

Pronóstico de ventas para toma de decisiones en comercio electrónico usando *Machine Learning*

Ing. Jonathan Matias Borda

Carrera de Especialización en Inteligencia Artificial

Director: Dr. Diego Onna (Universidad de Buenos Aires)

Jurados:

Jurado 1 (pertenencia)
Jurado 2 (pertenencia)
Jurado 3 (pertenencia)

Resumen

En la presente memoria se describe el desarrollo de un modelo de inteligencia artificial para el pronóstico de ventas en la tienda en línea de la empresa Latech, dedicada a la producción y comercialización de barritas alimenticias. El trabajo tiene como finalidad mejorar la planificación de la producción y la gestión del inventario mediante predicciones precisas basadas en datos históricos de ventas y publicidad.

Para su realización se aplicaron conocimientos de análisis de datos, modelado de series temporales, aprendizaje automático, integración de APIs y desarrollo de software, con el propósito de generar una herramienta que sirva de apoyo a la toma de decisiones estratégicas en la empresa.

Índice general

Resumen	I
1. Introducción general	1
1.1. Contexto del Trabajo	1
1.2. Motivación	2
1.3. Estado del arte	2
1.4. Objetivo y Alcance	3
1.4.1. Objetivo del trabajo	3
1.4.2. Alcance del trabajo	4
1.4.3. Condiciones del trabajo	4
2. Introducción específica	7
2.1. Requerimientos	7
2.1.1. Descripción técnica y alcances	8
2.2. Modelos de inteligencia artificial utilizados	12
2.3. Tratamiento de datos (herramientas)	12
3. Diseño e implementación	13
3.1. Arquitectura del sistema	13
3.2. Análisis de datos	13
3.3. Desarrollo de modelos	13
3.4. Desarrollo de un framework modular	13
3.5. Despliegue con microservicios e integración con bases de datos	13
4. Ensayos y resultados	15
4.1. Ensayos de modelos	15
4.2. Desempeño del modelo	15
4.3. Desempeño de la predicción	15
4.4. Comparación de algoritmos	15
4.5. Validación de cumplimiento de requerimientos	15
5. Conclusiones	17
5.1. Resultados obtenidos	17
5.2. Trabajo futuro	17

Índice de figuras

1.1. Diagrama en bloques.	1
-----------------------------------	---

Índice de tablas

2.1. Requerimientos	7
2.2. Requerimientos	8

Capítulo 1

Introducción general

En este capítulo se presenta el contexto general del trabajo y la problemática que motivó su desarrollo. Se exponen las razones y necesidades que impulsaron su realización, junto con una descripción de las soluciones existentes y los enfoques actuales relacionados con la temática abordada. Asimismo, se detallan los propósitos principales del trabajo y se delimitan los alcances y límites de su implementación.

1.1. Contexto del Trabajo

El presente trabajo forma parte de la Carrera de Especialización en Inteligencia Artificial y tiene como propósito el desarrollo de un modelo de inteligencia artificial capaz de pronosticar las ventas de una tienda en línea perteneciente a la empresa Latech. La empresa fabrica y comercializa barritas alimenticias en distintos sabores, lo que requiere una planificación de producción precisa para evitar tanto faltantes como excedentes de stock. En este contexto, contar con pronósticos confiables de ventas constituye un factor estratégico clave para la toma de decisiones operativas y comerciales.

El trabajo se desarrolló con financiamiento de la empresa Latech y tuvo un presupuesto estimado de 684 horas y un costo aproximado de \$12.825.000, con fecha de inicio el 24 de junio de 2025 y presentación pública prevista para abril de 2026.

Para la elaboración del modelo se dispuso del acceso a la API de Shopify, plataforma en la nube que aloja la tienda del cliente, y a la API de Triple Whale, una herramienta especializada en la gestión y análisis de inversión publicitaria. Estas fuentes de datos permitieron obtener información histórica de aproximadamente dos años sobre ventas e inversión en publicidad digital.

