



Pesaje al paso

Autor:

Ing. Esp. Jorge Cuenca

Director:

Nombre del Director (FIUBA)

Codirector:

Codirector (FIUBA)

*Esta planificación fue realizada en el curso de Gestión de proyectos
entre el 14 de marzo de 2023 y el 18 de julio 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	10
10. Diagrama de Activity On Node.	11
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	16

Registros de cambios

Revisión	Detalles de los cambios realizados	Fecha
0	Creación del documento	14 de marzo de 2023
1	Desarrollo hasta la sección 9	19 de marzo de 2023

Acta de constitución del proyecto

Buenos Aires, 14 de marzo de 2023

Por medio de la presente se acuerda con el Ing. Ing. Esp. Jorge Cuenca que su Trabajo Final de la Carrera de Especialización en Inteligencia artificial se titulará “Pesaje al paso”, consistirá esencialmente en la implementación de un algoritmo para obtener un resultado válido de peso de un animal que está circulando sobre la plataforma de pesaje, y tendrá un presupuesto preliminar estimado de 600 h de trabajo y \$500.000 colombianos, con fecha de inicio 14 de marzo de 2023 y fecha de presentación pública 15 de agosto de 2023.

Se adjunta a esta acta la planificación inicial.

Dr. Ing. Ariel Lutenberg
Director posgrado FIUBA

Ing. Guillermo
Emtech

Nombre del Director
Director del Trabajo Final

1. Descripción técnica-conceptual del proyecto a realizar

Análisis exploratorio de datos: Convertir los datos recolectados a un formato adecuado para su análisis, utilizando los datos adquiridos hasta el momento. Hacer una inspección de los datos rotulados para determinar la cantidad de muestras útiles para entrenamiento y evaluación. Sobre los datos útiles, se analizará la variedad de las etiquetas (ground-truth de los pesos) y la representatividad del espacio muestral que habría una vez desplegado en campo. Se sacarán algunas métricas estadísticas buscando sesgos y varianzas.

Esto servirá para estimar el tipo de modelo para realizar la regresión y su capacidad.

Especificación de nuevos datos a recolectar para ampliar el dataset: En función del estudio de la aplicación y de lo que se observe en el paso anterior y del o los modelos que se propongan, se debe realizar una especificación para la ampliación de las muestras. La especificación incluirá casos particulares de interés.

Conformación del dataset de entrenamiento y evaluación: Definir un entorno de desarrollo para el modelo a entrenar. Construir el pipeline para la transformación de los datos crudos al dataset depurado, para entrenar y evaluar el modelo. Esto incluirá la etapa de pre-procesamiento destinada sólo a la limpieza de datos.

Definición del pre-procesamiento a realizar a las muestras: Este pre-procesamiento es el que habría que implementar en el pipeline de inferencia (estimación del peso) por cada muestra que se quiera procesar. Este pre-procesamiento, a priori, se implementará como lo mínimo necesario que ayude a incrementar el desempeño del modelo. Tener en cuenta que es distinto al pre-procesamiento de limpieza mencionado en el punto anterior. Contempla el modelado del pre-procesamiento, no su implementación en la plataforma.

Modelos: Se implementarán y evaluarán al menos dos modelos, uno de mayor capacidad (para estimación off-line) que demuestre la viabilidad de la estimación, y otro de menor capacidad en función de las capacidades de cómputo disponibles para la estimación en tiempo real (que degrade el desempeño de manera tolerable).

Cuantización de un modelo para despliegue sobre la plataforma: Este punto aplicaría si se implementará el modelo en una plataforma que no utilice punto flotante (procesadores de baja capacidad de cómputo o FPGAs), o bien que no soporte las librerías utilizadas en los modelos. Podría no ser necesario en una primera instancia de desarrollo.

Despliegue: Si la plataforma no soporta las librerías usadas para ML, o bien no tiene una capacidad de cómputo medianamente aceptable (punto flotante, matricial), este paso puede demandar mucho trabajo.

Análisis de resultados: Este análisis se refiere a la evaluación del desempeño una vez desplegado el modelo. Incluiría sugerencias para mejoras de desempeño, ya sea modificando y re-entrenando los modelos, como adquisiciones de nuevas muestras. Comparación de los datos obtenidos cotejados con pesos estáticos realizados en campo.

