



Desarrollo de un Modelo Predictivo basado en Aprendizaje Automático para Trading Algorítmico en el Futuro E-mini Nasdaq-100 (MNQ)

Autor:

Ing. Gustavo Uñapillco

Director:

Título y Nombre del director (pertenencia)

*Esta planificación fue realizada en el curso de Gestión de proyectos
entre el 29 de abril de 2025 y el 17 de junio de 2025.*

Índice

1. Descripción técnica-conceptual del proyecto a realizar	5
2. Identificación y análisis de los interesados	6
3. Propósito del proyecto	7
4. Alcance del proyecto	7
5. Supuestos del proyecto.	8
6. Requerimientos	8
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	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	29 de abril de 2025
1	Se completa hasta el punto 5 inclusive	10 de Mayo de 2025

Acta de constitución del proyecto

Buenos Aires, 29 de abril de 2025

Por medio de la presente se acuerda con el Ing. Gustavo Uñapillco que su Trabajo Final de la Carrera de Especialización en Inteligencia Artificial se titulará “Desarrollo de un Modelo Predictivo basado en Aprendizaje Automático para Trading Algorítmico en el Futuro E-mini Nasdaq-100 (MNQ)” y consistirá en el desarrollo de un modelo predictivo basado en aprendizaje automático para trading algorítmico en el Futuro E-mini Nasdaq-100 (MNQ). El trabajo tendrá un presupuesto preliminar estimado de 600 horas y un costo estimado de \$ XXX, con fecha de inicio el 29 de abril de 2025 y fecha de presentación pública el 15 de mayo de 2025.

Se adjunta a esta acta la planificación inicial.

Dr. Ing. Ariel Lutenberg
Director posgrado FIUBA

Nombre del cliente
Empresa del cliente

Título y Nombre del director
Director del Trabajo Final

1. Descripción técnica-conceptual del proyecto a realizar

La transformación digital en el ámbito financiero ha dado lugar a nuevas formas de operar en los mercados, donde la velocidad, la precisión y la capacidad de procesar grandes volúmenes de datos se vuelven esenciales para la toma de decisiones. En este contexto, el *trading* algorítmico se ha consolidado como una herramienta clave para automatizar estrategias de inversión, reducir el sesgo humano y aumentar la eficiencia operativa. No obstante, los desafíos inherentes a la alta volatilidad y la complejidad de los mercados financieros siguen demandando enfoques más robustos e inteligentes.

Este proyecto se enmarca como un emprendimiento personal, impulsado por el interés del autor en la intersección entre inteligencia artificial, programación y mercados financieros. Su objetivo principal es desarrollar un sistema predictivo basado en modelos de aprendizaje automático, orientado a la anticipación de movimientos intradía del contrato de futuros E-mini Nasdaq-100 (MNQ), un instrumento financiero caracterizado por su alta liquidez y comportamiento dinámico minuto a minuto.

La propuesta consiste en construir un modelo que, a partir de datos históricos de alta frecuencia, permita predecir la evolución del precio en horizontes de muy corto plazo. Para ello, se emplearán series temporales que contienen los valores OHLCV (Open, High, Low, Close, Volume), extraídas desde la plataforma NinjaTrader, complementadas con indicadores técnicos tales como RSI, MACD y medias móviles exponenciales. Además, se explorará la generación de factores compuestos o transformados, a partir de la combinación de estos indicadores, conocidos como *alpha factors*, con el objetivo de capturar relaciones no lineales y patrones de mayor valor predictivo. Estas series serán sometidas a procesos de normalización, estructuración temporal y reducción de ruido, con el fin de capturar la dinámica subyacente del mercado.

Desde el punto de vista metodológico, se desarrollarán y compararán modelos de aprendizaje supervisado. Se priorizará el uso de redes neuronales recurrentes, como LSTM y GRU, por su capacidad de modelar secuencias temporales y capturar dependencias de corto y largo plazo. También se evaluarán modelos como XGBoost, ampliamente utilizados por su capacidad de generalización y eficiencia. El enfoque no se limitará exclusivamente a estos algoritmos, sino que también se dejará abierta la posibilidad de explorar arquitecturas más complejas, como modelos basados en *Transformers* para series temporales, siempre que los recursos disponibles y el tiempo del proyecto lo permitan.