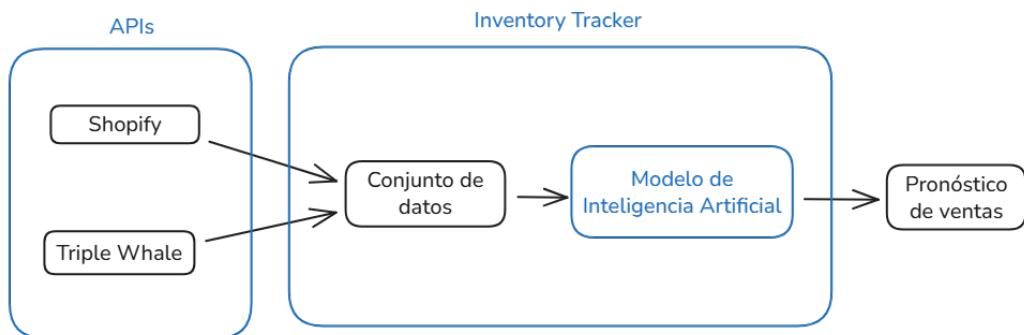


FIGURA 1.1. Diagrama en bloques.

El sistema desarrollado se integró dentro del entorno interno del cliente, denominado Inventory Tracker, a fin de generar pronósticos de ventas a partir de los datos recopilados. En la Figura 1 se presenta el diagrama en bloques del sistema, donde se observa el flujo de información desde las APIs externas hasta la generación de los resultados.

1.2. Motivación

El desarrollo de este trabajo surge de la necesidad de la empresa Latech de disponer de una herramienta que le permita anticipar el comportamiento de las ventas de su tienda en línea y, de esta manera, mejorar la planificación de la producción y la gestión del stock.

La falta de un sistema de pronóstico confiable generaba dificultades en la toma de decisiones estratégicas y operativas, esto afectaba tanto la disponibilidad de productos como la eficiencia de las campañas de marketing. Por ello, se consideró necesario diseñar una solución basada en técnicas de inteligencia artificial que proporcionara predicciones con un alto grado de precisión y permitiera analizar la evolución de las ventas en el tiempo.

Asimismo, se buscó analizar el comportamiento de los usuarios. Se observó la cantidad de clientes recurrentes y los canales de venta utilizados. Los datos indicaron que las compras se realizaban a través de diferentes plataformas: Skio, un servicio que gestiona pagos y ventas adicionales en la tienda; TikTok, Facebook, Instagram y la propia tienda alojada en Shopify. Esta información resultó fundamental para comprender los patrones de compra y optimizar las predicciones del modelo.

Además, el uso de fuentes de datos integradas, como Shopify y Triple Whale, posibilitó aprovechar información histórica de ventas, inversión publicitaria y comportamiento de los usuarios, lo que resultó esencial para la construcción de un modelo predictivo robusto y adaptable a las necesidades del cliente.

En este sentido, el trabajo buscó no solo resolver una problemática específica de la empresa, sino también demostrar el potencial del aprendizaje automático aplicado al comercio electrónico como herramienta de apoyo a la toma de decisiones empresariales.

1.3. Estado del arte

En la actualidad, el pronóstico de ventas mediante técnicas de inteligencia artificial constituye un área de creciente interés en el ámbito del comercio electrónico. Los avances en el análisis de datos y en el aprendizaje automático han permitido el desarrollo de modelos capaces de anticipar la demanda con altos niveles de precisión. Esto contribuye a una mejor planificación de la producción, la gestión de inventarios y la toma de decisiones estratégicas.

Entre los enfoques más utilizados se encuentran los modelos de series temporales clásicos, como ARIMA, SARIMA y *Exponential Smoothing*, que permiten capturar patrones estacionales y tendencias a lo largo del tiempo. Sin embargo, estos métodos suelen presentar limitaciones cuando los datos incluyen múltiples factores

externos, como campañas publicitarias, canales de venta o comportamiento del usuario.

