

FORMULACIÓN DEL PROYECTO: MODELOS PARA LA PREDICCIÓN DE VENTAS DE PRODUCTOS DE WALMART

VICENTE MORENO GARRIDO[‡], JOAQUÍN RAMIREZ VILLALOBOS[‡], NICOLÁS RODRIGUEZ ANSALDO[‡],
AND YUNIS VIDAL CATALÁN[‡].

[‡] DEPARTAMENTO DE MATEMÁTICA, UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA, SANTIAGO, CHILE

CONCURSO M5 2020

La presente postulación constará de tres secciones, las que deberán cubrir los siguientes aspectos:

1. Resultados previos y contenido científico y tecnológico.

En esta sección el postulante deberá describir la problemática que busca resolver, plantear cual sería la solución a ese problema y cómo a través de un proyecto de investigación aplicada, que cuenta con resultados previos, propone avanzar hacia esa solución.

A través del proyecto presentado deberá alcanzar, en el plazo de ejecución de dos años, resultados que representen un desarrollo tecnológico que permita validar su aplicación futura como solución al problema planteado.

Deberá hacer un ejercicio de moverse desde un análisis macro del problema a la descripción en detalle de la metodología y su respectiva programación y organización de actividades que le permitirá alcanzar resultados que aportarán al logro de la solución propuesta.

2. Estrategia de desarrollo de producto e impacto ambiental y territorial.

En esta sección se deberá realizar una clara identificación del desarrollo tecnológico en la forma de prototipo de producto, proceso o servicio final que se derivará de los resultados del proyecto.

Sumado a lo anterior, la propuesta deberá presentar la identificación y caracterización del destinatario, o de la población o mercado objetivo, la identificación de las ventajas competitivas del producto, proceso o servicio final y su relevancia y potencial impacto para la población o mercado objetivo en términos económicos, sociales, ambientales y territoriales, dependiendo del objetivo y alcance del proyecto. Se deberá mencionar los atributos que diferencian a la solución propuesta de las alternativas disponibles y haciéndola más competitiva, indicando y justificando claramente si la novedad es a nivel nacional o internacional. Además, caracterizar el mercado. Por último, la propuesta deberá identificar el costo-beneficio del proyecto, así como las eventuales externalidades (entendidas como efectos más allá del bienestar de los beneficiarios que utilizan o producen la solución).

Además, se solicita describir la estrategia que enfrentaría para completar el desarrollo del producto, proceso o servicio final, identificando el rol de las entidades asociadas y de otros socios relevantes en este proceso y actores esenciales para lograr concretar un negocio.

3. Capacidades Científicas y tecnológicas y mecanismos de colaboración.

En esta sección se debe entregar la organización presupuestaria en formato ANID (planilla de costos) que permita abordar todas las actividades requeridas por el proyecto. Junto con esto, se debe entregar la definición de cargos, una descripción de las capacidades y habilidades para llevar a cabo el proyecto, el aporte de cada uno de los investigadores en las diferentes actividades programadas y los tiempos de dedicación del personal crítico en el equipo de trabajo (directores e investigadores principalmente).

Finalmente, se debe explicitar de qué manera el proyecto contribuirá a la formación de capacidades en pronósticos y en relación con la temática desarrollada.

1. CONTENIDO CIENTÍFICO Y TECNOLÓGICO

1.1. Resultados Previos.

TRL 1: Idea Básica

Se estudiaron modelos de pronóstico de series temporales, basándose en modelos de aprendizaje automático. Esto incluye los conceptos de redes neuronales y modelos de regresión.

TRL 2: Formulación del concepto

Se estudiaron modelos de pronóstico de series temporales, y como aplicarlos a la predicción de ventas de productos. Esto incluye el estudio del preprocesamiento de datos y de programación, junto con análisis sobre las variables que influyen en la venta de productos.

TRL 3: Prueba de concepto

Proponer varios modelos de aprendizaje automático basados en distintos métodos de aprendizaje, y seleccionando el modelo que muestra tener mejores métricas de error. Se verifica la validez del modelo comparando los datos predecidos por el mejor modelo encontrado con respecto a las ventas reales por producto.