El desempeño de los modelos será evaluado mediante un *pipeline* automatizado de entrenamiento, validación y *backtesting*, utilizando métricas estadísticas (MAE, RMSE, R^2) y métricas específicas del dominio financiero, tales como el retorno promedio, ratio de Sharpe y tasa de aciertos operativos. El sistema será validado en entorno *offline* con datos históricos no utilizados durante el entrenamiento, y se documentará todo el proceso de desarrollo técnico.

El proyecto aún no cuenta con un prototipo funcional. Hasta el momento, solo se ha dado un primer paso durante el trabajo práctico final de la materia *Análisis de Series Temporales*, correspondiente al curso de especialización. En esa instancia preliminar se exploraron modelos sencillos, como regresiones ARIMA, aplicados sobre una fracción de los datos, pero los resultados obtenidos fueron poco satisfactorios debido a la limitada capacidad de dichos modelos para capturar la complejidad del comportamiento del índice MNQ. El presente proyecto constituye así el primer intento formal y sistemático de abordar el problema desde un enfoque más robusto, con técnicas de aprendizaje automático, una arquitectura escalable y un proceso de validación más riguroso.

Es importante señalar que el objetivo inmediato de este proyecto es validar la capacidad predictiva del sistema. La generación automatizada de señales de compra y venta no será implementada en esta etapa, pero se contempla como una línea de desarrollo futuro, una vez que se haya verificado el rendimiento del modelo en condiciones controladas.

El proyecto se encuentra actualmente en una etapa de planificación y diseño. No existen acuerdos de confidencialidad ni restricciones de propiedad intelectual, dado que se trata de un desarrollo individual con datos de acceso público o comercial. La solución ha sido concebida con una arquitectura modular y replicable, lo que permitirá su futura adaptación a entornos reales o incluso su integración con fuentes de datos en tiempo real mediante APIs especializadas.

En la figura 1 se presenta el diagrama en bloques del sistema propuesto. Se observa que el flujo comienza con la recopilación de datos históricos de alta frecuencia del índice MNQ, los cuales son preprocesados y estructurados en secuencias temporales para su posterior uso en el entrenamiento del modelo de predicción. Una vez entrenado, el modelo es evaluado tanto desde un enfoque estadístico como financiero. Finalmente, se contempla una etapa futura donde las predicciones obtenidas podrían ser traducidas en señales operativas automatizadas de compra y venta, sujeto a la validación del rendimiento predictivo.



Figura 1. Diagrama en bloques del sistema.

2. Identificación y análisis de los interesados

- Orientador: El director aún no fue designado.

Rol	Nombre y Apellido	Organización	Puesto
Responsable	Ing. Gustavo Uñapillco	FIUBA	Alumno
Orientador	Título y Nombre del director	pertenencia	Director del Trabajo Final

3. Propósito del proyecto

El propósito de este proyecto es desarrollar un modelo predictivo basado en técnicas de aprendizaje automático que permita anticipar los movimientos de muy corto plazo del contrato de futuros E-mini Nasdaq-100 (MNQ), utilizando datos históricos de alta frecuencia. La intención es explorar el potencial de la inteligencia artificial como herramienta para el análisis de series temporales financieras, enfocándose en su aplicabilidad operativa y en la comprensión de patrones complejos del mercado. El proyecto busca validar si es posible construir una solución precisa y robusta que sirva como base para estrategias automatizadas en entornos financieros reales.

4. Alcance del proyecto

Este proyecto incluye la recolección y análisis de datos históricos de alta frecuencia del contrato de futuros E-mini Nasdaq-100 (MNQ), específicamente con resolución de un minuto, provenientes de la plataforma NinjaTrader. Los datos considerados abarcan los precios de apertura, cierre, máximo, mínimo y volumen negociado (OHLCV).

Dentro del alcance se encuentra el preprocesamiento de dichos datos, que implica tareas de limpieza, normalización y estructuración en ventanas temporales. También se incorporarán indicadores técnicos complementarios, tales como el RSI, medias móviles y MACD, junto con la generación de factores combinados o transformados conocidos como *alpha factors*, diseñados a partir de la combinación o transformación de variables existentes, con el objetivo de mejorar la capacidad predictiva del modelo.

