



Pronóstico de Ventas para Toma de Decisiones en Comercio Electrónico usando Machine Learning

Autor:

Ing. Borda Jonathan Matias

Director:

Título y Nombre del director (a confirmar) (pertenencia)

Esta planificación fue realizada en el curso de Gestión de proyectos entre el 30 de abril de 2023 y el 18 de junio 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>).	9
8. Entregables principales del proyecto	9
9. Desglose del trabajo en tareas	10
10. Diagrama de Activity On Node.	10
11. Diagrama de Gantt	12
12. Presupuesto detallado del proyecto	15
13. Gestión de riesgos	15
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	30 de abril de 2023

Acta de constitución del proyecto

Buenos Aires, 30 de abril de 2023

Por medio de la presente se acuerda con el Ing. Borda Jonathan Matias que su Trabajo Final de la Carrera de Especialización en Inteligencia Artificial se titulará “Pronóstico de Ventas para Toma de Decisiones en Comercio Electrónico usando Machine Learning” y consistirá en desarrollar un modelo de inteligencia artificial capaz de pronosticar las ventas de una tienda en línea de la empresa Latech. El trabajo tendrá un presupuesto preliminar estimado de **600** horas y un costo estimado de **\$ XXX**, con fecha de inicio el 30 de abril de 2023 y fecha de presentación pública el 15 de mayo de 2024.

Se adjunta a esta acta la planificación inicial.

Dr. Ing. Ariel Lutenberg
Director posgrado FIUBA

Juan Cruz Bonina
Latech

Título y Nombre del director (a confirmar)
Director del Trabajo Final

1. Descripción técnica-conceptual del proyecto a realizar

La empresa Latech fabrica y comercializa barritas alimenticias en distintos sabores. Como se trata de alimentos, la empresa necesita contar con pronósticos precisos de ventas, lo que le permite planificar la producción y evitar tanto faltantes como excedentes de stock.

A partir de esta problemática, se propone desarrollar un modelo de inteligencia artificial capaz de pronosticar las ventas de la tienda en línea de la empresa Latech.

El desafío es que la precisión de la predicción de las ventas sea alta para que el cliente pueda tomar decisiones estratégicas con un alto grado de confianza. A su vez, el cliente desea conocer la evolución de las ventas en el tiempo para poder ajustar la estrategia de marketing y de ventas en función de los resultados.

Se va a utilizar como insumo un conjunto de datos históricos, que abarca aproximadamente dos años, sobre ventas e inversión en publicidad. Estos datos se van a obtener desde una plataforma llamada Shopify para generar el modelo capaz de predecir con precisión. El propósito final es contar con una herramienta que facilite la toma de decisión estratégica basada en la predicción.

El cliente Latech cuenta con una tienda en línea alojada en Shopify. Shopify es plataforma de comercio electrónico basada en la nube que permite a empresas y particulares crear y gestionar tiendas en línea. Además, esta ofrece una *API* que facilita el acceso a datos como: los productos vendidos por día, los usuarios que realizan la compra, el canal de venta, etc. A través de esta *API* también se obtiene métricas de comportamiento de los usuarios tales como: la cantidad de páginas vistas, cuando agrega un producto al carrito de compra, cuando termina la compra y la tasa de conversión. Adicionalmente, utilizamos una plataforma llamada Triple Whale, especializada en la gestión y análisis de inversión publicitaria que proporciona una *API* donde se obtiene información sobre la inversión realizada por la tienda en sitios como: Google, Facebook, Instagram, Klaviyo, etc.

Para llevar a cabo el proyecto se cuenta con el financiamiento de la empresa Latech y acceso a la *API* de Shopify y a la *API* de Triple Whale. Asimismo, existe un acuerdo de confidencialidad con la empresa, que establece que los datos utilizados y los entregables generados durante el desarrollo del proyecto no podrán ser difundidos públicamente ni compartidos con terceros sin autorización expresa.