En respuesta a esas limitaciones, en los últimos años se ha incrementado el uso de técnicas de aprendizaje profundo (*Deep Learning*) y aprendizaje automático (*Machine Learning*), que ofrecen una mayor capacidad para modelar relaciones no lineales y para integrar múltiples fuentes de información. Modelos como Redes Neuronales Recurrentes (RNN), LSTM (*Long Short-Term Memory*) y *Random Forest Regressor* se han aplicado con éxito en la predicción de ventas en entornos de comercio electrónico, debido a su habilidad para capturar dependencias temporales y correlaciones entre variables.

Asimismo, el uso de plataformas integradas de datos, tales como Shopify y Triple Whale, ha impulsado el desarrollo de soluciones personalizadas que combinan información transaccional, métricas de marketing y comportamiento de los usuarios. Estas herramientas permiten no solo mejorar la precisión de los modelos de pronóstico, sino también ofrecer una visión integral del desempeño comercial y de la efectividad de las campañas publicitarias.

Entre las soluciones comerciales más reconocidas se destacan Amazon Forecast, un servicio basado en redes neuronales desarrollado por Amazon Web Services que automatiza la creación de modelos de predicción de demanda, y Prophet, una herramienta de código abierto creada por Meta (Facebook) que utiliza un enfoque aditivo para modelar tendencias y estacionalidades de manera flexible. Ambas herramientas representan referencias relevantes en el campo del pronóstico de series temporales con datos de negocios.

La tendencia actual en el campo apunta hacia el desarrollo de modelos híbridos, que combinan la capacidad de las redes neuronales para capturar relaciones no lineales y complejas con la robustez e interpretabilidad de las técnicas tradicionales (como ARIMA o SARIMA). Esta combinación permite alcanzar una mayor precisión y robustez en los pronósticos. El sector minorista ha sido el principal adoptante de estos enfoques, ya que sus ventas están muy influenciadas por factores externos como la publicidad digital, las promociones y la estacionalidad.

En este marco, el trabajo desarrollado se apoyó en los enfoques actuales de predicción de demanda mediante aprendizaje automático, datos históricos de ventas, inversión publicitaria y comportamiento del usuario. El objetivo fue construir un modelo que reflejara las tendencias reales del negocio y sirviera como herramienta de apoyo para la toma de decisiones estratégicas en la empresa Latech.

1.4. Objetivo y Alcance

En esta sección se mencionan los propósitos principales del trabajo y los límites de su implementación.

1.4.1. Objetivo del trabajo

El propósito de este trabajo 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, podrá reducir el riesgo de faltantes o excesos de inventario, y mejorar la eficiencia operativa y comercial.

1.4.2. Alcance del trabajo

A continuación, se enuncia los items que incluye el trabajo:

- 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 ofrece 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.

A modo de limitaciones en la implementación se enuncian los items que no incluye el trabajo:

- Implementaciones en tiempo real del modelo, es decir, sistemas que actualicen las predicciones de forma instantánea ante cada nuevo dato recibido. Si 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 precisión 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.

1.4.3. Condiciones del trabajo

El desarrollo del presente trabajo se realizó bajo las siguientes condiciones:

- La empresa Latech proporcionó acceso completo y continuo a las APIs externas de Shopify y Triple Whale, lo que permitió obtener datos históricos y actualizados.
- Se contó 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 fue suficiente para entrenar y validar los modelos de predicción.
- Durante el período de desarrollo no se produjeron cambios significativos en las políticas de acceso ni en la estructura de datos de las APIs utilizadas.
- No existieron restricciones legales o contractuales que impidieran el uso de los datos necesarios para el trabajo.

Capítulo 2

Introducción específica

2.1. Requerimientos

En esta sección se presentan y describen de manera ordenada los requerimientos establecidos para el trabajo. En cada requerimiento se detalla su objetivo, alcance y relevancia, con el fin de precisar las condiciones necesarias para el desarrollo y la correcta implementación del modelo propuesto.