- Este proyecto es parte del programa de vinculación con empresas del posgrado, en este caso se tiene la oportunidad de trabajar con la empresa EMTECH.

En la Figura 1 se presenta el diagrama en bloques del sistema. Se observa el proceso general para llevar a cabo el desarrollo del proyecto.

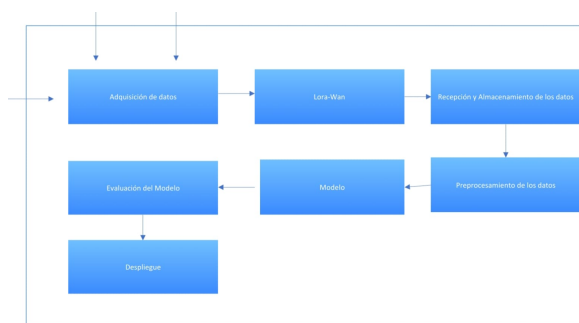


Figura 1. Diagrama en bloques del sistema

2. Identificación y análisis de los interesados

Rol	Nombre y Apellido	Organización	Puesto
Cliente	Ing. Guillermo	Emtech	
Responsable	Ing. Esp. Jorge Cuenca	FIUBA	Alumno
Orientador	-	FIUBA	Director Trabajo final

3. Propósito del proyecto

El propósito de este proyecto es crear un algoritmo que sea capaz de obtener un resultado válido de peso de un animal que está circulando sobre la plataforma de pesaje. Dependiendo del resultado del algoritmo se utilizará este valor de peso para ajustes de raciones, análisis de sanidad, etc.

4. Alcance del proyecto

El alcance del proyecto consiste en llevar a cabo un análisis exploratorio de datos con el objetivo de estimar el tipo de modelo para realizar la regresión y su capacidad. En primer lugar, se llevará a cabo la conversión de los datos recolectados a un formato adecuado para su análisis, seguido de una inspección de los datos rotulados para determinar la cantidad de muestras útiles para entrenamiento y evaluación. A continuación, se analizará la variedad de las etiquetas (ground-truth de los pesos) y la representatividad del espacio muestral que habría una vez desplegado en campo, extrayendo algunas métricas estadísticas para buscar sesgos y varianzas.

Posteriormente, se especificarán nuevos datos a recolectar para ampliar el dataset, en función del estudio de la aplicación y de lo que se observe en el paso anterior y del o los modelos que se propongan, incluyendo casos particulares de interés. Se definirá un entorno de desarrollo para el modelo a entrenar y se construirá el pipeline para la transformación de los datos crudos al dataset depurado para entrenar y evaluar el modelo. Esto incluirá la etapa de pre-procesamiento destinada sólo a la limpieza de datos.

Se definirá el pre-procesamiento a realizar a las muestras, el cual se implementará en el pipeline de inferencia por cada muestra que se quiera procesar. Este pre-procesamiento, a priori, se implementará como lo mínimo necesario que ayude a incrementar el desempeño del modelo, contemplando el modelado del pre-procesamiento, no su implementación en la plataforma.

Se implementarán y evaluarán al menos dos modelos, uno de mayor capacidad (para estimación off-line) que demuestre la viabilidad de la estimación, y otro de menor capacidad en función de las capacidades de cómputo disponibles para la estimación en tiempo real (que degrade el desempeño de manera tolerable). Además, se llevará a cabo la cuantización de un modelo para despliegue sobre la plataforma, si es necesario.

Finalmente, se realizará el despliegue del modelo y se llevará a cabo el análisis de resultados, evaluando el desempeño una vez desplegado el modelo e incluyendo sugerencias para mejoras de desempeño, ya sea modificando y re-entrenando los modelos, como adquisiciones de nuevas muestras. También se realizará una comparación de los datos obtenidos cotejados con pesos estáticos realizados en campo.

5. Supuestos del proyecto

Para el desarrollo del presente proyecto se supone que:

- El algoritmo proporciona mediciones precisas y consistentes del peso de los animales, independientemente del tamaño, raza o especie del animal
- El algoritmo es capaz de detectar y corregir errores y factores de confusión, como la influencia del viento o el movimiento de la plataforma de pesaje, para proporcionar mediciones precisas y confiables del peso de los animales.
- El algoritmo es capaz de proporcionar mediciones de peso en tiempo real y en un formato legible para su uso en ajustes de raciones alimenticias y análisis de sanidad de los animales.