TRL 4: Validación en el laboratorio

Se hace un estudio estadístico de los datos de ventas de alguna tienda. Por otra parte, se empiezan a estudiar posibles extensiones del modelo propuesto en la etapa 3 de manera que pueda incluir ventas de productos de múltiples tiendas, generando un desglose de ventas por cada tienda.

A finales de esta etapa se tiene listo un modelo de predicción versátil, que permite predecir las ventas futuras de los productos de múltiples tiendas con un margen de error razonable, durante una cantidad de días fija.

1.2. Problema u oportunidad.

Las competencias Makridakis son una serie de competencias que buscan evaluar y comparar la precisión de diferentes métodos de pronóstico (forecasting) de series temporales. La quinta edición de esta competencia comenzó el 3 de marzo de 2020 y los resultados se anunciaron el 1 de julio de 2020. Se utilizaron datos reales de Walmart y se llevó a cabo en la plataforma de Kaggle. Ofreció premios sustanciales por un total de US \$100.000 a los ganadores. Los datos, proporcionados por Walmart, consistieron en alrededor de 42,000 series temporales diarias jerárquicas, comenzando por los SKU y terminando con la demanda total de una amplia zona geográfica. Además de los datos de ventas, también se incluyó información sobre precios, actividad publicitaria/promocional y niveles de inventario, así como el día de la semana al que se refieren los datos.

Uno de los desafíos que enfrentan las grandes tiendas, como Walmart, es la capacidad de determinar de manera inequívoca la demanda de sus productos. A medida que la cantidad de consumo crece, los patrones de este mismo se vuelven dinámicos y difíciles de predecir. La necesidad de contar con modelos robustos de forecasting se vuelve fundamental para las gestiones operativas.

Naturalmente, surge la duda de hasta qué punto los modelos actuales son capaces de capturar variaciones estacionales, efectos promocionales, cambios de comportamiento del consumidor y, en general, todos los factores que influyen en la demanda diaria. Y son dudas como estas que motivó la creación de competencias como las de Makridakis.

Un eje central en el estudio de la demanda es el horizonte de predicción y el comportamiento que esta presenta en el tiempo. Si se logra obtener pronósticos confiables a largo plazo, las empresas pueden planificar de forma más eficiente su inventario, optimizar cadenas de suministro y minimizar costos asociados a quiebres de stock o sobre-producción. De igual manera, las predicciones a corto plazo, permiten ajustar decisiones tácticas como pedidos a proveedores, la distribución entre bodegas y la planificación de personal. Ambas escalas de predicción, aunque distintas, son críticas para el funcionamiento diario y estratégico de una organización de gran escala.

Por otra parte, el estudio del pronóstico de demanda no se limita a un único conjunto de productos o a un entorno específico. Las técnicas desarrolladas para series temporales de ventas pueden extrapolarse a otros contextos, como predicción de energía, tráfico, recursos hospitalarios, o incluso comportamiento financiero. En particular, el análisis realizado sobre las ventas diarias de Walmart en el contexto de la competencia M5 sirve como un laboratorio experimental que permite probar metodologías y obtener evidencia cuantitativa del rendimiento de distintos modelos, lo cual eventualmente puede aplicarse a múltiples áreas fuera del retail.

El problema presentado es el que buscamos resolver. Debido a que pretende aportar nuevo conocimiento a la literatura actual mediante la extensión de los trabajos existentes. En particular, se busca generar modelos basados en técnicas de Machine Learning y métodos de series temporales que permitan predicciones confiables de la demanda futura, incorporando distintos factores relevantes como la estacionalidad, efectos de calendario y las características propias de cada producto. Además, se pretende entregar un software para automatizar estas predicciones y facilitar su implementación en escenarios reales de toma de decisiones.

1.3. Análisis del estado del arte.

a. Estado actual de la investigación:

En la actualidad existen varias técnicas para el modelamiento y estudio del pronóstico de series temporales, entre las cuales destacamos el aprendizaje automático. Existen muchos tipos de modelos de predicción para series temporales, por ejemplo Naive o SNaive para series temporales estacionarias, ARIMA/SARIMA/ETS aplicados a toda la serie temporal o solo a segmentos, o incluso regresión con lags. Esta diversidad de modelos dan una amplia gama de modelos de predicción, y el aplicar estos diversos modelos frecuentemente resulta en alguno de ellos prediciendo correctamente el fenómeno que uno quiere modelar con alta precisión.