Requerimiento	Descripción
Requerimiento Funcional 1: Importación de datos de ventas desde Shopify	El sistema debe permitir la obtención de datos históricos y actualizados de ventas desde la plataforma Shopify, con el objetivo de utilizarlos en el proceso de entrenamiento del modelo de predicción de demanda.
Requerimiento Funcional 2: Importación de datos de inversión publicitaria desde Triple Whale	El sistema debe permitir la incorporación de datos de inversión publicitaria provenientes de la plataforma Triple Whale, con el propósito de integrar esta variable en el análisis y modelado de ventas.
Requerimiento Funcional 3: Consolidación de datos provenientes de Shopify y Triple Whale	El sistema debe unificar la información de ventas y de inversión publicitaria en un único conjunto de datos, con el fin de preparar una base consistente para el entrenamiento del modelo de predicción.
Requerimiento Funcional 4: Entrenamiento del modelo de predicción de ventas	El objetivo de este requerimiento fue desarrollar y entrenar un modelo de predicción que permita anticipar el comportamiento de las ventas futuras utilizando como base el <i>dataset</i> consolidado.
Requerimiento Funcional 5: Visualización de predicciones de ventas en Inventory Tracker	El objetivo de este requerimiento fue permitir a los usuarios de Inventory Tracker consultar y analizar las predicciones de ventas generadas por el modelo, integrándolas con los datos históricos disponibles para facilitar la planificación de la producción.

TABLA 2.1. Requerimientos

Requerimiento	Descripción
Requerimiento Funcional 6: Descarga de predicciones de ventas en formato CSV	El objetivo de este requerimiento fue permitir a los usuarios de Inventory Tracker exportar las predicciones de ventas a un archivo CSV para su análisis externo o integración con otras herramientas de planificación.
Requerimiento Funcional 7: Alertas por datos faltantes en el sistema	El objetivo de este requerimiento fue garantizar que los responsables del sistema puedan identificar oportunamente la falta de datos críticos en el <i>dataset</i> , de manera que la precisión de las predicciones no se vea comprometida.

TABLA 2.2. Requerimientos

2.1.1. Descripción técnica y alcances

A continuación se detalla la descripción técnica y alcance correspondientes a cada uno de los requerimientos especificados anteriormente.

Requerimiento funcional 1

Para cumplir este requerimiento, se desarrolló un módulo de integración que:

- Establece conexión con la API de Shopify mediante credenciales *API Key* y *Access Token*.
- Obtiene y procesa los siguientes atributos mínimos: fecha de compra, identificador de producto, cantidad vendida, precio unitario y canal de venta.
- Permite especificar intervalos de fechas para la descarga de datos.
- Almacena los datos importados en una base de datos PostgreSQL destinada a su posterior uso en el modelado.
- Implementa validaciones y manejo de errores en caso de fallos de autenticación, pérdida de conexión o respuestas vacías.
- Registra la trazabilidad del proceso de importación mediante un archivo de log, incluyendo volumen de datos procesados y errores detectados.
- Verifica duplicidades para evitar el doble procesamiento de registros previamente importados.

Requerimiento funcional 2

Para satisfacer este requerimiento se desarrolló un módulo de integración que:

- Establece conexión con la API de Triple Whale con una clave de acceso provista por la organización.
- Recupera la información de inversión por fecha, campaña y plataforma publicitaria.
- Permite seleccionar y configurar el intervalo temporal de los datos a importar.

- Almacena los datos obtenidos en una base de datos PostgreSQL, manteniendo una estructura consistente con el modelo de análisis utilizado.
- Implementa mecanismos de manejo de errores ante fallos de autenticación, interrupciones en la conexión o respuestas inválidas provenientes de la API.
- Registra el detalle del proceso mediante un archivo de log, incluyendo la cantidad de registros importados y los errores detectados durante la ejecución.
- Permite ejecutar el proceso de importación tanto de forma manual como programada mediante tareas periódicas.
- Incluye validaciones para evitar la duplicación de registros en caso de importaciones sobre el mismo rango de fechas.