En la figura 1 se presenta el diagrama en bloques del sistema. Se observa que las *APIs* externas proveen los datos que alimentan el conjunto de datos del proyecto. Tanto este conjunto de datos como el modelo de predicción se integrarán dentro del sistema del cliente, denominado Inventory Tracker. Como resultado, el sistema generará el pronóstico de ventas.

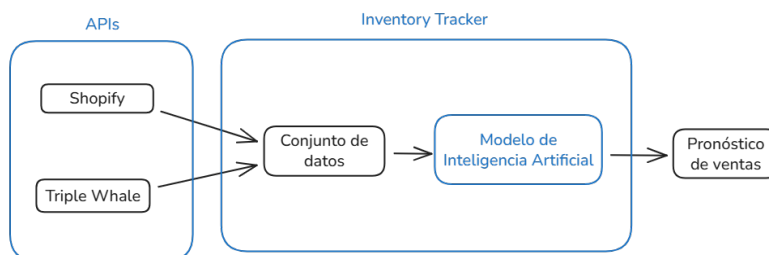


Figura 1. Diagrama en bloques del sistema.

2. Identificación y análisis de los interesados

Cuadro 1. Identificación de los interesados

Rol	Nombre y Apellido	Organización	Puesto
Cliente	Juan Cruz Bonina	Latech	CEO
Responsable del Proyecto	Ing. Borda Jonathan Matias	FIUBA	Alumno
Orientador	Título y Nombre del director (a confirmar)	pertenencia	Director del Trabajo Final
Usuario Final	Juan Willink	Latech	Gerente de Ventas

Cliente: Juan Cruz Bonina es el CEO de la empresa Latech quien va a definir los requerimientos. Suele estar disponible en cualquier momento para responder preguntas y brindar comentarios o aclaraciones. Usuario final: Juan Willink es el Gerente de Ventas de la empresa Latech quien va a utilizar el modelo de inteligencia artificial para tomar decisiones estratégicas. Vive en Estados Unidos y trabaja en el horario de la tarde.

3. Propósito del proyecto

El propósito de este proyecto es desarrollar un modelo de inteligencia artificial que permita predecir con precisión las ventas del producto alimenticio de la empresa Latech, a partir de datos de ventas y de inversión publicitaria. Esto permitirá a la empresa: anticipar sus necesidades de producción, optimizar la planificación del inventario y tomar decisiones estratégicas basadas en datos. Asimismo, reducir el riesgo de faltantes o excesos de inventario y mejorar la eficiencia operativa y comercial.

4. Alcance del proyecto

El proyecto incluye:

- Relevamiento y análisis de los datos históricos de ventas de la tienda en línea obtenidos mediante la *API* de Shopify.
- Relevamiento y análisis de los datos históricos de inversión publicitaria provenientes de la *API* de Triple Whale.
- Limpieza, transformación y consolidación de datos provenientes de ambas plataformas.
- Análisis exploratorio de datos (EDA) para identificar patrones, tendencias y relaciones entre las variables.
- Desarrollo y entrenamiento de modelos de inteligencia artificial orientados a la predicción de ventas.
- Evaluación comparativa de distintos modelos para seleccionar el que ofrezca la mejor precisión y capacidad predictiva.

- Implementación de un módulo funcional integrado al sistema Inventory Tracker de la empresa, que permita realizar predicciones de ventas para períodos futuros.
- Documentación técnica del proceso, del modelo elegido y de su uso.
- Entrega de reportes que expliquen los hallazgos y recomendaciones derivadas del análisis.

