelo

información y conocimiento de la organización a fin de apoyar.		para el análisis y modelado				
Implementar soluciones sistémicas integrales para automatizar y optimizar procesos de negocio de acuerdo a las necesidades de la organización.	Desarrollar modelo predictivo	Construir uno o más modelos predictivos, utilizando técnicas de machine learning apropiadas para el proyecto.	Python	3 semanas	Cindy Cañete Patricia Aldunce Francisco González	Requerir de pruebas y ajustes
Realizar pruebas de certificación tanto de los productos como de los procesos utilizando buenas prácticas definidas por la industria.	Evaluar modelo	Probar el modelo con datos de validación para medir su precisión y efectividad.	Python	2 semanas	Francisco González	Ajustar en caso de que no cumpla con los objetivos
Implementar soluciones sistémicas integrales para automatizar y optimizar procesos de negocio de acuerdo a las necesidades de la organización.	Desplegar reportería	Crear y publicar reportes interactivos en Power BI, basados en los modelos desarrollados .	Power BI	3 semanas	Patricia Aldunce Cindy Cañete	Asegurar acceso y comprensión por parte del usuario final

Ofrecer propuestas de solución informática, analizando de forma integral los procesos de acuerdo a los requerimientos de la organización.	Definir caso de uso	Describir el caso de uso asociado a la reportería de DuocLaboral, especificando cómo interactúan los usuarios con el dashboard	Documentación de requisitos (reportería y data duoclaboral)	2 semanas	Cindy Cañete Patricia Aldunce	Puede requerir revisión por parte de los stakeholders.
Ofrecer propuestas de solución informática, analizando de forma integral los procesos de acuerdo a los requerimientos de la organización.	Identificar de stakeholders y sus necesidades	Identificar a las partes interesadas de Duoc UC y la VRA; y documentar sus expectativas y necesidades.	Reuniones, entrevistas.	2 semanas	Cindy Cañete	Involucrar a todas las áreas relevantes.

## 9. Carta Gantt

Busca un formato de Carta Gantt que te acomode y organiza en este las actividades planificadas en el punto anterior considerando el periodo asignado para el desarrollo de tu Proyecto APT. Debes mantener la temporalidad del periodo académico en el desarrollo de las tres fases que contempla la Asignatura de Portafolio de Título.

[illegible]