Requerimiento funcional 3

Para satisfacer este requerimiento se desarrolló un proceso de consolidación que integra los datos provenientes de Shopify (ventas) y Triple Whale (inversión publicitaria), y garantiza la correcta correspondencia temporal entre ambas fuentes. En particular:

- Se generó un único *dataset* en el que las ventas diarias y la inversión total diaria se vinculan a través del campo de fecha.
- Se realizó el alineamiento temporal de las fechas, y se tuvo en cuenta posibles diferencias de zona horaria entre ambas plataformas.
- En los casos en que alguna de las fuentes no presentaba información para una fecha determinada, se mantuvo el registro correspondiente utilizando valores nulos para evitar la pérdida de información histórica.
- El conjunto de datos final incluye las siguientes columnas: `fecha`, `ventas_diarias`, `inversion_diaria` y `plataforma`.
- Se aplicaron validaciones para detectar y corregir duplicados, inconsistencias o formatos inválidos en los registros procesados.
- El proceso de consolidación se diseñó de manera automatizada y reproducible para su ejecución periódica sin intervención manual.
- El *dataset* resultante se almacenó en un formato estructurado apto para ser consumido directamente en las etapas de modelado y entrenamiento.
- Se registraron las operaciones del proceso mediante archivos de log, incluyendo la cantidad total de registros consolidados y cualquier incidencia detectada durante la ejecución.

Requerimiento funcional 4

Para cumplir con este requerimiento se implementó un proceso de modelado y entrenamiento estructurado que incluyó las siguientes etapas:

- Se empleó el *dataset* consolidado, compuesto por variables de ventas y de inversión publicitaria, como fuente principal para el entrenamiento del modelo.

- El conjunto de datos fue dividido en subconjuntos de entrenamiento y validación, siguiendo una proporción aproximada de 80 % para entrenamiento y 20 % para validación.
- Se evaluaron distintos modelos de predicción utilizando métricas de error adecuadas para series temporales, tales como MAE, RMSE y MAPE, con el fin de comparar su desempeño.
- Se seleccionó como modelo final aquel que presentó el mejor desempeño sobre el conjunto de validación.
- El modelo entrenado quedó habilitado para generar predicciones futuras a partir de valores recientes de entrada.
- Se garantizó la reproducibilidad del proceso mediante la definición de un *pipeline* de entrenamiento que permite repetir la ejecución bajo las mismas configuraciones.
- El modelo seleccionado y sus parámetros finales fueron almacenados para su posterior uso dentro del sistema Inventory Tracker.
- Se documentó el proceso de entrenamiento de manera detallada, incluyendo la selección de variables, las técnicas de preprocesamiento aplicadas y los métodos de evaluación utilizados.
- Se realizó una validación visual comparando las predicciones generadas frente a los valores reales observados, mediante gráficos que permitieron analizar la coherencia del modelo respecto a la dinámica histórica de las ventas.

Requerimiento funcional 5

Para cumplir con este requerimiento se implementaron las siguientes funcionalidades:

- Las predicciones de ventas se muestran en la interfaz de Inventory Tracker junto a las fechas correspondientes.
- La visualización integra tanto las predicciones como los datos históricos de ventas para comparaciones temporales y análisis de tendencias.
- La información se presenta de manera clara y entendible, utilizando gráficos de líneas y tablas según corresponda.
- Los usuarios pueden seleccionar rangos de fechas específicos para observar predicciones detalladas de períodos concretos.
- Las predicciones se actualizan automáticamente al entrenarse un nuevo modelo o al actualizarse los datos de entrada y se garantiza que la información refleje siempre el estado más reciente.
- La interfaz muestra un indicador visual de la fecha y hora de generación de la última predicción.
- Se distingue claramente entre valores predichos y datos históricos que evitan confusiones en el análisis.

- La integración de las predicciones no afecta la funcionalidad ni el rendimiento general de la interfaz del módulo Inventory Tracker.
- El rendimiento de la interfaz se mantiene aceptable incluso cuando se manejan conjuntos de datos de gran tamaño.