El presente proyecto no incluye:

- Implementaciones en tiempo real del modelo, es decir, sistemas que actualicen las predicciones de forma instantánea ante cada nuevo dato recibido. Sí se contempla la posibilidad de programar ejecuciones periódicas del modelo (por ejemplo, una vez al día) para actualizar las predicciones de manera regular.
- Garantía de precisión absoluta en las predicciones, ya que la misma depende de la calidad y estabilidad de los datos futuros y de factores externos no controlables.
- Acciones o recomendaciones específicas sobre estrategias de marketing más allá de lo inferido de los análisis de datos.
- Optimización de procesos internos de producción o logística, salvo en lo que respecta a la estimación de demanda.

5. Supuestos del proyecto

Para el desarrollo del presente proyecto se supone que:

- La empresa Latech proveerá acceso completo y continuo a las *APIs* externas de Shopify y Triple Whale para la obtención de datos históricos y actualizados.
- Se contará con acceso al repositorio de código y a la infraestructura necesaria para integrar el módulo de predicción en el sistema existente Inventory Tracker.
- La calidad, integridad y consistencia de los datos obtenidos desde las plataformas externas será suficiente para entrenar y validar los modelos de predicción.
- No se producirán cambios significativos en las políticas de acceso o en la estructura de datos de las *APIs* de Shopify y Triple Whale durante el período de desarrollo.
- No existirán restricciones legales o contractuales que impidan el uso de los datos necesarios para el proyecto.
- Se contará con la disponibilidad de recursos computacionales necesarios para el procesamiento, entrenamiento y validación de los modelos de inteligencia artificial.

6. Requerimientos

1. Requerimientos funcionales:

- 1.1. El sistema debe permitir importar datos históricos de ventas desde la *API* de Shopify.(Prioridad alta).

- 1.2. El sistema debe permitir importar datos históricos de inversión publicitaria desde la *API* de Triple Whale. (Prioridad alta).
- 1.3. El sistema debe consolidar y procesar los datos provenientes de ambas plataformas para crear un conjunto de datos único. (Prioridad alta).
- 1.4. El sistema debe entrenar modelos de inteligencia artificial capaces de predecir las ventas del producto por período de tiempo (día, semana, mes). (Prioridad alta).
- 1.5. El sistema debe permitir actualizar periódicamente las predicciones, mediante una ejecución programada. (Prioridad alta).
- 1.6. El usuario debe poder consultar las predicciones de ventas a través del sistema Inventory Tracker, integrado de forma nativa. (Prioridad alta).
- 1.7. El sistema debe permitir seleccionar el período de tiempo sobre el cual realizar las predicciones. (Prioridad media).
- 1.8. El sistema debe mantener la confidencialidad y seguridad de la información procesada. (Prioridad alta).
2. Requerimientos de documentación:
 - 2.1. Documentación técnica que describa la arquitectura del sistema, el proceso de integración y el funcionamiento del modelo predictivo. (Prioridad alta).
 - 2.2. Manual de usuario para la operación del módulo dentro de Inventory Tracker. (Prioridad media).
 - 2.3. Documentación de instalación y configuración del módulo desarrollado. (Prioridad media).
3. Requerimiento de testing:
 - 3.1. El sistema debe contar con pruebas unitarias sobre los componentes críticos. (Prioridad alta).
 - 3.2. El sistema debe ser validado mediante pruebas funcionales integradas en Inventory Tracker. (Prioridad alta).
 - 3.3. El modelo predictivo debe ser evaluado usando métricas estadísticas para asegurar precisión aceptable. (Prioridad alta).
4. Requerimientos de la interfaz:
 - 4.1. La interfaz de usuario dentro de Inventory Tracker debe permitir visualizar las predicciones de ventas de manera clara y comprensible. (Prioridad alta).
 - 4.2. Las predicciones deben presentarse con opciones gráficas (por ejemplo, gráficos de series temporales) para facilitar su interpretación. (Prioridad media).
 - 4.3. El sistema debe permitir la descarga de las predicciones y datos consolidados en formato *CSV* para su análisis externo. (Prioridad media).
5. Requerimientos de interoperabilidad:
 - 5.1. El sistema debe integrarse sin problemas con la infraestructura existente de Inventory Tracker. (Prioridad alta).
 - 5.2. El sistema debe adaptarse a posibles cambios en las *APIs* de Shopify y Triple Whale, siempre que estos cambios no impliquen modificaciones sustanciales en los datos entregados. (Prioridad media).
6. Requerimientos legales y regulatorios:

- 6.1. El sistema debe cumplir con las leyes y regulaciones vigentes en materia de protección de datos personales. (Prioridad alta).
- 6.2. La información procesada y almacenada no debe ser divulgada a terceros, en cumplimiento con el acuerdo de confidencialidad firmado con Latech. (Prioridad alta).

7. Historias de usuarios (*Product backlog*)

1. Épica 1: Integración de fuentes de datos

- 1.1. “Como analista de datos quiero importar datos de ventas desde Shopify para poder entrenar el modelo de predicción.”
Story points: 5 (complejidad: 2, dificultad: 2, incertidumbre: 1)
- 1.2. “Como analista de datos quiero importar datos de inversión publicitaria desde Triple Whale para incluir la variable de inversión en el análisis de ventas.”
Story points: 5 (complejidad: 2, dificultad: 2, incertidumbre: 1)
- 1.3. “Como analista de datos quiero consolidar los datos de Shopify y Triple Whale para tener un único dataset listo para el entrenamiento del modelo.”
Story points: 8 (complejidad: 3, dificultad: 2, incertidumbre: 3)

2. Épica 2: Modelado predictivo

- 2.1. “Como analista de datos quiero entrenar un modelo que prediga las ventas para poder anticipar la demanda de producción.”
Story points: 13 (complejidad: 5, dificultad: 4, incertidumbre: 4)

3. Épica 3: Visualización y exportación de resultados

- 3.1. “Como usuario de Inventory Tracker quiero visualizar las predicciones de ventas dentro del sistema para poder planificar la producción.”
Story points: 8 (complejidad: 3, dificultad: 2, incertidumbre: 3)
- 3.2. “Como usuario de Inventory Tracker quiero descargar las predicciones en formato CSV para analizarlas externamente.”
Story points: 3 (complejidad: 1, dificultad: 1, incertidumbre: 1)

4. Épica 4: Monitoreo y alertas del sistema

- 4.1. “Como responsable del sistema quiero recibir alertas si hay datos faltantes que puedan afectar significativamente la precisión de las predicciones.”
Story points: 5 (complejidad: 2, dificultad: 2, incertidumbre: 1)

8. Entregables principales del proyecto

- Manual de usuario.
- Diagramas de arquitectura del sistema y de integración con Inventory Tracker.
- Código fuente del software para procesamiento de datos, entrenamiento de modelos y generación de predicciones.
- Diagrama de flujo del proceso de carga, procesamiento y predicción de datos.

- Memoria del trabajo final.
- Documentación técnica detallada sobre la instalación, configuración y mantenimiento del módulo desarrollado.

9. Desglose del trabajo en tareas

El WBS debe tener relación directa o indirecta con los requerimientos. Son todas las actividades que se harán en el proyecto para dar cumplimiento a los requerimientos. Se recomienda mostrar el WBS mediante una lista indexada:

1. Grupo de tareas 1 (suma h)

- 1.1. Tarea 1 (tantas h)
- 1.2. Tarea 2 (tantas h)
- 1.3. Tarea 3 (tantas h)

2. Grupo de tareas 2 (suma h)

- 2.1. Tarea 1 (tantas h)
- 2.2. Tarea 2 (tantas h)
- 2.3. Tarea 3 (tantas h)

3. Grupo de tareas 3 (suma h)

- 3.1. Tarea 1 (tantas h)
- 3.2. Tarea 2 (tantas h)
- 3.3. Tarea 3 (tantas h)
- 3.4. Tarea 4 (tantas h)
- 3.5. Tarea 5 (tantas h)

Cantidad total de horas: tantas.