Actualmente en todas partes del mundo, existen empresas que predicen ventas diarias de productos para otras empresas, basadas en los métodos anteriores. Algunos ejemplos de empresas actuales que sirven esta función mediante los métodos anteriores son:

- Retalon: Software de “demand forecasting” especializado en retail. Permite predecir la demanda por SKU y por tienda, considerando estacionalidad, canibalización, elasticidad de precio, afinidades de compra, ubicación geográfica, etc.
- Prodware Demand Forecasting: Solución de forecasting basada en machine-learning, lista para integrarse con ERP (u otros sistemas), pensada para retail, distribución y manufactura. Permite automatizar predicciones con múltiples modelos en paralelo.
- ElmoSoft (servicio de consultoría / outsourcing de forecasting): Ofrece servicios de forecast para retail, combinando analítica avanzada, regresión no lineal, detección de patrones complejos, para distintos tipos de productos — desde los más estables hasta los de demanda errática.

En el estudio internacional de 2023 llamado “Improved Sales Forecasting using Trend and Seasonality Decomposition with LightGBM (2023)”, muestran que usando el marco de aprendizaje LightGBM logran predecir con errores bastante bajos la venta diaria de productos, usando un método de descomposición de los “trends” y la estacionalidad de las series temporales.

En el caso de Chile, en el estudio también de 2023 llamado “Predicción de demanda de largo plazo para mejorar planificación logística en empresa del sector retail”, estudian el problema de ventas pero a largo plazo, usando también herramientas de aprendizaje automático mencionadas anteriormente.

La solución propuesta se diferencia esencialmente en el tiempo de predicción, y la zona de predicción, es decir, si bien ya se ha realizado en Chile modelos de predicción de venta de productos, se pretende trabajar en un marco a tiempo corto, haciendo predicciones diarias de las ventas. Además, se planea extenderlo en rango de aplicabilidad, abarcando cualquier tipo de tienda o combinación de tiendas en vez de solo retail, para obtener un desglose de ventas general aplicable para cualquier contexto.

b. Análisis de la propiedad intelectual e industrial y productos existentes en el mercado

Actualmente, a lo largo de todo el mundo, los acercamientos para resolver el problema de forecasting que reinan en la predicción de venta de productos a escalas largas y cortas usan estos modelos de aprendizaje automático. Hay varios modelos que utilizan técnicas probabilísticas y análisis estocástico profundo como herramientas suplementarias para mejorar las métricas de error de las predicciones.

La teoría anterior servirá de base para el estudio, y en particular, pueden extenderse al propósito de este proyecto. Además, los modelos explicados anteriormente están liberados al público, y pueden ser utilizados sin problema como base para utilizarse en el modelo a realizar.

c. Normativas.

Las siguientes normativas son de las de mayor influencia para el proyecto:

- Ley 21.719 de Protección de Datos Personales: Si tu sistema de predicción utiliza datos de clientes, compradores o individuos (por ejemplo: historial de compra, nombres, IDs, datos demográficos, datos sensibles, datos que permitan identificar personas), entonces estás sujeto a esta ley. Entre sus requisitos están: obtener un consentimiento informado del titular, definir claramente la finalidad del uso de los datos, garantizar minimización de datos (recolectar solo lo necesario), asegurar seguridad, integridad y confidencialidad, permitir los derechos de los titulares (acceso, rectificación, eliminación, portabilidad) y registrar las operaciones de tratamiento.

- Principio de “Protección desde el diseño y por defecto” (data protection by design / by default): La normativa en Chile exige que quien desarrolle sistemas que traten datos personales aplique medidas técnicas y organizativas apropiadas desde el diseño del sistema, considerando riesgos, naturaleza de los datos, minimización del tratamiento, limitación en la recolección/uso, conservación, acceso, etc.

1.4. Solución propuesta.

La solución que proponemos se puede entender como un “laboratorio de pronósticos” construido sobre los datos de la competencia M5. La idea central es ir desde modelos muy simples y transparentes, hasta modelos más sofisticados, pero siempre comparándolos de forma justa y usando métricas que sean coherentes con las reglas de la competencia.