Requerimiento funcional 6

Para cumplir con este requerimiento se implementaron las siguientes funcionalidades:

- Se incorporó un botón o enlace claramente identificado que permite la descarga de las predicciones en formato CSV.
- El archivo CSV generado incluye las fechas y los valores de predicción correspondientes.
- De manera opcional, el archivo puede incluir los datos históricos de ventas para permitir la comparación con las predicciones.
- El formato del CSV es compatible con herramientas comunes de análisis, con valores separados por comas y codificación UTF-8.
- El nombre del archivo contiene la fecha y hora de generación que lo identifica.
- La descarga se realiza correctamente en navegadores modernos, incluyendo Chrome y Firefox.
- El contenido del archivo refleja exactamente la información presentada en la interfaz del módulo de predicción.
- Si aún no existen predicciones generadas, el botón de descarga se muestra deshabilitado.

Requerimiento funcional 7

Para cumplir con este requerimiento se implementaron las siguientes funcionalidades:

- El sistema verifica automáticamente la integridad del *dataset* antes de cada entrenamiento del modelo.
- Si se detectan datos faltantes críticos como fechas sin registros de ventas o de inversión publicitaria se genera una alerta.
- La alerta incluye información detallada sobre el tipo de dato faltante (ventas, inversión, o ambos), el rango de fechas afectado y el nivel de severidad del impacto estimado.
- La alerta se muestra de forma visible dentro del sistema para que los responsables puedan tomar conocimiento inmediato.
- Se envía una notificación por correo electrónico al responsable del sistema, en caso de estar configurada esta opción.
- El sistema no bloquea el entrenamiento del modelo, pero advierte que la precisión de las predicciones puede verse afectada.

- La alerta desaparece únicamente cuando los datos faltantes son completados o cuando se marca como revisada manualmente.
- Se registra un log de todas las alertas emitidas, accesible desde una sección de administración o monitoreo del sistema.

2.2. Modelos de inteligencia artificial utilizados

2.3. Tratamiento de datos (herramientas)

Capítulo 3

Diseño e implementación

Todos los capítulos deben comenzar con un breve párrafo introductorio que indique cuál es el contenido que se encontrará al leerlo. La redacción sobre el contenido de la memoria debe hacerse en presente y todo lo referido al proyecto en pasado, siempre de modo impersonal.

- 3.1. Arquitectura del sistema**
- 3.2. Análisis de datos**
- 3.3. Desarrollo de modelos**
- 3.4. Desarrollo de un framework modular**
- 3.5. Despliegue con microservicios e integración con bases de datos**

Capítulo 4

Ensayos y resultados

Todos los capítulos deben comenzar con un breve párrafo introductorio que indique cuál es el contenido que se encontrará al leerlo. La redacción sobre el contenido de la memoria debe hacerse en presente y todo lo referido al proyecto en pasado, siempre de modo impersonal.

4.1. Ensayos de modelos

La idea de esta sección es explicar cómo se hicieron los ensayos, qué resultados se obtuvieron y analizarlos.

4.2. Desempeño del modelo

4.3. Desempeño de la predicción

4.4. Comparación de algoritmos

4.5. Validación de cumplimiento de requerimientos

Capítulo 5

Conclusiones

Todos los capítulos deben comenzar con un breve párrafo introductorio que indique cuál es el contenido que se encontrará al leerlo. La redacción sobre el contenido de la memoria debe hacerse en presente y todo lo referido al proyecto en pasado, siempre de modo impersonal.

5.1. Resultados obtenidos

La idea de esta sección es resaltar cuáles son los principales aportes del trabajo realizado y cómo se podría continuar. Debe ser especialmente breve y concisa. Es buena idea usar un listado para enumerar los logros obtenidos.

En esta sección no se deben incluir ni tablas ni gráficos.

Algunas preguntas que pueden servir para completar este capítulo:

- ¿Cuál es el grado de cumplimiento de los requerimientos?
- ¿Cuán fielmente se puedo seguir la planificación original (cronograma incluido)?
- ¿Se manifestó algunos de los riesgos identificados en la planificación? ¿Fue efectivo el plan de mitigación? ¿Se debió aplicar alguna otra acción no contemplada previamente?
- Si se debieron hacer modificaciones a lo planificado ¿Cuáles fueron las causas y los efectos?
- ¿Qué técnicas resultaron útiles para el desarrollo del proyecto y cuáles no tanto?

5.2. Trabajo futuro

Acá se indica cómo se podría continuar el trabajo más adelante.