¡Importante!: la unidad de horas es h y va separada por espacio del número. Es incorrecto escribir “23hs”.

Se recomienda que no haya ninguna tarea que lleve más de 40 h. De ser así se recomienda dividirla en tareas de menor duración.

10. Diagrama de Activity On Node

Armar el AoN a partir del WBS definido en la etapa anterior.

Una herramienta simple para desarrollar los diagramas es el Draw.io (<https://app.diagrams.net/>). Draw.io

Indicar claramente en qué unidades están expresados los tiempos. De ser necesario indicar los caminos semi críticos y analizar sus tiempos mediante un cuadro. Es recomendable usar colores y un cuadro indicativo describiendo qué representa cada color.



Figura 2. Diagrama de *Activity on Node*.

11. Diagrama de Gantt

Existen muchos programas y recursos *online* para hacer diagramas de Gantt, entre los cuales destacamos:

- Planner
- GanttProject
- Trello + *plugins*. En el siguiente link hay un tutorial oficial:
<https://blog.trello.com/es/diagrama-de-gantt-de-un-proyecto>
- Creately, herramienta online colaborativa.
<https://creately.com/diagram/example/ieb3p3ml/LaTeX>
- Se puede hacer en latex con el paquete *pgfgantt*
<http://ctan.dcc.uchile.cl/graphics/pgf/contrib/pgfgantt/pgfgantt.pdf>

Pegar acá una captura de pantalla del diagrama de Gantt, cuidando que la letra sea suficientemente grande como para ser legible. Si el diagrama queda demasiado ancho, se puede pegar primero la “tabla” del Gantt y luego pegar la parte del diagrama de barras del diagrama de Gantt.

Configurar el software para que en la parte de la tabla muestre los códigos del EDT (WBS).
Configurar el software para que al lado de cada barra muestre el nombre de cada tarea.
Revisar que la fecha de finalización coincida con lo indicado en el Acta Constitutiva.

En la figura 3, se muestra un ejemplo de diagrama de gantt realizado con el paquete de *pgfgantt*. En la plantilla pueden ver el código que lo genera y usarlo de base para construir el propio.

Las fechas pueden ser calculadas utilizando alguna de las herramientas antes citadas. Sin embargo, el siguiente ejemplo fue elaborado utilizando [esta hoja de cálculo](#).

Es importante destacar que el ancho del diagrama estará dado por la longitud del texto utilizado para las tareas (Ejemplo: tarea 1, tarea 2, etcétera) y el valor *x unit*. Para mejorar la apariencia del diagrama, es necesario ajustar este valor y, quizás, acortar los nombres de las tareas.