6. Requerimientos

1. Requerimientos funcionales

- 1.1. Convertir los datos recolectados a un formato adecuado para su análisis.
- 1.2. Inspeccionar los datos rotulados para determinar la cantidad de muestras útiles para entrenamiento y evaluación.
- 1.3. Analizar la variedad de las etiquetas (ground-truth de los pesos) y la representatividad del espacio muestral.
- 1.4. Sacar algunas métricas estadísticas buscando sesgos y varianzas.
- 1.5. Especificar nuevos datos a recolectar para ampliar el dataset.
- 1.6. Definir un entorno de desarrollo para el modelo a entrenar.
- 1.7. Construir el pipeline para la transformación de los datos crudos al dataset depurado, para entrenar y evaluar el modelo.
- 1.8. Definir el pre-procesamiento a realizar a las muestras.
- 1.9. Implementar y evaluar al menos dos modelos.
- 1.10. Cuantizar un modelo para despliegue sobre la plataforma.
- 1.11. Desplegar el modelo en la plataforma.
- 1.12. Analizar los resultados obtenidos y sugerir mejoras.

2. Requerimientos de documentación

- 2.1. Documentar el proceso de conversión de los datos.
- 2.2. Documentar el proceso de inspección de los datos.
- 2.3. Documentar el proceso de análisis de las etiquetas y de las métricas estadísticas.
- 2.4. Documentar la especificación de nuevos datos a recolectar.
- 2.5. Documentar el entorno de desarrollo para el modelo a entrenar.
- 2.6. Documentar el pipeline para la transformación de los datos crudos al dataset depurado.
- 2.7. Documentar el pre-procesamiento a realizar a las muestras.
- 2.8. Documentar los modelos implementados y evaluados.
- 2.9. Documentar la cuantización del modelo.
- 2.10. Documentar el despliegue del modelo en la plataforma.
- 2.11. Documentar el análisis de resultados y las sugerencias de mejora.

3. Requerimiento de testing

- 3.1. Realizar pruebas para garantizar la calidad del dataset depurado.
- 3.2. Realizar pruebas para garantizar la calidad del pre-procesamiento.
- 3.3. Realizar pruebas para garantizar la calidad de los modelos implementados.

4. Requerimientos interoperabilidad

- 4.1. Asegurarse de que el modelo sea compatible con la plataforma donde será desplegado.
- 4.2. Asegurarse de que el modelo cuantizado sea compatible con la plataforma donde será desplegado.

7. Historias de usuarios (*Product backlog*)

1. Como analista de datos, quiero poder importar archivos CSV para convertirlos a un formato adecuado para su análisis, para poder utilizar los datos adquiridos hasta el momento. Ponderación: 3 story points.

Criterio de ponderación: La tarea de importar archivos CSV es una tarea relativamente simple, pero requiere cierta cantidad de trabajo y tiempo. Por lo tanto, la historia se califica con un número moderado de story points.

2. Como usuario final, quiero poder ver un resumen estadístico de los datos, incluyendo métricas como la media, la desviación estándar y el rango, para poder tener una idea general de la distribución de los datos. Ponderación: 5 story points.

Criterio de ponderación: La tarea de implementar un resumen estadístico de los datos implica algunas tareas más complejas, como la implementación de algoritmos de cálculo de estadísticas y la presentación de los resultados en un formato fácil de leer. Por lo tanto, la historia se califica con un número moderado-alto de story points.

3. Como administrador del sistema, quiero poder agregar nuevas fuentes de datos al sistema para ampliar el conjunto de datos disponibles para el análisis, para poder mejorar la precisión de los modelos. Ponderación: 8 story points.

Criterio de ponderación: La tarea de agregar nuevas fuentes de datos implica varias tareas complejas, como la integración de nuevos formatos de archivo y la validación de los datos para asegurarse de que sean adecuados para su uso en el sistema existente. Por lo tanto, la historia se califica con un número alto de story points.