En una primera etapa trabajamos con modelos base como Naive y SNaive (que repiten el valor del día anterior o el de la misma semana anterior). Estos modelos no requieren muchas suposiciones y nos sirven como piso de rendimiento: cualquier modelo más complejo debería mejorar claramente estos resultados para justificar su uso.

Luego, construimos modelos de la familia SARIMA a nivel agregado (por estado y para la suma total de ventas). Estos modelos son más flexibles y permiten capturar patrones de tendencia y estacionalidad semanal de manera explícita. Para evitar “hacer trampa mirando el futuro”, entrenamos y evaluamos estos modelos con un esquema de walk-forward: avanzamos en el tiempo, re-entrenando y validando varias veces, como ocurriría en un uso real del sistema.

Sobre esa base, el siguiente paso es entrenar un modelo “global” de machine learning (por ejemplo, un modelo tipo LightGBM) que use muchas series al mismo tiempo. A este modelo no solo le entregamos las ventas pasadas, sino también un conjunto de variables explicativas construidas a partir de los datos de la competencia: rezagos de ventas (lags), promedios móviles, variaciones de precio, indicadores de promociones y eventos, información de calendario, cercanía a feriados, etc. La intuición es que este modelo pueda aprender patrones comunes entre productos y tiendas, aprovechando mejor la gran cantidad de series disponibles.

Finalmente, comparamos todos los modelos con métricas como RMSSE, RMSE y MAE, y analizamos el desempeño por subgrupos (por estado, por tienda y por tipo de producto). Con esto buscamos no solo “ganar puntos” en la métrica principal de M5, sino también entender dónde el modelo funciona bien o mal y por qué. Todo el trabajo se implementa en Python y quedará empaquetado en un prototipo de software que permita, de forma relativamente simple, generar pronósticos diarios para distintos niveles de agregación (producto, tienda, estado o total), facilitando su uso posterior en contextos reales.

Como cierre del proyecto, se desarrollará un módulo adicional que permita automatizar por completo el flujo de trabajo: desde la carga de datos crudos hasta la generación de pronósticos finales y reportes comparativos. Este módulo incluirá validaciones básicas de calidad de datos, diagnósticos de los modelos y un sistema de alertas que notifique cuando un modelo global o local presente desviaciones importantes respecto a su desempeño histórico. El objetivo es que el usuario final no solo pueda obtener pronósticos, sino también entender cuándo confiar en ellos y cuándo es necesario revisar la configuración o actualizar el modelo.

Además, se estudiará la extensión del sistema hacia modelos jerárquicos de pronóstico, de manera que las predicciones realizadas a nivel producto, tienda o estado sean coherentes entre sí. Esto es especialmente relevante para aplicaciones reales, en las que diferentes niveles de agregación deben alinearse para facilitar la gestión de inventarios, la planificación de demanda o la administración de recursos logísticos. Esta extensión requerirá experimentos adicionales con métodos de reconciliación jerárquica y evaluaciones sobre su impacto en las métricas principales.

Finalmente, se preparará una versión preliminar del software lista para ser utilizada en un entorno real, acompañada de una documentación detallada que explique su configuración, uso, limitaciones y posibles líneas de mejora. Junto con esto, se elaborarán ejemplos prácticos que ilustren cómo generar pronósticos en distintos escenarios, cómo comparar modelos entre sí y cómo interpretar las

métricas seleccionadas. El fin es que el prototipo sea suficientemente robusto para ser utilizado por analistas, estudiantes o equipos de ciencia de datos interesados en trabajar con datos de ventas y en comprender los distintos enfoques de modelamiento.

En conjunto, este proyecto busca no solo implementar modelos avanzados de pronóstico, sino también construir una herramienta didáctica y operativa que permita explorar, comparar y aplicar técnicas de forecasting de manera transparente y reproducible. Con ello esperamos entregar un aporte tanto técnico como pedagógico, dejando la base para futuras extensiones que incorporen nuevos tipos de datos, modelos más sofisticados o escenarios de aplicación más amplios.

1.5. Hipótesis y componente de investigación.

La hipótesis principal de nuestro proyecto es que la venta de productos por tienda presente patrones similarmente aproximables por el mismo modelo. Es por esta razón que la extensión del modelo base propuesta debería seguir aproximándose con métricas similares, ya que, en el caso de que esta hipótesis sea cierta, no debería haber mayor diferencia entre las predicciones de ambos modelos.