Se realizará el diseño, entrenamiento y evaluación de modelos de predicción basados en técnicas de aprendizaje automático. En particular, se utilizarán redes neuronales recurrentes (LSTM, GRU) y algoritmos de *boosting* como XGBoost. No obstante, el enfoque no se limitará exclusivamente a estos modelos, sino que se explorará también la viabilidad de arquitecturas más complejas y avanzadas, tales como *Transformers* adaptados a series temporales o modelos híbridos, en la medida en que los tiempos del proyecto y los recursos disponibles lo permitan.

La validación del desempeño se hará mediante un *pipeline* automatizado, utilizando métricas estadísticas (MAE, RMSE, R^2) y métricas financieras relevantes (retorno medio, ratio de Sharpe, tasa de aciertos).

El sistema será evaluado de forma *offline*, utilizando datos históricos no vistos, y se documentarán todos los resultados, decisiones técnicas y hallazgos. Se elaborará además la memoria técnica del proyecto y la presentación correspondiente para la defensa.

Quedan fuera del alcance del proyecto la integración en tiempo real con plataformas de *trading* o APIs financieras, así como la ejecución de operaciones reales en el mercado. Tampoco se contempla la generación de señales automatizadas de compra y venta como resultado final del trabajo. Esto se debe a que el proyecto contempla un máximo estimado de 600 horas de

dedicación, lo cual delimita el alcance a la construcción, entrenamiento y validación de modelos predictivos en entorno *offline*. La exploración de funcionalidades operativas en tiempo real se plantea como una posible línea de trabajo futuro, sujeta a la validación del rendimiento del sistema en condiciones controladas y a la disponibilidad de recursos técnicos adicionales.

5. Supuestos del proyecto

Para el desarrollo del presente proyecto se supone que:

- El responsable del proyecto contará con una disponibilidad horaria estimada de 600 horas, distribuidas de forma regular a lo largo del cronograma previsto por la carrera.
- Se podrá acceder sin restricciones a los datos históricos minuto a minuto del contrato E-mini Nasdaq-100 (MNQ) mediante la plataforma NinjaTrader, o por otros medios alternativos equivalentes.
- Se contará con los recursos computacionales personales necesarios (una notebook equipada con entorno de desarrollo Python y bibliotecas compatibles) para realizar el procesamiento, entrenamiento y evaluación de los modelos propuestos.
- No habrá impedimentos legales, comerciales o de derechos de autor que limiten el uso de los datos descargados ni la publicación de los resultados obtenidos.
- El entorno técnico y académico (orientador, repositorio, bibliografía, plataforma de desarrollo) estará disponible para consultas, revisiones periódicas y documentación durante toda la duración del proyecto.
- Los modelos de aprendizaje automático seleccionados podrán entrenarse de forma local o en la nube sin requerir infraestructura especializada adicional.
- No se prevén cambios drásticos en las condiciones regulatorias del acceso a datos de mercado, ni restricciones en el uso de herramientas *open source* empleadas.

6. Requerimientos

Los requerimientos deben enumerarse y de ser posible estar agrupados por afinidad, por ejemplo:

1. Requerimientos funcionales:

- 1.1. El sistema debe...
- 1.2. Tal componente debe...
- 1.3. El usuario debe poder...

2. Requerimientos de documentación:

- 2.1. Requerimiento 1.
- 2.2. Requerimiento 2 (prioridad menor)

3. Requerimiento de testing...

4. Requerimientos de la interfaz...
5. Requerimientos interoperabilidad...
6. etc...

Leyendo los requerimientos se debe poder interpretar cómo será el proyecto y su funcionalidad.

Indicar claramente cuál es la prioridad entre los distintos requerimientos y si hay requerimientos opcionales.

!!!No olvidarse de que los requerimientos incluyen a las regulaciones y normas vigentes!!!

Y al escribirlos seguir las siguientes reglas:

- Ser breve y conciso (nadie lee cosas largas).
- Ser específico: no dejar lugar a confusiones.
- Expresar los requerimientos en términos que sean cuantificables y medibles.

7. Historias de usuarios (*Product backlog*)

Descripción: en esta sección se deben incluir las historias de usuarios y su ponderación (*history points*). Recordar que las historias de usuarios son descripciones cortas y simples de una característica contada desde la perspectiva de la persona que desea la nueva capacidad, generalmente un usuario o cliente del sistema. La ponderación es un número entero que representa el tamaño de la historia comparada con otras historias de similar tipo.

Se debe indicar explícitamente el criterio para calcular los *story points* de cada historia.