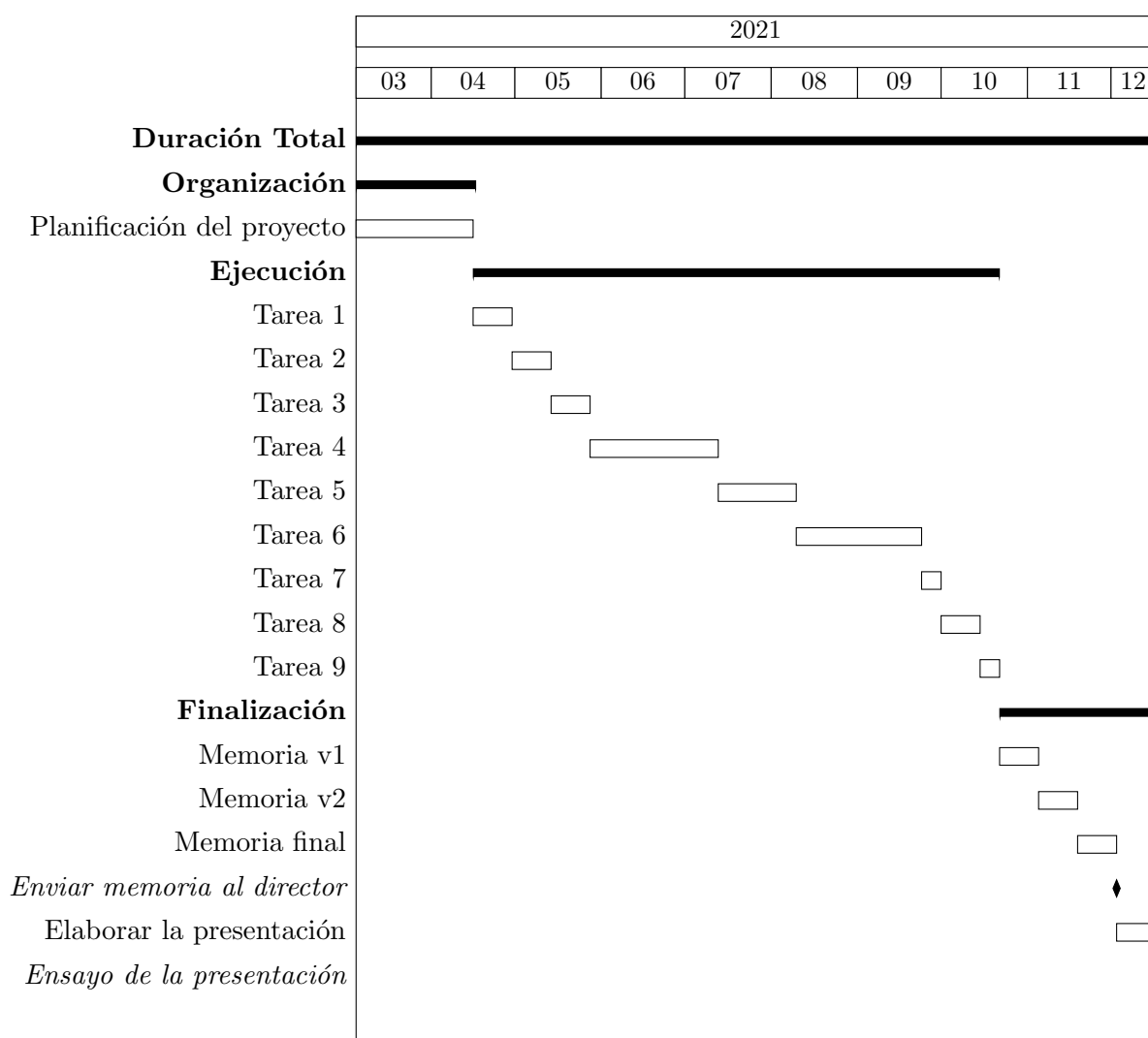


Figura 3. Diagrama de gantt de ejemplo



Figura 4. Ejemplo de diagrama de Gantt (apaisado).

12. Presupuesto detallado del proyecto

Si el proyecto es complejo entonces separarlo en partes:

- Un total global, indicando el subtotal acumulado por cada una de las áreas.
- El desglose detallado del subtotal de cada una de las áreas.

IMPORTANTE: No olvidarse de considerar los **COSTOS INDIRECTOS**.

Incluir la aclaración de si se emplea como moneda el peso argentino (ARS) o si se usa moneda extranjera (USD, EUR, etc). Si es en moneda extranjera se debe indicar la tasa de conversión respecto a la moneda local en una fecha dada.

COSTOS DIRECTOS			
Descripción	Cantidad	Valor unitario	Valor total
SUBTOTAL			
COSTOS INDIRECTOS			
Descripción	Cantidad	Valor unitario	Valor total
SUBTOTAL			
TOTAL			

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): X.
Justificación...

- Ocurriencia (O): Y.
Justificación...

Riesgo 3:

- Severidad (S): X.
Justificación...
- Ocurriencia (O): Y.
Justificación...

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 con su correspondiente número.

- 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, 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.