La componente de investigación científica corresponde a la realización y posterior análisis de los modelos que predigan información sobre las ventas de productos de ciertas tiendas de Walmart en EE.UU, donde este largo plazo corresponde a un plazo de 2 años. La pregunta científica asociada a esto es si estas predicciones logran predecir con un margen de error razonable las ventas reales, durante un periodo donde ya se tenga esta información. Nuestra hipótesis es que sí.

En este proyecto, el área de desarrollo de la tecnología y de impacto final corresponden al área de economía y finanzas, y no posee relación con el cambio climático.

1.6. Objetivos.

1.6.1. *Objetivo General.*

Predecir las ventas diarias de los productos de tiendas asociadas a una empresa, durante un periodo de tiempo fijo, utilizando modelos de aprendizaje automático e implementando softwares para lograr este fin.

1.6.2. *Objetivos Específicos.*

Objetivo 1.- Estudio estadístico de las variables presentes en las ventas de productos diaria de una tienda.

Objetivo 2.- Confección de modelo de aprendizaje automático simplificado, de predicción de productos de una tienda.

Objetivo 3. - Confección de modelo extendido, de predicción de productos de varias tiendas simultáneamente.

Objetivo 4. - Predicción aproximada de las ventas, mediante el modelo extendido.

Objetivo 5.- Confección de interfaz para el software de predicción de ventas diarias de productos de múltiples tiendas.

ÁREA DE DESARROLLO DE LA TECNOLOGÍA (Marcar una sola opción)		ÁREA DE IMPACTO FINAL(Marcar una opción)	
Agropecuaria		Agropecuaria	
Agua y Sector Sanitario		Agua y Sector Sanitario	
Alimentos Funcionales		Alimentos Funcionales	
Comercio y Servicios Financieros	X	Comercio y Servicios Financieros	X
Construcción e Infraestructura		Construcción e Infraestructura	
Cs. Sociales y Educación		Cs. Sociales y Educación	
Energía		Energía	
Forestal		Forestal	
Industria Creativa		Industria Creativa	
Logística y Transporte		Logística y Transporte	
Manufactura		Manufactura	
Minería		Minería	
Pesca y Acuicultura		Pesca y Acuicultura	
Salud		Salud	
Tecnología de la Información y Comunicación		Tecnología de la Información y Comunicación	
Turismo		Turismo	

FIGURE 1. Formulario: Áreas de Desarrollo e Impacto

Ciencias del Clima	Vulnerabilidad y Adaptación	Mitigación (factores de emisión de GEI, tecnologías de bajas emisiones de GEI, etc.)	
Atmósfera	Recursos hídricos	Eficiencia energética	
Criósfera	Biodiversidad	Energías renovables no convencionales (ERNC)	
Océanos	Silvoagropecuario	Procesos industriales	
	Pesca y Acuicultura	Transporte	
	Salud	Gestión de Residuos	
	Infraestructura, energía, vivienda, transportes	Sumideros (bosques)	
	Ciencias sociales y económicas	Estudios de variables para factores de emisión GEI	
	Riesgo de desastres		
No se relaciona con cambio climático			X