4. Como investigador científico, quiero poder realizar análisis de correlación para determinar si hay relaciones significativas entre las variables, para poder identificar patrones y tendencias en los datos. Ponderación: 13 story points.

Criterio de ponderación: La tarea de implementar un análisis de correlación es un proceso más complejo que implica la implementación de un algoritmo de correlación y la presentación de los resultados en un formato útil para el usuario. Por lo tanto, la historia se califica con un número alto de story points.

8. Entregables principales del proyecto

Los entregables del proyecto son:

- Informe de análisis exploratorio de datos: Este informe contendría los resultados de la exploración de los datos recolectados, incluyendo una descripción del formato adecuado para su análisis, el número de muestras útiles para entrenamiento y evaluación, la variedad de las etiquetas y la representatividad del espacio muestral.
- Especificación de nuevos datos a recolectar: Este entregable incluiría una lista de casos particulares de interés para la ampliación de las muestras, basados en el estudio de la aplicación y lo que se observó en el análisis exploratorio de datos.
- Dataset de entrenamiento y evaluación: Este entregable consistiría en el conjunto de datos depurados y listos para ser utilizados en el entrenamiento y evaluación de los modelos. Además, incluiría el pipeline de transformación de los datos crudos al dataset depurado.
- Definición del pre-procesamiento a realizar a las muestras: Este entregable consistiría en la especificación del pre-procesamiento a realizar a cada muestra para incrementar el desempeño del modelo.
- Implementación y evaluación de los modelos: Este entregable consistiría en la implementación y evaluación de al menos dos modelos, uno de mayor capacidad y otro de menor capacidad en función de las capacidades de cómputo disponibles. Se proporcionarían los resultados de la evaluación del desempeño de cada modelo.
- Cuantización de un modelo para despliegue sobre la plataforma: Este entregable incluiría la cuantización de un modelo para ser desplegado en una plataforma que no utilice punto flotante o que no soporte las librerías utilizadas en los modelos.
- Despliegue del modelo: Este entregable incluiría la implementación del modelo en la plataforma destinada para su uso y el informe de los resultados obtenidos en el despliegue.
- Informe final: Este entregable sería un informe completo del proyecto, incluyendo una descripción detallada de cada uno de los entregables mencionados anteriormente, los resultados obtenidos y las recomendaciones para mejoras de desempeño, ya sea modificando y re-entrenando los modelos, como adquisiciones de nuevas muestras. Además, se incluiría una comparación de los datos obtenidos cotejados con pesos estáticos realizados en campo.

9. Desglose del trabajo en tareas

1. Grupo de tareas 1

- 1.1. Convertir los datos recolectados a un formato adecuado para su análisis. (tantas h)
- 1.2. Inspeccionar los datos rotulados para determinar la cantidad de muestras útiles para entrenamiento y evaluación. (tantas hs)
- 1.3. Analizar la variedad de las etiquetas y la representatividad del espacio muestral para estimar el tipo de modelo para realizar la regresión y su capacidad. (tantas h)
- 1.4. Especificar nuevos datos a recolectar para ampliar el dataset.
- 1.5. Definir un entorno de desarrollo para el modelo a entrenar.
- 1.6. Construir el pipeline para la transformación de los datos crudos al dataset depurado, para entrenar y evaluar el modelo.
- 1.7. Definir el pre-procesamiento a realizar a las muestras para incrementar el desempeño del modelo.
- 1.8. Implementar y evaluar al menos dos modelos, uno de mayor capacidad y otro de menor capacidad.
- 1.9. Cuantizar un modelo para despliegue sobre la plataforma
- 1.10. Desplegar el modelo en una plataforma que soporte las librerías utilizadas para ML y tenga una capacidad de cómputo medianamente aceptable.
- 1.11. Realizar el análisis de resultados y sugerir mejoras de desempeño, modificando y re-entrenando los modelos, o adquiriendo nuevas muestras.