El formato propuesto es:

1. “Como [rol] quiero [tal cosa] para [tal otra cosa].”
Story points: 8 (complejidad: 3, dificultad: 2, incertidumbre: 3)

8. Entregables principales del proyecto

Los entregables del proyecto son (ejemplo):

- Manual de usuario.
- Diagrama de circuitos esquemáticos.
- Código fuente del firmware.
- Diagrama de instalación.

- Memoria del trabajo final.
- etc...

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.

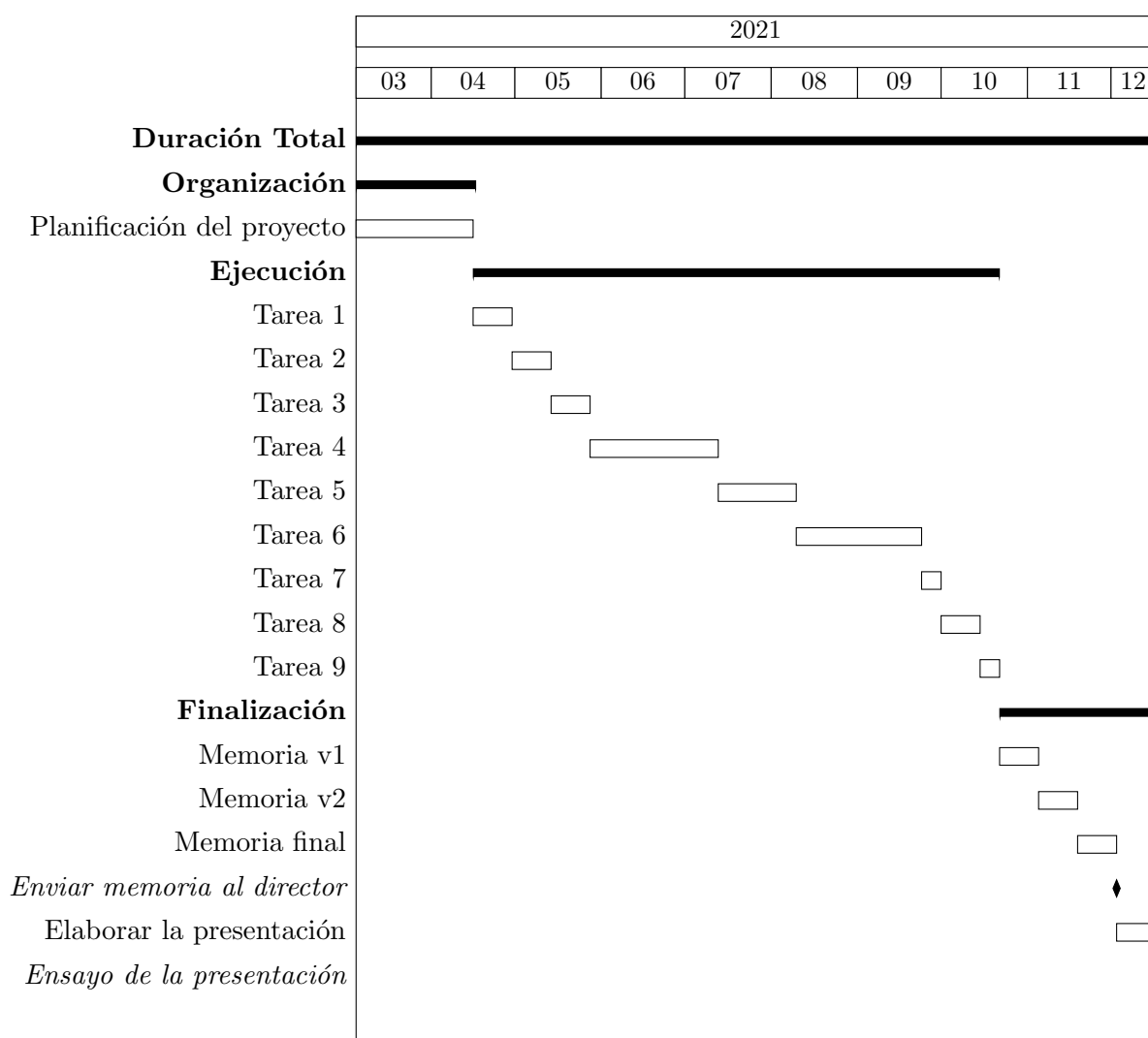


Figura 3. Diagrama de gantt de ejemplo



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.