FIGURE 2. Formulario: Relación Cambio Climático

Objetivo 6. - Confección de Software Final

1.7. Metodologías de investigación y desarrollo.

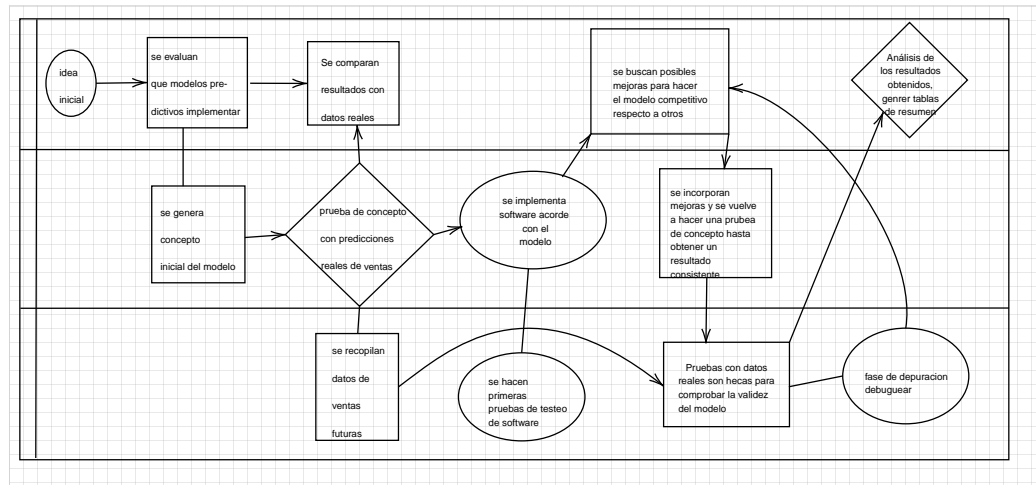


FIGURE 3. Diagrama Metodología

CERTIFICADO	Marque su opción con una "X"
Ética	
Bioética	
Bioseguridad	
Sitio Arqueológico	
Área Silvestre	
Introducción de especies protegidas	
Bases de datos que contengan información sensible	X
Otro (señale cual)	
No necesita certificaciones o permisos especiales	

FIGURE 4. Formulario: Certificados Propuestas

Declare las certificaciones o permisos especiales con que debe cumplir su propuesta. Seleccione con "X" la opción correspondiente.

1.8. Resultado de producción e hitos.

1.9. Planificación de actividades (carta Gantt).

Título del Resultado de Producción e Hitos	Breve Descripción
Objetivo 1: EDA más limpieza	Análisis exploratorio de ventas, detección de outliers, imputación de valores faltantes y estandarización para preparar los datos.
Objetivo 2: Confección de features y de baselines	Generación de variables predictoras y construcción de modelos base para comparación.
Objetivo 3: Implementación de modelos y tuning	Implementación de modelos ML y búsqueda de hiperparámetros para optimizar rendimiento.
Objetivo 4: Validación walk-forward e interpretación	Validación temporal con ventanas móviles e interpretación de resultados y variables relevantes.
Objetivo 5: Evaluación	Evaluación final del modelo en datos no vistos y comparación con baselines.

FIGURE 5. Hitos

2. ESTRATEGIA DE DESARROLLO Y DE TRANSFERENCIA TECNOLÓGICA Y ESTIMACIÓN DEL IMPACTO ECONÓMICO Y SOCIAL.

2.1. Producto, proceso o servicio a desarrollar.

Se desarrollará un modelo de aprendizaje automático de predicción de ventas diarias de productos de tiendas de una empresa, durante un periodo de tiempo fijo, en el lenguaje de programación de Python. Este podrá ser utilizado por cualquier persona natural o institución que lo adquiera desde una página web que se creará dentro de los 2 años del proyecto. El producto está pensado para ayudar a la toma de decisiones de cualquier persona natural o institución que necesite planear operaciones referente a la compra y venta de productos.

2.2. Ventajas competitivas.

Dentro del mercado de los modelos de predicción de demanda y ventas, distintas empresas tecnológicas y centros de investigación mantienen plataformas online donde se publican proyecciones de consumo, pronósticos macroeconómicos o estimaciones de demanda agregada. Sin embargo, estas herramientas suelen enfocarse en indicadores generales, sin entregar predicciones específicas, automatizadas y de alta resolución temporal para productos particulares dentro de una empresa o cadena logística.

Considerando estos principales sustitutos, las principales ventajas competitivas de nuestro desarrollo son:

- La utilización de un enfoque híbrido que integra modelos de machine learning (como ARIMA y SARIMA, Gradient Boosting y Redes Neuronales) con técnicas avanzadas de series temporales. Se selecciona de manera objetiva el modelo con mejor desempeño mediante criterios estadísticos estandarizados y métricas de error robustas (MAE, RMSE, RMSSE), garantizando predicciones de alta exactitud.