2. Grupo de tareas 2

- 2.1. Documentar el proceso de transformación de los datos crudos al dataset depurado, incluyendo los detalles de pre-procesamiento y limpieza de datos.
- 2.2. Documentar la especificación para la ampliación de las muestras.
- 2.3. Documentar la implementación de los modelos, incluyendo detalles sobre los hiper-parámetros y la arquitectura utilizada.
- 2.4. Documentar la cuantización del modelo
- 2.5. Documentar el proceso de despliegue del modelo en la plataforma, incluyendo detalles sobre las librerías y capacidades de cómputo utilizadas.
- 2.6. Documentar los resultados obtenidos y las sugerencias para mejoras de desempeño.

3. Grupo de tareas 3

- 3.1. Realizar las pruebas necesarias para verificar que todo este funcionando.
- 3.2. Diseño de casos de prueba
- 3.3. Preparación del entorno de prueba
- 3.4. Ejecución de pruebas
- 3.5. Reporte de resultados
- 3.6. Corrección de errores
- 3.7. Investigación de requisitos de la plataforma
- 3.8. Análisis de compatibilidad del modelo
- 3.9. Ajustes necesarios para cumplir con los requisitos

- 3.10. Verificación de compatibilidad
- 3.11. Documentación de compatibilidad

Cantidad total de horas: (tantas h)

10. Diagrama de Activity On Node

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

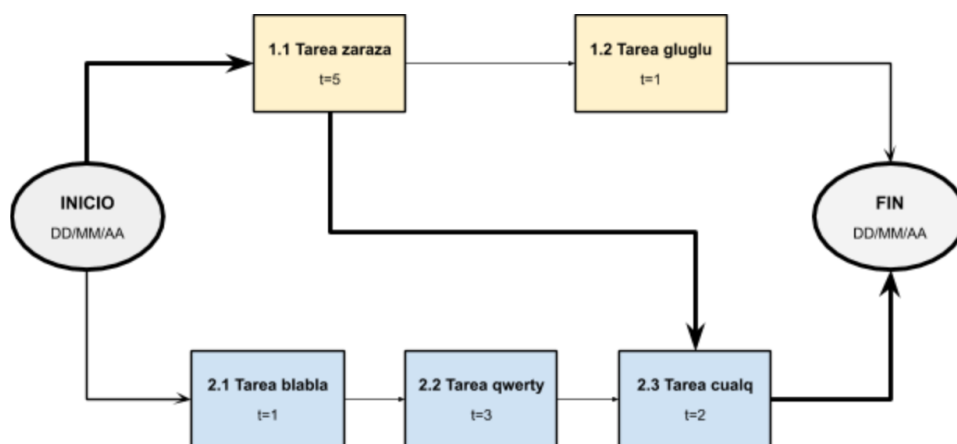


Figura 2. Diagrama de *Activity on Node*.

Indicar claramente en qué unidades están expresados los tiempos. De ser necesario indicar los caminos semicríticos y analizar sus tiempos mediante un cuadro. Es recomendable usar colores y un cuadro indicativo describiendo qué representa cada color, como se muestra en el siguiente ejemplo:

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.

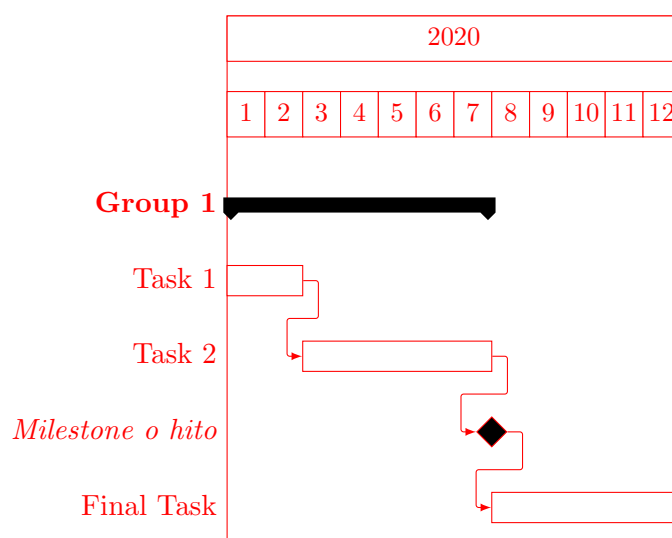


Figura 3. Diagrama de Gantt de ejemplo



Figura 4. Ejemplo de diagrama de Gantt rotado

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**.

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):
- 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

Para cada uno de los requerimientos del proyecto indique:

- 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.