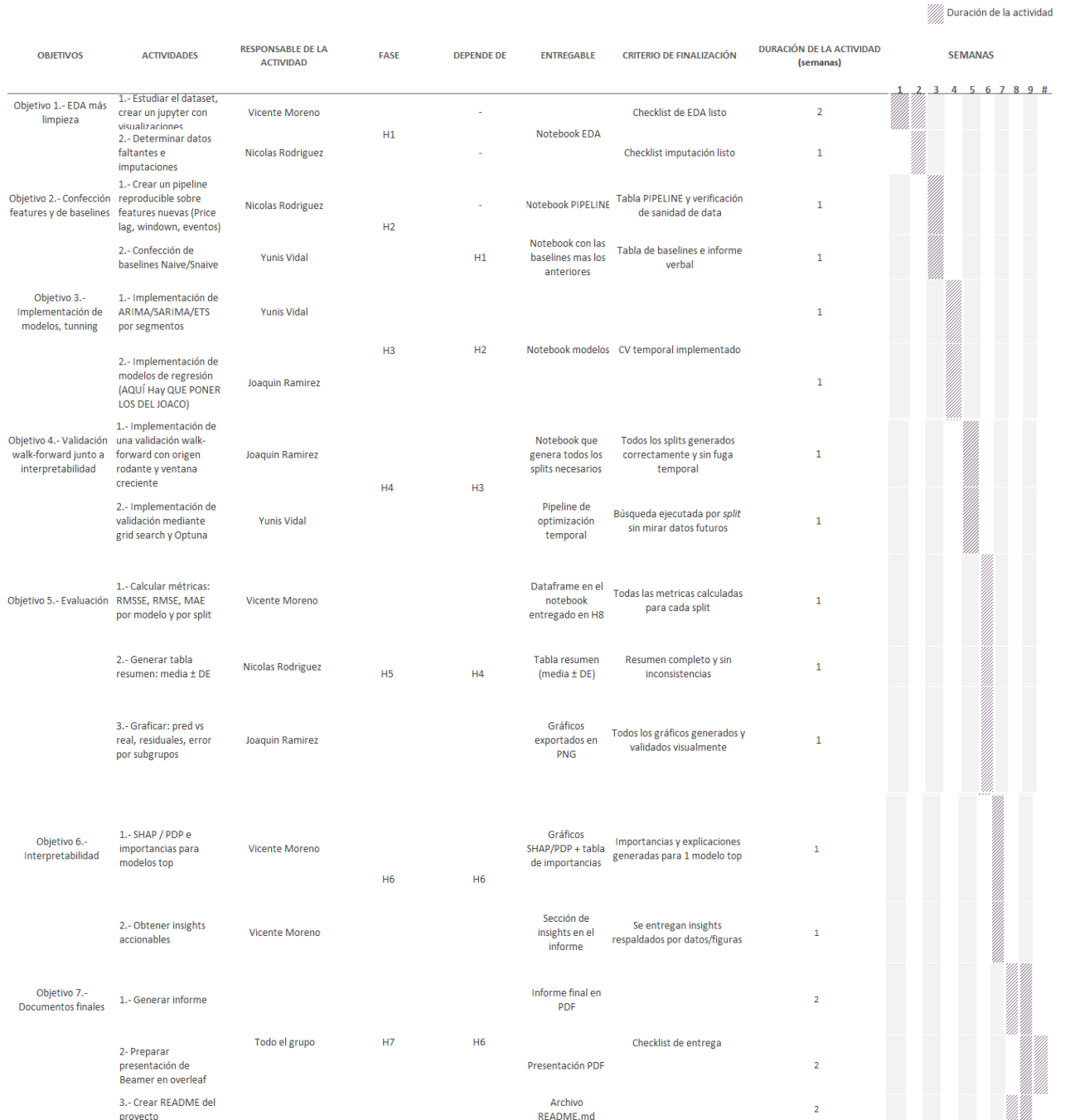


FIGURE 6. Carta Gantt

• La incorporación de datos actualizados y multifuente, integrando ventas históricas, calendarios y variables externas relevantes. Esto permite capturar dinámicas comerciales reales y adaptarse continuamente a cambios en el comportamiento del consumidor.

- La transparencia metodológica del modelo, que permite comprender la contribución de cada variable a las predicciones mediante técnicas de interpretabilidad (SHAP, importancia de características), facilitando la toma de decisiones estratégicas por parte de áreas comerciales, de abastecimiento y operaciones.

2.3. Identificación de etapas futuras y tiempo estimado para la comercialización y/o masificación. ,

El modelo y el software asociado estarán completamente desarrollados y listos para ser comercializados al término de los primeros 12 meses del proyecto. Durante este período se contempla la implementación de una plataforma web en la cual los usuarios podrán acceder al sistema, cargar sus datos y obtener predicciones de ventas de manera personalizada.

Posteriormente, durante los meses 12 a 24, se llevará a cabo una fase de monitoreo y mejora continua, orientada a validar el desempeño del modelo en entornos reales, optimizar la experiencia de usuario y ajustar el producto a los requerimientos de distintos sectores comerciales.

Luego de los 24 meses iniciales, se destinará un período adicional de un año para ampliar las capacidades del sistema hacia predicciones multiescala (horarias, semanales y mensuales), así como para incorporar módulos avanzados de optimización de inventario, detección de anomalías y recomendación de niveles de reposición.

Tras estos avances, y aproximadamente 2 años después del cierre del desarrollo inicial, se espera contar con una plataforma integral aplicable a múltiples industrias y mercados internacionales, junto con iniciar estudios de replicabilidad del modelo en distintos países y contextos comerciales.

2.4. Mercado potencial o población objetivo.

El segmento de mercado objetivo son todas las personas naturales y/o instituciones que requieran tomar decisiones respecto a la compra y venta de productos de diversas tiendas.

El mercado que se menciona está en constante evolución y permanente búsqueda de tecnologías de predicción de ventas, lo que conlleva una relevancia permanente del producto a través del tiempo. Dada la naturaleza del modelo a desarrollar, se puede extender el mercado objetivo al estudiar predicciones de precios de productos, logrando así una inserción en mercado cada vez más grandes.

Esto último respalda el proyecto para que a futuro se extienda no solo al mercado nacional y se estudie la viabilidad de inserción del producto en el mercado internacional.

2.5. Entidades Asociadas y Colaboradoras.

La entidad asociada a este proyecto es la Universidad Técnica Federico Santa María, que es la institución beneficiaria de la creación de este proyecto.

La Universidad Técnica Federico Santa María posee estudiantes y profesionales del área de Modelación Matemática destacados nacional y mundialmente. Esta universidad ha participado previamente en modelación y predicción de ventas para múltiples organizaciones, y aportará personal con experiencia en modelación para este proyecto.

Nombre / RUT	Institución	Cargo en el proyecto	Funciones y capacidades críticas que aporta al proyecto	Dedicación HH/mes	\$/HH	Actividades a desarrollar en el proyecto (individualizadas en la carta Gantt)
Vicente Moreno Garrido	Universidad Técnica Federico Santa María	Director	Experiencia en análisis de datos, EDA y gestión de personas	80	9376	EDA, evaluación de métricas, SHAP y obtener insights adicionales.
Joaquín Ramírez Villalobos	Universidad Técnica Federico Santa María	Investigador	Experiencia en entrenamiento de modelos de regresión	100	7500.8	Implementación de modelos de regresión y de validación, visualización
Nicolás Rodríguez Ansaldo	Universidad Técnica Federico Santa María	Investigador	Experiencia en creación de modelos matemáticos y	80	9376	Limpieza de datos, creación de pipeline, generar tablas de resumen.
Yunis Catalina Vidal	Universidad Técnica Federico Santa María	Investigadora	Experiencia en entrenamiento de modelos de series temporales.	100	7500.8	Implementación de ARIMA/SARIMA/ETS, validación GridSearch y Optuna.

FIGURE 7. Cargos y Funciones

3. CAPACIDADES CIENTÍFICAS Y TECNOLÓGICAS, DE GESTIÓN Y MECANISMOS DE COLABORACIÓN

3.1. Definición detallada de cargos y funciones.

3.2. Aporte a la formación de capital humano.

La propuesta del proyecto implica una gran carga de investigación para estar a la par con el estado del arte, y en específico en modelos de predicción de ventas a corto plazo. Es por esto que se decidió incorporar a un tesista y a varios jóvenes investigadores que su fin sea exactamente este, estudiar, investigar y generar conocimiento.

La investigación aplicada interuniversitaria del proyecto ayuda al desarrollo de las entidades asociadas pues se generan nuevas tecnologías, en forma del producto entregado, y se fomentan sus áreas de investigación.

Al generar modelos de predicción actualizados, nuevos y de mayor tiempo de predicción, se fomenta la innovación en las entidades asociadas mediante sus intermediarios en el proyecto.